

# MAT4006: Introduction to Coding Theory

## Lecture 01: Codes, Channels, Decoding Rules

Instructor: Zitan Chen

Scribe: Siqi Yao

Spring 2025

## 1 Basic Definitions

Let  $A$  be a finite set of size  $q$ . We will refer to  $A$  as a *code alphabet*, and an element  $a \in A$  is called a *code symbol*.

### Definition 1.1

1. An element  $x \in A^n$  is called a  *$q$ -ary word of length  $n$*  over  $A$ .
2. Let  $C \subseteq A^n$  be a nonempty subset of  $A^n$ . The set  $C$  is called a  *$q$ -ary block code of length  $n$*  over  $A$ .
3. An element  $c \in C$  is called a *codeword*.
4. The number of codewords in  $C$ , denoted by  $|C|$ , is called the *size of  $C$* .
5. The *rate* of  $C$  is defined by

$$\text{rate}(C) = \frac{\log_q |C|}{n}$$

**Note.** This comes from:

$$\frac{\log_q |C|}{\log_q |A^n|} = \frac{\log_q |C|}{\log_q (q^n)} = \frac{\log_q |C|}{n}.$$

Think about its meaning.

6. A code of length  $n$  and size  $M$  is called an  $(n, M)$  code.

### Examples

Let  $A = \{0, 1\}$ . Define  $C_1 \subset A^2$  by

$$C_1 = \{00, 01, 10, 11\}.$$

Then  $C_1$  is a  $(2, 4)$  code with  $\text{rate}(C_1) = 1$ .

Define  $C_2 \subset A^3$  by

$$C_2 = \{000, 011, 101, 110\}.$$

Then  $C_2$  is a  $(3, 4)$  code with  $\text{rate}(C_2) = \frac{\log_2 4}{3} = \frac{2}{3}$ .  
 $C_1, C_2$  are binary codes since  $|A| = 2$ .

## 2 Model of Channels

### Definition 2.1 (Communication Channel)

A communication channel consists of:

- A finite input alphabet  $A$ ,
- A finite output alphabet  $B$ ,
- And a set of conditional probabilities

$$\{\Pr(b \text{ received} \mid a \text{ sent})\}_{a \in A, b \in B}$$

satisfying

$$\sum_{b \in B} \Pr(b \text{ received} \mid a \text{ sent}) = 1, \quad \forall a \in A$$

### Remarks

- This model is known as a *discrete probabilistic channel*.
- The input and output alphabets are called *channel alphabets*, and the conditional probabilities are called the *channel probabilities*.
- The conditional probabilities can be represented by a *stochastic matrix*:

$$\mathcal{P} = (p_{ij}) \quad \text{for } i \in A, j \in B$$

where

$$p_{ij} := \Pr(j \text{ received} \mid i \text{ sent}).$$

In matrix form:

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1,|B|} \\ \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots \end{pmatrix} \quad \text{with each row summing to 1.}$$

### Definition 2.2 (Memoryless Channel)

A channel  $(A, B, \mathcal{P})$  is said to be *memoryless* if it satisfies the following condition: for any  $\mathbf{x} = (x_1, \dots, x_n) \in A^n$  and  $\mathbf{y} = (y_1, \dots, y_n) \in B^n$ , both of length  $n$ ,

$$\Pr(\mathbf{y} \text{ received} \mid \mathbf{x} \text{ sent}) = \prod_{i=1}^n \Pr(y_i \text{ received} \mid x_i \text{ sent}).$$

### Definition 2.3 (q-ary Symmetric Channel)

Let  $p \in [0, 1]$ . A  $q$ -ary symmetric channel is a memoryless channel  $(A, A, \mathcal{P})$  such that  $|A| = q$ , and

$$\Pr(b \text{ received} \mid a \text{ sent}) = \begin{cases} p & \text{if } a \neq b, \\ 1 - (q-1)p & \text{if } a = b, \end{cases} \quad \text{for all } a, b \in A.$$

### Remarks

A  $q$ -ary symmetric channel, often denoted as  $\text{qSC}(p)$ , is commonly represented by the following diagram.

Let the input alphabet be  $A = \{0, 1, \dots, q-1\}$ .

