# On *extremal* auxiliaries in network information theory

Chandra Nair

Dept. of Information Engineering
The Chinese University of Hong Kong

Mar 30, 2015

## The mathematics of digital communication [Shannon '48]

A sender $X$ communicates to receiver $Y$ over a noisy channel $\mathfrak{q}(y|x)$.

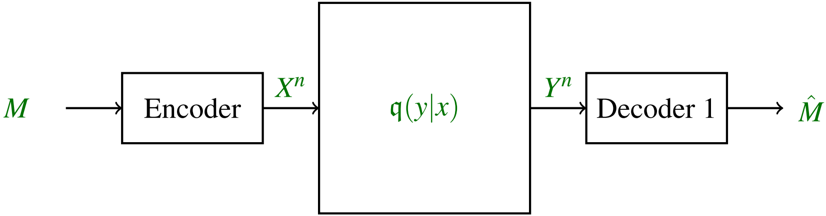



Figure: Discrete memoryless channel

The maximum rate that can be reliably transmitted (using blocks)

$$C = \max_{p(x)} I(X;Y).$$

# EXTENSION TO NETWORKS

What if there are more than one sender/receiver?

Can we obtain a similar *capacity region*?

# EXTENSION TO NETWORKS

What if there are more than one sender/receiver?

Can we obtain a similar *capacity region*?

The answer is mostly *NO*, i.e. we do not know the capacity regions.

- NOTABLE EXCEPTION: Multiple access channel

# Open setting 1: Broadcast channels [Cover '72]

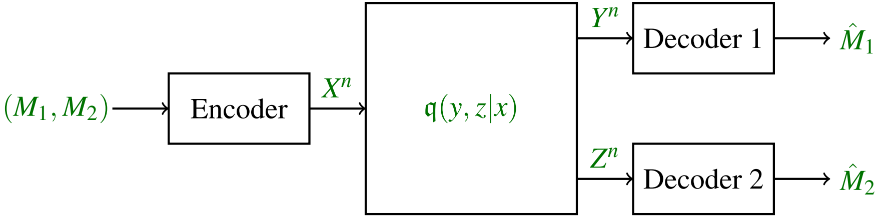



Figure: Discrete memoryless broadcast channel

- Goal: Compute *Capacity Region* or set of achievable rates $(R_1, R_2)$?

# Open setting 2: Interference channels

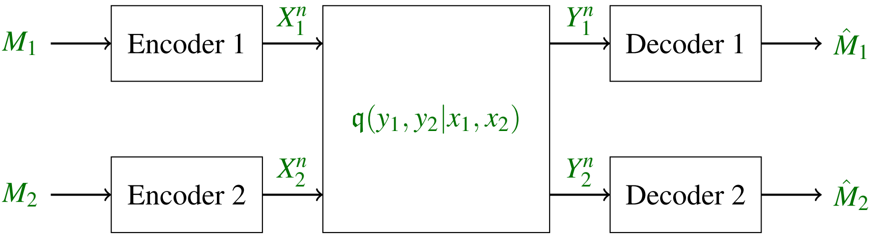



Figure: Discrete memoryless interference channel

Goal: compute *Capacity Region* or set of achievable rates $(R_1, R_2)$?

# An observation (folk-lore)

For these two problems

- there are achievable regions (one for each) whose optimality or sub-optimality had not been established for over 30 years !
- for both these regions, there is a way to test the optimality or sub-optimality

# An observation (folk-lore)

For these two problems

- there are achievable regions (one for each) whose optimality or sub-optimality had not been established for over 30 years !
- for both these regions, there is a way to test the optimality or sub-optimality
- the testing procedure: infinite computational resources
  - if suboptimal, the procedure terminates in finite time

# An observation (folk-lore)

For these two problems

- there are achievable regions (one for each) whose optimality or sub-optimality had not been established for over 30 years !
- for both these regions, there is a way to test the optimality or sub-optimality
- the testing procedure: infinite computational resources
  - if suboptimal, the procedure terminates in finite time

Testing strategy: Suppose some one gives you an achievable strategy

- for any channel $\mathfrak{q}$, it yields a computable region $\mathcal{A}(\mathfrak{q})$
- as $n \to \infty$, the normalized region $\frac{1}{n}\mathcal{A}(\underbrace{\mathfrak{q} \otimes \cdots \otimes \mathfrak{q}}_{n}) \to \mathcal{C}$

# An observation (folk-lore)

For these two problems

- there are achievable regions (one for each) whose optimality or sub-optimality had not been established for over 30 years !
- for both these regions, there is a way to test the optimality or sub-optimality
- the testing procedure: infinite computational resources
  - if suboptimal, the procedure terminates in finite time

Testing strategy: Suppose some one gives you an achievable strategy

- for any channel $\mathfrak{q}$, it yields a computable region $\mathcal{A}(\mathfrak{q})$
- as $n \to \infty$, the normalized region $\frac{1}{n}\mathcal{A}(\underbrace{\mathfrak{q} \otimes \cdots \otimes \mathfrak{q}}_{n}) \to \mathcal{C}$

then it is enough to test whether

$$\mathcal{A}(\mathfrak{q}) = \frac{1}{2}\mathcal{A}(\mathfrak{q} \otimes \mathfrak{q}) \quad \forall \mathfrak{q} \qquad \text{(optimal)}$$

$$\mathcal{A}(\mathfrak{q}) \subsetneq \frac{1}{2}\mathcal{A}(\mathfrak{q} \otimes \mathfrak{q}) \quad \text{for some } \mathfrak{q} \qquad \text{(sub-optimal)}$$

