Non-convex Optimization and Network Information Theory

Chandra Nair

The Chinese University of Hong Kong

3rd January, 2019
Non-convex problems and network information theory

- Introduction
  - Building blocks
  - How to test the optimality of coding schemes
  - Where do the optimization problems arise?

- Two problems to illustrate some ideas

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

**Question**: Is $R(W) = C(W)$? **(YES)** (Shannon '48)
A rate $R$ is achievable if there exists a sequence of encoding/decoding maps so that $P(M \neq \hat{M}) \to 0$ as $n \to \infty$. Capacity, $C(W) := \sup \{ R : R \text{ is achievable} \}$.

Random coding can be used to achieve

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

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

$I(X;Y)$: mutual information between $X$ and $Y$
A rate $R$ is achievable if there exists a sequence of encoding/decoding maps so that $P(M \neq \hat{M}) \to 0$ as $n \to \infty$. Capacity, $C(W) := \sup \{ R : R$ is achievable $\}$. Random coding can be used to achieve

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

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

$I(X; Y)$: 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 $P(M \neq \hat{M}) \to 0$ as $n \to \infty$. Capacity, $C(W) := \sup \{ R : R \text{ is achievable} \}$.

Random coding can be used to achieve

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

where

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

$I(X;Y)$: 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

\[
R(W \otimes W) = 2R(W) \quad \forall W.
\]

If \( \exists W \) such that \( \frac{1}{2}R(W \otimes W) > R(W) \) then

\[
C(W) \geq \frac{1}{2}R(W \otimes W) > R(W)
\]

(Hence equality is necessary)
Testing optimality

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

\[
R(W \otimes W) = 2R(W) \quad \forall \ W.
\]

Given \( \epsilon > 0 \) there is a sequence of codes such that

\[
\frac{1}{n} I(X^n; Y^n) \geq C(W) - \epsilon, \quad \forall n > n_0
\]

- Fano’s inequality
- Data processing inequality
Testing optimality

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

\[
R(W \otimes W) = 2R(W) \quad \forall W.
\]

Given \( \epsilon > 0 \) there is a sequence of codes such that

\[
\frac{1}{n} I(X^n; Y^n) \geq C(W) - \epsilon, \quad \forall n > n_0
\]

Hence, for \( k \) such that \( N = 2^k > n_0 \) we have

\[
R(W) \stackrel{\text{indc}}{=} \frac{1}{N} R(W \otimes \cdots \otimes W) = \frac{1}{N} I(X^N; Y^N) \geq C(W) - \epsilon.
\]
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:

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

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

$$I(X_1, X_2; Y_1, Y_2) = I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1)$$

$$= I(X_1, X_2; Y_1) + I(Y_1, X_1, X_2; Y_2) - I(Y_1; Y_2)$$

$$= I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2)$$

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

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

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

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

The various ideas introduced by Shannon have led to an information revolution

Random coding and its optimality have directly inspired

- Low density parity check codes (LDPC)
- Polar codes
  - proof of sub-additivity
Successes

The various ideas introduced by Shannon have led to an information revolution

Random coding and its optimality have directly inspired

- Low density parity check codes (LDPC)
- Polar codes
  - proof of sub-additivity

We are now (fully immersed) in a wireless world

- Network of users sharing same medium

- Clear need to maximally utilize the limited resources (power, bandwidth, energy)

- Develop a similar understanding in network settings
  - But we first need to fully understand the basic building blocks
1. Multiple Access Channel (uplink) (Shannon ’61)

\[ W(y|x_1, x_2) \]

\[ Y^n \rightarrow \text{Decoder} \rightarrow (\hat{M}_1, \hat{M}_2) \]

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

\[ R_1 \leq I(X_1; Y | X_2, Q) \]
\[ R_2 \leq I(X_2; Y | X_1, Q) \]
\[ R_1 + R_2 \leq I(X_1, X_2; Y | Q) \]

for some \( p(q | p(x_1 | q), p(x_2 | q)) \). It suffices to consider \( |Q| \leq 2 \). Call this region \( R(W) \).

**Question:** Is this the capacity (optimal) region? (YES) (Ahlswede ’72)
Random coding can be used to achieve rate pairs \((R_1, R_2)\) that satisfy

\[
\begin{align*}
R_1 &\leq I(X_1; Y|X_2, Q) \\
R_2 &\leq I(X_2; Y|X_1, Q) \\
R_1 + R_2 &\leq I(X_1, X_2; Y|Q)
\end{align*}
\]

for some \(p(q)p(x_1|q)p(x_2|q)\); it suffices to consider \(|Q| \leq 2\). Call this region \(\mathcal{R}(W)\).
Random coding can be used to achieve rate pairs $(R_1, R_2)$ that satisfy

\[
R_1 \leq I(X_1; Y | X_2, Q) \\
R_2 \leq I(X_2; Y | X_1, Q) \\
R_1 + R_2 \leq I(X_1, X_2; Y | Q)
\]

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

**Question:** Is this the capacity (optimal) region?
1. Multiple Access Channel (uplink) (Shannon ’61)

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

\[
R_1 \leq I(X_1; Y|X_2, Q) \\
R_2 \leq I(X_2; Y|X_1, Q) \\
R_1 + R_2 \leq I(X_1, X_2; Y|Q)
\]

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

**Question:** Is this the capacity (optimal) region? (YES) (Ahlswede ’72)
Testing optimality

Define, for $\lambda \geq 1$,

$$S_\lambda(W) = \max_{(R_1, R_2) \in \mathcal{R}(W)} \left\{ \lambda R_1 + R_2 \right\}$$

$$= \max_{p_1(x_1)p_2(x_2)} \left\{ (\lambda - 1)I(X_1; Y|X_2) + I(X_1, X_2; Y) \right\}$$
Testing optimality

Define, for $\lambda \geq 1$,

$$S_\lambda(W) = \max_{(R_1, R_2) \in \mathcal{R}(W)} \left\{ \lambda R_1 + R_2 \right\}$$

$$= \max_{p_1(x_1)p_2(x_2)} \left\{ (\lambda - 1)I(X_1; Y|X_2) + I(X_1, X_2; Y) \right\}$$
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:

