6. Channel Coding

The primary goals of communication are:

1. Reliability

$$\text{Error probability: } \mathbb{P}[\text{received messgae } \neq \text{ transmitted message}]$$

2. Efficiency

$$\text{Rate: } R = \text{average number of information bits sent per unit time}$$

Unfortunately, these two goals are fundamentally opposed.

▼ Setup

- The message $W \in \{1, 2, \cdots, M\}$ is one of M possible numbers that we want to communicate. This message is generated from a random source and is distributed uniformly over all possibilities.
- An (M, n) coding scheme consists of an encoder (or codebook) $\mathcal{E}$ that maps the message W to an n-length sequence of channel inputs X^n:

$$\mathcal{E} : \{1, \cdots, M\} \rightarrow \mathcal{X}^n$$

- and a decoder that maps n-length sequences of channel outputs Y^n to an estimate $\hat{W}$ of the message.

$$\mathcal{D} : \mathcal{Y}^n \rightarrow \{1, 2, 3, .., M\}$$

- The **channel** specifies the (probabilistic) transformation from inputs to outputs:

$$\mathbb{P}[Y^n = y^n \mid X^n = x^n] = p_{Y^n|X^n}(y^n \mid x^n)$$

- The channel is **memoryless** if the outputs between channel uses are conditionally independent given the input, i.e.,

$$p_{Y^n|X^n}(y^n \mid x^n) = \prod_{i=1}^{n} p_{Y|X}(y_i \mid x_i)$$

The rate R of an (M, n) coding scheme is defined as

$$R = \frac{\log_2 M}{n} \text{ bits/transmission}$$

Alternatively, the number of messages for a given rate R and block-length n is given by

$$M = 2^{nR}$$

To specify a rate R code, we write $(2^{nR}, n)$ instead of (M, n).

- **conditional error probability** :

$$P_e^{(n)}(w) = \mathbb{P}[W \neq \hat{W} \mid W = w]$$

- **average error probability :**

$$P_e^{(n)} = \mathbb{P}[\hat{W} \neq W] = \frac{1}{M}\sum_{w=1}^{M} P_e^{(n)}(w)$$

- **maximum error probability :**

$$P_{\mathrm{e,max}}^{(n)} = \max_{w\in\{1,\cdots,M\}} P_e^{(n)}(w)$$

▼ Discrete Memoryless Channel (DMC)

- Input alphabet $\mathcal{X}$
- Output alphabet $\mathcal{Y}$
- a conditional probability distribution $p_{Y|X}(\cdot|x)$ for all $x \in \mathcal{X}$

$$p_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} p_{Y|X}(y_i|x_i)$$

i**nformation capacity:**

$$C = \max_{p_X(x)} I(X;Y)$$

$$C = \max_{p(x)} \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x)p(y \mid x)\log\left(\frac{p(y \mid x)}{\sum_{x'} p\left(y \mid x'\right) p\left(x'\right)}\right)$$

Channel Coding Theorem: The operational capacity of a discrete memoryless channel is equal to the information capacity

$$C_{op} = \max_{p(x)} I(X;Y)$$

- **Achievability**: Every rate $R < C$ is achievable
- **Converse**: Any sequence of $(2^{nR}, n)$ coding schemes with maximum error probability $P_{e,max}^{(n)}$ converging to zero as the block-length n increase must have rate $R \leq C$.

Lemma : For any input distribution $p_{X^n}(x^n)$, the mutual information between the input X^n and output Y^n of a discrete memoryless channel with capacity C obeys

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

- ▼ Proof

$$
\begin{aligned}
I\left(X^n ; Y^n\right) & =H\left(Y^n\right)-H\left(Y^n \mid X^n\right) \\
& =H\left(Y^n\right)-\sum_{i=1}^n H\left(Y_i \mid Y_1^{i-1}, X^n\right) \quad \text{(Chain rule)} \\
& =H\left(Y^n\right)-\sum_{i=1}^n H\left(Y_i \mid X_i\right) \quad \text{(Channel is Memoryless)} \\
& =\sum_{i=1}^n H\left(Y_i \mid Y_1^{i-1}\right)-\sum_{i=1}^n H\left(Y_i \mid X_i\right) \quad \text{(Chain rule)} \\
& \leq \sum_{i=1}^n H\left(Y_i\right)-\sum_{i=1}^n H\left(Y_i \mid X_i\right) \quad \text{(Conditioning cannot increase entropy)} \\
& =\sum_{i=1}^n I\left(X_i ; Y_i\right) \\
& \leq n C \quad \text{(Definition of } C \text{)}
\end{aligned}
$$

