Sub-additive functionals, information theory, and non-convex optimization

Chandra Nair

The Chinese University of Hong Kong

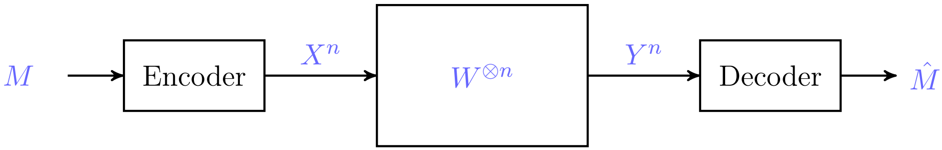
ETH Zurich

January 17, 2020

Sub-additive functionals, information theory, and non-convex optimization

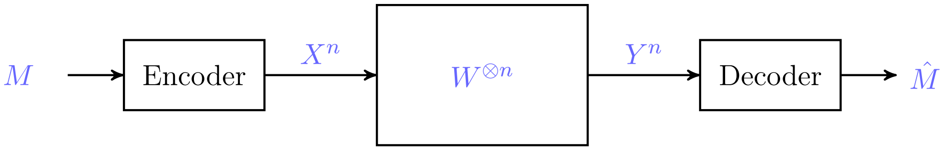
- Introduction
  - ⋆ Building blocks
  - ⋆ How to test the optimality of coding schemes
  - ⋆ Where do the optimization problems arise?
- Sub-additive functionals
  - ⋆ Gallager-style proofs of sub-additivity
  - ⋆ Consequences: inequalities and Gaussian optimality
- Observations and potential future directions

Point-to-point communication

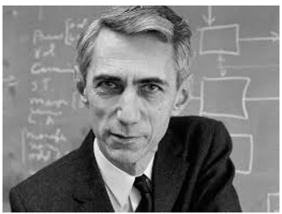


A rate R is achievable if there exists a sequence of encoding/decoding maps so that $\mathbf{P}(M \neq \hat{M}) \to 0$ as $n \to \infty$. Capacity, $C(W) := \sup\{R : R \text{ is achievable } \}$.

Point-to-point communication



A rate R is achievable if there exists a sequence of encoding/decoding maps so that $\mathbf{P}(M \neq \hat{M}) \to 0$ as $n \to \infty$. Capacity, $C(W) := \sup\{R : R$ is achievable $\}$.



Shannon

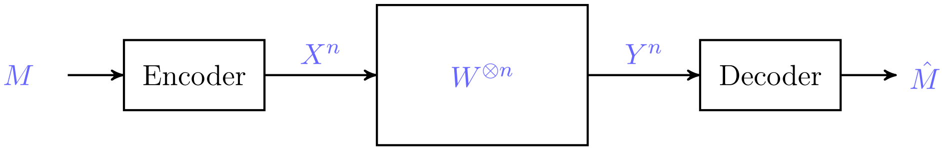
Random coding can be used to achieve

$$R(W) = \sup_{\mu(x)} I(X;Y)$$

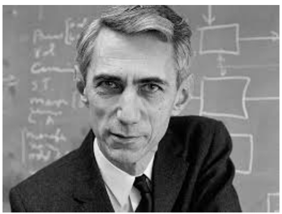
where $\quad I(X;Y) := \sum_{x,y} \mu_{X,Y}(x,y) \log\left(\frac{\mu_{X,Y}(x,y)}{\mu_X(x)\mu_Y(y)}\right)$

$I(X;Y)$: mutual information between X and Y

Point-to-point communication



A rate R is achievable if there exists a sequence of encoding/decoding maps so that $\mathbf{P}(M \neq \hat{M}) \to 0$ as $n \to \infty$. Capacity, $C(W) := \sup\{R : R \text{ is achievable }\}$.



Shannon

Random coding can be used to achieve

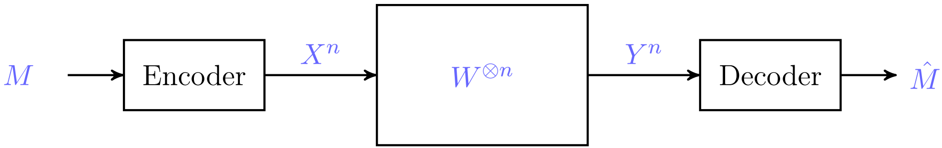
$$R(W) = \sup_{\mu(x)} I(X;Y)$$

where $\quad I(X;Y) := \sum_{x,y} \mu_{X,Y}(x,y) \log\left(\frac{\mu_{X,Y}(x,y)}{\mu_X(x)\mu_Y(y)}\right)$

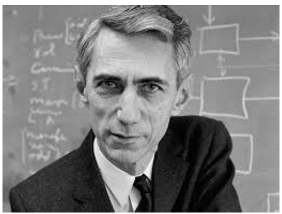
$I(X;Y)$: mutual information between X and Y

Question: Is $R(W) = C(W)$?

Point-to-point communication



A rate R is achievable if there exists a sequence of encoding/decoding maps so that $\mathbf{P}(M \neq \hat{M}) \to 0$ as $n \to \infty$. Capacity, $C(W) := \sup\{R : R$ is achievable $\}$.



Shannon

Random coding can be used to achieve

$$R(W) = \sup_{\mu(x)} I(X;Y)$$

where $\quad I(X;Y) := \sum_{x,y} \mu_{X,Y}(x,y) \log\left(\frac{\mu_{X,Y}(x,y)}{\mu_X(x)\mu_Y(y)}\right)$

$I(X;Y)$: mutual information between X and Y

Question: Is $R(W) = C(W)$? **(YES)** (Shannon '48)

Testing optimality

It is easy (why?) to see that $R(W)$ is optimal `if and only if`

$$R(W \otimes W) = 2R(W) \quad \forall\, W.$$

Testing optimality

It is easy (why?) to see that $R(W)$ is optimal `if and only if`

$$R(W \otimes W) = 2R(W) \quad \forall\, W.$$

The above equality (additivity) follows if the following sub-additivity holds:

$$I(X_1, X_2; Y_1, Y_2) \le I(X_1; Y_1) + I(X_2; Y_2).$$

Testing optimality

It is easy (why?) to see that $R(W)$ is optimal `if and only if`

$$R(W \otimes W) = 2R(W) \quad \forall\ W.$$

The above equality (additivity) follows if the following sub-additivity holds:

$$I(X_1, X_2; Y_1, Y_2) \leq I(X_1; Y_1) + I(X_2; Y_2).$$

Sub-additivity

A functional defined over a probability simplex is said to be sub-additive if

$$F_{12}(\mu_{X_1,X_2}) \leq F_1(\mu_{X_1}) + F_2(\mu_{X_2}) \quad \forall\ \mu_{X_1,X_2}.$$

In above, since W is fixed, $I(X;Y)$ is a functional over μ_X, the space of input distributions.

Testing optimality

It is easy (why?) to see that $R(W)$ is optimal `if and only if`

$$R(W \otimes W) = 2R(W) \quad \forall\, W.$$

The above equality (additivity) follows if the following sub-additivity holds:

$$I(X_1, X_2; Y_1, Y_2) \leq I(X_1; Y_1) + I(X_2; Y_2).$$

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\ &= I(X_1, X_2; Y_1) + I(Y_1, X_1, X_2; Y_2) - I(Y_1; Y_2) \\ &= I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &\leq I(X_1; Y_1) + I(X_2; Y_2). \end{aligned}$$

Testing optimality

It is easy (why?) to see that $R(W)$ is optimal `if and only if`

$$R(W \otimes W) = 2R(W) \quad \forall\, W.$$

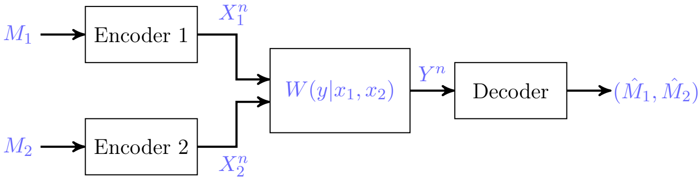
The above equality (additivity) follows if the following sub-additivity holds:

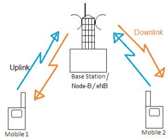
$$I(X_1, X_2; Y_1, Y_2) \leq I(X_1; Y_1) + I(X_2; Y_2).$$

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2 | Y_1) \\ &= I(X_1, X_2; Y_1) + I(Y_1, X_1, X_2; Y_2) - I(Y_1; Y_2) \\ &= I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &\leq I(X_1; Y_1) + I(X_2; Y_2). \end{aligned}$$

Note: Computing $R(W) = \sup_{p(x)} I(X; Y)$ is relatively easy, since $I(X; Y)$ is a concave function of $p(x)$.

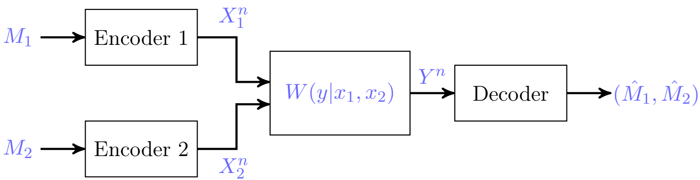
1. Multiple Access Channel (uplink) (Shannon '61)

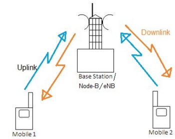




rfwireless-world

1. Multiple Access Channel (uplink) (Shannon '61)





rfwireless-world

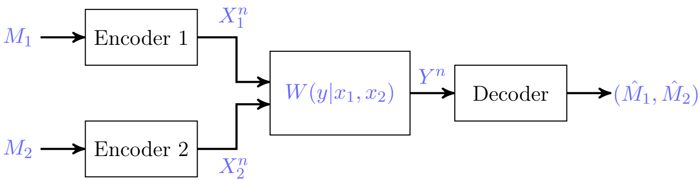
Ahlswede

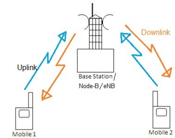
Random coding can be used to achieve rate pairs (R_1, R_2) that satisfy

$$
\begin{aligned}
R_1 &\leq I(X_1; Y | X_2, Q) \\
R_2 &\leq I(X_2; Y | X_1, Q) \\
R_1 + R_2 &\leq I(X_1, X_2; Y | Q)
\end{aligned}
$$

for some $p(q)p(x_1|q)p(x_2|q)$; it suffices to consider $|\mathcal{Q}| \leq 2$. Call this region $\mathcal{R}(W)$.