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

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

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

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

Define, for $\lambda \geq 1$,

$$S_\lambda(W) = \max_{(R_1, R_2) \in \mathcal{R}(W)} \left\{ \lambda R_1 + R_2 \right\}$$

$$= \max_{p_1(x_1)p_2(x_2)} \left\{ (\lambda - 1)I(X_1; Y|X_2) + I(X_1, X_2; Y) \right\}$$

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:

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

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

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

**Note:** Computing $S_\lambda(W)$ is relatively easy since

$$\left\{ (\lambda - 1)I(X_1; Y|X_2) + I(X_1, X_2; Y) \right\}$$ is concave in $p_1(x_1), p_2(x_2)$. 
2. Broadcast channel (downlink) (Cover ’72)

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

\[
R_0 \leq \min\{I(Q; Y_1), I(Q; Y_2)\}
\]

\[
R_0 + R_1 \leq I(U, Q; Y_1)
\]

\[
R_0 + R_2 \leq I(V, Q; Y_2)
\]

\[
R_0 + R_1 + R_2 \leq \min\{I(Q; Y_1), I(Q; Y_2)\} + I(U; Y_1|Q) + I(V; Y_2|Q) - I(U; V|Q)
\]

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

Question: Is this the capacity (optimal) region? (Open) (since Marton ’79)
Superposition coding and random hashing can be used to achieve rate triples \((R_0, R_1, R_2)\) that satisfy

\[
R_0 \leq \min\{I(Q; Y_1), I(Q; Y_2)\}
\]

\[
R_0 + R_1 \leq I(U, Q; Y_1)
\]

\[
R_0 + R_2 \leq I(V, Q; Y_2)
\]

\[
R_0 + R_1 + R_2 \leq \min\{I(Q; Y_1), I(Q; Y_2)\} + I(U; Y_1|Q) + I(V; Y_2|Q) - I(U; V|Q)
\]

for some \(p(q, u, v, x)\). Call this region \(\mathcal{R}(W_a, W_b)\).
Superposition coding and random hashing can be used to achieve rate triples $(R_0, R_1, R_2)$ that satisfy

\[ R_0 \leq \min\{I(Q; Y_1), I(Q; Y_2)\} \]
\[ R_0 + R_1 \leq I(U, Q; Y_1) \]
\[ R_0 + R_2 \leq I(V, Q; Y_2) \]
\[ R_0 + R_1 + R_2 \leq \min\{I(Q; Y_1), I(Q; Y_2)\} + I(U; Y_1|Q) + I(V; Y_2|Q) - I(U; V|Q) \]

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)

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

\[
\begin{align*}
R_0 &\leq \min\{I(Q;Y_1), I(Q;Y_2)\} \\
R_0 + R_1 &\leq I(U,Q;Y_1) \\
R_0 + R_2 &\leq I(V,Q;Y_2) \\
R_0 + R_1 + R_2 &\leq \min\{I(Q;Y_1), I(Q;Y_2)\} + I(U;Y_1|Q) \\
&\quad + I(V;Y_2|Q) - I(U;V|Q)
\end{align*}
\]

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 \(\lambda \geq 1\),

\[
S_\lambda(W) = \max_{(R_1, R_2) \in \mathcal{R}(W_a, W_b)} \{ \lambda R_1 + R_2 \}
\]

\[
= \max_{p(u,v,w,x)} \left\{ (\lambda - 1) I(U, Q; Y_1) + \min\{ I(Q; Y_1), I(Q; Y_2) \} + I(U; Y_1|Q) 
+ I(V; Y_2|Q) - I(U; V|Q) \right\}
\]

\[
= \min_{\alpha \in [0, 1]} \max_{p(u,v,w,x)} \left\{ (\lambda - \alpha) I(Q; Y_1) + \alpha I(Q; Y_2) + \lambda I(U; Y_1|Q) 
+ I(V; Y_2|Q) - I(U; V|Q) \right\}
\]

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 \(\lambda \geq 1\),

\[
S_\lambda(W) = \max_{(R_1, R_2) \in \mathcal{R}(W_a, W_b)} \{\lambda R_1 + R_2\}
\]

\[
= \max_{p(u,v,w,x)} \left\{ (\lambda - 1)I(U, Q; Y_1) + \min\{I(Q; Y_1), I(Q; Y_2)\} + I(U; Y_1|Q)
\right.
\]

\[
+ I(V; Y_2|Q) - I(U; V|Q) \}
\]

\[
= \min_{\alpha \in [0,1]} \max_{p(u,v,w,x)} \left\{ (\lambda - \alpha)I(Q; Y_1) + \alpha I(Q; Y_2) + \lambda I(U; Y_1|Q)
\right.
\]

\[
+ I(V; Y_2|Q) - I(U; V|Q) \}
\]

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

- $R(W_a, W_b)$ is optimal on $R_1 = 0$ (or $R_2 = 0$)
  - Degrade message sets: Korner and Marton ('77)

- $R(W_a, W_b)$ 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)

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

\[
\begin{align*}
R_1 &< I(X_1; Y_1|U_2, Q), \\
R_2 &< I(X_2; Y_2|U_1, Q), \\
R_1 + R_2 &< I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q), \\
R_1 + R_2 &< I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q), \\
R_1 + R_2 &< I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q),
\end{align*}
\]

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

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 \(\mathcal{R}(W_a, W_b)\).
Superposition coding, message splitting, coded time-sharing can be used to achieve rate pairs \((R_1, R_2)\) that satisfy

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

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

\[
R_1 < I(X_1; Y_1|U_2, Q), \\
R_2 < I(X_2; Y_2|U_1, Q), \\
R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q), \\
R_1 + R_2 < I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q), \\
R_1 + R_2 < I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q), \\
2R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|U_2, Q), \\
R_1 + 2R_2 < I(X_2, U_1; Y_2|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|U_1, Q)
\]

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?

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:

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

- Broadcast channel: Establishing optimality of Marton’s region for MIMO broadcast channel

- Interference channel: Sub-optimality of the Han–Kobayashi region

- Family of non-convex optimization problems
  - Relation to problems of interest in other fields
  - Unifying observations and some conjectures
MIMO (Vector) Gaussian broadcast channel

\[ Y_1^n = AX^n + Z, \quad Y_2^n = BX^n + Z \]

where \( Z \sim \mathcal{N}(0, I) \) denotes the additive noise.

