Introduction to algebraic codings
Lecture Notes for MTH 416
Fall 2012

Ulrich Meierfrankenfeld

December 7, 2012

Preface

These are the Lecture Notes for the class MTH 416 in Fall 2012 at Michigan State University.

The Lecture Notes will be frequently updated throughout the semester.

Contents

Chapter I

Coding

I.1 Matrices

Definition I.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) Let $n \in \mathbb{N}$, where $\mathbb{N} = \{0, 1, 2, 3, \dots, \}$ denotes the set of non-negative integers. Then an n-tuple is an $\{1, 2, \dots, n\}$-tuple.

Notation I.2. *Notation for tuples.*

1.

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

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

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

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

$y_1 = a, y_2 = a, y_3 = b$ and $y_4 = c$

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

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

Definition I.3. *Let I, J and R be sets.*

(a) 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.

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

Notation I.4. *Notations for matrices*

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

M	x	y	z
a	0	1	2
b	1	2	3
c	2	3	4
d	3	4	5

stands for the $\{a, b, c, d\} \times \{x, y, z\}$ matrix M with coefficients in $\mathbb{Z}$ such that $m_{ax} = 0$, $m_{ay} = 1$, $m_{bx} = 1$, $m_{cz} = 4, \dots$.

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 $m_{11} = 0, m_{12} = 2$, $m_{21} = 4$, $m_{23} = 6, \dots$.

Definition I.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 $\operatorname{Row}_i(M)$ 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 $\operatorname{Col}_j(M)$

Definition I.6. *Let A be $I \times J$-matrix, B an $J \times K$ 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 I.7. *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 $\begin{array}{c|cccc} B & \alpha & \beta & \gamma & \delta \\ \hline x & 0 & 0 & 1 & 0 \\ y & 1 & 0 & 0 & 1 \\ x & 1 & 1 & 0 & 0 \end{array}$

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

$$\begin{array}{c|cccc} AB & \alpha & \beta & \gamma & \delta \\ \hline a & 3 & 2 & 0 & 1 \\ b & 4 & c & 1 & 2 \end{array}$$

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

   Then Ax is the 2-tuple

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

3. Since we may represent tuples also as rows, (2) can be restated as:

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

   Then Ax is the 2-tuple

$$(3, 5)$$

4. 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 $u : \begin{array}{cc} a & b \\ \hline 2 & -1 \end{array}$

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

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

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

$$xy = 3$$

I.2 Basic Definitions

Definition I.8. *An alphabet is a finite set. The elements of an alphabet are called symbols.*

Example I.9. *(a)* $\mathbb{A} = \{A, B, C, D, \ldots, X, Y, Z, \sqcup\}$ *is the alphabet consisting of the regular 27 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 I.10. *Let S be an alphabet and n a non-negative integer.*

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

(b) $\varnothing$ denote the unique message of length 0 *in S.*

(c) S^n is the set of all messages of length n in S.

(d) S^* is the set of all messages in S, so

$$S^* = S_0 \cup S_1 \cup S_2 \cup S_3 \cup \ldots \cup S_k \cup \ldots$$

Example I.11. *1.* ROOM$\sqcup$C206WH$\sqcup$IS$\sqcup$NOW$\sqcup$A216WH *is a message in the alphabet* $\mathbb{A}$.

2. 1001110111011001 *is a message in the alphabet* $\mathbb{B}$.

3. $\mathbb{B}^0 = \{\varnothing\}$, $\mathbb{B}^1 = \mathbb{B} = \{0, 1\}$. $\mathbb{B}^2 = \{00, 01, 10, 11\}$, $\mathbb{B}^3 = \{000, 001, 010, 011.100, 101, 110, 111\}$, *and* $\mathbb{B}^* = \{\varnothing, 0, 1, 00, 01, 10, 11, 000, \ldots, 111, 0000. \ldots, 1111, \ldots\}$

Definition I.12. *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 $c(s)$ in 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 I.13. *1. The function $c : \mathbb{A} \to \mathbb{A}$ such that*

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

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

$$x \to 0, y \to 01, z \to 10$$

is a code for $\{x, y, z\}$ using $\mathbb{B}$. The set of codewords is $C = \{0, 01, 10\}$.

Definition I.14. *Let $c : S \to T^*$ be a code. Then the function $c^* : S^* \to T^*$ defined by*

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

for all $s_1 \ldots s_n \in S^*$ is called the concatenation of c. We also call c^* 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 I.15. *1. Let $c : \mathbb{A} \to \mathbb{A}^*$ be the code from I.13(1). Then c*(MTH)=PWK.

2. Let $c : \{x, y, z\} \to \mathbb{B}^*$ be the code from I.13(2). Then $c(xzzx)$ = 010100 *and $c(yyxx)$* = 010100.

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

Definition I.16. *A code $c : S \to T^*$ is called uniquely decodable (UD) if the extended function $c : S^* \to T^*$ is 1-1.*

Example I.17. *(a) The code from I.15(1) is* UD.

(b) The code from I.15(2) is not UD.

I.3 Coding for economy

Example I.18. *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	0	P	Q	R
$\bullet - - - \odot$	$- \bullet - \odot$	$\bullet - \bullet \bullet \odot$	$- - \odot$	$- \bullet \odot$	$- - -$	$\bullet - - \bullet \odot$	$- - \bullet - \odot$	$\bullet - \bullet \odot$
S	T	U	V	W	X	Y	Z	$\sqcup$
$\bullet \bullet \bullet \odot$	$- \odot$	$\bullet \bullet - \odot$	$\bullet \bullet \bullet - \odot$	$\bullet - - \odot$	$- \bullet \bullet - \odot$	$- \bullet - - \odot$	$- - \bullet \bullet$	$\odot \odot$

I.4 Coding for reliability

Example I.19. *Codes for* $\{\text{Buy}, \text{Sell}\}$ *using* $\{B, S\}$.

1. Buy → *B,* Sell → S.

2. Buy → *BB,* Sell → SS.

3. Buy → *BBB,* Sell → SSS.

4. Buy → *BBBBBBBBB,* Sell → SSSSSSSSS.

I.5 Coding for security

Example I.20. *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.* $\sqcup$ *is unchanged. For example the code in I.13(1), is* c_3.

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

Chapter II

Prefix-free codes

II.1 The decoding problem

Definition II.1. *Let C be a code using the alphabet T.*

(a) Let $a, b \in T^*$. Then a is called a prefix of b if there exists a message r in T^* with $b = ar$.

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

(c) C is called prefix-free (PF) if no codeword is a prefix of a different codeword.

Remark II.2. *T be an alphabet and $b = t_1 \dots t_m$ a message of length m in T.*

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

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

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

Proof. Let $a = s_1 \dots s_n$ be a prefix of length n of b. Then $b = ar$ for some $r \in T^*$ or length say k. Let $r = u_1 \dots u_k$. Then $b = s_1 \dots s_n u_1 \dots u_k$. Thus $m = n + k$ and $n \leq m$ and (a) holds. Also $t_i = s_i$ for $1 \leq i \leq n$ and so $a = t_1 \dots t_n$ and (b) holds. If a is a parent then $r \in T$, that is $k = 1$ and $n = m - 1$. So (c) follows from (a). □

Example II.3. *Which of the following codes are PF? UD?*

1. Let $C = \{10, 01, 11, 011\}$. 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 the 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?
∅	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?
∅	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 II.4. *Any regular PF code is UD.*

Proof. Let $n, m \in \mathbb{N}$ and $c_1, \ldots, c_n, d_1, \ldots, d_m$ be codewords with

$$c_1 \ldots c_n = d_1 \ldots d_m$$

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

Suppose first that $n = 0$. Then $e = \varnothing$ and so $d_1 \ldots d_m = \varnothing$, By definition of a regular code each d_i is not empty. Hence also $m = 0$ and we are done in this case.

So we may assume that $n > 0$ and similarly $m > 0$. Let k and l be the lengths of c_1 and d_1 respectively.

We may assume that $k \leq l$. Since $e = c_1 \ldots c_n$, c_1 is the prefix of length k of e. Also $e = d_1 \ldots d_m$ implies that d_1 is the prefix of length l of e. Since $k \leq l$ this shows that c_1 is the prefix of length k of d_1. Since C is prefix free we conclude that $c_1 = d_1$. Since $c_1 c_2 \ldots c_n = d_1 d_2 \ldots d_m$ this implies

$$c_2 \ldots c_n = d_2 \ldots d_m$$

From the induction assumption we conclude that $n - 1 = m - 1$ and $c_2 = d_2, \ldots, c_n = d_m$. □

Lemma II.5. *(a) All UD codes are regular.*

(b) Let $c : S \to T^*$ be not regular. Then c is PF if and only if $|S| = 1$ and $c(s) = \varnothing$ for $s \in C$.

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 and (a) holds.

If $|S| > 1$, then c has at least two codewords. Since $c(s) = \varnothing$ is the prefix of any message we conclude that S is not PF. This proves the forward direction in (b).

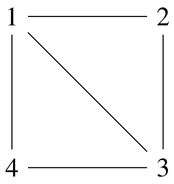
Any code with only one codeword is PF and so the backwards direction in (b) holds. □

II.2 Representing codes by trees

Definition II.6. *A graph G is a pair (V, E) such that V and E are sets and each elements e of E is a subset of size two of V. The elements of V are called the vertices of V. The elements of E are called the edges of V. We say that the vertex a is adjacent to b in G if $\{a, b\}$ is an edge.*

Example II.7. *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 II.8. *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 $1 \leq i \leq n$.

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

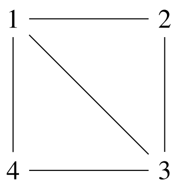
(c) A path is called a cycle if the first and the last vertex are the same.

(d) A cycle is called simple if all the vertices except the last one are pairwise distinct.

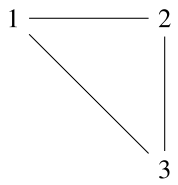
(e) A tree is a connected graph with no simple cycles of length larger than two.

Example II.9. *Which of the following graphs are trees?*

1.

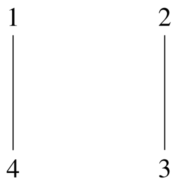


is connect, but



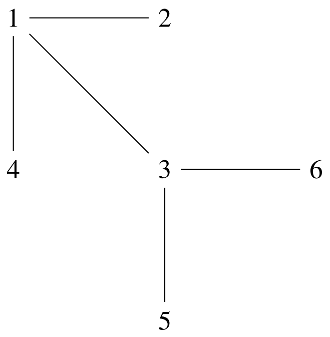
is simple circle of length three in G. So G is not a tree.

2.



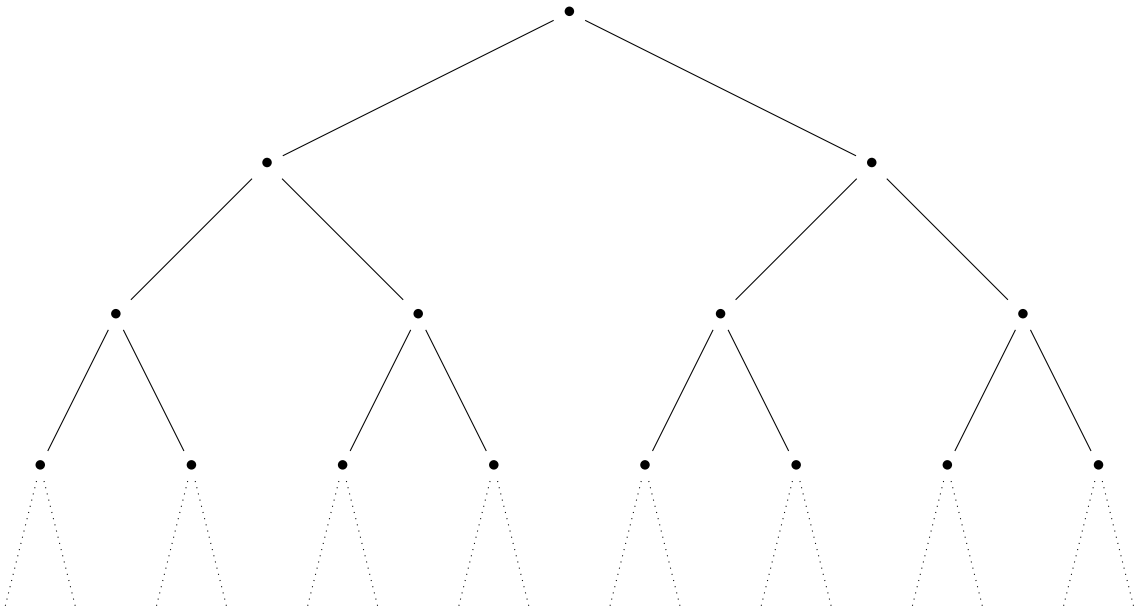
has no simple circle of length larger than two. But it is not connected since there is no path from 1 to 2. So G is not a tree.

3.



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

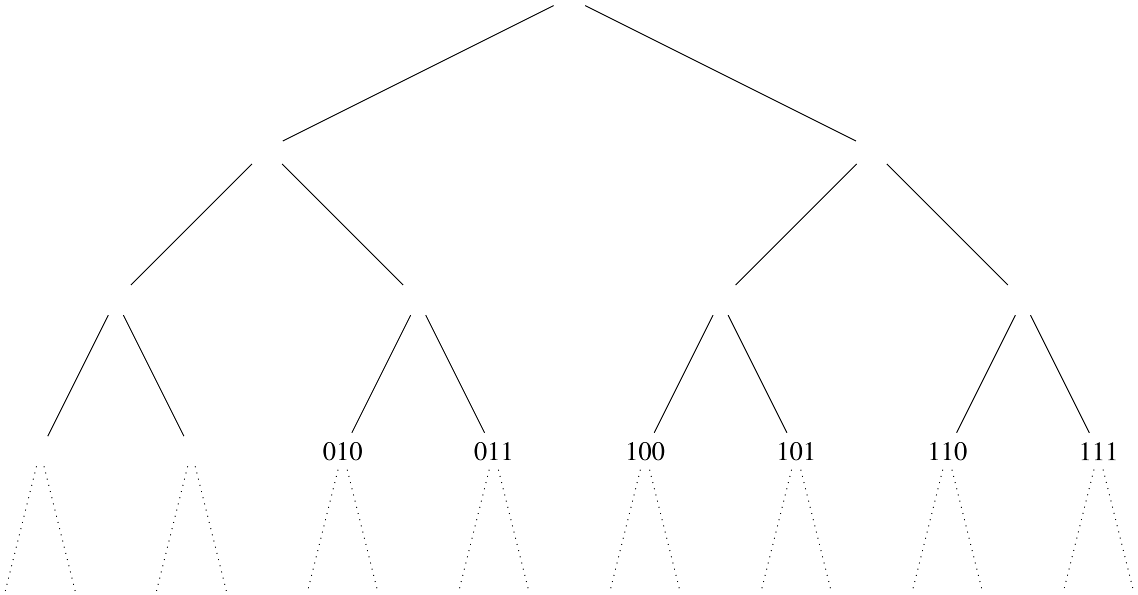
Example II.10. *The infinite binary tree*



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

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

So the infinite binary tree now looks like:

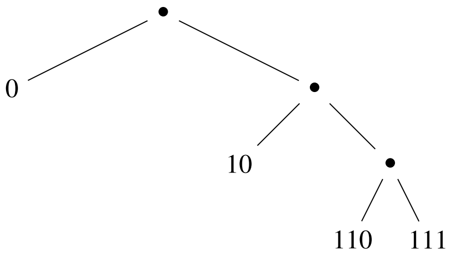


Definition II.11. *Let C be a code. Then $G(C)$ is the graph (V,E), where V is the set of prefixes of codewords and*

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

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

Example II.12. *Determine the tree associated to the code $C = \{0,10,110,111\}$.*



Definition II.13. *Let G be a tree. A leave of G is a vertex which is adjacent to at most one vertex of G.*

Theorem II.14. *Let C be a code and let $G(C) = (V,E)$ be the graph associate to C.*

(a) Let $c \in V$ be of length n. For $0 \le k \le n$ let c_k be the prefix of length k of c. Then $(c_0,c_1,\dots,c_n)$ is a path from $\varnothing$ to c in $G(C)$.

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

(c) Let c be a codeword. Then c is prefix of another codeword if and only if c 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) Just observe that $c_i = c_{i-1}t_i$ where t_i is the i-symbol of c. Hence c_{i-1} is adjacent to c_i and $(c_0,\dots,c_n)$ is a path. 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:

1°. *Let $(a_0,a_1\dots,a_m)$ be path in $G(C)$ with pairwise distinct vertices of length at least 1. Let l_i be length of a_i and suppose that $l_0 \le l_1$. Then for all $1 \le i \le m$, a_{i-1} is the parent of a_i. In particular, $l_i = l_0 + i$ for all $0 \le i \le m$ and a_i is a prefix of a_j for all $0 \le i \le j \le m$.*

Since a_0 is adjacent to a_1, either one of a_0 and a_1 is the parent of the other. Since $l_0 \le l_1$ we conclude that a_0 is parent of a_1. So if $m = 1$, (1°) holds. Suppose $m \ge 2$. Since $a_2 \ne a_0$ and a_0 is the parent of a_1, 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, $l_1 \le l_2$. Induction applied to the path $(a_1,\dots,a_m)$ shows that a_{i-1} is a parent of a_i for all $2 \le i \le m$ and so the first statement in
rf 1 is proved.

The remaining statements follow from the first.

Now suppose for a contradiction that there exists a simple circle $(v_0,v_1,\dots,v_n)$ in $G(C)$ with $n \ge 3$. Let $l = \min_{0\le i\le n} \mathrm{l}(v_i)$ and choose $0 \le k \le n$ such that $\mathrm{l}(v_k) = l$.

Assume that $0 < k < n$. Then we can apply (1°) to the paths $(v_k,\dots,v_n)$ and $(v_k,v_{k-1},\dots v_0)$. It follows that v_{k+1} is the prefix of length $l+1$ of v_n and v_{k-1} is the prefix of length $l+1$ of v_0. Since $(v_0,\dots,v_n)$ is a circle we have $v_0 = v_n$ and we conclude that $v_{k-1} = v_{k+1}$. Thus implies $k-1 = 0$ and $k+1 = n$. So $n = 2$, a contradiction.

Assume next that $k = 0$ or $k = n$. Since $v_0 = v_n$ we conclude that v_0 and v_n have length l. By (1°) applied to (v_0, v_1), v_1 has length $l+1$ and by (1°) applied to $(v_n, v_{n-1}, \dots, v_1)$. v_1 has length $l + (n-1)$. Thus $n-1 = 1$ and again $n = 2$.

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.

(c): Suppose first that a, b are distinct codewords and a is the prefix of b. Let $l = \mathrm{l}(a)$. Let c be the prefix of length $l+1$ of b. Then a is the parent of c. Since c is prefix of the codeword b, $c \in V$. So c is adjacent to a. By definition of a code, $a \neq \varnothing$ and so a has a parent d. Then d is in V, d is adjacent to a and $d \neq c$. So 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. Since c is a V, c is the prefix of some codeword b. Since a is the parent of c, a is a prefix of b and $a \neq b$. So a is the prefix of another codeword.

(d) C is prefix free if and only if no codeword is a prefix of another codeword and so by (a) if and only if no codewords is adjacent two different vertices of $G(C)$, that is very codewords is adjacent to at most one vertex and so if and only if each codeword is a leave. □

II.3 The Kraft-McMillan number

Definition II.15. *Let C be a code using the alphabet T.*

(a) C is called a b-ary code, where $b = |T|$.

(b) Let i be a non-negative integer. Then C_i is the set of codewords of length i and $n_i = |C_i|$. So n_i is the number of codewords of length i.

(c) Let M be the maximal length of a codeword. (So $n_M \neq 0$ and $n_i = 0$ for all $i > M$). The M-tuple $n = (n_0, n_1, \dots, n_M)$ is called the parameter of C. More generally if $N \geq M$ we will also call $(n_0, n_1, \dots, n_N)$ the parameter of C.

(d) The number

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

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

Example II.16. *Compute the Kraft-McMillan number of the binary code $C = \{10, 01, 11, 011\}$.*

We have $C_0 = \{\}$, $C_1 = \{\}$, $C_2 = \{10, 01, 11\}$ and $C_3 = \{011\}$. So the parameter is $(0, 0, 3, 1)$ and

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

Lemma II.17. *Let C be a code with Kraft-MacMillan number K using the alphabet T. Let M be a positive integer such that every code word has length less than N. Let D be the set message of length M in T which have a codeword as a prefix. Then* *(a)* $|D| \le Kb^M$.

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

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

Proof. (a) Let c be a codeword of length i. Then any message of length M in T is of the from cr where r is a message of length $M - i$. Note that where are b^{M-i} such messages and so there are exactly b^{M-i} message of length M which have c as a prefix. It follows that there are at most $n_i b^{M-i}$ message of length M which have a codeword of length i as a prefix. Thus

$$|D| \le \sum_{i=0}^{M} n_i b^{M-i} = b^m \sum_{i=1}^{M} n_i b^{-i} = b^M K$$

(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) Note that the numbers of message of length M is b^M. Thus $|D| \le b^M$ and so by (b), $b^M K \le b^M$. Hence $K \le 1$. □

Theorem II.18. *Let b and M be integers with* $M \ge 0$ *and* $b \ge 1$. *Let* $n = (n_0, n_1, \ldots, n_M)$ *be tuple of non-negative integers such that* $K \le 1$, *where K is the Kraft-McMillan number associated to parameter n and the base b. Then there exists a b-ary PF code C with parameter n.*

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

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

Suppose next that $M \ge 1$ and that the theorem holds for $M - 1$ in place of M.

Put $\tilde{K} = \sum_{i=0}^{M-1} \frac{n_i}{b^i}$. Then

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

and so

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

By the induction assumption there exists a PF code $\tilde{C}$ with parameter $(n_0, n_1, \ldots, n_{M-1})$. Let $\tilde{D}$ be the set of messages of length M in T which have a codeword from $\tilde{C}$ as a prefix. By II.17 $|\tilde{D}| = b^M \tilde{K}$. Multiplying (*) with b^M gives $b^M \tilde{K} \le b^M - n_M$. Thus $|\tilde{D}| \le b^M - n_m$ and so $n_m \le b^M - |D|$. Since b^M is the number of messages of length b^M, $b^M - |D|$ is the number of messages of length M which do not have a codeword as a prefix. Thus there exists a set D of message of length M such that $|D| = n_M$ and none of the messages has a codeword from $\tilde{C}$ as a prefix. Put $C = \tilde{C} \cup D$.

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

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

Suppose that a and b are both in D. Then$a$ and b have the same length and so a is not a prefix of b.

Suppose that $a \in \tilde{C}$ and $b \in D$. Then the choice of D shows that a is not a prefix of b.

Suppose that $a \in D$ and $b \in \tilde{C}$. a has larger length than b 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 D and a has length M. Thus the parameter of C are $(n_0, n_1, \ldots, n_{M-1}, n_M)$ and so C has all the required properties. □

II.4 A counting principal

Lemma II.19. *Let $c : S \to T^*$ and $d : R \to T^*$ be codes. Define the function $cd : S \times R \to T^*$ by*

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

for all $s \in S, r \in R$. Also let C and D be the set of codewords of c and d, respectively, and define

$$CD = \{ab \mid a \in C, b \in D\}$$

Then cd is a code if and only for each $e \in CD$, there exists unique $a \in C$ and $b \in D$ with $e = cd$. In this case CD is set of codewords of cd.

Proof. Let $s \in S$ and $r \in R$. Since c is a code, $c(s)$ is not the empty message and so also $c(s)d(r)$ is not the empty message. Hence $(cd)(r,s) \neq \varnothing$ for all $(s,r) \in S \times R$.

Thus cd is a code if and only of cd is 1-1. We have

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

In particular, CD is the set of codewords of cd.

Moreover, cd is 1-1 if and only if for each $e \in CD$ there 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 $e = ab$. □

Definition II.20. *Let C be a code with 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.

Example II.21. *Compute the generating function of the code $C = \{01, 10, 110, 1110, 1101\}$.*

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

So

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

Theorem II.22 (The Counting Principal). *(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. For positive integer r define c^r inductively by $c^1 = c$ and $c^{r+1} = c^r c$. Then c^r is a code and

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

Proof. Let $(n_0, \ldots, n_M), (p_0, \ldots, p_U), (q_0, \ldots, q_V)$ be the parameter of c, d and cd respectively.

Let $0 \leq i \leq V$. Let $a \in C$ and $b \in D$. Then ab has length i if and only if a has length j for some $0 \leq j \leq i$ and b has length $i - j$. Given j, there are n_j choices for a and p_{i-j} choices for b. Since cd is a code, a different choice for the pair (a, b) yields a different ab. 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.

Since c is a UD code the extended function $c^* : S^* \to T^*$ is 1-1. Observe that c^r is just the restriction of c^* to S^r. So also c^r is 1-1 and thus c^r is a code. Applying (a) $r-1$ times gives

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

□

II.5 Unique decodability implies $K \leq 1$

Lemma II.23. *Let c be a b-ary code with maximal codeword length M and Kraft McMillan number K. Then*

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

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

Proof. (a) Since there are b^i messages of length i, $n_i \leq b^i$ and $\frac{n_i}{b_i} \leq 1$. So each summand in the Kraft McMillan number is bounded by 1. Note that there are $M+1$ summand and so $K \leq M+1$. If c is regular, $\varnothing$ is not a codeword and $n_0 = 0$. So $K \leq M$ in this case.

(b)

$$K = \sum_{i=0}^{M} \frac{n_0}{b^i} = \sum_{i=0}^{M} n_i \left(\frac{1}{b}\right)^i = Q_c\left(\frac{1}{b}\right)$$

□

Lemma II.24. *(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 c be a UD-code with Kraft McMillan number K. Then the Kraft McMillan number of c^r *is* K^r.

Proof. (a) 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) follows from (a) and induction. □

Theorem II.25. *Let C be a UD-code with Kraft McMillan number K. Then* $K \leq 1$.

Proof. Let M be the maximal codeword length of C. Then the maximal codeword length of C^r is rM and -by II.24(b) - the Kraft McMillan number of C^r is K^r. By II.23 the Kraft-McMillan number is bounded by the maximal codeword length plus 1 and so

$$K^n \leq rM + 1$$

for all $r \in \mathbb{Z}^+$. Thus

$$r \ln K \leq \ln(rM+1) \text{ and } \ln K \leq \frac{\ln(rM+1)}{r}$$

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

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

and so $\ln K \leq 0$ and $K \leq 1$. □

Corollary II.26. *Given a parameter $(n_0, \ldots, n_M)$ and a base b with Kraft-McMillan number K. Then the following statements are equivalent.*

(a) Either $|S| = 1$ and $(n_0, \ldots, n_M) = (1, 0, \ldots, 0)$, or there exists a b-ary UD code with parameter $(n_0, \ldots, n_M)$.

(b) $K \leq 1$.

(c) There exists a b-ary PF code with parameter $(n_0, \ldots, n_M)$.

Proof. By II.25, (a) implies (b). By II.18 (b) implies (c). Suppose c is a b-ary PF-code with parameter $(n_0, \ldots, n_M)$. If c is regular, II.4 shows that c is UD and (a) holds. If c is not regular, then $\varnothing$ is a codeword and since c is prefix free, $\varnothing$ is the only codeword. So the parameter of c is $(1, 0, \ldots, 0)$ and again (a) holds. □

Chapter III

Economical coding

III.1 Probability distributions

Definition III.1. *Let S be an alphabet. Then a probabilty distribution on S is an S-tuple with coefficients in the interval $[0,1]$ such that*

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

Notation III.2. *Suppose S is an alphabet with exactly m-symbols $s_1, s_2, \dots, s_m$ and that*

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

is an S-tuple.

Then we will denote t by $(t_1, \dots, t_m)$. Note that this is slightly ambiguous, since t does not only depended on the n-tuple $(t_1, \dots, t_m)$ but also on the order of the elements $s_1, \dots, s_m$.

Example III.3.

(a)

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

is a probability distribution on $\{w, x, y, z\}$.

(b) Using the notation from III.2 Example (b) can be stated as: $p = \left(\frac{1}{2}, \frac{1}{3}, 0, \frac{1}{6}\right)$ is a probability distribution on $\{w, x, y, z\}$. Note that $p_x = \frac{1}{3}$.

(c) $p = \left(\frac{1}{2}, \frac{1}{3}, 0, \frac{1}{6}\right)$ is a probability distribution on $\{x, w, z, y\}$. Note that $p_x = \frac{1}{2}$.

(d)

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.

(e) Let S be a set with m elements and put $p = \left(\frac{1}{m}\right)_{s\in S}$. Then $p_s = \frac{1}{m}$ for all $s \in S$ and p is probability distribution on S. p is called the equal probability distribution on S.

III.2 The optimization problem

Definition III.4. *Let $c : S \to T^*$ be a code and p a probability distribution on S.*

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

(b) The average length 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 the depends of L on p and c we will sometimes us the notations $L(c)$ and $L_p(c)$ for L

Note that the average codeword length only depends of the length of the codewords with non-zero probability. So we will often assume that p is positive, that is $p_s > 0$ for all $s \in S$.

Example III.5. *Compute L if $S = \{s_1, s_2, s_3\}$, $p = (0.2, 0.6, 0.2)$ and c is the binary code with*

$$s_1 \to 0, \quad s_2 \to 10, \quad s_3 \to 11$$

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

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

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

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

$$s_1 \to 01, \quad s_2 \to 0, \quad s_3 \to 11$$

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

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

Definition III.6. *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 $\mathcal{C}$.

Definition III.7. *Let y be an S-tuple of non-negative integers and b a positive integer.*

(a) Let $M = \max_{s\in S} y_s$ and for $0 \leq i \leq M$ let n_i be the number of $s \in S$ with $y_s = i$. Then $(n_0, \dots, n_m)$ are called the parameter of the codewords length y.

