Coding theory

Elena Berardini, Anestis Tzogias

April 13, 2026

Contents

1 Introduction

Information theory deals with the **efficient, reliable**, and **secure** transmission of information. Founded by Shannon in 1948 [4], information theory constructs and studies mathematical models, primarily based on probability, to provide

intuitive insights into the behaviour of communication channels. It has since become indispensable in the design of any communication system.

Its main domains are:

- information encoding
    1. data compression (**efficiency**);
    2. error correction (**reliability**);
- cryptography (**security**).

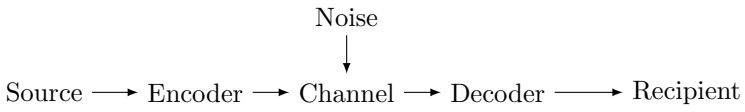


Figure 1: Representation of message transmission.
Encoder = operations performed on the information emitted by the source, prior to transmission.
Noise = the channel is typically disturbed, which can lead to errors (loss of reliability) or intrusions (loss of security).
Decoder = restores the information provided by the source in an acceptable manner.

The encoder has a dual purpose: to represent the information emitted by the source concisely (compression), and to protect the information from channel noise (error correction). We will address these two tasks separately, referring to them as source encoding and channel encoding. (Figure 2).

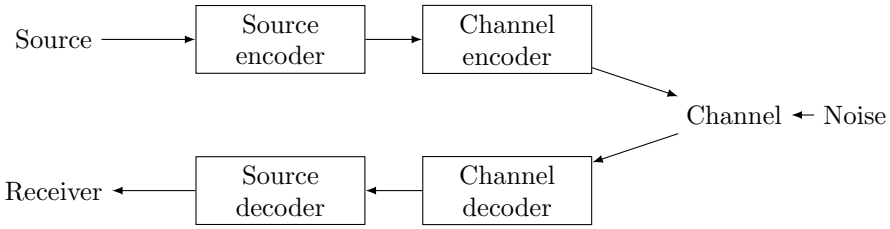


Figure 2: Source coding and channel coding.

Example 1.1. Minimal example of source and channel coding with $\{yes, no\}$, $\{1, 0\}$ and $\{111, 000\}$.

2 Discrete Probability Theory and Entropy

Definition 2.1. A discrete probability space (Ω, P) consists of a finite (or countable) set Ω equipped with a probability measure, i.e., an additive function

$$P : \Omega \to [0, 1],$$

satisfying $\sum_{\omega\in\Omega} P(\omega) = 1$. The probability P is said to be *uniform* if $P(\omega) = 1/|\Omega|$ for all $\omega \in \Omega$.

Definition 2.2. A subset A of Ω is called an event. The singleton $\{\omega\}$ is called an elementary event. We extend the probability measure on Ω to the subsets of Ω, $2^\Omega = \{A \subseteq \Omega\}$, by setting $P(\{\omega\}) = P(\omega)$. The probability of A is thus $P(A) = \sum_{\omega\in A} P(\omega)$. The complement of A, $\Omega \setminus A$, is denoted A^c. Its probability is $P(A^c) = 1 - P(A)$.

Example 2.1. If the probability measure on Ω is uniform, then $P(A) = |A|/|\Omega|$.

Exercise 2.1. If $A, B \subset \Omega$ and $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

Remark 2.1. In general probability theory (Ω not necessarily finite), we define a measure μ on Ω and the probability of $A \subset \Omega$ is $P(A) = \int_A d\mu$.

2.1 Conditional Probabilities

Definition 2.3. Let A, B be two events. We define the probability of A given B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

If $P(A|B) = P(A)$, or $P(B|A) = P(B)$, or $P(A \cap B) = P(A)P(B)$, we say that A and B are *independent*.

Lemma 2.1. *The function $P(\,\cdot\,|B) : \Omega \to [0, 1]$ is a probability function on Ω. Moreover, $(B, P(\,\cdot\,|B))$ is a probability space.*

Exercise 2.2. Prove Lemma 2.1.

Theorem 2.1 (Bayes' Formula)**.** *We have*

$$P(A) = P(A|B)P(B) + P(A|B^c)(1 - P(B)).$$

Proof. We have $A = (A \cap B) \cup (A \cap B^c)$ with $(A \cap B) \cap (A \cap B^c) = \emptyset$. Since $P(A|B)P(B) = P(A \cap B)$ and $P(A|B^c)(1 - P(B)) = P(A \cap B^c)$, the result follows. □

Exercise 2.3. A student answers a multiple-choice questionnaire. There are m possible answers for each question. We assume that when the student does not know the answer, they randomly select an answer uniformly. If we observe a proportion x of correct answers, we want to estimate the proportion y of questions the student actually knows the answer to. Let C be the event "the student knows the answer" and R the event "the student answers correctly". We want to estimate $y = P(C)$ as a function of $x = P(R)$ and m.

Solution. Using Bayes' formula:

$$P(R) = P(R|C)P(C) + P(R|C^c)(1 - P(C)) = 1 \cdot y + \frac{1}{m}(1 - y).$$

Therefore, $y = (x - 1/m)/(1 - 1/m)$. ■

2.2 Random Variables

The idea: a random variable is a function defined on Ω. Its value is determined by the outcome of the experiment, so we can associate probabilities with its values via the probability function P defined on Ω.

Definition 2.4. Let (Ω, P) be a probability space. A random variable is a function

$$X : \Omega \to \mathbb{R}^n.$$

If $n = 1$, we speak of a *real* random variable.

We denote $P(X = x) = P(X^{-1}(x)) = P(\{\omega \in \Omega \mid X(\omega) = x\})$.

For $S \subseteq \mathbb{R}^n$, we set $P(X \in S) = P(X^{-1}(S)) = P(\{\omega \in \Omega \mid X(\omega) \in S\})$.

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be the image of X. We call the law of X the data of $P(x)$ for all $x \in \mathcal{X}$. We say that X follows the uniform law if $P(X = x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$.

A random variable is said to be Bernoulli if $\mathcal{X} = \{0, 1\}$.

Definition 2.5. Two variables $X, Y : \Omega \to \mathcal{X}$ are *independent* if for all $x, y \in \mathcal{X}$

$$P(X = x, Y = y) = P(X = x)P(Y = y),$$

where $\{X = x, Y = y\}$ is the intersection of the events $\{X = x\}$ and $\{Y = y\}$. This definition generalizes to several variables $X_1, \ldots, X_n$.

Definition 2.6. Let (Ω, P) be a probability space. The *expectation* of a random variable X is

$$\mathbf{E}(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{x \in \mathcal{X}} xP(X = x).$$

This is a "weighted average" or sum weighted by the probability law.

Example 2.2. Let X be a Bernoulli variable. Then $\mathbf{E}[X] = P(X = 1)$ (it suffices to apply the definition).

Exercise 2.4. For an event A, we define its indicator function $\mathbf{1}_A$ by

$$\mathbf{1}_A = \begin{cases} 1 \text{ if } A \text{ occurs} \\ 0 \text{ if } A^c \text{ occurs.} \end{cases}$$

Calculate $\mathbf{E}[\mathbf{1}_A]$.

Solution. $\mathbf{E}[\mathbf{1}_A] = 1 \cdot P(\mathbf{1}_A = 1) + 0 \cdot P(\mathbf{1}_A = 0) = P(\mathbf{1}_A = 1) = P(A)$. ■

Remark 2.2. In the non-discrete setting, the expectation of the variable X is $\int_\Omega X(\omega)d\mu(\omega)$.

Theorem 2.2. *Let X, Y be two random variables on Ω. Then*

- $\boldsymbol{E}[X \pm Y] = \boldsymbol{E}[X] \pm \boldsymbol{E}[Y]$,
- $\boldsymbol{E}[\lambda X] = \lambda \boldsymbol{E}[X]$ *for all* $\lambda \in \mathbb{R}$.

Proof. Seen in class. □

Definition 2.7. Let X, Y be two random variables.

The *covariance* of X, Y is

$$\operatorname{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

The *variance* of X is defined by

$$\operatorname{var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2].$$

The *standard deviation* is

$$\sigma(X) = \sqrt{\operatorname{var}(X)}.$$

Lemma 2.2. *We have* $\operatorname{var}(X) = \boldsymbol{E}[X^2] - \boldsymbol{E}[X]^2 = \operatorname{cov}(X, X)$.

