Single letterization arguments in network information theory

Optimality of Gaussian random variables

Chandra Nair

Dept. of Information Engineering
The Chinese University of Hong Kong

Jan 28/30, 2015

Last time...

- Showed that for a point-to-point channel with additive Gaussian noise, the optimal input distribution subject to a power constraint is Gaussian
  - Used a characterization of Gaussian
  - Used the single-letterization arguments

This time:

- General broadcast channels
  - Vector Gaussian (MIMO) broadcast channels with private messages
  - Vector Gaussian (MIMO) broadcast channels with private and common messages
- Other applications of the technique

THE LAST TALK OF THE SCHOOL...

Last time...

- Showed that for a point-to-point channel with additive Gaussian noise, the optimal input distribution subject to a power constraint is Gaussian
  - Used a characterization of Gaussian
  - Used the single-letterization arguments

This time:

- General broadcast channels
  - Vector Gaussian (MIMO) broadcast channels with private messages
  - Vector Gaussian (MIMO) broadcast channels with private and common messages
- Other applications of the technique

The characterization of Gaussian random variables

Theorem (Bernstein '40, Darmois '51, Skitovic '54)

If X and Y are independent random variables such that $X+Y$ and $X-Y$ are independent, then X and Y must be Gaussian with the same covariance matrix.

BROADCAST CHANNELS [COVER '72]

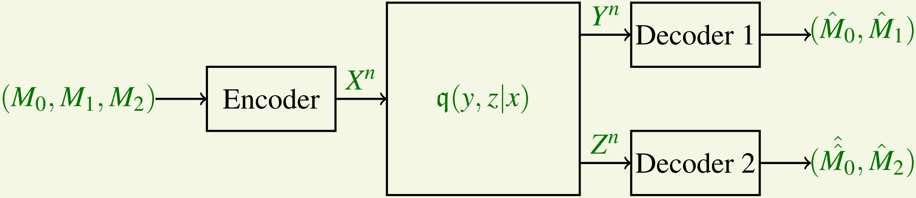

Figure: Discrete memoryless broadcast channel

- Goal: Compute *Capacity Region* or set of achievable rates (R_0, R_1, R_2)?
- Open for discrete memoryless channels
- Solved for vector Gaussian (MIMO) channels

BROADCAST CHANNELS [COVER '72]

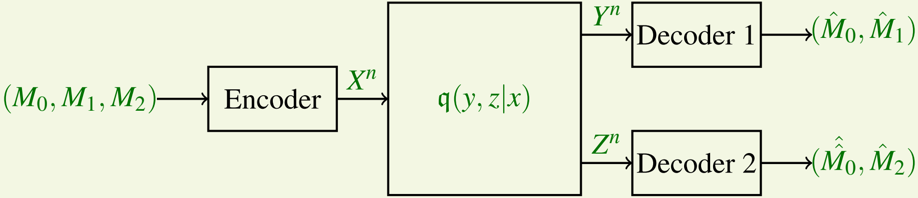

Figure: Discrete memoryless broadcast channel

- Goal: Compute *Capacity Region* or set of achievable rates (R_0, R_1, R_2)?

- Open for discrete memoryless channels
- Solved for vector Gaussian (MIMO) channels

Marton's achievable region

The set of rates (R_0, R_1, R_2) satisfying

$$\begin{aligned} R_0 &\leq \min\{I(W;Y), I(W;Z)\} \\ R_0 + R_1 &\leq I(U,W;Y) \\ R_0 + R_2 &\leq I(V,W;Z) \\ R_0 + R_1 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \end{aligned}$$

for any $(U,V,W) \to X \xrightarrow{\mathrm{q}} (Y,Z)$ 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
- Do not know whether this is optimal or not optimal