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

Lemma III.8. *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 and $(n_0, \dots, n_M)$ the parameter for the codeword length y.*

(a) Let c be a b-ary code on the set S with codeword length y. Then $(n_0, \dots, n_M)$ is the parameter of c.

(b) K is the Kraft McMillan number of $(n_0, \dots, n_M)$.

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

Proof. (a) 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. So $(n_0, \dots, n_M)$ are the parameter of c.

(b)

We compute

$$K = \sum_{s\in S}^{n} \frac{1}{b^{y_s}} = \sum_{i=0}^{M} \sum_{\substack{s\in S\\ y_s=i}} \frac{1}{b^{y_j}} = \sum_{i=0}^{M} \sum_{\substack{s\in S\\ y_j=i}} \frac{1}{b^i} = \sum_{i=0}^{M} n_i \frac{1}{b^i}$$

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

(c) For $0 \leq i \leq M$ define

$$D_i = \{d \in D \mid d \text{ has length } i\}.$$

Also put

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

Since $(n_0, \dots, n_M)$ are the parameter of c', $|D_i| = n_i$.

By definition of n_i, $|S_i| = n_i$. Thus $|D_i| = |S_i|$ and there exists a 1-1- and onto function $\alpha_i : S_i \to D_i$. Define a code c for S as follows:

Let $s \in S$ and put $i = y_s$. Then $s \in S_i$ and we define $c(s) = \alpha_i(s))$. Since $\alpha_i(s) \in D_i$, $\alpha_i(s)$ has length $i = y_s$. So y_s is exactly the length of $c(s)$ and thus y is the codeword length of c. By construction, the set of codewords of c is C. □

Lemma III.9. *Let S be an alphabet, b a positive integer and y an S-tuple with coefficients in $\mathbb{Z}^+$. Let K be the Kraft-McMillan number of the codeword length y. Then there exists a b-ary PF code with codeword length y if and only if $K \leq 1$.*

Proof. This follows from III.8 and II.18. □

In view of the preceding lemma, the optimization problem for b-ary UD codes with respect to the probability distribution $(p_1, \dots, p_m)$ can be restated as follows:

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

$$p_1y_1 + p_2y_2 + \dots + p_my_m$$

is minimal subject to

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

III.3 Entropy

Notation III.10. *(a) Let S, I and J be sets, $f : I \to J$ be a function and t an S tuple with coefficients in I. Then $f(t)$ denotes the S-tuple $\big(f(s)\big)_{s\in S}$ with coefficients in J.*

(b) Let S be alphabet. Then $\mathbb{1}_S$ is the S tuple all of whose coefficients equal to 1. *We will often just write $\mathbb{1}$ for $\mathbb{1}_S$. $\mathbb{1}$ is called the all one tuple on S.*

Note that if t is an S-tuple of real numbers, $\mathbb{1}t = \sum_{s\in S} t_s$. Also $\frac{1}{b^y} = \left(\frac{1}{b^{y_s}}\right)_{s\in S}$ and so we have

$$\mathbb{1} \cdot \frac{1}{b^y} = \sum_{s\in S} \frac{1}{b^{y_s}}.$$

This notation is elegant, but can be confusing. So we will use it sparsely.

Theorem III.11. *S be an alphabet, p a probability distribution on S and b a real number larger than* 1. *Let y be an S tuple of real numbers with $\sum_{s\in S} \frac{1}{b^{y_s}} \leq 1$.*

Then

$$\sum_{\substack{s\in S\\ p_s\neq 0}} p_s \log_b\left(\frac{1}{p_s}\right) \leq \sum_{s\in S} p_s y_s$$

with equality if and only if $p_s \neq 0$ and $y_s = \log_b\left(\frac{1}{p_s}\right)$ for all $s \in S$.

Moreover, $\sum_{s\in S} \frac{1}{b^{y_s}} = 1$ in case of equality.

Proof. Let $\tilde{S} = \{s \in S \mid p_s \neq 0\}$ and let $\tilde{p}$ and $\tilde{y}$ be the restriction of p and y to $\tilde{S}$. Suppose we proved the theorem holds for $\tilde{p}$ and $\tilde{y}$. Then the inequality also holds for p and y. In case of equality we get

$$1 = \sum_{s\in\tilde{S}} \frac{1}{b^{y_s}} \leq \sum_{s\in S} \frac{1}{b^{y_s}} \leq 1$$

Since $\frac{1}{b^{y_s}} > 0$ for all $s \in S$, this implies $S = \tilde{S}$ and so the theorem also holds for p and y.

We may assume from now on that $p_s \neq 0$ for all $s \in S$. We will present two proofs for this case.

Proof 1 (Using Lagrange's Multipliers)

If $\sum_{s\in S} \frac{1}{b^{y_s}} < 1$ we can replace one of the y_s by a smaller real number to get an S-tuple y^* with

$$\sum_{s\in S} \frac{1}{b^{y^*_s}} = 1 \quad \text{and} \quad \sum_{s\in S} p_s y^*_s < \sum_{s\in S} p_s y_s.$$

So we only have to consider y's with $\sum_{s\in S} \frac{1}{b^{y_s}} = 1$. Since $\sum_{s\in S} p_s \cdot y_s \to \infty$ as $\sum_{s\in S} y_s^2 \to \infty$, $\sum_{s\in S} p_s \cdot y_s$ must obtain a minimum value on the region $\sum_{s\in S} \frac{1}{b^{y_s}} = 1$. From Calculus we know that an extrema for the function $f(y)$ subject to the condition $g(y) = k$, k a constant, occurs at a point where $\nabla f = \lambda\nabla g$ for some $\lambda \in \mathbb{R}$. Since

$$\frac{\partial}{\partial y_s}\left(\sum_{s\in S} p_s y_s\right) = p_s \quad \text{and} \quad \frac{\partial}{\partial y_s}\left(\sum_{s\in S} \frac{1}{b^{y_s}}\right) = (-\ln b)\frac{1}{b^{y_s}}$$

we see that minimum occur at points y such that $p_s = \mu\frac{1}{b^{y_s}}$ for some all $s \in S$ and some $\mu \in \mathbb{R}$. Since $\sum_{s\in S} y_s = 1$ and $\sum_{s\in S} \frac{1}{b^{y_s}} = 1$ we get $\mu = 1$ and so $p_s = \frac{1}{b^{y_s}}$. Thus $y_s = \log_b\left(\frac{1}{p_s}\right)$ and the minimum value is

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

Proof 2 (Using basic Calculus to show that $\ln x \leq x - 1$)

We will first show that

1°. *Let x be a positive real number. Then* $\ln x \leq x - 1$ *with equality if and only* $x = 1$

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

2°. *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* $y_s = \ln\left(\frac{1}{p_s}\right)$.

Using $x = \frac{1}{p_s b^{y_s}}$ in (1°) 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 of $\frac{1}{p_s b^{y_s}} = 1$, that is $y_s = \log_b\left(\frac{1}{y_s}\right)$.

Hence

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

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

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

and so (2°) holds.

Summing (2°) over all $s \in S$ gives

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

with equality if and only of $y_s = \log_b\left(\frac{1}{p}\right)_s$ for all $s \in S$.

Since $\sum_{s\in S} \frac{1}{b^{y_s}} \leq 1$ and $\sum_{s\in S} p_s = 1$ we get

$$\sum_{s\in S} p_s \log_b\left(\frac{1}{p_s}\right) - \sum_{s\in S} p_s y_s \leq 0$$

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 $y_s = \log_b\left(\frac{1}{y_s}\right)$, $\frac{1}{b^{y_s}} = p_s$ and so $\sum_{s\in S}\frac{1}{b^{y_s}} = 1$. Thus

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

with equality if and only if $y_s = \log_b\left(\frac{1}{p}\right)_s$ for all $s \in S$. □

Definition III.12. *Let p be a probability distribution on the set S and b a positive integer. 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* $\log(a) = \log_2(a)$

We will usually interpret the undefined expression $0\log_b\left(\frac{1}{0}\right))$ as 0 and just write

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

Example III.13. *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

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

III.4 Optimal codes – the fundamental theorems

Theorem III.14 (Fundamental Theorem, Lower Bound). *Let p be probability distribution on the set S and c a b-ary PF code for S. Let y the codewords length of c and L the average codeword length of c with respect to p. Then*

$$H_b(p) \leq L$$

with equality if and only if $p_s \neq 0$ and $y_s = \log_b\left(\frac{1}{p_s}\right)$ for all $s \in S$.

Proof. Let K be the Kraft-McMillan number of c. By II.25, $K \leq 1$ and by III.8 $K = \sum_{s\in S}\frac{1}{b^{y_s}}$. Thus $\sum_{s\in S}\frac{1}{b^{y_s}} \leq 1$ and the conditions of III.11 are fulfilled. Since $L = p\cdot y$ and $H_b(p) = p\cdot\log_b\left(\frac{1}{y}\right)$ the theorem now follows from III.11. □

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 $y_s = \left\lceil \log_b\left(\frac{1}{p_s}\right)\right\rceil$, where $\lceil r\rceil$ is the smallest integer larger or equal to r. In our compact notation this means $y = \left\lceil \log_b\left(\frac{1}{p}\right)\right\rceil$.

Definition III.15. *Let c be a b-ary code for the set S with codeword lengths y and let p be positive probability distribution on S. c is called a b-ary Shannon-Fano (SF) code 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 III.16 (Fundamental Theorem, Upper Bound)**.** *Let S be a set, p a probability distribution on S and $b > 1$ and integer.*

(a) Let c be any b-ary SF code. Then p.

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

(b) There exists b-ary SF code with respect to p, unless p is not positive and $\log_b\left(\frac{1}{p_s}\right)$ is an integer for all $s \in S$ with $p_s \neq 0$.

(c) Suppose $\log_b\left(\frac{1}{p_s}\right)$ is an integer for all $s \in S$ with $p_s \neq 0$. Then there exists a b-ary PF code c for S such that

$$L_p(c) \leq H_b(p) + \min_{\substack{s \in S \\ p_s \neq 0}} p_s$$

In particular, $L_p(c) < H_b(p) + 1$ unless $p_s = 1$ for some $s \in S$.

Proof. Put $\tilde{S} = \{s \in S \mid p_s \neq 0\}$.

(a) Let $s \in \tilde{S}$. Then $y_s = \left\lceil \log_b\left(\frac{1}{p_s}\right)\right\rceil$ and $y_s < \log_b\left(\frac{1}{p_s}\right) + 1$. Thus

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

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

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

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

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

Suppose now that $\tilde{K} < 1$ or $S = \tilde{S}$. Then we can choose $y_s \in \mathbb{Z}^+$ for $s \in S \setminus \tilde{S}$ such that

$$K := \sum_{s \in S} \frac{1}{b^{y_s}} = \tilde{K} + \sum_{s \in S \setminus \tilde{S}} \frac{1}{b^{y_s}} \leq 1$$

Let $(n_0, \ldots, n_M)$ be the parameters for the codeword length $y = (y_s)_{s \in S}$. By III.8, K is the Kraft McMillan number of $(n_0, \ldots, n_M)$. Since $K \leq 1$, II.18 shows that there exists a b-ary PF code with parameter $(n_0, \ldots, n_M)$. Thus by III.8 there also exists b-ary PF code c for S with codewords length y. Then c is a SF code and so (b) holds.

Suppose that $\tilde{K} = 1$ and $S \neq \tilde{S}$. Then p is not positive. Also equality holds in (**) and so also in (*). Thus $\log_b\left(\frac{1}{p_s}\right) = y_s$ is an integer for each $s \in \tilde{S}$. So the exceptional case in (b) holds.

(c) Suppose that $\log_b\left(\frac{1}{p_s}\right)$ is an integer for all $s \in \tilde{S}$. We will use a slight modification of the arguments in (b):

Choose $t \in \tilde{S}$ with p_t minimal. Define $y_s = \log_b\left(\frac{1}{p_s}\right)$ if $s \in \tilde{S}$ with $s \neq t$ and $y_t = \log_b\left(\frac{1}{p_t}\right) + 1$. Then (**) is a strict inequality again. So we can choose $y_s, s \in S \setminus \tilde{S}$ and a b-ary PF-code with codeword lengths y, just as in (b). Moreover,

$$L_p(c) = \sum_{s\in\tilde{S}} p_s y_s = p_t + \sum_{s\in\tilde{S}} p_s \log_b\left(\frac{1}{p_s}\right) = p_t + H_b(p)$$

and (c) is proved. □

Example III.17. *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* $L < H_2(p) + 1$.

We have

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

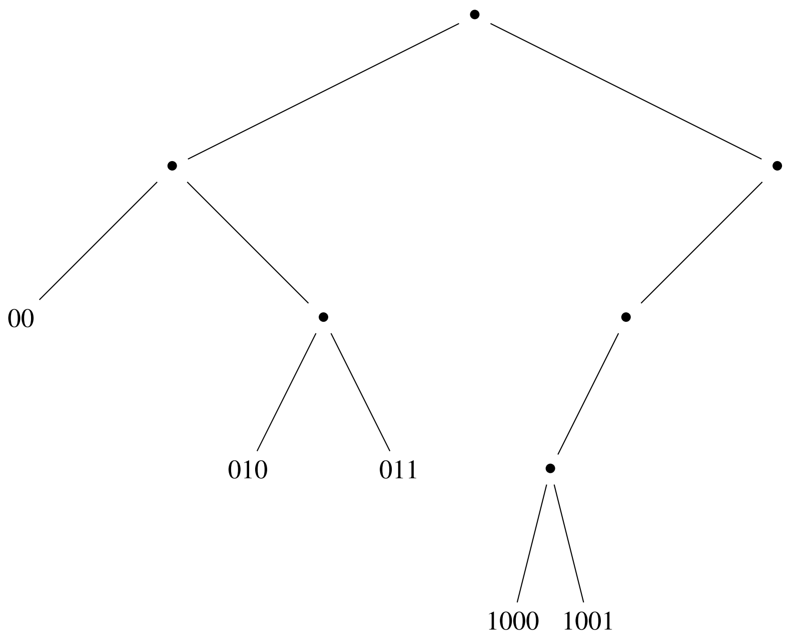
$$\log_2\left(\frac{1}{p}\right) \approx (3.3, 1.3, 2.3, 3.3, 2.3)$$

and so

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

Hence there are zero codewords of length 1, one length 2, two of length 3 and 2 of length 4.

We now use the tree method to construct a tree with these parameters



Since $y = (4,2,3,4,3)$ we can choose the code c as follows

$$(1000, 00, 010, 1001, 011)$$

We have

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

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

Theorem III.18 (Fundamental Theorem)**.** *Let p be a probability distribution on the set S and c an optimal b-ary PF code for S with respect to p. Let L be the average codeword length of c. Then*

$$H_b(p) \le L \le H_b(p) + 1$$

Moreover, the second inequality is strict unless $|S| > 1$ and there exists $s \in S$ with $p_s = 1$.

Proof. By III.14, $H_b(p) \le L$.

By III.16 there exists a b-ary PF code d with $L_p(d) \le H_p(d) + 1$, where the inequality is strict, unless $|S| > 1$ and there exists $s \in S$ with $p_s = 1$. Since c is optimal we have $L \le L_p(d)$ and so the Theorem is proved. □

III.5 Hufman rules

Lemma III.19. *Let p be a positive probability distribution on the set S and c on optimal b-ary PF-code for S with respect to p. Suppose $|S| \ge 2$.*

(a) Let y be the codeword length of c. If $d, e \in S$ with $y_d < y_e$, then $p_d \ge p_e$.

(b) Among the codewords of maximal length there exist two which have the same parent.

Proof. (a): Suppose that $p_d < p_e$. Then

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

Hence $p_e y_e + p_d y_d > p_e y_d + p_d y_e$ and so interchanging the codewords for d and e gives a code with smaller average length, a contradiction.

(b): Let M be the maximal length of a codeword.

Suppose that $M = 1$. Then any codeword has length 1 and so has the empty message $\varnothing$ as its parent. So any two codewords are codewords of maximal length and have the same parent. Since $|S| \ge 2$ we see that that (b) b holds in this case.

Suppose next tat $M > 1$ and by way of contradiction, assume that no two codewords of length M have the same parent. Let x be a codeword of maximal length, let y be the parent of x and let z be any codeword. Suppose that y is a prefix of z. If $z = y$,then z would be a prefix x and $x \ne z$, a contradiction. Thus $z \ne y$ and since y has length $M - 1$, we conclude that z has length at least M. Since M is the maximal codeword length this means that z has length M and y is the parent of z. Since no two codewords of maximal length have the same parent, this means $z = x$.

We proved that x is the only codeword with y as a prefix. Hence, replacing each codewords of maximal length by its parent, gives a PF code of smaller average codeword length than c. Since c is an optimal code, this is a contradiction. So (b) is proved. □

Theorem III.20. *Let p a positive probability distribution on the alphabet S with $|S| \geq 2$. Let d, e be distinct symbols in S such that $p_d \leq p_e \leq p_s$ for all $s \in S$ with $s \neq d$. Let $\tilde{S} = S \setminus \{d\}$ and let $\tilde{c} : \tilde{S} \to \mathbb{B}^*$ be a binary PF-code.*

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

$$(H1) \qquad \tilde{p}_{\tilde{s}} = \begin{cases} p_{\tilde{s}} & \text{if } \tilde{s} \neq e \\ p_d + p_e & \text{if } \tilde{s} = e \end{cases}$$

Also define the code $c : S \to T^*$ by

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

Then

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

(b) $L(c) = L(\tilde{c}) + \tilde{p}_e$, where $L(x)$ is the codeword length of a code x.

(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) Since $\tilde{c}$ is prefix free it is readily verified that also c is prefix free.

(b) Put $\hat{S} = S \setminus \{d, e\}$. Then $\tilde{S} = \hat{S} \cup \{e\}$ and $S = \hat{S} \cup \{d, e\}$. Let y and $\tilde{y}$ be the codeword length for c and $\tilde{c}$, respectively. Then

$$y_s = \tilde{y}_s \text{ and } p_s = \tilde{p}_s \text{ for all } s \in \hat{S}$$

and

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

Hence

$$\begin{aligned} L(c) &= \textstyle\sum_{s \in S} p_s y_s \\ &= p_d y_d + p_e y_e + \textstyle\sum_{s \in \hat{S}} p_e y_e \\ &= p_d(\tilde{y}_e + 1) + p_e(\tilde{y}_e + 1) + \textstyle\sum_{s \in \hat{S}} p_s y_s \\ &= (p_d + p_e) + (p_d + p_e)\tilde{y}_e + \textstyle\sum_{s \in \hat{S}} p_s y_s \\ &= \tilde{p}_e + \tilde{p}_e \tilde{y}_e + \textstyle\sum_{s \in \hat{S}} \tilde{p}_s \tilde{y}_s \\ &= \tilde{p}_e + \textstyle\sum_{s \in \tilde{S}} \tilde{p}_s \tilde{y}_s \\ &= \tilde{p}_e + L(\tilde{c}) \end{aligned}$$

So (b) holds.

(c) $\Longrightarrow$: Suppose c is an optimal binary PFcode. Let $\tilde{a}$ be an optimal binary PF code for $\tilde{p}$ and let a be the code constructed from $\tilde{a}$ using rule H2 with $\tilde{a}$ in place of $\tilde{c}$.

Since c is an optimal, $L(c) \leq L(a)$. Hence also $L(c) - \tilde{p}_e \leq L(a) - \tilde{p}_e$. Applying (b) twice (once to c and once to a) gives $L(\tilde{c}) \leq L(\tilde{a})$. Since $\tilde{a}$ is optimal, $L(\tilde{a}) \leq L(\tilde{c})$. Thus $L(\tilde{a}) = L(\tilde{c})$. Since $\tilde{a}$ is optimal, it follows that $\tilde{c}$ is optimal.

(c) $\Longleftarrow$: Suppose $\tilde{c}$ is an optimal binary PF code. Let a be a optimal binary PF code for $\tilde{p}$. Let M be the maximal length of a codeword of a.

Suppose that there exists $f \in \{d, e\}$ with such that $a(f)$ as length less than M. By III.19(b) there exists at least two codewords of length M and so we can choose $g \in \hat{S}$ such that $a(g)$ has length l. Then $a(l)$ has larger length than $a(g)$ and III.19(a) shows that $p_g \leq p_f$. By choices of e and d, $p_f \leq p_g$ and so $p_f = p_g$. Thus interchanging the codewords for f and g does not change the average length of a. We therefore may assume that both $a(d)$ and $a(e)$ have maximal length.

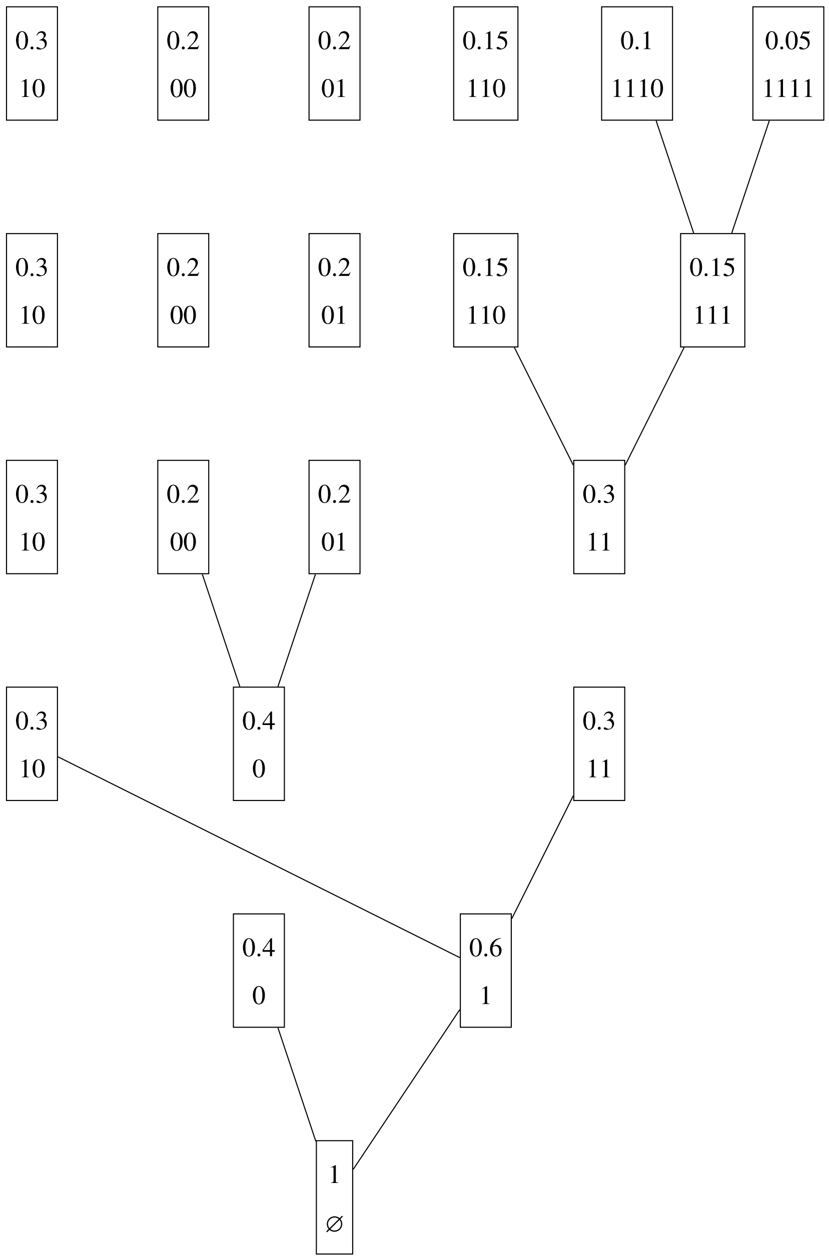
By III.19(b) there exists two codewords of a of maximal length with a common parent. Permuting codewords of equal length does not change the average codeword length. So we may and do choose a such that $a(d)$ and $a(e)$ have a common parent u and that $a(d) = u0$ and $a(e) = u1$. Let $\tilde{a}$ be the code for $\tilde{S}$ defined by

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

Since a is PF, u is not a codeword of a. Hence any codeword of a with u as a prefix must have length M and since a is binary code must be $u0$ or $u1$. Hence u is not the prefix of any $a(s), s \in \hat{S}$. 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 $L(\tilde{a}) = L(\tilde{c})$. Thus also $L(\tilde{a}) + \tilde{p}_e = L(\tilde{c}) + \tilde{p}_e$. From (b) applied to a and c we conclude that $L(a) = L(c)$. Since a is optimal, this means that also c is optimal. □

Example III.21. *Use Hufman's Rules H1 and H2 to construct an optimal code with respect to the probability distribution* $(0.3, 0.2, 0.2, 0.15, 0.1, 0.05)$


0.3
10
0.2
00
0.2
01
0.15
110
0.1
1110
0.05
1111
0.3
10
0.2
00
0.2
01
0.15
110
0.15
111
0.3
10
0.2
00
0.2
01
0.3
11
0.3
10
0.4
0
0.3
11
0.4
0
0.6
1
1
∅

Chapter IV

Data Compression

IV.1 The Comparison Theorem

Theorem IV.1 (The comparison theorem)**.** *Let p and q probability distributions on the alphabet S and let b a real number larger than* 1*. Then*

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

with equality and only if $p = q$.

(If $p_s = q_s = 0$ we interpret $p_s \log_b \left(\frac{1}{q_s} \right)$ as 0*. If $p_s \neq 0$ and $q_s = 0$ we interpret $p_s \log_b \left(\frac{1}{q_s} \right)$ as ∞ and so also $\sum_{s \in S} p_s \log_b \left(\frac{1}{q_s} \right) = \infty$.)*

Proof. If $q_s = 0$ for some $s \in S$ with $p_s \neq 0$, the sum on the right sides is ∞. Thus $H_b(p) < \sum_{s \in S} p_s \log_b \left(\frac{1}{q_s} \right)$ and $q \neq p$ in this case.

So we may assume that $q_s \neq 0$ for all $s \in S$ with $p_s \neq 0$. We also may remove all s from S for which $p_s = q_s = 0$. So $q_s \neq 0$ for all $s \in S$ and we can define $y_s = \log_b \left(\frac{1}{q_s} \right)$. Then $\frac{1}{b^{y_s}} = q_s$ and so

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

Thus by Theorem III.11 $\sum_{s \in S} p_s \log_b \left(\frac{1}{p_s} \right) \leq \sum_{s \in S} p_s y_s$ with equality if and only if $y_s \neq 0$ and $y_s = \log_b \left(\frac{1}{p_s} \right)$ for all $s \in S$. Hence $H_b(p) \leq \sum_{s \in S} p_s \log_b \left(\frac{1}{q_s} \right)$ 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 IV.2. *Let p be a probability distribution on the alphabet S with m symbols. Let $b > 1$. Then*

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

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

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

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

IV.2 Coding in pairs

Lemma IV.3. *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 IV.4. *Let I be an alphabet and p an I-tuple with coefficients in $\mathbb{R}$. Suppose that $p_i \geq 0$ for all $i \in I$ and $t = \sum_{i \in I} p_i \neq 0$. 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 $\frac{p_i}{t} \geq 0$. Also

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

and so by IV.3, $(\frac{p_i}{t})_{i \in I}$ is a probability distribution on I. □

Definition IV.5. *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 $j \in J$. Then f' is called the (first) marginal tuple of f an I. f'' is called the (second) marginal tuple of f on J.

Note that $f' = \sum_{j \in J} \mathrm{Col}_j(f)$ 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.

Lemma IV.6. *Let I and J be alphabets and p be a $I \times J$-matrix with coefficients in $\mathbb{R}^+$ and marginal tuples p' and p''. Then p is a probablity distribution if and only if p' is a probability distribution and if and only if p'' is a probality distribution.*

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

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

If one p, p' and p'' is a probability distributions, then this sum is equal to 1 and so by IV.3 all of p, p', p'' are probability distributions. □

Definition IV.7. *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

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

Lemma IV.8. *Let p' and p'' be probability distributions on I and J, respectively.*

(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 I.

(b) By (a) the marginal tuples of $p' \otimes p''$ are probability distributions and so by IV.6 also $p' \otimes p''$ is a probability distribution.

(c) Follows immediately from the definition of independent. □

Theorem IV.9. *Let I and J be alphabets and p be a probability distribution on $I \times J$ with marginal distributions p' and p''. Then*

$$H(p) \leq H(p') + H(p'')$$

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

Proof. We have

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

Thus the comparison theorem shows that $H(p) \le H(p') + H(p'')$ 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' and p''. □

IV.3 Coding in blocks

Definition IV.10. *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 $\xi_1\xi_2 \ldots, \xi_n, \ldots$ of symbols from S. $P(a_1a_2 \ldots a_n)$ is the probability that $\xi_1 = a_1, \xi_2 = a_2, \ldots, \xi_{n-1} = a_{n-1}$ and $\xi_n = a_n$.

Definition IV.11. *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 $x = x_1x_2 \ldots x_n$ is a string of length n in the alphabet S^r, then each x_i is a string of length rin the alphabet S, and so x is a string of length nr in the alphabet S, so $(S^r)^* = \bigcup_{n=0}^{\infty} S^{nr} \subseteq S^*$.

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

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

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

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

for all $x \in S^r$.