Proof.

$$\begin{aligned}
\operatorname{var}(X) &= \mathbf{E}[(X - \mathbf{E}[X]))^2] \\
&= \sum_{x \in \mathcal{X}} (x - \mathbf{E}[X])^2 p(x) \\
&= \sum_{x \in \mathcal{X}} (x^2 + \mathbf{E}[X]^2 - 2x\mathbf{E}[X]) p(x) \\
&= \sum_{x \in \mathcal{X}} x^2 p(x) + \sum_{x \in \mathcal{X}} \mathbf{E}[X]^2 p(x) - \sum_{x \in \mathcal{X}} 2x\mathbf{E}[X] p(x) \\
&= \mathbf{E}[X^2] + \mathbf{E}[X]^2 \sum_{x \in \mathcal{X}} p(x) - 2\mathbf{E}[X] \sum_{x \in \mathcal{X}} x p(x) \\
&= \mathbf{E}[X^2] + \mathbf{E}[X]^2 - 2\mathbf{E}[X]^2 \\
&= \operatorname{cov}(X, X).
\end{aligned}$$

□

Theorem 2.3 (Chebyshev's Inequality)**.** *Let $X : \Omega \to \mathcal{X}$ be a random variable, $\epsilon \in \mathbb{R}$. Then,*

$$P(|X - \boldsymbol{E}(X)| \geq \epsilon) \leq \frac{\sigma(X)^2}{\epsilon^2}.$$

Proof.

$$\begin{aligned} P(|X - \mathbf{E}(X)| \geq \epsilon) &= \mathbf{E}\left[\mathbf{1}_{|X-\mathbf{E}(X)|\geq\epsilon}\right] \\ &\leq \mathbf{E}\left[\mathbf{1}_{|X-\mathbf{E}(X)|\geq\epsilon}\frac{|X-\mathbf{E}(X)|^2}{\epsilon^2}\right] \\ &\leq \mathbf{E}\left[\frac{|X-\mathbf{E}(X)|^2}{\epsilon^2}\right] = \frac{\mathrm{var}(X)}{\epsilon^2}. \end{aligned}$$

□

Chebyshev's inequality shows that it is unlikely for a random variable to deviate from its expectation by much more than $\sigma(X)$, hence the name *standard deviation* for the latter.

Lemma 2.3. *Let $X, Y : \Omega \to \mathcal{X}$ be two independent variables. Then*

$$\boldsymbol{E}[XY] = \boldsymbol{E}[X]\boldsymbol{E}[Y].$$

Proof.

$$\begin{aligned} \mathbf{E}[X]\mathbf{E}[Y] &= \sum_{x\in\mathcal{X}} xP(X=x)\sum_{x\in\mathcal{X}} yP(Y=y) \\ &= \sum_{x,y\in\mathcal{X}} xyP(X=x)P(Y=y) \\ &= \sum_{x,y\in\mathcal{X}} xyP(X=x,Y=y) \\ &= \sum_{z} z\sum_{xy=z} P(X=x,Y=y) \\ &= \sum_{z} zP(XY=z) = \mathbf{E}[XY], \end{aligned}$$

since $\{XY = z\} = \bigcup_{x,y,z=xy}\{X = x, Y = y\}$, and this union is disjoint. □

Corollary 2.1. *Let $X_1, \ldots, X_n$ be pairwise independent variables. Then*

$$\mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n).$$

Proof. Exercise. It is a matter of using the previous lemma and the additivity of expectation. □

2.3 Entropy

Recalls on the Logarithm

In the following, we will use the following properties of the logarithm:

- For any real $z > 0$, we have $\log_2 z \leq z - 1$,

- $\log_2(a \cdot b) = \log_2(a) + \log_2(b)$,
- $\log_2 \frac{a}{b} = \log_2 a - \log_2 b$.

In what follows, X denotes a random variable with law p. Let $\mathcal{X} = \{x_1, \ldots, x_m\}$ be its image, and $p_i = P(X = x_i)$, then $p = (p_1, \ldots, p_m)$.

Definition 2.8. The *entropy* of X is defined by

$$H(X) = \sum_{x \in \mathcal{X}} P(X = x) \log_2 \frac{1}{P(X = x)} = \sum_{i=1}^{m} p_i \log_2 \frac{1}{p_i}.$$

The entropy depends only on p, the law of X, and is thus also denoted $H(p)$. It can also be seen as the expectation of the variable $F = -\log_2(P(X = x))$.

Remark 2.3. If $p_i = 0$, we define $p_i \log_2 \frac{1}{p_i} = 0$. This is consistent with the limit $\lim_{t \to 0} t \log_2 \frac{1}{t} = 0$.

Entropy, introduced by Shannon, measures the amount of information in a probability space. We call a bit of information the unit of entropy.

Lemma 2.4. *Let X have the uniform law, and let $m = |\mathcal{X}|$. Then, $H(X) = \log_2 m$.*

Proof. Let $|\mathcal{X}| = m$, then the uniform law is $p_i = 1/m$ for each i. By the definition of entropy, we then obtain

$$H(X) = \sum_{i=1}^{m} p_i \log_2 \frac{1}{p_i} = \sum_{i=1}^{m} \frac{1}{m} \log_2 m = \log_2 m.$$

□

Definition 2.9. For X, Y two random variables, we define the entropy of the pair (X, Y) by

$$H(X, Y) = \sum_{x,y} P(X = x, Y = y) \log_2 \frac{1}{P(X = x, Y = y)},$$

where $\{X = x, Y = y\}$ is the intersection of the events $\{X = x\}$ and $\{Y = y\}$.

Definition 2.10. We define the *Kullback divergence* between two laws p and q by

$$D(p \parallel q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}.$$

Remark 2.4. Often we speak of the *Kullback distance*. However, the previous definition does not define a distance. Why? A distance d must satisfy the following properties by definition:

1. $d(x, y) \geq 0$ with equality if and only if $x = y$;

2. $d(x,y) = d(y,x)$;

3. $d(x,y) \geq d(x,z) + d(z,y)$.

First, $D(p \parallel q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$ and $D(q \parallel p) = \sum_x q(x) \log_2 \frac{q(x)}{p(x)}$ are different. Moreover, the Kullback divergence does not satisfy the triangle inequality either. However, it does satisfy the first property of a distance, as shown in the following lemma.

Lemma 2.5. *We have $D(p \parallel q) \geq 0$, with equality if, and only if, $p = q$.*

Corollary 2.2. *The maximum entropy is achieved with the uniform law.*

Proof. We have seen that if p is the uniform law, then $H(p) = \log_2 m$. For any other law q, we have

$$\begin{aligned}
\log_2 m - H(q) &= \log_2 m - \sum_{i=1}^{m} q_i \log_2 \frac{1}{q_i} \\
&= \sum_{i=1}^{m} q_i \log_2 m + \sum_{i=1}^{m} q_i \log_2 q_i \\
&= \sum_{i=1}^{m} q_i \log_2 \frac{q_i}{1/m} \\
&= D(q \parallel 1/m) \geq 0,
\end{aligned}$$

where the last inequality relies on Lemma 2.5. □

3 Channels and Error Correcting Codes

To define a transmission channel, one must describe the set of possible inputs and outputs of the channel, as well as the noise that will disrupt the transmission. In the case of a discrete channel, the simplest example, the sets of inputs and outputs are finite non-empty sets $\mathcal{X}$ and $\mathcal{Y}$, viewed as the images of two random variables X and Y, and the noise is modelled by a conditional probability law of Y given X. In the following, we will often forget the random variables X and Y and work with their image sets $\mathcal{X}$ and $\mathcal{Y}$.

Definition 3.1. A *channel (of communication)* is given by $(\mathcal{X}, \mathcal{Y}, P)$, where $\mathcal{X}$ and $\mathcal{Y}$ are non-empty sets called the input and output alphabets, respectively, and $P : \mathcal{Y} \times \mathcal{X} \to [0,1], (y,x) \mapsto P(y|x)$ is a conditional probability function.

Remark 3.1. Each channel can be associated with a collection of probability spaces as follows. Let $x \in \mathcal{X}$ be an element of the input alphabet. We can then define $P_x : 2^{\mathcal{Y}} \to \mathbb{R}$ by $P_x(\mathcal{Y}') = \sum_{y \in \mathcal{Y}'} P(Y = y | X = x)$, for $\mathcal{Y}' \in 2^{\mathcal{Y}}$. Then $(\mathcal{Y}, P_x)$ is a probability space.

Definition 3.2. Let $\mathcal{X}, \mathcal{Y} = \{0,1\} = \mathbb{F}_2$ and let $0 \leq p \leq 1$ be a real number. The *binary symmetric channel* $(\mathcal{X}, \mathcal{Y}, P)$ with probability p is defined by

$$P(y|x) = \begin{cases} 1-p \text{ if } x = y, \\ p \text{ if } x \neq y. \end{cases}$$

This can be generalised to q-ary alphabet $\mathbb{F}_q$.