(U, V, W)-OUTER BOUND

The set of rates (R_0, R_1, R_2) satisfying

$$\begin{aligned} R_0 &\leq \min\{I(W;Y), I(W;Z)\} \\ R_0 + R_1 &\leq \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) \\ R_0 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} + I(V;Z|W) \\ R_0 + R_1 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} + \min\{I(U;Y|W) \\ &\qquad + I(X;Z|U,W), I(V;Z|W) + I(X;Y|V,W)\} \end{aligned}$$

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

- Simple hard problem (unknown capacity region)

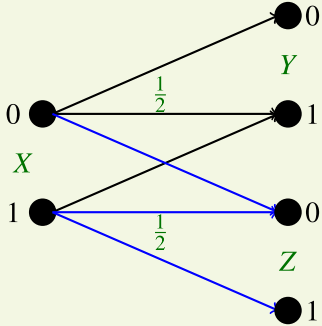

Figure: Binary skew-symmetric broadcast channel

The inner and outer bounds presented earlier differ for this channel

A simple channel with unknown capacity region

- Simple hard problem (unknown capacity region)

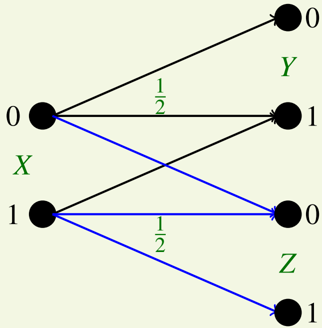

Figure: Binary skew-symmetric broadcast channel

The inner and outer bounds presented earlier differ for this channel

Establish the capacity region

Idea: Show that the inner and outer bound evaluate to the same region

- In words, it seems to be simple: show that the regions coincide
- Difficulty: Evaluation of the regions (union over auxiliaries)
  - I have spent a good part of last 7 years trying to evaluate various regions and understand *extremal auxiliaries*

Vector Gaussian (MIMO) broadcast channels

Establish the capacity region

Idea: Show that the inner and outer bound evaluate to the same region

- In words, it seems to be simple: show that the regions coincide
- Difficulty: Evaluation of the regions (union over auxiliaries)
  - I have spent a good part of last 7 years trying to evaluate various regions and understand *extremal auxiliaries*
    - Showed that the bounds in general are different
    - Showed that the outer bound is strictly sub-optimal
    - Found new capacity regions
    - Discovered new information inequalities

Establish the capacity region

Idea: Show that the inner and outer bound evaluate to the same region

- In words, it seems to be simple: show that the regions coincide
- Difficulty: Evaluation of the regions (union over auxiliaries)
  - I have spent a good part of last 7 years trying to evaluate various regions and understand *extremal auxiliaries*
    - Showed that the bounds in general are different
    - Showed that the outer bound is strictly sub-optimal
    - Found new capacity regions
    - Discovered new information inequalities
- Use the single-letterization in outer bound to argue that Gaussian is maximal

Case 1: Private Messages only: Single letterization (outer bound)

Claim: For $\lambda \geq 1$ we have

$$R_1 + \lambda R_2 \leq \max_{p(v,x):\mathrm{E}(XX^T)\preceq K} I(X;Y|V) + \lambda I(V;Z).$$

Proof: From Fano's inequality

$$\begin{aligned}
R_1 + \lambda R_2 &\leq \frac{1}{n}\big(I(M_1;Y^n|M_2) + \lambda I(M_2;Z^n)\big) \\
&\leq \frac{1}{n}\big(I(X^n;Y^n|M_2) + \lambda I(M_2;Z^n)\big) && D-P\ ineq \\
&\leq \frac{1}{n} \max_{p(v,x^n):\frac{1}{n}\sum_i \mathrm{E}(X_iX_i^T)\preceq K} \big(I(X^n;Y^n|V) + \lambda I(V;Z^n)\big) && \text{set } V = M_2
\end{aligned}$$

SINGLE LETTERIZATION

Goal: Show that for $\lambda > 1$

$$\frac{1}{2} \max_{p(v,x_1,x_2)} I(X_1, X_2; Y_1, Y_2|V) + \lambda I(V; Z_1, Z_2) \\ \leq \max_{p(v,x)} I(X; Y|V) + \lambda I(V; Z)$$

