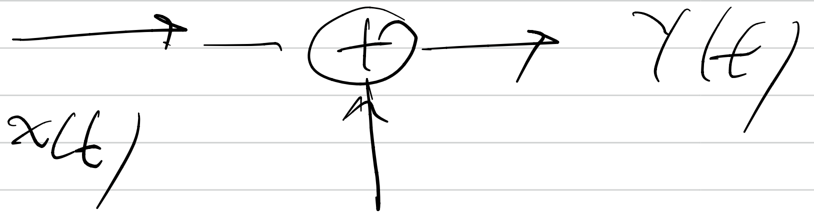
Information Theory & Coding

Nov 16th, 2020

Last week:

- waveform channels



$Z(t)$: white Gaussian noise, intensity $\frac{N_0}{2}$

bandlimited: $X(f)$ = Fourier Xform of $x(t)$

$$X(f) = 0 \quad |f| > W$$
← bandwidth

power constrained: $\frac{1}{2T}\int_{-T}^{T} x^2(t)\,dt \le P$

We found:

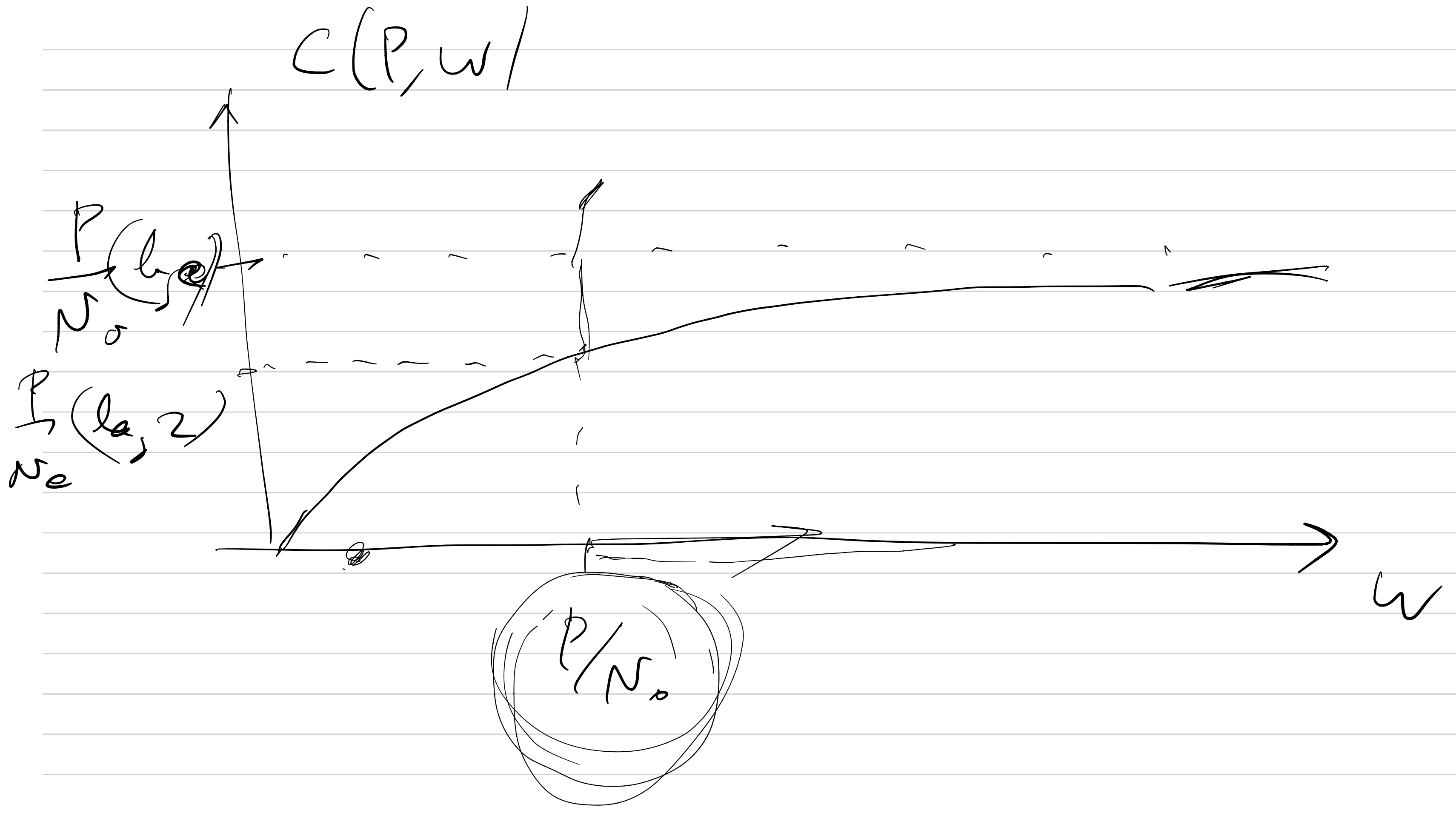
$\Rightarrow$ capacity: $W \log\left(1 + \frac{P}{N_0 W}\right)$ bits/sec.

