

ECE 587 / STA 563: Lecture 8 – Gaussian Channel

Information Theory
Duke University, Fall 2024

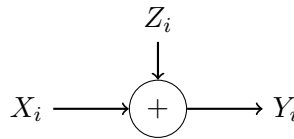
History: Written by Galen Reeves (-2023). Updated by Henry Pfister (2024-).
Last Modified: October 28, 2024

Outline of lecture:

8.1 Gaussian channel	1
8.1.1 Parallel Gaussian Channels	4

8.1 Gaussian channel

- In many real-world applications, the difference between what is sent X and what is received Y can be modeled as additive white Gaussian noise.
- The discrete time Gaussian channel is given by



where $Z_i \sim N(0, N)$ is independent of X_i .

- Without any constraints, the capacity is infinite!
- To model real-world constraint, impose *average power constraint* on codewords (x_1, x_2, \dots, x_n)

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

- **Example:** A simple strategy for communication on the AWGN channel
 - Send $X = +\sqrt{P}$ to communicate 1 and $-\sqrt{P}$ to communicate 0. This obeys the average power constraint.
 - The received signal is

$$Y = \pm\sqrt{P} + Z$$

- Since noise is symmetric, the optimal decoder is given by

$$\text{If } Y \geq 0 \implies \text{decide } +\sqrt{P}$$

$$\text{If } Y < 0 \implies \text{decide } -\sqrt{P}$$

- The probability of error is

$$\begin{aligned}
 P_e &= \mathbb{P}[\text{error} | X = \sqrt{P}] \frac{1}{2} + \mathbb{P}[\text{error} | X = -\sqrt{P}] \frac{1}{2} \\
 &= \frac{1}{2} \mathbb{P}[Z < -\sqrt{P}] + \frac{1}{2} \mathbb{P}[Z \geq \sqrt{P}] \\
 &= \mathbb{P}[Z > \sqrt{P}] \\
 &= \mathbb{P}[N(0, 1) > \sqrt{P/N}] \\
 &= 1 - \Phi(\sqrt{P/N})
 \end{aligned}$$

where $\Phi(x)$ is the CDF of the standard Gaussian distribution

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

- In this example, we have converted Gaussian channel into a discrete BSC with $p = P_e$. It turns out we can do much better (at least when the SNR P/N is large).

- **Definition:** The *information capacity* of the Gaussian channel is

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

- **Theorem:** The information capacity of the Gaussian channel with additive noise power N and power constraint P is

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

- Proof:

- The mutual information can be expressed as

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

- The maximum of the first term occurs when Y is Gaussian:

$$\begin{aligned}
 \max_{f(x) : \mathbb{E}[X^2] \leq P} h(Y) &= \max_{f(x) : \mathbb{E}[X^2] \leq P} h(X + Z) \\
 &\leq \max_{f(y) : \mathbb{E}[Y^2] \leq P+N} h(Y) \\
 &= \frac{1}{2} \log(2\pi e(N + P))
 \end{aligned}$$

- Putting everything together gives

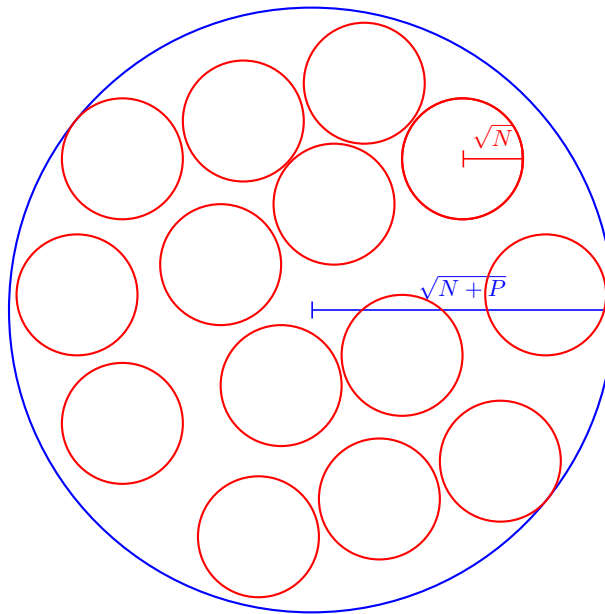
$$I(X; Y) \leq \frac{1}{2} \log \left(\frac{N + P}{N} \right) = \frac{1}{2} \log \left(1 + \frac{P}{N} \right)$$