1. Multiple Access Channel (uplink) (Shannon '61)





rfwireless-world

Ahlswede

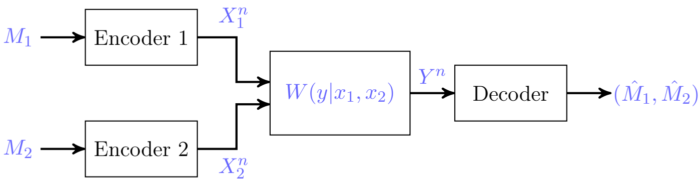
Random coding can be used to achieve rate pairs (R_1, R_2) that satisfy

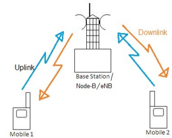
$$
\begin{aligned}
R_1 &\le I(X_1; Y | X_2, Q) \\
R_2 &\le I(X_2; Y | X_1, Q) \\
R_1 + R_2 &\le I(X_1, X_2; Y | Q)
\end{aligned}
$$

for some $p(q)p(x_1|q)p(x_2|q)$; it suffices to consider $|\mathcal{Q}| \le 2$. Call this region $\mathcal{R}(W)$.

Question: Is this the capacity (optimal) region?

1. Multiple Access Channel (uplink) (Shannon '61)





rfwireless-world

Ahlswede

Random coding can be used to achieve rate pairs (R_1, R_2) that satisfy

$$
\begin{aligned}
R_1 &\leq I(X_1; Y|X_2, Q) \\
R_2 &\leq I(X_2; Y|X_1, Q) \\
R_1 + R_2 &\leq I(X_1, X_2; Y|Q)
\end{aligned}
$$

for some $p(q)p(x_1|q)p(x_2|q)$; it suffices to consider $|\mathcal{Q}| \leq 2$. Call this region $\mathcal{R}(W)$.

Question: Is this the capacity (optimal) region? **(YES)** (Ahlswede '72)

Testing optimality

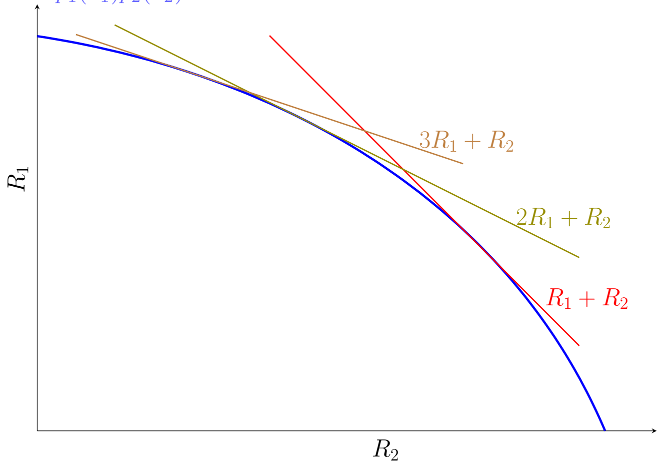
Define, for $\lambda \geq 1$,

$$\begin{aligned} S_\lambda(W) &= \max_{(R_1,R_2)\in\mathcal{R}(W)} \Big\{\lambda R_1 + R_2\Big\} \\ &= \max_{p_1(x_1)p_2(x_2)} \Big\{(\lambda-1)I(X_1;Y|X_2) + I(X_1,X_2;Y)\Big\} \end{aligned}$$

Testing optimality

Define, for $\lambda \geq 1$,

$$S_\lambda(W) = \max_{(R_1,R_2)\in\mathcal{R}(W)} \left\{\lambda R_1 + R_2\right\}$$
$$= \max_{p_1(x_1)p_2(x_2)} \left\{(\lambda-1)I(X_1;Y|X_2) + I(X_1,X_2;Y)\right\}$$



Supporting hyperplanes

Testing optimality

Define, for $\lambda \geq 1$,

$$S_\lambda(W) = \max_{(R_1,R_2)\in\mathcal{R}(W)} \left\{\lambda R_1 + R_2\right\}$$
$$= \max_{p_1(x_1)p_2(x_2)} \left\{(\lambda-1)I(X_1;Y|X_2) + I(X_1,X_2;Y)\right\}$$

As before, $\mathcal{R}(W)$ is optimal `if and only if`

$$S_\lambda(W \otimes W) = 2S_\lambda(W) \quad \forall\ W, \lambda \geq 1.$$

The above equality (additivity) follows if the following sub-additivity holds:

$$(\lambda-1)I(X_{11},X_{12};Y_1,Y_2|X_{21},X_{22}) + I(X_{11},X_{12},X_{21},X_{22};Y_1,Y_2)$$
$$\leq (\lambda-1)I(X_{11};Y_1|X_{21}) + I(X_{11},X_{21};Y_1)$$
$$+ (\lambda-1)I(X_{12};Y_2|X_{22}) + I(X_{12},X_{22};Y_2)$$

Testing optimality

Define, for $\lambda \geq 1$,

$$S_\lambda(W) = \max_{(R_1,R_2)\in\mathcal{R}(W)} \left\{\lambda R_1 + R_2\right\}$$
$$= \max_{p_1(x_1)p_2(x_2)} \left\{(\lambda - 1)I(X_1;Y|X_2) + I(X_1,X_2;Y)\right\}$$

As before, $\mathcal{R}(W)$ is optimal `if and only if`

$$S_\lambda(W \otimes W) = 2S_\lambda(W) \quad \forall\, W, \lambda \geq 1.$$

The above equality (additivity) follows if the following sub-additivity holds:

$$\begin{aligned}
&(\lambda - 1)I(X_{11},X_{12};Y_1,Y_2|X_{21},X_{22}) + I(X_{11},X_{12},X_{21},X_{22};Y_1,Y_2) \\
&\quad \leq (\lambda - 1)I(X_{11};Y_1|X_{21}) + I(X_{11},X_{21};Y_1) \\
&\qquad + (\lambda - 1)I(X_{12};Y_2|X_{22}) + I(X_{12},X_{22};Y_2)
\end{aligned}$$

One can establish this in same way as point-to-point setting.

Testing optimality

Define, for $\lambda \geq 1$,

$$S_\lambda(W) = \max_{(R_1,R_2)\in\mathcal{R}(W)} \left\{\lambda R_1 + R_2\right\}$$
$$= \max_{p_1(x_1)p_2(x_2)} \Big\{(\lambda-1)I(X_1;Y|X_2) + I(X_1,X_2;Y)\Big\}$$

As before, $\mathcal{R}(W)$ is optimal `if and only if`

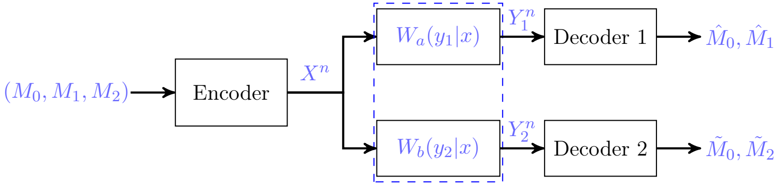
$$S_\lambda(W \otimes W) = 2S_\lambda(W) \quad \forall\, W, \lambda \geq 1.$$

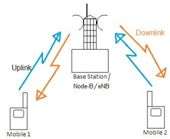
The above equality (additivity) follows if the following sub-additivity holds:

$$(\lambda-1)I(X_{11},X_{12};Y_1,Y_2|X_{21},X_{22}) + I(X_{11},X_{12},X_{21},X_{22};Y_1,Y_2)$$
$$\leq (\lambda-1)I(X_{11};Y_1|X_{21}) + I(X_{11},X_{21};Y_1)$$
$$+ (\lambda-1)I(X_{12};Y_2|X_{22}) + I(X_{12},X_{22};Y_2)$$

Note: Computing $S_\lambda(W)$ is relatively easy since
$\Big\{(\lambda-1)I(X_1;Y|X_2) + I(X_1,X_2;Y)\Big\}$ is concave in $p_1(x_1)$ and $p_2(x_2)$.

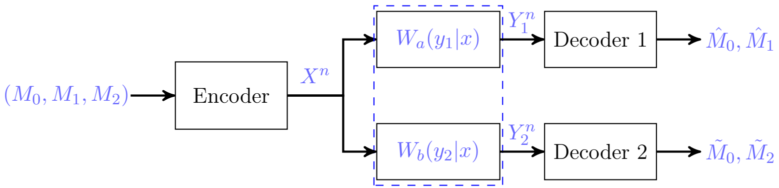
2. Broadcast channel (downlink) (Cover ’72)

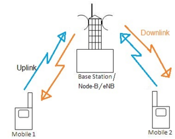




rfwireless-world

2. Broadcast channel (downlink) (Cover '72)





rfwireless-world

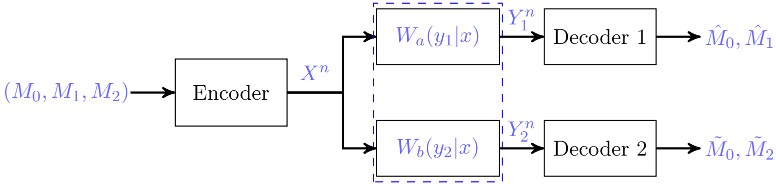
Marton

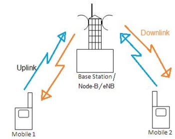
Superposition coding and random hashing can be used to achieve rate triples (R_0, R_1, R_2) that satisfy

$$\begin{aligned}
R_0 &\le \min\{I(Q;Y_1), I(Q;Y_2)\} \\
R_0 + R_1 &\le I(U,Q;Y_1) \\
R_0 + R_2 &\le I(V,Q;Y_2) \\
R_0 + R_1 + R_2 &\le \min\{I(Q;Y_1), I(Q;Y_2)\} + I(U;Y_1|Q) \\
&\quad + I(V;Y_2|Q) - I(U;V|Q)
\end{aligned}$$

for some $p(q,u,v,x)$. Call this region $\mathcal{R}(W_a, W_b)$.