We interpret $p^l(s_1s_2\dots s_r)$ as the probabilty that $\xi_{l_1} = s_1, \xi_{l_2} = s_2, \dots, \xi_{l_r} = s_r$.
Note also that

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

Lemma IV.12. *Let (S,P) be a source, $r \in \mathbb{N}$ and $l = (l_1,\dots,l_r)$ be an 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) Put $t = l_r - r$ (and so $l_r = r + t$) and let $k = (k_1,\dots,k_t)$ be the increasing t-tuple of positive integers with $\{1,\dots,t+r\} = \{l_1,\dots,l_l,k_1,\dots,k_t\}$. Then we can identify S^{t+r} with $S^t \times S^r$ via the map

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

and, after this identification, 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

$$\sum_{x\in S^{r+1}} P(ax) = \sum_{y\in S^r, s\in S} P\big(a(ys)\big) = \sum_{y\in S^r}\left(\sum_{s\in S} P\big((ay)s\big)\right) = \sum_{y\in S^r} P(ay) = P(a)$$

where the third equality holds by the definition of the source and the last by the induction assumption.

(b) Using $a = \varnothing$ in (a) we get $1 = P(\varnothing) = \sum_{x\in S^r} P(\varnothing x) = \sum_{x\in S^r} P(x)$. Hence by IV.3 is probability distribution on S^r.

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

(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 map

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

is inverse to the map

$$S^{t+r} \to (S^t, S^r), d \to (d_k, d_l)$$

Let $y \in S^{t+r}$ and $x \in S^r$. Then $y = \rho(a,b)$ for some $a \in S^t$ and $b \in S^a$. Then $y_l = x$ if and only if $b = x$ and so if and only if $y = \rho(a,x)$ for unique $a \in S^t$. Thus

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

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

(c) By (b), p^{r+t} is a probability distribution on S^{r+t} and by (d), p^l is a marginal tuple of p^{t+r}. So by IV.6 p^l is a probability distribution on S^r. □

Lemma IV.13. *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 by IV.12(a)

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

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

IV.4 Memoryless Sources

Definition IV.14. *A source (S,P) is called memoryless if*

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

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

Lemma IV.15. *Let p be a probability distribution on the alphabet S. Define*

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

inductively by

$$P(\varnothing) = 1$$

and

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

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

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

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

Proof. Since $0 \le p_s \le 1$ for all $s \in S$, also $0 \le P(a) \le 1$ 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.

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 IV.16. *Let (S, P) be a memoryless source and let $r, t \in \mathbb{N}$. Then $p^{t+r} = p^t \otimes p^r$ and so p^t and p^r are independent with respect to p^{t+r}.*

Proof. Let $a = s_1 \dots s_r \in S^t$ and $b = s_{t+1} \dots s_{t+r} \in S^r$. Then

$$\begin{aligned} p^{t+r}(ab) &= p^{t+r}(s_1 \dots s_r s_{r+1} \dots s_{r+t}) \\ &= p_{s_1} \dots p_{s_r} p_{s_{r+1}} \dots p_{t+r} \\ &= \big(p_{s_1} \dots p_{s_r}\big)\big(p_{s_{r+1}} \dots p_{t+r}\big) \\ &= p^t(s_1 \dots s_r) p^r(s_{r+1} \dots s_{r+t}) \\ &= p^t(a) p^r(b) \end{aligned}$$

□

IV.5 Coding a stationary source

Definition IV.17. *A source (S, P) is called stationary if*

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

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

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

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

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

Lemma IV.18. *Let (S, P) be a source. Let $r, t \in \mathbb{N}$.*

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

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

Proof. (a) By IV.12(a),

$$p^t(a) = P(a) = \sum_{b \in S^r} P(ab) = \sum_{b \in S^r} p^{t+r}(ab)$$

Hence p^t is the marginal distribution of p^{t+r} on S^t. The special case $l = (t + 1, r + 2, \dots, t + r)$ and $k = (1, \dots .t)$ in IV.12(d) shows that $p^{(t+1,\dots,t+r)}$ is the the marginal distributions of p^{t+r} on S^r.

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

Lemma IV.19. *A memoryless source is stationary.*

Proof. Let (S,P) be a memoryless source and $r,t \in \mathbb{N}$. By IV.16 $p^{t+r} = p^t \times p^r$ and so by IV.8(c) , p^r is the marginal distribution of p^{t+r} on p^r. By IV.18(a), $p^{(t+1,\dots,t+r)}$ is also the marginal distribution of p^{t+r} on S^r. Thus $p^r = p^{(t+1,\dots,t+r)}$ and P is stationary. □

Theorem IV.20. *Let (S,P) be a source and 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^{r+t}.

(c) Suppose (S,P) is stationary, and r divides t. Then

$$\frac{H_b(p^t)}{t} \leq \frac{H_b(p^r)}{r}.$$

(d) Supppose (S,P) is memoryless. Then

$$\frac{H_b(p^t)}{t} = H_b(p).$$

Proof. By IV.18(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 IV.9

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

Before proving (c) and (d) we first show:

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

For $q = 1$, (*) is obviously true. Suppose now that (*) holds for q. Then by (b)

$$H_b(p^{(q+1)r}) = H_b(p^{qr+r}) \leq H_b(p^{qr}) + H_b(q^r) \leq qH_b(p^r) + H_b(p^r) = (q+1)H_b(p^r)$$

with equality if (S,P) is memoryless. (Note here that by IV.16 p^{qr} and p^r are independent with respect to p^{qr+r} for memoryless sources.) So $(*)$ holds for $q+1$ and hence by the Principal of induction, for all $q \in \mathbb{Z}$.

(c) Since r divides t, $t = qr$ for some $q \in \mathbb{Z}^+$. Thus using (*)

Dividing by qr gives

$$\frac{H_b(p^t)}{t} = \frac{H_b(p^{qr})}{qr} \leq \frac{qH_b(p^r)}{qr} = \frac{H_b(p^r)}{r}$$

.

(d) Suppose (S,P) is memoryless. By (*) applied with $q = t$ and $r = 1$.

$$H_b(p^t) = H_b(p^{t\cdot 1}) = tH_b(p^1) = H_b(p)$$

and so $\frac{H_b(p^t)}{t} = H_b(p)$. □

Definition IV.21. *Let (S, P) be a source and $b > 1$. Then the entropy of P to the base b is the real number*

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

Recall here that the limit inferior of a sequence $(a_n)_{n=1}^{\infty}$ of real numbers is defined as

$$\liminf_{m\to\infty} a_m = \lim_{m\to\infty} \left(\inf_{n=m}^{\infty} a_n \right)$$

Lemma IV.22. *Let (S, P) be a stationary source. Then*

$$H_b(p^n) = \inf_{n=1}^{\infty} \frac{H_b(p^n)}{n}$$

Proof. Let $m \in \mathbb{Z}^+$. If $1 \le n < m$ we have $\frac{H_b(p^{nm})}{nm} \le \frac{H_b(p^n)}{n}$ and so

$$\inf_{n=m}^{\infty} \frac{H_b(p^n)}{n} = \inf_{n=1}^{\infty} \frac{H_b(p^n)}{n}$$

□

Lemma IV.23. *Let (S, P) be a memoryless source and $b > 1$. Then $H_b(P) = H_b(p)$.*

Proof. Just recall that by IV.20(c) we have $\frac{H_b(p^r)}{r} = H_b(p)$ for all $r \in \mathbb{Z}^+$. □

Theorem IV.24 (Coding Theorem for memoryless sources)**.** *Let (S, P) be a memoryless source, b an integer with $b > 1$ and $\epsilon > 0$. Let n be an integer with $n > \frac{1}{\epsilon}$. Then there exists a b-ary prefix-free code for S^n for which the average codeword length L with respect to p^n satisfies*

$$\frac{L}{n} < H_b(p) + \epsilon$$

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

By the Fundamental Theorem, there exists a prefix-free code b-ary code c for S^n with average length L respect to p^n satisfying $L \le H_b(p^n) + 1$. Then

$$\frac{L}{n} \le \frac{H_b(p^n) + 1}{n} = \frac{H_b(p^n)}{n} + \frac{1}{n} < H_b(p) + \epsilon$$

□

Theorem IV.25 (Coding Theorem for Sources)**.** *Let (S, P) be a source, $b > 1$ an integer and $\epsilon > 0$. Then there exists a positive integer n and a b-ary prefix-free code for S^n for which the average codeword length L with respect to p^n satisfies*

$$\frac{L}{n} < H_b(P) + \epsilon.$$

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

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

for all $m \ge r$. Thus for all $m \ge r$ there exists $n \ge m$ with

$$(1) \qquad \frac{H_b(p^n)}{n} < H_b(P) + \frac{\epsilon}{2}$$

Choose an integer m such that $m \geq r$ and $m > \frac{2}{\epsilon}$. Choose $n \geq m$ as in (1). Then

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

By the Fundamental Theorem, there exists a prefix-free code b-ary code c for S^n with average length L respect to p^n satisfying

$$(3) \qquad L \leq H_b(p^n) + 1$$

Combining (1), (2) and (3) we obtain

$$\frac{L}{n} \leq \frac{H_b(p^n) + 1}{n} = \frac{H_b(p^n)}{n} + \frac{1}{n} < H_b(P) + \frac{\epsilon}{2} + \frac{\epsilon}{2} = H_b(P) + \epsilon$$

□

IV.6 Arithmetic codes

Definition IV.26. *Let S be set. Then 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 an alphabet together with an ordering '<' on S.

Example IV.27. *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 $1 \leq i < n$. Indeed we have $s_i < s_j$ if and only if $i < j$.*

Notation IV.28. *In the following,* "Let $S = \{s_1, \dots, s_m\}$ be an ordered alphabet" *will mean that the ordering is given by $s_i < s_j$ if and only if $i < j$.*

Definition IV.29. *Let S be an ordered alphabet and p a positive probability distribution in S. Fix $s \in S$ and define:*

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

where $\alpha = 0$ if s is the smallest element in 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

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

Then $c : S \to \mathbb{B}^*, s \to c(s)$ is called the arithmetic code for S with respect to p.

Example IV.30. *Determine the arithmetic code for the order 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	$2^n\alpha$	c'	$\sum_{i=1}^{n} z_i 2^{n-i}$	c
a	0.1	0	10	4	5	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	0.8	1	$0 \cdot 4 + 0 \cdot 2 + 1$	001
e	0.2	0.4	5	3	4	6.4	7	$0 \cdot 8 + 1 \cdot 4 + 1 \cdot 2 + 1$	0111
b	0.15	0.6	6.3…	3	4	9.6	10	$1 \cdot 8 + 0 \cdot 4 + 1 \cdot 2 + 0$	1010
c	0.25	0.75	4	2	3	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 IV.31. *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 $0_* z$.

Observe that $0_* z \leq \sum_{i=1}^{n} \frac{1}{2^i} < \sum_{i=1}^{\infty} \frac{1}{2^i} = 1$. So $0 \leq 0_* z < 1$.

Lemma IV.32. *Let $n \in \mathbb{Z}^+$ and α a real number such that $0 \leq \alpha < \alpha + \frac{1}{2^n} < 1$.*

(a) There exists a unique $z \in \mathbb{B}^n$ with $0_* z \in [\alpha, \alpha + \frac{1}{2^n})$.

(b) If $c = \lceil 2^n \alpha \rceil$ and $z = z_1 \dots z_n$, then $c = \sum_{i=1}^{n} z_i 2^{n-i}$.

(c) $0_* zy = 0_* z + \frac{1}{2^n} 0_* y \in [\alpha, \alpha + \frac{1}{2^{n-1}})$ for all $y \in \mathbb{B}^*$.

Proof. (a) and (b): Let $z = z_1 \dots z_n \in \mathbb{B}^n$ and put $d = \sum_{i=1}^{n} z_i 2^{n-i}$. Then $d \in \mathbb{N}$ and

$$\frac{d}{2^n} = \sum_{i=1}^{n} \frac{z_i}{2^i} = 0_* z.$$

Hence

$$\begin{aligned} & 0_* z \in [\alpha, \alpha + \tfrac{1}{2^n}) \\ \iff \quad & d \in [2^n \alpha, 2^n \alpha + 1) \\ \iff \quad & 2^n \alpha \leq d \text{ and } d < 2^n \alpha + 1 \\ \iff \quad & d = \lceil 2^n \alpha \rceil \end{aligned}$$

This proves (a) and (b).

(c) Let $y = y_1 y_2 \dots y_m$. Then y_i is the $n+i$-coefficient of zy and so

$$\begin{array}{rcccc}
\alpha \leq 0_* z \leq 0_* zy & = & 0_* z + \sum_{i=1}^{m} \frac{y_i}{2^{n+i}} & = & 0_* z + \frac{1}{2^n} \sum_{i=1}^{m} \frac{y_i}{2^i} \\
 & = & 0_* z + \frac{1}{2^n} 0_* y & < & 0_* z + \frac{1}{2^n} \\
 & < & \left(\alpha + \frac{1}{2^n}\right) + \frac{1}{2^n} & = & \alpha + \frac{1}{2^{n-1}}.
\end{array}$$

Thus (c) holds. □

Theorem IV.33. *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 $I_s = \left[\alpha(s), \alpha(s) + p_s\right)$. 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 all $s \in S$ and $y \in \mathbb{B}^*$.

(c) c is a prefix-free code.

Proof. (a) Let $S = \{s_1, s_2, \dots, s_m\}$ with $s_1 < s_2 < \dots < s_{m-1} < s_m$. 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 $p_{s_i} > 0$ we have

$$0 = \alpha(s_1) < \alpha(s_2) < \dots < \alpha(s_{m-1}) < 1$$

and for each $r \in [0, 1)$ there exists a unique $1 \leq i \leq m$ with $\alpha(s_i) \leq r < \alpha(s_{i+1})$ (if $i < m$) and $\alpha(s_i) \leq r < 1$ (if $i = m$). Then $s = s_i$ is the unique element of S with $r \in I_s$.

(b) Note that $\alpha + \frac{1}{2^n} < \alpha + \frac{1}{2^{n'}} \leq \alpha + p_s \leq 1$ and so we can apply IV.32. It follows that

$$0_* c(s) y \in \left[\alpha, \alpha + \frac{1}{2^{n-1}}\right) = \left[\alpha, \alpha + \frac{1}{2^{n'}}\right) \subseteq \left[\alpha, \alpha + p_s\right) = I_s$$

for all $s \in S$, $y \in \mathbb{B}^*$.

(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 $0_* c(s) y \in I(s)$. Hence by (a) $c(t) \neq c(s) y$ and so $c(s)$ is neither equal to $c(t)$ nor to a prefix of $c(t)$. Hence c is a prefix-free code. □

Theorem IV.34. *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

$$L_p(c) < \sum_{s \in S} p_s \left(\log_2 \left(\frac{1}{p_s} \right) + 2 \right) = \sum_{s \in S} p_s \left(\log_2 \left(\frac{1}{p_s} \right) \right) + \sum_{s \in S} 2 p_s = H_2(p) + 2$$

□

Corollary IV.35 (Coding Theorem for Arithmetic codes)**.** *Let (S, P) be a source with p positive. Let $\epsilon > 0$. Then there exists an integer n with the following property:*

$$\frac{L_{p^n}(c)}{n} < H_2(P) + \epsilon$$

for every arithmetic code c for S^n with respect to p^n.

Moreover, if P is memoryless, any integer $n > \frac{2}{\epsilon}$ has this property.

Proof. Follow the proof for Coding Theorem for Sources (IV.25) with the following modifications:

After Equation (1): Choose m such that $m \geq r$ and $m > \frac{4}{\epsilon}$. So (2) becomes:

$$(2*) \qquad n > \frac{4}{\epsilon}$$

After Equation (2): Let c be an arithmetic code on S^n with respect to p^n. By IV.34 we get

$$(3*) \qquad L_{p^n}(c) < H(p^n) + 2$$

The changes in (2) and (3) 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. □

IV.7 Coding with a dynamic dictionary

Definition IV.36. *An dictionary D based on the alphabet S is a 1-1 N-tuple $D = (d_1, \dots, d_N)$ with coefficients in S^*. (Here 1-1 means D is 1-1 as a function from $\{1, \dots, N\}$ to S^*, that is $d_i \neq d_j$ for all $1 \leq i < j \leq N$).*

Let $d \in S^*$. If $d = d_i$ for some $1 \leq i \leq N$, we say that d appears in D and call i the index of d in D.

Example IV.37. *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 IV.38 (LZW encoding)**.** *Let $S = \{s_1, s_2, \dots, s_m\}$ be an ordered alphabet and X be a non-empty message in S. Define*

- *a positive integer n;*
- *non-empty message Y_k, $1 \leq k \leq n$ in S;*
- *positive integer c_k, $1 \leq k \leq n$;*
- *non-empty messages X_k, $0 \leq k < n$ in S;*
- *symbols z_k, $1 \leq k < n$ in S; and*
- *dictionaries D_k, $0 \leq k < n$, based on S*

inductively as follows:

For $k = 0$ define

$$D_0 = (s_0, s_1 \dots, s_m) \quad \text{and} \quad X_0 = X$$

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}.* [1]
- *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 in S with $X_{k-1} = Y_k X_k$.*
- *z_k is the first symbol of X_k.*
- *$D_k = (D_{k-1}, Y_k z_k)$.* [2]

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 S*

Example IV.39. *Given the ordered set $\{a, b, c, d, e\}$. Determine the LZW encoding of bdddaad.*

$$\begin{array}{c|c|c|c|c|c} b & d & dd & a & a & d \\ 2 & 4 & 7 & 1 & 1 & 4 \end{array}$$

$m+k$	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 247114.

Lemma IV.40. *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 $0 \leq k \leq l < n$.

(d) Y_k is the element of index c_k in D_l for all $k - 1 \leq l < n$.

(e) $y_k z_k$ is the element appearing with index $m + k$ in D_k for all $1 \leq k < n$.

(f) z_k is the first symbol of Y_{k+1} for all $1 \leq k < n$.

Proof. (a) By construction $X_{n-1} = Y_n$. So (a) holds for $k = n - 1$. By construction $X_{k-1} = Y_k X_k$ and (a) holds 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 $k - 1 \leq l < n$, (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. So by (a), z_k is the first symbol of Y_{k+1}. □

Algorithm IV.41 (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 $u \neq \varnothing$ and $c_k < m + k$ for all $1 \leq k \leq n$ define*

[1] Note that all symbols of S appear in D_{k-1} so such a prefix exists and has length at least 1

[2] Note that by maximality of Y_k, $Y_k z_k$ does not appear in D_k. So D_k is a dictionary

⋆ *a message* Y_k, $1 \leq k \leq n$ *in* S*;*

⋆ *symbols* z_k, $1 \leq k < n$ *in* S*; and*

⋆ $m + k$*-tuples* D_k, $0 \leq k < n$, *with coefficients in* S

inductively as follows:

For $k = 0$ *define*

$$D_0 = (s_1 \ldots, s_m)$$

Suppose $k \geq 1$ *and that* D_{k-1} *already has been defined.*

- Y_k *is the the message with index* c_k *appearing in* D_{k-1}.

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}.

  ◇ *If* $c_{k+1} = m + k$*, let* z_k *be the first symbol of* Y_k.[3]

- $D_k = (D_{k-1}, Y_k z_k)$.

Put $e(u) = Y_1 \ldots Y_n$.
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 IV.42. *Find the LZW decoding of the message* 4.7.4.3.9.11 *for the ordered alphabet* $\{b, a, e, f, d, c\}$.

u							4	7	4	3	9	11
Y_k							f	ff	f	e	fe	fef
z_k							f	f	e	f	f	fef
d_{m+k}	b	a	e	f	d	c	ff	fff	fe	ef	fef	
$m + k$	1	2	3	4	5	6	7	8	9	10	11	

So the decoding is fffefefef.

[3] In this case $Y_{k+1} = Y_k z_k$

Chapter V

Noisy Channels

V.1 The definition of a channel

Definition V.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 V.2.

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

is a channel with input alphabet $\{a,b,c,d\}$ and output alphabet $\{a,b,c,d\}$.

Lemma V.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 V.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 V.5. *Let e be the bit error of a binary symmetric channel Γ. Then*

$$\Gamma = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-e & e \\ 1 & e & 1-e \end{array}$$

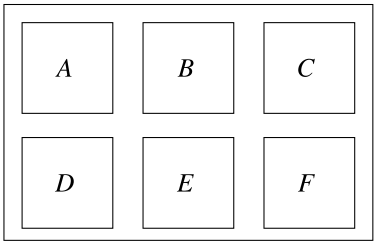
Proof. By definition $\Gamma_{01} = e$. Since $\Gamma_{00} + \Gamma_{01} = 1$, $\Gamma_{00} = 1 - e$. Since Γ is symmetric the sum of each column must be $\frac{|\mathbb{B}|}{|\mathbb{B}|} = 1$. So $\Gamma_{10} = 1 - \Gamma_{00} = e$ and $\Gamma_{11} = 1 - \Gamma_{01} = 1 - e$. □

Notation V.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 V.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

V.2 Transmitting a source through a channel

Definition V.8. *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) $\mathrm{Diag}(p)$ *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 V.9. *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*

$$t = \mathrm{Diag}(p)\Gamma \quad (\text{ so } t_{ij} = p_i\Gamma_{ij} \text{ and } t_i = p_i\Gamma_i)$$

(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)

$$\begin{aligned} & p \text{ is the marginal distribution of } t \text{ on } I \\ \iff & \textstyle\sum_{j\in J} t_{ij} = p_i \text{ for all } i \in I \\ \iff & \textstyle\sum_{j\in J} p_i\Gamma_{ij} = p_i \text{ for all } i \in I \\ \iff & \textstyle\sum_{j\in J} \Gamma_{ij} = 1 \text{ for all } i \in I \text{ with } p_i \neq 0 \\ \iff & \Gamma_i \text{ is a probability distribution on } J \text{ for all } i \in I \text{ with } p_i \neq 0 \end{aligned}$$

So (a) holds.
(b) Follows from (a).
(c)

$$\begin{aligned} & q \text{ is linked to } p \text{ via } \Gamma \\ \iff & q = p\Gamma \\ \iff & q_j = \textstyle\sum_{i\in I} p_i\Gamma_{ij} \text{ for all } j \in J \\ \iff & q_j = \textstyle\sum_{i\in I} t_{ij} \text{ for all } j \in J \\ \iff & q \text{ is the marginal distribution of } t \text{ on } J \end{aligned}$$

□

Definition V.10. *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* $\mathrm{Pr}^{\Sigma}(j|i)$*.*

Assuming that there is no doubt underlying channel system, we will usually drop the superscript Σ*.*

Example V.11. *Compute the channel system for the channel* $\mathrm{BSC}(e)$ *and the probability distribution* $(p, 1-p)$*.*

$$\begin{aligned} t &= \mathrm{Diag}\left(p, 1-p)\right)\mathrm{BSC}(e) \\ &= \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix}\begin{bmatrix} 1-e & e \\ e & 1-e \end{bmatrix} \\ &= \begin{bmatrix} p(1-e) & pe \\ (1-p)e & (1-p)(1-e) \end{bmatrix} \\ &= \begin{bmatrix} p-pe & pe \\ e-pe & 1-p-e+pe \end{bmatrix} \end{aligned}$$

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 V.12. *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 $\mathrm{Diag}(p)^{-1}\Gamma$ is the unique channel associated to associated to t.

Proof. (a) Suppose first that Γ is a channel associated to t. Then $t = \mathrm{Diag}(p)\Gamma$ 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 $t = \mathrm{Diag}(p)\Gamma$. From V.9 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). □

V.3 Conditional Entropy

Definition V.13. *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 V.14. *(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 V.15. $(0, 1, 0, 3, 0.4, 0, 2)$ *is a permutation of* $(0.4, 0.2, 0.3, 0.1)$.

Example V.16. $\mathrm{BSC}(e)$ *is additive with row distribution $E = (e, 1 - e)$.*

Lemma V.17. *(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

$$\begin{aligned} H(t) &= H(p) + H(E) & H(p|q) &= H(p) + H(E) - H(q) \\ H(q|p) &= H(E) & I(p, q) &= H(q) - H(E) \end{aligned}$$

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 IV.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 V.18. *Let* $(\mathrm{BSC}(e), t, p, q)$ *be a channel system. Then*

$$H\big(p|q\big) = H\big(p\big) + H\big((e, 1-e)\big) - H\big(q\big)$$

V.4 Capacity of a channel

Theorem V.19. *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 IV.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 V.20. *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 V.21. *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 V.17

$$I(p,q) = H(q) - H(E)$$

Since $\gamma(\Gamma) = \max_p I(p,q)$ the first equality holds. By IV.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 V.22. *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 V.21 □

Corollary V.23. $\gamma(\mathrm{BSC}(e)) = 1 - H((e, 1-e)) = 1 - e\log\left(\frac{1}{e}\right) - (1-e)\log\left(\frac{1}{1-e}\right)$.

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 V.22. □

Lemma V.24. *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

$$p_i = \sum_{j\in J} t_{ij} \tag{1}$$

Also

$$\frac{p_i}{t_{ij}} = \frac{p_i}{p_i\Gamma_{ij}} = \frac{1}{\Gamma_{ij}}. \tag{2}$$

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 VI

The noisy coding theorems

VI.1 The probability of a mistake

Definition VI.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 VI.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 VI.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

$$\Pr(i|j) \leq \Pr(\sigma(j)|j) \text{ for all } j \in J \text{ with } q_j \neq 0$$

Example VI.4. *Find the Ideal Observer Rule for the channel* $\mathrm{BSC}(0.3)$ *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 VI.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 $j \in J$*

$$\Gamma_{ij} \leq \Gamma_{\sigma(j)j}$$

Since $\Gamma_{ij} = \Pr(j|i)$, this is the same as

$$\Pr(j|i) \leq \Pr(j|\sigma(j))$$

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

Example VI.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

$$M = p_0 M(0) + p_1 M(1) = 0.2 \cdot 0.3 + 0.8 \cdot 0.3 = 0.3$$

VI.2 Fano's inequality

Lemma VI.7. *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*

$$(I \times J) \smallsetminus K = \{(\sigma(j), j) \mid j \in J\},$$

$$M = \sum_{(i,j)\in K} t_{ij} \quad \text{and } 1 - M = \sum_{j\in J} t_{\sigma(j)j}$$

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 VI.8 (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 VI.7 $M = \sum_{(i,j)\in K} t_{ij}$ and $1 - M = \sum_{j\in J} t_{\sigma(j)j}$. Thus

$$\begin{array}{ccccc} H\big((M, 1-M)\big) & = & M \log\left(\frac{1}{M}\right) & + & (1-M)\log\left(\frac{1}{1-M}\right) \\ & = & \sum_{(i,j)\in K} t_{ij} \log\left(\frac{1}{M}\right) & + & \sum_{j\in J} t_{\sigma(j)j} \log\left(\frac{1}{1-M}\right) \end{array}$$

Also by VI.7

$$I \times J = K \uplus \{(\sigma(j), j) \mid j \in J\}.$$

Let $\Delta = \left[\frac{t_{ij}}{q_j}\right]_{\substack{j\in J \\ i\in I}}$. By Exercise 9(f) on Homework 3

$$\begin{aligned} H(\Gamma; p) = H(p \mid q) \qquad & = \sum_{j\in J} q_j H(\Delta_j) \\ = \sum_{j\in J} q_j \left(\sum_{i\in I} \Delta_{ji} \log\left(\frac{1}{\Delta_{ji}}\right)\right) \qquad & = \sum_{j\in J}\sum_{i\in I} q_j \frac{t_{ij}}{q_j} \log\left(\frac{1}{\frac{t_{ij}}{q_j}}\right) \\ = \sum_{(i,j)\in I\times J} t_{ij} \log\left(\frac{q_j}{t_{ij}}\right) \qquad & = \sum_{(i,j)\in K} t_{ij} \log\left(\frac{q_j}{t_{ij}}\right) + \sum_{j\in J} t_{\sigma(j)j} \log\left(\frac{q_j}{t_{\sigma(j)j}}\right) \end{aligned}$$

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 $S_1 \le M\log(|I|-1)$ and $S_2 \le 0$.

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 IV.1

$$\begin{aligned}
0 &\ge \textstyle\sum_{(i,j)\in K} v_{ij} \log\left(\frac{1}{v_{ij}}\right) - \sum_{(i,j)\in K} v_{ij} \log\left(\frac{1}{w_{ij}}\right) \\
&= \textstyle\sum_{(i,j)\in K} v_{ij} \log\left(\frac{w_{ij}}{v_{ij}}\right) \\
&= \textstyle\sum_{(i,j)\in K} \frac{t_{ij}}{M} \log\left(\frac{q_j M}{t_{ij}(|I|-1)}\right) \\
&= \textstyle\sum_{(i,j)\in K} \frac{t_{ij}}{M} \log\left(\frac{q_j M}{t_{ij}}\right) - \sum_{(i,j)\in K} \frac{t_{ij}}{M} \log\left(|I|-1\right)) \\
&= \textstyle\frac{1}{M} S_1 + \log(|I|-1)
\end{aligned}$$

Thus indeed $S_1 \le M\log(|I|-1)$.

