On Subadditivity, Global, and Local Optimizers of Information Functionals

Chandra Nair

The Chinese University of Hong Kong

14 March, 2022

Introduction

This talk is about **capacity regions** in multiuser settings

Introduction

This talk is about **capacity regions** in multiuser settings

Since capacity regions are usually convex subsets,
I will focus on *optimal* **weighted sum-rates**

Introduction

This talk is about **capacity regions** in multiuser settings

Since capacity regions are usually convex subsets,
I will focus on *optimal* **weighted sum-rates**

If I am very specific:
this talk is about *information functionals* whose maximum value yields the optimal weighted sum-rates

Introduction

This talk is about **capacity regions** in multiuser settings

Since capacity regions are usually convex subsets,
I will focus on *optimal* **weighted sum-rates**

If I am very specific:
this talk is about *information functionals* whose maximum value yields the optimal weighted sum-rates

To make it clear

For a point-to-point memoryless channel $W_{Y|X}$, we know that

$$C(W) := \max_{p_X} I(X;Y)$$

The talk is not about $C(W)$ but rather about

$$(W, p_X) \mapsto I(X;Y)$$

This talk is about **capacity regions** in multiuser settings

Since capacity regions are usually convex subsets,
I will focus on *optimal* **weighted sum-rates**

If I am very specific:
this talk is about *information functionals* whose maximum value yields the optimal weighted sum-rates

To make it clear

For a broadcast channel with degraded message sets $W_{Y_1,Y_2|X}$, we know that, for $\lambda \geq 1$

$$\begin{aligned}\max_{(R_0,R_1)\in\mathcal{C}(W)} R_1 + \lambda R_0 &= \max_{p_{V,X}} \min_{\alpha\in[0,1]} \{(\lambda-\alpha)I(V;Y_2) + (1-\alpha)I(X;Y_1|V) + \alpha I(X;Y_1)\} \\ &= \min_{\alpha\in[0,1]} \max_{p_X} (\lambda-\alpha)I(X;Y_2) + \alpha I(X;Y_1) \\ &\qquad + \mathsf{C}_{p_X}[(1-\alpha)I(X;Y_1) - (\lambda-\alpha)I(X;Y_2)]\end{aligned}$$

In this talk we are interested in functionals like

$$(W, p_X) \mapsto \mathsf{C}_{p_X}[(1-\alpha)I(X;Y_1) - (\lambda-\alpha)I(X;Y_2)]$$

Introduction

This talk is about **capacity regions** in multiuser settings

Since capacity regions are usually convex subsets,
I will focus on *optimal* **weighted sum-rates**

If I am very specific:
this talk is about *information functionals* whose maximum value yields the optimal weighted sum-rates

To make it clear

There is a class of interference channels $W_{Y_1,Y_2|X_1,X_2}$, for which

$$\max_{(R_0,R_1)\in\mathcal{C}(W)} R_1 + R_2 = \max_{p_{X_1}p_{X_2}} I(X_1;Y_1) + I(X_2;Y_2)$$

However we are not interested in

$$(W, p_{X_1}p_{X_2}) \mapsto I(X_1;Y_1) + I(X_2;Y_2)$$

This talk is about **capacity regions** in multiuser settings

Since capacity regions are usually convex subsets,

I will focus on *optimal* **weighted sum-rates**

If I am very specific:

this talk is about *information functionals* whose maximum value yields the optimal weighted sum-rates

To make it clear

There is a class of interference channels $W_{Y_1,Y_2|X_1,X_2}$, for which

$$\max_{(R_0,R_1)\in\mathcal{C}(W)} R_1 + R_2 = \max_{p_{X_1}p_{X_2}} I(X_1;Y_1) + I(X_2;Y_2)$$

However we are not interested in

$$(W, p_{X_1}p_{X_2}) \mapsto I(X_1;Y_1) + I(X_2;Y_2)$$

I will show one of the functionals of interest here later

Additivity and Sub-additivity

It is clear that capacity or weighted-sum-capacity, $\mathcal{C}(W)$, satisfies

$$\mathcal{C}(W \otimes W) = 2\,\mathrm{C}(W)$$

Additivity and Sub-additivity

It is clear that capacity or weighted-sum-capacity, $\mathcal{C}(W)$, satisfies

$$\mathcal{C}(W \otimes W) = 2\,\mathrm{C}(W)$$

Given a distribution p_X on the input (or inputs), once can also look at the capacity or weighted-sum-capacity, $\mathcal{C}(W, p_X)$, conditioned on codes having the distribution, p_X.

Additivity and Sub-additivity

It is clear that capacity or weighted-sum-capacity, $\mathcal{C}(W)$, satisfies

$$\mathcal{C}(W \otimes W) = 2\,\mathrm{C}(W)$$

Given a distribution p_X on the input (or inputs), once can also look at the capacity or weighted-sum-capacity, $\mathcal{C}(W, p_X)$, conditioned on codes having the distribution, p_X.

It is also clear that capacity or weighted-sum-capacity, $\mathcal{C}(W, p_X)$, satisfies

$$\mathcal{C}(W \otimes W, p_{X_1,X_2}) \leq 2\,\mathrm{C}(W, p_X)$$

where $p_X = \frac{1}{2}p_{X_1} + \frac{1}{2}p_{X_2}$.