2. Broadcast channel (downlink) (Cover '72)





rfwireless-world

Marton

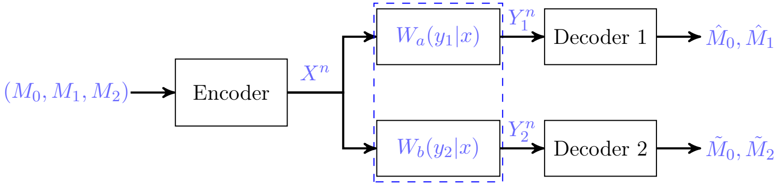
Superposition coding and random hashing can be used to achieve rate triples (R_0, R_1, R_2) that satisfy

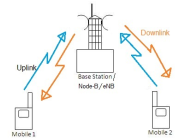
$$
\begin{aligned}
R_0 &\le \min\{I(Q;Y_1), I(Q;Y_2)\} \\
R_0 + R_1 &\le I(U,Q;Y_1) \\
R_0 + R_2 &\le I(V,Q;Y_2) \\
R_0 + R_1 + R_2 &\le \min\{I(Q;Y_1), I(Q;Y_2)\} + I(U;Y_1|Q) \\
&\quad + I(V;Y_2|Q) - I(U;V|Q)
\end{aligned}
$$

for some $p(q,u,v,x)$. Call this region $\mathcal{R}(W_a, W_b)$.

Question: Is this the capacity (optimal) region?

2. Broadcast channel (downlink) (Cover '72)





rfwireless-world

Marton

Superposition coding and random hashing can be used to achieve rate triples (R_0, R_1, R_2) that satisfy

$$
\begin{aligned}
R_0 &\le \min\{I(Q;Y_1), I(Q;Y_2)\} \\
R_0 + R_1 &\le I(U,Q;Y_1) \\
R_0 + R_2 &\le I(V,Q;Y_2) \\
R_0 + R_1 + R_2 &\le \min\{I(Q;Y_1), I(Q;Y_2)\} + I(U;Y_1|Q) \\
&\quad + I(V;Y_2|Q) - I(U;V|Q)
\end{aligned}
$$

for some $p(q,u,v,x)$. Call this region $\mathcal{R}(W_a, W_b)$.

Question: Is this the capacity (optimal) region? **(Open)** (since Marton '79)

Testing optimality ($R_0 = 0$)

Define, for $\lambda \geq 1$,

$$\begin{aligned}
S_\lambda(W) &= \max_{(R_1,R_2)\in\mathcal{R}(W_a,W_b)} \{\lambda R_1 + R_2\} \\
&= \max_{p(u,v,w,x)} \Big\{(\lambda-1)I(U,Q;Y_1) + \min\{I(Q;Y_1), I(Q;Y_2)\} + I(U;Y_1|Q) \\
&\qquad\qquad + I(V;Y_2|Q) - I(U;V|Q)\Big\} \\
&= \min_{\alpha\in[0,1]} \max_{p(u,v,w,x)} \Big\{(\lambda-\alpha)I(Q;Y_1) + \alpha I(Q;Y_2) + \lambda I(U;Y_1|Q) \\
&\qquad\qquad + I(V;Y_2|Q) - I(U;V|Q)\Big\}
\end{aligned}$$

As before, $\mathcal{R}(W_a, W_b)$ is optimal `if and only if`

$$S_\lambda(W_a \otimes W_a, W_b \otimes W_b) = 2S_\lambda(W_a, W_b) \quad \forall\ W_a, W_b, \lambda \geq 1.$$

Testing optimality ($R_0 = 0$)

Define, for $\lambda \geq 1$,

$$\begin{aligned}
S_\lambda(W) &= \max_{(R_1,R_2)\in\mathcal{R}(W_a,W_b)} \{\lambda R_1 + R_2\} \\
&= \max_{p(u,v,w,x)} \Big\{(\lambda-1)I(U,Q;Y_1) + \min\{I(Q;Y_1), I(Q;Y_2)\} + I(U;Y_1|Q) \\
&\qquad\qquad + I(V;Y_2|Q) - I(U;V|Q)\Big\} \\
&= \min_{\alpha\in[0,1]} \max_{p(u,v,w,x)} \Big\{(\lambda-\alpha)I(Q;Y_1) + \alpha I(Q;Y_2) + \lambda I(U;Y_1|Q) \\
&\qquad\qquad + I(V;Y_2|Q) - I(U;V|Q)\Big\}
\end{aligned}$$

As before, $\mathcal{R}(W_a, W_b)$ is optimal `if and only if`

$$S_\lambda(W_a \otimes W_a, W_b \otimes W_b) = 2S_\lambda(W_a, W_b) \quad \forall\ W_a, W_b, \lambda \geq 1.$$

Note: Computing $S_\lambda(W_a, W_b)$ is a non-convex optimization problem.

Successes

In spite of the underlying problem being intrinsically non-convex

- $\mathcal{R}(W_a, W_b)$ is optimal on $R_1 = 0$ (or $R_2 = 0$)
  - ⋆ Degraded message sets: Korner and Marton ('77)
- $\mathcal{R}(W_a, W_b)$ is optimal for some classes of channels
  - ⋆ Gallager '74, Korner and Marton ('75, '77, '79), Gelfand and Pinsker ('78), Poltyrev ('78), El Gamal ('79, '80)
  - ⋆ Weingarten and Steinberg and Shamai '06, Nair '10, Geng and Gohari and Nair and Yu '14, Geng and Nair '14
- Novel ideas and techniques were needed to establish these capacity regions

Successes

In spite of the underlying problem being intrinsically non-convex

- $\mathcal{R}(W_a, W_b)$ is optimal on $R_1 = 0$ (or $R_2 = 0$)
  - ⋆ Degraded message sets: Korner and Marton ('77)
- $\mathcal{R}(W_a, W_b)$ is optimal for some classes of channels
  - ⋆ Gallager '74, Korner and Marton ('75, '77, '79), Gelfand and Pinsker ('78), Poltyrev ('78), El Gamal ('79, '80)
  - ⋆ Weingarten and Steinberg and Shamai '06, Nair '10, Geng and Gohari and Nair and Yu '14, Geng and Nair '14
- Novel ideas and techniques were needed to establish these capacity regions
  - ⋆ **Cover** '72: development of superposition coding strategy

Successes

In spite of the underlying problem being intrinsically non-convex

- $\mathcal{R}(W_a, W_b)$ is optimal on $R_1 = 0$ (or $R_2 = 0$)
  - ⋆ Degraded message sets: Korner and Marton ('77)
- $\mathcal{R}(W_a, W_b)$ is optimal for some classes of channels
  - ⋆ Gallager '74, Korner and Marton ('75, '77, '79), Gelfand and Pinsker ('78), Poltyrev ('78), El Gamal ('79, '80)
  - ⋆ Weingarten and Steinberg and Shamai '06, Nair '10, Geng and Gohari and Nair and Yu '14, Geng and Nair '14
- Novel ideas and techniques were needed to establish these capacity regions
  - ⋆ **Cover** '72 : development of superposition coding strategy
  - ⋆ **Gallager** '74 : converse to the degraded broadcast channel (sub-additivity)

Successes

In spite of the underlying problem being intrinsically non-convex

- $\mathcal{R}(W_a, W_b)$ is optimal on $R_1 = 0$ (or $R_2 = 0$)
  - ⋆ Degraded message sets: Korner and Marton ('77)
- $\mathcal{R}(W_a, W_b)$ is optimal for some classes of channels
  - ⋆ Gallager '74, Korner and Marton ('75, '77, '79), Gelfand and Pinsker ('78), Poltyrev ('78), El Gamal ('79, '80)
  - ⋆ Weingarten and Steinberg and Shamai '06, Nair '10, Geng and Gohari and Nair and Yu '14, Geng and Nair '14

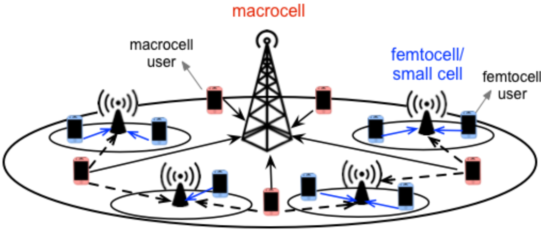
- Novel ideas and techniques were needed to establish these capacity regions
  - ⋆ **Cover** '72 : development of superposition coding strategy
  - ⋆ **Gallager** '74 : converse to the degraded broadcast channel (sub-additivity)
  - ⋆ **Weingarten-Steinberg-Shamai** '06 : Optimality of $\mathcal{R}(W_a, W_b)$ (on $R_0 = 0$) for Gaussian broadcast channel; developing a family of tight convex relaxations to compute the optimal value of a non-convex optimization problem

Successes

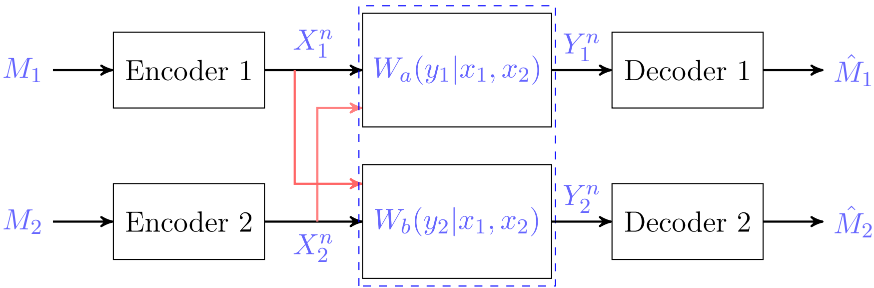
In spite of the underlying problem being intrinsically non-convex

- $\mathcal{R}(W_a, W_b)$ is optimal on $R_1 = 0$ (or $R_2 = 0$)
  - ⋆ Degraded message sets: Korner and Marton ('77)
- $\mathcal{R}(W_a, W_b)$ is optimal for some classes of channels
  - ⋆ Gallager '74, Korner and Marton ('75, '77, '79), Gelfand and Pinsker ('78), Poltyrev ('78), El Gamal ('79, '80)
  - ⋆ Weingarten and Steinberg and Shamai '06, Nair '10, Geng and Gohari and Nair and Yu '14, Geng and Nair '14
- Novel ideas and techniques were needed to establish these capacity regions
  - ⋆ **Cover** '72 : development of superposition coding strategy
  - ⋆ **Gallager** '74 : converse to the degraded broadcast channel (sub-additivity)
  - ⋆ **Weingarten-Steinberg-Shamai** '06 : Optimality of $\mathcal{R}(W_a, W_b)$ (on $R_0 = 0$) for Gaussian broadcast channel; developing a family of tight convex relaxations to compute the optimal value of a non-convex optimization problem
  - ⋆ **Geng-Nair** '14 : Optimality of $\mathcal{R}(W_a, W_b)$ for Gaussian broadcast channel: Technique for establishing extremality of Gaussian distributions using sub-additivity of functionals

