Lecture 4: Gaussian channel. Multi-user channels.

Course instructor: Alexey Frolov

al.frolov@skoltech.ru

Teaching Assistant: Stanislav Kruglik

stanislav.kruglik@skolkovotech.ru

February 7, 2017

Outline

Outline

1. Differential entropy
2. Gaussian channel
3. Multi-user channels

Differential entropy

Definition

Let X be a random variable with cumulative distribution function $F(x) = \Pr(X \leq x)$. If $F(x)$ is continuous, the random variable is said to be continuous. Let $f(x) = F'(x)$ when the derivative is defined, $f(x)$ is called the probability density function for X. The set where $f(x) > 0$ is called the support set of X.

Definition

The differential entropy $h(X)$ of a continuous random variable X with a density $f(x)$ is defined as

$$h(X) = -\int_S f(x) \log f(x) dx,$$

where S is a support of X.

Properties

- $D(f||g) = \int f \log \frac{f}{g} \geq 0$
- $h(X|Y) \leq h(X)$
- $h(aX) = h(X) + \log |a|$
- $I(X;Y) = \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} \geq 0$

Normal distribution

Example

Let $X \sim \phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$.

$$
\begin{aligned}
h(X) &= -\int \phi \ln \phi \\
&= -\int \phi(x) \left[-\frac{x^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2}\right] dx \\
&= \frac{\mathbb{E}[X^2]}{2\sigma^2} + \frac{1}{2} \ln 2\pi\sigma^2 \\
&= \frac{1}{2} \ln 2\pi e\sigma^2 \ \text{ nats.}
\end{aligned}
$$

Multivariate normal distribution

Example

Let

$$\mathbf{X} = [X_1, X_2, \ldots, X_n],$$

have a multivariate normal distribution with mean

$$\mu = \mathbb{E}[\mathbf{X}] = [\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_n]]$$

and a covariance matrix

$$K = \mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T],$$

i.e

$$f(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^n sqrt|K|} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T K^{-1}(\mathbf{x} - \mu)\right).$$

Multivariate normal distribution

Example

$$
\begin{aligned}
h(\mathbf{X}) &= -\int f(\mathbf{x})\left[-\frac{1}{2}(\mathbf{x}-\mu)^T K^{-1}(\mathbf{x}-\mu) - \ln \sqrt{2\pi}^n \sqrt{|K|}\right] d\mathbf{x} \\
&= \frac{1}{2}\mathbb{E}\left[\sum_{i,j}(x_i-\mu_i)^T K_{i,j}^{-1}(x_j-\mu_j)\right] + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}\sum_{i,j}\mathbb{E}\left[(x_i-\mu_i)^T(x_j-\mu_j)\right] K_{i,j}^{-1} + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}\sum_{i,j} K_{j,i}K_{i,j}^{-1} + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{n}{2} + \frac{1}{2}\ln(2\pi)^n|K| = \frac{1}{2}\ln(2\pi e)^n|K| \ \text{ nats.}
\end{aligned}
$$

Multivariate normal distribution maximizes the entropy

Theorem

Let the random vector $\mathbf{X} \in \mathbb{R}^m$ *have zero mean and covariance* $K = \mathbb{E}[\mathbf{X}\mathbf{X}^T]$, *i.e.* $K_{i,j} = \mathbb{E}[X_i X_j]$. *Then*

$$h(\mathbf{X}) \leq \frac{1}{2} \log(2\pi e)^n |K|$$

with equality iff $\mathbf{X} \sim N(0, K)$.

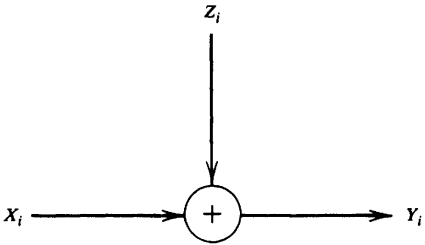
Proof.

$$\begin{aligned} 0 \leq D(g||\phi) &= \int g \log \frac{g}{\phi} = -h(g) - \int g \log \phi \\ &= -h(g) - \int \phi \log \phi = -h(g) + h(\phi). \end{aligned}$$