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 IV.1

$$\begin{aligned}
0 &\ge \textstyle\sum_{j\in J} u_j \log\left(\frac{1}{u_j}\right) - \sum_{j\in J} u_j \log\left(\frac{1}{q_j}\right) \\
&= \textstyle\sum_{j\in J} u_j \log\left(\frac{q_j}{u_j}\right) \\
&= \textstyle\sum_{j\in J} \frac{t_{\sigma(j)j}}{1-M} \log\left(\frac{q_j(1-M)}{t_{\sigma(j)j}}\right) \\
&= \textstyle\frac{1}{1-M} S_2
\end{aligned}$$

and so indeed $S_2 \le 0$. □

VI.3 A lower bound for the probability of a mistake

Theorem VI.9. *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 VI.8

$$(*) \qquad H(\Gamma; p) \leq H\big((M, 1-M)\big) + M\log(|I|-1) \leq \log 2 + M\log(|I|-1)) < 1 + M\log(|I|)$$

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 IV.2 $H(p) = \log(|I|)$ and so

$$M > \frac{\log(|I|) - \gamma(\Gamma) - 1}{\log|I|} = 1 - \frac{\gamma(\Gamma) + 1}{\log|I|}$$

□

VI.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 VI.10. *(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

$$M^{\otimes 1} = M \quad \text{and } M^{\otimes(n+1)} = M^{\otimes n} \otimes M$$

Example VI.11. *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 VI.12. *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 IV.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 VI.13. *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

$$q_{jj'} = \sum_{i\in I, i'\in I'} p''_{ii'}\Gamma''_{ii',jj'} = \sum_{i\in I, i'\in I'} p_ip'_{i'}\Gamma'_{ij}\Gamma''_{jj'} = \left(\sum_{i\in I} p_i\Gamma_{ij}\right)\left(\sum_{i'\in I'} p'_{i'}\Gamma'_{i'j'}\right) = q_jq'_{j'}$$

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 VI.12(a),

$$\Gamma_{ii'} = \Gamma_i \otimes \Gamma'_{i'}$$

and so (c) follows from IV.9.

(d) By V.24

$$H(q''|p'') = \sum_{i\in I, i'\in I'} p''_{ii'}H(\Gamma_{ii'})$$

and so by (c)

$$H(q''|p'') = \sum_{i\in I}\left(\sum_{i'\in I'} p''_{ii'}\right)H(\Gamma_i) + \sum_{i'\in I'}\left(\sum_{i\in I} p''_{ii'}\right)H(\Gamma'_{i'}) = \sum_{i\in I} p_iH(\Gamma_i) + \sum_{i'\in I'} p'_{i'}H(\Gamma'_{i'})$$

Two more applications of V.24 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'' IV.9 gives

$$(3) \qquad H(q'') \leq H(q) + H(q')$$

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

$$(4) \qquad \begin{array}{rcc} f_{\Gamma''}(p'') & \overset{(2)}{=} & H(q'') - (H(q''|p'')) \\ & \overset{(3)}{\leq} & H(q) + H(q') - (H(t'') - H(q'')) \\ & \overset{(c)}{=} & H(q) + H(q') - \big(H(q|p)\big) + \big(H(q'|p')\big) \\ & = & \big((H(q) - (H(q|p))\big) + \big((H(q') - (H(q'|p'))\big) \\ & \overset{(2)}{=} & f_\Gamma(p) + f_{\Gamma'}(p') \end{array}$$

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 $p \in \mathcal{P}$,

$$(5) \qquad \gamma(\Gamma'') = \max_{p'' \in \mathcal{P}} f_\Gamma(p'') \leq \gamma(\Gamma) + \gamma(\Gamma').$$

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 VI.14. *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). □

VI.5 Coding at a given rate

Definition VI.15. *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 VI.16 (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 VI.9

$$M_i > 1 - \frac{\gamma(\Gamma_i) + 1}{\log(|C_i|)} \tag{1}$$

By C.1 in the Appendix, $\gamma(\Gamma_i) \leq \gamma(\Gamma^{\otimes n_i})$ and by VI.14 $\gamma(\Gamma^{n_i}) = n_i \gamma(\Gamma)$. 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 VI.17 (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. □

VI.6 Minimum Distance Decision Rule

Definition VI.18. *Let $x, y \in \mathbb{B}^n$. Then $d(x,y) = \{i \mid 1 \le i \le n, x_i \neq y_i\}$. $d(x,y)$ is called the Hamming distance of x and y.*

Definition VI.19. *Let $C \subseteq \mathbb{B}^n$ be a code (so C is the set of codewords of a binary code all of whose codewords have the same length). Let $a \in C$ and $z \in \mathbb{B}^n$.*

(a) A decision rule for C is decision rule for $C \times \mathbb{B}^n$, that is a function $\sigma : \mathbb{B}^n \to C$.

(b) Let σ be a decision rule for C. We say that σ is a Minimum Distance rule if for all $a \in C$ and $z \in \mathbb{B}$

$$d(a,z) \ge d(\sigma(z), z)$$

Example VI.20. *Let σ be a Minimum Distance rule for the code*

$$\{0011\,1000, 1100\,0001, 0000\,1110, 1100\,1011\}$$

What is $\sigma(11001001)$?

0011 1000	1100 0001	0000 1110	1100 1011
1100 1001	1100 1001	1100 1001	1100 1001
5	1	5	1

So $\sigma(1100\,1001)$ is either 1100 0001 or 1100 1011.

Lemma VI.21. *Let $\Gamma = \mathrm{BSC}(e)$, $n \in \mathbb{Z}^+$ and $x, y \in \mathbb{B}^n$. Then*

$$\Gamma^{\otimes n}_{xy} = e^d(1-e)^{n-d}$$

Proof. We have

$$\Gamma^{\otimes n}_{xy} = \prod_{k=1}^{n} \Gamma_{x_k y_k}$$

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 $\Gamma^{\otimes n}_{xy} = e^d(1-e)^{n-d}$ □

Lemma VI.22. *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 $d = d(a,z)$ and $d' = d(a',z)$ and $f = \frac{1-e}{e}$. Since $e < \frac{1}{2}$, $1 - e > \frac{1}{2} > e$ and so $f > 1$. We compute

$$\frac{\Gamma_{a'z}}{\Gamma_{az}} = \frac{e^{d'}(1-e)^{n-d'}}{e^{d}(1-e)^{n-d}} = \left(\frac{1-e}{e}\right)^{d-d'} = f^{d-d'}$$

It follow that

$$\Gamma_{az} \le \Gamma_{a'z} \Longleftrightarrow d \ge d'$$

So $\Gamma_{a'z}$ is maximal if and only if d' is minimal. □

Example VI.23. *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 $M_{000} \le 4e^2$ and also $M \le 4e^2$.

Chapter VII

Error Correcting

Definition VII.1. *Let $C \subseteq \mathbb{B}^n$ be a code and σ a decision rule for C. Let $a \in C$ and $z \in \mathbb{B}^n$.*

(a) Let $k \in \mathbb{N}$. (a, z) is called a k-bit error 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 $0 \leq k \leq r$.

Example VII.2. *Given the code $C = \{000, 110, 101, 011\}$ (so C consist of all even messages of length 3) and the decision rule*

$$\sigma : \begin{array}{cccccccc} 000 & 100 & 010 & 001 & 110 & 101 & 011 & 111 \\ \hline 000 & 011 & 000 & 101 & 110 & 101 & 011 & 110 \end{array}$$

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 error 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 not corrected by σ. So σ is not 1-error correcting.

Example VII.3. *Given the code $C = \{000\ 000, 110\ 110, 101\ 101, 011\ 011\}$. (Note that C is obtained by doubling the code in VII.2). Define the decision rule σ for C by*

$$\sigma(xy) = \begin{cases} xx & \textit{if x is even} \\ yy & \textit{if x is odd} \end{cases}$$

for all $x, y \in \mathbb{B}^3$. Show that σ is 1-error correcting.

Let (a, z) be k-bit error for $k \leq 1$. 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 $\sigma(xy) = xx = bb = a$. If $b \neq x$, b must differ in exactly one place from x. So x is odd and $\sigma(xy) = yy = bb = a$. So σ is indeed 1-error correcting.

Lemma VII.4. *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) $d(a,z) \geq r+1$ *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 $d(a,z) \leq r$. 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 $d = d(a,z)$, 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) $M = \sum_{a\in C} p_a M_a \leq \sum_{a\in C} p_a|F(a)|e^{r+1} = \Big(\sum_{a\in C} p_a|F(a)|\Big)e^{r+1}$. □

VII.1 Minimum Distance

Definition VII.5. *Let $C \subseteq \mathbb{B}^n$ be a binary code.*

(a) $\delta = \delta(C) = \min\{d(a,b) \mid a,b \in C, a \neq b\}$. *$\delta$ is called the minimum distance of C.*

(b) For $a,b \in \mathbb{B}^n$ define $D(a,b) = \{i \mid a_i \neq b_i, 1 \leq i \leq n\}$.

Example VII.6. *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

$$\begin{array}{ccc} 111\,000 & 111\,000 & 001\,110 \\ 001\,110 & 110\,011 & 110\,011 \\ 4 & 3 & 5 \end{array}$$

So the minimum distance is 3.

Definition VII.7. *Let A and B be sets. Then*

$$A + B = (A \cup B) \setminus (A \cap B)$$

$A+B$ is called the symmetric difference of A and B.

Lemma VII.8. *Let A and B be sets.*

(a) $A + B = (A \setminus B) \mathbin{\dot\cup} (B \setminus A)$.

(b) $|A+B| = |A \setminus B| + |A \setminus B| = |A| + |B| - 2|A \cap B|$.

Proof. Readily verified. □

Lemma VII.9. *Let $a, b, c \in \mathbb{B}^n$.*

(a) $D(a,c) = D(a,b) + D(b,c)$.

(b) $d(a,c) = d(a,b) + d(a,c) - 2|D(a,b) \cap D(b,c)|$.

(c) $d(a,c) \le d(a,b) + d(b,c)$, with equality if and only if $D(a,b) \cap D(b,c) = \varnothing$.

Proof. (a) Let $1 \le i \le n$.

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 $i \in D(a,b) \setminus D(b,c)$, then $a_i \ne b_i = c_i$ and so $a_i \ne c_i$ and $i \in D(a,c)$.

If $i \in D(b,c) \setminus D(a,b)$, then $a_i = b_i \ne c_i$ and so $a_i \ne c_i$ and $i \in D(a,c)$.

If $i \in D(a,b) \cap D(b,c)$, then $a_i \ne b_i \ne c_i$. 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 of $i \in (D(a,b) \setminus D(b,c)) \cup (D(b,c) \setminus D(a,b))) = D(a,b) + D(b,c)$

(b) and (b):

$$\begin{aligned} d(a,c) &= |D(a,c)| = |D(a,b) + D(b,c)| = |D(a,b)| + |D(b,c)| - 2|D(a,b) \cap D(b,c)| \\ &= d(a,b) + d(b,c) - 2|D(a,b) \cap D(b,c)| \le d(a,b) + d(b,c) \end{aligned}$$

with equality if and only if $|D(a,b) \cap D(b,c)| = 0$, that is if $D(a,b) \cap D(b,c) = \varnothing$. □

Lemma VII.10. *Let $a, b \in \mathbb{B}^n$. Put $d = d(a,b)$ and let $0 \le e \le d$. Then there exists $x \in \mathbb{B}^n$ with $d(a,x) = e$ and $d(x,b) = d - e$.*

Proof. Let J be a subset of $D(a,b)$ with $|J| = e$. Define $x \in \mathbb{B}^n$ by $x_i = b_i$ if $i \in J$ and $x_i = a_i$ if $i \notin J$. Since $a_i \ne b_i$ for all $i \in J$ we have $D(a,x) = J$ and so $d(a,x) = |J| = e$. Since $D(a,x) = J \subseteq D(a,b)$ we conclude from VII.9 that

$$D(x,b) = D(x,a) + D(a,b) = J + D(a,b) = D(a,b) \setminus J$$

and

$$d(x,b) = |D(a,b) \setminus J| = |D(a,b)| - |J| = d - e.$$

□

Definition VII.11. *Let $r, n \in \mathbb{N}$, $x \in \mathbb{B}^n$ and $X \subseteq \mathbb{B}^n$. Then*

$$\mathrm{N}_r(x) = \{y \in \mathbb{B}^n \mid d(x,y) \le r\}$$

and

$$\mathrm{N}_r(X) = \{y \in \mathbb{B}^n \mid d(x,y) \le r \textit{ for some } x \in X\}$$

$\mathrm{N}_r(x)$ *is called the neighborhood of radius r around x.*

Example VII.12. *Compute* $\mathrm{N}_1(0110)$.

$$\mathrm{N}_1(0110) = \{0110, 1110, 0010, 0100, 0111\}$$

Lemma VII.13. *Let $C \subseteq \mathbb{B}^n$ be a code, σ an decision rule for σ. 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 no k-bit error (a,z) with $k \leq r$ is a mistake. This holds if and only if $\sigma(z) = a$ for all $a \in C$ and $z \in \mathbb{B}^n$ with $d(a,z) \leq r$ and so if and only if $\sigma(z) = a$ for all $a \in C$ and $z \in \mathrm{N}_r(a)$. □

Definition VII.14. *A binary code C is called an r-error correcting code if $\delta \geq 2r+1$, that is $d(a,b) \geq 2r+1$ for all $a,b \in C$ with $a \neq b$.*

Theorem VII.15. *Let $C \subseteq \mathbb{B}^n$ and $r \in \mathbb{N}$. Then the following 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 $d(a,z) \leq r$.

(c) $N_r(a) \cap N_r(b) = \varnothing$ for all $a,b \in C$ with $a \neq b$.

(d) Any minimum distance decision rule for C is 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 C has minimal distance at least $2r+1$. Let $z \in \mathbb{B}^n$ and $a,b \in C$ with $d(a,z) \leq r$ and $d(b,z) \leq r$. Then $d(a,b) \leq d(a,z) + d(z,b) \leq r + r = 2r < 2r+1$. Since C is minimal distance at least $2r+1$, $a = b$. So there exists at most one codeword of distance less or equal to r from z.

(b) $\Longrightarrow$ (c): Suppose $N_r(a) \cap N_r(b) \neq \varnothing$ and let $z \in N_r(a) \cap N_r(b)$. Then $d(a,z) \leq r$ and $d(b,z) \leq r$, contradiction to (b).

(c) $\Longrightarrow$ (d): Let σ be minimal distance decision rule and (a,z) be a k-bit error with $k \leq r$. Then $d(a,z) = k \leq r$. Put $b = \sigma(z)$. Since σ is a minimal distance decision rule $d(b,z) \leq d(a,z) \leq r$. Thus $z \in N_r(a) \cap N_r(b)$ and (c) implies that $a = b$. So σ corrects (a,z) and σ is r-error correcting.

(d) $\Longrightarrow$ (e): Obvious.

(e) $\Longrightarrow$ (a): Let $a,b \in C$ with $a \neq b$ and put $d = d(a,b)$. Let $d = 2e + \epsilon$ where $\epsilon \in \{0,1\}$ and $e \in \mathbb{N}$. By VII.10 there exists $z \in \mathbb{B}^n$ with $d(a,z) = e$ and $d(x,b) = d - e = e + \epsilon$. Since $a \neq b$, $\sigma(a) \neq a$ or $\sigma(z) \neq b$. So at least one of (a,z) and (b,z) is a mistake. Hence σ is not $(e+\epsilon)$-error-correcting and so $r < e + \epsilon$. It follows that $r \leq e$, $d = e + (e+\epsilon) > r + r = 2r$ and $d \geq 2r+1$. Thus $\delta(C) \geq 2r+1$ and C is r-error correcting, □

VII.2 The Packing Bound

Lemma VII.16. *Let $r,n \in \mathbb{N}$ with $r \leq n$ and $x \in \mathbb{B}^n$.*

(a) $|\{y \in \mathbb{B}^n \mid d(x,y) = r\}| = \binom{n}{r}$.

(b) $|\mathrm{N}_r(x)| = \sum_{i=0}^{r} \binom{n}{i}$.

Proof. (a) The map

$$y \to D(x,y)$$

is bijection between $\{y \in \mathbb{B}^n \mid d(x,y) = r\}$ and the set of subsets of size r of $\{1,\dots,n\}$.

(b) Note that $\mathrm{N}_r(x)$ is the disjoint union of the sets $\{y \in \mathbb{B}^n \mid d(x,y) = i\}$, $0 \leq i \leq r$. So (a) follows from (b). □

Lemma VII.17. *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

$$\left| \bigcup_{a \in C} \mathrm{N}_r(a) \right| \leq \sum_{a \in C} |\mathrm{N}_r(a)| \tag{1}$$

with equality if and only if $\mathrm{N}_r(a) \cap \mathrm{N}_r(b) = \varnothing$ for all $a, b \in C$ with $a \neq b$. So by VII.15 with equality if and only if C is r-error-correcting.

By VII.16(b)

$$\sum_{a \in C} |\mathrm{N}_r(a)| = \sum_{a \in C} \sum_{i=0}^{r} \binom{n}{k} = |C| \sum_{i=0}^{r} \binom{n}{i} \tag{2}$$

Combining (1) and (2) gives (b). □

Theorem VII.18 (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 so

$$|C| \leq \frac{2^n}{\sum_{o=0}^{r} \binom{n}{i}}$$

Proof. By VII.17

$$|C| \cdot \sum_{i=0}^{r} \binom{n}{i} = |N_r(C)| \leq 2^n$$

□

Definition VII.19. *Let $C \subseteq \mathbb{B}^n$ be a code. Then the information rate $\rho(C)$ of C is the information rate of C relative to $|\mathbb{B}^n$. That is*

$$\rho(C) = \log_{|\mathbb{B}^n|} |C| = \frac{\log_2 |C|}{n}$$

Example VII.20. *Use packing theorem to find an upper bound for the information rate of a 2-error correcting code $C \subseteq \mathbb{B}^n$ of size 100?*

We will first find a lower bound for n.

We need $100\left(1 + n + \binom{n}{2}\right) \leq 2^n$, that is

$$100\left(\frac{2 + 2n + n^2 - n}{2}\right) \leq 2^n$$

and

$$25(n^2+n+2) \leq 2^{n-1}$$

Thus also $25n^2 \leq 2^{n-1}$ and $5n \leq 2^{\frac{n-1}{2}}$. Put $m = \frac{n-1}{2}$. Then $n = 2m+1$ and so $5(2m+1) \leq 2^m$ and

$$2^m - 10m - 5 > 0$$

Consider the function $f(x) = 2^x - 10x - 5$. Then

$$f'(x) = \ln(2)\,2^x - 10 > 0 \text{ if and only if } x > \log_2\left(\frac{10}{\ln 2}\right) \approx 3.85$$

Since $f(0) = 1 - 0 - 5 < 0$ and $f(6) = 64 - 60 - 6 < 0$, $f(x) < 0$ 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 VII.21. *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 $d(z,a) = r$.*

Lemma VII.22. *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 an r-error correcting code and*

$$|C|\sum_{i=0}^{r}\binom{n}{i} = 2^n$$

Proof. By VII.15 C is r-error correcting the following holds for any perfect code and for any r-error correcting code:

(*) For each z in $\mathbb{B}^n$, there exists at most one $a \in C$ with $d(a,z) \leq r$.

Note that this condition holds if C is perfect code. So we may assume that $(*)$ holds and so C is r-error correcting. Then

$$\begin{aligned}
& \text{C is a perfect code} \\
\iff\ & \text{for all } z \in \mathbb{B}^n \text{ there exists a unique } a \in C \text{ with } d(a,z) \leq r \\
\iff\ & \text{for all } z \in \mathbb{B}^n \text{ there exists at least one } c \in C \text{ with } d(c,z) \leq r \quad \text{— by (*)} \\
\iff\ & \mathbb{B}^n = N_r(C) \\
\iff\ & 2^n = |N_r(C)| \\
\iff\ & 2^n = |C|\textstyle\sum_{i=0}^{r}\binom{n}{i} \qquad \text{VII.17(b)}
\end{aligned}$$

□

Definition VII.23. *Let $\delta, n \in \mathbb{N}$ with $\delta \leq n$. Then $A(n,\delta)$ is the largest possible size of code $C \subseteq \mathbb{B}^n$ with minimum distance at least δ. That is*

$$A(n,\delta) = \max_{\substack{C\subseteq\mathbb{B}^n \\ \delta(C)\geq\delta}} |C|$$

Note that the code with minimal distance δ will be $\lfloor \frac{\delta-1}{2} \rfloor$-error correcting. So the packing bound will provide an upper bound for $A(n,\delta)$. But if δ is fairly large this upper bound can be easily improved:

Lemma VII.24. *Let $\delta, n \in \mathbb{N}$ with $\frac{2}{3}n < \delta \le n$. Then $A(n,\delta) = 2$.*

Proof. The code $\{00\dots0, 11\dots1\}$ show that $A(n,\delta) \ge 2$. We will now show that any code C with at least three codewords has minimal distance

$$\delta(C) \le \frac{2}{3}n.$$

So suppose a, b, c are three distinct codewords in C and assume without loss $d(a,b) \ge \frac{2}{3}n$ and $d(b,c) \ge \frac{2}{3}n$. Then

$$n \ge |D(a,b) \cup D(b,c)| = |D(a,b)| + |D(b,c)| - |D(a,b) \cup D(b,c)| \ge \frac{2}{3}n + \frac{2}{3}n - |D(a,b) \cap D(b,c)|$$

and so

$$|D(a,b) \cap D(b,c)| \ge \frac{1}{3}n$$

Since $D(a,c) \cap \big(D(a,b) \cap D(b,c)\big) = \varnothing$ this gives

$$|D(a,c)| \le n - |D(a,b) \cap D(b,c)| \le n - \frac{1}{3}n = \frac{2}{3}n$$

So indeed $\delta(C) \le \frac{2}{3}n$. □

The packing bound can be much higher: Consider $n = 10$ and $\delta = 7$. Then $r = \lfloor \frac{7-1}{2} \rfloor = 3$ and

$$|C| \le \frac{2^{10}}{1 + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}} = \frac{1024}{1 + 10 + 45 + 120} = \frac{1024}{176} < 6$$

So we only get $|C| \le 5$ from the packing bound.

Chapter VIII

Linear Codes

VIII.1 Introduction to linear codes

Definition VIII.1. *$\mathbb{F}_2$ is the set $\mathbb{B} = \{0, 1\}$ together with the following addition and multiplication:*

$+$	0	1
0	0	1
0	1	0

$\cdot$	0	1
0	0	0
0	0	1

So for example in $\mathbb{F}_2$, $1 + 1 = 0$ and $1 \cdot 0 = 0$.

Definition VIII.2. *$\mathbb{F}_2^n$ is $\mathbb{B}^n$ together with the following addition and scalar multiplication*

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \qquad l \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{pmatrix}$$

for all $l, x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{F}_2$.

If $l = 0$, then $lx_i = 0$ for all i and so $0x = \vec{0}$ for all $x \in \mathbb{F}_2^n$. Here $\vec{0} = 00\ldots0$. If $l = 1$, then $lx_i = x_i$ for all i and so $1x = x$ for all $x \in \mathbb{F}_2^n$.

Recall here that we are viewing

$$x_1x_2\ldots x_n, \quad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and} \quad (x_1, x_2, \ldots, x_n)$$

as three different notations for the exact same n-tuple with coefficients in $\mathbb{B}$.

(The book considers these to be different objects. If $x = x_1 \ldots x_n$ is a message in $\mathbb{B}$ then x' denotes the corresponding column vector.)

Example VIII.3. *Compute* 11011 + 10110

We have

$$\begin{array}{rl} & 11011 \\ + & 10110 \\ = & 01101 \end{array}$$

Definition VIII.4. *A subspace of $\mathbb{F}_2^n$ is a subset C of $\mathbb{F}_2^n$ such that*

(a) $\vec{0} \in C$.

(b) $lx \in C$ *for all* $l \in \mathbb{F}_2, x \in \mathbb{F}_2$[1].

(c) $x + y \in C$ *for all* $x, y \in C$.

A subspace of $\mathbb{F}_2^n$ is also called a binary linear code of length n

Definition VIII.5. *Let C be a subspace of $\mathbb{F}_2^n$. A basis for C is a k-tuple*

$$(v_1, \dots, v_k)$$

in C such that for each c in C there exists a unique k-tuple

$$(l_1, \dots, l_k)$$

in $\mathbb{F}_2$ with

$$c = l_1 v_1 + \dots + l_k v_k.$$

Lemma VIII.6. *Let C be a subspace of $\mathbb{F}_2^n$.*

(a) *C has a basis* $(v_1, \dots, v_k)$.

(b) $|C| = 2^k$ *and so* $k = \log |C|$.

(c) $0 \le k \le n$.

Proof. (a) See your favorite linear algebra book.

(b) By definition of a basis and a subspace the map

$$\mathbb{F}_2^k \to C, \quad (l_1, \dots, l_k) \to l_1 v_1 + \dots + l_k v_k$$

is a well-defined bijection. Thus $|C| = |\mathbb{F}_2^k| = 2^k$. (c) Since $1 \le |C| \le |\mathbb{F}_2^n| = 2^n$, this follows from (b). □

Example VIII.7. *Let* $C = \{000\,000, 111\,000, 000\,111, 111\,111\}$. *Find a basis for C and determine the parameters of C.*

Both

$$(111\,000, 000\,111)$$

and

$$(111\,000, 111\,111)$$

are a basis for C and so $\dim C = 2$.

[1] Since $lx = \vec{0}$ or $lx = x$, this condition is redundant

Definition VIII.8. *Let C be a binary linear code of length n. The length k of a basis of C is called the dimension of C and is denoted by* $\dim C$. *The triple* $(n, \dim C, \delta(C))$ *are called the parameters of C.*

Remark VIII.9. *The information rate $\rho(C)$ of a binary linear code C of length n and dimension k is $\frac{k}{n}$.*

Proof. Let $C \subseteq \mathbb{F}_2^n$ be a linear code. Then

$$\rho(C) = \frac{\log_2 |C|}{\log_2 |\mathbb{F}_2^n|} = \frac{\log_2 2^k}{\log_2 2^n} = \frac{k}{n}$$

□

Definition VIII.10. *Let $x = x_1 \dots x_n \in \mathbb{F}_2^n$. Then* $\mathrm{wt}(x) = |\{1 \leq i \leq n \mid x_i \neq 0\}|$ *is called the weight of x.*

Lemma VIII.11. *Let $x, y \in \mathbb{F}_2^n$.*

(a) $d(x,y) = \mathrm{wt}(x-y) = \mathrm{wt}(x+y)$.

(b) $\mathrm{wt}(x) = d(x, \vec{0})$.

(c) Let C be a binary linear code. Then $\delta(C)$ is the minimal weight of a non-zero codeword.

Proof. (a)
$$d(x,y) = |\{1 \leq i \leq n \mid x_i \neq y_i\}| = \{|\{1 \leq i \leq n \mid x_i - y_i \neq 0\} = \mathrm{wt}(x-y)$$

(b) $d(x, \vec{0}) = \mathrm{wt}(x - \vec{0}) = \mathrm{wt}(x)$.

(c) Let w be the minimal weight of a non-zero codeword.

Let $x, y \in C$ with $x \neq y$. Then $x - y \neq \vec{0}$ and so $d(x,y) = \mathrm{wt}(x-y) \geq w$. Since C is subspace of $\mathbb{F}_2^n$, $x - y \in C$ and so $\delta(C) \geq w$.

Let x be non-zero codeword. Since $\vec{0} \in C$, we have $\mathrm{wt}(x) = d(\vec{0}, x) \geq \delta(C)$. Hence $w \geq \delta(C)$ and so $\delta(C) = w$. □

Example VIII.12. *Let* $D = \{000\,000, 111\,000, 000\,111, 111\,111\}$. *Determine the parameters of D.*

The length of C is 6. The dimension of C is $\log_2 |C| =$
$log_2 4 = 2$. The weights of the codewords are $0, 3, 3, 6$. So the minimum non-zero weight is 3. Thus $\delta(C) = 3$ and the parameters are

$$(6, 4, 3)$$

VIII.2 Construction of linear codes using matrices

Definition VIII.13. *(a) Let E be an $n \times k$ matrix over $\mathbb{F}_2$. 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* $\mathrm{Col}(E)$.

(b) Let H be a $m \times n$-matrix. Then

$$\mathrm{Nul}(H) = \{x \in \mathbb{F}_2^n \mid Hx = \vec{0}\}$$

H is called a check matrix for $\mathrm{Nul}(H)$.

Lemma VIII.14. *(a) Let E be an $n \times k$ matrix over $\mathbb{F}_2$. Then* $\operatorname{Col}(E)$ *is a subspace of* $\mathbb{F}_2^n$.

(b) Let H be a $m \times n$-matrix over $\mathbb{F}_2$. Then $\operatorname{Nul}(H)$ *is a subspace of* $\mathbb{F}_2^n$.

Proof. (a) $E\vec{0} = \vec{0}$. So $\vec{0} \in \operatorname{Col}(E)$. Let $a, b \in \operatorname{Col}(E)$. 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 $a + b \in \operatorname{Col}E$.

(b) $H\vec{0} = \vec{0}$ and so $\vec{0} \in \operatorname{Nul}(H)$. Let $a, b \in \operatorname{Nul}H$. The $Ha = \vec{0}$ and $Hb = \vec{0}$. Hence

$$H(a + b) = Ha + Hb = \vec{0} + \vec{0} = \vec{0}$$

and so $a + b \in \operatorname{Nul}H$. □

Example VIII.15. *Find the minimal distance of*

$$\operatorname{Col}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Let x be in the Column space of the matrix. Then

$$x \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ a + b \\ b + c \\ a + c \end{pmatrix}$$

for some $a, b, c \in \mathbb{F}_2$.

If exactly one of a, b and c are non-zero, one of the first three coefficient is non-zero and two of the last three coefficient are non-zero. So $\operatorname{wt}(x) = 3$

If exactly two of the a, b, c are not zero, two of the first three coefficient are non-zero and two of the last three coefficient are non-zero, so $\operatorname{wt}(x) = 4$

If all of a, b, c are non-zero, the first three coefficient are non-zero and the last three coefficient are zero. So $\operatorname{wt}(x) = 3$.

Thus $\operatorname{Col}(E)$ has minimum weight three and so is 1-error correcting.

