LECTURE 17

Last time:

- Differential Entropy
- AEP for continuous random variables
- Coding Theorem
- Gaussian Channels

Lecture outline

- Gaussian Channel Capacity as Sphere Packing
- Parallel Gaussian Channels
- Waveform Channels
- Wide-band limit.

Review

- $$Y = X + W$$

  $$C = \max_{f_X : E[X^2] \le \sigma_X^2} I(X;Y) = \frac{1}{2} \log \left(1 + \frac{\sigma_X^2}{\sigma_W^2} \right)$$

- $$Y_i = X_i + W_i$$

  with power constraint $\frac{1}{n} \sum_{i=1}^{n} x_i^2 \le \sigma_X^2$.

Maximum achievable rate is C.

- **Key:** treat average power constraint in proving the coding theorem.

Discussions on Gaussian Channel Capacity

- Consider an arbitrary random vector $\underline{Y}$ with power constraint $E[Y^2] \le \sigma_Y^2$. By WLLN, $\underline{Y}$ lies in a sphere of radius $\sqrt{n\sigma_Y^2}$ with high probability.

$$
\begin{aligned}
\text{Vol}(S^{(n)}(\sqrt{n\sigma_Y^2})) &= \frac{\sqrt{n\sigma_Y^2}^n \pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \\
&= A_n (n\sigma_Y^2)^{\frac{n}{2}}
\end{aligned}
$$

- Consider an i.i.d. Gaussian random vector $\underline{W}$ with variance σ_W^2. What is the typical set $A_\epsilon^{(n)}$ look like?

$$
\begin{aligned}
& \underline{W} \in A_\epsilon^{(n)}(W) \\
\Leftrightarrow\ & \left| -\frac{1}{n} \log f_{\underline{W}}(\underline{w}) - h(W) \right| \le \epsilon \\
\Leftrightarrow\ & \left| \frac{1}{2} \log 2\pi\sigma_W^2 + \frac{\|\underline{w}\|^2}{2n\sigma_W^2} - \frac{1}{2} \log 2\pi e \sigma_W^2 \right| \le \epsilon \\
\Leftrightarrow\ & \left| \frac{\|\underline{w}\|^2}{n\sigma_W^2} - 1 \right| \le \epsilon
\end{aligned}
$$

The typical set is a shell of the sphere $S^{(n)}(\sqrt{n\sigma_W^2})$.

A few easy conclusions

- For the Gaussian vector, since every $\underline{w}$ inside the sphere has higher pdf, the volume of the sphere is concentrated on the shell.
- The typical set of a Gaussian r.v. covers almost all the volume that a power limited r.v. can cover.

Sphere Packing in Gaussian channel

- for a transmitted codeword $\underline{x}$, the set of $\underline{y}$ that are joint typical with $\underline{x}$ is a sphere centered at $\underline{x}$ with radius $\sqrt{n\sigma_W^2}$.
- the maximum data rate is

$$R = \frac{1}{n}\log\frac{A_n(n\sigma_Y^2)^{\frac{n}{2}}}{A_n(n\sigma_W^2)^{\frac{n}{2}}} = \frac{1}{2}\log\left(\frac{\sigma_Y^2}{\sigma_W^2}\right)$$

- Need $\underline{Y}$ to be i.i.d. Gaussian.
- For $\underline{x}$ and $\underline{y}$ independently chosen according to the marginal distribution, the probability that they are jointly typical is

$$\frac{A_n(n\sigma_W^2)^{\frac{n}{2}}}{A_n(n\sigma_Y^2)^{\frac{n}{2}}} \approx 2^{-nI(X;Y)}$$

- For a code book of $\underline{x}_i, i = 1, \ldots, 2^{nR}$, and an independently chosen $\underline{y}$, the probability that there exists i, such that $(\underline{x}_i, \underline{y})$ are jointly Gaussian

$$2^{nR}\frac{A_n(n\sigma_W^2)^{\frac{n}{2}}}{A_n(n\sigma_Y^2)^{\frac{n}{2}}} \approx 2^{-n(I(X;Y)-R)} \to 0$$

Extensions of the Sphere Packing Argument

- Can use the sphere packing argument in the $\mathcal{X}$ space. For any fixed $\underline{y}$, the set of $\underline{x}$ that are jointly typical with $\underline{y}$ is a sphere of radius

$$\left(n\frac{\sigma_W^2\sigma_X^2}{\sigma_X^2+\sigma_W^2}\right)^{\frac{1}{2}}$$

centered at $\hat{X}$.