Observe that (exercise)

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2|V) + \lambda I(V; Z_1, Z_2) \\ = I(X_1; Y_1|V, Z_2) + \lambda I(V, Z_2; Z_1) \\ + I(X_2; Y_2|V, Y_1) + \lambda I(V, Y_1; Z_2) \\ -\lambda I(Z_1; Z_2) - (\lambda - 1) I(Y_1; Z_2|V) \end{aligned}$$

Optimality of Gaussian Variables

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{G}_1$$
$$\mathbf{Z} = \mathbf{B}\mathbf{X} + \mathbf{G}_2$$

here $\mathbf{G}_1, \mathbf{G}_2$ are i.i.d. Gaussian noise vectors.
Let (V_*, X_*) be a maximizer of

$$D = \max_{\substack{p(v,x) \\ \mathrm{E}(XX^T) \preceq K}} I(X;Y|V) + \lambda I(V;Z)$$

Let $(V_1, X_1), (V_2, X_2)$ be i.i.d. distributed according to (V_*, X_*).

$$2D = I(X_1, X_2; Y_1, Y_2|V_1, V_2) + \lambda I(V_1, V_2; Z_1, Z_2)$$

As before, let

$$X_\pm = \frac{X_1 \pm X_2}{\sqrt{2}} \qquad Y_\pm = \frac{Y_1 \pm Y_2}{\sqrt{2}} \qquad Z_\pm = \frac{Z_1 \pm Z_2}{\sqrt{2}}$$

Optimality of Gaussian Variables

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{G}_1$$
$$\mathbf{Z} = \mathbf{B}\mathbf{X} + \mathbf{G}_2$$

here $\mathbf{G}_1, \mathbf{G}_2$ are i.i.d. Gaussian noise vectors.
Let (V_*, X_*) be a maximizer of

$$D = \max_{\substack{p(v,x) \\ \mathrm{E}(XX^T) \preceq K}} I(X;Y|V) + \lambda I(V;Z)$$

Let $(V_1, X_1), (V_2, X_2)$ be i.i.d. distributed according to (V_*, X_*).

$$2D = I(X_1, X_2; Y_1, Y_2 | V_1, V_2) + \lambda I(V_1, V_2; Z_1, Z_2)$$

As before, let

$$X_\pm = \frac{X_1 \pm X_2}{\sqrt{2}} \qquad Y_\pm = \frac{Y_1 \pm Y_2}{\sqrt{2}} \qquad Z_\pm = \frac{Z_1 \pm Z_2}{\sqrt{2}}$$

Single letterization - revisited

Note that

$$\begin{aligned}
2D &= I(X_1, X_2; Y_1, Y_2 | V_1, V_2) + \lambda I(V_1, V_2; Z_1, Z_2) \\
&= I(X_+, X_-; Y_+, Y_- | V_1, V_2) + \lambda I(V_1, V_2; Z_+, Z_-) \\
&= I(X_+; Y_+ | V_1, V_2, Z_-) + \lambda I(V_1, V_2, Z_-; Z_+) \\
&\quad + I(X_-; Y_- | V_1, V_2, Y_+) + \lambda I(V_1, V_2, , Y_+; Z_-) \\
&\quad -\lambda I(Z_+; Z_-) - (\lambda - 1) I(Y_+; Z_- | V_1, V_2)
\end{aligned}$$

Hence, we obtain that,

$$I(Z_+; Z_-) = 0, \quad I(Y_+; Z_- | V_1, V_2) = 0.$$

The latter equality implies that (recall)

$$I(X_+; X_- | V_1, V_2) = 0.$$

Implies $X | V = v \sim \mathcal{N}(\mu_v, K_*)$, for some $K_* \preceq K$.

