Lecture 06

Note Title 1/31/2012

* The Shannon capacity for an i.i.d. channel is

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

where $I(X;Y) = D(P_{XY} \| P_X P_Y)$

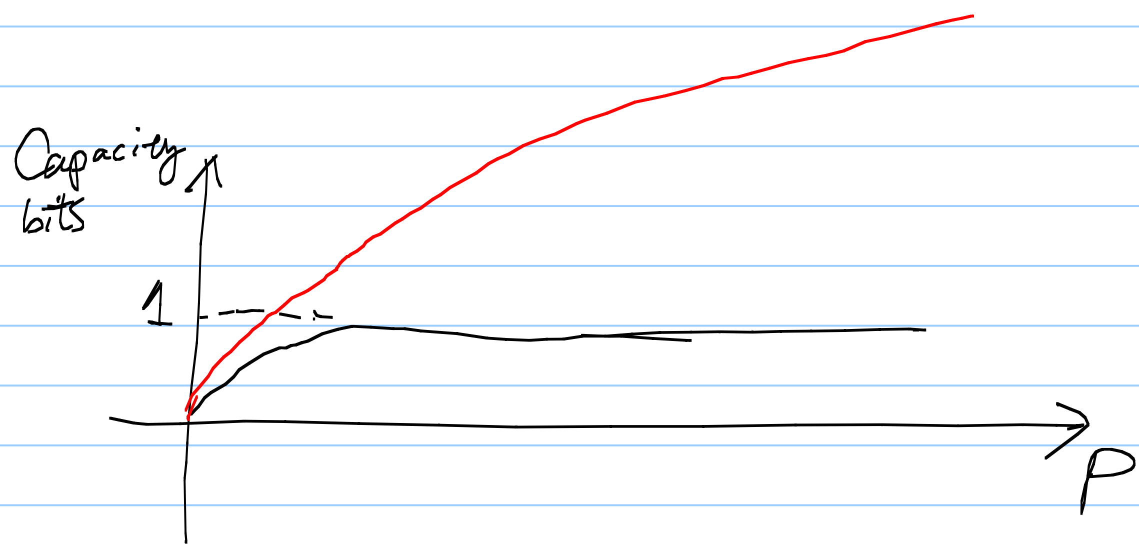
$$= E_X\left(D(P_{Y|X}(\cdot|X) \| P_Y)\right)$$

Example 1

* $X_i = (-1)^{b_i}\sqrt{P}$ the BPSK transmission

Example 2:

$E(X_i^2) \le P$, the power-constrained transmission.



Question: How to achieve the "capacity" of BPSK?

How to achieve the capacity of arbitrary P_X?

Given any Signal to Noise Ratio

$\Rightarrow I(X;Y)$ is a function of SNR

and any code with good performance must have rate $r < I(X;Y)$

* If we use logarithm with base 2, then we have 2^{rn} distinct codewords in S_X^n

Alternatively.

given any code of rate r.

the good performance happens when SNR satisfying $I(X;Y) > r$

* In error-control coding, we usually study the alternative question by ploting the error rate vs. SNR curve

Illustration:

SNR vs error-rates for different codes of the same rate

Constructing Capacity-achieving ECCs.

* Binary codes:

each codeword (a point in the high-dim space) $\vec{x} \in \{0,1\}^n$

$\{\vec{x} : \text{all codewords}\}$ is termed the codebook.

* Binary Random codes:

Fix the design parameter $P_X(x) = \begin{cases} p & \text{if } x=0 \\ 1-p & \text{if } x=1 \end{cases}$

$$\vec{x}_1 = (x_{1,1}, x_{1,2}, \cdots, x_{1,n})$$

$$\vdots$$

$$\vec{x}_{2^{rn}} = (x_{2^{rn},1}, \cdots, x_{2^{rn},n})$$

We have 2^{rn} distinct codewords, Each entry is chosen w. i.i.d. P_X

r: Code rate (bits/symbol usage

n: codeword length

① Tractable Analysis

That is if $n \to \infty$ and the random code has $r < I_{P_X}(X;Y)$ then the error-rate $\to 0$

② We can choose arbitrary P_X

ex: For Z-channel,

$$P_X(x) = \begin{cases} \dfrac{1-\varepsilon^{\frac{1}{1-\varepsilon}}}{1+(1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}} & \text{if } x=0 \\ \dfrac{\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}}{1+(1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}} & \text{if } x=1 \end{cases}$$

③ Not practical for long n

Encoding complexity : $O(2^{rn})$

Decoding complexity : $O(2^{rn})$

Binary linear codes

Definition 1: $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_{2^{rn}}\}$ form a binary linear subspace of $(GF(2))^n$ where $GF(2)$ is the binary field.