Definition 3.3. Let $\mathcal{X}, \mathcal{Y} = \mathbb{F}_q$ and let $0 \leq p \leq 1$ be a real number. The *q-ary symmetric channel* $(\mathcal{X}, \mathcal{Y}, P)$ with probability p is defined by

$$P(y|x) = \begin{cases} 1-p \text{ if } x = y, \\ p/(q-1) \text{ if } x \neq y. \end{cases}$$

If $p \geq \frac{q-1}{q}$, then reliable communication is impossible (Shannon Theorem). Hence, we can assume from now on that $p < \frac{q-1}{q}$.

Definition 3.4. Let $\mathcal{X} = \{0,1\}$, let $\mathcal{Y} = \mathcal{X} \cup \{?\}$, and let $0 \leq p \leq 1$ be a real number. The *erasure channel* $(\mathcal{X}, \mathcal{Y}, P)$ with erasure probability p is defined by

$$P(y|x) = \begin{cases} 1-p \text{ if } y = x, \\ p \text{ if } y =?, \\ 0 \text{ otherwise.} \end{cases}$$

The symbol "?" is called the erasure symbol.

Definition 3.5. Let $(\mathcal{X}, \mathcal{Y}, P)$ be a channel and n a positive integer. We define the *discrete memoryless channel* by $(\mathcal{X}^n, \mathcal{Y}^n, P^n)$, where P^n satisfies the following two hypotheses:

- The channel is *memoryless*, i.e.,

$$P^n(Y_n = y_n | X^n = x^n, Y^{n-1} = y^{n-1}) = P^n(Y_n = y_n | X_n = x_n),$$

- The channel is used *without feedback*, i.e., the n-th symbol X_n does not depend on previous outputs:

$$P^n(X_n = x_n | X^{n-1} = x^{n-1}, Y^{n-1} = y^{n-1}) = P^n(X_n = x_n | X^{n-1} = x^{n-1}).$$

Memoryless indicates that what happens in one use of the channel does not influence what happens in another use. Indeed, the two hypotheses imply the decomposition of the following lemma.

Lemma 3.1. *Let P^n be the probability function of the memoryless discrete channel. Then for all $y \in \mathcal{Y}^n$ and all $x \in \mathcal{X}^n$ we have*

$$P^n(y|x) = \prod_{i=1}^{n} P(y_i|x_i).$$

Equivalently, if X and Y are the two random variable of image set $\mathcal{X}$ and $\mathcal{Y}$, then

$$P^n(Y^n = y^n | X^n = x^n) = \prod_{i=1}^{n} P(Y_i = y_i | X_i = x_i).$$

Proof.

$$\begin{aligned} P^n(Y^n = y^n | X^n = x^n) &= P^n(Y_n = y_n | X^n = x^n, Y^{n-1} = y^{n-1}) P^n(Y^{n-1} = y^{n-1} | X^n = x^n) \\ &= P^n(Y_n = y_n | X_n = x_n) P^n(Y^{n-1} = y^{n-1} | X^n = x^n), \end{aligned} \tag{1}$$

using the fact that the channel is memoryless. Moreover,

$$\begin{aligned} P^n(Y^{n-1} = y^{n-1} | X^n = x^n) &= \frac{P^n(Y^{n-1} = y^{n-1}, X^n = x^n)}{P^n(X^n = x^n)} \\ &= \frac{P^n(X_n = x_n | Y^{n-1} = y^{n-1}, X^{n-1} = x^{n-1}) P^n(Y^{n-1} = y^{n-1}, X^{n-1} = x^{n-1})}{P^n(X^n = x^n)} \\ &= \frac{P^n(X_n = x_n | X^{n-1} = x^{n-1}) P^n(Y^{n-1} = y^{n-1}, X^{n-1} = x^{n-1})}{P^n(X^n = x^n)}, \end{aligned}$$

by the *no feedback* hypothesis. We thus obtain

$$\begin{aligned} P^n(Y^{n-1} = y^{n-1} | X^n = x^n) &= \frac{P^n(X^n = x^n) P^n(Y^{n-1} = y^{n-1}, X^{n-1} = x^{n-1})}{P^n(X^n = x^n) P^n(X^{n-1} = x^{n-1})} \\ &= \frac{P^n(Y^{n-1} = y^{n-1}, X^{n-1} = x^{n-1})}{P^n(X^{n-1} = x^{n-1})} \\ &= P^n(Y^{n-1} = y^{n-1} | X^{n-1} = x^{n-1}). \end{aligned}$$

Using this last equality in (1), we deduce

$$P^n(Y^n = y^n | X^n = x^n) = P^n(Y_n = y_n | X_n = x_n) P^n(Y^{n-1} = y^{n-1} | X^{n-1} = x^{n-1}).$$

We conclude by induction on n. □

3.1 Coding and decoding

For the sake of these notes, we will restrict to the case in which the input and output alphabet are $\mathbb{F}_q$.

Definition 3.6. A subset $C \subseteq \mathbb{F}_q^n$ is called an *error-correcting code* over the alphabet $\mathbb{F}_q$. The elements of C are called codewords. The *length* n of C is the length of any codeword in it.

The support of a vector $x \in \mathbb{F}_q^n$, denoted $\mathrm{supp}(x)$, is the set of indices on which it has non-zero coordinates, *i.e.,*

$$\mathrm{supp}(x) = \{i \in \{1, \ldots, n\} \,|\, x_i \neq 0\}.$$

For two words x and y, we write $x \subset y$ if $\mathrm{supp}(x) \subset \mathrm{supp}(y)$. In this case, we say that "y covers x".

Definition 3.7. The Hamming distance between $x, y \in C \subseteq \mathbb{F}_q^n$ is defined by

$$d_H(x, y) = \#\{i \in \{1, \ldots, n\} \mid x_i \neq y_i\}.$$

The Hamming weight of $x \in C$ is the number of its non-zero coordinates, *i.e.,*

$$\mathrm{wt}(x) = |\mathrm{supp}(x)|.$$

Definition 3.8. The *minimum distance* $d(C)$ of a code $C \subseteq \mathbb{F}_q^n$ is the smallest non-zero distance between two words of the code C:

$$d(C) = \min\{d_H(x, y) \mid x, y \in C, x \neq y\}.$$

By convention, if $|C| = 1$, then $d(C) = n + 1$.

Remark 3.2.

- The Hamming distance $d_H(\cdot, \cdot)$ is a distance. The verification is left as an exercise.
- If $D \subseteq C \subseteq \mathbb{F}_q^n$, then $d(D) \geq d(C)$.

We say that C is an (n, M, d)-code if C has length n, minimum distance d, and $|C| = M$.

Definition 3.9. Let C be an (n, M, d)-code. A *decoder* for C is a mapping

$$D : \mathbb{F}_q^n \to \mathbb{F}_q^n \cup \{\perp\}.$$

Here $\perp$ represents the failure message which is sent when the decoder is not able to identify the originally sent message. The decoder is said to be *complete* if it always returns a word from C, *i.e.,* if $\mathrm{Im}D \subseteq C$.

In the binary symmetric channel, an error corresponds to a bit-flip: $0 \to 1$, $1 \to 0$. Let X and Y be random variables taking values in $\mathcal{X}^n$ and $\mathcal{Y}^n$, respectively, which denote the transmitted and received messages. When a message $y \in \mathcal{Y}^n$ is received, the decoding problem consists of finding the codeword $c \in C \subseteq \mathcal{X}^n$ that is most likely, *i.e.,* the one that maximises

$$P(c \mid y).$$

Another example of a decoder is the nearest codeword decoder, which relies on the Hamming distance:

$$D : \mathbb{F}_q^n \to C$$

such that for all $y \in \mathbb{F}_q^n$,

$$D(y) = c \in \{c \in C \mid \forall c' \in C, d_H(c, y) \leq d_H(c', y)\}.$$

The two aforementioned decoders are, in fact, equivalent, as shown in the following proposition.

Proposition 3.1. *Let y be the received message. Assume that the transmitted codeword follows a uniform distribution over the code $\mathcal{C}$. Then, the most likely codeword $c \in \mathcal{C}$ is the codeword closest to the received word y in terms of Hamming distance.*

Proof. We have

$$P(X = c \,|\, Y = y) = \frac{P(X = c, Y = y)}{P(Y = y)} = \frac{P(X = c)P(Y = y \,|\, X = c)}{P(Y = y)}.$$

Under the assumptions of the proposition, $\frac{P(X=c)}{P(Y=y)}$ is constant. Therefore, maximizing $P(X = c \,|\, Y = y)$ is equivalent to maximizing $P(Y = y \,|\, X = c)$. We have

$$P(Y = y \,|\, X = c) = p^{d(y,c)}(1-p)^{n-d(y,c)}.$$

This quantity is maximised when $d(y, c)$ is minimised. □

Remark 3.3. It is possible that there are multiple codewords at the minimal distance from y, and thus multiple equally likely codewords.

