INFORMATION THEORY & CODING

Week 12 : Gaussian Channel

Dr. Rui Wang

Department of Electrical and Electronic Engineering
Southern Univ. of Science and Technology (SUSTech)

Email: wang.r@sustech.edu.cn

December 1, 2020

Multivariate Gaussian maximizes the entropy

Theorem

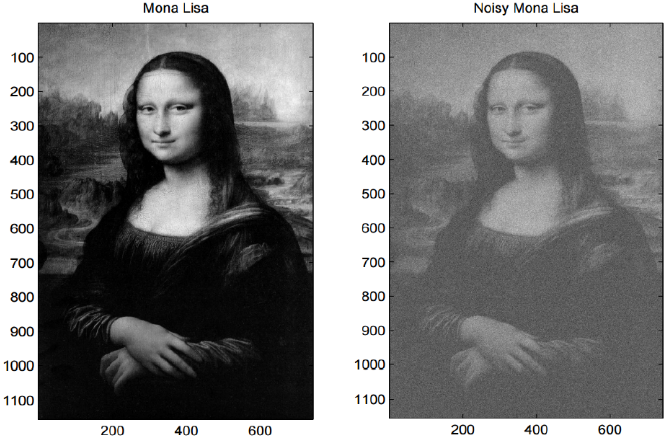
Let the random vector $\mathbf{X} \in \mathbb{R}^n$ have zero mean and covariance $K = \mathbb{E}\mathbf{X}\mathbf{X}^t$ (i.e., $K_{ij} = \mathbb{E}X_iX_j, 1 \le i, j \le n$). Then

$$h(\mathbf{X}) \le \frac{1}{2} \log(2\pi e)^n |K|$$

with equality iff $\mathbf{X} \sim \mathcal{N}(0, K)$.

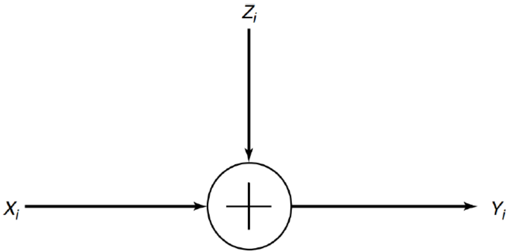
Communication with noise

- Mona Lisa in AWGN



Gaussian channel

- The most important continuous alphabet channel: Additive White Gaussian Noise (AWGN) channel
- Given the input X_i, the noise $Z_i \sim \mathcal{N}(0, N)$ independent of X_i, the channel output can be written as $Y_i = X_i + Z_i$
- a model for communication channels: wireless phone, satellite links



Channel capacity of Gaussian channel

- Intuition: $C = \log \#$ of distinguishable signals
- If $N = 0$, $C = \infty$ (receives the transmission perfectly)
- If no power constraint on the input, $C = \infty$ (can choose an infinite subset of inputs arbitrarily far apart)
- The most common limitation average power constraint: for any codeword $(x_1, x_2, \ldots x_n)$

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

Naive way of using Gaussian channel

- Binary phase-shift keying (BPSK)
- transmit 1 bit over the channel
- $1 \to x = +\sqrt{P}$, $0 \to x = -\sqrt{P}$
- $Y = \pm\sqrt{P} + Z$
- Probability of error

$$P_e = 1 - \Phi\left(\sqrt{\frac{P}{N}}\right) = Q\left(\sqrt{\frac{P}{N}}\right),$$

  where $\Phi(x)$ is the cumulative normal function of standard normal distribution:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) \mathrm{d}t.$$

- Convert Gaussian channel into a discrete BSC with $p = P_e$. Lose information in quantization, but make processing of the output signal easy.

Gaussian channel capacity

Definition

The capacity of the Gaussian channel with power constraint P is

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

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

with equality attained when $X \sim \mathcal{N}(0,P)$.

- We will show that C the the supremum of the rates achievable for AWGN. (Similar to a discrete channel)

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 $x : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, yielding codewords $x^n(1), x^n(2), \ldots, x^n(M)$, satisfying the power constraint P:

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

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

C as maximum data rate

- We will show that C the the supremum of the rates achievable for AWGN channel. (simlilar to a discrete channel)

Definition

A rate R is achievable for a Gaussian channel with a power constraint P if there exists a $(2^{nR}, n)$ codes with maximum probability of error

$$\lambda^{(n)} = \max_{i=1,2,\ldots,2^{nR}} \lambda_i \to 0 \quad \text{as} \quad n \to \infty.$$

Sphere packing

- Why we may be able to construct $(2^{nR}, n)$ codes with low probability of error?

