In the last Chapters, we learned…..

How to extend the concepts of entropy, Kullback Leibler distance and mutual information to continuous random variable cases.

```
Continuous values
```

Now, we are ready to derive Channel Coding Theorem for the Gaussian channel case:

1. There exists a \( (2^{nR}, n) \) rate \( R \) code such that the maximum error probability \( \lambda^{(n)} \) can be made arbitrarily small, if the code rate is lower than the capacity \( R < C \).
2. Conversely, any \( (2^{nR}, n) \) rate \( R \) code that can achieve arbitrarily small \( \lambda^{(n)} \) must satisfy \( R < C \).

We start with the derivation of the channel coding theorem in the Gaussian channel.
Outline

1. Gaussian Channel
   - Definition
2. Capacity of Gaussian Channel
3. Channel Coding Theorem for Gaussian Channel
4. Parallel Channels
   - Water Filling Technique

Gaussian Channel (1)

Definition 10.1.1: Gaussian Channel Model
If the channel’s output $Y$ is given as a sum of its input $X$ and noise $Z$, as

$$ Y = X + Z $$

where all variables $Y$, $X$, and $Z$ take continuous values,

and if the noise term $Z$ follows an i.i.d. Gaussian distribution with variance $N$, $N(0, N)$, the channel is called Gaussian channel.

Definition 10.1.2: Power Constraint on Code Word
Message is transmitted in the form of code word $(x_1, x_2, \ldots, x_n)$. Any code word has to satisfy the power limitation, defined as:

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

where $P$ is the power of the code words.
**Theorem 10.1.1: Error Probability**

For binary message, i.e., \(X = 1\) or \(-1\), probability of error in the Gaussian channel is given by:

\[
P_e = 1 - \Phi\left(\frac{P}{\sqrt{N}}\right)
\]

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

where the appearance of the binary message is assumed to be equiprobable.

**Proof:**

\[
P_e = \frac{1}{2} \text{Pr}(Y < 0 \mid X = \sqrt{P}) + \frac{1}{2} \text{Pr}(Y > 0 \mid X = -\sqrt{P})
\]

\[
= \frac{1}{2} \text{Pr}(Z < -\sqrt{P} \mid X = \sqrt{P}) + \frac{1}{2} \text{Pr}(Z > -\sqrt{P} \mid X = -\sqrt{P}) = \text{Pr}(Z > \sqrt{P}) = 1 - \Phi\left(\frac{P}{\sqrt{N}}\right)
\]

---

**Theorem 10.2.1: Capacity**

Capacity of the Gaussian channel is given by:

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

**Proof:**

By definition, \(C = \max_{p(x)} I(X;Y)\)

where account has been taken of the power constraint.

Now, \(I(X;Y) = h(Y) - h(Y \mid X) = h(Y) - h(X + Z \mid X) = h(Y) - h(Z \mid X)\) (Theorem 9.4.6)

However, since \(Z\) is independent of \(X\), \(h(Z \mid X) = h(Z)\)

Therefore, \(I(X;Y) = h(Y) - h(Z) = h(Y) - \frac{1}{2} \log 2\pi e N\)

Also, we know that

\[
EY^2 = E(X + Z)^2 = EX^2 + 2\frac{EXZ}{-0 \text{ because they are independent}} + EZ^2 = P + N
\]

\(h(y)\) is maximized if \(Y\) follows the Gaussian distribution among those distributions having the same variance (because of Theorem 9.4.7).
Proof *Continued*

Therefore, \( h(Y) = \frac{1}{2} \log 2 \pi e (P + N) \)

Combining all, we have:

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

Hence, \( C = \max_{p(x)} \max_{E^Y \leq P} I(X;Y) = \frac{1}{2} \log (1 + \frac{P}{N}) \)

---

**Definition 10.3.1: System Model**

Assume that a message \( W \) to be transmitted over the Gaussian channel is drawn from the index set \( \{1, 2, \ldots, M\} \), i.e., each message is indexed by a number from this set. The following summarizes how the communication system we analyze works:

1. The message \( W \) is encoded into a continuous code word \( x^j(W), W = 1, 2, \ldots, M \), satisfying the power constraint:
   \[
   \frac{1}{n} \sum_{j=1}^{n} x_j^2 (w) \leq P
   \]

2. Given the received sample vector \( y^j \), the receiver finds the message considered most likely to have been transmitted. The receiver’s operation described above is denoted as \( g(y^j) \).

