Introduction to algebraic codings
Lecture Notes for MTH 416
Fall 2024

Ulrich Meierfrankenfeld

August 27, 2024

Preface

These are the Lecture Notes for the class MTH 416 in Fall 2024 at Michigan State University.

The Lecture Notes will be frequently updated throughout the semester.

Contents

Chapter 1

Coding

1.1 Matrices

Definition 1.1.1. *Let I and R be sets.*

(a) *An I-tuple with coefficients in R is a function $x : I \to R$. We will write x_i for the image of i under x and denote x by $(x_i)_{i \in I}$. x_i is called the i-coefficient of x.*

(b) $\mathbb{N} := \{0, 1, 2, 3, \dots,\}$ *denotes the set of non-negative integers.*

(c) $\mathbb{Z}^+ := \{1, 2, 3, \dots,\}$ *denotes the set of positive integers.*

(d) *Let $n \in \mathbb{N}$. Then an n-tuple is a $\{1, 2, \dots, n\}$-tuple.*

(e) $\mathbb{R}$ *denotes the set of real numbers.*

Notation 1.1.2. *Notation for tuples.*

(1)

$$x : \quad \begin{array}{cccc} a & b & c & d \\ \hline 0 & \pi & 1 & \frac{1}{3} \end{array}$$

denotes the $\{a, b, c, d\}$-tuple with coefficients in $\mathbb{R}$ such that

$$x_a = 0, \quad x_b = \pi, \quad x_c = 1 \quad \textit{and} \quad x_d = \frac{1}{3}$$

We denote this tuple also by

$$x: \quad \begin{array}{c|c} a & 0 \\ b & \pi \\ c & 1 \\ d & \frac{1}{3} \end{array} .$$

(2)
$$y = (a, a, b, c)$$
denotes the 4*-tuple with coefficients in* $\{a, b, c, d, e, \dots, z\}$ *such that*

$$y_1 = a, \quad y_2 = a, \quad y_3 = b, \quad \text{and} \quad y_4 = c.$$

We will denote such a 4*-tuple also by*

$$y = \begin{pmatrix} a \\ a \\ b \\ c \end{pmatrix}.$$

Definition 1.1.3. *Let* I, J *and* R *be sets.*

(a) $I \times J := \{(i, j) \mid i \in I, j \in J\}$.

(b) *An* $I \times J$*-matrix with coefficients in* R *is a function* $M : I \times J \to R$. *We will write* m_{ij} *for the image of* (i, j) *under* M *and denote* M *by* $[m_{ij}]_{\substack{i \in I \\ j \in J}}$. m_{ij} *is called the* ij*-coefficients of* M.

(c) *Let* $n, m \in \mathbb{N}$. *An* $n \times m$*-matrix is an* $\{1, 2, \dots, n\} \times \{1, 2, \dots, m\}$*-matrix.*

Notation 1.1.4. *Notations for matrices*

(1) *We will often write an* $I \times J$*-matrix as an array. For example*

$$\begin{array}{c|ccc} M & x & y & z \\ \hline a & 0 & 1 & 2 \\ b & 1 & 2 & 3 \\ c & 2 & 3 & 4 \\ d & 3 & 4 & 5 \end{array}$$

stands for the $\{a,b,c,d\}\times\{x,y,z\}$ matrix M with coefficients in $\mathbb{Z}$ such that

$m_{ax}=0$, $m_{ay}=1$, $m_{bx}=1$, $m_{cz}=4$

(2) *$n\times m$-matrices are denoted by an $n\times m$-array in square brackets. For example*

$$M=\begin{bmatrix}0&1&2\\4&5&6\end{bmatrix}$$

denotes the 2×3 matrix M with

$m_{11}=0, m_{12}=1$, $m_{21}=4$, $m_{23}=6$,....

Definition 1.1.5. *Let M be an $I\times J$-matrix.*

(a) *Let $i\in I$. Then row i of M is the J-tuple $(m_{ij})_{j\in J}$. We denote row i of M by $\mathrm{Row}_i(M)$ or by M_i.*

(b) *Let $j\in J$. Then column j of J is the I-tuple $(m_{ij})_{i\in I}$. We denote column j of M by $\mathrm{Col}_j(M)$*

Example 1.1.6. Let

$$M=\begin{bmatrix}a&\pi&3\\0&\alpha&x\end{bmatrix}.$$

Compute M_2, $\mathrm{Row}_2(M)$ and $\mathrm{Col}_3(M)$.

$$M_2=\mathrm{Row}_2(M)=(0,\alpha,x)$$

and

$$\mathrm{Col}_3(M)=\begin{pmatrix}3\\x\end{pmatrix}.$$

Definition 1.1.7. *Let A be an $I\times J$-matrix, B a $J\times K$ matrix and x and y J-tuples with coefficients in $\mathbb{R}$. Suppose J is finite.*

(a) *AB denotes the $I\times K$ matrix whose ik-coefficient is*

$$\sum_{j\in J}a_{ij}b_{jk}$$

(b) *Ax denotes the I-tuple whose i-coefficient is*

$$\sum_{j\in J} a_{ij}x_j$$

(c) *xB denotes the K-tuple whose k-coefficient is*

$$\sum_{j\in J} x_j b_{jk}$$

(d) *xy denotes the real number*

$$\sum_{j\in J} x_j y_j$$

Example 1.1.8. Examples of matrix multiplication.

(1) Given the matrices $\begin{array}{c|ccc} A & x & y & z \\ \hline a & 0 & 1 & 2 \\ b & 1 & 2 & 3 \end{array}$ and $\begin{array}{c|cccc} B & \alpha & \beta & \gamma & \delta \\ \hline x & 0 & 0 & 1 & 0 \\ y & 1 & 0 & 0 & 1 \\ z & 1 & 1 & 0 & 0 \end{array}$

Compute AB

AB is the $\{a, b\} \times \{\alpha, \beta, \gamma, \delta\}$ matrix

$$\begin{array}{c|cccc} AB & \alpha & \beta & \gamma & \delta \\ \hline a & 3 & 2 & 0 & 1 \\ b & 5 & 3 & 1 & 2 \end{array}$$

(2) Given the matrix 2×3-matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$ and the 3-tuple $x = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.

Compute Ax

Ax is the 2-tuple

$$\begin{pmatrix}3\\5\end{pmatrix}$$

(3) Given the matrix $\begin{array}{c|ccc} A & x & y & z \\ \hline a & 0 & 1 & 2 \\ b & 1 & 2 & 3 \end{array}$ and the tuple $u : \begin{array}{cc} a & b \\ \hline 2 & -1 \end{array}$

Compute uA.

uA is the $\{x,y,z\}$-tuple

$$\begin{array}{ccc} x & y & z \\ \hline -1 & 0 & 1 \end{array}$$

(4) Given the 4-tuples $x = (1,1,2,-1)$ and $y = (-1,1,2,1)$. Compute xy.

$$xy = 3$$

1.2 Basic Definitions

Definition 1.2.1. *An alphabet is a non-empty finite set. The elements of an alphabet are called symbols.*

Example 1.2.2. (a) $\mathbb{A} = \{A, B, C, D, \dots, X, Y, Z, \sqcup\}$ is the alphabet consisting of the regular 26 uppercase letters and a space (denoted by $\sqcup$).

(b) $\mathbb{B} = \{0, 1\}$ is the alphabet consisting of the two symbols 0 and 1.

Definition 1.2.3. *Let S be an alphabet and n a non-negative integer.*

(a) *A message of length n using S is an n-tuple $(s_1, \dots, s_n)$ with coefficients in S. We denote such an n-tuple by $s_1 s_2 \dots s_n$. A message of length n is also called a string of length n.*

(b) *The length of a message m is denoted by $\ell(m)$.*

(c) *$\varnothing$ denotes the unique message of length 0 using S.*

(d) *S^n is the set of all messages of length n using S.*

(e) *S^* is the set of all messages using S, that is*

$$S^* := \bigcup_{k \in \mathbb{N}} S_k = S_0 \cup S_1 \cup S_2 \cup S_3 \cup \ldots \cup S_k \cup \ldots$$

Example 1.2.4. (1) What is the length of the message
WATCH⊔OUT⊔FOR⊔POKEMONS

using the alphabet $\mathbb{A}$?

22

(2) What is the length of the message 10011101111 using the alphabet $\mathbb{B}$?

11

(3) Compute $\mathbb{B}^0, \mathbb{B}^1, \mathbb{B}^2, \mathbb{B}^3$ and $\mathbb{B}^*$.

$\mathbb{B}^0 = \{\varnothing\}$, $\mathbb{B}^1 = \mathbb{B} = \{0, 1\}$. $\mathbb{B}^2 = \{00, 01, 10, 11\}$, $\mathbb{B}^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$, and $\mathbb{B}^* = \{\varnothing, 0, 1, 00, 01, 10, 11, 000, \ldots, 111, 0000, \ldots, 1111, \ldots\}$

Definition 1.2.5. *Let S and T be alphabets.*

(a) *A code c for S using T is a 1-1 function from S to T^*. So a code assigns to each symbol $s \in S$ a message $c(s)$ using T, and different symbols are assigned different messages*

(b) *The set $C = \{c(s) \mid s \in S\}$ is called the set of codewords of c. Often (somewhat ambiguously) we will also call C a code. To avoid confusion, a code which is function will always be denoted by a lower case letter, while a code which is a set of codewords will be denoted by an upper case letter.*

(c) *A code is called regular if the empty message $\varnothing$ is not a codeword.*

Example 1.2.6. (1) The function $c : \mathbb{A} \to \mathbb{A}^*$ such that

$$A \mapsto D,\ B \mapsto E,\ C \mapsto F,\ \ldots,\ W \mapsto Z,\ X \mapsto A,\ Y \mapsto B,\ Z \mapsto C,\ \sqcup \mapsto \sqcup$$

is a code for $\mathbb{A}$ using $\mathbb{A}$. The set of codewords is

$$C = \mathbb{A}.$$

(2) The function $c : \{x, y, z\} \to \mathbb{B}^*$ such that

$$x \mapsto 0, \quad y \mapsto 01, \quad z \mapsto 10$$

is a code for $\{x, y, z\}$ using $\mathbb{B}$. The set of codewords is

$$C = \{0, 01, 10\}.$$

Definition 1.2.7. *Let $c : S \to T^*$ be a code. Then the function $c^* : S^* \to T^*$ defined by*

$$c^*(s_1 s_2 \dots s_n) = c(s_1)c(s_2)\dots c(s_n)$$

for all $s_1 \dots s_n \in S^*$ is called the concatenation of c. The function c^* is also called the extension of c. Often we will denote c^* by c rather than c^*. Since c is uniquely determined by c^* and vice versa, this ambiguous notation should not lead to any confusion.

Example 1.2.8. (1) Let $c : \mathbb{A} \to \mathbb{A}^*$ be the code from (1.2.6)(1). Then $c(ABC) = DEF$.

(2) Let $c : \{x, y, z\} \to \mathbb{B}^*$ be the code from (1.2.6)(2). Then

$$c(xzzx) = 010100$$

and

$$c(yyxx) = 010100$$

So $xzzx$ and $yyxx$ are encoded to the same message in $\mathbb{B}^*$.

Definition 1.2.9. *A code $c : S \to T^*$ is called uniquely decodable (UD) if the extended function $c^* : S^* \to T^*$ is 1-1.*

Example 1.2.10. (a) Is the code from (1.2.6)(1) a *UD*-code?

Yes, to decode a message, just shift each letter backwards by three letters.

(b) Is the code from (1.2.6)(2) a *UD*-code?

No, since $c^*(xzzx) = 010100 = c^*(yyxx)$.

1.3 Coding for economy

Example 1.3.1. The Morse alphabet $\mathbb{M}$ has three symbols $\bullet, -$ and $\odot$, called dot, dash and pause. The Morse code is the code for $\mathbb{A}$ using $\mathbb{M}$ defined by

A	B	C	D	E	F	G	H	I
$\bullet - \odot$	$- \bullet \bullet \bullet \odot$	$- \bullet - \bullet \odot$	$- \bullet \bullet\odot$	$\bullet\odot$	$\bullet \bullet - \bullet \odot$	$- - \bullet\odot$	$\bullet \bullet \bullet \bullet \odot$	$\bullet \bullet \odot$
J	K	L	M	N	O	P	Q	R
$\bullet - - - \odot$	$- \bullet -\odot$	$\bullet - \bullet \bullet \odot$	$- - \odot$	$- \bullet \odot$	$- - -\odot$	$\bullet - - \bullet \odot$	$- - \bullet - \odot$	$\bullet - \bullet\odot$
S	T	U	V	W	X	Y	Z	$\sqcup$
$\bullet \bullet \bullet\odot$	$-\odot$	$\bullet \bullet -\odot$	$\bullet \bullet \bullet - \odot$	$\bullet - -\odot$	$- \bullet \bullet - \odot$	$- \bullet - - \odot$	$- - \bullet \bullet \odot$	$\odot\odot$

So for example the messages SOS in $\mathbb{A}$ encodes to
$\bullet \bullet \bullet \odot - - - \odot \bullet \bullet \bullet\odot$.

Exercises 1.3:

1.3#1. (a) Is the Morse code, as defined in Example 1.3.1 (or page 7 in the book), prefix-free?

(b) Consider the modified Morse code obtained by removing the pause at the end of every codeword. Is this code prefix free? Is it uniquely decodable?

1.4 Coding for reliability

Example 1.4.1. Define some codes for $\{\text{Buy}, \text{Sell}\}$ using $\{B, S\}$.

(1) Sell $\mapsto S$, Buy $\mapsto B$.

(2) Buy $\mapsto BB$, Sell $\mapsto SS$.

(3) Buy $\mapsto BBB$, Sell $\mapsto SSS$.

(4) Buy $\mapsto BBBBBBBBB$, Sell $\mapsto SSSSSSSSS$.

1.5 Coding for security

Example 1.5.1. Let k be an integer with $0 \leq k \leq 25$. Then c_k is the code from $\mathbb{A} \to \mathbb{A}$ obtained by shifting each letter by k places to the right. ⊔ is unchanged. For example the code in Example (1.2.6)(1) is c_3.

This code is not very secure, in the sense that given an encoded message it is not very difficult to determine the original message, even if one does not know what parameter was used. See Exercise 1.5#1

Exercises 1.5:

1.5#1. The code c_k from Example 1.5.1 (or page 9 in the book) – shifting each letter by k-places– is used to encode a message. If the encoded messages is

QY⊔QBOOX⊔QY⊔GRSDO

what value for 'k' was used and what was the original message?

Chapter 2

Prefix-free codes

2.1 The decoding problem

Definition 2.1.1. *Let T be an alphabet and let $a, b \in T^*$.*

(a) *If $a = t_1 \dots t_m$ and $b = s_1 \dots s_n$, then $ab := t_1 \dots t_m s_1 \dots s_n$.*

(b) *a is called a prefix of b if there exists a message r using T with $b = ar$.*

(c) *a is called a proper prefix of b if a is a prefix of b and $a \neq b$.*

(d) *a is called a parent of b if there exists $r \in T$ with $b = ar$.*

(e) *A code C is called prefix-free (PF) if no codeword of C is a proper prefix of a codeword.*

Example 2.1.2. Consider the following binary messages.

(1) Is 10111 a prefix of 10111011? Is it a parent?
It is a prefix but not a parent.

(2) Find the parent of 0110111.

$$011011$$

Remark 2.1.3. *Let T be an alphabet and let $b = t_1 \dots t_m$ be a message of length m using T.*

(a) *$\varnothing b = b$ and $b\varnothing = b$. In particular, both $\varnothing$ and b are prefixes of b.*

(b) *Let $a, r \in T^*$. Then $\ell(ar) = \ell(a) + \ell(r)$.*

(c) *Any prefix of b has length less or equal to m.*

(d) *Let $0 \le n \le m$. Then $t_1 \dots t_n$ is the unique prefix of length n of b.*

(e) *If $m \ne 0$, then $t_1 \dots t_{m-1}$ is the unique parent of b.*

(f) *Let c and d be prefixes of b. Then c is prefix of d if and only if $\ell(c) \le \ell(d)$.*

Proof. (a): Clearly $\varnothing b = b$ and $b\varnothing = b$. So (a) holds.

(b): Let $a = s_1 \dots s_n$ and $r = u_1 \dots u_k$. Then $ar = s_1 \dots s_n u_1 \dots u_k$. So $\ell(ar) = n + k = \ell(a) + \ell(r)$.

(c): Let a be a prefix of b. Then $b = ar$ for some $r \in T^*$. By (b) we have $\ell(b) = \ell(ar) = \ell(a) + \ell(r) \ge \ell(a)$.

(d) and (e): Let $a = s_1 \dots s_n$ be a prefix of b of length n. Then $b = ar$ for some $r \in T^*$ Let $r = u_1 \dots u_k$. Then $b = s_1 \dots s_n u_1 \dots u_k$. It follows that $t_i = s_i$ for $1 \le i \le n$ and so $a = t_1 \dots t_n$ and (d) holds. If a is a parent, then $r \in T$, that is $k = 1$ and $n = m - 1$. By (d) a is the unique prefic of lenghth $m - 1$ of v an so (e) holds. a.

(f): If c is a prefix of d, then (d) shows that $\ell(c) \le \ell(d)$. So suppose $\ell(c) \le \ell(d)$. Put $n = \ell(c)$ and $k = \ell(d)$. Then $n \le k$ and by (a) $c = t_1 \dots t_n$ and $d = t_1 \dots t_k$. Hence $d = (t_1 \dots t_n)(t_{n+1} \dots t_k) = c(t_{n+1} \dots t_m)$. So c is a prefix of d. □

Example 2.1.4. Which of the following codes are PF? UD?

(1) $C := \{10, 01, 11, 011\}$.

Since 01 is a prefix of 011, C is not prefix-free. Also

$$011011 = (01)(10)(11) = (011)(011)$$

and so C is not uniquely decodable.

(2) Let $C := \{021, 2110, 10001, 21110\}$. Observe that C is prefix free.

This can be used to recover any sequence $c_1, c_2 \dots, c_n$ of codewords from their concatenation $e = c_1 \dots c_n$. Consider for example the string

$$e = 2111002110001.$$

We will look at prefixes of increasing length until we find a codeword:

Prefix	codeword?
$\varnothing$	no
2	no
21	no
211	no
2111	no
21110	yes

No longer prefix can be a codeword since it would have the codeword 21110 as a prefix.

So $c_1 = 21110$.

We now remove 21110 from e to get

$$02110001.$$

The prefixes are

Prefix	codeword?
$\varnothing$	no
0	no
02	no
021	yes

So $c_2 = 021$. Removing 021 gives

$$10001.$$

This is a codeword and so $c_3 = 10001$. Thus the only decomposition of e into codewords is

$$2111002110001 = (21110)(021)(1001).$$

This example indicates that C is UD. The next theorem confirms this.

Theorem 2.1.5. *Any regular PF code is UD.*

Proof. Let C be a regular PF-code. Let $n, m \in \mathbb{N}$ and $c_1, \dots, c_n$, $d_1, \dots, d_m$ be codewords with

$$c_1 \dots c_n = d_1 \dots d_m.$$

We need to show that $n = m$ and $c_1 = d_1, c_2 = d_2, \dots, c_n = d_n$. The proof is by induction on $\min(n, m)$. Put $e = c_1 \dots c_n$ and so also $e = d_1 \dots d_m$.

Suppose first that $\min(n, m) = 0$. Then $n = 0$ or $m = 0$. Since the setup is symmetric in n and m we may assume that $n = 0$. Then $e = c_1 \dots c_n = \varnothing$ and so $d_1 \dots d_m = \varnothing$. By definition of a regular code $d_i \neq \varnothing$ for all $1 \leq i \leq m$. Hence also $m = 0$ and we are done in this case.

Suppose next that $\min(n, m) \neq 0$. Then $n > 0$ and $m > 0$. We may assume that $\ell(c_1) \leq \ell(d_1)$.

Since $e = c_1 \dots c_n$, c_1 is a prefix of e and since $e = d_1 \dots d_m$, d_1 is a prefix of e. As $\ell(c_1) \leq \ell(d_1)$ this implies that c_1 is a prefix of d_1, see (2.1.3)(f). As C is prefix-free we conclude that $c_1 = d_1$. Since $c_1 c_2 \dots c_n = d_1 d_2 \dots d_m$ this gives

$$c_2 \dots c_n = d_2 \dots d_m.$$

By induction we conclude that $n - 1 = m - 1$ and $c_2 = d_2, \dots, c_n = d_n$. Hence $n = m$ and $c_1 = d_1, \dots, c_n = d_n$. □

Lemma 2.1.6. (a) *All UD codes are regular.*

(b) *Let $c : S \to T^*$ be non-regular code. Then c is PF if and only if $|S| = 1$ and $c(s) = \varnothing$ for $s \in S$.*

Proof. Let $c : S \to T^*$ be non-regular code. Then there exists $s \in S$ with $c(s) = \varnothing$ for $s \in S$. Thus

$$c^*(ss) = \varnothing\varnothing = \varnothing = c^*(s).$$

and so c is not UD. We proved that a non-regular code is not UD and so every UD-code is regular. Thus (a) holds.

Suppose in addition that c is prefix-free and let $t \in S$. Since $c(s) = \varnothing$ we know that $c(s)$ is a prefix of $c(t)$, see (2.1.3)(a). Since c is prefix-free this implies that $c(s) = c(t)$. By definition of a code, c is 1-1 and so $s = t$. This shows that $S = \{s\}$ and hence the forward direction in (b) holds.

Any code with only one codeword is PF and so the backwards direction in (b) holds. □

Exercises 2.1:

2.1#1. Suppose the code $c : S \to T^*$ is such that every codeword has the same length. Is the code uniquely decodable?

2.1#2. Consider the code $c : \{0,1,2,3\} \to \mathbb{B}^*$ with

$$0 \mapsto 0011, \quad 1 \mapsto 100, \quad 2 \mapsto 010 \quad 3 \mapsto 1001$$

(a) Is c prefix-free? (b) Is c uniquely decodable?

2.2 Representing codes by trees

Definition 2.2.1. *A (undirected) graph G is a pair (V, E) such that V and E are sets and each element of E is a subset of size two of V.*

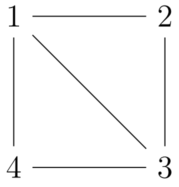
The elements of V are called the vertices of G.

The elements of E are called the edges of G.

We say that the vertex a is adjacent to the vertex b in G if $\{a, b\}$ is an edge.

Example 2.2.2. Let $V = \{1,2,3,4\}$ and $E = \Big\{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{3,4\}\Big\}$. Then $G = (V, E)$ is a graph

It is customary to represent a graph by a picture: Each vertex is represented by a node and a line is drawn between any two adjacent vertices. For example the above graph can be visualized by the following picture:



Definition 2.2.3. *Let $G = (V, E)$ be a graph.*

(a) *Let a and b be vertices. A path of length n from a to b in G is a tuple $(v_0, v_1, \ldots, v_n)$ of vertices such that $a = v_0$, $b = v_n$ and v_{i-1} is adjacent to v_i for all $1 \le i \le n$.*

(b) *G is called connected if for each $a, b \in V$, there exists a path from a to b in G.*

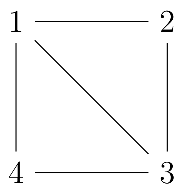
(c) *A path $(v_0, \ldots, v_n)$ is called a cycle if $v_0 = v_n$.*

(d) *A cycle $(v_0, \ldots, v_n)$ called simple $v_i \neq v_j$ for all $1 \le i < j \le n$.*

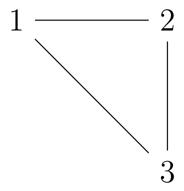
(e) *A tree is a connected graph with no simple cycles of length larger than two.*

Example 2.2.4. Which of the following graphs are trees?

(1)

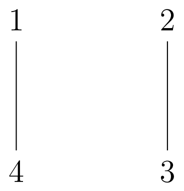


is connect, but



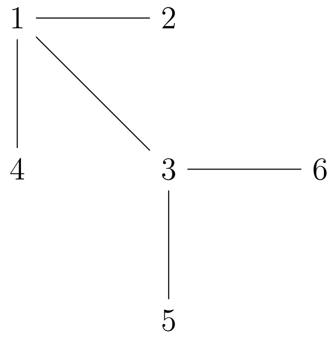
is simple circle of length three in G. So G is not a tree.

(2)



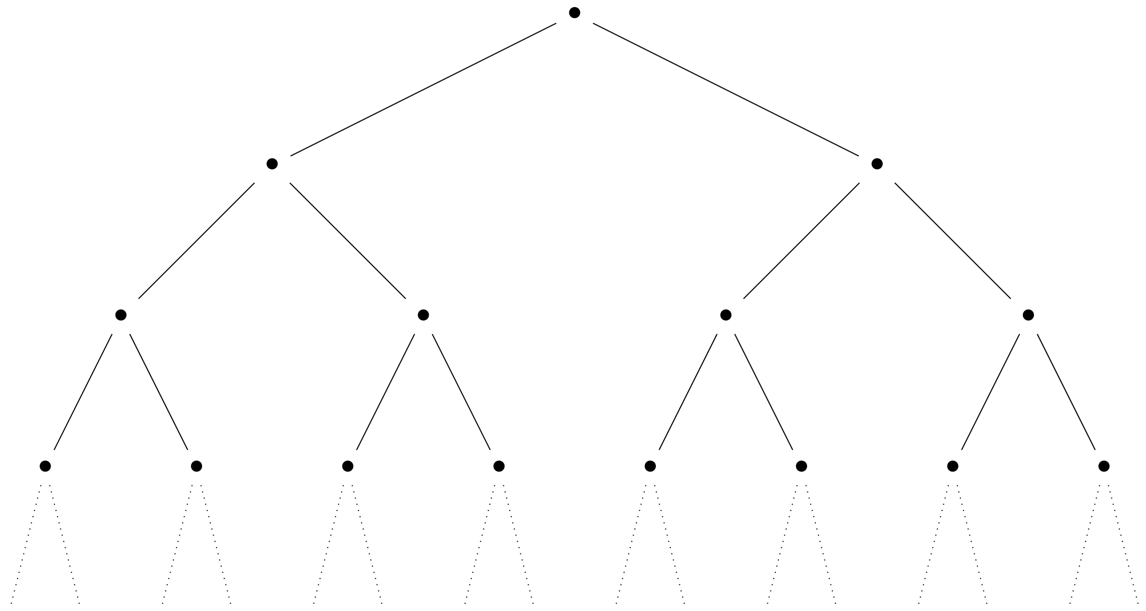
has no simple circle of length larger than two. But it is not connected since there is no path from 1 to 2. So G is not a tree.

(3)



is connected and has no simple circle of length larger than two. So it is a tree.

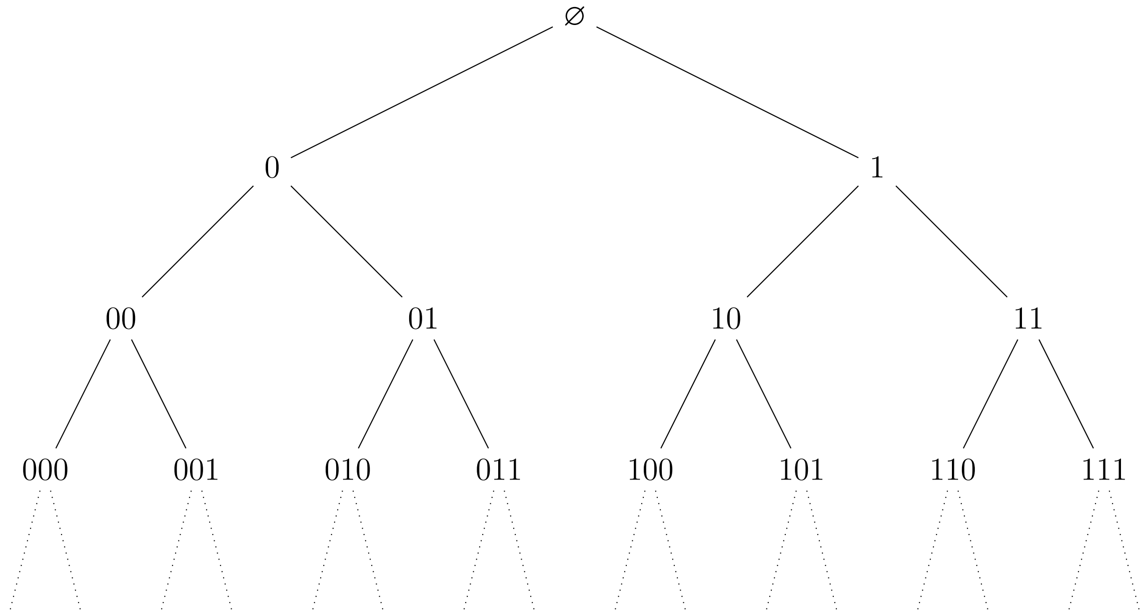
Example 2.2.5. The infinite binary tree



How can one describe this graph in terms of a pair (V, E)?. The vertices are all the binary messages and a message a is adjacent to message b if and only if a is the parent of b or b is the parent of a.

$$V = \mathbb{B}^* \text{ and } E = \Big\{\{a, as\}\Big| a \in \mathbb{B}^*, s \in \mathbb{B}\Big\} = \Big\{\{a, b\}\Big| a, b \in \mathbb{B}^*, a \text{ is the parent of } b\}$$

So the infinite binary tree now looks like:

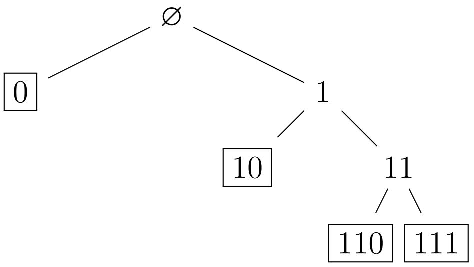


Definition 2.2.6. *Let C be a code. Then $G(C)$ is the graph (V, E), where V is the set of prefixes of codewords and*

$$E = \left\{ \{a, b\} \middle| a, b \in V, a \text{ is a parent of } b \right\}.$$

$G(C)$ is called the graph associated to C.

Example 2.2.7. Determine the graph associated to the code $C = \{0, 10, 110, 111\}$.



Definition 2.2.8. *Let G be a graph. A leaf of G is a vertex which is adjacent to at most one vertex of G.*

Theorem 2.2.9. *Let C be a code using the alphabet T and let $G(C) = (V, E)$ be the graph associated to C.*

(a) *Let $c \in V$ and put $n := \ell(c)$. For k in $\mathbb{N}$ with $k \leq n$ let c_k be the prefix of length k of c. Then $(c_0, c_1, \ldots, c_n)$ is a path from $\varnothing$ to c in $G(C)$.*

(b) *$G(C)$ is tree*

Suppose in addition that C is regular.

(c) *Let a be a codeword. Then a is a proper prefix of a codeword if and only if a is adjacent to at least two vertices of $G(C)$.*

(d) *C is prefix-free if and only if all codewords of C are leaves of $G(C)$.*

Proof. (a) Let $c = t_1 \ldots t_n$ with $t_i \in T$. Then $c_i = t_1 \ldots t_{i-1} t_i$ and so $c_i = c_{i-1} t_i$. Thus c_{i-1} is a parent of c_i. Hence c_{i-1} is adjacent to c_i and $(c_0, \ldots, c_n)$ is a path in $G(C)$. Also $c_0 = \varnothing$ and $c_n = c$.

(b): By (a) there exists a path from each vertex of $G(C)$ to $\varnothing$. It follows that G is connected.

We prove next:

$(*)$ *Let $m \in \mathbb{Z}^+$ and let $(a_0, a_1 \ldots, a_m)$ be path in $G(C)$ such that $a_{i-1} \neq a_{i+1}$ for all $i \in \mathbb{N}$ with $0 < i < m$ and $\ell(a_0) \leq \ell(a_1)$. Then, for all $i \in \mathbb{N}$ with $1 \leq i \leq m$, a_{i-1} is a parent of a_i. In particular, $a_0 \neq a_m$, is a proper prefix of a_m and $\ell(a_0) < \ell(a_m)$.*

Since a_0 is adjacent to a_1, one of a_0 and a_1 is the parent of the other. Since $\ell(a_0) \leq \ell(a_1)$, a_1 is not a parent of a_0 and we conclude that a_0 is parent of a_1. So if $m = 1$, $(*)$ holds. Suppose $m \geq 2$. Since $a_2 \neq a_0$ and a_0 is the unique parent of a_1 we know that a_2 is not the parent of a_1. Since a_2 is adjacent to a_1 we conclude that a_1 is the parent of a_2. In particular, $\ell(a_1) \leq \ell(a_2)$. Induction, applied to the path $(a_1, \ldots, a_m)$, shows that a_{i-1} is a parent of a_i for all $2 \leq i \leq m$ and so again $(*)$ holds.

Now suppose for a contradiction that there exists a simple circle $(v_0, v_1, \ldots, v_n)$ in $G(C)$ with $n \geq 3$. Since $n \geq 3$ we have $v_{i-1} \neq v_{i+1}$ for all $i \in \mathbb{N}$ with $i < n$.

Let $l := \min_{0 \leq i \leq n} \ell(v_i)$ and choose $0 \leq k \leq n$ such that $\ell(v_k) = l$.

Assume that $0 < k < n$. Then $\ell(v_k) = l \leq \ell(v_{k-1})$ and $\ell(v_k) = l \leq \ell(v_{k+1})$. Hence we can apply $(*)$ to the paths $(v_k, v_{k+1} \ldots, v_n)$ and $(v_k, v_{k-1}, \ldots v_0)$. It follows that v_{n-1} is the parent of v_n and v_1 is a parent of v_0. As $v_0 = v_n$ this shows that $v_{n-1} = v_1$. But then $n - 1 = 1$ and so $n = 2$, a contradiction.

Assume next that $k = 0$ or $k = n$. Since $v_0 = v_n$ we conclude that $v_0 = v_k = v_n$. Hence $\ell(v_0) = \ell(v_k) = l \leq \ell(v_1)$. But now $(*)$ shows that $v_0 \neq v_n$, again a contradiction.

So $G(C)$ has no simple circle of length at least three. We already proved that $G(C)$ is connected and so $G(C)$ is a tree.

Suppose from now on that C is regular.

(c) Suppose first that a, b are codewords and a is a proper prefix of b. Then $\ell(b) \geq \ell(a) + 1$. Thus b has a unique prefix c of length $\ell(a) + 1$. Then a is the parent of c. Since c is a prefix of the codeword b, we know that $c \in V$. So c is adjacent to a in $G(C)$. By definition of a regular code, $a \neq \varnothing$ and so a has a parent d. Then d is in V and d is adjacent to a. As a is the parent of c, and d is the parent of a we have $\ell(d) < \ell(a) < \ell(c)$. Hence $d \neq c$. Since both c and d are adjacent to a this shows that a is adjacent to at least two distinct vertices of $G(C)$.

Suppose next that a is a codeword and a is adjacent to two distinct vertices c and d in V. One of c and d, say c is not the parent of a. Since a is adjacent to c, this means that a is a parent of c. As c is in V, c is the prefix of some codeword b. Since a is the parent of c, we see that a is a prefix of b and $a \neq b$. So a is proper prefix of a codeword.

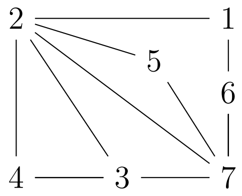
(d) :

	C is prefix-free	
$\iff$	no codeword is a proper prefix of a codeword	– Definition of prefix-free
$\iff$	no codewords is adjacent two different vertices of $G(C)$	– (c)
$\iff$	every codewords is adjacent to at most one vertex	– Basic Logic
$\iff$	each codeword is a leaf	– Definition of a leaf

□

Exercises 2.2:

2.2#1. Find a simple cycle of maximal length in the graph



2.2#2. Construct a tree representation of the ternary code using the alphabet $T = \{0, 1, 2\}$ and codewords 20, 121, 102, 001 and 000. Is it possible to extend this code without destroying the PF property?

2.2#3. Design a PF binary code $c : \{1, 2, 3, 4, 5, 6\} \to \mathbb{B}^*$ such that the sum of the lengths of the codewords is as small as possible. Construct the tree representation of such a code.

2.3 The Kraft-McMillan number

Definition 2.3.1. *Let C be a code using the alphabet T and put $b := |T|$.*

(a) *C is called a b-nary code.*

(b) *C is called binary if $T = \{0,1\}$ and ternary if $T = \{0,1,2\}$.*

(c) *Let $i \in \mathbb{N}$. Then C_i denotes the set of codewords of length i and $n_i := |C_i|$. So n_i is the number of codewords of length i.*

(d) *Let $M \in \mathbb{N}$ such that every code word has length at most M, that is $n_i = 0$ for all $i > M$. The $M+1$-tuple $n = (n_0, n_1, \dots, n_M)$ is called a parameter of C. Note that the parameters are unique up to trailing zeroes.*

(e) *The number*

$$K := \sum_{i=0}^{M} \frac{n_i}{b^i} = n_0 + \frac{n_1}{b} + \frac{n_2}{b^2} + \dots + \frac{n_M}{b^M}$$

is called the Kraft-McMillan number of the parameter $(n_0, n_1, \dots, n_M)$ to the base b, and also is called the Kraft-McMillan number of C.

Example 2.3.2. Compute the Kraft-McMillan number of the binary code $C = \{10, 01, 11, 011\}$.

Since C is a binary code we have $b = 2$. Also

$$C_0 = \{\}, \quad C_1 = \{\}, \quad C_2 = \{10, 01, 11\} \quad \text{and} \quad C_3 = \{011\}.$$

So the parameter of C is

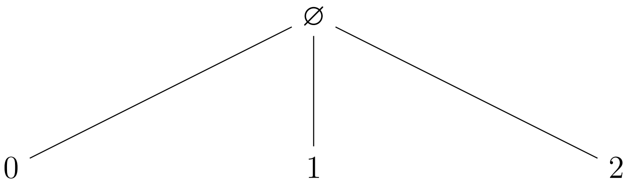
$$(0, 0, 3, 1)$$

and the Kraft-McMillan number is

$$K = 0 + \frac{0}{2} + \frac{3}{4} + \frac{1}{8} = \frac{6+1}{8} = \frac{7}{8}.$$

Example 2.3.3. Construct a ternary PF-code with parameter $(0, 1, 4, 4)$. How many ternary codewords of length 3 can be added to this code with still being PF?

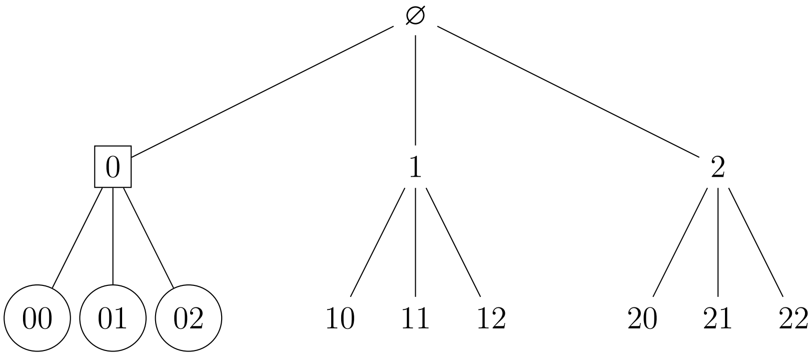
We first look at all the messages of length 1:



Since $n_1 = 1$ we need to pick one of these three messages for our code C. Note that there is no real difference between 0, 1 and 2. So we might as well pick 0. Then

$$C_1 = \{0\}.$$

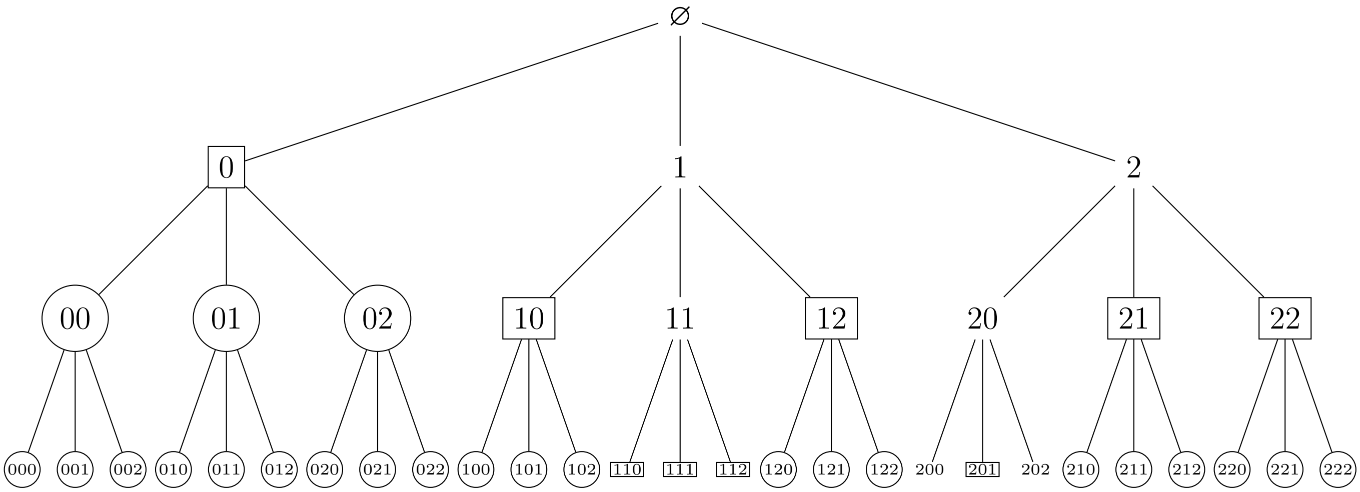
Next we look at all messages of length 2:



The three circled messages have the codeword 0 as a prefix and so cannot be used as codewords. Note that three are $3^2 = 9$ messages of length 3. So this leaves $9 - 3 = 6$ messages of length two for use in the code C. Since $n_2 = 4$ we pick four of these messages by random:

$$C_2 = \{10, 12, 21, 22\}.$$

Finally we look at all the ternary messages of length 3:



The one codeword of length 1, namely 0, is the prefix of $1 \cdot 3^2 = 9$ messages of length 3, the four codewords of length 2, namely $10, 12, 21, 22$, are the prefix of $4 \cdot 3 = 12$ messages of length 3. This leaves

$$3^3 - 1 \cdot 3^2 - 4 \cdot 3 = 27 - 9 - 12 = 6$$

messages of length 3 which can be used for the code C. We choose four of them by random:

$$C_3 := \{110, 111, 112, 201\}$$

This leaves exactly two codewords of length three, namely 200 and 202, which do not have a codeword as a prefix and which could be used to further extend the code. This number (two) can be computed from the Kraft-McMillan number. The total number of messages of length three which have a codeword as prefix is

$$1 \cdot 3^2 + 4 \cdot 3 + 4 \cdot 3^0 = 3^3 \left(\frac{1}{3^2} + \frac{4}{3^2} + \frac{4}{3^3} \right) = 3^3 \cdot K.$$

So the number of messages of length 3 which do not have a codeword as a prefix is $3^3 - 3^3 \cdot K = 3^3(1 - K)$.

Lemma 2.3.4. *Let C be a b-nary code with Kraft-McMillan number K using the alphabet T. Let $M \in \mathbb{N}$ such that every code word has length at most M. Let D be the set of messages of length M using T which have a codeword as a prefix. Then*

(a) $|D| \leq b^M K$.

(b) *If C is PF, then* $|D| = b^M K$.

(c) *If C is PF, then* $K \leq 1$.

Proof. We first show

$(*)$ *Let $r \in \mathbb{N}$. Then there are exactly b^r messages of length r using T.*

Indeed any such messages is of the form

$$t_1 t_2 \dots t_r$$

with $t_1, t_2, \dots t_r \in T$. Since C is b-nary, we know that $|T| = b$. So there are b choices for t_1, b-choices for t_2,..., b-choices for t_r. Thus there are b^r messages of length r.

(a) Let c be a codeword of length i. Then any message of length M using T with c as a prefix is of the form cr where r is a message of length $M - i$. By $(*)$ there are b^{M-i} such r and so there are exactly b^{M-i} message of length M which have c as a prefix. It follows that there are $n_i b^{M-i}$ message of length M which have a codeword of length i as a prefix. Adding over the possible length of codewords gives:

$$|D| \leq \sum_{i=0}^{M} n_i b^{M-i} = b^M \sum_{i=1}^{M} n_i b^{-i} = b^M \sum_{i=1}^{M} \frac{n_i}{b^i} = b^M K.$$

(b) Suppose C is prefix-free. Then a message of length M can have at most one codeword as a prefix. So the estimate in (a) is exact.

(c) By $(*)$ the numbers of message of length M is b^M. Thus $|D| \leq b^M$. By (b) we have $b^M K = |D|$, so $b^M K \leq b^M$. Hence $K \leq 1$. □

Theorem 2.3.5. *Let $b \in \mathbb{Z}^+$ and $M \in \mathbb{N}$. Let $n = (n_0, n_1, \dots, n_M)$ be a tuple of non-negative integers such that $K \leq 1$, where K is the Kraft-McMillan number of the parameter n to the base b. Then there exists a b-nary PF code C with parameter n.*

Proof. The proof is by induction on M. Let T be any set of size b.

Suppose first that $M = 0$. Then $n_0 = K \leq 1$. If $n_0 = 0$ put $C := \{\}$ and if $n_0 = 1$ put $C := \{\varnothing\}$. Then C is a PF code with parameter $n = (n_0)$.

Suppose next that $M \geq 1$ and that the theorem holds for $M - 1$ in place of M. Put $\tilde{n} := (n_0, \dots, n_{M-1})$ and let $\tilde{K}$ be the Kraft-McMillan number of the parameter $\tilde{n}$ to the base b. Then $\tilde{K} = \sum_{i=0}^{M-1} \frac{n_i}{b^i}$. Hence

$$\tilde{K} + \frac{n_M}{b^M} = K \leq 1,$$

and so

$$(*) \qquad \tilde{K} = K - \frac{n_M}{b^M} \leq 1 - \frac{n_M}{b^M} \leq 1.$$

By the induction assumption there exists a PF code $\tilde{C}$ with parameter $\tilde{n}$. Note that the codewords in $\tilde{C}$ have length at most $M-1$. Let $\tilde{D}$ be the set of messages of length M using T which have a codeword from $\tilde{C}$ as a prefix. As $\tilde{C}$ is PF, 2.3.4 shows that $|\tilde{D}| = b^M\tilde{K}$. Multiplying (*) with b^M gives $b^M\tilde{K} \leq b^M - n_M$. Thus $|\tilde{D}| \leq b^M - n_m$ and so $n_M \leq b^M - |\tilde{D}|$. Since b^M is the number of messages of length M, $b^M - |\tilde{D}|$ is the number of messages of length M which do not have a codeword from $\tilde{C}$ as a prefix. Thus there exists a set E of messages of length M using T such that $|E| = n_M$ and none of the messages has a codeword from $\tilde{C}$ as a prefix. Put $C := \tilde{C} \cup E$.

We claim that C is prefix-free. For this let a and d be distinct elements of C.

Suppose that a and d are both in $\tilde{C}$. Since $\tilde{C}$ is PF, a is not a prefix of d.

Suppose that a and d are both in E. Then a and d have the same length, namely M, and so a is not a prefix of d.

Suppose that $a \in \tilde{C}$ and $d \in E$. By choice of E no codeword of $\tilde{C}$ is a prefix of a message in E. So a is not a prefix of d.

Suppose that $a \in E$ and $b \in \tilde{C}$. Then a has length M, while b has length less than M, and so again a is not a prefix of b.

Thus C is indeed PF.

If a is a codeword of length i with $i < M$, then a is one of the n_i codewords of $\tilde{C}$ of length i. If a is a codeword of length at least M, then a is one of the n_M-codewords in E and a has length M. Thus the parameter of C is $(n_0, n_1, \ldots, n_{M-1}, n_M) = n$ and so C has all the required properties. □

Exercises 2.3:

2.3#1. Does there exist a prefix-free ternary code with the following parameter:
(a) $(0,1,3,10)$ (b) $(0,0,1,3,39)$.

2.3#2. A code is called *complete* if it is PF and $K = 1$. Show that if there exists a complete b-nary code for an alphabet of size m, then $b-1$ divides $m-1$.

2.4 A counting principle

Definition 2.4.1. *Let $c : S \to T^*$ and $d : R \to T^*$ be codes using the same alphabet T.*

(a) *Let $f : X \to Y$ be a function. Then* $\operatorname{Im} f := \{f(x) \mid x \in X\}$.

(b) *Define the function $cd : S \times R \to T^*$ by*

$$(cd)(s,r) := c(s)d(r)$$

for all $s \in S, r \in R$.

(c) *Let $A, B \subseteq T^*$. Then $AB := \{ab \mid a \in A, b \in B\} \subseteq T^*$.*

(d) *Let $r \in \mathbb{Z}^+$. Define the function $c^r : S^r \to T^*$ recursively by $c^1 := c$ and $c^{r+1} := c^r c$.*

Lemma 2.4.2. *Let $c : S \to T^*$ and $d : R \to T^*$ be codes. Let C and D be the set of codewords of c and d respectively.*

(a) $\operatorname{Im} cd = CD$ *and* $\operatorname{Im} c^r = C^r$ *for all $r \in \mathbb{Z}^+$.*

(b) *The function $cd : S \times R \to T^*$ is a code if and only for each $e \in CD$ there exist unique $a \in C$ and $b \in D$ with $e = ab$.*

(c) *If cd is a code, then CD is the set of codewords of cd.*

(d) *If c or d is regular, then cd is regular.*

Proof. (a): We have

$$\operatorname{Im} cd = \left\{(cd)(s,r) \middle| (s,r) \in S \times R\right\} = \left\{c(s)d(r) \middle| s \in S, r \in R\right\} = \left\{ab \middle| a \in C, d \in D\right\} = CD.$$

The second statement follows from the first and induction on r.

(b): cd is a code if and only if cd is 1-1, if and only if for each $e \in \operatorname{Im} cd$ there exist unique $s \in S$ and $r \in R$ with $e = (cd)(s,r)$, and if and only if for each $e \in CD$ there exist unique $s \in S$ and $r \in R$ with $e = c(s)d(r)$. Since c and d are 1-1, this holds if and only if for each $e \in CD$ there exist unique $a \in C$ and $b \in D$ with $e = ab$.

(c): The set of codewords of cd is $\operatorname{Im} cd$. So (b) follows from (a).

(d) Suppose c or d is regular. Let e be a codeword of cd. By (c) $e \in CD$ and so $e = ab$ with $a \in C$ and $b \in D$. If c is regular, then $a \neq \varnothing$ and if d is regular, $b \neq \varnothing$. In either case $ab \neq \varnothing$, so $e \neq \varnothing$ and cd is regular. □

Definition 2.4.3. *Let c be a code with set of codewords C and parameter $(n_0, n_1, \ldots, n_M)$. Then*

$$Q_c(x) = n_0 + n_1 x + n_2 x^2 + \ldots + n_M x^M$$

is called the generating function of c. Note that $Q_c(x)$ only depends on C. So we will also write $Q_C(x)$ for $Q_c(x)$.

Example 2.4.4. Compute the generating function of the code $C = \{01, 10, 110, 1110, 1101\}$.

$$n_0 = 0, \quad n_1 = 0, \quad n_2 = 2, \quad n_3 = 1, \quad \text{and} \quad n_4 = 2.$$

So

$$Q_C(x) = 2x^2 + x^3 + 2x^4.$$

Theorem 2.4.5 (The Counting Principle)**.**

(a) *Let c and d be codes using the same alphabet T such that cd is a code. Then*

$$Q_{cd}(x) = Q_c(x)Q_d(x)$$

(b) *Let c be a UD-code. Then c^r is a code and*

$$Q_{c^r}(x) = Q_c^r(x).$$

Proof. (a): Let $(n_0, \dots, n_M), (p_0, \dots, p_U)$ and $(q_0, \dots, q_V)$ be the parameters of c, d and cd respectively.

Let $i \in \mathbb{N}$ and $e \in CD$. Since cd is a code, (2.4.2)(b) shows that $e = ab$ for unique $a \in C$ and $b \in D$. Then e has length i if and only if a has length j for some $j \in \mathbb{N}$ with $j \leq i$ and b has length $i - j$. For a given j, there are n_j choices for a and p_{i-j} choices for b. So

$$q_i = n_0 p_i + n_1 p_{i-1} + n_2 p_{i-2} + \ldots + n_{i-1} p_1 + n_i p_0.$$

Note that this is exactly the coefficient of x^i in $Q_c(x)Q_d(x)$ and so (a) is proved.

(b): Since c is a UD code the extended function $c^* : S^* \to T^*$ is 1-1. Note that

$$c^*(s_1 \dots s_r) = c(s_1) \dots c(s_r) = c^r(s_1, s_2, \dots, s_r).$$

Hence c^r is just the restriction of c^* to S^r. As c^* is 1-1, also c^r is 1-1. Thus c^r is a code. Applying (a) $r - 1$ times gives

$$Q_{c^r}(x) = Q_{\underbrace{cc\dots c}_{r-\text{times}}}(x) = \underbrace{Q_c(x)Q_c(x)\dots Q_c(x)}_{r-\text{times}} = Q_c^r(x).$$

□

Example 2.4.6. Let c be UD-code with parameter $(0, 1, 0, 2)$. Compute the generating function and the parameter of c^3.

The generating function of C is $Q_c(x) = 0 + 1x + 0x^2 + 2x^3 = x + 2x^3$. Hence

$$Q_{c^3}(x) = Q_c^3(x) = (x + 2x^3)^3 = x^3 + 3 \cdot x^2 \cdot 2x^3 + 3 \cdot x \cdot (2x^3)^2 + (2x^3)^3 = x^3 + 6x^5 + 12x^7 + 8x^9.$$

The parameter of c^3 is formed by the coefficients of $Q_{c^3}(x)$. So the parameter of c^3 is

$$(0, 0, 0, 1, 0, 6, 0, 12, 0, 8).$$

Exercises 2.4:

2.4#1. Let $S = \{a, b, c, d, e, f, g\}$ and let $c : S \to \mathbb{B}^*$ be the binary code with

$$a \mapsto 00, \quad b \mapsto 010, \quad c \mapsto 011, \quad d \mapsto 1000, \quad e \mapsto 1001, \quad f \mapsto 1101, \quad g \mapsto 1111.$$

For a positive integer i define $Q_i(x) = Q_{c^i}(x)$.

(a) Write down $Q_1(x)$ and then compute $Q_2(x)$ and $Q_3(x)$.

(b) What do the coefficients of x^7 in $Q_2(x)$ and $Q_3(x)$ represent? Verify your answers by making a list of the corresponding messages in S^*.

2.5 Unique decodability implies $K \le 1$

Lemma 2.5.1. *Let c be a b-nary code with maximal codeword length M and Kraft-McMillan number K. Then*

(a) *$K \le M + 1$ and if c is regular, $K \le M$.*

(b) *$K = Q_c\left(\frac{1}{b}\right)$.*

Proof. (a): Since there are b^i messages of length i we have $n_i \le b^i$ and so $\frac{n_i}{b^i} \le 1$. Thus

$$K = \sum_{i=0}^{M} \frac{n_i}{b^i} \le \sum_{i=0}^{M} 1 = M + 1.$$

If c is regular, $\varnothing$ is not a codeword and thus $n_0 = 0$. So $K \le M$ in this case.

(b):

$$K = \sum_{i=0}^{M} \frac{n_i}{b^i} = \sum_{i=0}^{M} n_i \left(\frac{1}{b}\right)^i = Q_c\left(\frac{1}{b}\right).$$

□

Lemma 2.5.2. (a) *Let $c : S \to T^*$ and $d : R \to T^*$ be codes such that cd is a code. Let K and L be the Kraft-McMillan number of c and d, respectively. Then KL is the Kraft-McMillan number of cd.*

(b) *Let $r \in \mathbb{Z}^+$ and let c be a UD-code with Kraft-McMillan number K. Then the Kraft-McMillan number of c^r is K^r.*

Proof. (a) By (2.4.5)(a) we have $Q_{cd}(x) = Q_c(x)Q_d(x)$. Also by (2.5.1)(b) the Kraft-McMillan number of cd is

$$Q_{cd}\left(\frac{1}{b}\right) = Q_c\left(\frac{1}{b}\right)Q_d\left(\frac{1}{b}\right) = KL.$$

(b) By (2.4.5)(b) we have $Q_{c^r}(x) = Q_c^r(x)$. Also by (2.5.1)(b) the Kraft-McMillan number of c^r is

$$Q_{c^r}\left(\frac{1}{b}\right) = Q_c\left(\frac{1}{b}\right)^r = K^r.$$

□

Theorem 2.5.3. *Let c be a UD-code with Kraft-McMillan number K. Then $K \le 1$.*

Proof. Let M be the maximal length of a codeword of c. Then the maximal length of a codeword of c^r is rM. By (2.5.2)(b) the Kraft-McMillan number of c^r is K^r. Since c is a UD-code, (2.1.6)(a) shows that c is regular. Hence also c^r is regular. By 2.5.1 the Kraft-McMillan number of a regular code is bounded by the maximal codeword length. Thus

$$K^r \le rM$$

for all $r \in \mathbb{Z}^+$. Applying ln to both sides gives:

$$r \ln K = \ln(K^r) \le \ln(rM) \quad \text{and so} \quad \ln K \le \frac{\ln(rM)}{r}$$

The derivative of x is 1 and of $\ln(xM)$ is $\frac{M}{xM} = \frac{1}{x}$. Thus L'Hôpital's Rule gives

$$\lim_{r\to\infty} \frac{\ln(rM+1)}{r} = \lim_{r\to\infty} \frac{\frac{1}{r}}{1} = 0.$$

Hence $\ln K \le 0$ and so $K \le 1$. □

Corollary 2.5.4. *Given a parameter $n = (n_0, \dots, n_M)$ (where $M, n_0, \dots, n_M \in \mathbb{N}$) and a base $b \in \mathbb{Z}^+$ with Kraft-McMillan number K. Then the following three statements are equivalent.*

(a) *Either $n = (1, 0, \dots, 0)$ or there exists a b-nary UD-code with parameter n.*

(b) $K \le 1$.

(c) *There exists a b-nary PF-code with parameter n.*

Proof. (a) $\Longrightarrow$ (b): If $n = (1, 0, \dots, 0)$ then $K = 1 \leq 1$. If C is UD-code with parameter $(n_0, \dots, n_M)$ then 2.5.3 shows that $K \leq 1$.

(b) $\Longrightarrow$ (c): If $K \leq 1$, then 2.3.5 shows that there exists a b-nary PF-code with parameter n.

(c) $\Longrightarrow$ (a): Suppose c is a b-nary PF-code with parameter n. If c is regular, 2.1.5 shows that c is UD and (a) holds. If c is not regular, then (2.1.6)(b) shows that $\varnothing$ is the only codeword. Hence the parameter of c is $(1, 0, \dots, 0)$ and again (a) holds. □

Chapter 3

Economical coding

3.1 Probability distributions

Definition 3.1.1. *Let S be an alphabet. Then a probabilty distribution on S is an S-tuple $p = (p_s)_{s \in S}$ with coefficients in the interval $[0,1]$ such that*

$$\sum_{s \in S} p_s = 1.$$

A probability distribution p on S is called positive if $p_s > 0$ for all $s \in S$.

Definition 3.1.2. *Let S be set. An ordering on S is a relation '$<$' on S such that for all $a, b, c \in S$:*

(i) *Exactly one of $a < b, a = b$ and $b < a$ holds; and*

(ii) *if $a < b$ and $b < c$, then $a < c$.*

An ordered alphabet is a pair $(S, <)$, where S is an alphabet and $<$ is an ordering on S.

Example 3.1.3. Suppose $S = \{s_1, s_2, \dots, s_m\}$ is an alphabet of size m. Then there exists a unique ordering on S with $s_i < s_{i+1}$ for all $1 \leq i < m$. Indeed we have $s_i < s_j$ if and only if $i < j$.

Notation 3.1.4. *We will often say that S is an ordered alphabet rather than that $(S, <)$ is an ordered alphabet. We also will say that $S = (s_1, \dots, s_m)$ is an ordered alphabet if $(S, <)$ is the order alphabet with $S = \{s_1, \dots, s_m\}$ and $s_i < s_j$ for all $1 \leq i < j \leq m$.*

Notation 3.1.5. *Suppose $S = (s_1, s_1, s_2, \dots, s_m)$ is an ordered alphabet, and that*

$$t : \quad \begin{array}{cccc} s_1 & s_2 & \dots & s_m \\ \hline t_1 & t_2 & \dots & t_m \end{array}$$

is an S-tuple.

Then we will denote t by $(t_1, \ldots, t_m)$. Note that this is slightly ambiguous, since t does not only depended on the n-tuple $(t_1, \ldots, t_m)$ but also on the ordered alphabet S.

Example 3.1.6.

(a) Verify that

$$p: \quad \begin{array}{cccc} w & x & y & z \\ \hline \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{6} \end{array}$$

is a probability distribution on $\{w, x, y, z\}$. $\frac{1}{2}, \frac{1}{3}, 0, \frac{1}{6}$ all are interval $[0, 1]$. Also

$$\frac{1}{2} + \frac{1}{3} + 0 + \frac{1}{6} = \frac{3 + 2 + 0 + 1}{6} = \frac{6}{6} = 1,$$

So p is indeed a probability distribution.

(b) Using the notation from 3.1.5 we can say that $p = \left(\frac{1}{2}, \frac{1}{3}, 0, \frac{1}{6}\right)$ is a probability distribution on the ordered alphabet (w, x, y, z) with $p_x = \frac{1}{3}$.

(c) $q = \left(\frac{1}{2}, \frac{1}{3}, 0, \frac{1}{6}\right)$ is a probability distribution on (x, w, z, y) with $q_x = \frac{1}{2}$.

(d) $(01, 111, 011, 1110)$ is the code on (c, a, d, b) with

$$a \mapsto 111, \quad b \mapsto 1110, \quad c \mapsto 01, \quad d \mapsto 1110.$$

(e)

A	B	C	D	E	F	G	H	I
8.167%	1.492%	2.782%	4.253%	12.702%	2.228%	2.015%	6.094%	6.966%

J	K	L	M	N	O	P	Q	R
0.153%	0.747%	4.025%	2.406%	6.749%	7.507%	1.929%	0.095%	5.987%

S	T	U	V	W	X	Y	Z	$\sqcup$
6.327%	9.056%	2.758%	1.037%	2.365%	0.150%	1.974%	0.074%	0

is a probability distribution on $\mathbb{A}$. It lists how frequently a letter is used in the English language.

(f) Let S be an ordered alphabet with m elements and put

$$p := \left(\frac{1}{m}\right)_{s\in S} = \left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right).$$

Then $p_s = \frac{1}{m}$ for all $s \in S$ and p is a probability distribution on S. p is called the equal probability distribution on S.

3.2 The optimization problem

Definition 3.2.1. *Let $c : S \to T^*$ be a code and p a probability distribution on S.*

(a) *For $s \in S$ let y_s be the length of the codeword $c(s)$. The S-tuple $y = (y_s)_{s\in S}$ is called the codewords length of c.*

(b) *The average codeword length of c with respect to p is the number*

$$L = p \cdot y = \sum_{s\in S} p_s y_s$$

To emphasize that L depends on p and c we will sometimes use the notations $L(c)$ and $L_p(c)$ for L

Note that the average codeword length only depends on the length of the codewords with non-zero probability. So we will often assume that p is positive.