# Marton's region (Broadcast)

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

$$R_1 \leq I(U,W;Y)$$
$$R_2 \leq I(V,W;Z)$$
$$R_1 + R_2 \leq \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$$

for any $(U,V,W) \to X \stackrel{\mathfrak{q}}{\to} (Y,Z)$ is achievable

Remarks:

- An interesting (and natural generalization) of a strategy for deterministic broadcast channels [Marton '79]
- No reason to believe that it may be optimal or its optimality was worth investigating
- Even for a single channel $\mathfrak{q}(y,z|x)$ there were no bounds on $|U|$ or $|V|$, which made the region incomputable

▸ Marton

# Han and Kobayashi's region (Interference)

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

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

for some pmf $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$, where $|U_1| \leq |X_1| + 4$, $|U_2| \leq |X_2| + 4$, and $|Q| \leq 7$.

- Seems complicated to evaluate and use the 1-letter vs 2-letter strategy for testing optimality

▸ CZI

# Summary of talk: On evaluation of regions

## Statutory Disclaimer

Know more about evaluation of Marton's region than that of Han-Kobayashi

## Han-Kobayashi region

Main: Strict sub-optimality of the Han-Kobayashi region

- Restrict to a class of channels where evaluation is easy
- Show that 2-letter (dependence over time) beats 1-letter (independent over time)

## Marton's region

- Cardinality bounds for evaluation of Marton's region for broadcast channel
- Evaluation of Marton's region for **any** *binary* input broadcast channel
- Other results that helps evaluate Marton's region for broadcast channels

# Summary of talk: On evaluation of regions

## Statutory Disclaimer

Know more about evaluation of Marton's region than that of Han-Kobayashi

## Han-Kobayashi region

Main: Strict sub-optimality of the Han-Kobayashi region

- Restrict to a class of channels where evaluation is easy
- Show that 2-letter (dependence over time) beats 1-letter (independent over time)

## Marton's region

- Cardinality bounds for evaluation of Marton's region for broadcast channel
- Evaluation of Marton's region for any *binary* input broadcast channel
- Other results that helps evaluate Marton's region for broadcast channels

# Summary of talk: On evaluation of regions

## Statutory Disclaimer

Know more about evaluation of Marton's region than that of Han-Kobayashi

## Han-Kobayashi region

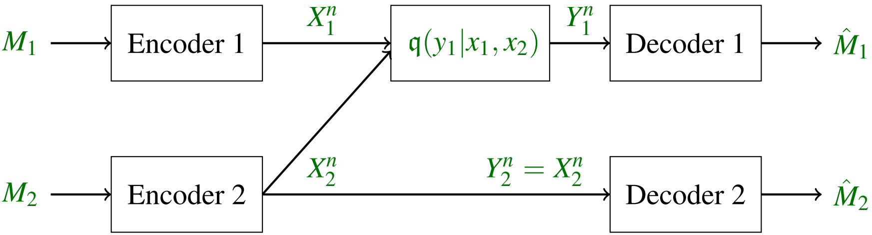
Main: Strict sub-optimality of the Han-Kobayashi region

- Restrict to a class of channels where evaluation is easy
- Show that 2-letter (dependence over time) beats 1-letter (independent over time)

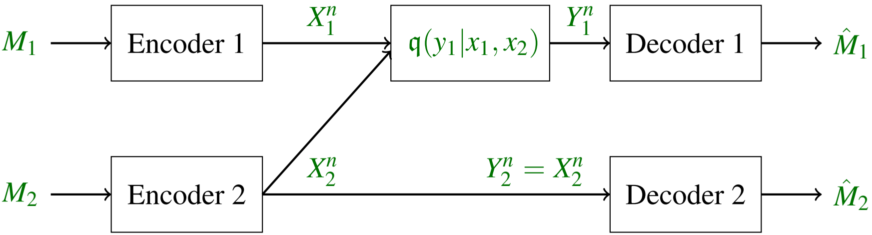
## Marton's region

- Cardinality bounds for evaluation of Marton's region for broadcast channel
- Evaluation of Marton's region for **any** *binary* input broadcast channel
- Other results that helps evaluate Marton's region for broadcast channels

# Clean-Z-Interference (CZI) channels (N-X-Y '15)



# Clean-Z-Interference (CZI) channels (N-X-Y '15)




## Proposition

The HK region of a CZI channel is the set of rate pairs $(R_1, R_2)$ that satisfy

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

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

**Proposition**

For a CZI channel, for any $\lambda \leq 1$

$$\max_{\mathcal{R}_{HK}}(\lambda R_1 + R_2) = \max_{\mathcal{C}}(\lambda R_1 + R_2) = \max_{p_1(x_1)p_2(x_2)} \lambda I(X_1;Y_1) + H(X_2).$$

Proof is rather straightforward and uses standard converse techniques

**Lemma**

For a CZI channel, for all $\lambda > 1$ $\max_{\mathcal{R}_{HK}}(\lambda R_1 + R_2)$ is

$$\max_{p_1(x_1)p_2(x_2)} \left\{ I(X_1,X_2;Y_1) + \mathop{\mathfrak{C}}_{p_2(x_2)} \left[ H(X_2) - I(X_2;Y_1|X_1) + (\lambda - 1) I(X_1;Y_1) \right] \right\}.$$

where $\mathop{\mathfrak{C}}_x[f(x)]$ of $f(x)$ denotes the upper concave envelope of $f(x)$ over $x$.