- This holds with equality when $X \sim N(0, P)$

- **Definition:** A rate R is *achievable* for the Gaussian channel with power constraint P if there exists a sequence of $(2^{nR}, n)$ coding schemes satisfying the power constraint such that the maximal probability of error converges to zero as n becomes large. The *capacity* of the channel is the supremum of the achievable rates.
- **Theorem:** The capacity of the Gaussian channel with additive noise power N and power constraint P is equal to the information capacity:

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

- Proof:
 - (1) Achievability: $R < C \implies R$ is achievable.
 - (2) Converse: R is achievable $\implies R < C$.
- For an intuitive explanation of the channel capacity, consider *sphere packing*.



- Suppose codeword $x^n(i)$ is sent. The received vector $Y^n = x^n(i) + Z^n$ satisfies

$$\frac{1}{n} \mathbb{E}[\|Y^n - x^n(i)\|^2] = \frac{1}{n} \mathbb{E}[\|Z^n\|^2] = N$$

Thus, with high probability, the received vector is contained in a sphere of radius $\sqrt{n(N + \epsilon)}$ around the true codeword.

$$\frac{1}{\sqrt{n}} \|Y^n - x^n(i)\| \leq \sqrt{N + \epsilon}$$

- Suppose the decoder assigns everything in this region to the i -th message. An error occurs under the following events:
 - (1) the received error falls outside the ball (i.e., big noise);
 - (2) the sphere of another codeword $x^n(j)$ overlaps with the i -th sphere.

- Since the codewords satisfy an average power constraint, the average power of Y^n obeys

$$\frac{1}{n} \mathbb{E}[\|Y^n\|^2] = \frac{1}{n} \mathbb{E}[\|X^n(W) + Z^n\|^2] \leq P + N$$

and thus, with high probability,

$$\frac{1}{\sqrt{n}} \|Y^n\| \leq \sqrt{P + N + \epsilon}$$

- The number M of messages we can send reliably is given by the number of spheres of radius $\approx \sqrt{N}$ that can we pack inside sphere of radius $\approx \sqrt{P + N}$
- The volume of an n -dimensional sphere of radius r is

$$\frac{\pi^{n/2}}{\Gamma(n/2 + 1)} r^n = C_n r^n$$

- Thus, the maximum number of nonintersecting spheres is upper bounded by the ratio of the volumes

$$M \leq \frac{\text{volume large sphere}}{\text{volume of small sphere}} = \frac{C_n (N + P)^{n/2}}{C_n (N)^{n/2}} = \left(1 + \frac{P}{N}\right)^{n/2}$$

- The rate of the corresponding code is

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

8.1.1 Parallel Gaussian Channels

- Consider k independent Gaussian channels in parallel with a common power constraint.

$$\begin{aligned} Y_1 &= X_1 + Z_1 \\ Y_2 &= X_2 + Z_2 \\ &\vdots \\ Y_k &= X_k + Z_k \end{aligned}$$

- **Examples:**

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

- The noise Z_i is independent with

$$Z_i \sim N(0, N_i)$$

- The total power constraint across the channels

$$\sum_{j=1}^k P_i \leq P, \quad P_i = \mathbb{E}[X_i^2]$$

- The goal is to distribute the power amongst the channels to maximize the total capacity

- For any allocation of powers obeying the power constraint, the mutual information obeys

$$\begin{aligned}
I(X^k; Y^k) &= h(Y^k) - h(Y^k | X^k) \\
&= h(Y^k) - h(Z^k | X^k) && \text{(shift invariance)} \\
&= h(Y^k) - h(Z^k) && (Z^k \text{ independent of } X^k) \\
&= h(Y^k) - \sum_{i=1}^k \frac{1}{2} \log(2\pi e N_i) \\
&\leq \sum_{i=1}^k h(Y_i) - \sum_{i=1}^k \frac{1}{2} \log(2\pi e N_i) && \text{(independence bound)} \\
&\leq \sum_{i=1}^k \frac{1}{2} \log(2\pi e (P_i + N_i)) - \sum_{i=1}^k \frac{1}{2} \log(2\pi e N_i) && \text{(Gaussian has max entropy)} \\
&= \sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right)
\end{aligned}$$

- This upper bound is achieved when X_i are independent with

$$X_i \sim N(0, P_i)$$

and thus the capacity is

$$\begin{aligned}
C &= \max_{X^k : \sum_i \mathbb{E}[X_i^2] \leq P} I(X^k; Y^k) \\
&= \max_{P_i} \sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right), \quad \text{subject to } \sum_i P_i \leq P, \quad P_i \geq 0
\end{aligned}$$