Definition 2: $\exists$ a binary $n \times rn$ generating matrix G st. any codeword $\vec{x}$ (column vector) is expressed

$$\vec{x} = G\vec{m}$$

where $\vec{m} \in (GF(2))^{rn}$ is the information bit string of length rn

$\boxed{\text{Encoding Complexity: } O(n^2)}$ matrix multiplication

① $\vec{x}, \vec{m}$ are column vectors.

② All operations are in the binary field:

$1+1=0 \qquad 1 \cdot 1 = 1$

$0+1=1 \qquad 1 \cdot 0 = 0$

$0+0=0$

Lemma 1: Assuming the "message vector" $\vec{m}$ is uniformly distributed over $[GF(2)]^{rn}$.

For any linear code, consider the marginal distribution of the i-th bit X_i

$\Rightarrow P_{X_i}$ is Bernoulli distributed with either $P(X_i = 1) = \frac{1}{2}$ or $P(X_i = 1) = 0$ #

Proof: For any linear code, consider its generating matrix representation

$$\vec{X} = G\vec{m} = (\vec{g}_1, \vec{g}_2, \cdots \vec{g}_{rn})\vec{m}$$

Consider the i-th bit X_i, & use $g_{i,j}$ to denote the i-th coordinate of the j-th vector $\vec{g}_j$

Case 1: If one of g_{ij} is non-zero, (say $g_{i,j_0} = 1$, then

$$x_i = m_{j_0} + (g_{i1}, g_{i,2}, \cdots g_{i,j_0-1}, g_{i,j_0+1}, \cdots, g_{i,rn}) \cdot (m_1, \cdots, m_{j_0-1}, m_{j_0+1}, \cdots, m_{rn})$$

Since m_{j_0} is uniform Bernoulli, so is X_i

Case 2: All $g_{i,j}$ are zero.

Then $x_i = (0, \cdots, 0) \cdot (m_1, \cdots, m_{rn}) = 0$ #

* Corollary: Linear codes do not achieve the capacity of a z-channel, which requires non-uniform P_X.

* Fortunately, linear codes can achieve most capacity when choosing each entry of G uniformly randomly.

Definition 3: $\exists$ a $((1-r)n) \times n$ ~~parity~~ ~~check~~ matrix H s.t. $\forall \vec{x}$ being a codeword, we have

$$H \vec{x} = 0$$

Example: Hamming code

$$n = 7, \quad rn = 4 \quad (\text{or } k = 4)$$

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

ex: $\vec{x} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ is a codeword.

$$\boxed{\text{Total \# of codewords: } 2^{n - \text{RowRank}(H)}}$$

Converting between G & H

$$H \xrightarrow{\text{row operation}} H' = (I_3 \; P)$$

$$G = \begin{pmatrix} P \\ I_4 \end{pmatrix}$$

$$\Rightarrow H' G = 0$$

$$\Rightarrow \quad G = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Now we can index each codeword $\vec{x}$

$$\vec{x}_0 = \vec{x}_{0000} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \vec{x}_{14} = \vec{x}_{1110} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\vec{x}_1 = \vec{x}_{0001} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \qquad \vec{x}_{15} = \vec{x}_{1111} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\vec{x}_2 = \vec{x}_{0010} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

$\vdots$

$$\vec{x}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \vec{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \vec{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

$$\vec{x}_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{x}_5 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \quad \vec{x}_6 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad \vec{x}_7 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\vec{x}_8 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \vec{x}_9 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \vec{x}_{10} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{x}_{11} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