3. Interference Channel (Ahlswede '74)



Credit: *www.personal.psu.edu/bxg215/research.html*



3. Interference Channel (Ahlswede '74)

Han

Superposition coding, message splitting, coded time-sharing can be used to achieve rate pairs (R_1, R_2) that satisfy

$$\begin{aligned}
R_1 &< I(X_1; Y_1 | U_2, Q), \\
R_2 &< I(X_2; Y_2 | U_1, Q), \\
R_1 + R_2 &< I(X_1, U_2; Y_1 | Q) + I(X_2; Y_2 | U_1, U_2, Q), \\
R_1 + R_2 &< I(X_2, U_1; Y_2 | Q) + I(X_1; Y_1 | U_1, U_2, Q), \\
R_1 + R_2 &< I(X_1, U_2; Y_1 | U_1, Q) + I(X_2, U_1; Y_2 | U_2, Q), \\
2R_1 + R_2 &< I(X_1, U_2; Y_1 | Q) + I(X_1; Y_1 | U_1, U_2, Q) + I(X_2, U_1; Y_2 | U_2, Q), \\
R_1 + 2R_2 &< I(X_2, U_1; Y_2 | Q) + I(X_2; Y_2 | U_1, U_2, Q) + I(X_1, U_2; Y_1 | U_1, Q)
\end{aligned}$$

for some pmf $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$, where $|U_1| \leq |X_1| + 4$, $|U_2| \leq |X_2| + 4$, and $|Q| \leq 7$. Call this region $\mathcal{R}(W_a, W_b)$.

Kobayashi

3. Interference Channel (Ahlswede '74)

Han

Kobayashi

Superposition coding, message splitting, coded time-sharing can be used to achieve rate pairs (R_1, R_2) that satisfy

$$
\begin{aligned}
R_1 &< I(X_1; Y_1 | U_2, Q), \\
R_2 &< I(X_2; Y_2 | U_1, Q), \\
R_1 + R_2 &< I(X_1, U_2; Y_1 | Q) + I(X_2; Y_2 | U_1, U_2, Q), \\
R_1 + R_2 &< I(X_2, U_1; Y_2 | Q) + I(X_1; Y_1 | U_1, U_2, Q), \\
R_1 + R_2 &< I(X_1, U_2; Y_1 | U_1, Q) + I(X_2, U_1; Y_2 | U_2, Q), \\
2R_1 + R_2 &< I(X_1, U_2; Y_1 | Q) + I(X_1; Y_1 | U_1, U_2, Q) + I(X_2, U_1; Y_2 | U_2, Q), \\
R_1 + 2R_2 &< I(X_2, U_1; Y_2 | Q) + I(X_2; Y_2 | U_1, U_2, Q) + I(X_1, U_2; Y_1 | U_1, Q)
\end{aligned}
$$

for some pmf $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$, where $|U_1| \leq |X_1| + 4$, $|U_2| \leq |X_2| + 4$, and $|Q| \leq 7$. Call this region $\mathcal{R}(W_a, W_b)$.

Question: Is this the capacity region?

3. Interference Channel (Ahlswede '74)

Han

Superposition coding, message splitting, coded time-sharing can be used to achieve rate pairs (R_1, R_2) that satisfy

$$\begin{aligned}
R_1 &< I(X_1; Y_1|U_2, Q),\\
R_2 &< I(X_2; Y_2|U_1, Q),\\
R_1 + R_2 &< I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q),\\
R_1 + R_2 &< I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q),\\
R_1 + R_2 &< I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q),\\
2R_1 + R_2 &< I(X_1, U_2; Y_1|Q) + I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|U_2, Q),\\
R_1 + 2R_2 &< I(X_2, U_1; Y_2|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|U_1, Q)
\end{aligned}$$

Kobayashi

for some pmf $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$, where $|U_1| \le |X_1| + 4$, $|U_2| \le |X_2| + 4$, and $|Q| \le 7$. Call this region $\mathcal{R}(W_a, W_b)$.

Question: Is this the capacity region?

Had been **open** (since Han and Kobayashi '81)

Successes

In spite of the underlying problem being intrinsically non-convex

- $\mathcal{R}(W_a, W_b)$ is optimal for some classes of channels
  - ⋆ Carleial '75, Sato '81, El Gamal and Costa ('81 and '86)
- $\mathcal{R}(W_a, W_b)$ is close to optimal for Gaussian Interference channel
  - ⋆ Etkin and Tse and Wang ('09)
- Novel ideas and mathematical results came out from the investigation of optimality
  - ⋆ Concavity of entropy power (Costa '85)
  - ⋆ Genie based approach to prove sub-additivity (El Gamal and Costa '81, Kramer '02)

Successes

In spite of the underlying problem being intrinsically non-convex

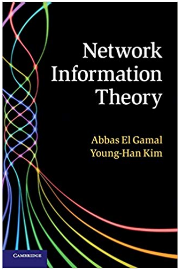
- $\mathcal{R}(W_a, W_b)$ is optimal for some classes of channels
  - ⋆ Carleial '75, Sato '81, El Gamal and Costa ('81 and '86)
- $\mathcal{R}(W_a, W_b)$ is close to optimal for Gaussian Interference channel
  - ⋆ Etkin and Tse and Wang ('09)
- Novel ideas and mathematical results came out from the investigation of optimality
  - ⋆ Concavity of entropy power (Costa '85)
  - ⋆ Genie based approach to prove sub-additivity (El Gamal and Costa '81, Kramer '02)
- $\mathcal{R}(W_a, W_b)$ is not optimal in general (Nair, Xia, Yazdanpanah '15)

Broadcast and interference channels are far too important

- To let non-convexity dissuade us
- To not investigate simple classes that require new ideas

A class of open problems

A sub-collection of the 15 open problems listed in Chaps. 5-9.



5.1 `What is the capacity region of less noisy broadcast-channels with four or more receivers?` (two-receiver: Korner-Marton '76, three-receiver: Nair-Wang '10)

5.2 `What is the capacity region of more capable broadcast-channels with three or more receivers?` (two-receiver: El Gamal '79)

6.1 `What is the capacity region of the Gaussian Interference channel with weak interference?` (strong-interference: Sato '79; mixed-interference corner-points: Sato' 81, Costa'85; weak-interference corner-points: rate-sum (partial): three-groups '09)

6.4 `Is the Han-Kobayashi inner bound tight in general for interference channels?`

8.2 `Is superposition coding optimal for the general 3-receiver broadcast channel with one message to all three receivers and another message to two receivers?`

8.3 `What is the sum-capacity of the binary skew-symmetric broadcast channel?`

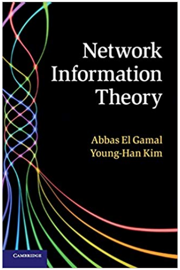
8.4 `Is the Marton inner bound tight in general for broadcast channels?`

9.2 `Can the converse for the Gaussian broadcast channel be proved directly by optimizing the Nair-El Gamal outer bound?`

9.3 `What is the capacity region of the 2-receiver Gaussian broadcast channel with common message?`

A class of open problems

My reformulations of a few of them.



5.1 Is superposition coding optimal for less-noisy broadcast channels with four or more receivers?

5.2 Is superposition coding optimal for more-capable broadcast channels with three or more receivers?

6.1 Is the Han-Kobayashi scheme with Gaussian signaling tight for the Gaussian Interference channel with weak interference?

6.4 Is the Han-Kobayashi inner bound tight in general for interference channels?

8.2 Is superposition coding optimal for the general 3-receiver broadcast channel with one message to all three receivers and another message to two receivers?

8.3 Does the Marton inner bound achieve the sum-capacity of the binary skew-symmetric broadcast channel?

8.4 Is the Marton inner bound tight in general for broadcast channels?

9.2 Can the converse for the Gaussian broadcast channel be proved directly by optimizing the Nair-El Gamal outer bound?

9.3 Does the Marton inner bound achieve the capacity region of the 2-receiver Gaussian broadcast channel with common message?

The common theme to these (reformulated) questions

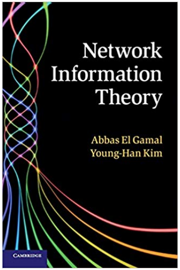
Common theme

Is a candidate rate region optimal?

