1 Review

1.1 Gaussian Channel

\[ Y = X + Z \]

where

\[ Z \sim \mathcal{N}(0, \sigma^2) \]

1.2 Average Power Constraint

Codewords \( \underline{x} \) must have average power \( \leq P \), i.e.

\[ \frac{1}{n} \sum_{i=1}^{n} x_i^2 \leq P \]

1.3 Capacity

\[ C_p(W) = \frac{1}{2} \log \left( 1 + \frac{P}{\sigma^2} \right) \]

1.4 Achievability

- Random code selection was used
- Select Codewords as rows of the following matrix

\[
\begin{bmatrix}
X_{11} & \cdots & X_{1n} \\
X_{21} & \cdots & X_{2n} \\
\vdots & \ddots & \vdots \\
X_{N1} & \cdots & X_{Nn}
\end{bmatrix}
\]

where \( X_{ij} \) are iid \( \sim \mathcal{N}(0, P - \delta) \)
1.5 Decoding
Consider $A_\delta$ (Typical set)

$$A_\delta = \left\{ (x, y) \mid \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{g(y_i - x_i, \sigma^2)}{g(y_i, P - \delta + \sigma^2)} \right) > \mathbb{E} \log \left( \frac{g(Y - X, \sigma^2)}{g(Y, P - \delta + \sigma^2)} \right) - \delta \right\}$$

Decoding is done as follows:

- $\phi(y) = i$, if $i$ is the unique index such that $(x_i, y) \in A_\delta$
- Declare error if no, or more than one $i$ are found

This “code” achieves rate $\frac{1}{2} \log \left( 1 + \frac{P - \delta}{\sigma^2} \right) - \delta$ where $\delta > 0$ is arbitrary

$\Rightarrow \frac{1}{2} \log \left( 1 + \frac{P}{\sigma^2} \right)$ is an achievable rate.

2 Converse proof

2.1 Converse proof using sphere packing argument

- **Step1 Estimate** $\text{Vol}\left( \bigcup_{i=1}^{N} D'_i \right)$
  
  We need to find an upper bound over the union of the sets $D'_i$. Since we have no control over $D_i$ (actual decoding sets), we estimate the volume using $D'_i$ which is defined as

  $$D'_i = D_i \cap T_\delta$$

  We need to figure out what $T_\delta$ is. For discrete case this was Typical set.

- **Step2 Estimate** $\text{Vol}\left( D'_i \right)$
  
  Here we need to find a lower bound on volume of set $D_i$

Consider,

$$T_\delta^{(1)} = \left\{ y \in \mathbb{R}^n \mid \frac{1}{n} \| y \|_2^2 \leq P + \delta + \sigma^2 \right\}$$

(1)

where

$$\| y \|_2^2 = \sum_{i=1}^{n} y_i^2$$

$$T_\delta^{(2)} = \left\{ z \in \mathbb{R}^n \mid \frac{1}{n} \| z \|_2^2 \geq \sigma^2(1 - \delta) \right\}$$

(2)
• **Comments**

1. For $Z_1, \ldots, Z_n \text{iid} \sim \mathcal{N}(0, \sigma^2)$

   \[
   \Pr(z \in T^{(2)}_\delta) \to 1 \text{ as } n \to \infty \text{ (by Weak law of large numbers)}
   \]

2. Suppose $Y_i = X_i + Z_i$ where $Z_1, \ldots, Z_n \text{iid} \sim \mathcal{N}(0, \sigma^2)$ . Then,

   \[
   \frac{1}{n} \sum_{i=1}^{n} EY_i^2 = E\left(\frac{1}{n} \| y \|^2 \right) = \frac{1}{n} \sum_{i=1}^{n} (x_i^2 + \sigma^2)
   \]

   \[
   = \frac{1}{n} \| x \|^2 + \sigma^2
   \]

   for $x$ such that $\frac{1}{n} \| x \|^2 \leq P$

   \[
   \therefore E\left(\frac{1}{n} \| y \|^2 \right) \leq P + \sigma^2
   \]

   So, by Weak Law of Large Numbers

   \[
   \Pr(y \in T^{(1)}_\delta) \to 1 \text{ as } n \to \infty
   \]

• **Proof**: Consider a code $(x_i, D_i)_{i=1}^{N}$ (codewords and decoding set) with maximum probability of error $\leq \epsilon$

   \[
   \implies W^n(D_i \mid x_i) \geq 1 - \epsilon, \ 1 \leq i \leq N
   \]

   and

   \[
   \frac{1}{n} \| x \|^2 \leq P
   \]

   Note that by Comments 1 and 2 discussed above

   1. $\Pr(y \in T^{(1)}_\delta \mid x_i \text{ is sent }) \to 1 \text{ as } n \to \infty$

   2. $\Pr(y - x_i \in T^{(2)}_\delta \mid x_i \text{ is sent }) = \Pr(z \in T^{(2)}_\delta) \to 1 \text{ as } n \to \infty$