□

Outline

Gaussian channel



$$Y_i = X_i + Z_i, \;\; Z_i \sim N(0, N).$$

Energy or power constraint

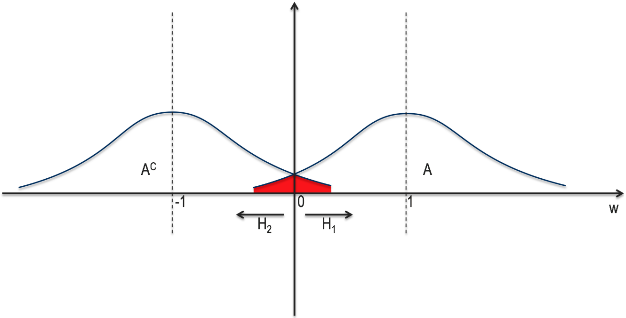
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P.$$

Signal-to-noise ratio

$$\begin{aligned} SNR &= \frac{P}{N} \quad \text{(linear scale)} \\ SNR &= 10\log_{10}\frac{P}{N} \quad \text{(dB)} \end{aligned}$$

Connection to BSC

Assume that we want to send 1 bit over the channel in 1 use of the channel. Given the power constraint, the best that we can do is to send one of two levels $+\sqrt{P}$ and $-\sqrt{P}$.



Connection to BSC

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

Recall, that

$$
Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{+\infty} \exp(-x^2/2).
$$

Using such a scheme, we have converted the Gaussian channel into a BSC with crossover probability P_e.

Likelihood ratio

Assume again, that 0 is sent with the level $+\sqrt{P}$ and 0 is sent with the level $-\sqrt{P}$. Let us calculate the log likelihood ratio given y

$$
\begin{aligned}
LLR &= \log \frac{\Pr(X = +\sqrt{P}|y)}{\Pr(X = -\sqrt{P}|y)} \\
&= \log \frac{\Pr(y|X = +\sqrt{P})\Pr(X = +\sqrt{P})\Pr(y)}{\Pr(y|X = -\sqrt{P})\Pr(X = \sqrt{P}\Pr(y)} \\
&= \log \frac{f(y| + \sqrt{P})}{f(y| - \sqrt{P})} \\
&= \frac{(y + \sqrt{P})^2 - (y - \sqrt{P})^2}{2\sigma^2} \\
&= \frac{2\sqrt{P}}{\sigma^2} y
\end{aligned}
$$

Capacity

Definition

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

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

Theorem

$$C = \frac{1}{2}\log(1+SNR).$$

Capacity

Proof.

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

$$
\mathbb{E}[Y^2] = \mathbb{E}[(X+Z)^2] = \mathbb{E}[X^2] + \mathbb{E}[Z^2] + 2\mathbb{E}[X]\mathbb{E}[Z] = P + N
$$

Thus,

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

□

Definitions

Definition (Code)

An (M, n) code consists of the following:

- message set $\{1, 2, \ldots, M\}$;
- encoding function *enc*: $\{1, 2, \ldots, M\} \to \mathcal{X}^n$;

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

- decoding function *dec*: $\mathcal{Y}^n \to \{1, 2, \ldots, M\}$.

Definitions

Definition (Probability of error)

$$\lambda_i = \Pr(dec(Y^n) \neq i | X^n = X^n(i)).$$

$$\lambda^{(n)} = \max_{i \in \{1,2,\ldots,M\}} \lambda_i$$

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$$

Definitions

Definition

A rate R of (M, n) code is

$$R = \frac{\log M}{n}.$$

Definition

A rate R is achievable is there exists a sequence of $(2^{Rn}, n)$ codes, such that $\lambda^{(n)} \to 0$ as $n \to \infty$.

Definition

The capacity Gaussian channel is the supremum of all achievable rates.

Achievability

Definition (Ensemble of codes)

$$\mathcal{C} = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ x_1(2) & x_2(2) & \dots & x_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{Rn}) & x_2(2^{Rn}) & \dots & x_n(2^{Rn}) \end{bmatrix}$$

We generate the codewords with each element i.i.d. according to a normal distribution with variance $P - \varepsilon$.

