Gaussian Channels

Introduction

- A discrete-time Gaussian channel is defined by

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

  The output Y_i is the sum of input X_i and noise Z_i. The noise Z is assumed to be independent of X.

- If the noise variance N is 0 or the power of input X is not constrained, then the capacity of the channel is infinite.

- Therefore we assume a constraint on the average power: for any codeword $(x_1, \ldots, x_n)$ we require

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

A BSC Using Gaussian Channel

- Send $\sqrt{P}$ and $-\sqrt{P}$ for bit 0 and bit 1 respectively.
- Decode $X_i = 0$ if $Y_i > 0$ and $X_i = 1$ if $Y_i \leq 0$.
- Assuming both levels are equally likely, the probability of decoding error is

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

where $\Phi(x)$ is the cumulative normal function.

- Thus we have used a Gaussian channel as a binary symmetric channel.

Information Capacity

- The information capacity of a Gaussian channel with power constraint P is

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

- Since $EY^2 = EX^2 + EZ^2 = P + N$, we have

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

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

(M, n) Code

- A (M, n) code for a Gaussian channel with power constraint consists of
  - An index set $\{1, 2, \ldots, M\}$
  - An encoder $f : \{1, 2, \ldots, M\} \to \mathcal{X}^n$
  - A decoder $g : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$
- The rate of a code is defined by $R = \frac{\log M}{n}$
- A rate R is achievable if there exists a sequence of $(2^{nR}, n)$ code such that the maximal probability of error tends to zero.

Capacity of A Gaussian Channel

- The capacity of a channel is the supremum of all achievable rates.
- The capacity of a Gaussian channel with power constraint P and noise variance P is

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

- So

capacity = information capacity.

Channel Coding Theorem

- Let C be the information capacity.
  - (α) If $R < C$, then R is achievable.
  - (β) If R is achievable, then

$$R \leq C$$

Proof (α)

- Randomly generate a codebook

$$P(x) \sim \mathcal{N}(0, P-\epsilon) \ \ (\text{so } \frac{1}{n}\sum_i x_i^2 \sim P)$$

  for codewords of $w \in \{1, \ldots, 2^{nR}\}$.

- Joint typical set decoding. The probability of error is $(E_0 = \{\frac{1}{n}\sum_i X_i(1)^2 > P\}, E_i = \{(X^n(i), Y^n) \in A_\epsilon^{(n)}\})$

$$Pr(E) = Pr(E|1) \leq Pr(E_0) + Pr(E_1^c) + \sum_{i=2}^{2^{nR}} Pr(E_i)$$

$$\rightarrow 0 \text{ if } R < I(X;Y) \leq C$$

Proof (β)

- From the Fano's inequality

$$H(W|Y^n) \leq 1 + nRP_e^{(n)} = n\epsilon_n$$

- Using similar arguments as in the discrete case

$$\begin{aligned} nR = H(W) &= I(W;Y^n) + H(W|Y^n) \\ &\leq I(X^n;Y^n) + n\epsilon_n \leq \sum_i h(Y_i) - \sum_i h(Z_i) + n\epsilon_n \\ &\leq \sum_i \frac{1}{2}\log(1+\frac{P_i}{N}) + n\epsilon_n \\ &\leq \frac{n}{2}\log(1+\frac{1}{n}\sum_i \frac{P_i}{N}) + n\epsilon_n \to R \leq \frac{1}{2}\log(1+\frac{P}{N}) \end{aligned}$$

See the text for details.

Band-limited Channels

- A common model for communication over radio network is a continuous-time **band-limited channel** with white noise: $Y(t) = (X(t) + Z(t)) * h(t)$.
- **Sampling theorem:** If the signal is band-limited to W, then $2W$ samples per second suffice to reconstruct the original signal.
- While a general function has a degree of freedom of infinity, a band-limited function has a degree of freedom of $2W$ per second.

Band-limited, Time-limited Functions

- We consider functions with most of their energy in a bandwidth of W and a time interval $[0, T]$. Such functions can be seen as points in a vector space of $2WT$ dimensions.
- Let the noise power spectral density be $\frac{N_0}{2}$. Then the noise variance per sample is $\frac{N_0 WT}{2WT} = \frac{N_0}{2}$.
- The capacity of the channel per sample is

$$C = \frac{1}{2} \log \left(1 + \frac{\frac{P}{2W}}{\frac{N_0}{2}} \right) = \frac{1}{2} \log \left(1 + \frac{P}{N_0 W} \right)$$

Channel Capacity Per Second

- Since there are $2W$ samples per second, the capacity of the channel per second is

$$C = W \log(1 + \frac{P}{N_0 W}).$$

- If we let $W \to \infty$,

$$C = \frac{P}{N} \log_2 e \text{ bits per second.}$$

  In other words, the capacity is proportional to P and inversely proportional to N_0. See Example 10.3.1 for the example of telephone line capacity.

