CS258: Information Theory

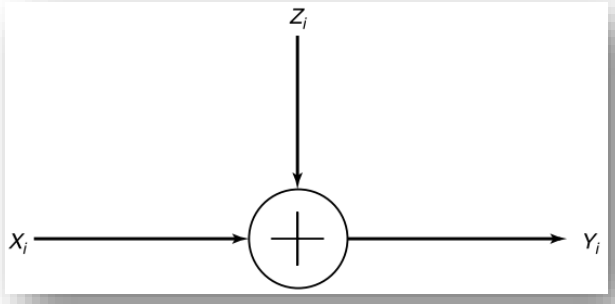
Fan Cheng
Shanghai Jiao Tong University

http://www.cs.sjtu.edu.cn/~chengfan/
chengfan@sjtu.edu.cn
Spring, 2023

Outline

- Channel Model
- Coding Theorem for Gaussian Channels
- Water-filling for Parallel Gaussian Channels
- Channel with Worst Additive Noise

Gaussian Channel



Gaussian channel

Continuous alphabet channel

- The channel could be use at each time i
- The input X_i, noise Z_i, output Y_i are continuous

- The most important continuous alphabet channel is the **Gaussian channel.** For example, wireless telephone channels and satellite links
- The noise Z_i is drawn i.i.d. from a Gaussian distribution with variance N
- The noise Z_i is assumed to be independent of the signal X_i
- This is a time-discrete channel with output Y_i at time i, where Y_i is the sum of the input X_i and the noise Z_i

$$Y_i = X_i + Z_i, \qquad Z_i \sim \mathcal{N}(0, N).$$

- Without further conditions, the capacity of this channel may be ∞.
  - The values of X may be very sparse
  - Assume the variance of noise N is neglected compared to the distances of the values of X. Then $Y = X + Z \approx X$. Thus $I(X;Y) \approx H(X)$, which may be ∞.

Energy Constraint

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

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

  - within the sphere $\sqrt{nP}$
  - P per channel use

- This communication channel models many practical channels, including radio and satellite links.

The information capacity of the Gaussian channel with power constraint P is

$$C = \max_{f(x):EX^2 \le P} I(X;Y)$$

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

$$h(Z) = \frac{1}{2}\log 2\pi eN$$

$$EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2 \le P + N$$

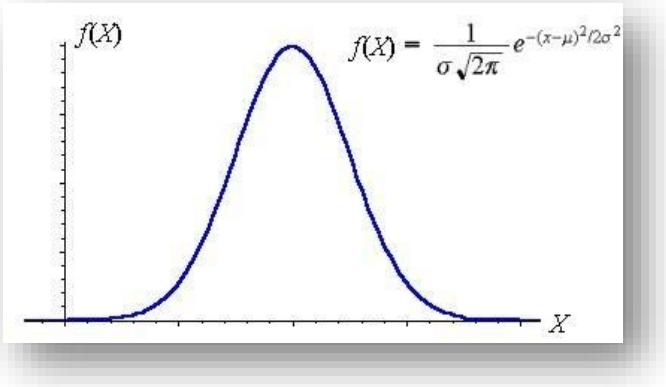
$$h(Y) \le \frac{1}{2}\log 2\pi e(P+N)$$

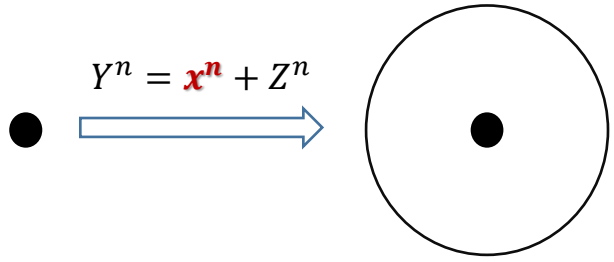
$$\begin{aligned} I(X;Y) = h(Y) - h(Z) &\le \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi eN \\ &= \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \end{aligned}$$