Idea for testing optimality:

- $S_\lambda(W \otimes W) \stackrel{?}{=} 2S_\lambda(W)$
- Determine sub-additivity of an associated non-convex functional

Status of the open problems



5.1 *Is superposition coding optimal for less-noisy broadcast channels with four or more receivers?* **(OPEN)**

5.2 Is superposition coding optimal for more-capable broadcast channels with three or more receivers? (**NO**: **Nair-Xia** '12)

6.1 *Is the Han-Kobayashi scheme with Gaussian signaling tight for the Gaussian Interference channel with weak interference?* (OPEN) (**yes**: **corner points** using ideas in measure transportation by **Polyanskiy-Wu** '15)

6.4 Is the Han-Kobayashi inner bound tight in general for interference channels? (**NO**: **Nair-Xia-Yazdanpanah** '15)

8.2 Is superposition coding optimal for the general 3-receiver broadcast channel with one message to all three receivers and another message to two receivers? (**NO**: **Nair-Yazdanpanah** '17)

8.3 *Does the Marton inner bound achieve the sum-capacity of the binary skew-symmetric broadcast channel?* **(OPEN)**

8.4 *Is the Marton inner bound tight in general for broadcast channels?* **(OPEN)**

9.2 Can the converse for the Gaussian broadcast channel be proved directly by optimizing the Nair-El Gamal outer bound? (**YES**: **Geng-Nair** '14)

9.3 Does the Marton inner bound achieve the capacity region of the 2-receiver Gaussian broadcast channel with common message? (**YES**: **Geng-Nair** '14)

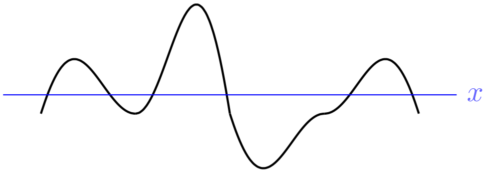
Part II: Sub-additive functionals

- Two examples of sub-additive functionals
  - ⋆ Gallager-style proofs of sub-additivity
  - ⋆ Consequences: inequalities and Gaussian optimality

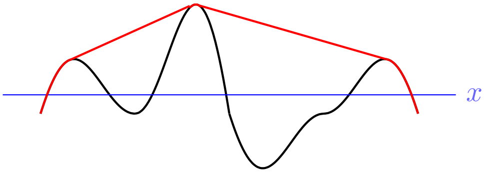
- Family of non-convex optimization problems
  - ⋆ Relation to problems of interest in other fields
  - ⋆ Unifying observations and some conjectures

The "Lieb-Stam" functional (j.w.: Anantharam and Jog '19)

Upper-concave envelope



$f(x)$



upper concave envelope: $\mathcal{C}_x[f]$

Not easy to compute (in general)

The "Lieb-Stam" functional (j.w.: Anantharam and Jog '19)

Let $X \sim \mu_X$ and $Y \sim \nu_Y$

$$F_{LS}(\mu_X, \nu_Y) := \sup_{U: X - U - Y} ah(X|U) + bh(Y|U) - h(X+Y|U)$$
$$= \mathcal{C}[ah(X) + bh(Y) - h(X+Y)]|_{\mu_X, \nu_Y}$$

The "Lieb-Stam" functional (j.w.: Anantharam and Jog '19)

Let $X \sim \mu_X$ and $Y \sim \nu_Y$

$$F_{LS}(\mu_X, \nu_Y) := \sup_{U: X-U-Y} ah(X|U) + bh(Y|U) - h(X+Y|U)$$
$$= \mathcal{C}[ah(X) + bh(Y) - h(X+Y)]|_{\mu_X, \nu_Y}$$

Proposition

The "Lieb-Stam" functional is sub-additive, i.e.

$$F_{LS}(\mu_{X_1X_2}, \nu_{Y_1Y_2}) \le F_{LS}(\mu_{X_1}, \nu_{Y_1}) + F_{LS}(\mu_{X_2}, \nu_{Y_2})$$

The "Lieb-Stam" functional (j.w.: Anantharam and Jog '19)

Let $X \sim \mu_X$ and $Y \sim \nu_Y$

$$F_{LS}(\mu_X, \nu_Y) := \sup_{U: X-U-Y} ah(X|U) + bh(Y|U) - h(X+Y|U)$$
$$= \mathcal{C}[ah(X) + bh(Y) - h(X+Y)]|_{\mu_X, \nu_Y}$$

Proposition

The "Lieb-Stam" functional is sub-additive, i.e.

$$F_{LS}(\mu_{X_1 X_2}, \nu_{Y_1 Y_2}) \leq F_{LS}(\mu_{X_1}, \nu_{Y_1}) + F_{LS}(\mu_{X_2}, \nu_{Y_2})$$

"Gallager-style" proof of sub-additivity

$$\begin{aligned}
&ah(X_1, X_2|U) + bh(Y_1, Y_2|U) - h(X_1 + Y_1, X_2 + Y_2|U) \\
&\quad = ah(X_1|U) + bh(Y_1|U) - h(X_1 + Y_1|U) \\
&\qquad + ah(X_2|U, X_1) + bh(Y_2|U, Y_1) - h(X_2 + Y_2|U, X_1 + Y_1) \\
&\quad \leq ah(X_1|U) + bh(Y_1|U) - h(X_1 + Y_1|U) \\
&\qquad + ah(X_2|U, X_1, Y_1) + bh(Y_2|U, X_1, Y_1) - h(X_2 + Y_2|U, X_1, Y_1) \\
&\quad \leq F_{LS}(\mu_{X_1}, \nu_{Y_1}) + F_{LS}(\mu_{X_2}, \nu_{Y_2}).
\end{aligned}$$

The "Lieb-Stam" functional (j.w.: Anantharam and Jog '19)

Let $X \sim \mu_X$ and $Y \sim \nu_Y$

$$F_{LS}(\mu_X, \nu_Y) := \sup_{U: X-U-Y} ah(X|U) + bh(Y|U) - h(X+Y|U)$$
$$= \mathcal{C}[ah(X) + bh(Y) - h(X+Y)]|_{\mu_X, \nu_Y}$$

Proposition

The "Lieb-Stam" functional is sub-additive, i.e.

$$F_{LS}(\mu_{X_1X_2}, \nu_{Y_1Y_2}) \leq F_{LS}(\mu_{X_1}, \nu_{Y_1}) + F_{LS}(\mu_{X_2}, \nu_{Y_2})$$

One can also define, more genreally, for a channel (Markov operator) $W(z|x,y)$

$$F_{LS}^{W}(\mu_X, \nu_Y) := \sup_{U: X-U-Y} ah(X|U) + bh(Y|U) - h(Z|U)$$
$$= \mathcal{C}[ah(X) + bh(Y) - h(Z)]|_{\mu_X, \nu_Y}$$

One can establish similarly that

$$F_{LS}^{W_1 \otimes W_2}(\mu_{X_1X_2}, \nu_{Y_1Y_2}) \leq F_{LS}^{W_1}(\mu_{X_1}, \nu_{Y_1}) + F_{LS}^{W_2}(\mu_{X_2}, \nu_{Y_2})$$

Korner-Marton functional (j.w.: Geng '14)

Let $X \sim \mu_X$ $W_a(y|x)$ and $W_b(z|x)$ be two channels and $\lambda \geq 1$.

$$F_{KM}^{\lambda, W_a, W_b}(\mu_X) := \sup_{U: U-X-(Y,Z)} I(X;Y|U) - \lambda I(X;Z|U)$$
$$= \mathcal{C}[I(X;Y) - \lambda I(X;Z)]|_{\mu_X}$$

Korner-Marton functional (j.w.: Geng '14)

Let $X \sim \mu_X$ $W_a(y|x)$ and $W_b(z|x)$ be two channels and $\lambda \geq 1$.

$$F_{KM}^{\lambda, W_a, W_b}(\mu_X) := \sup_{U: U - X - (Y,Z)} I(X;Y|U) - \lambda I(X;Z|U)$$
$$= \mathcal{C}[I(X;Y) - \lambda I(X;Z)]|_{\mu_X}$$

Proposition

The "Korner-Marton" functional is sub-additive, i.e.

$$F_{KM}^{\lambda, W_a \otimes \hat{W}_a, W_b \otimes \hat{W}_b}(\mu_{X_1 X_2}) \leq F_{KM}^{\lambda, W_a, W_b}(\mu_{X_1}) + F_{KM}^{\lambda, \hat{W}_a, \hat{W}_b}(\mu_{X_2})$$

Korner-Marton functional (j.w.: Geng '14)

Let $X \sim \mu_X$ $W_a(y|x)$ and $W_b(z|x)$ be two channels and $\lambda \geq 1$.

$$F_{KM}^{\lambda,W_a,W_b}(\mu_X) := \sup_{U:U-X-(Y,Z)} I(X;Y|U) - \lambda I(X;Z|U)$$
$$= \mathcal{C}[I(X;Y) - \lambda I(X;Z)]|_{\mu_X}$$

Proposition

The "Korner-Marton" functional is sub-additive, i.e.

$$F_{KM}^{\lambda,W_a\otimes\hat{W}_a,W_b\otimes\hat{W}_b}(\mu_{X_1X_2}) \leq F_{KM}^{\lambda,W_a,W_b}(\mu_{X_1}) + F_{KM}^{\lambda,\hat{W}_a,\hat{W}_b}(\mu_{X_2})$$

"Gallager-style" proof of sub-additivity

$$\begin{aligned}
&I(X_1,X_2;Y_1,Y_2|U) - \lambda I(X_1,X_2;Z_1,Z_2|U)\\
&\quad = I(X_1;Y_1|U,Z_2) - \lambda I(X_1;Z_1|U,Z_2)\\
&\qquad + I(X_2;Y_2|U,Y_1) - \lambda I(X_2;Z_2|U,Y_1) - (\lambda-1)I(Y_1;Z_2|U)\\
&\quad \leq F_{KM}^{\lambda,W_a,W_b}(\mu_{X_1}) + F_{KM}^{\lambda,\hat{W}_a,\hat{W}_b}(\mu_{X_2}).
\end{aligned}$$

Remarks: Sub-additive functionals

One can extract from

- outer bounds (channel coding)
- lower bounds (source coding)

many sub-additive functionals

Remarks: Sub-additive functionals

One can extract from

- outer bounds (channel coding)
- lower bounds (source coding)

many sub-additive functionals

Remarks

- Proofs of sub-additivity are almost exclusively "Gallager-style"
- The "art" has been in determining which functional is sub-additive
  - ⋆ Lift to a higher-dimensional space (Genie approach)
  - ⋆ Work with "extremal" families of U

Remarks: Sub-additive functionals

One can extract from

- outer bounds (channel coding)
- lower bounds (source coding)

many sub-additive functionals

Remarks

- Proofs of sub-additivity are almost exclusively "Gallager-style"
- The "art" has been in determining which functional is sub-additive
  - ⋆ Lift to a higher-dimensional space (Genie approach)
  - ⋆ Work with "extremal" families of U

Wish List: It would be nice to have a "repository" of sub-additive functionals

- like OEIS but much smaller magnitude obviously

Remarks: Sub-additive functionals

One can extract from

- outer bounds (channel coding)
- lower bounds (source coding)

many sub-additive functionals

Remarks

- Proofs of sub-additivity are almost exclusively "Gallager-style"
- The "art" has been in determining which functional is sub-additive
  - ⋆ Lift to a higher-dimensional space (Genie approach)
  - ⋆ Work with "extremal" families of U

Wish List: It would be nice to have a "repository" of sub-additive functionals

- like OEIS but much smaller magnitude obviously

Next: Gaussian Optimality from sub-additivity

Gaussian optimality via sub-additivity (Geng-Nair '14)

Korner-Marton functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)]$$

over $X : \mathrm{E}(XX^T) \preceq K$, where A, B are invertible matrices and $Z \sim \mathcal{N}(0, I)$.