VIII.3 Standard form of check matrix

Notation VIII.16. *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 $I = \dot\bigcup_{a\in A} I_a$ and $J = \dot\bigcup_{b\in B} J_b$. Then we will view M as an $I \times J$-matrix with*

$$M_{ij} = (M_{ab})_{ij}$$

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 VIII.17. *Given*

$$X_{11} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad X_{12} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

$$X_{21} = \begin{bmatrix} 9 & 10 & 11 \end{bmatrix}, \quad \textit{and} \quad X_{22} = \begin{bmatrix} 12 \end{bmatrix}.$$

Then

$$[X_{11}, X_{12}] = \begin{bmatrix} 1 & 2 & 3 & 7 \\ 4 & 5 & 6 & 8 \end{bmatrix},$$

$$\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 9 & 10 & 11 \end{bmatrix},$$

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 VIII.18. *Let A be a set. Then I_A denotes the $A \times A$ matrix with ab-coefficient equal to*

$$\delta_{ab} = \begin{cases} 1 & \textit{if } a = b \\ 0 & \textit{if } a \neq b \end{cases}$$

I_A is called the $A \times A$ identity matrix. If $n \in \mathbb{N}$, then $\mathrm{I}_n = \mathrm{I}_{\{1,\ldots,n\}}$.

Example VIII.19.

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \textit{and} \quad I_{\{x,y\}} = \begin{array}{c|cc} & x & y \\ \hline x & 1 & 0 \\ y & 0 & 1 \end{array}$$

Lemma VIII.20. *Let A be an $m \times k$ matrix in $\mathbb{F}_2$. Put*

$$E = \begin{bmatrix} I_k \\ A \end{bmatrix} \text{ and } H = [A\ I_m]$$

Then

(a) $\operatorname{Col}(E) = \operatorname{Nul}(H)$.

(b) The columns of E form a basis for $\operatorname{Col}(E)$.

(c) $\operatorname{Col}(E)$ *is a code of length $m + k$ and dimension k.*

Proof. (a) Let $x \in \mathbb{F}_2^{k+m}$. Then $x = \begin{pmatrix} a \\ b \end{pmatrix}$ for unique $a \in \mathbb{F}_2^k$ and $b \in \mathbb{F}_2^m$. We compute:

$$\begin{aligned}
& \begin{pmatrix} a \\ b \end{pmatrix} \in \operatorname{Col}\left(\begin{bmatrix} I_k \\ A \end{bmatrix}\right) \\
\iff\quad & \begin{pmatrix} a \\ b \end{pmatrix} = \begin{bmatrix} I_k \\ A \end{bmatrix} y \quad \text{for some } y \in \mathbb{F}_2^k \\
\iff\quad & \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y \\ Ay \end{pmatrix} \quad \text{for some } y \in \mathbb{F}_2^k \\
\iff\quad & a = y \text{ and } b = Ay \quad \text{for some } y \in \mathbb{F}_2^k \\
\iff\quad & b = Aa \\
\iff\quad & Aa + b = \vec{0} \\
\iff\quad & [A\ I_l] \begin{pmatrix} a \\ b \end{pmatrix} = \vec{0} \\
\iff\quad & \begin{pmatrix} a \\ b \end{pmatrix} \in \operatorname{Nul}([A\ I_l])
\end{aligned}$$

So $x \in \operatorname{Col} E$ if and only if $x \in \operatorname{Nul} H$.

(b)

Consider the map

$$c : \mathbb{F}_2^k \to \mathbb{F}_2^{m+k}, y \to Ey = \begin{pmatrix} y \\ Ay \end{pmatrix}$$

Note that

$$\operatorname{Col}(E) = \{Ey \mid y \in \mathbb{F}_2^k\} = \{c(y) \mid y \in \mathbb{F}_2^k\}$$

and so $\operatorname{Col}(E) = \operatorname{Im} c$. Looking at the first k bits of Ey we see that c is 1-1. So α is a bijection from $\mathbb{F}_2^k$ to $\operatorname{Col} E$. Let $v_i = \operatorname{Col}(E)$. Then

$$c(y) = Ey = y_1v_1 + y_2v_2 + \ldots + y_kv_k$$

and since c is a bijection, for each $a \in \mathrm{Col}(E)$ there exists a unique $(y_1, \ldots, y_k) \in \mathbb{F}_2^k$ with $a = y_1v_1 + \ldots + y_kv_k$. Hence $(v_1, \ldots, v_k)$ is basis for $\mathrm{Col}(E)$.

(c) By (b) $\dim \mathrm{Col}E = k$. Since $\mathrm{Col}E \subseteq \mathbb{F}_2^{m+k}$, $\mathrm{Col}E$ has length $m + k$. □

We may view c has code for $\mathbb{F}_2^k$ in the alphabet $\mathbb{B}_2$. Since $c(y) = Ey = (y, Ay)$ the first k-bits of $c(y)$ (namely y) are called the message bits, and the last m bits (namely Ay) are called the check bits.

Definition VIII.21. *Let $C \subseteq \mathbb{B}^n$ 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 $k \times m$ 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 $k \times m$-matrix A.

Definition VIII.22. *Let $C, D \subseteq \mathbb{F}_2^n$. We say that D is a permutation of C if the exist a bijection $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ with*

$$D = \{a_{\pi(1)}a_{\pi(2)} \ldots a_{\pi(n)} \mid a_1a_2 \ldots a_n \in C\}$$

Example VIII.23. *Show that*

$$D = \{000, 100, 001, 101\}$$

is a permutation of

$$C = \{000, 010, 001, 011\}$$

D is obtained from C via the permutation:

$$\pi : \frac{1 \quad 2 \quad 3}{2 \quad 1 \quad 3}$$

Notation VIII.24. *Let I, J be sets, $i, k \in I$, A an $I \times J$ matrix and x and y J-tuples.*

(a) $\mathrm{R}_i xA$ denotes the $I \times J$ matrix B with $\mathrm{Row}_i B = x$ and $\mathrm{Row}_l A = \mathrm{Row}_l A$ for all $l \in I$ with $l \neq i$. (So $\mathrm{R}_i xA$ is the matrix obtained from A by replacing Row i by x.)

(b) $\mathrm{R}_{ik}xyA = \mathrm{R}_i x(\mathrm{R}_k yA)$. (So $R_{ik}xyA$ is the matrix obtained from A by replacing Row k by y and then replacing Row i by x.)

Definition VIII.25. *Let I and J be sets and $i, k \in I$ with $i \neq k$. Then '$R_i \leftrightarrow R_j$' and '$\mathrm{R}_i + \mathrm{R}_k \to R_k$' are the functions with domain the binary $I \times J$-matrix defined as follows:*

(a) $(R_i \leftrightarrow R_k)(A) = R_{ik}a^ka^iA$. So $R_i \to R_k$ interchangings row i and row k of A

(b) $(R_i + R_k \to R_k)(A) = R_k(a^i + a^k)A$. So $(R_i + R_k \to R_k)$ adds row i to row k of A.

An elementary row operation is one of the function $R_i \leftrightarrow R_j$ and $R_i + R_k \to R_k$.

Elementary column operations are defined similarly, using the symbol C *in place of* R *and using columns rather that rows.*

Lemma VIII.26. *(a) Let H and G be $m \times n$-matrices in $\mathbb{F}_2$ and suppose G is obtain from H by sequence of elementary row operation. Then* $\mathrm{Nul}H = \mathrm{Nul}G$.

(b) Let E and F be $m \times n$-matrices in $\mathbb{F}_2$ 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 if $h_i x = 0$ for all $1 \le i \le m$

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 $x \in \mathrm{Nul}H$ 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 $x \in \mathrm{Nul}H$. If $l \neq k$, 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 $x \in \mathrm{Nul}G$. Note that $h_k = h_i + (h_i + h_k) = g_i + g_k$ and by symmetry, $\mathrm{Nul}G \subseteq \mathrm{Nul}H$. Thus (a) holds.

(b) Note that $x \in \mathrm{Col}E$. Let $v_j = \mathrm{Col}_j(E)$. Then $x \in \mathrm{Col}E$ if and only if

$$x = \sum_{j \in J} y_j v_j$$

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 $x \in \mathrm{Col}E$ 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

$$x = \sum_{l \in J} y_l v_l = y_j v_j + y_k v_k + \sum_{l \neq j,k} y_l v_l = (y_j + y_k)v_j + y_k(v_j + v_k) + \sum_{l \neq j,k} y_l v_l$$

and so $x \in \mathrm{Col}F$. By symmetry $\mathrm{Col}F \subseteq \mathrm{Col}E$ and so (b) holds. □

Proposition VIII.27. *Let C be a subspace of $\mathbb{B}^n$. Then there exists a permutation D of C such that D has a generating matrix and a check matrix in standard form.*

Proof. Observe first that E has $n \times k$- generating matrix. For example we can choose a matrix whose columns consists of all the codewords. Deleting all the zero columns we may assume that E has non zero column. Then $e_{1i} = 1$ for some i. Applying the permutation $1 \leftrightarrow i$ to the code C we may assume that $e_{11} = 1$. Adding the first column to column j for each $2 \le j \le k$ with $e_{1j} = 1$ we may assume that $e_{1j} = 0$ for all $2 \le j \le k$. Consider the matrix F obtained by deleting row 1 and column 1 from E. By induction we may assume that F is in standard form. Adding column i to the first column for each $2 \le i \le k$ with $e_{i1} = 1$ we obtain a generating matrix in standard form. □

Example VIII.28. *Find a generating matrix and a check matrix in standard form for*

$$C = \{000, 110, 101, 011\}$$

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

So

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \qquad \text{and} \qquad H = [1, 1, 1]$$

are a generating matrix and check matrix in standard form for C

VIII.4 Constructing 1-error-correcting linear codes

Definition VIII.29. *Let $(v_1, \ldots, v_k)$ be a k-tuple 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.

Note that $(v_1, \ldots, v_k)$ is linearly independent 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 VIII.30. *Let H be 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 any sum of $2r$ or less columns of H is not equal to $\vec{0}$.

Proof. Let $x = x_1 x_2 \ldots x_n \in \mathbb{F}^n$ with $x \neq \vec{0}$ and let h_i be the i-th column of H. Then

$$Hx = x_1 h_1 + x_2 h_2 + \ldots + x_n h_n.$$

Let $1 \leq i_1 < i_2 < \ldots < i_d \leq n$ such that $x_{i_j} \neq 0$ for all $1 \leq j \leq d$ and $x_i = 0$ for all other i. Then $x_{i_j} = 1$ and so

$$Hx = h_{i_1} + h_{i_2} + \ldots + h_{i_d}.$$

Observe that $d = \mathrm{wt}(x)$. Since $x \in C$ if only if $Hx = \vec{0}$, we conclude that there exists a vector of weight d in C if and only if there exists d columns of H whose sum is equal to 0.

This proves the first statement. Let $1 \leq k_1 < k_2 < \ldots < k_e \leq n$. Then the columns $h_{k_1}, \ldots, h_{k_e}$ are linearly dependent if and only if there exists a non-zero $(l_1, l_2, \ldots, l_e) \in \mathbb{F}_2^e$ with

$$(*) \qquad l_1 h_{k_1} + l_2 h_{k_2} + \ldots l_e h_{k_e} = \vec{0}.$$

In a minimal linear dependent set of columns, all the l_j will be non-zero (since the columns with $l_j \neq 0$ are still linear dependent). So $l_j = 1$ for all $1 \leq j \leq e$ and (*) becomes

$$h_{k_1} + h_{k_2} + \ldots + h_{k_e} = \vec{0}.$$

Thus the minimal number of linear dependent columns is also the minimal numbers of columns whose sum is equal to $\vec{0}$.

(b) C is r-error correcting if and only if the minimal distance is at least $2r + 1$. By (a) C has minimum distance less or equal to $2r$ if and only if there exists $2r$ or less columns whose sum is zero. So (b) holds. □

Corollary VIII.31. *Let $n, k \in \mathbb{Z}$ with $k \leq n$ and put $m = n - k$. Let $C \subseteq \mathbb{F}_2$ and H be an $m \times n$-matrix in $\mathbb{F}_2$. Then the following two statements are equivalent:*

(a) C is binary linear 1-error correcting code with check matrix H and dimension k.

(b) $C = \mathrm{Nul}(H)$, $\mathrm{Col}H = \mathbb{F}_2^m$ and the columns of H are non-zero and pairwise distinct.

Proof. By definition H is a check matrix for C if and only if $C = \mathrm{Nul}H$. So we may assume $C = \mathrm{Nul}H$. Then

$$\dim C = \dim \mathrm{Nul}H = n - \dim \mathrm{Col}H = k + (m - \dim \mathrm{Col}H)$$

Thus $\dim C = k$ if and only if $\dim \mathrm{Col}H = m$ and so if and only if $\mathrm{Col}H = \mathbb{F}^m$.

We now apply VIII.30 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 $\vec{0}$. Since $x + y = \vec{0}$ if and only if $x = y$, we see that C is 1-error correcting if and only if the columns of H are non-zero and pairwise distinct. □

Lemma VIII.32. *(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*

$$k \leq n - \lceil \log_2(n+1) \rceil$$

(b) The maximal information rate of an 1-error correcting, binary linear code of length n is

$$1 - \frac{\lceil \log_2(n+1) \rceil}{n}$$

Proof. Suppose first that there exist an 1-error-correcting code of length n and dimension n. Then the Packing bound shows

$$2^k(1+n) \leq 2^n$$

and so $k + \log_2(1+n) \leq n$ and $k \leq n - \log_2(1+n)$.

Suppose next that $k \leq n - \log_2(1+n)$. Put $m = n - k$. Then $m \leq m + k = n \leq 2^m - 1$. Put $e_i = \mathrm{Col}_i(I_m)$. Then $|\mathbb{F}_2^k \smallsetminus \{\vec{0}, e_1, \dots, e_m\}| = 2^m - m - 1 \geq k$. So there exists distinct vectors $a_1, \dots, a_k \in \mathbb{F}_2^k \smallsetminus \{\vec{0}, e_1, \dots, e_m\}$ Let A be the $k \times m$ matrix with $\mathrm{Col}_i(A) = a_i$ and put $H = [A, I_m]$. Since $\mathrm{Col}I_m = \mathbb{F}_2^m$, also $\mathrm{Col}H = \mathbb{F}_2^m$. VIII.31 now shows that $\mathrm{Nul}H$ is a k-dimensional 1-error correcting code of length n. □

Note that

$$\lim_{n\to\infty} 1 - \frac{\lceil \log_2(n+1) \rceil}{n} = 1$$

So one can construct 1-error-correcting binary linear codes of information rates arbitrarily close to 1.

Definition VIII.33. *A Hamming code is a perfect-1-error correcting binary linear code.*

Theorem VIII.34. *Let $C \subseteq \mathbb{F}_2^n$ be a linear code with $n \times m$ check matrix H such that $\dim C = n - m$. 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. By VII.22

1°. *C is a Hamming code if and only if C is 1-error correcting and $|C|(1+n) = 2^n$*

Since $\dim C = n - m$

$$|C|(1+n) = 2^n \iff 2^{n-m}(1+n) = 2^n \iff 1+n = 2^m \iff n = 2^m - 1$$

By VIII.31 C is 1-error correcting if and only if the columns of H are non-zero and pairwise distinct. Thus

2°. *C is a Hamming code if and only if and only if the columns of H are non-zero and pairwise distinct and $n = 2^m - 1$.*

Since $\mathbb{F}_2^m$ has exactly $2^m - 1$ non-zero vectors (2°) implies the theorem. □

Corollary VIII.35. *Let $n \in \mathbb{N}$. Then there exists a Hamming code of length n if and only if $n = 2^m - 1$ for some $m \in \mathbb{N}$, that is if and only if $n + 1$ is a power of 2. If this is the case, a Hamming code of length n is unique up to permutation.*

Proof. The first statement follows directly from Theorem VIII.34. The same Theorem shows that an $n \times m$-check matrix of an Hamming code of length n is unique up to a permutation of the columns. So also the Hamming code is unique up to permutation. □

Example VIII.36. *Find a Hamming code of length 7.*

We have $7 = 2^3 - 1$. So $m = 3$ and we can choose the check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

So

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

is a generating matrix. The codewords are

$$\begin{array}{cccc} 0000000 & 0001011 & 1111111 & 1110100 \\ 1000111 & 1100001 & 0111000 & 0011110 \\ 0100110 & 1010010 & 0111001 & 0101101 \\ 0010101 & 1001100 & 1101010 & 0110011 \end{array}$$

Lemma VIII.37. *Let C be a linear code with check matrix H. Let (c, z) be a 1-bit error for C.*

(a) If the error occurred in bit i, then $Hz = \mathrm{Col}_i H$.

(b) If C is 1-error-correcting, i is the uniquely determined by $\mathrm{Col}_i H = Hz$.

Proof. (a) Let $e_i = \mathrm{Col}_i(I_n)$. Then $z = c + e_i$ and so $Hz = H(c + e_i) = Hc + He_i = \vec{0} + \mathrm{Col}_i(H) = \mathrm{Col}_i(H)$.

(b) Just recall no two columns of the check matrix of a 1-error correcting binary linear code are the same. □

Example VIII.38. *Let C be the binary linear code with check matrix*

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

(a) Is C 1-error correcting?

(b) Suppose that (a, z) is a 1-bit error with $z =$ 101100. Find z.

(a) The columns of H are non-zero and pairwise distinct. So C is 1-error correcting.

(b)

$$Hz = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

So Hz is the third column of H. Hence the error occurred in bit 3 and $a =$ 100110. To confirm

$$Hz = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

So 100110 is indeed a codeword.

VIII.5 Decoding linear codes

Definition VIII.39. *Let H be a check matrix for the linear code $C \subseteq \mathbb{F}_2^n$ and $z \in \mathbb{F}_2^n$.*

(a) Hz is called the syndrome of z with respect to H.

(b) $z + C = \{z + c \mid c \in C\}$ *is called the coset of* C *containing* z.

Observe that $z \in C$ if and only of the syndrome of z is $\vec{0}$.

Lemma VIII.40. *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 are equivalent:*

(a) y *and* z *have the same syndrome with respect to* H.

(b) $y + z \in C$.

(c) $z = y + c$ *for some codeword* $c \in C$.

(d) $z \in y + C$.

(e) $(y + C) \cap (z + C) \neq \varnothing$.

(f) $y + C = z + C$.

Proof. We have

$$\begin{array}{lll}
 & Hy = Hz & \text{(a)} \\
\Longleftrightarrow & Hz + Hy = \vec{0} & \\
\Longleftrightarrow & H(z + y) = \vec{0} & \\
\Longleftrightarrow & z + y \in C & \text{(b)} \\
\Longleftrightarrow & z + y = c \text{ for some } c \in C & \\
\Longleftrightarrow & z = y + c \text{ for some } c \in C & \text{(c)} \\
\Longleftrightarrow & z \in y + C & \text{(d)}
\end{array}$$

So the first four statements are equivalent.

(d) $\Longrightarrow$ (e): Note that $z = z + \vec{0} \in z + C$. So if $z \in y + C$, then $z \in z + C \cap y + C$ and $z + C \cap y + C \neq \varnothing$ and (e) holds.

(e) $\Longrightarrow$ (f): Let $u \in y + C \cap z + C$. Then since (d) implies (a), $Hu = Hy = Hz$ Let $v \in \mathbb{F}_2^n$. Since (a) and (d) are equivalent, $v \in y + Z$ if and only of $Hv = Hy$, if and only if $Hv = Hz$ and if and only if $v \in z + C$. So $y + C = z + C$.

(f) $\Longrightarrow$ (d): If $y + C = z + C$, then $z = z + \vec{0} \in z + C = y + C$. □

Corollary VIII.41. *Let* $C \subseteq \mathbb{F}_2^n$ *be a linear code and* $z \in \mathbb{F}_2^n$. *Then* z *lies in a unique coset of* C, *namely* $z + C$. *In particular, distinct cosets are disjoint.*

Proof. If $z \in y + C$, then VIII.40 show that $y + C = z + C$. □

Example VIII.42. *Let* C *be the code with check matrix*

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Find the cosets of C *and the corresponding syndromes.*

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. So

$$C = \{0000, 1010, 0101, 1111\}$$

We compute

$0000 + C$	$1000 + C$	$0100 + C$	$1100 + C$
0000	1000	0100	1100
1010	0010	1110	0110
0101	1101	0001	1001
1111	0111	1011	0011
00	10	01	11

The last row lists the common syndrome of the elements in the coset.

Which of the cosets contains 1101?

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

So 1101 lies in the coset with symdrome 10, that is $1000 + C$.

Lemma VIII.43. *Let C be the 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 C with respect to H is $\mathrm{Col}(H)$.

(b) The numbers of syndromes for C with respect to H is 2^{n-k}.

(c) $n - k \le m$.

(d) If $m = n - k$ (for example if H is in standard form), every element of $\mathbb{F}_2^m$ is a syndrome.

Proof. (a) s is a syndrome 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 u be the numbers of syndromes. by VIII.40 u is also the number of cosets of C. Each cosets of C contains 2^k elements and each element of $\mathbb{F}_2^n$ lies in unique coset so

$$u \cdot 2^k = 2^n.$$

Thus $u = 2^{n-k}$.

(c) The syndromes are contained in $\mathbb{F}_2^m$. So there are at most 2^m syndromes. Thus $n - k \le m$,

(d) If $m = n - k$, C has 2^m syndromes and so every element of $\mathbb{F}_2^m$ is a syndrome.

□

Remark VIII.44. *Let H be an $m \times n$-matrix in $\mathbb{F}_2$. Then* $\dim \mathrm{Nul} H + \dim \mathrm{Col} H = n$.

Proof. Put $k = \dim \mathrm{Nul} H$. By VIII.43 $|\mathrm{Col} H|$ is equal to the number of syndromes of $\mathrm{Nul} H$ and so equal to 2^{n-k}. Thus $\dim \mathrm{Col} H = n - k = n - \dim \mathrm{Nul} H$. □

Definition VIII.45. *Let C be linear code with an $m \times n$ check matrix H. A syndrome look-up table for C with respect to H is a function*

$$\tau : \mathbb{F}_2^m \to \mathbb{F}_2^n$$

such that for all syndrome s of H,

(i) $\tau(s)$ has syndrome s,

(ii) $\tau(s)$ is a vector of minimal weight in $\tau(s) + C$.

Note that (i) and (ii) just mean that $\tau(s)$ is a of minimal weight among the vectors with syndrome s.

Definition VIII.46. *Let C be linear code of length n with check matrix H and τ a syndrome look-up table for C with respect to H. Then the function*

$$\sigma : \mathbb{R}^n \to C, \; z \to z + \tau(Hz)$$

is called the decision rule for C with respect to H and τ.

Note here that by definition of a syndrome table $\tau(Hz)$ has syndrome Hz and so by VIII.40 $z + \tau(Hz) \in C$. So this is well-defined.

Lemma VIII.47. *Let C be linear code of length n with check matrix H, τ a syndrome look-up table for C with respect to H and σ the corresponding decision rule. Then σ is a Minimum Distance Rule*

Proof. Let $z \in \mathbb{R}^n$ and $c \in C$. Put $s = Hz$. Be definition of τ, $\tau(s)$ has syndrome s. Since also z has syndrome s, we conclude from VIII.40 that $z + c \in z + C = \tau(s) + C$. By definition of τ, $\tau(s)$ is of minimal weight in $\tau(s) + C$. So

$$\mathrm{wt}(z + c) \leq \mathrm{wt}(\tau(s)).$$

We compute

$$d(z, \sigma(z)) = d(z, z + \tau(s)) = \mathrm{wt}(z + (z + \tau(s)) = \mathrm{wt}(\tau(s)) \geq \mathrm{wt}(z + c) = d(z, c).$$

□

Algorithm VIII.48. *Let C be linear code with an $m \times n$ check matrix H. Choose an ordering $\{z_1, \ldots, z_{2^n}\}$ of $\mathbb{F}_2^n$ such that $\mathrm{wt}(z_i) \leq \mathrm{wt}(z_j)$ for all $1 \leq i < j \leq 2^n$. Define functions $\tau_l : S_l \to \mathbb{F}_2^m$ inductively as follows:*

For $l = 0$ let $S_0 = \varnothing$ and $\tau_0 = \varnothing$.

Suppose $l > 0$ and τ_{l-1} has been defined. Compute $s_l = Hz_l$.

- *If $s_l \in S_{l-1}$, put $\tau_l = \tau_{l-1}$*
- *If $s_l \notin S_l$, put $S_l = S_{l-1} \cup \{s_l\}$ and extend τ_{l-1} to a function τ_l on S_l by $\tau_l(s_l) = z_l$.*

Stop the algorithm if $l = 2^n$ or $|S_l| = 2^m$.

The last τ_l is a syndrome look up table for C with respect to H.

Example VIII.49. *Let C be the 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 and compute $\sigma(11001)$, where σ is the decision rule for C with respect to H and τ.

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
$\|S_l\|$	1	2	3	4	5	6	7	–	–	8

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\tau(100) = 00100$$

and so

$$\sigma(11001) = 11001 + 00100 = 11101$$

To double check

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Chapter IX

Algebraic Coding Theory

IX.1 Classification and properties of cyclic codes

Definition IX.1. *A code $C \subseteq \mathbb{F}_2^n$ is called cyclic if*

(a) C is linear, and

(b) $a_n a_1 \dots a_{n-1} \in C$ for all $a = a_1 \dots a_n \in C$.

Example IX.2. *Which of following codes are cyclic?*

(a) $\{000, 100, 111, 011\}$.

(b) $\{000, 100, 010, 001\}$

(c) $\{000, 1010, 0101, 1111\}$

(a): Not cyclic, since 100 is in the code, but 010 is not.
(b): Not cyclic, since its not linear: 100 and 010 are in the code, but $100 + 010 = 110$ is not.
(c): Is cyclic.

Definition IX.3. *An ideal in a ring R is a subset S of R such that*

(a) $0 \in S$.

(b) If $s, t \in S$ then $s + t \in S$.

(c) If $a \in R$ and $s \in S$, then $as \in S$ and $sa \in S$.

Example IX.4. *The set of even integers is an ideal in the integers.*

Indeed 0 is even. Sums of even integers are even and any multiple of an even integer is even.

Lemma IX.5. *Let $\mathbb{F}$ be a field and $f, g \in \mathbb{F}[x]$ with $g \neq 0$. Then there exists uniquely determined $q, r \in \mathbb{F}[x]$ with*

$$f = qh + r \text{ and } \deg r < \deg h$$

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 $\deg f \geq h$. 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 $m = \deg h \geq \deg h = n$. Put

$$\tilde{f} = f - \frac{b_m}{a_n} x^{m-n} h.$$

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

$$\tilde{f} = \tilde{q}h + \tilde{r} \text{ and } \deg \tilde{r} < \deg h$$

We have

$$f = \tilde{f} + \frac{b_m}{a_n} x^{m-n} h = \tilde{q}h + \tilde{r} + \frac{b_m}{a_n} x^{m-n} h = (\tilde{q} + \frac{b_m}{a_n} x^{m-n})h + \tilde{r}.$$

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

$$f = qh + r = q^* h + r^*, \quad \deg r < \deg h \quad \text{and} \quad \deg r^* < \deg h$$

for some $q, q^*, r, r^* \in \mathbb{F}[x]$. Then

$$(q - q^*)h = r^* - r.$$

By A.8, $\deg(r^* - r) \leq \max(\deg r, \deg r^*) < \deg h$ and so $\deg(q - q^*)h < \deg h$. Hence A.8 shows that $q - q^* = 0$ and so also $r^* - r = (q - q^*)h = 0h = 0$.

Hence $q = q^*$ and $r = r^*$. So q and r are unique. □

Definition IX.6. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$ with $h \neq 0$. Let $f, g \in \mathbb{F}[x]$.*

$\overline{f}$ is the remainder of f 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

$$f^{\odot 0} = \overline{1} \qquad \text{and} \qquad f^{\odot(n+1)} = f^{\odot n} \odot f.$$

Let

$$\mathbb{F}^h[x] = \{f \in \mathbb{F}[x] \mid \deg f < \deg h\}$$

$(\mathbb{F}^h[x], \oplus, \odot)$ is called the ring of polynomials modulo h with coefficients in $\mathbb{F}$.

Example IX.7. *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

$$(a+bx)\oplus(c+dx)=(a+c)+(b+dx)$$

and

$$\begin{aligned}(a+bx)\odot(c+dx) &= \overline{ac+adx+bcx+bdx^2}\\ &= \overline{(ac-bd)+(ad+bc)x+bd(x^2+1)}\\ &= (ac-bd)+(ad+bc)x\end{aligned}$$

So $\mathbb{R}^{x^2+1}[x]$ has the same addition and multiplication as the ring of complex numbers

$$\mathbb{C}=\{a+bi \mid a,b\in\mathbb{R}\}.$$

Lemma IX.8. *Let $\mathbb{F}$ be a field and $h\in\mathbb{F}[x]$ with $h\neq 0$. Let $f,g\in\mathbb{F}^h[x]$ and $a\in\mathbb{F}$.*

(a) $\overline{f}=f$.

(b) $f\oplus g=f+g$

(c) $a\odot f=af$.

Proof. (a) $f=0h+f$ and $\deg f<\deg h$. So f is the remainder of f when divided by h.

(b) Observe that $\deg(f+g)\leq\max(\deg f,\deg g)<\deg h$ and so by (a) $\overline{f+g}=f+g$.

(c) Since $\deg af\leq\deg f<\deg h$, (a) gives $\overline{af}=af$. □

Remark IX.9. *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 IX.10. *Let R be a commutative ring and $a,b\in R$. We say that a divides b and write $a\mid b$ if $b=ra$ for some $r\in R$.*

Lemma IX.11. *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) $\overline{f}\in\mathbb{F}^h[x]$.