Parallel Gaussian Channels

- Consider independent Gaussian channels with a joint power constraint.
- The objective is to distribute the total power among channels so as to maximize the total capacity.
- This models a **non-white additive Gaussian noise** channel where each parallel component represents a different frequency.

Assumptions and Constraints

- For channel j,

$$Y_j = X_j + Z_j, \quad Z_j \sim \mathcal{N}(0, N_j), \quad j = 1, \ldots, k$$

- We assume that the noises are independent from channel to channel.
- The information channel capacity is

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

The Optimal Distribution

- The optimal distribution of X is a multi-variate normal distribution with a diagonal covariance matrix

$$P_j = (\nu - N_j)^+ = \begin{cases} \nu - N_j & \text{if } \nu \geq N_j \\ 0 & \text{otherwise,} \end{cases}$$

where ν is chosen such that

$$\sum_j (\nu - N_j)^+ = P.$$

- This is referred to as **water-filling** and is illustrated in Figure 10.4.

Gaussian Channels with Memory

- Dependencies among channel noises, characterized by K_Z, the noise covariance matrix.
- Information channel capacity

$$C = \max_{f(\mathbf{x}); E\mathbf{X}^2 \leq P} I(\mathbf{X}; \mathbf{Y}) = \max h(\mathbf{Y}) - h(\mathbf{Z})$$

- The optimal solution maximize $h(Y)$, which we know is the same as maximizing $|K_Y| = |K_X + K_Z|$, subject to the constraint that $tr(K_X) \leq nP$.
- Can be solved by simultaneously diagonalizing K_Z and K_X.

Optimal Solution

- First diagonalize $K_Z = Q\Lambda Q^T$. Then

$$|K_X + K_Z| = |A + \Lambda|,$$

where $A = Q^T K_X Q$. Note $tr(A) = tr(K_X)$.

- By Hadamard inequality, $|K| \leq \prod K_{ii}$,

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

Combining the constraint $\frac{1}{n} tr(A) \leq P$, we have

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

where ν is chosen to satisfy the power constraint.

Spectral Domain Interpretation

- Consider a channel in which the additive Gaussian noise forms a stochastic process with finite dimensional covariance matrix $K_Z^{(n)}$.
- If the process is stationary, then the eigenvalues tend to a limit as $n \to \infty$ and the density of eignevalues tends to the power spectrum of the process.
- The input should be chosen to be a Gaussian process with a spectrum which is large at frequencies where the noise spectrum is small.

Feedback on Gaussian Channels

- For Gaussian channels with memory, where noises are correlated, feedback does increase capacity.
- Feedback allows the code to depend on the previous Ys. So the constraint is

$$E\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2(w, Y^{i-1})\right] \leq P.$$

- With feedback: $C_{n,FB} = \max_{\frac{1}{n}tr(K_X^{(n)})\leq P} \frac{1}{2n}\log\frac{|K_{X+Z}^{(n)}|}{|K_Z^{(n)}|}$
- Without feedback: $C_n = \max_{\frac{1}{n}tr(K_X^{(n)})\leq P} \frac{1}{2n}\log\frac{|K_X^{(n)}+K_Z^{(n)}|}{|K_Z^{(n)}|}$

Lemmas

- (Lemma 10.6.1) For any random vectors X, Z,

$$K_{X+Z} + K_{X-Z} = 2K_X + 2K_Z.$$

- (Lemma 10.6.2) For any positive definite matrix A, B, if $A - B$ is positive definite, then $|A| \geq |B|$.

- (Lemma 10.6.3) For any n-dimensional random vectors X, Z

$$|K_{X+Z}| \leq 2^n |K_X + K_Z|.$$

Theorems

- The achievable rate on Gaussian channel with feedback is bounded by

$$R_{n,FB} \leq \frac{1}{n}\frac{1}{2}\log\frac{|K_Y^{(n)}|}{|K_Z^{(n)}|} + \epsilon_n$$

- The capacities of a Gaussian channel of codes with and without feedback is related by

$$C_{n,FB} \leq C_n + \frac{1}{2}$$