3. If \( g(y^j) \neq W \), decoding error happens.
Theorem 10.3.1: Channel Coding Theorem

(1) There exists a \( (2^{nR}, n) \) rate \( R \) code satisfying the power constraint such that the decoding error probability \( P_e(n) \) can be made arbitrarily small, if the code rate is lower than the capacity \( R < C \), with
\[
C = \frac{1}{2} \log(1 + \frac{P}{N})
\]

(2) Conversely, any \( (2^{nR}, n) \) rate \( R \) code that can achieve arbitrarily small \( P_e(n) \) must satisfy \( R < C \).

Proof of (1):
The proof uses the properties of random coding, as in the finite alphabet case. The detailed proof is far beyond the expected level of this course. Instead, a proof outline is described below:

The received vector is normally distributed with mean equal to the true code word, and the variance equal to the noise variance. With the high probability, the received vector is contained within a sphere of radius \( \sqrt{n(N + \varepsilon)} \)

where \( \varepsilon \) is an arbitrarily small constant.

Proof of (1) Continued-1:
If we assign every code word distinctly in position within this sphere, then, error happens if the received vector falls outside the sphere corresponding to the transmitted code word. However, this probability can be made arbitrarily small, if the code rate \( R \) is determined.

(The error probability is bounded by a multiple of \( \varepsilon \). However, proof of this requires a bounding technique using asymptotic equi-partition property (AEP)).

The volume of an \( n \)-dimensional sphere is given in a form of \( A_n r^n \), where \( A_n \) is a constant. Each code word’s decoding sphere has radius \( \sqrt{nN} \). The received vector has energy no larger than \( n(N + P) \), which means that the whole sphere has its radius \( \sqrt{n(N + P)} \).

Now, we know that the each code word’s decoding sphere has a volume \( A_n (nM)^{n/2} \), and the whole sphere has a volume \( A_n (n(N+P))^{n/2} \). Therefore, it can be concluded that the maximum number \( 2^{nR} \) of the code word than the ratio of the volume, given by
\[
2^{nR} \leq \frac{A_n (n(P + N))^{n/2}}{A_n (nN)^{n/2}} = 2^{\frac{1}{2} \log(1 + \frac{P}{N})}
\]
the decoding spheres for those code words do not intersect each other.
Hence, the decoding error can be made as small as a multiple of \( \epsilon \), which is arbitrarily small, if the code rate satisfy:

\[
R \leq C = \frac{1}{2} \log(1 + \frac{P}{N})
\]

Proof of (1) Continued-2:

Similarly to the proof for Channel Coding Theorem for Finite Alphabet case, we apply Fano’s inequality to the conditional entropy \( H(W|Y^n) \):

\[
H(W|Y^n) \leq 1 + nR \epsilon_e(n) = n \epsilon_e \quad \text{with} \quad \epsilon_e = (\frac{1}{n} + R \epsilon_e^{(n)})
\]

Hence, \( \epsilon_e \to 0 \), as \( \epsilon_e^{(n)} \to 0 \) and \( n \to \infty \).

Then,

\[
nR = H(W) = I(W;Y^n) + H(W|Y^n) \leq I(W;Y^n) + n \epsilon_e \leq I(X^n;Y^n) + n \epsilon_e
\]

\[
= h(Y^n) - h(Y^n|X^n) + n \epsilon_e = h(Y^n) - h(Z^n) + n \epsilon_e \leq \sum_{i=1}^{n} h(Y_i) - h(Z_i) + n \epsilon_e
\]

\[
= \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Z_i) + n \epsilon_e = \sum_{i=1}^{n} I(X_i;Y_i) + n \epsilon_e
\]
Channel Coding Theorem for Gaussian Channel (5)