# Results on CZI

## Proposition

For a CZI channel, for any $\lambda \leq 1$

$$\max_{\mathcal{R}_{HK}}(\lambda R_1 + R_2) = \max_{\mathcal{C}}(\lambda R_1 + R_2) = \max_{p_1(x_1)p_2(x_2)} \lambda I(X_1;Y_1) + H(X_2).$$

Proof is rather straightforward and uses standard converse techniques

## Lemma

For a CZI channel, for all $\lambda > 1$ $\max_{\mathcal{R}_{HK}}(\lambda R_1 + R_2)$ is

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

where $\mathop{\mathcal{C}}_{x}[f(x)]$ of $f(x)$ denotes the upper concave envelope of $f(x)$ over $x$.

# SUB-OPTIMALITY OF HK

For $\lambda > 1$ it turns out that there are examples where

$$\max_{\mathcal{R}_{HK}}(\lambda R_1 + R_2) < \max_{\mathcal{C}}(\lambda R_1 + R_2)$$

### An example (CZI), i.e. $Y_2 = X_2$



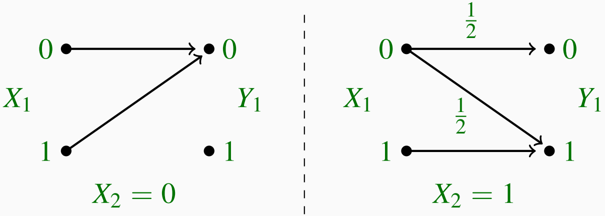

$$\max_{\mathcal{R}_{HK}}(2R_1 + R_2) = 1.1075163.. < 1.108035632 \leq \max_{2\text{-}\mathcal{R}_{HK}}(2R_1 + R_2)$$

# SUB-OPTIMALITY OF HK

For $\lambda > 1$ it turns out that there are examples where

$$\max_{\mathcal{R}_{HK}}(\lambda R_1 + R_2) < \max_{\mathcal{C}}(\lambda R_1 + R_2)$$

## An example (CZI), i.e. $Y_2 = X_2$




$$\max_{\mathcal{R}_{HK}}(2R_1 + R_2) = 1.1075163.. < 1.108035632 \leq \max_{2\text{-}\mathcal{R}_{HK}}(2R_1 + R_2)$$

Tab. 1: Table of counter-examples

| $\lambda$ | channel | $\max_{\mathcal{R}_{hk}}(\lambda R_1 + R_2)$ | $\max_{\mathcal{R}_{two}}(\lambda R_1 + R_2)$ |
|---|---|---|---|
| 2 | $\begin{bmatrix} 1 & 0.5 \\ 1 & 0 \end{bmatrix}$ | 1.107516 | 1.108141 |
| 2.5 | $\begin{bmatrix} 0.204581 & 0.364813 \\ 0.030209 & 0.992978 \end{bmatrix}$ | 1.159383 | 1.169312 |
| 3 | $\begin{bmatrix} 0.591419 & 0.865901 \\ 0.004021 & 0.898113 \end{bmatrix}$ | 1.241521 | 1.255814 |
| 3 | $\begin{bmatrix} 0.356166 & 0.073253 \\ 0.985504 & 0.031707 \end{bmatrix}$ | 1.292172 | 1.311027 |
| 3 | $\begin{bmatrix} 0.287272 & 0.459966 \\ 0.113711 & 0.995405 \end{bmatrix}$ | 1.117253 | 1.123151 |
| 4 | $\begin{bmatrix} 0.429804 & 0.147712 \\ 0.948192 & 0.002848 \end{bmatrix}$ | 1.181392 | 1.196189 |
| 4 | $\begin{bmatrix} 0.068730 & 0.443630 \end{bmatrix}$ [illegible] | 1.223409 | 1.243958 |

# Evaluation of Marton's region

## Extremal auxiliaries

Mar 30, 2015

Evaluating Marton's region

- Simple hard problem (unknown capacity region)

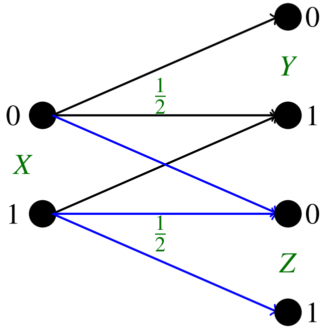



Figure: Binary skew-symmetric broadcast channel

Evaluating Marton's region

- Simple hard problem (unknown capacity region)