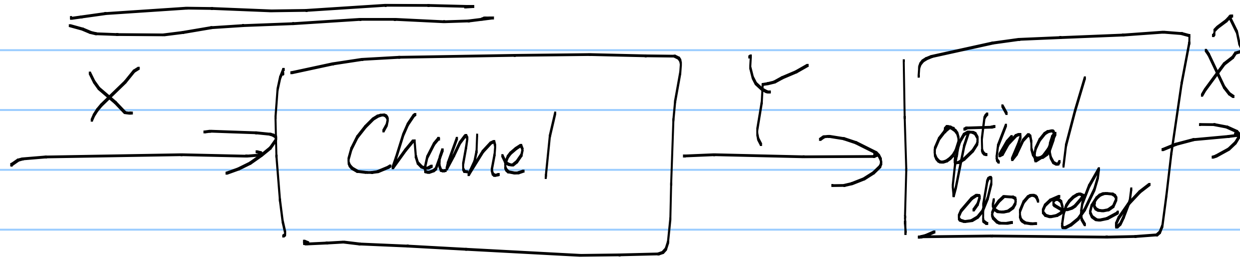
$$\vec{x}_{12} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{x}_{13} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \quad \vec{x}_{14} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad \vec{x}_{15} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Q: What is the optimal decoder?

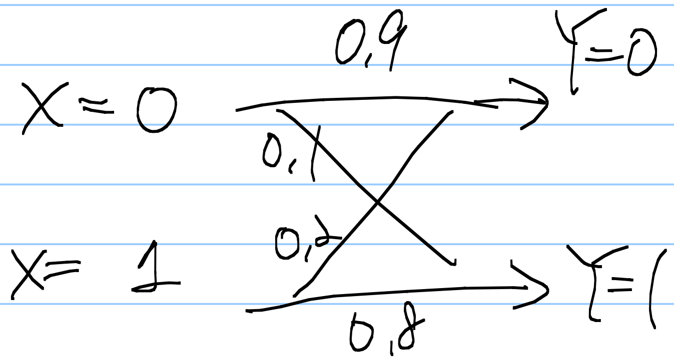
Syndrome decoder, Error-Trapping decoder, ... ? Which is optimal (mimizing the decoding error.)

Or the minimal distance decoder.

* The answer depends on the channel model.



Example: The i.i.d channel follows



for each x_i

Q: What is the optimal decoder?

$\hat{X}(\vec{y})$ say when the observation is

$\vec{y}=(1110110)$

Ans: It is no different than a hypothesis testing problem with 16 competing candidates.

H_0 : $Y_1, \cdots, Y_n$ follow $P_{\vec{Y}|\vec{X}}(\cdot|\vec{x}_0)$

H_1 $\vdots$

$\vdots$

H_{15} $P_{\vec{Y}|\vec{X}}(\cdot|\vec{x}_{15})$

with $P(\vec{X}=\vec{x}_i)=\frac{1}{16}$ for $i=0,\cdots,15$

The optimal decoder is simply the

$\hat{X}_{MAP}(\vec{y})$ MAP decoder

(or the ML decoder

since $P(\vec{X}=\vec{x}_i)$ is uniform

The 16 likelihood values are

$L_0=(0.1)^5\times 0.9^2$ $P_{\vec{Y}|\vec{X}}(1110110|0000000)$

$L_1=0.1^2\times 0.9\times 0.8^3\times 0.2$ $P_{\vec{Y}|\vec{X}}(1110110|1110001)$

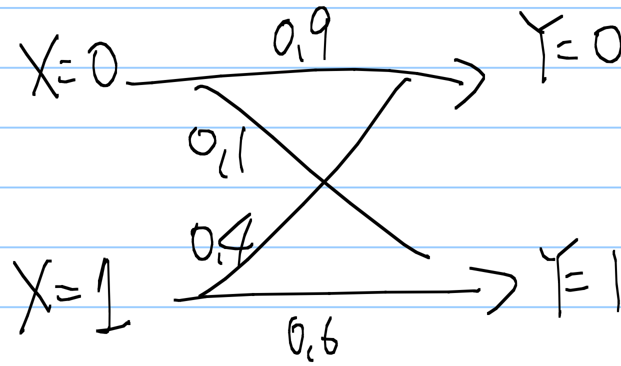
$L_2=0.1^2\times 0.9^2\times 0.8^3$ $P_{\vec{Y}|\vec{X}}(1110110|0110010)$