Proposition 3.1 shows why we consider the Hamming distance in coding theory, and particularly highlights its importance in correcting errors.

4 Linear Codes

The most studied model of error-correcting codes is that of *linear codes*.

Notation 4.1. We will denote

- $\mathbf{0}$ the vector $(0, \ldots, 0) \in \mathbb{F}_q^n$;
- $\boldsymbol{I}_n$ the $n \times n$ identity matrix.

Definition 4.1. A *linear code* $C \subseteq \mathbb{F}_q^n$ is a vector subspace of $\mathbb{F}_q^n$. Its *length* is $n = \dim_{\mathbb{F}_q} \mathbb{F}_q^n$. Its *dimension* is $k = \dim_{\mathbb{F}_q} C$.

When $q = 2$, we speak of a binary code.

Definition 4.2. The *generator matrix* G of a linear code $C \subseteq \mathbb{F}_q^n$ of dimension k is a $k \times n$ matrix with coefficients in $\mathbb{F}_q$ whose rows form a basis of the vector space C.

If G is a generator matrix of the code C, an encoding function of the set of messages $\mathcal{M}$ is given by

$$\begin{aligned} c : \mathcal{M} &\to \mathbb{F}_q^n \\ x &\mapsto xG. \end{aligned}$$

Definition 4.3. A generator matrix is said to be in *systematic form* if it is written as

$$G = [\boldsymbol{I}_k | A]$$

with A a $k \times (n-k)$ matrix.

In this case, the encoding function is of the form

$$(x_1, \ldots, x_k) \mapsto (x_1, \ldots, x_k, x_{k+1}, \ldots, x_n).$$

The first k symbols are called *information bits* and the other $n-k$ symbols are called *parity bits*.

4.1 Hamming Distance and Error Correction and Detection Capacity

We recall some definitions from the previous section and interpret them in the case of linear codes.

Definition 4.4. The Hamming distance between $x, y \in C \subseteq \mathbb{F}_q^n$ is defined by

$$d_H(x, y) = \#\{i \in \{1, \ldots, n\} \,|\, x_i \neq y_i\}.$$

The Hamming weight of $x \in C$ is the number of its non-zero coordinates, i.e.,

$$\mathrm{wt}(x) = |\mathrm{supp}(x)|.$$

If the code C is linear, then in particular we have $\mathrm{wt}(x-y) = d(x, y)$.

Definition 4.5. The *minimum distance* $d(C)$ of a code C is the smallest non-zero distance between two words of the code C. When the code $C \subseteq \mathbb{F}_q^n$ is linear, we have

$$d(C) = \min\{\mathrm{wt}(x) \,|\, x \in C, x \neq \mathbf{0}\}.$$

If C is linear, we speak of an $[n, k, d]_q$-code to indicate that C has length n, its *dimension* as a vector space is k, and its minimum distance is d. An $[n, k, d]_q$-linear code C contains q^k words. Indeed, let $\{c_1, \ldots, c_k\}$ be a basis of C. Then all elements of the code are linear combinations of the c_i, and there are q^k of these.

Example 4.1.

1. **Trivial codes.** $\mathbb{F}_q^n$ and all sets $C \subseteq \mathbb{F}_q^n$ of cardinality 1 are trivial codes of length n. They have dimensions n and 1, respectively, and minimum distances 1 and $n+1$, respectively. The only trivial linear codes are $\mathbb{F}_q^n$ and $\{0\}$.

2. **Repetition code.**

   $$C = \{(\alpha, \ldots, \alpha) \,|\, \alpha \in \mathbb{F}_q\} \subseteq \mathbb{F}_q^n.$$

   It has dimension 1 (what is a basis? And its generator matrix?), and minimum distance equal to n.

3. **Parity code.** For $n \geq 2$, the set

$$C = \left\{ (x_1, \ldots, x_n) \in \mathbb{F}_q^n \,|\, x_n = -\sum_{i=1}^{n-1} x_i \right\} \subseteq \mathbb{F}_q^n,$$

defines a code of length n and dimension $n-1$. What is its minimum distance?

We will now explore the relationship between minimum distance and the ability to detect and correct errors.

Proposition 4.1. *A code with minimum distance d can detect up to $d-1$ errors.*

Proof. Let $c \in C$ be a codeword sent over a channel. Suppose it undergoes $t < d$ errors and is transformed into $x \in \mathbb{F}_q^n$. Then $d(c, x) = t < d$. Since the minimum distance is d, we detect that $x \notin C$. □

Proposition 4.2. *A code with minimum distance d can correct up to $\lfloor \frac{d-1}{2} \rfloor$ errors.*

Proof. Let C be a code with minimum distance d. Suppose the codeword $c \in C$ is transmitted and undergoes t errors, with $t < d/2$. Let $x \in \mathbb{F}_q^n$ be the received word. We then have $d(c, x) = t$, and for any other codeword $c' \in C$, $c' \neq c$, we have $d(c', x) > d(c, x)$. Indeed, the triangle inequality gives

$$d \leq d(c, c') \leq d(c, x) + d(x, c'),$$

and thus, since $d(c, x) < d/2$, we have $d(x, c') > d/2$. In conclusion, the closest codeword to x, which is unique if $t < d/2$, is the codeword c that was transmitted. Therefore, we correct x to c. □

We are now interested in the erasure correction capacity of a code.

Definition 4.6. An *erasure vector* is a vector v_E of $\{0,1\}^n$ whose support corresponds to the set E of erased coordinates in a transmitted codeword.

Proposition 4.3. *A linear code C can correct erasures in a set E if and only if the erasure vector v_E does not cover any non-zero codeword of C, i.e., if for all $c \in C, c \neq \mathbf{0}$, we have $c \not\subset v_E$.*

Proof. Let E be an erasure configuration that is not correctable by the code C. Then, there exist at least two codewords $c, c' \in C$, $c \neq c'$ that coincide outside the erased positions. Then, the vector $c - c'$, which is a codeword of C by linearity, is covered by v_E, i.e., $c - c' \subset v_E$. □

Corollary 4.1. *A linear code with minimum distance d can correct up to $d-1$ erasures.*

It follows that we would like to have codes with dimension and minimum distance as large as possible.

4.2 Dual Code

The vector space $\mathbb{F}_q^n$ is equipped with a symmetric bilinear form $\mathbb{F}_q^n \times \mathbb{F}_q^n \to \mathbb{F}_q$, called the dot product, given by

$$(x, y) \mapsto x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

When $x \cdot y = 0$, we say that x and y are orthogonal. The following properties are easily verifiable.

Proposition 4.4. *Let $x, x', y, y' \in \mathbb{F}_q^n$ and $\alpha \in \mathbb{F}_q$. Then*

- $(x + x') \cdot (y + y') = x \cdot y + x \cdot y' + x' \cdot y + x' \cdot y'$,
- $\alpha(x \cdot y) = (\alpha x) \cdot y = x \cdot (\alpha y)$.

Definition 4.7. Let C be an $[n, k]_q$ code. The orthogonal code of C is called the code

$$C^\perp = \{x \in \mathbb{F}_q^n \mid x \cdot c = 0 \, \forall c \in C\}.$$

By definition, it is clear that $C^\perp$ is a linear code.

Definition 4.8. A code C is *self-orthogonal* if $C \subseteq C^\perp$. It is *self-dual* if $C = C^\perp$. The *hull* of C is $C \cap C^\perp$.

Proposition 4.5. *Let $G = [I_k|A]$ be a generator matrix of the $[n, k]$-code C in systematic form (A is a $k \times (n - k)$ matrix). Then, a generator matrix of the dual code $C^\perp$ is the $(n - k) \times n$ matrix H given by*

$$H = [-A^t|I_{n-k}].$$

Proof. We write $a_{i,j}$ for the entries of the matrix A. We verify that the dot product of the i-th row of G with the j-th row of H is $-a_{i,j} + a_{i,j} = 0$. Thus, the code generated by H is indeed included in $C^\perp$. Let $x \in C^\perp$, then we can form

$$y = x - \sum_{j=n-k+1}^{n} x_j H_j$$

where H_j is the j-th row of H, such that

- $y \in C^\perp$,
- for all j such that $n - k + 1 \leq j \leq n$, we have $y_j = 0$.