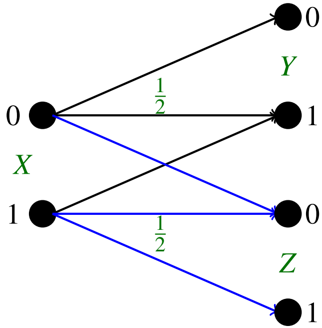



Figure: Binary skew-symmetric broadcast channel

Conjecture: [Nair-Wang ITA '08] For every $(U,V) \to X \to (Y,Z)$

$$I(U;Y) + I(V;Z) - I(U;V) \le \max\{I(X;Y), I(X;Z)\}$$

# Historical remarks: Perturbation approach

- The conjecture caught the attention of Amin Gohari and Venkat Anantharam
- Amin [2009] developed the perturbation approach to show that one can restrict one's attention to $|U|, |V| \leq 2$
- More generally, they used the ideas to show that one can restrict ones attention to $|U| \leq |X|, |V| \leq |X|$ while computing Marton's achievable region
- [Jog and Nair *ITA* 2010] extended the perturbation approach to show that the conjecture was true
- [Geng, Nair, and Wang 2010] showed that the information inequality is true for all broadcast channels when $|X| = 2$

Perturbation approach: A technique to reduce the search space (bounding cardinalities and more) of *extremal* auxiliary distributions

# Historical remarks: Perturbation approach

- The conjecture caught the attention of Amin Gohari and Venkat Anantharam
- Amin [2009] developed the perturbation approach to show that one can restrict one's attention to $|U|, |V| \leq 2$
- More generally, they used the ideas to show that one can restrict ones attention to $|U| \leq |X|, |V| \leq |X|$ while computing Marton's achievable region
- [Jog and Nair ITA 2010] extended the perturbation approach to show that the conjecture was true
- [Geng, Nair, and Wang 2010] showed that the information inequality is true for all broadcast channels when $|X| = 2$

Perturbation approach: A technique to reduce the search space (bounding cardinalities and more) of *extremal* auxiliary distributions

# Historical remarks: Perturbation approach

- The conjecture caught the attention of Amin Gohari and Venkat Anantharam
- Amin [2009] developed the perturbation approach to show that one can restrict one's attention to $|U|, |V| \leq 2$
- More generally, they used the ideas to show that one can restrict ones attention to $|U| \leq |X|, |V| \leq |X|$ while computing Marton's achievable region
- [Jog and Nair ITA 2010] extended the perturbation approach to show that the conjecture was true
- [Geng, Nair, and Wang 2010] showed that the information inequality is true for all broadcast channels when $|X| = 2$

Perturbation approach: A technique to reduce the search space (bounding cardinalities and more) of *extremal* auxiliary distributions

# Aside: Extremal distributions and their uses

Achievable regions (or outer bounds) are usually written as a union of regions - each corresponding to a distribution over random variables (including auxiliary random variables)

Distributions of random variables that give rise to points in the boundary (of the union) form *extremal distributions*

Uses of characterizing extremal distributions:

- If we can show that *extremal distributions* $\subseteq S$ *(a proper subset of all distributions)*, this makes computations of achievable regions (or outer bounds) simpler
  - Is $\mathcal{A}(q) \stackrel{?}{=} \frac{1}{2}\mathcal{A}(q \otimes q)$
- We could utilize properties of extremal distributions to show that inner and outer bounds match for classes of channels
  - The (famous) MIMO Gaussian broadcast channel [Weingarten-Steinberg-Shamai 2007]
  - The capacity of BSC/BEC broadcast channel [Nair 2012]
    - representation using concave envelopes

# ASIDE: EXTREMAL DISTRIBUTIONS AND THEIR USES

Achievable regions (or outer bounds) are usually written as a union of regions - each corresponding to a distribution over random variables (including auxiliary random variables)

Distributions of random variables that give rise to points in the boundary (of the union) form *extremal distributions*

Uses of characterizing extremal distributions

- If we can show that *extremal distributions* $\subseteq \mathcal{S}$ *(a proper subset of all distributions)*, this makes computations of achievable regions (or outer bounds) simpler
  - Is $\mathcal{A}(\mathfrak{q}) \stackrel{?}{=} \frac{1}{2}\mathcal{A}(\mathfrak{q} \otimes \mathfrak{q})$
- We could utilize properties of extremal distributions to show that inner and outer bounds match for classes of channels
  - The (famous) MIMO Gaussian broadcast channel [Weingarten-Steinberg-Shamai 2007]
  - The capacity of BSC/BEC broadcast channel [Nair 2012]
    - representation using concave envelopes

# Aside: Extremal distributions and their uses

Achievable regions (or outer bounds) are usually written as a union of regions - each corresponding to a distribution over random variables (including auxiliary random variables)

Distributions of random variables that give rise to points in the boundary (of the union) form *extremal distributions*

Uses of characterizing extremal distributions

- If we can show that *extremal distributions* $\subseteq \mathcal{S}$ *(a proper subset of all distributions)*, this makes computations of achievable regions (or outer bounds) simpler
  - Is $\mathcal{A}(\mathfrak{q}) \stackrel{?}{=} \frac{1}{2}\mathcal{A}(\mathfrak{q} \otimes \mathfrak{q})$
- We could utilize properties of extremal distributions to show that inner and outer bounds match for classes of channels
  - The (famous) MIMO Gaussian broadcast channel [Weingarten-Steinberg-Shamai 2007]
  - The capacity of BSC/BEC broadcast channel [Nair 2012]
    - representation using concave envelopes

