Chapter 9

Gaussian Channel

Peng-Hua Wang

Graduate Inst. of Comm. Engineering

National Taipei University

Chapter Outline

Chap. 9 Gaussian Channel

- **9.1** Gaussian Channel: Definitions
- **9.2** Converse to the Coding Theorem for Gaussian Channels
- **9.3** Bandlimited Channels
- **9.4** Parallel Gaussian Channels
- **9.5** Channels with Colored Gaussian Noise
- **9.6** Gaussian Channels with Feedback

9.1 Gaussian Channel: Definitions

Introduction

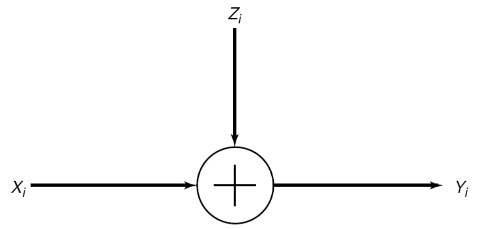

FIGURE 9.1. Gaussian channel.

$$Y_i = X_i + Z_i, \quad Z_i \sim N(0, N)$$

- X_i: input, Y_i:output, Z_i: noise. Z_i is independent of X_i.
- Without further constraint, the capacity of this channel may be infinite.
  - If the noise variance N is zero, the channel can transmit an arbitrary real number with no error.
  - If the noise variance N is nonzero, we can choose an infinite subset of inputs arbitrary far apart, so that they are distinguishable at the output with arbitrarily small probability of error.

Introduction

- The most common limitation on the input is an energy or power constraint.
- We assume an average power constraint. For any codeword $(x_1, x_2, \ldots, x_n)$ transmitted over the channel, we require that

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

Information Capacity

Definition 1 (Capacity) *The information capacity of the Gaussian channel with power P is*

$$C = \max_{f(x):E[X^2]\leq P} I(X;Y)$$

We can calculate the information capacity as follows.

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) = h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z|X) = h(Y) - h(Z) \\ &\leq \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi eN \\ &= \frac{1}{2}\log\left(1+\frac{P}{N}\right) \end{aligned}$$

Note that $E[Y^2] = E[(X+Z)^2] = P+N$ and the entropy of gaussian with variance σ^2 is $\frac{1}{2}\log 2\pi e\sigma^2$.

Information Capacity

Therefore, the information capacity of the Gaussian channel is

$$C = \max_{E[X^2]\leq P} I(X;Y) = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

and the equality holds when $X \sim N(0, P)$.

- Next, we will show that this capacity is achievable.

Code for Gaussian Channel

Definition 2 ((M, n) code for Gaussian Channel) *An (M, n) code for the Gaussian channel with power constraint P consists the following:*

1. *An index set $\{1, 2, \ldots, M\}$.*
2. *An encoding function $x : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, yielding codewords $x^n(1), x^n(2), \ldots, x^n(M)$, satisfying the power constraint P*

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \leq P, \quad w = 1, 2, \ldots, M.$$

3. *A decoding function $g : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$.*

Definitions

Definition 3 (Conditional probability of error)

$$\lambda_i = \Pr(g(Y^n) \neq i | X^n = x^n(i)) = \sum_{g(y^n) \neq i} p(y^n | x^n(i))$$

$$= \sum_{y^n} p(y^n | x^n(i)) I(g(y^n) \neq i)$$

- $I(\cdot)$ is the indicator function.

Definitions

Definition 4 (Maximal probability of error)

$$\lambda^{(n)} = \max_{i \in \{1,2,\ldots,M\}} \lambda_i$$

Definition 5 (Average probability of error)

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$$

- The decoding error is

$$\Pr(g(Y^n) \neq W) = \sum_{i=1}^{M} \Pr(W = i) \Pr(g(Y^n) \neq i | W = i)$$

If the index W is chosen uniformly from $\{1, 2, \ldots, M\}$, then $P_e^{(n)} = \Pr(g(Y^n) \neq W)$.

