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

The mathematics of digital communication [Shannon '48]

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

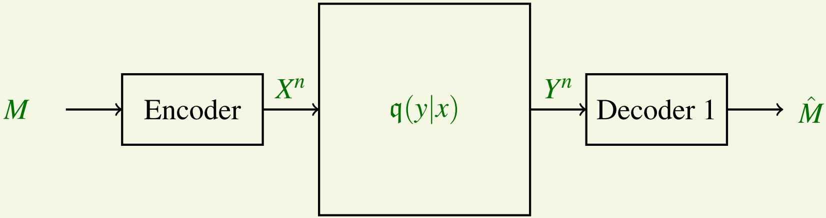

Figure: Discrete memoryless channel

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

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} \sum_{x,y} p(x)\mathfrak{q}(y|x) \log \frac{\mathfrak{q}(y|x)}{p(y)}$$

The mathematics of digital communication [Shannon '48]

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

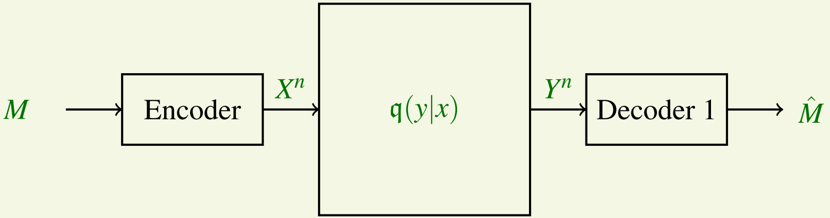

Figure: Discrete memoryless channel

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

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} \sum_{x,y} p(x)\mathfrak{q}(y|x) \log \frac{\mathfrak{q}(y|x)}{p(y)}.$$

Proof of channel capacity - I

A codebook of rate R, with block length n, is said to achieve a probability of error at most ϵ if: for $M \sim U[0 : 2^{nR} - 1]$, we have $\mathrm{P}(\hat{M} \neq M) \leq \epsilon$.

The proof of capacity uses two mathematical tools:

1. Fano's inequality
2. Data-processing inequality

Fano's inequality

If $M, \hat{M} \in \{0, 1, \ldots, 2^{nR} - 1\}$ and $\mathrm{P}(\hat{M} \neq M) \leq \epsilon$ then

$$H(M|\hat{M}) \leq 1 + \epsilon nR.$$

Proof: Let

$$E = \begin{cases} 0 & M = \hat{M} \\ 1 & M \neq \hat{M} \end{cases}.$$

Then

$$\begin{aligned} H(M|\hat{M}) &\leq H(M, E|\hat{M}) = H(E|\hat{M}) + H(M|E, \hat{M}) \\ &\leq 1 + \epsilon \log_2 2^{nR} = 1 + \epsilon nR. \end{aligned}$$

Fano's inequality

Fano's inequality

If $M, \hat{M} \in \{0, 1, \ldots, 2^{nR} - 1\}$ and $\mathrm{P}(\hat{M} \neq M) \leq \epsilon$ then

$$H(M|\hat{M}) \leq 1 + \epsilon nR.$$

Proof: Let

$$E = \begin{cases} 0 & M = \hat{M} \\ 1 & M \neq \hat{M} \end{cases}.$$

Then

$$\begin{aligned} H(M|\hat{M}) &\leq H(M, E|\hat{M}) = H(E|\hat{M}) + H(M|E, \hat{M}) \\ &\leq 1 + \epsilon \log_2 2^{nR} = 1 + \epsilon nR. \end{aligned}$$

Data processing inequality

Data-processing inequality

If $X \to Y \to Z$ forms a Markov chain then

$$I(X;Z) \leq I(X;Y).$$

Proof: Since $X \to Y \to Z$ is Markov, we have $I(X;Z|Y) = 0$. Observe that

$$\begin{aligned} I(X;Y) &= I(X;Y) + I(X;Z|Y) = I(X;Y,Z) \\ &= I(X;Z) + I(X;Y|Z) \geq I(X;Z). \end{aligned}$$

Data processing inequality

Data-processing inequality

If $X \to Y \to Z$ forms a Markov chain then

$$I(X;Z) \leq I(X;Y).$$

Proof: Since $X - Y - Z$ is Markov, we have $I(X;Z|Y) = 0$. Observe that

$$\begin{aligned} I(X;Y) &= I(X;Y) + I(X;Z|Y) = I(X;Y,Z) \\ &= I(X;Z) + I(X;Y|Z) \geq I(X;Z). \end{aligned}$$

Proof of channel capacity - II

Given: A sequence (increasing block length) of codebooks with rate R with probability of error at most ϵ.
Goal: Bound the value of R.

