Introduction to algebraic codings
Lecture Notes for MTH 416
Fall 2015

Ulrich Meierfrankenfeld

December 9, 2015
Preface

These are the Lecture Notes for the class MTH 416 in Fall 2015 at Michigan State University. The Lecture Notes will be frequently updated throughout the semester.
## Contents

I Coding ........................................... 7  
  I.1 Matrices ........................................ 7  
  I.2 Basic Definitions .............................. 11  
  I.3 Coding for economy ........................... 13  
  I.4 Coding for reliability ........................ 13  
  I.5 Coding for security ........................... 13  

II Prefix-free codes .............................. 15  
  II.1 The decoding problem ...................... 15  
  II.2 Representing codes by trees ............... 18  
  II.3 The Kraft-McMillan number ............... 23  
  II.4 A counting principal ...................... 27  
  II.5 Unique decodability implies $K \leq 1$ .......... 29  

III Economical coding .......................... 33  
  III.1 Probability distributions ................. 33  
  III.2 The optimization problem .............. 34  
  III.3 Entropy ....................................... 37  
  III.4 Optimal codes – the fundamental theorems 39  
  III.5 Huffman rules ............................. 42  

IV Data Compression ......................... 47  
  IV.1 The Comparison Theorem .................. 47  
  IV.2 Coding in pairs ............................. 48  
  IV.3 Coding in blocks ............................ 51  
  IV.4 Memoryless Sources ...................... 55  
  IV.5 Coding a stationary source ............. 56  
  IV.6 Arithmetic codes ........................... 60  
  IV.7 Coding with a dynamic dictionary ....... 64  

V Error Correcting ............................. 69  
  V.1 Decision rules .............................. 69  
  V.2 The Packing Bound ........................... 74  

VI Linear Codes .................................. 79
VI.1 Introduction to linear codes .................................................. 79
VI.2 Construction of linear codes using matrices ................................. 83
VI.3 Standard form of check matrix ................................................. 84
VI.4 Constructing 1-error-correcting linear codes ................................. 92
VI.5 Decoding linear codes .......................................................... 98
VII Algebraic Coding Theory ......................................................... 105
VII.1 Classification and properties of cyclic codes ............................... 105
VII.2 Definition of a family of BCH codes ..................................... 124
VII.3 Properties of BCH codes ...................................................... 132
VIII The RSA cryptosystem .......................................................... 139
VIII.1 Public-key cryptosystems ...................................................... 139
VIII.2 The Euclidean Algorithm ..................................................... 140
VIII.3 Definition of the RSA public-key cryptosystem .......................... 144
IX Noisy Channels ........................................................................ 147
IX.1 The definition of a channel ...................................................... 147
IX.2 Transmitting a source through a channel .................................... 149
IX.3 Conditional Entropy ............................................................... 152
IX.4 Capacity of a channel ............................................................. 153
X The noisy coding theorems ........................................................ 157
X.1 The probability of a mistake ...................................................... 157
X.2 Fano’s inequality ..................................................................... 159
X.3 A lower bound for the probability of a mistake ............................. 161
X.4 Extended Channels .................................................................. 162
X.5 Coding at a given rate ............................................................. 167
X.6 Minimum Distance Decision Rule .......................................... 168
XI Cryptography in theory and practice ........................................ 171
XI.1 Encryption in terms of a channel .............................................. 171
XI.2 Perfect secrecy ................................................................. 174
XI.3 The one-time pad ................................................................. 175
XI.4 Iterative methods ................................................................. 175
XI.5 The Double-Locking Procedure ............................................. 177
A Rings and Field ........................................................................ 181
A.1 Basic Properties of Rings and Fields ....................................... 181
A.2 Polynomials .......................................................................... 183
A.3 Irreducible Polynomials ......................................................... 185
A.4 Primitive elements in finite field ............................................. 187
B Constructing Sources ................................................................ 191
B.1 Marginal Distributions on Triples .......................................... 191
B.2 A stationary source which is not memory less ............................ 192
B.3 Matrices with given Margin ..................................................... 193
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C More On channels</td>
<td>195</td>
</tr>
<tr>
<td>C.1 Sub channels</td>
<td>195</td>
</tr>
<tr>
<td>D Examples of codes</td>
<td>197</td>
</tr>
<tr>
<td>D.1 A 1-error correcting binary code of length 8 and size 20</td>
<td>197</td>
</tr>
</tbody>
</table>
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) $\mathbb{N} := \{0, 1, 2, 3, \ldots\}$ denotes the set of non-negative integers.

(c) Let $n \in \mathbb{N}$. Then an $n$-tuple is an $\{1, 2, \ldots, n\}$-tuple.

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

Notation I.2. Notation for tuples.

(1)

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

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

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

We denote this tuple also by

7
CHAPTER I. CODING

\[ a \quad 0 \\
\begin{array}{c}
 b \\
 c \\
 d \\
\end{array} \quad \pi
\]
\[ x:\]
\[ \begin{array}{c}
 a \\
 b \\
 c \\
 d \\
\end{array} \quad \frac{1}{3} \]

(2)
\[ y = (a, a, b, c) \]
denotes the 4-tuple with coefficients in \( \{a, b, c, d, e, \ldots, z\} \) such that
\[ y_1 = a, \quad y_2 = a, \quad y_3 = b, \quad \text{and} \quad y_4 = c. \]
We will denote such a 4-tuple also by
\[ y = \begin{pmatrix}
 a \\
 a \\
 b \\
 c \\
\end{pmatrix}. \]

Definition 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 \( [m_{ij}]_{i \in I, j \in J} \). \( m_{ij} \) is called the \( ij \)-coefficients of \( M \).

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

Notation I.4. Notations for matrices

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

\[
\begin{array}{c|ccc}
 M & x & y & z \\
 \hline
 a & 0 & 1 & 2 \\
 b & 1 & 2 & 3 \\
 c & 2 & 3 & 4 \\
 d & 3 & 4 & 5 \\
\end{array}
\]
stands for the \{a, b, c, d\} \times \{x, y, z\} matrix \(M\) with coefficients in \(\mathbb{Z}\) such that \(m_{ax} = 0, m_{ay} = 1, m_{bx} = 1, m_{cz} = 4, \ldots\)

(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 \(2 \times 3\) matrix \(M\) with \(m_{11} = 0, m_{12} = 2, m_{21} = 4, m_{23} = 6, \ldots\)

**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 \(\text{Row}_i(M)\) 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 \(\text{Col}_j(M)\)

**Example I.6.** Let

\[
M = \begin{bmatrix}
a & \pi & 3 \\
0 & \alpha & x
\end{bmatrix}
\]

Then

\[M_2 = \text{Row}_2(M) = (0, \alpha, x)\]

and

\[\text{Col}_3(M) = \begin{bmatrix}3 \\ x\end{bmatrix}.\]

**Definition I.7.** Let \(A\) be \(I \times J\)-matrix, \(B\) an \(J \times K\) matrix and \(x, y\) \(J\)-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.8. Examples of matrix multiplication.

(1) Given the matrices
\[
\begin{array}{ccc}
A & x & y \\
& a & 0 & 1 & 2 \\
b & 1 & 2 & 3 \\
\end{array}
\begin{array}{cccc}
B & \alpha & \beta & \gamma & \delta \\
x & 0 & 0 & 1 & 0 \\
y & 1 & 0 & 0 & 1 \\
x & 1 & 1 & 0 & 0 \\
\end{array}
\]

Then $AB$ is the $\{a, b\} \times \{\alpha, \beta, \gamma, \delta\}$ matrix
\[
\begin{array}{cccc}
AB & \alpha & \beta & \gamma & \delta \\
a & 3 & 2 & 0 & 1 \\
b & 4 & c & 1 & 2 \\
\end{array}
\]

(2) Given the matrix $2 \times 3$-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) Given the matrix
\[
\begin{array}{ccc}
A & x & y \\
& a & 0 & 1 & 2 \\
b & 1 & 2 & 3 \\
\end{array}
\begin{array}{cccc}
& a & b \\
2 & -1 \\
\end{array}
\]

and the tuple $u = \begin{pmatrix} \frac{a}{2} \\ \frac{b}{-1} \end{pmatrix}$
Then $uA$ is the $\{x, y, z\}$-tuple

\[
\begin{array}{c}
  x & y & z \\
  -1 & 0 & 1 \\
\end{array}
\]

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

\[xy = 3\]

### I.2 Basic Definitions

**Definition I.9.** An alphabet is a finite set. The elements of an alphabet are called symbols.

**Example I.10.** (a) $A = \{A, B, C, D, \ldots, X, Y, Z, \}$ is the alphabet consisting of the regular 27 uppercase letters and a space (denoted by $\$\$).

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

**Definition I.11.** 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 $s_1s_2\ldots s_n$. A message of length $n$ is sometimes also called a string of length $n$.

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

(c) $\emptyset$ denote the unique message of length 0 in $S$.

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

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

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

**Example I.12.** (1) WATCH\$OUT\$FOR\$MUSKRATS is a message of length 22 in the alphabet $A$.

(2) 100111011011001 is a message of length 16 in the alphabet $B$.

(3) $B^0 = \{\emptyset\}$, $B^1 = B = \{0, 1\}$, $B^2 = \{00, 01, 10, 11\}$, $B^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$, and $B^* = \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \ldots, 111, 0000, \ldots, 1111, \ldots\}$.
Definition I.13. 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 $\emptyset$ is not a codeword.

Example I.14. (1) The function $c: A \to A^*$ such that

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

is a code for $A$ using $A$. The set of codewords is

$$C = A.$$  

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

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

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

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

Definition I.15. Let $c: S \to T^*$ be a code. Then the function $c^*: S^* \to T^*$ defined by

$$c^*(s_1s_2\ldots s_n) = c(s_1)c(s_2)\ldots c(s_n)$$

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

Example I.16. (1) Let $c: A \to A^*$ be the code from (14.1). Then $c(MTH) = PWK$.

(2) Let $c: \{x, y, z\} \to B^*$ be the code from (14.2). Then $c(xzzx) = 010100$ and $c(yyxx) = 010100$.

So $xzzx$ and $yyxx$ are encoded to the same message in $B$.

Definition I.17. A code $c: S \to T^*$ is called uniquely decodable (UD) if the extended function $c: S^* \to T^*$ is 1-1.

Example I.18. (a) The code from (16.1) is UD.

(b) The code from (16.2) is not UD since $c^*(xzzx) = c^*(yyxx)$. 