From the orthogonality of y with all the rows of G, we deduce that $(y_1, \ldots, y_k)$ is orthogonal to the vectors of the canonical basis of $\mathbb{F}_q^k$ and thus that $y_1 = y_2 = \ldots y_k = 0$, i.e., $y = \mathbf{0}$ by the hypothesis on y. The vector x is therefore generated by the matrix H. □

Definition 4.9. The generator matrix H of the code $C^\perp$ is called the *parity-check matrix* or control matrix of the code C.

Corollary 4.2. *Let C be an $[n,k]$-code. Then the dimension of $C^\perp$ is $n-k$.*

Corollary 4.3. *We have $(C^\perp)^\perp = C$.*

Proof. The inclusion $C \subseteq (C^\perp)^\perp$ is easy to show. Then we conclude by a dimension argument using Corollary 4.2. □

Definition 4.10. Let $H = (h_1|\dots|h_n)$ be the parity-check matrix of an $[n,k]$-code C. We define the *syndrome* application associated with the matrix H by

$$\begin{aligned}\sigma : \mathbb{F}_q^n &\to \mathbb{F}_q^{n-k} \\ x &\mapsto \sigma(x) = xH^t = \sum_{i=1}^{n} x_i h_i.\end{aligned}$$

Proposition 4.6. *Let H be the parity-check matrix of a code C and σ the associated syndrome. Then*

$$C = \{x \in \mathbb{F}_q^n \mid \sigma(x) = 0\}.$$

Moreover, the minimum distance of C is the smallest number of columns of H summing to zero, i.e., the smallest number of linearly dependent columns.

4.3 Syndrome Decoding

Let $H = (h_1|\dots|h_n)$ be the parity-check matrix of an $[n,k]$-code C. We have defined (Definition 4.10) the syndrome application associated with the matrix H by

$$\begin{aligned}\sigma : \mathbb{F}_q^n &\to \mathbb{F}_q^{n-k} \\ x &\mapsto \sigma(x) = xH^t = \sum_{i=1}^{n} x_i h_i.\end{aligned}$$

We will now use it to decode in the code C.

Let $c \in C$ be a transmitted codeword and let $y = c + e$ be the received word. We want to find the closest codeword in terms of Hamming distance. To do this, we look for a vector $e \in \mathbb{F}_q^n$ of minimal Hamming weight such that $y - e \in C$, then we decode y to $y - e$.

We compute $s = \sigma(y)$. Then we look for the smallest set $I \subset \{1,\dots,n\}$ such that there exist $\lambda_i \in \mathbb{F}_q^\times$, $i \in I$, and

$$\sum_{i \in I} \lambda_i h_i = s.$$

Since $\sigma(c) = 0$, we have $\sigma(y) = \sigma(e)$. Therefore, the closest codeword to y, which is unique if $|I| < d/2$, is the word c that was transmitted. Thus, we correct y to c.

In practice, we can prepare a list S of syndromes and a list E of minimal weight vectors such that for each $s \in S$, there is an $e \in E$ such that $\sigma(e) = s$. Then, upon receiving y, we compute $s = \sigma(y)$, find $s \in S$, find the corresponding vector $e \in E$, and decode y to $y - e$.

Exercise 4.1. Let $C \subset \mathbb{F}_2^n$ be the binary code with generator matrix

$$G = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

1. Write the generator matrix of C in systematic form and provide the parity-check matrix of the code C.
2. Determine the number of codewords in C.
3. What are the parameters of C? And of the dual code of C?
4. The word (111001) is a codeword $\boldsymbol{c'}$ of C that has undergone an error. Compute its syndrome. Is it possible to recover $\boldsymbol{c'}$? Justify and, if affirmative, determine $\boldsymbol{c'}$.
5. The word (10?000) is a codeword $\boldsymbol{c}$ of C that has undergone an erasure. Is it possible to recover $\boldsymbol{c}$? Justify and, if affirmative, determine $\boldsymbol{c}$.

4.4 New Codes from Old Ones

For a subset $I \subseteq \{1, \ldots, n\}$, we denote by

$$\pi_I : \mathbb{F}_q^n \to \mathbb{F}_q^{|I|}$$

the projection onto the coordinates indexed by I.

Definition 4.11. Let $C \subseteq \mathbb{F}_q^n$ be a code and $I \subseteq \{1, \ldots, n\}$. The *punctured code* of C at I is $\pi_{I^c}(C)$.

Let $C(I) \coloneqq \{x \in C \mid \operatorname{supp}(x) \subseteq I\}$ be the set of codewords of C whose support is included in I.

Definition 4.12. Let $C \subseteq \mathbb{F}_q^n$ be a code and $I \subseteq \{1, \ldots, n\}$. The *shortened code* of C at I is $\pi_{I^c}(C(I^c))$.

Proposition 4.7. *Let $C \subseteq \mathbb{F}_q^n$ be a linear code, and let $I \subseteq \{1, \ldots, n\}$. Then*

1. *$\pi_I(C)$ and $\pi_I(C(I))$ are linear codes.*
2. *$d(C(I)) = d(\pi_I(C(I))) \geq d(C)$.*

Proposition 4.8. *Let $C \in \mathbb{F}_q^n$ be a linear code of dimension $k \geq 2$ and minimum distance d. Let $x \in C$ be a codeword of weight w such that $1 \leq w < \min\{n, qd/(q-1)\}$. Let $I = \{1, \ldots, n\} \setminus \operatorname{supp}(x)$. Then the code punctured on $\operatorname{supp}(x)$, $\pi_I(C)$, has dimension $k-1$ and minimum distance at least $d - w + \lceil w/q \rceil$.*

Example 4.2. to-do

Definition 4.13. For $n_1, n_2 \in \mathbb{N}_{>0}$, consider the two codes $C_1 \subseteq \mathbb{F}_q^{n_1}$ and $C_2 \subseteq \mathbb{F}_q^{n_2}$. The *product of codes* C_1 and C_2 is the code

$$C_1 \times C_2 \coloneqq \{(x, y) \mid x \in C_1, y \in C_2\} \subseteq \mathbb{F}_q^{n_1+n_2}.$$

Proposition 4.9. *Suppose $\mid C_1 \mid, \mid C_2 \mid \geq 2$. Then the code $C_1 \times C_2$ is an $[n_1 + n_2, \mid C_1 \mid \cdot \mid C_2 \mid, \min\{d(C_1), d(C_2)\}]$ code.*

Definition 4.14. Let $C_1, C_2 \subseteq \mathbb{F}_q^n$ be two linear codes. The *Plotkin sum* of C_1 and C_2 is the linear code

$$C_1 \oplus_P C_2 \coloneqq \{(x, x+y) \mid x \in C_1, y \in C_2\} \subseteq \mathbb{F}_q^{2n}.$$

Proposition 4.10. *Let $C_1, C_2 \subseteq \mathbb{F}_q^n$ be two non-zero linear codes with minimum distances d_1 and d_2, respectively. Then $C_1 \oplus_P C_2$ is a code of length $2n$ and minimum distance $\min\{2d_1, d_2\}$.*

5 Bounds on the Parameters

In this section, we will study bounds for the parameters of linear codes.

The Gilbert–Varshamov bound shows that there always exists a code with "large enough" cardinality and minimum distance.

Definition 5.1. We define the Hamming ball of radius $0 \leq r \leq n$ centred at $x \in \mathbb{F}_q^n$ by

$$B(x, r) = \{c \in \mathbb{F}_q^n \mid d(c, x) \leq r\} \subseteq \mathbb{F}_q^n.$$

The cardinality of $B(x, r)$ does not depend on the choice of x. Indeed, we have

$$|B(x, r)| = \sum_{i=0}^{r} \binom{n}{i} (q-1)^i.$$

Recall: $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$.

Proposition 5.1 (Gilbert–Varshamov Bound)**.** *There exists a code $C \in \mathbb{F}_q^n$ with minimum distance $d(C) \geq d$ for $0 < d \leq n+1$ integer, and*

$$|C| \geq \frac{q^n}{\sum_{i=0}^{d-1} \binom{n}{i} (q-1)^i} = \frac{q^n}{|B(x, d-1)|}.$$

Proof. We start by noting that if $d = n + 1$, then $|B(x, n)| = q^n$ and thus the statement of the theorem becomes $|C| \geq 1$, for which we can take $C = \{0\}$. So for the rest, we will assume $d \leq n$.