Very important channel class in wireless communication: multi-antenna transmitter/receivers (downlink).
MIMO (Vector) Gaussian broadcast channel:

\[ Y_1 = AX + Z \]
\[ Y_2 = BX + Z \]

where \( Z \sim \mathcal{N}(0, I) \) denotes the additive noise.

Very important channel class in wireless communication

Models: multi-antenna transmitter/receivers (downlink)
History

Optimality of Marton’s bound, $\mathcal{R}(W_a, W_b)$, was established:

- Scalar case (Bergmans ’73) (Entropy Power Inequality)

- Reversely degraded setting (Poltyrev ’78, El Gamal ’81)

- Optimality on $R_0 = 0$ (Weingarten and Steinberg and Shamai ’06)
  - Builds on ideas in Poltyrev
  - Tour de force in optimization
Optimality of Marton’s bound, $\mathcal{R}(W_a, W_b)$, was established:

- Scalar case (Bergmans ’73) *(Entropy Power Inequality)*
- Reversely degraded setting (Poltyrev ’78, El Gamal ’81)
- Optimality on $R_0 = 0$ (Weingarten and Steinberg and Shamai ’06)
  - Builds on ideas in Poltyrev
  - Tour de force in optimization
  - **Gist**: Showing optimality of Gaussian random variables for a non-convex optimization problem
  - **Remark**: Ideas do not extend to show optimality when there is common message, i.e. $R_0 \neq 0$
Optimality of Marton’s bound, $\mathcal{R}(W_a, W_b)$, was established:

- Scalar case (Bergmans ’73) (Entropy Power Inequality)
- Reversely degraded setting (Poltyrev ’78, El Gamal ’81)
- Optimality on $R_0 = 0$ (Weingarten and Steinberg and Shamai ’06)
  - Builds on ideas in Poltyrev
  - Tour de force in optimization
  - **Gist**: Showing optimality of Gaussian random variables for a non-convex optimization problem
  - **Remark**: Ideas do not extend to show optimality when there is common message, i.e. $R_0 \neq 0$

- Optimality in general (Geng and Nair ’14)
  - **Gist**: Develop a technique for proving optimality of Gaussian random variables (from sub-additivity)
Optimality of Marton’s bound, $\mathcal{R}(W_a, W_b)$, was established:

- Scalar case (Bergmans ’73) (Entropy Power Inequality)

- Reversely degraded setting (Poltyrev ’78, El Gamal ’81)

- Optimality on $R_0 = 0$ (Weingarten and Steinberg and Shamai ’06)
  - Builds on ideas in Poltyrev
  - Tour de force in optimization
  - Gist: Showing optimality of Gaussian random variables for a non-convex optimization problem
  - Remark: Ideas do not extend to show optimality when there is common message, i.e. $R_0 \neq 0$

- Optimality in general (Geng and Nair ’14)
  - Gist: Develop a technique for proving optimality of Gaussian random variables (from sub-additivity)

Explain our technique on $R_0 = 0$ (for simplicity)
Outer bound (Korner-Marton ’79)

The set of rate pairs \((R_1, R_2)\) satisfying

\[
R_2 \leq I(U; Y_2)
\]
\[
R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1|U)
\]

for some \(p(u, x)\), where \(E(\|X\|^2) \leq P\) forms an outer bound to the capacity region. Denote this region as \(\mathcal{O}(W_a, W_b)\).
Outer bound (Korner-Marton ’79)

The set of rate pairs \((R_1, R_2)\) satisfying

\[
R_2 \leq I(U; Y_2) \\
R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1|U)
\]

for some \(p(u, x)\), where \(E(\|X\|^2) \leq P\) forms an outer bound to the capacity region. Denote this region as \(\mathcal{O}(W_a, W_b)\).

For \(\lambda > 1\), let

\[
S_\lambda(W_a, W_b) := \max_{(R_1, R_2) \in \mathcal{O}} R_1 + \lambda R_2 \\
= \max_{p(u, x)} \lambda I(U; Y_2) + I(X; Y_1|U)
\]
Outer bound (Korner-Marton ’79)

The set of rate pairs \((R_1, R_2)\) satisfying

\[
R_2 \leq I(U; Y_2)
\]
\[
R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1|U)
\]

for some \(p(u, x)\), where \(E(\|X\|^2) \leq P\) forms an outer bound to the capacity region. Denote this region as \(\mathcal{O}(W_a, W_b)\).

For \(\lambda > 1\), let

\[
S_\lambda(W_a, W_b) := \max_{(R_1, R_2) \in \mathcal{O}} R_1 + \lambda R_2
\]
\[
= \max_{p(u, x)} \lambda I(U; Y_2) + I(X; Y_1|U)
\]
\[
= \max_{p(x)} \left\{ \lambda I(X; Z) + C_{\mu_X}[I(X; Y) - \lambda I(X; Z)] \right\}
\]

(Nair ’13)
Outer bound (Korner-Marton ’79)

The set of rate pairs \((R_1, R_2)\) satisfying

\[
R_2 \leq I(U; Y_2) \\
R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1|U)
\]

for some \(p(u, x)\), where \(E(\|X\|^2) \leq P\) forms an outer bound to the capacity region. Denote this region as \(\mathcal{O}(W_a, W_b)\).

For \(\lambda > 1\), let

\[
S_\lambda(W_a, W_b) := \max_{(R_1, R_2) \in \mathcal{O}} R_1 + \lambda R_2
\]

\[
= \max_{p(u, x)} \lambda I(U; Y_2) + I(X; Y_1|U)
\]

\[
= \max_{p(x)} \left\{ \lambda I(X; Z) + C_{\mu_X} [I(X; Y) - \lambda I(X; Z)] \right\}
\]

(Nair ’13)

Not easy to compute (in general)
One can show that if Gaussians maximize

\[ C_{\mu_X}[h(Y_1) - \lambda h(Y_2)] \]

then Marton’s inner bound is optimal (on \( R_0 = 0 \))

Here \( h(X) \) is the differential entropy:

\[ h(X) := - \int f(x) \log f(x) \, dx, \]

where \( f(x) \) is the density function of \( X \).