Definitions

Definition 6 (Rate) *The rate R of an (M, n) code is*

$$R = \frac{\log M}{n} \quad \textit{bits per transmission}$$

Definition 7 (Achievable rate) *A rate R is said to be achievable for a Gaussian channel with a power constraint P if there exists a $(\lceil 2^{nR} \rceil, n)$ code with codewords satisfying the power constraint such that the maximal probability of error $\lambda^{(n)}$ tends to 0 as $n \to \infty$.*

Definition 8 (Channel capacity) *The capacity of a channel is the supremum of all achievable rates.*

Capacity of a Gaussian Channel

Theorem 1 (Capacity of a Gaussian Channel) *The capacity of a Gaussian channel with power constraint P and noise variance N is*

$$\frac{1}{2}\log\left(1+\frac{P}{N}\right) \text{ bits per transmission.}$$

Sphere Packing Argument

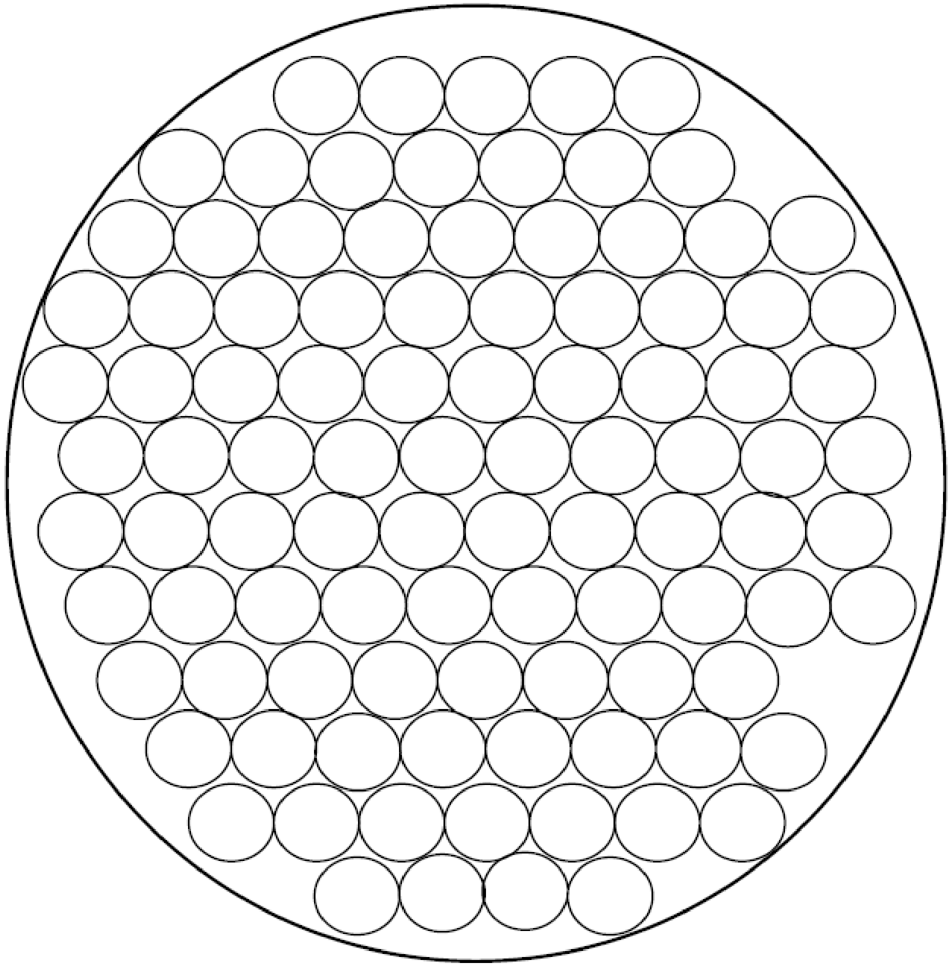

FIGURE 9.2. Sphere packing for the Gaussian channel.

Sphere Packing Argument

For each sent codeword, the received codeword is contained in a sphere of radius $\sqrt{nN}$. The received vectors have energy no grater than $n(P+N)$, so they lie in a sphere of radius $\sqrt{n(P+N)}$. How many codeword can we use without intersection in the decoding sphere?

$$M = \frac{A_n\left(\sqrt{n(P+N)}\right)^n}{A_n(\sqrt{nN})^n} = \left(1+\frac{P}{N}\right)^{n/2}$$

where A the constant for calculating the volume of n-dimensional sphere. For example, $A_2 = \pi$, $A_3 = \frac{4}{3}\pi$. Therefore, the capacity is