# Current tools - I

## Perturbation based arguments

Mar 30, 2014

# THE PERTURBATION ARGUMENT (GOHARI-ANANTHARAM)

$\max_{p(u,v|x)} I(U;Y) + I(V;Z) - I(U;V)$

**Theorem** (Gohari-Anantharam)

*Suffices to consider* $|U|, |V| \leq |X|$

Observe: Bunt-Caratheodory does not work here

Proof:

Suppose $p_*(u,v|x)$ is a maximizer.

$$p_\epsilon(u,v|x) := p_*(u,v|x)(1+\epsilon L(u)).$$

For $p_\epsilon(u,v|x)$ to be a valid distribution it is necessary that

$$\sum_u p_*(u|x) L(u) = 0 \quad \forall x.$$

A non-zero $L(u)$ exists when $|U| > |X|$.

# THE PERTURBATION ARGUMENT (GOHARI-ANANTHARAM)

$\max_{p(u,v|x)} I(U;Y) + I(V;Z) - I(U;V)$

**Theorem** (Gohari-Anantharam)

*Suffices to consider* $|U|, |V| \leq |X|$

Observe: Bunt-Caratheodory does not work here

Proof:

Suppose $p_*(u,v|x)$ is a maximizer.

$$p_\epsilon(u,v|x) := p_*(u,v|x)(1 + \epsilon L(u)).$$

For $p_\epsilon(u,v|x)$ to be a valid distribution it is necessary that

$$\sum_u p_*(u|x) L(u) = 0 \quad \forall x.$$

A non-zero $L(u)$ exists when $|U| > |X|$.

# THE PERTURBATION ARGUMENT (GOHARI-ANANTHARAM)

$\max_{p(u,v|x)} I(U;Y) + I(V;Z) - I(U;V)$

**Theorem (Gohari-Anantharam)**

*Suffices to consider* $|U|, |V| \leq |X|$

Observe: Bunt-Caratheodory does not work here

Proof:

Suppose $p_*(u,v|x)$ is a maximizer.

$$p_\epsilon(u,v|x) := p_*(u,v|x)(1+\epsilon L(u)).$$

For $p_\epsilon(u,v|x)$ to be a valid distribution it is necessary that

$$\sum_u p_*(u|x)L(u) = 0 \quad \forall x.$$

A non-zero $L(u)$ exists when $|U| > |X|$.

# Argument..

$I(U;Y)+I(V;Z)-I(U;V) = H(Y)+H(Z)+H(U,V)-H(U,Y)-H(V,Z)$

$p_\epsilon(u,v|x) := p_*(u,v|x)(1+\epsilon L(u)).$

$$S(\epsilon) := H_{p_\epsilon}(U,V) - H_{p_\epsilon}(U,Y) - H_{p_\epsilon}(V,Z)$$

Since $p_*(u,v|x)$ is a maximizer

- $\frac{d}{d\epsilon}S(\epsilon)\Big|_{\epsilon=0} = 0, \quad \frac{d^2}{d\epsilon^2}S(\epsilon)\Big|_{\epsilon=0} \le 0$

These two conditions imply that $S(\epsilon)$ has to be a constant.

Choose $\epsilon$ large enough to reduce support of $U$ by one

- Repeat till $|U| \le |X|$, and similarly $|V| \le |X|$
- This perturbation argument has been generalized to
  - prove information inequalities
  - restrict space of extremal distributions

# ARGUMENT..

$I(U;Y)+I(V;Z)-I(U;V) = H(Y)+H(Z)+H(U,V)-H(U,Y)-H(V,Z)$

$p_\epsilon(u,v|x) := p_*(u,v|x)(1+\epsilon L(u)).$

$$S(\epsilon) := H_{p_\epsilon}(U,V) - H_{p_\epsilon}(U,Y) - H_{p_\epsilon}(V,Z)$$

Since $p_*(u,v|x)$ is a maximizer

- $\frac{d}{d\epsilon}S(\epsilon)\Big|_{\epsilon=0} = 0, \quad \frac{d^2}{d\epsilon^2}S(\epsilon)\Big|_{\epsilon=0} \leq 0$

These two conditions imply that $S(\epsilon)$ has to be a constant.

