Lecture 11: Channel Capacity

Course instructor: Alexey Frolov

al.frolov@skoltech.ru

Teaching Assistant: Stanislav Kruglik

stanislav.kruglik@skolkovotech.ru

February 22, 2018

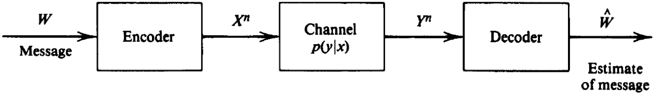
Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

Discrete memoryless channels



Definition

Discrete channel is specified by:

- $\mathcal{X}$ – input alphabet (finite);
- $\mathcal{Y}$ – output alphabet (finite);
- probability transition matrix $P(y|x)$.

Definition

Channel is *memoryless* if

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x_i).$$

Capacity of DMC

Definition (Capacity of DMC)

$$C = \max_{P_X} \{I(X;Y)\},$$

where the maximum is taken over all possible input distributions P_X.

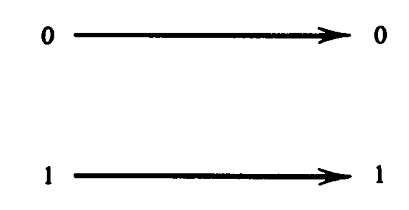
Channel capacity is the highest rate in bits per channel use at which the information can be sent with arbitrarily low probability of error.

Duality between Data Compression and Data Transmission

During compression, we remove all the redundancy in the data to form the most compressed version possible, whereas during data transmission, we add redundancy in a controlled fashion to combat errors in the channel. A general communication system can be broken into two parts and that the problems of data compression and data transmission can be considered separately.

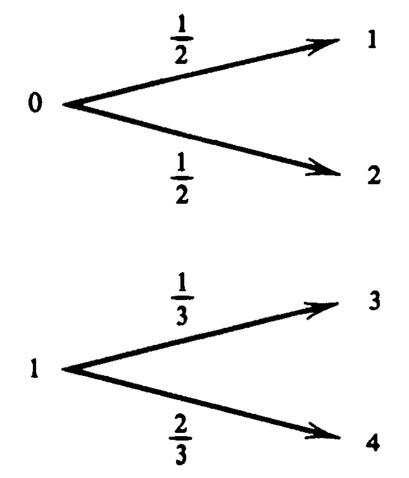
Examples of channels

Example (Noiseless binary channel)



Examples of channels

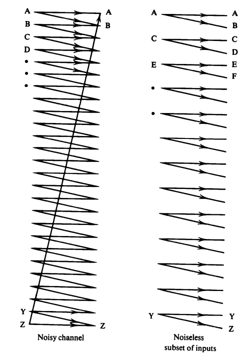
Example (Noisy channel with non-overlapping outputs)



Examples of channels

Example (Noisy typewriter)

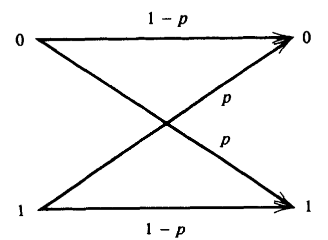
$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - 1 \\ &= \log 26 - 1 = \log 13 \end{aligned}$$



Examples of channels

Example (Binary symmetric channel)

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum_{x\in\mathcal{X}} P(x)H(Y|X=x) \\
&= H(Y) - \sum_{x\in\mathcal{X}} P(x)h(p) \\
&\leq 1 - h(p).
\end{aligned}
$$



Example (Binary erasure channel)

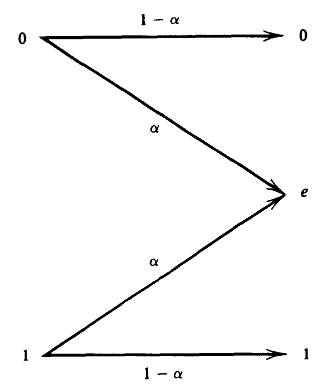
$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum_{x \in \mathcal{X}} P(x) H(Y|X=x) \\
&= H(Y) - h(\alpha)
\end{aligned}
$$

Let E be an indicator $I_{\{Y=e\}}$ and $\pi = \Pr(X=1)$.

$$
\begin{aligned}
H(Y) &= H(Y,E) = H(E) + H(Y|E) \\
&= h(\alpha) + (1-\alpha) h(\pi).
\end{aligned}
$$

Thus,

$$
I(X;Y) = (1-\alpha) h(\pi).
$$



Binary Erasure Channel

Main example: internet traffic.