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

$R < C \to$ Achievable

- **Codebook.** Let $X_i(w), i = 1, 2, \dots, n, w = 1, 2, \dots, 2^{nR}$ be i.i.d. $\sim \mathcal{N}(0, P - \epsilon)$. For large n,

$$\frac{1}{n} \sum X_i^2 \to P - \epsilon.$$

- **Encoding.** The codebook is revealed to both the sender and the receiver. To send the message index w, the transmitter sends the wth codeword $X^n(w)$ in the codebook.
- **Decoding.** The receiver searches for the one that is jointly typical with the received vector. If there is one and only one such codeword $X^n(w)$, the receiver declares $\hat{W} = w$. Otherwise, the receiver declares an error. If the power constraint is not satisfied, the receiver also declare an error.

$R < C \to$ Achievable

- **Probability of error.** Assume that codeword 1 was sent. $Y^n = X^n(1) + Z^n$. Define the events

$$E_0 = \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j^2(1) > P \right\}$$

and

$$E_i = \left\{ \left(X^n(i), Y^n(i) \text{ is in } A_\epsilon^{(n)} \right) \right\}.$$

Then an error occurs if

- The power constraint is violate. $\Rightarrow E_0$ occurs.
- The transmitted codeword and the received sequence are not jointly typical. $\Rightarrow E_1^c$ occurs.
- Wrong codeword is jointly typical with the received sequence. $\Rightarrow$ $E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}$ occurs.

$R < C \to$ Achievable

Let W be uniformly distributed. We have

$$
\begin{aligned}
P_e^{(n)} &= \frac{1}{2^{nR}} \sum \lambda_i = P(\mathcal{E}) = \Pr(\mathcal{E}|W=1) \\
&= P(E_0 \cup E_a^c \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}) \\
&\leq P(E_0) + P_(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i) \\
&\leq \epsilon + \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)} \\
&\leq 2\epsilon + 2^{-n(I(X;Y)-R-3\epsilon)} \leq 3\epsilon
\end{aligned}
$$

for n sufficient large and $R < I(X;Y) - 3\epsilon$.

$R < C \rightarrow$ Achievable, final part

- Since the average probability of error over codebooks is less then 3ϵ, there exists at least one codebook $\mathcal{C}^*$ such that $\Pr(\mathcal{E}|\mathcal{C}^*) < 3\epsilon$.
  - $\mathcal{C}^*$ can be found by an exhaustive search over all codes.
- Deleting the worst half of the codewords in $\mathcal{C}^*$, we obtain a code with low maximal probability of error. The codewords that violates the power constraint is definitely deleted. (why?) Hence, we have construct a code that achieves a rate arbitrarily close to C.

9.2 Converse to the Coding Theorem for Gaussian Channels

Achievable $\rightarrow R < C$

We will prove that if $P_e^{(n)} \rightarrow 0$ then $R \leq C = \frac{1}{2}\log(1+\frac{P}{N})$. Let W be distributed uniformly. We have $W \rightarrow X^n \rightarrow Y^n \rightarrow \hat{W}$. By Fano's inequality,

$$H(W|\hat{W}) \leq 1 + nRP_e^{(n)} = n\epsilon_n, \quad \text{where } \epsilon_n = \frac{1}{n} + RP_e^{(n)} \rightarrow 0$$

as $P_e^{(n)} \rightarrow 0$. Now,

$$\begin{aligned} nR = H(W) &= I(W;\hat{W}) + H(W|\hat{W}) \\ &\leq I(W;\hat{W}) + n\epsilon_n \leq I(X^n;Y^n) + n\epsilon_n \text{(data processing ineq.)} \\ &= h(Y^n) - h(Y^n|X^n) + n\epsilon_n = h(Y^n) - h(Z^n) + n\epsilon_n \\ &\leq \sum_{i=1}^n h(Y_i) - h(Z^n) + n\epsilon_n \leq \sum_{i=1}^n h(Y_i) - \sum_{i=1}^n h(Z_i) + n\epsilon_n \end{aligned}$$