Let C be the code of maximum cardinality among the codes with minimum distance $\geq d$. For every $x \in \mathbb{F}_q^n$, there exists a $c \in C$ such that $d(x, c) \leq d - 1$, otherwise the code $C \cup \{x\}$ would have a larger cardinality than C and minimum distance $\geq d$. Therefore, $\mathbb{F}_q^n$ is contained in the union of Hamming balls of radius $d - 1$ centered at the codewords of C, i.e.,

$$\mathbb{F}_q^n \subseteq \bigcup_{c \in C} B(c, d - 1).$$

In conclusion, we have

$$q^n = |\mathbb{F}_q^n| \leq |\bigcup_{c \in C} B(c, d - 1)| \leq \sum_{c \in C} |B(c, d - 1)| = |C| \sum_{i=0}^{d-1} \binom{n}{i} (q - 1)^i,$$

hence $|C| \geq \frac{q^n}{|B(x, d-1)|}$. □

Proposition 5.2 (Singleton Bound)**.** *For an (n, M, d) code $C \subseteq \mathbb{F}_q^n$, we have $M \leq q^{n+1-d}$. In particular, if C is linear, then*

$$d + k \leq n + 1.$$

Proof. Seen in class. □

Definition 5.2. A code whose parameters reach the Singleton bound is called MDS (Maximum Distance Separable).

Proposition 5.3 (Hamming Bound)**.** *Let $C \subseteq \mathbb{F}_q^n$ be a code of cardinality ≥ 2 and minimum distance $d(C) \geq d$. Let $t = \lfloor (d - 1)/2 \rfloor$. Then*

$$|C| \leq \frac{q^n}{\sum_{i=0}^{t} \binom{n}{i} (q - 1)^i}.$$

Proof. The Hamming balls of radius t centered at the codewords of C are pairwise disjoint. Indeed, if $x \in B(c, t) \cap B(d, t)$, then we would have

$$d(c, x) + d(x, c') \leq 2t < d \leq d(C).$$

Therefore, we have

$$\sum_{c \in C} |B(c, t)| = |\bigcup_{c \in C} B(c, t)| \leq |\mathbb{F}_q^n| = q^n.$$

Since $\sum_{c \in C} |B(c, t)| = |C| \sum_{i=0}^{t} \binom{n}{i} (q - 1)^i$, we conclude. □

Definition 5.3. A code whose parameters reach the Hamming bound is called *perfect*.

Definition 5.4. Let $r \geq 3$ and let H be an $r \times (q^r - 1)/(q-1)$ matrix whose columns are all the non-zero vectors of $\mathbb{F}_q^r$ up to scalar multiplication. The code having H as its parity-check matrix is called the *Hamming code* of redundancy r.

Proposition 5.4. *The Hamming code is perfect.*

Proof. Seen in class. □

5.1 Griesmer and Plotkin Bounds

The Griesmer bound is a bound on the length of a linear code, depending on its dimension and minimum distance.

Theorem 5.1 (Griesmer Bound)**.** *Let $C \subseteq \mathbb{F}_q^n$ be a linear code of dimension $k \geq 1$. Let d be its minimum distance. Then*

$$n \geq \sum_{i=0}^{k-1} \lceil d/q^i \rceil.$$

Proof. By induction on the dimension. If $k = 1$, the result $n \geq d$ is immediate. Let $k \geq 2$. In particular, $d < n$ and $d < qd/(q-1)$. Let $x \in C$ be a word of minimal weight, i.e., $\mathrm{wt}(x) = d$. Let $I = \{1, \ldots, n\} \setminus \mathrm{supp}(x)$. Then, the code $\pi_I(C)$ has dimension $k-1$ and minimum distance $d(\pi_I(C)) \geq d - d + \lceil d/q \rceil = \lceil d/q \rceil$. By applying the induction hypothesis to the code $\pi_I(C)$, we obtain

$$n - d \geq \sum_{i=0}^{k-2} \lceil d(\pi_I(C))/q^i \rceil \geq \sum_{i=0}^{k-2} \lceil d/q^{i+1} \rceil$$

and thus

$$n \geq d + \sum_{i=0}^{k-2} \lceil d/q^{i+1} \rceil = d + \sum_{i=1}^{k-1} \lceil d/q^i \rceil = \sum_{i=0}^{k-1} \lceil d/q^i \rceil.$$

□

Example 5.1. The Golay code $\mathcal{G}_{11} \subseteq \mathbb{F}_3^{11}$ whose parity-check matrix is given by

$$\begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 2 & 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 2 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

reaches the Griesmer bound. Indeed, from the matrix H, we deduce that $\mathcal{G}_{11}$ is an $[11, 6, 5]_2$-linear code. The Griesmer bound then reads:

$$11 \geq \sum_{i=0}^{5} \lceil d/3^i \rceil = 11.$$

Theorem 5.2 (Plotkin Bound)**.** *Let $C \subseteq \mathbb{F}_q^n$ be a code of cardinality $\mid C \mid \geq 2$. Let d be its minimum distance and suppose $qd > (q-1)n$. Then*

$$\mid C \mid \leq \left\lfloor \frac{qd}{qd-(q-1)n} \right\rfloor .$$

Proof. Let

$$\Sigma := \sum_{x \in C} \sum_{y \in C} d(x,y).$$

We have

$$\Sigma \geq d \mid C \mid (\mid C \mid -1).$$

Let $\delta : \mathbb{F}_q \times \mathbb{F}_q \to [0,1]$ be the function defined as follows:

$$\delta(x,y) = \begin{cases} 1 \text{ if } x \neq y, \\ 0 \text{ otherwise.} \end{cases}$$

Then, we can write

$$d(x,y) = \sum_{i=1}^{n} \delta(x_i, y_i).$$

Therefore,

$$\Sigma = \sum_{i=1}^{n} \sum_{x \in C} \sum_{y \in C} \delta(x_i, y_i) = \sum_{i=1}^{n} \sum_{\alpha \in \mathbb{F}_q} \sum_{x \in C, x_i = \alpha} \sum_{y \in C} \delta(\alpha, y_i).$$

For $\alpha \in \mathbb{F}_q$ and $i \in [n]$, let $\epsilon_{\alpha,i} = \mid \{x \in C \mid x_i = \alpha\} \mid$. Then

$$\Sigma = \sum_{i=1}^{n} \sum_{\alpha \in \mathbb{F}_q} \mid \{(x,y) \in C \times C \mid x_i = \alpha, y_i \neq \alpha\} \mid =$$

$$= \sum_{i=1}^{n} \sum_{\alpha \in \mathbb{F}_q} \epsilon_{\alpha,i} (\mid C \mid - \epsilon_{\alpha,i}) = n \mid C \mid^2 - \sum_{i=1}^{n} \sum_{\alpha \in \mathbb{F}_q} \epsilon_{\alpha,i}^2.$$

Now, we apply the Cauchy-Schwarz inequality to the vectors $v = (1, \ldots, 1) \in \mathbb{F}_q^q$ and $w = (\epsilon_{\alpha,i} \mid \alpha \in \mathbb{F}_q) \in \mathbb{F}_q^q$, obtaining

$$q \sum_{\alpha \in \mathbb{F}_q} \epsilon_{\alpha,i}^2 \geq (\sum_{\alpha \in \mathbb{F}_q} \epsilon_{\alpha,i})^2 = \mid C \mid^2,$$

hence

$$\Sigma \leq n \mid C \mid^2 - n \mid C \mid^2 / q = n \mid C \mid^2 (q-1)/q$$

and thus

$$d \mid C \mid (\mid C \mid -1) \leq n \mid C \mid^2 (q-1)/q,$$

hence

$$\mid C \mid \leq \frac{qd}{qd-(q-1)n}.$$

□

Definition 5.5. Let $r \geq 3$. The *simplex code* of dimension r, denoted $\mathcal{S}_r$, is the dual code of the Hamming code of redundancy r.