Observe that

$$
\begin{aligned}
R &= \frac{1}{n} H(M) && M \sim U[0 : 2^{nR} - 1] \\
&= \frac{1}{n}\left(I(M; \hat{M}) + H(M|\hat{M})\right) \\
&\leq \frac{1}{n}\left(I(M; \hat{M}) + 1 + nR\epsilon\right) && \text{Fano's inequality} \\
&\leq \frac{1}{n} I(X^n; Y^n) + \frac{1}{n} + R\epsilon && \text{Markov}: (M - X^n - Y^n - \hat{M}) \\
&\leq \frac{1}{1-\epsilon} \limsup_{n\to\infty} \frac{1}{n} I(X^n; Y^n) && \text{Letting } n \to \infty \\
&\leq \limsup_{n\to\infty} \frac{1}{n} I(X^n; Y^n) && \text{Letting } \epsilon \to 0.
\end{aligned}
$$

Proof of channel capacity - II

Given: A sequence (increasing block length) of codebooks with rate R with probability of error at most ϵ.
Goal: Bound the value of R.
Observe that

$$\begin{aligned}
R &= \frac{1}{n}H(M) && M \ \sim U[0:2^{nR}-1] \\
&= \frac{1}{n}\big(I(M;\hat{M}) + H(M|\hat{M})\big) && \\
&\leq \frac{1}{n}\big(I(M;\hat{M}) + 1 + nR\epsilon\big) && \text{Fano's inequality} \\
&\leq \frac{1}{n}I(X^n;Y^n) + \frac{1}{n} + R\epsilon && \text{Markov}: \ (M - X^n - Y^n - \hat{M}) \\
&\leq \frac{1}{1-\epsilon}\limsup_{n\to\infty}\frac{1}{n}I(X^n;Y^n) && \text{Letting } n \to \infty \\
&\leq \limsup_{n\to\infty}\frac{1}{n}I(X^n;Y^n) && \text{Letting } \epsilon \to 0.
\end{aligned}$$

SINGLE LETTERIZATION

We established via Fano's inequality and data-processing that

$$R \le \limsup_{n\to\infty} \frac{1}{n} I(X^n; Y^n).$$

Unfortunately, this value is not computable (infinite dimensional optimization problem).

Goal: Show that

$$\frac{1}{n} I(X^n; Y^n) \le \max_{p(x)} I(X; Y).$$

1-LETTER TO 2-LETTER

1-letter to 2-letter

Claim: It suffices to show that for every pair of channels $\mathfrak{q}(y|x)$, and for every $p(x_1, x_2)$

$$\frac{1}{2} I_{\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2; Y_1, Y_2) \leq \max_{p(x)} I(X; Y).$$

Proof: If true by induction, we would have

$$\begin{aligned}\frac{1}{4} I_{\mathfrak{q}\otimes\mathfrak{q}\otimes\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2, X_3, X_4; Y_1, Y_2, Y_3, Y_4) &\leq \max_{p(x_1,x_2)} \frac{1}{2} I_{\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2; Y_1, Y_2) \\ &\leq \max_{p(x)} I(X; Y)\end{aligned}$$

and so on for higher powers of 2.

Exercise: Why does powers of 2 suffice (and not every n)?
Hint: Try a sandwiching argument... (superadditivity).

1-letter to 2-letter

Claim: It suffices to show that for every pair of channels $\mathfrak{q}(y|x)$, and for every $p(x_1, x_2)$

$$\frac{1}{2} I_{\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2; Y_1, Y_2) \leq \max_{p(x)} I(X; Y).$$

Proof: If true by induction, we would have

$$\begin{aligned}\frac{1}{4} I_{\mathfrak{q}\otimes\mathfrak{q}\otimes\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2, X_3, X_4; Y_1, Y_2, Y_3, Y_4) &\leq \max_{p(x_1,x_2)} \frac{1}{2} I_{\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2; Y_1, Y_2) \\ &\leq \max_{p(x)} I(X; Y)\end{aligned}$$

and so on for higher powers of 2.

Exercise: Why does powers of 2 suffice (and not every n)?
Hint: Try a sandwiching argument... (superadditivity).

1-LETTER TO 2-LETTER - II

To Show:

$$\frac{1}{2} I_{\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2; Y_1, Y_2) \leq \max_{p(x)} I(X;Y).$$

Proof: Note

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\
&= I(X_1; Y_1) + I(Y_1, X_1, X_2; Y_2) - I(Y_1; Y_2) && (X_2 - X_1 - Y_1) \\
&= I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) && ((X_1, Y_1) - X_2 - Y_2) \\
&\leq 2 \max_{p(x)} I(X;Y).
\end{aligned}$$

Call this *a single-letterization argument.*

My talks: Optimality of Gaussian distributions in additive Gaussian noise channels as a consequence of single-letterization arguments.