Typical set decoding

The receiver declares, that the index $\hat{W}$ was transmitted if the following conditions are satisfied:

- the pair $(X^n(\hat{W}), Y^n)$ is jointly typical;
- there is no other index i, such that $(X^n(i), Y^n)$ is jointly typical.

Achievability

Define

$$E_0 = \left\{ \frac{1}{n} \sum_{i=1}^{n} X_i^2 > P \right\}$$

and

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

$$\begin{aligned} \Pr(\mathcal{E} | W = 1) &\leq \Pr(E_0) + \Pr(E_1^c) + \sum_{i=2}^{2^{Rn}} E_i \\ &\leq 2\varepsilon + \left(2^{Rn} - 1\right) 2^{-n[I(X;Y) - 3\varepsilon]} \end{aligned}$$

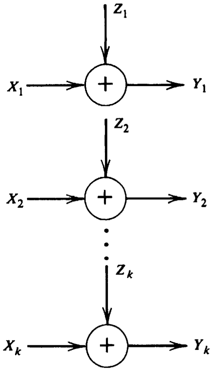
Thus, if $R < I(X; Y)$ we can choose ε and n, such that $\Pr(\mathcal{E})$ less, then ε'.

Parallel Gaussian channels

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

$$Z_j \sim N(0, N_j)$$

$$\mathbb{E}\left[\sum_{i=1}^{k} X_i^2\right] \leq P$$



We wish to distribute the power among the various channels so as to maximize the total capacity.

Parallel Gaussian channels

$$I(X_1, X_2, \ldots, X_k; Y_1, Y_2, \ldots, Y_k) \leq \sum_i \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right),$$

where $P_i = \mathbb{E}[X_i]$ and $\sum P_i = P$.

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

Parallel Gaussian channels

$$J(P_1, P_2, \dots, P_k) = \sum_i \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right) + \lambda \sum P_i$$

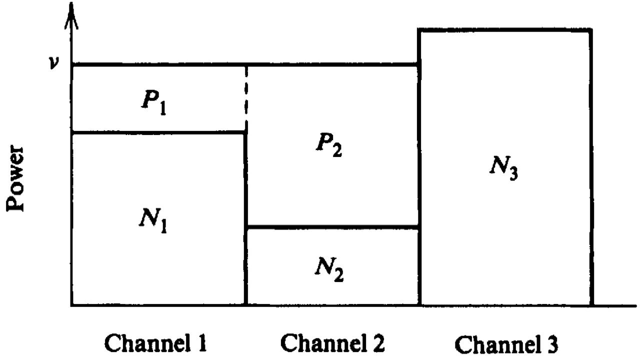
and differentiating with respect to P_i, we have

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

or

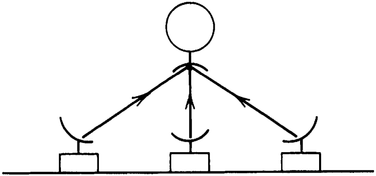
$$P_i = (\nu - N_i)^+$$

Analogy to water filling

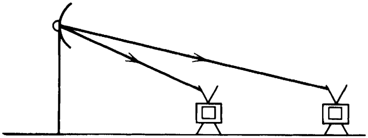


Outline

Multi-user channels

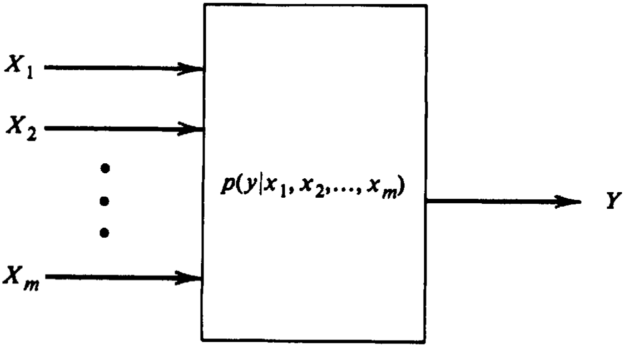


(a) Multiple access channel (MAC)



(b) Broadcast channel

m-user Multiple access channel



m-user Multiple access channel

Let $S \subseteq \{1, 2, \ldots, m\}$, $R(S) = \sum_{i \in S} R_i$ and $X(S) = \{X(i), i \in S\}$.