Lemma: (Fano's Inequality) Let $\hat{W}$ be an estimate of the message $W \in \left\{1, \cdots, 2^{nR}\right\}$. The conditional entropy of W given $\hat{W}$ is related to the average error probability $P_e^{(n)}=\mathbf{P}(W \neq \hat{W})$ via the inequality

$$
H(W \mid \hat{W}) \leq 1+P_e^{(n)} n R
$$

▼ Proof of Channel Coding Theorem via Random Coding

Any $\left(2^{nR}, n\right)$ encoder $\mathcal{E}$ can be represented by a codebook $\mathcal{C}$, i.e., a massive lookup table whose rows are length-n vectors:

$$
\mathcal{C}=\left[\begin{array}{c}
x^n(1) \\
x^n(2) \\
\vdots \\
x^n\left(2^{nR}\right)
\end{array}\right]=\left[\begin{array}{cccc}
x_1(1) & x_2(1) & \cdots & x_n(1) \\
x_1(2) & x_2(2) & \cdots & x_n(2) \\
\vdots & \vdots & \ddots & \vdots \\
x_1\left(2^{nR}\right) & x_2\left(2^{nR}\right) & \cdots & x_n\left(2^{nR}\right)
\end{array}\right]
$$

To communicate message w, the encoder sends the w'th codeword:

$$
\mathcal{E}(w)=x^n(w)
$$

▼ Optimal Decoder

The the **probability** of **error** is minimized by the maximum a posteriori (**MAP**) decoder:

$$
\begin{aligned}
\mathcal{D}^{\mathrm{MAP}}\left(y^n\right) & =\arg \max _{w \in\left\{1, \cdots, 2^{nR}\right\}} p_{W \mid Y^n}\left(w \mid y^n\right) \\
& =\arg \max _{w \in\left\{1, \cdots, 2^{nR}\right\}} \frac{p_W(w) p_{Y^n \mid W}\left(y^n \mid w\right)}{p_{Y^n}\left(y^n\right)} \quad \text{Bayes' theorem} \\
& =\arg \max _{w \in\left\{1, \cdots, 2^{nR}\right\}} p_W(w) p_{Y^n \mid W}\left(y^n \mid w\right) \quad p_{Y^n}\left(y^n\right) \text{ does not depend on } w \\
& =\arg \max _{w \in\left\{1, \cdots, 2^{nR}\right\}} p_{Y^n \mid W}\left(y^n \mid w\right) \quad \text{message is uniform} \\
& =\arg \max _{w \in\left\{1, \cdots, 2^{nR}\right\}} p_{Y^n \mid X^n}\left(y^n \mid x^n(w)\right) \quad \text{because } W \rightarrow \mathcal{E}(W) \rightarrow Y^n \text{ forms a Markov chain}
\end{aligned}
$$

The information density associated with the joint distribution of $\left(X^n, Y^n\right)$ is defined as :

$$i\left(x^n ; y^n\right)=\log \left(\frac{p_{X^n, Y^n}\left(x^n, y^n\right)}{p_{X^n}\left(x^n\right) p_{Y^n}\left(y^n\right)}\right)=\log \left(\frac{p_{Y^n \mid X^n}\left(y^n \mid x^n\right)}{p_{Y^n}\left(y^n\right)}\right)$$

Because the input is iid and the channel is memoryless, the information density can be decomposed as

$$i\left(x^n ; y^n\right)=\sum_{k=1}^n i\left(x_k ; y_k\right), \quad \text { where } \quad i(x ; y)=\log \left(\frac{p_{Y \mid X}(y \mid x)}{p_Y(y)}\right)$$

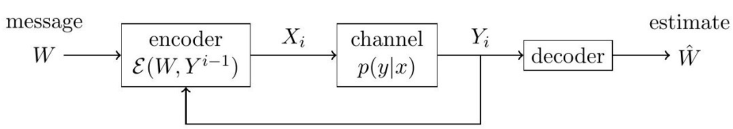
The optimal decoder can be expressed in terms of the information density:

$$\begin{aligned}
D^{\mathrm{MAP}}\left(y^n\right) & =\arg \max _w p_{Y^n \mid X^n}\left(y^n \mid x^n(w)\right) \\
& =\arg \max _w \frac{p_{Y^n \mid X^n}\left(y^n \mid x^n(w)\right)}{p_{Y^n}\left(y^n\right)} \\
& =\arg \max _w i\left(x^n(w) ; y^n\right) \quad \text { because logarithm is increasing }
\end{aligned}$$

S**up-optimal thresholding decoder:** For a given threshold T we define the decoding rule as follows:

$$\mathcal{D}\left(y^n\right)= \begin{cases}\hat{w}, & \text { if } i\left(x^n(\hat{w}) ; y^n\right)>T \text { and } i\left(x^n(w) ; y^n\right) \leq T \text { for all } w \neq \hat{w} \\ 0, & \text { otherise }\end{cases}$$

▼ Channel Coding with Feedback



Encoder $\mathcal{E}\left(W, Y^{i-1}\right)$ can use previous channel outputs

Theorem: Feedback cannot increase capacity. For a discrete memoryless channel, the capacity with feedback, C_{FB}, is the same as the capacity without feedback:

$$C_{\mathrm{FB}}=C$$

▼ Error Correction Codes

Hamming Codes:

Hamming code is an error correction system that can detect and correct errors when data is stored or transmitted.