1-LETTER TO 2-LETTER - II

To Show:

$$\frac{1}{2} I_{\mathfrak{q}\otimes\mathfrak{q}}(X_1, X_2; Y_1, Y_2) \leq \max_{p(x)} I(X; Y).$$

Proof: Note

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2 | Y_1) \\ &= I(X_1; Y_1) + I(Y_1, X_1, X_2; Y_2) - I(Y_1; Y_2) && (X_2 - X_1 - Y_1) \\ &= I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) && ((X_1, Y_1) - X_2 - Y_2) \\ &\leq 2 \max_{p(x)} I(X; Y). \end{aligned}$$

Call this *a single-letterization argument*.

My talks: Optimality of Gaussian distributions in additive Gaussian noise channels as a consequence of single-letterization arguments.

Outline - for the next part

- A characterization of Gaussian random variables
- Application to point-to-point channel capacity (warm up)
- Some mathematical preliminaries
- Proof of the characterization

Later: Move on to multiuser settings.

A 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.

Additive white Gaussian Channel [Shannon '48]

A sender X communicates to receiver Y over a noisy channel $Y = X + Z$.

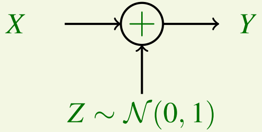

Figure: Additive white Gaussian noise channel

Given a power constraint P on the codebook
The maximum rate that can be reliably transmitted (using blocks)

$$C = \sup_{X:\mathrm{E}(X^2) \leq P} I(X;Y) = \frac{1}{2}\log(1+P).$$

Note: $X \sim \mathcal{N}(0, P)$ is the maximizing distribution.

OUTLINE

Setting:

$$Y = X + Z, \quad Z \sim \mathcal{N}(0,1).$$

Goal: Determine extremal distribution: $\sup_{X:\mathrm{E}(X^2)\leq P} I(X;Y)$.

Let X_* be a maximizing distribution (existence: later)

Let $C = I(X;Y)\big|_{X\sim X_*}$.

X_1, X_2 be two i.i.d. copies of X_*

Take two independent realizations (2-letter) of the channel

$$2C = I(X_1, X_2; Y_1, Y_2).$$

Observation

Two channel realizations

$$Y_1 = X_1 + Z_1$$
$$Y_2 = X_2 + Z_2$$

Define

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

Note: Z_+, Z_- are i.i.d. $\mathcal{N}(0, 1)$.

When X_+, X_- are inputs to the same two-letter channel we get

$$Y_+ = X_+ + Z_+$$
$$Y_- = X_- + Z_-$$

RECALL: SINGLE LETTERIZATION

$$
\begin{aligned}
2C &= I(X_1, X_2; Y_1, Y_2) \\
&= I(X_+, X_-; Y_+, Y_-) && \text{(bijection)} \\
&= I(X_+, X_-; Y_+) + I(X_+, X_-; Y_- | Y_+) \\
&= I(X_+; Y_+) + I(X_-; Y_-) - I(Y_+; Y_-) \\
&\le C + C = 2C
\end{aligned}
$$

Since end-to-end equality holds, we must have

$$
\begin{aligned}
I(X_+; Y_+) &= I(X_-; Y_-) = C \\
I(Y_+; Y_-) &= 0
\end{aligned}
$$

Thus Y_1, Y_2 are independent and $Y_1 + Y_2, Y_1 - Y_2$ are independent.
Hence Y_1, Y_2 are Gaussians, implying X_1, X_2 are Gaussians.

Why is the Gaussian characterization valid?

Mathematical Preliminaries: Characteristic functions

Give a random variable X we have the characteristic function

$$\Phi_X(t) := \mathrm{E}(e^{itX}).$$

Properties of the characteristic function:

- $|\Phi_X(t)| \leq 1 \; \forall t$
- $\Phi_X(t)$ is a continuous function of t (in fact *uniformly continuous*)
- $\Phi_X(t)$ uniquely characterizes the distribution of a random variable (inversion)
- $\Phi_X(-t) = \Phi_X^*(t)$ (conjugate symmetry)

Given: X and Y are independent and $X+Y$ and $X-Y$ are also independent.