I.3 Coding for economy

Example I.19. The Morse alphabet \( \mathbb{M} \) has three symbols \( \bullet, - \) and \( \circ \), called dot, dash and pause. The Morse code is the code for \( A \) using \( \mathbb{M} \) defined by

\[
\begin{array}{cccccccccccc}
A & B & C & D & E & F & G & H & I \\
\bullet & - & \circ & - \bullet \circ & - \bullet & \circ & \bullet & \circ & \bullet & \circ & \bullet & \circ \\
J & K & L & M & N & 0 & P & Q & R \\
\bullet & - & - & - & \circ & - & \bullet & \circ & \bullet & \circ & \bullet & \circ \\
S & T & U & V & W & X & Y & Z & \square \\
\bullet & \bullet & \bullet & \circ & - \circ & \bullet & - \circ & \bullet & - \circ & \bullet & - \circ & \bullet & - \circ \\
\end{array}
\]

So for example the messages \( SOS \) in \( A \) encodes to \( \bullet \bullet \circ - - - \circ \bullet \bullet \circ \).

I.4 Coding for reliability

Example I.20. Codes for \{Buy, Sell\} using \{B, S\}.

1. Sell \( \rightarrow \) S, Sell \( \rightarrow \) S.
2. Buy \( \rightarrow \) BB, Sell \( \rightarrow \) SS.
3. Buy \( \rightarrow \) BBB, Sell \( \rightarrow \) SSS.
4. Buy \( \rightarrow \) BBBBBBBBB, Sell \( \rightarrow \) SSSSSSSSS.

I.5 Coding for security

Example I.21. Let \( k \) be an integer with \( 0 \leq k \leq 25 \). Then \( c_k \) is the code from \( A \rightarrow A \) obtained by shifting each letter by \( k \) places to the right. \( \square \) is unchanged. For example the code in Example I.14(1) is \( c_3 \).

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

II.1 The decoding problem

Definition II.1. Let $T$ be an alphabet.

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

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

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

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

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

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

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

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

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

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

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

(b): Let $a = s_1 \ldots s_n$ and $r = u_1 \ldots u_k$. Then $ar = s_1 \ldots s_n u_1 \ldots u_k$. So $\ell(ar) = n + k = \ell(a) + \ell(r)$.
[c]: Let $a$ be a prefix of $b$. Then $b = ar$ for some $r \in T^*$. By [b] we have $\ell(b) = \ell(ar) = \ell(a) + \ell(r) \geq \ell(a)$.

[d] and [e]: Let $a = s_1 \ldots s_n$ be a prefix of $b$ of length $n$. Then $b = ar$ for some $r \in T^*$. Let $r = u_1 \ldots u_k$. Then $b = s_1 \ldots s_n u_1 \ldots u_k$. It follows that $t_i = s_i$ for $1 \leq i \leq n$ and so $a = t_1 \ldots t_n$ and (d) holds. If $a$ is a parent, then $r \in T$, that is $k = 1$ and $n = m - 1$. So (e) follows from (c).

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

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 $c_1, c_2, \ldots, c_n$ of codewords from their concatenation $e = c_1 \ldots c_n$. Consider for example the string $e = 2111002110001$

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

<table>
<thead>
<tr>
<th>Prefix</th>
<th>codeword?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
</tr>
<tr>
<td>21</td>
<td>no</td>
</tr>
<tr>
<td>211</td>
<td>no</td>
</tr>
<tr>
<td>2111</td>
<td>no</td>
</tr>
<tr>
<td>21110</td>
<td>yes</td>
</tr>
</tbody>
</table>

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
II.1. THE DECODING PROBLEM

02110001

The prefixes are

<table>
<thead>
<tr>
<th>Prefix</th>
<th>codeword?</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>no</td>
</tr>
<tr>
<td>0</td>
<td>no</td>
</tr>
<tr>
<td>02</td>
<td>no</td>
</tr>
<tr>
<td>021</td>
<td>yes</td>
</tr>
</tbody>
</table>

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

$$211102110001 = (21110)(021)(1001)$$

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

**Theorem II.4.** Any regular PF code is UD.

**Proof.** Let $C$ be a regular PF-code. 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 induction on $\min(n, m)$. Put $e = c_1 \ldots c_n$ and so also $e = d_1 \ldots d_m$.

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

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

Since $e = c_1 \ldots c_n$, $c_1$ is a prefix of $e$ and since $e = d_1 \ldots d_m$, $d_1$ is a prefix of $e$. As $\ell(c_1) \leq \ell(d_1)$ this implies that $c_1$ is a prefix of $d_1$, see II.2[4]. As $C$ 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 gives
By induction we conclude that \( n - 1 = m - 1 \) and \( c_2 = d_2, \ldots, c_n = d_n \). Hence \( n = m \) and 
\[ c_1 = d_1, \ldots, c_n = d_n. \]

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

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

**Proof.** Let \( c : S \to T^* \) be non-regular code. Then there exists \( s \in S \) with \( c(s) = \emptyset \) for \( s \in S \). Thus 
\[ c^*(ss) = \emptyset \emptyset = \emptyset = c^*(s). \]
and so \( c \) is not UD. We proved that a non-regular code is not UD and so every UD-code is regular. Thus (a) holds.

Suppose in addition that \( c \) is prefix-free and let \( t \in S \). Since \( c(s) = \emptyset \) we know that \( c(s) \) is a prefix of \( c(t) \), see II.2(a). Since \( c \) is prefix-free this implies that \( c(s) = c(t) \). By definition of a code, \( c \) is 1-1 and so \( s = t \). This shows that \( S = \{s\} \) and hence the forward direction in (b) holds.

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

### 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 element \( e \) 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 the vertex \( b \) in \( G \) if \( \{a, b\} \) is an edge.

**Example II.7.** Let \( V = \{1, 2, 3, 4\} \) and \( E = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{3, 4\}\} \). 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:

```
    1   2
   / \ /  \
  4   3
```
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 $a, b \in V$, there exists a path from $a$ to $b$ in $G$.

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

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

(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)

```
1 -- 2
 |   |
|   |
4 -- 3
```

is connected, but

```
1 -- 2
 |   |
|   |
3
```

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

(2)

```
1 | 2
 | |
| |
4 | 3
```
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)

![Graph](image)

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 a pair \((V, E)\)? The vertices are all the binary messages and a message \(a\) is adjacent to message \(b\) if \(a\) is the parent of \(b\) or \(b\) is the parent of \(a\).
II.2. REPRESENTING CODES BY TREES

\[ V = \mathbb{B}^* \text{ and } E = \{ \{a, as\} \mid a \in \mathbb{B}^*, s \in \mathbb{B}\} = \{ \{a, b\} \mid a, b \in \mathbb{B}^*, a \text{ is the parent of } b\} \]

So the infinite binary tree now looks like:

\[
\begin{array}{c}
\emptyset \\
0 & 1 \\
00 & 01 & 10 & 11 \\
000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\
\end{array}
\]

**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

\[ E = \{ \{a, b\} \mid a, b \in V, a \text{ is a parent of } b\} \]

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

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

\[
\begin{array}{c}
\emptyset \\
0 & 1 \\
10 & 11 \\
110 & 111 \\
\end{array}
\]

**Definition II.13.** Let \( G \) be a graph. A leaf 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 $c \in V$ and put $n := \ell(c)$. For $0 \leq k \leq n$ let $c_k$ be the prefix of length $k$ of $c$. Then $(c_0, c_1, \ldots, c_n)$ is a path from $\varnothing$ to $c$ in $G(C)$.

(b) $G(C)$ is tree

Suppose in addition that $C$ is regular.

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

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

Proof. (a) Let $c = s_1 \ldots s_n$. Then $c_i = s_1 \ldots s_{i-1}s_i$ and so $c_i = c_{i-1}s_i$. Thus $c_{i-1}$ is a parent if $c_i$. Hence $c_{i-1}$ is adjacent to $c_i$ and $(c_0, \ldots, c_n)$ is a path in $G(C)$. Also $c_0 = \varnothing$ and $c_n = c$.

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

We prove next:

(*) Let $m \in \mathbb{Z}^+$ and let $(a_0, a_1, \ldots, a_m)$ be path in $G(C)$ with pairwise distinct vertices. Suppose that $\ell(a_0) \leq \ell(a_1)$. Then for all $1 \leq i \leq m$, $a_{i-1}$ is the parent of $a_i$. In particular, $\ell(a_i) = \ell(a_0) + i$ for all $0 \leq i \leq m$ and $a_i$ is a prefix of $a_j$ for all $0 \leq i \leq j \leq m$.

Since $a_0$ is adjacent to $a_1$, one of $a_0$ and $a_1$ is the parent of the other. Since $\ell(a_0) \leq \ell(a_1)$, $a_1$ is not a parent of $a_0$ and we conclude that $a_0$ is parent of $a_1$. So if $m = 1$, (*) holds. Suppose $m \geq 2$. Since $a_2 \neq a_0$ and $a_0$ is the unique parent of $a_1$ we know that $a_2$ is not the parent of $a_1$. Since $a_2$ is adjacent to $a_1$ we conclude that $a_1$ is the parent of $a_2$. In particular, $\ell(a_1) \leq \ell(a_2)$.

Induction, applied to the path $(a_1, \ldots, a_m)$, shows that $a_{i-1}$ is a parent of $a_i$ for all $2 \leq i \leq m$ and so the first statement in (*) is proved.

Note that remaining statement in (*) follow from the first and so (*) is proved.

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

Assume that $0 < k < n$. Then $\ell(v_k) = l \leq \ell(v_{k-1})$ and $\ell(v_k) = l \leq \ell(v_{k+1})$. Hence can apply (*) to the paths $(v_k, v_{k+1}, \ldots, v_n)$ and $(v_k, v_{k-1}, \ldots, v_0)$. It follows that $v_{k+1}$ 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 $(v_0, \ldots, v_n)$ is a circle we have $v_0 = v_n$. So both $v_{k-1}$ and $v_{k+1}$ are the parent of $v_0$ and we conclude that $v_{k-1} = v_{k+1}$. As $(v_0, v_1, \ldots, v_n)$ is a simple circle, this implies $k-1 = 0$ and $k+1 = n$. So $n = 2$, a contradiction to $n \geq 3$.

Assume next that $k = 0$ or $k = n$. Since $v_0 = v_n$ we conclude that $v_0 = v_k = v_n$. In particular, $v_0$ and $v_n$ have length $l$. In particular, $\ell(v_0) \leq \ell(v_1)$ and $\ell(v_n) \leq \ell(v_{n-1})$. By (*) applied to $(v_0, v_1)$, $v_1$ has length $l+1$ and by (*) applied to $(v_n, v_{n-1}, \ldots, v_1)$, $v_1$ has length $l + (n-1)$. Thus $l+1 = l + (n-1)$ and again $n = 2$, a contradiction.
So $G(C)$ has no simple circle of length at least three. We already proved that $G(C)$ is connected and so $G(C)$ is a tree.