Capacity is $\nearrow$ in P

also $\nearrow$ in W.

$$\ln(1+z) \le z \Rightarrow \log\left(1 + \frac{P}{N_0 W}\right) \le (\log e)\frac{P}{N_0 W}$$

$\Rightarrow$ capacity $\le \frac{P}{N_0}(\log e)$ bits/sec



P = power = energy/sec

C = bits/sec

$\frac{C}{P}$: bits/energy $= \dfrac{W \log\left(1 + \frac{P}{N_0 W}\right)}{P}$

$\leq (\log e)\frac{1}{N_0}$

attained at $W \to \infty$

or $P \downarrow 0$.

$\Rightarrow$ minimum energy/bit for reliable comm over an AWGN channel with noise intensity $N_0/2$

is $\dfrac{N_0}{(\log e)} = N_0(\ln 2)$.

Parallel Additive noise channels:

$$\begin{cases} x_1 \longrightarrow & x_1 + Z_1 =: Y_1 & Z_1 \sim N(0,\sigma_1^2) \\ x_2 \longrightarrow & x_2 + Z_2 =: Y_2 & Z_2 \sim N(0,\sigma_2^2) \\ \vdots & & \\ x_K \longrightarrow & x_K + Z_K =: Y_K & Z_K \sim N(0,\sigma_K^2) \end{cases}$$

input is $(x_1 .. x_K) \in \mathbb{R}^K$

output " $(Y_1 .. Y_K) \in \mathbb{R}^K$

with $(Z_1 .. Z_K) \sim N\left(0, \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_K^2 \end{bmatrix}\right)$

Input is globally constrained:

$$E\big(\underbrace{X_1^2 + .. + X_K^2}\big) \le \beta.$$

$$\underline{b(x_1 .. x_K)} = \sum_{i=1}^K x_i^2$$

What is the capacity of this channel?

$$\max_{\sum_{i=1}^K E(X_i^2) \le \beta} I(X_1 .. X_K ; Y_1 .. Y_K)$$

first note:

$$I(X_1 \ldots X_k ; Y_1 \ldots Y_k)$$

$$= h(Y_1 \ldots Y_k) - h(Y_1 \ldots Y_k \mid X_1 \ldots X_k)$$

$$= h(Y_1 \ldots Y_k) - h(Z_1 \ldots Z_k \mid X_1 \ldots X_k)$$

$$= \quad '' \quad - h(Z_1 \ldots Z_k)$$

$$\leq \sum_{i=1}^{k} h(Y_i) - \sum_{i=1}^{k} h(Z_i)$$

$\leftarrow$ equals

equals if $X_1 \ldots X_k$ are indep

so: $I(X_1 \ldots X_k ; Y_1 \ldots Y_k) \leq \sum_{i=1}^{k} I(X_i ; Y_i)$

equals if $X_1 \ldots X_k$ indep

$$I(X_i ; Y_i) \leq \frac{1}{2} \log \left(1 + \frac{E(X_i^2)}{\sigma_i^2}\right)$$