$$\boldsymbol{C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)}$$

The maximum is attained when $X \sim \mathcal{N}(0, P)$

Gaussian Channel: Intuition



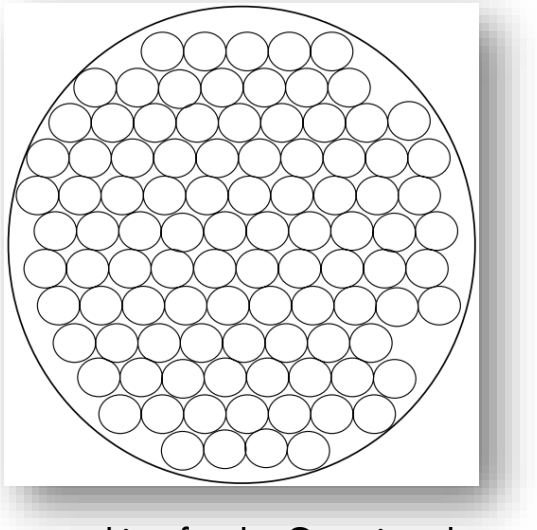


- Consider any codeword x^n of length n.

$$Y^n = \boldsymbol{x^n} + Z^n$$

- The received vector is normally distributed with mean equal to the true codeword and variance equal to the noise variance.
- With high probability, the received vector is contained in a sphere of radius $\sqrt{n(N+\epsilon)}$ around the true codeword.
- If we assign everything within this sphere to the given codeword, when this codeword is sent there will be an error only if the received vector falls outside the sphere, which has low probability.
  - Each codeword is represented by a sphere
  - Low decoding error requires no intersection between any spheres

Gaussian Channel: Intuition (cont'd)



Sphere packing for the Gaussian channel

The maximum number of nonintersecting decoding spheres is no more than

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

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

$$C = \sup R = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

- The received vectors $(Y = X + Z)$ have energy no greater than $n(P+N)$, so they lie in a sphere of radius $\sqrt{n(P+N)}$
- The volume of an n-dimensional sphere is of the form $C_n r^n$, where r is the radius of the sphere.

$$2\pi r,\ \pi r^2 \text{ and } \frac{4}{3}\pi r^3$$

- The volumes are approximated by

$$C_n(nN)^{\frac{n}{2}} \text{ and } C_n\left(n(P+N)\right)^{\frac{n}{2}}$$

Gaussian Channel: Definition

- Definition. An (M, n) code for the Gaussian channel with power constraint P consists of the following:
  1. An index set $\{1, 2, \ldots, M\}$.
  2. An encoding function $x: \{1, 2, \ldots, M\} \rightarrow \mathcal{X}^n$, yielding codewords $x^n(1), x^n(2), \ldots, x^n(M)$, satisfying the power constraint P; that is, for every codeword
  $$\sum_{i=1}^{n} x_i^2(w) \leq nP, \qquad w = 1,2, \ldots, M.$$
  3. A decoding function
  $$g: \mathcal{Y}^n \rightarrow \{1,2, \ldots, M\}$$
- The arithmetic average of the probability of error is defined by
$$P_e^{(n)} = \frac{1}{2^{nR}} \sum \lambda_i$$

A rate R is said to be *achievable* for a Gaussian channel with a power constraint P if there exists a sequence of $(2^{nR}, n)$ codes with codewords satisfying the power constraint such that the maximal probability of error $\lambda^{(n)}$ tends to zero. The capacity of the channel is the supremum of the achievable rates.

Gaussian Channel: Code Construction

Random codes and joint typicality: Power constraints and the random variables are continuous

- **Generation of the codebook**

  We generate the codewords $(x_1, x_2, \ldots, x_n)$ with each element i.i.d. according to a normal distribution with variance $P - \epsilon$. Since for large n,

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

  The probability that a codeword does not satisfy the power constraint will be small.

  Let $X_i(w), \quad i = 1,2,\ldots,n, w = 1,2,\ldots,2^{nR}$ be i.i.d. $\sim \mathcal{N}(0, P-\epsilon)$, forming codewords $X^n(1), X^n(2), \ldots, X^n(2^{nR}) \in \mathcal{R}^n$

- **Encoding**:
  - The codebook is revealed to both the sender and the receiver.
  - To send the message index w, sends the wth codeword $X^n(w)$ in the codebook.
- **Decoding**: The receiver looks down the list of codewords $\{X^n(w)\}$ and searches for one that is jointly typical with the received vector.
  - If there is one and only one such codeword $X^n(w)$, the receiver declares $\hat{W} = w$ to be the transmitted codeword.
  - Otherwise, the receiver declares an error. The receiver also declares an error if the chosen codeword does not satisfy the power constraint.