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

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

<table>
<thead>
<tr>
<th>(d)</th>
<th>$C$ is prefix-free</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\iff$</td>
<td>no codeword is a prefix of another codeword</td>
</tr>
<tr>
<td>$\iff$</td>
<td>no codewords is adjacent two different vertices of $G(C)$</td>
</tr>
<tr>
<td>$\iff$</td>
<td>every codewords is adjacent to at most one vertex</td>
</tr>
<tr>
<td>$\iff$</td>
<td>each codeword is a leaf</td>
</tr>
</tbody>
</table>

## II.3 The Kraft-McMillan number

### Definition II.15. Let $C$ be a code using the alphabet $T$ and put $b := |T|$.

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

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

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

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

(e) The number
CHAPTER II. PREFIX-FREE CODES

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

is called the Kraft-McMillan number of the parameter \((n_0, n_1, \ldots, n_M)\) to 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 \(b = 2\), \(C_0 = \{\}\), \(C_1 = \{\}\), \(C_2 = \{10, 01, 11\}\) and \(C_3 = \{011\}\). So the parameter is \((0, 0, 3, 1)\) and

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

**Example II.17.** Construct a ternary code with parameter \((0, 1, 4, 4)\).

We first look at all the messages of length 1:

\[
\begin{array}{c}
\emptyset \\
0 \\
1 \\
2 
\end{array}
\]

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

\[ C_1 = \{0\}. \]

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

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

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

```
00 01 02 10 11 12 20 21 22
000 001 002 010 011 012 020 021 022
100 101 102 110 111 112 120 121 122
200 201 202 210 211 212 220 221 222
```

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

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

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

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

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

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

So the number of messages of length 3 which do not have a codeword as a prefix is

$$3^3 - 3^3 \cdot K = 3^3(1 - K).$$
Lemma II.18. Let \( C \) be a \( b \)-nary code with Kraft-McMillan number \( K \) using the alphabet \( T \). Let \( M \in \mathbb{N} \) such that every code word has length at most \( M \). Let \( D \) be the set of messages of length \( M \) in \( T \) which have a codeword as a prefix. Then

(a) \(|D| \leq Kb^M\).

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

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

Proof. (a)

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

\[
|D| \leq \sum_{i=0}^{M} n_i b^{M-i} = b^M \sum_{i=1}^{M} n_i b^{-i} = b^M K
\]

(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 \(|D| \leq b^M \) and so by (b), \( b^M K \leq b^M \). Hence \( K \leq 1 \).

Theorem II.19. Let \( b \in \mathbb{Z}^+ \) and \( M \in \mathbb{N} \). Let \( n = (n_0, n_1, \ldots, n_M) \) be a tuple of non-negative integers such that \( K \leq 1 \), where \( K \) is the Kraft-McMillan number of the parameter \( n \) to the base \( b \). Then there exists a \( b \)-nary PF code \( C \) with parameter \( n \).

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

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

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

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

and so

\[
(*) \quad \tilde{K} = K - \frac{n_M}{b^M} \leq 1 - \frac{n_M}{b^M} \leq 1.
\]
By the induction assumption there exists a PF code $\tilde{C}$ with parameter $\tilde{n}$. Note that the codewords in $\tilde{C}$ have length at most $M - 1$. Let $\tilde{D}$ be the set of messages of length $M$ in $T$ which have a codeword from $\tilde{C}$ as a prefix. By II.18 $|\tilde{D}| = b^M \tilde{K}$. Multiplying (*) with $b^M$ gives $b^M \tilde{K} \leq b^M - n_m$. Thus $|\tilde{D}| \leq b^M - n_m$ and so $n_M \leq b^M - |\tilde{D}|$. Since $b^M$ is the number of messages of length $M$, $b^M - |\tilde{D}|$ is the number of messages of length $M$ which do not have a codeword from $\tilde{C}$ as a prefix. Thus there exists a set $E$ of messages of length $M$ such that $|E| = n_M$ and none of the messages has a codeword from $\tilde{C}$ as a prefix. Put $C := \tilde{C} \cup E$.

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

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

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

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

Thus $C$ is indeed PF.

II.4 A counting principal

**Definition II.20.** Let $c : S \to T^*$ and $d : R \to T^*$ be codes using the same alphabet $T$.

(a) $cd$ is the function from $S \times R$ to $T^*$ defined by $(cd)(s, r) = c(s)d(r)$ for all $s \in S, r \in R$.

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

(c) Let $r \in \mathbb{Z}^+$. Then $c^r$ is the function from $S^r \to T^*$ inductively defined by $c^1 := c$ and $c^{r+1} := c^r c$.

**Lemma II.21.** Let $c : S \to T^*$ and $d : R \to T^*$ be codes. Let $C$ and $D$ be the set of codewords of $c$ and $d$ respectively.
(a) $\text{Im} cd = CD$.

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

(c) If $\text{cd}$ is a code, then $\text{CD}$ is set of codewords of $\text{cd}$.

(d) If $c$ or $d$ is regular, then $\text{cd}$ is regular.

Proof. (a): We have

$$\text{Im} cd = \{ (cd)(s,r) \mid (s,r) \in S \times R \} = \{ c(s)d(r) \mid s \in S, r \in R \} = \{ ab \mid a \in C, d \in D \} = CD$$

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

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

(d) Suppose $c$ is regular and let $s \in S$ and $r \in R$. Since $c$ is regular, $c(s)$ is not the empty message and so also $c(s)d(r)$ is not the empty message. Hence $(cd)(r,s) \neq \emptyset$ for all $(s,r) \in S \times R$ and $\text{cd}$ is regular.

A similar argument shows that $\text{cd}$ is regular of $d$ is regular. \qed

Definition II.22. Let $c$ be a code with set of codewords $C$ and parameter $(n_0, n_1, \ldots, n_M)$. Then

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

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

Example II.23. Compute the generating function of the code $C = \{01, 10, 110, 1110, 1101\}$.

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

So

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

Theorem II.24 (The Counting Principal).

(a) Let $c$ and $d$ be codes using the same alphabet $T$ such that $\text{cd}$ is a code. Then

$$Q_{\text{cd}}(x) = Q_c(x)Q_d(x)$$
II.5. UNIQUE DECODABILITY IMPLIES $K \leq 1$

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

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

Proof. Let $(n_0, \ldots, n_M), (p_0, \ldots, p_V), (q_0, \ldots, q_V)$ 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

$$Q_{c^r}(x) = Q_{c^{r-1}}(x) = Q_c(x)Q_c(x) \cdots Q_c(x) = Q_{c^r}(x).$$

\[ \square \]

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

Lemma II.25. Let $c$ be a $b$-nary 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, $\emptyset$ is not a codeword and thus $n_0 = 0$. So $K \leq M$ in this case.

(b):

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

\[ \square \]

Lemma II.26. (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. \( \square \)

**Example II.27.** Let \( C \) be UD-code with parameter \((0,1,0,2)\). Compute the generating function and the parameter of \( C^3 \).

The generating function of \( C \) is

\[
Q_C(x) = 0 + 1x + 0x^2 + 2x^3 = x + 2x^3.
\]

Hence

\[
Q_{C^3}(x) = Q_C^3(x) = (x + 2x^3)^3 = x^3 + 3 \cdot x^2 \cdot 2x^3 + 3 \cdot x \cdot (2x^3)^2 + (2x^3)^3 = x^3 + 6x^5 + 12x^7 + 8x^9
\]

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

\((0,0,1,0,6,0,12,0,8)\)

**Theorem II.28.** Let \( C \) be a UD-code with Kraft-McMillan number \( K \). Then \( K \leq 1 \).

**Proof.** Let \( M \) be the maximal length of a codeword of \( C \). Then the maximal length of a codeword of \( C^r \) is \( rM \). By [II.26](b) the Kraft-McMillan number of \( C^r \) is \( K^r \), and by [II.25] the Kraft-McMillan number is bounded by the maximal codeword length plus 1. Thus

\[
K^r \leq rM + 1
\]

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

\[
K \ln K = \ln(K^r) \leq \ln(rM + 1) \quad \text{and so} \quad \ln K \leq \frac{\ln(rM + 1)}{r}
\]

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

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

and so \( \ln K \leq 0 \) and \( K \leq 1 \). \( \square \)

**Corollary II.29.** Given a parameter \( n = (n_0, \ldots, n_M) \) and a base \( b \) with Kraft-McMillan number \( K \). Then the following three statements are equivalent.

(a) Either \( n = (1,0,\ldots,0) \) or there exists a \( b \)-nary UD-code with parameter \( n \).
II.5. UNIQUE DECODABILITY IMPLIES $K \leq 1$

(b) $K \leq 1$.

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

Proof. (a) $\implies$ (b): If $n = (1,0,\ldots,0)$ then $K = 1$. If $C$ is UD- code with parameter $(n_0,\ldots,n_M)$ then II.28 shows that $K \leq 1$.

(b) $\implies$ (c): If $K \leq 1$, then II.19 shows that there exists a $b$-nary PF-code with parameter $n$.

(c) $\implies$ (a): Suppose $c$ is a $b$-nary PF-code with parameter $n$. If $c$ is regular, II.4 shows that $c$ is UD and (a) holds. If $c$ is not regular, then II.5(b) shows that $\emptyset$ is the only codeword. Hence the the parameter of $c$ is $(1,0,\ldots,0)$ and again (a) holds. $\square$
Chapter III
Economical coding

III.1 Probability distributions

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

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

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

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

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

is an $S$-tuple.

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

Example III.3.

(a) 

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

is a probability distribution on $\{w, x, y, z\}$ with $p_x = \frac{1}{3}$.

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

33
(c) \( p = \left( \frac{1}{2}, \frac{1}{3}, 0, \frac{1}{6} \right) \) is a probability distribution on \( \{x, w, z, y\} \) with \( p_x = \frac{1}{2} \).

(d) \((01, 111, 011, 111)\) is the code on \( \{a, b, c, d\} \) with \( a \rightarrow 01, b \rightarrow 111, c \rightarrow 011, d \rightarrow 111 \).

(e) \[
\begin{array}{cccccccccc}
A & B & C & D & E & F & G & H & I \\
\end{array}
\]
\[
\begin{array}{cccccccccc}
J & K & L & M & N & O & P & Q & R \\
0.153\% & 0.747\% & 4.025\% & 2.406\% & 6.749\% & 7.507\% & 1.929\% & 0.959\% & 5.987\%
\end{array}
\]
\[
\begin{array}{cccccccccc}
S & T & U & V & W & X & Y & Z & \Box \\
6.327\% & 9.056\% & 2.758\% & 1.037\% & 2.365\% & 0.150\% & 1.974\% & 0.074\% & 0
\end{array}
\]
is a probability distribution on \( A \). It lists how frequently a letter is used in the English language.