Example 3.2.2. Compute L if $S = (s_1, s_2, s_3)$, $p = (0.2, 0.6, 0.2)$ and c is the binary code $(0, 10, 11)$.

Does there exist a code with the same codewords but smaller average codeword length?

We have $y = (1, 2, 2)$ and so

$$L = p \cdot y = (0.2, 0.6, 0.2) \cdot (1, 2, 2) = 0.2 \cdot 1 + 0.6 \cdot 2 + 0.2 \cdot 2 = 0.2 + 1.2 + 0.4 = 1.8.$$

To improve the average length, we will assign the shortest codeword, 0, to the most likeliest symbol s_2:

$$s_1 \mapsto 01, \quad s_2 \mapsto 0, \quad s_3 \mapsto 11.$$

Then $y = (2, 1, 2)$ and

$$L = p \cdot y = (0.2, 0.6, 0.2) \cdot (2, 1, 2) = 0.2 \cdot 2 + 0.6 \cdot 1 + 0.2 \cdot 2 = 0.4 + 0.6 + 0.4 = 1.4.$$

Definition 3.2.3. *Given an alphabet S, a probability distribution p on S and a class $\mathcal{C}$ of codes for S. A code c in $\mathcal{C}$ is called an optimal $\mathcal{C}$-code with respect to p if*

$$L_p(c) \le L_p(\tilde{c})$$

for all codes $\tilde{c}$ in $\mathcal{C}$.

Example 3.2.4. List some classes of codes:

binary codes
PF-codes
ternary UD-codes.
For $b \in \mathbb{Z}^+$: b-nary codes.

Example 3.2.5. Suppose $(3,3,3,4,4,6)$ is the codewords length of a binary code C. What is the parameter and the Kraft-McMillan number of C?

There are 0 codewords of length $0, 1$ or 2, there are 3 codewords of length 3, 2 of length 4, 0 of length 5 and 1 of length 6. So the parameter is

$$(0,0,0,3,2,0,1)$$

Since C is binary, $b = 2$ and so

$$\begin{aligned} K &= \frac{0}{1}+\frac{0}{2^1}+\frac{0}{2^2} \quad +\frac{3}{2^3} \quad +\frac{2}{2^4}+ \quad \frac{0}{2^5} \quad +\frac{1}{2^6} \\ &= \quad \frac{1}{2^3}+\frac{1}{2^3}+\frac{1}{2^3} \quad +\frac{1}{2^4}+\frac{1}{2^4} \quad +\frac{1}{2^6} \end{aligned}$$

Definition 3.2.6. *Let S be an alphabet, y an S-tuple with coefficients in $\mathbb{N}$ and b a positive integer.*

(a) *Let $M \in \mathbb{N}$ with $y_s \le M$ for all $s \in S$. For $i \in \mathbb{N}$ with $i \le M$ let n_i be the number of $s \in S$ with $y_s = i$. Then $(n_0, \dots, n_M)$ is called the parameter of the codewords length y.*

(b) *$\sum_{s \in S} \frac{1}{b^{y_s}}$ is called the Kraft-McMillan numbers of the codewords length y to the base b.*

Lemma 3.2.7. *Let S be an alphabet, b a positive integer and y an S-tuple with coefficients in $\mathbb{N}$. Let K be the Kraft-McMillan number to the base b and $(n_0, \dots, n_M)$ the parameter of the codewords length y.*

(a) *K is the Kraft-McMillan number of the parameter $(n_0, \dots, n_M)$ to the base b.*

(b) *Let c be a b-nary code for the set S with codewords length y. Then $(n_0, \dots, n_M)$ is the parameter of c and K is the Kraft-McMillan number of c.*

(c) *Suppose there exists a b-nary code C with parameter $(n_0, \dots, n_M)$. Then there exists a b-nary code c for S with codewords length y such that C is the set of codewords of c.*

Proof. Let $i \in \mathbb{N}$ with $i \le M$ define $S_i := \{s \in S \mid y_s = i\}$. Then for $s \in S$

$$(*) \qquad s \in S_i \iff y_s = i$$

The definition of n_i gives

$$(**) \qquad n_i = |S_i|$$

Note also that

$(***)$ $(S_0, S_1, \dots, S_M)$ *is a partition of S, that is for each $s \in S$ there exists a unique $i \in \mathbb{N}$ with $0 \le i \le M$ and $s \in S_i$ (namely $i = y_s$).*

(a) We compute

$$\begin{aligned} K &= \sum_{s \in S} \frac{1}{b^{y_s}} && - \text{definition of } K \\ &= \sum_{i=0}^{M} \sum_{s \in S_i} \frac{1}{b^{y_s}} && - (***) \\ &= \sum_{i=0}^{M} \sum_{s \in S_i} \frac{1}{b^{i}} && - y_s = i \text{ for } s \in S_i \\ &= \sum_{i=0}^{M} |S_i| \frac{1}{b^{i}} \\ &= \sum_{i=0}^{M} n_i \frac{1}{b^{i}} && - (**) \end{aligned}$$

and so K is the Kraft-McMillan number of the parameter $(n_0, \dots, n_M)$.

(b) Note that $c(s)$ has length i if and only if $y_s = i$. So n_i is the number of codewords of length i. Thus $(n_0, \dots, n_M)$ is the parameter of c. Hence by (a), K is the Kraft-McMillan number of c.

(c) Let C be a b-nary code with parameter $(n_0, \dots, n_M)$. Recall from Definition (2.3.1)(c) that

$$C_i = \{d \in C \mid \ell(d) = i\}.$$

and note that $(C_0, C_1, \dots, C_M)$ is a partition of C.

Also put

$$S_i := \{s \in S \mid y_s = i\}.$$

Since $(n_0, \ldots, n_M)$ is the parameter of C, we have $|C_i| = n_i$.

Recall from $(**)$ that $|S_i| = n_i$. Thus $|C_i| = |S_i|$ and there exists a bijection $\alpha_i : S_i \to C_i$.

Define a function $c : S \to C$ has follows: Let $s \in S$ and put $i := y_s$. By $(*)$ $s \in S_i$ and we define $c(s) := \alpha_i(s)$. As $(S_0, \ldots, S_M)$ is a partition of S and $(C_0, \ldots, C_M)$ a partition of C, we conclude that $c : S \to C$ is a bijection. Since $\alpha_i(s) \in C_i$, $\alpha_i(s)$ has length i. As $i = y_s$ this shows that y_s is the length of $c(s)$ and thus y is the codewords length of c. As $c : S \to C$ is a bijection, we know that c is 1-1 and $\operatorname{Im} c = C$. Thus c code with set of codewords C. □

Lemma 3.2.8. *Let S be an alphabet, $b \in \mathbb{Z}^+$ and y an S-tuple with coefficients in $\mathbb{N}$. Let K be the Kraft-McMillan number of the codewords length y to the base b. Then there exists a b-nary PF-code with codewords length y if and only if $K \le 1$.*

Proof. Suppose first that there exists a PF-code c with codewords length y. Then by (3.2.7)(b) K is the Kraft-McMillan number of c. Since c is PF we conclude from (2.3.4)(c) that $K \le 1$.

Suppose next that $K \le 1$ and let n be the parameter of the codewords length y. By (3.2.7)(a), K is the Kraft-McMillan number of the parameter n to the base b. Since $K \le 1$ we conclude from 2.3.5 that there exists a b-nary PF-code C with parameter n. Hence by (3.2.7)(c) there exists a code c with codewords length y and set of codewords C. As C is PF, so is c. □

Remark 3.2.9. *In view of the preceding lemma, the problem of finding the codewords length of an optimal b-nary PF-code with respect to a given probability distribution $(p_1, \ldots, p_m)$ can be restated as follows:*

Find non-negative integers $y_1, \ldots, y_m$ such that

$$p_1 y_1 + p_2 y_2 + \ldots + p_m y_m$$

is minimal subject to

$$\frac{1}{b^{y_1}} + \frac{1}{b^{y_2}} + \ldots + \frac{1}{b^{y_m}} \le 1.$$

3.3 Entropy

Theorem 3.3.1. *Let S be an alphabet, p a positive probability distribution on S and $b > 1$ (that is $b \in \mathbb{R}$ with $b > 1$). Let y be an S-tuple of real numbers with $\sum_{s \in S} \frac{1}{b^{y_s}} \le 1$.*

Then

$$\sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) \leq \sum_{s\in S} p_s y_s$$

with equality if and only if

$$y_s = \log_b\left(\frac{1}{p_s}\right) \textit{ for all } s \in S.$$

Proof. We will first show that

$(*)$ *Let* $x \in \mathbb{R}^+$*. Then* $\ln x \leq x - 1$ *with equality if and only if* $x = 1$.

Put $f = x - 1 - \ln x$. Then $f' = 1 - \frac{1}{x} = \frac{x-1}{x}$. Thus $f'(x) < 0$ for $0 < x < 1$ and $f'(x) > 0$ for $x > 1$. So f is strictly decreasing on $(0,1]$ and strictly increasing on $[1,\infty)$. It follows that $f(1) = 0$ is the minimum value for f on $(0,\infty)$ and that $x - 1 - \ln x \geq 0$ with equality if and only if $x = 1$. This gives $(*)$.

$(**)$ *Let* $s \in S$*. Then* $\frac{1}{p_s b^{y_s}} = 1 \iff \frac{1}{b^{y_s}} = p_s \iff y_s = \log_b\left(\frac{1}{p_s}\right)$.

$$\frac{1}{p_s b^{y_s}} = 1 \quad\iff\quad \frac{1}{b^{y_s}} = p_s \quad\iff\quad b^{y_s} = \frac{1}{p_s} \quad\iff\quad y_s = \log_b\left(\frac{1}{p_s}\right).$$

$(***)$ *Let* $s \in S$*. Then* $p_s \log_b\left(\frac{1}{p_s}\right) - p_s y_s \leq \frac{1}{\ln b}\left(\frac{1}{b^{y_s}} - p_s\right)$ *with equality if and only if* $y_s = \log_b\left(\frac{1}{p_s}\right)$

Using $x = \frac{1}{p_s b^{y_s}}$ in $(*)$ we get

$$\ln\left(\frac{1}{p_s b^{y_s}}\right) \leq \frac{1}{p_s b^{y_s}} - 1$$

with equality if and only if $\frac{1}{p_s b^{y_s}} = 1$. By $(**)$ the latter holds if and only if $y_s = \log_b\left(\frac{1}{p_s}\right)$.

As $\ln\left(\frac{1}{p_s b^{y_s}}\right) = \ln\left(\frac{1}{p_s}\right) - (\ln b) y_s$ we conclude that

$$\ln\left(\frac{1}{p_s}\right) - (\ln b) y_s \leq \frac{1}{p_s b^{y_s}} - 1.$$

Multiplying with $\frac{1}{\ln b} p_s$ gives

$$p_s \frac{1}{\ln b} \ln\left(\frac{1}{p_s}\right) - p_s y_s \leq \frac{1}{\ln b}\left(\frac{1}{b^{y_s}} - p_s\right)$$

and so

$$p_s \log_b\left(\frac{1}{p_s}\right) - p_s y_s \leq \frac{1}{\ln b}\left(\frac{1}{b^{y_s}} - p_s\right).$$

Thus $(***)$ holds.

Summing $(***)$ over all $s \in S$ gives

$$(+) \qquad \sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) - \sum_{s\in S} p_s y_s \le \frac{1}{\ln b}\left(\sum_{s\in S}\frac{1}{b^{y_s}} - \sum_{s\in S} p_s\right).$$

By Hypothesis $\sum_{s\in S}\frac{1}{b^{y_s}} \le 1$ and, since p is a probability distribution, $\sum_{s\in S} p_s = 1$. Hence

$$(++) \qquad \sum_{s\in S}\frac{1}{b^{y_s}} - \sum_{s\in S} p_s \le 1 - 1 = 0$$

with equality if and only if $\sum_{s\in S}\frac{1}{b^{y_s}} = 1$.

As $b > 1$ we have $\ln b > 0$. So we can divide $(++)$ by $\ln b$ and get

$$\frac{1}{\ln b}\left(\sum_{s\in S}\frac{1}{b^{y_s}} - \sum_{s\in S} p_s\right) \le 0$$

Together with $(+)$ we get

$$\sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) - \sum_{s\in S} p_s y_s \le \frac{1}{\ln b}\left(\sum_{s\in S}\frac{1}{b^{y_s}} - \sum_{s\in S} p_s\right) \le 0.$$

Thus

$$\sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) \le \sum_{s\in S} p_s y_s$$

with equality if and only if $y_s = \log_b\left(\frac{1}{p_s}\right)$ and $\sum_{s\in S}\frac{1}{b^{y_s}} = 1$ for all $s \in S$.

Suppose that $y_s = \log_b\left(\frac{1}{p_s}\right)$ for all $s \in S$. Then by $(**)$ $\frac{1}{b^{y_s}} = p_s$ and so

$$\sum_{s\in S}\frac{1}{b^{y_s}} = \sum_{s\in S} p_s = 1.$$

□

Notation 3.3.2. *Let S, I and J be sets, let $f : I \to J$ be a function and let $t = (t_s)_{s\in S}$ an S-tuple with coefficients in I. Then $f(t)$ denotes the S-tuple $\big(f(t_s)\big)_{s\in S}$ with coefficients in J.*

Example 3.3.3. Consider the 4-tuple $t = (1, 3, 4, 7)$ of real numbers. Compute $\frac{1}{t}$ and t^3.

$$\frac{1}{t} = \left(\frac{1}{1}, \frac{1}{3}, \frac{1}{4}, \frac{1}{7}\right)$$

and

$$t^3 = \left(1^3, 3^3, 4^3, 7^3\right)$$

Definition 3.3.4. *Let p be a probability distribution on the alphabet S and $b > 1$. The entropy of p to the base b is defined as*

$$H_b(p) := \sum_{\substack{s \in S \\ p_s \neq 0}} p_s \log_b \left(\frac{1}{p_s} \right).$$

If no base is mentioned, the base is assumed to be 2. $H(p)$ means $H_2(p)$ and $\log(a) = \log_2(a)$

Note that $\lim_{x \to 0} x \log_b \left(\frac{1}{x} \right) = \lim_{y \to \infty} \frac{\log_b y}{y} = 0$. So we will usually interpret the undefined expression $0 \log_b \left(\frac{1}{0} \right)$ as 0 and just write

$$H_b(p) = \sum_{s \in S} p_s \log_b \left(\frac{1}{p_s} \right)$$

Using notation 3.3.2 we have $p = (p_s)_{s \in S}$, $\log_b \left(\frac{1}{p} \right) = \left(\log_b \left(\frac{1}{p_s} \right) \right)_{s \in S}$ and

$$H_b(p) = p \cdot \log_b \left(\frac{1}{p} \right).$$

Example 3.3.5. Compute the entropy to the base 2 for $p = (0.125, 0.25, 0.5, 0.125)$.

$$\frac{1}{p} = (8, 4, 2, 8),$$

$$\log_2 \left(\frac{1}{p} \right) = (3, 2, 1, 3),$$

and

$$H_2(p) = p \cdot \log_2 \left(\frac{1}{p} \right) = 0.375 + 0.5 + 0.5 + 0.375 = 1.75.$$

Exercises 3.3:

3.3#1. What is the entropy to the base 2 of probability distribution $(0.2, 0.3, 0.2, 0.2, 0.1)$?

3.4 The Comparison Theorem

Theorem 3.4.1 (Comparison theorem)**.** *Let p and q be positive probability distributions on the alphabet S and let $b > 1$. Then*

$$H_b(p) \leq \sum_{s \in S} p_s \log_b \left(\frac{1}{q_s} \right)$$

with equality if and only if $p = q$.

Proof. For $s \in S$ define $y_s := \log_b\left(\frac{1}{q_s}\right)$. Then $b^{y_s} = \frac{1}{q_s}$. Hence $\frac{1}{b^{y_s}} = q_s$ and so

$$\sum_{s\in S} \frac{1}{b^{y_s}} = \sum_{s\in S} q_s = 1.$$

Theorem 3.3.1 now shows that

$$\sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) \le \sum_{s\in S} p_s y_s$$

with equality if and only if $y_s = \log_b\left(\frac{1}{p_s}\right)$ for all $s \in S$. Hence $H_b(p) \le \sum_{s\in S} p_s \log_b\left(\frac{1}{q_s}\right)$ with equality if and only if $\log_b\left(\frac{1}{q_s}\right) = \log_b\left(\frac{1}{p_s}\right)$ for all $s \in S$, that is if and only if $q_s = p_s$, for all $s \in S$ □

Theorem 3.4.2. *Let p be a positive probability distribution on the alphabet S with m symbols. Let $b > 1$. Then*

$$H_b(p) \le \log_b m$$

with equality if and only if p is the equal probability distribution.

Proof. Let $q := \left(\frac{1}{m}\right)_{s\in S}$ be the equal probability distribution on S. Then

$$\sum_{s\in S} p_s \log_b\left(\frac{1}{q_s}\right) = \sum_{s\in S} p_s \log_b\left(\frac{1}{\frac{1}{m}}\right) = \sum_{s\in S} p_s \log_b m = \left(\sum_{s\in S} p_s\right)\log_b m = \log_b m$$

and so by the comparison theorem

$$H_b(p) \le \log_b m$$

with equality if and only if $p = q$. □

3.4#1. Consider the probability distributions $(0.6, 0.3, 0.1)$ and $(0.5, 0.3, 0.2)$ on the alphabet $S = (a, b, c)$. Which one has larger entropy? What probability distribution on S produces the largest entropy?

3.5 Optimal codes – the fundamental theorems

Theorem 3.5.1 (Fundamental Theorem, Lower Bound). *Let p be a positive probability distribution on the alphabet S, let b be an integer with $b > 1$ and let c be a b-nary PF code for S. Then*

$$H_b(p) \le L_p(c).$$

Proof. Let K be the Kraft-McMillan number of c. Since c is PF we know that $K \le 1$, see 2.5.3. Let y be the codewords length of c. By 3.2.7 $K = \sum_{s\in S} \frac{1}{b^{y_s}}$. Thus $\sum_{s\in S} \frac{1}{b^{y_s}} \le 1$ and the conditions of 3.3.1 are fulfilled. Thus

$$\sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) \le \sum_{s\in S} p_s y_s$$

and so

$$H_b(p) \le L_p(c).$$

□

Definition 3.5.2. *Let $k \in \mathbb{R}$. Then $\lceil k \rceil$ is the smallest integer larger or equal to k. Note that $\lceil k \rceil$ is the unique integer with $k \le \lceil k \rceil < k+1$.*

By the Fundamental Theorem a code with $y = \log_b\left(\frac{1}{p}\right)$ will be optimal. Unfortunately, $\log_b\left(\frac{1}{p_s}\right)$ need not be an integer. Choosing y_s too small will cause K to be larger than 1. This suggest to choose $y_s = \left\lceil \log_b\left(\frac{1}{p_s}\right)\right\rceil$. Using notation 3.3.2 this means $y = \left\lceil \log_b\left(\frac{1}{p}\right)\right\rceil$.

Definition 3.5.3. *Let c be a b-nary code for the alphabet S with codewords length y and let p be a probability distribution on S. Then c is called a b-nary Shannon-Fano (SF) code for S with respect to p if c is a PF-code and $y_s = \left\lceil \log_b\left(\frac{1}{p_s}\right)\right\rceil$ for all $s \in S$ with $p_s \neq 0$.*

Theorem 3.5.4 (Fundamental Theorem, Upper Bound)**.** *Let S be an alphabet, p a positive probability distribution on S and $b > 1$ an integer.*

(a) *Let c be any b-nary SF code with respect to p. Then*

$$L_p(c) < H_b(p) + 1.$$

(b) *There exists b-nary SF code with respect to p.*

Proof. (a): Let c be a SF-code with codewords length y. Let $s \in S$, then the definition of an SF-code gives $y_s = \left\lceil \log_b\left(\frac{1}{p_s}\right)\right\rceil$ and so $y_s < \log_b\left(\frac{1}{p_s}\right) + 1$. Thus

$$L_p(c) = p \cdot y = \sum_{s\in S} p_s \left\lceil \log_b\left(\frac{1}{p_s}\right)\right\rceil < \sum_{s\in S} p_s \left(\log_b\left(\frac{1}{p_s}\right) + 1\right) = \sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) + \sum_{s\in S} p_s = H_b(p) + 1.$$

(b) Let $s \in S$ and define $y_s := \left\lceil \log_b\left(\frac{1}{p}\right)\right\rceil$. Then

$$(*) \qquad \log_b\left(\frac{1}{p_s}\right) \le y_s$$

and so $\frac{1}{p_s} \leq b^{y_s}$. Then $\frac{1}{b^{y_s}} \leq p_s$ and thus

$$(**) \qquad K = \sum_{s \in S} \frac{1}{b^{y_s}} \leq \sum_{s \in S} p_s = 1.$$

Hence by 3.2.8 there exists a b-nary PF-code c for S with codewords length $(y_s)_{s \in S}$. Then c is a SF-code and so (b) holds in this case. □

Example 3.5.5. Find a binary SF-code c with respect to the probability distribution $p = (0.1, 0.4, 0.2, 0.1, 0.2)$. Verify that $H_2(p) \leq L < H_2(p) + 1$.

We have

$$\frac{1}{p} = (10, 2.5, 5, 10, 5),$$

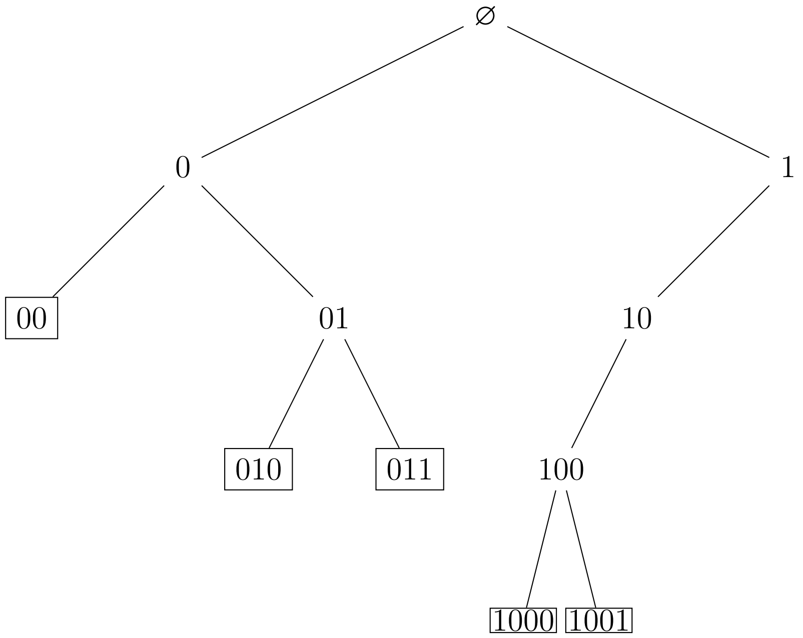
$$\log_2\left(\frac{1}{p}\right) \approx (3.3, 1.3, 2.3, 3.3, 2.3),$$

and so

$$y = \left\lceil \log_2\left(\frac{1}{p}\right) \right\rceil = (4, 2, 3, 4, 3).$$

Hence c has parameter $(0, 0, 1, 2, 2)$.

We now use the tree method to construct a code with this parameter:



Since $y = (4,2,3,4,3)$ we can choose

$$c = (1000, 00, 010, 1001, 011).$$

We have

$$\begin{aligned} H_2(p) &= p \cdot \log_2\left(\frac{1}{p}\right) \approx (0.1, 0.4, 0.2, 0.1, 0.2) \cdot (3.3, 1.3, 2.3, 3.3, 2.3) \\ &= 0.33 + 0.52 + 0.46 + 0.33 + 0.46 = 2.08 \approx 2.1 \end{aligned}$$

and

$$L = p \cdot y = (0.1, 0.4, 0.2, 0.1, 0.2) \cdot (4, 2, 3, 4, 3) = 0.4 + 0.8 + 0.6 + 0.4 + 0.6 = 2.8$$

Since $2.1 \leq 2.8 < 2.1 + 1$, the inequality $H_2(p) \leq L < H_2(p) + 1$ does indeed hold.

Note that this code is not an optimal PF-code for S with respect to p. Indeed,

$$(10, 00, 010, 11, 011).$$

is a PF-code with smaller average codeword length and

$$(010, 00, 10, 011, 11)$$

is even better. The latter code as average codeword length

$$0.1 \cdot 3 + 0.4 \cdot 2 + 0.2 \cdot 2 + 0.1 \cdot 3 + 0.2 \cdot 2 = 0.3 + 0.8 + 0.4 + 0.3 + 0.4 = 2.2$$

which is very close to the entropy of 2.1

Theorem 3.5.6 (Fundamental Theorem)**.** *Let p be a positive probability distribution on the alphabet S and c an optimal b-nary PF code for S with respect to p. Then*

$$H_b(p) \leq L_p(c) < H_b(p) + 1.$$

Proof. By 3.5.1, $H_b(p) \leq L_p(c)$.

By (3.5.4)(b) there exists an b-nary SF-code d for S with respect to p. Then (3.5.4)(a) shows that $L_p(d) < H_b(p) + 1$. Since c is an optimal b-nary PF-code for S with respect to p we have $L_p(c) \leq L_p(d)$ and so also $L_p(c) < H_b(p) + 1$. □

Exercises 3.5:

3.5#1. Construct a binary Shannon-Fano code with respect to the probability distribution $(0.5, 0.3, 0.2)$. Is this code an optimal binary PF-code?

3.5#2. Let $p = (\alpha, \beta, \gamma)$ be a probability distribution on the ordered alphabet $S = (a, b, c)$ with $\alpha > \beta > \gamma$. Show that the average codeword length of an optimal binary UD-code for S with respect to p is $2 - \alpha$.

3.6 Huffman rules

Lemma 3.6.1. *Let p be a probability distribution on the alphabet S and let $c : S \to T^*$ be an optimal b-nary PF-code for S with respect to p. Suppose $|S| \geq 2$. Let y be the codewords length of c and let M the maximal length of a codeword.*

(a) *c is a regular code and $M \geq 1$.*

(b) *If $d, e \in S$ with $y_d < y_e$, then $p_d \geq p_e$.*

Suppose in addition that p is positive. Then

(c) *Let $t \in S$ and $w \in T^*$ with $\ell(w) < \ell\big(c(t)\big)$. Then there exists $s \in S \smallsetminus \{t\}$ such that $c(s)$ is a prefix of w or w is a prefix of $c(s)$.*

(d) *Let $w \in T^*$ and suppose w is a proper prefix of a codeword. Then w is the proper prefix of at least two codewords.*

(e) *Let $w \in T^*$. Suppose that $\ell(w) < M$ and that no prefix of w is a codeword. Then w is a proper prefix of at least two codewords.*

(f) *Any parent of a codeword of length M is the parent of two codewords of length M.*

(g) *There exist two codewords of length M with a common parent.*

Proof. (a): Since c is a prefix-free code with $|S| \geq 2$ we know that c is regular, see (2.1.6)(b). Thus $\varnothing$ is not a codeword and so $M \geq 1$.

(b): Suppose for a contradiction that $p_d < p_e$. Then $p_e - p_d > 0$ and since $y_d < y_e$, we have $y_e - y_d > 0$. Hence

$$0 < (p_e - p_d)(y_e - y_d) = (p_e y_e + p_d y_d) - (p_e y_d + p_d y_e)$$

Thus $p_e y_e + p_d y_d > p_e y_d + p_d y_e$ and so interchanging the codewords for d and e gives a b-nary PF-code with smaller average codeword length. But this is a contradiction, since c is optimal.

(c) Suppose for a contradiction that for each $s \in S \smallsetminus \{t\}$ neither $c(s)$ is a prefix of w nor w is a prefix of $c(s)$. Define $d : S \to T^*$ by

$$d(s) = \begin{cases} c(s) & \text{if } s \neq t \\ w & \text{if } s = t \end{cases}$$

Let $s, r \in S$ such that $d(r)$ is a prefix of $d(s)$. We will show that $s = r$.

Suppose that $s \neq t$ and $r \neq t$. Then $c(r)$ is a prefix of $c(s)$ and since c is prefix-free, we conclude that $s = r$.

Suppose that $s = t$ and $r \neq t$. Then $c(r)$ is a prefix of w, a contradiction to initial assumption.

Suppose that $s \neq t$ and $r = t$. Then w is a prefix of $c(s)$, again a contradiction to the initial assumption.

Suppose that $s = t$ and $r = t$. Then $s = r$.

We proved that $s = r$ and so d is a PF-code. Since c is an optimal b-nary code with respect to p this gives $L_p(c) \leq L_p(d)$. Recall that p is positive, that $\ell(c(s)) = \ell(d(s))$ for $s \neq t$ and that $\ell(d(t)) = \ell(w)$. It follows that $\ell\big(c(t)\big) \leq \ell(w)$, contrary to the hypothesis of (c).

(d) Let $t \in S$ such that w is a proper prefix of $c(t)$. Then $\ell(w) < \ell\big(c(t)\big)$ and (c) shows that there exists $s \in S \smallsetminus \{t\}$ such that $c(s)$ is a prefix of w or w is a prefix of $c(s)$.

Suppose for a contradiction that $c(s)$ is a prefix of w. Since w is proper prefix of $c(t)$ we conclude that $c(s)$ is a proper prefix of $c(t)$, a contradiction as c is PF.

Thus $c(s)$ is not prefix of w. It follows that $c(s) \neq w$ and that w is prefix of $c(s)$. Thus w is a proper prefix of $c(s)$ and of $c(t)$, and (d) is proved.

(e) Since M is the maximal length of a codeword, there exists $t \in S$ with $\ell\big(c(t)\big) = M$. By assumption $\ell(w) < M$ and so $\ell(w) < \ell\big(c(t)\big)$. Hence (c) shows that there exists $s \in S \smallsetminus t$ such that $c(s)$ is a prefix of w or w is a prefix of $c(s)$. By hypothesis of (e) no codeword is a prefix of w and we conclude that w is proper prefix of $c(s)$. So we can apply (d) and conclude that (e) holds.

(f) and (g) Let u a codeword of length M. By (a) $M \geq 1$ and so u has a parent w. Then $\ell(w) = M - 1$. Note that any prefix of w is proper prefix of u. As c is PF this shows that no prefix of w is a codeword. Hence (d) shows that w is the proper prefix of two codewords b_1 and b_2. Let $i \in \{1,2\}$. Then $M - 1 = \ell(w) < \ell(b_i) \leq M$ and so $\ell(b_i) = M$ and w is the parent of b_i. In particular, w is a common parent of b_1 and b_2.

□

Theorem 3.6.2. *Let p a positive probability distribution on the alphabet S with $|S| \geq 2$. Let d, e be distinct symbols in S such that $p_d \leq p_s$ and $p_e \leq p_s$ for all $s \in S \smallsetminus \{d, e\}$. Define $\tilde{S} := S \smallsetminus \{d\}$ and let $\tilde{c}$ be a binary PF-code on $\tilde{S}$.*

Define a probability distribution $\tilde{p}$ on $\tilde{S}$ by

$$\tilde{p}_s = \begin{cases} p_d + p_e & \text{if } s = e \\ p_s & \text{if } s \neq e \end{cases} \tag{H1}$$

Also define a binary code c on S by

$$c(s) = \begin{cases} \tilde{c}(e)0 & \text{if } s = d \\ \tilde{c}(e)1 & \text{if } s = e \\ \tilde{c}(s) & \text{otherwise} \end{cases} \tag{H2}$$

Then

(a) *c is a binary prefix-free code for S.*

(b) $L_p(c) = L_{\tilde{p}}(\tilde{c}) + \tilde{p}_e$.

(c) *c is an optimal binary PF code for S with respect to p if and only if $\tilde{c}$ is a optimal binary PF code for $\tilde{S}$ with respect to $\tilde{p}$.*

Proof. (a): Let $s, t \in S$ such that $c(s)$ is a prefix of $c(t)$. We need to show that $s = t$.

Suppose first that $s \in S \smallsetminus \{d, e\}$ and $t \in S \smallsetminus \{d, e\}$. Then $\tilde{c}(s) = c(s)$ and $\tilde{c}(t) = c(t)$. Hence $\tilde{c}(s)$ is a prefix of $\tilde{c}(t)$ and since $\tilde{c}$ is prefix-free we get $s = t$.

Suppose next that $s \in S \smallsetminus \{d, e\}$ and $t \notin S \smallsetminus \{d, e\}$. Then $\tilde{c}(s) = c(s)$, $t = d$ or e, and $c(t) = \tilde{c}(e)i$ where $i = 0$ or 1. Hence $\tilde{c}(s)$ is a prefix of $\tilde{c}(e)i$. Since $\tilde{c}$ is PF and $s \neq e$ we know that $\tilde{c}(s)$ is not a prefix of $\tilde{c}(e)$. It follows that $\tilde{c}(s) = \tilde{c}(e)i$ and so $\tilde{c}(e)$ is proper prefix of $\tilde{c}(s)$, a contradiction since $\tilde{c}$ is PF.

Suppose next that $s \notin S \smallsetminus \{d, e\}$ and $t \in S \smallsetminus \{d, e\}$. Then $\tilde{c}(s) = \tilde{c}(e)i$, $i = 0$ or 1, and $c(t) = \tilde{c}(t)$. Thus $\tilde{c}(e)i$ is a prefix of $\tilde{c}(s)$. Hence $\tilde{c}(e)$ is a proper prefix of $\tilde{c}(s)$, a contradiction since $\tilde{c}$ is PF.

Suppose finally that $s \notin S \smallsetminus \{d, e\}$ and $t \notin S \smallsetminus \{d, e\}$. Then $\tilde{c}(s) = \tilde{c}(e)i$ and $c(t) = \tilde{c}(e)j$ where $i, j \in \{0, 1\}$. Hence $\tilde{c}(e)i$ is prefix of $\tilde{c}(e)j$. As $\tilde{c}(e)i$ and $\tilde{c}(e)j$ have the same length, this gives $\tilde{c}(e)i = \tilde{c}(e)j$, so $i = j$. Thus either $i = j = 0$ and $s = t = d$ or $i = j = 1$ and $s = t = e$. In either case $s = t$.

(b) Note that $\tilde{S} = (S \smallsetminus \{d, e\}) \cup \{e\}$ and $S = (S \smallsetminus \{d, e\}) \cup \{d, e\}$. Let y and $\tilde{y}$ be the codewords length of c and $\tilde{c}$, respectively. Then

$$y_s = \tilde{y}_s \quad \text{and} \quad p_s = \tilde{p}_s \quad \text{for all } s \in S \smallsetminus \{d, e\}$$

and

$$\tilde{p}_e = p_d + p_e, \qquad y_d = \tilde{y}_e + 1, \qquad y_e = \tilde{y}_e + 1$$

Hence

$$\begin{aligned}
L_p(c) &= \textstyle\sum_{s \in S} p_s y_s \\
&= p_d y_d + p_e y_e + \textstyle\sum_{s \in S \smallsetminus \{d,e\}} p_s y_s \\
&= p_d(\tilde{y}_e + 1) + p_e(\tilde{y}_e + 1) + \textstyle\sum_{s \in S \smallsetminus \{d,e\}} p_s y_s \\
&= (p_d + p_e) + (p_d + p_e)\tilde{y}_e + \textstyle\sum_{s \in S \smallsetminus \{d,e\}} p_s y_s \\
&= \tilde{p}_e + \tilde{p}_e \tilde{y}_e + \textstyle\sum_{s \in S \smallsetminus \{d,e\}} \tilde{p}_s \tilde{y}_s \\
&= \tilde{p}_e + \textstyle\sum_{s \in \tilde{S}} \tilde{p}_s \tilde{y}_s \\
&= \tilde{p}_e + L_{\tilde{p}}(\tilde{c})
\end{aligned}$$

So (b) holds.

(c) $\Longrightarrow$: Suppose c is an optimal binary PF-code with respect to p. Let $\tilde{a}$ be an optimal binary PF-code with respect to $\tilde{p}$ and let a be the binary PF-code for S constructed from $\tilde{a}$ using rule H2 with $\tilde{a}$ in place of $\tilde{c}$. As c is optimal, $L_p(c) \le L_p(a)$. By (b) applied to c and a, $L_p(c) = L_{\tilde{p}}(\tilde{c}) + \tilde{p}_e$ and $L_p(a) = L_{\tilde{p}}(\tilde{a}) + \tilde{p}_e$. As $L_p(c) \le L_p(a)$ this gives $L_{\tilde{p}}(\tilde{c}) \le L_{\tilde{p}}(\tilde{a})$. Since $\tilde{a}$ is optimal, we have $L_{\tilde{p}}(\tilde{a}) \le L_{\tilde{p}}(\tilde{c})$. Thus $L_{\tilde{p}}(\tilde{a}) = L_{\tilde{p}}(\tilde{c})$. As $\tilde{a}$ is optimal, it follows that $\tilde{c}$ is optimal.

(c) $\Longleftarrow$: Suppose $\tilde{c}$ is an optimal binary PF-code.

We will first find an optimal binary PF-code a with respect to $\tilde{p}$ such that $a(d)$ and $a(e)$ are codewords of maximal length and such $a(d)$ and $a(e)$ have a common parent. For this we start with any optimal PF-code a with respect to p and then modify a. Let M be the maximal length of a codeword of a.

Suppose that there exists $f \in \{d, e\}$ such that $\ell\big(a(f)\big) < M$. By (3.6.1)(g) there exist at least two codewords of a of length M. As $\ell\big(a(f)\big) < M$ at most one of these two codewords can be in $\{a(d), a(e)\}$, so we can choose $g \in S \setminus \{d, e\}$ with $\ell\big(a(g)\big) = M$. Then $\ell\big(a(f)\big) < \ell\big(a(g)\big)$ and (3.6.1)(b) shows that $p_g \le p_f$. By choices of e and d, $p_d \le p_g$ and $p_e \le p_g$, so $p_f \le p_g$. Thus $p_f = p_g$ and interchanging the codewords for f and g does not change the average code length of a. We therefore may and do choose the optimal code a such that both $a(d)$ and $a(e)$ have length M.

By (3.6.1)(g) there exists two codewords of a of length M with a common parent. Permuting codewords of equal length does not change the average codeword length. So we may and do choose a such that $a(d)$ and $a(e)$ have a common parent u and that

$$a(d) = u0 \qquad \text{and} \qquad a(e) = u1.$$

Let $\tilde{a}$ be the code for $\tilde{S}$ defined by

$$\tilde{a}(s) = \begin{cases} u & \text{if } s = e \\ a(s) & \text{if} s \in S \setminus \{d, e\} \end{cases}$$

We will now show that $\tilde{a}$ is PF. Since a is PF and u is a proper prefix of $a(d)$, we know that no prefix of u is codeword of a. Let w be a codeword of a with u as a prefix. Since u is not a codeword, $u \neq w$. Thus $M - 1 = \ell(u) < \ell(w) \le M$. So $\ell(w) = M$ and thus $w = u0 = a(d)$ or $w = u1 = a(e)$. Hence u is not the prefix of any $a(s), s \in S \setminus \{d, e\}$. It follows that $\tilde{a}$ is PF-code for $\tilde{S}$.

Note that a is the code constructed from $\tilde{a}$ via Rule H2. Since a is optimal, the already proven forward direction of (c) shows that $\tilde{a}$ is optimal. Since $\tilde{c}$ is optimal this gives $L_{\tilde{p}}(\tilde{a}) = L_{\tilde{p}}(\tilde{c})$. It follows that $L_{\tilde{p}}(\tilde{a}) + \tilde{p}_e = L_{\tilde{p}}(\tilde{c}) + \tilde{p}_e$. From (b) applied to a and c we conclude that $L_p(a) = L_p(c)$. Since a is optimal, this means that also c is optimal. □

Definition 3.6.3. *Let p positive probabilty distribution on alphabet S. Recursivey we define a binary code c for S to be a Huffman code for S with respect to p if either $|S| = 1$ and $c(s) = \varnothing$ for $s \in S$, or $|S| \geq 2$ and c is constructed via Huffman's Rules H1 and H2 from a Huffman code $\tilde{c}$ on $\tilde{S}$ with respect to $\tilde{p}$.*

Theorem 3.6.4. *Let c be an Huffman code for the alphabet S with respect to the positive probabilty distribution p. Then c is an optimal binary PF-code for S with respect to p.*

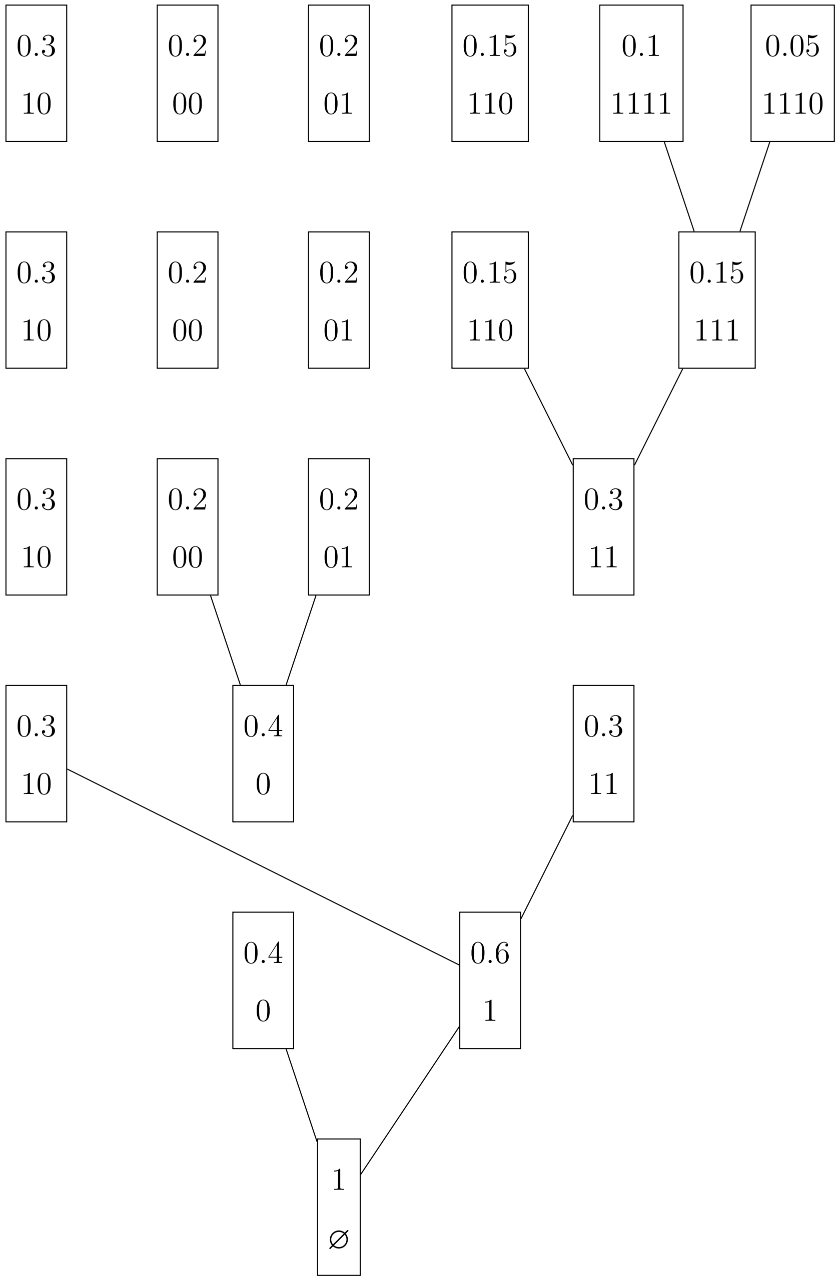
Proof. The proof is by induction on $|S|$.

Suppose $|S| = 1$ and let $s \in S$. By definition of a Huffman code, $c(s) = \varnothing$. So $\ell(c(s)) = 0$ and $L_p(c) = 0$. As $L_p(d) \geq 0$ for all codes d for S this shows that c is an optimal binary PF-code for S with respect to p.

Suppose next that $|S| \geq 2$. Then by definition of a Huffman code, c is constructed via Huffman's Rules H1 and H2 from an Huffman code $\tilde{c}$ on $\tilde{S}$ with respect tp $\tilde{p}$. By induction $\tilde{c}$ is an optimal binary code PF-for $\tilde{S}$ with respect to $\tilde{p}$. Now (3.6.2)(c) shows that c is an optimal binary PF-code for S with respect to p. □

Example 3.6.5. Find a Huffman code with respect to the probability distribution

$$(0.3, 0.2, 0.2, 0.15, 0.1, 0.05).$$



Exercises 3.6:

3.6#1. Find a Huffman code for the probability distribution $(0.4, 0.3, 0.1, 0.1, 0.06, 0.04)$.

3.6#2. While computing a Huffman code c, in each application of the Huffman rule **H1** a new probability $\tilde{p}(e)$ is introduced. Show that the average codeword length L of c is the sum of these $\tilde{p}(e)$.

3.6#3. For a code c let $\mathrm{TL}(c)$ be the sum of the lengths of the codewords and $\mathrm{M}(c)$ the maximal length of a codeword.

(a) Let $\tilde{c}$ and c be codes as in Huffman rule **H2**. Show that

$$\mathrm{M}(c) \leq \mathrm{M}(\tilde{c}) + 1 \qquad \text{and} \qquad \mathrm{TL}(c) \leq \mathrm{TL}(\tilde{c}) + M(\tilde{c}) + 2.$$

(b) If c is a Huffman code on an alphabet of size N, show that

$$\mathrm{M}(c) \leq N - 1 \qquad \text{and} \qquad \mathrm{TL}(c) \leq \frac{N^2 + N - 2}{2}.$$

Chapter 4

Data Compression

4.1 Coding in pairs

Lemma 4.1.1. *Let I be an alphabet and p an I-tuple with coefficients in $\mathbb{R}$. Then p is a probabilty distribution if and only if*

(i) $p_i \geq 0$ *for all* $i \in I$.

(ii) $\sum_{i\in I} p_i = 1$.

Proof. If p is a probabilty distribution, then p is a function from I to $[0,1]$ with $\sum_{i\in I} p_i = 1$. So (i) and (ii) holds.

Suppose now that (i) and (ii) holds. Then $p_j \geq 0$ for all $j \in I$ and so

$$p_i \leq p_i + \sum_{\substack{j\in I\\ j\neq i}} p_j = \sum_{j\in I} p_j = 1.$$

Thus $0 \leq p_i \leq 1$ and so $p_i \in [0,1]$. Hence p has coefficients in $[0,1]$ and by (ii) p is a probability distribution. □

Corollary 4.1.2. *Let I be an alphabet and p an I-tuple with coefficients in $\mathbb{R}$. Put $t := \sum_{i\in I} p_i$. Suppose that $p_i \geq 0$ for all $i \in I$ and that $t \neq 0$. Then $(\frac{p_i}{t})_{i\in I}$ is a probability distribution on I.*

Proof. Since $p_i \geq 0$ for all $i \in I$, also $t \geq 0$ and $\frac{p_i}{t} \geq 0$. We compute

$$\sum_{i\in I} \frac{p_i}{t} = \frac{\sum_{i\in I} p_i}{t} = \frac{t}{t} = 1$$

and so by 4.1.1, $(\frac{p_i}{t})_{i\in I}$ is a probability distribution on I. □

Definition 4.1.3. *Let I and J be alphabets and f an $I \times J$-matrix with coefficients in $\mathbb{R}$. Define the I-tuple f' by*

$$f'_i := \sum_{j \in J} f_{ij}$$

for all $i \in I$ and the J-tuple f'' by

$$f''_j := \sum_{i \in I} f_{ij}$$

for all $j \in J$. (So $f' = \sum_{j \in J} \mathrm{Col}_j(f)$ is the sum of the columns of f, while $f'' = \sum_{i \in I} \mathrm{Row}_i(f)$ is the sum of the rows of f.)

Then f' is called the (first) marginal tuple of f on I. f'' is called the (second) marginal tuple of f on J.

Example 4.1.4. Compute the marginal tuples of

f	a	b	c
d	0.1	0.2	0.3
e	0.2	0.1	0.1

f	a	b	c	f'
d	0.1	0.2	0.3	0.6
e	0.2	0.1	0.1	0.4
f''	0.3	0.3	0.4	

Lemma 4.1.5. *Let I and J be alphabets and let p be an $I \times J$-matrix with non-negative real coefficients and marginal tuples p' and p''. Then p is a probability distribution if and only if p' is a probability distribution and if and only if p'' is a probability distribution.*

Proof. Since $p_{ij} \geq 0$ we also have $p'_i \geq 0$ and $p''_j \geq 0$ for all $i \in I, j \in J$. We compute

$$\sum_{i \in I} p'_i = \sum_{i \in I} \left(\sum_{j \in J} p_{ij} \right) = \sum_{(i,j) \in I \times J} p_{ij} = \sum_{j \in J} \left(\sum_{i \in I} p_{ij} \right) = \sum_{j \in J} p''_j . \tag{$*$}$$

Suppose one p, p' and p'' is a probability distributions. Then one of the sums in $(*)$ is equal to 1, and so all of the the sums are equal to 1. Hence by 4.1.1 each of p, p', p'' is a probability distribution. □

Definition 4.1.6. *Let I and J be alphabets.*

(a) *Let f' and f'' be I- and J-tuples, respectively, with coefficients in $\mathbb{R}$. Then $f' \otimes f''$ is the $I \times J$-matrix defined by*

$$(f' \otimes f'')_{ij} = f'_i f''_j.$$

for all $i \in I, j \in J$.

(b) *Let p be a probability distribution on $I \times J$ with marginal distribution p' and p''. Then p' and p'' are called independent with respect to p if*

$$p = p' \otimes p'',$$

that is $p_{ij} = p'_i p''_j$ for all $i \in I, j \in J$.

Example 4.1.7. (a) Let $f' = \begin{array}{cc} d & e \\ \hline 0.6 & 0.4 \end{array}$ and $f'' = \begin{array}{ccc} a & b & c \\ \hline 0.3 & 0.3 & 0.4 \end{array}$. Compute $f' \otimes f''$:

$f' \otimes f''$	a	b	c	f'
d	0.18	0.18	0.24	0.6
e	0.12	0.12	0.16	0.4
f''	0.3	0.3	0.4	

(b) Consider

p	a	b	c
d	0.1	0.2	0.3
e	0.2	0.1	0.1

Are p' and p'' independent with respect to p?

We first compute p' and p'':

p	a	b	c	p'
d	0.1	0.2	0.3	0.6
e	0.2	0.1	0.1	0.4
p''	0.3	0.3	0.4	

We have

$$p'_d \cdot p''_a = 0.3 \cdot 0.6 = 0.18 \neq 0.1 = p_{da}$$

and so p' and p'' are not independent with respect to p.

Lemma 4.1.8. *Let p' and p'' be probability distributions on I and J, respectively. Then:*

(a) *p' and p'' are the marginal tuples of $p' \otimes p''$.*

(b) *$p' \otimes p''$ is a probability distribution on $I \times J$.*

(c) *p' and p'' are independent with respect to $p' \otimes p''$.*

Proof. (a) We have

$$\sum_{j \in J} (p' \otimes p'')_{ij} = \sum_{j \in J} p'_i p''_j = p'_i \left(\sum_{j \in J} p''_j \right) = p'_i \cdot 1 = p'_i$$

and so p' is the marginal tuple of $p' \otimes p''$ on I. Similarly, p'' is the marginal tuple of $p' \otimes p''$ on J.

(b) By (a) the marginal tuples of $p' \otimes p''$ are p' and p'' and so are probability distributions. Hence by 4.1.5 also $p' \otimes p''$ is a probability distribution.

(c) In view of (a), this immediately from the definition of independent. □

Theorem 4.1.9. *Let I and J be alphabets, let $b > 1$ and let p be a positive probability distribution on $I \times J$ with marginal distributions p' and p''. Then*

$$H_b(p) \le H_b(p') + H_b(p'')$$

with equality if and only if p' and p'' are independent with respect to p.

Proof. We have

$$\begin{aligned}
H_b(p') + H_b(p'') &= \sum_{i \in I} p'_i \log_b \left(\frac{1}{p'_i} \right) + \sum_{j \in J} p''_j \log_b \left(\frac{1}{p''_j} \right) \\
&= \sum_{i \in I} \left(\sum_{j \in J} p_{ij} \right) \log_b \left(\frac{1}{p'_i} \right) + \sum_{j \in J} \left(\sum_{i \in I} p_{ij} \right) \log_b \left(\frac{1}{p''_j} \right) \\
&= \sum_{i \in I, j \in J} p_{ij} \left(\log_b \left(\frac{1}{p'_i} \right) + \log_b \left(\frac{1}{p''_j} \right) \right) \\
&= \sum_{i \in I, j \in J} p_{ij} \log_b \left(\frac{1}{p'_i p''_j} \right) \\
&= \sum_{s \in I \times J} p_s \log_b \left(\frac{1}{(p' \otimes p'')_s} \right)
\end{aligned}$$

Thus the comparison theorem applied with $q := p' \otimes p''$ shows that $H_b(p) \le H_b(p') + H_b(p'')$ with equality if and only if $p = p' \otimes p''$, that is if and only if p' and p'' are independent with respect to p. □

4.2 Coding in blocks

Definition 4.2.1. *A source is a pair (S, P), where S is an alphabet and P is a function*

$$P : S^* \to [0, 1],$$

such that

(i) $P(\varnothing) = 1$, *and*

(ii) *for all $a \in S^*$:*

$$P(a) = \sum_{s \in S} P(as)$$

We interpret a source as a device which emits an infinite stream $\xi_1\xi_2 \dots \xi_n \dots$ of symbols from S. $P(a_1a_2 \dots a_n)$ is the probability that $\xi_1 = a_1, \xi_2 = a_2, \dots, \xi_{n-1} = a_{n-1}$ and $\xi_n = a_n$.

Notation 4.2.2. *Let (S, P) be a source and let $r \in \mathbb{N}$.*

(a) *p^r is the restriction of P to S^r, so p^r is the function from S^r to the interval $[0, 1]$ with $p^r(a) = P(a)$ for all $a \in S^r$.*

(b) *$p = p^1$, so p is the restriction of P to S.*

(c) *P^r is the restriction of P to $(S^r)^*$. Note here that if $x = x_1x_2 \dots x_n$ is a string of length n using the alphabet S^r, then each x_i is a string $x_{i1}x_{i2} \dots x_{ir}$ of length r using the alphabet S. So*

$$x = x_{11} \dots x_{1r}x_{21} \dots x_{2r} \dots x_{n1} \dots x_{nr}$$

is a string of length nr using the alphabet S. Hence $(S^r)^n = S^{rn}$ and

$$(S^r)^* = \bigcup_{n=0}^{\infty} (S^r)^n = \bigcup_{n=0}^{\infty} S^{nr} \subseteq S^*.$$

(d) *Let $l = (l_1, l_2, \dots l_r)$ be a strictly increasing r-tuple of positive integers, that is $l_i \in \mathbb{Z}^+$ and $l_1 < l_2 < \dots < l_r$. Let $u \in \mathbb{Z}$ with $u \geq l_r$. For $y = y_1y_2 \dots y_u \in S^u$ define*

$$y_l := y_{l_1}y_{l_2} \dots y_{l_r}$$

and note that $y_l \in S^r$. Define the function p^l from S^r to $\mathbb{R}$ via

$$p^l(x) = \sum_{\substack{y \in S^{l_r} \\ y_l = x}} P(y)$$

for all $x \in S^r$. So

$$p^{(l_1,\dots,l_r)}(x_1x_2\dots x_r) = \sum_{\substack{y_1y_2\dots y_{l_r}\in S^{l_r} \\ y_{l_1}=x_1, y_{l_2}=x_2,\dots,y_{l_r}=x_r}} P(y_1y_2\dots y_u)$$

We interpret $p^{(l_1,l_2,\dots,l_r)}(x_1x_2\dots x_r)$ as the probability that $\xi_{l_1} = x_1, \xi_{l_2} = x_2, \dots, \xi_{l_r} = x_r$.

Example 4.2.3. Suppose $(\mathbb{B}, P)$ is a binary source with

$$\begin{array}{llll} P(000) = 0.1 & P(001) = 0.05 & P(010) = 0.2 & P(011) = 0.1 \\ P(100) = 0.25 & P(101) = 0.1 & P(110) = 0.15 & P(111) = 0.05 \end{array}$$

(1) If $y = 110$ and $l = (1,3)$, compute y_l.

Since $y_1 = 1$ and $y_3 = 0$, $y_l = 10$.

(2) Compute $p^{(1,3)}(01)$.

There are two $y \in \mathbb{B}^3$ with $y_1 = 0$ and $y_3 = 1$, namely $y = 001$ and $y = 011$. So by definition of $p^{(1,3)}$:

$$p^{(1,3)}(01) = P(001) + P(011) = 0.05 + 0.1 = 0.15.$$

(3) Compute $P(11)$.

By Condition (4.2.1)(ii) in the definition of a source, $P(a) = \sum_{s\in S} P(as)$ and so

$$P(11) = P(110) + P(111) = 0.1 + 0.15 = 0.2.$$

Lemma 4.2.4. *Let (S,P) be a source, $r \in \mathbb{N}$ and $l = (l_1,\dots,l_r)$ be a strictly increasing r-tuple of positive integers.*

(a) *Let $a \in S^*$. Then*

$$P(a) = \sum_{x\in S^r} P(ax).$$

(b) *$\sum_{x\in S^r} P(x) = 1$ and p^r is a probability distribution on S^r.*

(c) *p is a probability distribution on S.*

(d) *Let t be integer with $t+r \geq l_r$ and let $k = (k_1, \dots, k_t)$ be the increasing t-tuple of positive integers with $\{1, \dots, t+r\} = \{l_1, \dots, l_r, k_1, \dots, k_t\}$. Define*

$$\mu : \qquad S^{t+r} \to S^t \times S^r, \ d \mapsto (d_k, d_l)$$

Then μ is a bijection, and, after identifying S^{t+r} with $S^t \times S^r$ via μ, p^l is the marginal tuple of p^{t+r} on S^r.

(e) *p^l is a probability distribution on S^r.*

Proof. (a) We will prove (a) by induction on r. If $r = 0$, then the empty message $\varnothing$ is the only element of S^r. Hence

$$\sum_{x \in S^r} P(ax) = P(a\varnothing) = P(a)$$

and (a) holds in this case. Now suppose that (a) holds for r. Since every $x \in S^{r+1}$ can be uniquely written as $x = ys$ with $y \in S^r$ and $s \in S$ we get

$$\begin{aligned}
\sum_{x \in S^{r+1}} P(ax) &= \sum_{y \in S^r, s \in S} P\big(a(ys)\big) \\
&= \sum_{y \in S^r} \left(\sum_{s \in S} P\big((ay)s\big) \right) && \text{– Commutative Law} \\
&= \sum_{y \in S^r} P(ay) && \text{– Definition of the source} \\
&= P(a) && \text{– Induction hypothesis}
\end{aligned}$$

(b) Since P is a function from S^* to $[0,1]$, and since p^r is the restriction of S^* to S^r we know that p^r is a function from S^r to $[0,1]$. By (a) applied with $a = \varnothing$ we have

$$P(\varnothing) = \sum_{x \in S^r} P(\varnothing x).$$

By definition of a source $P(\varnothing) = 1$. Also $\varnothing x = x$ and so

$$1 = \sum_{x \in S^r} P(x) = \sum_{x \in S^r} p^r(x).$$

Hence p^r is a probability distribution on S^r.

(c) This is the special case $r = 1$ in (b).

(d) Before starting the proof of (d), let's consider an example. Suppose $l = (2,3,7)$ and $t = 4$. Then $r = 3$, $t + r = 7$, $k = (1,4,5,6)$ and

$$\mu : \quad S^7 \to S^4 \times S^3, \quad s_1s_2s_3s_4s_5s_6s_7 \mapsto (s_1s_4s_5s_6, s_2s_3s_7).$$

We now start the proof of (d): Let $a = a_1 \dots a_t \in S^t$ and $b = b_1 \dots b_r \in S^r$. Define $d \in S^{t+r}$ by $d_i = a_j$ if $i = k_j$ for some $1 \le j \le t$ and $d_i = b_j$ if $i = l_j$ for some $1 \le j \le r$. Then $d_k = a$ and $d_l = b$. So the function

$$\rho : \qquad S^t \times S^r \to S^{t+r}, (a,b) \mapsto d$$

is inverse to function

$$\mu : \qquad S^{t+r} \to (S^t, S^r), d \mapsto (d_k, d_l).$$

Hence ρ and μ are bijections. Put $m := t + r$. By the hypothesis of (d) $m \ge l_r$ and so any $y \in S^m$ can be uniquely written as wu with $w \in S^{l_r}$ and $u \in S^{m-l_r}$. Moreover, $y_l = w_l$. Thus for $x \in S^r$:

$$\begin{aligned} p^l(x) &= \sum_{\substack{w \in S^{l_r} \\ w_l = x}} P(w) && \text{– Definition of } p^l \\ &= \sum_{\substack{w \in S^{l_r} \\ w_l = x}} \sum_{u \in S^{m-l_r}} P(wu) && \text{– (a)} \\ &= \sum_{\substack{y \in S^m \\ y_l = x}} P(y) && \text{– Substitution } y = wu \end{aligned}$$

Let $y \in S^{t+r}$ and $x \in S^r$. Since ρ is a bijection with inverse μ there exist unique $a \in S^t$ and $b \in S^r$ with $y = \rho(a,b)$, namely $a = y_k$ and $b = y_l$. Hence $y_l = x$ if and only if $b = x$ and so if and only if $y = \rho(a,x)$ for some $a \in S^t$. Thus

$$p^l(x) = \sum_{\substack{y \in S^m \\ y_l = x}} P(y) = \sum_{a \in S^t} P\big(\rho(a,x)\big) = \sum_{a \in S^t} p^{t+r}\big(\rho(a,x)\big)$$

So p^l is the marginal tuple of p^{t+r} on S^r.

(e): By (b), p^{t+r} is a probability distribution on S^{t+r} and by (d), p^l is the marginal tuple of p^{t+r} on S^r. So by 4.1.5 p^l is a probability distribution on S^r. □

Lemma 4.2.5. *Let (S,P) be a source and $r \in \mathbb{N}$. Then (S^r, P^r) is a source.*

Proof. Let $a \in (S^r)^*$. Then

$$P^r(a) = P(a) \overset{(4.2.4)(a)}{=} \sum_{b \in S^r} P(ab) = \sum_{b \in S^r} P^r(ab).$$

Moreover, $P^r(\varnothing) = P(\varnothing) = 1$ and so (S^r, P^r) is a source. □

4.3 Memoryless Sources

Definition 4.3.1. *A source (S, P) is called memoryless if*

$$P(as) = P(a)P(s)$$

for all $a \in S^*$ and $s \in S$.