Choose $\epsilon$ large enough to reduce support of $U$ by one

- Repeat till $|U| \leq |X|$, and similarly $|V| \leq |X|$
- This perturbation argument has been generalized to
  - prove information inequalities
  - restrict space of extremal distributions

# Current tools - II

## Concave envelopes and extremal distributions

Mar 30, 2015

# USING CONCAVE ENVELOPES

Superposition coding region for degraded broadcast channels: the union of rate pairs satisfying:

$$R_2 \le I(V;Z)$$
$$R_1 \le I(X;Y|V)$$

for some pmf $p(v,x) : V \to X \to (Y,Z)$

**Characterization of boundary: using supporting hyperplanes**

For $\lambda \ge 1$, observe that

$$\begin{aligned}
\max_{(R_1,R_2)\in C} \lambda R_2 + R_1 &\le \max_{p(v,x)} \lambda I(V;Z) + I(X:Y|V) \\
&= \max_{p(v,x)} \lambda\big(I(X;Z) - I(X;Z|V)\big) + I(X:Y|V) \\
&= \max_{p(x)} \left( \lambda I(X;Z) + \max_{p(v|x)} \big(I(X;Y|V) - \lambda I(X;Z|V)\big) \right) \\
&= \max_{p(x)} \lambda I(X;Z) + C[I(X;Y) - \lambda I(X;Z)]
\end{aligned}$$

# USING CONCAVE ENVELOPES

Superposition coding region for degraded broadcast channels: the union of rate pairs satisfying:

$$R_2 \leq I(V;Z)$$
$$R_1 \leq I(X;Y|V)$$

for some pmf $p(v,x) : V \to X \to (Y,Z)$

## Characterization of boundary: using supporting hyperplanes

For $\lambda \geq 1$, observe that

$$\begin{aligned}
\max_{(R_1,R_2)\in\mathcal{C}} \lambda R_2 + R_1 &\leq \max_{p(v,x)} \lambda I(V;Z) + I(X:Y|V) \\
&= \max_{p(v,x)} \lambda\big(I(X;Z) - I(X;Z|V)\big) + I(X:Y|V) \\
&= \max_{p(x)} \left(\lambda I(X;Z) + \max_{p(v|x)} \big(I(X;Y|V) - \lambda I(X;Z|V)\big)\right) \\
&= \max_{p(x)} \lambda I(X;Z) + \mathcal{C}[I(X;Y) - \lambda I(X;Z)]
\end{aligned}$$

# Application: degraded BSC broadcast channel

Proposition: When $X \to Y \to Z$ is a degraded BSC broadcast channel, it suffices to consider $(V, X) \sim DSBS(s)$ to compute, for any $\lambda \geq 1$,

$$\max_{(R_1, R_2) \in \mathcal{C}} \lambda R_2 + R_1.$$

- Conjectured by Cover and established by Wyner-Ziv (Mrs. Gerber's Lemma)

From previous slide, we saw that we wish to compute

$$\max_{p(x)} \lambda I(X;Z) + \mathcal{C}[I(X;Y) - \lambda I(X;Z)]$$

Claim: The maximum happens at $P(X = 0) = \frac{1}{2}$.

# Application: degraded BSC broadcast channel

Proposition: When $X \to Y \to Z$ is a degraded BSC broadcast channel, it suffices to consider $(V, X) \sim DSBS(s)$ to compute, for any $\lambda \geq 1$,

$$\max_{(R_1,R_2)\in\mathcal{C}} \lambda R_2 + R_1.$$

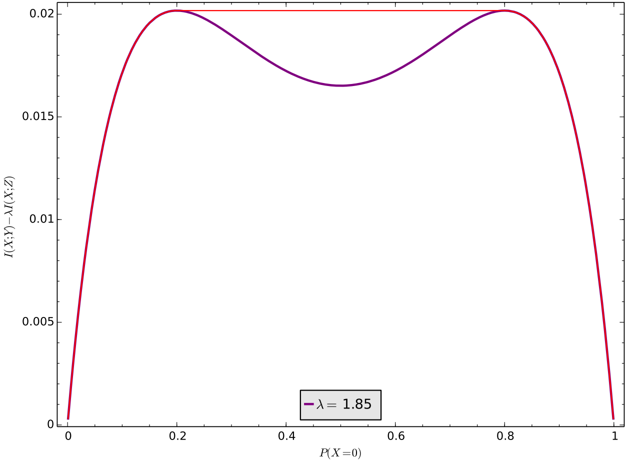
- Conjectured by Cover and established by Wyner-Ziv (Mrs. Gerber's Lemma)

From previous slide, we saw that we wish to compute

$$\max_{p(x)} \lambda I(X;Z) + \mathcal{C}[I(X;Y) - \lambda I(X;Z)]$$

Claim: The maximum happens at $P(X=0) = \frac{1}{2}$.

# DEGRADED BSC BROADCAST, $p = 0.1, q = 0.2$




Observe that: The plot of $I(X;Y) - \lambda I(X;Z)$ vs $\mathrm{P}(X = 0)$ is symmetrical about $\mathrm{P}(X = 0) = \frac{1}{2}$. Implies $U \to X \sim BSC$ (Q.E.D.)