Additivity and Sub-additivity

It is clear that capacity or weighted-sum-capacity, $\mathcal{C}(W)$, satisfies

$$\mathcal{C}(W \otimes W) = 2\,\mathrm{C}(W)$$

Given a distribution p_X on the input (or inputs), once can also look at the capacity or weighted-sum-capacity, $\mathcal{C}(W, p_X)$, conditioned on codes having the distribution, p_X.

It is also clear that capacity or weighted-sum-capacity, $\mathcal{C}(W, p_X)$, satisfies

$$\mathcal{C}(W \otimes W, p_{X_1,X_2}) \leq 2\,\mathrm{C}(W, p_X)$$

where $p_X = \frac{1}{2} p_{X_1} + \frac{1}{2} p_{X_2}$.

Motivated by this observation, let

$$\mathcal{I} = \{(W, p_X)\}$$

be the collection of pairs of channels and input distributions (p_X is consistent with W).

A function $F : \mathcal{I} \mapsto \mathbb{R}$ is called **sub-additive** if

$$F(W_1 \otimes W_2, p_{X_1,X_2}) \leq F(W_1, p_{X_1}) + F(W_2, p_{X_2})$$

Additivity and Sub-additivity

It is clear that capacity or weighted-sum-capacity, $\mathcal{C}(W)$, satisfies

$$\mathcal{C}(W \otimes W) = 2\,\mathrm{C}(W)$$

Given a distribution p_X on the input (or inputs), once can also look at the capacity or weighted-sum-capacity, $\mathcal{C}(W, p_X)$, conditioned on codes having the distribution, p_X.

It is also clear that capacity or weighted-sum-capacity, $\mathcal{C}(W, p_X)$, satisfies

$$\mathcal{C}(W \otimes W, p_{X_1,X_2}) \leq 2\,\mathrm{C}(W, p_X)$$

where $p_X = \frac{1}{2}p_{X_1} + \frac{1}{2}p_{X_2}$.

Motivated by this observation, let

$$\mathcal{I} = \{(W, p_X)\}$$

be the collection of pairs of channels and input distributions (p_X is consistent with W).

A function $F : \mathcal{I} \mapsto \mathbb{R}$ is called **concave sub-additive** if it is sub-additive and $F(W, p_X)$ is concave in p_X for a fixed W.

Sub-additivity and single-letterization

In many network information theory (channel coding) instances, we have multi-letter characterizations of weighted-sum-capacity

- Broadcast channel $W_{Y_1,Y_2|X}$, the weighted sum-capacity $\lambda R_1 + R_2$ for a distribution p_X, is given by

$$\mathcal{C}(W, p_X) = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n} \in \mathcal{M}_n} \sup_{p_U p_V} \lambda I(U; Y_{11}^n) + I(V; Y_{21}^n)$$

- Interference channel $W_{Y_1,Y_2|X_1,X_2}$, the weighted sum-capacity $\lambda R_1 + R_2$ for a distribution $p_{X_1}p_{X_2}$, is given by

$$\mathcal{C}(W, p_{X_1}p_{X_2}) = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X_{11}^n} p_{X_{21}^n} \in \mathcal{M}_n} \lambda I(X_{11}^n; Y_{11}^n) + I(X_{21}^n; Y_{21}^n)$$

- Relay channel $W_{Y,Y_{\mathrm{r}}|X,X_{\mathrm{r}}}$, the capacity R for a distribution p_X, is given by

$$\mathcal{C}(W, p_X) = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n} \in \mathcal{M}_n} I(X^n; Y^n).$$

Sub-additivity and single-letterization

In many network information theory (channel coding) instances, we have multi-letter characterizations of weighted-sum-capacity

Let $G(W^{\otimes n}, p_{X^n})$ be a multi-letter characterization of a weighted-sum-capacity, i.e.

$$\mathcal{C}(W, p_X) = \lim_n \frac{1}{n} \sup_{p_{X^n} \in \mathcal{M}_n} G(W^{\otimes n}, p_{X^n}).$$

If one could find a *concave sub-additive* function F that dominates pointwise, i.e.

$$G(W^{\otimes n}, p_{X^n}) \leq F(W^{\otimes n}, p_{X^n}) \ \forall \ W, p_{X^n}$$

then note that

$$\frac{1}{n} G(W^{\otimes n}, p_{X^n}) \leq \frac{1}{n} F(W^{\otimes n}, p_{X^n}) \overset{s.a.}{\leq} \frac{1}{n} \sum_{i=1}^{n} F(W, p_{X_i}) \overset{ccv}{\leq} F(W, p_X).$$

Sub-additivity and single-letterization

In many network information theory (channel coding) instances, we have multi-letter characterizations of weighted-sum-capacity

Let $G(W^{\otimes n}, p_{X^n})$ be a multi-letter characterization of a weighted-sum-capacity, i.e.

$$\mathcal{C}(W, p_X) = \lim_n \frac{1}{n} \sup_{p_{X^n} \in \mathcal{M}_n} G(W^{\otimes n}, p_{X^n}).$$

If one could find a *concave sub-additive* function F that dominates pointwise, i.e.

$$G(W^{\otimes n}, p_{X^n}) \leq F(W^{\otimes n}, p_{X^n}) \; \forall \; W, p_{X^n}$$

then note that

$$\frac{1}{n} G(W^{\otimes n}, p_{X^n}) \leq \frac{1}{n} F(W^{\otimes n}, p_{X^n}) \overset{s.a.}{\leq} \frac{1}{n} \sum_{i=1}^{n} F(W, p_{X_i}) \overset{ccv}{\leq} F(W, p_X).$$

Hence $F(W, p_X)$ will be a single-letter upper bound to the weighted sum capacity.