- Let $m = 2n$,

$$\begin{aligned}
&\frac{1}{m}\log \mathrm{Vol}(S^{(m)}(\sqrt{m\sigma_W^2})) \\
= \;& \frac{1}{m}\log\frac{(m\sigma_W^2)^n\pi^n}{\Gamma(n+1)} \\
\approx \;& \frac{1}{2n}\log\left(\frac{2n\sigma_W^2\pi}{ne^{-1}}\right)^n \\
= \;& \frac{1}{2}\log(2\pi e\sigma_W^2)
\end{aligned}$$

Summary for Sphere Packing

- Useful geometric approach to the channel capacity, and the optimal input distribution
- Not limited to the AWGN channels.
- Very hard to prove rigorously.

Example: In the AWGN channel, if the average power constraint is replaced by

$$x_i^2 \leq \sigma_W^2, \quad \forall i$$

optimal input should be "uniform" in the cube.

Example: In the AWGN channel, assume the signal-to-noise ratio is very high. Suppose now the receiver can only receive $\underline{Y} / \|\underline{Y}\|$, i.e., the norm of $\underline{Y}$ is completely lost.

optimal input should have "uniform" distribution on the direction of $\underline{X}$.

A Greedy way to Spend Power

Consider the AWGN channel, with power constraint

$$\frac{1}{n}\sum x_i^2 \leq P$$

The transmitter

- Divide the power in to $P = P_1 + P_2$, and the data stream into $R = R_1 + R_2$, and assign to two virtual users
- Each user has a random code book, generate codewords $\underline{X}_1$ and $\underline{X}_2$. The superposition $\underline{X} = \underline{X}_1 + \underline{X}_2$ is transmitted.

The receiver receives

$$\underline{Y} = \underline{X}_1 + \underline{X}_2 + \underline{W}$$

- First treat $\underline{X}_2 + \underline{W}$ as the noise, can support

$$nR_1 \leq I(\underline{X}_1; \underline{Y}) = \frac{n}{2} \log \left(1 + \frac{P_1}{P_2 + N}\right)$$

- Upon decoding $\underline{X}_1$, subtract $\underline{X}_1$ from $\underline{Y}$, then decode $\underline{X}_2$. Can support rate

$$nR_2 \leq I(\underline{X}_2; \underline{Y}|\underline{X}_1) = \frac{n}{2} \log \left(1 + \frac{P_2}{N}\right)$$

- The over all data rate is $R_1 + R_2$, the channel capacity is achieved.

Discussions

- To generalize, we can have many virtual users. Useful in multiple access channels.
- Highly depends on the fact that the sum of independent Gaussian r.v.'s is Gaussian
- Can be used when the capacity is the only concern.

Example Consider a Gaussian channel with $nR = 10bits$, random coding need $2^{nR} = 1024$ code words. If divided into two virtual users of $nR_i = 5bits$ each, need only two codebooks of 32 codewords each. Does this mean the complexity is greatly decreased?

Parallel Gaussian Channels

Assume we have a set of Gaussian channels in parallel,

$$Y_i = X_i + W_i, \quad i = 1, \ldots, k$$

where the noise W_i are independent of X's and are independent of each other, with power σ_i^2.