(b) $0\oplus f=1\odot f=\overline{f}=\overline{\overline{f}}=f\odot 1=f\oplus 0$

(c) $f\oplus g=\overline{f+g}=\overline{f}+\overline{g}=\overline{f}\oplus\overline{g}=\overline{f}\oplus g=f\oplus\overline{g}$

(d) $f\odot g=\overline{fg}=\overline{\overline{f}\,\overline{g}}=\overline{f}\odot\overline{g}=\overline{f}\odot g=f\odot\overline{g}$

(e) $\overline{f^n}=f^{\odot n}=\left(\overline{f}\right)^{\odot n}=\overline{\left(\overline{f}\right)^n}$

(f) $-\overline{f}=\overline{-f}$.

(g) $\overline{f}=0$ *if and only if h divides f.*

(h) $\overline{f}=\overline{g}$ *if and only if h divides $g-f$.*

(i) $a\odot f=\overline{af}=a\overline{f}$

Proof. Recall first that by definition of the remainder $f = ph + \overline{f}$ and $g = qh + \overline{g}$ for some $p, q \in \mathbb{F}[x]$. Also $\deg \overline{f} < \deg h$ and $\deg \overline{g} < \deg h$.

(a) Since $\deg \overline{f} < \deg h$, $\overline{f} \in F^h[x]$.

(b) $1 \odot f = \overline{1f} = \overline{f} = \overline{0 + f} = 0 \oplus f$. Since $\overline{f} \in \mathbb{F}^h[x]$, IX.8(a) gives $\overline{\overline{f}} = \overline{f}$.

(c) We compute

$$f + g = (ph + \overline{f}) + (qh + \overline{g}) = (p + q)h + (\overline{f} + \overline{g})$$

Note that $\overline{f} + \overline{g}$ has degree less than h and so $\overline{f} + \overline{g}$ is the remainder of $f + g$ when divided by h. Hence $f \oplus g = \overline{f + g} = \overline{f} + \overline{g}$.

By IX.8(b), $\overline{f} \oplus \overline{g} = \overline{f} + \overline{g}$. The remaining parts of (c) now follow from $\overline{\overline{f}} = \overline{f}$ and $\overline{\overline{g}} = \overline{g}$.

(d) By definition $\overline{f}\,\overline{g} = th + \overline{\overline{f}\,\overline{g}}$ for some $t \in \mathbb{F}[x]$ and $\deg \overline{\overline{f}\,\overline{g}} < \deg h$. We have

$$fg = (ph + \overline{f})(qh + \overline{g}) = (phq + \overline{f}q + p\overline{g})h + \overline{f}\,\overline{g} = (phq + \overline{f}q + p\overline{g} + t)h + \overline{\overline{f}\,\overline{g}}$$

So $\overline{\overline{f}\,\overline{g}}$ is the remainder of fg when divided by h. Hence $f \odot g = \overline{fg} = \overline{\overline{f}\,\overline{g}}$. (d) is now readily verified.

(e) For $n = 0$ all four expression are equal to $\overline{1}$.

Suppose (e) holds for n. Then using (d)

$$\overline{f^{(n+1)}} = \overline{f^n f} = f^n \odot f = \overline{f^n} \odot f = f^{\odot n} \odot f = f^{\odot(n+1)}$$

$$\left(\overline{f}\right)^{\odot(n+1)} = \left(\overline{f}\right)^{\odot n} \odot \overline{f} = \overline{f^n} \odot \overline{f} = \overline{f^n} \odot f$$

and

$$\overline{\left(\overline{f}\right)^{(n+1)}} = \overline{\left(\overline{f}\right)^n f} = \overline{f^{\otimes n} f} = f^{\otimes n} \odot f$$

(f) $-f = -ph + (-\overline{f})$ and so $-\overline{f}$ is the remainder of f when divided by h.

(g) h divides f if and only if $f = sh$ for some $s \in F[x]$ and so if and only if $f = sh + 0$ for some $s \in F[x]$. Since $\deg 0 < \deg h$, this holds if and only if 0 is the remainder of f when divided by 0.

(h) $\overline{f} = \overline{g}$ if and only if $\overline{f} - \overline{g} = 0$ and if and only if $\overline{f - g} = 0$. So (h) follow from (g).

(i) By (d), $a \odot f = \overline{af} = a \odot \overline{f}$ and by IX.8(c) $a \odot \overline{f} = a\overline{f}$. □

Lemma IX.12. *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 in $\mathbb{F}^h[x]$ 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 in $\mathbb{F}^h[x]$. Then by definition of an ideal, $0 \in I$ and $f + g \in I$. Also since $a \in \mathbb{F}^h[x])$, $af = a \odot f \in I$. Thus I is an $\mathbb{F}$-subspace of $\mathbb{F}^h[x]$. Also $x \odot f = \overline{x} \odot f \in I$ 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 $f \in I$. Then by definition of a subspace $0 \in I$ and $f + g \in I$.

We claim that $x^i \odot f \in I$ for all $i \in \mathbb{N}$. $x^0 \odot f = 1 \odot f = f$ and so the claim holds for $i = 0$. Suppose inductively that $x^i \odot f \in I$. Then by assumption also $x \odot (x^i \odot f) \in I$ and hence

$$x^{i+1} \odot f = \overline{x^{i+1} f} = \overline{x(x^i f)} = x \odot \overline{x^i f} = x \odot (x^i \odot f) \in I.$$

So indeed $x^i \odot f \in I$ for all $i \in \mathbb{N}$ and $f \in I$.

Let $g \in \mathbb{F}^h[x]$. Then $g = \sum_{i=0}^n a_i x^i$ for some $a_i \in \mathbb{F}_2$ and so

$$g \odot f = \left(\sum_{i=0}^{n-1} a_i x^i\right) \odot f = \sum_{i=0}^{n-1} (a_i x^i) \odot f = \sum_{i=0}^{n-1} a_i \left(x^i \odot f\right)$$

Since each $x^i \odot f \in I$ and I is an $\mathbb{F}$-subspace, $g \odot f \in I$. Thus I is an ideal in $\mathbb{F}^h[x]$. □

Definition IX.13. *Let $n \in \mathbb{Z}^+$. Then $V^n[x] = \mathbb{F}_2^{x^n-1}[x]$.*

Lemma IX.14. *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

$$f = \sum_{i=0}^{r} \sum_{j=0}^{n-1} a_{ij} x^{in+j}$$

for some $a_{ij} \in \mathbb{F}$. Then

$$\overline{f} = \sum_{j=0}^{n-1} \left(\sum_{i=0}^{r} a_{ij}\right) x^i$$

Proof. Since $x^n = (x^n - 1) + 1$, $\overline{x^n} = 1$ and so

$$\overline{x^{ni+j}} = \overline{(x^n)^i x^j} = \overline{\overline{x^n}^i x^j} = \overline{1x^j} = \overline{x^j} = x^j$$

So (a) holds.
From (a) we get

$$\overline{\sum_{i=0}^{r} \sum_{j=0}^{n-1} a_{ij} x^{in+j}} = \sum_{i=0}^{r} \sum_{j=0}^{n-1} a_{ij} \overline{x^{in+j}} = \sum_{i=0}^{r} \sum_{j=0}^{n-1} a_{ij} x^j = \sum_{j=0}^{n-1} \left(\sum_{i=0}^{r} a_{ij}\right) x^j$$

and so (b) holds. □

Example IX.15. *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 $\mathbb{F}_2[x]$.*

$$f = 1 + x + x^4 + x^5 + x^{6+1} + x^{6+5} + x^{12+5} + x^{24+4}$$

and so

$$\overline{f} = 1 + x + x^4 + x^5 + x + x^5 + x^5 + x^4 = 1 + 2x + 2x^4 + 3x^5 = 1 + x^5$$

Lemma IX.16. *Let $C \subseteq \mathbb{F}_2^n$. For $a = a_0 \ldots a_{n-1} \in \mathbb{F}_2^n$ define*

$$a(x) = \sum_{i=0}^{n-1} a_i x^i = a_0 + a_1 x + \ldots a_{n-1} x^{n-1} \in V^n[x].$$

Put $C(x) = \{a(x) \mid a \in C\}$. Then C is a cyclic code if and only if $C(x)$ is an ideal in $V^n[x]$.

Proof. 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 $V^n[x]$.

Put $\hat{a} = a_{n-1}a_0 \ldots a_{n-2}$. Then

$$x \odot a(x) = \overline{x \sum_{i=0}^{n-1} a_i x^i} = \overline{\sum_{i=0}^{n-1} a_i x^{i+1}} = a_{n-1} + a_0 x + \ldots + a_{n-2}x^{n-1} = \hat{a}(x)$$

Thus $\hat{a} \in C$ if and only if $x \odot a(x) \in C(x)$. So by IX.12 C is cyclic if and only if $C(x)$ is an ideal. □

Definition IX.17. *Let R be a commutative ring and $a \in R$. Define $\langle a \rangle = \{ra \mid r \in R\}$. $\langle a \rangle$ is called the ideal in R generated by a.*

Lemma IX.18. *Let R be a commutative ring and $a \in R$.*

(a) $\langle a \rangle$ is an ideal in R.

(b) $\langle a \rangle \subseteq I$ for any ideal I of R with $a \in I$.

Proof. (a) 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 $s(ra) = (sr)a \in \langle a \rangle$.

(b) By definition of an ideal, $ra \in I$ for all $r \in R$ and so $\langle a \rangle \subseteq I$. □

Lemma IX.19. *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 in $\mathbb{F}^h[x]$ generated by f. Then*

(a) $\langle f \rangle$ is the $\mathbb{F}$-subspace of $\mathbb{F}^h[x]$ spanned by

$$1 \odot f, x \odot f, \ldots, x^{n-1} \odot f$$

where $n = \deg h$.

(b) $\langle f \rangle = \{g \odot f \mid g \in \mathbb{F}[x]\}$.

Proof. (a) Let $g = \sum_{i=0}^{n-1} a_i x^i \in \mathbb{F}^h[x]$. Then

$$g \odot f = a_0(1 \odot f) + a_1(x \odot f) + \ldots + a_{n-1}(x^{n-1} \odot f)$$

and so the elements of $\langle f \rangle$ are exactly the $\mathbb{F}$-linear combinations of $1 \odot f, x \odot f, \ldots, x^{n-1} \odot f$.

(b)

$$\langle f \rangle = \{g \odot f \mid g \in \mathbb{F}^h[x]\} = \{\overline{g} \odot f \mid g \in \mathbb{F}[x]\} = \{g \odot f \mid g \in \mathbb{F}[x]\}.$$

□

Example IX.20. *Find the ideal in $V^3[x]$ generated by $f = 1 + x^2$ and determined the corresponding cyclic code.*

$$\begin{aligned} 1 \odot (1 + x^2) &= 1 + x^2 \\ x \odot (1 + x^2) &= \overline{x + x^3} = x + 1 \\ x^2 \odot (1 + x^2) &= \overline{x^2 + x^4} = x + x^2 \end{aligned}$$

Note that $(1 + x^2) + (1 + x) = x^2 + x$. So

$$\langle x^2+1\rangle = \{0, 1+x^2, 1+x, x+x^2\}.$$

The corresponding cyclic code is

$$\{000, 101, 110, 011\}$$

Lemma IX.21. *Let $\mathbb{F}$ be a field and $f, g, t \in \mathbb{F}[x]$ with $t \neq 0$. Then $f \mid g$ if and only if $ft \mid gt$.*

Proof.

$$\begin{aligned}
& && ft \mid gt \\
&\Longleftrightarrow && gt = ftl \text{ for some } l \in \mathbb{F}[x] \\
&\Longleftrightarrow && gt - ftl = 0 \text{ for some } l \in \mathbb{F}[x] \\
&\Longleftrightarrow && (g - fl)t = 0 \text{ for some } l \in \mathbb{F}[x] \\
&\Longleftrightarrow && g - fl = 0 \text{ for some } l \in \mathbb{F}[x] \\
&\Longleftrightarrow && g = fl \text{ for some } l \in \mathbb{F}[x] \\
&\Longleftrightarrow && f \mid g
\end{aligned}$$

□

Lemma IX.22. *Let $\mathbb{F}$ be a field, $h \in \mathbb{F}[x]$ with $h \neq 0$ and I an ideal in $\mathbb{F}^h[x]$. Let $g \in \mathbb{F}[x]$ be of minimal degree with $\overline{g} \in I$ 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) $I = \langle \overline{g} \rangle$.

(c) g divides h in $\mathbb{F}[x]$ and so $h = tg$ for some $t \in \mathbb{F}[x]$.

(d) Let $f, f^* \in \mathbb{F}[x]$. Then $f \odot g = f^* \odot g$ if and only if t divides $f^* - f$.

(e) Let $k = \deg t$. Then $I = \{sg \mid s \in \mathbb{F}[x], \deg s < k\}$.

(f) $(g, xg, x^2g, \dots, x^{k-1}g)$ is an $\mathbb{F}$-basis for I.

(g) $\dim_{\mathbb{F}} I = k$.

(h) Let $f \in \mathbb{F}[x]$. Then $\overline{f} \in I$ if and only if $f \odot t = 0$.

(i) 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. Let $n = \deg h$. Since $\overline{h} = 0 \in I$, $\deg g \leq \deg h$

(a) Let $f = qg + r$ with $q, r \in \mathbb{F}[x]$ and $\deg r < \deg g$. If $g \mid f$, then $f = qg$ and so $\overline{f} = \overline{q} \odot g \in I$.

Suppose $\overline{f} \in I$. Since $\deg r < \deg g \leq \deg h$, we have $r = \overline{r} = \overline{f - qg} = \overline{f} \oplus \overline{-q} \odot \overline{g} \in I$. The minimal choice of $\deg g$ shows that $r = 0$ and so $g \mid f$.

(b) Let $f \in I$. Then $\deg f < n$ and so $\overline{f} = f \in I$. Thus by (a), $f = eg$ for some $e \in \mathbb{F}[x]$. Therefore $f = \overline{f} = \overline{e} \odot \overline{g}$ and $f \in \langle \overline{g} \rangle$. Since $\overline{g} \in I$, IX.18(b) shows $\langle \overline{g} \rangle \subseteq I$ and thus $I = \langle \overline{g} \rangle$.

(c) Note that $\overline{h} = 0 \in I$ and so by (a) , $h = tg$ for some $t \in \mathbb{F}[x]$.

(d) We have

$$\begin{array}{lll} & f \odot g = f^* \odot g & \\ \iff & \overline{fg} = \overline{f^* g} & - \text{ definition of } \odot \\ \iff & h \text{ divides } fg - f^* g & - \text{ IX.11(h)} \\ \iff & tg \text{ divides } (f - f^*)g & - \text{ (c)} \\ \iff & t \text{ divides } f - f^* & - \text{IX.21} \end{array}$$

(e) Let $f \in \mathbb{F}^h[x]$ and let s be the remainder of f when divided by t. Then t divides $f - s$ and so by (d), $f \odot \overline{g} = f \odot g = s \odot g$. Note that $\deg sg < k + (n-k) = n$ and so $s \odot g = \overline{sg} = sg$. Hence

$$I = \langle \overline{g} \rangle = \{ f \odot \overline{g} \mid f \in \mathbb{F}^h[x] \} = \{ sg \mid s \in \mathbb{F}[x], \deg s < k \}$$

(f) Note that $\deg g = n - k$. Let $s \in \mathbb{F}[x]$ with $\deg r < k$ and let $s = \sum_{i=0}^{k-1} a_i x^i$. Then

$$(*) \qquad sg = \sum_{i=0}^{k-1} a_i x^i g.$$

Together with (e) we see that I is the set of $\mathbb{F}$-linear combinations of $(g, xg, x^2 g, \dots, x^{k-1} g)$.
If $(a_0, \dots, a_{k-1}) \neq \vec{0}$, then $s \neq 0$ and $sg \neq 0$. By (*) $\sum_{i=0}^{k-1} a_i x^i \neq 0$ and so $(g, xg, x^2 g, \dots x^{k-1} g)$ is linearly independent.

(g) follows immediately from (f).
(h)

$$\begin{array}{lll} & \overline{f} \in I & \\ \iff & g \mid f & -(a) \\ \iff & gt \mid ft & - \text{IX.21} \\ \iff & h \mid ft & - \text{ (c)} \\ \iff & \overline{ft} = 0 & - \text{ IX.11(h)} \\ \iff & f \odot t = 0 & - \text{ definition of } \odot \end{array}$$

(i) Let $f \in \mathbb{F}^h[x]$ and write $f = qg + r$ with $\deg r < \deg g = n - k$. By (a),

$$f = \overline{f} \in I \qquad \iff \qquad r = 0.$$

Since $\deg f < n$ and $\deg g = n - k$, we have

$$\deg q < k$$

Also

$$\deg rt = \deg r + \deg t < (n-k) + k = n.$$

Note that $r = 0$ if and only if $\deg rt < \deg t$, that is if only if $\deg rt < k$. Since $\deg rt < n$, this gives

$$r = 0 \qquad \iff \qquad \text{the coefficient of } x^i \text{ in } rt \text{ is 0 for } k \leq i < n$$

We compute

$$ft = (qg + r)t = qgt + rt = qh + rt = q \cdot (x^n - 1) + rt = qx^n - q + rt$$

Let $k \leq i < n$. Since $i < n$, the coefficient of x^i in qx^n is 0. Since $i \geq k$ and $\deg q < k$, the coefficient of x^i in q also 0. Thus the coefficent of x^i in ft is the same as the coefficient of x^i in rt. Hence (i) holds. □

Definition IX.23. *Let $\mathbb{F}$ be a field.*

(a) Let $0 \neq f = \sum_{i=0}^{m} a_i x^i \in \mathbb{F}[x]$ with $a_m \neq 0$. Then a_m is called the leading coefficient of f. If $f = 0$, we call 0 the leading coefficient of f. We denote the leading coefficient of f be $\mathrm{lead}(f)$.

(b) A monic polynomial is a polynomial with leading coefficient 1.

(c) Let $0 \neq h \in \mathbb{F}[x]$ and I an ideal in $\mathbb{F}^h[x]$. Let $g \in \mathbb{F}[x]$ be a monic polynomial of minimal degree with $\overline{g} \in I$. Then g is called the canonical generator for I. (If $I \neq 0$, then g is the monic polynomial of minimal degree in I, and if $I = 0$, $g = \frac{1}{\mathrm{lead}(h)}h$).

Lemma IX.24. *Let $\mathbb{F}$ be a field and $0 \neq h \in \mathbb{F}[x]$.*

(a) Let I an ideal in $\mathbb{F}^h[x]$. Then the canonical generator of I is unique.

(b) Let $g \in \mathbb{F}[x]$ be monic. Then g is the canonical generator for the ideal $\langle\overline{g}\rangle$ in $\mathbb{F}^h[x]$ if and only if $g \mid h$.

Proof. (a) Let g_1 and g_2 be canonical generators of I. Then $\deg g_1 = \deg g_2$ and g_1 and g_2 are monic. Thus $\deg g_1 - g_2 < \deg g_1$. Also $\overline{g_1 - g_2} = \overline{g_1} - \overline{g_2} \in I$ and so by minimal choice of $\deg g_i$, $g_1 - g_2 = 0$ and $g_1 = g_2$.

(b) If g is the canonical generator for $\langle\overline{g}\rangle$, then IX.22 shows that $g \mid h$.

Conversely suppose that $g \mid h$ and let $\tilde{g}$ be the canonical generator for $\langle\overline{g}\rangle$. Since $\overline{\tilde{g}} = \langle\overline{g}\rangle$,

$$\overline{\tilde{g}} = f \odot \overline{g} = \overline{fg}$$

for some $f \in \mathbb{F}^h[x]$. Thus IX.11 shows that h divides $\tilde{g} - fg$. Thus $\tilde{g} - fg = kh$ for some $k \in \mathbb{F}[x]$. So $\tilde{g} = fg + kh$. Since $g \mid h$ this shows that $g \mid \tilde{g}$. Thus $\deg g \leq \deg \tilde{g}$. The minimal choice of $\deg \tilde{g}$ now shows that $\deg g = \deg \tilde{g}$ and so also g is a canonical generator of $\langle\overline{g}\rangle$. □

Definition IX.25. *Let $a = a_0 \ldots a_{n-1} \in \mathbb{F}_2^n$. For $0 \leq i < n$ define*

$$a^{(i)} = a_{n-i} \ldots a_{n-1} a_0 a_1 \ldots a_{n-1-i}$$

$a^{(i)}$ is called the cyclic i-shift of a.

$\langle a\rangle$ is the linear subspace of $\mathbb{F}_2^n$ spanned by $a^{(0)}, a^{(1)}, \ldots, a^{(n-1)}$. $\langle a\rangle$ is called the cyclic code generated by a.

Corollary IX.26. *Let $C \subseteq \mathbb{F}_2^n$ be cyclic code. Then there exists $a \in C$ with $C = \langle a\rangle$.*

Proof. Let $C(x)$ be the ideal of $V^n[x]$ correponding to C. Then $C(x) = \langle g\rangle$ for some $g \in V[x]$. Let $a \in C$ be the codeword corresponding to g. Then $a^{(i)}$ corresponds to $x^i \odot g$ and since $C(x)$ is spanned by $x^i \odot g$, $0 \leq i < n$, C is spanned by $a^{(i)}$. □

Example IX.27. *Determine the cyclic code generated by* 0110 *and find the canonical generator for the corresponding ideal in $V^4[x]$.*

The cyclic shifts of 0110 are

$$0110, 0011, 1001, 1100$$

Since $1001 = 1100 + 0110 + 0011$, $\langle 1100\rangle$ is spanned by

$$1100, 0110, 0011.$$

Sums of 0 of these word:

$$0000$$

Sums of two of these words:

$$1010, 0101, 1111.$$

Sum of all three:

$$1001.$$

So

$$\langle 1100\rangle = \{0000, 1100, 0110, 0011, 1001, 1010, 0101, 1111\}.$$

The non-zero codeword with the most trailing zeros is 1100. So the canonical generator is $1 + x$.

Theorem IX.28. *Let $C \subseteq \mathbb{F}_2^n$ be a cyclic code. Let g be the canonical generator of the ideal $C(x)$ in $V^n[x]$. Let $h \in \mathbb{F}[x]$ with $gh = x^n - 1$. Put $k = \deg h$ and $m = n - k = \deg g$.*

(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 matric for C.

(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) By IX.22 $C(x)$ 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 $d = d_0 d_1 \dots d_{n-1} \in \mathbb{F}_2^n$ and $d(x)$ the corresponding polynomial. By IX.22, $d(x) \in C(x)$ if and only if the coefficent a_s of x^s in $h \cdot d(x)$ is equal to 0 for all $k \le s < n$. We compute

$$a_s = \sum_{l=0}^{s} h_l d_{s-l}$$

Since $\deg h = k$, $h_l = 0$ for $l \ge k$. Put $i = s - k$. Then $0 \le i < n - k = m$ and

$$a_s = \sum_{l=0}^{k} h_l d_{s-l} = h_k d_i + h_{k-1} d_{i+1} + \dots + h_1 d_{s-1} + h_0 d_s$$

Hence

$$\begin{array}{lcl} a_k & = & h_k d_0 + h_{k-1} d_1 + h_{k-2} d_2 + \dots + h_1 d_{k-1} + h_0 d_k \\ a_{k+1} & = & \quad h_k d_1 + h_{k-1} d_2 + \dots + h_2 d_{k-1} + h_1 d_k + h_0 d_{k+1} \\ a_{k+2} & = & \quad\quad h_k d_2 + \dots + h_3 d_{k-1} + h_2 d_k + h_1 d_{k+1} + h_0 d_{k+2} \\ \vdots & \vdots & \quad\quad\quad \ddots \; \ddots \; \ddots \; \ddots \; \ddots \; \ddots \; \ddots \; \ddots \; \ddots \; \ddots \\ a_{n-2} & = & \quad\quad\quad\quad h_k d_{n-k-2} + \dots + \dots + h_1 d_{n-3} + h_0 d_{n-2} \\ a_{n-1} & = & \quad\quad\quad\quad\quad h_k d_{n-k-1} + \dots + h_2 d_{n-3} + h_1 d_{n-2} + h_0 d_{n-1} \end{array}$$

So $a_{k+i} = 0$ for $0 \le i < m$ if and only if $Hd = \vec{0}$. So H is indeed a check matrix for C.

□

Example IX.29. *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?*

We have $n = 7$, $m = \deg g = 4$ and $k = n - m = 7 - 4 = 3$. Since $g = 1 + x^2 + x^3 + x^4$ the first column of E is 1011100. So

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

To determine H we first compute $h = \frac{x^7+1}{g}$:

$$\begin{array}{r|llllllll}
 & x^3 & +\ x^2 & & +\ 1 & & & & \\ \cline{2-9}
x^4+x^3+x^2+1 & x^7 & & & & & & & +\ 1 \\
 & x^7 & +\ x^6 & +\ x^5 & & +\ x^3 & & & \\ \cline{2-9}
 & & x^6 & +\ x^5 & & +\ x^3 & & & +\ 1 \\
 & & x^6 & +\ x^5 & +\ x^4 & & +\ x^2 & & \\ \cline{3-9}
 & & & & x^4 & +\ x^3 & +\ x^2 & & +\ 1 \\
 & & & & x^4 & +\ x^3 & +\ x^2 & & +\ 1 \\ \cline{5-9}
 & & & & & & & & 0
\end{array}$$

Hence $h = x^3 + x^2 + 1$ and 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 :

$$Hd = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1+1+0+0 \\ 0+0+0+0 \\ 0+1+1+0 \\ 0+1+0+1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

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 and so $d \in C$.

IX.2 Definition of a family of BCH codes

Irreducible polynomials

Definition IX.30. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$. Then h is called irreducible if*

(i) $\deg h > 0$.

(ii) if $h = fg$ for some $f, g \in \mathbb{F}[x]$, then $\deg f = 0$ *or* $\deg g = 0$.

Remark IX.31. *Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$. Then h is irreducible if and only*

(i) $\deg h > 0$.

(ii) If $f \mid h$ for some $f \in \mathbb{F}[x]$, then $\deg f = 0$ *or* $\deg f = h$.

Proof. We may assume that $\deg h > 0$.

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 $\deg f = \deg h - \deg g = \deg f$.

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. □

Lemma IX.32. *Let R be a commutative ring and $r \in R$. Then $sr = 1$ for some $s \in R$ if and only if $\langle r \rangle = R$.*

Proof. 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 $R = \langle r \rangle$. Suppose that $R = \langle r \rangle$. Then $1 \in \langle r \rangle$ and so $1 = sr$ for some $s \in R$. □

Lemma IX.33. *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) $\mathbb{E}$ is a field.

(c) Let $h = \sum_{i=0}^n h_i x^i$. 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) Let $a, b \in \mathbb{F}$. Then $a, b, a+b$ and ab are polynomials of degree at most 0. Since $\deg h > 0$, we get $a \oplus b = a + b$ and $a \odot b = ab$. So $\mathbb{F}$ is a subfield of $\mathbb{F}^h[x]$.

(b) Let $0 \neq f \in \mathbb{F}^h[x]$ and put $I = \langle f \rangle$. Let g be the canonical generator for I. Then g divides h. Since $0 \neq f \in I$ we have $g \mid f$ and so $\deg g \leq \deg f < \deg h$ and $g \neq 0$. Since h is irreducible, this implies $\deg g = 0$. Hence $g \in \mathbb{F}$ and g^{-1} is a polynomial of degree 0 with $1 = g^{-1}g = g^{-1} \cdot g$. Thus by IX.32, $\langle g \rangle = \mathbb{F}^h[x]$. Since $\langle f \rangle = I = \langle g \rangle = \mathbb{F}^h[x]$, another application of IX.32 shows that $s \cdot f = 1$ for some $s \in \mathbb{F}^h[x]$. Hence $\mathbb{F}^h[x]$ is a field.

(c) Let $e \in \mathbb{E}$. Note that in E we have to use the operations $\oplus$ and $\odot$. So e is a root of $\sum_{i=0}^n h_i y^i$ in $\mathbb{E}$ if and only if

$$h_0 \oplus h_1 \odot e \oplus h_1 \odot e^{\odot 2} \oplus \ldots \oplus h_n \odot e^{\odot n} = 0$$

and if and only if

$$\overline{\sum_{i=0}^{m} h_i e^i} = 0$$

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 IX.34. *We will investigate $\mathbb{F}_2^h[x]$, where $h = 1 + x + x^3 \in \mathbb{F}_2[x]$.*

Note that neither x nor $x + 1$ divide h, so h is irreducible. Put $\mathbb{E} = \mathbb{F}_2^h[x]$. 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 we view α as element of the the field $\mathbb{E}$ (rather than an element of $\mathbb{F}[x]$). Then

$$1 + \alpha + \alpha^3 = \overline{1 + x + x^3} = \overline{h} = 0.$$

Also every element in $\mathbb{E}$ 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.

$$\begin{array}{lllllllllllll}
\alpha^0 & & & & & & & = & 1 & & & & \\
\alpha^1 & & & & & & & = & & & \alpha & & \\
\alpha^2 & & & & & & & = & & & & & \alpha^2 \\
\alpha^3 & & & & & & & = & 1 & + & \alpha & & \\
\alpha^4 & = & \alpha\alpha^3 & & & & & = & & & \alpha & + & \alpha^2 \\
\alpha^5 & = & \alpha\alpha^4 & = & a^2 + \alpha^3 & & & = & 1 & + & \alpha & + & \alpha^2 \\
\alpha^6 & = & \alpha\alpha^5 & = & \alpha + \alpha^2 + a^3 & = & \alpha + \alpha^2 + 1 + \alpha & = & 1 & & & + & \alpha^2 \\
\alpha^7 & = & \alpha\alpha^6 & = & \alpha + \alpha^3 & = & \alpha + 1 + \alpha & = & 1 & & & &
\end{array}$$

Hence

$$\mathbb{E}^\sharp = \mathbb{E} \smallsetminus \{0\} = \{\alpha^i \mid 0 \le i < 7\}$$

Since $h(\alpha) = 0$ we $h(\alpha^2) = h(\alpha)^2 = 0$. Let's verify this by direct computation

$$h(\alpha^2) = 1 + \alpha^2 + (\alpha^2)^3 = 1 + \alpha^2 + \alpha^6 = 1 + \alpha^2 + (1 + \alpha^2) = 0$$

also

$$h(\alpha^4) = 1 + \alpha^4 + (\alpha^4)^3 = 1 + \alpha^4 + \alpha^{12} = 1 + \alpha^4 + \alpha^5 = 1 + (\alpha + \alpha^2) + (1 + \alpha + \alpha^2) = 0$$

Thus $\alpha, \alpha^2, \alpha^4$ are the roots of $1 + x + x^3$.