Two directions

Direction 1: Design new concave sub-additive functionals that serve as outer bounds to the capacity region (or weighted-sum-capacity)

Direction 2: Devise techniques to determine whether a given function $F : \mathcal{I} \mapsto \mathbb{R}$ is concave sub-additive

Two directions

Direction 1: Design new concave sub-additive functionals that serve as outer bounds to the capacity region (or weighted-sum-capacity)

- These have led to new outer bounds that have strictly improved on existing bounds
- Solved some open problems and new results regarding capacity regions

Direction 2: Devise techniques to determine whether a given function $F : \mathcal{I} \mapsto \mathbb{R}$ is concave sub-additive

Two directions

Direction 1: Design new concave sub-additive functionals that serve as outer bounds to the capacity region (or weighted-sum-capacity)

- These have led to new outer bounds that have strictly improved on existing bounds
- Solved some open problems and new results regarding capacity regions

Direction 2: Devise techniques to determine whether a given function $F : \mathcal{I} \mapsto \mathbb{R}$ is concave sub-additive

- Main aim: to test optimality of existing inner bounds
- Solved some of these optimality questions
- Some intriguing observations *(unpublished work)*

Two directions

Direction 1: Design new concave sub-additive functionals that serve as outer bounds to the capacity region (or weighted-sum-capacity)

- These have led to new outer bounds that have strictly improved on existing bounds
- Solved some open problems and new results regarding capacity regions

Direction 2: Devise techniques to determine whether a given function $F : \mathcal{I} \mapsto \mathbb{R}$ is concave sub-additive

- Main aim: to test optimality of existing inner bounds
- Solved some of these optimality questions
- Some intriguing observations *(unpublished work)*

Common: Understanding global and local optimizers of certain non-convex functionals

Direction 1: Starting points of single-letterization

Traditionally, one starts from expressions like

- $I(M;Y^n)$ or $I(X^n;Y^n)$: for point-to-point channel
- $\alpha I(M_0,M_1;Y_{11}^n)+(1-\alpha)I(M_1;Y_{11}^n|M_0)+(\lambda-1)I(M_0;Y_{21}^n)$: for broadcast channel with degraded message sets

Traditionally, one starts from expressions like

- $I(M;Y^n)$ or $I(X^n;Y^n)$: for point-to-point channel
- $\alpha I(M_0,M_1;Y_{11}^n) + (1-\alpha)I(M_1;Y_{11}^n|M_0) + (\lambda-1)I(M_0;Y_{21}^n)$: for broadcast channel with degraded message sets

Rather recent: Genie-aided starting points (interference channel)

- $\lambda I(M_1;Y_{11}^n,T_1^n) + I(M_2;Y_{21}^n,T_2^n)$

Here T_1^n and T_2^n are the genie sequences.

Rather surprisingly, there are concave sub-additive F from this starting point (for non-trivial T_1^n and T_2^n) that turn out to be **equal to capacity** for some non-trivial interference channels.

Direction 1: Starting points of single-letterization

Traditionally, one starts from expressions like

- $I(M;Y^n)$ or $I(X^n;Y^n)$: for point-to-point channel
- $\alpha I(M_0,M_1;Y_{11}^n) + (1-\alpha)I(M_1;Y_{11}^n|M_0) + (\lambda-1)I(M_0;Y_{21}^n)$: for broadcast channel with degraded message sets

Rather recent: Genie-aided starting points (interference channel)

- $\lambda I(M_1;Y_{11}^n,T_1^n) + I(M_2;Y_{21}^n,T_2^n)$

Here T_1^n and T_2^n are the genie sequences.

Rather surprisingly, there are concave sub-additive F from this starting point (for non-trivial T_1^n and T_2^n) that turn out to be **equal to capacity** for some non-trivial interference channels.

Very recent: Auxiliary receiver starting points (broadcast, interference, and relay channel)

- $I(M;J^n) + \big(I(M;Y^n) - I(M;J^n)\big)$

Here J^n is an auxiliary receiver sequence

The main tools that we have

Chain-rule

$$H(Y^n|U) = \sum_{i=1}^{n} H(Y_i|U, Y^{i-1}) = \sum_{i=1}^{n} H(Y_i|U_i)$$

where $U_i = (U, Y^{i-1})$ (Gallager '74)

Cśiszar-Körner-Marton identity

$$\begin{aligned} H(Y^n|U) - H(J^n|U) &= \sum_{i=1}^{n} \Big(H(Y_i|U, Y^{i-1}, J_{i+1}^n) - H(J_i|U, Y^{i-1}, J_{i+1}^n) \Big) \\ &= \sum_{i=1}^{n} \Big(H(Y_i|U_i) - H(J_i|U_i) \Big) \end{aligned}$$

where $U_i = (U, Y^{i-1}, J_{i+1}^n)$ (Körner-Marton '77)