Lemma 4.3.2. *Let p be a probability distribution on the alphabet S. Define the function*

$$P : \quad S^* \to [0,1]$$

recursively by

$$P(\varnothing) = 1$$

and

$$P(as) = P(a)p_s$$

for all $a \in S^*, s \in S$. So

$$P(s_1 \ldots s_n) = p_{s_1} p_{s_2} \ldots p_{s_n}$$

for any $s_1, s_2, \ldots, s_n \in S$.

Then (S, P) is the unique memoryless source with $P(s) = p_s$ for all $s \in S$.

Proof. Since $0 \leq p_s \leq 1$ for all $s \in S$, also $0 \leq P(a) \leq 1$ for all $a \in S^*$. So P is indeed a function from S^* to $[0,1]$. By definition of P, $P(\varnothing) = 1$. Let $a \in S^*$. Then

$$\sum_{s \in S} P(as) = \sum_{s \in S} P(a)p_s = P(a) \sum_{s \in S} p_s = P(a)1 = P(a)$$

so P is source. Let $s \in S$. Then

$$P(s) = P(\varnothing s) = P(\varnothing)p_s = 1p_s = p_s$$

and so

$$P(as) = P(a)p_s = P(a)P(s).$$

Thus P is indeed memoryless.

Now let Q be any memoryless source with $Q(s) = p_s$ for all $s \in S$. We need to prove that $Q(a) = P(a)$ for all $a \in S^*$. The proof is by induction on the length n of a. If $n = 0$, then $a = \varnothing$ and $Q(a) = 1 = P(a)$. Suppose now $Q(a) = P(a)$ holds for all messages a of length n and let b be a message of length $n + 1$. Then $b = as$ for some message a of length n and some $s \in S$. Thus

$$Q(b) = Q(as) = Q(a)Q(s) = P(a)p_s = P(b)$$

Thus $Q = P$ and Q is uniquely determined by p. □

Lemma 4.3.3. *Let (S, P) be a memoryless source and let $r, t \in \mathbb{N}$. Then*

(a) $p^{t+r} = p^t \otimes p^r$.

(b) p^t *and* p^r *are the marginal distributions of* p^{t+r} *on* S^t *and* S^r*, respectively.*

(c) p^t *and* p^r *are independent with respect to* p^{t+r}.

Proof. (a) Let

$$(*) \qquad a = s_1 \dots s_t \in S^t \qquad \text{and} \qquad b = s_{t+1} \dots s_{t+r} \in S^r$$

Then

$$\begin{aligned}
p^{t+r}(ab) &= P(ab) && - \text{definition of } p^{t+r}\\
&= P(s_1 \dots s_t s_{t+1} \dots s_{t+r}) && - \ (*)\\
&= p_{s_1} \dots p_{s_t} p_{s_{t+1}} \dots p_{t+r} && - 4.3.2\\
&= \big(p_{s_1} \dots p_{s_t}\big)\big(p_{s_{t+1}} \dots p_{t+r}\big) && \\
&= P(s_1 \dots s_t) P(s_{t+1} \dots s_{t+r}) && - 4.3.2\\
&= P(a)P(b) && - (*)\\
&= p^t(a) p^r(b) && - \text{definition of } p^t \text{ and } p^r\\
&= (p^t \otimes p^s)(ab) && - \text{definition of } p^t \otimes p^s
\end{aligned}$$

Hence $p^{t+r} = p^t \otimes p^r$.

Since $p^{t+r} = p^t \otimes p^r$, we conclude from 4.1.8 that (b) and (c) hold. □

Exercises 4.3:

4.3#1. Let $(\mathbb{B}, P)$ be a source. Suppose that

(i) $P(000) = \frac{1}{12}$;

(ii) $p^{(l)}(0) = p^{(l)}(1)$ for all integers l with $1 \le l \le 3$.

(iii) whenever (l_1, l_2) is a pair of integers with $1 \le l_1 < l_2 \le 3$, then $p^{(l_1)}$ and $p^{(l_2)}$ are independent with respect to $p^{(l_1, l_2)}$.

(a) Compute $p^{(l)}$ for $1 \le l \le 3$.

(b) Compute $p^{(l_1, l_2)}$ for $1 \le l_1 < l_2 \le 3$.

(c) Show that, for $x_1, x_2, x_3 \in \mathbb{B}$,

$$P(x_1x_2x_3) = \begin{cases} \frac{1}{12} & \text{if } x_1 + x_2 + x_3 \text{ is even} \\ \frac{1}{6} & \text{if } x_1 + x_2 + x_3 \text{ is odd} \end{cases}$$

(d) Show that $p^{(1)}$ and $p^{(2,3)}$ are not independent with respect to p^3.

(e) Show that $(\mathbb{B}, P)$ is not memoryless.

4.4 Entropy of a source

Definition 4.4.1. *A source (S, P) is called stationary if*

$$p^r = p^{(t+1,\ldots,t+r)}$$

for all $r, t \in \mathbb{N}$.

Since $p^r = p^{(1,\ldots,r)}$ we can rewrite the conditions on a stationary source as follows:

$$p^{(1,\ldots,r)} = p^{(t+1,\ldots,t+r)}$$

Intuitively, this means that the probability of a string $s_1 \ldots s_r$ to appear at the positions $1, 2 \ldots, r$ of the infinite stream $\xi_1\xi_2 \ldots \xi_n \ldots$ is the same as the probability of the string to appear at positions $t+1, t+2, \ldots, t+r$.

Lemma 4.4.2. *Let (S, P) be a source. Let $r, t \in \mathbb{N}$.*

(a) *p^t and $p^{(t+1,\ldots,t+r)}$ are the marginal distributions of p^{t+r} on S^t and S^r, respectively.*

(b) *Suppose P is stationary. Then p^t and p^r are the marginal distributions of p^{t+r} on S^t and S^r, respectively.*

Proof. (a) These are the special cases $l = (1, \ldots, t)$ and $l = (t+1, t+2, \ldots, t+r)$ in (4.2.4)(d).

(b) Since P is stationary, we have $p^r = p^{(t+1,\ldots,t+r)}$ and so (a) implies (b). □

Lemma 4.4.3. *Any memoryless source is stationary.*

Proof. Let (S, P) be a memoryless source and $r, t \in \mathbb{N}$. By 4.3.3 we know that p^r is the marginal distribution of p^{t+r} on S^r. By (4.4.2)(a) also $p^{(t+1,\ldots,t+r)}$ is the marginal distribution of p^{t+r} on S^r. Thus $p^r = p^{(t+1,\ldots,t+r)}$ and so P is stationary. □

Theorem 4.4.4. *Let (S, P) be a source, let $r, t \in \mathbb{Z}^+$ and let $b > 1$.*

(a)
$$H_b(p^{t+r}) \leq H_b(p^t) + H_b(p^{(t+1,\dots,t+r)})$$
with equality if and only if p^t and $p^{(t+1,\dots,t+r)}$ are independent with respect to p^{t+r}.

(b) *Suppose P is stationary. Then*
$$H_b(p^{t+r}) \leq H_b(p^t) + H_b(p^r)$$
with equality if and only if p^t and p^r are independent with respect to p^{t+r}.

(c) *Suppose P is stationary and let $q \in \mathbb{Z}^+$. Then $H_b(p^{qr}) \leq qH_b(p^r)$ with equality if (S,P) is memoryless*

(d) *Suppose (S,P) is stationary and that r divides t. Then*
$$\frac{H_b(p^t)}{t} \leq \frac{H_b(p^r)}{r}.$$

(e) *Supppose (S,P) is memoryless. Then*
$$\frac{H_b(p^t)}{t} = H_b(p).$$

Proof. (a): By (4.4.2)(b) p^r and $p^{(t+1,\dots,p^{t+r})}$ are the marginal distributions of p^{t+r} on S^t and S^r. Thus (a) follows from 4.1.9.

(b): Since P is stationary, $p^r = p^{(t+1,\dots,t+r)}$. So (b) follows from (a).

(c): For $q = 1$ (c) is obviously true. Suppose now that (c) holds for q. Then by (b)
$$H_b(p^{(q+1)r}) = H_b(p^{qr+r}) \overset{\text{(b)}}{\leq} H_b(p^{qr}) + H_b(p^r) \overset{\text{Ind}}{\leq} qH_b(p^r) + H_b(p^r) = (q+1)H_b(p^r)$$

Suppose (S,P) is memoryless. By 4.3.3 p^{qr} and p^r are independent with respect to p^{qr+r} and so by (b) the first inequality is an equality. By induction also the second inequality is an equality.

So (c) holds for $q+1$ and hence by the Principle Of Induction, for all $q \in \mathbb{Z}$.

(d) Since r divides t, $t = qr$ for some $q \in \mathbb{Z}^+$. Thus
$$\frac{H_b(p^t)}{t} = \frac{H_b(p^{qr})}{qr} \overset{\text{(c)}}{\leq} \frac{qH_b(p^r)}{qr} = \frac{H_b(p^r)}{r}.$$

(e) Suppose (S,P) is memoryless. By 4.4.3 (S,P) is stationary. Hence by (c) applied with $q = t$ and $r = 1$ we get
$$H_b(p^t) = H_b(p^{t\cdot 1}) = tH_b(p^1) = tH_b(p).$$
Thus $\frac{H_b(p^t)}{t} = H_b(p)$. □

Definition 4.4.5. (a) *Let A be a set of real numbers and $m \in \mathbb{R}$. Then we say that m is the infimum of A*[1] *and write $m = \inf A$ if*

(i) *$m \leq a$ for all $a \in A$, and*

(ii) *for all $k \in \mathbb{R}$ with $m < k$ there exists $a \in S$ with $a < k$.*

(b) *Let $a = (a_i)_{i=1}^{\infty}$ be infinite sequence of real numbers. Then*

$$\liminf a := \liminf_{m\to\infty} a_m := \lim_{m\to\infty} \left(\inf_{n\geq m} a_n \right).$$

Example 4.4.6. Let

$$A := \{1,4,2,5,7,8,-3,-5,-1\}, \quad \mathbb{R}^+ := \{r \in \mathbb{R} \mid r > 0\}, \quad b = (2, \frac{1}{2}, 3, \frac{1}{3}, 4, \frac{1}{4}, 5, \frac{1}{5}, 6, \frac{1}{6}, \dots)$$

Compute $\inf A$, $\inf \mathbb{R}^+$ and $\liminf b$.

$$\inf A = -5, \quad \inf \mathbb{R}^+ = 0, \quad \text{and} \quad \liminf b = 0.$$

Definition 4.4.7. *Let (S,P) be a source and $b > 1$. Then the entropy of P to the base b is the real number*

$$H_b(P) := \liminf_{m\to\infty} \frac{H_b(p^m)}{m}.$$

Lemma 4.4.8. *Let (S,P) be a memoryless source and $b > 1$. Then $H_b(P) = H_b(p)$.*

Proof. By (4.4.4)(e) we have $\frac{H_b(p^n)}{n} = H_b(p)$ for all $n \in \mathbb{Z}^+$. Thus $\inf_{n\geq m} \frac{H_b(p^n)}{n} = H_b(p)$ and

$$H_b(P) = \lim_{m\to\infty} \left(\inf_{n\geq m} H_b(p^n) \right) = \lim_{m\to\infty} H_b(p) = H_b(p).$$

□

Lemma 4.4.9. *Let (S,P) be a stationary source and $b > 1$. Then*

$$H_b(P) = \inf_{n\geq 1} \frac{H_b(p^n)}{n}.$$

[1] m is also called the largest lower bound of A

Proof. Let $m \in \mathbb{Z}^+$. We will first show that

$$(*) \qquad \inf_{k \geq m} \frac{H_b(p^k)}{k} = \inf_{n \geq 1} \frac{H_b(p^n)}{n}.$$

For this let $n \in \mathbb{Z}^+$. By (4.4.4)(d) we have $\frac{H_b(p^{nm})}{nm} \leq \frac{H_b(p^n)}{n}$. Since $nm \geq m$ this gives

$$\inf_{k \geq m} \frac{H_b(p^k)}{k} \leq \frac{H_b(p^{nm})}{nm} \leq \frac{H_b(p^n)}{n}.$$

As this holds for all $n \in \mathbb{Z}^+$ we conclude that

$$\inf_{k \geq m} \frac{H_b(p^k)}{k} \leq \inf_{n \geq 1} \frac{H_b(p^n)}{n}.$$

Clearly $\inf_{k \geq m} \frac{H_b(k)}{k} \geq \inf_{n \geq 1} \frac{H_b(p^n)}{n}$ and so $(*)$ holds.
Hence

$$H_b(P) = \liminf_{m \to \infty} \frac{H_b(p^m)}{m} = \lim_{m \to \infty} \left(\inf_{k \geq m} \frac{H_b(p^k)}{k} \right) \overset{(*)}{=} \lim_{m \to \infty} \left(\inf_{n \geq 1} \frac{H_b(p^n)}{n} \right) = \inf_{n \geq 1} \frac{H_b(p^n)}{n}.$$

□

Theorem 4.4.10 (Coding Theorem for Memoryless Sources)**.** *Let (S, P) be a memoryless source, let b an integer with $b > 1$ and let $\epsilon > 0$. Let n be any integer with $n \geq \frac{1}{\epsilon}$. Then there exists a b-nary PF-code c_n for S^n such that*

$$\frac{L_{p^n}(c_n)}{n} < H_b(p) + \epsilon.$$

Proof. Note that $\frac{1}{n} \leq \epsilon$. Also since P is memoryless, $\frac{H_b(p^n)}{n} = H_b(p)$, see (4.4.4)(e).

By the Fundamental Theorem 3.5.6 there exists a b-nary PF-code c_n for S^n with $L_{p^n}(c_n) < H_b(p^n) + 1$. Then

$$\frac{L_{p^n}(c_n)}{n} < \frac{H_b(p^n) + 1}{n} = \frac{H_b(p^n)}{n} + \frac{1}{n} \leq H_b(p) + \epsilon.$$

□

Theorem 4.4.11 (Coding Theorem for Sources)**.** *Let (S, P) be a source, $b > 1$ an integer, $\epsilon > 0$ and $k > 0$. Then there exists an integer n with $n > k$ and a b-nary prefix-free code c_n for S^n such that*

$$\frac{L_{p^n}(c_n)}{n} < H_b(P) + \epsilon.$$

Proof. Since $H_b(P) = \liminf_{m\to\infty} \frac{H_b(p^m)}{m} = \lim_{m\to\infty}\left(\inf_{n\geq m}\frac{H_b(p^n)}{n}\right)$ there exists a positive integer r such that

$$\inf_{n\geq m}\frac{H_b(p^n)}{n} < H_b(P) + \frac{\epsilon}{2}$$

for all $m \geq r$. Thus for all $m \geq r$ there exists $n \in \mathbb{Z}^+$ with $n \geq m$ and

$$(*) \qquad \frac{H_b(p^n)}{n} < H_b(P) + \frac{\epsilon}{2}.$$

Choose an integer m such that

$$m \geq r, \quad m > k \quad \text{and} \quad m > \frac{2}{\epsilon}.$$

Choose $n \geq m$ as in $(*)$. Note that

$$(**) \qquad \frac{1}{n} \leq \frac{1}{m} < \frac{1}{\frac{2}{\epsilon}} = \frac{\epsilon}{2}.$$

By the Fundamental Theorem, there exists a prefix-free b-nary code c_n for S^n with

$$(***) \qquad L_{p^n}(c_n) \leq H_b(p^n) + 1.$$

Combining $(*)$, $(**)$ and $(***)$ we obtain

$$\frac{L_{p^n}(c_n)}{n} \overset{(***)}{\leq} \frac{H_b(p^n)+1}{n} = \frac{H_b(p^n)}{n} + \frac{1}{n} \overset{(*)}{<} H_b(P) + \frac{\epsilon}{2} + \frac{1}{n} \overset{(**)}{<} H_b(P) + \frac{\epsilon}{2} + \frac{\epsilon}{2} = H_b(P) + \epsilon.$$

□

Exercises 4.4:

4.4#1. Given a stationary source (S, P) with entropy H. Let n, q, r, s be integers with $n = qr + s$, $0 \leq s < r$, $n \geq 1$ and $q \geq 1$. Show that

(a)

$$\frac{H(p^n)}{n} \leq \frac{H(p^r)}{r}\left(1 - \frac{s}{n}\right) + \frac{H(p^s)}{n}$$

(b)
$$\lim_{n\to\infty} \frac{H(p^n)}{n} = H$$

(*Hint: Given $\epsilon > 0$, show that there exists $r \in \mathbb{Z}^+$ such that $\frac{H(p^r)}{r} < H + \frac{\epsilon}{2}$. Let K be the maximum value of $H(p^s)$ for $0 \le s < r$ and choose a positive integer n_0 such that $\frac{K}{n_0} < \frac{\epsilon}{2}$.*)

4.4#2. Given a memoryless source $(\mathbb{B}, P)$ with probability distribution $p = (0.2, 0.8)$.

(a) Compute the entropy $H_2(P)$ of the source to the base 2.

(b) Use the Coding Theorem for Memoryless Sources 4.4.10 to find a positive integer n such that there exists a prefix-free binary code for $\mathbb{B}^n$ with average word-length L such that $\frac{L}{n}$ is within 10% of $H_2(P)$.

4.5 Arithmetic codes

Definition 4.5.1. *Let $(S, <)$ be an ordered alphabet and p a positive probability distribution on S. For $s \in S$ define:*

$$\alpha := \alpha(s) := \sum_{\substack{t\in S\\ t<s}} p_t,$$

where $\alpha = 0$ if s is the smallest element of S;

$$n' := n'(s) := \left\lceil \log_2\left(\frac{1}{p_s}\right)\right\rceil;$$

$$n := n(s) := n' + 1;$$

$$c' := c'(s) := \lceil 2^n \alpha \rceil;$$

$$c(s) := z_1 z_2 \dots z_n \in \mathbb{B}^*,$$

where

$$z_1, \dots, z_n \in \mathbb{B} \quad \textit{with} \quad c' = \sum_{i=1}^{n} z_i 2^{n-i} = z_1 2^{n-1} + z_2 2^{n-2} + \dots + z_{n-1} 2 + z_n.$$

Note here that $0 \le \alpha \le 1 - p_s$ and $n > n' \ge \log_2\left(\frac{1}{p_s}\right)$. Thus $2^n p_s > 1$ and so

$$0 \le 2^n \alpha \le 2^n (1 - p_s) = 2^n - 2^n p_s < 2^n - 1.$$

It follows that $0 \le c' \le 2^n - 1$ and so there exist unique such $z_1, \dots, z_n$.

Then the function

$$c: \quad S \to \mathbb{B}^*, \quad s \mapsto c(s)$$

is called the arithmetic code for the ordered alphabet S with respect to p.

Example 4.5.2. Determine the arithmetic code for the ordered alphabet $S = (a, d, e, b, c)$ with respect to the probability distribution $p = (0.1, 0.3, 0.2, 0.15, 0.25)$.

s	p	α	$\frac{1}{p}$	n'	n	2^n	$2^n\alpha$	c'	$\sum_{i=1}^n z_i 2^{n-i}$	c
a	0.1	0	10	4	5	32	0	0	$0\cdot 16 + 0\cdot 8 + 0\cdot 4 + 0\cdot 2 + 0$	00000
d	0.3	0.1	3.3...	2	3	8	0.8	1	$0\cdot 4 + 0\cdot 2 + 1$	001
e	0.2	0.4	5	3	4	16	6.4	7	$0\cdot 8 + 1\cdot 4 + 1\cdot 2 + 1$	0111
b	0.15	0.6	6.3...	3	4	16	9.6	10	$1\cdot 8 + 0\cdot 4 + 1\cdot 2 + 0$	1010
c	0.25	0.75	4	2	3	8	6	6	$1\cdot 4 + 1\cdot 2 + 0$	110

In the remainder of the section we will prove that arithmetic codes are prefix-free and find an upper bound for their average codeword length.

Definition 4.5.3. *Let $z = z_1 z_2 \dots z_n \in \mathbb{B}^n$. Then the rational number*

$$\sum_{i=1}^{n} \frac{z_i}{2^i}$$

is called the rational number associated to z and is denoted by $0_* z$.

Example 4.5.4. Compute 0_*1011.

$0_*1011 = \frac{1}{2} + \frac{0}{4} + \frac{1}{8} + \frac{1}{16} = \frac{8+2+1}{16} = \frac{11}{16}$.

Lemma 4.5.5. *Let $z \in \mathbb{B}^*$. Then $0_* z \in [0, 1)$.*

Proof. Let $z = z_1 \dots z_n$. Then

$$0 \le 0_* z = \sum_{i=1}^{n} \frac{z_i}{2^i} \le \sum_{i=1}^{n} \frac{1}{2^i} < \sum_{i=1}^{\infty} \frac{1}{2^i} = 1.$$

□

Lemma 4.5.6. *Let $\alpha \in [0, 1)$ and $n \in \mathbb{Z}^+$ with $\alpha + \frac{1}{2^n} < 1$. Put $c' := \lceil 2^n \alpha \rceil$ and let $z = z_1 \dots z_n$ with $z_i \in \mathbb{B}$ and $c' = \sum_{i=1}^n z_i 2^{n-i}$.*

(a) *There exists a unique $x \in \mathbb{B}^n$ with $0_* x \in [\alpha, \alpha + \frac{1}{2^n})$, namely $x = z$.*

(b) *Let $y \in \mathbb{B}^*$. Then $0_* zy = 0_* z + \frac{1}{2^n} 0_* y \in [\alpha, \alpha + \frac{1}{2^{n-1}})$.*

Proof. (a) Let $x = x_1 \dots x_n$ be any element of $\mathbb{B}^n$ and put $d := \sum_{i=1}^n x_i 2^{n-i}$. Then $d \in \mathbb{N}$ and x is uniquely determined by d. We need to show that $0_* x \in [\alpha, \alpha + \frac{1}{2^n})$ if and only if $x = z$.

We compute

$$(*) \qquad 2^n \cdot 0_* x = 2^n \cdot \sum_{i=1}^n \frac{x_i}{2^i} = \sum_{i=1}^n x_i 2^{n-i} = d.$$

Hence

$$\begin{array}{rcl} & 0_* x \in [\alpha, \alpha + \frac{1}{2^n}) & \\ \iff & \alpha \le 0_* x < \alpha + \frac{1}{2^n} & \\ \iff & 2^n \alpha \le d < 2^n \alpha + 1 & \text{– multiplication by } 2^n, \text{ and } (*) \\ \iff & d = \lceil 2^n \alpha \rceil & \text{– } d \in \mathbb{Z}, \text{ definition of } \lceil\ \rceil \\ \iff & d = c' & \text{– definition of } c' \\ \iff & x = z & \text{– } x \text{ is uniquely determined by } d \end{array}$$

This proves (a).

(b) Let $y = y_1 y_2 \dots y_m$. Then $zy = z_1 \dots z_n y_1 \dots y_m$. So y_i is the $(n+i)$-coefficient of zy. We compute

$$(**) \qquad 0_* zy = \sum_{i=1}^n \frac{z_i}{2^i} + \sum_{i=1}^m \frac{y_i}{2^{n+i}} = 0_* z + \frac{1}{2^n} \sum_{i=1}^m \frac{y_i}{2^i} = 0_* z + \frac{1}{2^n} 0_* y$$

By (a)

$$(***) \qquad \alpha \le 0_* z < \alpha + \frac{1}{2^n}$$

and by 4.5.5

$$0 \le 0_* y < 1.$$

Dividing by 2^n now shows

$$(+) \qquad 0 \le \frac{1}{2^n} 0_* y < \frac{1}{2^n}$$

Adding the inequalities $(***)$ and $(+)$ gives

$$\alpha \le 0_* z + \frac{1}{2^n} 0_* y < \alpha + \frac{1}{2^n} + \frac{1}{2^n} = \alpha + \frac{1}{2^{n-1}}$$

Using $(**)$ we conclude

$$\alpha \leq 0_* zy < \alpha + \frac{1}{2^{n-1}}$$

Thus (b) holds. □

Theorem 4.5.7. *Let c be the arithmetic code for the ordered alphabet S with respect to the positive probability distribution p.*

(a) *For $s \in S$ put $I_s := \big[\alpha(s), \alpha(s) + p_s\big)$. Then $(I_s)_{s\in S}$ is a partition of $[0,1)$, that is for each $r \in [0,1)$, there exists a unique $s \in S$ with $r \in I_s$.*

(b) *$0_* c(s) y \in I_s$ for each $s \in S$ and $y \in \mathbb{B}^*$.*

(c) *c is a prefix-free binary code.*

Proof. (a) Let $S = (s_1, s_2, \dots, s_m)$ with $s_1, s_2, \dots, s_m \in S$. Observe that $\alpha(s_i) + p_{s_i} = \alpha(s_{i+1})$ for all $1 \leq i < m$ and $\alpha(s_m) + p_{s_m} = \sum_{s\in S} p_s = 1$. Since $p_{s_i} > 0$ this gives

$$0 = \alpha(s_1) < \alpha(s_2) < \dots < \alpha(s_{m-1}) < \alpha(s_m) < 1.$$

Let $r \in [0,1)$. Then there exists a unique $i \in \mathbb{Z}^+$ with $i \leq m$ such that $\alpha(s_i) \leq r < \alpha(s_{i+1})$ (if $i < m$) and $\alpha(s_i) \leq r < 1$ (if $i = m$). Observe that $I_{s_i} = \big[\,\alpha(s_i), \alpha(s_{i+1})\,\big)$ if $1 \leq i \leq m-1$ and $I_{s_m} = [\alpha(s_m), 1)$. Hence there exists a unique element of S with $r \in I_s$, namely $s = s_i$.

(b) Let $s \in S$ and $y \in \mathbb{B}^*$. Recall that $n = n' + 1$ and $n' = \left\lceil \log_2\left(\frac{1}{p_s}\right)\right\rceil$. Thus $2^{n'} \geq \frac{1}{p_s}$ and so $\frac{1}{2^{n'}} \leq p_s$. Therefore

$$(*) \qquad \alpha + \frac{1}{2^n} < \alpha + \frac{1}{2^{n'}} \leq \alpha + p_s \leq 1.$$

Let $c' := \lceil 2^n \alpha \rceil$ and $c(s) = z_1 \dots z_n$ with $z_i \in \mathbb{B}$. By definition of c we have $c' = \sum_{i=1}^n z_i 2^{n-i}$. Hence we can apply (4.5.6)(a) with $z = c(s)$ and conclude that

$$0_* c(s) y \in \left[\alpha, \alpha + \frac{1}{2^{n-1}}\right) = \left[\alpha, \alpha + \frac{1}{2^{n'}}\right) \overset{(*)}{\subseteq} \Big[\alpha, \alpha + p_s\Big) = I_s.$$

(c) Let $s, t \in S$ with $s \neq t$ and let $y \in \mathbb{B}^*$. By (b), $0_* c(t) \in I_t$ and $0_* c(s) y \in I_s$. By (a) $I_t \cap I_s = \varnothing$. Hence $c(t) \neq c(s) y$ and so $c(s)$ is not a prefix of $c(t)$. Thus c is a prefix-free code. □

Theorem 4.5.8. *Let c be the arithmetic code for the ordered alphabet S with respect to the positive probability distribution p. Then*

$$L_p(c) < H_2(p) + 2.$$

Proof. Let $s \in S$. Then $c(s)$ has length

$$n(s) = n'(s) + 1 = \left\lceil \log_2\left(\frac{1}{p_s}\right)\right\rceil + 1 < \log_2\left(\frac{1}{p_s}\right) + 2.$$

So

$$L_p(c) < \sum_{s\in S} p_s\left(\log_2\left(\frac{1}{p_s}\right) + 2\right) = \sum_{s\in S} p_s\left(\log_2\left(\frac{1}{p_s}\right)\right) + 2\sum_{s\in S} p_s = H_2(p) + 2.$$

We can also prove this theorem by comparing c with a SF-code d for S with respect to d. By definition with an SF-code $\ell(d(s)) = \left\lceil \log_2\left(\frac{1}{p_s}\right)\right\rceil$ and so $\ell(c(s) = \left\lceil \log_2\left(\frac{1}{p_s}\right)\right\rceil + 1 = \ell(d(s)) + 1$. It follows that $L_p(c) = L_p(d) + 1$. By (3.5.4)(b) $L_p(d) \leq H_b(p) + 1$ and so $L_p(c) \leq H_b(p) + 2$. □

Corollary 4.5.9 (Coding Theorem for Arithmetic codes). *Let (S, P) be a source such that p is positive. Let $\epsilon > 0$ and $k > 0$. Then there exists an integer $n \geq k$ such that*

$$\frac{L_{p^n}(c)}{n} < H_2(P) + \epsilon$$

for every arithmetic code c for S^n with respect to p^n.

Moreover, if P is memoryless, this holds for any integer n with $n > \frac{2}{\epsilon}$.

Proof. Follow the proof for Coding Theorem for Sources (4.4.11) with the following modifications:

After Equation $(*)$: Choose m such that $m \geq r$ and $m > \frac{4}{\epsilon}$. So $(**)$ becomes:

$$(**') \qquad n > \frac{4}{\epsilon}$$

After Equation $(**)$: Let c be an arithmetic code on S^n with respect to p^n. By 4.5.8 we get

$$(***') \qquad L_{p^n}(c) < H(p^n) + 2$$

The changes in (**) and (***) cancel in the last computation in the proof of the Coding Theorem.

A similar change in the proof of the Coding Theorem for Memoryless sources gives the extra statement on memoryless sources. □

Exercises 4.5:

4.5#1. Compute $0_* z$ for $z = 00100101$ and for $z = 10111001$.

For exercises 4.5#2 and 4.5#3 let S be an ordered alphabet and (S, P) a memoryless source. For $r \in \mathbb{N}$ view S^r as an ordered alphabet via the lexicographic order. So if $x, y \in S^r$ then $x < y$ if there exists $1 \le k \le r$ with $x_i = y_i$ for $1 \le i < k$ and $x_k < y_k$.) Let c_r be the arithmetic code for S^r with respect to p^r.

4.5#2. Let $y = y_1 \dots y_r \in S^r$ and $s \in S$. Prove that

$$\alpha(ys) = \alpha(y) + P(y)\alpha(s) = \alpha(y) + p(y_1)p(y_2)\cdots p(y_r)\alpha(s).$$

4.5#3. Suppose $S = (a, b, c, d)$ and $p = (0.4, 0.3, 0.2, 0.1)$.

(a) Compute the average codeword length L of c_2 with respect to p^2 and compare $\frac{L}{2}$ with $H_2(p)$.

(b) Compute $c_2(ac)$ and $c_3(acd)$.

4.6 Coding with a dynamic dictionary

Definition 4.6.1. *A dictionary D based on the alphabet S is a 1-1 N-tuple $D = (d_1, \dots, d_N)$ with coefficients in S^* for some $N \in \mathbb{N}$. (Here 1-1 means D is 1-1 as a function from $\{1, \dots, N\}$ to S^*, that is $d_i \ne d_j$ for all $1 \le i < j \le N$).*

Let $d \in S^*$. If $d = d_i$ for some $1 \le i \le N$, we say that d appears in D with the index i.

Example 4.6.2. Let $S = (s_1, s_2, \dots, s_m)$ be an ordered alphabet. Then $D = (s_1, \dots, s_m)$ is a dictionary based on S.

Algorithm 4.6.3 (LZW encoding)**.** *Let $S = (s_1, s_2, \dots, s_m)$ be an ordered alphabet and let X be a non-empty message using S. Define*

(∗) *a positive integer n;*

(∗) *non-empty message Y_k, $1 \le k \le n$ using S;*

(∗) *positive integer c_k, $1 \le k \le n$;*

(∗) *non-empty messages X_k, $0 \le k < n$ using S;*

(∗) *symbols z_k, $1 \le k < n$ in S; and*

(∗) *dictionaries D_k, $0 \le k < n$, based on S*

recursively as follows:

For $k = 0$ *define*

$$D_0 := (s_1 \dots, s_m) \quad \text{and} \quad X_0 := X.$$

Suppose $k \geq 1$ *and that* D_{k-1} *and* X_{k-1} *have been defined.*

(•) Y_k *is the longest prefix of* X_{k-1} *such that* Y_k *appears in* D_{k-1}*.* [2]

(•) c_k *is the index of* Y_k *in* D_{k-1}*.*

If $Y_k = X_{k-1}$*, put* $n = k$ *and terminate the algorithm. If* $Y_k \neq X_{k-1}$*:*

(•) X_k *is the (non-empty) message using* S *with* $X_{k-1} = Y_k X_k$*.*

(•) z_k *is the first symbol of* X_k*.*

(•) $D_k := (D_{k-1}, Y_k z_k)$.[3]

Put $c(X) := c_1 c_2 \dots c_n$*. Also define* $c(\varnothing) = \varnothing$*. The function*

$$c : S^* \to \mathbb{N}^*, X \to c(X)$$

is called the LZW*-encoding function for the ordered alphabet* S*.*

Example 4.6.4. Given the ordered alphabet (a, b, c, d, e). Determine the LZW-encoding of *bdddaadda*.

$$\begin{array}{c|c|c|c|c|c} b & d & d\,d & a & a & dda \\ 2 & 4 & 7 & 1 & 1 & 8 \end{array}$$

$m+k$	1	2	3	4	5	6	7	8	9	10
d_{m+k}	a	b	c	d	e	bd	dd	dda	aa	ad

So the encoding is 247118.

Lemma 4.6.5. *With the notation as in the LZW-encoding algorithm:*

(a) $X_k = Y_{k+1} Y_{k+2} \dots Y_n$ *for all* $0 \leq k < n$*.*

(b) $X = Y_1 \dots Y_n$*.*

(c) D_k *has length* $m + k$ *and* D_k *is a prefix of* D_l *for all* $0 \leq k \leq l < n$*.*

[2] Note that all symbols of S appear in D_{k-1}, so such a prefix exists and has length at least 1.

[3] Note that by maximality of $\ell(Y_k)$, $Y_k z_k$ does not appear in D_k. So D_k is a dictionary.

(d) *Y_k is the element appearing with index c_k in D_l for all $k - 1 \le l < n$.*

(e) *$Y_k z_k$ is the element appearing with index $m + k$ in D_k for all $1 \le k < n$.*

(f) *z_k is the first symbol of Y_{k+1} for all $1 \le k < n$.*

Proof. (a) By definition of n we have $X_{n-1} = Y_n$. So (a) holds for $k = n - 1$. Suppose (a) holds for k, that is $X_k = Y_{k+1} \dots Y_n$. By construction $X_{k-1} = Y_k X_k$, so $X_{k-1} = Y_k Y_{k+1} \dots Y_k$ and (a) holds for $k - 1$. Thus (a) holds for all $0 \le k \le n - 1$, by downwards induction.

(b) Since $X = X_0$, this is the case $k = 0$ in (a).

(c) By construction, D_0 has length m and D_{k-1} is the parent of D_k. So (c) holds by induction.

(d) By construction, this holds for $l = k - 1$. Since D_{k-1} is a prefix of D_l for all $k - 1 \le l < n$, (c) follows.

(e) By construction, $Y_k z_k$ is the last element of D_k. Since D_k has length $m + k$, (d) holds.

(f) By construction, z_k is the first symbol of X_k. By (a), $X_k = Y_{k+1} \dots Y_n$ and so z_k is the first symbol of Y_{k+1}. □

Algorithm 4.6.6 (LZW decoding)**.** *Let $S = (s_1, s_2, \dots, s_m)$ be an ordered alphabet and let $u = c_1 \dots c_n$ be a message in $\mathbb{N}$. If $u \ne \varnothing$ and $c_k \le m + k - 1$ for all $1 \le k \le n$ define*

(∗) *a message Y_{k+1}, $0 \le k < n$,*

(∗) *symbols z_k, $1 \le k \le n$.*

(∗) *$m + k$-tuples D_k, $0 \le k \le n$, with coefficients in S*

recursively as follows:

For $k = 0$ define

$$D_0 := (s_1 \dots, s_m).$$

and let Y_1 be the message of index c_1 in D_0.

Suppose $k \ge 1$ and that D_{k-1} and Y_k already have been defined.

If $k = n$, the algorithm stops. If $k < n$:

(•) (⋄) *If $c_{k+1} < m + k$, let Y_{k+1} be the message with index c_{k+1} in D_{k-1} and let z_k be the first symbol of Y_{k+1}.* [4]

[4] Recall from 4.6.5 that in LZW-encoding, Y_{k+1} is the message of index c_k in D_{k-1} and z_k is the first symbol of Y_{k+1}.

($\diamond$) *If $c_{k+1} = m + k$, let z_k be the first symbol of Y_k and let $Y_{k+1} = Y_k z_k$*[5]

($\bullet$) $D_k = (D_{k-1}, Y_k z_k)$.

Put $e(u) = Y_1 \dots Y_n$.
If $u = \varnothing$ or $c_k \geq m + k$ for some $1 \leq k \leq n$, define $e(u) = \varnothing$
The function $e : \mathbb{N}^* \to S^*, u \to e(u)$ is called the LZW*-decoding for S.*

Example 4.6.7. Find the LZW-decoding of the message 3.4.6.7.9.10.8 for the ordered alphabet $S = (B, A, D, E, F)$.

u						3	4	6	7	9	10	8
Y_k						D	E	DE	ED	EDE	EDEE	DEE
z_k						E	D	E	E	E	D	
d_{m+k}	B	A	D	E	F	DE	ED	DEE	EDE	EDEE	EDEED	
$m + k$	1	2	3	4	5	6	7	8	9	10	11	

So the decoding is $DEDEEDEDEEDEEDEE$.

4.6#1.

4.6#2. Use the LZW-decoding rules for the ordered alphabet $S = (\sqcup, B, D, E, N, O, P, R, S, T)$ to decode

10.6.1.2.4.1.6.8.1.5.6.10.1.11.13.4

4.6#3. Compute the LZW-encoding of MISSISSIPPI for ordered alphabet $S = (I, M, P, S)$.

[5] Recall from 4.6.5 that in LZW-encoding $Y_k z_k$ is the message of index $m + k$ in D_k

Chapter 5

Error Correcting

5.1 Decision rules

Definition 5.1.1. *Let $n \in \mathbb{N}$ and $x, y \in \mathbb{B}^n$.*

(a) $D(x,y) \coloneqq \{i \mid x_i \neq y_i, 1 \leq i \leq n\}$.

(b) $\mathrm{d}(x,y) \coloneqq |D(x,y)|$ *(so* $\mathrm{d}(x,y)$ *is the number of positions in which x and y differ.)* $\mathrm{d}(x,y)$ *is called the Hamming distance of x and y.*

(c) *x is called even if $\sum_{i=1}^{n} x_i$ is even. Otherwise x is called odd.*

Example 5.1.2. Let $x = 0100100$ and $y = 1101011$.

(1) Compute $D(x,y)$ and $\mathrm{d}(x,y)$.

(2) Which of x and y are even? odd?

(1)

$$\begin{array}{ccccccc}
0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
\hline
\text{✘} & & & \text{✘} & \text{✘} & \text{✘} & \text{✘} \\
1 & 2 & 3 & 4 & 5 & 6 & 7
\end{array}$$

So $D(x,y) = \{1,4,5,6,7\}$ and $\mathrm{d}(x,y) = 5$.

(2) x has a 1 in two positions and so is even. y has a 1 in five positions and so is odd.

Definition 5.1.3. *Let $C \subseteq \mathbb{B}^n$.*

(a) *C is called (the set of codewords of) a binary code of length n.*

(b) *A decision rule for C is a function $\sigma : \mathbb{B}^n \to C$.*

Definition 5.1.4. *Let $C \subseteq \mathbb{B}^n$ and σ a decision rule for C. Let $a \in C$ and $z \in \mathbb{B}^n$. (Think of a as the codeword transmitted via some channel, and z as the message received).*

(a) *Let $k \in \mathbb{N}$. (a, z) is called a k-bit error of C if $d(a, z) = k$.*

(b) *We say that σ corrects (a, z) if $a = \sigma(z)$.*

(c) *We say that σ is r-error correcting if σ corrects all k-bit errors for $0 \le k \le r$.*

Example 5.1.5. Given the binary code $C = \{000, 110, 101, 011\}$ (so C consist of all even binary messages of length 3) and the decision rule

$$\sigma : \quad \begin{array}{cccccccc} \boxed{000} & 100 & 010 & 001 & \boxed{110} & \boxed{101} & \boxed{011} & 111 \\ \hline 000 & 110 & 000 & 101 & 110 & 101 & 011 & 110 \end{array}$$

Does σ correct 0-bit errors? Does σ correct 1-bit errors?

0-bit errors are of the form $(a, a), a \in C$. Since $\sigma(a) = a$ for all $a \in C$, all 0-bit errors are corrected. So σ is 0-error correcting.

σ corrects some 1-bit errors but not all:

$$\sigma(001) = 101 \neq 011$$

Thus $(101, 001)$ is a 1-bit error corrected by σ, but $(011, 001)$ is a 1-bit error which is not corrected by σ. So σ is not 1-error correcting.

Example 5.1.6. Given the code $C = \{000\,000, 110\,110, 101\,101, 011\,011\}$. (Note that C is obtained by doubling the code in 5.1.5). Define the decision rule σ for C by

$$\sigma(xy) = \begin{cases} xx & \text{if } x \text{ is even} \\ yy & \text{if } x \text{ is odd} \end{cases}$$

for all $x, y \in \mathbb{B}^3$. Show that σ is 1-error correcting.

Let (a, z) be k-bit error for $k \le 1$. Since $a \in C$, $a = bb$ for some even b in $\mathbb{B}^3$. Let $z = xy$ with $x, y \in \mathbb{B}^3$. Since bb and xy differ in at most 1 place, $b = x$ or $b = y$.

Suppose $b = x$. Since b is even we get $\sigma(xy) = xx = bb = a$.

Suppose $b \neq x$. Then b must differ in exactly one bit from x, and $y = b$. So x is odd and $\sigma(xy) = yy = bb = a$. So σ is indeed 1-error correcting.

Definition 5.1.7. *Let $C \subseteq \mathbb{B}^n$. Define*

$$\delta := \delta(C) := \min\{\mathrm{d}(a,b) \mid a, b \in C, a \neq b\}.$$

δ is called the minimum distance of C.

Example 5.1.8. Compute the minimum distance of the code

$$\{000\ 000, 111\ 000, 001\ 110, 110\ 011\}$$

The distances of 000 000 to the other codewords are 3, 3 and 4. Also

a	111 000	111 000	001 110
b	001 110	110 011	110 011
$\mathrm{d}(a,b)$	4	3	5

So the minimum distance is 3.

Definition 5.1.9. *Let A and B be sets. Then*

$$A + B = (A \smallsetminus B) \cup (B \smallsetminus A)$$

$A + B$ is called the symmetric difference of A and B.

Lemma 5.1.10. *Let A and B be sets.*

(a) $A + B = (A \cup B) \smallsetminus (A \cap B)$.

(b) $|A + B| = |A \smallsetminus B| + |B \smallsetminus A| = |A| + |B| - 2|A \cap B|$.

Proof. Readily verified. □

Lemma 5.1.11. *Let $a, b, c \in \mathbb{B}^n$.*

(a) $D(a,c) = D(a,b) + D(b,c)$.

(b) $\mathrm{d}(a,c) = \mathrm{d}(a,b) + \mathrm{d}(b,c) - 2|D(a,b) \cap D(b,c)|$.

(c) [**Triangular Inequality**] $\mathrm{d}(a,c) \leq \mathrm{d}(a,b) + \mathrm{d}(b,c)$, with equality if and only if $D(a,b) \cap D(b,c) = \varnothing$.

Proof. (a) Let $1 \leq i \leq n$.

If $i \notin D(a,b) \cup D(b,c)$, then $a_i = b_i = c_i$ and so $i \notin D(a,c)$.
If $i \in D(a,b) \smallsetminus D(b,c)$, then $a_i \neq b_i = c_i$ and so $a_i \neq c_i$ and $i \in D(a,c)$.
If $i \in D(b,c) \smallsetminus D(a,b)$, then $a_i = b_i \neq c_i$ and so $a_i \neq c_i$ and $i \in D(a,c)$.

If $i \in D(a,b) \cap D(b,c)$, then $a_i \neq b_i \neq c_i$. Since $\mathbb{B}$ only has two elements this gives $a_i = c_i$ and so $i \notin D(a,c)$.

Hence $i \in D(a,c)$ if and only if

$$i \in (D(a,b) \smallsetminus D(b,c)) \cup (D(b,c) \smallsetminus D(a,b)) \overset{(5.1.10)(a)}{=} D(a,b) + D(b,c).$$

(b):

$$\begin{aligned}
\mathrm{d}(a,c) =& |D(a,c)| && - \text{Definition of } \mathrm{d}(a,c)\\
=& |D(a,b) + D(b,c)| && - \text{(a)}\\
=& |D(a,b)| + |D(b,c)| - 2|D(a,b) \cap D(b,c)| && - (5.1.10)\text{(b)}\\
=& \mathrm{d}(a,b) + \mathrm{d}(b,c) - 2|D(a,b) \cap D(b,c)| && - \text{Definition of } \mathrm{d}(\cdot,\cdot), \text{twice}
\end{aligned}$$

(c): Note that $|D(a,b) \cap D(b,c)| \geq 0$ with equality if and only if $D(a,b) \cap D(b,c) = \varnothing$. So (c) follows from (b). □

Lemma 5.1.12. *Let $a, b \in \mathbb{B}^n$. Put $d := \mathrm{d}(a,b)$ and let $e \in \mathbb{N}$ with $e \leq d$. Then there exists $x \in \mathbb{B}^n$ with $\mathrm{d}(a,x) = e$ and $\mathrm{d}(b,x) = d - e$.*

Proof. Since $e \leq d = \mathrm{d}(a,b) = |D(a,b)|$ we can choose a subset J of $D(a,b)$ with $|J| = e$. Define $x \in \mathbb{B}^n$ by

$$x_i = \begin{cases} b_i & \text{if } i \in J \\ a_i & \text{if } i \notin J. \end{cases}$$

Then

$$\begin{aligned}
x_i &= b_i \neq a_i && \text{if } i \in J\\
x_i &= a_i \neq b_i && \text{if } i \in D(a,b) \smallsetminus J\\
x_i &= a_i = b_i && \text{if } i \notin D(a,b)
\end{aligned}$$

Thus $D(a,x) = J$ and $D(b,x) = D(a,b) \smallsetminus J$. So $\mathrm{d}(a,x) = |J| = e$ and $\mathrm{d}(b,x) = |D(a,b) \smallsetminus J| = |D(a,b)| - |J| = d - e$. □

Example 5.1.13. Let $a = 110100111$ and $b = 101010101$. Show that $\mathrm{d}(a,b) = 5$ and find $x \in \mathbb{B}^9$ with $\mathrm{d}(a,x) = 2$ and $\mathrm{d}(b,x) = 3$.

a:	1	1	0	1	0	0	1	1	1
b:	1	0	1	0	1	0	1	0	1
		☒	✘	☒	✘			✘	
x:	1	0	0	0	0	0	1	1	1

Definition 5.1.14. *Let $r, n \in \mathbb{N}$, $x \in \mathbb{B}^n$ and $X \subseteq \mathbb{B}^n$. Then*

$$\mathrm{N}_r(x) := \{y \in \mathbb{B}^n \mid \mathrm{d}(x,y) \leq r\}$$

and

$$\mathrm{N}_r(X) := \{y \in \mathbb{B}^n \mid \mathrm{d}(x,y) \leq r \text{ for some } x \in X\} = \bigcup_{x \in X} \mathrm{N}_r(x)$$

$\mathrm{N}_r(x)$ is called the neighborhood of radius r around x.

Example 5.1.15. Compute $\mathrm{N}_1(0110)$.

$$\mathrm{N}_1(0110) = \{0110, 1110, 0010, 0100, 0111\}$$

Lemma 5.1.16. *Let $C \subseteq \mathbb{B}^n$ be a code and σ a decision rule for C. Then σ is r-error correcting if and only if $\sigma(z) = a$ for all $a \in C$ and all $z \in \mathrm{N}_r(a)$.*

Proof. σ is r-error correcting if and only if for all $0 \leq k \leq r$, σ corrects all k-bit errors (a,z) of C.

This holds if and only if $\sigma(z) = a$ for all $a \in C$ and $z \in \mathbb{B}^n$ with $\mathrm{d}(a,z) \leq r$ and so if and only if $\sigma(z) = a$ for all $a \in C$ and $z \in \mathrm{N}_r(a)$. □

Definition 5.1.17. *Let $C \subseteq \mathbb{B}^n$ be a binary code.*

(a) *C is called an r-error correcting code if $\delta(C) \geq 2r+1$, that is $\mathrm{d}(a,b) \geq 2r+1$ for all $a, b \in C$ with $a \neq b$.*

(b) *Let σ be a decision rule for C. We say that σ is a Minimum Distance decision rule (MD-rule) if*

$$\mathrm{d}\big(\sigma(z), z\big) \leq \mathrm{d}(a,z)$$

for all $a \in C$ and all $z \in \mathbb{B}^n$. (So $\sigma(z)$ is a codeword of minimal distance from z.)

Remark 5.1.18. *Let $C \subseteq \mathbb{B}^n$ be a binary code. Then there exists a MD-rule for C.*

Proof. For each $z \in \mathbb{B}^n$ choose $z' \in C$ with $\mathrm{d}(z', z)$ minimal. Then

$$\sigma: \quad \mathbb{B}^n \to C, \quad z \mapsto z'$$

is a MD- rule for C. □

Theorem 5.1.19. *Let $C \subseteq \mathbb{B}^n$ and $r \in \mathbb{N}$. Then the following statements are equivalent:*

(a) *C is an r-error correcting code.*

(b) *For each $z \in \mathbb{B}^n$, there exists at most one $a \in C$ with $\mathrm{d}(a, z) \leq r$.*

(c) $\mathrm{N}_r(a) \cap \mathrm{N}_r(b) = \varnothing$ *for all $a, b \in C$ with $a \neq b$.*

(d) *All MD-rules for C are r-error correcting.*

(e) *There exists an r-error correcting decision rule for C.*

Proof. (a) $\Longrightarrow$ (b): Suppose C is an r-error correcting code. Then $\delta(C) \geq 2r+1$. Let $z \in \mathbb{B}^n$ and $a, b \in C$ with $\mathrm{d}(a, z) \leq r$ and $\mathrm{d}(b, z) \leq r$. Then

$$\mathrm{d}(a, b) \overset{(5.1.11)(c)}{\leq} \mathrm{d}(a, z) + \mathrm{d}(z, b) \leq r + r = 2r < 2r + 1.$$

Since $\delta(C) \geq 2r + 1$ this implies that $a = b$. So there exists at most one codeword of distance less or equal to r from z.

(b) $\Longrightarrow$ (c): Suppose $\mathrm{N}_r(a) \cap \mathrm{N}_r(b) \neq \varnothing$ and let $z \in \mathrm{N}_r(a) \cap \mathrm{N}_r(b)$. Then $\mathrm{d}(a, z) \leq r$ and $\mathrm{d}(b, z) \leq r$, contradiction to (b).

(c) $\Longrightarrow$ (d): Let σ be minimal distance decision rule and let (a, z) be a k-bit error of C with $k \leq r$. Then $\mathrm{d}(a, z) = k \leq r$. Since σ is a minimal distance decision rule, we know that $\mathrm{d}(\sigma(z), z) \leq \mathrm{d}(a, z) \leq r$. Thus $z \in \mathrm{N}_r(a) \cap \mathrm{N}_r(\sigma(z))$ and (c) implies that $a = \sigma(z)$. So σ corrects (a, z) and σ is r-error correcting.

(d) $\Longrightarrow$ (e): By 5.1.18 there exists an MD-rule for C. So if all such rules are r-error-correcting there exists an r-error-correcting decision rule for C.

(e) $\Longrightarrow$ (a): Let σ be an r-error-correcting decision rule for C. Let $a, b \in C$ with $a \neq b$ and put $d := \mathrm{d}(a, b)$. Let $d = 2e + \epsilon$ with $\epsilon \in \{0, 1\}$ and $e \in \mathbb{N}$. Then $0 \leq e \leq d$ and so by 5.1.12 there exists $z \in \mathbb{B}^n$ with

$$\mathrm{d}(a, z) = e \quad \text{and} \quad \mathrm{d}(b, z) = d - e = e + \epsilon.$$

Since $a \neq b$, we have $\sigma(z) \neq a$ or $\sigma(z) \neq b$. So at least one of (a, z) and (b, z) is not corrected by σ. Since d is r-error correcting this implies that $\mathrm{d}(a, z) > r$ or $\mathrm{d}(b, z) > r$. Thus $e > r$ or $e + \epsilon > r$. In either case $e + \epsilon > r$. So $e + 1 \geq e + \epsilon > r$ and thus $e \geq r$. We proved that $e \geq r$ and $e + \epsilon > r$. Hence $d = 2e + \epsilon = e + (e + \epsilon) > r + r = 2r$ and so $d \geq 2r + 1$. Thus $\delta(C) \geq 2r + 1$ and C is r-error-correcting, □

Exercises 5.1:

5.1#1. Find a Minimum Distance decision rule for the binary code

$$C := \{0000, 1100, 0011, 1111\}$$

5.1#2. Let $D := \{00000, 11100, 10011\}$. Find all $a \in \mathbb{B}^5 \setminus D$ such that $D \cup \{a\}$ is a 1-error-correcting binary code.

5.1#3. Let n and r be positive integers, let $D \subseteq \mathbb{B}^n$ be an r-error-correcting code and let $a \in \mathbb{B}^n \setminus D$. Show that $D \cup \{a\}$ is an r-error-correcting code if and only if $a \notin N_{2r}(D)$.

5.1#4. Let $n \in \mathbb{N}$.

(a) Let $a, b, c \in \mathbb{B}^n$. Show that $\mathrm{d}(a,b) + \mathrm{d}(b,c) + \mathrm{d}(a,c) \le 2n$.

(b) Let $C \subseteq \mathbb{B}^n$ be a binary code with minimum distance δ. Suppose that $|C| \ge 3$. Show that $\mathrm{d}(a,b) \le 2(n-\delta)$ for all $a, b \in C$.

5.2 The Packing Bound

Lemma 5.2.1. *Let $r, n \in \mathbb{N}$ with $r \le n$ and let $x \in \mathbb{B}^n$.*

(a) *Let $A \subseteq \{1, 2, 3 \ldots, n\}$. Then there exists a unique $y \in \mathbb{B}^n$ with $D(x,y) = A$, namely y is given by*

$$y_i = \begin{cases} 1 - x_i & \textit{if } i \in A \\ x_i & \textit{if } i \notin A \end{cases}$$

for all $1 \le i \le n$.

(b) $|\{y \in \mathbb{B}^n \mid \mathrm{d}(x,y) = r\}| = \binom{n}{r}$.

(c) $|\mathrm{N}_r(x)| = \sum_{i=0}^{r} \binom{n}{i}$.

Proof. (a) Let $y \in \mathbb{B}^n$. Then $D(x,y) = A$ if and only if $y_i \neq x_i$ for $i \in A$ and $y_i = x_i$ for $i \notin A$. Since $\mathbb{B} = \{0,1\}$, $1 - x_i$ is the unique element of $\mathbb{B}$ unequal to x_i. Thus $D(x,y) = A$ if and only if

$$y_i = \begin{cases} 1 - x_i & \text{if } i \in A \\ x_i & \text{if } i \in A \end{cases}$$

(b) Note that $\mathrm{d}(x,y) = r$ if and only if $|D(x,y)| = r$. Together with (a) this shows that function

$$y \mapsto D(x,y)$$

is bijection between $\{y \in \mathbb{B}^n \mid \mathrm{d}(x,y) = r\}$ and the set of subsets of size r of $\{1, \ldots, n\}$. Note that there are exactly $\binom{n}{r}$ subsets of size r of $\{1, \ldots, n\}$ and so (b) holds.

(c) Note that $\mathrm{N}_r(x)$ is the disjoint union of the sets $\{y \in \mathbb{B}^n \mid \mathrm{d}(x,y) = i\}$, $0 \le i \le r$. So (b) follows from (c). □

Lemma 5.2.2. *Let $C \subseteq \mathbb{B}^n$ and $r \in \mathbb{N}$. Then*

(a) $|\mathrm{N}_r(C)| \leq 2^n$ *with equality if and only if* $\mathrm{N}_r(C) = \mathbb{B}^n$.

(b) $|\mathrm{N}_r(C)| \leq |C| \sum_{i=0}^{r} \binom{n}{i}$ *with equality if and only if C is an r-error correcting code.*

Proof. (a) Just observe that $\mathrm{N}_r(C) \subseteq \mathbb{B}^n$ and $|\mathbb{B}^n| = 2^n$.

(b) Note that $\mathrm{N}_r(C) = \bigcup_{a \in C} \mathrm{N}_r(a)$ and so

$$(*) \qquad |\mathrm{N}_r(C)| = \left| \bigcup_{a \in C} \mathrm{N}_r(a) \right| \leq \sum_{a \in C} |\mathrm{N}_r(a)|$$

with equality if and only if $\mathrm{N}_r(a) \cap \mathrm{N}_r(b) = \varnothing$ for all $a, b \in C$ with $a \neq b$. By 5.1.19 the latter holds if and only if C is r-error correcting.

By (5.2.1)(c)

$$(**) \qquad \sum_{a \in C} |\mathrm{N}_r(a)| = \sum_{a \in C} \sum_{i=0}^{r} \binom{n}{i} = |C| \sum_{i=0}^{r} \binom{n}{i}.$$

Combining $(*)$ and $(**)$ gives (b). □

Theorem 5.2.3 (The Packing Bound). *Let $C \subseteq \mathbb{B}^n$ be an r-error-correcting code. Then*

$$|C| \cdot \sum_{i=0}^{r} \binom{n}{i} \leq 2^n$$

and

$$|C| \leq \frac{2^n}{\sum_{i=0}^{r} \binom{n}{i}}.$$

Proof.

$$|C| \cdot \sum_{i=0}^{r} \binom{n}{i} \overset{(5.2.2)(\mathrm{b})}{=} |N_r(C)| \overset{(5.2.2)(\mathrm{a})}{\leq} 2^n$$

□

Definition 5.2.4. *Let $C \subseteq \mathbb{B}^n$ be a code. Then the information rate of C is the real number*

$$\rho(C) := \log_{|\mathbb{B}^n|} |C| = \frac{\log_2 |C|}{n}.$$

Example 5.2.5. Use the packing bound to find an upper bound for the information rate of a 2-error correcting code $C \subseteq \mathbb{B}^n$ of size 100.

By definition,

$$\rho(C) = \frac{\log_2(|C|)}{n} = \frac{\log_2(100)}{n}$$

So finding a upper bound for $\rho(C)$ is equivalent to finding lower bound for n. According to the packing bound,

$$|C| \cdot \sum_{i=0}^{r} \binom{n}{i} \leq 2^n$$

As $|C| = 100$ and $r = 2$ this gives

$$100\left(1 + n + \binom{n}{2}\right) \leq 2^n.$$

Hence

$$100\left(\frac{2 + 2n + n^2 - n}{2}\right) \leq 2^n$$

and

$$25(n^2 + n + 2) \leq 2^{n-1}.$$

Thus also $25n^2 \leq 2^{n-1}$ and so

$$5n \leq 2^{\frac{n-1}{2}}$$

Put $m := \frac{n-1}{2}$. Then $n = 2m + 1$ and so $5(2m + 1) \leq 2^m$ and

$$2^m - 10m - 5 \geq 0.$$

Consider the function $f(x) = 2^x - 10x - 5$. Then

$$f'(x) = \ln(2)\, 2^x - 10 \quad \text{and} \quad f''(x) = \frac{\ln(2)}{x}$$

Thus $f''(x) > 0$ for all $x \in (0, 6)$. Hence the maximum value of $f(x)$ on $[0, 6]$ occurs at one of the endpoints, that is at $x = 0$ or $x = 6$.

Since $f(0) = 1 - 0 - 5 < 0$ and $f(6) = 64 - 60 - 6 < 0$ we conclude $f(x) < 0$ on the interval $[0, 6]$. Thus $m > 6$ and $n = 2m + 1 > 13$. Hence $n \geq 14$ and

$$\rho(C) \leq \frac{\log_2 100}{14} \approx 0.228$$

Definition 5.2.6. *Let $\delta, n \in \mathbb{Z}^+$ with $\delta \le n$. Then $A(n,\delta)$ is the largest possible size of a code $C \subseteq \mathbb{B}^n$ with minimum distance at least δ. That is*

$$A(n,\delta) = \max_{\substack{C\subseteq\mathbb{B}^n \\ \delta(C)\ge\delta}} |C|$$

Lemma 5.2.7. *Let $\delta, n \in \mathbb{N}$ with $\delta \le n$ and put $r := \left\lfloor \frac{\delta-1}{2} \right\rfloor$. Then*

$$A(n,\delta) \le \frac{2^n}{\sum_{i=0}^{r}\binom{n}{i}}.$$

Proof. Let C be a code with $\delta(C) \ge \delta$. Note that $r \le \frac{\delta-1}{2}$, so $\delta \ge 2r+1$ and $\delta(C) \ge 2r+1$. Thus C is r-error correcting. Hence the Packing Bound shows that

$$|C| \le \frac{2^n}{\sum_{i=0}^{r}\binom{n}{i}}.$$

□

Lemma 5.2.8. *Let $\delta, n \in \mathbb{N}$ with $\frac{2}{3}n < \delta \le n$. Then $A(n,\delta) = 2$.*

Proof. The code $\{00\ldots0, 11\ldots1\}$ shows that $A(n,\delta) \ge 2$. Suppose for a contradiction that $A(n,\delta) \ge 3$. Then there exists a binary code $C \subseteq \mathbb{B}^n$ with $|C| \ge 3$ and

$$\delta(C) \ge \delta > \frac{2}{3}n.$$

Since $|C| \ge 3$ we can choose three distinct codewords a, b, c in C. Then

$$(*) \qquad \mathrm{d}(a,b) \ge \delta(C) > \frac{2}{3}n \qquad \text{and} \qquad \mathrm{d}(b,c) \ge \delta(C) > \frac{2}{3}n.$$

Put $I := \{1, 2, \ldots, n\}$. We compute

$$\begin{aligned}
\mathrm{d}(a,c) &= |D(a,c)| && - \text{definition of } \mathrm{d}(a,c) \\
&= |D(a,b) + D(a,c)| && - (5.1.11)(\mathrm{a}) \\
&= |D(a,b) \smallsetminus D(b,c)| + |D(b,c) \smallsetminus D(a,b)| && - (5.1.10)(\mathrm{b}) \\
&\le |I \smallsetminus D(b,c)| + |I \smallsetminus D(a,b)| && - D(a,b) \subseteq I \text{ and } D(b,c) \subseteq I \\
&\le (n - \mathrm{d}(b,c)) + (n - \mathrm{d}(a,b)) && - |I| = n, |D(b,c)| = \mathrm{d}(b,c), |D(a,b)| = \mathrm{d}(a,b) \\
&< (n - \frac{2}{3}n) + (n - \frac{2}{3}n) && - (*) \\
&= \frac{2}{3}n.
\end{aligned}$$

Thus $\mathrm{d}(a,c) < \frac{2}{3}n$, a contradiction to $\mathrm{d}(a,c) \ge \delta(C) > \frac{2}{3}n$. Hence $A(n,\delta) = 2$. □

Example 5.2.9. Consider a binary code $C \subseteq \mathbb{B}^{10}$ with $\delta(C) \geq 7$. Then $\delta(C) \geq 2 \cdot 3 + 1$ and so C is an 3-error-correcting code. Hence the Packing Bound shows that

$$|C| \leq \frac{2^{10}}{1 + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}} = \frac{1024}{1 + 10 + 45 + 120} = \frac{1024}{176} < 6$$

Thus $|C| \leq 5$. On the other hand, $\frac{2}{3} \cdot 10 < 7 \leq 10$ and so 5.2.8 shows that $A(10,7) = 2$. Thus $|C| \leq 2$. This shows that, in general, the packing bound is not the best possible bound.