A similar (more-involved) problem shows up when \( R_0 \neq 0 \)

An identical technique (to the one I am going to demonstrate) establishes that also
Gaussian optimality via sub-additivity (Geng-Nair ’14)

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

$$C_{\mu X}[h(AX + Z) - \lambda h(BX + Z)]$$

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

We will see that the maximum value is

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

where $X_\ast \sim \mathcal{N}(0, K')$ for some $K' \preceq K$. 
Maximize, for $\lambda > 1$, the value of the functional

$$
C_{\mu X}[h(AX + Z) - \lambda h(BX + Z)]
$$

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

We will see that the maximum value is

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

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

**Lemma:** $C_{\mu X}[h(AX + Z) - \lambda h(BX + Z)]$ is sub-additive.

**Proof:** For any $\mu_{X_1, X_2}$

$$
\begin{align*}
&h(AX_1 + Z_1, AX_2 + Z_2|U) - \lambda h(BX_1 + Z_1, BX_2 + Z_2|U) \\
&= h(AX_1 + Z_1|U, AX_2 + Z_2) - \lambda h(BX_1 + Z_1|U, AX_2 + Z_2) \\
&\quad + h(AX_2 + Z_2|U, BX_1 + Z_1) - \lambda h(BX_2 + Z_2|U, BX_1 + Z_1) \\
&\quad - (\lambda - 1)I(AX_2 + Z_2; BX_1 + Z_1|U)
\end{align*}
$$
Maximize, for $\lambda > 1$, the value of the functional

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

over $X : E(XX^T) \preceq K$, 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:** $C_{\mu_X}[h(AX + Z) - \lambda h(BX + Z)]$ is sub-additive.

**Proof:** For any $\mu_{X_1, X_2}$

$$C_{\mu_{X_1, X_2}}[h(AX_1 + Z_1, AX_2 + Z_2) - \lambda h(BX_1 + Z_1, BX_2 + Z_2)]$$

$$\leq C_{\mu_{X_1}}[h(AX_1 + Z_1) - \lambda h(BX_1 + Z_1)]$$

$$+ C_{\mu_{X_2}}[h(AX_2 + Z_2) - \lambda h(BX_2 + Z_2)]$$
Gaussian optimality via sub-additivity (Geng-Nair ’14)

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

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

over $X : E(XX^T) \preceq K$, 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:** $C_{\mu_X}[h(AX + Z) - \lambda h(BX + Z)]$ is sub-additive.

**Proof:** For any $\mu_{X_1, X_2}$

$$h(AX_1 + Z_1, AX_2 + Z_2 | U) - \lambda h(BX_1 + Z_1, BX_2 + Z_2 | U)$$

$$= h(AX_1 + Z_1 | U, AX_2 + Z_2) - \lambda h(BX_1 + Z_1 | U, AX_2 + Z_2)$$

$$+ h(AX_2 + Z_2 | U, BX_1 + Z_1) - \lambda h(BX_2 + Z_2 | U, BX_1 + Z_1)$$

$$-(\lambda - 1)I(AX_2 + Z_2; BX_1 + Z_1 | U)$$
Let \((U, X)\) be a maximizer, i.e.
\[
V = \max_{\mu_X} \mathcal{C}_{\mu_X}[h(AX + Z) - \lambda h(BX + Z)] = h(AX + Z|U) - \lambda h(BX + Z|U).
\]

Let \((X_a, U_a)\) and \((X_b, U_b)\) be i.i.d. according to \((U, X)\).
Gaussian optimality: ctd..

Let \((U \dagger, X \dagger)\) be a maximizer, i.e.

\[
V = \max_{\mu_X} C_{\mu_X}[h(AX + Z) - \lambda h(BX + Z)] = h(AX \dagger + Z|U \dagger) - \lambda h(BX \dagger + Z|U \dagger).
\]

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

\[
2V = C_{\mu_{X_1, X_2}}[h(AX_1 + Z_1, AX_2 + Z_2) - \lambda h(BX_1 + Z_1, BX_2 + Z_2)]
\]

\[
\leq C_{\mu_{X_1}}[h(AX_1 + Z_1) - \lambda h(BX_1 + Z_1)]_{\mu_{X_+}}
\]

\[
+ C_{\mu_{X_2}}[h(AX_2 + Z_2) - \lambda h(BX_2 + Z_2)]_{\mu_{X_-}}
\]

\[
-(\lambda - 1)I(AX_- + Z_2; BX_+ + Z_1|U_a, U_b)
\]

\[
\leq V + V
\]
Gaussian optimality: ctd..

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

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

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

\[
2V = C_{\mu_{X_1}, X_2} [h(AX_1 + Z_1, AX_2 + Z_2) - \lambda h(BX_1 + Z_1, BX_2 + Z_2)] \bigg|_{\mu_{X_+, X_-}}
\]

\[
\leq C_{\mu_{X_1}} [h(AX_1 + Z_1) - \lambda h(BX_1 + Z_1)] \bigg|_{\mu_{X_+}}
\]

\[
+ C_{\mu_{X_2}} [h(AX_2 + Z_2) - \lambda h(BX_2 + Z_2)] \bigg|_{\mu_{X_-}}
\]

\[
-(\lambda - 1) I(AX_- + Z_2; BX_+ + Z_1|U_a, U_b)
\]

\[
\leq V + V
\]

**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.

\[
V = \max_{\mu_X} C_{\mu_X} \left[ h(AX + Z) - \lambda h(BX + Z) \right] = h(AX\dagger + Z|U\dagger) - \lambda h(BX\dagger + Z|U\dagger).
\]

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.

\[
V = \max_{\mu_X} C_{\mu_X} [h(AX + Z) - \lambda h(BX + Z)] = h(AX^\dagger + Z|U^\dagger) - \lambda h(BX^\dagger + Z|U^\dagger).
\]

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)

This technique has subsequently been used by others in various other instances.

Note: There are some similarities with work of Lieb and Barthe (90s)
They also use rotations (but not information measures and its algebra)
Gaussian optimality: ctd..

Let \((U^\dagger, X^\dagger)\) be a maximizer, i.e.

\[
V = \max_{\mu_X} C_{\mu_X} [h(AX + Z) - \lambda h(BX + Z)] = h(AX^\dagger + Z|U^\dagger) - \lambda h(BX^\dagger + Z|U^\dagger).
\]

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)

This technique has been subsequently used by others in various other instances.