The main tools that we have

Chain-rule

$$H(Y^n|U) = \sum_{i=1}^{n} H(Y_i|U, Y^{i-1}) = \sum_{i=1}^{n} H(Y_i|U_i)$$

where $U_i = (U, Y^{i-1})$ (Gallager '74)

Cśiszar-Körner-Marton identity

$$H(Y^n|U) - H(J^n|U) = \sum_{i=1}^{n} \Big(H(Y_i|U, Y^{i-1}, J_{i+1}^n) - H(J_i|U, Y^{i-1}, J_{i+1}^n) \Big)$$
$$= \sum_{i=1}^{n} \Big(H(Y_i|U_i) - H(J_i|U_i) \Big)$$

where $U_i = (U, Y^{i-1}, J_{i+1}^n)$ (Körner-Marton '77)

Potential sub-optimality

- Data-processing inequality (and drop some terms)
- Identify auxiliaries in terms of distributions induced by good codes, but then take union over all possible distributions

Traditional outer bounds such as

- Korner-Marton or UVW outer bound for broadcast channel
- Cutset bound for the relay channel

have been well studied and now even *automated*
(talk (Wed) by Cheuk Ting Li)

On Traditional and Genie-based Outer bounds

Genie-based outer bounds were popularized by the Gaussian interference channel

Here is a *family of* outer bound functionals for *sum-rate* of an interference channel using the *genie* approach

$$
\begin{aligned}
I(X_1;T_1,Y_1|S_2) &+ \lambda I(X_2;T_2,Y_2|S_1) \\
&+ \mathsf{C}_{p_{X_1}p_{X_2}}[I(X_1;T_1|X_2,T_2,S_1) - \lambda I(X_1;Y_2|X_2,T_2,S_1)] \\
&- I(X_1;T_1|X_2,T_2,S_1) + \lambda I(X_1;Y_2|X_2,T_2,S_1) \\
&+ \mathsf{C}_{p_{X_1}p_{X_2}}[I(X_2;T_2|X_1,T_1,S_2) - I(X_2;Y_1|X_1,T_1,S_2)] \\
&- I(X_2;T_2|X_1,T_1,S_2) + I(X_2;Y_1|X_1,T_1,S_2)
\end{aligned}
$$

Here

$$p(y_1,t_1,s_1,y_2,t_2,s_2|x_1,x_2) = p(t_1,s_1|x_1)p(t_2,s_2|x_2)p(y_1,y_2|t_1,t_2,s_1,s_2,x_1,x_2).$$

Further we require that

- $p(y_1|x_1,x_2) = w(y_1|x_1,x_2)$ and $p(y_2|x_1,x_2) = w(y_2|x_1,x_2)$.
- for each $i = 1,2$, T_i, S_i has degraded order, *i.e.* either $X_i \to T_i \to S_i$ or $X_i \to S_i \to T_i$ holds.

On Traditional and Genie-based Outer bounds

Genie-based outer bounds were popularized by the Gaussian interference channel

Here is a *family of* outer bound functionals for *sum-rate* of an interference channel using the *genie* approach

$$
\begin{aligned}
&I(X_1;T_1,Y_1|S_2) + \lambda I(X_2;T_2,Y_2|S_1) \\
&\quad + \mathsf{C}_{p_{X_1}p_{X_2}}[I(X_1;T_1|X_2,T_2,S_1) - \lambda I(X_1;Y_2|X_2,T_2,S_1)] \\
&\quad - I(X_1;T_1|X_2,T_2,S_1) + \lambda I(X_1;Y_2|X_2,T_2,S_1) \\
&\quad + \mathsf{C}_{p_{X_1}p_{X_2}}[I(X_2;T_2|X_1,T_1,S_2) - I(X_2;Y_1|X_1,T_1,S_2)] \\
&\quad - I(X_2;T_2|X_1,T_1,S_2) + I(X_2;Y_1|X_1,T_1,S_2)
\end{aligned}
$$

Why is this interesting?

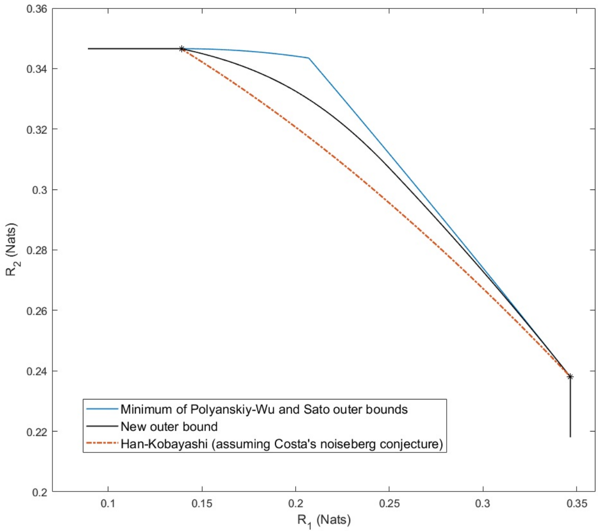
- Only known way to obtain sum-capacity of a certain class of discrete memoryless interference channels (by choosing a suitable genie)
- Recovers many of the known capacity results in the Gaussian interference channel

Sida Liu

Auxiliary receiver based outer bounds [GN22]

Disclaimer: The functionals are less exciting than some of their implications

Gaussian Z-Interference channel



Amin Gohari

Illustration for $a = 0.8$, $P_1 = P_2 = 1$.