Definition 5.2.10. *Let $C \subseteq \mathbb{B}^n$. Then C is called a perfect code if there exists $r \in \mathbb{N}$ such that for all $z \in \mathbb{B}^n$ there exists a unique $a \in C$ with $\mathrm{d}(z,a) \leq r$.*

Lemma 5.2.11. *Let $C \subseteq \mathbb{B}^n$. Then C is a perfect code if and only if there exists $r \in \mathbb{N}$ such that C is r-error correcting and*

$$|C| \sum_{i=0}^{r} \binom{n}{i} = 2^n.$$

Proof. Let $C \subseteq \mathbb{B}^n$. Then

$$\begin{aligned}
& \text{C is a perfect code} && \\
\iff\ & \exists r \in \mathbb{N}\Big(\forall z \in \mathbb{B}^n \text{ there exists a unique } a \in C \text{ with } \mathrm{d}(a,z) \leq r\Big) && -\text{definition of a perfect code}\\
\iff\ & \exists r \in \mathbb{N}\Big(\forall z \in \mathbb{B}^n \text{ there exists at most one } a \in C \text{ with } \mathrm{d}(a,z) \leq r && \\
& \qquad\qquad \text{and } \forall z \in \mathbb{B}^n \text{ there exists } a \in C \text{ with } \mathrm{d}(a,z) \leq r\Big) && \\
\iff\ & \exists r \in \mathbb{N}\Big(C \text{ is r-error-correcting} && -5.1.19\\
& \qquad\qquad \text{and } \forall z \in \mathbb{B}^n \text{ there exists } a \in C \text{ with } \mathrm{d}(a,z) \leq r\Big) && \\
\iff\ & \exists r \in \mathbb{N}\Big(C \text{ is r-error-correcting and } \mathbb{B}^n = N_r(C)\Big) && -\text{definition of } N_r(C)\\
\iff\ & \exists r \in \mathbb{N}\Big(C \text{ is r-error-correcting and } 2^n = |N_r(C)|\Big) && -(5.2.2)(\text{a})\\
\iff\ & \exists r \in \mathbb{N}\Big(C \text{ is r-error-correcting and } 2^n = |C| \textstyle\sum_{i=0}^{r} \binom{n}{i}\Big) && -(5.2.2)(\text{b})
\end{aligned}$$

□

Exercises 5.2:

5.2#1. Let C be a 3-error-correcting code with $C \subseteq \mathbb{B}^{12}$ and $|C| = 8$. Determine $|N_3(C)|$.

5.2#2. Let $n \in \mathbb{N}$ and suppose $C \subseteq \mathbb{B}^n$ is a perfect, 1-error-correcting binary code. Show that there exists $l \in \mathbb{N}$ such that $n = 2^l - 1$ and $|C| = 2^{2^l - l - 1}$.

Chapter 6

Linear Codes

6.1 Introduction to linear codes

Definition 6.1.1. $\mathbb{F}_2$ *denotes the set* $\mathbb{B} = \{0, 1\}$ *together with the following addition and multiplication:*

$+$	0	1
0	0	1
1	1	0

and

$\cdot$	0	1
0	0	0
1	0	1

Example 6.1.2. Compute $1 + 1$ and and $1 \cdot 0$ in $\mathbb{F}_2$.

$$1 + 1 = 0 \qquad \text{and} \qquad 1 \cdot 0 = 0$$

Remark 6.1.3. *Let* $a, b \in \mathbb{F}_2$*. Then* $a = b$ *if and only of* $a + b = 0$*.*

Proof. Follows immediately from the definition of the addition. □

Definition 6.1.4. *Let* $n \in \mathbb{N}$*.*

(a) $\mathbb{F}_2^n$ *is* $\mathbb{B}^n$ *together with the following addition and scalar multiplication:*

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \qquad \textit{and} \qquad l \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} lx_1 \\ lx_2 \\ \vdots \\ lx_n \end{pmatrix}$$

for all $l, x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{F}_2$.

(b) $\vec{0} := \underbrace{00\dots0}_{n-times}$.

Recall here that we are viewing

$$x_1x_2\dots x_n, \quad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and} \quad (x_1, x_2, \dots, x_n)$$

as three different notations for the exact same n-tuple.

(The book considers these to be different objects. If $x = x_1 \dots x_n$ is a message using $\mathbb{B}$ then x' denotes the corresponding column vector.)

Example 6.1.5. (1) Compute $11011 + 10110$.

(2) Let $n \in \mathbb{N}$, $l \in \mathbb{F}_2$ and $x \in \mathbb{F}_2^n$. Note that $l = 0$ or $l = 1$ and compute lx in each case.

(1)

$$\begin{array}{rr} & 11011 \\ + & 10110 \\ \hline = & 01101 \end{array}$$

(2) If $l = 0$, then $lx_i = 0$ for all i and if $l = 1$, then $lx_i = x_i$. Thus

$$lx = \begin{cases} \vec{0} & \text{if } l = 0 \\ x & \text{if } l = 1 \end{cases}$$

Definition 6.1.6. *A subspace of* $\mathbb{F}_2^n$ *is a subset* C *of* $\mathbb{F}_2^n$ *such that*

(i) $\vec{0} \in C$.

(ii) $x + y \in C$ *for all* $x, y \in C$. *[Closure under addition]*

(iii) $lx \in C$ *for all* $l \in \mathbb{F}_2, x \in \mathbb{F}_2$[1]. *[Closure under scalar multiplication]*

A subspace of $\mathbb{F}_2^n$ *is also called a binary linear code of length* n

[1] Since $lx = \vec{0}$ or $lx = x$, this condition is redundant

Example 6.1.7. (1) Is $\{0000, 1101, 1011, 0111\}$ a subspace of $\mathbb{F}_2^4$?

$$\begin{array}{rr} & 1101 \\ + & 1011 \\ \hline = & 0110 \end{array}$$

Since $0110 \notin C$, C is not closed under addition. So C is not a subspace.

(2) Is $\{011, 101, 110\}$ a subspace of $\mathbb{F}_2^3$?

No since $000 \notin C$.

(3) Is $\{000, 011, 101, 110\}$ is subspace of $\mathbb{F}_2^3$?

$000 \in C$. Also

$$\begin{array}{rr} & 011 \\ + & 101 \\ \hline = & 110 \\ & \in C \end{array} \quad \begin{array}{rr} & 011 \\ + & 110 \\ \hline = & 101 \\ & \in C \end{array} \quad \begin{array}{rr} & 101 \\ + & 110 \\ \hline = & 011 \\ & \in C \end{array}$$

Thus C is subspace.

(4) Is $C := \{x \in \mathbb{F}_2^n \mid x_1 + x_2 + \ldots + x_n = 0\}$ a subspace of $\mathbb{F}_2^n$.(Note that C consists of all the even $x \in \mathbb{F}_2^n$.)

Yes: Clearly $\vec{0} \in C$ and if $x, y \in C$, then $\sum_{i=1}(x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i = 0 + 0 = 0$ and so $x + y \in C$.

Definition 6.1.8. *Let C be a subspace of $\mathbb{F}_2^n$.*

(a) *A basis for C is a k-tuple*

$$(v_1, \ldots, v_k)$$

with coefficients in C such that for each a in C there exists a unique k-tuple

$$(l_1, \ldots, l_k)$$

with coefficients in $\mathbb{F}_2$ such that

$$a = l_1 v_1 + \ldots + l_k v_k.$$

(b) $\dim C := \log_2 |C|$.

(c) *The triple* $\big(n, \dim C, \delta(C)\big)$ *is called the linear parameter of* C.

Lemma 6.1.9. *Let* C *be a subspace of* $\mathbb{F}_2^n$.

(a) *If* $(v_1, \dots, v_k)$ *is a basis for* C [2]*, then* $|C| = 2^k$. *In particular,* $k = \log_2 |C| = \dim C$.

(b) $0 \le \dim C \le n$. *Moreover,* $\dim C = 0$ *if and only of* $C = \{\vec{0}\}$*, and* $\dim C = n$ *if and only if* $C = \mathbb{F}_2^n$.

(c) $\rho(C) = \frac{\dim C}{n}$.

Proof. (a) Let $l_1, \dots l_k \in \mathbb{F}_2$ and $v_1, \dots v_k \in C$. Since C is closed under addition and scalar multiplication we see that $\sum_{i=1}^k l_i v_i \in C$. Hence

$$\mathbb{F}_2^k \to C, \quad (l_1, \dots, l_k) \to l_1 v_1 + \dots + l_k v_k$$

is a well-defined function. The definition of a basis shows that this function is a bijection. Thus $|C| = |\mathbb{F}_2^k| = 2^k$.

(b) Since $1 \le |C| \le |\mathbb{F}_2^n| = 2^n$, we get $0 \le \log_2(|C|) \le n$. By definition of $\dim C$, we have $\dim C = \log_2(|C|)$ and so (b) holds.

(c) By definition of $\rho(C)$, we have $\rho(C) = \frac{\log_2(|C|)}{n}$. As $\dim C = \log_2(|C|)$ this gives (c). □

Example 6.1.10. Let $C = \{000\,000, 111\,000, 000\,111, 111\,111\}$. Find a basis for C. Determine $\dim C$ and $\rho(C)$.

Both

$$(111\,000, 000\,111)$$

and

$$(111\,000, 111\,111)$$

are bases for C. Thus $\dim C = 2$ and $\rho(C) = \frac{2}{6} = \frac{1}{3}$.

Definition 6.1.11. *Let* $x = x_1 \dots x_n \in \mathbb{F}_2^n$. *Then*

$$\mathrm{wt}(x) := |\{1 \le i \le n \mid x_i \ne 0\}|.$$

$\mathrm{wt}(x)$ *is called the weight of* x.

Lemma 6.1.12. *Let* $x, y \in \mathbb{F}_2^n$.

[2] We will prove in 6.3.11 that any subspace has a basis.

(a) *$x = y$ if and only if $x + y = \vec{0}$.*

(b) $\mathrm{d}(x, y) = \mathrm{wt}(x + y)$.

(c) $\mathrm{wt}(x) = \mathrm{d}(x, \vec{0})$.

(d) *Let C be a binary linear code of length n. Then $\delta(C)$ is the minimal weight of a non-zero codeword in C.*

Proof. (a): By 6.1.3 we know that $x_i = y_i$ if and only if $x_i + y_i = 0$. This gives (a).

(b) :

$$\mathrm{d}(x, y) = |\{1 \le i \le n \mid x_i \neq y_i\}| \overset{6.1.3}{=} |\{1 \le i \le n \mid x_i + y_i \neq 0\}| = \mathrm{wt}(x + y)$$

(c) $\mathrm{d}(x, \vec{0}) = \mathrm{wt}(x + \vec{0}) = \mathrm{wt}(x)$.

(d) Let w be the minimal weight of a non-zero codeword, and let $x, y \in C$ with $x \neq y$ and $\mathrm{d}(x, y) = \delta(C)$. Since $x \neq y$, (b) shows that $x + y \neq \vec{0}$. As C is subspace of $\mathbb{F}_2^n$ we have $x + y \in C$ and so $x + y$ is a non-zero codeword. Thus $\mathrm{wt}(x + y) \ge w$, and so

$$\delta(C) = \mathrm{d}(x, y) \overset{\text{(b)}}{=} \mathrm{wt}(x + y) \ge w.$$

Let $a \in C$ be non-zero codeword with $\mathrm{wt}(a) = w$. Since $\vec{0} \in C$, we have $\mathrm{d}(a, \vec{0}) \ge \delta(C)$. Thus

$$w = \mathrm{wt}(a) \overset{\text{(c)}}{=} \mathrm{d}(a, \vec{0}) \ge \delta(C).$$

We proved that $\delta(C) \ge w$ and $w \ge \delta(C)$, so $\delta(C) = w$. □

Example 6.1.13. Let $C := \{000\,000, 111\,000, 000\ 111, 111\,111\}$ be the linear code from Example 6.1.10. Determine the linear parameter of C.

The length of C is 6. The dimension of C is $\log_2 |C| = \log_2 4 = 2$. The weights of the non-zero codewords are $3, 3, 6$. So the minimum weight of a non-zero codeword is 3. Thus $\delta(C) = 3$ and the linear parameter is

$$(6, 2, 3).$$

6.2 Construction of linear codes using matrices

Definition 6.2.1. (a) *A binary matrix is a matrix with coefficients in $\mathbb{F}_2$*

(b) *Let E be a binary $n \times k$ matrix. Then*

$$\mathrm{Col}(E) := \{Ey \mid y \in \mathbb{F}_2^k\}.$$

$\mathrm{Col}(E)$ is called the linear code generated by E, and E is called a generating matrix for $\mathrm{Col}(E)$. $\mathrm{Col}(E)$ is also called the column space of E.

(c) *Let H be a binary $m \times n$-matrix. Then*

$$\mathrm{Nul}(H) := \{x \in \mathbb{F}_2^n \mid Hx = \vec{0}\}$$

$\mathrm{Nul}(H)$ is called the null space of H. H is called a check matrix for $\mathrm{Nul}(H)$.

Lemma 6.2.2. (a) *Let E be a binary $n \times k$- matrix and $y \in \mathbb{F}_2^k$. For $1 \le j \le k$ let $e_j := \mathrm{Col}_j(E)$ (so e_j is the j'th column of E.) Then*

$$Ey = \sum_{j=1}^{k} y_j e_j = y_1 e_1 + \ldots + y_k e_k = \sum_{\substack{j=1 \\ y_j=1}}^{k} e_j.$$

(b) *Let E be a binary $n \times k$ matrix. Then $\mathrm{Col}(E)$ is a subspace of $\mathbb{F}_2^n$.*

(c) *Let H be a binary $m \times n$-matrix. Then $\mathrm{Nul}(H)$ is a subspace of $\mathbb{F}_2^n$.*

Proof. (a) Let $1 \le i \le n$. Then i-coefficient of Ey is $\sum_{j=1}^{k} e_{ij} y_j$. The i-coefficient of e_j is e_{ij} and so the i-coefficient of $\sum_{j=1}^{k} y_j e_j$ is $\sum_{j=1}^{k} y_j e_{ij}$. Note that $e_{ij} y_j = y_j e_{ij}$ and so i-coefficients of Ey and $\sum_{j=1}^{k} y_j e_j$ are the same. Thus

$$\begin{aligned} Ey &= \sum_{j=1}^{k} y_j e_j \\ &= \sum_{\substack{j=1 \\ y_j=0}}^{k} y_j e_j + \sum_{\substack{j=1 \\ y_j=1}} y_j e_j \\ &= \sum_{\substack{j=1 \\ y_j=1}}^{k} e_j && \mid 0e_j = \vec{0}, 1e_j = e_j \text{ by } (6.1.5)(2) \end{aligned}$$

So (a) holds.

(b) $E\vec{0} = \vec{0}$. So $\vec{0} \in \mathrm{Col}(E)$. Let $a, b \in \mathrm{Col}(E)$. Then $a = Ex$ and $b = Ey$ for some $x, y \in \mathbb{F}_2^k$. Thus

$$a + b = Ex + Ey = E(x + y)$$

and so $a + b \in \mathrm{Col}(E)$.

(c) $H\vec{0} = \vec{0}$ and so $\vec{0} \in \mathrm{Nul}(H)$. Let $a, b \in \mathrm{Nul}H$. The $Ha = \vec{0}$ and $Hb = \vec{0}$. Hence

$$H(a + b) = Ha + Hb = \vec{0} + \vec{0} = \vec{0}$$

and so $a + b \in \mathrm{Nul}H$. □

Example 6.2.3. Consider the binary matrix

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

(1) Find the minimal distance of $\mathrm{Col}(E)$.

(2) Is $\mathrm{Col}(E)$ 1-error correcting?

We will determine the minimum weight of a non-zero codeword. For this let $x \in \mathrm{Col}(E)$ with $x \neq \vec{0}$. Then $x = Ey$ for some $y \in F_2^3$ with $y \neq \vec{0}$. Let $y = abc$. Then

$$x = \begin{array}{c} \begin{pmatrix} a & b & c \end{pmatrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \end{array} = \begin{pmatrix} a \\ b \\ c \\ a + b \\ b + c \\ a + c \end{pmatrix}$$

Observe that $\mathrm{wt}(x)$ does not depend on the order of a, b and c. So we may assume that $a \geq b \geq c$. This gives three cases:

$y = 100$, $x = 100101$ and $\mathrm{wt}(x) = 3$.
$y = 110$, $x = 110011$ and $\mathrm{wt}(x) = 4$.
$y = 111$, $x = 111000$ and $\mathrm{wt}(x) = 3$.
Thus $\mathrm{Col}(E)$ has minimum distance $\delta(C) = 3 = 2 \cdot 1 + 1$ and $\mathrm{Col}(E)$ is 1-error correcting.

6.3 Standard form of check matrix

Notation 6.3.1 (Matrix of Matrices). *Let $(I_a)_{a\in A}$ and $(J_b)_{b\in B}$ be tuples of sets. Let $M = [M_{ab}]_{\substack{a\in A\\ b\in B}}$ be an $A \times B$ matrix such that each M_{ab} is an $I_a \times J_b$-matrix. Put $I := \dot\bigcup_{a\in A} I_a$ (the disjoint union of I_a, $a \in A$) and $J := \dot\bigcup_{b\in B} J_b$. Then we identify M with the $I \times J$-matrix N defined by*

$$N_{ij} = (M_{ab})_{ij}$$

for all $i \in I, j \in J$, where a is the unique element of A with $i \in I_a$ and b is the unique element of B with $j \in J_b$.

Example 6.3.2. Given the matrices

$$X_{11} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad X_{12} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

$$X_{21} = \begin{bmatrix} 9 & 10 & 11 \end{bmatrix}, \quad \text{and} \quad X_{22} = \begin{bmatrix} 12 \end{bmatrix}.$$

Apply 6.3.1 to $[X_{11}, X_{12}]$, to $\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix}$ and to $\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$.

$$[X_{11}, X_{12}] = \begin{bmatrix} 1 & 2 & 3 & 7 \\ 4 & 5 & 6 & 8 \end{bmatrix},$$

$$\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 9 & 10 & 11 \end{bmatrix},$$

and

$$\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 7 \\ 4 & 5 & 6 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$$

Definition 6.3.3. *Let A be a set. Then I_A is the $A \times A$-matrix*

$$I_A := [\delta_{ab}]_{\substack{a\in A \\ b\in A}} \qquad \textit{where} \quad \delta_{ab} := \begin{cases} 1 & \textit{if } a = b \\ 0 & \textit{if } a \neq b \end{cases}$$

I_A is called the $A \times A$ identity matrix. If $n \in \mathbb{N}$, then

$$I_n := I_{\{1,\dots,n\}}.$$

Example 6.3.4. Determine I_4 and $I_{\{x,y\}}$.

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad I_{\{x,y\}} = \begin{array}{c|cc} & x & y \\ \hline x & 1 & 0 \\ y & 0 & 1 \end{array}$$

Lemma 6.3.5. *Let A be binary $m \times k$ matrix. Define*

$$n := m + k, \qquad E := \begin{bmatrix} I_k \\ A \end{bmatrix}, \qquad H := [A \ I_m]$$

and

$$c_E : \quad \mathbb{F}_2^k \to \mathbb{F}_2^n, \quad y \mapsto Ey.$$

Let $a \in \mathbb{F}_2^n$ and write $a = vw$ with $v \in \mathbb{F}_2^k$ and $w \in \mathbb{F}_2^m$. (v is called the message part of a and w the check part of a)

(a) *$c_E(y) = Ey = (y, Ay)$ for all $y \in \mathbb{F}_2^k$.*

(b) *c_E is a bijection from $\mathbb{F}_2^k$ to* $\mathrm{Col}(E)$. *In particular, c_E is a binary code for $\mathbb{F}_2^k$ with set of codewords* $\mathrm{Col}(E)$.

(c) *The columns of E form a basis for* $\mathrm{Col}(E)$.

(d) $\mathrm{Col}(E)$ *is a binary linear code of length* n *and dimension* k.

(e) $a \in \mathrm{Col}(E) \quad\iff\quad w = Av \quad\iff\quad a \in \mathrm{Nul}(H) \quad\iff\quad a = c_E(v)$.

(f) $\mathrm{Col}(E) = \mathrm{Nul}(H)$. *In particular, H is a check matrix for* $\mathrm{Col}(E)$.

Proof. (a):

$$c_E(y) = Ey = \begin{bmatrix} I_k \\ A \end{bmatrix} y = \begin{pmatrix} I_k y \\ Ay \end{pmatrix} = \begin{pmatrix} y \\ Ay \end{pmatrix} = (y, Ay)$$

(b): We need to show that c_E is 1-1 and $\mathrm{Im}\, c_E = \mathrm{Col}(E)$. Let $y, z \in \mathbb{F}_2^k$ with $c_E(y) = c_E(z)$. Then

$$(y, Ay) \overset{\text{(a)}}{=} c_E(y) = c_E(z) \overset{\text{(a)}}{=} (z, Ay)$$

and so $y = z$. Hence c_E is 1-1. Also

$$\mathrm{Col}(E) = \{Ey \mid y \in \mathbb{F}_2^k\} = \{c_E(y) \mid y \in \mathbb{F}_2^k\} = \mathrm{Im}\, c_E.$$

(c): Set $e_i := \mathrm{Col}_i(E)$ and let $y \in \mathbb{F}_2^k$. Then

$$c_E(y) = Ey \overset{(6.2.2)(\text{a})}{=} y_1 e_1 + y_2 e_2 + \ldots + y_k e_k.$$

Since c_E is a bijection from $\mathbb{F}_2^k$ to $\mathrm{Col}(E)$, we conclude that for each $a \in \mathrm{Col}(E)$ there exists a unique $y = (y_1, \ldots, y_k) \in \mathbb{F}_2^k$ with $a = y_1 e_1 + \ldots + y_k e_k$. Hence $(e_1, \ldots, e_k)$ is basis for $\mathrm{Col}(E)$.

(d): By (c) $(e_1, \ldots, e_k)$ is a basis for $\mathrm{Col}(E)$, so (6.1.9)(a) shows that $\dim \mathrm{Col}(E) = k$. Note that $\mathrm{Col}(E) \subseteq \mathbb{F}_2^{m+k} = \mathbb{F}_2^n$ and so $\mathrm{Col}(E)$ has length n.

(e):

$$\begin{array}{clll}
 & a \in \mathrm{Col}(E) & & \\
\iff & (v,w) \in \mathrm{Col}(E) & & \mid \text{Definition of } v,w \\
\iff & (v,w) = Ey & \text{for some } y \in \mathbb{F}_2^k & \mid \text{Definition of } \mathrm{Col}(E) \\
\iff & (v,w) = (y, Ay) & \text{for some } y \in \mathbb{F}_2^k & \mid \text{(a)} \\
\iff & v = y \text{ and } w = Ay & \text{for some } y \in \mathbb{F}_2^k & \\
\iff & w = Av & & \\
\iff & Av + w = \vec{0} & & \mid (6.1.12)(\text{a}) \\
\iff & [A\ I_m] \begin{pmatrix} v \\ w \end{pmatrix} = \vec{0} & & \mid \text{Definition of Matrix Multiplication} \\
\iff & Ha = \vec{0} & & \mid \text{Definition of } H \\
\iff & a \in \mathrm{Nul}(H) & & \mid \text{Definition of } \mathrm{Nul}(H)
\end{array}$$

Hence $a \in \operatorname{Col}(E) \iff w = Av \iff a \in \operatorname{Nul}(H)$. Recall that $a = vw = (v, w)$ and by (a) $c_E(v) = (v, Av)$. Thus

$$a = c_E(v) \quad\iff\quad (v, w) = (v, Av) \quad\iff\quad w = Av.$$

and (e) is proved.

(f): By (e) $a \in \operatorname{Col}(E)$ if and only if $a \in \operatorname{Nul}(H)$, so $\operatorname{Col}(E) = \operatorname{Nul}(H)$. □

Definition 6.3.6. *Let $C \subseteq \mathbb{F}_2^n$ be a linear code.*

(a) *A check matrix H for C is said to be in standard form if $H = [A\ I_m]$ for some $m \times k$ matrix A.*

(b) *A generating matrix E for C is said to be in standard form if $E = \begin{bmatrix} I_k \\ A \end{bmatrix}$ for some $m \times k$-matrix A.*

Definition 6.3.7. *Let $C, D \subseteq \mathbb{F}_2^n$. We say that D is a permutation of C if there exists a bijection $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ with*

$$D = \{a_1 a_2 \dots a_n \in \mathbb{F}_2^n \mid a_{\pi(1)} a_{\pi(2)} \dots a_{\pi(n)} \in C\}$$

Example 6.3.8. Show that

$$D = \{000, 100, 001, 101\}$$

is a permutation of

$$C = \{000, 010, 001, 011\}$$

D is obtained from C by interchanging the first two bits, that is via the permutation:

$$\pi : \frac{1\ \ 2\ \ 3}{2\ \ 1\ \ 3}.$$

Definition 6.3.9. *Let I and J be sets and let $i, k \in I$ with $i \neq k$. Then*

'$\mathrm{R}_i \leftrightarrow R_k$' and '$\mathrm{R}_i + \mathrm{R}_k \to R_k$' are the functions with domain and codomain the binary $I \times J$-matrices defined as follows: Let A be a binary $I \times J$-matrix.

(a) $(\mathrm{R}_i \leftrightarrow \mathrm{R}_k)(A)$ *is the matrix obtained by interchanging row i and row k of A*

(b) $(\mathrm{R}_i + \mathrm{R}_k \to \mathrm{R}_k)(A)$ *is the matrix obtained by adding row i to row k of A.*

An elementary row operation is any of the functions $\mathrm{R}_i \leftrightarrow \mathrm{R}_k$ *and* $\mathrm{R}_i + \mathrm{R}_k \to \mathrm{R}_k$.

Elementary column operations are defined similarly, using the symbol C *in place of* R *and using columns rather that rows.*

Lemma 6.3.10. (a) *Let H and G be binary $m \times n$-matrices and suppose G is obtain from H by a sequence of elementary row operation. Then* $\mathrm{Nul}(H) = \mathrm{Nul}(G)$.

(b) *Let E and F be binary $m \times n$-matrices and suppose F is obtain from E by a sequence of elementary column operation. Then* $\mathrm{Col}(E) = \mathrm{Col}(F)$.

Proof. (a): Let $h^i := \mathrm{Row}_i(H)$. Note that $x \in \mathrm{Nul}(H)$ if and only $Hx = 0$ and if and only of if $h^i x = 0$ for all $1 \leq i \leq m$.

Suppose that $G = (\mathrm{R}_i \leftrightarrow \mathrm{R}_k)(H)$. Then H and G have the same rows (just in a different order) and so $x \in \mathrm{Nul}(H)$ if and only if $x \in \mathrm{Nul}(G)$.

Suppose next that $G = (\mathrm{R}_i + \mathrm{R}_k \to \mathrm{R}_k)(H)$. Let $x \in \mathrm{Nul}(H)$. Let $l \in \mathbb{Z}$ with $1 \leq l \leq n$ and $l \neq k$, then $g^l = h^l$ and so $g^l x = 0$. Also $g^k x = (h^i + h^k)x = h^i x + h^k x = 0 + 0$. So $x \in \mathrm{Nul}(G)$ and thus $\mathrm{Nul}(H) \subseteq \mathrm{Nul}(G)$.

Note that $h^k = h^i + (h^i + h^k) = g^i + g^k$. Hence $(R_i + R_k \to R_k)(G) = H$ so $\mathrm{Nul}(G) \subseteq \mathrm{Nul}(H)$ by the result of the previous paragraph, applied with G and H interchanged. Thus $\mathrm{Nul}(G) = \mathrm{Nul}(H)$ and (a) holds.

(b): Put $e_j := \mathrm{Col}_j(E)$. Then $x \in \mathrm{Col}(E)$ if and only if $x = Ey$ for some $y \in \mathbb{F}_2^k$ and so if and only if

$$x = \sum_{l=1}^{n} y_l e_l$$

for some $(y_1, \ldots, y_k) \in \mathbb{F}_2^k$.

Suppose that $F = (\mathrm{C}_j \leftrightarrow \mathrm{C}_k)(E)$. Then E and F have the same columns (just in a different order) and so $x \in \mathrm{Col}(E)$ if and only if $x \in \mathrm{Col}(F)$.

Suppose next that $F = (\mathrm{C}_j + \mathrm{C}_k \to \mathrm{C}_k)(E)$ and let $x \in \mathrm{Col}(E)$. Then, for some $y \in \mathbb{F}_2^k$,

$$\begin{aligned} x = \sum_{l=1}^{n} y_l e_l \qquad\qquad &= y_j e_j + y_k e_k + \sum_{\substack{l=1 \\ l \neq j,k}}^{n} y_l e_l \\ = (y_j + y_k)e_j + y_k(e_j + e_k) + \sum_{\substack{l=1 \\ l \neq j,k}}^{n} y_l e_l &= (y_j + y_k)f_j + y_k f_k + \sum_{\substack{l=1 \\ l \neq j,k}}^{n} y_l f_l \end{aligned}$$

and so $x \in \mathrm{Col}(F)$.

Note that $E = (\mathrm{C}_j + \mathrm{C}_k \to \mathrm{C}_k)(F)$ and so the preceding result, applied with E and F interchanged, gives $\mathrm{Col}(F) \subseteq \mathrm{Col}(E)$. Thus $\mathrm{Col}(F) = \mathrm{Col}(E)$ and (b) holds. □

Theorem 6.3.11. *Let C be a subspace of $\mathbb{F}_2^n$. Then there exists a permutation D of C such that D has a generating matrix in standard form and a check matrix in standard form. In particular, C has a basis.*

Proof. We will first show that there exists a permutation of D with generating matrix E in standard form. For this we will construct a finite sequence of subspaces $C_k \subseteq \mathbb{F}_2^n$ and matrices E_k, $k = 0, 1, 2\ldots$, such that

(i) C_k is a permutation of C,

(ii) E_k is a generating matrix for C_k, and

(iii) E_k is of the form

$$E_k = \begin{bmatrix} I_k & 0 \\ * & * \end{bmatrix}.$$

Put $C_0 := C$ and let E_0 be any generating matrix for C_0, for example one can choose E_0 to be a matrix whose columns consists of all the codewords of C (in some order). Suppose now that C_k and E_k have already been constructed. Put $l := k + 1$.

Step 1. *Let A_l be the matrix obtained from E_k by deleting the zero columns of E_k.*

Observe that A_l is still a generating matrix for C_k and, since none of the first k-columns of E_k have been deleted, A_l still is of the form $\left[\begin{smallmatrix} I_k & 0 \\ * & * \end{smallmatrix}\right]$. Let m be the number of columns of A_l.

If $m = k$, then A_l is in standard from and we can choose $D = C_k$ and $E = A_l$ and we are done.

So suppose $m > k$. Then $l \leq m$ and Column l of A_l is not zero. So we can:

Step 2. *Choose $1 \leq i \leq n$ such that the il-coefficient of A_l is non-zero. Let C_l by the code obtain from C_k via the permutation $l \leftrightarrow i$. Put $B_l = (R_l \leftrightarrow R_i)(A_l)$.*

Since A_l is of shape $\left[\begin{smallmatrix} I_k & 0 \\ * & * \end{smallmatrix}\right]$ we have $a_{lj} = 0$ for $1 \leq j \leq k$. Thus $l \leq i \leq n$. So the first k rows of B_l are the same as the first k rows of A_l. Also the ll-coefficient of B_l is the li-coefficient of A_l and so is equal to 1. Thus B_l has shape

$$B_l = \begin{bmatrix} I_k & 0 & 0 \\ * & 1 & * \\ * & * & * \end{bmatrix}.$$

Since A_l is a generating matrix for C_k and C_l is a permuation of C_k, we see that B_l is a generating matrix for C_l.

Step 3. *Let E_l be the matrix obtained from B_l by adding Column l of B_l to Column j of B_l for each $1 \le j \le m$ such that $j \neq l$ and the jl-coefficient of B_l is equal to* 1.

By (6.3.10)(b) elementary column operations do not change the columns space. As B_l is a generating matrix for C_l we conclude that also E_l is a generating matrix for C_l. Note that E_l has shape

$$E_l = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 1 & 0 \\ * & * & * \end{bmatrix} = \begin{bmatrix} I_l & 0 \\ * & * \end{bmatrix}.$$

Observe that the number of columns of E_l is the number of non-zero columns of E_k. So the numbers of columns of E_l is less or equal to the number of columns of E_0. Thus the above algorithm will terminate in a find number of iterations.

This completes the proof that there exists a permutation D of C with a generating matrix $E = \left[\begin{smallmatrix} I_k \\ A \end{smallmatrix}\right]$ in standard form.

Put $m := n - k$ and $H = [A\, I_m]$. By 6.3.5 $\mathrm{Nul}(H) = \mathrm{Col}(E) = D$ and so H is check matrix in standard form for D. Moreover, the columns of E form a basis for $\mathrm{Col}(E) = D$. Hence D has a basis, and since C is a permutation of D, also C has a basis. □

Example 6.3.12. Consider the binary linear code

$$C := \{0000, 0011, 1010, 1001\}.$$

of length 4.

(a) Find a permutation D of C, a generating matrix E for D in standard form, and check matrix H for D in standard form.

(b) Find a basis for C and a basis for D.

l	A_l	$\mathrm{R}_l \leftrightarrow \mathrm{R}_i$	B_l	E_l	C_l
0				$\begin{bmatrix}0&0&1&1\\0&0&0&0\\0&1&1&0\\0&1&0&1\end{bmatrix}$	$\left\{\begin{matrix}0&0&1&1\\0&0&0&0\\0&1&1&0\\0&1&0&1\end{matrix}\right\}$
1	$\begin{bmatrix}0&1&1\\0&0&0\\1&1&0\\1&0&1\end{bmatrix}$	$\mathrm{R}_1 \leftrightarrow \mathrm{R}_3$	$\begin{bmatrix}1&1&0\\0&0&0\\0&1&1\\1&0&1\end{bmatrix}$	$\begin{bmatrix}1&0&0\\0&0&0\\0&1&1\\1&1&1\end{bmatrix}$	$\left\{\begin{matrix}0&1&1&0\\0&0&0&0\\0&0&1&1\\0&1&0&1\end{matrix}\right\}$
2	E_1	$\mathrm{R}_2 \leftrightarrow \mathrm{R}_3$	$\begin{bmatrix}1&0&0\\0&1&1\\0&0&0\\1&1&1\end{bmatrix}$	$\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\\1&1&0\end{bmatrix}$	$\left\{\begin{matrix}0&1&1&0\\0&0&1&1\\0&0&0&0\\0&1&0&1\end{matrix}\right\}$
3	$\begin{bmatrix}1&0\\0&1\\0&0\\1&1\end{bmatrix}$				

Put

$$E := A_3 = \begin{bmatrix}1&0\\0&1\\0&0\\1&1\end{bmatrix} \qquad H := \begin{bmatrix}0&0&1&0\\1&1&0&1\end{bmatrix} \qquad \text{and} \qquad D := C_2 = \left\{\begin{matrix}0&1&1&0\\0&0&1&1\\0&0&0&0\\0&1&0&1\end{matrix}\right\}$$

Then D is a permutation of C with generating matrix E and check matrix H.

Moreover, $(1001, 0101)$ (the columns of E) is a basis for D and since C is obtained from D via the permutation $2 \leftrightarrow 3$ followed by $1 \leftrightarrow 3$, that is via the permutation $1 \to 3 \to 2 \to 1$, we conclude that

$$(0011, 1001)$$

is a basis for C.

6.4 Constructing 1-error-correcting linear codes

Definition 6.4.1. *Let $v_1, \ldots, v_k \in \mathbb{F}_2^n$. Then $(v_1, \ldots, v_k)$ is called linearly dependent if there exists a non-zero $(l_1, \ldots, l_k) \in \mathbb{F}_2^k$ with*

$$l_1 v_1 + l_2 v_2 + \ldots + l_k v_k = \vec{0}.$$

$(v_1, \ldots, v_k)$ is called linearly independent, if its is not linearly dependent.

Remark 6.4.2. *A k–tuple $(v_1, \ldots, v_k)$ with $v_i \in \mathbb{F}_2^n$ is linearly independent if and only if for all $(l_1, \ldots, l_k) \in \mathbb{F}_2^k$:*

$$l_1 v_1 + l_2 v_2 + \ldots + l_k v_k = \vec{0} \quad \Longrightarrow \quad l_1 = 0, l_2 = 0, \ldots, l_k = 0.$$

Lemma 6.4.3. *Let H be a check matrix for the binary linear code C.*

(a) *The minimum distance of C is the minimal number of columns of H whose sum is equal to $\vec{0}$. It is also equal to the minimal number of linearly dependent columns of H.*

(b) *Let $r \in \mathbb{Z}^+$. Then C is r-error-correcting if and only if no (non-empty) sum of $2r$ or less columns of H is equal to $\vec{0}$.*

Proof. (a):

- Let δ_1 be the minimum weight of a non-zero codeword, so $\delta_1 = \delta(C)$ by (6.1.12)(d).
- Let δ_2 be the minimum number of columns whose sum is equal to $\vec{0}$.
- Let δ_3 be the minimum number of linear dependent columns of H.
- Let $h_i := \mathrm{Col}_i(H)$ be the i-th column of H.

We will show that $\delta_1 \le \delta_2 \le \delta_3 \le \delta_1$.

By definition of δ_2, there exists a non-empty set $I \subseteq \{1, \ldots, n\}$ of size δ_2 such that $\sum_{i \in I} h_i = \vec{0}$. Define $a \in \mathbb{F}_2^n$ by $a_i = \begin{cases} 1 & \text{if } i \in I \\ 0 & \text{if } i \notin I \end{cases}$. Then $\mathrm{wt}(a) = |I| = \delta_2$ and

$$Ha \overset{(6.2.2)(a)}{=} \sum_{i=1}^{n} a_i h_i = \sum_{i \in I} h_i = \vec{0}.$$

Since H is a check matrix of C this gives $a \in C$. Hence $\delta_1 \le \mathrm{wt}(a) = \delta_2$.

By definition of δ_3, there exists a set $J \subseteq \{1, \dots, n\}$ of size δ_3 such that the columns $(h_i)_{i\in J}$ are linearly dependent. Then $\sum_{i\in J} \lambda_i h_i = 0$ for some $\lambda_i \in \mathbb{F}_2$, not all zero. Put $K := \{i \in J \mid \lambda_i = 1\} = \{i \in J \mid \lambda_i \neq 0\}$. Then

$$\sum_{i\in K} h_i = \sum_{i\in K} \lambda_i h_i \sum_{\substack{i\in J\\ \lambda_i\neq 0}} \lambda_i h_i = \sum_{i\in J} \lambda_i h_i = \vec{0}.$$

Thus $(h_i)_{i\in K}$ are $|K|$ columns of H whose sum is $\vec{0}$. So $\delta_2 \leq |K| \leq |J| = \delta_3$.

Choose $a \in C$ with $a \neq \vec{0}$ and $\mathrm{wt}(a) = \delta_1$. Let $L := \{i \in \mathbb{Z} \mid 1 \leq i \leq n, a_i \neq 0\}$. Then $|L| = |\mathrm{wt}(a)| = \delta_1$. Since H is a check matrix for C we get $Ha = \vec{0}$. Hence

$$\sum_{i\in L} a_i h_i = \sum_{\substack{i=1\\ a_i\neq 0}}^{n} a_i h_i = \sum_{i=1}^{n} a_i h_i = Ha = \vec{0}.$$

Hence the columns $(h_i)_{i\in L}$ are linearly dependent. So $\delta_3 \leq |L| = \delta_1$.

We proved that $\delta_1 \leq \delta_2 \leq \delta_3 \leq \delta_1$ and so $\delta_1 = \delta_2 = \delta_3$.

(b):

$$\begin{array}{lcl}
 & C \text{ is } r\text{-error correcting} & \\
\iff & \delta(C) \geq 2r + 1 & \text{— Definition of } r \text{ -error correcting.} \\
\iff & \delta(C) > 2r & \text{— } \delta(C) \text{ and } 2r \text{ are integers} \\
\iff & \delta_2 > 2r & \text{— (a)} \\
\iff & \text{No non-mepty sum of } 2r \text{ or less columns of } H \text{ is zero.} & \text{— Definition of } \delta_2.
\end{array}$$

□

Corollary 6.4.4. *Let H be a check matrix for the binary linear code C. Then C is 1-error-correcting if and only if the columns of H are non-zero and pairwise distinct.*

Proof. We apply 6.4.3 with $r = 1$. We conclude that C is 1-error correcting if and only the sum of one or two of the columns is never $\vec{0}$.

The sum of one column is never zero if and only if the columns are non-zero. Since $x + y = \vec{0}$ if and only if $x = y$, the sum of two columns is never zero if and only if the columns are pairwise distinct. Hence C is 1-error correcting if and only if the columns of H are non-zero and pairwise distinct. □

Example 6.4.5. Let C be the binary linear code with check matrix

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Is C 1-error correcting? Is C 2-error-correcting?

The columns of H are non-zero and pairwise distinct. So by 6.4.4 C is 1-error correcting.

Note that the sum of first three columns of H is the fifth column. Thus the sum of these four columns is $\vec{0}$. Since $4 = 2 \cdot 2$, we conclude from (6.4.3)(b) C is not 2-error correcting.

It might be interesting to observe that since the sum of Columns 1,2,3 and 5 is $\vec{0}$ we know that 111010 is in C. So C has minimum weight at most 4.

Lemma 6.4.6. (a) *Let $n, k \in \mathbb{Z}^+$. Then there exists a binary linear 1-error correcting code of dimension k and length n if and only if*

$$k \le n - \lceil \log_2(n+1) \rceil .$$

(b) *The maximal information rate of a 1-error correcting, binary linear code of length n is*

$$1 - \frac{\lceil \log_2(n+1) \rceil}{n}.$$

Proof. (a) Note that $k \le n - \lceil \log_2(n+1) \rceil$ if and only if $\lceil \log_2(n+1) \rceil \le n-k$ and since $n-k$ is an integer, if and only if

$$\log_2(n+1) \le n-k.$$

Suppose first that there exists a 1-error-correcting code C of length n and dimension k. Then the Packing bound shows that

$$|C| \sum_{i=0}^{1} \binom{n}{i} \le 2^n$$

Since $\dim C = k$ and so $|C| = 2^k$, this gives

$$2^k(1+n) \le 2^n$$

As $\log_2(x)$ is an increasing function, we can apply $\log_2()$ to each side of this equation and conclude that $k + \log_2(n+1) \le n$. Thus $\log_2(n+1) \le n-k$.

Suppose next that $\log_2(1+n) \leq n-k$ and put $m := n-k$. Then $\log_2(1+n) \leq m$ and so $1+n \leq 2^m$. Hence

$$2^m - (1+m) \geq (1+n) - (1+m) = n - m = k.$$

Put $e_i := \mathrm{Col}_i(I_m)$. Then

$$|\mathbb{F}_2^m \smallsetminus \{\vec{0}, e_1, \dots, e_m\}| = 2^m - (1+m) \geq k.$$

So there exists k pairwise distinct vectors $a_1, \dots, a_k \in \mathbb{F}_2^m \smallsetminus \{\vec{0}, e_1, \dots, e_m\}$. Let $A = [a_1\, a_2\, \dots\, a_k]$ be the $m \times k$ matrix with $\mathrm{Col}_i(A) = a_i$ and put

$$H := [A\ I_m] = [a_1\, a_2\, \dots a_k\, e_1\, e_2\, \dots e_m.]$$

Note that $k+m=n$. Thus H is a binary $m \times n$ matrix with pairwise distinct non-zero columns. Hence 6.4.4 show that $\mathrm{Nul}(H)$ is 1-error-correcting code of length n. As H is in standard form 6.3.5 shows that $\dim \mathrm{Nul}(H) = k$.

(b): By (6.1.9)(c) we have $\rho(C) = \frac{\dim C}{n} = \frac{k}{n}$. Dividing the equation in (a) by n we obtain (b). □

Corollary 6.4.7. *Let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \in \mathbb{Z}^+$ with $n \geq N$ there exists a binary, linear, 1-error-correcting code of length n and information rate at least $1-\epsilon$.*

Proof. Observe that

$$\lim_{n\to\infty} 1 - \frac{\lceil \log_2(n+1) \rceil}{n} = 1,$$

and so there exists $N \in \mathbb{N}$ such that

$$1 - \frac{\lceil \log_2(n+1) \rceil}{n} \geq 1 - \epsilon$$

for all $n \geq N$. Let $n \in \mathbb{N}$ with $n \geq N$. By (6.4.6)(b) there exists a binary, linear 1-error-correcting code C of length n with $\rho(C) = 1 - \frac{\lceil \log_2(n+1) \rceil}{n}$. Thus $\rho(C) \geq 1-\epsilon$ and the corollary holds. □

Definition 6.4.8. *A Hamming code is a perfect, 1-error correcting, binary, linear code.*

Theorem 6.4.9. *Let $C \subseteq \mathbb{F}_2^n$ be a linear code with an $m \times n$ check matrix H in standard form. Then C is a Hamming code if and only if $n = 2^m - 1$ and the columns of H are the non-zero vectors of $\mathbb{F}_2^m$ (in some order).*

Proof. Since C is binary and linear, C is a Hamming code if and only if C is 1-error-correcting and perfect. So by 5.2.11

$(*)$ *C is a Hamming code if and only if C is 1-error correcting and $|C|(1+n) = 2^n$.*

Since H is in standard form, we know that $\dim C = n - m$, see (6.3.5)(d). So $|C| = 2^{n-m}$. Thus

$$|C|(1+n) = 2^n \iff 2^{n-m}(1+n) = 2^n \iff 1+n = 2^m \iff n = 2^m - 1.$$

By 6.4.4 C is 1-error correcting if and only if the columns of H are non-zero and pairwise distinct.

Thus

$(**)$ *C is a Hamming code if and only if $n = 2^m - 1$ and the columns of H are non-zero and pairwise distinct.*

Since $\mathbb{F}_2^m$ has exactly $2^m - 1$ non-zero vectors, the theorem follows from $(**)$. □

Corollary 6.4.10. *Let $n \in \mathbb{N}$. Then there exists a Hamming code of length n if and only if $n = 2^m - 1$ for some $m \in \mathbb{N}$. If $n = 2^m - 1$, a Hamming code of length n is unique up to permutations.*

Proof. Suppose C is a Hamming code of length n. By 6.3.11 there exists a permutation D of C with an $m \times n$ check matrix H in standard form. Then 6.4.9 shows that $n = 2^m - 1$ and the columns of H are the non-zero vectors of $\mathbb{F}_2^m$ (in some order). This determines H up to a permutation of the columns. Thus D and so also C is unique up to permutation.

Suppose now that $n = 2^m - 1$ for some $m \in \mathbb{N}$ and let H be binary $m \times n$ check matrix in standard form such that the columns of H are the non-zero vectors of $\mathbb{F}_2^m$ (in some order). Now 6.4.9 implies that $\mathrm{Nul}(H)$ is a Hamming code of length n.

□

Example 6.4.11. Find a check matrix for a Hamming code of length 7. List all the codewords.

We have $7 = 2^3 - 1 = 2^m - 1$ with $m = 3$. So we choose the check matrix

$$H := \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Observe that H is in standard form and thus by 6.3.5

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

is a generating matrix for C. We obtain all codewords by computing the sums of any i of the columns of E for $0 \leq i \leq 4$.

$$\begin{array}{lllllll}
i = 0: & & & 0000000 & & & \\
i = 1: & & 1000111 & 0100110 & 0010101 & 0001011 & \\
i = 2: & 1100001 & 1010010 & 1001100 & 0110011 & 0101101 & 0011110 \\
i = 3: & & 1110100 & 1101010 & 1011001 & 0111000 & \\
i = 4: & & & & 1111111 & &
\end{array}$$

Lemma 6.4.12. *Let $C \subseteq \mathbb{F}_2^n$ be a binary linear code with check matrix H. Let $z \in \mathbb{F}_2^n$ and let $i \in \mathbb{Z}^+$ with $i \leq n$. Define $a \in \mathbb{F}_2^n$ by*

$$a_j = \begin{cases} z_j & \textit{if } j \neq i \\ z_j + 1 & \textit{if } j = i \end{cases}$$

Then (a, z) is a 1-bit error of C if and only if $a \in C$ and if and only if $Hz = \mathrm{Col}_i(H)$.

Proof. By definition, (a, z) is a 1-bit error of C if and only if $a \in C, z \in \mathbb{F}_2^n$ and $\mathrm{d}(a, z) = 1$. As $z \in \mathbb{F}_2^n$ and $\mathrm{d}(a, z) = 1$ this holds if and only if $a \in C$. Put $e_i := \mathrm{Col}_i(I_n)$ and note that $a = z + e_i$. Then

$$\begin{array}{lll}
& a \in C & \\
\Longleftrightarrow & Ha = \vec{0} & - H \text{ is a check matrix for } C \\
\Longleftrightarrow & H(z + e_i) = \vec{0} & - a = z + e_i \\
\Longleftrightarrow & Hz + He_i = \vec{0} & - \text{Distributive Law} \\
\Longleftrightarrow & Hz = He_i & - (6.1.12)(\text{a}) \\
\Longleftrightarrow & Hz = \text{Col}_i(H) & - He_i = H\text{Col}_i(I_n) = \text{Col}_i(HI_n) = \text{Col}_i(H)
\end{array}$$

□

Example 6.4.13. Let C be the binary linear code with check matrix

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

(a) Is C 1-error correcting?

(b) For $z = 101100$ and $z = 101110$, does there exist $a \in C$ such that (a, z) is a 1-bit error? If yes, find a.

(a) The columns of H are non-zero and pairwise distinct. So C is 1-error correcting by 6.4.4

(b) Consider first that case $z = 101100$. We compute

$$\begin{matrix} \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} \\ Hz = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \end{matrix}$$

Thus Hz is not a column of H and so z cannot by the result of 1-bit error.
Now consider $z = 101110$. Then

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

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

Thus Hz is the third column of H. Put $a := z + e_3 = 100110$. Then by 6.4.12 (a, z) is a 1-bit error. To confirm that a is codeword we compute:

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

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

So $Ha = \vec{0}$ and $a = 100110$ is indeed a codeword.

6.5 Decision Rules for Linear Codes

Definition 6.5.1. *Let H be a check matrix for the linear code $C \subseteq \mathbb{F}_2^n$ and let $z \in \mathbb{F}_2^n$.*

(a) *Hz is called the syndrome of z with respect to H.*

(b) *$z + C := \{z + a \mid a \in C\}$ is called the coset of C containing z.*

Remark 6.5.2. *Let H be a check matrix for the linear code $C \subseteq \mathbb{F}_2^n$ and let $z \in \mathbb{F}_2^n$.*

(a) *$z \in C$ if and only of the syndrome of z with respect to H is $\vec{0}$.*

(b) *$z \in z + C$.*

Proof. (a) Since H is a check matrix C we have $C = \mathrm{Nul}H$ and so $z \in C$ if and only if $Hz = \vec{0}$.

(b) Since C is a linear code, $\vec{0} \in C$. Thus $z = z + \vec{0} \in \{z + a \mid a \in C\} = z + C$. □

Lemma 6.5.3. *Let H be a check matrix for the linear code $C \subseteq \mathbb{F}_2^n$. Let $y, z \in \mathbb{F}_2^n$. Then the following statements are equivalent:*

(a) *y and z have the same syndrome with respect to H.*

(b) $Hy = Hz$

(c) $y + z \in C$.

(d) $z = y + a$ *for some codeword* $a \in C$.

(e) $z \in y + C$.

(f) $(y + C) \cap (z + C) \neq \varnothing$.

(g) $y + C = z + C$.

Proof. We have

$$\begin{array}{lll}
 & y \text{ and } z \text{ have the same syndrome with respect to } H. & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \mid \text{(a)} \\
\iff & Hy = Hz & \mid \text{definition of a syndrome} \quad \mid \text{(b)} \\
\iff & Hz + Hy = \vec{0} & \mid (6.1.12)(b) \\
\iff & H(z + y) = \vec{0} & \mid \text{Distributive Law} \\
\iff & z + y \in C & \mid C = \text{Nul}(H) \quad \mid \text{(c)} \\
\iff & z + y = a \text{ for some } a \in C & \mid \text{Set } a := z + y \\
\iff & z = y + a \text{ for some } a \in C & \mid +y \quad \mid \text{(d)} \\
\iff & z \in y + C & \mid \text{definition of } y + C \quad \mid \text{(e)}
\end{array}$$

So the first five statements are equivalent.

(e) $\Longrightarrow$ (f): By (6.5.2)(b) $z \in z + C$. So if $z \in y + C$, then $z \in (z + C) \cap (y + C)$ and $(z + C) \cap (y + C) \neq \varnothing$. Thus (f) holds.

(f) $\Longrightarrow$ (g): Let $u \in (y + C) \cap (z + C)$. Then $u \in y + C$. Since (e) implies (b), we conclude that $Hu = Hy$. Similarly as $u \in z + C$ we have $Hu = Hz$. So

$$(*) \qquad Hy = Hz.$$

Let $v \in \mathbb{F}_2^n$. Since (b) and (e) are equivalent, we get

$$v \in y + C \quad \overset{(b) \iff (e)}{\iff} \quad Hv = Hy \quad \overset{(*)}{\iff} \quad Hv = Hz \quad \overset{(b) \iff (e)}{\iff} \quad v \in z + C.$$

So $y + C = z + C$.

(g) $\Longrightarrow$ (e): By (6.5.2)(b) $z \in z + C$. So if $y + C = z + C$, then $z \in y + C$. □

Corollary 6.5.4. *Let $C \subseteq \mathbb{F}_2^n$ be a linear code and let $z \in \mathbb{F}_2^n$. Then z lies in a unique coset of C, namely $z + C$. In particular, distinct cosets are disjoint.*

Proof. Let $y \in \mathbb{F}_2^n$. Then by 6.5.3 $z \in y + C$ if and only if $y + C = z + C$. So $z + C$ is the unique coset of C containing z. □

Example 6.5.5. Let C be the binary linear code with check matrix

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

(a) Compute the cosets of C and the corresponding syndromes.

(b) Use syndromes to find the coset containing 1101.

Since H is in standard form, $E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ is a generating matrix for C, see 6.3.5. Computing the sum of each set of columns of E we obtain

$$C = \{0000, 1010, 0101, 1111\}.$$

We calculate

coset	$0000 + C$	$1000 + C$	$0100 + C$	$1100 + C$
elements	0000	1000	0100	1100
	1010	0010	1110	0110
	0101	1101	0001	1001
	1111	0111	1011	0011
syndrome	00	10	01	11

Here in last row we listed the common syndrome of the elements in the coset. Observe that each of the 16 elements of $\mathbb{F}_2^4$ appears in one of these cosets. So the above four cosets are all the cosets of C. Which of the cosets contains 1101?

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

So 1101 is contained in the coset with syndrome 10, that is in $1000+C$.

Lemma 6.5.6. *Let C be a binary linear code of length n and dimension k. Let H be an $m\times n$-check matrix for C.*

(a) *The set of syndromes of H is $\mathrm{Col}(H)$.*

(b) *The function*

$$\alpha_H: \qquad \{z+C \mid z\in\mathbb{F}_2^n\} \;\to\; \mathrm{Col}(H), \quad z+C\mapsto Hz$$

is a well-defined bijection.

(c) *The numbers of syndromes for C with respect to H, the number of cosets of C and $|\mathrm{Col}(H)|$ all are equal to 2^{n-k}. In particular, $\dim\mathrm{Col}(H)=n-k$.*

(d) *$n-k\le m$ with equality if and only if and only if $\mathrm{Col}(H)=\mathbb{F}_2^m$.* [3]

Proof. (a) s is a syndrome of H if and only if $s=Hz$ for some $z\in\mathbb{F}_2^n$ and so if and only if $s\in\mathrm{Col}(H)$.

(b) Let $y,z\in\mathbb{F}_2^n$. By 6.5.3:

$$y+C=z+C \qquad\Longleftrightarrow\qquad Hy=Hz.$$

The forward direction shows that the function α_H is well-defined. The backward direction shows that α_H is 1-1. By definition, $\mathrm{Col}(H)=\{Hz\mid z\in\mathbb{F}_2^n\}=\mathrm{Im}\,\alpha_H$ and so α_H is surjective.

(c) Let u be the numbers of cosets of C in $\mathbb{F}_2^n$. Note that for $z\in\mathbb{F}_2^n$ the function

$$C\to z+C, \quad a\mapsto z+a$$

is a bijection (with inverse $b\mapsto z+b$). Thus $|z+C|=|C|=2^k$. Thus each cosets of C contains 2^k elements. Since there are u coset of C and each of the 2^n elements of $\mathbb{F}_2^n$ lies in a unique coset of C we conclude that

$$u\cdot 2^k=2^n.$$

[3] Note that $n-k=m$ if H is standard form.

Thus $u = 2^{n-k}$. By (b) $u = |\mathrm{Col}(H)|$ and by (a) $\mathrm{Col}(H)$ is the set of syndromes of H. Thus the number of syndrome of H is 2^{n-k} and (c) is proved.

(d) Note that $\mathrm{Col}(H) \subseteq \mathbb{F}_2^m$ and so

$$2^{n-k} \overset{(c)}{=} |\mathrm{Col}(H)| \leq |\mathbb{F}_2^m| = 2^m.$$

with equality if and only if $|\mathrm{Col}(H)| = |\mathbb{F}_2^m|$. Thus $n-k \leq m$ with equality if and only if $\mathrm{Col}(H) = \mathbb{F}_2^m$.

□

Remark 6.5.7. *Let H be binary $m \times n$-matrix. Then* $\dim \mathrm{Nul}(H) + \dim \mathrm{Col}(H) = n$.

Proof. Put $C := \mathrm{Nul}(H)$ and $k := \dim C$. Then

$$\dim \mathrm{Col}(H) \overset{(6.5.6)(\mathrm{d})}{=} n-k = n - \dim C = n - \dim \mathrm{Nul}(H).$$

□

Definition 6.5.8. *Let C be binary linear code with an $m \times n$ check matrix H. A syndrome look-up table for H is a function*

$$\tau : \quad \mathrm{Col}(H) \to \mathbb{F}_2^n$$

such that for each syndrome s of H:

(i) *$\tau(s)$ has syndrome s with respect to H, i.e $H\tau(s) = s$, and*

(ii) *$\tau(s)$ is a vector of minimal weight in $\tau(s)+C$, i.e* $\mathrm{wt}\big(\tau(s)\big) \leq \mathrm{wt}(z)$ *for all $z \in \tau(s)+C$.*

Remark 6.5.9. *Let C be binary linear code of length n with check matrix H and τ a syndrome look-up table for H.*

(a) *Let s be a syndrome of H. Then $\tau(s)+C$ is the set of vectors in $\mathbb{F}_2^n$ which have syndrome s with respect to H.*

(b) *Let $z \in \mathbb{F}_2^n$. Then $z + \tau(Hz) \in C$ and $z + C = \tau(Hz) + C$.*

Proof. (a) Let $z \in \mathbb{F}_2^n$. By definition of a syndrome look-up-table, $\tau(s)$ has syndrome s. Thus z has syndrome s if and only z and $\tau(s)$ have the same syndrome, and so by 6.5.3 if and only if $z \in \tau(s)+C$.

(b) Note that z and $\tau(Hz)$ both have syndrome Hz. Thus by 6.5.3 $z + \tau(Hz) \in C$ and $z + C = \tau(Hz) + C$. □

Definition 6.5.10. *Let C be binary linear code of length n with check matrix H and let τ a syndrome look-up table for H. Then the function*

$$\sigma : \quad \mathbb{F}_2^n \to C, \quad z \mapsto z + \tau(Hz)$$

is called the decision rule for C corresponding to H and τ.

Note that by the preceding remark $z + \tau(Hz) \in C$, so σ is well-defined.

Lemma 6.5.11. *Let C be linear code of length n with check matrix H, let τ be a syndrome look-up table for H and let σ be the corresponding decision rule. Then σ is a Minimum Distance decision rule for C.*

Proof. Let $z \in \mathbb{F}_2^n$ and $a \in C$. Put $s := Hz$. By (6.5.9)(b) $z + C = \tau(s) + C$. Thus

$$z + a \in z + C = \tau(s) + C.$$

By definition of syndrome look-up table, $\tau(s)$ is of minimal weight in $\tau(s) + C$. So

$$(*) \qquad \operatorname{wt}\big(\tau(s)\big) \le \operatorname{wt}(z + a).$$

We compute

$$\begin{aligned}
\mathrm{d}\big(\sigma(z), z\big) &= \mathrm{d}\big(z + \tau(Hz), z\big) && \mid \text{definition of } \sigma \\
&= \mathrm{d}\big(z + \tau(s), z\big) && \mid s = Hz \\
&= \operatorname{wt}\big(z + \tau(s) + z\big) && \mid \text{(6.1.12)(b)} \\
&= \operatorname{wt}\big(\tau(s)\big) && \mid z + z = \vec{0} \\
&\le \operatorname{wt}(z + a) && \mid (*) \\
&= \mathrm{d}(a, z) && \mid \text{(6.1.12)(b)}
\end{aligned}$$

Thus σ is Minimum-Distance decision rule. □

Algorithm 6.5.12. *Let C be binary linear code with an $m \times n$ check matrix H. Choose an ordering $(z_1, \dots, z_{2^n})$ of $\mathbb{F}_2^n$ such that $\operatorname{wt}(z_i) \le \operatorname{wt}(z_j)$ for all $1 \le i < j \le 2^n$. Define sets S_l and functions $\tau_l : S_l \to \mathbb{F}_2^n$ recursively as follows:*

For $l = 0$ let $S_0 = \varnothing$ and $\tau_0 = \varnothing$.

Suppose $l > 0$ and that S_{l-1} and τ_{l-1} have been defined. Compute $s_l := Hz_l$.

- *If $s_l \in S_{l-1}$, put $S_l := S_{l-1}$ and $\tau_l := \tau_{l-1}$*
- *If $s_l \notin S_{l-1}$, put $S_l := S_{l-1} \cup \{s_l\}$ and extend τ_{l-1} to a function τ_l on S_l by $\tau_l(s_l) := z_l$.*

If $|S_l| = |\mathrm{Col}(H)|$, $|S_l| = 2^m$ or $l = 2^n$, the algorithm stops. Otherwise continue with $l+1$ in place of l.

Then the last τ_l is a syndrome look-up table for H.