**Note:** There are some similarities with work of Lieb and Barthe (90s)
They also use rotations (but not information measures and its algebra)
An open question

We have seen (yesterday and today) how **sub-additivity** implies Gaussian optimality.

**Open question**

For $\alpha, a \in (0, 1)$, do Gaussians maximize the functional

$$\alpha h(X_2 + aX_1 + Z) + (1 - \alpha) h(X_1 + Z) - h(aX_1 + Z)$$

over $X_1 \perp X_2$, subject to $E(X_1^2) \leq P_1$, $E(X_2^2) \leq P_2$. Here $Z \sim \mathcal{N}(0, 1)$ is independent of $X_1, X_2$. 

**Why should someone care?**

- If true, solves the capacity region for the Gaussian $Z$-interference channel
- Related to reverse EPIs, hyperplane conjecture, etc.
An open question

We have seen (yesterday and today) how sub-additivity implies Gaussian optimality.

Open question

For $\alpha, a \in (0, 1)$, do Gaussians maximize the functional

$$\alpha h(X_2 + aX_1 + Z) + (1 - \alpha) h(X_1 + Z) - h(aX_1 + Z)$$

over $X_1 \perp X_2$, subject to $E(X_1^2) \leq P_1$, $E(X_2^2) \leq P_2$. Here $Z \sim \mathcal{N}(0, 1)$ is independent of $X_1, X_2$.

Affirmative if the following functional sub-additive?

$$C_{\mu_{X_1}} [\alpha h(X_2 + aX_1 + Z) + (1 - \alpha) h(X_1 + Z) - h(aX_1 + Z)]$$
An open question

We have seen (yesterday and today) how sub-additivity implies Gaussian optimality

Open question

For $\alpha, a \in (0, 1)$, do Gaussians maximize the functional

$$\alpha h(X_2 + aX_1 + Z) + (1 - \alpha) h(X_1 + Z) - h(aX_1 + Z)$$

over $X_1 \perp X_2$, subject to $E(X_1^2) \leq P_1$, $E(X_2^2) \leq P_2$. Here $Z \sim \mathcal{N}(0, 1)$ is independent of $X_1, X_2$.

Affirmative if the following functional sub-additive?

$$C_{\mu_{X_1}}[\alpha h(X_2 + aX_1 + Z) + (1 - \alpha) h(X_1 + Z) - h(aX_1 + Z)]$$

Why should someone care?

- If true, solves the capacity region for the Gaussian $Z$-interference channel
- Related to reverse EPIs, hyperplane conjecture, etc.
Outline

- Broadcast channel: Establishing optimality of Marton’s for MIMO broadcast channel
- Interference channel: Sub-optimality of the Han–Kobayashi region
- Family of non-convex optimization problems
  - Relation to problems of interest in other fields
  - Unifying observations and some conjectures
Interference Channel (Ahlswede ’74)

\[ M_1 \xrightarrow{X_1^n} W_a(y_1|x_1, x_2) \xrightarrow{Y_1^n} \hat{M}_1 \]

\[ M_2 \xrightarrow{X_2^n} W_b(y_2|x_1, x_2) \xrightarrow{Y_2^n} \hat{M}_2 \]
A rate-pair \((R_1, R_2)\) is achievable for the interference channel if

\[
R_1 < I(X_1; Y_1|U_2, Q),
\]
\[
R_2 < I(X_2; Y_2|U_1, Q),
\]
\[
R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q),
\]
\[
R_1 + R_2 < I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q),
\]
\[
R_1 + R_2 < I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q),
\]
\[
2R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|U_2, Q),
\]
\[
R_1 + 2R_2 < I(X_2, U_1; Y_2|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|U_1, Q)
\]

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 \(|Q| \leq 7\). Denote the (closure of) region as \(\mathcal{R}(W_a, W_b)\).

Numerically infeasible to compute \(\mathcal{R}(W_a W_b)\) even for generic binary-input binary-output interference channels.
Han-Kobayashi achievable region (1981) á la Chong et. al.

A rate-pair \((R_1, R_2)\) is achievable for the interference channel if

\[
R_1 < I(X_1; Y_1|U_2, Q),
\]
\[
R_2 < I(X_2; Y_2|U_1, Q),
\]
\[
R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q),
\]
\[
R_1 + R_2 < I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q),
\]
\[
R_1 + R_2 < I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q),
\]
\[
2R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|U_2, Q),
\]
\[
R_1 + 2R_2 < I(X_2, U_1; Y_2|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|U_1, Q)
\]

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 \(|Q| \leq 7\). Denote the (closure of) region as \(\mathcal{R}(W_a, W_b)\).

Numerically infeasible to compute \(\mathcal{R}(W_aW_b)\) even for generic binary-input binary-output interference channels

First step:

- Find a channel class where HK region simplifies AND yet not too trivial
Clean Z Interference Channel (CZIC) Model

![Diagram of CZIC model]

**Lemma:** A rate-pair $(R_1, R_2)$ belongs to Han-Kobayashi region if and only if

\[
R_1 < I(X_1; Y_1|U_2, Q),
\]
\[
R_2 < H(X_2|Q),
\]
\[
R_1 + R_2 < I(X_1, U_2; Y_1|Q) + H(X_2|U_2, Q),
\]

for some pmf $p(q)p(x_1|q)p(u_2, x_2|q)$, where $|U_2| \leq |X_2|$ and $|Q| \leq 2$.