equality if $X_i \sim N(0, \cdot)$

i.e.: $\max I(X_1 \dots X_k ; Y_1 \dots Y_k)$ /

$\sum E(X_i^2) \le \beta$

$$= \max_{\substack{\beta_1 \dots \beta_k \ge 0 \\ \sum \beta_i \le \beta}} \sum_{i=1}^{k} \frac{1}{2} \log\left(1 + \beta_i / \sigma_i^2\right)$$

$$= \max_{\substack{q_1 \dots q_k \ge 0 \\ \sum_{i=1}^{k} q_i = 1}} \sum_{i=1}^{k} \frac{1}{2} \log\left(1 + \beta \frac{q_i}{\sigma_i^2}\right) \qquad q_i = \frac{\beta_i}{\beta}$$

∴ this a maximization problem over the k-simplex:

$$\max \{ \underbrace{f(q_1 \dots q_k)}_{\text{concave}} : (q_1 \dots q_k) \in S_k \}$$

⇒ we know that the KKT conditions are necessary & sufficient.

KKT: $\exists \mu$ s.t

$$\frac{\partial f}{\partial q_i} \leq \mu \quad \forall i = 1..k$$

$$= \mu \quad \forall i : q_i > 0.$$

$$\frac{\partial f}{\partial q_i} = \frac{1}{2} \frac{(\beta/\sigma_i^2)}{1 + \beta q_i/\sigma_i^2}$$

$$\text{KKT} \equiv \frac{\sigma_i^2}{\beta} + q_i \geq \lambda \quad \forall i$$

$$= \quad \forall i : q_i > 0.$$

$$q_i = \lambda - \frac{\sigma_i^2}{\beta} \quad i \text{ s.t } q_i > 0$$

$$q_i = 0 \geq \lambda - \frac{\sigma_i^2}{\beta} \quad i \text{ s.t } q_i = 0.$$

$$\equiv \quad q_i = \left(\lambda - \sigma_i^2/\beta\right)^+$$

$$a^+ = \begin{cases} a & a \geq 0 \\ 0 & \text{else.} \end{cases}$$

The optimal $(q_1 \dots q_k)$ is given by

$$q_i = \left(\lambda - \sigma_i^2/\beta\right)^+ \qquad \lambda \text{ chosen s.t.}$$

$$\sum_{i=1}^{k} q_i = 1.$$

$$\frac{\beta q_i}{\sigma_i^2} = \left(\frac{\beta\lambda}{\sigma_i^2} - \frac{\sigma_i^2}{\sigma_i^2}\right)^+$$

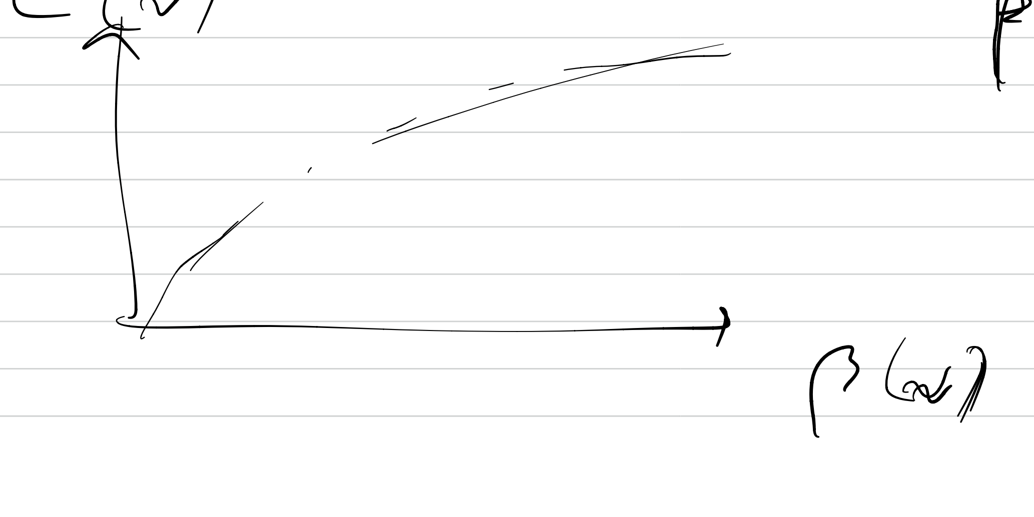
$$C = \sum_{i=1}^{k} \frac{1}{2}\log\left(1 + \left(\frac{\beta\lambda}{\sigma_i^2} - 1\right)^+\right)$$