The capacity has an intuitive meaning. Since a proportion α of the bits are lost in the channel, we can recover (at most) a proportion $1-\alpha$ of the bits.

We can achieve the capacity if we use feedback: if a bit is lost, retransmit it until it gets through.

Symmetric Channel

Example

$$P(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Symmetric Channel

Assume all the rows of the probability transition matrix are permutations of each other, then (by $\mathbf{r}$ we denote some row)

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(\mathbf{r}) \\ &\leq \log|\mathcal{Y}| - H(\mathbf{r}). \end{aligned}$$

Symmetric Channel

Now assume the sum of the entries in each column of the probability transition matrix is the same and equal to c, then $P(x) = 1/|\mathcal{X}|$ achieves a uniform distribution on Y, i.e.

$$P(y) = \sum_{x \in \mathcal{X}} P(y|x)p(x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} P(y|x) = \frac{c}{|\mathcal{X}|}.$$

Symmetric Channel

Definition

A channel is said to be *symmetric* if the rows of the channel transition matrix are permutations of each other, and the columns are permutations of each other.

A channel is said to be *weakly symmetric* if the rows of the channel transition matrix are permutations of each other, and the column sums are equal.

Symmetric Channel

Example (Weakly Symmetric Channel)

$$P(y|x) = \left[\begin{array}{ccc} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{array} \right]$$

Symmetric Channel

Theorem

For a weakly symmetric channel,

$$C = \log |\mathcal{Y}| - H(\mathbf{r})$$

and this is achieved by a uniform distribution on the input alphabet.

Properties of channel capacity

- $C \geq 0$;
- $C \leq \log |\mathcal{X}|$;
- $C \leq \log |\mathcal{Y}|$;
- $I(X; Y)$ is a continuous function of $P(x)$;
- $I(X; Y)$ is a concave function of $P(x)$;

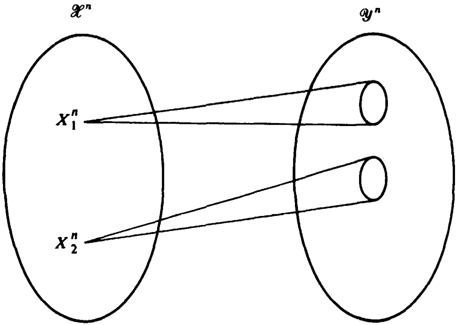
Local maximum is global maximum, maximum is finite.

Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

Intuitive idea

Main idea: the channel has a subset of inputs, that produce essentially disjoint sequences at the output. produce essentially disjoint sequences at the output. For each (typical) input n-sequence, there are approximately $2^{nH(Y|X)}$ possible Y sequences, all of them equally likely. The total number of possible (typical) Y sequences is $2^{nH(Y)}$ and, thus, the total number of disjoint sets is approx. $2^{nI(X;Y)}$.



Definitions

- $\{1, 2, \ldots, M\}$ – message set;
- $W \in \{1, 2, \ldots, M\}$ – message;
- $X^n(W) = enc(W)$ – codeword;
- $Y^n \sim P(y^n|x^n)$ – received sequence;
- $\hat{W} = dec(Y^n)$ – decoding rule.

Definitions

Definition (Code)

An (M, n) code for the channel $(\mathcal{X}, P(y|x), \mathcal{Y})$ consists of the following:

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

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 of a discrete memoryless channel is the supremum of all achievable rates.

Jointly typical sequences

$$
\begin{aligned}
A_{\varepsilon}^{(n)} = & \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \\
& \left| -\frac{1}{n} \log P(x^n) - H(X) \right| < \varepsilon \\
& \left| -\frac{1}{n} \log P(y^n) - H(Y) \right| < \varepsilon \\
& \left| -\frac{1}{n} \log P(x^n, y^n) - H(X, Y) \right| < \varepsilon \Big\},
\end{aligned}
$$

where $P(x^n, y^n) = \prod_{i=1}^{n} P(x_i, y_i)$.

