EE 231A Information Theory
Lecture 11
Gaussian Channel Capacity

A. Gaussian distribution has maximum entropy.
B. Maximum Gaussian Channel Mutual Information
C. Sphere packing argument
D. Joint AEP for Continuous R.V.'s
E. Gaussian Channel Coding Theorem

Part 11 A: Gaussian distribution has maximum entropy.

Gaussian has Maximum Entropy

- Let X be any random variable with
  $E[X]=0 \quad E\left[X^2\right]=\sigma^2.$
- The maximum value of $h(X)$ is $\frac{1}{2}\log\left(2\pi e\sigma^2\right).$
- And is achieved only if X is normal (Gaussian).

Proof that Gaussian Maximizes Entropy

$$\text{Let } \phi(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}.$$

$$E[X]=0 \qquad E\left[X^2\right]=\sigma^2.$$

Let $g(x)$ be any other distribution with

$$E[X]=0 \qquad E\left[X^2\right]=\sigma^2.$$

$$
\begin{aligned}
0 &\le D(g\|\phi) \\
&= \int g(x)\ln\frac{g(x)}{\phi(x)}dx \\
&= \int g(x)\ln g(x)dx - \int g(x)\ln\phi(x)dx \\
&= \int g(x)\ln g(x)dx + \int g(x)\left[\frac{1}{2}\ln\left(2\pi\sigma^2\right)+\frac{x^2}{2\sigma^2}\right]dx \\
&= \int g(x)\ln g(x)dx + \int \phi(x)\left[\frac{1}{2}\ln\left(2\pi\sigma^2\right)+\frac{x^2}{2\sigma^2}\right]dx \\
&= \int g(x)\ln g(x)dx - \int \phi(x)\ln\phi(x)\,dx \\
&= -h(g)+h(\phi) \qquad \Rightarrow \qquad \boxed{h(\phi)\ge h(g)}
\end{aligned}
$$

Same argument works for multivariate Gaussian.

- Let X^n be any n-dimensional random variable with

$$
E\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mu = 0 \qquad E\left[\begin{bmatrix} x_1-\mu_1 \\ \vdots \\ x_n-\mu_n \end{bmatrix}\begin{bmatrix} x_1-\mu_1 & \cdots & x_n-\mu_n \end{bmatrix}\right] = K
$$

- The maximum value of $h(X^n)$ is $\frac{1}{2}\log(2\pi e)^n|K|$ bits,
- And is achieved only if X^n is jointly Gaussian.

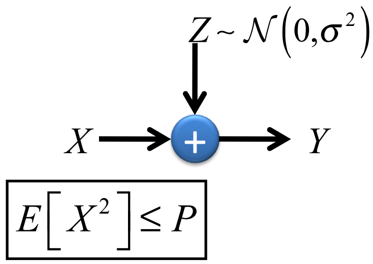
$$\phi(\bar{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)}$$

$g(\bar{x})$: any distribution with $\mu = 0,\ E\left[XX^T\right] = K$

$$
\begin{aligned}
0 &\le D(g \| \phi) \\
&= \int g(\bar{x}) \log \frac{g(\bar{x})}{\phi(\bar{x})} dx_1 \dots dx_n \\
&= \int g(\bar{x}) \log g(\bar{x}) dx_1 \dots dx_n - \int g(\bar{x}) \log \phi(\bar{x}) dx_1 \dots dx_n \\
&= \int g(\bar{x}) \log g(\bar{x}) dx_1 \dots dx_n - \int \phi(\bar{x}) \log \phi(\bar{x}) dx_1 \dots dx_n \\
&= -h(g) + h(\phi) \qquad \Rightarrow \qquad \boxed{h(\phi) \ge h(g)}
\end{aligned}
$$

Part 11 B:
Maximizing Gaussian Channel Mutual Information

Gaussian Channel Mutual Information



$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Y \mid X) \\
&= h(Y) - h(Y - X \mid X) \\
&= h(Y) - h(Z \mid X) \\
&= h(Y) - h(Z)
\end{aligned}
$$

- $h(Z)$ is $\frac{1}{2}\log 2\pi e\sigma^2$ regardless of choice of X.