Gaussian Channel: Probability of Error

Without loss of generality, assume that codeword 1 was sent. Thus,

$$Y^n = X^n(1) + Z^n.$$

Define the following events:

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

and

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

$$\Pr(\mathcal{E}|W = 1) = P\left(E_0 \cup E_1^c \cup E_2 \cup E_3 \ldots \cup E_{2^{nR}}\right) \leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$

$$P(E_0) \to 0$$
$$P(E_1^c) \leq \epsilon$$

$$\sum_{i=2}^{2^{nR}} P(E_i) = (2^{nR} - 1)2^{-n(I(X;Y)-3\epsilon)} \leq 2^{-n(I(X;Y)-R-3\epsilon)}$$

$$P_e^{(n)} \leq 3\epsilon$$

Now choosing a good codebook and deleting the worst half of the codewords, we obtain a code with low maximal probability of error $\lambda^{(n)} \to 0$. (DMC)

Gaussian Channel: Converse

Let W be distributed uniformly over $\{1, 2, \ldots, 2^{nR}\}$.

$$W \to X^n(W) \to Y^n \to \hat{W}$$

By Fano's inequality

$$H(W|\hat{W}) \le 1 + nRP_e^{(n)} = n\epsilon_n$$

where $\epsilon_n \to 0$ as $P_e^{(n)} \to 0$.

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

Let P_i be the average power of the ith column of the codebook

$$P_i = \frac{1}{2^{nR}} \sum_{w} x_i^2(w) \quad \text{and} \quad \frac{1}{n}\sum_{i} P_i \le P$$

Since X_i and Z_i are independent, then

$$EY_i^2 = P_i + N, h(Y_i) \le \frac{1}{2}\log 2\pi e(P_i + N)$$

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

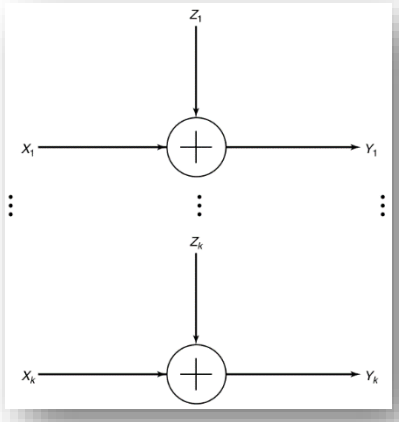
$f(x) = \frac{1}{2}\log(1+x)$ is concave

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\log\left(1 + \frac{P_i}{N}\right) \\ &\le \frac{1}{2}\log\left(1 + \frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{N}\right) \le \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \end{aligned}$$

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

$$\mathcal{C} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{nR}) & x_2(2^{nR}) & \cdots & x_n(2^{nR}) \end{bmatrix}$$

Parallel Gaussian Channel



$$C = \max_{f(x_1,x_2,\ldots,x_k):E\sum X_i^2 \le P} I(X_1, X_2, \ldots, X_k; Y_1, Y_2, \ldots, Y_k)$$

- Assume that we have a set of Gaussian channels in parallel. The output of each channel is the sum of the input and Gaussian noise. For channel j,

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

- The noise is assumed to be independent from channel to channel. We assume that there is a common power constraint on the total power used, that is

$$E\sum_{j=1}^{k} X_j^2 \le P$$

- We wish to distribute the power among the various channels so as to maximize the total capacity.
  - $P_i = EX_i^2$,and $\sum P_i \le P$

Parallel Gaussian Channel (cont'd)