(f) Let \( S \) be a set with \( m \) elements and put

\[
p := \left( \frac{1}{m} \right)_{s \in S} = \left( \frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m} \right).
\]

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 \( c : S \rightarrow T^* \) 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 codeword length of \( c \) with respect to \( p \) is the number

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

To emphasize that \( L \) depends on \( p \) and \( c \) we will sometimes use the notations \( L(c) \) and \( L_p(c) \) for \( L \).
III.2. THE OPTIMIZATION PROBLEM

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

**Example 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 $s_1 \rightarrow 0$, $s_2 \rightarrow 10$, $s_3 \rightarrow 11$.

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

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

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

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

$s_1 \rightarrow 01$, $s_2 \rightarrow 0$, $s_3 \rightarrow 11$.

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

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

**Definition 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

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

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

**Definition III.7.** Let $S$ be an alphabet, $y$ an $S$-tuple with coefficients in $\mathbb{N}$ and $b$ a positive integer.

(a) Let $M \in \mathbb{N}$ with $y_s \leq M$ for all $s \in S$. For $0 \leq i \leq M$ let $n_i$ be the number of $s \in S$ with $y_s = i$. Then $(n_0, \ldots, n_M)$ is called the parameter of the codewords length $y$.

(b) $\sum_{s \in S} \frac{1}{y_s}$ is called the Kraft-McMillan numbers of 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 $(n_0, \ldots, n_M)$ the parameter for the codewords length $y$.

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

(b) Let $c$ be a $b$-nary code for the set $S$ with codeword length $y$. Then $(n_0, \ldots, n_M)$ is the parameter of $c$ and $K$ is the Kraft-McMillan number of $c$. 
(c) Suppose there exists a \( b \)-nary code \( C \) with parameter \((n_0, \ldots, n_M)\). Then there exists a \( b \)-nary code \( c \) for \( S \) with codewords length \( y \) such that \( C \) is the set of codewords of \( c \).

Proof. \( \square \) We compute

\[
K = \sum_{s \in S} \frac{1}{b^{y_s}} = \sum_{i=0}^{M} \sum_{s : y_s = i} \frac{1}{b^i} = \sum_{i=0}^{M} \sum_{s : y_s = i} \frac{n_i}{b^i} = \sum_{i=0}^{M} n_i \frac{1}{b^i}
\]

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

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

\( \square \) Let \( C \) be a \( b \)-nary code with parameter \((n_0, \ldots, n_m)\). Recall from Definition \( \text{II.15}(c) \) that

\[
C_i = \{ d \in C \mid d \text{ has length } i \}.
\]

Also put

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

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

By definition of \( n_i \) we have \(|S_i| = n_i|\). Thus \( |C_i| = |S_i| \) and there exists a 1-1 and onto function \( \alpha_i : S_i \rightarrow C_i \). 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 \( \alpha_i(s) \in C_i \), \( \alpha_i(s) \) has length \( i = y_s \). So \( y_s \) is exactly the length of \( c(s) \) and thus \( y \) is the codeword length of \( c \). Note that \( c : S \rightarrow C \) is a bijections and so \( c \) is a code with set of codewords \( C \). \( \square \)

Lemma III.9. 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 of the codeword length \( y \). Then there exists a \( b \)-nary PF code with codeword length \( y \) if and only if \( K \leq 1 \).

Proof. Suppose first that there exists a PF-code \( c \) with codewords length \( y \). Then by \( \text{III.8}(b) \) \( K \) is the Kraft-McMillan number of \( c \). Since \( c \) is PF we conclude from \( \text{II.18}(c) \) that \( K \leq 1 \).

Suppose next that \( K \leq 1 \) and let \( n \) be the parameter of the codewords length \( y \). By \( \text{III.8}(a) \), \( K \) is the Kraft-McMillan number of the parameter \( n \) to the base \( b \). Since \( K \leq 1 \) we conclude from \( \text{II.19} \) that there exists a \( b \)-nary PF-code \( C \) with parameter \( n \). Hence by \( \text{III.8}(c) \) there exists a code \( c \) with codewords length \( y \) and set of codewords \( C \). As \( C \) is PF, so is \( c \). \( \square \)
Remark III.10. In view of the preceding lemma, the problem of finding the codeword length of an optimal $b$-nary PF-code with respect to a given probability distribution $(p_1, \ldots, p_m)$ can be restated as follows:

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

$$p_1y_1 + p_2y_2 + \ldots + p_my_m$$

is minimal subject to

$$\frac{1}{by_1} + \frac{1}{by_2} + \ldots + \frac{1}{by_m} \leq 1.$$

III.3 Entropy

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

Then

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

with equality if and only if

$$\sum_{s \in S} \frac{1}{bys} = 1 \quad \text{and} \quad y_s = \log_b \left( \frac{1}{p_s} \right) \text{ for all } s \in S.$$

Proof. We will first show that

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

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

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

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

$$\ln \left( \frac{1}{p_s bys} \right) \leq \frac{1}{p_s bys} - 1$$

with equality if and only if $\frac{1}{p_s bys} = 1$. Observe that $\frac{1}{p_s bys} = 1$ if and only if $bys = \frac{1}{p_s}$ and so if and only if $y_s = \log_b \left( \frac{1}{p_s} \right)$. 
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 \( \text{(**)} \) holds.

Summing \( \text{(**)} \) over all \( s \in S \) gives
\[
\sum_{s \in S} p_s \log_b \left( \frac{1}{p_s} \right) - \sum_{s \in S} p_s y_s \leq \frac{1}{\ln b} \left( \sum_{s \in S} \frac{1}{b^{y_s}} - \sum_{s \in S} p_s \right)
\]
By Hypothesis \( \sum_{s \in S} \frac{1}{b^{y_s}} \leq 1 \) and, since \( p \) is a probability distribution, \( \sum_{s \in S} p_s = 1 \). Hence
\[
\sum_{s \in S} p_s \log_b \left( \frac{1}{p_s} \right) - \sum_{s \in S} p_s y_s \leq 0.
\]
and so
\[
\sum_{s \in S} p_s \log_b \left( \frac{1}{p_s} \right) \leq \sum_{s \in S} p_s y_s
\]
with equality if and only if \( \sum_{s \in S} \frac{1}{b^{y_s}} = 1 \) and \( y_s = \log_b \left( \frac{1}{p_s} \right) \) for all \( s \in S \).

**Notation III.12.** 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 \( \left( f(s) \right)_{s \in S} \) with coefficients in \( J \).

**Definition III.13.** Let \( p \) be a probability distribution on the alphabet \( S \) and \( b > 1 \). The entropy of \( p \) to the base \( b \) is defined as
\[
H_b(p) = \sum_{s \in S \atop p_s \neq 0} p_s \log_b \left( \frac{1}{p_s} \right).
\]
If no base is mentioned, the base is assumed to be 2. \( H(p) \) means \( H_2(p) \) and \( \log(a) = \log_2(a) \)

Note that \( \lim_{x \to 0} x \log_b \left( \frac{1}{x} \right) = \lim_{y \to \infty} \log_b \frac{y}{y} = 0 \). So we will usually interpret the undefined expression \( 0 \log_b \left( \frac{1}{0} \right) \) as 0 and just write
\[
H_b(p) = \sum_{s \in S} p_s \log_b \left( \frac{1}{p_s} \right)
\]
Using notation III.12 we have $p = (p_s)_{s \in S}$, $\log_b \left( \frac{1}{p} \right) = \left( \log_b \left( \frac{1}{p_s} \right) \right)_{s \in S}$ and

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

**Example III.14.** 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.15** (Fundamental Theorem, Lower Bound). Let $p$ be a positive probability distribution on the alphabet $S$, let $b$ be an integer with $b > 1$ and let $c$ a $b$-nary PF code for $S$. Let $y$ be the codewords length of $c$. Then

$$H_b(p) \leq L_p(c)$$

**Proof.** Let $K$ be the Kraft-McMillan number of $c$. Since $c$ is PF we know that $K \leq 1$, see II.28. By III.8 $K = \sum_{s \in S} \frac{1}{p_s}$. Thus $\sum_{s \in S} \frac{1}{p_s} \leq 1$ and the conditions of III.11 are fulfilled. Since $L_p(c) = p \cdot y$ 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 $y_s = \left\lfloor \log_b \left( \frac{1}{p_s} \right) \right\rfloor$, where $\lfloor r \rfloor$ is the smallest integer larger or equal to $r$. Using notation III.12 this means $y = \left\lfloor \log_b \left( \frac{1}{p} \right) \right\rfloor$.

**Definition III.16.** Let $c$ be a $b$-nary code for the alphabet $S$ with codeword lengths $y$ and let $p$ be a probability distribution on $S$. $c$ is called a $b$-nary Shannon-Fano (SF) code for $S$ with respect to $p$ if $c$ is a PF-code and $y_s = \left\lfloor \log_b \left( \frac{1}{p_s} \right) \right\rfloor$ for all $s \in S$ with $p_s \neq 0$.

**Theorem III.17** (Fundamental Theorem, Upper Bound). Let $S$ be an alphabet, $p$ a positive probability distribution on $S$ and $b > 1$ an integer.
(a) Let $c$ be any $b$-nary SF code. Then $p$.

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

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

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

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

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

$$\log_b \left( \frac{1}{p_s} \right) \leq y_s$$

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

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

Hence by III.9 there exists $b$-nary PF code $c$ for $S$ with codewords length $(y_s)_{s \in S}$. Then $c$ is a SF code and so (b) holds in this case.

Example III.18. 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

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

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

We now use the tree method to construct a code with this parameter:
Since \( y = (4, 2, 3, 4, 3) \) we can choose the code \( c \) as follows

\[
(1000, 00, 010, 1001, 011)
\]

We have

\[
H_b(p) = p \cdot \log_2 \left( \frac{1}{p} \right) \approx (0.1, 0.4, 0.2, 0.1, 0.2) \cdot (3.3, 1.3, 2.3, 3.3, 2.3)
\]

\[
= 0.33 + 0.52 + 0.46 + 0.33 + 0.46 = 2.08 \approx 2.1
\]

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

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

**Proof.** By [III.15], \( H_b(p) \leq L \).

Let \( d \) be any \( b \)-nary PF-code for \( S \) with respect to \( p \). Since \( c \) is an optimal \( b \)-nary PF-code for \( S \) with respect to \( p \) we have \( L_p(c) \leq L_p(d) \).