# RESULTS

Capacity results using extremal distributions

- MIMO Gaussian broadcast channel [Weingarten-Steinberg-Shamai '2006]
- BSC-BEC broadcast channel [Nair '10]

Capacity results using concave envelopes

- BSC-BEC broadcast channel [Nair '10]
- Classes of product broadcast channels [ Geng-Gohari-Nair-Yu '2012]
- MIMO Gaussian BC with common message [Geng-Nair 2014]

Other results using concave envelopes

- Strict sub-optimality of UV outer bound

# RESULTS

Capacity results using extremal distributions

- MIMO Gaussian broadcast channel [Weingarten-Steinberg-Shamai '2006]
- BSC-BEC broadcast channel [Nair '10]

Capacity results using concave envelopes

- BSC-BEC broadcast channel [Nair '10]
- Classes of product broadcast channels [ Geng-Gohari-Nair-Yu '2012]
- MIMO Gaussian BC with common message [Geng-Nair 2014]

Other results using concave envelopes

- Strict sub-optimality of UV outer bound

# RESULTS

Capacity results using extremal distributions

- MIMO Gaussian broadcast channel [Weingarten-Steinberg-Shamai '2006]
- BSC-BEC broadcast channel [Nair '10]

Capacity results using concave envelopes

- BSC-BEC broadcast channel [Nair '10]
- Classes of product broadcast channels [ Geng-Gohari-Nair-Yu '2012]
- MIMO Gaussian BC with common message [Geng-Nair 2014]

Other results using concave envelopes

- Strict sub-optimality of UV outer bound

# Combining perturbation and concave envelopes

New cardinality bounds on Marton's achievable region
[Anantharam-Gohari-Nair 2013]

- $|U| + |V| \leq |X| + 1$ suffices
- Further, can restrict to $X = f(U, V)$

**Theorem**

For a binary input broadcast channel, the maximum of $\lambda R_1 + R_2$ in Marton's region, when $\lambda \geq 1$ is,

$$\min_{\alpha \in [0,1]} \max_{p(x)} (\lambda - \alpha) I(X;Y) + \alpha I(X;Z) + \mathcal{C}_{p(x)} \big[ - (\lambda - \alpha) I(X;Y) - \alpha I(X;Z) + \max\{\lambda I(X;Y), I(X:Z)\} \big]$$

# IDEA OF PROOF

Suppose $p(u,v,x)$ is an *extremal distribution* such that

$$\begin{aligned}&\mathcal{C}[-(\alpha-\lambda)H(Y)-\lambda H(Z)+T_{\mathfrak{q},\alpha}(X)]\\&\quad=-(\alpha-\lambda)H(Y)-\lambda H(Z)+\alpha I(U;Y)+I(V;Z)-I(U;V),\end{aligned}$$

then the right hand side is *locally concave* with respect to all perturbations of $p(u,v,x)$.

Rearrange the right hand side as

$$\lambda(H(Y)-H(Z))-\alpha H(Y|U)+H(V|U)-H(Z|V)$$

Consider a perturbation of the form

$$p_\epsilon(u,v,x)=p(u,v,x)(1+\epsilon f(u)), \qquad (\sum_u p(u)f(u)=0).$$

For the second derivative to be negative, we need

$$\frac{d^2}{d\epsilon^2}\left[H(Y)-H(Z)\right]_{\epsilon=0}\le 0$$

# Idea of proof

Suppose $p(u,v,x)$ is an *extremal distribution* such that

$$\begin{aligned}&\mathcal{C}[-(\alpha-\lambda)H(Y)-\lambda H(Z)+T_{\mathfrak{q},\alpha}(X)]\\&\quad=-(\alpha-\lambda)H(Y)-\lambda H(Z)+\alpha I(U;Y)+I(V;Z)-I(U;V),\end{aligned}$$

then the right hand side is *locally concave* with respect to all perturbations of $p(u,v,x)$.

Rearrange the right hand side as

$$\lambda(H(Y)-H(Z))-\alpha H(Y|U)+H(V|U)-H(Z|V)$$

Consider a perturbation of the form

$$p_\epsilon(u,v,x)=p(u,v,x)(1+\epsilon f(u)),\qquad \Big(\sum_u p(u)f(u)=0\Big).$$

For the second derivative to be negative, we need

$$\frac{d^2}{d\epsilon^2}\left[H(Y)-H(Z)\right]_{\epsilon=0}\le 0$$

# IDEA OF PROOF (CNTD...)

Alternately, rearrange the right hand side as

$$(1-\lambda)(H(Z)-H(Y)) - H(Z|V) + H(U|V) - H(U|Y) - (\alpha-1)H(Y|U)$$

Consider a perturbation of the form

$$\hat{p}_\epsilon(u,v,x) = p(u,v,x)(1+\epsilon g(v)), \qquad \Big(\sum_v p(v)g(v) = 0\Big).$$

For the second derivative to be negative, we need

$$\frac{d^2}{d\epsilon^2}\left[H(Z)-H(Y)\right]_{\epsilon=0} \leq 0$$

# OBSERVATION

For a fixed channel $\mathfrak{q}(y,z|x)$ the term $H(Y) - H(Z)$ depends only on $p(x)$.

Hence, if there exists $f(u)$ and $g(v)$ such that $p_\epsilon(x) = \hat{p}_\epsilon(x)$ for all $x \in \mathcal{X}$, then one would need to have

$$\frac{d^2}{d\epsilon^2}\left[H(Y) - H(Z)\right]_{\epsilon=0} = 0.$$

This will in turn force the convex terms to have zero second derivative as well.

As a consequence, it will turn out that the expression

$$-(\alpha - \lambda)H(Y) - \lambda H(Z) + \alpha I(U;Y) + I(V;Z) - I(U;V)$$

will remain unchanged by either of these perturbations.

