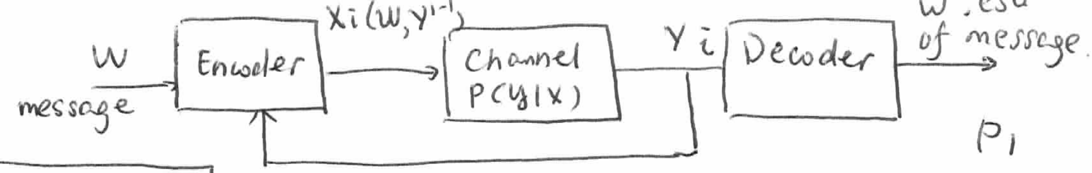

Lecture 16

Feedback capacity theorem

feedback : $(2^{nR}, n)$ feedback code

$$x_i(\underbrace{w}_{\in\{1,\cdots,2^{nR}\}}, Y^{i-1}) \quad Y^{i-1} \to \{y_1, \cdots, y_{i-1}\}$$

decoding function: $g : \mathcal{Y}^n \to \{1, 2 \cdots, 2^{nR}\}$

Define $P_e^{(n)} = Pr\{g(Y^n) \neq W\}$

W : uniformly distributed over $\{1, \cdots, 2^{nR}\}$

Def C_{FB} = sup {all rates achieved by feedback codes}

Theorem: $C_{FB} = C = \max\limits_{p(x)} I(X;Y)$

1. Since nonfeedback code is a special case of a feedback code, any rate that can be achieved w/o feedback can be achieved w. feedback:

$$C_{FB} \geq C$$

2. Converse cannot be proved the same way as we prove converse of channel capacity theorem:

$$I(X^n; Y^n) \not\leq nC$$

Since X_i^n depends on past received symbols

However, we are not too far away from desired proof.

instead of using X^n, we'll use W

W unif over $\{1, \dots, 2^{nR}\}$

$\Rightarrow P(W \neq \hat{W}) = P_e^{(n)}$

$$
\begin{aligned}
nR = H(W) &= H(W \mid \hat{W}) + I(W; \hat{W}) \\
&\leq 1 + P_e^{(n)} nR + I(W; \hat{W}) \\
&\qquad\qquad \text{(Fano)} \\
&\leq 1 + P_e^{(n)} nR + I(W; Y^n) \\
&\qquad\qquad \text{(data processing)}
\end{aligned}
$$

$$
\begin{aligned}
I(W; Y^n) &= H(Y^n) - H(Y^n \mid W) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_1, \dots, Y_{i-1}, W) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_1, \dots, Y_{i-1}, X_i, W) \\
&\qquad \text{(Since } X_i \text{ is function of } Y_1 \dots Y_{i-1} \text{ and } W) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) \\
&\qquad \text{(conditioning on } X_i\text{, } Y_i \text{ independent of } Y_{i-1}, \dots Y_1 \text{ and } W) \\
&= \sum_{i=1}^{n} I(X_i; Y_i) \\
&\leq nC
\end{aligned}
$$

$\Rightarrow nR \leq 1 + P_e^{(n)} nR + nC$

$\Rightarrow R \leq \frac{1}{n} + P_e^{(n)} R + C \xrightarrow{n \to \infty} R \leq C$

Source channel separation theorem

sourse $V \in \mathcal{V}$: finite alphabet,
satisfies AEP

(e.g. iid-random variables
- sequence of states of a stationary irreducible Markov chain)

(any stationary ergodic sources satisfy AEP)

Send a sequence of symbols
- $V^n = V_1 V_2 \cdots V_n$
- map this into a channel codeword $X^n(V^n)$
- receives y^n,
- estimate $\hat{V}^n$ using y^n
- make an error if $V^n \neq \hat{V}^n$

$$P(V^n \neq \hat{V}^n) = \sum_{y^n} \sum_{v^n} P(v^n) P(y^n | x^n(v^n)) \cdot I(g(y^n) \neq v^n)$$

$$I_x = \begin{cases} 1 & \text{if } x \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

g : decoding function.

Achievability

two-stage encoder :

Since input seq. satisfies AEP
→ exists a typical set $A_\varepsilon^{(n)}$ of size $\leq 2^{n(H(\mathcal{V})+\varepsilon)}$
which contains most of the prob.

- we only encode sequence in typical set
all others just err (incur at most ε
of prob. of err)

- indexing seq in $A_\varepsilon^{(n)}$
using $n(H+\varepsilon)$ bits

- we can transmit desired index to receiver
with error $< \varepsilon$, if

$$H(\mathcal{V}) + \varepsilon \doteq R < C$$

- receiver can reconstruct V^n by enumerating
the typical set $A_\varepsilon^{(n)}$

$$Pe = P(V^n \neq \hat{V}^n) \leq P(V^n \notin A_\varepsilon^{(n)}) + \underbrace{P(g(Y^n) \neq V^n \mid V^n \in A_\varepsilon^{(n)})}_{\text{joint typicality \& channel-coding theorem.}}$$
$$\leq \varepsilon + \varepsilon = 2\varepsilon$$

⇒ we can reconstruct the seq w. low
prob. of error for n large, and

$$H(\mathcal{V}) < C$$

Converse

Goal prove that $P(\hat{V}^n \neq V^n) \to 0$ implies that $H(\mathcal{V}) \leq C$ for the pair of source-channel codes

$$X^n(V^n) : \mathcal{V}^n \to \mathcal{X}^n \quad : \text{encoding}$$
$$g_n(Y^n) : \mathcal{Y}^n \to \mathcal{V}^n \quad : \text{decoding}$$

by Fano's inequality

$$H(V^n | \hat{V}^n) \leq 1 + P_e \log |\mathcal{V}^n|$$
$$= 1 + P_e n \log |\mathcal{V}|$$

$$H(\mathcal{V}) \leq \frac{H(V_1, \cdots, V_n)}{n}$$ (def. of entropy rate)

$$= \frac{1}{n} \sum_{i=1}^{n} \underbrace{H(X_i | X_{i-1}, \cdots, X_1)}_{\text{decreasing}}$$

$$= \frac{H(V^n)}{n}$$

$$= \frac{1}{n} H(V^n | \hat{V}^n) + \frac{1}{n} I(V^n ; \hat{V}^n)$$

$$\leq \frac{1}{n}(1 + P_e n \log |\mathcal{V}|) + \frac{1}{n} I(V^n ; \hat{V}^n)$$

(Fano)

$$\leq \frac{1}{n}(1 + P_e n \log |\mathcal{V}|) + \frac{1}{n} I(X^n ; Y^n)$$

(data processing)

$$V^n \to X^n \to Y^n \to \hat{V}^n$$

$$\leq \frac{1}{n} + P_e \log |\mathcal{V}| + C$$

(memoriless of channel, no feedback)

$n \to \infty \qquad H(\mathcal{V}) \leq C$