By [III.17] a) there exists an \( b \)-nary SF-code \( d \) for \( S \) with respect to \( p \). Then [III.17] b) shows that \( L_p(d) < H_b(p) + 1 \) and so also \( L_p(c) < H_b(p) + 1 \).
III.5  Hufman rules

**Lemma III.20.** Let \( p \) be a positive probability distribution on the alphabet \( S \) and let \( c : S \to T^* \) be an optimal \( b \)-nary PF-code for \( S \) with respect to \( p \). Suppose \( |S| \geq 2 \). Let \( y \) be the codeword length of \( c \) and let \( M \) the maximal length of a code word.

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

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

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

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

(e) Any parent of a codeword of length \( M \) is the parent of at least two codewords of length \( M \).

(f) There exist two codewords of length \( M \) with a common parent.

**Proof.**

(a): Suppose for a contradiction 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 codeword length. But this is a contradiction, since \( c \) is optimal.

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

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

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

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

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

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

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

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

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

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

(d) Let \( t \in S \) with \( \ell(c(t)) = M \). By assumption \( \ell(w) < M \) and so \( \ell(w) < \ell(c(t)) \). Hence (b) shows that there exists \( s \in S \setminus t \) such that \( c(s) \) is a prefix of \( w \) or \( w \) is a prefix of \( c(s) \). By hypothesis of (d) no codeword is a prefix of \( w \) and we conclude that \( w \) is proper prefix of \( c(s) \). So we can apply (c) with \( s \) in place of \( t \) and (c) holds.

(e) Let \( w \) be the parent of a codeword \( u \) of length \( M \). Then \( \ell(w) = M - 1 \). Note that any prefix of \( w \) is proper prefix of \( u \). As \( c \) is PF this shows that no prefix of \( w \) is a codeword. Hence (c) shows that \( w \) is the proper prefix of two codewords \( b_1 \) and \( b_2 \). Then \( M - 1 = \ell(w) < \ell(b_i) \leq M \) and so \( \ell(b_i) = M \) and \( w \) is the parent of \( b_i \).

(f) Let \( b \) a codeword of length \( M \). Since \( c \) is a prefix-free code with \( |S| \geq 2 \) we know that \( c \) is regular, see (1.15(b)). Thus \( b \neq \emptyset \) and so \( b \) has a parent \( w \). By (e) \( w \) is the parent of two codewords of length \( M \) and so (f) holds. \( \square \)

**Theorem III.21.** Let \( p \) a positive probability distribution on the alphabet \( S \) with \( |S| \geq 2 \). Let \( d,e \) be distinct symbols in \( S \) such that \( p_d \leq p_e \) and \( p_e \leq p_s \) for all \( s \in S \setminus \{d,e\} \). Define \( S := S \setminus \{d\} \) and let \( \tilde{c} \) be a binary PF-code on \( S \).

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

\[
(H1) \quad \tilde{p}_s = \begin{cases} p_d + p_e & \text{if } s = e \\ p_s & \text{if } s \neq e \end{cases}
\]

Also define a binary code \( c \) on \( S \) by

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

Then

(a) \( c \) is a binary prefix-free code for \( S \).
(b) \( L_p(c) = L_{\tilde{p}}(\tilde{c}) + \tilde{p}_e. \)

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

**Proof.** Put \( \hat{S} := S \setminus \{d, e\}. \)

[A] Let \( s, t \in \hat{S} \) such that \( c(s) \) is a prefix of \( c(t) \). We need to show that \( s = t \).

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

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

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

Suppose finally that \( s \notin \hat{S} \) and \( t \notin \hat{S} \). Then \( \tilde{c}(s) = \tilde{c}(e)i \) and \( \tilde{c}(t) = \tilde{c}(e)j \) where \( i, j \in \{0, 1\} \). Hence \( \tilde{c}(e)i \) is prefix of \( \tilde{c}(e)j \). So \( \tilde{c}(e)i = \tilde{c}(e)j \) and it follows that \( i = j \). Thus either \( i = j = 0 \) and \( s = t = d \) or \( i = j = 1 \) and \( s = t = e \). In either case \( s = t \).

[B] Then \( \hat{S} = \hat{S} \cup \{e\} \) and \( S = \hat{S} \cup \{d, e\} \). Let \( y \) and \( \tilde{y} \) be the codewords length of \( c \) and \( \tilde{c} \), respectively. Then

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

and

\[
\tilde{p}_e = p_d + p_e, \quad \tilde{y}_d = \tilde{y}_e + 1, \quad \tilde{y}_e = \tilde{y}_e + 1
\]

Hence

\[
L_p(c) = \sum_{s \in S} p_s y_s
\]

\[
= p_d y_d + p_e y_e + \sum_{s \in \hat{S}} p_s y_s
\]

\[
= p_d (\tilde{y}_e + 1) + p_e (\tilde{y}_e + 1) + \sum_{s \in \hat{S}} p_s y_s
\]

\[
= (p_d + p_e) (\tilde{y}_e + 1) + \sum_{s \in \hat{S}} p_s y_s
\]

\[
= \tilde{p}_e + \tilde{p}_e \tilde{y}_e + \sum_{s \in \hat{S}} \tilde{p}_s \tilde{y}_s
\]

\[
= \tilde{p}_e + \sum_{s \in \hat{S}} \tilde{p}_s \tilde{y}_s
\]

\[
= \tilde{p}_e + L_{\tilde{p}}(\tilde{c})
\]

So [B] holds.

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

$[c] \iff$ Suppose $\tilde{c}$ is an optimal binary PF code.

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

Suppose that there exists $f \in \{d, e\}$ such that $\ell(a(f)) < M$. By [III.20][f] there exists at least two codewords of $a$ of length $M$. As $\ell(a(f)) < M$ at most one of these two codewords can be in $\{d, e\}$ and and so we can choose $g \in S$ with $\ell(a(g)) = M$. Then $\ell(a(f)) < \ell(a(g))$ and [III.20][a] shows that $p_g \leq p_f$. By choices of $e$ and $d$, $p_d \leq p_e \leq p_g$ and so $p_f \leq p_g$. Thus $p_f = p_g$ and interchanging the codewords for $f$ and $g$ does not change the average code length of $a$. We therefore may and do choose the optimal code $a$ such that both $a(d)$ and $a(e)$ have length $M$.

By [III.20][f] there exists two codewords of $a$ of length $M$ with a common parent. Permuting codewords of equal length does not change the average codeword length. So we may and do choose $a$ such that $a(d)$ and $a(e)$ have a common parent $u$ and that $a(d) = u0$ and $a(e) = u1$. Let $\tilde{a}$ be the code for $\tilde{S}$ defined by

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

Since $a$ is PF and $u$ is a prefix of $a(d)$ we know that $u$ is not a codeword of $a$. Let $w$ be a codeword of $a$ with $u$ as a prefix. Since $u$ is not a codeword, $u \neq w$. Thus $M - 1 = \ell(u) < \ell(w) \leq M$. So $\ell(w) = M$ and thus $w = u0 = a(d)$ or $w = u1 = a(e)$. Hence $u$ is not the prefix of any $a(s), s \in \tilde{S}$. 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 $L_{\tilde{p}}(\tilde{a}) = L_{\tilde{p}}(\tilde{c})$. It follows that $L_{\tilde{p}}(\tilde{a}) + \tilde{p}e = L_{\tilde{p}}(\tilde{c}) + \tilde{p}e$. From (b) applied to $a$ and $c$ we conclude that $L_{\tilde{p}}(a) = L_{\tilde{p}}(c)$. Since $a$ is optimal, this means that also $c$ is optimal. $\Box$

**Example III.22.** Use Huffman’s Rules H1 and H2 to construct an optimal code with respect to the probability distribution $(0.3, 0.2, 0.2, 0.15, 0.1, 0.05)$
Chapter IV

Data Compression

IV.1 The Comparison Theorem

Theorem IV.1 (The comparison theorem). Let \( p \) and \( q \) be positive probability distributions on the alphabet \( S \) and let \( b > 1 \). Then

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

with equality if and only if \( p = q \).

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

\[
\sum_{s \in S} \frac{1}{y_s} = \sum_{s \in S} q_s = 1.
\]

Theorem III.11 now shows that

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

with equality if and only if \( p_s \neq 0 \) 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 \( s \in S \).

\[\Box\]

Theorem IV.2. Let \( p \) be a positive 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 \( q := (\frac{1}{m})_{s \in S} \) 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 probability 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 probability 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
\[
p_i \leq p_i + \sum_{j \in I, j \neq i} p_j = \sum_{j \in I} p_j = 1.
\]

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

Corollary IV.4. Let \( I \) be an alphabet and \( p \) an \( I \)-tuple with coefficients in \( \mathbb{R} \). Put \( t = \sum_{i \in I} p_i \). Suppose that \( p_i \geq 0 \) for all \( i \in I \) and that \( t \neq 0 \). Then \( \left( \frac{p_i}{t} \right)_{i \in I} \) 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 \( \left( \frac{p_i}{t} \right)_{i \in I} \) 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 $j \in J$. (So $f' = \sum_{j \in J} \text{Col}_j(f)$ is the sum of the columns of $f$, while $f'' = \sum_{i \in I} \text{Row}_i(f)$ is the sum of the rows of $f$.)

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

**Example IV.6.**

<table>
<thead>
<tr>
<th>f</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>f'</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>e</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>f''</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

**Lemma IV.7.** Let $I$ and $J$ be alphabets and $p$ be an $I \times J$-matrix with non-negative real coefficients and marginal tuples $p'$ and $p''$. Then $p$ is a probability distribution if and only if $p'$ is a probability distribution and if and only if $p''$ is a probability distribution.

**Proof.** Since $p_{ij} \geq 0$ we also have $p'_i \geq 0$ and $p''_j \geq 0$ for all $i \in I, j \in J$. We compute
\[
\sum_{i \in I} p'_i = \sum_{i \in I} \left( \sum_{j \in J} p_{ij} \right) = \sum_{(i,j) \in I \times J} p_{ij} = \sum_{j \in J} \left( \sum_{i \in I} p_{ij} \right) = \sum_{j \in J} p''_j.
\]

If one $p, p', p''$ 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.8.** 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
\[
p = p' \otimes p''
\]
Example IV.9. (a) Compute \( f' \otimes f'' \):

<table>
<thead>
<tr>
<th>( f' \otimes f'' )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
<th>( f' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>0.18</td>
<td>0.18</td>
<td>0.24</td>
<td>0.6</td>
</tr>
<tr>
<td>( e )</td>
<td>0.12</td>
<td>0.12</td>
<td>0.16</td>
<td>0.4</td>
</tr>
<tr>
<td>( f'' )</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

(b) Consider

<table>
<thead>
<tr>
<th>( p )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
<th>( p' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>( e )</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>( p'' )</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

Since \( 0.3 \cdot 0.6 \neq 0.1 \), \( p' \) and \( p'' \) are not independent with respect to \( p \).