Denote region: $\mathcal{R}(W_a)$
Testing optimality

Equivalent to test if

\[ S_\lambda(W_a \otimes W_a) = 2S_\lambda(W_a), \forall W_a, \lambda \geq 0, \]

where

\[ S_\lambda(W_a) := \max_{(R_1, R_2) \in \mathcal{R}(W_a)} \lambda R_1 + R_2. \]
Testing optimality

Equivalent to test if

\[ S_\lambda(W_a \otimes W_a) = 2S_\lambda(W_a), \quad \forall W_a, \lambda \geq 0, \]

where

\[ S_\lambda(W_a) := \max_{(R_1, R_2) \in \mathcal{R}(W_a)} \lambda R_1 + R_2. \]

For \( \lambda \in [0, 1] \), \( S_\lambda(W_a) \) is given by

\[
\begin{align*}
\max_{p_1(x_1)p_2(u_2,x_2)} \left\{ (1 - \lambda)H(X_2) + \lambda I(X_1, U_2; Y_1) + \lambda H(X_2|U_2) \right\} \\
= \max_{p_1(x_1)p_2(x_2)} \left\{ H(X_2) + \lambda I(X_1; Y_1) \right\}
\end{align*}
\]
Testing optimality

Equivalent to test if

\[ S_\lambda(W_a \otimes W_a) = 2S_\lambda(W_a), \ \forall \ W_a, \lambda \geq 0, \]

where

\[ S_\lambda(W_a) := \max_{(R_1,R_2) \in \mathcal{R}(W_a)} \lambda R_1 + R_2. \]

For \( \lambda \in [0,1] \), \( S_\lambda(W_a) \) is given by

\[
\max_{p_1(x_1)p_2(u_2,x_2)} \left\{ (1 - \lambda)H(X_2) + \lambda I(X_1,U_2;Y_1) + \lambda H(X_2|U_2) \right\} = \max_{p_1(x_1)p_2(x_2)} \left\{ H(X_2) + \lambda I(X_1;Y_1) \right\}
\]

**Lemma (sub-additivity)** (Nair-Xia-Yazdanpanah ’15):

\[
H(X_{21},X_{22}) + \lambda I(X_{11},X_{12};Y_{11},Y_{12}) \leq \left\{ H(X_{21}) + \lambda I(X_{11};Y_{11}) \right\} + \left\{ H(X_{22}) + \lambda I(X_{12};Y_{12}) \right\} - (1 - \lambda)I(X_{21};X_{22}).
\]

Implies optimality of \( S_\lambda(W_a), \lambda \in [0,1] \).
What about $\lambda > 1$?

For $\lambda \geq 1$, $S_\lambda(W_a)$ is given by

$$
\max_{p_1(x_1)p_2(u_2,x_2)} \left\{ I(X_1, U_2; Y_1) + H(X_2|U_2) + (\lambda - 1)I(X_1; Y_1|U_2) \right\}
$$

$$
= \max_{p_1(x_1)p_2(x_2)} \left\{ I(X_1, X_2; Y_1) + C_{X_2}[(\lambda - 1)I(X_1; Y_1) + H(X_2) - I(X_2; Y_1|X_1)] \right\}
$$
What about $\lambda > 1$?

For $\lambda \geq 1$, $S_{\lambda}(W_a)$ is given by

$$
\max_{p_1(x_1)p_2(u_2,x_2)} \left\{ I(X_1, U_2; Y_1) + H(X_2|U_2) + (\lambda - 1)I(X_1; Y_1|U_2) \right\}
$$

$$
= \max_{p_1(x_1)p_2(x_2)} \left\{ I(X_1, X_2; Y_1) + C_{X_2}[(\lambda - 1)I(X_1; Y_1) + H(X_2) - I(X_2; Y_1|X_1)] \right\}
$$

**Question**: Can we numerically test if $S_{\lambda}(W_a \otimes W_a) = 2S_{\lambda}(W_a)$?

$X_2$ is a binary random variable (i.e. concave envelope over single variable)

$(\lambda - 1)I(X_1; Y_1) + H(X_2) - I(X_2; Y_1|X_1)$: has at most 2 inflexion points
What about $\lambda > 1$?

The shape of concave envelope for a generic binary CZIC
Sub-optimality of the Han-Kobayashi region

| λ   | $W(Y_1 = 0|X_1, X_2)$ | $\mathcal{A}^{HK}_{\lambda}(W)$ | $\frac{1}{2}\mathcal{A}^{\text{TIN}}_{\lambda}(W \otimes 2)$ |
|-----|------------------------|---------------------------------|------------------------------------------------------|
| 2   | $\begin{bmatrix} 1 & 0.5 \\ 1 & 0 \end{bmatrix}$ | 1.107516                        | 1.108141                                             |
| 9   | $\begin{bmatrix} 0.12 & 0.89 \\ 0.21 & 0.62 \end{bmatrix}$ | 1.074484                        | 1.075544                                             |
| 12  | $\begin{bmatrix} 0.01 & 0.58 \\ 0.20 & 0.74 \end{bmatrix}$ | 1.289830                        | 1.293760                                             |
| 14  | $\begin{bmatrix} 0.78 & 0.07 \\ 0.46 & 0.05 \end{bmatrix}$ | 1.426526                        | 1.432419                                             |
| 15  | $\begin{bmatrix} 0.91 & 0.22 \\ 0.66 & 0.15 \end{bmatrix}$ | 1.323766                        | 1.339065                                             |
| 16  | $\begin{bmatrix} 0.91 & 0.13 \\ 0.62 & 0.06 \end{bmatrix}$ | 1.515421                        | 1.534724                                             |
| 18  | $\begin{bmatrix} 0.38 & 0.87 \\ 0.12 & 0.79 \end{bmatrix}$ | 1.449959                        | 1.468577                                             |

Counterexamples to the optimality of Han-Kobayashi region.

**Note:** For the first example, we can calculate the concave envelope analytically.
Particular Channel

- We compute $\max_{HK} \lambda R_1 + R_2$ for $\lambda = 2$

$$
\max_{p_1(x_1)p_2(x_2)} \left( H(Y_1) + \frac{C}{p_2(x_2)} \left[ H(X_2) + 2H(Y_1) - H(Y_1|X_1) \right] \right)
$$
Particular Channel

We compute \( \max_{HK} \lambda R_1 + R_2 \) for \( \lambda = 2 \)

\[
\max_{p_1(x_1)p_2(x_2)} \left( H(Y_1) + \mathcal{C} \left[ H(X_2) + 2H(Y_1) - H(Y_1|X_1) \right] \right)
\]

Let \( p \) and \( q \) respectively denote \( Pr(X_1 = 0) \) and \( Pr(X_2 = 0) \)

\[
f(p, q) = (1 - 2\bar{p})h_b(q) + h_b(q + \frac{p}{2\bar{q}}) - 2ph_b\left(\frac{q + 1}{2}\right)
\]

where \( h_b(.) \) is the binary entropy function.
Particular channel continued

$f(p, q)$ is concave in $q$ for $p \geq \frac{1}{2}$ and for $0 \leq p < \frac{1}{2}$

$$C_q[f(p, q)] = \begin{cases} 
  \frac{f(p, q)}{1 - 2p}q + f(p, 0) & q > 1 - 2p \\
  \frac{f(p, 1 - 2p) - f(p, 0)}{1 - 2p}q + f(p, 0) & q \in [0, 1 - 2p]
\end{cases}$$
Particular channel continued