Disclaimer: The functionals are less exciting than some of their implications

Gaussian Z-Interference channel

Amin Gohari

Theorem

Let $C_2 = \frac{1}{2}\log(1+P_2)$ *and* $R_1^* = \frac{1}{2}\log\left(1+\frac{a^2 P_1}{1+P_2}\right)$. *Then*

$$\max_{(R_1,R_2)\in\mathcal{R}_{OB}} \lambda R_2 + R_1 = \lambda C_2 + R_1^*$$

when

$$\lambda \geq 1 + \begin{cases} \frac{(1+P_2)(1-a^2)}{a^2 P_2}\frac{\left(1+\sqrt{1+4a^2(1-a^2)P_2}\right)^2}{4a^2(1-a^2)P_2} & a^2 < \frac{1}{2} \\ \frac{(1+P_2)(1-a^2)}{a^2 P_2}\frac{\left(1+\sqrt{1+P_2}\right)^2}{P_2} & a^2 \geq \frac{1}{2} \end{cases}.$$

Auxiliary receiver based outer bounds [GN22]

Disclaimer: The functionals are less exciting than some of their implications

Gaussian Z-Interference channel

Amin Gohari

Theorem

Let $C_2 = \frac{1}{2}\log(1+P_2)$ *and* $R_1^* = \frac{1}{2}\log\left(1+\frac{a^2P_1}{1+P_2}\right)$. *Then*

$$\max_{(R_1,R_2)\in\mathcal{R}_{OB}} \lambda R_2 + R_1 = \lambda C_2 + R_1^*$$

when

$$\lambda \geq 1 + \begin{cases} \frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\left(1+\sqrt{1+4a^2(1-a^2)P_2}\right)^2}{4a^2(1-a^2)P_2} & a^2 < \frac{1}{2} \\ \frac{(1+P_2)(1-a^2)}{a^2P_2}\frac{\left(1+\sqrt{1+P_2}\right)^2}{P_2} & a^2 \geq \frac{1}{2} \end{cases}.$$

Remark: This kind of behavior at the *Costa-Polyanskiy-Wu* corner point does not follow from Talagrand's HWI inequality.

Auxiliary receiver based outer bounds [GN22]

Deterministic Broadcast channel

Korner-Marton Outer bound yields the capacity region

$$R_1 \leq H(Y_1),$$
$$R_2 \leq H(Y_2),$$
$$R_1 + R_2 \leq H(Y_1, Y_2).$$

Amin Gohari

Auxiliary receiver based outer bounds [GN22]

Deterministic Broadcast channel

Korner-Marton Outer bound yields the capacity region

Amin Gohari

$$R_1 \leq H(Y_1),$$
$$R_2 \leq H(Y_2),$$
$$R_1 + R_2 \leq H(Y_1, Y_2).$$

Erased Deterministic Broadcast channel (pass each output through a $BEC(\epsilon)$): Korner-Marton Outer bound (or UVW outer bound) yields

$$R_1 \leq (1-\epsilon)H(Y_1),$$
$$R_2 \leq (1-\epsilon)H(Y_2),$$
$$R_1 + R_2 \leq (1-\epsilon)H(Y_1, Y_2).$$

Auxiliary receiver based outer bounds [GN22]

Deterministic Broadcast channel

Korner-Marton Outer bound yields the capacity region

$$R_1 \leq H(Y_1),$$
$$R_2 \leq H(Y_2),$$
$$R_1 + R_2 \leq H(Y_1, Y_2).$$

Amin Gohari

Erased Deterministic Broadcast channel (pass each output through a $BEC(\epsilon)$): Korner-Marton Outer bound (or UVW outer bound) yields

$$R_1 \leq (1-\epsilon)H(Y_1),$$
$$R_2 \leq (1-\epsilon)H(Y_2),$$
$$R_1 + R_2 \leq (1-\epsilon)H(Y_1, Y_2).$$

Auxiliary Receiver based outer bounds show that capacity region can be strictly inside this region

- Erased Blackwell broadcast channel (strict inclusion)

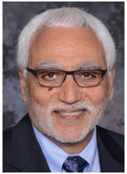
Auxiliary receiver based outer bounds [GN22]

Relay Channel (also including results from an about to be published paper)

Just by setting $J = Y_{\mathrm{r}}$, we obtain an outer bound that

- Strictly improves on cutset outer bound for the scalar Gaussian relay channel for **all non-zero channel gains**
- Solves Kim's conjecture (for deterministic relay channels)
- Solves Cover's open problem for a large class of channels (complement lies in a smaller dimensional space)
- Recovers the outer bound in the Gaussian case obtained using spherical rearrangement (Wu et. al.)
- Strictly improves on the outer bound for the BSC case (Wu et. al., Barnes e. al.)

Amin Gohari



El Gamal

Auxiliary receiver based outer bounds [GN22]

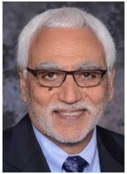
Relay Channel (also including results from an about to be published paper)

Just by setting $J = Y_{\mathrm{r}}$, we obtain an outer bound that

- Strictly improves on cutset outer bound for the scalar Gaussian relay channel for **all non-zero channel gains**
- Solves Kim's conjecture (for deterministic relay channels)
- Solves Cover's open problem for a large class of channels (complement lies in a smaller dimensional space)
- Recovers the outer bound in the Gaussian case obtained using spherical rearrangement (Wu et. al.)
- Strictly improves on the outer bound for the BSC case (Wu et. al., Barnes e. al.)