Lemma IV.10. Let \( p' \) and \( p'' \) be probability distributions on \( I \) and \( J \), respectively. Then:

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

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

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

Proof. (a) We have

\[
\sum_{j \in J} (p' \otimes p'')_{i,j} = \sum_{j \in J} p'_i p''_j = p'_i \left( \sum_{j \in J} p''_j \right) = p'_i \cdot 1 = p'_i
\]

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

(b) By (a) the marginal tuples of \( p' \otimes p'' \) are probability distributions and so by IV.7 also \( p' \otimes p'' \) is a probability distribution.

(c) Follows immediately from the definition of independent. \( \square \)

Theorem IV.11. Let \( I \) and \( J \) be alphabets, let \( b > 1 \) and let \( p \) be a positive probability distribution on \( I \times J \) with marginal distributions \( p' \) and \( p'' \). Then

\[
H_b(p) \leq H_b(p') + H_b(p'')
\]

with equality if and only if \( p' \) and \( p'' \) are independent with respect to \( p \).
Proof. We have

\[ H_b(p') + H_b(p'') = \sum_{i \in I} p'_i \log_b \left( \frac{1}{p'_i} \right) + \sum_{j \in J} p''_j \log_b \left( \frac{1}{p''_j} \right) \]

\[ = \sum_{i \in I} \left( \sum_{j \in J} p_{ij} \right) \log_b \left( \frac{1}{p'_i} \right) + \sum_{j \in J} \left( \sum_{i \in I} p_{ij} \right) \log_b \left( \frac{1}{p''_j} \right) \]

\[ = \sum_{i \in I, j \in J} p_{ij} \log_b \left( \frac{1}{p'_i p''_j} \right) \]

\[ = \sum_{s \in I \times J} p_s \log_b \left( \frac{1}{(p' \otimes p'')_s} \right) \]

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

\[ \square \]

IV.3 Coding in blocks

Definition IV.12. A source is a pair \((S, P)\), where \( S \) is an alphabet and \( P \) is a function

\[ P : S^* \rightarrow [0, 1], \]

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 \( S \). \( P(a_1 a_2 \ldots a_n) \) 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.13. 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 $r$ in the alphabet $S$. So $x$ is a string of length $nr$ in the alphabet $S$. Hence $(S^r)^n = S^{rn}$ and $(S^r)^* = \bigcup_{n=0}^{\infty} (S^r)^n = \bigcup_{n=0}^{\infty} S^{nr} \subseteq S^*.$

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

$$y_l := y_1y_2 \ldots y_r$$

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

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

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

$$p^{(l_1, \ldots, l_r)}(x_1x_2 \ldots x_r) = \sum_{\substack{(y_1y_2 \ldots y_u) \in S^u \mid y_1 = x_1, y_2 = x_2, \ldots, y_r = x_r}} P(y_1y_2 \ldots y_u)$$

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

**Example IV.14.** Suppose $(\mathbb{B}, P)$ is a binary source with

<table>
<thead>
<tr>
<th></th>
<th>$P(000)$</th>
<th>$P(001)$</th>
<th>$P(010)$</th>
<th>$P(011)$</th>
<th>$P(100)$</th>
<th>$P(101)$</th>
<th>$P(110)$</th>
<th>$P(111)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.05</td>
<td>0.2</td>
<td>0.1</td>
<td>0.25</td>
<td>0.1</td>
<td>0.15</td>
<td>0.05</td>
</tr>
</tbody>
</table>

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

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

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

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

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

(3) Compute $P(11)$.

By Condition IV.12(ii) in the definition of a source, $P(as) = \sum_{s \in S} P(as)$ and so

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

Lemma IV.15. Let $(S, P)$ be a source, $r \in \mathbb{N}$ and $l = (l_1, \ldots, l_r)$ be a strictly increasing $r$-tuple of positive integers.

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

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

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

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

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

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

Then $\mu$ is a bijection, and, after identifying $S^{t+r}$ with $S^t \times S^r$ via $\mu$, $p^l$ 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 $\emptyset$ is the only element of $S^r$. Hence

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

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

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

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

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

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

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

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

So the function

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

is inverse to function

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

Hence \( \rho \) and \( \mu \) are bijections. Put \( m := t + r \). Then \( m \geq l_r \). Hence any \( y \in S^m \) can be uniquely written as \( wu \) with \( w \in S^{t_r} \) and \( u \in S^{m-t_r} \). Moreover, \( y_i = w_i \). Thus for \( x \in S^r \):

\[
p^l(x) = \sum_{\substack{w \in S^{t_r} \\ w_i = x}} P(w) - \text{Definition of } p^l
\]

\[
= \sum_{\substack{w \in S^{t_r} \\ w_i = x}} \sum_{u \in S^{m-t_r}} P(wu) - (a)
\]

\[
= \sum_{y \in S^m \\ y_i = x} P(y) - \text{Substitution } y = wu
\]

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

\[
p^l(x) = \sum_{y \in S^m \\ y_i = x} P(y) = \sum_{a \in S^t} P(\rho(a, x)) = \sum_{a \in S^t} p^{t+r}(\rho(a, x))
\]

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

By (b), \( p^{t+r} \) is a probability distribution on \( S^{t+r} \) and by (d), \( p^l \) is the marginal tuple of \( p^{t+r} \) on \( S^r \). So by IV.7, \( p^l \) is a probability distribution on \( S^r \).

Lemma IV.16. 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.15(a)

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

Moreover, \( P^r(\emptyset) = P(\emptyset) = 1 \) and so \( P^r \) is a source on \( S^r \).
IV.4 Memoryless Sources

Definition IV.17. 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.18. Let \(p\) be a probability distribution on the alphabet \(S\). Define

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

inductively by

\[ P(\emptyset) = 1 \]

and

\[ P(as) = P(a)p_s \]

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

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

for any \(s_1, s_2, \ldots, s_n \in S\). Then \((S, P)\) is the unique memoryless source with \(P(s) = p_s\) for all \(s \in S\).

Proof. Since \(0 \leq p_s \leq 1\) for all \(s \in S\), also \(0 \leq P(a) \leq 1\) for all \(a \in S^*\). So \(P\) is indeed a function from \(S^*\) to \([0, 1]\). By definition of \(P\), \(P(\emptyset) = 1\). 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 \(a = \emptyset\) 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.19. 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 \(p^{t+r}\).\]
CHAPTER IV. DATA COMPRESSION

Proof. Let

\[(*) \quad a = s_1 \ldots s_t \in S^t \quad \text{and} \quad b = s_{t+1} \ldots s_{t+r} \in S^r\]

Then

\[p^{t+r}(ab) = P(ab) \quad \text{\text{ -- definition of } p^{t+r}}\]
\[= P(s_1 \ldots s_t s_{t+1} \ldots s_{t+r}) \quad \text{\text{ -- IV.18}}\]
\[= p_s \ldots p_s p_{s_{t+1}} \ldots p_{t+r} \quad \text{\text{ -- IV.18}}\]
\[= (p_s \ldots p_s)(p_{s_{t+1}} \ldots p_{t+r}) \quad \text{\text{ -- IV.18}}\]
\[= P(s_1 \ldots s_t)P(s_{t+1} \ldots s_{t+r}) \quad \text{\text{ -- IV.18}}\]
\[= p'(a)p'(b) \quad \text{\text{ -- definition of } p' \text{ and } p'^r}\]

Hence \(p^{t+r} = p^t \otimes p^r\). \(\square\)

IV.5 Coding a stationary source

Definition IV.20. A source \((S, P)\) is called stationary if

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

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

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

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

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

Lemma IV.21. Let \((S, P)\) be a source. Let \(r, t \in \mathbb{N}\) and identifying \(S^{t+r}\) with \(S^t \times S^r\).

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

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

Proof. (a) These are the special cases \(l = (1, \ldots, t)\) and \(l = (t+1, t+2, \ldots, t+r)\) in IV.15(d).

(b) Since \(P\) is stationary, we have \(p^r = p^{(t+1, \ldots, t+r)}\) and so (a) implies (b). \(\square\)
IV.5. CODING A STATIONARY SOURCE

Lemma IV.22. A memoryless source is stationary.

Proof. Let \((S, P)\) be a memoryless source and \(r, t \in \mathbb{N}\). By IV.19 \(p^{t+r} = p^t \otimes p^r\) and so by IV.10(c) \(p^r\) is the marginal distribution of \(p^{t+r}\) on \(p^r\). By IV.21(a) \(p^{(t+1, \ldots, t+r)}\) is also the marginal distribution of \(p^{t+r}\) on \(S^r\). Thus \(p^r = p^{(t+1, \ldots, t+r)}\) and so \(P\) is stationary.

Theorem IV.23. Let \((S, P)\) be a source and let \(r, t \in \mathbb{Z}^+\) and let \(b > 1\).

(a) \[ H_b(p^{t+r}) \leq H_b(p^t) + H_b(p^r) \]

with equality if and only if \(p^t\) and \(p^{(t+1, \ldots, t+r)}\) are independent with respect to \(p^{t+r}\).

(b) Suppose \(P\) is stationary. Then

\[ H_b(p^{t+r}) \leq H_b(p^t) + H_b(p^r) \]

with equality if and only if \(p^t\) and \(p^r\) are independent with respect to \(p^{t+r}\).

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

(d) Suppose \((S, P)\) is stationary and that \(r\) divides \(t\). Then

\[ \frac{H_b(p^t)}{t} \leq \frac{H_b(p^r)}{r} \]

(e) Suppose \((S, P)\) is memoryless. Then

\[ \frac{H_b(p^t)}{t} = H_b(p^r) \]

Proof. (a): By IV.21(b) \(p^r\) and \(p^{(t+1, \ldots, t+r)}\) are the marginal distributions of \(p^{t+r}\) on \(S^t\) and \(S^r\). Thus (a) follows from IV.11.

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

(c): For \(q = 1\) (d) is obviously true. Suppose now that (c) holds for \(q\). Then by (b)

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

with equality if \((S, P)\) is memoryless. (Note here that by IV.19 \(p^{qr}\) and \(p^r\) are independent with respect to \(p^{qr+r}\) for memoryless sources.) So (d) holds for \(q + 1\) and hence by the Principal of induction, for all \(q \in \mathbb{Z}\).