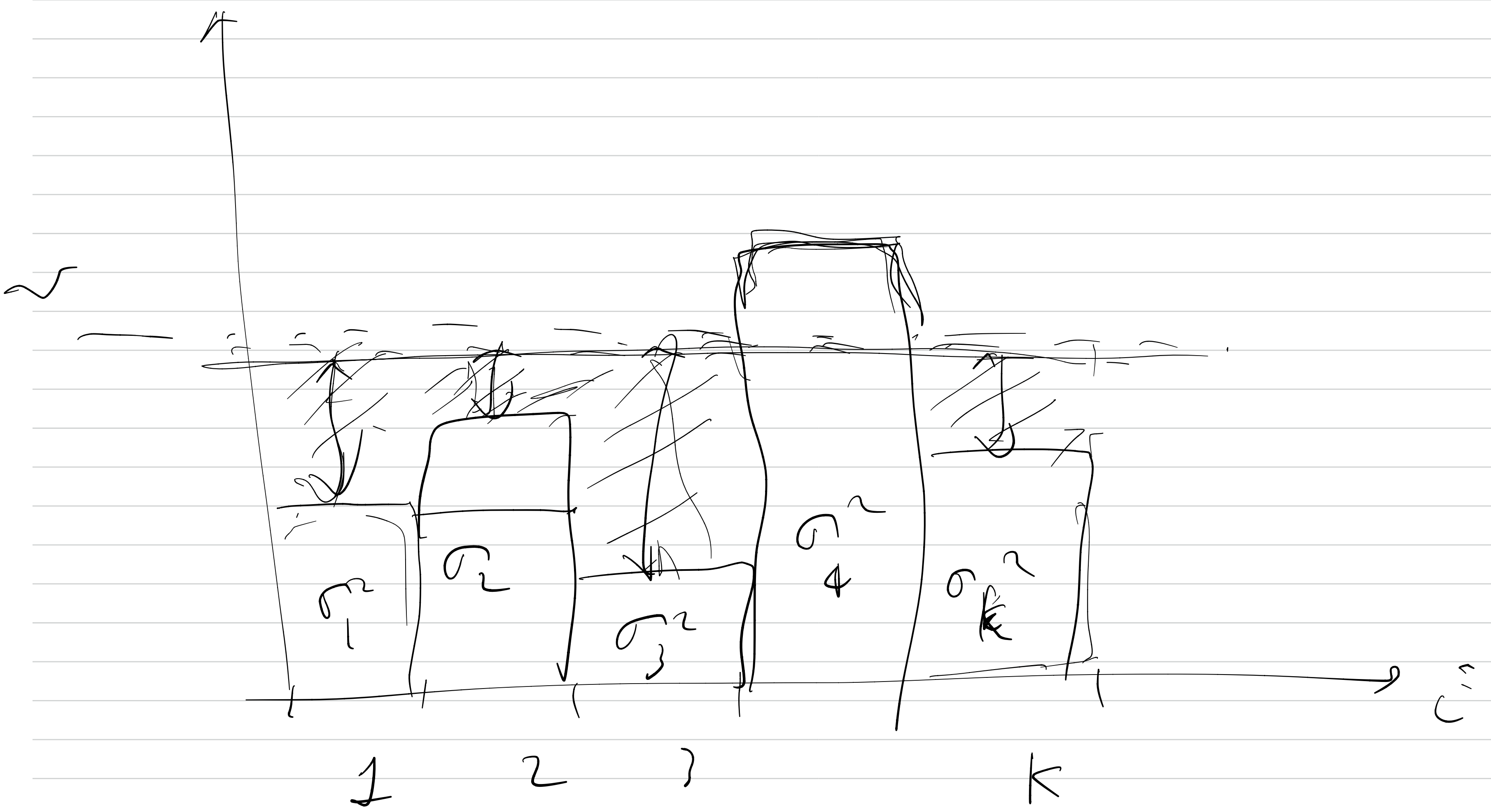
$$\beta = \sum_{i=1}^{k} \left(\beta\lambda - \sigma_i^2\right)^+$$