We also show that this bound can be improved by using other auxiliary receivers

Amin Gohari



El Gamal

Auxiliary receiver based outer bounds [GN22]

The main concerns of the auxiliary receiver approach

- Every choice of auxiliary receivers yields an outer bound
- One can incorporate many auxiliary receivers into a single bound
- But how do we know which auxiliary receivers lead to strict improvements

Amin Gohari

Auxiliary receiver based outer bounds [GN22]

The main concerns of the auxiliary receiver approach

- Every choice of auxiliary receivers yields an outer bound
- One can incorporate many auxiliary receivers into a single bound
- But how do we know which auxiliary receivers lead to strict improvements

Amin Gohari

One of the main contributions

- Show that there exists auxiliary receivers that lead to strict improvements

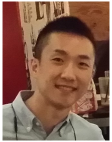
But this is something that needs lot more investigation

Direction 2: Testing for sub-additivity

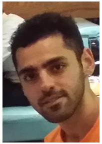
This is motivated by inner bounds whose multi-letter expressions tend to capacity

Finding counterexamples to the optimality of certain achievable regions were motivated by the ideas (which I am going to present)

- Han-Kobayashi region for the interference channel
- Superposition coding for a three receiver broadcast channel with degraded message sets



Lingxiao



Babak

This is motivated by inner bounds whose multi-letter expressions tend to capacity

Finding counterexamples to the optimality of certain achievable regions were motivated by the ideas (which I am going to present)

- Han-Kobayashi region for the interference channel
- Superposition coding for a three receiver broadcast channel with degraded message sets

However in some other settings we could not find counterexamples

For instance: Marton's inner bound for the two-receiver broadcast channel

Direction 2: Testing for sub-additivity

This is motivated by inner bounds whose multi-letter expressions tend to capacity

Finding counterexamples to the optimality of certain achievable regions were motivated by the ideas (which I am going to present)

- Han-Kobayashi region for the interference channel
- Superposition coding for a three receiver broadcast channel with degraded message sets

However in some other settings we could not find counterexamples

For instance: Marton's inner bound for the two-receiver broadcast channel

A functional $F(W, p_X)$ of interest here (Marton's region) is

$$\mathsf{C}_{p_X}\left[-(\lambda-\alpha)H(Y_1)-\alpha H(Y_2)+\max_{p_{UV|X}}\left\{\lambda I(U;Y_1)+I(V;Y_2)-I(U;V)\right\}\right]$$

An Observation

We wish to test the sub-additivity of $F(W, p_X)$ of the form

$$\mathsf{C}_{p_X}\left[\sup_{p_{U|X}}\left\{\sum_S \alpha_S H(Y_S)\right\}\right], \text{ for } \alpha_s \in \mathbb{R}.$$

By Fenchel duality

$$\mathsf{C}_{p_X}\left[\max_{p_{U|X}}\left\{\sum_S \alpha_S H(Y_S)\right\}\right] = \inf_{\gamma(X)}\left\{\hat{F}^{(\gamma)}(W) + \mathrm{E}_{p_X}(\gamma(X))\right\}$$

where

$$\hat{F}^{(\gamma)}(W) = \sup_{q_{U,X}}\left\{\sum_S \alpha_S H_q(Y_S) - \mathrm{E}_{q_X}(\gamma(X))\right\}$$

Lemma

The sub-additivity of $F(W, p_X)$ is equivalent to requiring that for every $\gamma_1(X_1)$ and $\gamma_2(X_2)$ a **product distribution** $q_{U_1,X_1}q_{U_2,X_2}$ maximizes

$$\sum_S \alpha_S H_{q_{X_1X_2}}(Y_{1S}, Y_{2S}) - \mathrm{E}_{q_{X_1}}(\gamma_1(X_1)) - \mathrm{E}_{q_{X_2}}(\gamma_2(X_2))$$

Global and Local tensorization

Definition

A functional $\sum_S \alpha_S H(Y_S)$ is said to satisfy **global tensorization** if a `product distribution maximizes` $\hat{F}^{(\gamma_1,\gamma_2)}(W_1 \otimes W_2, p_{UX_1X_2})$ for all $\gamma_1(X_1), \gamma_2(X_2)$, where

$$\hat{F}^{(\gamma_1,\gamma_2)}(W_1 \otimes W_2, p_{UX_1X_2}) := \sum_S \alpha_S H(Y_{1S}, Y_{2S}) - \mathrm{E}(\gamma_1(X_1)) - \mathrm{E}(\gamma_2(X_2))$$

Global and Local tensorization

Definition

A functional $\sum_S \alpha_S H(Y_S)$ is said to satisfy **global tensorization** if a `product distribution maximizes` $\hat{F}^{(\gamma_1,\gamma_2)}(W_1 \otimes W_2, p_{UX_1X_2})$ for all $\gamma_1(X_1), \gamma_2(X_2)$, where

$$\hat{F}^{(\gamma_1,\gamma_2)}(W_1 \otimes W_2, p_{UX_1X_2}) := \sum_S \alpha_S H(Y_{1S}, Y_{2S}) - \mathrm{E}(\gamma_1(X_1)) - \mathrm{E}(\gamma_2(X_2))$$