Thus maximizer is

$$X = U + V, \quad U \sim \mathcal{N}(0, K_*), \ V \sim \mathcal{N}(0, K - K_*)$$

Single letterization - revisited

Note that

$$\begin{aligned}2D &= I(X_1, X_2; Y_1, Y_2 | V_1, V_2) + \lambda I(V_1, V_2; Z_1, Z_2)\\ &= I(X_+, X_-; Y_+, Y_- | V_1, V_2) + \lambda I(V_1, V_2; Z_+, Z_-)\\ &= I(X_+; Y_+ | V_1, V_2, Z_-) + \lambda I(V_1, V_2, Z_-; Z_+)\\ &\quad + I(X_-; Y_- | V_1, V_2, Y_+) + \lambda I(V_1, V_2, , Y_+; Z_-)\\ &\quad -\lambda I(Z_+; Z_-) - (\lambda - 1) I(Y_+; Z_- | V_1, V_2)\end{aligned}$$

Hence, we obtain that,

$$I(Z_+; Z_-) = 0, \quad I(Y_+; Z_- | V_1, V_2) = 0.$$

The latter equality implies that (recall)

$$I(X_+; X_| V_1, V_2) = 0.$$

Implies $X|V = v \sim \mathcal{N}(\mu_v, K_*)$, for some $K_* \preceq K$.

Thus maximizer is

$$X = U + V, \quad U \sim \mathcal{N}(0, K_*),\ V \sim \mathcal{N}(0, K - K_*).$$

IMPLICATION

For some $K' \preceq K$ and $X = U + V, \quad U \sim \mathcal{N}(0, K_*),\ V \sim \mathcal{N}(0, K - K_*)$

$$R_2 = I(V;Z), R_1 = I(X;Y_1|V)$$

lies on or outside the boundary of the outer bound to the capacity region.

Question: Can one always achieve this rate pair

Can one construct a U' (jointly distributed with V such that

$$I(U';Y) - I(U';V) = I(X;Y|V),$$

when (V, X) satisfy the relationship above.

Reason: If such a U' exists, then one can substitute in Marton's achievable region and show that the rate pair is achievable.

Answer: Yes (thanks to Max Costa ('81)) because this is exactly the Dirty Paper Coding choice. $\square$

Implication

For some $K' \preceq K$ and $X = U + V, \quad U \sim \mathcal{N}(0, K_*),\ V \sim \mathcal{N}(0, K - K_*)$

$$R_2 = I(V;Z), R_1 = I(X;Y_1|V)$$

lies on or outside the boundary of the outer bound to the capacity region.

Question: Can one always achieve this rate pair

Can one construct a U' (jointly distributed with V such that

$$I(U';Y) - I(U';V) = I(X;Y|V),$$

when (V, X) satisfy the relationship above.

Reason: If such a U' exists, then one can substitute in Marton's achievable region and show that the rate pair is achievable.

Answer: Yes (thanks to Max Costa ('81)) because this is exactly the Dirty Paper Coding choice. □

IMPLICATION

For some $K' \preceq K$ and $X = U + V, \quad U \sim \mathcal{N}(0, K_*),\ V \sim \mathcal{N}(0, K - K_*)$

$$R_2 = I(V;Z), R_1 = I(X;Y_1|V)$$

lies on or outside the boundary of the outer bound to the capacity region.

Question: Can one always achieve this rate pair

Can one construct a U' (jointly distributed with V such that

$$I(U';Y) - I(U';V) = I(X;Y|V),$$

when (V,X) satisfy the relationship above.

Reason: If such a U' exists, then one can substitute in Marton's achievable region and show that the rate pair is achievable.

Answer: Yes (thanks to Max Costa ('81)) because this is exactly the Dirty Paper Coding choice. □

Dirty paper coding choice

The channel was

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{G}_1$$

Let $U \sim \mathcal{N}(0, K_*), \;\; V \sim \mathcal{N}(0, K - K_*), \;\; X = U + V.$

Set $U' = U + AV_*$ where $A = K_*\mathbf{A}^T(\mathbf{A}K_*\mathbf{A}^T + I)^{-1}$.