Achievable $\rightarrow R < C$

$$
\begin{aligned}
nR &\le \sum_{i=1}^{n} (h(Y_i) - h(Z_i)) + n\epsilon_n \\
&\le \sum \left(\frac{1}{2} \log\left(2\pi e (P_i + N)\right) - \frac{1}{2} \log 2\pi e N \right) + n\epsilon_n \\
&= \sum \frac{1}{2} \log\left(1 + \frac{P_i}{N}\right) + n\epsilon_n \\
&\le \frac{n}{2} \log\left(1 + \frac{P}{N}\right) + n\epsilon_n
\end{aligned}
$$

since every codeword satisfies the power constraint. Thus,

$$
R \le \frac{1}{2} \log\left(1 + \frac{P}{N}\right) + \epsilon_n .
$$

9.3 Bandlimited Channels

Capacity of Bandlimited Channels

- Suppose the output of a band-limited channel can be represented by

$$Y(t) = (X(t) + N(t)) * h(t)$$

  where $X(t)$ is the input signal, $Z(t)$ is the white Gaussian noise, and $h(t)$ is the impulse response of the channel with bandwidth W.

- The sampling frequency is $2W$. If the channel be used over the time interval $[0, T]$, then there are $2WT$ samples transmitted.

Capacity of Bandlimited Channels

- If the noise has power spectral density $N_0/2$ watts/Hz, the noise power is $(N_0/2)(2W) = N_0W$. The noise energy per sample is $N_0W * T/2WT = N_0/2$. If the signal power is P. The signal energy per sample is $PT/2WT = P/2W$.
- The capacity is $\frac{1}{2}\log\left(1+\frac{P/2W}{N_0/2}\right)$ bits/sample or

$$C = W\log\left(1+\frac{P}{N_0W}\right) \text{ bits/second}$$

9.4 Parallel Gaussian Channels

Capacity of Bandlimited Channels

- In this section we consider k independent Gaussian channels in parallel with a common power constraint. The objective is to distribute the total power among the channels so as to maximize the capacity. The channels are modeled as

$$Y_j = X_j + Z_j, j = 1, 2, \ldots, k.$$

with $Z_j \sim \mathcal{N}(0, N_j)$. There is a common power constraint

$$E\left[\sum_{j=1}^{k} X_j^2\right] \leq P.$$

Capacity of Bandlimited Channels

The information capacity is

$$C = \max_{f(X-1,\ldots,x_n):EX_i^2<P} I(X_1, X_2 \ldots, X_k; Y_1, Y_2, \ldots, Y_k)$$

Since $Z_1, Z_2, \ldots, Z_k$ are independent,

$$\begin{aligned}
& I(X_1, X_2 \ldots, X_k; Y_1, Y_2, \ldots, Y_k) \\
=& h(Y_1, Y_2, \ldots, Y_k) - h(Y_1, Y_2, \ldots, Y_k | X_1, X_2 \ldots, X_k) \\
=& h(Y_1, Y_2, \ldots, Y_k) - h(Z_1, Z_2, \ldots, Z_k | X_1, X_2 \ldots, X_k) \\
=& h(Y_1, Y_2, \ldots, Y_k) - h(Z_1, Z_2, \ldots, Z_k) \\
=& h(Y_1, Y_2, \ldots, Y_k) - \sum_i h(Z_i) \\
\leq& \sum_i h(Y_i) - \sum_i h(Z_i) \leq \sum_i \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right)
\end{aligned}$$

where $P_i = EX_i^2$ and $\sum P_i = P$

Capacity of Bandlimited Channels

Therefore, we have a constrained optimization problem

$$\max \sum_i \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right) \text{ subject to } \sum_i P_i \le P, P_i \ge 0.$$

This can be solved by Lagrange multiplier together with the Kuhn-Tucker condition.

$$-\frac{1}{2}\frac{1/N_i}{1+P_i/N_i} - \mu_i + \lambda = 0$$

$$-P_i \le 0, \sum_i P_i - P \le 0$$

$$\mu_i P_i = 0, \lambda(\sum_i P_i - P) = 0$$

$$\mu_i \ge 0, \lambda \ge 0$$

Capacity of Bandlimited Channels

Case I. $\lambda = 0$. We have

$$P_i + N_i = -\frac{1}{2\mu_i}, \quad P_i = -\frac{1}{2\mu_i} - N_i$$

This violates the condition $-P_i \leq 0$ since $N_i > 0$ and $\mu_i \geq 0$.