Fix one codeword

- consider any codeword of length n
- received vector is normally distributed $\sim \mathcal{N}_n(\text{true codeword}, N\mathbf{I}_n)$
- with high probability, received vector contained in a sphere of radius $\sqrt{n(N+\epsilon)}$ around true codeword
- assign everything within a sphere to a given codeword

Sphere packing

- Why we may be able to construct $(2^{nR}, n)$ codes with low probability of error?

Fix one codeword

- consider any codeword of length n
- received vector is normally distributed $\sim \mathcal{N}_n(\text{true codeword}, N\mathbf{I}_n)$
- with high probability, received vector contained in a sphere of radius $\sqrt{n(N+\epsilon)}$ around true codeword
- assign everything within a sphere to a given codeword

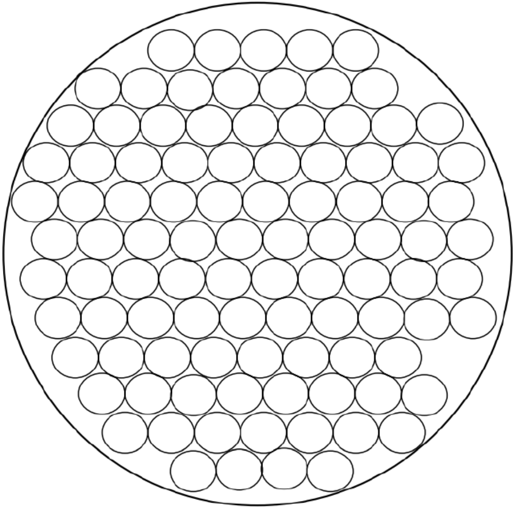
Sphere packing

- with power constraint, with high probability the space of received vectors is a sphere with radius $\sqrt{n(P+N)}$
- volume of n-dimensional sphere $= C_n r^n$, for constant C_n and radius r
- the maximum number of nonintersection decoding spheres is

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

- rate of this codebook $= \frac{\log_2(\text{size of the codewords})}{n} = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$

Sphere packing



sphere packing

Theorem

The capacity of a Gaussian channel with power constraint P and noise variance N is

$$C = \frac{1}{2}\log\left(1+\frac{P}{N}\right) \quad \textit{bits per transmission}$$

Proof.

Use the same ideas as in the proof of the channel coding theorem in the discrete case to prove:
1) achievability; 2) converse

Two main differences:
1) the power constraint P;
2) the variables are continuous

Gaussian channel capacity theorem

Theorem

The capacity of a Gaussian channel with power constraint P and noise variance N is

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \quad \textit{bits per transmission}$$

Proof.

Use the same ideas as in the proof of the channel coding theorem in the discrete case to prove:
1) achievability; 2) converse

□

Two main differences:
1) the power constraint P;
2) the variables are continuous

Gaussian channel capacity theorem

Theorem

The capacity of a Gaussian channel with power constraint P and noise variance N is

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \quad \textit{bits per transmission}$$

Proof.

Use the same ideas as in the proof of the channel coding theorem in the discrete case to prove:
1) achievability; 2) converse

$\square$

Two main differences:
1) the power constraint P;
2) the variables are continuous

New stuff in proof

- **Achievability**:
  - codeword elements generated i.i.d. according $X_j(i) \sim \mathcal{N}(0, P-\epsilon)$. So

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

  - Probability error : w.l.o.g., assume that codeword 1 was sent. Define

$$E_0 = \left\{\frac{1}{n}\sum_{j=1}^{n} X_j^2(1) > P\right\} \quad \text{and} \quad E_i = \{(X^n(i), Y^n) \text{ is in } A_\epsilon^{(n)}\}.$$

Then an error occurs if E_0 occurs or E_1^c occurs or $\cup_{i=2}^{2^{nR}} E_i$ occurs. The error probability is small according to law of large numbers.

New stuff in proof

- **Converse**: Gaussian distribution has maximum entropy. Parallel to the arguments for a discrete channel. Please read the proof in the textbook.

Bandlimited channels

- More common channel model: bandlimited continuous AWGN:

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

where "$*$" denotes convolution
$X(t)$–signal waveform
$Z(t)$–white Gaussian noise
$h(t)$–impulse response of an ideal bandpass filter, which cuts off all frequencies $> W$.