Let $\Phi_X(t) = \mathrm{E}(e^{itX})$ and $\Phi_Y(t) = \mathrm{E}(e^{itY})$

$$
\begin{aligned}
\Phi_X(t_1+t_2)\Phi_Y(t_1-t_2) &= \mathrm{E}(e^{i(t_1+t_2)X})\mathrm{E}(e^{i(t_1-t_2)Y}) \\
&= \mathrm{E}(e^{i(t_1+t_2)X+i(t_1-t_2)Y}) \\
&= \mathrm{E}(e^{it_1(X+Y)+it_2(X-Y)}) \\
&= \mathrm{E}(e^{it_1(X+Y)})\mathrm{E}(e^{it_2(X-Y)}) \\
&= \mathrm{E}(e^{it_1X})\mathrm{E}(e^{it_1Y})\mathrm{E}(e^{it_2X})\mathrm{E}(e^{-it_2Y}) \\
&= \Phi_X(t_1)\Phi_X(t_2)\Phi_Y(t_1)\Phi_Y(-t_2) \\
&= \Phi_X(t_1)\Phi_X(t_2)\Phi_Y(t_1)\phi_Y^*(t_2)
\end{aligned}
$$

A functional equation

Given: X and Y are independent and $X+Y$ and $X-Y$ are also independent.

Let $\Phi_X(t) = \mathrm{E}(e^{itX})$ and $\Phi_Y(t) = \mathrm{E}(e^{itY})$

$$\begin{aligned}
\Phi_X(t_1+t_2)\Phi_Y(t_1-t_2) &= \mathrm{E}(e^{i(t_1+t_2)X})\mathrm{E}(e^{i(t_1-t_2)Y}) \\
&= \mathrm{E}(e^{i(t_1+t_2)X+i(t_1-t_2)Y}) \\
&= \mathrm{E}(e^{it_1(X+Y)+it_2(X-Y)}) \\
&= \mathrm{E}(e^{it_1(X+Y)})\mathrm{E}(e^{it_2(X-Y)}) \\
&= \mathrm{E}(e^{it_1X})\mathrm{E}(e^{it_1Y})\mathrm{E}(e^{it_2X})\mathrm{E}(e^{-it_2Y}) \\
&= \Phi_X(t_1)\Phi_X(t_2)\Phi_Y(t_1)\Phi_Y(-t_2) \\
&= \Phi_X(t_1)\Phi_X(t_2)\Phi_Y(t_1)\phi_Y^*(t_2)
\end{aligned}$$

Solution of the functional equation

$$\Phi_X(t_1+t_2)\Phi_Y(t_1-t_2) = \Phi_X(t_1)\Phi_X(t_2)\Phi_Y(t_1)\phi_Y^*(t_2)$$

- $\Phi_X(t) = \phi_X^2(\frac{t}{2})|\Phi_Y(\frac{t}{2})|^2$: set $t_1 = t_2 = \frac{t}{2}$
- $\Phi_Y(t) = |\Phi_X(\frac{t}{2})|^2\phi_Y^2(\frac{t}{2})$: set $t_1 = -t_2 = \frac{t}{2}$

Implies $|\Phi_X(t)| = |\Phi_Y(t)|$, say $f(t)$ $\forall t$

Thus the absolute values satisfy the equation

$$f(t_1+t_2)f(t_1-t_2) = f^2(t_1)f^2(t_2)$$

- $f(2t) = f^4(t)$: set $t_1 = t_2 = t$
- Induction: $f(kt) = f(t)^{k^2}, k \in \mathbb{N}$ (exercise)
- $f(qt) = f(t)^{q^2}, q \in \mathbb{Q}$
- $f(t) = f(1)^{t^2} = e^{-ct^2}, t \in \mathbb{R}$ (by continuity of $\Phi_X(t)$ in t)

Solution of the functional equation

$$\Phi_X(t_1+t_2)\Phi_Y(t_1-t_2) = \Phi_X(t_1)\Phi_X(t_2)\Phi_Y(t_1)\phi_Y^*(t_2)$$

- $\Phi_X(t) = \phi_X^2(\frac{t}{2})|\Phi_Y(\frac{t}{2})|^2 :$ set $t_1 = t_2 = \frac{t}{2}$
- $\Phi_Y(t) = |\Phi_X(\frac{t}{2})|^2\phi_Y^2(\frac{t}{2}) :$ set $t_1 = -t_2 = \frac{t}{2}$

Implies $|\Phi_X(t)| = |\Phi_Y(t)|$, say $f(t)\ \forall t$

Thus the absolute values satisfy the equation

$$f(t_1+t_2)f(t_1-t_2) = f^2(t_1)f^2(t_2)$$

- $f(2t) = f^4(t) :$ set $t_1 = t_2 = t$
- Induction: $f(kt) = f(t)^{k^2}, k \in \mathbb{N}$ (exercise)
- $f(qt) = f(t)^{q^2}, q \in \mathbb{Q}$
- $f(t) = f(1)^{t^2} = e^{-ct^2}, t \in \mathbb{R}$ (by continuity of $\Phi_X(t)$in t)