Example 6.5.13. Let C be the binary linear code with check matrix

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

Construct a syndrome look-up table τ for H and compute $\sigma(11001)$, where σ is the decision rule for C corresponding to H and τ.

Since H is in standard form, $\mathrm{Col}(H) = \mathbb{F}_2^3$. Thus the algorithm stops once we found 8 syndromes.

z_l	0	1	0	0	0	0	1	1	1	1
	0	0	1	0	0	0	1	0	0	0
	0	0	0	1	0	0	0	1	0	0
	0	0	0	0	1	0	0	0	1	0
	0	0	0	0	0	1	0	0	0	1
s_l	0	1	0	1	0	0	1	0	1	1
	0	1	1	0	1	0	0	1	0	1
	0	0	1	0	0	1	1	0	0	1
$\lvert S_l \rvert$	1	2	3	4	5	6	7	=	=	8

Let $z := 11001$. Then

$$\begin{array}{c} \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \end{pmatrix} \\ Hz = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \end{array}$$

$$\tau(Hz) = \tau(100) = 00100,$$

and

$$\begin{array}{rcl} z: & & 11001 \\ \tau(Hz): & + & 00100 \\ \hline \sigma(11001): & = & 11101 \end{array}$$

To double check

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

Thus 11101 $\in C$.

Chapter 7

Algebraic Coding Theory

7.1 Classification and properties of cyclic codes

In this chapter we will denote a string $a \in \mathbb{F}_2^n$ by $a_0 \dots a_{n-1}$ rather than $a_1 \dots a_n$, since we will consider the associated polynomial $a_0 + a_1 x + \dots a_{n-1}x^{n-1} \in \mathbb{F}_2[x]$.

Definition 7.1.1. *A binary code $C \subseteq \mathbb{F}_2^n$ of length n is called cyclic if*

(i) *C is linear, and*

(ii) *$a_{n-1}a_0 \dots a_{n-2} \in C$ for all $a = a_0 \dots a_{n-1} \in C$.*

Example 7.1.2. Which of following codes are cyclic?

(1) $\{000, 100, 111, 011\}$.

(2) $\{000, 100, 010, 001\}$

(3) $\{0000, 1010, 0101, 1111\}$

(1): Not cyclic, since 100 is in the code, but 010 is not.

(2): Not cyclic, since its not linear: 100 and 010 are in the code, but $100 + 010 = 110$ is not.

(3): Is cyclic.

Definition 7.1.3. *Let R be a ring with identity.*

(a) *$R[x]$ is the ring defined as follows:*

(i) *The elements of $R[x]$ are the infinite sequences $f = (f_i)_{i=0}^{\infty}$ with coefficients in R such that there exists $n \in \mathbb{N}$ with $f_i = 0$ for all $i \in \mathbb{N}$ with $i > n$.*

(ii) $f+g := (f_i+g_i)_{i=0}^{\infty}$ *for all* $f,g\in R[x]$

(iii) $fg = \left(\sum_{k=0}^{i} f_k g_{i-k}\right)_{i=0}^{\infty}$ *for all* $f,g\in R[x]$

(b) *The elements of* $R[x]$ *are called polynomials.*

(c) x *denotes the polynomial*

$$x := (0,1,0,0,\dots)$$

(d) $R[x]$ *is called the polynomial ring in* x *with coefficients in* R.

(e) $\mathbb{N}^* := \mathbb{N}\cup\{-\infty\}$. *Extend the binary operation* $+$ *on* $\mathbb{N}$ *to a binary operation on* $\mathbb{N}^*$, *and extend the relation* $\leq$ *on* $\mathbb{N}$ *to a relation on* $\mathbb{N}^*$ *via*

$$-\infty + a = -\infty, \qquad a+(-\infty) = -\infty, \quad \textit{and} \quad -\infty \leq a$$

for all $a\in\mathbb{N}^*$.

(f) *Let* $f\in R[x]$. *Define*

$$\deg f := \min\{i\in\mathbb{N}^* \mid f_n = 0 \text{ for all } n\in\mathbb{N} \text{ with } n>i\} = \begin{cases}\max\{i\in\mathbb{N}\mid f_i\neq 0\} & \textit{if } f\neq 0\\ -\infty & \textit{if } f=0\end{cases}$$

Remark 7.1.4. *Let* $\mathbb{F}$ *be a field and* $f,g\in\mathbb{F}[x]$.

(a) $f = \sum_{i=0}^{\deg f} f_i x^i$.

(b) $\deg(f+g)\leq\max(\deg f,\deg g)$.

(c) *If* $\deg g<\deg f$, *then* $\deg(f+g)=\deg f$.

(d) $\deg(fg)=\deg f+\deg g$.

(e) $\deg(af)=\deg f$ *for all* $a\in\mathbb{F}$ *with* $a\neq 0$.

(f) $\deg(-f)=\deg f$.

Lemma 7.1.5. *Let* $\mathbb{F}$ *be a field and* $f,h\in\mathbb{F}[x]$ *with* $h\neq 0$. *Then there exist uniquely determined* $q,r\in\mathbb{F}[x]$ *with*

$$f = qh+r \qquad \textit{and} \qquad \deg r<\deg h.$$

r *is called the remainder of* f *when divided by* h.

Proof. We first prove the existence of q and r by induction on $\deg f$. Note that $f = 0h + f$, so if $\deg f < \deg h$, we can choose $q = 0$ and $r = f$.

Suppose now that $\deg f \geq \deg h$. Let $f = \sum_{i=0}^{m} a_i x^i$ and $h = \sum_{i=0}^{n} b_i x^i$ with $a_m \neq 0 \neq b_n$. Then $m = \deg f \geq \deg h = n$. Put

$$(*) \qquad \tilde{f} := f - \frac{b_m}{a_n} x^{m-n} h.$$

Observe that the coefficient of x^m in $\tilde{f}$ is $b_m - \frac{b_m}{a_n} a_n = b_m - b_m = 0$. Hence $\deg \tilde{f} < m = \deg f$. So by induction, there exist $\tilde{q}, \tilde{r} \in \mathbb{F}[x]$ with

$$(**) \qquad \tilde{f} = \tilde{q} h + \tilde{r} \qquad \text{and} \qquad \deg \tilde{r} < \deg h.$$

We have

$$f \overset{(*)}{=} \tilde{f} + \frac{b_m}{a_n} x^{m-n} h \overset{(**)}{=} \tilde{q} h + \tilde{r} + \frac{b_m}{a_n} x^{m-n} h = (\tilde{q} + \frac{b_m}{a_n} x^{m-n}) h + \tilde{r}.$$

So we can choose $q = \tilde{q} + \frac{b_m}{a_n} x^{m-n}$ and $r = \tilde{r}$.

This shows this existence of q and r. For the uniqueness suppose that

$$f = qh + r = q^* h + r^*, \qquad \deg r < \deg h \qquad \text{and} \qquad \deg r^* < \deg h$$

for some $q, q^*, r, r^* \in \mathbb{F}[x]$. Then

$$(q - q^*) h = r^* - r.$$

By (7.1.4)(b), $\deg(r^* - r) \leq \max(\deg r, \deg r^*) < \deg h$ and so $\deg(q - q^*) h < \deg h$. If $q - q^* \neq 0$, then (7.1.4)(d) implies that $\deg(q - q^*) h = \deg(q - q^*) + \deg h \geq \deg h$, a contradiction. Thus $q - q^* = 0$ and so also $r^* - r = (q - q^*) h = 0h = 0$.

Hence $q = q^*$ and $r = r^*$. So q and r are unique. □

Example 7.1.6. Consider the polynomials $f = x^4 + x$ and $h = x^2 + 1$ in $\mathbb{F}_2[x]$. Find $q, r \in \mathbb{F}_2[x]$ with $f = qh + r$ and $\deg r < \deg h$.

$$\begin{array}{r|ccccccc}
\multicolumn{1}{r}{} & x^2 & + & 1 & & & & \\ \cline{2-8}
x^2 + 1 & x^4 & & & + & x & & \\ \cline{1-1}
\multicolumn{1}{r}{} & x^4 & + & x^2 & & & & \\ \cline{2-8}
\multicolumn{1}{r}{} & & & x^2 & + & x & & \\
\multicolumn{1}{r}{} & & & x^2 & & & + & 1 \\ \cline{3-8}
\multicolumn{1}{r}{} & & & & & x & + & 1
\end{array}$$

Thus

$$x^4 + x = (x^2 + 1) \cdot (x^2 + 1) + (x + 1).$$

In more compact form the above 'long division of polynomials' can be written as follows:

$$\begin{array}{r|ccccc}
\multicolumn{1}{r}{} & 1 & 0 & 1 & & \\ \cline{2-6}
101 & 1 & 0 & 0 & 1 & 0 \\ \cline{1-1}
\multicolumn{1}{r}{} & 1 & 0 & 1 & & \\ \cline{2-6}
\multicolumn{1}{r}{} & & & 1 & 1 & 0 \\
\multicolumn{1}{r}{} & & & 1 & 0 & 1 \\ \cline{3-6}
\multicolumn{1}{r}{} & & & & 1 & 1
\end{array}$$

Definition 7.1.7. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$ with $h \neq 0$. Let $f, g \in \mathbb{F}[x]$.*

$$\overline{f} \text{ is the remainder of } f \text{ when divided by } h.$$

Define an addition $\oplus$ and multiplication $\odot$ on $\mathbb{F}[x]$ as follows:

$$f \oplus g = \overline{f + g}$$

and

$$f \odot g = \overline{fg}.$$

For $n \in \mathbb{N}$, define $f^{\odot n}$ inductively by

$$f^{\odot 0} = \overline{1} \qquad \text{and} \qquad f^{\odot(n+1)} = f^{\odot n} \odot f.$$

Define

$$\mathbb{F}^h[x] := \{f \in \mathbb{F}[x] \mid \deg f < \deg h\}$$

$(\mathbb{F}^h[x], \oplus, \odot)$ is called the ring of polynomials modulo h with coefficients in $\mathbb{F}$.

Example 7.1.8. Determine the addition and multiplication in $\mathbb{R}^{x^2+1}[x]$.

Since $\deg x^2 + 1 = 2$ we have $\mathbb{R}^{x^2+1}[x] = \{a + bx \mid a, b \in \mathbb{R}\}$. We compute

$$\begin{aligned}
(a + bx) \oplus (c + dx) &= \overline{(a + bx) + (c + dx)} \\
&= \overline{(a + c) + (b + dx)} \\
&= (a + c) + (b + dx)
\end{aligned}$$

and

$$\begin{aligned}(a+bx)\odot(c+dx) &= \overline{(a+bx)\cdot(c+dx)}\\ &= \overline{ac+(ad+bc)x+bdx^2}\\ &= \overline{(ac-bd)+(ad+bc)x+b\mathrm{d}(x^2+1)}\\ &= (ac-bd)+(ad+bc)x.\end{aligned}$$

Note that this is the same addition and multiplication as for the ring of complex numbers $\mathbb{C}=\{a+bi \mid a,b\in\mathbb{R}\}$:

$$(a+bi)+(c+di)=(a+c)+(b+d)i \qquad \text{and} \qquad (a+bi)(c+di)=(ac-bd)+(ad+bc)i.$$

Lemma 7.1.9. *Let $\mathbb{F}$ be a field and $h\in\mathbb{F}[x]$ with $h\neq 0$. Let $f,g\in\mathbb{F}[x]$.*

(a) $\overline{f}\in\mathbb{F}^h[x]$.

(b) $f=\overline{f}$ *if and only if* $\deg f<\deg h$ *and if and only if* $f\in\mathbb{F}^h[x]$.

(c) $\mathbb{F}^h[x]=\{\overline{f}\mid f\in\mathbb{F}[x]\}$.

(d) *Suppose that $f,g\in\mathbb{F}^h[x]$. Then $f\oplus g=f+g$*

(e) *Let $a\in\mathbb{F}$ and suppose that $f\in\mathbb{F}^h[x]$. Then $a\odot f=af$.*

Proof. (a) By definition of $\overline{f}$, $\deg\overline{f}<\deg h$. So $\overline{f}\in\mathbb{F}^h[x]$.

(b) Suppose $f=\overline{f}$. By (a), $\overline{f}\in\mathbb{F}^h[x]$, so $f\in\mathbb{F}^h[x]$.

Suppose $f\in\mathbb{F}^h[x]$. Then $f=0h+f$ and $\deg f<\deg h$. Hence f is the remainder of f when divided by h, so $f=\overline{f}$.

(c) If $f\in\mathbb{F}^h[x]$, then (b) gives $f=\overline{f}\in\{\overline{f}\mid f\in F[x]\}$. Also by (b) $\overline{f}\in\mathbb{F}^h[x]$ for all $f\in\mathbb{F}[x]$.

(d) By (7.1.4)(b) $\deg(f+g)\le\max(\deg f,\deg g)<\deg h$ and so by (b) $\overline{f+g}=f+g$.

(e) Since $a\in\mathbb{F}$, $\deg a\le 0$. So $\deg af=\deg a+\deg f\le\deg f<\deg h$. Thus (b) gives $\overline{af}=af$. □

Remark 7.1.10. *Let $\mathbb{F}$ be a field and $h\in\mathbb{F}[x]$ with $h\neq 0$. Then $\mathbb{F}^h[x]$ is vector space over $\mathbb{F}$.*

Definition 7.1.11. *Let R be a ring and $a,b\in R$. We say that a divides b in R and write $a|b$ if $b=ra$ for some $r\in R$.*

Lemma 7.1.12. *Let $\mathbb{F}$ be a field and $h\in\mathbb{F}[x]$ with $h\neq 0$. Let $f,g\in\mathbb{F}[x]$, $a\in\mathbb{F}$ and $n\in\mathbb{N}$. Then*

(a) $0\oplus f=1\odot f=\overline{f}=\overline{\overline{f}}=f\odot 1=f\oplus 0$.

(b) $f \oplus g = \overline{f+g} = \overline{f} + \overline{g} = \overline{f} \oplus \overline{g} = \overline{f} \oplus g = f \oplus \overline{g}$.

(c) $f \odot g = \overline{fg} = \overline{\overline{f}\,\overline{g}} = \overline{f} \odot \overline{g} = \overline{f} \odot g = f \odot \overline{g}$.

(d) $\overline{f^n} = f^{\odot n} = \left(\overline{f}\right)^{\odot n} = \overline{\left(\overline{f}\right)^n}$.

(e) $-\overline{f} = \overline{-f}$.

(f) $a \odot f = \overline{af} = a\overline{f}$.

(g) $\overline{f} = 0$ *if and only if* h *divides* f *in* $\mathbb{F}[x]$.

(h) $\overline{f} = \overline{g}$ *if and only if* h *divides* $g - f$ *in* $\mathbb{F}[x]$.

Proof. Recall first that by definition of the remainder

$$(*) \qquad f = ph + \overline{f}, \quad \deg \overline{f} < \deg h, \quad g = qh + \overline{g}, \quad \text{and} \quad \deg \overline{g} < \deg h$$

for some $p, q \in \mathbb{F}[x]$.

(a) $1 \odot f = \overline{1f} = \overline{f} = \overline{0+f} = 0 \oplus f = \overline{f+0} = f \oplus 0$ and $f \odot 1 = \overline{f1} = \overline{f}$. Since $\overline{f} \in \mathbb{F}^h[x]$, (7.1.9)(b) shows that $\overline{\overline{f}} = \overline{f}$.

(b) We compute

$$f + g = (ph + \overline{f}) + (qh + \overline{g}) = (p+q)h + (\overline{f} + \overline{g}).$$

Note that

$$\deg(\overline{f} + \overline{g}) \leq \max(\deg \overline{f}, \deg \overline{g}) < \deg h$$

and so $\overline{f} + \overline{g}$ is the remainder of $f + g$ when divided by h. Hence

$$\overline{f} + \overline{g} = \overline{f+g} = f \oplus g$$

This formula applied with $\overline{f}$ in place of f gives $\overline{\overline{f}} + \overline{g} = \overline{f} \oplus g$ and with $\overline{g}$ in place of g, $\overline{f} + \overline{\overline{g}} = f \oplus \overline{g}$. As $\overline{\overline{f}} = \overline{f}$ and $\overline{\overline{g}} = \overline{g}$ we conclude that

$$\overline{f} + \overline{g} = \overline{f} \oplus g \quad \text{and} \quad \overline{f} + \overline{g} = f \oplus \overline{g}.$$

Since $\overline{f}, \overline{g} \in \mathbb{F}^h[x]$, (7.1.9)(d) gives

$$\overline{f} + \overline{g} = \overline{\overline{f} + \overline{g}} = \overline{f} \oplus \overline{g}$$

Thus (b) is proved.

(c) By definition of $f \odot g$ we know that

$$(**) \qquad f \odot g = \overline{fg}.$$

By definition of $\overline{\overline{f}\,\overline{g}}$ we have $\overline{f}\,\overline{g} = th + \overline{\overline{f}\,\overline{g}}$ for some $t \in \mathbb{F}[x]$ and $\deg \overline{\overline{f}\,\overline{g}} < \deg h$. We compute

$$fg = (ph + \overline{f})(qh + \overline{g}) = (phq + \overline{f}q + p\overline{g})h + \overline{f}\,\overline{g} = (phq + \overline{f}q + p\overline{g} + t)h + \overline{\overline{f}\,\overline{g}}$$

So $\overline{\overline{f}\,\overline{g}}$ is the remainder of fg when divided by h, that is

$$(***) \qquad \overline{fg} = \overline{\overline{f}\,\overline{g}}.$$

By definition of $\overline{f} \odot \overline{g}$ we get

$$(+) \qquad \overline{\overline{f}\,\overline{g}} = \overline{f} \odot \overline{g}.$$

By definition of $\overline{f} \odot g$ we have $\overline{f} \odot g = \overline{\overline{f}g}$. By $(***)$ applied with $\overline{f}$ in place of f, we know that $\overline{\overline{f}g} = \overline{\overline{\overline{f}}\,\overline{g}} = \overline{\overline{f}\,\overline{g}}$ and so

$$(++) \qquad \overline{f} \odot g = \overline{\overline{f}\,\overline{g}}.$$

Similarly,

$$(+++) \qquad f \odot \overline{g} = \overline{\overline{f}\,\overline{g}}$$

From $(**)$-$(+++)$, we see that (c) hold.

(d) For $n = 0$ all four expression are equal to $\overline{1}$.
Suppose (d) holds for n. Then

$$(\#) \qquad \overline{f^{(n+1)}} \overset{\text{Def}}{=} \overline{f^n f} \overset{\text{(c)}}{=} \overline{f^n} \odot f \overset{\text{Ind}}{=} f^{\odot n} \odot f \overset{\text{Def}}{=} f^{\odot(n+1)}.$$

$$\left(\overline{f}\right)^{\odot(n+1)} \overset{\text{Def}}{=} \left(\overline{f}\right)^{\odot n} \odot \overline{f} \overset{\text{Ind.}}{=} \overline{f^n} \odot \overline{f} \overset{\text{(c)}}{=} \overline{f^n} \odot f \overset{(\#)}{=} f^{\odot(n+1)}$$

and

$$\overline{(\overline{f})^{(n+1)}} \overset{\text{Def}}{=} \overline{(\overline{f})^{n}\,\overline{f}} \overset{\text{Ind.}}{=} \overline{f^{\odot n}\overline{f}} \overset{(c)}{=} f^{\odot n} \odot f \overset{\text{Def}}{=} f^{\odot(n+1)}$$

(e) $-f = (-p)h + (-\overline{f})$ and $\deg(-f) = \deg f < \deg h$, so $-\overline{f}$ is the remainder of f when divided by h.

(f) By (c), $a \odot f = \overline{af} = a \odot \overline{f}$ and by (7.1.9)(e) $a \odot \overline{f} = a\overline{f}$.

(g) Suppose that h divides f. Then $f = sh$ for some $s \in F[x]$. Thus $f = sh + 0$ and since $\deg 0 < \deg h$, we conclude that 0 is the remainder of f when divided by 0.

Suppose that 0 is the remainder of f when divided by 0. Then $f = sh + 0 = s$ for some $s \in \mathbb{F}[x]$ and so h divides f.

(h) $\overline{f} = \overline{g}$ if and only if $\overline{f} - \overline{g} = 0$ and if and only if $\overline{f - g} = 0$. By (g) the latter holds if and only if h divides $f - g$ □

Definition 7.1.13. *An ideal of a ring R is a subset S of R such that*

(i) $0 \in S$.

(ii) *If $s, t \in S$ then $s + t \in S$ and $-s \in S$.*

(iii) *If $a \in R$ and $s \in S$, then $as \in S$ and $sa \in S$.*

Example 7.1.14. Show that the set of even integers is an ideal of the ring of integers.

0 is even. Sums of even integers are even and any multiple of an even integer is even.

Lemma 7.1.15. *Let $\mathbb{F}$ be a field, $0 \neq h \in \mathbb{F}[x]$ and $I \subseteq \mathbb{F}^h[x]$. Then I is an ideal of $\mathbb{F}^h[x]$ if and only if I is an $\mathbb{F}$-subspace of $\mathbb{F}^h[x]$ and $x \odot f \in I$ for all $f \in I$.*

Proof. Let $f, g \in I$ and $a \in \mathbb{F}$.

$\Longrightarrow$: Suppose that I is an ideal of $\mathbb{F}^h[x]$. Then by definition of an ideal, $0 \in I$ and $f \oplus g \in I$. Since $f \in I$ and I is an ideal we have $a \odot f \in I$. Thus I is an $\mathbb{F}$-subspace of $\mathbb{F}^h[x]$. Also $x \odot f \overset{(7.1.12)(c)}{=} \overline{x} \odot f \in I$ and so the forward direction is proved.

$\Longleftarrow$: Suppose I is an $\mathbb{F}$-subspace of $\mathbb{F}^h[x]$ and $x \odot f \in I$ for all $f \in I$. By definition of a subspace we have $0 \in I$ and $f \oplus g \in I$.

We claim that $x^i \odot f \in I$ for all $i \in \mathbb{N}$. We have $x^0 \odot f = 1 \odot f = \overline{1\overline{f}} = \overline{f} = f$ and so the claim holds for $i = 0$. Suppose inductively that $x^i \odot f \in I$. Then by assumption also

$$(*) \qquad x \odot (x^i \odot f) \in I.$$

Hence

$$x^{i+1} \odot f = \overline{x^{i+1} f} = \overline{x(x^i f)} = x \odot \overline{x^i f} \overset{(7.1.12)(c)}{=} x \odot (x^i \odot f) \overset{(*)}{\in} I.$$

So indeed $x^i \odot f \in I$ for all $i \in \mathbb{N}$ and $f \in I$.

Let $k \in \mathbb{F}^h[x]$. Then $\deg k < \deg h = n$ and so $k = \sum_{i=0}^{n-1} k_i x^i$ with $k_i \in \mathbb{F}$. Thus

$$k \odot f = \left(\sum_{i=0}^{n-1} k_i x^i \right) \odot f = \sum_{i=0}^{n-1} (k_i x^i) \odot f \overset{7.1.12}{=} \sum_{i=0}^{n-1} k_i \left(x^i \odot f \right).$$

Since each $x^i \odot f \in I$ and I is an $\mathbb{F}$-subspace of $\mathbb{F}^h[x]$ we conclude that $k \odot f \in I$. Since $\mathbb{F}^h[x]$ is commutative, also $f \odot k \in I$ and so I is an ideal of $\mathbb{F}^h[x]$. □

Definition 7.1.16. *Let $n \in \mathbb{Z}^+$. Then $V^n[x] := \mathbb{F}_2^{x^n-1}[x]$.*

Lemma 7.1.17. *Let $\mathbb{F}$ be a field, $n \in \mathbb{Z}^+$ and for $f \in \mathbb{F}[x]$ let $\overline{f}$ be remainder of f when divided by $x^n - 1$.*

(a) *Let $i, j \in \mathbb{N}$ with $0 \leq j < n$. Then $\overline{x^{ni+j}} = x^j$.*

(b) *Let $f \in \mathbb{F}[x]$ with*

$$f = \sum_{i=0}^{m} \sum_{j=0}^{n-1} a_{ij} x^{in+j}$$

for some $a_{ij} \in \mathbb{F}$. Then

$$\overline{f} = \sum_{j=0}^{n-1} \left(\sum_{i=0}^{m} a_{ij} \right) x^j$$

Proof. (a): We first compute $\overline{x^n}$: Since $x^n = (x^n - 1) + 1$ and $\deg 1 = 0 < n = \deg(x^n - 1)$ we have

$$(*) \qquad \overline{x^n} = 1.$$

$$\overline{x^{ni+j}} = \overline{(x^n)^i x^j} \overset{7.1.12}{=} \overline{\overline{x^n}^i x^j} \overset{(*)}{=} \overline{1^i x^j} = \overline{x^j} = x^j.$$

So (a) holds.

(b): We compute

$$\overline{\sum_{i=0}^{m} \sum_{j=0}^{n-1} a_{ij} x^{in+j}} \overset{7.1.12}{=} \sum_{i=0}^{m} \sum_{j=0}^{n-1} a_{ij} \overline{x^{in+j}} \overset{(a)}{=} \sum_{i=0}^{m} \sum_{j=0}^{n-1} a_{ij} x^j = \sum_{j=0}^{n-1} \left(\sum_{i=0}^{m} a_{ij} \right) x^j$$

and so (b) holds. □

Example 7.1.18. Compute the remainder of $f = 1 + x + x^4 + x^5 + x^7 + x^{11} + x^{17} + x^{28}$ when divided by $x^6 + 1$ in $\mathbb{F}_2[x]$.

$$f = 1 + x + x^4 + x^5 + x^{6+1} + x^{6+5} + x^{6\cdot 2+5} + x^{6\cdot 4+4}$$

and so

$$\overline{f} = 1 + x + x^4 + x^5 + x + x^5 + x^5 + x^4 = 1 + 2x + 2x^4 + 3x^5 = 1 + x^5.$$

Definition 7.1.19. *Let $a = a_0 \dots a_{n-1} \in \mathbb{F}_2^n$ and $C \subseteq \mathbb{F}_2^n$.*

(a) *For $0 \leq i < n$ define*

$$a^{(i)} = a_{n-i} \dots a_{n-1} a_0 a_1 \dots a_{n-1-i},$$

that is

$$\begin{array}{rcccccc} a = a^{(0)} = & a_0 & a_1 & a_2 & \dots & a_{n-2} & a_{n-1} \\ a^{(1)} = & a_{n-1} & a_0 & a_1 & \dots & a_{n-3} & a_{n-2} \\ a^{(2)} = & a_{n-2} & a_{n-1} & a_0 & \dots & a_{n-4} & a_{n-3} \\ & & & \vdots & & & \\ a^{(n-1)} = & a_1 & a_2 & a_3 & \dots & a_{n-1} & a_0 \end{array}$$

$a^{(i)}$ is called the cyclic i-shift of a.

(b) *$\langle a \rangle$ is the subspace of $\mathbb{F}_2^n$ spanned by $a^{(0)}, a^{(1)}, \dots, a^{(n-1)}$, that is*

$$\langle a \rangle = \left\{ \sum_{i=0}^{n-1} k_i a^{(i)} \,\middle|\, k_i \in \mathbb{F}_2 \right\} = \operatorname{Col}\left(\begin{bmatrix} a^{(0)} & a^{(1)} & \dots & a^{(n-1)} \end{bmatrix} \right).$$

$\langle a \rangle$ is called the binary cyclic code generated by a.

(c) $a(x) := \sum_{i=0}^{n-1} a_i x^i = a_0 + a_1 x + \dots + a_{n-1} x^{n-1} \in V^n[x]$.

(d) $C(x) := \{ a(x) \mid a \in C \} \subseteq V^n[x]$.

Example 7.1.20. Compute the binary cyclic code generate by 1100.

We have

$$\langle a \rangle = \operatorname{Col}\left(\begin{bmatrix} a^{(0)} & a^{(1)} & a^{(2)} & a^{(3)} \end{bmatrix} \right) = \operatorname{Col}\left(\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \right).$$

We use the method from 6.3.11 to compute a check matrix for $\langle a\rangle$ in standard form.

$$\begin{bmatrix}1&0&0&1\\1&1&0&0\\0&1&1&0\\0&0&1&1\end{bmatrix} \xrightarrow{C_1+C_4\to C_4} \begin{bmatrix}1&0&0&0\\1&1&0&1\\0&1&1&0\\0&0&1&1\end{bmatrix} \xrightarrow{\substack{C_2+C_1\to C_1\\ C_2+C_4\to C_4}} \begin{bmatrix}1&0&0&0\\0&1&0&0\\1&1&1&1\\0&0&1&1\end{bmatrix} \xrightarrow{\substack{C_3+C_1\to C_1\\ C_3+C_2\to C_2\\ C_3+C_4\to C_4}} \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\1&1&1&0\end{bmatrix} \to \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&1&1\end{bmatrix}$$

Thus

$$\langle a\rangle = \operatorname{Col}\left(\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&1&1\end{bmatrix}\right) = \left\{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&1&1\end{bmatrix}\cdot\begin{pmatrix}x\\y\\z\end{pmatrix}\,\middle|\,\begin{pmatrix}x\\y\\z\end{pmatrix}\in F_2^3\right\} = \left\{\begin{pmatrix}x\\y\\z\\x+y+z\end{pmatrix}\,\middle|\,\begin{pmatrix}x\\y\\z\end{pmatrix}\in F_2^3\right\}$$

$$= \{0000, 1001, 01001, 0011, 1100, 1011, 0110, 1111\}$$

Observe that the above generating matrix has standard form $E = \begin{bmatrix}I_3\\A\end{bmatrix}$ where $A = [1\ 1\ 1]$. Hence

$$H = [A\ I_1] = [1\ 1\ 1\ 1]$$

is a check matrix of $\langle a\rangle$. So $\langle a\rangle$ consists of all even binary strings of length 4.

Lemma 7.1.21. *Let $a\in\mathbb{F}_2^n$ and $C\subseteq\mathbb{F}_2^n$.*

(a) *The function $\mu:\mathbb{F}_2^n\to V^n[x], a\mapsto a(x)$ is an isomorphism of $\mathbb{F}_2$-vector spaces.*

(b) *$a\in C$ if and only if $a(x)\in C(x)$.*

(c) *Let $a\in\mathbb{F}_2^n$ and $0\le i<n$. Then $a^{(i)}(x) = x^i\odot a(x)$.*

(d) *C is a cyclic code if and only if $C(x)$ is an ideal of $V^n[x]$.*

Proof. (a): Any $a\in\mathbb{F}_2^n$ can be uniquely written as $a_0\dots a_{n-1}$ with $a_i\in\mathbb{F}_2$. Any $f\in V^n[x]$ can be uniquely written as $a_0+a_1x+\dots a_{n-1}x^{n-1}$. So μ is a bijection. Let $a,b\in\mathbb{F}_2^n$. Then

$$\mu(a+b) = (a+b)(x) = (a_0+b_0)+(a_1+b_1)x+\dots(a_{n-1}+b_{n-1})x^{n-1} = a(x)+b(x) = \mu(a)+\mu(b).$$

So μ is also $\mathbb{F}_2$-linear. Thus μ is an isomorphism.

(b) By (a) we know that μ is a bijection. So (b) holds.

(c): Let $a = a_0 \dots a_{n-1}$ with $a_j \in \mathbb{F}_2$. Then

$$\begin{array}{ll}
x^i \odot a(x) & \\
=\overline{x^i \cdot (a_0 + a_1 x + \dots + a_{n-1}x^{n-1})} & \mid \text{Definitions of } \odot \text{ and } a(x) \\
=\overline{a_0 x^i + a_1 x^{i+1} + \dots + a_{n-1-i}x^{n-1} + a_{n-i}x^n + \dots a_{n-1}x^{n-1+i}} & \mid \text{Definition of } \cdot \\
=a_0 x^i + a_1 x^{i+1} + \dots + a_{n-1-i}x^{n-1} + a_{n-i} + \dots + a_{n-1}x^{i-1} & \mid 7.1.17 \\
=a_{n-i} + \dots + a_{n-1}x^{i-1} + a_0 x^i + \dots a_{n-1-i}x^{n-1} & \mid \text{General Commutative Law} \\
=a^{(i)}(x). & \mid \text{Definitions of } a^i \text{ and } a^i(x)
\end{array}$$

(d): Note that $\vec{0} \in C$ if and only if $0 = \vec{0}(x) \in C(x)$. Let $a, b \in C$. Then $a + b \in C$ if and only if $(a+b)(x) \in C(x)$ and so if and only if $a(x) + b(x) \in C(x)$. Thus

$(*)$ *C is a subspace of $\mathbb{F}_2^n$ if and only if $C(x)$ is a subspace of $V^n[x]$.*

By (c) $a^{(1)}(x) = x \odot a(x)$ and so (b) implies:

$(**)$ *$a^{(1)} \in C$ if and only if $x \odot a(x) \in C(x)$.*

Thus

$$\begin{array}{lll}
 & C \text{ is cyclic} & \\
\iff & C \text{ is a subspace of } \mathbb{F}_2^n \text{ and } a^{(1)} \in C \text{ for all } a \in C & \mid \text{Definition of cyclic} \\
\iff & C(x) \text{ is a subspace of } V^n[x] \text{ and } x \odot a(x) \in C(x) \text{ for all } a(x) \in C(x) & \mid (*) \text{ and } (**) \\
\iff & C(x) \text{ is an ideal of } V^n[x] & \mid 7.1.15
\end{array}$$

□

Definition 7.1.22. *Let R be a commutative ring with identity and $a \in R$. Define*

$$\langle a \rangle := \{ ra \mid r \in R \}.$$

$\langle a \rangle$ is called the ideal of R generated by a.

Lemma 7.1.23. *Let R be a commutative ring with identity and $a \in R$. Then $\langle a \rangle$ is the smallest ideal of R containing a, that is*

(a) *$a \in \langle a \rangle$,*

(b) *$\langle a \rangle$ is an ideal of R, and*

(c) *$\langle a\rangle \subseteq I$ for any ideal I of R with $a \in I$.*

Proof. (a) $a = 1a \in \langle a\rangle$.

(b) We have $0 = 0a \in \langle a\rangle$. Also for any $r, s \in R$, $ra + sa = (r + s)a \in \langle a\rangle$ and $(ra)s = s(ra) = (sr)a \in \langle a\rangle$.

(c) By definition of an ideal, $ra \in I$ for all $r \in R$ and so $\langle a\rangle \subseteq I$. □

Lemma 7.1.24. *Let $\mathbb{F}$ be a field, $h \in \mathbb{F}[x]$ with $h \neq 0$ and $f \in \mathbb{F}^h[x]$. Let $\langle f\rangle$ be the ideal of $\mathbb{F}^h[x]$ generated by f.*

(a) *Put $n := \deg(h)$. Then $\langle f\rangle$ is the $\mathbb{F}$-subspace of $\mathbb{F}^h[x]$ spanned by*

$$1 \odot f, \quad x \odot f, \quad \dots, \quad x^{n-1} \odot f.$$

(b) *$\langle f\rangle = \{g \odot f \mid g \in \mathbb{F}[x]\}$.*

Proof. (a) Let $g = \sum_{i=0}^{n-1} g_i x^i \in \mathbb{F}^h[x]$. Then

$$g \odot f = g_0(1 \odot f) + g_1(x \odot f) + \ldots + g_{n-1}(x^{n-1} \odot f)$$

and so the elements of $\langle f\rangle$ are exactly the $\mathbb{F}$-linear combinations of $1 \odot f, x \odot f, \dots, x^{n-1} \odot f$.

(b)

$$\begin{aligned}
\langle f\rangle &= \{g \odot f \mid g \in \mathbb{F}^h[x]\} && \mid \text{Definition of } \langle f\rangle \\
&= \{\overline{g} \odot f \mid g \in \mathbb{F}[x]\} && \mid \mathbb{F}^h[x] = \{\overline{g} \mid g \in \mathbb{F}[x]\} \text{ by(7.1.9)(c)} \\
&= \{g \odot f \mid g \in \mathbb{F}[x]\} && \mid \text{(7.1.12)(c)}
\end{aligned}$$

□

Example 7.1.25. Find the ideal of $V^3[x]$ generated by $f = 1 + x^2$ and determined the corresponding cyclic code.

$$\begin{aligned}
1 \odot (1 + x^2) &= 1 + x^2 \\
x \odot (1 + x^2) &= \overline{x + x^3} = \overline{x + x^{3+0}} = 1 + x
\end{aligned}$$

Note that $(1 + x^2) + (1 + x) = x + x^2$. So

$$\langle 1 + x^2\rangle = \{0, 1 + x^2, 1 + x, x + x^2\}.$$

The corresponding cyclic code is

$$\{000, 101, 110, 011\}$$

Lemma 7.1.26. *Let $\mathbb{F}$ be a field and $f, g, t \in \mathbb{F}[x]$ with $t \neq 0$.*

(a) *$ft = gt$ if and only if $f = g$.*

(b) *$f \mid g$ if and only if $ft \mid gt$.*

Proof. (a): $\Longrightarrow$: Suppose $ft = gt$. Then $(f - g)t = ft - gt = 0$ and so

$$\deg(f - g) + \deg t = \deg\big((f - g)t\big) = \deg 0 = -\infty.$$

Since $t \neq 0$, we have $\deg t \in \mathbb{N}$. As $\deg(f - g) + \deg t = -\infty$ we conclude that $\deg(f - g) = -\infty$. Hence $f - g = 0$ and $f = g$.

$\Longleftarrow$: If $f = g$, then clearly $ft = gt$.

(b):

$$\begin{array}{llll} & ft \mid gt & & \\ \Longleftrightarrow & gt = l(ft) & \text{for some } l \in \mathbb{F}[x] & \mid \text{definition of `divide'} \\ \Longleftrightarrow & gt = (lf)t & \text{for some } l \in \mathbb{F}[x] & \\ \Longleftrightarrow & g = lf & \text{for some } l \in \mathbb{F}[x] & \mid \text{(a)} \\ \Longleftrightarrow & f \mid g & & \mid \text{definition of `divide'} \end{array}$$

□

Theorem 7.1.27. *Let $\mathbb{F}$ be a field, let $h \in \mathbb{F}[x]$ with $h \neq 0$ and let I be an ideal of $\mathbb{F}^h[x]$. Observe that $\overline{h} = 0 \in I$ and choose $g \in \mathbb{F}[x]$ of minimal degree subject to $\overline{g} \in I$ and $g \neq 0$.*

(a) *Let $f \in \mathbb{F}[x]$. Then $\overline{f} \in I$ if and only if g divides f in $\mathbb{F}[x]$.*

(b) *$I = \langle \overline{g} \rangle$ is the ideal of $\mathbb{F}^h[x]$ generated by $\overline{g}$.*

(c) *g divides h in $\mathbb{F}[x]$.*

According to (c) choose $t \in \mathbb{F}[x]$ with $h = tg$. Put $k := \deg t$.

(d) *$t \neq 0$, and $k \in \mathbb{N}$.*

(e) *Let $f, f^* \in \mathbb{F}[x]$. Then $f \odot g = f^* \odot g$ if and only if t divides $f^* - f$.*

(f) *$I = \{sg \mid s \in \mathbb{F}[x], \deg s < k\}$. Moreover, for each $f \in I$ there exists a unique $s \in \mathbb{F}[x]$ with $f = sg$ and $\deg s < k$.*

(g) *$(g, xg, x^2g, \ldots, x^{k-1}g)$ is an $\mathbb{F}$-basis for I.*

(h) $\dim_{\mathbb{F}} I = k$.

(i) *Let* $f \in \mathbb{F}[x]$. *Then* $\overline{f} \in I$ *if and only if* $f \odot t = 0$.

(j) *Suppose that* $h = x^n - 1$ *and let* $f \in \mathbb{F}^h[x]$. *Then* $f \in I$ *if and only if the coefficient of* x^i *in* ft *is* 0 *for all* $k \leq i < n$.

Proof. (a): $\Longrightarrow$: Suppose $\overline{f} \in I$. Let $f = qg + r$ with $q, r \in \mathbb{F}[x]$ and $\deg r < \deg g$. Then $r = f - qg$ and so

$$\overline{r} = \overline{f + (-q)g} \overset{7.1.12}{=} \overline{f} \oplus \overline{-q} \odot \overline{g}.$$

Since both $\overline{f}$ and $\overline{g}$ are in I and I is an ideal, this gives $\overline{r} \in I$. As $\deg r < \deg g$, the minimal choice of $\deg g$ shows that $r = 0$ and so $g \mid f$.

$\Longleftarrow$: Suppose that $g \mid f$. Then $f = qg$ for some $q \in F[x]$ and so $\overline{f} = \overline{qg} = \overline{q} \odot \overline{g}$. Recall that $\overline{q} \in \mathbb{F}^h[x]$, $\overline{g} \in I$ and I is an ideal of $\mathbb{F}^h[x]$. Thus $\overline{f} = \overline{q} \odot \overline{g} \in I$.

(b) Let $f \in I$. Then $f \in \mathbb{F}^h[x]$ and so $\overline{f} \overset{(7.1.9)(b)}{=} f \in I$. Hence (a) shows that g divides f and so $f = eg$ for some $e \in \mathbb{F}[x]$. Therefore $f = \overline{f} = \overline{e} \odot \overline{g}$ and so $f \in \langle \overline{g} \rangle$. Hence $I \subseteq \langle \overline{g} \rangle$.
Since $\overline{g} \in I$ and I is an ideal, (7.1.23)(c) shows $\langle \overline{g} \rangle \subseteq I$. Thus $I = \langle \overline{g} \rangle$.

(c): Note that $\overline{h} = 0 \in I$ and so $g | h$ by (a).

(d) Since $h \neq 0$ and $h = tg$ we get $t \neq 0$. As $t \neq 0$ we have $k = \deg(t) \in \mathbb{N}$.

(e): We have

$$\begin{array}{lll}
 & f \odot g = f^* \odot g & \\
\Longleftrightarrow & \overline{fg} = \overline{f^* g} & - \text{ definition of } \odot \\
\Longleftrightarrow & h \text{ divides } fg - f^* g & - \text{ (7.1.12)(h)} \\
\Longleftrightarrow & tg \text{ divides } (f - f^*)g & - h = tg \text{ by (c)} \\
\Longleftrightarrow & t \text{ divides } f - f^* & - \text{7.1.26}
\end{array}$$

(f): Let $s \in \mathbb{F}[x]$ with $\deg s < k$. Then

$$\deg sg \overset{(7.1.4)(d)}{=} \deg s + \deg g < \deg t + \deg g \overset{(7.1.4)(d)}{=} \deg tg = \deg h.$$

Thus $\overline{sg} = sg$ by (7.1.9)(b) and so

$$(*) \qquad s \odot \overline{g} \overset{7.1.12}{=} s \odot g \overset{\text{def.} \odot}{=} \overline{sg} = sg.$$

Hence $sg = s \odot \overline{g} \in \langle \overline{g} \rangle \in I$. Put

$$J := \{sg \mid s \in \mathbb{F}[x], \deg s < k\}$$

Then

$$(**) \qquad J \subseteq I.$$

Now let $f \in \mathbb{F}^h[x]$ and let $f = qt + s$ with $q, s \in \mathbb{F}[x]$ and $\deg s < \deg t$. Then $\deg s < k$, $f - s = qt$, and t divides $f - s$ in $\mathbb{F}[x]$. Thus (e) gives $f \odot g = s \odot g$. So

$$(***) \qquad f \odot \overline{g} \overset{7.1.12}{=} f \odot g = s \odot g \overset{(*)}{=} sg$$

and so

$$I \overset{\text{(b)}}{=} \langle \overline{g} \rangle = \{f \odot \overline{g} \mid f \in \mathbb{F}^h[x]\} \overset{(***)}{\subseteq} \{sg \mid s \in \mathbb{F}[x], \deg s < k\} = J.$$

By $(**)$ $J \subseteq I$ and so $I = J$. In particular, for each $f \in I$ there exists $s \in \mathbb{F}[x]$ with $\deg s < k$ and $f = sg$. Suppose that $\tilde{s} \in \mathbb{F}[x]$ with $\deg \tilde{s} < k$ and $f = \tilde{s}g$. Then $sg = \tilde{s}g$ and since $g \neq 0$ we get $s = \tilde{s}$, see (7.1.26)(a). Hence s is unique.

(g): Let $f \in I$. By (f) there exists a unique $s \in \mathbb{F}[x]$ with $\deg s < k$ and $f = sg$. Note that $s = \sum_{i=0}^{k-1} s_i x^i$ for unique $s_i \in \mathbb{F}$. Then

$$f = sg = \sum_{i=0}^{k-1} s_i x^i g.$$

So for each $f \in I$ there exists a unique $(s_0, \ldots, s_{k-1}) \in \mathbb{F}^k$ with $f = \sum_{i=0}^{k-1} s_i x^i g$. Thus (g) holds.

(h): Note that $\dim_{\mathbb{F}} I$ is the size of any $\mathbb{F}$-basis of I. So (h) follows from (g).

(i):

$$\begin{array}{lll}
 & \overline{f} \in I & \\
\iff & g \mid f & - (a) \\
\iff & gt \mid ft & - \text{7.1.26} \\
\iff & h \mid ft & - h = gt \text{ by (c)} \\
\iff & \overline{ft} = 0 & - \text{(7.1.12)(h)} \\
\iff & f \odot t = 0 & - \text{definition of } \odot
\end{array}$$

(j) Let $f \in \mathbb{F}^h[x]$ and choose $q, r \in \mathbb{F}[x]$ such that

$$(+) \qquad f = qg + r$$

and

(++) $$\deg r < \deg g.$$

Note that $f = \overline{f}$ and so

(+++) $$f \in I \iff \overline{f} \in I \overset{\text{(a)}}{\iff} g|f \overset{\text{(7.1.12)(g)}}{\iff} r = 0.$$

Next we show

(#) $$\deg(qg) \le \deg f$$

Indeed, if $q = 0$, then also $qg = 0$ and so $\deg(qg) = -\infty \le \deg f$. Suppose $q \neq 0$. Then $\deg q \ge 0$ and so

$$\deg(qg) \overset{\text{(7.1.4)(d)}}{=} \deg(q) + \deg(g) \ge \deg(g) > \deg r.$$

Hence (7.1.4)(c) shows that $\deg f = \deg(qg + r) = \deg(qg)$.

(##) $$\deg q < k.$$

Note that

$$\deg q + \deg g \overset{\text{(7.1.4)(d)}}{=} \deg(qg) \overset{(\#)}{\le} \deg f < \deg h = \deg(tg) \overset{\text{(7.1.4)(d)}}{=} \deg t + \deg g = k + \deg g.$$

As $g \neq 0$ we have $\deg g \in \mathbb{N}$ and we conclude that $\deg q < k$.

(###) $$\deg rt < n$$

We compute

$$\deg rt \overset{\text{(7.1.4)(d)}}{=} \deg r + \deg t \overset{(++)}{<} \deg g + \deg t \overset{\text{(7.1.4)(d)}}{=} \deg gt = \deg h = \deg(x^n - 1) = n.$$

($\diamond$) $r = 0$ *if and only if* $\deg rt < k$.

If $r \neq 0$, then $\deg r \ge 0$ and so $\deg rt \overset{\text{(7.1.4)(d)}}{=} \deg r + \deg t \ge \deg t = k$. If $r = 0$, then $\deg(rt) = \deg 0 = -\infty < k$.

($\diamond\diamond$) $\deg(rt) < k$ *if and only if the coefficient of* x^i *in* rt *is* 0 *for each* $i \in \mathbb{Z}$ *with* $k \le i < n$.

By (###) $\deg rt < n$ and so ($\diamond\diamond$) holds.

($\diamond \diamond \diamond$) *For each $i \in \mathbb{Z}$ with $k \le i < n$, the coefficient of x^i in ft is the same as the coefficient of x^i in rt.*

We compute

$$ft = (qg + r)t = qgt + rt = qh + rt = q \cdot (x^n - 1) + rt = qx^n - q + rt$$

Let $i \in \mathbb{Z}$ with $k \le i < n$. Since $i < n$, the coefficient of x^i in qx^n is 0. By (##) we have $\deg q < k$. Since $i \ge k$ we conclude that the coefficient of x^i in q is 0. Hence ($\diamond \diamond \diamond$) holds.

From (+++), ($\diamond$), ($\diamond\diamond$) and ($\diamond \diamond \diamond$) we see that $f \in I$ if and only if the coefficent of x^i in ft is 0 for all $i \in \mathbb{Z}$ with $k \le i < n$. Hence (j) holds. □

Definition 7.1.28. *Let $\mathbb{F}$ be a field.*

(a) *Let $f \in \mathbb{F}[x]$. If $f = \sum_{i=0}^m f_i x^i$ with $f_m \ne 0$, define* $\mathrm{lead}(f) := f_m$. *If $f = 0$ define* $\mathrm{lead}(f) = 0$. *Then* $\mathrm{lead}(f)$ *is called the leading coefficient of f.*

(b) *A monic polynomial is a polynomial with leading coefficient* 1.

(c) *Let $h \in \mathbb{F}[x]$ with $h \ne 0$ and let I be an ideal of $\mathbb{F}^h[x]$. Let $g \in \mathbb{F}[x]$ be a monic polynomial of minimal degree with $\overline{g} \in I$. Then g is called a canonical generator for I.*

Example 7.1.29. Determine the cyclic code generated by 0101 and find a canonical generator for the corresponding ideal of $V^4[x]$.

The cyclic shifts of 0101 are

$$0101, \quad 1010.$$

Thus

$$\langle 1010 \rangle = \{0000, 0101, 1010, 1111\}.$$

The non-zero codeword with the most trailing zeros is 1010. So the canonical generator is $1 + x^2$.

Lemma 7.1.30. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$ with $h \ne 0$.*

(a) *Let I be an ideal of $\mathbb{F}^h[x]$. Then there exists a unique canonical generator g of I.*

(b) *$I = \langle \overline{g} \rangle$ in $\mathbb{F}^h[x]$ and g divides h in $\mathbb{F}[x]$.*

(c) *If $I \ne \{0\}$, then* $\deg g < \deg h$ *and $0 \ne g = \overline{g} \in I$. If $I = \{0\}$, then* $\deg g = \deg h$, *$\overline{g} = 0$ and* $g = \mathrm{lead}(h)^{-1} h$.

(d) *Suppose $I \neq 0$. Then g is a non-zero polynomial of minimal degree in I. Moreover, g is the unique monic polynomial of minimal degree in I.*

Proof. Choose $g \in \mathbb{F}[x]$ of minimal degree subject to $g \neq 0$ and $\overline{g} \in I$. Since I is an ideal we get $\overline{(\text{lead}g)^{-1}g} = (\text{lead}g)^{-1}\overline{g} \in I$. So replacing g by $(\text{lead}g)^{-1}g$ we may assume that g is monic.

(a) To show that g is a canonical generator for I, let f be a monic polynomial with $\overline{f} \in I$. Then $f \neq 0$ and so $\deg f \geq \deg g$, by choice of g. Hence g is a canonical generator.

Let f be any canonical generator for I. Then the minimal choice of the degree of a canonical generator shows that $\deg g \leq \deg f$ and vice versa. Thus $\deg g = \deg f$. Since both g and f are monic we conclude that $\deg(g - f) < \deg g$. Also $\overline{g - f} = \overline{g} - \overline{f} \in I$ and the minimal choice of $\deg g$ gives $g - f = 0$. Thus $f = g$ and g is the unique canonical generator for I.

(b): See (7.1.27)(b),(c).

For (c) and (d) we prove:

$(*)$ *Let $f \in I$ with $f \neq 0$. Then $\overline{f} = f \in I$, $\deg g \leq \deg f < \deg h$ and $g = \overline{g} \in I$.*

Since $f \in I$ we have $f \in \mathbb{F}^h[x]$. Hence $\deg f < \deg h$ and $\overline{f} = f \in I$. As $f \neq 0$, the minimal choice of $\deg g$ implies that $\deg g \leq \deg f$. Thus $\deg g < \deg h$ and so $g = \overline{g} \in I$.

(c): Suppose that $I \neq 0$. Then we can choose $f \in I$ with $I \neq 0$. Then $(*)$ shows that $\deg g < \deg h$ and $g = \overline{g} \in I$.

Suppose that $I = 0$. Then $\overline{g} = 0$. Thus $h|g$ and so $\deg h \leq \deg g$. As $\overline{h} = 0 \in I$, the minimal choice of $\deg g$ shows that $\deg h = \deg g$. Hence $\text{lead}(h)^{-1}h$ fulfills the assumptions on g. It follows that $\text{lead}(h)^{-1}h$ is also a canonical generator of I and so $g = \text{lead}(h)^{-1}h$.

(d): Let $f \in I$ with $f \neq 0$. By $(*)$, $\deg f \leq \deg g$ and $g \in I$. As g is monic, g is monic polynomial of minimal degree in I. If f is a monic polynomial of minimal degree in I, we get $\deg f \leq \deg g$ As $\overline{f} = f \in I$ and g is a canonic generator, also f is a canonic generator. So $f = g$ by (a). □

Lemma 7.1.31. *Let $\mathbb{F}$ be a field, $h \in \mathbb{F}[x]$ with $h \neq 0$ and let $f \in \mathbb{F}[x]$ be monic. Then f is the canonical generator for the ideal $\langle\overline{f}\rangle$ in $\mathbb{F}^h[x]$ if and only if $f \mid h$ in $\mathbb{F}[x]$.*

Proof. If f is the canonical generator for $\langle\overline{f}\rangle$, then (7.1.27)(c) shows that $f \mid h$.

Conversely suppose that $f \mid h$ and let $e \in \mathbb{F}[x]$ with $\overline{e} \in \langle\overline{f}\rangle$. By definition of $< \overline{f}\rangle$ we get

$$\overline{e} = d \odot \overline{f} = \overline{df}$$

for some $d \in \mathbb{F}^h[x]$. Thus 7.1.12 shows that h divides $e - df$. Also $f|h$ and so

$$e - df = lh \quad \text{and} \quad h = kf$$

for some $k, l \in \mathbb{F}[x]$.

$$e = df + lh = df + l(kf) = (d + lk)f$$

Thus $f|e$ and so $\deg f \geq \deg e$ Hence f is a canonical generator of $\langle \overline{f} \rangle$. □

Corollary 7.1.32. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$ with $h \neq 0$. Then the function*

$$\alpha : \qquad g \quad \mapsto \quad \langle \overline{g} \rangle$$

is a 1-1 correspondence between monic divisors of h in $\mathbb{F}[x]$ and the ideals of $\mathbb{F}^h[x]$, with inverse

$$\beta : \qquad I \quad \mapsto \quad \text{the canonical generator for } I.$$

Proof. We just need to show β is an inverse of α.

Let g be a monic divisor of h. Then $\alpha(g) = \langle \overline{g} \rangle$. By 7.1.31 g is the canonical generator for $\langle \overline{g} \rangle$. Thus

$$g = \beta(\langle \overline{g} \rangle) = \beta(\alpha(g)).$$

Now let I be an ideal of $\mathbb{F}^h[x]$. Let g be the canonical generator for I, so $g = \beta(I)$. By (7.1.27)(c) g is a divisor of h and by (7.1.27)(b) we have $I = \langle \overline{g} \rangle$. Thus

$$I = \langle \overline{g} \rangle = \alpha(g) = \alpha(\beta(I)).$$

□

Theorem 7.1.33. *Let $C \subseteq \mathbb{F}_2^n$ be a binary cyclic code. Let g be the canonical generator of the ideal $C(x)$ in $V^n[x]$. Let $t \in \mathbb{F}_2[x]$ with $gt = x^n - 1$. Put $k := \deg t$ and $m := n - k = \deg g$. Let*

$$g = c_0 + c_1 x + \ldots + c_m x^m \qquad \textit{and} \qquad t = h_0 + h_1 x + \ldots h_k x^k.$$

with $c_i, h_i \in \mathbb{F}_2$.

(a) *The $n \times k$ matrix*

$$E = \begin{bmatrix} c_0 & 0 & 0 & \dots & 0 & 0 \\ c_1 & c_0 & 0 & \dots & 0 & 0 \\ \vdots & c_1 & c_0 & \dots & \vdots & \vdots \\ \vdots & \vdots & c_1 & \dots & 0 & \vdots \\ c_{m-1} & \vdots & \vdots & \dots & c_0 & 0 \\ c_m & c_{m-1} & \vdots & \dots & c_1 & c_0 \\ 0 & c_m & c_{m-1} & \dots & \vdots & c_1 \\ 0 & 0 & c_m & \dots & \vdots & \vdots \\ \vdots & 0 & 0 & \dots & c_{m-1} & \vdots \\ \vdots & \vdots & \vdots & \dots & c_m & c_{m-1} \\ 0 & 0 & 0 & \dots & 0 & c_m \end{bmatrix}$$

is a generating matrix for C. (Note here that if $g = x^n - 1$, then $m = n$, $k = 0$, $E = [\,]$ and $C = \{\vec{0}\}$.)

(b) *The $m \times n$ matrix*

$$H = \begin{bmatrix} h_k & h_{k-1} & h_{k-2} & \dots & h_1 & h_0 & 0 & 0 & \dots & 0 \\ 0 & h_k & h_{k-1} & \dots & h_2 & h_1 & h_0 & 0 & \dots & 0 \\ 0 & 0 & h_k & \dots & h_3 & h_2 & h_1 & h_0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & h_k & \dots & \dots & h_1 & h_0 & 0 \\ 0 & 0 & 0 & \dots & 0 & h_k & \dots & h_2 & h_1 & h_0 \end{bmatrix}$$

is a check matrix for C.

Proof. (a) Let $c \in \mathbb{F}_2^n$ with $c(x) = g$. So $c = c_0 \dots c_m 0 \dots 0$ and c is the first column of E. By 7.1.27 $C(x)$ is spanned by the polynomials $x^i g$, $0 \le i < k$. Thus C is spanned by the cyclic shifts $c^{(i)}$, $0 \le i < k$. Since the columns of E are the $c^{(i)}$, (a) holds.

(b) Let $d = d_0 d_1 \dots d_{n-1} \in \mathbb{F}_2^n$. By 7.1.27, $d(x) \in C(x)$ if and only if the coefficient a_s of x^s in $t \cdot d(x)$ is equal to 0 for all $s \in \mathbb{Z}$ with $k \le s < n$. Since $\deg t = k$ we have $h_l = 0$ for $l > k$. Thus

$$a_s = \sum_{l=0}^{s} h_l d_{s-l} = \sum_{l=0}^{k} h_l d_{s-l} = h_k d_{s-k} + h_{k-1} d_{s-k+1} + \ldots + h_1 d_{s-1} + h_0 d_s.$$

Hence

$$\begin{array}{llllllllll}
a_k & = h_kd_0 + h_{k-1}d_1 + h_{k-2}d_2 + \ldots + h_1d_{k-1} + & h_0d_k & & & & & & \\
a_{k+1} & = \quad h_kd_1 \;+ h_{k-1}d_2 + \ldots + h_2d_{k-1} + & h_1d_k & + & h_0d_{k+1} & & & & \\
a_{k+2} & = \quad\quad h_kd_2 \;+ \ldots + h_3d_{k-1} + & h_2d_k & + & h_1d_{k+1} & + h_0d_{k+2} & & & \\
\vdots & \vdots \quad \ddots\;\ddots\;\ddots\;\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\
a_{n-2} & = & h_kd_{n-k-2} + & \ldots & + & \ldots & + h_1d_{n-3} + h_0d_{n-2} & & \\
a_{n-1} & = & & & h_kd_{n-k-1} + & \ldots & + h_2d_{n-3} + h_1d_{n-2} + h_0d_{n-1} & &
\end{array}$$

and so

$$\begin{pmatrix} a_k \\ a_{k+1} \\ \vdots \\ a_{n-2} \\ a_{n-1} \end{pmatrix} = H \cdot \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{n-2} \\ d_{n-1} \end{pmatrix}$$

It follows that $a_s = 0$ for $k \leq s < n$ if and only if $Hd = \vec{0}$. Thus H is a check matrix for C. □

Example 7.1.34. Find a generating and a check matrix for the cyclic code C of length 7 with canonical generator $g = 1 + x^2 + x^3 + x^4$. Is 1001011 in the code?

Let E be the generating matrix from (7.1.33)(a). Since

$$g = 1 + x^2 + x^3 + x^4 = 1 + 0x + 1x^2 + 1x^3 + 1x^4$$

and $n = 7$, the first column of E is 1011100. So

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Let H be the check matrix from (7.1.33)(b). We first compute $t := \frac{x^7+1}{g}$:

$$(*)\qquad \begin{array}{r@{}l}
 & 1101 \\ \cline{2-2}
11101\big) & 10000001 \\
 & 11101 \\ \cline{2-2}
 & 1101001 \\
 & 11101 \\ \cline{2-2}
 & 11101 \\
 & 11101 \\ \cline{2-2}
 & 0
\end{array}\qquad\qquad \text{so}\quad t = x^3 + x^2 + 1$$

Also $n = 7$ and so the first row of H is 1101000. Thus

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

To check whether $d = 1001011$ is in the code we compute Hd:

$$\begin{array}{rcccc} & \begin{pmatrix}1 & 0 & 0 & 1 & 0 & 1 & 1\end{pmatrix} & & & \\ Hd = & \begin{bmatrix}1 & 1 & 0 & 1 & 0 & 0 & 0\\ 0 & 1 & 1 & 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 & 1 & 0 & 1\end{bmatrix} & = \begin{pmatrix}1+1+0+0\\ 0+0+0+0\\ 0+1+1+0\\ 0+1+0+1\end{pmatrix} & = & \begin{pmatrix}0\\0\\0\\0\end{pmatrix}\end{array}$$

So $Hd = \vec{0}$ and $d \in C$.

We could also have observed that d is the sum of the first and last column of E:

$$\begin{array}{cc} & 1011100 \\ + & 0010111 \\ \hline = & 1001011 \end{array}$$

and so $d \in C$.

7.2 Irreducible Polynomials and Extensions of fields

Definition 7.2.1. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$. Then h is called irreducible provided that*

(i) $\deg h > 0$*, and*

(ii) *if $h = fg$ for some $f, g \in \mathbb{F}[x]$, then* $\deg f = 0$ *or* $\deg g = 0$.

Lemma 7.2.2. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$. Then h is irreducible if and only*

(I) $\deg h > 0$*, and*

(II) *if $f \mid h$ for some $f \in \mathbb{F}[x]$, then* $\deg f = 0$ *or* $\deg f = \deg h$.

Proof. We may assume that $\deg h > 0$.

$\Longrightarrow$: Suppose first that h is irreducible and let $f \in \mathbb{F}[x]$ with $f \mid h$. Then $h = fg$ for some $g \in \mathbb{F}[x]$. Since h is irreducible, $\deg f = 0$ or $\deg g = 0$. If $\deg g = 0$, then $\deg f = \deg h - \deg g = \deg f$.