The power constraint is

$$E\left[\sum_{i=1}^{k} X_i^2\right] \leq P$$

Capacity

$$C = \max_{f_{X_1,\ldots,X_k}: \sum EX_i^2 \leq P} I(X_1, \ldots X_k; Y_1, \ldots, Y_k)$$

$$
\begin{aligned}
& I(X_1, \ldots, X_k; Y_1, \ldots, Y_k) \\
= \ & h(Y_1, \ldots, Y_k) - h(Y_1, \ldots, Y_k | X_1, \ldots, X_k) \\
= \ & h(Y_1, \ldots, Y_k) - \sum_{i=1}^{k} h(W_i) \\
\leq \ & \sum_i [h(Y_i) - h(W_i)] \\
\leq \ & \sum_i \frac{1}{2} \left(1 + \frac{P_i}{\sigma_i^2}\right)
\end{aligned}
$$

Equality achieved if X_i's are independent and with power $E[X_i^2] = P_i$, subject to power constraint $\sum P_i \leq P$.

Now we need the optimal power allocation

$$
\max_{P_1, \ldots, P_k : \sum P_i \leq P} \sum \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right)
$$

Water Filling

Define

$$J = \sum \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right) + \lambda \left(\sum P_i\right)$$

Need $\forall i$

$$\frac{1}{P_i + N_i} + \lambda = 0$$

or

$$P_i = \nu - N_i$$

However, since $P_i \geq 0$, so we have

$$P_i = (\nu - N_i)^+$$

and ν can be solved by

$$\sum_i (\nu - N_i)^+ = P$$

- Use the greedy approach to get the same result.

Waveform Channel: White Gaussian Noise

$$Y(t) = X(t) + W(t)$$

- $W(t)$ is a Gaussian random process. Any set of samples are i.i.d. Gaussian distributed with zero mean and variance $E[W(t)^2] = \frac{N_0}{2}$.
- Autocorrelation function $R(\tau) = E[W(t)W(t-\tau)] = \frac{N_0}{2}\delta(\tau)$.
- Power Spectral Density is flat $\frac{N_0}{2}$ over all frequency
- Do we really have white Gaussian noise?

Front end of the receiver

Goal: provide sufficient statistics for bandlimited transmitted signals.

Assume the $X(t)$ has non-zero frequency components only in $[-W, W]$, can sample at $T_s = \frac{1}{2W}$.

- can be thought as a bank of matched filter matched to

  $$\text{sinc}\left(t - \frac{n}{2W}\right)$$

  where

  $$\text{sinc}(t) = \frac{\sin(2\pi W t)}{2\pi W t}$$

- The front end provide sufficient statistics since
  - the out-of-band noise is independent of the transmitted signals
  - the out-of-band noise is independent of the in-band noise

Discrete-Time Channel Model

- $2W$ samples per second.
- The noise power in each sample is $\frac{N_0}{2}$.
- Average per symbol signal power $P/2W$.

We get back to

$$Y_i = X_i + W_i$$

The overall capacity in (bits/sec)

$$C = 2W\frac{1}{2}\log\left(1 + \frac{P/2W}{N_0/2}\right)$$

$$\boxed{C = W\log\left(1 + \frac{P}{N_0 W}\right)}$$

P has unit (watt), and N_0 (watt/hz).

Admiring Shannon's Formula

$$\boxed{C = W \log\left(1 + \frac{P}{N_0 W}\right)}$$

- **Question 1** Do we always want more bandwidth?
- **Question 2** What can we do with infinite bandwidth?

  $$C \rightarrow \frac{P}{N_0} \log_2 e (bits/sec)$$

  Capacity grows linearly with the power (very power efficient).
- The energy required to send 1 bit is E_b $= P/C$ watts $\times 1$ sec,

  $$\frac{E_b}{N_0} = \frac{1}{\log_2 e} = -1.59(dB)$$

- **Question 3** What can we do with infinite power?

  $$\lim_{W \rightarrow \infty} \frac{C}{W} \rightarrow \log \frac{P}{N_0}$$

  capacity linearly increases with W (spectral efficient)