Solution of the functional equation

Let $\Phi_X(t) = e^{-ct^2} e^{i\theta_1(t)}, \Phi_Y(t) = e^{-ct^2} e^{i\theta_2(t)}$

Phases (recall: odd functions) satisfies the recursion

$$\theta_1(t_1 + t_2) + \theta_2(t_1 - t_2) = \theta_1(t_1) + \theta_1(t_2) + \theta_2(t_1) - \theta_2(t_2)$$

Exercise: Show that $\theta_1(t) = t\theta_1(1)$ and $\theta_2(t) = t\theta_2(1)$

Thus

$$\Phi_X(t) = e^{-ct^2 + iat}, \Phi_Y(t) = e^{-ct^2 + ibt}$$

In other words, X and Y are Gaussians with same variance. Q.E.D.

Solution of the functional equation

Let $\Phi_X(t) = e^{-ct^2} e^{i\theta_1(t)}, \Phi_Y(t) = e^{-ct^2} e^{i\theta_2(t)}$

Phases (recall: odd functions) satisfies the recursion

$$\theta_1(t_1 + t_2) + \theta_2(t_1 - t_2) = \theta_1(t_1) + \theta_1(t_2) + \theta_2(t_1) - \theta_2(t_2)$$

Exercise: Show that $\theta_1(t) = t\theta_1(1)$ and $\theta_2(t) = t\theta_2(1)$

Thus

$$\Phi_X(t) = e^{-ct^2 + iat}, \Phi_Y(t) = e^{-ct^2 + ibt}$$

In other words, X and Y are Gaussians with same variance. Q.E.D.

Solution of the functional equation

Let $\Phi_X(t) = e^{-ct^2} e^{i\theta_1(t)}, \Phi_Y(t) = e^{-ct^2} e^{i\theta_2(t)}$

Phases (recall: odd functions) satisfies the recursion

$$\theta_1(t_1 + t_2) + \theta_2(t_1 - t_2) = \theta_1(t_1) + \theta_1(t_2) + \theta_2(t_1) - \theta_2(t_2)$$

Exercise: Show that $\theta_1(t) = t\theta_1(1)$ and $\theta_2(t) = t\theta_2(1)$

Thus

$$\Phi_X(t) = e^{-ct^2 + iat}, \Phi_Y(t) = e^{-ct^2 + ibt}$$

In other words, X and Y are Gaussians with same variance. Q.E.D.

Existence of maximizer (standard analysis)

Line of reasoning:

- A sequence of variables $\{X_n\}$ whose value $I(X_n; Y)$ converges to the supremum
- Implies (since their second moment is bounded) a subsequence $\{X_{n_i}\}$ that converges weakly. (Prokhorov)
- The densities of induced $\{Y_{n_i}\}$ converge pointwise (use the additive Gaussian noise and other standard results)
- Entropies $h(Y_{n_i})$ converge
  - densities are uniformly bounded (Gaussian noise)

$$\liminf_i h(Y_{n_i}) \geq h(Y_*)$$

  - uniformly bounded κ-moment ($\kappa > 1$)

$$\limsup_i h(Y_{n_i}) \leq h(Y_*)$$

A SIMPLE EXERCISE

Exercise

Let $Y_1 = X_1 + Z_1$ and $Y_2 = X_2 + Z_2$, where $Z_1 \sim \mathcal{N}(0, N_1),\ Z_2 \sim \mathcal{N}(0, N_2)$ are independent Gaussians and independent of (X_1, X_2).

1. If Y_1 and Y_2 are independent, then show that X_1 and X_2 are also independent.
2. If Y_1 and Y_2 are independent Gaussians, then show that X_1 and X_2 are also independent Gaussians.

Hint: Use characteristic functions.

What we did

- Showed that for the 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

- Seems more complicated than traditional technique.

In network information theory, the presented technique (program) does not get any more complicated

Traditional Techniques: Become very involved and sometimes does not work

Recap: Point-to-point channel

What we did

- Showed that for the 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

- Seems more complicated than traditional technique.

In network information theory, the presented technique (program) does not get any more complicated

Traditional Techniques: Become very involved and sometimes does not work

What we did

- Showed that for the 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
- Seems more complicated than traditional technique.