$\Longleftarrow$: : Suppose next that (II) holds and that $f, g \in \mathbb{F}[x]$ with $h = fg$. Then $f \mid h$ and so $\deg f = 0$ or $\deg f = \deg h$. If $\deg f = \deg h$, then $\deg g = \deg h - \deg f = 0$. So h is irreducible. □

Definition 7.2.3. *Let R be a ring with identity, $c = c_0 c_1 \dots c_{n-1} \in R^n$, $f \in R[x]$ and $a \in R$.*

(a) *$f(a) \coloneqq \sum_{i=0}^{\deg f} f_i a^i$ and $c(a) \coloneqq c_0 + c_1 a + \dots c_{n-1}a^{n-1}$.*

(b) *a is called a root of f if $f(a) = 0$.*

Lemma 7.2.4. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$ with $\deg(h) > 0$.*

(a) *There exists a monic irreducible polynomial $g \in \mathbb{F}[x]$ such that $g|h$ in $\mathbb{F}[x]$.*

(b) *h is irreducible if and only if there does not exist a monic irreducible polynomial $g \in \mathbb{F}[x]$ such that g divides h in $\mathbb{F}[x]$ and $\deg(g) \le \left\lfloor \frac{\deg(h)}{2} \right\rfloor$.*

(c) *Let $a \in \mathbb{F}$. Then $x - a$ divides h in $\mathbb{F}[x]$ if and only if $h(a) = 0$.*

(d) *Suppose that $2 \le \deg h \le 3$. Then h is irreducible in $\mathbb{F}[x]$ if and only if $h(a) \ne 0$ for all $a \in \mathbb{F}$.*

Proof. (a) Note that $h \in \mathbb{F}[x]$, $\deg h > 0$ and $h|h$. So we can choose $g \in \mathbb{F}[x]$ of minimal degree subject to $\deg g > 0$ and $g|h$. Replacing g by $\mathrm{lead}(g)^{-1}g$ we may assume the g is monic. We need to show that g is irreducible. For this let $f \in \mathbb{F}[x]$ with $f|g$ and $\deg(f) \ne 0$. Since $f|g$ and $g|h$ we get that $f|h$. The minimal choice of $\deg g$ now implies that $\deg g \le \deg f$. As $f|g$ we also have $\deg f \le \deg g$ and so $\deg f = \deg h$. Thus 7.2.2 shows that g is irreducible.

(b) We will prove the contrapositive of (b), that is we will show that there exist a g as in (b) if and only if h is not irreducible.

Suppose first that there does exist a monic irreducible polynomial $g \in \mathbb{F}[x]$ such that $g|h$ and $\deg(g) \le \left\lfloor \frac{\deg(h)}{2} \right\rfloor$. Then $g \mid h$, $\deg g \ne 0$ and $\deg g \ne \deg h$. Thus 7.2.2 shows that h is not irreducible.

Suppose next that h is not irreducible. Then $h = fk$ where $f, k \in \mathbb{F}[x]$ and $\deg(f) \ne 0$ and $\deg k \ne 0$. Without loss $\deg k \ge \deg f$. By (b) there exists a g be a monic irreducible polynomial in $\mathbb{F}[x]$ with $g \mid f$. Then g divides h and $\deg f \ge \deg g$. Hence

$$\deg h = \deg f + \deg k \ge 2\deg f \ge 2\deg g$$

and so $\deg g \le \left\lfloor \frac{\deg(h)}{2} \right\rfloor$.

(c) Let $h = q \cdot (x - a) + r$, where $q, r \in \mathbb{F}[x]$ with $\deg r < \deg(x - a) = 1$. Then $r \in \mathbb{F}$ and so $r(a) = r$. Thus

$$h(a) = q(a) \cdot (a - a) + r(a) = q(a) \cdot 0 + r = r$$

It follows that $h(a) = 0$ if and only if $r = 0$ and if and only if $x - a$ divides h.

(d) Let g be a monic polynomial with $\deg g > 0$. Since $2 \leq \deg h \leq 3$, we have $\left\lfloor \frac{\deg(h)}{2} \right\rfloor = 1$. Hence $\deg g \leq \left\lfloor \frac{\deg(h)}{2} \right\rfloor$ if and only if $\deg g \leq 1$ and so if and only if $g = x - a$ for some $a \in \mathbb{F}$. Observe that $x - a$ is irreducible for each $a \in \mathbb{F}$. Hence (b) shows that h is irreducible if and only if there does not exist $a \in \mathbb{F}$ such that $x - a$ divides h and so by (c) if and only if $h(a) \neq 0$ for all $a \in \mathbb{F}$. □

Example 7.2.5. Determine all irreducible polynomials in $\mathbb{F}_2[x]$ of degree at most three.

Let $f = \sum_{i=0}^n f_i x^i \in \mathbb{F}_2[x]$ with $f_n \neq 0$ and let $a \in \mathbb{F}_2$. Then $f(a) \neq 0$ iff $f(a) = 1$. Note that

$$f(0) = f_0 \quad \text{and} \quad f(1) = \sum_{i=0}^n f_i$$

Thus f has no roots in $\mathbb{F}_2$ if and only if $f_0 = 1$ and an odd number of coefficients of f are equal to 1. Since $f_n = 1$ this holds if and only if $f_0 = 1$ and an odd number of the coefficients $f_1, f_2, \ldots, f_{n-1}$ are equal to 1.

Suppose now that $\deg f = 2$ or 3. Then by (7.2.4)(d) f is irreducible if and only if f as no roots.

Suppose $\deg f = 2$. Then $f = x^2 + ax + b$ with $a, b \in \mathbb{F}_2$. So f is irreducible if and only if $b = 1$ and $a = 1$. Hence $x^2 + x + 1$ is the unique irreducible polynomial of degree 2 over $\mathbb{F}_2$.

Suppose $\deg f = 3$. Then $f = x^3 + ax^2 + bx + c$ with $a, b, c \in \mathbb{F}_2$. Hence f is irreducible if and only if $c = 1$ and exactly one of a and b is 1. Hence $x^3 + x^2 + 1$ and $x^3 + x + 1$ are the irreducible polynomial of degree 3 over $\mathbb{F}_2$.

Clearly all polynomials of degree 1 are irreducible. So x and $x + 1$ are the irreducible polynomial of degree 3 over $\mathbb{F}_2$. We summarize

Irreducible polynomials of degree at most three over $\mathbb{F}_2$:

Degree 1:	$x, \quad x + 1$
Degree 2:	$x^2 + x + 1$
Degree 3:	$x^3 + x^2 + 1, \quad x^3 + x + 1$

Lemma 7.2.6. *Let $\mathbb{F}$ be a field containing $\mathbb{F}_2$. Let $h \in \mathbb{F}_2[x]$ and $\alpha \in \mathbb{F}$. Then $h(\alpha^2) = h(\alpha)^2$. In particular, $h(\alpha) = 0$ if and only if $h(\alpha^2) = 0$.*

Proof. Let $h = \sum_{i=0}^n h_i x^i$ with $h_i \in \mathbb{F}_2$. Then $h_i = 0$ or $h_i = 1$ and in either case $h_i^2 = h_i$. Note also that $2 = 1 + 1 = 0$ in $\mathbb{F}_2$. Thus

$$h(\alpha)^2 = \left(\sum_{i=0}^n h_i \, \alpha^i \right)^2 = \sum_{i=0}^n h_i^2 \, (\alpha^i)^2 + 2 \sum_{0 \leq i < j \leq n} h_i \alpha^i h_j \alpha^j = \sum_{i=0}^n h_i \, (\alpha^2)^i = h(\alpha^2).$$

□

Lemma 7.2.7. *Let R be a commutative ring with identity and $r \in R$. Then $sr = 1$ for some $s \in R$ if and only if $\langle r \rangle = R$.*

Proof. $\Longrightarrow$: Suppose $sr = 1$ for some $s \in R$. Let $t \in R$. Then $t = t1 = t(sr) = (ts)r \in \langle r \rangle$ and so $R = \langle r \rangle$.

$\Longleftarrow$: Suppose that $R = \langle r \rangle$. Then $1 \in \langle r \rangle$ and so $1 = sr$ for some $s \in R$. □

Lemma 7.2.8. *Let $\mathbb{F}$ be a field and let $h \in \mathbb{F}[x]$ be irreducible. Put $\mathbb{E} := \mathbb{F}^h[x]$.*

(a) *$\mathbb{F}$ is a subfield of $\mathbb{E}$*

(b) *$(\mathbb{E}, \oplus, \odot)$ is a field.*

(c) *Let $h = \sum_{i=0}^n h_i x^i$ with $h_i \in \mathbb{F}$. Then x is a root in $\mathbb{E}$ of the polynomial $\sum_{i=0}^n h_i y^i$ in $\mathbb{E}[y]$.*

Proof. (a) Since h is irreducible, $\deg(h) > 0$. Let $a, b \in \mathbb{F}$. Then $a, b, a+b$ and ab are polynomials of degree at most 0, and so have degree less than $\deg(h)$. It follows that $a, b \in \mathbb{F}^h[x]$, $a \oplus b = a+b$ and $a \odot b = ab$. So $\mathbb{F}$ is a subfield of $\mathbb{F}^h[x]$.

(b) Let $0 \neq f \in \mathbb{F}^h[x]$. To show that $\mathbb{F}^h[x]$ is a field we need to find $s \in \mathbb{F}^h[x]$ with $s \odot f = 1$. For this let $I := \langle f \rangle$ be the ideal of $\mathbb{F}^h[x]$ generated by f. Let g be the canonical generator for I. As $I \neq 0$ we conclude from (7.1.30)(c) that $\deg g < \deg h$ and $g = \overline{g} \in I$. Moreover, by (7.1.30)(b) $g|h$. Since h is irreducible, we get $\deg g = 0$, see 7.2.2. As g is monic, this gives $g = 1$. So $1g = 1$ and 7.2.7 implies that $\langle g \rangle = \mathbb{F}^h[x]$. Hence $\langle f \rangle = I = \langle g \rangle = \mathbb{F}^h[x]$ and another application of 7.2.7 shows that $s \odot f = 1$ for some $s \in \mathbb{F}^h[x]$. Thus $\mathbb{F}^h[x]$ is a field.

(c) Let $e \in \mathbb{E}$. In $\mathbb{E}$ we have to use the operations $\oplus$ and $\odot$. Hence e is a root of $\sum_{i=0}^n h_i y^i$ in $\mathbb{E}$ if and only if

$$h_0 \oplus h_1 \odot e \oplus h_2 \odot e^{\odot 2} \oplus \ldots \oplus h_n \odot e^{\odot n} = 0,$$

and if and only if

$$\overline{\sum_{i=0}^n h_i e^i} = 0.$$

So x is a root of $\sum_{i=0}^n h_i y^i$ if and only if

$$\overline{\sum_{i=0} h_i x^i} = 0$$

But this just say $\overline{h} = 0$, which is true. □

Example 7.2.9. We will investigate $\mathbb{F}_2^h[x]$, where $h = 1 + x + x^3 \in \mathbb{F}_2[x]$.

Put $\mathbb{E} := \mathbb{F}_2^h[x]$. By 7.2.5 h is irreducible. Thus 7.2.8 shows that $\mathbb{E}$ is a field. To simplify notation, we will just write $f+g$ and fg for $f \oplus g$ and $f \odot g$. But to avoid confusion, we will write α for x to indicate that our computation are in the field $\mathbb{E}$ (rather than in $\mathbb{F}[x]$). Then

$$1 + \alpha + \alpha^3 = \overline{1 + x + x^3} = \overline{h} = 0.$$

Thus

$$\alpha^3 = -(1+\alpha) = 1 + \alpha.$$

Note that every element of $\mathbb{E}$ is a polynomial of degree less than $\deg(h)$ and so has degree at most 2. Thus every element of $\mathbb{E}$ can be uniquely written as

$$a + b\alpha + c\alpha^2$$

with $a, b, c \in \mathbb{F}_2$. We now compute all the powers of α.

$$\begin{array}{lllllllllllllll}
\alpha^0 & & & & & & & & & = & 1 & & & & \\
\alpha^1 & & & & & & & & & = & & & \alpha & & \\
\alpha^2 & & & & & & & & & = & & & & & \alpha^2 \\
\alpha^3 & & & & & & & & & = & 1 & + & \alpha & & \\
\alpha^4 & = & \alpha\alpha^3 & = & \alpha(1+\alpha) & & & & & = & & & \alpha & + & \alpha^2 \\
\alpha^5 & = & \alpha\alpha^4 & = & \alpha(\alpha+\alpha^2) & = & \alpha^2+\alpha^3 & = & \alpha^2+(1+\alpha) & = & 1 & + & \alpha & + & \alpha^2 \\
\alpha^6 & = & \alpha\alpha^5 & = & \alpha(1+\alpha+\alpha^2) & = & \alpha+\alpha^2+\alpha^3 & = & \alpha+\alpha^2+(1+\alpha) & = & 1 & & & + & \alpha^2 \\
\alpha^7 & = & \alpha\alpha^6 & = & \alpha(1+\alpha^2) & = & \alpha+\alpha^3 & = & \alpha+1+\alpha & = & 1. & & & &
\end{array}$$

Hence

$$\mathbb{E}^\sharp = \mathbb{E} \smallsetminus \{0\} = \{\alpha^i \mid 0 \le i < 7\}$$

Since α is a root of h, also α^2 and $\alpha^4 = (\alpha^2)^2$ are roots of h, see 7.2.6. We will verify this by direct computation:

$$h(\alpha^2) = 1 + \alpha^2 + (\alpha^2)^3 = 1 + \alpha^2 + \alpha^6 = 1 + \alpha^2 + (1 + \alpha^2) = 0$$

and

$$h(\alpha^4) = 1 + \alpha^4 + (\alpha^4)^3 = 1 + \alpha^4 + \alpha^{12} = 1 + \alpha^4 + \alpha^7\alpha^5 = 1 + \alpha^4 + \alpha^5 = 1 + (\alpha + \alpha^2) + (1 + \alpha + \alpha^2) = 0.$$

Thus $\alpha, \alpha^2, \alpha^4$ are the roots of $1 + x + x^3$.

From $\alpha^7 = 1$ we have $(\alpha^i)^7 = (\alpha^7)^i = 1$. So if $0 \ne e \in \mathbb{E}$, then $e^7 = 1$ and e is a root of $x^7 - 1$.

Note that $(1+x+x^3)(1+x) = 1+x+x^3+x+x^2+x^4 = 1+x^2+x^3+x^4$. In Example (7.1.34)(*) we computed that $\frac{x^7-1}{1+x^2+x^3+x^4} = 1+x^2+x^3$. Thus

$$(*) \qquad x^7 - 1 = (1+x)(1+x+x^3)(1+x^2+x^3).$$

1 is a root of $1+x$ and $\alpha, \alpha^2, \alpha^4$ are the roots of $1+x+x^3$. So α^3, α^4 and α^6 are the roots of $1+x^2+x^3$. To confirm

$$1 + (\alpha^3)^2 + (\alpha^3)^3 = 1 + \alpha^6 + \alpha^9 = 1 + \alpha^6 + \alpha^2 = 1 + (1+\alpha^2) + \alpha^2 = 0.$$

Since $(\alpha^3)^2 = \alpha^6$ and $(\alpha^6)^2 = \alpha^{12} = \alpha^5$, also α^6 and α^5 are roots of $1+x^2+x^3$.

Lemma 7.2.10. *Let $n \in \mathbb{Z}^+$ and write*

$$x^n - 1 = f_0^{\eta_0} f_1^{\eta_1} \dots f_l^{\eta_l}$$

where $f_0, f_1, \dots, f_l$ are pairwise distinct irreducible polynomials in $\mathbb{F}_2[x]$ and $\eta_i \in \mathbb{N}$. Let $C \subseteq \mathbb{F}_2^n$ be a cyclic code and let $g \in \mathbb{F}_2[x]$ be the canonic generator for the ideal $C(x)$ in $V^n[x]$. Then there exist $\epsilon_i \in \mathbb{N}$, for $0 \le i \le n$, with $\epsilon_i \le \eta_i$ and

$$g = f_0^{\epsilon_0} \dots f_l^{\epsilon_l}$$

Moreover, $x^n - 1 = gt$ where

$$t = f_0^{\delta_0} \dots f_l^{\delta_l}$$

and $\delta_i = \eta_i - \epsilon_i$.

Proof. By 7.1.27 g divides $x^n - 1$ in $\mathbb{F}_2[x]$. The lemma now follows from A.3.4. □

Example 7.2.11. Find a generating matrix and a check matrix for all the four dimensional binary cyclic codes of length 7.

Let C be a four dimensional binary cyclic code of length 7. Let g be the canonical generator of $C(x)$ and let $t \in \mathbb{F}_2[x]$ with $gt = x^7 + 1$. Then by (7.1.27)(h) $\deg t = \dim C = 4$ and so $\deg g = 7 - 4 = 3$. By (7.2.9)(*) $x^7 + 1 = (1+x)(1+x+x^3)(1+x^2+x^3)$. By 7.2.5 each of the three factors is irreducible. By 7.2.10

$$g = (1+x)^{\epsilon_1}(1+x+x^3)^{\epsilon_2}(1+x^2+x^3)^{\epsilon_3} \quad \text{and} \quad (1+x)^{\delta_1}(1+x+x^3\delta_2(1+x^2+x^3)^{\delta_3}$$

for some $\epsilon_i \in \{0,1\}$ and $\delta_i = 1 - \epsilon_i$. We have $3 = \deg g = \epsilon_1 + 3\epsilon_2 + 3\epsilon_3$. It follows that $3 \mid \epsilon_1$. Thus $\epsilon_1 = 0$ and $3 = 3\epsilon_2 + 3\epsilon_3$. Hence $1 = \epsilon_2 + \epsilon_3$. It follows that either ($\epsilon_2 = 1$ and $\epsilon_3 = 0$) or ($\epsilon_2 = 0$ and $\epsilon_3 = 1$). Therefore,

$$g = 1 + x + x^3 \quad \text{and} \quad t = (1+x)(1+x^2+x^3) = (1+x^2+x^3) + (x+x^3+x^4) = x^4+x^2+x+1$$

or

$$g = 1 + x^2 + x^3 \quad \text{and} \quad t = (1+x)(1+x+x^3) = (1+x+x^3) + (x+x^2+x^4) = x^4+x^3+x^2+1$$

Thus by 7.1.33

$$E = \begin{bmatrix} 1&0&0&0\\ 1&1&0&0\\ 0&1&1&0\\ 1&0&1&1\\ 0&1&0&1\\ 0&0&1&0\\ 0&0&0&1 \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} 1&0&1&1&1&0&0\\ 0&1&0&1&1&1&0\\ 0&0&1&0&1&1&1 \end{bmatrix}$$

or

$$E = \begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 1&0&1&0\\ 1&1&0&1\\ 0&1&1&0\\ 0&0&1&1\\ 0&0&0&1 \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} 1&1&1&0&1&0&0\\ 0&1&1&1&0&1&0\\ 0&0&1&1&0&1&1 \end{bmatrix}$$

Observe that for both codes the columns of H are the non-zero elements of $\mathbb{F}_2^3$. Hence by 6.4.9 both codes are Hamming codes.

Lemma 7.2.12. *Let $\mathbb{F}$ be a field and $0 \neq f \in \mathbb{F}[x]$. Set $a := \operatorname{lead}(f)$ and $n := \deg f$. Then there exists a field $\mathbb{E}$ containing $\mathbb{F}$ and elements $\alpha_1, \alpha_2 \ldots, \alpha_n$ in $\mathbb{E}$ such that*

$$f = a(x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_n)$$

Moreover, if $\mathbb{F}$ is finite we can choose $\mathbb{E}$ to be finite.

Proof. The proof is by induction on $\deg f$. If $\deg f = 0$, then $f = a$ and the lemma holds with $\mathbb{E} = \mathbb{F}$. So suppose $\deg f > 0$. Then $f = gh$ with $g, h \in \mathbb{F}[x]$ and h irreducible. Put $\mathbb{K} := \mathbb{F}^h[x]$. Then by 7.2.8 $\mathbb{K}$ is a field and there exists a root α of h in $\mathbb{K}$. Moreover, $\mathbb{K}$ is finite if $\mathbb{F}$ is finite. Since α is a root of h in $\mathbb{K}$ we know that $x - \alpha$ divides h in $\mathbb{K}[x]$. As $f = gh$ we conclude that $x - \alpha$ divides f and so $f = d \cdot (x - \alpha)$ for some $d \in \mathbb{K}[x]$. Observe that $\deg d = n - 1$ and $\operatorname{lead}(d) = \operatorname{lead}(f) = a$. By induction there exist a field $\mathbb{E}$ containing $\mathbb{K}$ and elements $\alpha_1, \alpha_2 \ldots, \alpha_{n-1}$ in $\mathbb{E}$ such that

$$d = a(x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_{n-1})$$

and $\mathbb{E}$ is finite if $\mathbb{K}$ is finite. Hence

$$f = d \cdot (x - \alpha) = a(x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_{n-1})(x - \alpha_n),$$

where $\alpha_n = \alpha$. □

Lemma 7.2.13. *Let $\mathbb{F}$ and $\mathbb{E}$ be a fields with $\mathbb{F} \subseteq \mathbb{E}$. Let $\alpha \in \mathbb{E}$ and suppose that α is a root of some non-zero polynomial in $\mathbb{F}[x]$. Let $m \in \mathbb{F}[x]$ be a monic polynomial of minimal degree with respect to $m(\alpha) = 0$. Put $\mathbb{F}[\alpha] := \{f(\alpha) \mid f \in \mathbb{F}[x]\}$. Let $f, g \in \mathbb{F}[x]$ and let r and s be the remainders of f and g, respectively, when divided by m, Then*

(a) $f(\alpha) = r(\alpha)$.

(b) $f(\alpha) = 0$ *if and only $r = 0$ and if and only if m divides f in $\mathbb{F}[x]$*

(c)
$$f(\alpha) = g(\alpha) \iff r = s \iff m \mid f - g.$$

(d) *For each $e \in \mathbb{F}[a]$ there exists a unique $r \in \mathbb{F}^m[x]$ with $e = r(\alpha)$.*

(e) *Let $n = \deg m$. Then $(1, \alpha, \alpha^2, \ldots, \alpha^{n-1})$ is an $\mathbb{F}$-basis for $\mathbb{F}[\alpha]$. In particular, $\dim_{\mathbb{F}} \mathbb{F}[\alpha] = n$.*

(f) *m is the unique monic irreducible polynomial in $\mathbb{F}[x]$ with $m(\alpha) = 0$.*

(g) *The function*
$$\Phi : \qquad \mathbb{F}^m[x] \to \mathbb{F}[\alpha], \qquad f \mapsto f(\alpha)$$
is an isomorphism of rings. In particular, $\mathbb{F}[\alpha]$ is a subfield of $\mathbb{E}$ isomorphic to $\mathbb{F}^m[x]$.

m is called the minimal polynomial of α over $\mathbb{F}$ and is denoted by $m_\alpha^{\mathbb{F}}$ or m_α.

Proof. (a) Let $f = qm + r$ with $q, r \in \mathbb{F}[x]$ and $\deg r < \deg m$. Then $f(\alpha) = q(\alpha)m(\alpha) + r(\alpha) = q(\alpha)0 + r(\alpha) = r(\alpha)$. So (a) holds.

(b): We have $f(\alpha) = 0$ if and only if $r(\alpha) = 0$. Since $\deg r < \deg m$, the minimality of $\deg m$ shows that $r(\alpha) = 0$ if and only if $r = 0$. By (7.1.12)(e) $r = 0$ if and only if $m|f$. So (b) holds.

(c) Note that $r - s$ is the remainder of $f - g$ when divided by m. Hence by (b) applied to $f - g$ in place of f:

$$(f-g)(\alpha) = 0 \quad \Longleftrightarrow \quad r - s = 0 \quad \Longleftrightarrow \quad m \mid f - g.$$

Observe that $(f - g)(\alpha) = 0$ if and only if $f(\alpha) = g(\alpha)$. Also $r - s = 0$ if and only if $r = s$. So (c) is proved.

(d) Let $e \in \mathbb{F}[\alpha]$. By definition of $\mathbb{F}[\alpha]$, $e = f(\alpha)$ for some $f \in \mathbb{F}[x]$. By (a) we have $f(\alpha) = r(\alpha)$. Since $\deg r < \deg m$ we know that $r \in \mathbb{F}^m[x]$. This shows that existence of r. Suppose $e = g(\alpha)$ for some $g \in \mathbb{F}^m[x]$. Then $g(\alpha) = e = f(\alpha)$. So (c) implies that $r = s$. As $g \in \mathbb{F}^m[x]$ we have $g = s$. Hence $g = r$ and r is unique.

(e) Let $e \in \mathbb{F}[a]$.

By (d) there exists a unique $r \in \mathbb{F}^m[x]$ with $e = r(\alpha)$. Then $r = \sum_{i=0}^{n-1} a_i x^n$ for unique $a_0, \ldots, a_{n-1} \in \mathbb{F}$.

This shows that there exists unique $a_0, \ldots, a_{n-1} \in F$ with $e = a_0 + a_1\alpha + \ldots \alpha_{n-1}\alpha^{n-1}$. Thus $(1, \alpha, \ldots, \alpha^{n-1})$ is an $\mathbb{F}$-basis for $\mathbb{F}[\alpha]$.

(f) Suppose that $m = gh$ for some $g, h \in \mathbb{F}[x]$. Then $g(\alpha)h(\alpha) = m(\alpha) = 0$ and since $\mathbb{E}$ is a field, $g(\alpha) = 0$ or $h(\alpha) = 0$. Without loss $g(\alpha) = 0$. Then the minimality of $\deg m$ implies $\deg g = \deg m$ and so $\deg h = 0$. Thus m is irreducible.

Let g be any irreducible monic polynomial with $g(\alpha) = 0$. By (b), $m|g$ and since g is irreducible, A.3.2 implies $m = g$. Thus m is unique.

(g) By (d) Φ is a bijection. Also

$$\begin{aligned}\Phi(f+g) &= (f+g)(\alpha) = f(\alpha) + g(\alpha) = \Phi(f) + \Phi(g),\\ \Phi(f \cdot g) &= (f \cdot g)(\alpha) \ = f(\alpha) \cdot g(\alpha) \ = \Phi(f) \cdot \Phi(g).\end{aligned}$$

So Φ is a isomorphism. □

Definition 7.2.14. *Let $\mathbb{E}$ be a field containing $\mathbb{F}_2$, let $C \subseteq \mathbb{F}_2^n$ be a binary linear code and let H be an $m \times n$-matrix with coefficients in $\mathbb{E}$. We say that H is a check matrix for C over $\mathbb{E}$ if*

$$C = \{a \in \mathbb{F}_2^n \mid Ha = \vec{0}\}.$$

Lemma 7.2.15. *Let $\mathbb{F}$ be a field and let α_i, $1 \le i \le d$, be elements in $\mathbb{F}$. Let H be the $d \times n$ matrix*

$$H = [\alpha_i^j]_{\substack{1 \le i \le d \\ 0 \le j < n}} = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_d & \alpha_d^2 & \dots & \alpha_d^{n-1} \end{bmatrix}.$$

and let $c = c_0 c_1 \dots c_{n-1} \in \mathbb{F}^n$. Then the i-th coefficient of Hc is

$$c(\alpha_i) = c_0 + c_1\alpha_i + c_2\alpha_i^2 + \dots + c_{n-1}\alpha_i^{n-1} = \sum_{j=0}^{n-1} c_j \alpha_i^j$$

In particular, $Hc = \vec{0}$ if and only if $\alpha_1, \alpha_2, \dots, \alpha_d$ all are roots of $c(x)$.

Proof. Just note that the i'th coefficient of Hc is

$$\sum_{j=0}^{n-1} \alpha_i^j c_j = \sum_{j=0}^{n-1} c_j a_i^j = c(\alpha_i).$$

□

Lemma 7.2.16. *Let $C \subseteq \mathbb{F}_2^n$ be a binary cyclic code with canonical generator $g \in \mathbb{F}_2[x]$. Suppose that $g = m_1 \dots m_s$, where $m_1, \dots, m_s$ are pairwise distinct, irreducible, monic polynomials in $\mathbb{F}_2[x]$. Let $\mathbb{E}$ be a field containing $\mathbb{F}_2$ and let $\alpha_1, \dots, \alpha_d$ be pairwise distinct elements in $\mathbb{E}$. Suppose that*

(i) *for each $1 \le i \le d$ there exists $1 \le j \le s$ with $m_j(\alpha_i) = 0$, and*

(ii) *for each $1 \le j \le s$ there exists $1 \le i \le d$ with $m_j(\alpha_i) = 0$.*

Put

$$H := [\alpha_i^j]_{\substack{1 \le i \le d \\ 0 \le j \le n-1}}.$$

Then H is a check matrix for C over $\mathbb{E}$ and

$$C = \{c \in \mathbb{F}_2^n \mid c(\alpha_i) = 0 \text{ for all } 1 \le i \le d\}$$

Proof. Let $c \in \mathbb{F}_2^n$. We will first show that:

$$(\ast) \qquad c \in C \qquad \text{if and only if} \qquad g \text{ divides } c(x) \text{ in } \mathbb{F}_2[x].$$

Note first that $c \in C$ if and only if $c(x) \in C(x)$. Since g is a canonical generator for $C(x)$, this is the case if and only if g divides $c(x)$, see (7.1.27)(a).

Next we show

$$(**) \qquad g \text{ divides } c(x) \qquad \text{if and only if} \qquad c(\alpha_i) = 0 \text{ for all } 1 \le i \le d.$$

Suppose g divides $c(x)$. Let $1 \le i \le d$. By (i) there exists $1 \le j \le s$ with $m_j(\alpha_i) = 0$. Since m_j divides g, we have $g(\alpha_i) = 0$ and since g divides $c(x)$ we get $c(\alpha_i) = 0$.

Suppose $c(\alpha_i) = 0$ for all $1 \le i \le d$. Let $1 \le j \le s$. By (ii) there exists $1 \le i \le d$ with $m_j(\alpha_i) = 0$. As $c(\alpha_i) = 0$ we conclude from 7.2.13 that m_j divides $c(x)$. Since $g = m_1 m_2 \dots m_s$ and the m_j's are pairwise distinct monic irreducible polynomials we conclude from A.3.4 that g divides $c(x)$.

From $(*)$ and $(**)$ we get

$$c \in C \qquad \text{if and only if} \qquad c(\alpha_i) = 0 \text{ for all } 1 \le i \le d.$$

and so

$$C = \{c \in \mathbb{F}_2^n \mid c(\alpha_i) = 0 \text{ for all } 1 \le i \le d\}$$

Hence 7.2.15 implies that $C = \{c \in \mathbb{F}_2^n \mid Hc = \vec{0}\}$. Thus H is check matrix for C over $\mathbb{E}$. □

7.3 Definition and Properties of BCH-codes

Definition 7.3.1. *Let $\mathbb{E}$ be a finite field and $\alpha \in \mathbb{E}$. Put $n := |\mathbb{E}| - 1$. Then α is called a primitive element for $\mathbb{E}$ if*

$$\alpha^n = 1 \qquad \text{and} \quad \mathbb{E} \smallsetminus \{0\} = \{\alpha^i \mid 0 \le i \le n-1\}.$$

Lemma 7.3.2. *Every finite field has a primitive element.*

Proof. For a proof see A.4.5 in the appendix. □

Definition 7.3.3. *Let $\mathbb{E}$ be a finite field containing $\mathbb{F}_2$. Put $n := |\mathbb{E}| - 1$ and let α be a primitive element for $\mathbb{E}$. Let $1 \le d \le n-1$ and put*

$$g := \operatorname{lcm}(m_{\alpha^i} \mid 1 \le i \le d).$$

Let $C \subseteq \mathbb{F}_2^n$ be the binary cyclic code with canonical generator g. Then C is called the BCH-code of length n and designated distance $d+1$ with respect to α.

Lemma 7.3.4. *Let $\mathbb{F}$ be a field and let $\beta_j, 1 \le j \le e$, be pairwise distinct non-zero elements in $\mathbb{F}$. Let $1 \le d \le e$ and let A be the $d \times e$ matrix*

$$A := [\beta_j^i]_{\substack{1\le i\le d\\ 1\le j\le e}} = \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_e \\ \beta_1^2 & \beta_2^2 & \dots & \beta_e^2 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1^d & \beta_2^d & \dots & \beta_e^d \end{bmatrix}$$

Then any d columns of A are linearly independent over $\mathbb{F}$.

Proof. Replacing A by d of its columns we may assume that $e = d$. Put

$$B := [\beta_j^i]_{\substack{0\le i\le d-1\\ 1\le j\le d}} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \beta_1 & \beta_2 & \dots & \beta_d \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1^{d-1} & \beta_2^{d-1} & \dots & \beta_{d-1}^{d-1} \end{bmatrix}.$$

Then $\beta_j \cdot \mathrm{Col}_j(B) = \mathrm{Col}_j(A)$. Since $\beta_j \neq 0$ for all $1 \le j \le d$, the columns of B are linearly independent if and only if the columns of A are linearly independent.

Since B is a square matrix, the columns of B are linearly independent if and only if B is invertible, if and only if B^{T} is invertible, and if and only if $c = 0$ for all $c \in \mathbb{F}^d$ with $B^{\mathrm{T}}c = 0$. Note that

$$B^{\mathrm{T}} = [\beta_i^j]_{\substack{1\le i\le d\\ 0\le j\le d-1}}.$$

By 7.2.15, $B^{\mathrm{T}}c = 0$ if and only if $\beta_1, \beta_2, \dots, \beta_d$ are roots of $c(x)$. Observe that a non-zero polynomial of degree less or equal to $d-1$ has at most $d-1$ roots. Since $c(x)$ is a polynomial of degree at most $d-1$ and since $\beta_1, \beta_2, \dots \beta_d$ are d distinct elements of $\mathbb{F}$, we conclude that $\beta_1, \beta_2, \dots, \beta_d$ are roots of $c(x)$ if and only if $c(x) = 0$ and so if and only if $c = 0$. □

Theorem 7.3.5. *Let $\mathbb{E}$ be a finite field containing $\mathbb{F}_2$. Put $n := |\mathbb{E}| - 1$ and let $1 \le d \le n-1$. Let C be the BCH-code of length n and designated distance $d+1$ with respect to the primitive element α in $\mathbb{E}$. Let*

$$\{m_{\alpha^i} \mid 1 \le i \le d\} = \{m_1, \dots, m_s\}.$$

where the m_i's, $1 \le i \le s$, are pairwise distinct.

(a) *Let g be the canonical generator of C. Then $g = m_1 m_2 \dots m_s$.*

(b) *Put* $H := [\alpha^{ij}]_{\substack{1\le i\le d\\ 0\le j\le n-1}}$. *Then* H *is a check matrix for* C *over* $\mathbb{E}$ *and*

$$C = \{c \in \mathbb{F}_2^n \mid c(\alpha^i) = 0 \text{ for all } 1 \le i \le d\}.$$

(c) *For* $1 \le i \le s$ *let* α_i *be a root of* m_i *in* $\mathbb{E}$. *Put* $\tilde{H} := [\alpha_i^j]_{\substack{1\le i\le s\\ 0\le j\le n-1}}$. *Then* $\tilde{H}$ *is a check matrix for* C *over* $\mathbb{E}$ *and*

$$C = \{c \in \mathbb{F}_2^n \mid c(\alpha_i) = 0 \text{ for all } 1 \le i \le s\}.$$

(d) C *has minimum distance at least* $d+1$.

(e) $\dim C \ge n - \lceil \frac{d}{2} \rceil \log_2(n+1)$.

(f) *Suppose* d *is even and let* $d = 2r$ *with* $r \in \mathbb{Z}$. *Then* C *is an* r*-error correcting code and* $\dim C \ge n - r\log_2(n+1)$.

Proof. (a) By definition of a BCH-code we have $g = \operatorname{lcm}(m_{\alpha^i} | 1 \le i \le d)$. Since

$$\{m_{\alpha^i} \mid 1 \le i < d\} = \{m_1, \dots, m_s\}$$

we get $g = \operatorname{lcm}(m_i | 1 \le i \le s)$. As the m_i are pairwise distinct monic irreducible polynomials we conclude from A.3.4 that $g = m_1 m_2 \dots m_s$.

(b) We will verify that the conditions (7.2.16)(i) and (ii) are fulfilled for $\alpha_i = \alpha^i$:

(i): Let $1 \le i \le d$. Choose $1 \le j \le s$ with $m_j = m_{\alpha^i}$. As α^i is a root of m_{α^i} we conclude that $m_j(\alpha^i) = 0$.

(ii): Let $1 \le j \le s$. Then $m_j = m_{\alpha^i}$ for some $1 \le i \le d$ and so $m_j(\alpha^i) = 0$.

Thus we can apply 7.2.16 and so (b) holds.

(c) Note that $m_i(\alpha_i) = 0$ for all $1 \le i \le s$. So both conditions in 7.2.16 are fulfilled (with $d = s$) and thus (c) holds.

(d) By 7.3.4 applied with $\beta_j = \alpha^{j-1}$, $1 \le j \le n$, any d-columns of H are linearly independent over $\mathbb{E}$. Hence the sum of any d-columns of H is non-zero and so $Hc \ne \vec{0}$ for any $\vec{0} \ne c \in \mathbb{F}_2^n$ with $\operatorname{wt}(c) \le d$. Thus C has minimum weight at least $d+1$.

(e) Put $l := \log_2(n+1)$. Then $|\mathbb{E}| = n+1 = 2^l$. Put $t_i := \deg m_i$. Since α_i is a root of m_i and m_i is irreducible, we conclude from (7.2.13)(f) that m_i is the minimal polynomial of α_i. Part (e) of the same lemma now shows that $\dim_{\mathbb{F}_2} \mathbb{F}_2[\alpha_i] = t_i$ and so $|\mathbb{F}_2[\alpha_i]| = 2^{t_i}$. Since $|\mathbb{F}_2[\alpha_i]| \le |\mathbb{E}| = 2^l$ we have $t_i \le l$. Thus

$$\deg g = \deg(m_1 m_2 \dots m_s) = \sum_{i=1}^{s} \deg m_i = \sum_{i=1}^{s} t_i \le \sum_{i=1}^{s} l = sl.$$

By (7.1.27)(g),

$$(*) \qquad \dim C = n - \deg g \geq n - sl.$$

Since α^j is a root of m_{α^j} and $\alpha^{2j} = (\alpha^j)^2$, also α^{2j} is a root of m_{α^j}, see 7.2.6. Thus m_{α^j} is an irreducible monic polynomial with root α^{2j}. By (7.2.13)(f) $m_{\alpha^{2j}}$ is the unique such polynomial. Hence $m_{\alpha^{2j}} = m_{\alpha^j}$. If $i = 2^k j$ with j odd, we conclude that $m_{\alpha^i} = m_{\alpha^j}$. Therefore,

$$\{m_{\alpha^i} \mid 1 \leq i \leq d\} = \{m_{\alpha^j} \mid 1 \leq j \leq d, j \text{ odd}\}$$

If d is even, the number of odd integers j with $1 \leq j \leq d$ is $\frac{d}{2}$ and if d is odd the number of such integers is $\frac{d+1}{2}$. So in either case the number of such integers is $\lceil \frac{d}{2} \rceil$. Hence $s \leq \lceil \frac{d}{2} \rceil$. Thus

$$\dim C \overset{(*)}{\geq} n - sl \geq n - \left\lceil \frac{d}{2} \right\rceil \log_2(n+1).$$

(f) Suppose that $d = 2r$. By (d) C has minimum distance at least $d + 1 = 2r + 1$ and so C is r-error-correcting. Also $\lceil \frac{d}{2} \rceil = r$ and so by (e) $\dim C \geq n - r\log_2(n+1)$. □

Example 7.3.6. Put $\mathbb{E} := \mathbb{F}_2^{x^4+x+1}[x]$ and $\alpha := x \in \mathbb{E}$. Let C be the BCH-code of length 15 and designated distance 7 with respect to α. Verify that $\mathbb{E}$ is a field and that α is a primitive element of $\mathbb{E}$. Determine the dimension of C, the minimal distance of C, the canonical generator for C and a check matrix for C over $\mathbb{E}$.

To show that $\mathbb{E}$ is a field it suffices to show that $x^4 + x + 1$ is irreducible in $\mathbb{F}_2[x]$, (see Lemma 7.2.8).

Let $h := x^4 + x + 1$. By (7.2.4)(b) h is irreducible if and only if no irreducible polynomial of degree at most $\frac{4}{2} = 2$ divides h. By 7.2.5 the irreducible polynomial of degree at most 2 are $x, x+1$ and $x^2 + x + 1$. As $h(0) = 1$ and $h(1) = 1$, neither x nor $x + 1$ divides h. We compute

```
    110
111|10011
    111
     1111
     111
        1
```

Hence also $x^2 + x + 1$ does not divide h. Thus h is irreducible.

In order to shows that α is a primitive element for $\mathbb{E}$, we will compute the powers of α. Recall first from (7.2.8)(c) that α is a root of $x^4 + x + 1$. So $\alpha^4 = -(1 + \alpha) = 1 + \alpha$.

$$\begin{array}{rclclclclrclclclcl}
\alpha^0 & = & 1 & & & & & & & \qquad\alpha^8 & = & 1 & & & + & \alpha^2 & & \\
\alpha^1 & = & & & \alpha & & & & & \qquad\alpha^9 & = & & & \alpha & & & + & \alpha^3 \\
\alpha^2 & = & & & & & \alpha^2 & & & \qquad\alpha^{10} & = & 1 & + & \alpha & + & \alpha^2 & & \\
\alpha^3 & = & & & & & & & \alpha^3 & \qquad\alpha^{11} & = & & & \alpha & + & \alpha^2 & + & \alpha^3 \\
\alpha^4 & = & 1 & + & \alpha & & & & & \qquad\alpha^{12} & = & 1 & + & \alpha & + & \alpha^2 & + & \alpha^3 \\
\alpha^5 & = & & & \alpha & + & \alpha^2 & & & \qquad\alpha^{13} & = & 1 & & & + & \alpha^2 & + & \alpha^3 \\
\alpha^6 & = & & & & & \alpha^2 & + & \alpha^3 & \qquad\alpha^{14} & = & 1 & & & & & + & \alpha^3 \\
\alpha^7 & = & 1 & + & \alpha & & & + & \alpha^3 & \qquad\alpha^{15} & = & 1 & & & & & &
\end{array}$$

Hence α is a primitive element. Since C has designated distance 7, $d = 6$. To find the canonical generator g we need to determine m_{α^i} for $1 \le i \le 6$.

α, α^2 and α^4 are roots of $1 + x + x^4$ and so

$$m_\alpha = m_{\alpha^2} = m_{\alpha^4} = 1 + x + x^4.$$

We now introduce a method to compute m_β for $\beta \in \mathbb{E}$. Note that $\beta = b_0 + b_1\alpha + b_2\alpha^2 + b_3\alpha^3$ for some $b_0, b_1, b_2, b_3 \in \mathbb{F}_2$. We call $b_0b_1b_2b_3$ the string associated to β and write $\beta \mapsto b_0b_1b_2b_3$.

Let s_i be the string associated to β^i. Let $l \in \mathbb{N}$ be minimal such that

$$a_0s_0 + a_1s_2 + \dots \alpha_l s_l = \vec{0}$$

for some $a_0, \dots, \alpha_l \in \mathbb{F}_2$ with $\alpha_l = 1$. Then l is also minimal in $\mathbb{N}$ with respect to

$$a_0 + a_1\beta + \dots + a_l\beta^l = 0.$$

for some $a_0, \dots, \alpha_l \in \mathbb{F}_2$ with $\alpha_l = 1$, so $m_\beta = a_0 + a_1x + \dots + a_lx^l$.

Hence we can compute m_β by computing the strings $s_0, s_1, s_2 \dots$ associated to $1, \beta, \beta^2 \dots$ until we reach $l \in \mathbb{N}$ such that $s_0, s_2, \dots s_l$ are linearly dependent.

For $\beta = \alpha^3$ we have

$$\begin{array}{lcrcl}
(\alpha^3)^0 & = 1 & & \mapsto & 1000 \\
(\alpha^3)^1 & = & \alpha^3 & \mapsto & 0001 \\
(\alpha^3)^2 = \alpha^6 & = & \alpha^2{+}\alpha^3 & \mapsto & 0011 \\
(\alpha^3)^3 = \alpha^9 & = & \alpha \quad {+}\alpha^3 & \mapsto & 0101 \\
(\alpha^3)^4 = \alpha^{12} & = & 1{+}\alpha{+}\alpha^2{+}\alpha^3 & \mapsto & 1111
\end{array}$$

The first four strings are linearly independent, and the sum of all five is zero.

So

$$m_{\alpha^3} = 1 + x + x^2 + x^3 + x^4 \quad \text{and} \quad m_{\alpha^6} = m_{\alpha^3}$$

For $\beta = \alpha^5$ we get

$$\begin{array}{lcrcl}
(\alpha^5)^0 & & = 1 & \mapsto & 1000 \\
(\alpha^5)^1 & & = \quad \alpha{+}\alpha^2 & \mapsto & 0110 \\
(\alpha^5)^2 & = \alpha^{10} & = 1{+}\alpha{+}\alpha^2 & \mapsto & 1110
\end{array}$$

The first two strings are linearly independent and the sum of all three is zero. So

$$m_{\alpha^5} = 1 + x + x^2.$$

Thus we can choose

$$m_1 = 1 + x + x^4, \quad m_2 = 1 + x + x^2 + x^3 + x^4, \quad \text{and} \quad m_3 = 1 + x + x^2.$$

and so

$$g = m_1 m_2 m_3 = (1 + x + x^4)(1 + x + x^2 + x^3 + x^4)(1 + x + x^2)$$

We compute

$$\begin{array}{rcccccccccccccl}
 & & 0&1&2&3&4&5&6&7&8&9&10& \\ \cline{3-13}
1+x+x^2+x^3+x^4 & \mapsto & 1&1&1&1&1& & & & & & &\cdot 1 \\
 & & &1&1&1&1&1& & & & & &\cdot x \\
 & & & & & &1&1&1&1&1& & &\cdot x^4 \\ \cline{3-13}
 & &1&0&0&0&1&0&1&1&1& & &\cdot 1 \\
 & & &1&0&0&0&1&0&1&1&1& &\cdot x \\
 & & & &1&0&0&0&1&0&1&1&1&\cdot x^2 \\ \cline{3-13}
g & \mapsto &1&1&1&0&1&1&0&0&1&0&1&
\end{array}$$

Thus $g = 1 + x + x^2 + x^4 + x^5 + x^8 + x^{10}$. Since $\deg g = 10$, $\dim C = 15 - 10 = 5$. Note that $g = \overline{g} \in C(x)$ and so the string 111011001010000 corresponding to g is a codeword. This string has weight 7 and so C has minimum distance at most 7. By 7.3.5 C has minimum weight at least 7 and so $\delta(C) = 7$.

Note that α, α^3 and α^5 are roots of m_1, m_2 and m_3 respectively. Hence (7.3.5)(b) provides the following check matrix for C over $\mathbb{E}$:

$$\tilde{H} := \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 & \alpha^6 & \alpha^7 & \alpha^8 & \alpha^9 & \alpha^{10} & \alpha^{11} & \alpha^{12} & \alpha^{13} & \alpha^{14} \\ 1 & \alpha^3 & \alpha^6 & \alpha^9 & \alpha^{12} & 1 & \alpha^3 & \alpha^6 & \alpha^9 & \alpha^{12} & 1 & \alpha^3 & \alpha^6 & \alpha^9 & \alpha^{12} \\ 1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10} \end{bmatrix}$$

Lemma 7.3.7. *Let $C \subseteq \mathbb{F}_2^n$ be the BCH-code with respect to the primitive element α. Let (a, z) be a 1-bit error for C. Then $z(\alpha) \neq 0$. Moreover, if $i \in \mathbb{N}$ is the unique element with $0 \leq i < n$ and $z(\alpha) = \alpha^i$, then $a_i \neq z_i$.*

Proof. Since (a, z) is 1-bit error, $a_j \neq z_j$ for a unique $j \in \mathbb{N}$ with $0 \leq j < n$. Then $z_j = a_j + 1$ and so $z(x) = a(x) + x^j$. By 7.2.16 $c(\alpha) = 0$ for all $c \in C$. As $a \in C$ this gives

$$z(\alpha) = a(\alpha) + \alpha^j = \alpha^j.$$

Since α is a primitive element, the elements $\alpha^s, 0 \leq s < n$, are pairwise distinct. Hence $j = i$ and so $z_i \neq a_i$. □

Example 7.3.8. Let C be the BCH-code with respect to the primitive element $\alpha = x \in \mathbb{F}_2^{x^4+x+1}[x]$ from Example 7.3.6. Does there exist a 1-bit error (a, z) of C with

$$z = 011111110000000?$$

We first calculate the remainder r of $z(x)$ when divided by $x^4 + x + 1$.

```
           111
     10011|11111110
           10011
            1100110
            10011
             101010
             10011
               1100
```

So $r = x^3 + x^2$. Then (7.2.13)(a) shows that $z(\alpha) = r(\alpha) = \alpha^3 + \alpha^2 = \alpha^2(1 + \alpha) = \alpha^2\alpha^4 = \alpha^6$. Note here that by (7.2.8)(c) α is a root of $x^4 + 1$ and so $\alpha^4 = \alpha + 1$.(Alternatively, $\alpha^2(1 + \alpha) = \alpha^2 + \alpha^3$ and as seen in Example 7.3.6 $\alpha^2 + \alpha^3 = \alpha^6$). Hence, if (a, z) is a 1-bit error, the error occurred in z_6. So

$$a = 0111110100000000.$$

By (7.3.5)(c) $d \in C$ if and only if α, α^3 and α^5 are roots of $a(x) = x + x^2 + x^3 + x^4 + x^5 + x^7$. We compute

$$a(\alpha^5) = \alpha^5 + \alpha^{10} + \alpha^{15} + \alpha^{20} + \alpha^{25} + \alpha^{35} = \alpha^5 + \alpha^{10} + 1 + \alpha^5 + \alpha^{10} + \alpha^5 = 1 + \alpha^5 \neq 0.$$

(since $\alpha^5 \neq 1$). So a is not in the code. Thus z cannot be the result of a 1-bit-error. (Alternatively, C has minimum weight at least 7 and a has weight 6 so a is not in the code.)

Chapter 8

The RSA cryptosystem

8.1 Public-key cryptosytems

Definition 8.1.1. *A cryptosystem* Ω *is quadruple* $(\mathcal{M},\mathcal{C},(E_k)_{k\in\mathcal{K}},(D_l)_{l\in\mathcal{K}})$*, where* $\mathcal{M}$*,* $\mathcal{C}$ *and* $\mathcal{K}$ *are alphabets and for* $k\in\mathcal{K}$*,* $E_k:\mathcal{M}\to\mathcal{C}$ *and* $D_k:\mathcal{C}\to\mathcal{M}$ *are functions such that for each* $k\in\mathcal{K}$ *there exists* $k^*\in\mathcal{K}$ *with* $D_{k^*}\circ E_k=\mathrm{id}_{\mathcal{M}}$*.*

The elements of $\mathcal{M}$ *are called plaintext messages, the elements of* $\mathcal{C}$ *are called ciphertext messages, the elements of* $\mathcal{K}$ *are called keys, each* E_k *is called an encryption function and each* D_l *is called a decryption function. If* $k,k^*\in\mathcal{K}$ *with* $D_{k^*}\circ E_k=\mathrm{id}_{\mathcal{M}}$*, then* k^* *is called an inverse key for* k*.*

Example 8.1.2. Let $\mathcal{M}=\mathcal{C}=\mathbb{A}=\{A,B,C,D,\ldots,Z,\sqcup\}$, $\mathcal{K}=\{0,1,\ldots,25\}$ and, for $k\in\mathcal{K}$, let $E_k=D_k=c_k$, where c_k is the shift by k-letters defined in Example 1.5.1. Note that D_{26-k} is the inverse of E_k, so this is indeed a cryptosystem.

Definition 8.1.3. *A public-key cryptosystem is pair* (Ω,ξ) *where* Ω *is a cryptosystem and* ξ *is a function*

$$\xi:A\to\mathcal{K}\times\mathcal{K},\ a\to(k_a,k_a^*)$$

such that, for each $a\in A$*,* k_a^* *is an inverse key for* k_a*.* k_a *is called a public key and* k_a^* *a private key.*

In public key cryptography one picks $a\in A$ and uses ξ to compute the public key k_a and the private key k_{a^*}. The key k_a is publicized, but a and k_{a^*} are kept secret. Anybody then can encrypt a message using the public key k_a and the publicly known function E_{k_a}. But only oneself knows the privat key k_a^* and can decrypt the encrypted message using the function $D_{k_a^*}$.

This can only work if it is virtually impossible to determine k_{a^*} from k_a. In particular, A has to be really large, since otherwise one can just compute all the possible pairs of keys using the publicly known function ξ.

In this sections we will describe a public-key cryptosystem discovered by Rivest, Shamir and Adleman in 1977 known as the RSA cryptosystem. But we first need to prove a couple of lemmata about the ring of integers.

8.2 The Euclidean Algorithim

Lemma 8.2.1. *Let a, b, q and r be integers with $a = qb + r$. Then $\gcd(a, b) = \gcd(b, r)$.*

Proof. Let $d = \gcd(a, b)$ and $e = \gcd(b, r)$. Then d divides a and b and so also $r = a - qb$. Hence d is a common divisor of b and r. Thus $d \le e$.

Similarly, e divides b and r and so also $a = qb + r$. Thus e is a common divisor of a nd b and so $e \le d$. Hence $e = d$. □

Theorem 8.2.2 (Euclidean Algorithm)**.** *Let a and b be integers and let E_0 and E_1 be the equations*

$$\begin{array}{ccccccc} E_0 & : & a & = & 1a & + & 0b \\ E_1 & : & b & = & 0a & + & 1b \end{array},$$

and suppose inductively we defined equation $E_k, -1 \le k \le i$ of the form

$$E_k \ : \ r_k \ = \ x_k a \ + \ y_k b \ .$$

If $r_i \neq 0$, let E_{i+1} be equation obtained by subtracting q_{i+1} times equation E_i from E_{i-1} where q_{i+1} is the integer quotient of r_{i-1} when divided by r_i (so $q_{i+1} = \lfloor \frac{r_{i-1}}{r_i} \rfloor$). Let $m \in \mathbb{N}$ be minimal with $r_m = 0$ and put $d = r_{m-1}$, $x = x_{m-1}$ and $y = y_{m-1}$. Then

(a) $\gcd(a, b) = |d|$

(b) *$x, y \in \mathbb{Z}$ and $d = xa + yb$.*

Proof. Observe that $r_{i+1} = r_{i-1} - q_{i+1} r_i$, $x_{i+1} = x_{i-1} - q_{i+1} x_i$ and $y_{i+1} = y_{i-1} - q_{i+1} x_i$. So inductively $r_{i+1}, x_{i+1}, y_{i+1}$ are integers and r_{i+1} is the remainder of r_{i-1} when divided by r_i. So $r_{i+1} < |r_i|$ and the algorithm will terminate in finitely many steps.

From $r_{i-1} = q_{i+1} r_i + r_{i+1}$ and 8.2.1 we have $\gcd(r_{i-1}, r_i) = \gcd(r_i, r_{i+1})$ and so

$$\gcd(a, b) = \gcd(r_{-1}, r_0) = \gcd(r_0, r_1) = \ldots = \gcd(r_{m-1}, r_m) = \gcd(d, 0) = |d|$$

So (a) holds. Since each x_i and y_i are integers, x and y are integers. $d = xa + yb$ is just the equation E_{m-1}. □

Example 8.2.3. Let $a = 1492$ and $b = 1066$. Compute $\gcd(a, b)$ and find $x, y \in \mathbb{Z}$ with $\gcd(a, b) = xa + yb$.

$$\begin{array}{llll}
E_0: & 1492 = & 1\cdot a + 0\cdot b & a = 1492 \\
E_1: & 1066 = & 0\cdot a + 1\cdot b & b = 1066 \\
E_2: & 426 = & 1\cdot a - 1\cdot b & E_0 - E_1 \\
E_3: & 214 = & -2\cdot a + 3\cdot b & E_1 - 2E_2 \\
E_4: & 212 = & 3\cdot a - 4\cdot b & E_2 - E_3 \\
E_5: & 2 = & -5\cdot a + 7\cdot b & E_3 - E_4 \\
E_6: & 0 & & E_4 - 106E_5
\end{array}$$

So $\gcd(1492, 1066) = 2$ and $2 = -5\cdot 1492 + 7\cdot 1066$.

Definition 8.2.4. *Let $n \in \mathbb{Z}^+$ and $a, b \in \mathbb{Z}$. Then*

$$\mathbb{Z}_n := \{a \in \mathbb{Z} \mid 0 \leq a < n\},$$

$$\mathbb{Z}_n^* := \{a \in \mathbb{Z}_n \mid \gcd(a, n) = 1\},$$

and

$$\phi(n) = |\mathbb{Z}_n^*|.$$

If $a \in \mathbb{Z}$, then $[a]_n$ denotes the remainder of a when divided by n.

The relation '$\equiv \pmod{n}$' on $\mathbb{Z}$ is defined by

$$a \equiv b \pmod{n} \qquad \Longleftrightarrow \qquad n \mid b - a$$

Lemma 8.2.5. *Let $a, b, a', b', n \in \mathbb{Z}$ with $n > 0$. If*

$$a \equiv a' \pmod{n} \qquad \text{and} \qquad b \equiv b' \pmod{n}$$

then

$$ab \equiv a'b' \pmod{n}$$

Proof. Since $a \equiv a'$ there exists $k \in \mathbb{Z}$ with $a' - a = kn$. So $a' = a + kn$. By symmetry, $b' = b + ln$ for some $l \in \mathbb{Z}$. Thus

$$a'b' - ab = (a + kn)(b + ln) - ab = (al + kb + kln)n$$

So n divides $a'b' - ab$ the lemma holds. □

Lemma 8.2.6. *Let $n \in \mathbb{Z}^+$. Let $d \in \mathbb{Z}$ with $\gcd(d, n) = 1$. Then there exists $e \in \mathbb{Z}_n^*$ with $de \equiv 1 \pmod{n}$.*

Proof. By the Euclidean algorithm there exist $r, s \in \mathbb{Z}$ with

$$1 = \gcd(d, n) = rd + sn.$$

Hence $rd \equiv 1 \pmod{n}$. Put $e = [r]_n$. Then $0 \leq e < n$ and so $e \in \mathbb{Z}_n$. Note that $e \equiv r \pmod{n}$ and so by 8.2.5 $ed \equiv rd \equiv 1 \pmod{n}$.

In particular, $n \mid ed - 1$ and so $ed - 1 = mn$ for some $m \in \mathbb{Z}$. Hence $1 = ed - mn$ and any common divisor of n and e will divide 1. Thus $\gcd(n, e) = 1$ and so $e \in \mathbb{Z}_n^*$. □

Lemma 8.2.7. *Let $n, m \in \mathbb{Z}$ with $n > 0$. Then* $\gcd(n, m) = \gcd(n, [m]_n)$.

Proof. Observe that $m = qn + [m]_n$ for some $q \in \mathbb{Z}$ and so the lemma follows from 8.2.1. □

Lemma 8.2.8. *Let $n \in \mathbb{Z}^+$, $x, y \in \mathbb{Z}$, $a, b \in \mathbb{Z}_n$ and $d \in \mathbb{Z}$ with* $\gcd(d, n) = 1$.

(a) *If $a \equiv b \pmod{n}$, then $a = b$.*

(b) *If $xd \equiv yd \pmod{n}$, then $x \equiv y \pmod{n}$*

(c) *If $[ad]_n = [bd]_n$, then $a = b$.*

Proof. (a) Since $a \equiv b \pmod{n}$, $n \mid a - b$. As $a, b \in \mathbb{Z}_n^*$ we have $0 < a, b < n$ and so $|a - b| < n$. Thus $a - b = 0$ and $a = b$.

(b) By 8.2.6 there exists $e \in \mathbb{Z}_n^*$ with $de \equiv 1 \pmod{n}$. Since $xd \equiv yd \pmod{n}$ have $exd \equiv eyd \pmod{n}$ and so also $(de)x \equiv (de)y \pmod{n}$. As $de \equiv 1 \pmod{n}$ this gives $x \equiv y \pmod{n}$.

(c): From $[ad]_n = [bd]_n$ we get $ad \equiv bd \pmod{n}$. Hence (b) shows that $a \equiv b \pmod{n}$ and then (a) gives $a = b$. □

Lemma 8.2.9. *Let $n, m \in \mathbb{Z}^+$ with* $\gcd(n, m) = 1$. *Then* $\phi(nm) = \phi(n)\phi(m)$.

Proof. Consider the function

$$\alpha : \quad \mathbb{Z}_{nm} \to \mathbb{Z}_n \times \mathbb{Z}_m, \quad a \mapsto ([a]_n, [a]_m)$$

We claim that α is $1-1$ and onto. Let $a, b \in \mathbb{Z}_{nm}$ with $[a]_n = [b]_n$ and $[a]_m = [b]_m$. Then n and m divide $a - b$ and since $\gcd(n, m) = 1$ we get $nm \mid a - b$. Thus $a \equiv b \pmod{nm}$. As $a, b \in \mathbb{Z}_{nm}$ we conclude that $a = b$ by (8.2.8)(a). So α is 1-1. Since $|\mathbb{Z}_{nm}| = nm = |\mathbb{Z}_n \times \mathbb{Z}_m|$, α is also onto.

Note that $\gcd(a, nm) = 1$ if and only if $\gcd(a, n) = 1$ and $\gcd(a, m) = 1$. By 8.2.7 this holds if and only if $\gcd([a]_n, n) = 1$ and $\gcd([a]_m, m) = 1$. It follows that $a \in \mathbb{Z}_{nm}^*$ if and only if $([a]_n, [b]_m) \in \mathbb{Z}_n^* \times \mathbb{Z}_m^*$. Thus α induces a bijection

$$\alpha^* : \quad \mathbb{Z}_{nm}^* \to \mathbb{Z}_n^* \times \mathbb{Z}_m^*, \quad a \mapsto ([a]_n, [a]_m).$$

So

$$\phi(nm) = |\mathbb{Z}_{nm}^*| = |\mathbb{Z}_n^* \times \mathbb{Z}_m^*| = \phi(n)\phi(m).$$

□

Corollary 8.2.10. *Let p and q be distinct positive prime integers. Then*

$$\phi(p) = p - 1 \qquad \textit{and} \qquad \phi(pq) = (p-1)(q-1) = pq + 1 - (p+q).$$

Proof. Note that $\mathbb{Z}_p^* = \{1, 2, \dots, p-1\}$ and so $\phi(p) = p - 1$. By 8.2.9 $\phi(pq) = \phi(p)\phi(q) = (p-1)(q-1) = pq + 1 - (p+q)$. □

Lemma 8.2.11. *Let $a \in \mathbb{Z}_n^*$. Then $a^{\phi(n)} \equiv 1 \pmod{n}$.*

Proof. For $d \in \mathbb{Z}$ put $\overline{d} := [d]_n$. Let $b \in \mathbb{Z}_n^*$. Note that $\gcd(ab, n) = 1$ and so by 8.2.7, also $\gcd(\overline{ab}, n) = 1$. Thus $\overline{ab} \in \mathbb{Z}_n^*$ and we obtain a well-defined function

$$\alpha : \quad \mathbb{Z}_n^* \to \mathbb{Z}_n^*, \quad b \mapsto \overline{ab}.$$

Let $b, c \in \mathbb{Z}_n^*$ with $\alpha(b) = \alpha(c)$. Then $\overline{ab} = \overline{ac}$ and 8.2.8 shows that $b = c$. Thus α is 1-1. As α is function from a finite set to itself, this shows that α is bijection.