We will see that the maximum value is

$$h(AX_* + Z) - \lambda h(BX_* + Z),$$

where $X_* \sim \mathcal{N}(0, K')$ for some $K' \preceq K$.

Gaussian optimality via sub-additivity (Geng-Nair '14)

Korner-Marton functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)]$$

over $X : \mathrm{E}(XX^T) \preceq K$, where A, B are invertible matrices and $Z \sim \mathcal{N}(0, I)$.

We will see that the maximum value is

$$h(AX_* + Z) - \lambda h(BX_* + Z),$$

where $X_* \sim \mathcal{N}(0, K')$ for some $K' \preceq K$.

Lemma: $\mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)]$ is sub-additive.

Proof: For any μ_{X_1,X_2}

$$\begin{aligned}
&h(AX_1+Z_1, AX_2+Z_2|U) - \lambda h(BX_1+Z_1, BX_2+Z_2|U) \\
&= h(AX_1+Z_1|U, AX_2+Z_2) - \lambda h(BX_1+Z_1|U, AX_2+Z_2) \\
&\quad + h(AX_2+Z_2|U, BX_1+Z_1) - \lambda h(BX_2+Z_2|U, BX_1+Z_1) \\
&\qquad - (\lambda-1) I(AX_2+Z_2; BX_1+Z_1|U)
\end{aligned}$$

Gaussian optimality via sub-additivity (Geng-Nair '14)

Korner-Marton functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X}[h(AX + Z) - \lambda h(BX + Z)]$$

over $X : \mathrm{E}(XX^T) \preceq K$, where A, B are invertible matrices and $Z \sim \mathcal{N}(0, I)$.

We will see that the maximum value is

$$h(AX_* + Z) - \lambda h(BX_* + Z),$$

where $X_* \sim \mathcal{N}(0, K')$ for some $K' \preceq K$.

Lemma: $\mathcal{C}_{\mu_X}[h(AX + Z) - \lambda h(BX + Z)]$ is sub-additive.

Proof: For any μ_{X_1,X_2}

$$\begin{aligned}
&\mathcal{C}_{\mu_{X_1,X_2}}[h(AX_1 + Z_1, AX_2 + Z_2) - \lambda h(BX_1 + Z_1, BX_2 + Z_2)] \\
&\leq \mathcal{C}_{\mu_{X_1}}[h(AX_1 + Z_1) - \lambda h(BX_1 + Z_1)] \\
&\quad + \mathcal{C}_{\mu_{X_2}}[h(AX_2 + Z_2) - \lambda h(BX_2 + Z_2)]
\end{aligned}$$

Gaussian optimality via sub-additivity (Geng-Nair '14)

Korner-Marton functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)]$$

over $X : \mathrm{E}(XX^T) \preceq K$, where A, B are invertible matrices and $Z \sim \mathcal{N}(0, I)$.

We will see that the maximum value is

$$h(AX_* + Z) - \lambda h(BX_* + Z),$$

where $X_* \sim \mathcal{N}(0, K')$ for some $K' \preceq K$.

Lemma: $\mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)]$ is sub-additive.

Proof: For any μ_{X_1,X_2}

$$\begin{aligned}
&h(AX_1+Z_1, AX_2+Z_2|U) - \lambda h(BX_1+Z_1, BX_2+Z_2|U) \\
&= h(AX_1+Z_1|U, AX_2+Z_2) - \lambda h(BX_1+Z_1|U, AX_2+Z_2) \\
&\quad + h(AX_2+Z_2|U, BX_1+Z_1) - \lambda h(BX_2+Z_2|U, BX_1+Z_1) \\
&\quad\quad -(\lambda-1) I(AX_2+Z_2; BX_1+Z_1|U)
\end{aligned}$$

Gaussian optimality: ctd..

Let $(U_\dagger, X_\dagger)$ be a maximizer, i.e.

$$V = \max_{\mu_X} \mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)] = h(AX_\dagger + Z|U_\dagger) - \lambda h(BX_\dagger + Z|U_\dagger).$$

Let (X_a, U_a) and (X_b, U_b) be i.i.d. according to $(U_\dagger, X_\dagger)$.

Gaussian optimality: ctd..

Let $(U_\dagger, X_\dagger)$ be a maximizer, i.e.

$$V = \max_{\mu_X} \mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)] = h(AX_\dagger + Z|U_\dagger) - \lambda h(BX_\dagger + Z|U_\dagger).$$

Let (X_a, U_a) and (X_b, U_b) be i.i.d. according to $(U_\dagger, X_\dagger)$.

Setting $U = (U_a, U_b)$, $X_+ = \frac{X_a + X_b}{\sqrt{2}}$ and $X_- = \frac{X_a - X_b}{\sqrt{2}}$ the proof of sub-additivity yields

$$\begin{aligned}
2V &= \mathcal{C}_{\mu_{X_1,X_2}}[h(AX_1+Z_1, AX_2+Z_2) - \lambda h(BX_1+Z_1, BX_2+Z_2)]\Big|_{(\mu_{X_+,X_-})} \\
&\le \mathcal{C}_{\mu_{X_1}}[h(AX_1+Z_1) - \lambda h(BX_1+Z_1)]\Big|_{\mu_{X_+}} \\
&\quad + \mathcal{C}_{\mu_{X_2}}[h(AX_2+Z_2) - \lambda h(BX_2+Z_2)]\Big|_{\mu_{X_-}} \\
&\quad\quad -(\lambda-1) I(AX_- + Z_2; BX_+ + Z_1 | U_a, U_b) \\
&\le V + V
\end{aligned}$$

Gaussian optimality: ctd..

Let $(U_\dagger, X_\dagger)$ be a maximizer, i.e.

$$V = \max_{\mu_X} \mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)] = h(AX_\dagger + Z|U_\dagger) - \lambda h(BX_\dagger + Z|U_\dagger).$$

Let (X_a, U_a) and (X_b, U_b) be i.i.d. according to $(U_\dagger, X_\dagger)$.

Setting $U = (U_a, U_b)$, $X_+ = \frac{X_a + X_b}{\sqrt{2}}$ and $X_- = \frac{X_a - X_b}{\sqrt{2}}$ the proof of sub-additivity yields

$$\begin{aligned}
2V &= \mathcal{C}_{\mu_{X_1,X_2}}[h(AX_1+Z_1, AX_2+Z_2) - \lambda h(BX_1+Z_1, BX_2+Z_2)]\Big|_{(\mu_{X_+,X_-})} \\
&\le \mathcal{C}_{\mu_{X_1}}[h(AX_1+Z_1) - \lambda h(BX_1+Z_1)]\Big|_{\mu_{X_+}} \\
&\quad + \mathcal{C}_{\mu_{X_2}}[h(AX_2+Z_2) - \lambda h(BX_2+Z_2)]\Big|_{\mu_{X_-}} \\
&\qquad -(\lambda-1)I(AX_- + Z_2; BX_+ + Z_1|U_a, U_b) \\
&\le V + V
\end{aligned}$$

Therefore: we get that conditioned on (U_a, U_b): $X_+ \perp X_-$.

Gaussian optimality: ctd..

Let $(U_\dagger, X_\dagger)$ be a maximizer, i.e.

$$V = \max_{\mu_X} \mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)] = h(AX_\dagger + Z|U_\dagger) - \lambda h(BX_\dagger + Z|U_\dagger).$$

Let (X_a, U_a) and (X_b, U_b) be i.i.d. according to $(U_\dagger, X_\dagger)$.

Note: Thus, conditioned on (U_a, U_b):

- $X_a \perp X_b$ (from construction)
- $(X_a + X_b) \perp (X_a - X_b)$ (from proof of sub-additivity)

Gaussian optimality: ctd..

Let $(U_\dagger, X_\dagger)$ be a maximizer, i.e.

$$V = \max_{\mu_X} \mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)] = h(AX_\dagger + Z|U_\dagger) - \lambda h(BX_\dagger + Z|U_\dagger).$$

Let (X_a, U_a) and (X_b, U_b) be i.i.d. according to $(U_\dagger, X_\dagger)$.

Note: Thus, conditioned on (U_a, U_b):

- $X_a \perp X_b$ (from construction)
- $(X_a + X_b) \perp (X_a - X_b)$ (from proof of sub-additivity)
- Implies that conditioned on (U_a, U_b): X_a, X_b are Gaussian
  - ⋆ Characterization of Gaussians (Bernstein '40s)
  - ⋆ Proof: Using characteristic functions (Fourier transforms)

Gaussian optimality: ctd..

Let $(U_\dagger, X_\dagger)$ be a maximizer, i.e.

$$V = \max_{\mu_X} \mathcal{C}_{\mu_X}[h(AX+Z) - \lambda h(BX+Z)] = h(AX_\dagger + Z|U_\dagger) - \lambda h(BX_\dagger + Z|U_\dagger).$$

Let (X_a, U_a) and (X_b, U_b) be i.i.d. according to $(U_\dagger, X_\dagger)$.

Note: Thus, conditioned on (U_a, U_b):

- $X_a \perp X_b$ (from construction)
- $(X_a + X_b) \perp (X_a - X_b)$ (from proof of sub-additivity)
- Implies that conditioned on (U_a, U_b): X_a, X_b are Gaussian
  - ⋆ Characterization of Gaussians (Bernstein '40s)
  - ⋆ Proof: Using characteristic functions (Fourier transforms)

Note: There are some similarities with work of Lieb and Barthe (90s)

They also use rotations (but not information measures and its algebra)

Lieb-Stam functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X,\nu_Y}[ah(X) + bh(Y) - h(X+Y)]$$

over $X : \mathrm{E}(XX^T) \preceq K_a$, and $Y : \mathrm{E}(YY^T) \preceq K_b$.

In a very similar fashion, the maximum value is

$$ah(X_*) + bh(Y_*) - h(X_* + Y_*),$$

where $X_* \sim \mathcal{N}(0, K')$ and $X_* \sim \mathcal{N}(0, K^\dagger)$.

Lieb-Stam functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X,\nu_Y}[ah(X) + bh(Y) - h(X+Y)]$$

over $X : \mathrm{E}(XX^T) \preceq K_a$, and $Y : \mathrm{E}(YY^T) \preceq K_b$.