Case II. $\lambda \neq 0$. We have

$$P_i + N_i = \frac{1}{2(\lambda - \mu_i)} = \begin{cases} \frac{1}{2\lambda} = \text{constant}, & P_i > 0 (\text{ imply } \mu_i = 0) \\ \frac{1}{2(\lambda - \mu_i)}, & P_i = 0. \end{cases}$$

We can solve λ by $\sum_i P_i = \sum_i (\frac{1}{2\lambda} - N_i)^+ = P$

Capacity of Bandlimited Channels

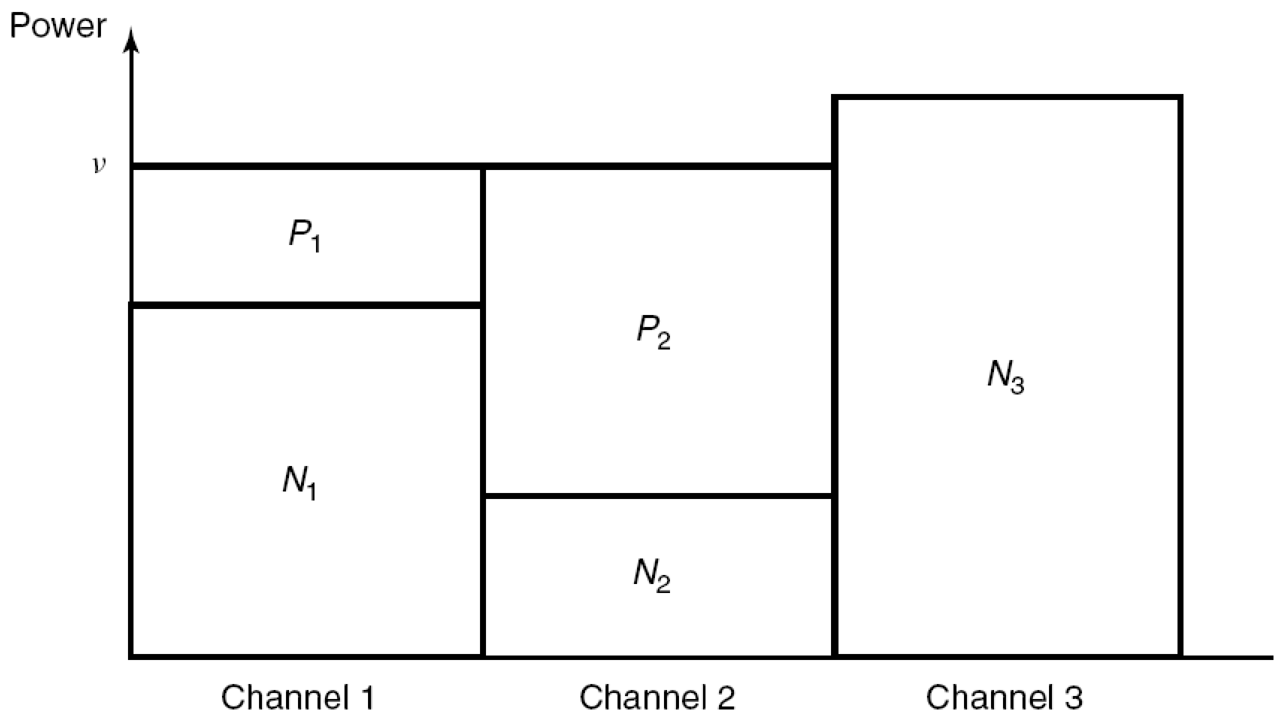

FIGURE 9.4. Water-filling for parallel channels.

Nonlinear Optimization

For the problem

$$\min f(x_1, x_2, \ldots, x_n)$$

subject to

$$g_j(x_1, x_2, \ldots, x_n) \le 0, j = 1, 2, \ldots m$$

The necessary conditions for optimization are

$$\begin{aligned}
\frac{\partial f}{\partial x_i} + \sum_j \mu_j \frac{\partial g_j}{\partial x_i} &= 0,\ i = 1, 2, \ldots, n \\
g_j(x_1, x_2, \ldots, x_n) &\le 0,\ j = 1, 2, \ldots, m \\
\mu_j g_j(x_1, x_2, \ldots, x_n) &= 0,\ j = 1, 2, \ldots, m \\
\mu_j &\ge 0,\ j = 1, 2, \ldots, m
\end{aligned}$$