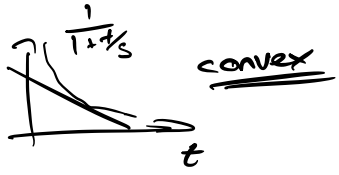
6.2.3 Achievability and random codes

Lemma: Let $t \geq 0$, $s \in [0,1]$. Then, $1 - \frac{s}{s+t} \leq 1 - s + t$.

Proof: $\iff 0 \leq t - s + \frac{s}{s+t}$



i) if $t \geq s$ ✓

ii) if $t < s$ then $\frac{s}{s+t} = \frac{1}{1+\frac{t}{s}}$

$\geq 1 - \frac{t}{s} = \frac{s-t}{s} \geqslant s - t$ ✓ □

(tangent at $t=0$)

Proposition: For any $\varepsilon, \delta \in (0,1)$ s.t. $\varepsilon + \delta < 1$ there exists an $(\varepsilon+\delta, |\mathcal{M}|, 1)$-channel code for W as long as

$$|\mathcal{M}| \leq \delta \cdot \frac{1}{\beta_\varepsilon^*(P_{XY} \| P_X \times P_Y)} \quad \text{for some } P_X \in \mathcal{P}(\mathcal{X})$$

Recall meta-converse: Always have $|\mathcal{M}| \leq \max_{P_X} \frac{1}{\beta_\varepsilon^*(P_{XY} \| P_X \times P_Y)}$

Proof: Fix P_X. We generate $|\mathcal{M}|$ codewords by picking them at random (sampling them) from P_X.

- The Encoder $E(m)$ is a random variable.
- The Decoder is constructed using a hypothesis test

$H_0: P_{XY}$ $\quad H_1: P_X \times P_Y$

There exist a test $A \subset \mathcal{X} \times \mathcal{Y}$ such that

$$P_{XY}(A) \leq \varepsilon \quad \text{and} \quad (P_X \times P_Y)(A^c) = \beta_\varepsilon^*(P_{XY} \| P_X \times P_Y)$$

Define $A_x = \{y \in \mathcal{Y} : (x,y) \notin A\}$

The decoder is probabilistic! For a fixed encoder $E = e$

$$\Pr[\hat{M} = m \mid Y = y] = \frac{\mathbb{1}\{y \in A_{e(m)}\}}{\sum_{m'} \mathbb{1}\{y \in A_{e(m')}\}}$$

$$= \frac{\mathbb{1}\{y \in A_{e(m)}\}}{\mathbb{1}\{y \in A_{e(m)}\} + \sum_{m' \neq m} \mathbb{1}\{y \in A_{e(m')}\}} \quad (*)$$

Probability of error:

$$\Pr[\hat{M} \neq M \mid M = m, E = e] = 1 - \sum_y W(y \mid e(m)) \Pr[\hat{M} = m \mid Y = y]$$

$$= \sum_{y \in \mathcal{Y}} W(y \mid e(m)) \left(1 - (*)\right)$$

$$\leq \sum_{y \in \mathcal{Y}} W(y \mid e(m)) \left(\underbrace{1 - \mathbb{1}\{y \in A_{e(m)}\}}_{= \mathbb{1}\{y \notin A_{e(m)}\}} + \sum_{m' \neq m} \mathbb{1}\{y \in A_{e(m')}\} \right)$$

Now we take the average over encoders e:

$$\Pr[\hat{M} = M \mid M = m] \leq \sum_{x,y} P_x(x) W_y(y \mid x) \Big(\mathbb{1}\{y \notin A_x\} + \underbrace{(|\mathcal{M}| - 1)}_{\leq |\mathcal{M}|} \sum_{x' \in \mathcal{X}} P_x(x') \mathbb{1}\{y \in A_{x'}\} \Big)$$

upper bound does not depend on m!

1) $\sum P_{XY}(x,y) \mathbb{1}\{y \notin A_x\} = P_{XY}(A) \leq \varepsilon$

$$2) \sum_{x,y} P_{xy}(x,y) \sum_{x'} P_x(x') \mathbb{1}\{y \in A_{x'}\}$$

$$= \sum_{x',y} P_x(x') P_y(y) \mathbb{1}\{y \in A_{x'}\}$$

$$= (P_x \times P_y)(A^c) = \beta_\varepsilon^*(P_{xy} \| P_x \times P_y)$$

$$\Rightarrow \Pr[\hat{M} \neq M] \leq \varepsilon + |M| \cdot \beta_\varepsilon^*(P_{xy} \| P_x \times P_y)$$

$\Rightarrow$ if we choose $|M| \leq \delta \cdot \frac{1}{\beta_\varepsilon^*(P_{xy} \| P_x \times P_y)}$

then $\Pr[\hat{M} \neq M] \leq \varepsilon + \delta$! $\square$

Proof of channel coding theorem (achievability).

We need to shows that for any rate $R < I(W)$ there exists a sequence of $(\varepsilon_n, 2^{\lceil nR \rceil}, n)$-codes with $\lim_{n \to \infty} \varepsilon_n = 0$ for some sequence ε_n.