$$\nu = (\beta\lambda)$$

$$\Rightarrow \quad C(\nu) = \frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \left(\frac{\nu}{\sigma_i^2} - 1\right)^+\right)$$

$$\beta(\nu) = \sum_{i=1}^{k} \underbrace{\left(\nu - \sigma_i^2\right)^+}_{\beta_i}$$





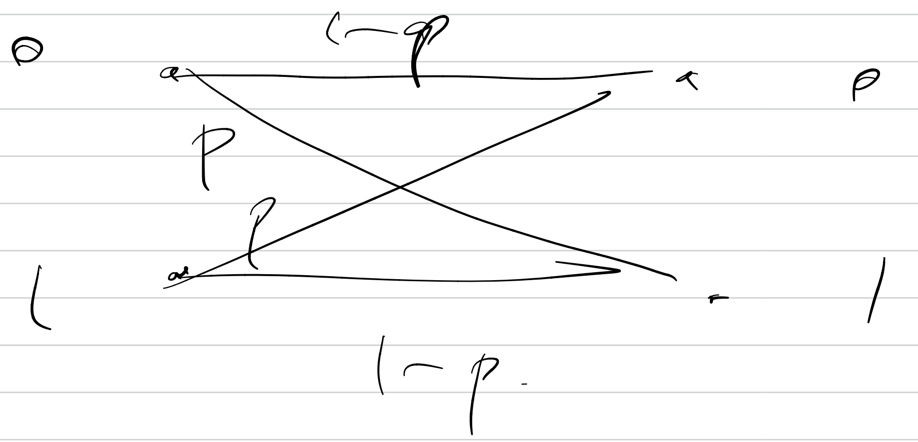
shaded area $= P$.

"water-filling solution".

Back to the discrete world:

Binary Symmetric Channel $BSC(p)$.

$\mathcal{X} = \{0,1\}$ $\mathcal{Y} = \{0,1\}$



$$Y = X \oplus Z$$ mod 2

$$Z = \begin{cases} 0 & \text{w.p. } 1-p \\ 1 & \text{w.p. } p \end{cases}$$

$\perp\!\!\!\perp$ of X.

We can assume, without loss of generality that $p \leq \frac{1}{2}$.

Coding for the BSC, decoding for BSC.