Proposition 5.5. *Every non-zero word of $\mathcal{S}_r$ has a Hamming weight equal to q^{r-1}. In particular, $\mathcal{S}_r$ reaches the Plotkin bound.*

Proof. The code $\mathcal{S}_r$ has as its generator matrix the parity-check matrix of the Hamming code, i.e., the matrix whose columns are the vectors of $\mathbb{F}_q^r$ up to scalar multiplication. Let $x \in \mathcal{S}_r$ be a non-zero codeword. Then $x = y \cdot H$ for a non-zero vector $y \in \mathbb{F}_q^r$. For every non-zero vector $y \in \mathbb{F}_q^r$, there are exactly $(q^{r-1}-1)/(q-1)$ non-zero vectors (up to scalar multiplication) $h \in \mathbb{F}_q^r$ such that $y \cdot h = 0$. Therefore, there are exactly $(q^{r-1}-1)/(q-1)$ columns h of H for which $y \cdot h = 0$. Therefore, the Hamming weight of x is

$$\frac{q^r - 1}{q-1} - \frac{q^{r-1}-1}{q-1} = q^{r-1}.$$

□

5.2 Typical minimum distance of linear codes: the asymptotic Gilbert–Varshamov bound

Theorem 5.3. *For any rational $0 \leq \kappa \leq 1$, there exists a family of linear codes $\{C_n\}_{n=1}^{\infty}$ with rate κ such that the quantity $\delta = \liminf_{n\to\infty} d(C_n)/n$ satisfies*

$$\kappa \geq 1 - h(\delta),$$

where $h(\delta)$ is the entropy function that is, in the binary case,

$$h(\delta) = \delta \log_2(1/\delta) + (1-\delta)\log_2(1/(1-\delta)),$$

and in the q-ary case

$$h(\delta) = \delta \log_q(q-1) + \delta \log_q(1/\delta) + (1-\delta)\log_q(1/(1-\delta)).$$

Remark 5.1. Let C be the code of rate κ generated by a random matrix G of size $k \times n$ with $k = \kappa n$, in systematic form, i.e., of the form $[I_k|A]$, where A is chosen uniformly in the space of binary matrices of size $k \times (n-k)$. It can be shown that with probability 1, we have $\liminf_{n\to\infty} d(C)/n = h^{-1}(1-\kappa)$.

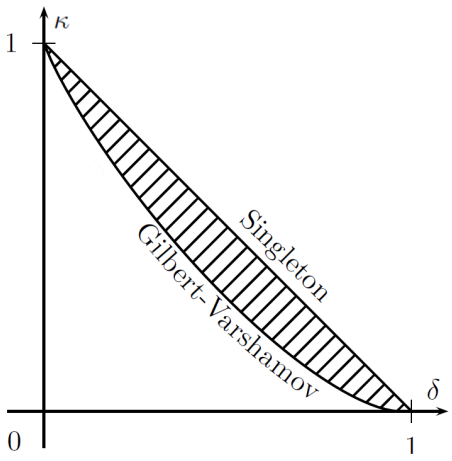


Figure 3: The domain of linear codes.

6 Evaluation codes

6.1 Reed–Solomon codes and their relatives

In this section, we are going to study a systematic construction of linear codes attaining the Singleton bound, known under the name of Reed–Solomon codes, and some codes related to them.

Definition 6.1 (Reed–Solomon codes)**.** Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be a set of distinct elements in $\mathbb{F}_q$. Let k be an integer. Consider the following evaluation map

$$\begin{aligned} \mathrm{ev} : \mathbb{F}_q[T]_{<k} &\to \mathbb{F}_q^n \\ P &\to (P(x_1), \ldots, P(x_n)) \end{aligned}.$$

The Reed–Solomon code $\mathrm{RS}_k(\mathbf{x})$ is the image of the above evaluation map. In other terms, we have

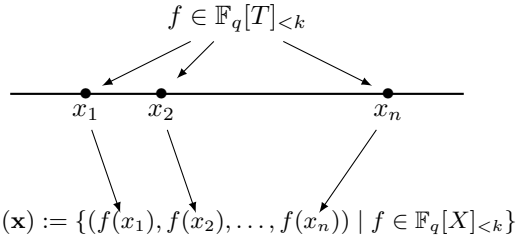
$$\mathrm{RS}_k(\mathbf{x}) := \{(P(x_1), \ldots, P(x_n)) \mid P \in \mathbb{F}_q[T]_{<k}\}.$$

Equivalently, $\mathrm{RS}_k(\mathbf{x})$ is the code with generator matrix

$$G = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{k-1} & x_2^{k-1} & \cdots & x_n^{k-1} \end{pmatrix}.$$

Theorem 6.1. *Let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ be a set of distinct elements in $\mathbb{F}_q$. Let $k \leq n$. Then the Reed–Solomon code $\mathrm{RS}_k(\boldsymbol{x})$ is a linear $[n, k, n+1-k]$ code. In particular, it is an MDS code.*

Figure 4: Reed–Solomon codes



Proof. The code $\mathrm{RS}_k(\mathbf{x})$ is linear since it is the image of a linear map. Its length is clearly n. A basis of $\mathbb{F}_q[T]_{<k}$ is given by $\{1, T, \ldots, T^{k-1}\}$, hence the dimension of $\mathbb{F}_q[T]_{<k}$ is k. Hence, we only need to prove that the evaluation map is injective. To do so, suppose that there are two distinct polynomials P and Q such that $\mathrm{ev}(P) = \mathrm{ev}(Q)$. Then $\mathrm{ev}(P-Q)$ We now turn to the minimum distance. Thus, the polynomial $P - Q$ vanishes in at least $n \geq k$ points and has degree at most $k-1$. Since a polynomial of degree $k-1$ has at most $k-1$ zeros, $P - Q$ must be the zero polynomial, hence $P = Q$. Now Hence

$$d(\mathrm{ev}(P)(\boldsymbol{x}), \mathbf{0}) = \#\{P(x_i) \neq 0\} = n - \#\{P(x_i) = 0\} \geq n - (k-1).$$

On the other hand, the Singleton bound ensures that $d \leq n - k + 1$. We finally conclude that the minimum distance is $n - k + 1$. Thus $\mathrm{RS}_k(\alpha)$ reaches the Singleton bound. □

Definition 6.2 (Generalized Reed–Solomon codes)**.** Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be a set of distinct elements in $\mathbb{F}_q$. Let $\mathbf{v} = \{v_1, \ldots, v_n\}$ be a set of nonzero elements in $\mathbb{F}_q$. The Generalized Reed–Solomon (GRS) code $\mathrm{GRS}_k(\mathbf{x}, \mathbf{v})$ is defined as

$$\mathrm{GRS}_k(\mathbf{x}, \mathbf{v}) := \{(v_1 P(x_1), \ldots, v_n P(x_n)) \mid P \in \mathbb{F}_q[T]_{<k}\}.$$

The proof of Theorem 6.1 applies to prove that GRS codes are MDS codes too.

Theorem 6.2. *Let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ be a set of distinct elements in $\mathbb{F}_q$. Let $\boldsymbol{v} = \{v_1, \ldots, v_n\}$ be a set of nonzero elements in $\mathbb{F}_q$. Let $k \leq n$. Then the Generalized Reed–Solomon code $\mathrm{GRS}_k(\boldsymbol{x}, \boldsymbol{v})$ is a $[n, k, n+1-k]$ code. In particular, it is an MDS code.*

Note that a generator matrix of $\mathrm{GRS}_k(\mathbf{x}, \mathbf{v})$ is

$$G = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \\ v_1 x_1 & v_2 x_2 & \cdots & v_n x_n \\ v_1 x_1^2 & v_2 x_2^2 & \cdots & v_n x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ v_1 x_1^{k-1} & v_2 x_2^{k-1} & \cdots & v_n x_n^{k-1} \end{pmatrix}.$$

Remark 6.1. Since we can only choose up to q distinct elements in $\mathbb{F}_q$, the length n is bounded by q. Therefore, to construct long Reed–Solomon codes, one needs to work over big finite fields.

Theorem 6.3 (Dual of a Generalized Reed–Solomon code)**.** *Let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ be a set of distinct elements in $\mathbb{F}_q$ and $\boldsymbol{v} = \{v_1, \ldots, v_n\}$ be a set of nonzero elements in $\mathbb{F}_q$. The dual of the GRS code $\mathrm{GRS}_k(\boldsymbol{x}, \boldsymbol{v})$ is given by*

$$\mathrm{GRS}_k(\boldsymbol{x}, \boldsymbol{v})^{\perp} = \mathrm{GRS}_{n-k}(\boldsymbol{x}, \boldsymbol{w}),$$

where $\boldsymbol{w} = \{w_1, \ldots, w_n\}$ is defined by

$$w_i = \frac{1}{v_i \prod_{j \neq i}(x_i - x_j)} \quad \text{for } i = 1, \ldots, n.$$

Proof. Let $\mathrm{GRS}_k(\mathbf{x}, \mathbf{v})$ be a GRS code as defined, and let $\mathbf{c} = (v_1 P(x_1), \ldots, v_n P(x_n))$ be a codeword in $\mathrm{GRS}_k(\mathbf{x}, \mathbf{v})$ for some polynomial $P \in \mathbb{F}_q[X]_{<k}$. We want to show that $\mathrm{GRS}_k(\mathbf{x}, \mathbf{v})^{\perp} = \mathrm{GRS}_{n-k}(\mathbf{x}, \mathbf{w})$, where $\mathbf{w}$ is defined as above.