Verify:

$$I(U';Y) - I(U';V) = I(X;Y|V).$$

Thus outer and inner bounds match.

The channel was

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{G}_1$$

Let $U \sim \mathcal{N}(0, K_*),\ \ V \sim \mathcal{N}(0, K - K_*),\ \ X = U + V.$

Set $U' = U + AV_*$ where $A = K_*\mathbf{A}^T(\mathbf{A}K_*\mathbf{A}^T + I)^{-1}$.

Verify:

$$I(U';Y) - I(U';V) = I(X;Y|V).$$

Thus outer and inner bounds match.

Recap

We established the capacity region for MIMO Gaussian broadcast channel with private messages

We did this by showing that inner and outer bound coincided

- Used single-letterization to show optimality of Gaussian random variables in the outer bound
- Showed that the outer bound is achievable, using a Dirty-Paper-Coding-inspired auxiliary variable construction

The old method:

- Tour de force in optimization
- Channel enhancement idea (to be able to use EPI)
- MAC-BC duality

A long paper by Weingarten-Steingberg-Shamai ('2006) (Best IT paper award)

RECAP

We established the capacity region for MIMO Gaussian broadcast channel with private messages

We did this by showing that inner and outer bound coincided

- Used single-letterization to show optimality of Gaussian random variables in the outer bound
- Showed that the outer bound is achievable, using a Dirty-Paper-Coding-inspired auxiliary variable construction

The old method:

- Tour de force in optimization
- Channel enhancement idea (to be able to use EPI)
- MAC-BC duality

A long paper by Weingarten-Steingberg-Shamai (‘2006) (Best IT paper award)

Historical Remarks

The technique by Weingarten-Steinberg-Shamai could not be extended to the case with private and common messages

- Despite concerted collaborative efforts by a good group of researchers

What we did

- Looked for a more direct proof of this (open problem in NIT)
- Once we obtained this technique, we could extend it to private and common messages

We need one other insights (which we had at that time)

- a min-max theorem that we had earlier established for discrete channels

Consider the U, V, W outer bound

$$\begin{aligned} R_0 &\leq \min\{I(W;Y), I(W;Z)\} \\ R_0 + R_1 &\leq \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) \\ R_0 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} + I(V;Z|W) \\ R_0 + R_1 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} + \min\{I(U;Y|W) \\ &\quad + I(X;Z|U,W), I(V;Z|W) + I(X;Y|V,W)\} \end{aligned}$$

Goal: Evaluate the outer bound along some directions

Let $\lambda_0 > (\lambda_1 + \lambda_2)$ and $\lambda_1, \lambda_2 > 0$.

$$\max_{(R_0,R_1,R_2)} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1 + \lambda_2) R_2$$

A min-max claim

$$\max_{p(u,v,w,x)} \lambda_0 \min\{I(W;Y), I(W;Z)\} + (\lambda_1 + \lambda_2) I(V;Z|W) + \lambda_1 I(X;Y|V,W)$$

$$= \min_{\alpha \in [0,1]} \max_{p(u,v,w,x)} \lambda_0 \Big(\alpha I(W;Y) + (1-\alpha) I(W;Z)\Big) + (\lambda_1 + \lambda_2) I(V;Z|W) + \lambda_1 I(X;Y|V,W)$$

A max-min theorem

Theorem (Terkelsen '72)

Let X be a compact connected space, let Y be a set, and let $f : X \times Y \mapsto \mathbb{R}$ be a function satisfying:

(i) *For any $y_1, y_2 \in Y$ there exists $y_0 \in Y$ such that*

$$f(x, y_0) \geq \frac{1}{2}\left(f(x, y_1) + f(x, y_2)\right), \forall x \in X.$$

(ii) *Every finite intersection of sets of the form $\{x \in X : f(x, y) \leq \alpha)\}$ with $(y, \alpha) \in Y \times R$ is closed and connected.*

Then

$$\sup_{y \in Y} \min_{x \in X} f(x, y) = \min_{x \in X} \sup_{y \in Y} f(x, y).$$