\( f(p, q) \) is concave in \( q \) for \( p \geq \frac{1}{2} \) and for \( 0 \leq p < \frac{1}{2} \)

\[
\mathcal{C}_q[f(p, q)] = \begin{cases} 
    f(p, q) & q > 1 - 2p \\
    \frac{f(p, 1 - 2p) - f(p, 0)}{1 - 2p} q + f(p, 0) & q \in [0, 1 - 2p]
\end{cases}
\]
Particular channel continued

\( f(p, q) \) is concave in \( q \) for \( p \geq \frac{1}{2} \) and for \( 0 \leq p < \frac{1}{2} \)

\[
\mathcal{Q}_q[f(p, q)] = \begin{cases} 
  f(p, q) & q > 1 - 2p \\
  \frac{f(p, 1 - 2p) - f(p, 0)}{1 - 2p}q + f(p, 0) & q \in [0, 1 - 2p]
\end{cases}
\]

Corollary

Maximum of \( 2R_1 + R_2 \) for the Han–Kobayashi region is equal to the maximum of \( T(p, q) \) for \( (p, q) \in [0, 1] \times [0, 1] \), where

\[
T(p, q) = \begin{cases} 
  h_b(q + \frac{p}{2} \bar{q}) + f(p, q) & q \geq \min\{0, 1 - 2p\} \\
  h_b(q + \frac{p}{2} \bar{q}) + \frac{f(p, 1 - 2p) - f(p, 0)}{1 - 2p}q + f(p, 0) & \text{o.w.,}
\end{cases}
\]

where \( f(p, q) = (1 - 2\bar{p})h_b(q) + h_b(q + \frac{p}{2} \bar{q}) - 2ph_b(\frac{q+1}{2}) \)
Numerical search indicates: $\max_{p,q} T(p, q) = 1.107516..$ at $p = 0.5078..$ and $q = 0.4365..$
Particular channel continued

- **Interval arithmetic** is a method to obtain formal bounds for functions consisting of basic arithmetic functions and commonly used functions such as logarithms and trigonometric functions.

- $T(p, q)$ only includes basic arithmetic functions and logarithm.

- We used Julia based implementation of this formal method to obtain
  \[
  \max T(p, q) \in [1.10751, 1.10769]
  \]

- The 2-letter TIN achieves $2R_1 + R_2 = 1.108141$ at the distribution
  \[
  \begin{align*}
  P((X_{11}, X_{12}) = (0, 0)) &= p \\
  P((X_{11}, X_{12}) = (1, 1)) &= 1 - p \\
  P((X_{21}, X_{22}) = (0, 0)) &= 0.36q \\
  P((X_{21}, X_{22}) = (1, 1)) &= 1 - 1.64q \\
  P((X_{21}, X_{22}) = (0, 1)) &= 0.64q \\
  P((X_{21}, X_{22}) = (1, 0)) &= 0.64q 
  \end{align*}
  \]

  where $p = 0.507829413$, $q = 0.436538150$

- **Repetition coding** across time seems to outperform memoryless coding
What about 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 \( \lambda \geq 1 \).

\[
C_{\mu_X} \left[ (\lambda - \alpha)H(Y) - \alpha H(Z) + \max_{p(u,v|x)} \{ \lambda I(U;Y) + I(V;Z) - I(U;V) \} \right]
\]
What about 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
What about 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-additivity for binary input broadcast channels
- Can prove sub-additivity when \( \alpha = 0 \) or \( \alpha = 1 \).
Remarks

- **Idea**: To demonstrate **super-additivity**

- **Difficulty**: Identify a sufficiently simple class where
  - Evaluation of the region is possible: non-convex optimization
  - Super-additivity holds
Remarks

- **Idea:** To demonstrate **super-additivity**

- **Difficulty:** Identify a sufficiently simple class where
  - Evaluation of the region is possible: non-convex optimization
  - Super-additivity holds

This idea was also used to resolve

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

**NO** (Nair, Yazdanpanah ’17)
Outline

- **Broadcast channel**: Establishing optimality of Marton’s for MIMO broadcast channel

- **Interference channel**: Sub-optimality of the Han–Kobayashi region

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

\[ C_{\nu_X} \left[ \sum_{S \subseteq [n]} \alpha_S H(X_S) \right], \alpha_S \in \mathbb{R}. \]
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

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

Using Fenchel duality this is same as

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

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

\[
G_{12}(\gamma_1, \gamma_2) := \max_{\mu_{X_1, X_2}} \sum_{S \subseteq [n]} \alpha_S H(X_{1S}, X_{2S}) - E(\gamma_1(X_1)) - E(\gamma_2(X_2))
\]

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

\[ C_{\nu_{\mu}} \left[ \sum_{S \subseteq [n]} \alpha_S H(X_S) \right], \alpha_S \in \mathbb{R}. \]

Using Fenchel duality this is same as

\[
G_1(\gamma_1) := \max_{\mu_X} \sum_{S \subseteq [n]} \alpha_S H(X_S) - E(\gamma_1(X)) \\
G_2(\gamma_2) := \max_{\mu_X} \sum_{S \subseteq [n]} \alpha_S H(X_S) - E(\gamma_2(X)) \\
G_{12}(\gamma_1, \gamma_2) := \max_{\mu_{X_1}, \mu_{X_2}} \sum_{S \subseteq [n]} \alpha_S H(X_{1S}, X_{2S}) - E(\gamma_1(X_1)) - E(\gamma_2(X_2))
\]

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 \(\|Z\|_p = \mathbb{E}(|Z|^p)^{\frac{1}{p}}\).
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 evaluating 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

\[C_{\nu_{X,Y}} \left[H(X, Y) - (1 - \frac{1}{p})H(X) - \frac{1}{q}H(Y)\right]_{\mu_{X,Y}} = H(X, Y) - (1 - \frac{1}{p})H(X) - \frac{1}{q}H(Y)\]
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 \(1 \leq q \leq p\) if

