Lecture 12 Continuous Channel

Textbook 9.1-9.5

December 24 and 26, 2024

The Guassian channel is a time-discrete channel with output Y_i at time i, where Y_i is the sum of the input X_i and the noise Z_i. The noise Z_i is drawn i.i.d from a Gaussian distribution with variance N. Thus,

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

The noise Z_i is assumed to be independent of the signal X_i.

- This channel is a model for some common communication channels, such as wired and wireless telephone channels and satellite links.
- Without further conditions the capacity of this channel may be infinite.
- If the noise variance is zero, the receiver receives the transmitted symbol perfectly.
- Since X can take on any real value, the channel can transmit an arbitrary real number with no error.

- If the noise variance is nonzero and there is no constraint on the input, we can choose an infinite subset of inputs arbitrarily far apart, so that they are distinguishable at the output with arbitrarily small probability of error.
- Such a scheme has an infinite capacity as well.
- Thus if the noise variance is zero or the input is unconstrained, the capacity of the channel is infinite.

- The most common limitation on the input is an energy or power constraint.
- We assume an average power constraint.
- For any codeword $(x_1, x_2, \cdots, x_n)$ transmitted over the channel, we require that

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

- This communication channel models many practical channels, including radio and satellite links.
- The additive noise in such channels may be due to a variety of causes. However, by the central limit theorem, the cumulative effect of a large number of small random effects will be approximately normal, so the Gaussian assumption is valid in a large number of situations.

- We first analyze a simple suboptimal way to use this channel.
- Assume that we want to send 1 bit over the channel in one use of the channel.
- Given the power constraint, the best that we can do is to send one of two levels, $+\sqrt{P}$ or $-\sqrt{P}$.
- The receiver looks at the corresponding Y received and tries to decide which of the two levels was sent. Assuming that both levels are equally likely (this would be the case if we wish to send exactly 1 bit of information), the optimum decoding rule is to decide that $+\sqrt{P}$ was sent if $Y > 0$ and decide $-\sqrt{P}$ was sent if $Y < 0$.

- The probability of error with such a decoding scheme is

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

- Here $\Phi(x)$ is the cumulative normal function

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

- Using such a scheme, we have converted the Gaussian channel into a discrete binary symmetric channel with crossover probability P_e.
- Similarly, by using a four-level input signal, we can convert the Gaussian channel into a discrete four-input channel.
- In some practical modulation schemes, similar ideas are used to convert the continuous channel into a discrete channel.
- The main advantage of a discrete channel is ease of processing of the output signal for error correction, but some information is lost in the quantization.

Definition

The information capacity of the Gaussian channel with power constraint P is

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

Information capacity of the Gaussian channel

We can calculate the information capacity as follows: Expanding $I(X;Y)$, we have

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

Now, $h(Z) = \frac{1}{2}\log 2\pi e N$. Also, since X and Z are independent and $EZ = 0$, we have that

$$EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2 = P + N.$$

Given $EY^2 = P + N$, the entropy of Y is bounded by $\frac{1}{2}\log 2\pi e(P+N)$.

Applying this result to bound the mutual information, we obtain

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

Hence, the information capacity of the Gaussian channel is

$$C = \max_{EX^2 \leq P} I(X;Y) = \frac{1}{2}\log(1+\frac{P}{N}),$$

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

An (M, n) code for the Gaussian channel with power constraint P consists of the following:

1. An index set $\{1, 2, 3, \cdots, M\}$.
2. An encoding function $x : \{1, 2, \cdots, M\} \to \mathcal{X}^n$ yielding codewords $x^n(1), x^n(2), \cdots, x^n(M)$, satisfying

$$\sum_{i=1}^{n} x_i^2(\omega) \leq nP, \ \omega = 1, 2, \cdots, M.$$

3. A decoding function

$$g : \mathcal{Y}^n \to \{1, 2, \cdots, M\}.$$

The rate and probability of error of the code are defined as in the discrete case. The arithmetic average of the probability of the error is defined by by $P_e^{(n)} = \frac{1}{2^{nR}} \sum \lambda_i$.