   Combining these two, let $T_\delta(x_i) = T^{(1)}_\delta \cap \{x_i \mid T^{(2)}_\delta\}$

   \[
   \Pr(y \in T_\delta(x_i) \mid x_i \text{ was sent }) = W^n(T_\delta(x_i) \mid x_i) \to 1 \text{ as } n \to \infty
   \]
Define $D'_i = D_i \cap T_\delta(x_i)$

Then,

$$W^n(D'_i \mid x_i) \geq \frac{1 - \epsilon}{2}$$

for $n$ large

$$\therefore \frac{1 - \epsilon}{2} \leq \frac{1}{N} \sum_{i=1}^{N} W^n(D'_i \mid x_i) = \frac{1}{N} W^n \left( \bigcup_{i=1}^{N} D'_i \mid x_i \right)$$

Since $D_i$ are disjoint $D'_i$ are also disjoint ($\because$ subsets of disjoint sets are disjoint)

Note that

$$T_\delta(x_i) \subseteq T_\delta^{(1)}$$

$$\implies D'_i \subseteq T_\delta(x_i) \subseteq T_\delta^{(1)} \ \forall \ i$$

$$\therefore \bigcup_{i=1}^{N} D'_i \subseteq T_\delta^{(1)}$$

$$\implies Vol \left( \bigcup_{i=1}^{N} D'_i \right) \leq Vol \left( T_\delta^{(1)} \right) \leq 2 \left[ (2\pi e)^{n/2} \frac{(P + \sigma^2 + \delta)^{n/2}}{\sqrt{n\pi}} \right]$$

Also,

$$\frac{1 - \epsilon}{2} \leq W^n(D'_i \mid x_i)$$

$$= \int_{D'_i} g^n(y, x_i, \sigma^2) \ dy$$

$$= \int_{D'_i} e^{-||y-x_i||^2/2\sigma^2} \left( \frac{2\pi\sigma^2}{n/2} \right) \ dy$$

$$\leq \int_{D'_i} e^{-n(1-\delta)/2} \left( \frac{2\pi\sigma^2}{n/2} \right) \ dy$$

$$= \frac{e^{-n(1-\delta)/2}}{Vol(D'_i)}$$

$$\left( \because \text{ in } T_\delta^{(2)} \frac{1}{\pi} \ || z ||^2 \geq \sigma^2(1-\delta) \right)$$

$$\therefore Vol(D'_i) \geq \left( \frac{1 - \epsilon}{2} \right) e^{-n(1-\delta)/2} \left( \frac{2\pi\sigma^2}{n/2} \right)$$
Therefore,

\[
N \leq \frac{\text{Vol} \left( \bigcup_{i=1}^{N} D'_i \right)}{\text{Vol}(D'_i)} \leq \frac{4(P + \sigma^2 + \delta)^{n/2}}{(1 - \epsilon)(\sigma^2)^{n/2}e^{-nd/2}\sqrt{n\pi}}
\]

\[
R = \frac{1}{n} \log N \leq \frac{1}{2} \log(1 + (P + \delta)/\sigma^2) + \frac{1}{n} \log \frac{4}{1 - \epsilon} + \frac{\delta}{2} + \frac{1}{2n} \log n\pi
\]

\[
\Rightarrow R \leq \frac{1}{2} \log(1 + (P + \delta)/\sigma^2) + \delta, \quad \forall \ n \text{ large}
\]

Geometrically typical sets under Gaussian distribution look like spheres and power constraint is also a sphere. That is why most of the terms cancel up.

### 2.2 Converse proof 2 (using Fano’s inequality)

Before going to the proof we need to answer the following questions.

1. What do we mean by “mutual information” for continuous random variable?

2. Why do we care?

   - Second question’s answer would be to bound size of large probability sets similar to discrete case.

   - To answer first question consider,

\[
\mathbb{E} \log \left[ \frac{g(Y - X, \sigma^2)}{g(Y, P + \sigma^2)} \right] = \mathbb{E} \log \left[ \frac{f_{Y|X}(y|x)}{f_Y(y)} \right] = \mathbb{E} \log \left[ \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} \right]
\]

In Discrete case,

\[
\mathbb{E} \log \left[ \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \right] = I(X \land Y)
\]

i.e. we need to find a similar definition as in discrete case.

**A general definition of mutual information** (valid for continuous as well as discrete case)

Given rvs X, Y taking values in \(\mathcal{X}\) and \(\mathcal{Y}\) respectively.