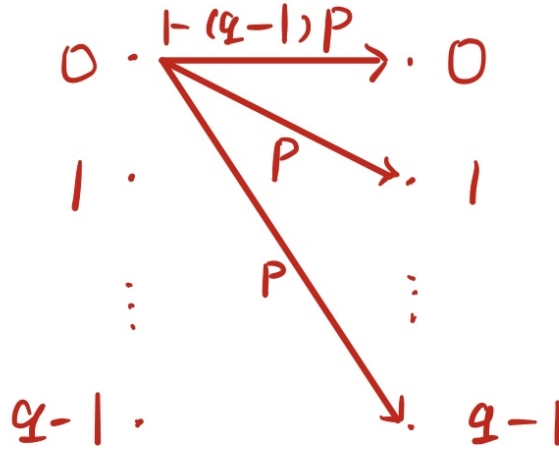


Figure 1:  $q$ -ary Symmetric Channel

Each symbol  $a \in A$  is:

- Received as  $a$  with probability  $1 - (q-1)p$
- Received as  $b \neq a$  with equal probability  $p$

This defines the following symmetric stochastic matrix:

$$\mathcal{P} = \begin{pmatrix} 1 - (q-1)p & p & \cdots & p \\ p & 1 - (q-1)p & \cdots & p \\ \vdots & \vdots & \ddots & \vdots \\ p & p & \cdots & 1 - (q-1)p \end{pmatrix}$$

## Example: Binary Symmetric Channel (BSC)

Let  $q = 2$ . The *binary symmetric channel* (BSC) has binary channel alphabet  $\{0, 1\}$ , and

$$\Pr(1 \text{ received} \mid 0 \text{ sent}) = \Pr(0 \text{ received} \mid 1 \text{ sent}) = p.$$

The channel transition diagram is:

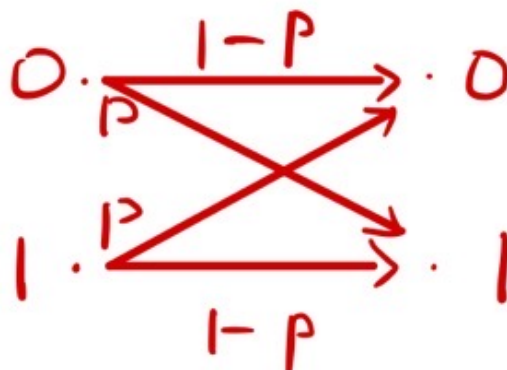


Figure 2: Binary Symmetric Channel

Alternatively, the BSC transition matrix is:

$$\mathcal{P} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

Here,  $p$  is called the *crossover probability* (also known as the *bit error probability*).

## Example

Let  $C = \{000, 111\}$ , and assume the channel is a binary symmetric channel (BSC) with crossover probability  $p = 0.05$ . Assume a uniform distribution on the code  $C$ , i.e.,

$$\Pr(000) = \Pr(111) = \frac{1}{2}.$$

Suppose the received word is 110. Which codeword is more likely to be the one that was sent? This is the procedure of **decoding**.

We compare the posterior probabilities:

$$\Pr(000 \text{ sent} \mid 110 \text{ received}) \quad \text{vs.} \quad \Pr(111 \text{ sent} \mid 110 \text{ received}).$$

Using Bayes' rule:

$$\Pr(000 \mid 110) = \frac{\Pr(000) \cdot \Pr(110 \mid 000)}{\Pr(110)}, \quad \Pr(111 \mid 110) = \frac{\Pr(111) \cdot \Pr(110 \mid 111)}{\Pr(110)}.$$

$\Pr(000) = \Pr(111)$ , we compare the likelihoods:

$$\Pr(110 \mid 000) = (1 - p) \cdot p^2, \quad \Pr(110 \mid 111) = (1 - p)^2 \cdot p.$$

Substituting  $p = 0.05$ , we get:

$$(1 - p)^2 \cdot p > (1 - p) \cdot p^2.$$

Therefore,

$$\Pr(111 \mid 110) > \Pr(000 \mid 110),$$

so the most likely codeword sent is **111**. This is called the **Maximum Likelihood Decoding** method which we will see in detail next lecture.

### 3 Decoding Rules

Decoding is the process of determining which codeword was sent based on the received word.

#### Definition 3.1 (Decoder)

Let  $C$  be an  $(n, M)$  code over an alphabet  $A$ , and let  $W = (A, B, \mathcal{P})$  be a channel. A *decoder* for  $C$  with respect to  $W$  is a function:

$$\mathcal{D} : B^n \rightarrow C.$$

The *average decoding error probability* of  $\mathcal{D}$  is defined as:

$$P_e = \sum_{c \in C} P_e(c) \cdot \Pr(c),$$

where

$$P_e(c) := \sum_{y: \mathcal{D}(y) \neq c} \Pr(y \text{ received} \mid c \text{ sent}).$$

#### Remarks

$\Pr(c)$  is the probability distribution of  $c$  and  $P_e(c)$  is the *decoding error probability*. The *maximum decoding error probability* is defined as:

$$P_{e, \max} := \max_{c \in C} P_e(c).$$

The goal of decoding is to have small  $P_e$  or  $P_{e, \max}$ .