Proof of (2) Continued:
Let the average power of the $i$th column of the codebook be denoted as:

$$\frac{1}{2\pi} \sum_{\omega} X_i^j(w) = P_i$$

Since $X_i$ and $Z_i$ are independent, $Y_i$ has an average power $P_i+N$. Entropy of $Y_i$ is maximized when it has normal distribution, as

$$h(Y_i) \leq \frac{1}{2} \log 2\pi e (P_i + N)$$

Thus,

$$nR \leq \sum_{i=1}^{n} \left[ h(Y_i) - h(Z_i) \right] + n\varepsilon_n \leq \sum_{i=1}^{n} \left\{ \frac{1}{2} \log 2\pi e (P_i + N) - \frac{1}{2} \log 2\pi e N \right\} + n\varepsilon_n$$

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

Because of the concavity of the logarithmic function,

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

Therefore,

$$R \leq \frac{1}{2} \log(1 + \frac{P}{N}) + \varepsilon_n \rightarrow \frac{1}{2} \log(1 + \frac{P}{N}) = C, \quad \text{as} \quad P^{\varepsilon_n} \rightarrow 0 \quad \text{and} \quad n \rightarrow \infty$$

Parallel Gaussian Channels (1)

Definition 10.4.1: Parallel Gaussian Channels

Assume that there are independent $k$ channels, each modeled as Gaussian channel, as

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

where all variables $Y_j$, $X_j$, and $Z_j$ take continuous values, and the noise term $Z_j$ follows an i.i.d. Gaussian distribution with variance $N_j$. $\mathcal{N}(0, N_j)$.

The channel has a constraint on the total power, which is:

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

This channel is called parallel Gaussian channel.

Note: Each sub-channel has different noise variance.
Definition 10.4.2: Capacity
Capacity of the parallel Gaussian channel is defined as:

\[ C = \max_{f(x_1, x_2, \ldots, x_k)} \sum_{x_1, x_2, \ldots, x_k} I(X_1, X_2, \ldots, X_k; Y_1, Y_2, \ldots, Y_k) \]

Theorem 10.4.1: Water Filling Theorem
The mutual information in the capacity equation for the parallel Gaussian channel is maximized if:

\[ P_i = (v - N_i)^+ = \begin{cases} v - N_i & \text{if } v - N_i \geq 0 \\ 0 & \text{if } v - N_i < 0 \end{cases} \]

where \( v \) is determined so that

\[ \sum_{j=1}^{k} (v - N_i)^+ = P \]

Proof:
\[ I(X_1, X_2, \ldots, X_k; Y_1, Y_2, \ldots, Y_k) = h(Y_1, Y_2, \ldots, Y_k) - h(Y_1, Y_2, \ldots, Y_k|X_1, X_2, \ldots, X_k) \]
\[ = h(Y_1, Y_2, \ldots, Y_k) - h(Z_1, Z_2, \ldots, Z_k|X_1, X_2, \ldots, X_k) = h(Y_1, Y_2, \ldots, Y_k) - h(Z_1, Z_2, \ldots, Z_k) \]
\[ \leq \sum_{j=1}^{k} \frac{1}{2} \log(1 + \frac{P_i}{N_i}) \]
By introducing Lagrange multiplier, the index function that should be maximized is given by:
\[ J(P_1, P_2, \ldots, P_k) = \sum_{j=1}^{k} \frac{1}{2} \log(1 + \frac{P_i}{N_i}) + \lambda \sum_{j=1}^{k} P_i \]
Taking the first derivative of this index function and setting the result at zero’s, we have
\[ \frac{1}{2} \frac{1}{P_i + N_i} + \lambda = 0, \quad \text{or} \quad P_i = v - N_i \text{ with } v = \frac{1}{2\lambda} \]
The parameter \( v \) is determined so that the power constraint is satisfied.
This theorem states that we should allocate more power to the channels having lower noise power, and do not allocate any power to those too noisy channels; By NOT allocating any powers to the too noisy channels, even more powers should be allocated to the less noisy channels. By doing this, we can maximize the mutual information of each channel, resulting in the capacity as a whole of the parallel channels.

Summary

We have visited.....

1. Gaussian Channel
   - Definition
2. Capacity of Gaussian Channel
3. Channel Coding Theorem for Gaussian Channel
4. Parallel Channels
   - Water Filling Technique