$\vdots$

$$\hat{\vec{X}}_{MAP}(\vec{y}) = \hat{\vec{X}}_{ML}(\vec{y}) = \underset{\vec{x}_i}{\arg\max}\ P_{\vec{Y}|\vec{X}}(\vec{y}|\vec{x})$$

$$= \vec{x}_6 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

Q: if the channel model becomes, what is the optimal decoder?



Ans:

The 16 likelihood values are

$L_0 = (0{,}1)^5 \times 0{,}9^2 \qquad P_{\vec{Y}|\vec{X}}(1110110 \mid 0000000)$

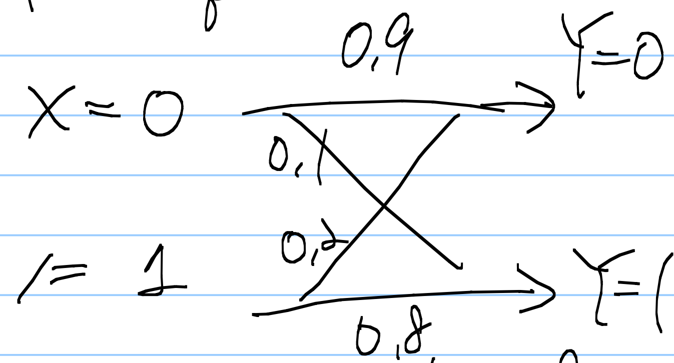
$L_1 = 0{,}1^2 \times 0{,}9 \times 0{,}6^3 \times 0{,}4 \qquad P_{\vec{Y}|\vec{X}}(1110110 \mid 1110001)$

$L_2 = 0{,}1^2 \times 0{,}9^2 \times 0{,}6^3 \qquad P_{\vec{Y}|\vec{X}}(1110110 \mid 0110010)$

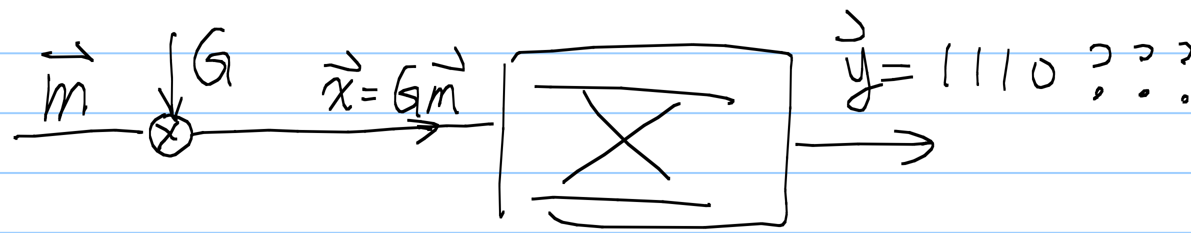
$\vdots$

$$\vec{X}_{MAP}(1110110) = \vec{x}_{15} = (1111111)$$

Q: Assume the same channel model as in the first question:



Suppose now only the first 4 bits can be observed



Find the most likely codeword:

The 16 likelihood values are

$L_0 = (0.1)^3 \times 0.9 \qquad P_{\vec{Y}|\vec{X}}(1110\square\square\square \mid 0000000)$

$L_1 = 0.8^3 \times 0.9 \qquad P_{\vec{Y}|\vec{X}}(1110\square\square\square \mid 1110001)$

$L_2 = 0.1 \times (0.8)^2 \times 0.9 \qquad P_{\vec{Y}|\vec{X}}(1110\square\square\square \mid 0110010)$

$\vdots$

$$\hat{X}_{MAP}(\vec{y}) = \arg\max_{\vec{x}_m} P_{\vec{X}|\vec{Y}}(\vec{x}_m \mid \vec{y})$$

$$= \vec{x}_1$$