- This is constrained optimization problem. Consider the Lagrangian

$$J(P_1, \dots, P_k, \lambda) = \sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right) - \lambda \left(\sum_i P_i - P\right)$$

- Since

$$\inf_{\lambda > 0} \left(\sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right) - \lambda \left(\sum_i P_i - P\right) \right) = \begin{cases} \sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right), & \sum_i P_i \leq P \\ -\infty, & \sum_i P_i > P \end{cases}$$

the objective function can be expressed as

$$C = \sup_{P_i \geq 0} \inf_{\lambda \geq 0} J(P_1, \dots, P_k, \lambda)$$

- Swapping the order of the max and sup provides an *upper bound*:

$$C \leq \inf_{\lambda \geq 0} \sup_{P_i \geq 0} J(P_1, \dots, P_k, \lambda)$$

- For each λ , the derivative of $J(P_1, \dots, P_k, \lambda)$ with respect to P_i is given by

$$\frac{\partial}{\partial P_i} J(P^k, \lambda) = \frac{1}{2(N_i + P_i)} - \lambda$$

thus

$$P_i < \frac{1}{2\lambda} - N_i \implies J(P_1, \dots, P_k, \lambda) \text{ is increasing with respect to } P_i$$

$$P_i > \frac{1}{2\lambda} - N_i \implies J(P_1, \dots, P_k, \lambda) \text{ is decreasing with respect to } P_i$$

Recalling that P_i are nonnegative, it thus follows that $J(P_1, \dots, P_k, \lambda)$ is maximized when

$$P_i^*(\lambda) = \begin{cases} \frac{1}{2\lambda} - N_i, & \frac{1}{2\lambda} - N_i > 0 \\ 0, & \frac{1}{2\lambda} - N_i \leq 0 \end{cases} = \max\left(\frac{1}{2\lambda} - N_i, 0\right)$$

and thus, for each $\lambda > 0$,

$$\max_{P_i \geq 0} J(P_1, \dots, P_k, \lambda) = \sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i^*(\lambda)}{N_i}\right) - \lambda \left(\sum_i P_i^*(\lambda) - P\right)$$

- The above expression provides an upper bound on C . Let λ^* be the value of λ such the power constraint is met with equality, i.e.

$$\sum_i P_i^*(\lambda^*) = P$$

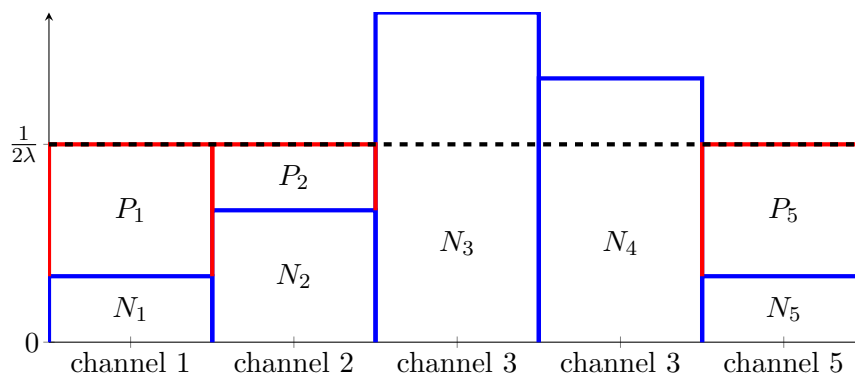
This value is guaranteed to exist because $P_i^*(\lambda)$ is continuous and increasing in λ . Then,

$$C \leq J(P_1^*(\lambda^*), \dots, P_k^*(\lambda^*), \lambda^*) = \sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i^*(\lambda^*)}{N_i}\right)$$

- By construction, the power allocation $P_i^*(\lambda^*)$ satisfies the constraints of our problem and thus is a feasible solution. Since it matches the upper bound on the objective function, we know that it is the optimal solution. We conclude that

$$C = \sum_{i=1}^k \frac{1}{2} \log\left(1 + \frac{P_i^*(\lambda^*)}{N_i}\right)$$

- This solution is known as *water filling*.



- This same story also holds true if the noise is correlated between channels. We apply water filling to the the *eigenvalues* of the covariance matrix K of the noise.