It follows that

$$\prod_{b \in \mathbb{Z}_n^*} b = \prod_{b \in \mathbb{Z}_n^*} \overline{ab}.$$

Since $\overline{ab} \equiv ab \pmod{n}$ we conclude from 8.2.5 that

$$(*) \qquad \prod_{b \in \mathbb{Z}_n^*} b \equiv \prod_{b \in \mathbb{Z}_n^*} \overline{ab} \equiv \prod_{b \in \mathbb{Z}_n^*} ab = \prod_{b \in \mathbb{Z}_n^*} a \prod_{b \in \mathbb{Z}_n^*} b \equiv a^{\phi(n)} \prod_{b \in \mathbb{Z}_n^*} b \pmod{n}.$$

Put $e := \prod_{b \in \mathbb{Z}_n^*} b$. Then

$$1e \equiv e \overset{(*)}{\equiv} a^{\phi(n)} e \pmod{n}.$$

Observe that $\gcd(e, n) = 1$ and so 8.2.8 implies that

$$1 \equiv a^{\phi(n)} \pmod{n}.$$

□

8.3 Definition of the RSA public-key cryptosystem

Definition 8.3.1. *Let*

$\mathcal{M}$ *be an alphabet* (plaintext messages)

Let $c \in \mathbb{Z}^+$. Put

$$\begin{aligned} \mathcal{C} &:= \big\{ n \in \mathbb{Z}^+ \mid n \leq c \big\}, && \text{(ciphertext messages)} \\ N &:= \big\{ n \in \mathcal{C} \mid \phi(n) \geq |\mathcal{M}| \big\} \end{aligned}$$

and

$$\mathcal{K} := \big\{ (n,d) \mid n \in N, d \in \mathbb{Z}^*_{\phi(n)} \big\}. \qquad \text{(keys)}$$

For $n \in N$ let

$\alpha_n : \mathcal{M} \to \mathbb{Z}^*_n$ *be a code and let* $\beta_n : \mathbb{Z}^*_n \to \mathcal{M}$ *be a function*

with

$$\beta_n \circ \alpha_n = \mathrm{id}_{\mathcal{M}}.$$

Given a key $(n,d) \in \mathcal{K}$. Define

$$\begin{aligned} E_{n,d} :& \quad \mathcal{M} \to \mathcal{C}, \quad m \mapsto \big[\alpha_n(m)^d\big]_n && \text{(encryption functions)} \\ D_{n,d} :& \quad \mathcal{C} \to \mathcal{M}, \quad z \mapsto \beta_n\big([z^d]_n\big) && \text{(decryption functions)} \end{aligned}$$

Let π be a set of primes such that

$$|\mathcal{M}| \leq (p-1)^2 \leq c$$

for all $p \in \pi$. Define

$$A := \big\{ \ (p,q,d,e) \mid p,q \in \pi, \ p \neq q, \ d,e \in \mathbb{Z}^*_{\phi(pq)}, \ de \equiv 1 (\operatorname{mod} \phi(pq)) \ \big\}$$

and

$$\xi : \quad A \to \mathcal{K} \times \mathcal{K}, \quad (p,q,d,e) \mapsto \big((pq,d),(pq,e)\big)$$

Then $\big(\mathcal{M}, \mathcal{C}, (E_{n,d})_{(n,d)\in\mathcal{K}}, (D_{n,d})_{(n,d)\in\mathcal{K}}, \xi\big)$ is called an RSA public-key cryptosystem.

Theorem 8.3.2. *Any RSA public-key cryptosystem is a public-key cryptosystem.*

Proof. Let $(n,d) \in \mathcal{K}$. By definition of an RSA public-key cryptosystem we have $d \in \mathbb{Z}^*_{\phi(n)}$ and so by 8.2.6 there exists $e \in \mathbb{Z}^*_{\phi(n)}$ with

$$(*) \qquad de \equiv 1 \ (\operatorname{mod} \phi(n)).$$

We will show (n,e) is an inverse key for (n,d), that is $D_{n,e}\circ E_{n,d} = \mathrm{id}_{\mathcal{M}}$. For this let $m\in\mathcal{M}$ and put

$$(**) \qquad w := \alpha_n(m).$$

By definition of RSA public-key cryptosystem we have $w\in\mathbb{Z}_n^*$ and

$$(***) \qquad E_{n,d}(m) = [w^d]_n.$$

From $(*)$ we get $de = 1 + q\phi(n)$ for some $q\in\mathbb{Z}$. Moreover, by 8.2.11 $w^{\phi(n)} \equiv 1 \pmod{n}$. Hence

$$w^{de} \equiv w^{1+q\phi(n)} \equiv w\cdot(w^{\phi(n)})^q \equiv w1^q \equiv w \pmod{n}.$$

As $w\in\mathbb{Z}_n^*$ and so $1\le w\le n-1$ we conclude that

$$(+) \qquad w = \left[w^{de}\right]_n.$$

We compute

$$(++) \qquad \left[\left([w^d]_n\right)^e\right]_n = \left[w^{de}\right]_n \overset{(+)}{=} w$$

and so

$$D_{n,e}\Big(E_{n,d}(m)\Big) \overset{\text{def: } D_{n,e}}{=} \beta_n\Big(\left[E_{n,d}(m)^e\right]_n\Big) \overset{(***)}{=} \beta_n\bigg(\left[\left([w^d]_n\right)^e\right]_n\bigg)$$

$$\overset{(++)}{=} \beta_n(w) \overset{(**)}{=} \beta_n\big(\alpha_n(m)\big) \overset{\beta_n\circ\alpha_n=\mathrm{id}_{\mathcal{M}}}{=} m.$$

Thus (n,e) is an inverse key for (n,d). This shows that any RSA public-key cryptosystem is cryptosystem.

Let $(p,q,d,e)\in A$ and put $n := pq$. By definition of an RSA public-key cryptosystem we have $\xi(p,q,d,e) = \Big((n,d),(n,e)\Big)$, $d,e\in\mathbb{Z}_{\phi(n)}^*$ and $de\equiv 1 \pmod{\phi(n)}$. As just seen, this means that (n,e) is an inverse key to (n,d). Hence any RSA public-key cryptosystem is a public-key cryptosystem. □

Example 8.3.3. Let

$$\mathcal{M} = \mathbb{A} = \{\sqcup, A, \dots, Z\}$$

and $c = 1000$. Then $N = \{i\in\mathbb{Z}^+ \mid i\le 1000, \phi(i)\ge 27\}$. Let $n\in N$.

Define $a_{i,n}$ for $1 \leq i \leq \phi(n)$ by

$$\mathbb{Z}_n^* = \{a_{1,n}, \dots, a_{\phi(n),n}\}$$

and $a_{i,n} < a_{i+1,n}$, if $i < \phi(n)$.

Define l_i for $0 \leq i \leq 26$ by

$$(\sqcup, A, \dots, Z) = (l_0, l_1, \dots l_{26})$$

.

Define

$$\alpha_n : \quad \mathcal{M} \to \mathbb{Z}_n^*, \quad l_i \mapsto a_{i+1,n}$$

and

$$\beta_n : \quad \mathbb{Z}_n^* \to \mathcal{M}, \quad j \mapsto l_{[i-1]_{27}}, \quad \text{where} \quad 1 \leq i \leq \phi(n) \text{ with } j = a_{i,n}$$

(a) Compute $u := E_{(667,5)}(K)$.

(b) Find the inverse key $(667, e)$ for $(667, 5)$.

(c) Verify that $D_{667,e}(u) = K$.

Recall first that $E_{667,5}(K) = [\alpha_{667}(K)^5]_{667}$. Note that $K = l_{11}$ and so $\alpha_{667} = a_{12,667}$. Thus α_{667} is the twelfth positive integer coprime to 667. As $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ all are coprime to 667 we see that

$$\alpha_{667}(K) = 12$$

To compute $[12^5]_{667}$ we determined 12^5 modulo 667

$$12^2 \equiv 144$$

$$12^3 \equiv \begin{array}{r} 1440 \\ +\ 288 \\ \hline 1728 \end{array}$$

$$12^3 \equiv 1728 - 3 \cdot 667 \equiv -(2001 - 1728) \equiv -273$$

$$12^4 \equiv -273 \cdot 12 \equiv - \begin{array}{r} 2730 \\ +\ 546 \\ \hline 3286 \end{array}$$

$$12^4 \equiv -3286 + 5 \cdot 667 \equiv 3335 - 3286 \equiv 59$$

$$12^5 \equiv 59 \cdot 12 \equiv 720 - 12 \equiv 708 \equiv 778 - 667 \equiv 41$$

Hence

$$E_{667,5}(K) = E_{667,5}(K) = \left[\alpha_{667}(K)^5\right]_{667} = [12^5]_{667} = 41.$$

Recall from the proof of 8.3.2 that $(667, e)$ will be an inverse key for $(667, 5)$ provided that $e \in \mathbb{Z}_{\phi(667)}$ with $5 \cdot e \equiv 1 \pmod{\phi(667)}$. To compute $\phi(667)$ we need to factorize 667 as a product of primes. None of $3, 5, 11$ divides 667, $667 - 7 = 660$, $667 + 13 = 680$, $667 - 17 = 650$, $667 + 23 = 690$. So $667 = 23 \cdot 29$. Thus by 8.2.10 $\phi(667) = 667 - (23 + 29) + 1 = 668 - 52 = 616$. Note that

$$1 = 616 - 123 \cdot 5$$

So $e = [-123]_{616} = 616 - 123 = 493$. A long calculation by hand or a quick computer calculation shows that $41^{493} \equiv 12 \pmod{667}$. Hence

$$D_{667,493}(41) = \beta_{667}\left(\left[41^{493}\right]_{667}\right) = \beta_{667}(12) = l_{[12-1]_{27}} = l_{11} = K.$$

Chapter 9

Noisy Channels

9.1 The definition of a channel

Definition 9.1.1. *Let I and J be alphabets. An $I \times J$- channel is $I \times J$-matrix $\Gamma = [\Gamma_{ij}]_{\substack{i\in I\\ j\in J}}$ with coefficients in $[0,1]$ such that*

$$\sum_{j\in J} \Gamma_{ij} = 1$$

for all $i \in I$.

I is called the input alphabet of Γ and J the output alphabet.

We interpret Γ_{ij} as the probability that the symbol j is received when the symbol i is send through the channel Γ.

Example 9.1.2. is a channel with input alphabet $\{a,b,c,d\}$ and output alphabet $\{a,b,c,d\}$.

Γ	a	b	c	d	e
a	0.3	0.2	0.1	0.1	0.3
b	0.4	0.2	0.1	0	0.3
c	0.7	0	0.1	0	0.1
d	0.1	0.2	0.3	0	0.4

Lemma 9.1.3. *Let I and J be alphabets and Γ an $I \times J$-matrix with coefficients in $\mathbb{R}$. Then Γ is a channel if and only if each row of Γ is a probability distribution on J.*

Proof. Both conditions just say that $\Gamma_{ij} \in [0,1]$ for all $i \in I, j \in J$ and $\sum_{j\in J} \Gamma_{ij} = 1$ for all $i \in I$. □

Definition 9.1.4. (a) *The transpose of an* $I \times J$*-matrix* $M = [m_{ij}]_{\substack{i\in I\\ j\in J}}$ *is the* $J \times I$ *matrix* $M^{\mathrm{Tr}} = [m_{ij}]_{\substack{j\in J\\ i\in I}}$.

(b) *An* $I \times J$*-channel* Γ *is called symmetric if* $\frac{|J|}{|I|}\Gamma^{\mathrm{Tr}}$ *is a channel. Note that this is the case if and only if* $\sum_{j\in J}\Gamma_{ij} = \frac{|I|}{|J|}$ *for all* $j \in J$ *and if and only if all columns of* Γ *have the same sum.*

(c) *A binary symmetric channel* BSC *is a symmetric channel with input and output alphabet* $\mathbb{B}$.

(d) *Let* Γ *be a binary symmetric channel. Then* $e = \Gamma_{01}$ *is called the bit error of* Γ.

Lemma 9.1.5. *Let* e *be the bit error of a binary symmetric channel* Γ. *Then*

$$\Gamma = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-e & e \\ 1 & e & 1-e \end{array}$$

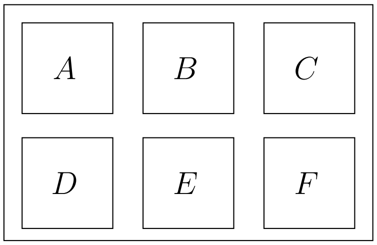
Proof. By definition $\Gamma_{01} = e$. Since $\Gamma_{00} + \Gamma_{01} = 1$, $\Gamma_{00} = 1 - e$. Since Γ is symmetric the sum of each column must be $\frac{|\mathbb{B}|}{|\mathbb{B}|} = 1$. So $\Gamma_{10} = 1 - \Gamma_{00} = e$ and $\Gamma_{11} = 1 - \Gamma_{01} = 1 - e$. □

Notation 9.1.6. *We will usually write an* $I \times J$*-matrix just as an* $|I| \times |J|$*-matrix, that is we do not bother to list the header row and column. Of course this simplified notation should only be used if a fixed ordering of elements in* I *and* J *is given.*

For example we will denote the BSC with error bit e by

$$\mathrm{BSC}(e) = \begin{bmatrix} 1-e & e \\ e & 1-e \end{bmatrix}$$

Example 9.1.7. Consider the simplified keypad



Two keys are called adjacent if an edge of the one is next to an edge of the other.

Suppose that for any two adjacent keys x and y there is a 10% chance that y will be pressed when intending to press x.

The channel is

Γ	A	B	C	D	E	F
A	0.8	0.1	0	0.1	0	0
B	0.1	0.7	0.1	0	0.1	0
C	0	0.1	0.8	0	0	0.1
D	0	0	0	0.8	0.1	0
E	0	0.1	0	0.1	0.7	0.1
F	0	0	0.1	0	0.1	0.8

9.2 Transmitting a source through a channel

Definition 9.2.1. *Let I and J be alphabets, let Γ and t be $I \times J$ matrices , let p be an I-tuple and let q be an J-tuple.*

(a) *We say that say that q is linked to p via Γ if $q = p\Gamma$, that is $q_j = \sum_{i \in I} p_i \Gamma_{ij}$.*

(b) $\operatorname{Diag}(p)$ *is the $I \times I$ matrix $[d_{ik}]$ where*

$$d_{ik} = \begin{cases} p_i & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}$$

for all $i, k \in I$.

Lemma 9.2.2. *Let I and J be alphabets,let Γ and t be $I \times J$-matrices, p an I-tuple and q a J-tuple, all with coefficients in $\mathbb{R}^{\geq 0}$. Suppose that*

$$t = \operatorname{Diag}(p)\Gamma \quad (\text{ so } t_{ij} = p_i \Gamma_{ij} \text{ and } t_i = p_i \Gamma_i)$$

(a) *p is the marginal tuple of t on I if and only if Γ_i is a probability distributions on J for all $i \in I$ with $p_i \neq 0$.*

(b) *If p is positive, p is the marginal tuple of t on I if and only if Γ is a channel.*

(c) *q is linked to p via Γ if and only if q is the marginal tuple of t on J.*

Proof. (a)

$$\begin{array}{rc} & p \text{ is the marginal distribution of } t \text{ on } I \\ \iff & \sum_{j\in J} t_{ij} = p_i \text{ for all } i \in I \\ \iff & \sum_{j\in J} p_i\Gamma_{ij} = p_i \text{ for all } i \in I \\ \iff & \sum_{j\in J} \Gamma_{ij} = 1 \text{ for all } i \in I \text{ with } p_i \neq 0 \\ \iff & \Gamma_i \text{ is a probability distribution on } J \text{ for all } i \in I \text{ with } p_i \neq 0 \end{array}$$

So (a) holds.
(b) Follows from (a).
(c)

$$\begin{array}{rc} & q \text{ is linked to } p \text{ via } \Gamma \\ \iff & q = p\Gamma \\ \iff & q_j = \sum_{i\in I} p_i\Gamma_{ij} \text{ for all } j \in J \\ \iff & q_j = \sum_{i\in I} t_{ij} \text{ for all } j \in J \\ \iff & q \text{ is the marginal distribution of } t \text{ on } J \end{array}$$

□

Definition 9.2.3. *Let I and J be alphabets.*

(a) *$I \times J$-channel system is a tuple (Γ, t, p, q) such that Γ is a $I \times J$-channel, t, p and q are probability distribution on $I \times J$, I and J respectively, $t = \mathrm{Diag}(p)\Gamma$ and $q = p\Gamma$.*

(b) *Let Γ be a $I \times J$-channel and p a probability distribution on T. Then $t = \mathrm{Diag}(p)\Gamma$ is called the joint distribution for Γ and p. $(\Gamma, t, p, p\Gamma)$ is called the Channel system for Γ and p and is denoted by $\Sigma(\Gamma, p)$.*

(c) *Let t be a probability distribution on $I \times J$ with marginal distribution p and q and Γ an $I \times J$-channel, Γ is called a channel associated to t (and (Γ, t, p, q) is called a channel system associated to t) if $t = \mathrm{Diag}(p)\Gamma$.*

(d) *Let $\Sigma = (\Gamma, t, p, q)$ be an $I \times J$ channel system. Let $i \in I$ and $j \in J$. Then*

- *t_{ij} is called the probability that i is send and j is received and is denoted by $\mathrm{Pr}^{\Sigma}(i, j)$.*
- *p_i is called probability that i is send, and is denoted by $\mathrm{Pr}^{\Sigma}(i, *)$.*
- *q_j is called the probability that j is received, and is denoted by $\mathrm{Pr}^{\Sigma}(*, j)$.*

- Γ_{ij} *is called the probability that* j *is received when* i *is send and is denoted by* $\Pr^{\Sigma}(j|i)$.

Assuming that there is no doubt underlying channel system, we will usually drop the superscript Σ.

Example 9.2.4. Compute the channel system for the channel BSC(e) and the probability distribution $(p, 1-p)$.

$$\begin{aligned}
t &= \operatorname{Diag}(p, 1-p))\,\mathrm{BSC}(e)\\
&= \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix}\begin{bmatrix} 1-e & e \\ e & 1-e \end{bmatrix}\\
&= \begin{bmatrix} p(1-e) & pe \\ (1-p)e & (1-p)(1-e) \end{bmatrix}\\
&= \begin{bmatrix} p-pe & pe \\ e-pe & 1-p-e+pe \end{bmatrix}
\end{aligned}$$

Since q is the column sum of t:

$$q = \big(p(1-e) + (1-p)e, pe + (1-p)(1-e)\big) = \big(p + e - 2pe, 1 + 2pe - p - e\big)$$

As a more concrete example consider the case $e = 0.1$ and $p = 0.3$. Then

$$\Gamma = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix},$$

$$t = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.7 \end{bmatrix}\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.27 & 0.03 \\ 0.07 & 0.63 \end{bmatrix}$$

and since q is the column sum of t

$$q = (0.34, 0.66)$$

Lemma 9.2.5. *Let I and J be alphabets and Let t be a probability distribution, p the marginal distribution of t on I and Γ an $I \times J$-matrix.*

(a) *Γ is a channel associated to t if and only if*

(i) *$\Gamma_i = \frac{1}{p_i} t_i$ for all $i \in I$ with $p_i \neq 0$, and*

(ii) *Γ_i is a probability distribution on J for all $i \in I$ with $p_i = 0$*

(b) *There exists a channel associated to t.*

(c) *If p is positive, then $\operatorname{Diag}(p)^{-1}\Gamma$ is the unique channel associated to associated to t.*

Proof. (a) Suppose first that Γ is a channel associated to t. Then $t = \operatorname{Diag}(p)\Gamma$ and so $t_i = p_i\Gamma_i$. Thus (a:i) holds. Since Γ is channel, also (a:ii) holds.

Suppose next that (a:i) and (a:ii) holds. If $p_i = 0$, then since $p_i = \sum_{j\in J} t_{ij}$ and $t_{ij} = 0$, also $t_{ij} = 0$ for all $j \in J$. Thus $t_i = p_i\Gamma_i$ for all $i \in I$. Hence $t = \operatorname{Diag}(p)\Gamma$. From 9.2.2 we conclude that Γ_i is a probability distribution on J for all $i \in I$ with $p_i \neq 0$. Together with (a:ii) this shows that Γ is a channel.

(b) Let Γ be the $I \times J$-matrix such that $\Gamma_i = \frac{1}{p_i} t_i$ if $p_i \neq 0$ and Γ_i is the equal probability distribution on J if $p_i = 0$. Then by (a) Γ is a channel associated to t.

(c) Follows immediately from (a). □

9.3 Conditional Entropy

Definition 9.3.1. *Let $\Sigma = (\Gamma, t, p, q)$ be a channel system.*

(a) *$H(t)$ is called the joint entropy of p and q and is denoted by $H^\Sigma(p,q)$.*

(b) *$H(t) - H(q)$ is called the conditional entropy of p given q with respect to t, and is denoted by $H^\Sigma(p|q)$ and $H(\Gamma; p)$.*

(c) *$H(t) - H(p)$ is called the conditional entropy of q given p with respect to t, and is denoted by $H^\Sigma(q|p)$.*

(d) *$H(p) + H(q) - H(t)$ is called the mutual information of p and q with respect to t and is denoted by $I^\Sigma(p,q)$.*

Definition 9.3.2. (a) *Let f be an I tuple and g a J-tuple. We say that f is a permutation of g if there exists a bijection $\pi : J \to I$ with $g_j = f_{\pi(j)}$ for all $j \in J$.*

(b) *Let Γ be a $I \times J$-channel and E a probability distribution on J. We say that Γ is additive with row distribution E if each row of Γ is a permutation of E.*

Example 9.3.3. $(0,1,0,3,0.4,0,2)$ is a permutation of $(0.4,0.2,0.3,0.1)$.

Example 9.3.4. BSC(e) is additive with row distribution $E=(e,1-e)$.

Lemma 9.3.5. (a) *Let p and p' be probability distributions. If p is a permutation of p', then $H(p)=H(p')$.*

(b) *Let (Γ,t,p,q) be a channel system. Suppose that Γ is additive with row distribution E. Then t is a permutation of $p\otimes E$ and*

$$\begin{array}{lcl} H(t) & = & H(p)+H(E) \\ H(q|p) & = & H(E) \end{array} \qquad \begin{array}{lcl} H(p|q) & = & H(p)+H(E)-H(q) \\ I(p,q) & = & H(q)-H(E) \end{array}$$

Proof. (a) $H(p)$ and $H(p')$ are sums of the same numbers $p_i\log\left(\frac{1}{p_i}\right)$, just in a different order.

(b) The i'th row of t is $p_i\Gamma_i$ and the i'th row of $p\otimes E$ is p_iE. Since Γ_i is a permutation of E, also $p_i\Gamma_i$ is a permutation of p_iE. So t is a permutation of $p\otimes E$. Thus by (a) and 4.1.9

$$H(t)=H(p\otimes E)=H(p)+H(E)$$

Hence

$$H(p|q)=H(t)-h(q)=H(p)+H(E)-H(q),$$

$$H(q|p)=H(t)-h(p)=H(E)$$

and

$$I(p,q)=H(p)+H(q)-H(t)=H(q)-H(E)$$

□

Corollary 9.3.6. *Let (BSC(e)$,t,p,q$) be a channel system. Then*

$$H\big(p|q\big)=H\big(p\big)+H\big((e,1-e)\big)-H\big(q\big)$$

9.4 Capacity of a channel

Theorem 9.4.1. *Let (Γ,t,p,q) be a channel system. Then*

$$I(p,q)\geq 0$$

with equality of if p and q are independent with respect to t.

Proof. Note that $I(p,q) \geq 0$ if and only if $H(t) \leq H(p) + H(q)$. Since p and q are the marginal distribution of t, the result now follows from 4.1.9 □

Of course one does not want the output of the channel to be independent of the input. So one likes $I(p,q)$ to be as large as possible. This leads to the following definition:

Definition 9.4.2. *Let Γ be a $I \times J$ channel and let $\mathcal{P}(I)$ be set probability distribution on I. Define*

$$f_\Gamma : \mathcal{P}(I) \to \mathbb{R}, \quad p \to I^{\Sigma(\Gamma,p)}(p, p\Gamma)$$

and

$$\gamma(\Gamma) = \max f_\Gamma = \max_{p \in \mathcal{P}(I)} f_\Gamma(p)$$

Then $\gamma(\Gamma)$ is called the capacity of the channel Γ.

In little less precise notation

$$\gamma(\Gamma) = \max_p I(p,q)$$

Theorem 9.4.3. *Let Γ be an additive $I \times J$ channel with row distribution E. Then*

$$\gamma(\Gamma) = \left(\max_{p \in \mathcal{P}(I)} H(p\Gamma) \right) - H(E) \leq \log |J| - H(E)$$

with equality if and only if $p\Gamma$ is the equal probability distribution on J for some probability distribution p on I.

Proof. Let $p \in \mathcal{P}(I)$ and (Γ, t, p, q) the channel system for Γ and p. So $q = p\Gamma$ and $t = \mathrm{Diag}(p)\Gamma$. By 9.3.5

$$I(p,q) = H(q) - H(E)$$

Since $\gamma(\Gamma) = \max_p I(p,q)$ the first equality holds. By 3.4.2 $H(q) \leq \log |J|$ with equality if and only if q is the equal probability distribution. Hence also the second inequality holds. □

Corollary 9.4.4. *Let Γ be an symmetric, additive $I \times J$ channel with row distribution E. Then*

$$\gamma(\Gamma) = \log |J| - H(E)$$

Proof. Let $p = (\frac{1}{|I|})_{i\in I}$ be the equal probability distribution on I. Let $j \in J$. We compute

$$q_j = \sum_{i\in I} p_i \Gamma_{ij} = \frac{1}{|I|}\sum_{i\in I}\Gamma_{ij} = \frac{1}{|I|}\frac{|I|}{|J|} = \frac{1}{|J|}$$

where the second equality holds since Γ is symmetric. So q is the equal probability distribution on J. Since Γ is additive, the Corollary now follows from 9.4.3 □

Corollary 9.4.5. $\gamma(\mathrm{BSC}(e)) = 1 - H((e, 1-e)) = 1 - e\log\left(\frac{1}{e}\right) - (1-e)\log\left(\frac{1}{1-e}\right)$.

Proof. Since $\mathrm{BSC}(e)$ is a symmetric, additive channel with row distribution $(e, 1-e)$ and output alphabet of size 2, this follows immediately from 9.4.4. □

Lemma 9.4.6. *Let (Γ, t, p, q) be a channel system. Then*

$$H(q|p) = \sum_{i\in I} p_i\, H(\Gamma_i) = \sum_{(i,j)\in I\times J} t_{ij}\log\left(\frac{1}{\Gamma_{ij}}\right)$$

Proof. Recall that p is the marginal distribution for t on I and so

(1) $$p_i = \sum_{j\in J} t_{ij}$$

Also

(2) $$\frac{p_i}{t_{ij}} = \frac{p_i}{p_i\Gamma_{ij}} = \frac{1}{\Gamma_{ij}}.$$

We compute

$$\begin{aligned}
H(q|p) &= H(t) - H(p) \\
&= \sum_{(i,j)\in I\times J} t_{ij}\log\left(\frac{1}{t_{ij}}\right) - \sum_{i\in I} p_i \log\left(\frac{1}{p_i}\right) \\
&\overset{(1)}{=} \sum_{(i,j)\in I\times J} t_{ij}\log\left(\frac{1}{t_{ij}}\right) - \sum_{i\in I}\left(\sum_{j\in J} t_{ij}\right)\log\left(\frac{1}{p_i}\right) \\
&= \sum_{(i,j)\in I\times J} t_{ij}\log\left(\frac{p_i}{t_{ij}}\right) \\
&\overset{(2)}{=} \sum_{(i,j)\in I\times J} t_{ij}\log\left(\frac{1}{\Gamma_{ij}}\right) \\
&= \sum_{(i,j)\in I\times J} p_i\Gamma_{ij}\log\left(\frac{1}{\Gamma_{ij}}\right) \\
&= \sum_{i\in I} p_i\left(\sum_{j\in J}\Gamma_{ij}\log\left(\frac{1}{\Gamma_{ij}}\right)\right) \\
&= \sum_{i\in I} p_i\, H(\Gamma_i)
\end{aligned}$$

□

Chapter 10

The noisy coding theorems

10.1 The probability of a mistake

Definition 10.1.1. *Let I and J be alphabets.*

(a) *An $I \times J$-decision rule is a function $\sigma : J \to I$.*

(b) *Let $i \in I$ and j in J. Then (i, j) is called a mistake for σ if $i \neq \sigma(j)$.*

(c) *An $I \times J$-decision system is a tuple $(\Gamma, t, p, q, \sigma)$, where (Γ, t, p, q) is a $I \times J$-channel system and σ an $I \times J$- decision rule.*

(d) *Let Γ be a $I \times J$-channel, p a probability distribution of I and σ and $I \times J$ decision rule. Let (Γ, t, p, q) be the channel system for Γ and p. Then $(\Gamma, t, p, q, \sigma)$ is called the decision system for Γ, p and σ and is denoted by $\Sigma(\Gamma, p, \sigma)$.*

Definition 10.1.2. *Let $\Sigma = (\Gamma, t, p, q, \sigma)$ be an $I \times J$-decision system. Let $i \in I$.*

(a) *Then $F^\sigma(i) = \{j \in J \mid \sigma(j) \neq i\}$.*

(b) *$M^\Sigma(i) = \sum_{j \in F^\sigma(i)} \Gamma_{ij}$. $M^\Sigma(i)$ is called the probability of a mistake if i is send.*

(c) *$M^\Sigma = \sum_{i \in I} p_i M^\Sigma(i)$. M^Σ is called the probability of a mistake.*

Of course we will often drop the superscripts.

Definition 10.1.3 (Ideal Observer Rule)**.** *Let $\Sigma = (\Gamma, t, p, q, \sigma)$ be an $I \times J$-decision system. We say that σ is an Ideal observer rule with respect to Σ if for all $i \in I$ and $j \in J$,*

$$t_{ij} \leq t_{\sigma(j)j}$$

Since $\Pr(i|j) = \frac{t_{ij}}{q_j}$, this is equivalent to

$$\Pr(i|j) \le \Pr(\sigma(j)|j) \text{ for all } j \in J \text{ with } q_j \ne 0$$

Example 10.1.4. Find the Ideal Observer Rule for the channel BSC(0.3) and probability distribution (0.2, 0.8). What is the probability of a mistake?

We have

$$\Gamma = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \quad \text{and} \quad t = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.14 & 0.06 \\ 0.24 & 0.56 \end{bmatrix}$$

Let σ be an ideal observer rule. The largest entry in the first column of t occurs in the second row. So $\sigma(0) = 1$. The largest entry in the second column of t occurs in the second row. So $\sigma(1) = 1$. So the receiver always decides that 1 was send, regardless on what was received.

The mistakes are $(0,0)$ and $(0,1)$. Thus

$$F(0) = \{0,1\} \quad \text{and} \quad F(1) = \{\}$$

So

$$M(0) = \Gamma_{00} + \Gamma_{01} = 1 \quad \text{and} \quad M(1) = 0$$

Hence

$$M = p_0 M(0) + p_1 M(1) = 0.2 \cdot 1 + 0.8 \cdot 0 = 0.2$$

Definition 10.1.5 (Maximum Likelihood Rule). *Let $\Sigma = (\Gamma, t, p, q, \sigma)$ be a $I \times J$-decision system. We say that σ is a Maximum Likelihood Rule with respect to Γ if for all $i \in I$ and $j \in J$*

$$\Gamma_{ij} \le \Gamma_{\sigma(j)j}$$

Since $\Gamma_{ij} = \Pr(j|i)$, this is the same as

$$\Pr(j|i) \le \Pr(j|\sigma(j))$$

for all $i \in I$ and $j \in J$.

Example 10.1.6. Find the Maximum Likelihood rule for channel BSC(0.3). What is the probability of a mistake with respect to the probability distribution (0.2, 0.8)?

We have

$$\Gamma = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

Let σ be a Maxumim Likelihood rule. The largest entry in the first column of Γ occurs in the first row, So $\sigma(0) = 0$. The largest entry in the second column of t occurs in the second row. So $\sigma(1) = 1$. So the receiver always decides that the symbol received was the symbol send.

The mistakes are $(0,1)$ and $(1,0)$. Thus

$$F(0) = \{1\} \quad \text{and} \quad F(1) = \{0\}$$

So

$$M(0) = \Gamma_{01} = 0.3 \quad \text{and} \quad M(1) = \Gamma_{10} = 0.3$$

Hence

$$M = p_0 M(0) + p_1 M(1) = 0.2 \cdot 0.3 + 0.8 \cdot 0.3 = 0.3$$

10.2 Fano's inequality

Lemma 10.2.1. *Let $(\Gamma, t, p, q, \sigma)$ be an $I \times J$-decision system. M the probability of a mistake and K the set of mistakes. Then*

$$(I \times J) \smallsetminus K = \{(\sigma(j), j) \mid j \in J\},$$

$$M = \sum_{(i,j)\in K} t_{ij} \qquad \textit{and } 1 - M = \sum_{j\in J} t_{\sigma(j)j}$$

Proof. Note that $K = \{(i,j) \in I \times J \mid \sigma(j) \neq i\}$. So

$$(I \times J) \smallsetminus K = \{(i,j) \in I \times J \mid i = \sigma(j)\} = \{(\sigma(j), j) \mid j \in J\}$$

and the first statement is proved.

We compute

$$\begin{array}{ccccc} \sum_{(i,j)\in K} t_{ij} & = & \sum_{\substack{(i,j)\in I\times J \\ \sigma(j)\neq i}} t_{ij} & = & \sum_{i\in I} \left(\sum_{j\in F(i)} t_{ij} \right) \\ & = & \sum_{i\in I} \left(\sum_{j\in F(i)} p_i \Gamma_{ij} \right) & = & \sum_{i\in I} p_i \left(\sum_{j\in F(i)} \Gamma_{ij} \right) \\ & = & \sum_{i\in I} p_i M(i) & = & M \end{array}$$

Thus the second statement holds. Since $1 = \sum_{(i,j)\in I\times J} t_{ij}$, the first two statement imply the third. □

Theorem 10.2.2 (Fano's inequality)**.** *Let $(\Gamma, t, p, q, \sigma)$ be an $I \times J$-decision system and M the probability of a mistake. Then*

$$H(\Gamma; p) \leq H\big((M, 1-M)\big) + M \log\big(|I| - 1\big)$$

Proof. Let K be the sets of mistakes. By 10.2.1 $M = \sum_{(i,j)\in K} t_{ij}$ and $1 - M = \sum_{j\in J} t_{\sigma(j)j}$. Thus

$$\begin{array}{rcccc} H\big((M,1-M)\big) & = & M\log\left(\frac{1}{M}\right) & + & (1-M)\log\left(\frac{1}{1-M}\right) \\ & = & \sum_{(i,j)\in K} t_{ij}\log\left(\frac{1}{M}\right) & + & \sum_{j\in J} t_{\sigma(j)j}\log\left(\frac{1}{1-M}\right) \end{array}$$

Also by 10.2.1

$$I \times J = K \uplus \{(\sigma(j), j) \mid j \in J\}.$$

Let $\Delta = \left[\frac{t_{ij}}{q_j}\right]_{\substack{j\in J \\ i\in I}}$. By Exercise 9(f) on Homework 3

$$\begin{aligned} H(\Gamma;p) = H(p\mid q) \quad & = \sum_{j\in J} q_j H(\Delta_j) \\ = \sum_{j\in J} q_j \left(\sum_{i\in I} \Delta_{ji}\log\left(\frac{1}{\Delta_{ji}}\right)\right) \quad & = \sum_{j\in J}\sum_{i\in I} q_j \frac{t_{ij}}{q_j}\log\left(\frac{1}{\frac{t_{ij}}{q_j}}\right) \\ = \sum_{(i,j)\in I\times J} t_{ij}\log\left(\frac{q_j}{t_{ij}}\right) \quad & = \sum_{(i,j)\in K} t_{ij}\log\left(\frac{q_j}{t_{ij}}\right) + \sum_{j\in J} t_{\sigma(j)j}\log\left(\frac{q_j}{t_{\sigma(j)j}}\right) \end{aligned}$$

Put

$$S_1 = \sum_{(i,j)\in K} t_{ij}\log\left(\frac{q_j}{t_{ij}}\right) - \sum_{((i,j)\in K} t_{ij}\log\left(\frac{1}{M}\right) = \sum_{(i,j)\in K} t_{ij}\log\left(\frac{Mq_j}{t_{ij}}\right)$$

and

$$S_2 = \sum_{j\in J} t_{\sigma(j)j}\log\left(\frac{q_j}{t_{\sigma(j)j}}\right) - \sum_{j\in J} t_{\sigma(j)j}\log\left(\frac{1}{1-M}\right) = \sum_{j\in J} t_{\sigma(j)j}\log\left(\frac{(1-M)q_j}{t_{\sigma(j)j}}\right).$$

Then

$$H(\Gamma;p) - H\big((M,1-M)\big) = S_1 + S_2.$$

So it suffices to show $S_1 \leq M\log(|I|-1)$ and $S_2 \leq 0$.

For $(i,j) \in K$ put

$$v_{ij} = \frac{t_{ij}}{M} \quad \text{and} \quad w_{ij} = \frac{q_j}{|I|-1}.$$

Since $\sum_{(i,j)\in K} t_{ij} = M$, $(v_{ij})_{(i,j)\in K}$ is a probability distribution on K. Also

$$\sum_{(i,j)\in K} q_j = \sum_{j\in J} \sum_{\substack{i\in I\\ i\neq\sigma(j)}} q_j = \sum_{j\in J} (|I|-1)q_j = (|I|-1)\sum_{j\in J} q_j = |I|-1,$$

and so also $(w_{ij})_{(i,j)\in K}$ is a probability distribution on K . Thus by the Comparison Theorem 3.4.1

$$\begin{array}{rcl} 0 & \geq & \sum_{(i,j)\in K} v_{ij} \log\left(\frac{1}{v_{ij}}\right) - \sum_{(i,j)\in K} v_{ij} \log\left(\frac{1}{w_{ij}}\right) \\ & = & \sum_{(i,j)\in K} v_{ij} \log\left(\frac{w_{ij}}{v_{ij}}\right) \\ & = & \sum_{(i,j)\in K} \frac{t_{ij}}{M} \log\left(\frac{q_j M}{t_{ij}(|I|-1)}\right) \\ & = & \sum_{(i,j)\in K} \frac{t_{ij}}{M} \log\left(\frac{q_j M}{t_{ij}}\right) - \sum_{(i,j)\in K} \frac{t_{ij}}{M} \log\left(|I|-1)\right) \\ & = & \frac{1}{M} S_1 + \log(|I|-1) \end{array}$$

Thus indeed $S_1 \leq M \log(|I|-1)$.

For $j \in J$, put

$$u_j = \frac{t_{\sigma(j)j}}{1-M}.$$

Since $\sum_{j\in J} t_{\sigma(j)j} = 1 - M$ both q and $(u_j)_{j\in J}$ are probability distributions on J. So by the Comparison Theorem 3.4.1

$$\begin{array}{rcl} 0 & \geq & \sum_{j\in J} u_j \log\left(\frac{1}{u_j}\right) - \sum_{j\in J} u_j \log\left(\frac{1}{q_j}\right) \\ & = & \sum_{j\in J} u_j \log\left(\frac{q_j}{u_j}\right) \\ & = & \sum_{j\in J} \frac{t_{\sigma(j)j}}{1-M} \log\left(\frac{q_j(1-M)}{t_{\sigma(j)j}}\right) \\ & = & \frac{1}{1-M} S_2 \end{array}$$

and so indeed $S_2 \leq 0$. □

10.3 A lower bound for the probability of a mistake

Theorem 10.3.1. *Let $(\Gamma, t, p, q, \sigma)$ be an $I \times J$-decision system and M the probability of a mistake. Then*

$$M > \frac{H(p) - \gamma(\Gamma) - 1}{\log(|I|)}$$

In particular, if p is the equal probability distribution, then

$$M > 1 - \frac{\gamma(\Gamma) + 1}{\log(|I|)}$$

Proof. By the Fano inequality 10.2.2

$$(*) \qquad H(\Gamma; p) \leq H\big((M, 1 - M)\big) + M\log(|I| - 1) \leq \log 2 + M\log(|I| - 1)) < 1 + M\log(|I|)$$

By definition of the capacity,

$$\gamma(\Gamma) \geq I(p, q) = H(p) + H(q) - H(t) = H(p) - H(\Gamma; p)$$

and so

$$H(p) - \gamma(\Gamma) \leq H(\Gamma; p)$$

So (*) implies

$$H(p) - \gamma(\Gamma) < 1 + M\log(|I|)$$

and thus

$$M > \frac{H(p) - \gamma(\Gamma) - 1}{\log|I|}$$

So the first statement holds. If p is the equal probability distribution, then by 3.4.2 $H(p) = \log(|I|)$ and so

$$M > \frac{\log(|I|) - \gamma(\Gamma) - 1}{\log|I|} = 1 - \frac{\gamma(\Gamma) + 1}{\log|I|}$$

□

10.4 Extended Channels

Let Γ be an $I \times J$ channel and Γ' be an $I' \times J'$ channel. Suppose the two channel are 'unrelated'. The pair of channels is used to send a pair of symbols ii', namely i is send via Γ and i' via Γ'. Then the probability that the pair of symbols jj' is received is $\Gamma_{ij}\Gamma'_{i'j'}$. So the combined channel Γ'' has input $I \times I'$, output $J \times J'$ and

$$\Gamma''_{ii',jj'} = \Gamma_{ij}\Gamma'_{i'j'}.$$

This leads to the following definitions:

Definition 10.4.1. (a) *Let M be an $I \times J$-matrix and M' an $I' \times J'$ matrix Put $I'' = I \times I'$ and $J'' = J \times J'$. Then $M'' = M \otimes M'$ is the $I'' \times J''$-matrix defined by*

$$m''_{ii',jj'} = m_{ij}m'_{i'j'}$$

M'' is called tensor product of M and M'.

(b) *M be an $I \times J$-matrix and n a positive integer then $M^{\otimes n}$ is the $I^n \times J^n$ matrix inductively defined by*

$$M^{\otimes 1} = M \qquad \textit{and } M^{\otimes(n+1)} = M^{\otimes n} \otimes M$$

Example 10.4.2. Compute

$$\begin{array}{c|ccc} & a & b & c \\ \hline d & 0 & 1 & 2 \\ e & 3 & -1 & 0 \end{array} \bigotimes \begin{array}{c|cc} & v & w \\ \hline x & 4 & 5 \\ y & -1 & 3 \end{array} \quad \text{and} \quad \begin{bmatrix} 0.2 & 0.8 \\ 0.3 & 0.7 \end{bmatrix}^{\otimes 2}$$

$$\begin{array}{c|cccccc} & av & aw & bv & bw & cv & cw \\ \hline dx & 0 & 0 & 4 & 5 & 8 & 10 \\ dy & 0 & 0 & -1 & 3 & -2 & 6 \\ ex & 12 & 15 & -4 & -5 & 0 & 0 \\ ey & -3 & 9 & 1 & -3 & 0 & 0 \end{array} \quad \text{and} \quad \begin{bmatrix} 0.04 & 0.16 & 0.16 & 0.64 \\ 0.06 & 0.14 & 0.24 & 0.56 \\ 0.06 & 0.24 & 0.14 & 0.56 \\ 0.09 & 0.21 & 0.21 & 0.49 \end{bmatrix}$$

Lemma 10.4.3. *Let Γ be an $I \times J$-channel and Γ' an $I' \times J'$ channel. Put $I'' = I \times I'$, $J'' = J \times J'$ and $\Gamma'' = \Gamma \otimes \Gamma'$.*

Then

(a) $\Gamma''_{ii'} = \Gamma_i \otimes \Gamma'_{i'}$ *for all $i \in I, i' \in I'$.*

(b) *Γ'' is an $I'' \times J''$-channel.*

(c) *Let $n \in \mathbb{Z}^+$. Then $\Gamma^{\otimes n}$ is an $I^n \times J^n$- channel, called the n-fold extension of Γ.*

(d) *For all $x \in I^n, y \in J^n$, $\Gamma^{\otimes n}_{xy} = \prod_{k=1}^n \Gamma_{x_k y_k}$.*

Proof. (a)

$$(\Gamma''_{ii'})_{jj'} = \Gamma''_{ii',jj'} = \Gamma_{ij}\Gamma'_{i'j'} = (\Gamma_i)_j(\Gamma'_{i'})_{j'} = (\Gamma_i \otimes \Gamma_{i'})_{jj'}$$

(b) Let $i \in I$ and $i' \in I'$. Since Γ and Γ' are channels, Γ_i is a probability distribution on J and $\Gamma'_{i'}$ is a probability distribution on J'. Thus by 4.1.8, $\Gamma_i \otimes \Gamma'_{i'}$ is a probablity distribution on $J'' = J \times J'$. So by (a) all rows of Γ'' are probability distributions and hence Γ'' is a channel.

(c) Follows from (b) and induction on n.

(d) Can be proved using an easy induction argument. □

Lemma 10.4.4. *Let Γ be an $I \times J$-channel and Γ' an $I' \times J'$-channel. Put $\Gamma'' = \Gamma \otimes \Gamma'$ and suppose $\Sigma'' = (\Gamma'', t'', p'', q'')$ is a channel system. Let t and t' be the marginal distribution for t'' on $I \times J$ and $I' \times J'$, respectively. Let p and p' be the marginal distribution for p'' on I and I', respectively. Let q and q' be the marginal distribution for q'' on J and J', respectively*

(a) *$\Sigma = (\Gamma, t, p, q)$ and $\Sigma' = (\Gamma', t', p', q')$ are channel system.*

(b) *If p and p' are independent with respect to p'', then q and q' are independent with respect to q''.*

(c) *Let $i \in I$ and $i' \in I'$. , then $H(\Gamma''_{ii'}) = H(\Gamma_i) + H(\Gamma'_{i'})$.*

(d) *$H(q''|p'') = H(q|p) + H(q'|p')$*

(e) *$\gamma(\Gamma'') = \gamma(\Gamma) + \gamma(\Gamma')$.*

Proof. Since Σ'' is a channel system

$$t''_{ii',jj'} = p''_{ii'}\Gamma''_{ii',jj'} = p''_{ii'}\Gamma_{ij}\Gamma'_{i'j'} \tag{1}$$

Since Γ' is a channel, $\sum_{i \in I}\Gamma'_{i'j'} = 1$. Also t and p are marginal distributions of t'' and p''. So summing (1) over all $i' \in I'$, $j' \in J'$ gives

$$t_{ij} = \sum_{i' \in I', j \in J'} t''_{ii',jj'} = \left(\sum_{i' \in I'}\left(p''_{ii'}\sum_{j' \in J'}\Gamma'_{i'j'}\right)\right)\Gamma_{ij} = \left(\sum_{i' \in I'} p''_{ii'}\right)\Gamma_{ij} = p_i\Gamma_{ij}$$

So t is the joint distribution of p and Γ.

Since Σ'' is a channel system, q'' is the marginal distribution of t'' on $J'' = J \times J'$. Also q is the marginal distribution of q'' on J. It follows that q is the marginal distribution of t'' on J. Also t is the marginal distribution of t'' on $I \times J$. Hence the marginal distribution of t on J is the marginal distribution of t'' on J and so equal to q. Thus Σ is a channel system. By symmetry also Σ' is a channel system.

(b) Suppose that p and p' are independent with respect to p''. Then $p''_{ii'} = p_ip'_{i'}$ and

$$q_{jj'} = \sum_{i\in I, i'\in I'} p''_{ii'}\Gamma''_{ii',jj'} = \sum_{i\in I, i'\in I'} p_i p'_{i'}\Gamma'_{ij}\Gamma''_{jj'} = \left(\sum_{i\in I} p_i \Gamma_{ij}\right)\left(\sum_{i'\in I'} p'_{i'}\Gamma'_{i'j'}\right) = q_j q'_{j'}$$

Hence $q'' = q \otimes q'$ and q and q' are independent with respect to q''.

(c) Let $i \in I$ and $i' \in I'$. By (10.4.3)(a),

$$\Gamma_{ii'} = \Gamma_i \otimes \Gamma'_{i'}$$

and so (c) follows from 4.1.9.

(d) By 9.4.6

$$H(q''|p'') = \sum_{i\in I, i'\in I'} p''_{ii'} H(\Gamma_{ii'})$$

and so by (c)

$$H(q''|p'') = \sum_{i\in I}\left(\sum_{i'\in I'} p''_{ii'}\right) H(\Gamma_i) + \sum_{i'\in I'}\left(\sum_{i\in I} p''_{ii'}\right) H(\Gamma'_{i'}) = \sum_{i\in I} p_i H(\Gamma_i) + \sum_{i'\in I'} p'_{i'} H(\Gamma'_{i'})$$

Two more applications of 9.4.6 give (d).

(e) Let $\mathcal{P}$ be the set of probability distributions on $I \times I'$. Recall that

$$\gamma(\Gamma'') = \max_{p''\in\mathcal{P}} f_{\Gamma''}(p'')$$

and

$$(2) \qquad f_{\Gamma''}(p'') = I(p'', q'') = H(p'') + H(q'') - H(t'') = H(q'') - H(q''|p'').$$

Since q and q' are the marginal distributions of q'' 4.1.9 gives

$$(3) \qquad H(q'') \le H(q) + H(q')$$

with equality if q and q' are independent with respect to q'', and so by (b) with equality if p and p' are independent with respect to p''.

Thus

$$\begin{aligned}
f_{\Gamma''}(p'') &\overset{(2)}{=} H(q'') - (H(q''|p'')) \\
&\overset{(3)}{\leq} H(q) + H(q') - (H(t'') - H(q'')) \\
&\overset{(c)}{=} H(q) + H(q') - \big(H(q|p)\big) + \big(H(q'|p')\big) \\
&= \Big((H(q) - (H(q|p)))\Big) + \Big((H(q') - (H(q'|p')))\Big) \\
&\overset{(2)}{=} f_\Gamma(p) + f_{\Gamma'}(p')
\end{aligned} \tag{4}$$

with equality in the independent case. Since $f_\Gamma(p) \leq \gamma(\Gamma)$ and $f_{\Gamma'}(p') \leq \gamma(\Gamma')$ we conclude that

$$f_{\Gamma''}(p'') \leq \gamma(\Gamma) + \gamma(\Gamma').$$

Since this holds for all $p \in \mathcal{P}$,

$$\gamma(\Gamma'') = \max_{p'' \in \mathcal{P}} f_\Gamma(p'') \leq \gamma(\Gamma) + \gamma(\Gamma'). \tag{5}$$

Let $p_{\max}$ be a probability distribution on I with $f_\Gamma(p_{\max}) = \gamma(\Gamma)$, and $p'_{\max}$ be a probability distribution on I' with $f_{\Gamma'}(p'_{\max}) = \gamma(\Gamma')$. Put $p''_{\max} = p_{\max} \otimes p'_{\max}$. Then $p_{\max}$ and $p'_{\max}$ are independent with respect to $p''_{\max}$ and so by (4)

$$f_{\Gamma''}(p''_{\max}) = f_\Gamma(p_{\max}) + f_{\Gamma'}(p'_{\max}) = \gamma(\Gamma) + \gamma(\Gamma').$$

Since $\gamma(\Gamma'') \geq f_{\Gamma''}(p''_{\max})$, this gives

$$\gamma(\Gamma'') \geq \gamma(\Gamma) + \gamma(\Gamma'.)$$

Together with (5) this gives (e). □

Corollary 10.4.5. *Let n be a positive integer.*

(a) *Let Γ be a channel. Then $\gamma\big(\Gamma^{\otimes n}\big) = n\gamma(\Gamma)$.*

(b) $\gamma\Big(\mathrm{BSC}^{\otimes n}(e)\Big) = n\Big(1 - H\big((e, 1-e)\big)\Big)$.

Proof. (a) This clearly holds for $n = 1$. Suppose its true for n. Then

$$\gamma\big(\Gamma^{\otimes(n+1)}\big) = \gamma\big(\Gamma^{\otimes n} \otimes \Gamma\big) = \gamma\big(\Gamma^{\otimes n}\big) + \gamma(\Gamma) = n\gamma(\Gamma) + \gamma(\Gamma) = (n+1)\gamma(\Gamma)$$

Thus (a) also holds for $n+1$ and thus for all n,
(b) Since $H(\mathrm{BSC}(e)) = 1 - H\big((e, 1-e)\big)$, (b) follows from (a). □

10.5 Coding at a given rate

Definition 10.5.1. *Let I and J be alphabets with $|J| > 1$. Then the information rate of I relative to J is $\log_{|J|}|I|$.*

Note that $\log_{|J|}|I| = \frac{\log|I|}{\log|J|}$ and $\log_{|J^n|}|I| = \frac{\log_{|J|}|I|}{n}$.

Theorem 10.5.2 (Noisy Coding Theorem I)**.** *Let $\rho > 0$ be a real number and let Γ be an $I \times J$ channel. Let $(n_i)_{i=1}^{\infty}$ be an increasing sequence of integers, $(C_i)_{i=1}^{\infty}$ a sequence of sets $C_i \subseteq I^{n_i}$, and $(\sigma_i)_{i=1}^{\infty}$ a sequence of $C_i \times J^{n_i}$-decision rules σ_i*

Let M_i be the probability of a mistake for the decision system determined by the channel $\Gamma^{\otimes n_i}|_{C_i \times J^{n_i}}$, the equal probability distribution on C_i and the decision rule σ_i. Suppose that

(i) $\frac{\log|C_i|}{n_i} \geq \rho$ *for all $i \in \mathbb{Z}^+$, and*

(ii) $\lim_{i\to\infty} M_i = 0$.

Then $\rho \leq \gamma(\Gamma)$.

Proof. Let Γ_i be the channel $\Gamma^{\otimes n_i}|_{C_i \times J^{n_i}}$. By 10.3.1

$$M_i > 1 - \frac{\gamma(\Gamma_i) + 1}{\log(|C_i|)} \tag{1}$$

By C.1.1 in the Appendix, $\gamma(\Gamma_i) \leq \gamma(\Gamma^{\otimes n_i})$ and by 10.4.5 $\gamma(\Gamma^{n_i}) = n_i\gamma(\Gamma)$. Thus

$$\gamma(\Gamma_i) \leq n_i\gamma(\Gamma). \tag{2}$$

By (i)

$$\log|C_i| \geq \rho n_i. \tag{3}$$

Substituting (2) and (3) into (1) gives

$$M_i > 1 - \frac{n_i\gamma(\Gamma) + 1}{n_i\rho} = 1 - \frac{\gamma(\Gamma)}{\rho} - \frac{1}{n_i\rho}.$$

Thus

$$\frac{\gamma(\Gamma)}{\rho} > 1 + M_i - \frac{1}{n_i\rho}$$

By (ii) we have $\lim_{i\to\infty} M_i = 0$. Since $(n_i)_{i=1}^{\infty}$ is increasing, $\lim_{i\to\infty}\frac{1}{n_i\rho} = 0$. Hence $\frac{\gamma(\Gamma)}{\rho} \geq 1$ and so $\rho \leq \gamma(\Gamma)$. □

Theorem 10.5.3 (Noisy Coding Theorem II)**.** *Let $\rho > 0$ and Γ an $I \times J$ channel. Suppose that $\rho < \gamma(\Gamma)$. Then there exists an increasing sequence of positive integers $(n_i)_{i=1}^{\infty}$ and a sequence $(C_i)_{i=1}^{\infty}$ of sets $C_i \subseteq I^{n_i}$ such that*

(i) $\frac{\log |C_i|}{n_i} \geq \rho$ *for all* $i \in \mathbb{Z}^+$*,*

(ii) *If $(p_i)_{i=1}^{\infty}$ is a sequence of probability distributions p_i on C_i, then there exists a sequence $(\sigma_i)_{i=1}^{\infty}$ of $C_i \times J^{n_i}$ decision rules σ_i such that*

$$\lim_{i \to \infty} M_i = 0$$

where M_i is probability of a mistake for the decision system determined by the channel $\Gamma^{\otimes n_i}|_{C_i \times J^{n_i}}$, the probability distribution p_i and the decision rule σ_i.

Proof. Beyond the scope of these lecture notes. □

10.6 Minimum Distance Decision Rule

Lemma 10.6.1. *Let $\Gamma = \mathrm{BSC}(e)$, $n \in \mathbb{Z}^+$ and $x, y \in \mathbb{B}^n$. Put $d = \mathrm{d}(x, y)$. Then*

$$\Gamma_{xy}^{\otimes n} = e^{\mathrm{d}}(1-e)^{n-d}$$

Proof. We have

$$\Gamma_{xy}^{\otimes n} = \prod_{k=1}^{n} \Gamma_{x_k y_k}$$

Observe that $\Gamma_{x_k y_k} = e$ if $x_k \neq y_k$ and $\Gamma_{x_k y_k} = 1 - e$ if $x_k = y_k$. Note that there are d k's with $x_k \neq y_k$ and $n - d$ k's with $x_k = y_k$. So $\Gamma_{xy}^{\otimes n} = e^{\mathrm{d}}(1-e)^{n-d}$ □

Lemma 10.6.2. *Let $0 < e < \frac{1}{2}$, $C \subseteq \mathbb{B}^n$ and σ a decision rule for C. Then σ is a minimal distance rule if and only if σ is a maximum likelihood rule with respect to* $\mathrm{BSC}^{\otimes n}(e)$*.*

Proof. Let $z \in \mathbb{B}^n$, $a \in \mathbb{B}^n$ and put $a' = \sigma(z)$. Let $d = \mathrm{d}(a, z)$ and $d' = \mathrm{d}(a', z)$ and $f = \frac{1-e}{e}$. Since $e < \frac{1}{2}$, $1 - e > \frac{1}{2} > e$ and so $f > 1$. We compute

$$\frac{\Gamma_{a'z}}{\Gamma_{az}} = \frac{e^{d'}(1-e)^{n-d'}}{e^{\mathrm{d}}(1-e)^{n-d}} = \left(\frac{1-e}{e}\right)^{d-d'} = f^{d-d'}$$

It follow that

$$\Gamma_{az} \leq \Gamma_{a'z} \iff d \geq d'$$

So $\Gamma_{a'z}$ is maximal if and only if d' is minimal. □

Lemma 10.6.3. *Let $C \subseteq \mathbb{B}^n$. Let Σ be a decision system with channel $\mathrm{BSC}^{\otimes n}(e)$ and an r-error-correcting decision rule.*

(a) $\mathrm{d}(a,z) \geq r+1$ *for any mistake* (a,z).

(b) $\Gamma_{az} \leq e^{r+1}$ *for any* $a \in C, z \in F(a)$.

(c) $M_a \leq |F(a)|e^{r+1}$ *for any* $a \in C$.

(d) $M \leq \left(\sum_{a \in C} p_a |F(a)|\right) e^{r+1}$.

Proof. (a) Let $a \in C$ and $z \in \mathbb{B}^n$ with $\mathrm{d}(a,z) \leq r$. Since σ is r-correcting, $\sigma(z) = a$ and so (a,z) is not a mistake.

(b) Since $z \in F(a)$, (a,z) is a mistake. Put $d = \mathrm{d}(a,z)$, then by (a) $d \geq r+1$. Hence

$$\Gamma_{az} = e^d(1-e)^{n-d} = e^{r+1}e^{d-(r+1)}(1-e)^{n-d} \leq e^{r+1}$$

(c) $M_a = \sum_{z \in F(a)} \Gamma_{az} \leq \sum_{z \in F(a)} e^{r+1} = |F(a)|e^{r+1}$.

(d) $M = \sum_{a \in C} p_a M_a \leq \sum_{a \in C} p_a |F(a)| e^{r+1} = \Big(\sum_{a \in C} p_a |F(a)|\Big) e^{r+1}$. □

Example 10.6.4. Suppose $C = \{000, 111\}$. Determine a minimal distance rule σ for C. Compute $F(c)$, M_c and M for the decision system determined by $\mathrm{BSC}(e)$, $(p, 1-p)$ and σ.

We have

$$\sigma(z) = \begin{cases} 000 & \text{if at least two coordinates are zero} \\ 111 & \text{if at most one coordinate is zero} \end{cases}$$

Hence

$$F_{000} = \{011, 101, 110, 111\} \quad \text{and } F_{111} = \{000, 001, 010, 100\}.$$

So

$$\begin{aligned} M_{000} &= \Gamma_{000,011} + \Gamma_{000,101} + \Gamma_{000,110} + \Gamma_{000,111} \\ &= e^2(1-e) + e^2(1-e) + e^2(1-e) + e^3 \\ &= e^2(3(1-e) + e) \\ &= e^2(3-2e) \end{aligned}$$

By symmetry, also $M_{111} = e^2(3-2e)$ and so

$$M = pe^2(3-2e) + (1-p)e^2(3-2e) = e^2(3-2e).$$

Note that each of the four summand in M_{000} is at most e^2. So $M_{000} \leq 4e^2$ and also $M \leq 4e^2$.

Chapter 11

Cryptography in theory and practice

11.1 Encryption in terms of a channel

Definition 11.1.1. *Let $(\mathcal{M}, \mathcal{C}, (E_k)_{k\in\mathcal{K}}, (D_l)_{l\in\mathcal{K}})$ be a cryptosystem and p and r probabilty distribution on $\mathcal{M}$ and $\mathcal{K}$, respectively. Define*

$$\begin{array}{rccc} u: & \mathcal{M}\times\mathcal{K}\times\mathcal{C} & \to & [0,1] \\ & (m,k,c) & \to & \begin{cases} p_m r_k & \textit{if } c = E_k(m) \\ 0 & \textit{if } c \neq E_k(m) \end{cases} \end{array}$$

t, q, and s are defined to be the marginal distribution of u on $\mathcal{M}\times\mathcal{C}$, $\mathcal{C}$ and $\mathcal{K}\times\mathcal{C}$ respectively.

Define the $\mathcal{M}\times\mathcal{C}$-matrix Γ by

$$\Gamma_{mc} = \sum_{\substack{k\in\mathcal{K} \\ E_k(m)=c}} r_k$$

Γ is called the encryption channel for the cryptosystem with respect to r.

We interpreted u_{mkc} as the probability $\mathrm{Prob}(m,k,c)$ that the plain text message m was encrypted via the key k and the cipher text c was obtained.

Lemma 11.1.2. *With the notation as in 11.1.1*

(a) *u is a probability distribution.*

(b) *p, r and $p\otimes r$ are the marginal distribution of u on $\mathcal{M}, \mathcal{K}$ and $\mathcal{M}\times\mathcal{K}$, respectively.*

(c) *$t_{mc} = p_m\Gamma_{mc}$ for all $m\in\mathcal{M}$ and $c\in\mathcal{C}$.*

(d) *E_k is 1-1 for all $k\in\mathcal{K}$.*

(e)

$$q_c = \sum_{m\in\mathcal{M}} p_m\Gamma_{mc} = \sum_{\substack{(m,k)\in\mathcal{M}\times\mathcal{K}\\ E_k(m)=c}} p_m r_k.$$

Proof. Let $m\in\mathcal{M}$, $k\in\mathcal{K}$ and $c\in\mathcal{C}$.