Maximizing $I(X;Y) = h(Y) - h(Z)$

- To maximize $I(X;Y)$, maximize $h(Y)$

$$h(Y) \le \frac{1}{2}\log 2\pi e(E[Y^2])$$

$$
\begin{aligned}
EY^2 &= E(X+Z)^2 \\
&= EX^2 + 2\cancelto{0}{EXEZ} + EZ^2 \\
&= EX^2 + EZ^2 \\
&\le P + N
\end{aligned}
$$

- $h(Y) \le \frac{1}{2}\log 2\pi e(P+N)$ with equality for $Y \sim N(0, P+N)$

Gaussian channel capacity

$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Z) \\
&\leq \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi eN \\
&= \frac{1}{2}\log \frac{2\pi e(P+N)}{2\pi eN} \\
&= \frac{1}{2}\log(1+\frac{P}{N}) \\
&= \frac{1}{2}\log(1+SNR) \\
&= C \qquad \text{achieved with } X \sim N(0,P).
\end{aligned}
$$

We need a coding theorem for $C = \frac{1}{2}\log(1+SNR)$.

- C is the maximum mutual information for AWGN channel with $Z \sim \mathcal{N}(0,N)$ and power constraint P.
- Is it the largest achievable rate?
- We have not proven that yet, since our previous proof in Chapter 7 was for DISCRETE memoryless channels.

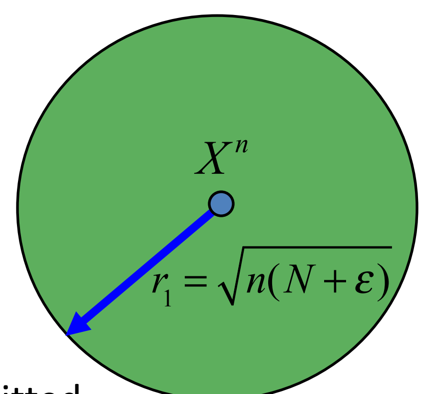
Part 11 C:
Sphere packing argument

Sphere of received Y^n's given X^n

$$Y^n = X^n + Z^n$$

$$\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \rightarrow EZ_i^2 \quad \text{in probability}$$

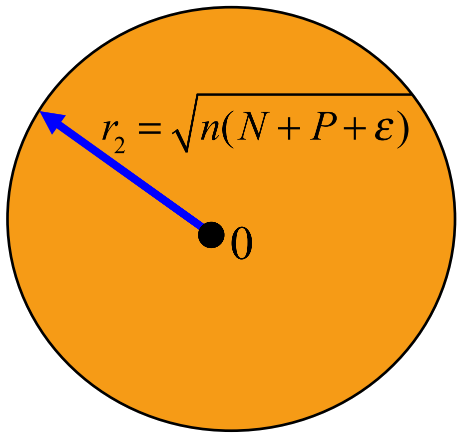
$$\sum_{i=1}^{n} Z_i^2 \approx nEZ_i^2 \quad \text{for large } n.$$



For large n. given a transmitted X^n, Y^n is very likely to be within a sphere of radius $\sqrt{n(N+\varepsilon)}$ centered on X^n.

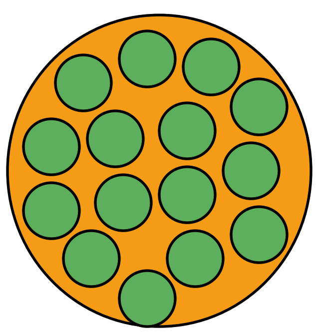
Sphere of received vectors Y^n

- Because $EY^2 \leq P + Nw$, Y^n is very likely to be in a sphere of radius $\sqrt{n(N+P+\epsilon)}$ centered at the origin.



Number of r_1 spheres in r_2 sphere

- How many non-intersecting r_1 spheres (They are essentially decoding spheres.) can we fit in the r_2 sphere?