$M = 2^{nR}$ messages,

$enc(1) = x_1(1) \ldots \quad x_n(1)$

$enc(2) = x_1(2) \ldots \quad x_n(2)$

$\vdots$

$enc(M) = x_1(M) \ldots \quad x_n(M).$

if the decoder receives $y_1 \ldots y_n$, assuming each message $\{1 \ldots M\}$ is equally likely, then the decoding rule that minimizing the error probability is

$$\underset{1 \le m \le M}{\arg\max} \underbrace{\Pr(Y_1 \dots Y_n = y_1 \dots y_n \mid m \text{ is sent})}$$

$\Pr(Y_1 \dots Y_n = y_1 \dots y_n \mid m \text{ is sent})$

$$= \Pr(\underline{Y_1 \dots Y_n = y_1 \dots y_n} \mid \underline{X_1 \dots X_n = x_1(m) \dots x_n(m)})$$

$$= \prod_{i=1}^{n} \begin{cases} p & y_i \ne x_i(m) \\ 1-p & y_i = x_i(m) \end{cases}$$

$$= p^{d(\bar{x}(m), \bar{y})} (1-p)^{n - d(\bar{x}(m), \bar{y})}.$$

$$\underline{d(\bar{x}, \bar{y})} = \sum_{i=1}^{n} \mathbb{1}\{x_i \ne y_i\}.$$

= Hamming distance between $\bar{x}$ & $\bar{y}$.

$\Rightarrow$

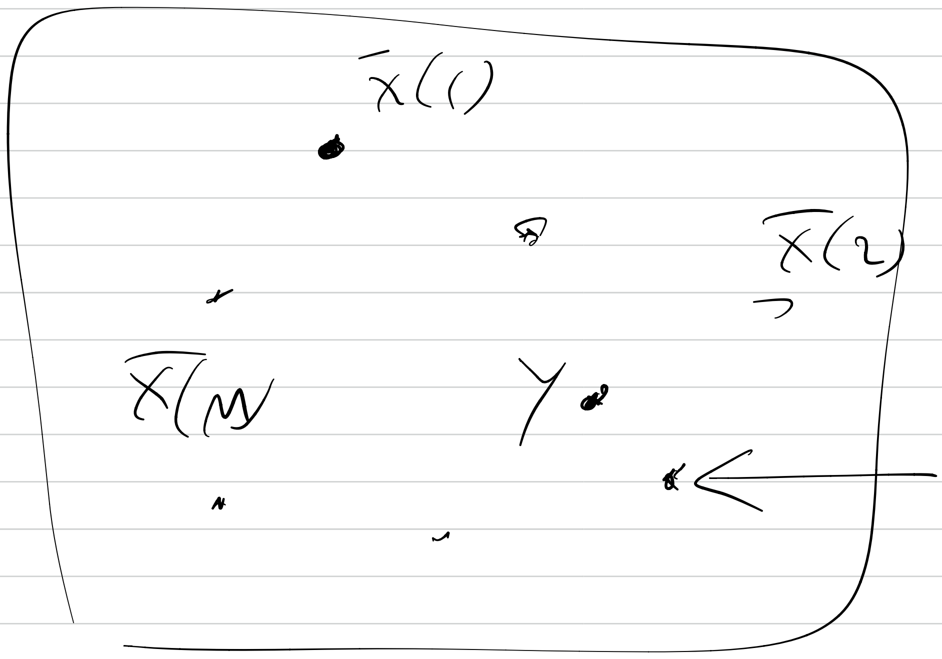
$P(Y_1 \dots Y_n = y_1 \dots y_n \mid m \text{ is sent})$

$$= \underbrace{\left(\frac{p}{1-p}\right)}_{<1}^{d(\bar{y}, \bar{x}(m))} (1-p)^n$$

a decreasing funct of $d(\bar{y}, \bar{x}(m))$.