Definition

A functional $\sum_S \alpha_S H(Y_S)$ is said to satisfy **local tensorization** if `the product of local maximizers` of $\hat{F}_1^{(\gamma_1)}(W_1, p_{UX_1})$ and $\hat{F}_2^{(\gamma_2)}(W_2, p_{UX_2})$ `is a local maximizer` of $\hat{F}_{12}^{(\gamma_1,\gamma_2)}(W_1 \otimes W_2, p_{UX_1X_2})$ for all $\gamma_1(X_1), \gamma_2(X_2)$, where

$$\hat{F}_1^{(\gamma_1)}(W_1, p_{UX_1}) := \sum_S \alpha_S H(Y_{1S}) - \mathrm{E}(\gamma_1(X_1))$$

$$\hat{F}_2^{(\gamma_2)}(W_2, p_{UX_2}) := \sum_S \alpha_S H(X_{2S}) - \mathrm{E}(\gamma_2(X_2))$$

$$\hat{F}_{12}^{(\gamma_1,\gamma_2)}(W_1 \otimes W_2, p_{UX_1X_2}) := \sum_S \alpha_S H(X_{1S}, X_{2S}) - \mathrm{E}(\gamma_1(X_1)) - \mathrm{E}(\gamma_2(X_2))$$

Observations

For all the functionals for which we can prove global tensorization:

we can *mechanically* convert our proofs to prove local tensorization

- First derivative conditions are rather immediate
- Second derivative conditions:

$$H(Y_s) \mapsto \mathrm{E}(\mathrm{Var}(f(X)|Y_S))$$

◇ For independent distributions, we can see that the key entropic equalities (and inequalities) have a correspondence with conditional variance equalities (and inequalities)

Observations

For all the functionals for which we can prove global tensorization:

we can *mechanically* convert our proofs to prove local tensorization

- First derivative conditions are rather immediate
- Second derivative conditions:

$$H(Y_s) \mapsto \mathrm{E}(\mathrm{Var}(f(X)|Y_S))$$

  - ◇ For independent distributions, we can see that the key entropic equalities (and inequalities) have a correspondence with conditional variance equalities (and inequalities)

For all the functionals for which we know that local tensorization fails

we can show that global tensorization also fails

This was how we

- guessed that Han–Kobayashi region and others may be sub-optimal
- identified counterexamples

A working hypothesis

Conjecture (broadly speaking)

Local tensorization **implies** Global tensorization

A working hypothesis

Conjecture (broadly speaking)

Local tensorization **implies** Global tensorization

Can show local tensorization for the functional (in Marton's inner bound) when one of the channels W_1 or W_2 has binary inputs

- Proof utilizes the knowledge of local maximizers in binary alphabet case
- If the conjecture is true, then we can show that Marton's inner bound matches the capacity region for binary input broadcast channels
- Whether local tensorization holds when both are non-binary is open
  - ◇ We do not have a good understanding of the local maximizers in the non-binary case

More importantly, it provides a new way to test if a functional is sub-additive or equivalently a **product distribution** $q_{X_1}q_{X_2}$ maximizes

$$\sum_S \alpha_S H_{q_{X_1X_2}}(Y_{1S}, Y_{2S}) - \mathrm{E}_{q_{X_1}}(\gamma_1(X_1)) - \mathrm{E}_{q_{X_2}}(\gamma_2(X_2))$$

Recap

Consider functionals of the form

$$\hat{F}^{(W,\gamma)}(p_{UX}) := \sum_S \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$

The quantities that we are interested in

- Local and global maximizers of $\hat{F}^{(W,\gamma)}(p_{UX})$
- Local and global maximizers of $\hat{F}^{(W_1 \otimes W_2, \gamma_1(X_1)+\gamma_2(X_2))}(p_{UX_1X_2})$

Recap

Consider functionals of the form

$$\hat{F}^{(W,\gamma)}(p_{UX}) := \sum_S \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$

The quantities that we are interested in

- Local and global maximizers of $\hat{F}^{(W,\gamma)}(p_{UX})$
- Local and global maximizers of $\hat{F}^{(W_1 \otimes W_2, \gamma_1(X_1)+\gamma_2(X_2))}(p_{UX_1X_2})$

Recap

Note that

- $F(W, p_X) := \mathsf{C}_{p_X}[\sum_S \alpha_S H(Y_S)]$ being (concave) sub-additive is directly related to single-letter outer bounds or capacity regions
- Testing sub-additivity is equivalent to requiring that $\hat{F}^{(W_1 \otimes W_2, \gamma_1(X_1)+\gamma_2(X_2))}(p_{UX_1X_2})$ has a product global maximizer

Recap

Consider functionals of the form

$$\hat{F}^{(W,\gamma)}(p_{UX}) := \sum_S \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$

The quantities that we are interested in

- Local and global maximizers of $\hat{F}^{(W,\gamma)}(p_{UX})$
- Local and global maximizers of $\hat{F}^{(W_1 \otimes W_2, \gamma_1(X_1)+\gamma_2(X_2))}(p_{UX_1X_2})$

