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.

• Illustration of discrete time Gaussian channel

\[
\begin{align*}
Z_i \\
X_i & \xrightarrow{+} Y_i
\end{align*}
\]

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

• Without any constraints, the capacity is infinite!

• To mode real-world constraint, impose average power constraint on codewords $(x_1, x_2, \cdots, x_n)$

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

• 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

\[
\begin{align*}
\text{If } Y \geq 0 & \implies \text{decide } +\sqrt{P} \\
\text{If } Y < 0 & \implies \text{decide } -\sqrt{P}
\end{align*}
\]
The probability of error is

\[ P_e = P[\text{error} | X = \sqrt{P}] \frac{1}{2} + P[\text{error} | X = -\sqrt{P}] \frac{1}{2} \]

\[ = \frac{1}{2} P[Z < -\sqrt{P}] + \frac{1}{2} P[Z \geq \sqrt{P}] \]

\[ = P[Z > \sqrt{P}] \]

\[ = P[N(0,1) > \sqrt{P/N}] \]

\[ = 1 - \Phi(\sqrt{P/N}) \]

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 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
  
  \[ I(X; Y) = h(Y) - h(Y|X) \]
  
  \[ = h(Y) - h(X + Z|X) \]
  
  \[ = h(Y) - h(Z) \]
  
  \[ = h(Y) - \frac{1}{2} \log(2\pi eN) \]

- The maximum of the first term occurs when \( Y \) is Gaussian:
  
  \[ \max_{f(x) : \mathbb{E}[X^2] \leq P} h(Y) = \max_{f(x) : \mathbb{E}[X^2] \leq P} h(X + Z) \]
  
  \[ \leq \max_{f(y) : \mathbb{E}[Y^2] \leq P+N} h(Y) \]
  
  \[ = \frac{1}{2} \log(2\pi e(N + P)) \]

- 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 $(2^n R, n)$ 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:

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

• **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 $Y^n = x^n(i) + Z^n$ obeys

    $$E[\|Y^n - x^n(i)\|^2] = E[\|Z^n\|^2] = nN$$

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

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

  - If our decoder assigns everything in this region to the $i$-th message, then 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 obey an average power constraint, the average power of $Y^n$ obeys

    $$E[\|Y^n\|^2] = E[\|X^n(W) + Z^n\|^2] \leq nP + nN$$

    and thus, with high probability,

    $$\|Y^n\| \leq \sqrt{n(P + N + \epsilon)}$$

  - 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 $\approx \sqrt{P + N}$

  - 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

    $$M \leq \frac{\text{volume large sphere}}{\text{volume of small sphere}} = \frac{C_n (n(N + P))^{n/2}}{C_n (nN)^{n/2}} = \left( 1 + \frac{P}{N} \right)^{n/2}$$
The rate of the corresponding code is
\[ R = \frac{\log M}{n} = \frac{1}{n} n \log \left( 1 + \frac{P}{N} \right) = \frac{1}{2} \log \left( 1 + \frac{P}{N} \right) \]

8.1.1 Parallel Gaussian Channels

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

\[
\begin{align*}
Y_1 &= X_1 + Z_1 \\
Y_2 &= X_2 + Z_2 \\
&\vdots \\
Y_k &= X_k + Z_k
\end{align*}
\]

- 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

\[
I(X^k; Y^k) = h(Y^k) - h(Y^k | X^k) = h(Y^k) - h(Z^k | X^k) = h(Y^k) - h(Z^k)
\]

(shift invariance)

\[
= h(Y^k) - \sum_{i=1}^{k} \frac{1}{2} \log(2\pi e N_i)
\]

\[
\leq \sum_{i=1}^{k} h(Y_i) - \sum_{i=1}^{k} \frac{1}{2} \log(2\pi e N_i)
\]

(independence bound)

\[
\leq \sum_{i=1}^{k} \frac{1}{2} \log(2\pi e(P_i + N_i)) - \sum_{i=1}^{k} \frac{1}{2} \log(2\pi e N_i)
\]

(Gaussian has max entropy)

\[
= \sum_{i=1}^{k} \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right)
\]

• This upper bound is achieved when \( X_i \) are independent with

\[
X_i \sim N(0, P_i)
\]

and thus the capacity is

\[
C = \max_{X^k : \sum_i E[X_i^2] \leq P} I(X^k; Y^k)
\]

\[
= \max_{P_i} \sum_{i=1}^{k} \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right), \quad \text{subject to} \quad \sum_i P_i \leq P, \quad P_i \geq 0
\]

• This is constrained optimization problem. Consider the Lagrangian

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

Since

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

the objective function can be expressed as

\[
C = \sup_{P_i \geq 0} \inf_{\lambda \geq 0} J(P_1, \cdots, P_k, \lambda)
\]

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

\[
C \leq \inf_{\lambda \geq 0} \sup_{P_i \geq 0} J(P_1, \cdots, P_k, \lambda)
\]
• For each $\lambda$, the derivative of $J(P_1, \cdots, P_k, \lambda)$ with respect to $\lambda_i$ is given by

$$\frac{\partial}{\partial P_i} J(P_1, \cdots, P_k, \lambda) = \frac{1}{2 N_i + P_i} - \lambda$$

thus

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

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

Recalling that $P_i$ are nonnegative, it thus follows that $J(P_1, \cdots, P_k, \lambda)$ 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 $\lambda > 0,$

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

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

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

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

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

• 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**.

![Diagram](image_url)

• 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.