So the decoding rule is

$$\hat{m} = \arg\min_{1 \le m \le M} d(\bar{y}, \bar{x}(m))$$



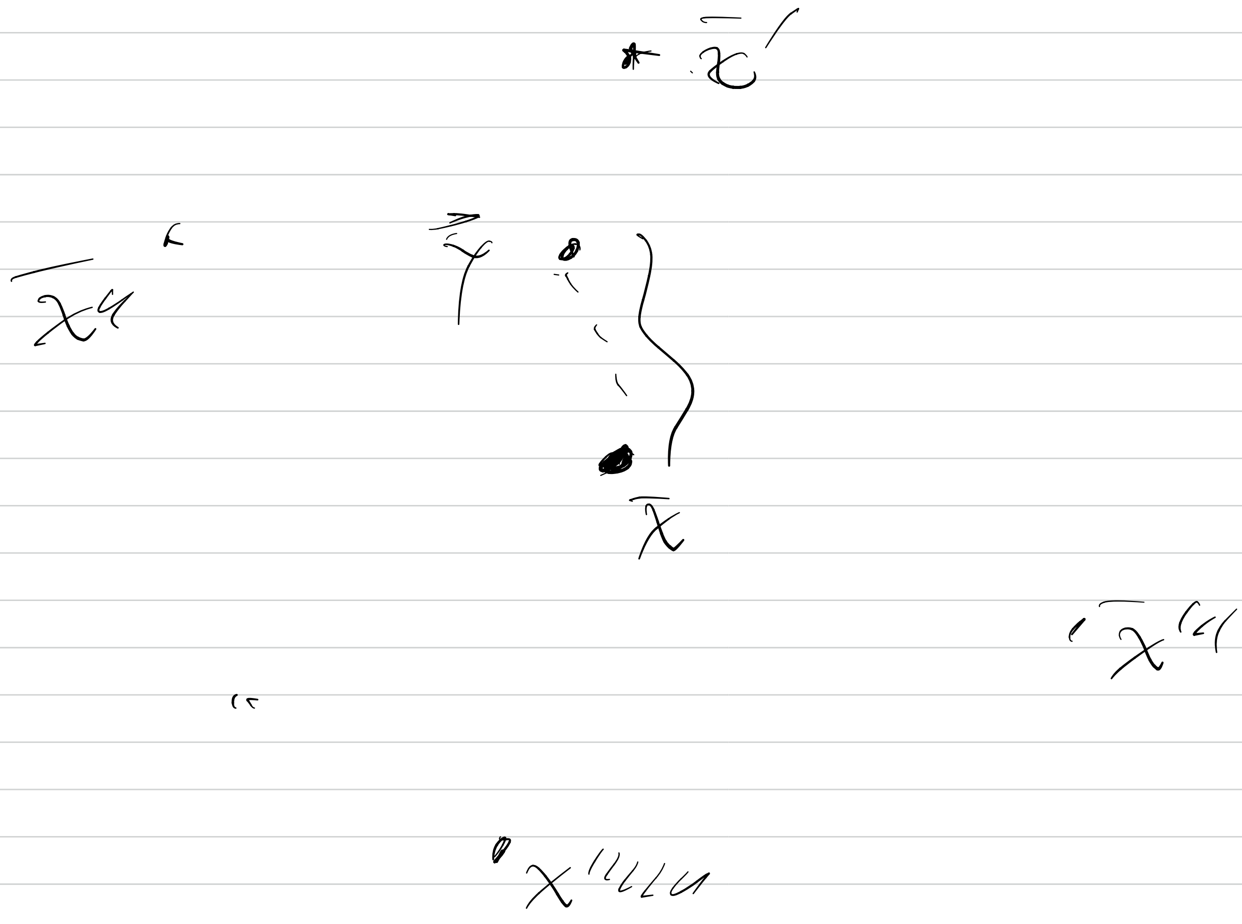
minimum distance decoder.

RCUR:

if $\bar{x}$ is sent.

$Y = \bar{x} \oplus \bar{z}$ is received

$$d(\bar{x}, \bar{Y}) = \underbrace{\sum_{i=1}^{n} \mathbb{1}\{z_i = 1\}}_{\text{Hamming weight of } (z_1 \dots z_n)}$$



if a code has a large value of

$d(\bar{x}(m), \bar{x}(m'))$ for each m, m'

then the probability of error will be small

So for the BSC, a "figure of merit" of a code with codewords $\bar{x}(1) \ldots \bar{x}(M)$ is

$$d_{\min} = \min_{\substack{m, m' \\ m \neq m'}} d(\underline{x(m)}, \underline{x(m')})$$

if $d_{\min}$ is large it suggests that the error probability is small.