In network information theory, the presented technique (program) does not get any more complicated

Traditional Techniques: Become very involved and sometimes does not work

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]

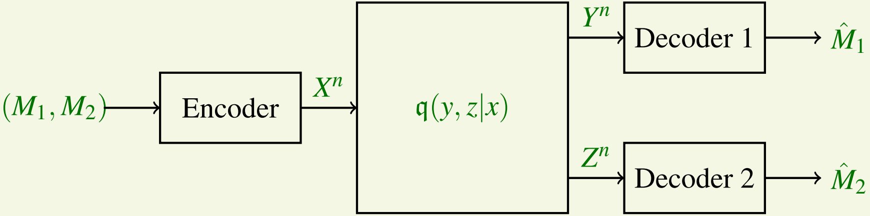

Figure: Discrete memoryless broadcast channel

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

Open setting 2: Interference channels

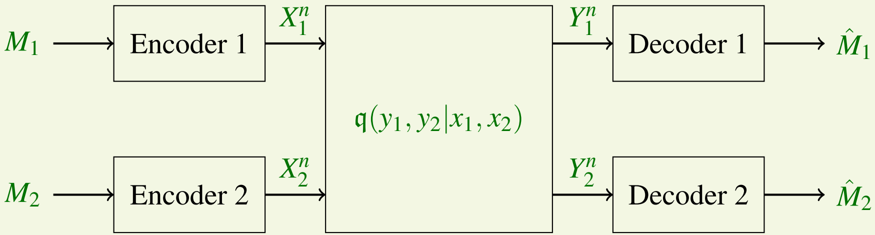

Figure: Discrete memoryless interference channel

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

An observation

For these two problems

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

An observation

For these two problems

- there are achievable regions (one for each) whose optimality or sub-optimality has 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

For these two problems

- there are achievable regions (one for each) whose optimality or sub-optimality has 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

For these two problems

- there are achievable regions (one for each) whose optimality or sub-optimality has 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

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

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

ASIDE: A NEW RESULT

Han-Kobayashi(HK) achievable region for the interference channel is strictly sub-optimal (‘2015)

- Idea: Showing 1-letter $\neq$ 2-letter for a class of channels
- Ingenuity: coming up with the class of channels for which HK-region is computable and sub-optimal.

Illustration of the power of computation and development of theoretical results that made such computation feasible.

Aside: A new result

Han-Kobayashi(HK) achievable region for the interference channel is strictly sub-optimal ('2015)

- Idea: Showing 1-letter $\neq$ 2-letter for a class of channels
- Ingenuity: coming up with the class of channels for which HK-region is computable and sub-optimal.

Illustration of the power of computation and development of theoretical results that made such computation feasible.

BROADCAST CHANNELS: RESULTS

The capacity region of the broadcast channel is unknown

However, for special classes of channel it is known

- Degraded (Gallager '74, Bergmans '73)
- Less noisy (Korner-Marton '75)
- More capable (El Gamal '78)
- Essentially less noisy and essentially more capable (Nair '10)
- Product of reversely degraded broadcast channel (El Gamal '81)
- Several classes of product broadcast channels (Geng-Gohari-Nair-Yu '14)
- Vector Gaussian broadcast channel with private messages (Weingarten-Steinberg-Shamai '06)
- Vector Gaussian broadcast channel with private and common messages (Geng-Nair '14)

Degraded broadcast channel

Channel setting: $X \to Y \to Z$ is Markov

Capacity Region (Gallager '74)

The union of rate pairs (R_1, R_2) satisfying

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

over $U \to X \to Y \to Z$ forms the capacity region. Further it suffices to consider $|U| \leq |X| + 1$.

Note: The freedom is in the choice of $p(u,x)$

Remarks

- Achievability: superposition coding (not focus of this talk)
- The capacity region is convex
- Can be characterized by intersection of supporting hyperplanes

Degraded broadcast channel

Channel setting: $X \to Y \to Z$ is Markov

Capacity Region (Gallager '74)

The union of rate pairs (R_1, R_2) satisfying

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

over $U \to X \to Y \to Z$ forms the capacity region. Further it suffices to consider $|U| \le |X| + 1$.

Note: The freedom is in the choice of $p(u,x)$

Remarks

- Achievability: superposition coding (not focus of this talk)
- The capacity region is convex
- Can be characterized by intersection of supporting hyperplanes

Alternate characterization

Obvious boundaries of capacity region

$$0 \leq R_1 \leq C_1, \quad 0 \leq R_2 \leq C_2.$$

Interested in characterizing supporting hyperplanes of the form

$$R_1 + \lambda R_2, \quad \lambda \geq 0.$$

Alternate expression of capacity region

For any $\lambda \geq 0$

$$\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \lambda R_2 = \max_{p(v,x)} I(X;Y|V) + \lambda I(V;Z).$$

Suffices to consider $|V| \leq |X|$.