(d): Since \(r\) divides \(t\), \(t = qr\) for some \(q \in \mathbb{Z}^+\). Thus
\[
\frac{H_b(p^t)}{t} = \frac{H_b(p^q)}{q} < qH_b(p^r) = \frac{H_b(p^r)}{r}.
\]

(e) Suppose \((S, P)\) is memoryless. By (c) applied with \(q = t\) and \(r = 1\).

\[
H_b(p^t) = H_b(p^{t+1}) = tH_b(p) = H_b(p)
\]

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

\[\square\]

\textbf{Definition IV.24.} Let \((S, P)\) be a source and \(b > 1\). Then the \emph{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

\[
\liminf a_m = \lim_{m \to \infty} \left( \inf_{n \geq m} a_n \right)
\]

\textbf{Lemma IV.25.} Let \((S, P)\) be a stationary source and \(b > 1\). Then

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

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

\[
\inf_{n \geq m} \frac{H_b(p^n)}{n} = \inf_{n \geq 1} \frac{H_b(p^n)}{n}.
\]

For this let \(n \in \mathbb{Z}^+\). By IV.23(d) we have \(\frac{H_b(p^{nm})}{nm} \leq \frac{H_b(p^n)}{n}\). Since \(nm \geq m\) this gives

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

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

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

Clearly \(\inf_{n \geq m} \frac{H_b(p^n)}{n} \geq \inf_{n \geq 1} \frac{H_b(p^n)}{n}\) and so (e) holds.

Hence

\[
H_b(P) = \liminf_{m \to \infty} \frac{H_b(p^m)}{m} = \lim_{m \to \infty} \left( \inf_{n \geq m} \frac{H_b(p^n)}{n} \right) = \lim_{m \to \infty} \inf_{n \geq 1} \frac{H_b(p^n)}{n} = \inf_{n \geq 1} \frac{H_b(p^n)}{n}.
\]

\[\square\]
Lemma IV.26. Let $(S, P)$ be a memoryless source and $b > 1$. Then $H_b(P) = H_b(p)$.  

*Proof.* Just recall that by IV.23(e) we have $\frac{H_b(p^r)}{r} = H_b(p)$ for all $r \in \mathbb{Z}^+$. \hfill $\Box$

Theorem IV.27 (Coding Theorem for Memoryless Sources). Let $(S, P)$ be a memoryless source, let $b$ an integer with $b > 1$ and let $\epsilon > 0$. Let $n$ be any integer with $n > \frac{1}{\epsilon}$. Then there exists a $b$-nary PF-code $c_n$ for $S^n$ such that

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

*Proof.* Note that $\frac{1}{n} < \epsilon$. Also since $P$ is memoryless, $\frac{H_b(p^n)}{n} = H_b(p)$, see IV.23(e). By the Fundamental Theorem III.19 there exists a $b$-nary PF-code $c_n$ for $S^n$ with $L_{p^n}(c_n) \leq H_b(p^n) + 1$. Then

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

\hfill $\Box$

Theorem IV.28 (Coding Theorem for Sources). Let $(S, P)$ be a source, $b > 1$ an integer, $\epsilon > 0$ and $k > 0$. Then there exists an integer $n$ with $n > k$ and a $b$-nary prefix-free code $c_n$ for $S^n$ such that

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

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

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

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

\[(\ast)\]

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

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

\[(\ast\ast)\]

$$
\frac{1}{n} \leq \frac{1}{m} < \frac{1}{\frac{2}{\epsilon}} = \frac{\epsilon}{2}.
$$

By the Fundamental Theorem, there exists a prefix-free code $b$-nary code $c_n$ for $S^n$ with
\[(\ast \ast \ast) \quad L_{p^n}(c_n) \leq H_b(p^n) + 1.\]

Combining \[(\ast), (\ast \ast) \text{ and } (\ast \ast \ast)\] we obtain

\[
\frac{L_{p^n}(c_n)}{n} \leq \frac{H_b(p^n)}{n} + \frac{1}{n} \leq H_b(P) + \frac{\epsilon}{2} + \frac{1}{n} \leq H_b(P) + \frac{\epsilon}{2} + \frac{\epsilon}{2} = H_b(P) + \epsilon.
\]

\[\square\]

### IV.6 Arithmetic codes

**Definition IV.29.** Let \(S\) be a 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 a pair \((S, <)\), where \(S\) is a set and \(<\) is an ordering on \(S\).

**Example IV.30.** Suppose \(S = \{s_1, s_2, \ldots, s_m\}\) is an alphabet of size \(m\). Then there exists a unique ordering on \(S\) with \(s_i < s_{i+1}\) for all \(1 \leq i < m\). Indeed we have \(s_i < s_j\) if and only if \(i < j\).

**Notation IV.31.** We will often say that \(S\) is an ordered alphabet rather than that \((S, <)\) is an ordered alphabet. We also will say that \(S = (s_1, \ldots, s_m)\) is an ordered alphabet if \((S, <)\) is the order alphabet with \(S = \{s_1, \ldots, s_m\}\) and \(s_i < s_j\) for all \(1 \leq i < j \leq m\).

**Definition IV.32.** Let \(S\) be an ordered alphabet and \(p\) a positive probability distribution on \(S\). For \(s \in S\) define:

\[\alpha = \alpha(s) = \sum_{t \leq s} p_t,\]

where \(\alpha = 0\) if \(s\) is the smallest element in \(S\);

\[n' = n'(s) = \left\lfloor \log_2 \left( \frac{1}{p_s} \right) \right\rfloor;\]

\[n = n(s) = n' + 1;\]

\[c' = c'(s) = \left\lfloor 2^n \alpha \right\rfloor;\]
IV.6. ARITHMETIC CODES

\[ c(s) = z_1 z_2 \ldots z_n \in \mathbb{B}^*, \]

where

\[ z_1, \ldots, z_n \in \mathbb{B} \quad \text{with} \quad c' = \sum_{i=1}^{n} z_i 2^{n-i} = z_1 2^{n-1} + z_2 2^{n-2} + \ldots + z_{n-1} 2 + z_n. \]

Then the function

\[ c : S \rightarrow \mathbb{B}^*, \quad s \mapsto c(s) \]

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

**Example IV.33.** Determine the arithmetic code for the ordered alphabet \( S = (a, d, e, b, c) \) with respect to the probability distribution \( p = (0.1, 0.3, 0.2, 0.15, 0.25) \).

<table>
<thead>
<tr>
<th>( s )</th>
<th>( p )</th>
<th>( \alpha )</th>
<th>( \frac{1}{p} )</th>
<th>( n' )</th>
<th>( n )</th>
<th>( 2^n )</th>
<th>( 2^n \alpha )</th>
<th>( c' )</th>
<th>( \sum_{i=1}^{n} z_i 2^{n-i} )</th>
<th>( c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>0.1</td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>32</td>
<td>0</td>
<td>0 \cdot 16 + 0 \cdot 8 + 0 \cdot 4 + 0 \cdot 2 + 0</td>
<td>00000</td>
<td></td>
</tr>
<tr>
<td>( d )</td>
<td>0.3</td>
<td>0.1</td>
<td>3.3\ldots</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>0.8</td>
<td>1</td>
<td>0 \cdot 4 + 0 \cdot 2 + 1</td>
<td>001</td>
</tr>
<tr>
<td>( e )</td>
<td>0.2</td>
<td>0.4</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>6.4</td>
<td>7</td>
<td>0 \cdot 8 + 1 \cdot 4 + 1 \cdot 2 + 1</td>
<td>0111</td>
</tr>
<tr>
<td>( b )</td>
<td>0.15</td>
<td>0.6</td>
<td>6.3\ldots</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>9.6</td>
<td>10</td>
<td>1 \cdot 8 + 0 \cdot 4 + 1 \cdot 2 + 0</td>
<td>1010</td>
</tr>
<tr>
<td>( c )</td>
<td>0.25</td>
<td>0.75</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>16</td>
<td>6</td>
<td>6</td>
<td>1 \cdot 4 + 1 \cdot 2 + 0</td>
<td>110</td>
</tr>
</tbody>
</table>

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.34.** Let \( z = z_1 z_2 \ldots z_n \in \mathbb{B}^n \). Then the rational number

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

is called the rational number associated to \( z \) and is denoted by \( 0_* z \).

**Example IV.35.** \( 0_* 1011 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{8 + 2 + 1}{16} = \frac{11}{16} \).

**Lemma IV.36.** Let \( z \in \mathbb{B}^* \). Then \( 0_* z \in [0, 1) \).

**Proof.** Let \( z = z_1 \ldots z_n \). Then

\[ 0 \leq 0_* z = \sum_{i=1}^{n} \frac{z_i}{2^i} \leq \sum_{i=1}^{\infty} \frac{1}{2^i} = \frac{1}{2^0} = 1 \]

\( \square \)
Lemma IV.37. Let $\alpha \in [0, 1)$ and $n \in \mathbb{Z}^+$ with $\alpha + \frac{1}{2^n} < 1$.

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

(b) Put $c':= \lceil 2^n \alpha \rceil$ and let $z = z_1 \ldots z_n$ with $z_i \in \mathbb{B}$. Then $c' = \sum_{i=1}^{n} z_i 2^{n-i}$.

(c) Let $y \in \mathbb{B}^\ast$. Then $0_z y \in [\alpha, \alpha + \frac{1}{2^n} - 1]$.

Proof. (a) and (b): Let $z = z_1 \ldots z_n$ be any element of $\mathbb{B}^n$. Put $d = \sum_{i=1}^{n} z_i 2^{n-i}$. Then $d \in \mathbb{Z}$ and $z$ is uniquely determined by $d$. We compute

$$2^n \cdot 0_z z = 2^n \cdot \sum_{i=1}^{n} z_i \geq \frac{1}{2^n} = \sum_{i=1}^{n} z_i 2^{n-i} = d.$$

Hence

$$0_z z \in [\alpha, \alpha + \frac{1}{2^n})$$

$$\iff \alpha \leq 0_z z < \alpha + \frac{1}{2^n}$$

$$\iff 2^n \alpha \leq d < 2^n \alpha + 1 \quad \text{multiplication by } 2^n$$

$$\iff d = \lceil 2^n \alpha \rceil \quad \text{since } d \in \mathbb{Z}$$

$$\iff d = c'.$$

This proves (a) and (b).