For the BSC, we might be interested in the following design problem:

Given M and n, design $\bar{x}(1), \ldots, \bar{x}(M)$ in $\underline{\{0,1\}^n}$, to maximize

$$\min_{\substack{m, m' \\ m \neq m'}} d(x(m), x(m')).$$

Easy example: $M = 2$

$\exists$ $x(1)$ & $x(2)$ complements of each other, e.g., $x(1) = 00\cdots0$

$x(2) = 11111$

will make $d_{min} = n$.

for $M = 3$ best $d_{min} = \left\lfloor \frac{n}{2} \right\rfloor$.

Formal definitions.

$\{0,1\}$: binary field.

$\{0,1\}^n$: binary sequences of length n.

(Hamming space of dimension n).

if $\bar{x}, \bar{y} \in \{0,1\}^n$

$$d_H(\bar{x}, \bar{y}) = \sum_{i=1}^{n} \mathbb{1}\{x_i \neq y_i\}$$ Hamming distance

$$w_H(\bar{x}) = \sum_{i=1}^{n} \mathbb{1}\{x_i \neq 0\}$$ Hamming weight

$$d_H(\bar{x}, \bar{y}) = w_H(\underbrace{\bar{x} \oplus \bar{y}})$$

Lemma: d_H is indeed a distance, namely,

① $d_H(\bar{x},\bar{y}) \geq 0$ & $d_H(\bar{x},\bar{y}) = 0 \iff \bar{x}=\bar{y}$.

② $d_H(\bar{x},\bar{y}) = d_H(\bar{y},\bar{x})$.

③ $d_H(\bar{x},\bar{y}) \leq d_H(\bar{x},\bar{z}) + d_H(\bar{z},\bar{y})$.

(Triangle ineq.)

Pf: ① & ② are obvious.

for ③ it is sufficient to prove for $n=1$.

because $\boxed{d(x_i,y_i) \leq d(x_i,z_i) + d(z_i,y_i)}$

$\Rightarrow \sum_i \leq \sum_i + \sum_i$

$$\mathbb{1}\{x\neq y\} \overset{?}{\leq} \mathbb{1}\{x\neq z\} + \mathbb{1}\{z\neq y\} \quad \forall x,y,z \in \{0,1\}$$

if $x=y \Rightarrow$ ok.

if $x\neq y \Rightarrow$ then either $z\neq x$ or $z\neq y$

$\Rightarrow$ ok //.

From these considerations:

Lemma:
if a code $enc: \{1 .. M\} \to \{0,1\}^n$

has minimum distance d,

$$d = \min_{\substack{m, m' \\ m \neq m'}} d_H(enc(m), enc(m'))$$

then $Pr(\underline{\underline{\mathcal{E}rror}}) \leq Pr(\overbrace{w_H(\underbrace{z_1 .. z_n}_{\text{iid } \{0,1\}}) \geq d/2})$.

on a BSC (p),
$p \leq 1/2$

$1-p$ $\quad$ p

Pf: if the channel noise $z_1 .. z_n$ has
$w_H(z_1 .. z_n) < d/2$ then

$$d_H(\overline{x(m)}, \overline{Y}) = w_H(Z_1 \dots Z_n) < d/2$$

true message

$$d \le d_H(x(m), x(m')) \le \underbrace{d_H(x(m), \overline{Y})}_{< d/2} + d_H(x(m'), \overline{Y})$$

some other message

$$\Rightarrow d_H(\overline{x(m')}, \overline{Y}) > d/2$$

$\Rightarrow$ $x(m)$ is closer to Y than any other $x(m')$

$\Rightarrow$ minimum distance decoder will not make an error. //

<u>Tomorrow</u>: tradeoff between (n, M, d).