A COROLLARY

Corollary (Geng-Gohari-Nair-Yu '14)

Let Λ_d be the d-dimensional simplex, i.e. $\lambda_i \geq 0$ and $\sum_{i=1}^{d} \lambda_i = 1$. Let $\mathcal{P}$ be a set of probability distributions $p(u)$. Let $T_i(p(u)), i = 1, .., d$ be a set of functions such that the set $\mathcal{A}$, defined by

$$\mathcal{A} = \{(a_1, a_2, ..., a_d) \in \mathbb{R}^d : a_i \leq T_i(p(u)) \text{ for some } p(u) \in \mathcal{P}\},$$

is a convex set.
Then

$$\sup_{p(u)\in\mathcal{P}} \min_{\lambda\in\Lambda_d} \sum_{i=1}^{d} \lambda_i T_i(p(u)) = \min_{\lambda\in\Lambda_d} \sup_{p(u)\in\mathcal{P}} \sum_{i=1}^{d} \lambda_i T_i(p(u)).$$

Remarks:

- The convexity of $\mathcal{A}$ in network information theory comes from a time-sharing argument.

A Corollary

Corollary (Geng-Gohari-Nair-Yu '14)

Let Λ_d be the d-dimensional simplex, i.e. $\lambda_i \geq 0$ and $\sum_{i=1}^{d} \lambda_i = 1$. Let $\mathcal{P}$ be a set of probability distributions $p(u)$. Let $T_i(p(u)), i = 1, .., d$ be a set of functions such that the set $\mathcal{A}$, defined by

$$\mathcal{A} = \{(a_1, a_2, ..., a_d) \in \mathbb{R}^d : a_i \leq T_i(p(u)) \text{ for some } p(u) \in \mathcal{P}\},$$

is a convex set.
Then

$$\sup_{p(u)\in\mathcal{P}} \min_{\lambda\in\Lambda_d} \sum_{i=1}^{d} \lambda_i T_i(p(u)) = \min_{\lambda\in\Lambda_d} \sup_{p(u)\in\mathcal{P}} \sum_{i=1}^{d} \lambda_i T_i(p(u)).$$

Remarks:

- The convexity of $\mathcal{A}$ in network information theory comes from a time-sharing argument.

Optimality of Gaussian

Let $\lambda_0 > (\lambda_1 + \lambda_2)$, $\lambda_i \geq 0$ and $\alpha \in [0, 1]$

Proposition

The value of the optimization problem

$$\sup_{X:\mathrm{E}(XX^T) \preceq K} \lambda_0 \Big(\alpha I(W;Y) + (1-\alpha) I(W;Z) \Big) + (\lambda_1 + \lambda_2) I(V;Z|W) + \lambda_1 I(X;Y|V,W)$$

is attained by a Gaussian distribution (and Gaussian auxiliaries).

Proof: Mimic the single letterization of U,V,W outer bound

Completion of capacity proof

To show that the outer bound is achievable: use the dirty paper coding choice for Marton's region.

Thus, we solved the capacity region of the
Vector Gaussian broadcast channel with private and common messages

This (optimality of Gaussian via single-letterization) technique was used

- to give an information theoretic proof of the celebrated Gaussian hypercontractivity region [Nair '14]
- to establish an inequality on *long Markov chains* [Courtade]
- A variety of network information theory settings [Chong et al.]
- A simpler proof of the secrecy capacity

Completion of capacity proof

To show that the outer bound is achievable: use the dirty paper coding choice for Marton's region.

Thus, we solved the capacity region of the
Vector Gaussian broadcast channel with private and common messages

This (optimality of Gaussian via single-letterization) technique was used

- to give an information theoretic proof of the celebrated Gaussian hypercontractivity region [Nair '14]
- to establish an inequality on *long Markov chains* [Courtade]
- A variety of network information theory settings [Chong et. al.]
- A simpler proof of the secrecy capacity

THANKS

Many Thanks

Muito Obrigado