Gaussian Channel

Information Theory

Antonio Sirignano

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione
Università degli studi Federico II di Napoli

October 15, 2025

Overview

1. **Gaussian Channel's Generality**
2. **Gaussian Channel Capacity**
3. **Implementation and Simulation**

Introduction

- The Gaussian channel is a time-discrete channel characterized by the input relationship at time i

$$Y_i = X_i + Z_i, \quad \mathcal{N}(0, N)$$

where Z_i's are i.i.d random variable which are assumed indipendent of the signal X_i

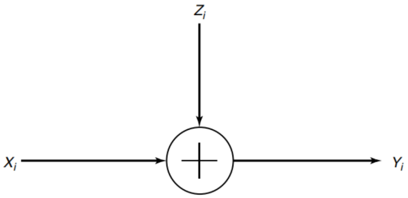

Figure: Gaussian Channel

- If the noise vairance is zero or the input is uncostrained, the capacity if the channel is infinity.

Introduction

- The most common limitation on the input is the power constraint, hence we assume an average power constraint.
  For any trasmitted codeword $(x_1, x_2, \ldots, x_n)$ over the channel, it requires that

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

- For example, assume that we want to sent one binary digit over the channel for each use of it. Given the power constraint, the best solution is to send one of two levels, $+\sqrt{P}$ or $-\sqrt{P}$.
- The receiver observes at the corresponding Y and tries to decide which of the two level was sent.

Introduction

- Assuming that both levels are equally likely and choosing the optimum decoding rule is to decide that $+\sqrt{P}$ was sent if $Y > 0$ and $-\sqrt{P}$ was sent if $Y < 0$, we can evaluate the probability of error with such a decoding schema:

$$\begin{aligned} P_e =& \frac{1}{2} \Pr\left(Y < 0 \mid X = +\sqrt{P}\right) + \frac{1}{2} \Pr\left(Y > 0 \mid X = -\sqrt{P}\right) \\ =& \frac{1}{2} \Pr\left(Z < -\sqrt{P}\right) + \frac{1}{2} \Pr\left(Z > \sqrt{P}\right) \\ =& \Pr\left(Z > \sqrt{P}\right) = 1 - \Phi\left(\sqrt{\frac{P}{N}}\right) \end{aligned}$$

where $\Phi(x)$ is. the comulative normal function $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$.

Information Capacity

- We define the information capacity of Gaussian channel as the maximum of the mutual information between the input and output over all distributions on the input that satisfy the average power constraint:

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

where P is the power constraint.

- We can observe that

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

being Z indipendent if X and the average does not effect the entropy.

Information Capacity

- We know that $h(Z) = \frac{1}{2} \log 2\pi e N$; moreover

$$E[Y^2] = E[(X + Z)^2] = E[X^2] + 2E[X]E[Z] + E[Z^2] = P + N$$

- As a consequence, the entropy of Y is boudned by $\frac{1}{2} \log 2\pi e(P + N)$, implying that

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

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

Information Capacity

Definition

An (M, n) code for a Gaussian channel with power constraint P is characterized by:

1. An encoding function

$$x : \{1, 2, \ldots, M\} \to \mathcal{X}^n,$$

   where M the number of messages to deliver, yielding codewords $x^n(1), x_2^n, \ldots, x^n(M)$, satisfying the power constraint, i.e., $\sum_{i=1}^n x_i^2(w) \leq nP, w = 1, 2, \ldots, M$.

2. A decoding function

$$g = \mathcal{Y}^n \to \{1, 2, \ldots, M\}$$

Information Capacity: Shannon's Second Theorem

- 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 λ_n tends to zero.
- The capacity of the channel is the supremum of the achievable rates.

Theorem

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

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

Information Capacity: Shannon's Second Theorem

- Consider the trassmission of any codeword of lenght n.
- The received vecotr is normally dstributed with mean equal to the true codeword and variance equal to the noice variance.
- Intuitively, we cane say that the received vector is contained in a sphere of radius $\sqrt{n(N+\epsilon)}$ around the true codeword.
- Then when at the sendig, there will be an error if the received vector is not in the sphere (with low probability).
- But, if we hava a set of n codewords, we have to consider a valume of a n-dimensional sphere with form $C_n r^n$, where r is the radius of the sphere and C_n is a constant depending on the space dimensionality.