From $\alpha^7 = 1$ we have $(\alpha^i)^7 = \alpha^{i7} = (\alpha^7)^i = 1$. So if $0 \neq e \in \mathbb{E}$, then $e^7 = 1$ and e is a root of $x^7 - 1$.

Note that $(1 + x + x^3)(1 + x) = 1 + x + x^3 + x + x^2 + x^4 = 1 + x^2 + x^3 + x^4$. So the long division in IX.29 shows that

$$x^7 - 1 = (1 + x)(1 + x + x^3)(1 + x^2 + x^3)$$

1 is a root of $1 + x$ and $\alpha, \alpha^2, \alpha^4$ are the root of $1 + x + x^3$. So α^3, α^4 and a^6 should be 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 a^5 are roots of $1 + x^2 + x^3$.

Lemma IX.35. *Let $n \in \mathbb{Z}^+$ and write*

$$x^n - 1 = f_0 f_1 \dots f_l$$

where $f_0, f_1, \dots, f_l$ are pairwise distinct irreducible polynomials in $\mathbb{F}_2[x]$ with $f_0 = x - 1 = 1 + x$. Let $C \subseteq \mathbb{F}_2^n$ be cyclic code and g the canonic generator for the ideal $C(x)$ in $V^n[x]$ corresponding to C. Then there exists uniquely determined $\epsilon_i \in \{0, 1\}$, $0 \leq i \leq l$ with

$$g = f_0^{\epsilon_0} \dots f_l^{\epsilon_l}$$

Moreover, $x^n - 1 = gh$ where

$$h = f_0^{\delta_0} \dots f_l^{\delta_l}$$

and $\delta_i = 1 - \epsilon_i$.

Proof. By IX.22 g divides $x^n - 1$ in $\mathbb{F}_2[x]$. The lemma now follows from A.13 □

Example IX.36. *Given that $x^7 - 1 = (1+x)(1+x+x^3)(1+x^2+x^3)$ is the factorization of $x^7 - 1$ into irreducible polynomials. Find a generating matrix and a check matrix for all the four dimensional cyclic codes of length 7.*

Let C be such a code, Let g be the canonical generator of C and $h \in \mathbb{F}_2[x]$ with $gh = x^7 - 1$. Then $\deg h = \dim C = 4$ and $\deg g = 7 - 4 = 3$. Then

$$g = (1 + x)^{\epsilon_1} (1 + x + x^3)^{\epsilon_2} (1 + x^2 + x^3)^{\epsilon_3}$$

for some $\epsilon_1, \epsilon_2, \epsilon_3 \in \{0, 1\}$. We have $3 = \deg g = \epsilon_1 + 3\epsilon_2 + 3\epsilon_3$ and we conclude that $g = 1 + x + x^3$ or $g = 1 + x^2 + x^3$. So either

$$g = 1 + x + x^3 \text{ and } h = (1 + x)(1 + x^2 + x^3) = (1 + x^2 + x^3) + (x + x^3 + x^4) = x^4 + x^2 + x + 1$$

or

$$g = 1 + x^2 + x^3 \text{ and } h = (1 + x)(1 + x + x^3) = (1 + x + x^3) + (x + x^2 + x^4) = x^4 + x^3 + x^2 + 1$$

Thus

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and } H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

or

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and } H = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

So both codes are Hamming codes.

Lemma IX.37. *Let $\mathbb{F}$ be a field and $0 \neq f \in \mathbb{F}[x]$. Let $a = \text{lead}(f)$ and $n = \deg f$. Then there exists a field $\mathbb{E}$ containing $\mathbb{F}$ and elements $\alpha_1, \alpha_2 \dots, \alpha_n$ in $\mathbb{E}$ such that*

$$f = a(x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_n)$$

Moreover, if $\mathbb{F}$ is finite we can choose $\mathbb{E}$ be be finite.

Such a field $\mathbb{E}$ is called a splitting field for f over $\mathbb{F}$.

Proof. The proof is by induction on $\deg f$. If $\deg f = 0$, 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 $\mathbb{K}$ is a field and there exists a root α of h in $\mathbb{K}$. Moreover, $\mathbb{K}$ is finite if $\mathbb{F}$ is finite. Then $x - \alpha$ divides h in $\mathbb{K}[x]$ and so $f = (x - \alpha)t$ for some $t \in \mathbb{K}[x]$. Since $\deg t = n - 1$, by induction there exist a field $\mathbb{E}$ containing $\mathbb{K}$ and elements $\alpha_1, \alpha_2 \dots, \alpha_{n-1}$ in $\mathbb{E}$ such that

$$t = a(x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_{n-1})$$

and $\mathbb{E}$ is finite if $\mathbb{K}$ is finite. Thus the lemma holds with $\alpha_n = \alpha$. □

Lemma IX.38. *Let $\mathbb{F}$ be a field and $\alpha_i, 1 \leq i \leq d$ be elements in $\mathbb{F}$. Let D be the $d \times n$ matrix*

$$D = [\alpha_i^j]_{\substack{1 \leq i \leq d \\ 0 \leq j < n}}.$$

and $c = c_0 c_1 \dots c_{n-1} \in \mathbb{F}^n$. The i-th coordinate of Dc is $\sum_{j=0}^{n-1} c_j \alpha_i^j$ and so $Dc = 0$ if and only if $\alpha_1, \alpha_2, \dots, \alpha_d$ all are roots of $c(x) = \sum_{j=0}^{n-1} c_j x^j$.

Proof. Just observe that the i'th coefficient of Dc 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 IX.39. *Let $\mathbb{F}$ and $\mathbb{E}$ be a fields with $\mathbb{F} \subseteq \mathbb{E}$. Let $f \in \mathbb{F}[x]$ and $\alpha \in \mathbb{E}$. Suppose that α is the root of some non-zero polynomial in $\mathbb{F}[x]$ and let $m \in \mathbb{F}[x]$ be a monic polynomial of minimal degree with respect to $m(\alpha) = 0$. Put $\mathbb{F}[\alpha] = \{f(\alpha) \mid f \in \mathbb{F}[x]\}$. Then*

(a) Let $f \in \mathbb{F}[x]$ and let r be the remainder of f when devided by m. Then $f(\alpha) = r(\alpha)$.

(b) $f(\alpha) = 0$ if and only if m divides f in $\mathbb{F}[x]$

(c) Let $f, g \in \mathbb{F}[x]$. Let r and s be the remainder of f and g when divided by m. Then

$$f(\alpha) = g(\alpha) \iff r = s \iff m \mid g - f.$$

(d) For each $e \in \mathbb{F}[a]$ there exists a unique $r \in \mathbb{F}^m[x]$ with $e = r(\alpha)$.

(e) Let $t = \deg m$. Then $(1, \alpha, \alpha^2, \dots, \alpha^{t-1})$ is an $\mathbb{F}$-basis for $\mathbb{F}[\alpha]$.

(f) m is the unique monic irreducible polynomial in $\mathbb{F}[x]$ with $m(\alpha) = 0$.

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 $r = 0$. So $f(\alpha) = 0$ if and only if $m \mid f$ and (b) holds.

(c)

$$\begin{array}{lcl}
 & f(\alpha) = g(\alpha) & \\
\iff & r(\alpha) = s(\alpha) & -\text{ (a)} \\
\iff & (r-s)(\alpha) = 0 & \\
\iff & m \mid r - s & -\text{ (b)} \\
\iff & r - s = 0 & -\deg(r-s) < \deg m \\
\iff & r = s &
\end{array}$$

(d) The existence of r follows from (a). The uniqueness from (c).

(e) Any $f \in \mathbb{F}^m[x]$ can be uniquely written as $f = \sum_{i=0}^{t-1} a_i x^i$ with $a_i \in \mathbb{F}$. Also $f(\alpha) = \sum_{i=0}^{t-1} a_i \alpha^i$. So (d) says that every $e \in \mathbb{F}[a]$ is a unique linear combination of $1, \alpha, \dots, \alpha^{t-1}$. Thus (e) holds.

(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 $h(\alpha) = 0$. Lets say $g(\alpha) = 0$. 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.11 implies $m = g$. Thus m is unique. □

Definition IX.40. *Let $\mathbb{E}$ be a field containing $\mathbb{F}_2$, $C \subseteq \mathbb{F}_2^n$ a linear code and H an $m \times n$-matrix with coefficents in $\mathbb{E}$. We say that H is a check matrix for C over $\mathbb{E}$ if*

$$C = \{c \in \mathbb{F}_2^n \mid Hc = 0\}$$

Lemma IX.41. *Let $C \in \mathbb{F}_2^n$ be a cyclic code with canonical generator $g \in \mathbb{F}_2[x]$. Suppose that $g = ag_1 \dots g_s$, where $g_1, \dots, g_s$ are pairwise distinct irreducible monic polynomials in $\mathbb{F}[x]$. 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) Each α_i is a root of one of the g_j's.

(ii) Each g_j has one of the α_i's as a root.

Put

$$H = [\alpha_i^j]_{\substack{1 \le i \le d \\ 0 \le j < n}}.$$

Then H is a check matrix for C over $\mathbb{E}$ and

$$C = \{c \in \mathbb{F}_2^n \mid c(\alpha_i) = 0 \textit{ for all } 1 \le i \le d\}$$

Proof. Let $c \in \mathbb{F}_2^n$, then $c \in C$ 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 devides $c(x)$.

Suppose g divides $c(x)$. Let $1 \leq i \leq d$. By (i) there exists $1 \leq j \leq s$ with $g_j(\alpha_i) = 0$. Since g_j divides g, $g(\alpha_i) = 0$ and since g divides $c(x)$, $c(\alpha_i) = 0$.

Suppose $c(\alpha_i) = 0$ for all $1 \leq i \leq d$. Let $1 \leq j \leq s$. By (ii) there exists $1 \leq i \leq d$ with $g_j(\alpha_i) = 0$. So by IX.39 g_j divides $c(x)$. Thus by A.13 g divides $c(x)$.

We proved that g divides $c(x)$ if and only if $c(\alpha_i) = 0$ for all $1 \leq i \leq d$. By IX.38 this holds if and only if $Hc = 0$. So H is check matrix for C. □

IX.3 Properties of BCH codes

Definition IX.42. *Let $\mathbb{E}$ be a finite field and put $n = |\mathbb{E}| - 1$. $\alpha \in \mathbb{E}$ is called a primitive element for $\mathbb{E}$ if*

$$\alpha^n = 1 \qquad \textit{and} \quad \mathbb{E} \setminus \{0\} = \{\alpha^i \mid 0 \leq i < n\}.$$

Lemma IX.43. *Every finite field has a primitive element.*

Proof. For a proof see A.18 in the appendix. □

Definition IX.44. *Let $\mathbb{E}$ be a finite field containing $\mathbb{F}_2$. Put $n = |\mathbb{E}^\sharp|$ and let α be a primitive element for $\mathbb{E}$. Let $1 \leq d < n$ and*

$$g = \operatorname{lcm}(m_{\alpha^i} \mid 1 \leq i < d).$$

Let $C \subseteq \mathbb{F}_2^n$ be the cyclic code with canonical generator g. Then C is called the BCH code of length n and designated distance d with respect to α.

Lemma IX.45. *Let $\mathbb{F}$ be a field and $\alpha_j, 0 \leq j < n$ pairwise distinct non-zero elements in $\mathbb{F}$. Let A be the $d \times n$ matrix*

$$A = [\alpha_j^i]_{\substack{1\leq i \leq d \\ 0 \leq j < n}}.$$

Then any d columns of A are linearly independent over $\mathbb{F}$.

Proof. Given $0 \leq s_1 < s_2 < \ldots < s_d < n$. Put $\beta_j = \alpha_{s_j}$ and consider the $d \times d$-matrix

$$B = [\beta_j^i]_{\substack{0\leq i < d \\ 1 \leq j \leq d}}.$$

Then β_j times column j of B is column s_j of A. Since $\beta_j \neq 0$ for all $1 \leq j \leq d$, the columns of B are linearly independent if and only if columns $s_1, s_2, \ldots s_d$ 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

$$B^{\mathrm{T}} = [\beta_i^j]_{\substack{1\leq i \leq d \\ 0 \leq j < d}}$$

By IX.38, $B^{\mathrm{T}}c = 0$ if and only if $\beta_1, \beta_2, \ldots, \beta_d$ are root of $c(x)$. Note that a non-zero polynomial of degree e has at most e roots. Since $c(x)$ is a polynomial of degree less than d, and the $\beta_1, \beta_2, \ldots \beta_d$ are d different elements of $\mathbb{F}$, we conclude that $\beta_1, \beta_2, \ldots, \beta_d$ are roots of $c(x)$ if and only if $c(x) = 0$ and so if and only if $c = 0$. □

Theorem IX.46. *Let $\mathbb{E}$ be a finite field containing $\mathbb{F}_2$ and $n = |\mathbb{E}| - 1$. Let C be the BCH-code of length n and designated distance d with respect to the primitive element α in $\mathbb{E}$. Let $\{m_{\alpha^i} \mid 1 \leq i < d\} = \{m_1, \ldots, m_s\}$, where the m_i's, $1 \leq i \leq s$ are pairwise distinct. Then*

(a) $g = m_1 m_2 \ldots m_s$.

(b) $H = [\alpha^{ij}]_{\substack{1 \leq i < d \\ 0 \leq j < n}}$ *is a check matrix for H over $\mathbb{E}$ and* $C = \{c \in \mathbb{F}_2^n \mid c(\alpha^i) = 0 \textit{ for all } 1 \leq i < d\}$.

(c) For $1 \leq i \leq s$ let α_i be a root of m_i in $\mathbb{E}$. Then $\tilde{H} = [\alpha_i^j]_{\substack{1 \leq i \leq s \\ 0 \leq j < n}}$ *is a check matrix for C and*

$$C = \{c \in \mathbb{F}_2^n \mid c(\alpha_i) = 0 \textit{ for all } 1 \leq i \leq s\}.$$

(d) C has minimum distance at least d

(e) $\dim C \geq n - \lfloor \frac{d}{2} \rfloor \log_2(n+1)$.

(f) If $d = 2r + 1$, then C is an r-error correcting code and $\dim C \geq n - r \log_2(n+1)$.

Proof. (a) Follows from $g = \operatorname{lcm}(m_{\alpha^i} | 1 \leq i < d)$.

(b) Since α^i is a root of m_{α^i} we conclude that each m_j has some $\alpha^i, 1 \leq i \leq d-1$ as a root and each α^i is a root of one the $m_j's$. So (b) follows IX.41.

(c) Since α_i is a root of m_i another application of IX.41 gives (c).

(d) By IX.45 applied with $\alpha_j = \alpha^j$, $0 \leq j < n$, any $d-1$-columns of H are linearly independent over $\mathbb{E}$. Hence the sum of any $d-1$-columns of H is non-zero and so $Hc \neq 0$ for any $0 \neq c \in \mathbb{F}_2^n$ with $\operatorname{wt}(c) \leq d-1$. Thus C has minimal weight at least d.

(e) Put $l = \log_2(n+1)$. Then $|E| = n + 1 = 2^l$. Put $t_i = \deg m_i$. By IX.39(e), $|\mathbb{F}_2[\alpha_i]| = 2^{t_i}$. Since $|\mathbb{F}_2[\alpha_i]| \leq |\mathbb{E}| = 2^l$ we have $t_i \leq l$. Thus

$$\deg g = \sum_{i=1}^{s} \deg m_i = \sum_{i=1}^{s} t_i \leq sl$$

By IX.22(f), $\dim C = n - \deg g \geq n - sl$.

Since α^{2j} is a root of m_{α^j}, we have $m_{\alpha^{2j}} = m_{\alpha^j}$ and so if $i = 2^k j$ with j odd, $m_{\alpha^i} = m_{\alpha^j}$. Thus

$$\{m_{\alpha^i} \mid 1 \leq i < d\} = \{m_{\alpha^j} \mid 1 \leq j < d, j \text{ odd}\}$$

Thus $s \leq \frac{(d-1)+1}{2}$ and $s \leq \lfloor \frac{d}{2} \rfloor$. So

$$\dim C \geq n - sl \geq n - \left\lfloor \frac{d}{2} \right\rfloor \log_2(n+1)$$

(f) Follows from (d) and (e) □

Example IX.47. *Let $\mathbb{E} = \mathbb{F}_2^{x^4+x+1}[x]$ and $\alpha = x \in \mathbb{E}$. Let C be the BCH-code of length 15 and designated distance 7 with respect to α. S Verify that $\mathbb{E}$ is a field and α is a primitive elements in $\mathbb{E}$. Determine the dimension of C, the minimal distance of C , the canonical generator for C and a check matrix over $\mathbb{E}$ for C.*

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 IX.33). Suppose that $x^4 + x + 1 = pq$ with $p, q \in \mathbb{F}_2[x]$. Then $\deg p + \deg q = 4$ and we may assume that $\deg p \leq 2$. We have $0^4 + 0 + 1 = 1$ and $1^4 + 1 + 1 = 1$ and so neither 0 nor 1 is a root of $x^4 + x + 1$. Hence neither 0 nor 1 is a root of p. Thus p is none of $x, x + 1, x^2 + 1, x^2 + x$ and x^2. To determined whether $x^2 + x + 1$ divides $x^4 + x + 1$ we use long devision of polynomials:

$$\begin{array}{rl}
 & 111 \\ \cline{2-2}
\multicolumn{1}{r|}{111} & 10011 \\
 & \underline{111} \\
 & 1111 \\
 & \underline{111} \\
 & 1
\end{array}$$

So the remainder of $x^4 + x + 1$ when divide by $x^2 + x + 1$ is 1. So $p \neq x^2 + x + 1$. We proved that p is neither a polynomial of degree 1 nor 2. Hence $\deg p = 0$ and $x^4 + x + 1$ is irreducible.

To show that α is a primitive element we compute the powers of α:

$$\begin{array}{lclclclcl@{\qquad}lclclclcl}
\alpha^0 & = & 1 & & & & & & & \alpha^8 & = & 1 & & & + & \alpha^2 & & \\
\alpha^1 & = & & & \alpha & & & & & \alpha^9 & = & & & \alpha & & & + & \alpha^3 \\
\alpha^2 & = & & & & & \alpha^2 & & & \alpha^{10} & = & 1 & + & \alpha & + & \alpha^2 & & \\
\alpha^3 & = & & & & & & & \alpha^3 & \alpha^{11} & = & & & \alpha & + & \alpha^2 & + & \alpha^3 \\
\alpha^4 & = & 1 & + & \alpha & & & & & \alpha^{12} & = & 1 & + & \alpha & + & \alpha^2 & + & \alpha^3 \\
\alpha^5 & = & & & \alpha & + & \alpha^2 & & & \alpha^{13} & = & 1 & & & + & \alpha^2 & + & \alpha^3 \\
\alpha^6 & = & & & & & \alpha^2 & + & \alpha^3 & \alpha^{14} & = & 1 & & & & & + & \alpha^3 \\
\alpha^7 & = & 1 & + & \alpha & & & + & \alpha^3 & \alpha^{15} & = & 1 & & & & & &
\end{array}$$

So α is indeed a primitive elements. To find the canonical generator g we need to determine m_{α^i} for $1 \leq i < 7$.

α, α^2 and α^4 are roots of $1 + x + x^4$ and so $m_\alpha = m_{\alpha^2} + m_{\alpha^4}$.

We look at the powers of α^3 as strings of length 4

$$\begin{aligned}
(\alpha^3)^0 &= 1 \leftrightarrow 1000 \\
(\alpha^3)^1 &= \alpha^3 \leftrightarrow 0001 \\
(\alpha^3)^2 &= \alpha^6 \leftrightarrow 0011 \\
(\alpha^3)^3 &= \alpha^9 \leftrightarrow 0101 \\
(\alpha^3)^4 &= \alpha^{12} \leftrightarrow 1111
\end{aligned}$$

The first four strings are linearly independent. The sum of all five is zero.

So α^3 and α^6 have minimal polynomial $1 + x + x^2 + x^3 + x^4$.

Powers of α^5:

$$\begin{aligned}
(\alpha^3)^0 &= 1 \leftrightarrow 1000\\
(\alpha^5)^1 &= \alpha^3 \leftrightarrow 0110\\
(\alpha^5)^2 &= \alpha^{10} \leftrightarrow 1110
\end{aligned}$$

The first two strings are linearly independent and the sum of all three is zero. So α^5 is a root of $1 + x + x^2$. Thus we can choose $m_1 = 1 + x + x^4$, $m_2 = 1 + x + x^2 + x^3 + x^4$ and $m_3 = 1 + x + x^2$. Hence

$$g = (1 + x + x^4)(1 + x + x^2 + x^3 + x^4)(1 + x + x^2)$$

We compute

$$\begin{array}{ccccccccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10\\
\hline
1 & 1 & 1 & 1 & 1 & & & & & & \\
 & 1 & 1 & 1 & 1 & 1 & & & & & \\
 & & & & 1 & 1 & 1 & 1 & 1 & & \\
\hline
1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & & \\
 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & \\
 & & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1\\
\hline
1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1
\end{array}$$

Thus $g = 1 + x + x^2 + x^4 + x^5 + x^8 + x^{10}$. Since $\deg g = 10$, $\dim C = 15 - 10 = 5$. The string of length 15 corresponding to g is 111011001010000. This is a codeword of weight 7 and so C has minimum distance at most 7. By IX.46 C has minimum weight at least 7 and so $\delta(C) = 7$.

Since α, α^3 and α^5 are roots of m_1, m_2 and m_3 respectively we obtain the following check matrix:

$$H = \begin{bmatrix}
1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 & \alpha^6 & \alpha^7 & \alpha^8 & \alpha^9 & \alpha^{10} & \alpha^{11} & \alpha^{12} & \alpha^{13} & \alpha^{14}\\
1 & \alpha^3 & \alpha^6 & \alpha^9 & \alpha^{12} & 1 & \alpha^3 & \alpha^6 & \alpha^9 & \alpha^{12} & 1 & \alpha^3 & \alpha^6 & \alpha^9 & \alpha^{12}\\
1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10} & 1 & \alpha^5 & \alpha^{10}
\end{bmatrix}$$

Lemma IX.48. *Let $C \subseteq \mathbb{F}_2^n$ be a BCH code with respect to the primitive element α. Let (c, z) be a 1-bit error for C. Then $z(\alpha) \neq 0$ and if $i \in \mathbb{N}$ with $0 \leq i < n$ and $z(\alpha) = \alpha^i$, then $c_i \neq z_i$.*

Proof. Since (c, z) is 1-bit error, $c_j \neq z_j$ for a unique $0 \leq j < n$. Thus $z(x) = c(x) + x^j$. By IX.41 $c(\alpha) = 0$ for all $c \in C$ and so

$$z(\alpha) = c(\alpha) + \alpha^j = \alpha^j.$$

Since α is a primitive element, the elements $\alpha^s, 0 \leq s < n$, are pairwise distinct. Hence $j = i$ and so $z_i \neq c_i$. □

Example IX.49. *Let C be the code from example IX.47. Does there exist a 1-bit error (c, z) with*

$$z = 011111110000000?$$

We have $z(\alpha) = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 + \alpha^6 + \alpha^7$. We calculate the remainder of this polynomial when divided by $x^4 + x + 1$.

$$\begin{array}{r|cccccccc}
\multicolumn{1}{r}{} & 1&1&1&&&&& \\ \cline{2-9}
10011 & 1&1&1&1&1&1&1&0 \\ \cline{1-1}
\multicolumn{1}{r}{} & 1&0&0&1&1&&& \\ \cline{2-9}
\multicolumn{1}{r}{} & &1&1&0&0&1&1&0 \\
\multicolumn{1}{r}{} & &1&0&0&1&1&& \\ \cline{3-9}
\multicolumn{1}{r}{} & &&1&0&1&0&1&0 \\
\multicolumn{1}{r}{} & &&1&0&0&1&1& \\ \cline{4-9}
\multicolumn{1}{r}{} & &&&&1&1&0&0
\end{array}$$

Thus $z(\alpha) = \alpha^3 + \alpha^2 = \alpha^2(1+\alpha) = \alpha^2\alpha^4 = \alpha^6$. So if (c, z) is a 1-bit error, then the error occurred in bit 7:

$$c = 0111110100000000$$

We compute $c(\alpha^3)$ and $c(\alpha^5)$ to check whether c is in the code:

$$c(\alpha^5) = 0 + \alpha^5 + \alpha^{10} + 1 + \alpha^5 + \alpha^{10} + 0 + \alpha^5 = 1 + \alpha^5 \neq 0$$

So c is not in the code. Thus z cannot be the result of a 1-bit-error.

Chapter X

Cryptography in theory and practice

X.1 Encryption in terms of a channel

Definition X.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 an decryption function.

Example X.2. *Let $\mathcal{M} = \mathcal{C} = \mathbb{A} = \{A, B, C, D, \dots, Z, \sqcup\}$, $\mathcal{K} = \{0, 1, \dots, 25\}$ and for $k \in \mathcal{K}$, $E_k = D_k = c_k$, where c_k is the shift by k-letters defined in Example I.20. Note that D_{26-k} is the inverse of E_k, so this is indeed a cryptosystem.*

Definition X.3. *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*

$$\begin{array}{rccc} u: & \mathcal{M} \times \mathcal{K} \times \mathcal{C} & \to & [0,1] \\ & (m,k,c) & \to & \begin{cases} p_m r_k & \text{if } c = E_k(m) \\ 0 & \text{if } c \neq E_k(m) \end{cases} \end{array}$$

t, q, and s are defined to by 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 plain text message m was encrypted via the key k and the cipher text c was obtained.

Lemma X.4. *With the notation as in X.3*

(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) Γ is a channel associated to the distribution t on $\mathcal{M} \times \mathcal{C}$.

(d) E_k is 1-1 for all $k \in \mathcal{K}$.

(e)

$$q_c = \sum_{\substack{(m,k)\in\mathcal{M}\times\mathcal{K}\\ E_k(m)=c}} p_m r_k.$$

(f) $s_{kc} = 0$ if $c \notin E_k(\mathcal{M})$ and $s_{k,c} = p_m r_k$ if $c \in E_k(\mathcal{M})$ and m is the unique element of $\mathcal{M}$ with $E_k(m) = c$.

(g) $H(s) = H(p \otimes r) = H(p) + H(r)$.

(h) $H(r|q) = H(r) + H(p) - H(q)$.

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, then $c^* = E_k(m)$ is the only 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 IV.6 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 in the appendix.

(c) We will first verify that Γ is a channel: Let $m \in \mathcal{M}$, then

$$\sum_{c\in\mathcal{C}} \Gamma_{mc} = \sum_{c\in C} \Big(\sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k \Big) = \sum_{\substack{k\in\mathcal{K}, c\in\mathcal{C}\\ E_k(m)=c}} r_k = \sum_{k\in\mathcal{K}} r_k = 1$$

So Γ is indeed a channel.

Since $u_{mkc} = 0$ for $c \neq E_k(m)$ we have

$$t_{mc} = \sum_{k\in\mathcal{K}} u_{mkc} = \sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} u_{mkc} = \sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} p_m r_k = p_m \sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k = p_m \Gamma_{mc}$$

Thus $t = \mathrm{Diag}(p)\Gamma$ and so (see Definition V.10(d)) Γ is a channel associated to t.

(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 p and q are the marginal distributions of t, $p = q\Gamma$. So using (c),

$$q_c = \sum_{m\in\mathcal{M}} p_m \Gamma_{mc} = \sum_{m\in\mathcal{M}} p_m \Big(\sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k \Big) = \sum_{\substack{(m,k)\in\mathcal{M}\times\mathcal{K}\\ E_k(m)=c}} p_m r_k.$$

(f) Since s is the marginal distribution of u on $\mathcal{K} \times \mathcal{C}$, $s_{kc} = \sum_{m\in\mathcal{M}} u_{mkc} = \sum_{\substack{m\in\mathcal{M}\\ E_k(m)=c}} p_m r_k$

Since E_k is 1-1, there either exists a unique $m \in \mathcal{M}$ with $E_k(m) = c$ or there exists no element $m \in \mathcal{M}$ with $E_k(m) = c$. This gives (f).

(g) Since p and r are independent with respect to $p \otimes r$, IV.9 gives

$$H(p \otimes r) = H(p) + H(r)$$

Consider the function:

$$\pi : \mathcal{M} \times \mathcal{K} \to \mathcal{K} \times \mathcal{C}, \quad (m,k) \to (k, E_k(m))$$

Since each E_k is 1-1, also π is 1-1 and so π is a bijection from $\mathcal{M} \times \mathcal{K}$ to $\operatorname{Im}\pi$. Let $k \in \mathcal{K}$ and $c \in \mathcal{C}$. If $(k,c) = \pi(m,k)$ for some $m \in \mathcal{M}$, then (e) shows that

$$s_{\pi(m,k)} = s_{kc} = p_m r_k = (p \otimes r)_{mk}$$

Since π is a bijection from $\mathcal{M} \times \mathcal{K}$ to $\operatorname{Im}\pi$, V.17(a) gives

$$H(p \otimes r) = H(s|_{\operatorname{Im}\pi})$$

Also $s_{kc} = 0$ if $(k,c) \notin \operatorname{Im}\pi$. So

$$H(s|_{\operatorname{Im}\pi}) = H(s)$$

and (g) holds.