(a) and (b) Clearly $u_{mkc}\in[0,1]$. Given m and k. Define $c^* := E_k(m)$. Then is the unique element of $\mathcal{C}$ with $u_{mkc^*}\neq 0$. Thus

$$\sum_{c\in\mathcal{C}} u_{mkc} = u_{mkc^*} = p_m r_k.$$

So $p\otimes r$ is the marginal distribution of u on $\mathcal{M}\times\mathcal{K}$. Since $p\otimes r$ is a probability distribution, we conclude from 4.1.5 that also u is a probablity distribution. Since p and r are the marginal distributions of $p\otimes r$ on $\mathcal{M}$ and $\mathcal{K}$, respectively, they are also the marginal distributions of u on $\mathcal{M}$ and $\mathcal{K}$, see B.1.1 in the appendix.

(c)

Since $u_{mkc}=0$ for $c\neq E_k(m)$ we have

$$t_{mc} = \sum_{k\in\mathcal{K}} u_{mkc} = \sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} u_{mkc} = \sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} p_m r_k = p_m \sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k = p_m\Gamma_{mc}$$

(d) Let $k\in\mathcal{K}$ and $m_1,m_2\in\mathcal{M}$ with $E_k(m_1)=E_k(m_2)$. By definition of a cryptosystem there exists $k^*\in\mathcal{K}$ with $D_{k^*}\circ E_k = \mathrm{id}_{\mathcal{M}}$. Thus

$$m_1 = D_{k^*}(E_k(m_1)) = D_{k^*}(E_k(m_2)) = m_2$$

and so E_k is 1-1.

(e) Since q is the marginal distributions of t,

$$q_c = \sum_{m\in\mathcal{M}} t_{mc} \overset{\text{(c)}}{=} \sum_{m\in\mathcal{M}} p_m\Gamma_{mc} = \sum_{m\in\mathcal{M}} p_m\Big(\sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k\Big) = \sum_{\substack{(m,k)\in\mathcal{M}\times\mathcal{K}\\ E_k(m)=c}} p_m r_k.$$

□

Definition 11.1.3. *A cryptosytem is said to have* perfect secrecy *with respect to the probability distribution r on $\mathcal{K}$, if, for all probability distributions p on $\mathcal{M}$, p and q are independent with respect to t. Here t, and q are as define in 11.1.1.*

11.2 Perfect secrecy

Theorem 11.2.1. *Given a cryptosystem and a probability distribution r on $\mathcal{K}$. Then the following are equivalent:*

(a) *The cryptosystem has perfect secrecy with respect to r.*

(b) *There exists a positive probability distribution p on $\mathcal{M}$ such that p and q are independent with respect to t.*

(c) *There exists a positive probability distribution p on $\mathcal{M}$ such that $\Gamma_{mc} = q_c$ for all $m \in \mathcal{M}$ and all $c \in \mathcal{C}$,*

(d) *Each column of Γ is constant, that is there exist a $\mathcal{C}$-tuple $(l_c)_{c\in\mathcal{C}}$ with $\Gamma_{mc} = l_c$ for all $m \in \Gamma, c \in \mathcal{C}$.*

(e) $\Gamma_{mc} = q_c$ *for all probability distributions p on $\mathcal{M}$, all $m \in \mathcal{M}$ and all $c \in \mathcal{C}$:*

Proof. (a) $\Longrightarrow$ (b): Just choose p to be the equal probability distribution on $\mathcal{M}$.

(b) $\Longrightarrow$ (c): Since p and q independent with respect to t we have $t_{mc} = p_m q_c$ for all $m \in \mathcal{M}$ and $c \in \mathcal{C}$. Since Γ is a channel associated to t, $t_{mc} = p_m \Gamma_{mc}$. Thus $p_m \Gamma_{mc} = p_m q_c$ for all $m \in \mathcal{M}$ and all $c \in \mathcal{C}$. Since p is positive $p_m \neq 0$ and so $\Gamma_{mc} = q_c$.

(c) $\Longrightarrow$ (d): All entries in column c of Γ are equal to q_c.

(d) $\Longrightarrow$ (e): Let $c \in \mathcal{C}$. Since column c of Γ is constant there exists $l_c \in [0,1]$ with $\Gamma_{mc} = l_c$ for all $m \in \mathcal{M}$. Let p be any probability distribution on $\mathcal{M}$. Then

$$q_c = \sum_{m\in\mathcal{M}} p_m \Gamma_{mc} = \sum_{m\in\mathcal{M}} p_m l_c = \left(\sum_{m\in\mathcal{M}} p_m\right) l_c = 1 l_c = l_c$$

and so (e) holds.

(e) $\Longrightarrow$ (a): Let p be a probability distribution on $\mathcal{M}$. Then $t_{mc} = p_m \Gamma_{mc} = p_m q_c$ and so p and q are independent with respect to t. Thus (a) holds. □

Corollary 11.2.2. *If a cryptosystem has perfect secrecy (with respect to some probability distribution on $\mathcal{K}$), then the numbers of keys is greater or equal to the number of plaintext messages.*

Proof. Fix $c \in \mathcal{C}$ with $q_c \neq 0$. Then by 11.2.1 $\Gamma_{mc} = q_c > 0$ for all $m \in \mathcal{M}$. By 11.1.2

$$\sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k = \Gamma_{mc} > 0$$

and so for all $m \in \mathcal{M}$ there exists $k_m \in \mathcal{K}$ with $E_{k_m}(m) = c$. Suppose $k := k_m = k_{\tilde{m}}$ for some $m, \tilde{m} \in \mathcal{M}$. Then $E_k(m) = c = E_k(\tilde{m})$ and since E_k is 1-1, $m = \tilde{m}$. So the map $m \to k_m$ is 1-1 and thus $|\mathcal{M}| \leq |\mathcal{K}|$. □

11.3 The one-time pad

Definition 11.3.1. *Let $\mathbb{F}$ be a finite field (or let $(\mathbb{F},+)$ be finite group) and let $n \in \mathbb{Z}^+$. For $k \in \mathbb{F}^n$ define*

$$E_k = D_k : \mathbb{F}^n \to \mathbb{F}^n, m \to m + k$$

Then $\Omega(\mathbb{F}^n) = \big(\mathbb{F}^n, \mathbb{F}^n, (E_k)_{k\in\mathbb{F}^n}, (D_k)_{k\in\mathbb{F}^n}\big)$ is called the one-time pad determined by $\mathbb{F}^n$.

Lemma 11.3.2 (One-Time Pad)**.** *Any one-time pad is a cryptosystem and has perfect secrecy with respect to the equal-probability distribution r on the set of keys.*

Proof. Given a one-time pad $\Omega(\mathbb{F}^n)$. Since $(m+k)+(-k) = m+(k+(-k)) = m+0 = m$, $D_{-k} \circ E_k = \mathrm{id}_{\mathbb{F}^n}$. Thus the one-time pad is a cryptosystem.

Set $e := \frac{1}{|\mathbb{F}^n|}$. Then $r_k = e$ for all $k \in \mathcal{K}$.

Let $m \in \mathcal{M}$ and $c \in \mathcal{C}$. By definition of Γ and E_k:

$$\Gamma_{mc} = \sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k = \sum_{\substack{k\in\mathcal{K}\\ m+k=c}} r_k$$

For any $m \in \mathcal{M}$ and $c \in \mathcal{C}$ there exists a unique $k \in \mathcal{K}$ with $m + k = c$ namely $k = -m + c$. Thus $\Gamma_{mc} = r_{-m+c} = e$. Hence the columns of Γ are constant and so by 11.2.1 the one-time pad has perfect secrecy.

□

11.4 Iterative methods

Definition 11.4.1. *Let $\mathbb{F}$ be a finite field or $(\mathbb{F},+)$ a finite group. Let $n, r \in \mathbb{Z}^+$, K an alphabet and $F : K \times \mathbb{F}^n \to \mathbb{F}^n$ a function. Put $\mathcal{M} = \mathcal{C} = \mathbb{F}^n \times \mathbb{F}^n$ and $\mathcal{K} = K^r$. For $(X_0, X_1) \in \mathcal{M}$ and $k = (k_1, \dots, k_r) \in \mathcal{K}$ define $X_{i+1}, 1 \le i \le r$ inductively by*

$$X_{i+1} = X_{i-1} + F(k_i, X_i)$$

Define

$$E_k : \mathcal{M} \to \mathcal{C}, (X_0, X_1) \to (X_{r+1}, X_r).$$

For $(Y_0, Y_1) \in \mathcal{C}$ and $k = (k_1, \dots, k_r) \in \mathcal{K}$ define $Y_{i+1}, 1 \le i \le r$ inductively by

$$Y_{i+1} = Y_{i-1} - F(k_i, Y_i), 1 \le i \le r$$

Define

$$D_k : \mathcal{C} \to \mathcal{M}, (Y_0, Y_1) \to (Y_{r+1}, Y_r).$$

Put $\Omega(\mathbb{F}^n, K, r, F) = \big(\mathbb{F}^n \times \mathbb{F}^n, \mathbb{F}^n \times \mathbb{F}^n, (E_k)_{k\in K^r}, (D_k)_{k\in K^r}\big)$. Then $\Omega(\mathbb{F}^n, K, r, F)$ is called the Feistel system determined by $\mathbb{F}^n, K, r$ and F.

Lemma 11.4.2. *Any Feistel system is a cryptosystem.*

Proof. Let $k = (k_1, k_2, \ldots, k_r) \in \mathcal{K}$. Let $(X_0, X_1) \in \mathcal{M}$ and define X_i as above. Put $(Y_0, Y_1) = E_k(X_0, X_1) = (X_{r+1}, X_r)$.

Define $Y_i, 0 \le i \le r+1$ as above, but with respect to the key $k^* = (k_r, k_{r-1}, \ldots, k_1)$. Note that $k_i^* = k_{r+1-i}$. For $0 \le i \le r+1$, consider the statement

$$P(i): \qquad Y_i = X_{r+1-i}$$

Note $P(0)$ and $P(1)$ hold by definition of Y_0 and Y_1. Suppose that $P(i-1)$ and $P(i)$ hold. We will show that also $P(i+1)$ hold:

$$\begin{aligned}
Y_{i+1} &\overset{\text{def } Y_{i+1}}{=} Y_{i-1} - F(k_i^*, Y_i) \overset{P(i-1), P(i)}{=} X_{(r+1)-(i-1)} - F(k_{r+1-i}, X_{r+1-i}) \\
&= X_{(r+1-i)+1} - F(k_{r+1-i}, X_{r+1-i}) \overset{\text{def } X_{(r+1-i)+1}}{=} X_{(r+1-i)-1} + F(k_{r+1-i}, X_{r+1-i}) - F(k_{r+1-i}, X_{r+1-i}) \\
&= X_{(r+1)-(i+1)}
\end{aligned}$$

Hence $P(i)$ holds for all $0 \le i \le r+1$ and so

$$D_{k^*}(Y_0, Y_1) = (Y_{r+1}, Y_r) = (X_0, X_1)$$

Thus $D_{k^*} \circ E_k = \mathrm{id}_{\mathcal{M}}$ and the Feistel system is indeed a cryptosystem. □

Example 11.4.3. Consider the Feistel system with $\mathbb{F}^n = \mathbb{F}_2^3$, $K = \mathbb{F}_2^3$, $r = 3$ and

$$F : F_2^3 \times \mathbb{F}_2^3 \to F_2^3,\ (\alpha\beta\gamma, xyz) \to (\alpha x + yz, \beta y + xz, \gamma z + xy)$$

Compute $E_k(m)$ for $k = (100, 101, 001)$ and $m = (101, 110)$. Verify that $D_{k^*}(E_k(m)) = m$.

i	k_i	X_i	$F(k_i, X_i)$
0	–	101	–
1	100	110	$(1\cdot 1 + 1\cdot 0, 0\cdot 1 + 1\cdot 0, 0\cdot 0 + 1\cdot 1) = 101$
2	101	000	000
3	001	110	$(0\cdot 1 + 1\cdot 0, 0\cdot 1 + 1\cdot 0, 1\cdot 0 + 1\cdot 1) = 001$
4	–	001	

So $E_k(m) = (001, 110)$. To decrypt $(001, 110)$ we use the key $k^* = (001, 101, 100)$.

i	k_i^*	Y_i	$F(k_i^*, Y_i)$
0	–	001	–
1	001	110	$(0 \cdot 1 + 1 \cdot 0, 0 \cdot 1 + 1 \cdot 0, 1 \cdot 0 + 1 \cdot 1) = 001$
2	101	000	000
3	100	110	$(1 \cdot 1 + 1 \cdot 0, 0 \cdot 1 + 1 \cdot 0, 0 \cdot 0 + 1 \cdot 1) = 101$
4	–	101	

So $D_{k^*}(E_k(m)) = (101, 110) = m$.

11.5 The Double-Locking Procedure

Two cryptosystems Ω and $\tilde{\Omega}$ are called compatible if $\mathcal{M} = \mathcal{C} = \tilde{\mathcal{M}} = \tilde{\mathcal{C}}$.

Given compatible cryptosystems Ω and $\tilde{\Omega}$, keys k, k^* in Ω and keys $\tilde{k}, \tilde{k}^*$ in $\tilde{\Omega}$ with $D_{k^*} \circ E_k = \mathrm{id}_{\mathcal{M}}$ and $\tilde{D}_{\tilde{k}^*} \circ \tilde{E}_{\tilde{k}} = \mathrm{id}_{\mathcal{M}}$

Consider the following procedure to send a message $m_0 \in \mathcal{M}$ from person X to person $\tilde{X}$. ○○

(•) X computes $m_1 = E_k(m_0)$ and sends m_1 to $\tilde{X}$. ●○

(•) $\tilde{X}$ computes $m_2 = \tilde{E}_{\tilde{k}}(m_1)$ and sends m_2 to X. ●●

(•) X computes $m_3 = D_{k^*}(\tilde{m})$ and sends $\tilde{X}$. ○●

(•) $\tilde{X}$ computes $m_4 = \tilde{D}_{\tilde{k}^*}(m_3)$. ○○

$$\begin{tikzcd} m_0 \arrow[r, "E_k"] & m_1 \arrow[r, "\tilde{E}_{\tilde{k}}"] & m_2 \arrow[r, "D_{k^*}"] & m_3 \arrow[r, "\tilde{D}_{\tilde{k}^*}"] & m_4 \end{tikzcd}$$

Is $m_4 = m_0$?

Consider the following example $\mathcal{M} = \mathcal{C} = \{1, 2, 3\}$,

$$E_k = D_{k^*} : \quad \begin{array}{ccc} 1 & 2 & 3 \\ \hline 1 & 3 & 2 \end{array} \qquad\qquad \tilde{E}_{\tilde{k}} = \tilde{D}_{\tilde{k}^*} : \quad \begin{array}{ccc} 1 & 2 & 3 \\ \hline 2 & 1 & 3 \end{array}$$

and $m_0 = 1$

$$\begin{tikzcd} 1 \arrow[r, "E_k"] & 1 \arrow[r, "\tilde{E}_{\tilde{k}}"] & 2 \arrow[r, "D_{k^*}"] & 3 \arrow[r, "\tilde{D}_{\tilde{k}^*}"] & 3 \end{tikzcd}$$

So $m_4 \neq m_0$.

In general

$$m_4 = \big(\tilde{D}_{\tilde{k}^*} \circ D_{k^*} \circ \tilde{E}_{\tilde{k}} \circ E_k\big)(m_0)$$

If D_{k^*} commutes with $\tilde{E}_{\tilde{k}}$, that is $D_{k^*} \circ \tilde{E}_{\tilde{k}} = \tilde{E}_{\tilde{k}} \circ D_{k^*}$ then

$$\tilde{D}_{\tilde{k}^*} \circ D_{k^*} \circ \tilde{E}_{\tilde{k}} \circ E_k = \big(\tilde{D}_{\tilde{k}^*} \circ \tilde{E}_{\tilde{k}}\big) \circ \big(D_{k^*} \circ E_k\big) = \mathrm{id}_{\mathcal{M}} \circ \mathrm{id}_{\mathcal{M}} = \mathrm{id}_{\mathcal{M}}$$

and procedure works.

Since addition in finite field is commutative, one-time pads provide examples where D_{k^*} commutes with $\tilde{E}_{\tilde{k}}$.

Example 11.5.1. Suppose Ω and $\tilde{\Omega}$ both are the one-time pad determined by $\mathbb{F}_2^4$. Given the following public information:

$$\begin{tikzcd} m_0 \arrow[r, "E_k"] & 1101 \arrow[r, "\tilde{E}_{\tilde{k}}"] & 0110 \arrow[r, "D_{k^*}"] & 1100 \arrow[r, "\tilde{D}_{\tilde{k}^*}"] & m_4 \end{tikzcd}$$

What is m_0?

Since $0110 + k^* = 1100$ and so $k^* = \begin{smallmatrix} & 1100 \\ - & 0110 \end{smallmatrix} = 1010$. So $m_0 = D_k^*(m_1) = \begin{smallmatrix} & 1101 \\ + & 1010 \end{smallmatrix} = 0111$.

So one-time pads should not be used for the double-locking procedure.

Lemma 11.5.2. *Let Ω and $\tilde{\Omega}$ be compatible cryptosystem. Let β be an encryption function in $\tilde{\Omega}$ and let γ and γ' be decryption functions in Ω which commute with β, that is*

$$\gamma \circ \beta = \beta \circ \gamma, \qquad \textit{and} \qquad \gamma' \circ \beta = \beta \circ \gamma'$$

Let $m_1 \in \mathcal{M}$ and put $m_2 = \beta(m_1)$. Then

$$\gamma(m_2) = \gamma'(m_2) \qquad \Longrightarrow \qquad \gamma(m_1) = \gamma'(m_2)$$

Proof.

$$\beta\big(\gamma(m_1)\big) = (\beta \circ \gamma)(m_1) = (\gamma \circ \beta)(m_1) = \gamma\big(\beta(m_1)\big) = \gamma(m_2)$$

By symmetry, $\beta\big(\gamma'(m_1)\big) = \gamma'(m_2)$. So if $\gamma(m_2) = \gamma'(m_2)$ we conclude that

$$\beta\big(\gamma(m_1)\big) = \beta\big(\gamma'(m_1)\big)$$

Since encryption functions are 1-1, we get $\gamma(m_1) = \gamma'(m_1)$. □

The lemma shows that the double locking procedure is very vulnerable: Anybody who intercepts the message m_1, m_2 and m_3 and is able to find a decryption function D_l in Ω with $D_l(m_2) = m_3$ can compute m_0, namely $m_0 = D_l(m_1)$.

Appendix A

Rings and Field

A.1 Basic Properties of Rings and Fields

Definition A.1.1. *A* ring *is a triple* $(R,+,\cdot)$ *such that*

(i) *R is a set;*

(ii) *$+$ is a function (called* ring addition*), $R\times R$ is a subset of the domain of $+$ and for $(a,b)\in R\times R$, $a+b$ denotes the image of (a,b) under $+$;*

(iii) *$\cdot$ is a function (*called *ring multiplication), $R\times R$ is a subset of the domain of $\cdot$ and for $(a,b)\in R\times R$, $a\cdot b$ (and also ab) denotes the image of (a,b) under $\cdot$;*

and such that the following eight axioms hold:

(Ax 1) *$a+b\in R$ for all $a,b\in R$;* *[closure for addition]*

(Ax 2) *$a+(b+c)=(a+b)+c$ for all $a,b,c\in R$;* *[associative addition]*

(Ax 3) *$a+b=b+a$ for all $a,b\in R$.* *[commutative addition]*

(Ax 4) *there exists an element in R, denoted by 0_R and called 'zero R', such that $a+0_R=a=0_R+a$ for all $a\in R$;* *[additive identity]*

(Ax 5) *for each $a\in R$ there exists an element in R, denoted by $-a$ and called 'negative a', such that $a+(-a)=0_R$;* *[additive inverses]*

(Ax 6) *$ab\in R$ for all $a,b\in R$;* *[closure for multiplication]*

(Ax 7) *$a(bc)=(ab)c$ for all $a,b,c\in R$;* *[associative multiplication]*

(Ax 8) $a(b+c) = ab+ac$ *and* $(a+b)c = ac+bc$ *for all* $a, b, c \in R$. *[distributive laws]*

Definition A.1.2. *A ring* $(R, +, \cdot)$ *is called* commutative *if*

(Ax 9) $ab = ba$ *for all* $a, b \in R$. *[commutative multiplication]*

Definition A.1.3. *An element* 1_R *in a ring* $(R, +, \cdot)$ *is called an (multiplicative)* identity *if*

(Ax 10) $1_R \cdot a = a = a \cdot 1_R$ *for all* $a \in R$. *[multiplicative identity]*

Definition A.1.4. *A field is a commutative ring* $(\mathbb{F}, +, \cdot)$ *with identity* $1_F \neq 0_F$ *such that*

(Ax 11) *for each* $a \in R$ *with* $a \neq 0_{\mathbb{F}}$ *there exists an element in* R*, denoted by* a^{-1} *[multiplicative inverses]*

and called ' a inverse ', such that $a \cdot a^{-1} = 1_R = a^{-1} \cdot a$*;*

If $(R, +, \cdot)$ is a ring, we will often just say that R is a ring, assuming that there is no confusion about the underlying addition and multiplication. Also we will usually write 0 for 0_R and 1 for 1_R.

With respect to the usual addition and multiplication:
The real number and the rational numbers are fields. The integers are a commutative ring but not a field. $\mathbb{F}_2$ is a field.

Lemma A.1.5. *Let* R *be ring and* $a, b \in R$*. Define* $a - b = a + (-b)$*.*

(a) $a + 0 = a$.

(b) $(b + a) + (-a) = b$.

(c) *Let* $d \in R$*. Then* $a = b$ *if and only if* $d + a = d + b$ *if and only if* $a + d = b + d$

(d) $x = b - a$ *is the unique element in* R *with* $x + a = b$*.*

(e) $x = -a$ *is the unique element in* R *with* $x + a = 0$*.*

(f) $0a = 0 = a0$.

(g) $(-b)a = -(ba)$.

(h) $-(-a) = a$

(i) $-(a + b) = (-a) + (-b)$.

(j) $-(a - b) = b - a$.

(k) *If* R *has an identity,* $(-1)a = -a$*.*

Proof. (a) $a + 0 = 0 + a = a$.

(b) $(b + a) + (-a) = b + (a + (-a)) = b + 0 = b$.

(c) If $a = b$, then clearly $d + a = d + b$. If $d + a = d + b$, then since $d + a = a + d$ and $d + b = b + d$, we have $a + d = b + d$.

Suppose that $a + d = b + d$. Adding $(-d)$ to both sides of the equation gives $(a + d) + (-d) = (b + d) + (-d)$ and so by (b), $a = b$.

(d) We have $x + a = b$ if and only if $(x + a) + (-a) = b + (-a)$. bBy (b) and the definition of $b - a$ this holds if and only if $x = b - a$.

(e) Since $-a = 0 + (-a) = 0 - a$, this follows from (d) applied with $b = 0$.

(f) We have $0 + 0a = 0a = (0 + 0)a = 0a + 0a$ and so by (c), $0 = 0a$. A similar argument shows that $a0 = 0$.

(g) $ba + (-b)a = (b + (-b))a = 0a = 0$ and so $-(ba) = (-b)a$ by (e).

(h) By $0 = a + (-a) = (-a) + a$ and so by (e), $a = -(-a)$.

(i) $(a + b) + ((-a) + (-b)) = ((a + b) + (-a)) + (-b) = ((b + a) + (-a)) + (-b) = b + (-b) = 0$ and so $(-a) + (-b) = -(a + b)$ by (e).

(h) $-(a - b) = -(a + (-b)) \overset{(i)}{=} -a + (-(-b)) \overset{(h)}{=} -a + b = b + (-a) = b - a$.

(k) By (g) $-a = -(1a) = (-1)a$. □

A.2 Polynomials

Definition A.2.1. *Let R be ring. Then $R[x]$ is the set of $\mathbb{N}$-tuples $(a_i)_{i\in\mathbb{N}}$ with coefficients in R such that there exists $n \in \mathbb{N}$ with $a_i = 0$ for all $i > n$. We denote such an $\mathbb{N}$-tuple by*

$$a_0 + a_1x + \ldots + a_nx^n.$$

Let $f = \sum_{i=0}^{n} a_ix^i$ and $g = \sum_{i=0}^{m} b_ix^i$ be elements of $R[x]$ define

$$f + g = \sum_{i=0}^{l}(a_i + b_i)x^i,$$

where $l = \max(n, m)$, $a_i = 0$ for $i > n$ and $b_i = 0$ for $i > m$; and

$$fg = \sum_{k=0}^{n+m}\left(\sum_{i=0}^{k} a_ib_{k-i}\right)x^i$$

Define $\deg f = \max\{i \mid a_i \neq 0\}$ *with* $\deg f = -\infty$ *if* $f = 0$.

Lemma A.2.2. *Let R be a ring. Then $R[x]$ is a ring. If R is commutative, so is $R[x]$.*

Proof. Readily verified. □

Lemma A.2.3. *Let $\mathbb{F}$ be a field and $f, g \in \mathbb{F}[x]$. Then*

(a) $\deg(f+g) \le \max(\deg f, \deg g)$.

(b) $\deg fg = \deg f + \deg g$.

(c) *If* $f \neq 0$, *then* $\deg g \le \deg fg$.

Proof. Readily verified. □

Lemma A.2.4. *Let* $\mathbb{F}$ *be a field and* $h \in \mathbb{F}[x]$ *with* $h \neq 0$.

(a) $(\mathbb{F}^h[x], \oplus, \odot)$ *is a commutative ring with identity.*

(b) $(\mathbb{F}[x], \oplus, \odot)$ *fulfills Axioms (1)-(9) of a commutative ring, except for Axiom (4) (that is* $\oplus$ *has not additive identity).*

Proof. Let $e, f, g \in \mathbb{F}[x]$. Recall that $\overline{f}$ denotes the remainder of f when divided by h. By definition of $\oplus$ and $\odot$:

$(*)$ $\quad f \oplus g = \overline{f+g}$ *and* $f \odot g = \overline{fg}$.

By (7.1.9)(b)

$(**)$ $\quad \overline{f} = f$ *for all* $f \in \mathbb{F}^h[x]$.

By 7.1.12

$(***)$ $\quad \overline{f} \oplus \overline{g} = \overline{f} \oplus g = f \oplus g = \overline{f+g}$ *and* $\overline{f} \odot \overline{g} = \overline{f} \odot g = f \odot g = \overline{fg}$.

We now will verify all the conditions on a commutative ring (see A.1.1) Since

$$f \oplus g = \overline{f+g} = \overline{g+f} = g \oplus f,$$

condition (i) holds.

We have

$$e \oplus (f \oplus g) = e \oplus \overline{f+g} = \overline{e + (f+g))} = \overline{(e+f)+g}$$

and

$$(e \oplus f) \oplus g = \overline{e+f} \oplus g = \overline{(e+f)+g}.$$

Thus condition (ii) holds.

We have $0 \oplus f = \overline{0+f} = \overline{f}$ and so for $f \in F^h[x]$, $0 \oplus f = f$. Hence condition (iii) holds.

$$f \oplus -f = \overline{f + (-f)} = \overline{0} = 0$$

and condition (iv) is proved.

Since

$$e \odot (f \odot g) = e \odot \overline{fg} = \overline{e(fg)} = \overline{(ef)g}$$

and

$$(e \odot f) \odot g = \overline{ef} \odot g = \overline{(ef)g}$$

condition (v) is verified. From

$$e \odot (f \oplus g) = e \odot \overline{f+g} = \overline{e(f+g)} = \overline{ef+eg}$$

and

$$(e \odot f) \oplus (e \odot g) = \overline{ef} \oplus \overline{eg} = \overline{ef+eg}$$

we conclude that condition (vi) holds.. A similar argument (or using that $\odot$ is commutative) gives condition (vii).

We have

$$1 \odot f = \overline{1f} = \overline{f}$$

and so $1 \odot f = f$ for all $f \in \mathbb{F}^h[x]$. Thus condition (viii) is verified.

Finally

$$f \odot g = \overline{fg} = \overline{gf} = g \odot f$$

and condition (ix) holds. □

A.3 Irreducible Polynomials

Lemma A.3.1. *Let $\mathbb{F}$ be a field, $f, g, h \in \mathbb{F}[x]$ and suppose that h is irreducible and $h|fg$. Then $h|f$ or $h|g$.*

Proof. Since $h|fg$, the remainder of fg when divided by h is 0. So $\overline{f} \odot \overline{g} = f \odot g = 0$ in $\mathbb{F}^h[x]$. By 7.2.8, $\mathbb{F}^h[x]$ is a field and we conclude that $\overline{f} = 0$ or $\overline{g} = 0$. Hence $h \mid f$ or $h \mid g$. □

Lemma A.3.2. *Let $\mathbb{F}$ be a field and $f, g \in \mathbb{F}[x]$. Suppose f and g are monic,* $\deg f > 0$*, $f|g$ and g is irreducible. Then $f = g$.*

Proof. Since $f|g$, $g = fh$ for some $h \in \mathbb{F}[x]$. Since $\deg f > 0$ and g is irreducible, $\deg h = 0$. Since both f and g are monic, $h = 1$ and so $f = g$. □

Lemma A.3.3. *Let $\mathbb{F}$ be a field, $0 \neq a \in \mathbb{F}$, $r \in \mathbb{N}$ and let $g, f_1, \dots, f_r$ be irreducible monic polynomials in $\mathbb{F}[x]$. If g divides $af_1 \dots f_r$ in $F[x]$, then $r \geq 1$ and there exists $1 \leq i \leq r$ with $g = f_i$.*

Proof. Since $\deg g > 0$ we must have $r \geq 1$. Put $h = af_1 \dots f_{r-1}$. Then g divides hf_r and so by A.3.1, g divides h or f_r. If g divides f_r, then by A.3.2 $h = f_r$. So suppose g divides h. Then $r - 1 > 0$ and by induction on r, $g = f_i$ for some $1 \leq i \leq r - 1$. □

Lemma A.3.4. *Let $\mathbb{F}$ be a field and $0 \neq f \in \mathbb{F}[x]$. Put $a = \mathrm{lead}(f)$.*

(a) *There exists monic irreducible polynomials $f_1, f_2 \dots, f_r \in \mathbb{F}[x]$ with*

$$f = af_1 f_2 \dots f_r$$

Moreover, the $f_1, f_2, \dots f_r$ are unique up to reordering.

(b) *Let $g \in \mathbb{F}[x]$ and put $b = \mathrm{lead}(g)$. Then g divides f in $\mathbb{F}[x]$ if and only if $b \neq 0$ and there exist $\epsilon_i \in \{0,1\}$, $1 \leq i \leq r$ with*

$$g = bf_1^{\epsilon_1} f_2^{\epsilon_2} \dots f_r^{\epsilon_r}$$

Moreover, if g is of this form, then $f = gh$, where

$$h = cf_1^{\delta_1} f_2^{\delta_2} \dots f_r^{\delta_r}$$

with $c = \frac{a}{b}$ and $\delta_i = 1 - \epsilon_i$.

Proof. We prove (a) by induction on $\deg f$.

If $\deg f = 0$, then (a) holds with $r = 0$.

So suppose $\deg f > 0$ and that the lemma holds for all non-zero polynomials of smaller degree.

We will now show the existence of $f_1, \dots f_r$. If f is irreducible, we can choose $r = 1$ and $f_1 = \frac{1}{a} f$. Suppose f is not irreducible. Then $f = gh$ with $g, h \in \mathbb{F}[x]$ and $\deg g \neq 0 \neq \deg h$. Then $\deg g < \deg f$ and $\deg h < \deg f$. Hence by induction

$$g = bg_1 g_2 \dots g_s \text{ and } h = ch_1 \dots h_t$$

where $b = \mathrm{lead}(g)$, $c = \mathrm{lead}(h)$ and $g_1, \dots, g_s$, $h_1 \dots, h_t$ are monic irreducible polynomials. Since $a = bc$ we can choose $r = s + t$ and

$$f_1 = g_1, \dots, f_s = g_s, f_{s+1} = h_1, \dots, f_{s+t} = h_t$$

To prove the existence suppose that

$$f = af_1 \dots f_r = ag_1 \dots g_s$$

for some monic irreducible polynomials $f_1, \dots, f_r, g_1, \dots g_s$.

Then $f_1 | f = ag_1 \dots g_s$ and so A.3.3 show that $f_1 = g_i$ for some $1 \leq i \leq r$. Reordering the $g_i' s$ we may assume that $f_1 = g_1$. Hence also

$$af_2 \dots f_r = ag_2 \dots g_s$$

The induction assumptions implies that $r = s$ and after reordering $f_2 = g_2, \dots, f_r = g_r$. So the f_i's are unique up to reordering.

(b) If g and h are of the given form then $gh = f$ and so g is a divisor of f.

Suppose now that g divides f. Then $f = gh$ for some $h \in \mathbb{F}[x]$. If $\deg g = 0$, then (a) holds with $\epsilon_i = 0$ for all $1 \le i \le r$. So suppose $\deg g = 0$. By (a) we can write $g = t\tilde{g}$ where t is an irreducible monic polynomial. Since $f = gh = t\tilde{g}h$, t divides f and so by A.3.3, $t = f_i$ for some $1 \le i \le t$. Without loss $i = 1$. Then

$$\tilde{g}h = af_2 \dots f_n$$

Note that $\operatorname{lead}(\tilde{g}) = \operatorname{lead}(g) = b$. By induction

$$\tilde{g} = bf_2^{\epsilon_2} \dots f_r^{\epsilon_r} \text{ and } h = f_2^{\delta_2} \dots f_n^{\delta_n}$$

where $c = \frac{a}{b}$, $\epsilon_i \in \{0, 1\}$ and $\delta_i = 1 - \epsilon_i$ for $2 \le i \le r$. Thus (b) holds with $\epsilon_1 = 1$ and $\delta_1 = 0$. □

A.4 Primitive elements in finite field

Lemma A.4.1. *Let $\mathbb{E}$ be a finite field and put $t = |\mathbb{E}| - 1$. Let $e \in \mathbb{E}^\sharp$. Then*

(a) *There exists positive integer m with $e^m = 1$. The smallest such positive integer is called the order of e in $\mathbb{E}^\sharp$ and is denoted by $|e|$.*

(b) *The elements $e^i, 0 \le i < |e|$, are pairwise distinct.*

(c) *Let $n \in \mathbb{Z}$ and r the remainder of n, then divided by $|e|$. Then $e^n = e^r$.*

(d) *Let $n, m \in \mathbb{Z}$. Then $e^n = e^m$ if and only if n and m have the same remainder when divided by $|e|$ and if and only if $|e|$ divides $n - m$.*

(e) *$|e|$ divides t.*

Proof. We first prove:

$(*)$ *Let s be a positive integer, then $e^i, 0 \le i \le s$ are pairwise distinct if and only $e^i \ne 1$ for all $1 \le i \le s$.*

Indeed $e^i \ne e^j$ for all $0 \le i < j \le s$, if and only if $e^{j-i} \ne 1$ for all $0 \le i < j \le s$ and so if and only $e^i \ne 1$ for some $1 \le i \le s$.

(a): Since $|\mathbb{E}^\sharp| = t$, the elements $e^i, 0 \le i \le t$ cannot be pairwise distinct. So by $(*)$ there exists $1 \le m \le t$ with $e^m = 1$.

(b) By minimality of $|e|$, $e^i \neq 1$ for all $1 \leq i < |e|$. So (b) follows from $(*)$.

(c) Let r be the remainder of n. Then $n = q|e| + r$ for some $q \in \mathbb{Z}$ and so $e^{q|e|+r|} = (e^{|e|})^q e^r = 1^q e^r = e^r$.

(d) Let r and s be the remainders of n and m when divides by $|e|$. By (c), $e^n = e^r$ and $e^m = e^s$. By (b), $e^n = e^m$ if and only if $r = s$ and so if and only if $|e|$ divides $n - m$.

(e) Define a relation $\sim$ on $\mathbb{E}^\sharp$, by $a \sim b$ of $a = be^i$ for some $i \in \mathbb{Z}$. Since $a = ae^0$, $\sim$ is reflexive. If $a = be^i$, then $b = ae^{-i}$ and so $\sim$ is symmetric. If $a = be^i$ and $c = be^j$, then $c = ae^ie^j = ae^{i+j}$ and so $\sim$ is transitive. Thus $\sim$ is an equivalence relation. Note that $ae^i = ae^j$ if and only if $e^i = e^j$ and if and only if i and j have the same remainder then divided by $|e|$. Since there are $|e|$ such remainders, each equivalence class has exactly $|e|$ elements. Let d be the number of equivalence class of $\sim$. Since each element of $\mathbb{E}^\sharp$ lies in exactly one equivalence class and since each equivalence class has $|e|$ elements, $|\mathbb{E}^\sharp| = d|e|$. Thus $t = d|e|$ and $|e|$ divides t. □

Lemma A.4.2. *Let n and d be positive integers with $d \mid n$. Define*

$$D_{\mathrm{d}}(n) = \{m \mid 0 \leq m < n, \gcd(n,m) = d\}.$$

Then

(a) $D_{\mathrm{d}}(n) = \{ed \mid e \in \mathbb{Z}^*_{\frac{n}{d}}\}$.

(b) $|D_{\mathrm{d}}(n)| = \phi(\frac{n}{d})$.

(c) $n = \sum_{\substack{d \in \mathbb{Z}^+ \\ d|n}} \phi(d)$.

(d) $\phi(n) \geq 1$.

Proof. (a) Let $0 \leq m < n$. Suppose $\gcd(m,n) = d$. Then $d \mid m$ and so $m = ed$ for some $e \in \mathbb{Z}$. Since $0 \leq m < n$ we have $0 \leq e < \frac{n}{d}$. Since $\gcd(m,n) = d$ we have $\gcd(e, \frac{n}{d}) = 1$ and so $e \in \mathbb{Z}^*_{\frac{n}{d}}$.

Conversely, if $e \in \mathbb{Z}^*_{\frac{n}{d}}$, then $0 \leq ed < n$ and $\gcd(ed, n) = d\gcd(d, \frac{n}{d}) = d$. Thus $ed \in D_{\mathrm{d}}(n)$ and (a) holds.

(b) follows from (a).

(c) Let $0 \leq m < n$. Then there exists a unique divisor f of n with $m \in D_f(n)$, namely $f = \gcd(n,m)$. Thus

$$n = |\{m \mid 0 \leq m < n\}| = |\sum_{f|n} D_f(n)| = \sum_{f|n} |D_f(n)| = \sum_{f|n} \phi\left(\frac{n}{f}\right) = \sum_{d|n} \phi(d)$$

(d) Just note that $\gcd(n-1, n) = 1$ and so $n - 1 \in \mathbb{Z}^*_n$. □

Definition A.4.3. *Let $\mathbb{F}$ be a field and $n \in \mathbb{Z}^+$. The $\alpha \in \mathbb{F}$ is called a primitive root of $x^n - 1$ if $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$ are pairwise distinct root of $x^n - 1$.*

Note that if α is primitive root of $x^n - 1$. Then

$$x^n - 1 = (x-1)(x-\alpha)\dots(x-\alpha^{n-1}).$$

Lemma A.4.4. *Let $\mathbb{F}$ be a field, $n \in \mathbb{Z}^+$ and $e \in \mathbb{F}^\sharp$ an element of order n. Then*

(a) *e is a primitive root of $x^n - 1$.*

(b) *Let $m \in \mathbb{Z}$ and put $d = \gcd(m,n)$. Then e^m has order $\frac{n}{d}$.*

(c) *Let $d \in \mathbb{Z}^+$ with $d \mid n$. Then $\mathbb{F}^\sharp$ has exactly $\phi(d)$ elements of order d, namely the elements $e^i, i \in D_{\frac{n}{d}}(n)$.*

Proof. (a) Let $0 \le i < n$. Then $(e^i)^n = (e^n)^i = 1$ and so e^i is a root of $x^n - 1$. Since the $e^i, 0 \le i < n$ are pairwise distinct, (a) holds.

(b) Let $l \in \mathbb{Z}^+$. Then $(e^m)^l = 1$ if and only if $e^{ml} = 1 = e^0$, if and only if $n \mid ml$ and if and only if $\frac{n}{d} \big| l$. Thus e^m has order $\frac{n}{d}$.

(c) Let $a \in \mathbb{F}^\sharp$. If a has order d, then $a^n = 1$ and so a is root of $x^n - 1$. So by (a), $a = e^i$ for some $0 \le i < n$. By (b), e^i has order d if and only if $\gcd(i,n) = \frac{n}{d}$ and so if and only if $i \in D_{\frac{n}{d}}(n)$. Since $|D_{\frac{n}{d}}(n)| = \phi\left(\frac{n}{\frac{n}{d}}\right) = \phi(d)$, (c) holds. □

Lemma A.4.5. *Let $\mathbb{E}$ be a finite field and put $t = |\mathbb{E}| - 1$. Then there exists an element $\beta \in \mathbb{E}$ such $\beta^t = 1$ and that*

$$\mathbb{E}^\sharp = \{\alpha^i \mid 0 \le i < t\}$$

Such an β is called a primitive element in $\mathbb{E}$.

Proof. For $n \in \mathbb{Z}^+$, let A_n be set of elements of order n in $\mathbb{E}^\sharp$. If $e \in \mathbb{E}^\sharp$ then $|e|$ is a divisor of t and so

$$(*) \qquad t = |\mathbb{E}^\sharp| = |\sum_{n|t} A_n| = \sum_{n|t} |A_n|.$$

Let $n \mid t$. Suppose $A_n \ne \varnothing$. Then $\mathbb{E}^\sharp$ has an element of order n and so by A.4.4, $|A_n| = \phi(n)$. Hence either $|A_n| = 0$ or $|A_n| = \phi(n)$. Therefore

$$(**) \qquad \sum_{n|t} |A_n| \le \sum_{n|t} \phi(n) = t.$$

Together with (*) we conclude that equaliy must holds everywhere in (*) and (**). In particular, $|A_n| = \phi(n)$ for all $n \mid t$. Thus $A_t = \phi(t) \ne 1$ and so $\mathbb{E}$ has an element β of order t. Then $\{\beta^i \mid 0 \le i < t\}$ are t-pairwise distinct elements in $\mathbb{E}^\sharp$ and the the lemma is proved.. □

Lemma A.4.6. *Let n be a positive integer. Let $n = 2^k m$ where $m, k \in \mathbb{N}$ with m odd. Let $\mathbb{E}$ be a splitting field for $x^m - 1$ over $\mathbb{F}_2$ and let $\alpha_1, \alpha_2 \dots, \alpha_m \in \mathbb{E}$ with*

$$x^m - 1 = (x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_m).$$

Then

(a) $x^n - 1 = (x^m - 1)^{2^k}$.

(b) $\alpha_1, \alpha_2, \alpha_3, \dots \alpha_m$ *are pairwise distinct.*

Proof. (a) Note that $(x^m - 1)^2 = x^{2m} - 1$ in $\mathbb{F}_2[x]$ and so (a) follows by induction on k.

(b) Suppose that $\alpha_i = \alpha_j$ for some $1 \le i < k \le m$. Put $\alpha = \alpha_i$. Then

$$x^m - 1 = (x - \alpha)^2 g$$

for some $g \in \mathbb{E}[x]$. Taking derivatives gives

$$m x^{m-1} = 2(x - \alpha) g + (x - \alpha)^2 g'.$$

Obseerve that in $\mathbb{E}$, $2 = 0$ and, since m is odd, $m = 1$. Therefore,

$$x^{m-1} = (x - \alpha)^2 g'.$$

Hence α is a root of x^{m-1} and so $\alpha = 0$, a contradiction to $\alpha^m = 1$. □

Lemma A.4.7. *Let n be a positive odd integers and let $\mathbb{E}$ be a finite field containg $\mathbb{F}_2$. Put $t = |\mathbb{E}| - 1$ and let β be a primitive root for $\mathbb{E}$.*

(a) $\mathbb{E}$ *is a splitting field for $x^n - 1$ if and only if n divides t.*

(b) *Suppose n divides t and put $\alpha = \beta^{\frac{t}{n}}$. Then α is a primitive root of $x^n - 1$.*

Proof. Let $d = \gcd(t, n)$ and put $s = \frac{t}{d}$. Then β^m is a root of $x^n - 1$ if and only if $\beta^{mn} = 1$, if and only if $t \mid mn$ and if and only if $s \mid m$. Thus the roots of $x^n - 1$ in $\mathbb{E}$ are

$$\beta^{is}, \quad 0 \le i < d.$$

Therefore $\mathbb{E}$ contains exactly d roots of $x^n - 1$.

From A.4.6, $\mathbb{E}$ is a splitting field for $x^n - 1$ if and only if $\mathbb{E}$ contains exactly n-roots of $x^n - 1$, if and only of $d = n$, and if and only if $n \mid t$.

If $n \mid t$, then $\beta^s = \beta^{\frac{t}{n}} = \alpha$ and so the roots of $x^n - 1$ are $\alpha^i, 0 \le i < n$. □

Appendix B

Constructing Sources

B.1 Marginal Distributions on Triples

Lemma B.1.1. *Let I, J, K be finite sets. Let $f : I \times J \times K \to \mathbb{R}$ be function. $f_{I\times J}$ the marginal tuple of f on $I \times J$ $\big($via $I \times J \times K = (I \times J) \times K\big)$ and let f_I the marginal tuple of f on I $\big($via $I \times J \times K = I \times (J \times K)\big)$. Then f is the marginal tuple of $f_{I\times J}$ on I.*

Proof. Let g be the marginal tuple of $f_{I\times J}$ on I and let $i \in I$. Then

$$\begin{aligned} g(i) &= \textstyle\sum_{j\in J} f_{I\times J}(i,j) \\ &= \textstyle\sum_{j\in J} \Big(\sum_{k\in K} f(i,j,k)\Big) \\ &= \textstyle\sum_{(j,k)\in J\times K} f(i,j,k) \\ &= f_I(i) \end{aligned}$$

□

Lemma B.1.2. *A (S,P) be source. Then the following statements are equivalent:*

(a) *For all $r \in \mathbb{Z}^+$, all strictly increasing r-tuples $(l_1,\dots,l_r)$ of positive integers and all $t \in \mathbb{N}$*

$$p^{(l_1,\dots,l_r)} = p^{(l_1+t,\dots,l_r+t)}$$

(b) *(S,P) is stationary.*

(c) *For all $r \in \mathbb{Z}^+$, $p^r = p^{(2,\dots,r+1)}$.*

Proof. Suppose (a) holds. Choosing $(l_1,\dots,l_r) = (1,\dots,r)$ we see that

$$p^r = p^{(1,\dots,r)} = p^{(1+t,\dots,r+t)}$$

and so (S,P) is stationary.

Suppose (S,P) is stationary. Choosing $t = 1$ in the definition of stationary gives (c).

Suppose (c) holds. We need to prove that (a) holds. So let $r \in \mathbb{Z}^+$, let $(l_1,\dots,l_r)$ be increasing r-tuples of positive integers and let $t \in \mathbb{N}$. If $t = 0$, (a) obviously holds. By induction on t it suffices to consider the case $t = 1$. Put $u = l_r$, $v = u - l$ and let $k = (k_1,\dots,k_v)$ be the increasing t-tuple of positive integers with $\{1,2,\dots,u\} = \{l_1,\dots,l_r\} \cup \{k_1,\dots,k_v\}$. Let $l = (l_1,\dots,l_r)$. Identifying S^u with $S^r \times S^v$ via $s \to (s_l, s_k)$ we see that p^l is the marginal distribution of p^u on S^r.

Let $\tilde{k} = (k_1 + 1,\dots,k_t + 1)$ and $\tilde{l} = (l_1 + 1,\dots,l_r + 1)$

Identifying S^{u+n} with $S^r \times S^v \times S$ via $s \to (s_{\tilde{l}}, s_{\tilde{k}}, s_1)$ we see that $p^{\tilde{l}}$ is the marginal distribution of p^{u+1} on S^r. Also $p^{(2,\dots,u+1)}$ is the marginal distribution of p^{u+1} on $S^r \times S^v$. Thus B.1.1 shows that $p^{\tilde{l}}$ is the marginal distribution of $p^{(2,\dots,u+1)}$ on S^r.

Since (c) holds, $p^u = p^{(2,\dots,u+1)}$. Hence also the marginal distribution p^l and $p^{\tilde{l}}$ of these distribution on S^r are equal. □

B.2 A stationary source which is not memory less

Example B.2.1. An example of a stationary source which is not mememory less.

For $z = z_1 \dots z_n \in B^*$ define $u(z) = |\{i \mid 1 \le i < n, z_i = z_{i+1}\}|$. Define $P(\varnothing) = 1$ and if $n \ge 1$,

$$P(z) = \frac{1}{2}\frac{3^{u(z)}}{4^{n-1}}$$

Note that $P(0) = P(1) = \frac{1}{2}\frac{3^0}{4^{1-1}} = \frac{1}{2}$. Also $u(zs) = u(z) + 1$ if $s = s_n$ and $u(zs) = u(s)$ if $z_n \ne s$. Hence for $z \in \mathbb{B}^n$ with $n \ge 1$ and $s \in \mathbb{B}$:

$$P(zs) = \begin{cases} \frac{3}{4}P(z) \text{ if } s = z_n \\ \frac{1}{4}P(z) \text{ if } s \ne z_n \end{cases}$$

Thus $P(z) = P(z0) + P(z1)$ and P is source. Similarly

$$P(sz) = \begin{cases} \frac{3}{4}P(z) \text{ if } s = z_1 \\ \frac{1}{4}P(z) \text{ if } s \ne z_1 \end{cases}$$

Thus $P(z) = P(0z) + P(1z)$ and so

$$p^n(z) = P(z) = P(0z) + P(1z) = p^{(2,\dots,n+1)}(z)$$

and so by B.1.2 P is stationary.

B.3 Matrices with given Margin

Let I and J be non empty alphabets, f an I-tuple and g an J tuple with coefficients in $\mathbb{R}^{\geq 0}$ such that

$$(*) \qquad t := \sum_{i\in I} f_i = \sum_{j\in J} g_j$$

We will give an inductive construction to determine all $I \times J$ matrices h with coefficients in $\mathbb{R}^{\geq 0}$ whose marginal tuples are f and g.

Suppose first that $|I| = 1$ and let $i \in I$ and $j \in J$. Then $g_j = \sum_{i\in I} h_{ij} = h_{ij}$ and so only row of h is equal to g. So there is just one solution in this case.

Suppose next that $|J| = 1$. Then the only column of h is equal to equal to f.

Suppose that $|I| = |J| = 2$. Let $I = \{a, b\}$ and $J = \{c, d\}$ with $f_a \leq f_b$ and $g_c \leq g_d$. Let $u \in \mathbb{R}$ with $0 \leq u \leq \min(f_a, g_c)$. By (*)

$$f_a + f_b = k = g_c + g_d \qquad \text{and so } g_d - f_a = f_b - g_c$$

So we can define h as follows

h	c	d	f
a	u	$f_a - u$	f_a
b	$g_c - u$	$g_d - f_a + u = f_b - g_c + u$	f_b
g	g_c	g_d	t

By choice of u, $u \leq f_a$ and $u \leq g_c$. So both $f_a - u$ and $g_c - u$ are non-negative. Note that $t = f_a + f_b \geq 2f_a$ and $t = g_c + g_d \leq 2g_d$. Hence $g_d \geq \frac{k}{2} \geq f_a$ and so $g_d - f_a + u \geq u \geq 0$.

Suppose now that $|I| > 2$ or $|J| > 2$. By symmetry we may assume that $|J| > 2$.

Pick $u, v \in J$ with $u \neq v$ and put $\tilde{J} = J \setminus \{v\}$. Define a $\tilde{J}$-tuple $\tilde{g}$ on $\tilde{J}$ by

$$\tilde{g}_j = \begin{cases} g_j & \text{if } j \neq u \\ g_u + g_v & \text{if } j \neq u \end{cases}$$

Then

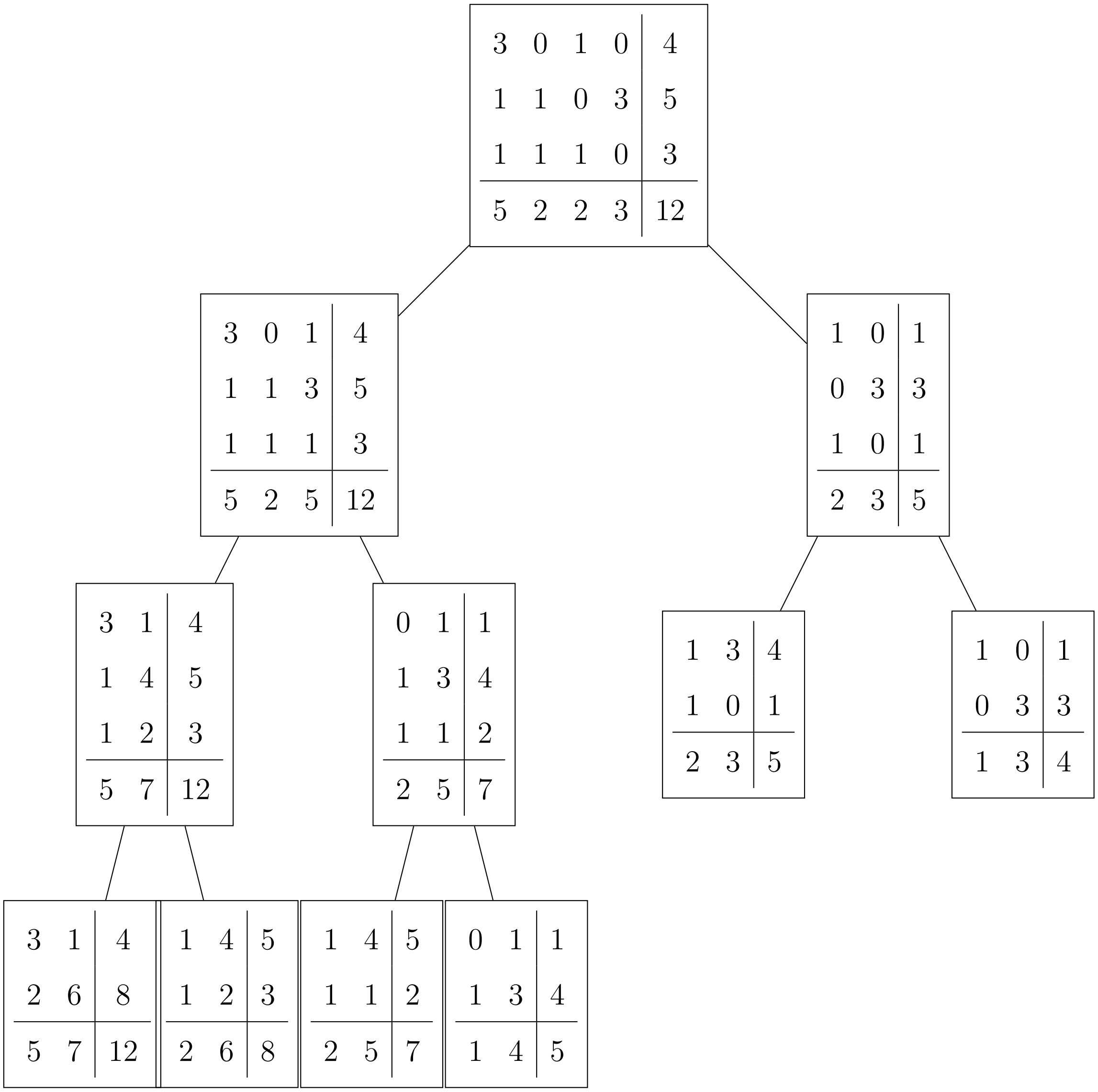
$$\sum_{j\in\tilde{J}} \tilde{g}_j = \tilde{g}_u + \sum_{\substack{j\in\tilde{J}\\ j\neq u}} \tilde{g}_j = g_u + g_v + \sum_{\substack{j\in J\\ j\neq u,v}} g_j = \sum_{j\in J} g_j = t = \sum_{i\in I} f_i$$

Inductively we may assume that we found all possible $I \times \tilde{J}$-matrices $\tilde{h}$ with coeffcients in $\mathbb{R}^{\geq 0}$ and marginal distribution f and $\tilde{g}$.

Put $\hat{J} = \{u, v\}$. Let $\hat{g} = g|_{\hat{J}}$ and let $\hat{f}$ be column u of $\tilde{h}$. So $\hat{g}$ is a $\hat{J}$-tuple and $\hat{f}$ is an I-tuple. Since $\tilde{g}$ is the marginal distribution of $\tilde{h}$, the sum of $\hat{f}$ is $\tilde{g}_u = g_u + g_v$. The sum of $\hat{g}$ is also $g_u + g_v$. Since $|\hat{J}| = 2 < |J|$ we may assume by induction that we found all $I \times \hat{J}$-matrices $\hat{h}$ with coeffcients in $\mathbb{R}^{\geq 0}$ and marginal distribution $\hat{f}$ and $\hat{g}$. Define the $I \times J$-matrix h by

$$h_{ij} = \begin{cases} \tilde{h}_{ij} & \text{if } j \in J \smallsetminus \tilde{J} \\ \hat{h}_{ij} & \text{if } j \in \hat{J} \end{cases}$$

So columns u and v of h come form $\hat{h}$, while the remaining columns come from $\tilde{h}$.



$$\begin{array}{cccc|c} 3 & 0 & 1 & 0 & 4 \\ 1 & 1 & 0 & 3 & 5 \\ 1 & 1 & 1 & 0 & 3 \\ \hline 5 & 2 & 2 & 3 & 12 \end{array}$$

$$\begin{array}{ccc|c} 3 & 0 & 1 & 4 \\ 1 & 1 & 3 & 5 \\ 1 & 1 & 1 & 3 \\ \hline 5 & 2 & 5 & 12 \end{array} \qquad \begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 3 & 3 \\ 1 & 0 & 1 \\ \hline 2 & 3 & 5 \end{array}$$

$$\begin{array}{cc|c} 3 & 1 & 4 \\ 1 & 4 & 5 \\ 1 & 2 & 3 \\ \hline 5 & 7 & 12 \end{array} \qquad \begin{array}{cc|c} 0 & 1 & 1 \\ 1 & 3 & 4 \\ 1 & 1 & 2 \\ \hline 2 & 5 & 7 \end{array} \qquad \begin{array}{cc|c} 1 & 3 & 4 \\ 1 & 0 & 1 \\ \hline 2 & 3 & 5 \end{array} \qquad \begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 3 & 3 \\ \hline 1 & 3 & 4 \end{array}$$

$$\begin{array}{cc|c} 3 & 1 & 4 \\ 2 & 6 & 8 \\ \hline 5 & 7 & 12 \end{array} \qquad \begin{array}{cc|c} 1 & 4 & 5 \\ 1 & 2 & 3 \\ \hline 2 & 6 & 8 \end{array} \qquad \begin{array}{cc|c} 1 & 4 & 5 \\ 1 & 1 & 2 \\ \hline 2 & 5 & 7 \end{array} \qquad \begin{array}{cc|c} 0 & 1 & 1 \\ 1 & 3 & 4 \\ \hline 1 & 4 & 5 \end{array}$$

Appendix C

More On channels

C.1 Sub channels

Lemma C.1.1. *Let $\Gamma : I \times J \to [0,1]$ be a channel. Let $K \subseteq I$ and let Ξ be the restriction of Γ to $K \times J$. (So Ξ is the function from $K \times I \to [0,1]$ with $\Xi_{kj} = \Gamma_{kj}$ for all $k \in K$, $j \in J$.) Let p be a probability distribution on K, and define $\hat{p} : I \to [0,1]$ by $\hat{p}_i = p_i$ if $i \in K$ and $\hat{p}_i = 0$ of $i \in I \smallsetminus K$. Put $q = p\Xi$. Then*

(a) *Ξ is a channel.*

(b) *$\hat{p}$ is a probability distribution on I.*

(c) *$q = p\Xi = \hat{p}\Gamma$.*

(d) *$H^{\Xi}(q \mid k) = H^{\Gamma}(q \mid k)$ for all $k \in K$.*

(e) *$H^{\Xi}(q \mid p) = H^{\Gamma}(q \mid \hat{p})$.*

(f) *$\gamma(\Xi) \leq \gamma(\Gamma)$*

Proof. (a) Let $k \in K$. Then $\mathrm{Row}_k(\Xi) = \mathrm{Row}_k(\Gamma)$ and so $\mathrm{Row}_k(\Xi)$ is a probability distribution on J.

(b) Since $\hat{p}_i = 0$ for all $i \in I \smallsetminus K$,

$$\sum_{i \in I} \hat{p}_i = \sum_{i \in K} \hat{p}_i = \sum_{i \in K} p_i = 1.$$

(c) $\hat{p}\Gamma = \sum_{i \in I} \hat{p}_i \Gamma_{ij} = \sum_{i \in K} p_i \Gamma_{ij} = \sum_{i \in K} p_i \Xi_{ij} = p\Xi$.

(d) $H^{\Xi}(q \mid k) = H(\mathrm{Row}_k(\Xi)) = H(\mathrm{Row}_k(\Gamma)) = H^{\Gamma}(q \mid k)$.

(e) Using 9.4.6 twice we have

$$H^{\Xi}(q \mid p) = \sum_{i \in K} p_i H^{\Xi}(q \mid i) = \sum_{i \in K} p_i H^{\Gamma}(q \mid i) = \sum_{i \in} \hat{p}_i H^{\Gamma}(q \mid i) = H^{\Gamma}(q \mid \hat{p}).$$

(f) Let $\mathcal{P}(I)$ and $\mathcal{P}(K)$ be the set of probability distribution on I and K respectively. Then using (c) and (e),

$$\gamma(\Xi) = \max_{p \in \mathcal{P}(K)} H(q) - H^{\Xi}(q \mid p) = \max_{p \in \mathcal{P}(K)} H(q) - H^{\Gamma}(q \mid \hat{p}) \leq \max_{p \in \mathcal{P}(I)} H(q) - H^{\Gamma}(q \mid p) = \gamma(\Gamma).$$

□

Appendix D

Examples of codes

D.1 A 1-error correcting binary code of length 8 and size 20

Example D.1.1. The rows of following matrix form a binary code of size 20, length 9 and minimum distance 4. Deleting any of the columns produces a 1-error correcting code of size 20 and length 8.

$$\begin{bmatrix} \vec{0} & \vec{0} & \vec{0} \\ J & P^2 & P \\ P & J & P^2 \\ P^2 & P & J \\ I & I+P & I+P^2 \\ I+P^2 & I & I+P \\ I+P & I+P^2 & I \\ \vec{1} & \vec{1} & \vec{1} \end{bmatrix}$$

Here

$$\vec{0} = 000, \quad \vec{1} = 111, \quad I = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

and so

$$P^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad I + P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad I + P^2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Bibliography

[Text Book] Norman L.Biggs *Codes, An introduction to information communication and cryptography* Springer UMS **2008**

Index