Information Capacity: Shannon's Second Theorem

- The received vectors have energy no greater than $n(P + N)$.
- Then, they are in a sphere of radius $\sqrt{n(P+N)}$. So the maximum numver of nonitersecting decoding sphere in this volume can not exceed

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

- Hence, the rate of the code is $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$.
- This idea is called sphere packing.

Information Capacity: Achievability Proof of Shannon's Second Theorem

- First steps:
  1. Codebook generation: let $X_i(w), i = 1, 2, \dots, n$, $w = 1, 2, \dots, 2^{nR}$, be i.i.d $\sim \mathcal{N}(0, P - \epsilon)$, forming codewords $X^n(1), X^n(2), \dots, X^n(2^{nR}) \in \mathcal{R}^n$. Since $\frac{1}{n} \sum X_i^2 \to P - \epsilon$, the probability that a codeword does not satisfy the power constraint will be arbitrary very small.
  2. Encoding: after the generation of the codebook, it is revealed to both the sender and the receiver. To send the message index w, the trasmitter send the w-th codeword $X^n(w)$ of the codebook.
  3. Decoding: the receiver search on the codebook 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 $\hat{W} = w$ to be the trasmitted codeword;
     - Otherwise, the receiver declares an error. It is declared an error also if the chosen codeword does not sotisfy the power constraint.

Information Capacity: Achievability Proof of Shannon's Second Theorem

- Without loss of generality, assume that codeword 1 was sent; so

$$\mathbf{Y}^n = \mathbf{X}^n(1) + \mathbf{Z}^n.$$

- Define the following error events

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

and

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

- Then an erros occurs if E_0 occurs, i.e., the power constraint is violeted, or E_1^c occurs, i.e., the transimetted codeword and the receiver sequence are not jointly typical, or $E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}$ occurs, i.em some wrong codewords are jointly typical with the receiver sequence, regardless of theri power constraint.

Information Capacity: Achievability Proof of Shannon's Second Theorem

- Let $\mathcal{E}$ denote the event $\left\{\hat{W} \neq W\right\}$. Then

$$\Pr(\mathcal{E}) = \Pr(\mathcal{E} \mid W = 1) = \Pr(E_0 \cup E_1^c \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}) \leq \Pr(E_0) + \Pr(E_1^c) + \sum_{i=2}^{2^{nR}} \Pr(E_i)$$

- By the law of large numbers,

$$P(E_0) \to 0 \mathrm{as} n \to \infty$$

  and by the joint AEP

$$P(E_1^c \leq \epsilon), \mathrm{for} n \mathrm{large\ enough.}$$

- Moreover Y^n are indipendent, because induced by $X^n(1)$ and $X^n(i)$. Hence, by the joint AEP, the probability that $X^n(i)$ and Y^n will be jointly typical is $\leq 2^{-n(I(X;Y)-3\epsilon)}$.

Information Capacity: Achievability Proof of Shannon's Second Theorem

- Now let W be uniformly distributed over $\left\{1, 2, \ldots, 2^{nR}\right\}$, and consequently,

$$\Pr(\mathcal{E}) = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \lambda_i = P_e^{(n)}.$$

- Then, for n sufficiently large and $R < I(X;Y) - 3\epsilon$, we have

$$P_e^{(n)} = \Pr(\mathcal{E}) = \Pr(\mathcal{E} \mid W = 1) \leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i) \leq$$

$$\leq \epsilon + \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)} = 2\epsilon + \left(2^{nR} - 1\right) 2^{-n(I(X;Y)-3\epsilon)} \leq 3\epsilon$$

- This allows to prove the existence of a good $(2^{nR}, n)$ code.
- The power constraint is satisfied by each of the selected codeword, but each codeword that does not sotisfy the power constraint is characterized by a conditional

Information Capacity: Converse Proof of Shannon's Second Theorem

- Let us shown noew that $R > \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ is unfeasible, i.e., that the achievable rate cannot exceed the capacity.
- This mean that if $P_e^{(n)} \to 0$ for a sequence of $(2^{nR}, n)$ code for a Gaussian channel with power constraint P, then