Let \(\mathcal{F}(\mathcal{X}) = \{\text{set of all finite valued functions on } \mathcal{X}\}\)

and \(\mathcal{F}(\mathcal{Y}) = \{\text{set of all finite valued functions on } \mathcal{Y}\}\)

\[
I(X \land Y) \overset{\text{def}}{=} \sup_{f \in \mathcal{F}(\mathcal{X}), \ g \in \mathcal{F}(\mathcal{Y})} I(f(X) \land g(Y))
\]

\[
\Pr(f(x) = i, g(y) = j) = \int_{f^{-1}(i) \times g^{-1}(j)} p_{xy} \, dx \, dy
\]
Theorem 1. (Gelfand, Yaglom, Perez)
Suppose $P_{XY}$ is absolutely continuous wrt $P_x P_y$, i.e. if $P_x P_y(S) = 0 \implies P_{XY}(S) = 0$.
Then,

$$I(X \wedge Y) = \mathbb{E} \log \left[ \frac{dP_{XY}}{dP_X P_Y} \right]$$

Corollary 1. For rvs $X, Y$ with a joint density $f_{XY}$,

$$I(X \wedge Y) = \int_{X \times Y} f_{XY}(x, y) \log \left[ \frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} \right] dxdy$$

Corollary 2. For random variables $X, Y$ such that $X$ is discrete and $Y$ given $X$ has a density $f_{Y|X}$,

$$I(X \wedge Y) = \mathbb{E} \log \left[ \frac{f_{Y|X}(y|x)}{\sum_x p_X(x) f_{Y|X}(y|x)} \right] = \mathbb{E} \log \left[ \frac{p_{X|Y}(x|y)}{p_X(x)} \right]$$

$$= \mathbb{E} \log \left[ \frac{p_X(x) f_{Y|X}(y|x)}{\sum_x p_X(x) f_{Y|X}(y|x)} \right] \text{ (by Bayes' rule)}$$

All expectations are with respect to joint distribution between $X$ and $Y$.

Exercise: Find $I(X \wedge Y)$, where

- $X = 0, Y = 0$ and
- $X = 1, Y \sim \mathcal{N}(0, \sigma^2)$

We need some definitions to do converse proof 2 (using Fano’s inequality)

2.3 Definitions

- **Differential Entropy**: For a random variable $X$ with density $f_X$, differential entropy of $X$, $h(X)$, is defined as

$$h(X) = \int f_X(x) \log \frac{1}{f_X(x)} dx$$

$$= \mathbb{E} \left( \log \frac{1}{f_X(x)} \right)$$
• **Joint Differential Entropy**: For rvs $X$, $Y$ with joint density $f_{XY}$, joint differential entropy, $h(X, Y)$, is defined as

$$h(X, Y) = \int \int f_{XY}(x, y) \log \frac{1}{f_{XY}(x, y)} \, dx \, dy$$

• **Conditional Differential Entropy**: For rvs $X$, $Y$ with joint density $f_{XY}$, conditional differential entropy, $h(X \mid Y)$, is defined as

$$h(X \mid Y) = \int \int f_{XY}(x, y) \log \frac{1}{f_{X|Y}(x \mid y)} \, dx \, dy$$

$$\implies h(X, Y) = h(Y) + h(X \mid Y)$$

**Example:**

$X \sim \mathcal{N}(0, \sigma^2)$

$$h(X) = \mathbb{E}\left(\log \frac{1}{f_X(x)}\right) = \mathbb{E}\left(\log(\sqrt{2\pi\sigma^2}e^{-x^2/2\sigma^2})\right)$$

$$= \frac{1}{2}\log(2\pi\sigma^2) + \log e^{\mathbb{E}(X^2)} \frac{1}{2\sigma^2}$$

$$= \frac{1}{2}\log(2\pi e\sigma^2)$$

**Remark**:

1. $h(X)$ can be negative! Therefore we have to carefully handle this quantity.
2. $I(X \land Y) \neq h(X)$. Since $X, X$ donot have joint density even though $X$ has a density.

**Example:**

$X \sim \mathcal{U}(0, a)$

$$h(X) = \mathbb{E}\left(\log \frac{1}{f_X(x)}\right) = \int f_X(x) \log \frac{1}{f_X(x)} \, dx$$

$$= \int_0^a \frac{1}{a} \log adx$$

$$= \log a$$

2.4 **Properties**

1. • **Case (1)**: Joint density $f_{XY}$ exists

$$I(X \land Y) = h(X) - h(X \mid Y)$$

$$= h(Y) - h(Y \mid X)$$
• **Case (2):** X is discrete and Y given X has density $f_{Y|X}$

$$I(X \land Y) = h(Y) - h(Y \mid X)$$

$$= H(X) - H(X \mid Y)$$

$$H(X \mid Y) = \int H(X \mid Y = y)f_Y(y)dy$$

2. **Chain rule :**

$$h(X, Y) = h(X) + h(Y \mid X)$$

$$= h(Y) + h(X \mid Y)$$

3. **Data Processing Inequality :** Suppose $f \in J(X)$, $g \in J(Y)$

$$I(f(X) \land g(Y)) \leq I(f(X) \land Y) \leq I(X \land Y)$$

4. **Non Negetivity of Mutual information :**

$$I(X \land Y) \geq 0$$

since $I(f(X) \land g(Y)) \geq 0 \ \forall \ f \in J(X), \ g \in J(Y)$

Equality holds iff

$$I(f(X) \land g(Y)) = 0$$

$$\iff f(X) \parallel g(Y)$$

$$\iff X \parallel Y$$

• **Corollary**

  – For Case (1)

  $$h(X) \geq h(X \mid Y)$$

  – For Case (2)

  $$H(X) \geq H(X \mid Y)$$

8