In a very similar fashion, the maximum value is

$$ah(X_*) + bh(Y_*) - h(X_* + Y_*),$$

where $X_* \sim \mathcal{N}(0, K')$ and $X_* \sim \mathcal{N}(0, K^\dagger)$.

Corollary: Entropy-Power inequality

Lieb-Stam functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X,\nu_Y}[ah(X) + bh(Y) - h(X+Y)]$$

over $X : \mathrm{E}(XX^T) \preceq K_a$, and $Y : \mathrm{E}(YY^T) \preceq K_b$.

In a very similar fashion, the maximum value is

$$ah(X_*) + bh(Y_*) - h(X_* + Y_*),$$

where $X_* \sim \mathcal{N}(0, K')$ and $X_* \sim \mathcal{N}(0, K^\dagger)$.

Corollary: Entropy-Power inequality

Remark: We consider a more general functional that yields EPI as one extreme and Brascamp-Lieb inequality as the other
`http://arxiv.org/abs/1901.06619`

Lieb-Stam functional j.w.: (Ananthram, Jog '19)

Lieb-Stam functional - extremal distribution

Maximize, for $\lambda > 1$, the value of the functional

$$\mathcal{C}_{\mu_X,\nu_Y}[ah(X) + bh(Y) - h(X+Y)]$$

over $X : \mathrm{E}(XX^T) \preceq K_a$, and $Y : \mathrm{E}(YY^T) \preceq K_b$.

In a very similar fashion, the maximum value is

$$ah(X_*) + bh(Y_*) - h(X_* + Y_*),$$

where $X_* \sim \mathcal{N}(0, K')$ and $X_* \sim \mathcal{N}(0, K^\dagger)$.

Corollary: Entropy-Power inequality

Observe: Sub-additivity results holds even in discrete spaces and with any channel $W(z|x,y)$.
Question: Can this be used to get new extremal inequalties (e.g. discrete EPIs)?

Another functional - (Courtade '17)

Let $X \sim \mu_X$ and Z be standard Gaussian independet of X

$$g(\mu_X) := \sup_{\substack{U,Y:(U,X)\perp Z,\\ I(X;Y|U,X+Z)=0}} \lambda h(X|U) + (1-\lambda)h(Z) - h(X+Z|U) + I(X;Y|U) - \lambda I(X+Z;Y|U)$$

Theorem

The functional $g_\epsilon(X)$ *is sub-additive.*

Proof: "Gallager-type"

Another functional - (Courtade '17)

Let $X \sim \mu_X$ and Z be standard Gaussian independet of X

$$g(\mu_X) := \sup_{\substack{U,Y:(U,X)\perp Z,\\ I(X;Y|U,X+Z)=0}} \lambda h(X|U) + (1-\lambda)h(Z) - h(X+Z|U) + I(X;Y|U) - \lambda I(X+Z;Y|U)$$

Theorem

The functional $g_\epsilon(X)$ is sub-additive.

Proof: "Gallager-type"

Corollary: Courtade's "strengthened-EPI"

$$2^{2(h(X+Z)-I(X;Y))} \geq 2^{2(h(X)-I(X+Z;Y))} + 2^{2h(Z)}$$

Open problem: Marton's region for the broadcast channel?

Is the following functional sub-additive or is there an example where it is super-additive?

Let $W_a(y|x)$ and $W_b(z|x)$ be given channels, $\alpha \in [0,1]$, and $\lambda \geq 1$.

$$\mathcal{C}_{\mu_X}\left[(\lambda-\alpha)H(Y)-\alpha H(Z)+\max_{p(u,v|x)}\{\lambda I(U;Y)+I(V;Z)-I(U;V)\}\right]$$

Open problem: Marton's region for the broadcast channel?

Is the following functional sub-additive or is there an example where it is super-additive?

Let $W_a(y|x)$ and $W_b(z|x)$ be given channels, $\alpha \in [0,1]$, and $\lambda \geq 1$.

$$\mathcal{C}_{\mu_X}\left[(\lambda-\alpha)H(Y)-\alpha H(Z)+\max_{p(u,v|x)}\{\lambda I(U;Y)+I(V;Z)-I(U;V)\}\right]$$

- If sub-additive, then Marton's region is optimal for broadcast channel
- If $\exists$ example where it is super-additive, then one should be able to deduce a channel where Marton's region is not optimal

Open problem: Marton's region for the broadcast channel?

Is the following functional sub-additive or is there an example where it is super-additive?

Let $W_a(y|x)$ and $W_b(z|x)$ be given channels, $\alpha \in [0,1]$, and $\lambda \geq 1$.

$$\mathcal{C}_{\mu_X}\left[(\lambda-\alpha)H(Y) - \alpha H(Z) + \max_{p(u,v|x)} \{\lambda I(U;Y) + I(V;Z) - I(U;V)\}\right]$$

- If sub-additive, then Marton's region is optimal for broadcast channel
- If $\exists$ example where it is super-additive, then one should be able to deduce a channel where Marton's region is not optimal

Remarks:

- Conjectured to be sub-additive (Anantharam-Gohari-Nair '13)
- To evaluate the concave envelope
  - ⋆ Suffices to consider (U,V): $|U| + |V| \leq |X| + 1$.
  - ⋆ We did not get any contradiction to sub-addivity for binary input broadcast channels
- Can prove sub-additivity when $\alpha = 0$ or $\alpha = 1$.

A specific family of non-convex optimization problems

Shows up: Testing the optimality of coding schemes

Testing optimality (usually) reduces to testing sub-additivity of

$$\mathcal{C}_{\nu_{\mathbf{X}}}\Big[\sum_{S\subseteq[n]} \alpha_S H(X_S)\Big], \ \alpha_S \in \mathbb{R}.$$

A specific family of non-convex optimization problems

Shows up: Testing the optimality of coding schemes

Testing optimality (usually) reduces to testing sub-additivity of

$$\mathcal{C}_{\nu_{\mathbf{X}}}\Big[\sum_{S\subseteq[n]} \alpha_S H(X_S)\Big],\ \alpha_S \in \mathbb{R}.$$

Using Fenchel duality this is same as (Anantharam, Gohari, Nair '13)

$$G_1(\gamma_1) := \max_{\mu_{\mathbf{X}}} \sum_{S\subseteq[n]} \alpha_S H(X_S) - \mathrm{E}(\gamma_1(\mathbf{X}))$$

$$G_2(\gamma_2) := \max_{\mu_{\mathbf{X}}} \sum_{S\subseteq[n]} \alpha_S H(X_S) - \mathrm{E}(\gamma_2(\mathbf{X}))$$

$$G_{12}(\gamma_1,\gamma_2) := \max_{\mu_{\mathbf{X}_1,\mathbf{X}_2}} \sum_{S\subseteq[n]} \alpha_S H(X_{1S}, X_{2S}) - \mathrm{E}(\gamma_1(\mathbf{X}_1)) - \mathrm{E}(\gamma_2(\mathbf{X}_2))$$

Is $G_{12}(\gamma_1,\gamma_2) = G_1(\gamma_1) + G_2(\gamma_2)\ \forall\ \gamma_1,\gamma_2$?
i.e. Is the maximizer of G_{12} a product distribution?

A specific family of non-convex optimization problems

Shows up: Testing the optimality of coding schemes

Testing optimality (usually) reduces to testing sub-additivity of

$$\mathcal{C}_{\nu_{\mathbf{X}}}\Big[\sum_{S\subseteq[n]} \alpha_S H(X_S)\Big],\ \alpha_S \in \mathbb{R}.$$

Using Fenchel duality this is same as (Anantharam, Gohari, Nair '13)

$$G_1(\gamma_1) := \max_{\mu_{\mathbf{X}}} \sum_{S\subseteq[n]} \alpha_S H(X_S) - \mathrm{E}(\gamma_1(\mathbf{X}))$$

$$G_2(\gamma_2) := \max_{\mu_{\mathbf{X}}} \sum_{S\subseteq[n]} \alpha_S H(X_S) - \mathrm{E}(\gamma_2(\mathbf{X}))$$

$$G_{12}(\gamma_1,\gamma_2) := \max_{\mu_{\mathbf{X}_1,\mathbf{X}_2}} \sum_{S\subseteq[n]} \alpha_S H(X_{1S}, X_{2S}) - \mathrm{E}(\gamma_1(\mathbf{X}_1)) - \mathrm{E}(\gamma_2(\mathbf{X}_2))$$

Are there other fields where the same family shows up?

Hypercontractivity

Studied in functional analysis, cs theory, etc.

Definition

$(X, Y) \sim \mu_{XY}$ is (p, q)-hypercontractive for $1 \leq q \leq p$ if

$$\|Tg\|_p \leq \|g\|_q \quad \forall g(Y)$$

where T is the Markov operator characterized by $\mu_{Y|X}$

Here $\|Z\|_p = E(|Z|^p)^{\frac{1}{p}}$.

Hypercontractivity

Studied in functional analysis, cs theory, etc.

Definition

$(X, Y) \sim \mu_{XY}$ is (p, q)-hypercontractive for $1 \leq q \leq p$ if

$$\|Tg\|_p \leq \|g\|_q \quad \forall g(Y)$$

where T is the Markov operator characterized by $\mu_{Y|X}$

Here $\|Z\|_p = E(|Z|^p)^{\frac{1}{p}}$.

There is a lot of interest in evaluting hypercontractivity parameters for distributions.

Theorem (Nair '14)

$(X, Y) \sim \mu_{XY}$ is (p, q)-hypercontractive for $1 \leq q \leq p$ if and only if

$$\mathcal{C}_{\nu_{X,Y}}\left[H(X,Y) - (1 - \frac{1}{p})H(X) - \frac{1}{q}H(Y)\right]\Big|_{\mu_{X,Y}}$$
$$= H(X,Y) - (1 - \frac{1}{p})H(X) - \frac{1}{q}H(Y)$$

Hypercontractivity

Studied in functional analysis, cs theory, etc.

Definition

$(X, Y) \sim \mu_{XY}$ is (p, q)-hypercontractive for $1 \leq q \leq p$ if

$$\|Tg\|_p \leq \|g\|_q \quad \forall g(Y)$$

where T is the Markov operator characterized by $\mu_{Y|X}$

Here $\|Z\|_p = E(|Z|^p)^{\frac{1}{p}}$.