Theorem (Nyquist-Shannon Sampling Theorem)

Suppose that a function $f(t)$ is bandlimited to W, namely, the spectrum of the function is 0 for all frequencies $> W$. Then the function is completely determined by samples of the functions spaced $\frac{1}{2W}$ seconds apart.

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Capacity of continuous-time bandlimited AWGN

- Thus, in each second, the transmission can be written as $Y(nT) = X(nT) + Z(nT)$, where $T = 1/2W$ and $n = 1, 2, ..., 2W$
- Noise has power spectral density $\frac{N_0}{2}$ watts/hertz, and bandwidth W hertz. The noise has power $= \frac{N_0}{2} 2W = N_0 W$ and each of the $2WT$ noise samples in time T has variance $\frac{N_0 WT}{2WT} = \frac{N_0}{2}$.
- Signal power P watts
- $2W$ samples each second
- Channel capacity

$$
\begin{aligned}
C &= 2W \frac{1}{2} \log\left(1 + \frac{P}{N}\right) && 2W \text{ samples per second} \\
&= 2W \frac{1}{2} \log\left(1 + \frac{\frac{P}{2W}}{\frac{N_0}{2}}\right) && P \text{ per sample } \frac{PT}{2WT} = \frac{P}{2W} \\
&= W \log\left(1 + \frac{P}{N_0 W}\right) && \text{bits per second}
\end{aligned}
$$

- channel capacity

$$C = 2W\frac{1}{2}\log\left(1+\frac{P}{N}\right) \quad 2W \text{ samples per second}$$

$$= 2W\frac{1}{2}\log\left(1+\frac{\frac{P}{2W}}{\frac{N_0}{2}}\right) \quad P \text{ per sample } \frac{PT}{2WT} = \frac{P}{2W}$$

$$= W\log\left(1+\frac{P}{N_0 W}\right) \quad \text{bits per second}$$

The capacity formula of a bandlimited Gaussian channel with noise spectral density $\frac{N_0}{2}$ watts/Hz and power P watts.

- when $W \to \infty$, $C \to \frac{P}{N_0}\log_2 e$ bits per second
  For infinite bandwidth channels, the capacity grows linearly with the power.

Capacity of continuous-time bandlimited AWGN

- channel capacity

$$
\begin{aligned}
C &= 2W \frac{1}{2} \log\left(1 + \frac{P}{N}\right) && 2W \text{ samples per second} \\
&= 2W \frac{1}{2} \log\left(1 + \frac{\frac{P}{2W}}{\frac{N_0}{2}}\right) && P \text{ per sample } \frac{PT}{2WT} = \frac{P}{2W} \\
&= W \log\left(1 + \frac{P}{N_0 W}\right) && \text{bits per second}
\end{aligned}
$$

  The capacity formula of a bandlimited Gaussian channel with noise spectral density $\frac{N_0}{2}$ watts/Hz and power P watts.

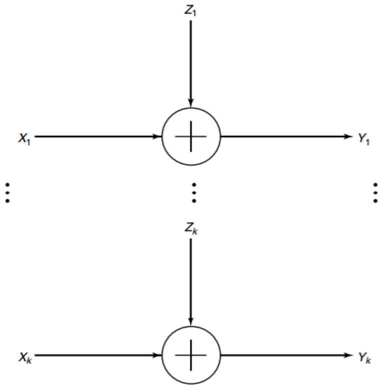
- when $W \to \infty$, $C \to \frac{P}{N_0} \log_2 e$ bits per second
  For infinite bandwidth channels, the capacity grows linearly with the power.

Example: telephone line

- telephone signals are bandlimited to 3300 Hz
- $\texttt{SNR} = 33\text{dB} : \frac{P}{N_0 W} = 2000$
- capacity $C = 36000$ bits per second
- practical modems achieve transmission rates up to 33600 bit per second uplink and downlink
- ADSL achieves $56kb/s$ downlink (asymmetric data rate)

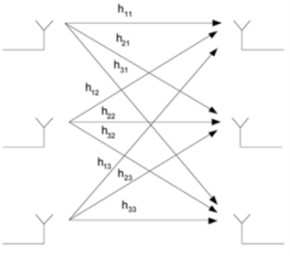
Parallel Gaussian channels

- Consider k independent Gaussian channels in parallel with a common power constraint
- Objective: to distribute the total power among the channels to maximize the capacity