Definition

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 $\lambda^{(n)}$ 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 and noise variance N is

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

Remark

- *We first present a plausibility argument as to why we may be able to construct* $(2^{nC}, n)$ *codes with a low probability of error.*
- *Consider any codeword of length* n*.*
- *The received vector is normally distributed with mean equal to the true codeword and variance equal to the noise variance.*
- *With high probability, the received vector is contained in a sphere of radius* $\sqrt{n(N+\varepsilon)}$ *around the true codeword.*
- *If we assign everything within this sphere to the given codeword, when this codeword is sent there will be an error only if the received vector falls outside the sphere, which has low probability.*

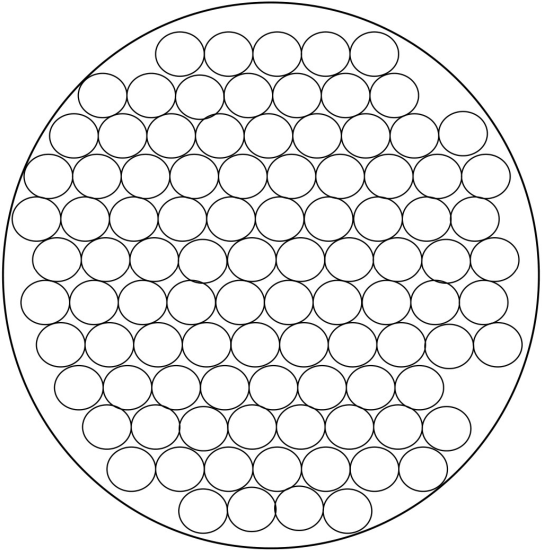
Remark

- *Similarly, we can choose other codewords and their corresponding decoding spheres.*
- *How many such codewords can we choose?*
- *The volume of an n-dimensional sphere is of the form $C_n r^n$, where r is the radius of the sphere.*
- *In this case, each decoding sphere has radius $\sqrt{nN}$.*
- *These spheres are scattered throughout the space of received vectors.*

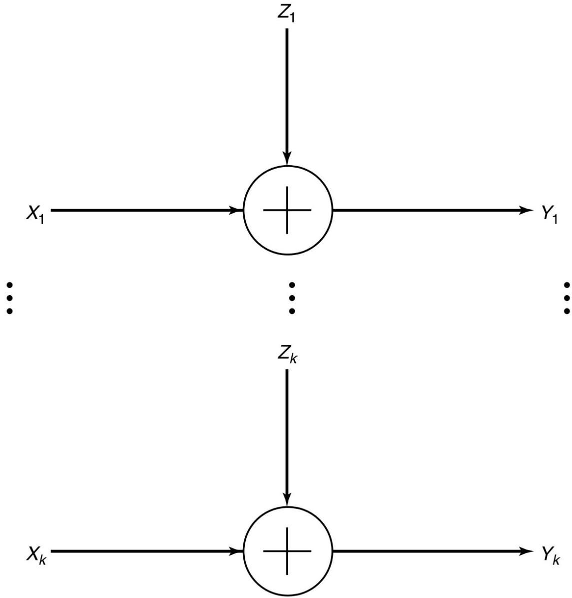
- The received vectors have energy no greater than $n(P+N)$, so they lie in a sphere of radius $\sqrt{n(P+N)}$.
- The maximum number of nonintersecting decoding spheres in this volume is no more than

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

- And the rate of the code is $\frac{1}{2}\log(1+\frac{P}{N})$.



- In this section we consider k independent Gaussian channels in parallel with a common power constraint.
- The objective is to distribute the total power among the channels so as to maximize the capacity.
- This channel models a nonwhite additive Gaussian noise channel where each parallel component represents a different frequency.
- Assume that we have a set of Gaussian channels in parallel as illustrated in the next figure.