Recap

Note that

- $F(W, p_X) := \mathsf{C}_{p_X}[\sum_S \alpha_S H(Y_S)]$ being (concave) sub-additive is directly related to single-letter outer bounds or capacity regions
- Testing sub-additivity is equivalent to requiring that $\hat{F}^{(W_1 \otimes W_2, \gamma_1(X_1)+\gamma_2(X_2))}(p_{UX_1X_2})$ has a product global maximizer
  - ◇ However, these are usually non-convex functions of p_{UX}
- From empirical observations it appears that a certain local property (product of local maximizers is a local maximizer) would imply the global maximizer property

On Local Maximizers

The conjecture that local tensorization implies global tensorization is true if

- For every W (or within a class that we are interested in, that is closed under $\otimes$) and γ,

$$\hat{F}^{(W,\gamma)}(p_{U,X}) = \sum_S \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$

  has a **unique** local maximizer (hence global maximizer)

On Local Maximizers

The conjecture that local tensorization implies global tensorization is true if

- For every W (or within a class that we are interested in, that is closed under $\otimes$) and γ,

$$\hat{F}^{(W,\gamma)}(p_{U,X}) = \sum_S \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$

  has a **unique** local maximizer (hence global maximizer)
  - ◇ This is trivially true when all α_S are non-negative, since the function is concave in p_X
  - ◇ Surprisingly, there are other non-trivial instances where this is true
  - ◇ In particular, when Gaussian optimality follows using rotation, sub-additivity arguments (for instance, MIMO Broadcast), you "automatically" get a proof of uniqueness of local maximizers in the space of Gaussian input distributions
    ("observation" made just four months ago)

On Local Maximizers

The conjecture that local tensorization implies global tensorization is true if

- For every W (or within a class that we are interested in, that is closed under $\otimes$) and γ,

$$\hat{F}^{(W,\gamma)}(p_{U,X}) = \sum_{S} \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$

  has a **unique** local maximizer (hence global maximizer)

  - ◇ This is trivially true when all α_S are non-negative, since the function is concave in p_X
  - ◇ Surprisingly, there are other non-trivial instances where this is true
  - ◇ In particular, when Gaussian optimality follows using rotation, sub-additivity arguments (for instance, MIMO Broadcast), you "automatically" get a proof of uniqueness of local maximizers in the space of Gaussian input distributions ("observation" made just four months ago)
  - ◇ For instance, say $\lambda > 1$, the following function of p.s.d matrices, in the domain $0 \preceq K \preceq K_0$, where $\Sigma_1, \Sigma_2 \succ 0$:

$$\log |K + \Sigma_1| - \lambda \log |K + \Sigma_2|$$

    has a unique local maximizer. The arguments are completely

On Local Maximizers

The conjecture that local tensorization implies global tensorization is true if

- For every W (or within a class that we are interested in, that is closed under $\otimes$) and γ,
$$\hat{F}^{(W,\gamma)}(p_{U,X}) = \sum_{S} \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$
has a **unique** local maximizer (hence global maximizer)

- However there are non-trivial discrete memoryless broadcast channels, for which
$$p_X \mapsto H(Y_1) - \lambda H(Y_2) - \mathrm{E}(\gamma(X))$$
has multiple local maximizers, for $\lambda > 1$
  - ◇ This is an instance for which we know that both local and global tensorizations hold

On Local Maximizers

The conjecture that local tensorization implies global tensorization is true if

- For every W (or within a class that we are interested in, that is closed under $\otimes$) and γ,

$$\hat{F}^{(W,\gamma)}(p_{U,X}) = \sum_S \alpha_S H(Y_S) - \mathrm{E}(\gamma(X))$$

  has a **unique** local maximizer (hence global maximizer)

- However there are non-trivial discrete memoryless broadcast channels, for which

$$p_X \mapsto H(Y_1) - \lambda H(Y_2) - \mathrm{E}(\gamma(X))$$

  has multiple local maximizers, for $\lambda > 1$

  - This is an instance for which we know that both local and global tensorizations hold

- Relaxed Guess: For every $W_1 \otimes W_2, \gamma_1, \gamma_2$, all local maximizers of

$$\sum_S \alpha_S H(Y_{1S}, Y_{2S}) - \mathrm{E}(\gamma(X_1) + \gamma(X_2))$$

  are product distributions

Remarks

For local-maximizers with full support we can show that they have to be product form for some functionals

- This is an interesting calculation
- But we cannot deal with "boundary" points yet

- Difference between additive Gaussian noise settings and discrete settings
- Perhaps we should view these functionals in a larger space (for instance using auxiliary receivers)

Remarks

For local-maximizers with full support we can show that they have to be product form for some functionals

- This is an interesting calculation
- But we cannot deal with "boundary" points yet

- Difference between additive Gaussian noise settings and discrete settings
- Perhaps we should view these functionals in a larger space (for instance using auxiliary receivers)

Acknowledgements

These ideas and intuition (and their refinements) is really based on many years of collective discussions with

Venkat Anantharam, Amin Gohari, Dustin Wang, Ken Lau, and others

As with any scientific method, the conjectures may need further refinement based on more empirical data

Remarks

For local-maximizers with full support we can show that they have to be product form for some functionals

- This is an interesting calculation
- But we cannot deal with "boundary" points yet

- Difference between additive Gaussian noise settings and discrete settings
- Perhaps we should view these functionals in a larger space (for instance using auxiliary receivers)

Thank You