Sphere packing argument

- Certainly we can upper bound by taking a ratio of the n-dimensional spheres.
- Volume of an n-dimensional sphere is $\Gamma(n)r^n$

$$\frac{\Gamma(n)r_2^n}{\Gamma(n)r_1^n} = \frac{r_2^n}{r_1^n} = \frac{[n(N+P+\epsilon)]^{n/2}}{[n(N+\epsilon)]^{n/2}} \rightarrow \left(\frac{N+P}{N}\right)^{n/2} \quad \text{as } \epsilon \rightarrow 0$$

$$= 2^{\frac{n}{2}\log\left(1+\frac{P}{N}\right)} = 2^{nC}$$

- So our rate is less than C by a sphere packing argument.

Part 11 D: Joint AEP for Continuous R.V.'s

Jointly Typical Sequences for Continuous R.V.'s

- The set $A_\epsilon^{(n)}$ of jointly typical sequences $\{x^n, y^n\}$ with respect to the distribution $f(x,y)$ is the set of n-sequences with empirical differential entropies ϵ-close to the true differential entropies:

$$A_\epsilon^{(n)} = \Big\{ (x^n, y^n) : \quad \left| -\frac{1}{n}\log f(x^n) - h(X) \right| < \epsilon$$

$$\left| -\frac{1}{n}\log f(y^n) - h(Y) \right| < \epsilon$$

$$\left| -\frac{1}{n}\log f(x^n, y^n) - h(X,Y) \right| < \epsilon \Big\}$$

where $f(x^n, y^n) = \prod_{i=1}^{n} f(x_i, y_i)$

Joint AEP for continuous R.V.s

- Let $\left(X^n, Y^n\right)$ be sequences of length n drawn according to $f(x^n, y^n) = \prod_{i=1}^{n} f(x_i, y_i)$. Then

1. $p\left((X^n, Y^n) \in A_\epsilon^{(n)}\right) \to 1 \quad \text{as } n \to \infty$
2. $\mathrm{Vol}\left(A_\epsilon^{(n)}\right) \le 2^{n\left(h(X,Y)+\epsilon\right)}$
3. If $(\tilde{X}^n, \tilde{Y}^n) \sim f(x^n) f(y^n)$ (i.e. $\tilde{X}^n$ and $\tilde{Y}^n$ are independent with the same marginals as $f(x^n, y^n)$) then $p\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) \le 2^{-n(I(X;Y)-3\epsilon)}$.

Proof of 1: $p\left((X^n, Y^n) \in A_\epsilon^{(n)}\right) \to 1 \quad \text{as } n \to \infty$

- Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables.

$$-\frac{1}{n}\log f(X_1,\ldots,X_n) \to h(X) \qquad \text{in prob.}$$

- Proof: $-\frac{1}{n}\log f(X_1,\ldots,X_n) = -\frac{1}{n}\sum_i \log f(X_i)$

$$\to -E\log f(X) \qquad \text{in probability}$$

$$= h(X)$$

$$-\frac{1}{n}\log f(Y_1,\ldots,Y_n) \to h(Y) \qquad \text{in prob.}$$

$$-\frac{1}{n}\log f(X_1,Y_1\ldots,X_n,Y_n) \to h(X,Y) \qquad \text{in prob.}$$

21

Proof of 2: $\mathrm{Vol}\left(A_\epsilon^{(n)}\right) \le 2^{n\left(h(X,Y)+\epsilon\right)}$

$$1 = \int_{S^n} f\left(x^n, y^n\right) dx_1 \ldots dx_n\, dy_1 \ldots dy_n$$

$$\ge \int_{A_\epsilon^{(n)}} f\left(x^n, y^n\right) dx_1 \ldots dx_n dy_1 \ldots dy_n$$

$$\ge \int_{A_\epsilon^{(n)}} 2^{-n\left(h(x,y)+\epsilon\right)}\, dx_1 \ldots dx_n dy_1 \ldots dy_n$$

$$= 2^{-n\left(h(x,y)+\epsilon\right)} \int_{A_\epsilon^{(n)}} dx_1 \ldots dx_n dy_1 \ldots dy_n$$

$$= 2^{-n\left(h(x,y)+\epsilon\right)}\, \mathrm{Vol}\left(A_\epsilon^{(n)}\right)$$