Z1
X1
Y1
Zk
Xk
Yk

- The output of each channel is the sum of input and Gaussian noise.
- For channel j,

$$Y_j + Z_j,\ j = 1, 2, \cdots, k,$$

with

$$Z_j \sim \mathcal{N}(0, N_j),$$

and the noise is assumed to be independent from channel to channel.

- We assume that there is a common power constraint on the total power used, that is,

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

- We wish to distribute the power among the various channels so as to maximize the total capacity.
- The information capacity of the channel C is

$$C = \max_{f(x_1, x_2, \cdots, x_k): \sum \mathsf{E} X_i^2 \leq P} I(X_1, X_2, \cdots, X_k; Y_1, Y_2, \cdots, Y_k).$$

- We calculate the distribution that achieves the information capacity for this channel.
- The fact that the information capacity is the supremum of achievable rates can be proved by methods identical to those in the proof of the capacity theorem for single Gaussian channels and will be omitted.

Since $Z_1, \cdots, Z_k$ are independent,

$$
\begin{aligned}
& I(X_1, X_2, \cdots, X_k; Y_1, Y_2, \cdots .Y_k) \\
=& h(Y_1, Y_2, \cdots, Y_k) - h(Y_1, Y_2, \cdots .Y_k | X_1, X_2, \cdots, X_k) \\
=& h(Y_1, Y_2, \cdots, Y_k) - h(Z_1, Z_2, \cdots .Z_k | X_1, X_2, \cdots, X_k) \\
=& h(Y_1, Y_2, \cdots, Y_k) - h(Z_1, Z_2, \cdots .Z_k) \\
=& h(Y_1, Y_2, \cdots, Y_k) - \sum_i h(Z_i) \\
\leq & \sum_i (h(Y_i) - h(Z_i)) \\
\leq & \sum_i \frac{1}{2} \log(1 + \frac{P_i}{N_i}),
\end{aligned}
$$

where $P_i = \mathsf{E} X_i^2$ and $\sum P_i = P$.

Equality is achieved by

$$(X_1, X_2, \cdots, X_k) \sim \mathcal{N}\left(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{bmatrix}\right).$$

So the problem is reduced to finding the power allotment that maximizes the capacity subject to the constraint that $\sum P_i = P$.

- This is a standard optimization problem and can be solved using Lagrange multipliers.
- Writing the functional as

$$J(P_1, \cdots, P_k) = \sum \frac{1}{2} \log(1 + \frac{P_i}{N_i}) + \lambda(\sum P_i).$$

- Differentiating with respect to P_i, we have

$$\frac{1}{2} \frac{1}{P_i + N_i} + \lambda = 0$$

or

$$P_i = \nu - N_i.$$

- However, since the P_i's must be nonnegative, it may not always be possible to find a solution of this form.
- In this case, we use the Kuhn-Tucker conditions to verify that the solutions

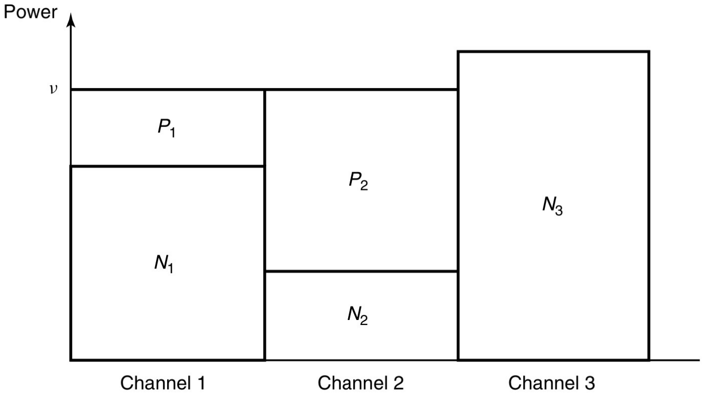
$$P_i = (\nu - N_i)^+$$

  is the assignment that maximizes capacity, where ν is chosen so that

$$\sum (\nu - N_i)^+ = P.$$

- Here $(x)^+$ denotes the positive part of x:

$$(x)^+ = \begin{cases} x & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$