From one-shot bound:

If $2^{\lceil nR \rceil} \leq \varepsilon \frac{1}{\beta_\varepsilon^*(P_{x^n y^n} \| P_{x^n} \times P_{y^n})}$ (*)

where P_{x^n} is i.i.d. such that $I(X:Y)_P = I(W)$.

then there exists a $(2\varepsilon, 2^{\lceil nR \rceil}, n)$-code for W

$\leftarrow R < \frac{1}{n} \left(\log \frac{1}{\beta_\varepsilon^*(P_{xy} \| P_x \times P_y)} + \log \varepsilon - 1 \right)$

$*$ $\Leftarrow$ [illegible] (***)

However, we know that (Stein's lemma)

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_\varepsilon^n(P_{XY^n} \| P_X \times P_{Y^n})} = D(P_{XY} \| P_X \times P_Y)$$
$$= I(X:Y)_P$$
$$= I(W)$$

$\Rightarrow$ If we choose $R = I(W) - \mu$ for some $\mu > 0$ then, for every $\varepsilon > 0$, there exists an n_0 s.t. for $n \geq n_0$ the condition (**) is satisfied.

Taking the supremum over all such rates we get $C(W) \geq I(W)$.

6.2.4 Maximum probability of error

so far: average error

$$\Pr[\hat{M} \neq M] = \sum_{m \in [|M|]} \frac{1}{|M|} P[\hat{M} \neq M | M = m]$$

maximum error: $\max_{m \in [|M|]} P[\hat{M} \neq M | M = m]$

Lemma: Given $(\varepsilon, |M|, 1)$-average error code, we can construct an $(2\varepsilon, \frac{|M|}{2}, 1)$-maximum error code.

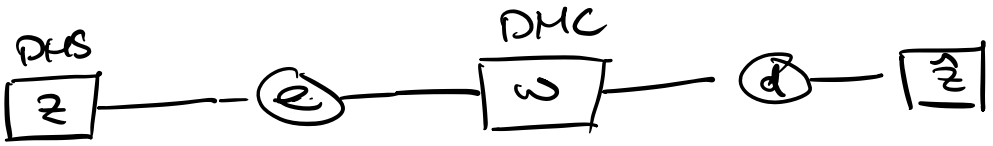
Proof:

Since $\sum_{m \in [|\mathcal{M}|]} \frac{1}{|\mathcal{M}|} P[\hat{M} \neq M | M = m] \leq \varepsilon \quad (*)$

There must be a subset $\mathcal{M}_{good} \subseteq [|\mathcal{M}|]$ of size at least $\frac{|\mathcal{M}|}{2}$ with $P_r[\hat{M} \neq M | M = m] \leq 2\varepsilon$.

as otherwise $(*)$ is violated.

6.3 Source-channel separation theorem



Source-channel separation (informal):
It is optimal to treat source compression and channel coding separately.

Theorem: Given DMS Z, DMC W, there exists a sequence of codes with asymptotically vanishing error if $H(Z) < I(W)$. Moreover, if $H(Z) > I(W)$ such a sequence cannot exist.

Proof: Achievability

channel code
source code

encoder: $z \longmapsto e(\underbrace{C(z)}_{??})$

$\longmapsto d(d(Y))$

decoder: $Y^n \to$ ( ) $\to \hat{m}$

source de-coding ↑ ; channel decoding ↑

$$\Pr[\text{error}] \le \Pr[\hat{M} \neq M] + \Pr[d(M) \neq Z]$$

$\Pr[\hat{M} \neq M]$: channel coding ; $\Pr[d(M) \neq Z]$: source coding

$$Z^n \xrightarrow{\text{enc}} X^n \to Y^n \xrightarrow{\text{dec}} \hat{Z}^n$$

Converse

Assume $H(Z) - I(\omega) = \nu > 0$. If $\exists$ sequences of codes with vanishing error, then $\forall \varepsilon > 0$, $\exists n_0$ s.t. $\forall n \ge n_0$ $\Pr[\hat{Z}^n \neq Z^n] \le \varepsilon$. By Fano's inequality

$$H(Z^n|Y^n) \le H(Z^n|\hat{Z}^n) \le 1 + \varepsilon n \log|\mathcal{Z}|$$

$$H(Z^n|Y^n) = H(Z^n) - I(Z^n : Y^n)$$

$$H(Z^n) - I(Z^n : Y^n) = H(Z^n) - I(X^n : Y^n)$$

$$\Rightarrow \quad n H(Z) - I(X^n : Y^n) \le 1 + \varepsilon n \log|\mathcal{Z}| \quad (*)$$

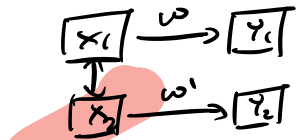
Lemma: $I(X^n : Y^n) \le n I(\omega)$ for any P_{X^n} (not necessarily i.i.d.)

Proof:

Note: $I(X^n : Y^n) \le I(\omega^n)$, so it remains to show additivity: $I(\omega^n) = n I(\omega)$!

For $n = 2$:

$$I(X_1X_2 : Y_1Y_2) = I(X_1 : Y_1Y_2) + I(X_2 : Y_1Y_2 \mid X_1)$$
$$= I(X_1 : Y_1) + I(X_2 : Y_2 \mid X_1)$$
$$\leq I(\omega) + I(\omega)$$



From (*): $H(Z) - I(\omega) \leq \frac{1}{n} + \varepsilon \log|Z|$

arbitrarily small

contradicts $H(Z) - I(\omega) = \nu$!