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

where $P_i = EX_i^2$,and $\sum P_i = P$. Equality is achieved by

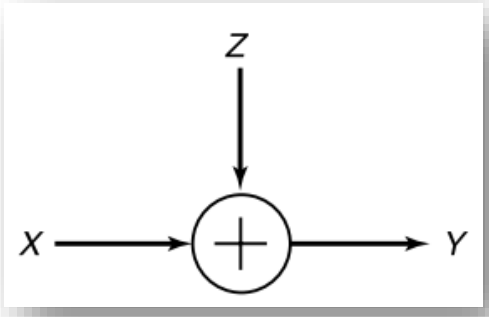
$$(X_1, X_2, \ldots, X_k) \sim \mathcal{N}(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{bmatrix})$$

- We need to optimize

$$\max \sum_i \log\left(1 + \frac{P_i}{N_i}\right)$$
$$\sum P_i = P$$

Lagrange and KKT ⇒ Water-filling

Worst Additive Noise



$$\boldsymbol{Y = X + Z}$$
$$\boldsymbol{C = \max_{X:EX^2 \le P} I(X; X+Z)}$$

- Under the energy constraint P, the channel capacity of **additive channel** $Y = X + Z$ is
$$C(Z) = \max_{X:EX^2 \le P} I(X;Y)$$
$$= \max_{X:EX^2 \le P} h(X+Z) - h(Z)$$
- What is the minimum of $C(Z)$, if we could choose $Z: EZ^2 \le N$.
  - That is, to play a max-min game between X and Z
$$\min_{Z:EZ^2 \le N} C(Z) := \min_{Z:EZ^2 \le N} \max_{X:EX^2 \le P} I(X; X+Z) = \min_{Z:EZ^2 \le N} \left(\max_{X:EX^2 \le P} I(X; X+Z) \right)$$
  - We need to find a Z^*. When $C(Z^*)$ is attained by X^*,
$$I(X^*; X^* + Z^*) \le \max_{X:EX^2 \le P} I(X; X+Z)$$
- **The $\min_{Z:EZ^2 \le N} C(Z)$ is attained iff $Z = Z_G \sim \mathcal{N}(0, \sigma^2)$ (Shannon, 1948)**

Worst Additive Noise

Entropy power inequality (EPI, Shannon 1948): If X and Y are independent random n-vectors with densities, then

$$e^{\frac{2}{n}h(X+Y)} \geq e^{\frac{2}{n}h(X)} + e^{\frac{2}{n}h(Y)}$$

- $I(X; X+Z) = h(X+Z) - h(Z)$
- By EPI,

$$h(X+Z) \geq \frac{1}{2}\log\left(e^{2h(X)} + e^{2h(Z)}\right)$$

- $I(X; X+Z) \geq \frac{1}{2}\log\left(e^{2h(X)} + e^{2h(Z)}\right) - h(Z)$
- $f(t,s) = \frac{1}{2}\log(e^{2t} + e^{2s}) - s$, where

$$t = h(X) \leq \frac{1}{2}\log 2\pi eP$$

$$s = h(Z) \leq \frac{1}{2}\log 2\pi eN$$

- $f(t,s)$ is increasing and convex in t, and is decreasing and convex in s
- Fix s, $f(t,s)$ is maximized if $t = \frac{1}{2}\log 2\pi eP$
- Fix t, $f(t,s)$ is minimized if $s = \frac{1}{2}\log 2\pi eN$
- $X^* \sim \mathcal{N}(0,P)$, $Z \sim \mathcal{N}(0,N^*)$

In Gaussian channel

$$\begin{aligned} I(X; X+Z^*) &\leq I(X^*; X^*+Z^*) = C(Z^*) \\ & I(X^*; X^*+Z) \\ &= h(X^*+Z) - h(Z) \\ &\geq \frac{1}{2}\log\left(e^{2h(X^*)} + e^{2h(Z)}\right) - h(Z) \\ &\geq \min_s f(t,s) \\ &= I(X^*; X^*+Z^*) \end{aligned}$$

$$\boldsymbol{I(X; X+Z^*) \leq I(X^*; X^*+Z^*) \leq I(X^*; X^*+Z)}$$

$$\min_Z \max_X I(X; X+Z) = \max_X \min_Z I(X; X+Z) = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

Summary

Cover: 9.1, 9.2, 9.4