Jointly typical sequences

Theorem

Let (X^n, Y^n) be sequences of length n drawn i.i.d. according to $P(x^n, y^n)$, then

1. $\Pr((x^n, y^n) \in A_\varepsilon^{(n)}) \to 1$ *as* $n \to \infty$;
2. $|A_\varepsilon^{(n)}| \le 2^{n(H(X,Y)+\varepsilon)}$;
3. *If $\hat{X}^n$ and $\hat{Y}^n$ are independent with the same marginals as $P(x^n, y^n)$, then*

$$\Pr((\hat{x}^n, \hat{y}^n) \in A_\varepsilon^{(n)}) \le 2^{-n(I(X;Y)-3\varepsilon)}$$

   and

$$\Pr((\hat{x}^n, \hat{y}^n) \in A_\varepsilon^{(n)}) \ge (1-\varepsilon)2^{-n(I(X;Y)+3\varepsilon)}$$

The channel coding theorem

Theorem

- *All rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(2^{Rn}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.*
- *Conversely, any sequence of $(2^{Rn}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.*

Achievability

Definition (Ensemble of codes)

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

Each entry in this matrix is generated i.i.d. according to $P(x)$.

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

Let $\Pr(\mathcal{E})$ be the average (over a random choice of codebook) probability of error.

$$\begin{aligned}
\Pr(\mathcal{E}) &= \sum_{\mathcal{C}} P(\mathcal{C}) P_e^{(n)}(\mathcal{C}) \\
&= \sum_{\mathcal{C}} P(\mathcal{C}) \frac{1}{2^{Rn}} \sum_{w=1}^{2^{Rn}} \lambda_w(\mathcal{C}) \\
&= \frac{1}{2^{Rn}} \sum_{w=1}^{2^{Rn}} \sum_{\mathcal{C}} P(\mathcal{C}) \lambda_w(\mathcal{C}) \\
&= \sum_{\mathcal{C}} P(\mathcal{C}) \lambda_1(\mathcal{C}) \text{ (by the symmetry of code construction)} \\
&= \Pr(\mathcal{E} | W = 1).
\end{aligned}$$

Define

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

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

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

Markov chain:

$$W \to X^n(W) \to Y^n \to \hat{W}.$$

Converse.

Fano's inequality

$$H(W|Y^n) \leq 1 + P_e^{(n)} nR$$

and $I(X^n; Y^n) \leq nC$

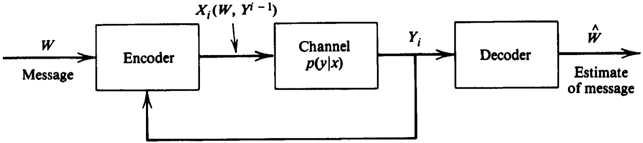
$$\begin{aligned}
nR &= H(W) = H(W|Y^n) + I(W; Y^n) \\
&\leq H(W|Y^n) + I(X^n(W); Y^n) \text{ (data processing ineq.)} \\
&\leq 1 + P_e^{(n)} nR + I(X^n(W); Y^n) \\
&\leq 1 + P_e^{(n)} nR + nC.
\end{aligned}$$

□

Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

Feedback capacity



Theorem (Feedback capacity)

$$C_{FB} = C.$$

Feedback does not increase capacity of discrete memoryless channels!

Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

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.

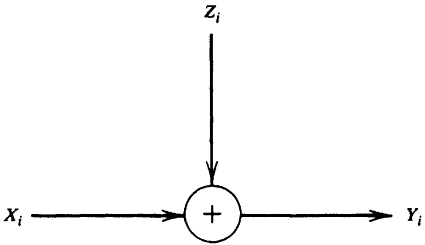
$$\begin{aligned} 0 \leq D(g||\phi) &= \int g \log \frac{g}{\phi} = -h(g) - \int g \log \phi \\ &= -h(g) - \int \phi \log \phi = -h(g) + h(\phi). \end{aligned}$$

□

Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

Gaussian channel



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

Energy or power constraint

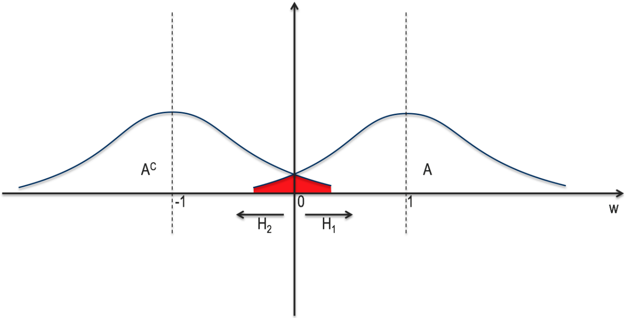
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P.$$

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

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

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

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

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

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

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

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

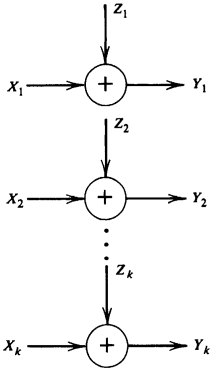
Thus, if $R < I(X;Y)$ we can choose ε and n, such that $\Pr(\mathcal{E})$ less, then ε'.