Consider a codeword $\mathbf{d} = (w_1 Q(x_1), \ldots, w_n Q(x_n))$ in $\mathrm{GRS}_{n-k}(\mathbf{x}, \mathbf{w})$ for some polynomial $Q \in \mathbb{F}_q[X]_{<n-k}$. The inner product of $\mathbf{c}$ and $\mathbf{d}$ is:

$$\langle \mathbf{c}, \mathbf{d} \rangle = \sum_{i=1}^{n} v_i P(x_i) \cdot w_i Q(x_i).$$

Substituting the expression for w_i:

$$\langle \mathbf{c}, \mathbf{d} \rangle = \sum_{i=1}^{n} v_i P(x_i) \cdot \frac{1}{v_i \prod_{j \neq i}(x_i - x_j)} Q(x_i) = \sum_{i=1}^{n} \frac{P(x_i) Q(x_i)}{\prod_{j \neq i}(x_i - x_j)}.$$

Let $L(X) = \prod_{j=1}^{n}(X - x_j)$. The derivative of $L(X)$ at x_i is:

$$L'(x_i) = \prod_{j \neq i}(x_i - x_j).$$

Thus, the inner product becomes:

$$\langle \mathbf{c}, \mathbf{d} \rangle = \sum_{i=1}^{n} \frac{P(x_i) Q(x_i)}{L'(x_i)}.$$

Since P and Q are polynomials of degree less than k and $n-k$ respectively, their product PQ has degree less than n. By the Lagrange interpolation formula, the sum $\sum_{i=1}^{n} \frac{P(x_i)Q(x_i)}{L'(x_i)}$ is the coefficient of X^{n-1} in PQ, which is zero because $\deg(PQ) < n$. Therefore, $\langle \mathbf{c}, \mathbf{d} \rangle = 0$.

This shows that $\mathrm{GRS}_{n-k}(\mathbf{x}, \mathbf{w}) \subseteq \mathrm{GRS}_k(\mathbf{x}, \mathbf{v})^{\perp}$. Equality follows from a dimension argument, as $\dim_{\mathbb{F}_q}(\mathrm{GRS}_k(\mathbf{x}, \mathbf{v})) = k$ and $\dim_{\mathbb{F}_q}(\mathrm{GRS}_{n-k}(\mathbf{x}, \mathbf{w})) = n - k$. □

Corollary 6.1. *Let $n = q$. Then*

$$\mathrm{RS}_k(\boldsymbol{x})^\perp := \mathrm{RS}_{n-k}(\boldsymbol{x}).$$

Proof. Exercise. *Hint:* use the fact that $L(X) = \prod_{i=1}^{q}(X - x_i) = X^q - x$ and that $P'(X) = -1$. □

Exercise 6.1. Extended Reed–Solomon Code. Let $\mathbf{x} = (x_1, \ldots, x_q) \in \mathbb{F}_q^q$ such that $\mathbb{F}_q = \{x_1, \ldots, x_q\}$. Let $k \leq q$ and $\mathbb{F}_q[X]_{<k}$ be the vector space of univariate polynomials of degree strictly less than k. We define the evaluation at infinity of a polynomial $f \in \mathbb{F}_q[X]_{<k}$, denoted $f(\infty)$, as the evaluation at 0 of $X^{k-1} f\left(\frac{1}{X}\right)$.

We consider the code $\mathrm{ERS}_k(\mathbf{x})$ as the image of the linear map

$$\mathrm{ev}_k : \begin{cases} \mathbb{F}_q[X]_{<k} & \to & \mathbb{F}_q^{q+1} \\ f & \mapsto & (f(x_1), \ldots, f(x_q), f(\infty)). \end{cases}$$

1. Prove that for any $f \in \mathbb{F}_q[X]_{<k}$, the evaluation at infinity $f(\infty)$ is equal to the coefficient of degree $k-1$.
2. Give the parameters $[n, k, d]$ of ERS_k and prove that for any $k \geq 0$, it is an MDS code (*Recall: A code is called MDS if its parameters meet the Singleton bound*).
3. What is the dimension of the dual of the code $\mathrm{ERS}_k(\mathbf{x})$?

6.2 Algebraic Geometry codes

In this subsection, we briefly introduce Algebraic Geometry codes from curves. We refer the reader to [5] and [6] for further details. In particular, the first reference has a more algebraic approach, while the second has a geometric flavour.

Let X be a smooth, projective, absolutely irreducible curve defined over a finite field $\mathbb{F}_q$. In the following, we will call such a curve a *nice* curve. Let D be a divisor on X, that is, a formal sum of the points on the curve with integer coefficients, almost all zero:

$$D = \sum_{P \in X} n_P P.$$

The degree of a divisor D as above is defined as

$$\deg(D) = \sum_{P \in X} n_P \deg(P),$$

where $\deg(P)$ is the minimum degree of the extension of $\mathbb{F}_q$ where the point is defined. The Riemann–Roch space associated with D is defined as follows:

$$L(D) = \{f \in \mathbb{F}_q(X)^* \mid \mathrm{ord}_P(f) + n_P \geq 0 \ \forall P \in X\} \cup \{0\}.$$

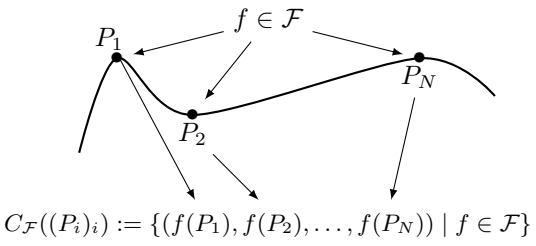


Figure 5: Algebraic Geometry codes.

Theorem 6.4 (Riemann's Inequality)**.** *Let X be a nice curve of genus g, and let D be a divisor on X. Then, the dimension $\ell(D)$ of the Riemann–Roch space $L(D)$ satisfies the following inequality:*

$$\ell(D) \geq \deg(D) + 1 - g.$$

In particular, if $\deg(D) \geq 2g - 1$, then equality holds.

Definition 6.3 (Algebraic Geometry (AG) Code)**.** Let X be a curve. Let $\mathcal{P} = \{P_1, \ldots, P_n\}$ be a set of n distinct $\mathbb{F}_q$-rational points on X, and let D be a divisor on X such that $\operatorname{supp}(D) \cap \mathcal{P} = \emptyset$. Consider the Riemann–Roch space $L(D)$ and the **evaluation map** defined as:

$$\begin{aligned} \operatorname{ev} : L(D) &\to \mathbb{F}_q^n \\ f &\mapsto (f(P_1), \ldots, f(P_n)). \end{aligned}$$

The **Algebraic Geometry (AG) code** $\mathrm{AG}_X(D, \mathcal{P})$ is the image of this evaluation map:

$$\mathrm{AG}_X(D, \mathcal{P}) := \{(f(P_1), \ldots, f(P_n)) \mid f \in \mathcal{L}(D)\}.$$

Theorem 6.5 (Parameters of the AG Code)**.** *Let X be a nice curve of genus g over $\mathbb{F}_q$. Let $\mathcal{P} = \{P_1, \ldots, P_n\}$ be a set of n distinct $\mathbb{F}_q$-rational points on X, and let D be a divisor on X with $\operatorname{supp}(D) \cap \mathcal{P} = \emptyset$. Let $\mathrm{AG}_X(D, \mathcal{P})$ denote the Algebraic Geometry code constructed from $L(D)$ and $\mathcal{P}$.*

The parameters of $\mathrm{AG}_X(D, \mathcal{P})$ are as follows:

- ***Length:*** n
- ***Dimension:*** $k = \ell(D)$*, where* $\ell(D) = \dim_{\mathbb{F}_q} L(D)$
- ***Minimum distance:*** $d \geq n - \deg(D)$

Furthermore, if $\deg(D) < n$, then $\mathrm{AG}_X(D, \mathcal{P})$ is an $[n, k, d]$ linear code with $k = \ell(D)$ and $d \geq n - \deg(D)$.

Bibliography

[1] Venkatesan Guruswami, Atri Rudra, et Madhu Sudan. *Essential coding theory.* Draft available at http://www. cse. buffalo. edu/atri/courses/coding-theory/book 2.1 (2012).

[2] Venkatesan Guruswami et Mahdi Cheraghchi. Information Theory and its applications in theory of computation, (2013).

[3] Alberto Ravagnani, Coding Theory (lecture notes) (2025).

[4] Claude Elwood Shannon. *A mathematical theory of communication.* The Bell System Technical Journal 27.3 (1948): 379-423.

[5] Henning Stichtenoth. *Algebraic function fields and codes.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[6] Michael A. Tsfasman, Serge G. Vlăduț, and Dmitry Nogin. *Algebraic Geometric Codes: Basic Notions* Vol. 1. American Mathematical Soc., 2007.