CONVERSE:

From Fano's inequality (ignoring ϵ terms)

$$
\begin{aligned}
R_2 &\leq \frac{1}{n} I(M_2; Z^n) \\
R_1 &\leq \frac{1}{n} I(M_1; Y^n | M_2) \leq \frac{1}{n} I(X^n; Y^n | M_2)
\end{aligned}
$$

The second inequality follows from: $(M_1, M_2) \to X^n \to Y^n$ being Markov (data-processing)

Hence

$$
\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \lambda R_2 = \frac{1}{n}\left(\max_{p(v,x^n)} I(X^n; Y^n | V) + \lambda I(V; Z^n) \right).
$$

Set $V = M_2$.

Next: Gallager's single-letterization argument

Converse:

From Fano's inequality (ignoring ϵ terms)

$$
\begin{aligned}
R_2 &\le \frac{1}{n} I(M_2; Z^n) \\
R_1 &\le \frac{1}{n} I(M_1; Y^n | M_2) \le \frac{1}{n} I(X^n; Y^n | M_2)
\end{aligned}
$$

The second inequality follows from: $(M_1, M_2) \to X^n \to Y^n$ being Markov (data-processing)

Hence

$$\max_{(R_1, R_2) \in \mathcal{C}} R_1 + \lambda R_2 = \frac{1}{n} \left(\max_{p(v, x^n)} I(X^n; Y^n | V) + \lambda I(V; Z^n) \right).$$

Set $V = M_2$.

Next: Gallager's single-letterization argument

Goal: Show that

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

Proof: Observe that

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

Goal: Show that

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

Proof: Observe that

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

Gaussian degraded broadcast channel

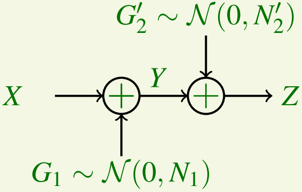

Figure: Degraded Gaussian broadcast channel

The capacity region is the union of rate pairs satisfying:

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

for some $V \to X \to (Y,Z)$ such that $\mathrm{E}(X^2) \leq P$.

[Bergmans '73]: Suffices to consider $X = U + V$,
$U \sim N(0, \alpha P), V \sim N(0, \bar{\alpha} P), U \perp V$

GAUSSIAN DEGRADED BROADCAST CHANNEL

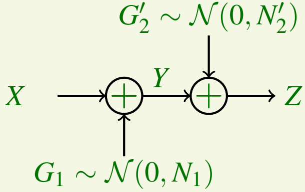

Figure: Degraded Gaussian broadcast channel

The capacity region is the union of rate pairs satisfying:

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

for some $V \to X \to (Y,Z)$ such that $\mathrm{E}(X^2) \leq P$.

[Bergmans '73]: Suffices to consider $X = U + V$,
$U \sim \mathcal{N}(0, \alpha P), V \sim \mathcal{N}(0, \bar{\alpha}P), U \perp V$

PROOF OF BERGMAN'S RESULT VIA GALLAGER'S SINGLE-LETTERIZATION

$$D = \sup_{X:\mathrm{E}(X^2)\leq P} I(X;Y|V) + \lambda I(V;Z)$$

and let (V_*, X_*) be a maximizer.

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

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

$$\begin{aligned}
2D &= I(X_+, X_-; Y_+, Y_-|V_1, V_2) + \lambda I(V_1, V_2; Z_+, Z_-) \\
&= I(X_+; Y_+|) + I(X_-; Y_-|V_1, V_2, Z_+) - I(Y_+; Y_-|V_1, V_2, Z_+) \\
&\quad + \lambda I(V_1, V_2; Z_+) + \lambda I(V_1, V_2, Z_+; Z_-) - \lambda I(Z_+; Z_-)
\end{aligned}$$

Implies

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

Thus Z_1, Z_2 are Gaussians and hence Y_1, Y_2 and X_1, X_2 are Gaussians. Since (check)

$$(V_1, V_2, Y_-) \to Y_+ \to Z_+$$

we have (by data-processing inequality)

$$I(Y_+; Y_-|V_1, V_2) = I(Z_+; Y_-|V_1, V_2)$$

If $Y_- - Y_+ - Z_+$ and $Y_- - Z_+ - Y_+$ is Markov (have positive joint density function) then Y_- is independent of (Y_+, Z_+) (conditioned on (V_1, V_2))

From Single-letterization

$$
\begin{aligned}
2D &= I(X_+, X_-; Y_+, Y_- | V_1, V_2) + \lambda I(V_1, V_2; Z_+, Z_-) \\
&= I(X_+; Y_+ |) + I(X_-; Y_- | V_1, V_2, Z_+) - I(Y_+; Y_- | V_1, V_2, Z_+) \\
&\quad + \lambda I(V_1, V_2; Z_+) + \lambda I(V_1, V_2, Z_+; Z_-) - \lambda I(Z_+; Z_-)
\end{aligned}
$$