Hypercontractivity parameters satisfies a property called tensorization:

If $(X_1, Y_1) \perp (X_2, Y_2)$ are both (p, q)-hypercontractive, then $((X_1, X_2), (Y_1, Y_2))$ is also (p, q)-hypercontractive

Gets around the curse of dimensionality.

Hypercontractivity

Studied in functional analysis, cs theory, etc.

Definition

$(X, Y) \sim \mu_{XY}$ is (p, q)-hypercontractive for $1 \le q \le p$ if

$$\|Tg\|_p \le \|g\|_q \quad \forall g(Y)$$

where T is the Markov operator characterized by $\mu_{Y|X}$

Here $\|Z\|_p = E(|Z|^p)^{\frac{1}{p}}$.

Rather immediate that sub-additivity, i.e.

$$\begin{aligned}
&\mathcal{C}_{\mu_{X_1Y_1X_2Y_2}}[H(X_1Y_1X_2Y_2) - \lambda_1 H(X_1X_2) - \lambda_2 H(Y_1Y_2)] \\
&\quad \le \mathcal{C}_{\mu_{X_1Y_1}}[H(X_1Y_1) - \lambda_1 H(X_1) - \lambda_2 H(Y_1)] + \mathcal{C}_{\mu_{X_2Y_2}}[H(X_2Y_2) - \lambda_1 H(X_2) - \lambda_2 H(Y_2)]
\end{aligned}$$

is equivalent to tensorization of hypercontractivity parameters

Hypercontractivity

Studied in functional analysis, cs theory, etc.

Definition

$(X, Y) \sim \mu_{XY}$ is (p, q)-hypercontractive for $1 \leq q \leq p$ if

$$\|Tg\|_p \leq \|g\|_q \quad \forall g(Y)$$

where T is the Markov operator characterized by $\mu_{Y|X}$

Here $\|Z\|_p = E(|Z|^p)^{\frac{1}{p}}$.

This (serendipitous) rediscovery of the link between hypercontractivity and information measures and these equivalent characterizations is spurring a lot of work

Consequences

Computation of hypercontractivity parameters is considered hard

- X is uniform and $\mu_{Y|X}$ is binary symmetric channel
  - ⋆ (Bonami-Beckner inequality '70s, Borrell '82)
- (X, Y) Jointly Gaussian (Gross '75)

Evaluation of achievable regions is of similar difficulty as determining hypercontractivity (same family and similar terms)

For testing optimality of schemes we had to develop tools for evaluating achievable regions for certain channels

Consequences

Computation of hypercontractivity parameters is considered hard

- X is uniform and $\mu_{Y|X}$ is binary symmetric channel
  - (Bonami-Beckner inequality '70s, Borrell '82)
- (X, Y) Jointly Gaussian (Gross '75)

Evaluation of achievable regions is of similar difficulty as determining hypercontractivity (same family and similar terms)

For testing optimality of schemes we had to develop tools for evaluating achievable regions for certain channels

Can we use our techniques to evaluate new hypercontractivity parameters?

Yes, we can.

E.g.: X is uniform and $\mu_{Y|X}$ is binary erasure channel (Nair-Wang '16,'17)

Consequences

Computation of hypercontractivity parameters is considered hard

- X is uniform and $\mu_{Y|X}$ is binary symmetric channel
  - (Bonami-Beckner inequality '70s, Borrell '82)
- (X, Y) Jointly Gaussian (Gross '75)

Evaluation of achievable regions is of similar difficulty as determining hypercontractivity (same family and similar terms)

For testing optimality of schemes we had to develop tools for evaluating achievable regions for certain channels

Can we use our techniques to evaluate new hypercontractivity parameters?

Yes, we can.

E.g.: X is uniform and $\mu_{Y|X}$ is binary erasure channel (Nair-Wang '16,'17)

Other techniques we used to solve these non-convex problems:

- Identify a lower dimensional manifold that contains all the stationary points
- Analyze the function directly on this manifold or
- Use properties of the points on this manifold to deduce sub-additivity

An Observation

Reminder: Family of functionals that showed up in network information theory

$$\sum_{S\subseteq[n]} \alpha_S H(X_S), \;\; \alpha_S \in \mathbb{R}.$$

Usually, one is interested in testing the sub-additivity of

$$\mathcal{C}_{\mu_X}[\alpha_S H(X_S)].$$

This is equivalent to testing a global tensorization property.

Definition

A functional $\sum_{S\subseteq[n]} \alpha_S H(X_S)$ is said to satisfy **global tensorization** if a `product distribution maximizes` $G^{\mu}_{12}(\gamma_1,\gamma_2)$ for all γ_1,γ_2, where

$$G^{\mu}_{12}(\gamma_1,\gamma_2) := \sum_{S\subseteq[n]} \alpha_S H(X_{1S},X_{2S}) - \mathrm{E}(\gamma_1(\mathbf{X}_1)) - \mathrm{E}(\gamma_2(\mathbf{X}_2))$$

An Observation

Definition

A functional $\sum_{S \subseteq [n]} \alpha_S H(X_S)$ is said to satisfy **local tensorization** if `the product of local maximizers` of $G^{\mu_1}(\gamma_1)$ and $G^{\mu_2}(\gamma_2)$ `is a local maximizer` of $G^{\mu}_{12}(\gamma_1, \gamma_2)$ for all γ_1, γ_2, where

$$G^{\mu}_{1}(\gamma_1) := \sum_{S \subseteq [n]} \alpha_S H(X_{1S}) - \mathrm{E}(\gamma_1(\mathbf{X}_1))$$

$$G^{\mu}_{2}(\gamma_2) := \sum_{S \subseteq [n]} \alpha_S H(X_{2S}) - \mathrm{E}(\gamma_2(\mathbf{X}_2))$$

$$G^{\mu}_{12}(\gamma_1, \gamma_2) := \sum_{S \subseteq [n]} \alpha_S H(X_{1S}, X_{2S}) - \mathrm{E}(\gamma_1(\mathbf{X}_1)) - \mathrm{E}(\gamma_2(\mathbf{X}_2))$$

An Observation

Definition

A functional $\sum_{S\subseteq[n]} \alpha_S H(X_S)$ is said to satisfy **local tensorization** if `the product of local maximizers` of $G^{\mu_1}(\gamma_1)$ and $G^{\mu_2}(\gamma_2)$ `is a local maximizer` of $G^{\mu}_{12}(\gamma_1,\gamma_2)$ for all γ_1,γ_2, where

$$G^{\mu}_1(\gamma_1) := \sum_{S\subseteq[n]} \alpha_S H(X_{1S}) - \mathrm{E}(\gamma_1(\mathbf{X}_1))$$

$$G^{\mu}_2(\gamma_2) := \sum_{S\subseteq[n]} \alpha_S H(X_{2S}) - \mathrm{E}(\gamma_2(\mathbf{X}_2))$$

$$G^{\mu}_{12}(\gamma_1,\gamma_2) := \sum_{S\subseteq[n]} \alpha_S H(X_{1S},X_{2S}) - \mathrm{E}(\gamma_1(\mathbf{X}_1)) - \mathrm{E}(\gamma_2(\mathbf{X}_2))$$

Observation (Conjecture)

For functionals in this family global tensorization holds if and only if local tensorization holds

Note: Similarity to testing concavity using a local (second derivative) condition

Notes

For some of the remaining open problems (mentioned earlier), we can establish local-tensorization

- Marton's inner bound for binary input broadcast channels
- Gaussian Z-interference channel

Therefore, if the Conjecture is true, then we would establish the capacity region for these settings

Notes

For some of the remaining open problems (mentioned earlier), we can establish local-tensorization

- Marton's inner bound for binary input broadcast channels
- Gaussian Z-interference channel

Therefore, if the Conjecture is true, then we would establish the capacity region for these settings

Question: How may these two phenomena be connected?

A possible answer is suggested by our computations in various examples

- Construct a path in the space of functionals and track the global maximizers

Theoretically: Is there some kind of geodesic convexity in some cases?

- Far-reaching and non-trivial extension of Mrs. Gerber's lemma
- Consistent with other conjectures

Optimization based approaches

Optimization based approaches have been game changers

First jump: Linear programming to convex optimization

Semi-definite program based algorithm design and analysis

- Compressive sensing
- Phase recovery
- Clustering
- Image processing

Optimization based approaches

Optimization based approaches have been game changers

First jump: Linear programming to convex optimization

Semi-definite program based algorithm design and analysis

- Compressive sensing
- Phase recovery
- Clustering
- Image processing

New Jump: Convex optimization to specific families of non-convex optimization

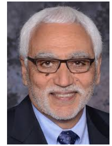
Studies on these families are already making impact in

- Machine learning and AI (Singular Value Decomposition)
- Graphical models and Statistical Physics based approaches (sum of energy and entropy terms)
- Communication networks (linear combination of entropies of subsets)

Acknowledgements (Rogues gallery)

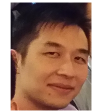
Amin Gohari

Max Costa



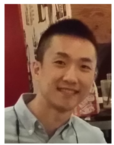
Abbas El Gamal

Yanlin Geng

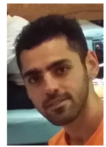


Vincent Wang

Salman Beigi



Lingxiao Xia

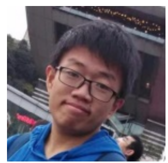


Babak Yazdanpanah

Janos Korner

Sida Liu

David Ng



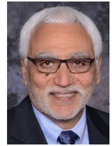
Yan Nan Wang

Venkat Anantharam

Acknowledgements (Rogues gallery)

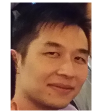
Amin Gohari

Max Costa



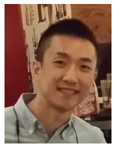
Abbas El Gamal

Yanlin Geng

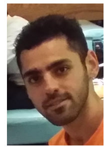


Vincent Wang

Salman Beigi



Lingxiao Xia

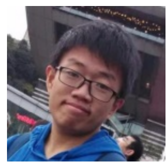


Babak Yazdanpanah

Janos Korner

Sida Liu

David Ng



Yan Nan Wang

Venkat Anantharam

THANK YOU