Power
ν
P_1
N_1
P_2
N_2
N_3
Channel 1
Channel 2
Channel 3

- We just considered the case of a set of parallel independent Gaussian channels in which the noise samples from different channels were independent.
- Now we will consider the case when the noise is dependent.
- This represents not only the case of parallel channels, but also the case when the channel has Gaussian noise with memory.
- For channels with memory, we can consider a block of n consecutive uses of the channel as n channels in parallel with dependent noise.

- Let K_Z be the covariance matrix of the noise, and let K_X be the input covariance matrix.
- The proper constraint on the input can then be written as

$$\frac{1}{n}\sum_i EX_i^2 \leq P,$$

or equivalently,

$$\frac{1}{n}\text{tr}(K_X) \leq P.$$

- Unlike the case of last section, the power constraint here depends on n; the capacity will have to be calculated for each n.

- Just as in the case of independent channels, we can write

$$I(X_1, \cdots, X_n; Y_1, \cdots, Y_n) = h(Y_1, \cdots, Y_n) - h(Z_1, \cdots, Z_n).$$

- Here $h(Z_1, \cdots, Z_n)$ is determines only by the distribution of the noise and is not dependent on the choice of input distribution. So finding the capacity amounts to maximizing $h(Y_1, \cdots, Y_n)$.
- The entropy of the output is maximized when Y is normal, which is achieved when the input is normal, which is achieved when the input is normal.

- Since the input and the noise are independent, the covariance of the output Y is $K_Y = K_X + K_Z$ and the entropy is

$$h(Y_1, \cdots, Y_n) = \frac{1}{2}\log((2\pi e)^n |K_X + K_Z|).$$

- Now the problem is reduced to choosing K_X so as to maximize $|K_X + K_Z|$, subject to a trace constraint on K_X.
- To do this, we decompose K_Z into diagonal form,

$$K_Z = Q\Lambda Q^t, \text{ where } QQ^t = I.$$

- Then

$$\begin{aligned}|K_X + K_Z| &= |K_X + Q\Lambda Q^t| \\ &= |Q||Q^t K_X Q + \Lambda||Q^t| \\ &= |Q^t K_X Q + \Lambda| \\ &= |A + \Lambda|,\end{aligned}$$

where $A = Q^t K_X Q$.

- Since for any matrices B and C, $\text{tr}(BC) = \text{tr}(CB)$, we have

$$\text{tr}(A) = \text{tr}(Q^t K_X Q) = \text{tr}(QQ^t K_X) = \text{tr}(K_X).$$

- Now the problem is reduced to maximizing $|A + \Lambda|$ subject to a trace constraint $\text{tr}(A) \leq nP$.

- Now we applied Hadamard's inequality.
- Hadamard's inequality states that the determinant of any positive definite matrix K is less than the product of its diagonal elements, that is,

$$|K| \leq \prod_i K_{ii}$$

  with equality if and only if the matrix is diagonal.
- Thus,

$$|A + \Lambda| \leq \prod_i (A_{ii} + \lambda_i)$$

  with equality if and only if A is diagonal.

- Since A is subject to a trace constraint,

$$\frac{1}{n}\sum_i A_{ii} \leq P,$$

  and $A_{ii} \geq 0$, the maximum value of $\prod_i (A_{ii} + \lambda_i)$ is attained when

$$A_{ii} + \lambda_i = \nu.$$

- However, given the constraints, it may not always be possible to satisfy this equation with positive A_{ii}.
- In such cases we can show by the standard Kuhn-Tucker conditions that the optimum solution corresponds to setting

$$A_{ii} = (\nu - \lambda_i)^+,$$

  where the water level ν is chosen so that $\sum A_{ii} = nP$.