Implies

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

Thus Z_1, Z_2 are Gaussians and hence Y_1, Y_2 and X_1, X_2 are Gaussians. Since (check)

$$(V_1, V_2, Y_-) \to Y_+ \to Z_+$$

we have (by data-processing inequality)

$$I(Y_+; Y_- | V_1, V_2) = I(Z_+; Y_- | V_1, V_2)$$

If $Y_- - Y_+ - Z_+$ and $Y_- - Z_+ - Y_+$ is Markov (have positive joint density function) then Y_- is independent of (Y_+, Z_+) (conditioned on (V_1, V_2))

From Single-letterization

$$
\begin{aligned}
2D &= I(X_+, X_-; Y_+, Y_- | V_1, V_2) + \lambda I(V_1, V_2; Z_+, Z_-) \\
&= I(X_+; Y_+|) + I(X_-; Y_- | V_1, V_2, Z_+) - I(Y_+; Y_- | V_1, V_2, Z_+) \\
&\quad + \lambda I(V_1, V_2; Z_+) + \lambda I(V_1, V_2, Z_+; Z_-) - \lambda I(Z_+; Z_-)
\end{aligned}
$$

Implies

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

Thus Z_1, Z_2 are Gaussians and hence Y_1, Y_2 and X_1, X_2 are Gaussians. Since (check)

$$(V_1, V_2, Y_-) \to Y_+ \to Z_+$$

we have (by data-processing inequality)

$$I(Y_+; Y_- | V_1, V_2) = I(Z_+; Y_- | V_1, V_2)$$

If $Y_- - Y_+ - Z_+$ and $Y_- - Z_+ - Y_+$ is Markov (have positive joint density function) then Y_- is independent of (Y_+, Z_+) (conditioned on (V_1, V_2))

From Single-letterization

$$\begin{aligned}2D &= I(X_+, X_-; Y_+, Y_- | V_1, V_2) + \lambda I(V_1, V_2; Z_+, Z_-)\\ &= I(X_+; Y_+|) + I(X_-; Y_-|V_1, V_2, Z_+) - I(Y_+; Y_-|V_1, V_2, Z_+)\\ &\quad + \lambda I(V_1, V_2; Z_+) + \lambda I(V_1, V_2, Z_+; Z_-) - \lambda I(Z_+; Z_-)\end{aligned}$$

Implies

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

Thus Z_1, Z_2 are Gaussians and hence Y_1, Y_2 and X_1, X_2 are Gaussians. Since (check)

$$(V_1, V_2, Y_-) \to Y_+ \to Z_+$$

we have (by data-processing inequality)

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

If $Y_- - Y_+ - Z_+$ and $Y_- - Z_+ - Y_+$ is Markov (have positive joint density function) then Y_- is independent of (Y_+, Z_+) (conditioned on (V_1, V_2))

Gaussian optimality

Since $I(Z_+; Y_- | V_1, V_2) = 0$, from the simple exercise we showed earlier

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

Let $X^v \sim X | \{V = v\}$ denote the conditional distribution.

Hence for every pair v_1, v_2, we have $X_1^{v_1}, X_2^{v_2}$ is independent and $X_1^{v_1} + X_2^{v_2}$ and $X_1^{v_1} - X_2^{v_2}$ are independent.

Hence $X_v \sim \mathcal{N}(\mu_v, P')$ with $P' \leq Q$; thus the conditional distribution of the maximizer is Gaussian.

Exercise: Reason why the above statement establishes Bergman's result.

Gaussian optimality

Since $I(Z_+; Y_-|V_1, V_2) = 0$, from the simple exercise we showed earlier

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

Let $X^v \sim X|\{V = v\}$ denote the conditional distribution.

Hence for every pair v_1, v_2, we have $X_1^{v_1}, X_2^{v_2}$ is independent and $X_1^{v_1} + X_2^{v_2}$ and $X_1^{v_1} - X_2^{v_2}$ are independent.

Hence $X_v \sim \mathcal{N}(\mu_v, P')$ with $P' \leq Q$; thus the conditional distribution of the maximizer is Gaussian.

Exercise: Reason why the above statement establishes Bergman's result.

REMARKS

Traditional Argument

- Uses (weak) Entropy Power Inequality (**)
- Works in n-letter forms (need not use single-letterization)

Questions about content so far

Next:

- 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

Questions about content so far

Next:

- 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