Theorem

The capacity region of the m-user multiple access channel is the closure of the convex hull of the rate vectors satisfying

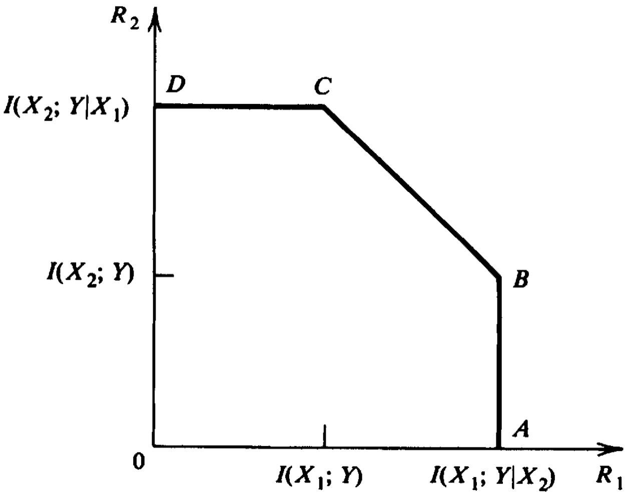
$$R(S) \leq I(X(S); Y | X(S^c)) \quad \forall S \subseteq \{1, 2, \ldots, m\}.$$

for some product distribution $P_1(x_1)P_2(x_2)\ldots P_m(x_m)$.

2-user Multiple access channel

$$
\begin{aligned}
R_1 &\leq I(X_1; Y | X_2) \\
R_2 &\leq I(X_2; Y | X_1) \\
R_1 + R_2 &\leq I(X_1, X_2; Y)
\end{aligned}
$$

2-user Multiple access channel



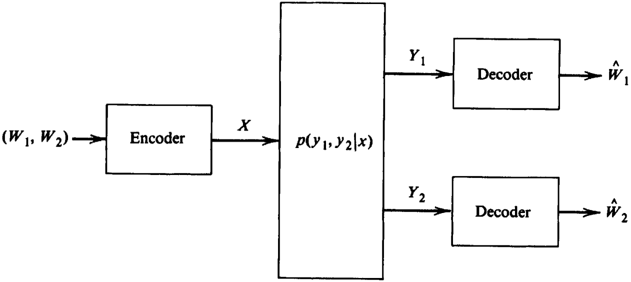
2-user Gaussian MAC

$$Y = X_1 + X_2 + Z.$$

$$C(x) = \frac{1}{2}\log(1+x).$$

$$\begin{aligned} R_1 &\le C\left(\frac{P_1}{N}\right) \\ R_2 &\le C\left(\frac{P_2}{N}\right) \\ R_1 + R_2 &\le C\left(\frac{P_1+P_2}{N}\right) \end{aligned}$$

Broadcast channel



Broadcast channel

Definition

A broadcast channel is said to be physically degraded if

$$p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1)$$

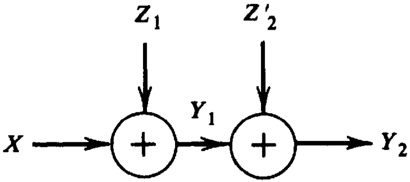
Theorem

The capacity region for sending independent information over the degraded broadcast channel $X \to Y_1 \to Y_2$ *is the convex hull of the closure of all* (R_1, R_2) *satisfying*

$$\begin{aligned} R_2 &\leq I(U; Y_2) \\ R_1 &\leq I(X; Y_1|U) \end{aligned}$$

for some joint distribution $p(u)p(x|u)p(y, z|x)$*, where the auxiliary random variable* U *has cardinality bounded by* $|\mathcal{U} \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}|$.

Broadcast channel



$$Y_1 = X + Z_1$$
$$Y_2 = X + Z_2 = Y_1 + Z_2'$$

$$\begin{aligned} R_1 &\leq C\left(\frac{\alpha P}{N_1}\right) \\ R_2 &\leq C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right) \end{aligned}$$

Thank you for your attention!