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

- Consider any $(2^{nR}, n)$ code that sotisfies the power constraint, that is,

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

Information Capacity: Converse Proof of Shannon's Second Theorem

- Let W be uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$, which induces as distribution ont he input codewords, which in turn induces marginal distributions over the input alphabet. More in general, a joint distribution on $W \to X^n(W) \to Y^n \to \hat{W}$ is specified.
- Now, recalling that $H(W \mid \hat{W}) \leq 1 + nRP_e^{(n)} = n\epsilon_n$, with $\epsilon_n = \left(\frac{1}{n} + RP_e^{(n)}\right) \to 0$, since $P_e^{(n)} \to 0$ as $n \to \infty$, we have

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

Information Capacity: Converse Proof of Shannon's Second Theorem

- Now let P_i be the average power of the i-th column of the codebook, that is

$$P_i = \frac{1}{2^{nR}} \sum_{e} x_i^2(w)$$

it follows $E\left[Y_i^2\right] = P_i + N$ and based on (1)

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

Information Capacity: Converse Proof of Shannon's Second Theorem

- Being $\frac{1}{n}\sum_i P_i \leq P$, i.e., each codeword has power smaller than or equal to P, and $f(x) = \frac{1}{2}\log(1+x)$ a concave and strictly increasing function of x, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(1\frac{P}{N}\right) \leq \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{N}\right) \leq \frac{1}{2}\log\left(1+\frac{P}{N}\right).$$

- Thus

$$R \leq \frac{1}{2}\log\left(1\frac{P}{N}\right) + \epsilon_n, \quad \epsilon_n \to 0.$$

- The proof is completed.

Implementation and Simulation

- Now we want to implement a Gaussian Channel, that fallows that design presented in the image (1) and the following equation

$$Y_i = X_i + Z_i, \quad \mathcal{N}(0, N)$$

.

- Our input signal is

$$f(t) = \sin(2\pi 5t)$$

that represents a sinusoidal wave with frequency $f = 5$ Hz.

- We will see how the output signal varies as the noise variance changes.

Implementation 1/4

```python
import numpy as np
import matplotlib.pyplot as plt

def gaussian_channel(signal, noise_std):
    """
    Adds white Gaussian noise (AWGN) to a signal,
    specifying the noise standard deviation.

    Parameters:
    -----------
    signal : np.ndarray
        Input signal.
    noise_std : float
        Standard deviation of the Gaussian noise.
```

Implementation 2/4

```
    Returns:
    --------
    noisy_signal : np.ndarray
        Signal after passing through the Gaussian channel.
    """
    noise = np.random.normal(0, noise_std, size=signal.shape)
    return signal + noise

# Test signal
t = np.linspace(0, 1, 1000)
signal = np.sin(2 * np.pi * 5 * t)  # 5 Hz sine wave

# Nois Levels
noise_levels = [0.05, 0.2, 0.5]  # low, medium, high noise
```

Implementation 3/4

```
# Generation of noisy signals
noisy_signals = [gaussian_channel(signal, std) for std in noise_levels]

# Noisy signal plotting
fig, axes = plt.subplots(3, 1, figsize=(12, 10), sharex=True)

for ax, noisy, std in zip(axes, noisy_signals, noise_levels):
    ax.plot(t, signal, label='Original Signal', linewidth=2, color='black')
    ax.plot(t, noisy, label=f'Noisy Signal  = {std}', color='red', alpha=0.5
    ax.set_ylabel('Amplitude')
    ax.set_title(f'Gaussian Channel with Noise  = {std}')
    ax.legend()
    ax.grid(True)

axes[-1].set_xlabel('Time [s]')
```

Implementation 4/4

```
plt.tight_layout()
plt.show()
```

Results

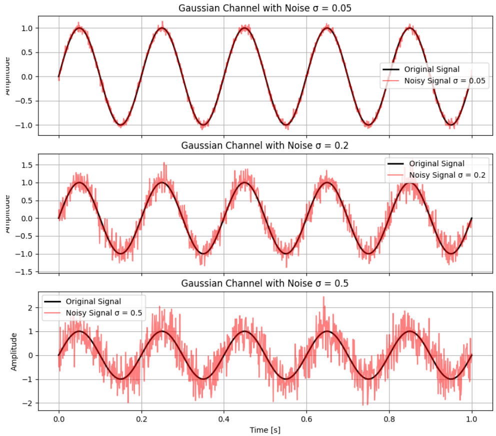

Figure: Results

The End