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

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

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

Rather immediate that sub-additivity, i.e.

\[
\begin{align*}
C_{\mu_{X_1Y_1X_2Y_2}}[H(X_1Y_1X_2Y_2) - \lambda_1 H(X_1X_2) - \lambda_2 H(Y_1Y_2)] \\
\leq C_{\mu_{X_1Y_1}}[H(X_1Y_1) - \lambda_1 H(X_1) - \lambda_2 H(Y_1)] + C_{\mu_{X_2Y_2}}[H(X_2Y_2) - \lambda_1 H(X_2) - \lambda_2 H(Y_2)]
\end{align*}
\]

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
  \hspace{1em} \star \hspace{1em} (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
Recap

Test the optimality of coding schemes in network information theory

- Resolved some open questions
- Many remain open

Computed the optimizers of several non-convex functionals
- Gaussian optimality via sub-additivity
- Optimal auxiliaries correspond to computation of concave envelopes
- Min-max theorem

More ideas and tools seem necessary

These (specific family) non-convex functionals also appear in other fields

The tools (already) developed can be used to get some new results
Recap

Test the optimality of coding schemes in network information theory
- Resolved some open questions
- Many remain open

Computed the optimizers of several non-convex functionals
- Developed some new tools and techniques
  - Gaussian optimality via sub-additivity
  - Optimal auxiliaries correspond to computation of concave envelopes
  - Min-max theorem
- More ideas and tools seem necessary
Recap

Test the optimality of coding schemes in network information theory

- Resolved some open questions
- Many remain open

Computed the optimizers of several non-convex functionals

- Developed some new tools and techniques
  - Gaussian optimality via sub-additivity
  - Optimal auxiliaries correspond to computation of concave envelopes
  - Min-max theorem
- More ideas and tools seem necessary

These (specific family) non-convex functionals also appear in other fields

- The tools (already) developed can be used to get some new results
Outline

- Broadcast channel: Establishing optimality of Marton’s for MIMO broadcast channel

- Interference channel: Sub-optimality of the Han–Kobayashi region

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

\[ C_{\mu_X} [\alpha_S H(X_S)]. \]

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 \( G_{12}^{\mu}(\gamma_1, \gamma_2) \) for all \( \gamma_1, \gamma_2 \), where

\[ G_{12}^{\mu}(\gamma_1, \gamma_2) := \sum_{S \subseteq [n]} \alpha_S H(X_{1S}, X_{2S}) - \mathbb{E}(\gamma_1(X_1)) - \mathbb{E}(\gamma_2(X_2)) \]
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 $\gamma_1, \gamma_2$, where

$$G^\mu_1(\gamma_1) := \sum_{S \subseteq [n]} \alpha_S H(X_1^S) - E(\gamma_1(X_1))$$

$$G^\mu_2(\gamma_2) := \sum_{S \subseteq [n]} \alpha_S H(X_2^S) - E(\gamma_2(X_2))$$

$$G^\mu_{12}(\gamma_1, \gamma_2) := \sum_{S \subseteq [n]} \alpha_S H(X_1^S, X_2^S) - E(\gamma_1(X_1)) - E(\gamma_2(X_2))$$

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.
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 $\gamma_1, \gamma_2$, where

$$G^\mu_1(\gamma_1) := \sum_{S \subseteq [n]} \alpha_S H(X_{1S}) - E(\gamma_1(X_1))$$

$$G^\mu_2(\gamma_2) := \sum_{S \subseteq [n]} \alpha_S H(X_{2S}) - E(\gamma_2(X_2))$$

$$G^\mu_{12}(\gamma_1, \gamma_2) := \sum_{S \subseteq [n]} \alpha_S H(X_{1S}, X_{2S}) - E(\gamma_1(X_1)) - E(\gamma_2(X_2))$$

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
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
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 (again) suggested by our computations in various examples.
Conjecture 2

Consider

\[ f_{\alpha}(\gamma) = \max_{\mu_X} \sum_{S \subseteq [n]} \alpha_S H(X_S) - E(\gamma(X)), \quad \alpha_S \in \mathbb{R}. \]

Suppose \( \alpha_S^{(0)} \) and \( \alpha_S^{(1)} \) have interior global maximizers.

Let \( \alpha_S^{(t)} = (1 - t)\alpha_S^{(0)} + t\alpha_S^{(1)}, \ t \in [0, 1] \). Then there exists a continuous path in the simplex such that \( \mu^{(t)} \) is a global maximizer of \( f_{\alpha^{(t)}}(\gamma) \) for all \( t \in [0, 1] \).
Conjecture 2

Consider

\[ f_\alpha(\gamma) = \max_{\mu_X} \sum_{S \subseteq [n]} \alpha_S H(X_S) - \mathbb{E}(\gamma(X)), \quad \alpha_S \in \mathbb{R}. \]

Suppose \( \alpha_S^{(0)} \) and \( \alpha_S^{(1)} \) have interior global maximizers.

Let \( \alpha_S^{(t)} = (1 - t)\alpha_S^{(0)} + t\alpha_S^{(1)}, \quad t \in [0, 1] \). Then there exists a continuous path in the simplex such that \( \mu^{(t)} \) is a global maximizer of \( f_{\alpha^{(t)}}(\gamma) \) for all \( t \in [0, 1] \).

Consequences:

- **Information theory:** Conjecture 2 (plus mild regularity conditions) implies the Conjecture that local tensorization implies global tensorization
- **Algorithms:** Suppose one wants to approximate hypercontractivity parameters
  - Start with \( p \to \infty \)
  - Approximate the maximizing distribution at this boundary value of norm.
  - Decrease \( p \) and track the global maximizers by local search.
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

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)
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)

Venkat Anantharam
Salman Beigi
Max Costa
Abbas El Gamal

Yanlin Geng
Amin Gohari
Varun Jog
Janos Korner

Sida Liu
David Ng
Vincent Wang
Yan Nan Wang

Lingxiao Xia
Babak Yazdanpanah
Acknowledgements (Rogues gallery)

Venkat Anantharam

Salman Beigi

Max Costa

Abbas El Gamal

Yanlin Geng

Amin Gohari

Varun Jog

Janos Korner

Sida Liu

David Ng

Vincent Wang

Yan Nan Wang

Lingxiao Xia

Babak Yazdanpanah