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

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

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

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

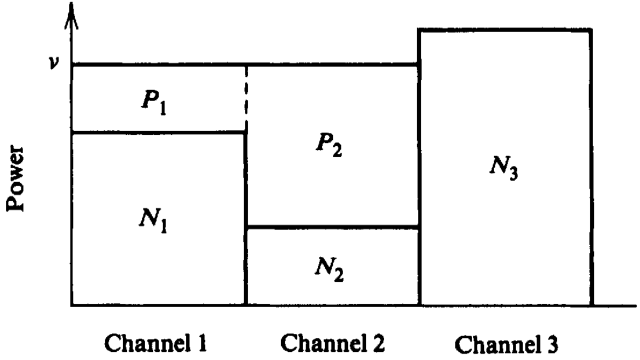
and differentiating with respect to P_i, we have

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

or

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

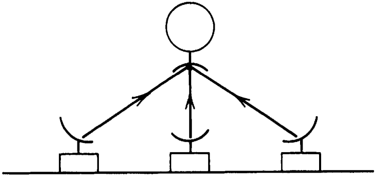
Analogy to water filling



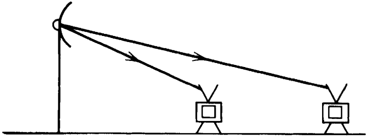
Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

Multi-user channels

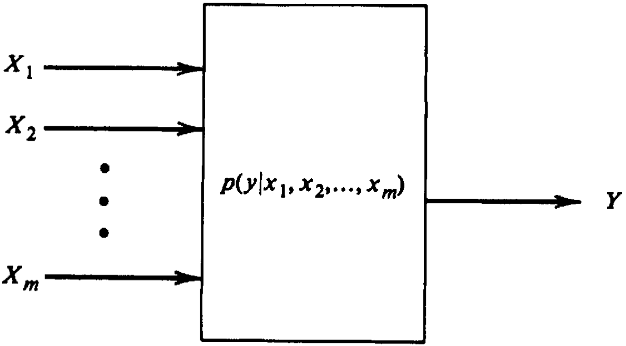


(a) Multiple access channel (MAC)



(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

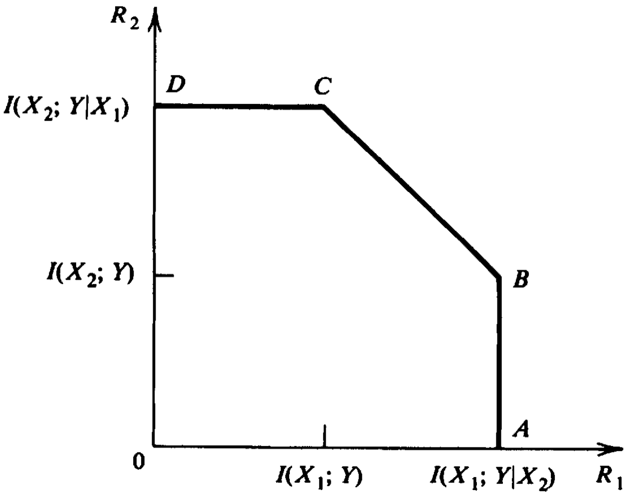
$$R(S) \leq I(X(S); Y | X(S^c)) \quad \forall S \subseteq \{1, 2, \ldots, m\}.$$

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

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

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

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

Thank you for your attention!

Outline

1. Discrete memoryless channels
2. Channel coding theorem
3. Feedback capacity
4. Differential entropy
5. Gaussian channel
6. Multi-user channels
7. Problems

Problem 1

Find the capacity of channel with transition matrix

$$P(y|x) = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix}$$

Problem 2

Find the capacity of channel with transition matrix and distribution of X corresponding to it

$$P(y|x) = \begin{bmatrix} 1/2 & 7/15 & 1/30 \\ 13/15 & 1/30 & 1/10 \end{bmatrix}$$

Problem 3

For channel in Problem 2 with $P_x(x = 0) = \frac{2}{5}$ and $P_x(x = 1) = \frac{3}{5}$ find the coefficient of channel usage.

Thank you for your attention!