Parallel channels are everywhere

- OFDM (orthogonal frequency-division multiplexing), parallel channels formed in frequency domain
- MIMO (multiple-input-multiple-output) - multiple antenna system
- DMT (discrete multi-tone systems)



Parallel independent channels

- k independent channels
- $Y_j = X_j + Z_j$, $j = 1, 2, \ldots, k$, $Z_j \sim \mathcal{N}(0, N_j)$
- total power constraint $\mathbb{E} \sum_{j=1}^{k} X_j^2 \leq P$
- Goal: distribute power among various channels to maximize the total capacity

Channel capacity

- channel capacity of parallel Gaussian channel

$$C = \max_{f(x_1,x_2,\ldots,x_k):\ \mathbb{E}\sum_{i=1}^k X_i^2 \le P} I(X_1.X_2,\ldots,X_k;Y_1,Y_2,\ldots,Y_k)$$
$$= \sum_{i=1}^{k} \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right)$$

where $P_i = \mathbb{E}X_i^2$, and $\sum_{i=1}^k P_i = P$.

- This is a standard optimization problem

$$\max_{P_1,P_2,\ldots,P_k} \sum_{i=1}^{k} \log(1+P_i/N_i)$$
$$\text{subject to } \sum_{i=1}^{k} P_i = P$$

- channel capacity of parallel Gaussian channel

$$\begin{aligned} C &= \max_{f(x_1,x_2,\ldots,x_k):\ \mathbb{E}\sum_{i=1}^k X_i^2 \le P} I(X_1.X_2,\ldots,X_k;Y_1,Y_2,\ldots,Y_k) \\ &= \sum_{i=1}^k \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right) \end{aligned}$$

where $P_i = \mathbb{E}X_i^2$, and $\sum_{i=1}^k P_i = P$.

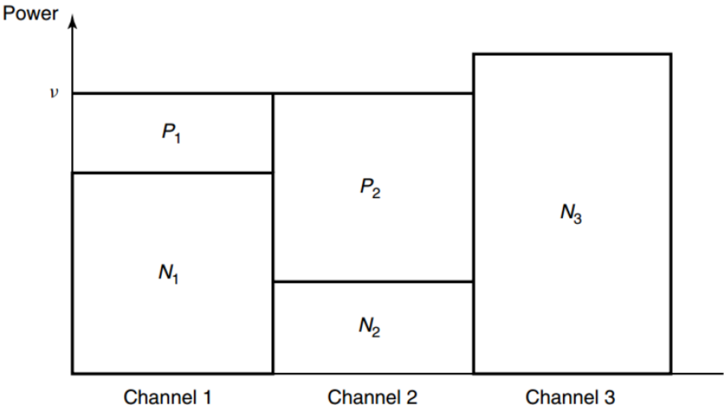
- This is a standard optimization problem

$$\max_{P_1,P_2,\ldots,P_k} \sum_{i=1}^k \log(1+P_i/N_i)$$

$$\text{subject to } \sum_{i=1}^k P_i = P$$

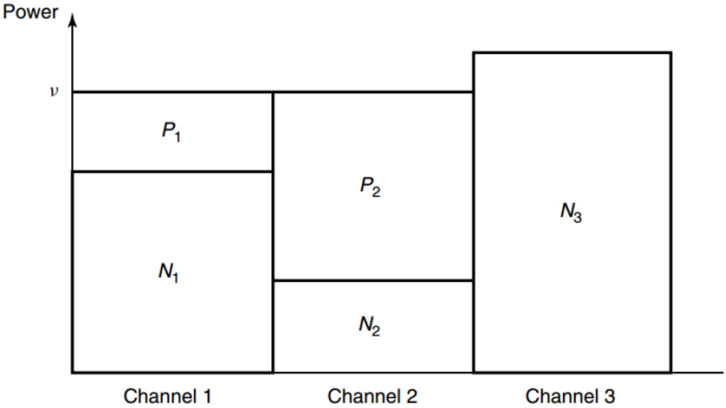
Water-filling for parallel channels

- allocate more power in less noisy channels
- very noisy channels are abandoned



Water-filling for parallel channels

- $P_i = (\nu - N_i)^+$, $(x)^+ = \max(x, 0)$
- ν is determined by power constraint: $\sum(\nu - N_i)^+ = P$



Reading & Homework

- Reading: Chapter 9.1 - 9.3
- Homework: Problems 9.4, 9.5