Set $\epsilon$ large enough so that the support of $U$ or $V$ reduces by one.

# Observation

For a fixed channel $\mathfrak{q}(y,z|x)$ the term $H(Y) - H(Z)$ depends only on $p(x)$.

Hence, if there exists $f(u)$ and $g(v)$ such that $p_\epsilon(x) = \hat{p}_\epsilon(x)$ for all $x \in \mathcal{X}$, then one would need to have

$$\frac{d^2}{d\epsilon^2}\left[H(Y) - H(Z)\right]_{\epsilon=0} = 0.$$

This will in turn force the convex terms to have zero second derivative as well.

As a consequence, it will turn out that the expression

$$-(\alpha - \lambda)H(Y) - \lambda H(Z) + \alpha I(U;Y) + I(V;Z) - I(U;V)$$

will remain unchanged by either of these perturbations.

Set $\epsilon$ large enough so that the support of $U$ or $V$ reduces by one.

# Observation

For a fixed channel $\mathfrak{q}(y,z|x)$ the term $H(Y)-H(Z)$ depends only on $p(x)$.

Hence, if there exists $f(u)$ and $g(v)$ such that $p_\epsilon(x)=\hat{p}_\epsilon(x)$ for all $x\in\mathcal{X}$, then one would need to have

$$\frac{d^2}{d\epsilon^2}\left[H(Y)-H(Z)\right]_{\epsilon=0}=0.$$

This will in turn force the convex terms to have zero second derivative as well.

As a consequence, it will turn out that the expression

$$-(\alpha-\lambda)H(Y)-\lambda H(Z)+\alpha I(U;Y)+I(V;Z)-I(U;V)$$

will remain unchanged by either of these perturbations.

Set $\epsilon$ large enough so that the support of $U$ or $V$ reduces by one.

# CONDITIONS FOR EXISTENCE OF $f(u), g(v)$

1. $\sum_{u,v} p(u,v,x)f(u) = \sum_{u,v} p(u,v,x)g(v) \ \ \forall x \in \mathcal{X}$.
    - From the condition: $p_\epsilon(x) = \hat{p}_\epsilon(x)$ for all $x \in \mathcal{X}$.
2. $\sum_{u,v,x} p(u,v,x)f(u) = 0$.
    - From the condition: $p_\epsilon(x)$ is a valid probability distribution.
3. $\sum_{u,v,x} p(u,v,x)g(v) = 0$.
    - From the condition: $\hat{p}_\epsilon(x)$ is a valid probability distribution.

So there are $|X| + 1$ linear constraints on a vector of size $|U| + |V|$.

A non-trivial solution exists when $|U| + |V| > |X| + 1$.

# Other results for computing Marton's region

From earlier slides we can restrict to:

- $|U| + |V| \leq |X| + 1$ and $X = f(U, V)$.

It turns out that we need not search over certain functions

1. XOR pattern: there is a $k \times k$ sub-matrix such that rows and columns are permutations in $\mathcal{S}_{|X|}$. For example, $X = f(U, V)$ has

$$\begin{array}{ccc} \text{U/V} & v_1 & v_2 \\ \begin{array}{c} u_1 \\ u_2 \end{array} & \multicolumn{2}{c}{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}} \end{array}$$

2. AND pattern: All entries in a row and all in entries in a column map to same entry.

Using above results one can estimate Marton's region for $|X| = 4$. Simulations are (as of yet) unable to find an example such that

$$\mathcal{A}(q) \subsetneq \frac{1}{2}\mathcal{A}(q \otimes q).$$

# Other results for computing Marton's region

From earlier slides we can restrict to:

- $|U| + |V| \leq |X| + 1$ and $X = f(U, V)$.

It turns out that we need not search over certain functions

1. XOR pattern: there is a $k \times k$ sub-matrix such that rows and columns are permutations in $\mathcal{S}_{|X|}$. For example, $X = f(U, V)$ has

$$\begin{array}{ccc} \text{U/V} & v_1 & v_2 \\ \begin{matrix} u_1 \\ u_2 \end{matrix} & \multicolumn{2}{c}{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}} \end{array}$$

2. AND pattern: All entries in a row and all in entries in a column map to same entry.

Using above results one can estimate Marton's region for $|X| = 4$.
Simulations are (as of yet) unable to find an example such that

$$\mathcal{A}(\mathfrak{q}) \subsetneq \frac{1}{2}\mathcal{A}(\mathfrak{q} \otimes \mathfrak{q}).$$

# SUMMARY

Computing regions in network information theory

- Understanding/restricting extremal distributions is the key
  - Going beyond the traditional representation [Cover] using auxiliary random variables
  - Perturbation ideas (calculus of variations)
  - Representation as concave envelopes

The above computations are useful

- To see if the current regions are optimal or not
- To establish capacity regions of some classes of channels

THANK YOU

# SUMMARY

Computing regions in network information theory

- Understanding/restricting extremal distributions is the key
  - Going beyond the traditional representation [Cover] using auxiliary random variables
  - Perturbation ideas (calculus of variations)
  - Representation as concave envelopes

The above computations are useful

- To see if the current regions are optimal or not
- To establish capacity regions of some classes of channels

THANK YOU