(h) Using the definition of $H(r \mid q)$ and (g),

$$H(r \mid q) = H(s) - H(q) = H(p) + H(r) - H(q).$$

□

Definition X.5. *(a) The* key equivocation *of a cryptosystem with respect to the probability distribution p and r on $\mathcal{M}$ and $\mathcal{K}$, respectively, is*

$$H(r|q)$$

(b) A cryptosytem is said to have perfect secrecy *with respect to the probability distribution r on $\mathcal{K}$ if*

$$I(p,q) = 0.$$

for all probability distributions p on $\mathcal{M}$.

Here Γ, r, s and q are as define in X.3.

$H(r|q)$ measures the amount of information obtained about the keys by observing ciphertext messages.

$I(p,q)$ measures the dependency of p and q. $I(p,q) = 0$ means that p and q are independent. So knowing the ciphertext does not reveal any information about the plaintext messages.

X.2 Perfect secrecy

Theorem X.6. *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 the probability distribution r.

(b) p and q are independent for all probability distributions p on $\mathcal{M}$.

(c) There exists a positive probability distribution p on $\mathcal{M}$ such that p and q are independent.

(d) There exists a positive probability distribution p on $\mathcal{M}$ such that $\Gamma_{mc} = q_c$ for all $m \in \mathcal{M}$ and all $c \in \mathcal{C}$,

(e) 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 $m \in \Gamma, c \in \mathcal{C}$.

(f) $\Gamma_{mc} = q_c$ for all probability distributions p on $\mathcal{M}$, all $m \in \mathcal{M}$ and all $c \in \mathcal{C}$:

Proof. (a) $\Longleftrightarrow$ (b) : By V.19 $I(p,q) = 0$ if and only if p and q are independent.

(b) $\Longrightarrow$ (c): Just choose p to be the equal probability distribution on $\mathcal{M}$.

(c) $\Longrightarrow$ (d): Since p and q independent 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 $\Gamma_{mc} = q_c$.

(d) $\Longrightarrow$ (e): All entries in column c of Γ are equal to q_c.

(e) $\Longrightarrow$ (f): 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

$$q_c = \sum_{m\in\mathcal{M}} p_m\Gamma_{mc} = \sum_{m\in\mathcal{M}} p_m l_c = \left(\sum_{m\in\mathcal{M}} p_m\right) l_c = 1l_c = l_c$$

and so (f) holds.

(f) $\Longrightarrow$ (b): 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 (b) holds. □

Corollary X.7. *If a cyrptosystem 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 X.6 $\Gamma_{mc} = q_c > 0$ for all $m \in \mathcal{M}$. By X.4

$$\sum_{\substack{k\in\mathcal{K}\\ E_k(m)=c}} r_k = \Gamma_{mc} > 0$$

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 $|\mathcal{M}| \leq |\mathcal{K}|$. □

X.3 The one-time pad

Definition X.8. *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 $\Omega(\mathbb{F}^n) = \left(\mathbb{F}^n, \mathbb{F}^n, (E_k)_{k\in\mathbb{F}^n}, (D_k)_{k\in\mathbb{F}^n}\right)$ is called the one-time pad determined by $\mathbb{F}^n$.

Lemma X.9 (One-Time Pad)**.** *Any one-time pad is a cryptosystem and has perfect secrecy with respect 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.

Put $e = \frac{1}{|\mathbb{F}^n|}$. Then $r_k = e$ for all $k \in \mathcal{K}$.

Let $m \in \mathcal{M}$ and $c \in \mathcal{C}$. By definition 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 X.6 the one-time pad has perfect secrecy.

□

X.4 Iterative methods

Definition X.10. *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 $X_{i+1}, 1 \leq i \leq r$ 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 $Y_{i+1}, 1 \leq i \leq r$ inductively by

$$Y_{i+1} = Y_{i-1} - F(k_i, Y_i), 1 \leq i \leq r$$

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 X.11. *Any Feistel system is a cryptosystem.*

Proof. Let $k = (k_1, k_2, \dots, k_r) \in \mathcal{K}$. 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 $Y_i, 0 \leq i \leq r + 1$ as above, but with respect to the key $k^* = (k_r, k_{r-1}, \dots, k_1)$. Note that $k^*_i = k_{r+1-i}$. For $0 \leq i \leq r + 1$, 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 $0 \leq i \leq r + 1$ 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 X.12. *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 $D_{k^*}(E_k(m)) = (101, 110) = m$.

X.5 The Double-Locking Procedure

Two cryptosystems Ω and $\widetilde{\Omega}$ are called compatible if $\mathcal{M} = \mathcal{C} = \widetilde{\mathcal{M}} = \widetilde{\mathcal{C}}$.

Given compatible cryptosystems Ω and $\widetilde{\Omega}$, 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

$$m_4 = \left(\tilde{D}_{\tilde{k}^*} \circ D_{k^*} \circ \tilde{E}_{\tilde{k}} \circ E_k\right)(m_0)$$

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

$$\tilde{D}_{\tilde{k}^*} \circ D_{k^*} \circ \tilde{E}_{\tilde{k}} \circ E_k = \left(\tilde{D}_{\tilde{k}^*} \circ \tilde{E}_{\tilde{k}}\right) \circ \left(D_{k^*} \circ E_k\right) = \mathrm{id}_{\mathcal{M}} \circ \mathrm{id}_{\mathcal{M}} = \mathrm{id}_{\mathcal{M}}$$

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 X.13. *Suppose Ω and $\widetilde{\Omega}$ 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 X.14. *Let Ω and $\widetilde{\Omega}$ be compatible cryptosystem. Let β be an encryption function in $\widetilde{\Omega}$ 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 $m_0 = D_l(m_1)$.

Chapter XI

The RSA cryptosystem

XI.1 Public-key cryptosytems

Definition XI.1. *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 all $a \in A$

$$D_{k_a^*} \circ E_{k_a} = \mathrm{id}_{\mathcal{M}}$$

k_a *is called a public key and* k_a^* *a private key.*

In public key cryptography, all the ingredients except the privat key are know to the public. Anybody then can encrypt a message using the public key k and the publicly known function E_k. But only somebody who knows the private key k^* is able to decrypt the encrypted message using the function D_l. This can only work if it is virtually impossible to determine k^* from k. In particular, A has to be really large, since otherwise one can just compute all the possible pairs (k, k^*) 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.

XI.2 The Euclidean Algorithim

Lemma XI.2. *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 $d \leq e$.

Similarly, e divides b and r and so also $a = qb + r$. Thus e is a common divisor of a nd b and so $e \leq d$. Hence $e = d$. □

Theorem XI.3 (Euclidean Algorithm)**.** *Let* a *and* b *be integers and let* E_{-1} *and* E_0 *be the equations*

$$\begin{array}{cccccc} E_{-1} & : & a & = & 1a & + & 0b \\ E_0 & : & b & = & 0a & + & 1b \end{array},$$

and suppose inductively we defined equation $E_k, -1 \leq k \leq i$ of the form

$$E_k \quad : \quad r_k \quad = \quad x_k a \quad + \quad y_k b \ .$$

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 algortithm will terminate in finitely many steps.

From $r_{i-1} = q_{i+1} r_i + r_{i+1}$ and XI.2 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 XI.4. *Let $a = 1492$ and $b = 1066$. Then*

$$\begin{array}{rll}
1492 = & 1 \cdot 1492 + 0 \cdot 1066 & \\
1066 = & 0 \cdot 1492 + 1 \cdot 1066 & q_1 = 1 \\
426 \ = & 1 \cdot 1492 - 1 \cdot 1066 & q_2 = 2 \\
214 \ = & -2 \cdot 1492 + 3 \cdot 1066 & q_3 = 1 \\
212 \ = & 3 \cdot 1492 - 4 \cdot 1066 & q_4 = 1 \\
2 \ = & -5 \cdot 1492 + 7 \cdot 1066 & q_5 = 106 \\
0 & &
\end{array}$$

So $\gcd(1492, 1066) = 2$ *and* $2 = -5 \cdot 1492 + 7 \cdot 1066$.

Definition XI.5. *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. We will sometimes also use the notation $\overline{a}$ for $[a]_n$.

The relation '$\equiv \pmod{n}$' on $\mathbb{Z}$ is defined by

$$a \equiv b \pmod{n} \qquad \Longleftrightarrow \qquad n \mid b - a$$

Lemma XI.6. *Let $a, b, a', b', n \in \mathbb{Z}$ with $n > 0$. If*

$$a \equiv a' \pmod{n} \qquad \textit{and} \qquad b \equiv b' \pmod{n}$$

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 XI.7. *Let $n \in \mathbb{Z}^+$ and $d \in \mathbb{Z}_n^*$. 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 XI.6 $ed \equiv rd \equiv 1 \pmod{n}$.

In particular, $n \mid ed - 1$ and any divisor of n and e will divide 1. Thus $\gcd(n, e) = 1$ and $e \in \mathbb{Z}_n^*$. □

Lemma XI.8. *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 XI.2. □

Lemma XI.9. *Let $n \in \mathbb{Z}^+$, $a, b \in \mathbb{Z}_n$ and $d \in \mathbb{Z}_n^*$.*

(a) If $a \equiv b \pmod{n}$, then $a = b$.

(b) If $[ad]_n = [bd]_n$, then $a = b$.

Proof. (a) Since $a \equiv b \pmod{n}$, $n \mid a - b$. Since $a, b \in \mathbb{Z}_n^*$ we have $|a - b| < n$. Thus $a - b = 0$ and $a = b$.

(b) By XI.7 there exists $e \in \mathbb{Z}_n^*$ with $de \equiv 1 \pmod{n}$. Note that $ad \equiv bd \pmod{n}$ and so $ead \equiv ebd \pmod{n}$ and $dea \equiv deb \pmod{n}$. Thus $a \equiv b \pmod{n}$ and so by (a) $a = b$. □

Lemma XI.10. *Let $n, m \in \mathbb{Z}^+$ with $\gcd(n, m) = 1$. Then $\phi(nm) = \phi(n)\phi(m)$.*

Proof. Consider the map

$$\alpha : \mathbb{Z}_{nm} \to \mathbb{Z}_n \times \mathbb{Z}_m, a \to ([a]_n, [a]_m)$$

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 divided $a - b$ and since $\gcd(n, m) = 1$, $nm \mid a - b$. Hence $a = b$ by XI.9(a). So α is 1-1. Since $|\mathbb{Z}_{nm}| = nm = |\mathbb{Z}_n \times \mathbb{Z}_m|$, α is also onto.

Since $\gcd(n, m) = 1$, $\gcd(a, nm) = \gcd(a, n)\gcd(a, m)$. Hence $\gcd(a, nm) = 1$ if and only if $\gcd(a, n) = 1$ and $\gcd(a, m)$ is 1. By XI.8 this holds if and only if $\gcd([a]_n, n) = 1$ and $\gcd([a]_m, =)1$. We proved 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

$$\alpha^* : \mathbb{Z}_{nm}^* \to \mathbb{Z}_n^* \times \mathbb{Z}_m^*,\ a \to ([a]_n, [a]_m).$$

So

$$\phi(nm) = |\mathbb{Z}_{nm}^*| = |\mathbb{Z}_n^* \times \mathbb{Z}_m^*| = \phi(n)\phi(m).$$

□

Corollary XI.11. *Let p and q be distinct primes. The $\phi(p) = p-1$ and $\phi(pq) = (p-1)(q-1) = pq+1-(p+q)$.*

Proof. $\mathbb{Z}_p^* = \{1, 2, \dots, p-1\}$ and so $\phi(p) = p-1$. By XI.10 $\phi(pq) = \phi(p)\phi(q) = (p-1)(q-1)$. □

Lemma XI.12. *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 XI.8, also $\gcd(\overline{ab}, n) = 1$. Thus $\overline{ab} \in \mathbb{Z}_n^*$ and we obtain an map

$$\alpha : \mathbb{Z}_n^* \to \mathbb{Z}_n^*, b \to \overline{ab}.$$

By XI.9, α is $1-1$ and so also onto.
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 XI.6 that

$$\prod_{b \in \mathbb{Z}_n^*} b \equiv \prod_{b \in \mathbb{Z}_n^*} ab \pmod{n}.$$

Put $e = \prod_{b \in \mathbb{Z}_n^*} b$. Then the above equation reads

$$1e = e \equiv a^{\phi(n)} e \pmod{n}.$$

Thus by XI.9

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

□

XI.3 Definition of the RSA public-key cryptosystem

Definition XI.13. *Let $\mathcal{M}$ be any alphabet of plain text messages and c a positive integer. Put*

$$\mathcal{C} = \left\{ n \in \mathbb{Z}^+ \mid n \le c \right\}, \quad N = \left\{ n \in \mathcal{C} \mid \phi(n) \ge |\mathcal{M}| \right\}, \quad \textit{and} \quad \mathcal{K} = \left\{ (n, d) \mid n \in N, d \in \mathbb{Z}_{\phi(n)}^* \right\}$$

For $n \in N$ let $\alpha_n : \mathcal{M} \to \mathbb{Z}_n^*$ be a code and $\beta_n : \mathbb{Z}_n \to \mathcal{M}$ a function with $\beta_n(\alpha_n(m)) = m$ for all $m \in \mathcal{M}$.

Given a key $(n, d) \in \mathcal{K}$. For $m \in \mathcal{M}$ define

$$E_{n,d}(m) = [\alpha_n(m)^d]_n$$

and for $z \in \mathcal{C}$ define

$$D_{n,d}(z) = \beta_n\left([z^d]_n\right)$$

Let π be a set of primes with $|\mathcal{M}| \le (p-1)^2 \le c$ for all $p \in \pi$. Put

$$A = \{(p, q, d, e) \mid p, q \in \pi,\ p \ne q,\ d, e \in \mathbb{Z}_{\phi(pq)}^*,\ de \equiv 1 \,(\mathrm{mod}\ \phi(pq))\}$$

Define

$$\xi : A \to \mathcal{K} \times \mathcal{K}, \quad (p, q, d, e) \to \left((pq, d), (pq, e)\right)$$

Then $\left(\mathcal{M}, \mathcal{C}, (E_{n,d})_{(n,d) \in \mathcal{K}}, (D_{n,d})_{(n,d) \in \mathcal{K}}, \xi\right)$ is called an RSA public-key cryptosystem.

Theorem XI.14. *Any RSA public-key cryptosystem is indeed a public-key cryptosystem.*

Proof. Let $(n,d) \in \mathcal{K}$ and $e \in \mathbb{Z}^*_{\phi(n)}$ with

$$de \equiv 1 \pmod{\phi(n)}.$$

We will show that $E_{(n,d)}$ is a left inverse of $D_{(n,d)}$. For this let $m \in \mathcal{M}$ and put $w = \alpha_n(m)$. Then $w \in \mathbb{Z}^*_n$ and $E_{n,d}(m) = [w^d]_n$. Note that $de = 1 + q\phi(n)$ for some $q \in \mathbb{Z}$ and so

$$w^{de} \equiv w^{1+q\phi(n)} \equiv w(w^{\phi(n)})^q \equiv w1^q \equiv w \pmod{n}$$

Hence

$$\Big[\big([w^d]_n\big)^e\Big]_n = \Big[w^{de}\Big]_n = w$$

and so

$$D_{n,e}\Big(E_{n,d}(m)\Big) = \beta_n\Big(\big[E_{n,d}(m)^e\big]_n\Big) = \beta_n\left(\Big[\big([w^d]_n\big)^e\Big]_n\right) = \beta_n(w) = \beta_n\big(\alpha_n(m)\big) = m.$$

Therefore $D_{n,e}$ is a left-inverse of $E_{n,d}$. Thus $\big(\mathcal{M}, \mathcal{C}, (E_{n,d})_{(n,d)\in\mathcal{K}}, (D_{n,d})_{(n,d)\in\mathcal{K}}\big)$ is cryptosystem.

Together with the definition of ξ we also see that $D_{n,e}$ is a left inverse of $E_{n,d}$, whenever $\big((n,d),(n,e)\big) = \xi(p,q,d,e)$ for some $(p,q,d,e) \in A$. Hence any RSA public-key cryptosystem is a public-key cryptosystem. □

Example XI.15. *Let*

$$\mathcal{M} = \mathbb{A} = \{\sqcup, A, \dots, Z\} = \{l_0, l_1 \dots, l_{26}\}$$

and $c = 1000$*. For* $n \in N$ *let*

$$\mathbb{Z}^*_n = \{a_{1,n}, \dots, a_{\phi(n),n}\}$$

with $a_{i,n} < a_{i+1,n}$ *for* $1 \le i < \phi(n)$ *. Define*

$$\alpha_n : \mathcal{M} \to \mathbb{Z}^*_n, l_i \to a_{i+1,n} \quad \textit{and} \quad \beta_n : \mathbb{Z}_n \to \mathcal{M} j \to \begin{cases} l_{[i-1]_{27}} & \textit{if } j \in \mathbb{Z}^*_n \textit{ and so} j = a_{n,i} \textit{ for some } 1 \le i \le \phi(n) \\ 0 & \textit{if } j \notin \mathbb{Z}^*_n \end{cases}$$

Compute $c = E_{(667,5)}(K)$*. Find the inverse key* k^* *for* $(667,5)$ *and verify that* $D_{k^*}(c) = K$*.*

$K = l_{11}$ and $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ all are coprime to 667. So

$$\alpha_{667}(K) = 12$$

We need to compute $[12^5]_{667}$:

$$12^2 = 144$$

$$12^3 = \begin{array}{r} 1440 \\ +\ \ 288 \\ \hline 1728 \end{array}$$

$$1728 - 3 \cdot 667 = -(2001 - 1728) = -273$$

$$-273 \cdot 12 = -\begin{array}{r} 2730 \\ +\ \ 546 \\ \hline 3286 \end{array}$$

$$-3286 + 5 \cdot 667 = 3335 - 3286 = 59$$

$$59 * 12 = 720 - 12 = 708$$

$$708 - 667 = 41$$

Hence $E_{667,5}(K) = 41$.

To compute e we need to factorize 29. 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 $\phi(667) = 667 - (23 + 29) + 1 = 668 - 52 = 616$.

$$1 = 616 - 123 \cdot 5$$

So $e = 616 - 123 = 493$. Since $41^{493} \equiv 12 \pmod{667}$, we get $D_{667,493}(41) = \beta_{667}(12) = l_{11} = K$.

Appendix A

Rings and Field

A.1 Basic Properties of Rings and Fields

Definition A.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',* *[additive identity]*
such that $a+0_R=a=0_R+a$ *for all* $a\in R$;

(Ax 5) *for each $a\in R$ there exists an element in R, denoted by $-a$* *[additive inverses]*
and called 'negative a', such that $a+(-a)=0_R$;

(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.2. *A ring* $(R,+,\cdot)$ *is called* commutative *if*

(Ax 9) $ab=ba$ *for all* $a,b\in R$. *[commutative multiplication]*

Definition A.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.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} *and called ' a inverse ', such that* $a \cdot a^{-1} = 1_R = a^{-1} \cdot a$*;* *[multiplicative inverses]*

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.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) $-(a - b) = -(a + (-b)) \overset{\text{(i)}}{=} -a + (-(-b)) \overset{\text{(h)}}{=} -a + b = b + (-a) = b - a$.

(k) By (g) $-a = -(1a) = (-1)a$. □

A.2 Polynomials

Definition A.6. *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*

$$a_0 + a_1 x + \ldots + a_n x^n.$$

Let $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{i=0}^{m} b_i x^i$ 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

$$fg = \sum_{k=0}^{n+m} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) x^i$$

Define $\deg f = \max\{i \mid a_i \neq 0\}$ with $\deg f = -\infty$ if $f = 0$.

Lemma A.7. *Let R be a ring. Then $R[x]$ is a ring. If R is commutative, so is $R[x]$.*

Proof. Readily verified. □

Lemma A.8. *Let $\mathbb{F}$ be a field and $f, g \in \mathbb{F}[x]$. Then*

(a) $\deg(f + g) \leq \max(\deg f, \deg g)$.

(b) $\deg fg = \deg f + \deg g$.

(c) If $f \neq 0$, then $\deg g \leq \deg fg$.

Proof. Readily verified. □

Lemma A.9. *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 $\odot$:

1°. *$f \oplus g = \overline{f + g}$ and $f \odot g = \overline{fg}$.*

By IX.8(a)

2°. *$\overline{f} = f$ for all $f \in \mathbb{F}^h[x]$.*

By IX.11

3°. *$\overline{f} \oplus \overline{g} = \overline{f} \oplus g = f \oplus g = \overline{f + g}$ 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) 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.10. *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 IX.33, $\mathbb{F}^h[x]$ is a field and we conclude that $\overline{f} = 0$ or $\overline{g} = 0$. Hence $h \mid f$ or $h \mid g$. □

Lemma A.11. *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.12. *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.10, g divides h or f_r. If g divides f_r, then by A.11 $h = f_r$. 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.13. *Let $\mathbb{F}$ be a field and $0 \neq f \in \mathbb{F}[x]$. Put $a =$* $\text{lead}(f)$.

(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 $b =$ $\text{lead}(g)$. *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 $b = \text{lead}(g)$, $c = \text{lead}(h)$ 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.12 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 $1 \leq i \leq r$. 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.12, $t = f_i$ for some $1 \leq i \leq t$. Without loss $i = 1$. Then

$$\tilde{g}h = af_2 \dots f_n$$

Note that $\text{lead}(\tilde{g}) = \text{lead}(g) = b$. 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 $2 \leq i \leq r$. Thus (b) holds with $\epsilon_1 = 1$ and $\delta_1 = 0$. □

A.4 Primitive elements in finite field

Lemma A.14. *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 $e^i, 0 \leq i < |e|$, 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:

1°. *Let s be a positive integer, then $e^i, 0 \leq i \leq s$ are pairwise distinct if and only $e^i \neq 1$ for all $1 \leq i \leq s$.*

Indeed $e^i \neq e^j$ for all $0 \leq i < j \leq s$, if and only if $e^{j-i} \neq 1$ for all $0 \leq i < j \leq s$ and so if and only $e^i \neq 1$ for some $1 \leq i \leq s$.

(a): Since $|\mathbb{E}^\sharp| = t$, the elements $e^i, 0 \leq i \leq t$ cannot be pairwise distinct. So by (1°) there exists $1 \leq m \leq t$ with $e^m = 1$.

(b) By minimality of $|e|$, $e^i \neq 1$ for all $1 \leq i < |e|$. So (b) follows from (1°).

(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.15. *Let n and d be positive integers with $d \mid n$. Define*

$$D_d(n) = \{m \mid 0 \leq m < n, \gcd(n,m) = d\}.$$

Then

(a) $D_d(n) = \{ed \mid e \in \mathbb{Z}^*_{\frac{n}{d}}\}$.

(b) $|D_d(n)| = \phi(\frac{n}{d})$.

(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 $ed \in D_d(n)$ 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.16. *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

$$x^n - 1 = (x-1)(x-\alpha)\ldots(x-\alpha^{n-1}).$$

Lemma A.17. *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.18. *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 $A_n \neq \emptyset$. Then $\mathbb{E}^\sharp$ has an element of order n and so by A.17, $|A_n| = \phi(n)$. 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 $A_t = \phi(t) \neq 1$ 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.19. *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 $\alpha_1, \alpha_2 \ldots, \alpha_m \in \mathbb{E}$ with*

$$x^m - 1 = (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_m).$$

Then

(a) $x^n - 1 = (x^m - 1)^{2^k}$.

(b) $\alpha_1, \alpha_2, \alpha_3, \ldots \alpha_m$ *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

$$mx^{m-1} = 2(x - \alpha)g + (x - \alpha)^2 g'.$$

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.20. *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

$$\beta^{is}, \quad 0 \leq i < d.$$

Therefore $\mathbb{E}$ contains exactly d roots of $x^n - 1$.

From A.19, $\mathbb{E}$ 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 $\alpha^i, 0 \leq i < n$. □

Appendix B

Constructing Sources

B.1 Marginal Distributions on Triples

Lemma B.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.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 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.3. *An example of a stationary source which is not mememory less.*

For $z = z_1 \dots z_n \in B^*$ define $u(z) = |\{i \mid 1 \leq i < n, z_i = z_{i+1}\}|$. Define $P(\emptyset) = 1$ and if $n \geq 1$,

$$P(z) = \frac{1}{2}\frac{3^{u(z)}}{4^{n-1}}$$

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 $z_n \neq s$. Hence for $z \in \mathbb{B}^n$ with $n \geq 1$ and $s \in \mathbb{B}$:

$$P(zs) = \begin{cases} \frac{3}{4}P(z) \text{ if } s = z_n \\ \frac{1}{4}P(z) \text{ if } s \neq z_n \end{cases}$$

Thus $P(z) = P(z0) + P(z1)$ and P is source. Similarly

$$P(sz) = \begin{cases} \frac{3}{4}P(z) \text{ if } s = z_1 \\ \frac{1}{4}P(z) \text{ if } s \neq z_1 \end{cases}$$

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.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 (*)

$$f_a + f_b = k = g_c + g_d \quad \text{and so } g_d - f_a = f_b - g_c$$

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

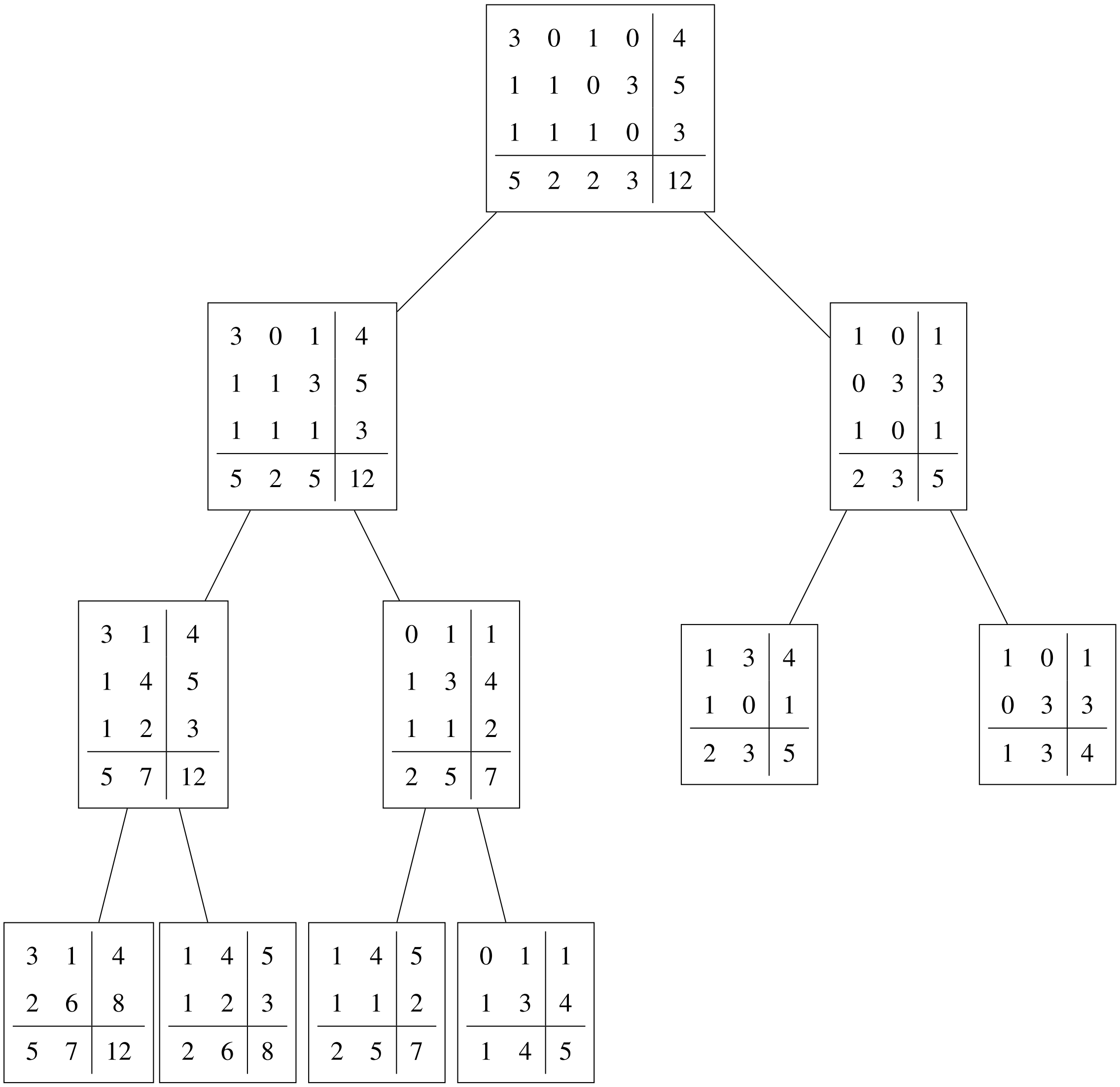
$$\sum_{j \in \tilde{J}} \tilde{g}_j = \tilde{g}_u + \sum_{\substack{j \in \tilde{J} \\ j \neq u}} \tilde{g}_j = g_u + g_v + \sum_{\substack{j \in J \\ j \neq u,v}} g_j = \sum_{j \in J} g_j = t = \sum_{i \in I} f_i$$

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

$$h_{ij} = \begin{cases} \tilde{h}_{ij} & \text{if } j \in J \setminus \tilde{J} \\ \hat{h}_{ij} & \text{if } j \in \hat{J} \end{cases}$$

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. *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 $i \in I \setminus K$. 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 $i \in I \setminus K$,

$$\sum_{i \in I} \hat{p}_i = \sum_{i \in K} \hat{p}_i = \sum_{i \in K} p_i = 1.$$

(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 V.24 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. *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**