Proof of 3:

$$(\tilde{X}^n, \tilde{Y}^n) \sim f(x^n) f(y^n)$$
$$p\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) \le 2^{-n(I(X;Y)-3\epsilon)}.$$

$$\begin{aligned}
p\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) &= \int_{(x^n, y^n) \in A_\epsilon^{(n)}} f(x^n) f(y^n) dx_1 \ldots dx_n dy_1 \ldots dy_n \\
&\le \int_{(x^n, y^n) \in A_\epsilon^{(n)}} 2^{-n(h(X)-\epsilon)} 2^{-n(h(Y)-\epsilon)} dx_1 \ldots dx_n dy_1 \ldots dy_n \\
&\le 2^{-n(h(X)-\epsilon)} 2^{-n(h(Y)-\epsilon)} \int_{(x^n, y^n) \in A_\epsilon^{(n)}} dx_1 \ldots dx_n \, dy_1 \ldots dy_n \\
&= 2^{-n(h(X)-\epsilon)} 2^{-n(h(Y)-\epsilon)} \operatorname{Vol}\left(A_\epsilon^{(n)}\right) \\
&\le 2^{n(h(X,Y)+\epsilon)} 2^{-n(h(X)-\epsilon)} 2^{-n(h(Y)-\epsilon)} \\
&= 2^{-n(I(X;Y)-3\epsilon)}
\end{aligned}$$

Part 11 E:
Gaussian Channel Coding Theorem

How many typical sequences?

- $\sim 2^{nh(X)}$ typical X sequences
- $\sim 2^{nh(Y)}$ typical Y sequences
- $\sim 2^{nh(X,Y)}$ typical (X,Y) sequences
- Not all pairings of a typical X sequence with a typical Y sequence produce a typical (X,Y) sequence.
- In fact, when the X and Y sequences are chosen independently, the probability is $\sim 2^{-nI(X;Y)}$

Outline of Proof

- For $R<C$ there is a sequence of $(2^{nR},n)$ codes such that $P_e \to 0$ as $n \to \infty$ and $EX \leq P$.
- Proof outline:
  - 1) generate a random code by drawing 2^{nR} blocks of n x values all IID $\sim N(0, P\text{-}\varepsilon)$
  - 2) use typical set decoding
  - 3) show $E[P_e] \to 0$ as $n \to \infty$

Defining Error Events

- If our power constraint $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \le P$ is violated, we consider this to be an error.
  - $E_0 = \left\{ \frac{1}{n}\sum_{i=1}^{n} X_i^2(1) > P \right\}$
  - $E_i = \left\{ (X^n(i), Y^n) \text{ are jointly typical according to } f(x,y) \right\}$
  - $f(x,y)$ is the distribution resulting from $\{X \sim \mathcal{N}(0, P\text{-}), Z \sim \mathcal{N}(0, N), Y = X + Z\}$.
  - If $X^n(i)$ and Y^n are independent, $P(E_i) \le 2^{-n(I(X;Y)-3\epsilon)}$

Probability of error computation

$$
\begin{aligned}
P(\hat{W} \ne W) &= P(\hat{W} \ne W \mid W = 1) \\
&= P(E_0 \cup E_1^c \cup E_2 \cup \cdots \cup E_{2^{nR}}) \\
&\le P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i) \\
&\le \epsilon + \epsilon + 2^{nR} 2^{-n(I(X;Y)-3\epsilon)} \\
&= 2\epsilon + 2^{-n(I-R-3\epsilon)} \\
&\to 0 \quad \text{as } n \to \infty
\end{aligned}
$$

If $R < C$, $P(\hat{W} \ne W) \to 0$

Proof of Converse (No Rate $R>C$ is achievable.)

Fano's inequality

$$H(W \mid \hat{W}) \le 1 + nRP(\hat{W} \ne W)$$

$$nR = H(W) = I(W;\hat{W}) + H(W \mid \hat{W})$$

$$\le I(X^n;Y^n) + 1 + nRP(\hat{W} \ne W)$$

$$\le nC + 1 + nRP(\hat{W} \ne W)$$

$$R \le C + \frac{1}{n} + RP(\hat{W} \ne W)$$

or $P(\hat{W} \ne W) \ge 1 - \frac{C}{R} - \frac{1}{nR}$

If $R > C,\ P(\hat{W} \ne W) > 0$