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

$$\alpha \leq 0_z z \leq 0_z y = \sum_{i=1}^{n} \frac{z_i}{2^n} + \sum_{i=1}^{m} \frac{y_i}{2^{n+i}} = 0_z z + \frac{1}{2^n} \sum_{i=1}^{m} \frac{y_i}{2^{n+i}}$$

$$= 0_z z + \frac{1}{2^n} 0_z y < 0_z z + \frac{1}{2^n}$$

$$< \lceil \alpha + \frac{1}{2^n} \rceil + \frac{1}{2^n} = \alpha + \frac{1}{2^{n-1}}.$$

Thus (c) holds.

Theorem IV.38. 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 $I_s := [\alpha(s), \alpha(s) + p_s]$. 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) $0_z c(s) y \in I(s)$ for each $s \in S$ and $y \in \mathbb{B}^\ast$.

(c) $c$ is a prefix-free binary code.
Proof. (a) Let \( S = \{ s_1, s_2, \ldots, s_m \} \) with \( s_1 < s_2 < \ldots < s_{m-1} < s_m \). 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 \( p_{s_i} > 0 \) this gives

\[
0 = \alpha(s_1) < \alpha(s_2) < \ldots < \alpha(s_{m-1}) < 1
\]

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

(b) Let \( s \in S \) and \( y \in \mathbb{B}^* \). Recall that \( n = n' + 1 \) and \( n' = \left\lceil \log_2 \frac{1}{p_s} \right\rceil \). Thus \( 2n' \geq \frac{1}{p_s} \) and so \( \frac{1}{2^{n'}} \leq p_s \). Therefore \( \alpha + \frac{1}{2^{n'}} < \alpha + \frac{1}{2^{n'}} \leq \alpha + p_s \leq 1 \). Hence we can apply Corollary IV.40 and it follows that

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

(c) Let \( s, t \in S \) with \( s \neq t \) and let \( y \in \mathbb{B}^* \). By (b), \( 0, c(t) \in I(t) \) and \( 0, c(s)y \in I(s) \). By (a) \( I(t) \cap I(s) = \emptyset \). Hence \( c(t) \neq c(s)y \) and so \( c(s) \) is not a prefix of \( c(t) \). Thus \( c \) is a prefix-free code.

**Theorem IV.39.** Let \( c \) be the arithmetic code for the ordered alphabet \( S \) with respect to the positive probability distribution \( p \). Then

\[
L_p(c) < H_2(p) + 2.
\]

**Proof.** Let \( s \in S \). Then \( c(s) \) has length

\[
n(s) = n'(s) + 1 = \left\lceil \log_2 \left( \frac{1}{p_s} \right) \right\rceil + 1 < \log_2 \left( \frac{1}{p_s} \right) + 2.
\]

So

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

**Corollary IV.40** (Coding Theorem for Arithmetic codes). Let \((S,P)\) be a source such that \( p \) is positive. Let \( \epsilon > 0 \). Then there exists an integer \( n \) such that

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

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

Moreover, if \( P \) is memoryless, this holds for any integer \( n \) with \( n > \frac{2}{\epsilon} \).
Proof. Follow the proof for Coding Theorem for Sources (IV.28) with the following modifications:

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

\[
(2*) \quad n > \frac{4}{\epsilon}
\]

After Equation (2): Let \( c \) be an arithmetic code on \( S^n \) with respect to \( p^n \). By IV.39 we get

\[
(3*) \quad L_{p^n}(c) < H(p^n) + 2
\]

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.

\[\square\]

IV.7 Coding with a dynamic dictionary

**Definition IV.41.** An dictionary \( D \) based on the alphabet \( S \) is a 1-1 \( N \)-tuple \( D = (d_1, \ldots, d_N) \) with coefficients in \( S^* \) for some \( N \in \mathbb{N} \). (Here 1-1 means \( D \) is 1-1 as a function from \( \{1, \ldots, N\} \) 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.42.** Let \( S = (s_1, s_2, \ldots, s_m) \) be an ordered alphabet. Then \( D = (s_1, \ldots, s_m) \) is a dictionary based on \( S \).

**Algorithm IV.43** (LZW encoding). Let \( S = (s_1, s_2, \ldots, s_m) \) be an ordered alphabet and let \( X \) be a non-empty message in \( S \). Define

\[
(*) \quad a \text{ positive integer } n;
\]

\[
(*) \quad \text{non-empty message } Y_k, \ 1 \leq k \leq n \text{ in } S;
\]

\[
(*) \quad \text{positive integer } c_k, \ 1 \leq k \leq n;
\]

\[
(*) \quad \text{non-empty messages } X_k, \ 0 \leq k < n \text{ in } S;
\]

\[
(*) \quad \text{symbols } z_k, \ 1 \leq k < n \text{ in } S; \text{ and}
\]
IV.7. CODING WITH A DYNAMIC DICTIONARY

(*) dictionaries $D_k$, $0 \leq k < n$, based on $S$

inductively as follows:
For $k = 0$ define
$$D_0 := (s_1, \ldots, s_m) \quad \text{and} \quad X_0 := X.$$
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 $c(X) := c_1 c_2 \ldots c_n$. Also define $c(\emptyset) = \emptyset$. The function
$$c : S^* \to \mathbb{N}^*, X \to c(X)$$
is called the LZW-encoding function for the ordered alphabet $S$.

Example IV.44. Given the ordered alphabet $(a, b, c, d, e)$. Determine the LZW encoding of $bdddad$.

<table>
<thead>
<tr>
<th>$m + k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{m+k}$</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>bd</td>
<td>dd</td>
<td>dda</td>
<td>aa</td>
<td>ad</td>
</tr>
</tbody>
</table>

So the encoding is 247114.

Lemma IV.45. With the notation as in the LZW encoding algorithm:

(a) $X_k = Y_{k+1} Y_{k+2} \ldots Y_n$ for all $0 \leq k < n$.

(b) $X = Y_1 \ldots Y_n$.

\[1\] 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
(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 definition of $n$ we have $X_{n-1} = Y_n$. 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.46 (LZW decoding). Let $S = (s_1, s_2, \ldots, s_m)$ be an ordered alphabet and let $u = c_1 \ldots c_n$ be a message in $\mathbb{N}$. If $u \neq \emptyset$ and $c_k < m+k$ for all $1 \leq k \leq n$ define

(*) 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 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 $Y_k$.

3In this case $Y_{k+1} = Y_k z_k$.
IV.7. CODING WITH A DYNAMIC DICTIONARY

(●) \( D_k = (D_{k-1}, Y_k, z_k) \).

\[ D_k = (D_{k-1}, Y_k, z_k) \]

Put \( e(u) = Y_1 \ldots Y_n \).

If \( u = \emptyset \) or \( c_k \geq m + k \) for some \( 1 \leq k \leq n \), define \( e(u) = \emptyset \).

The function \( e : \mathbb{N}^* \to S^* \), \( u \to e(u) \) is called the LZW-decoding for \( S \).

**Example IV.47.** Find the LZW-decoding of the message 4.7.4.3.9.11 for the ordered alphabet \( S = (b, a, e, f, d, c) \).

<table>
<thead>
<tr>
<th>( u )</th>
<th>4</th>
<th>7</th>
<th>4</th>
<th>3</th>
<th>9</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_k )</td>
<td>f</td>
<td>ff</td>
<td>f</td>
<td>e</td>
<td>fe</td>
<td>fef</td>
</tr>
<tr>
<td>( z_k )</td>
<td>f</td>
<td>f</td>
<td>e</td>
<td>f</td>
<td>f</td>
<td>fef</td>
</tr>
</tbody>
</table>

\( d_{m+k} \) | b | a | e | f | d | c | ff | fff | fe | ef | fef |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( m+k )</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

So the decoding is fffeefefef.
Chapter V

Error Correcting

V.1 Decision rules

Definition V.1. 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 $n$). Let $a \in C$ and $z \in \mathbb{B}^n$. A decision rule for $C$ is a function $\sigma : \mathbb{B}^n \to C$.

Definition V.2. Let $x, y \in \mathbb{B}^n$. Then $d(x, y) := \{i \mid 1 \leq i \leq n, x_i \neq y_i\}$. $d(x, y)$ is called the Hamming distance of $x$ and $y$.

Definition V.3. Let $C \subseteq \mathbb{B}^n$ be a binary code and $\sigma$ 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 of $C$ if $d(a, z) = k$.

(b) We say that $\sigma$ corrects $(a, z)$ if $a = \sigma(z)$.

(c) We say that $\sigma$ is $r$-error correcting if $\sigma$ corrects all $k$-bit errors for $0 \leq k \leq r$.

Example V.4. Given the binary code $C = \{000, 110, 101, 011\}$ (so $C$ consist of all even messages of length 3) and the decision rule

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>000</th>
<th>100</th>
<th>010</th>
<th>001</th>
<th>110</th>
<th>101</th>
<th>011</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>000</td>
<td>110</td>
<td>000</td>
<td>101</td>
<td>110</td>
<td>101</td>
<td>011</td>
<td>110</td>
</tr>
</tbody>
</table>

Does $\sigma$ correct 0-bit errors? Does $\sigma$ 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 $\sigma$ is 0-error correcting.

$\sigma$ corrects some 1-bit errors but not all:
\[ \sigma(001) = 101 \neq 011 \]

Thus \((101, 001)\) is a 1-bit error corrected by \(\sigma\), but \((011, 001)\) is a 1-bit error not corrected by \(\sigma\). So \(\sigma\) is not 1-error correcting.

**Example V.5.** Given the code \(C = \{000, 000, 110, 110, 101, 101, 011, 011\}\). (Note that \(C\) is obtained by doubling the code in [V.4].) Define the decision rule \(\sigma\) for \(C\) by

\[
\sigma(xy) = \begin{cases} 
  xx & \text{if } x \text{ is even} \\
  yy & \text{if } x \text{ is odd}
\end{cases}
\]

for all \(x, y \in \mathbb{B}^3\). Show that \(\sigma\) 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 \(\sigma\) is indeed 1-error correcting.

**Definition V.6.** 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 V.7.** 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

<table>
<thead>
<tr>
<th></th>
<th>111 000</th>
<th>111 000</th>
<th>001 110</th>
<th>110 110</th>
<th>001 001</th>
<th>110 011</th>
<th>110 011</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001 110</td>
<td>001 110</td>
<td>110 110</td>
<td>110 110</td>
<td>110 110</td>
<td>110 110</td>
<td>110 110</td>
</tr>
<tr>
<td>d(a, b)</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

So the minimum distance is 3.

**Definition V.8.** Let \(A\) and \(B\) be sets. Then

\[ A + B = (A \setminus B) \cup (B \setminus A) \]

\(A + B\) is called the symmetric difference of \(A\) and \(B\).

**Lemma V.9.** Let \(A\) and \(B\) be sets.
Lemma V.11. Let \( a, b \in \mathbb{B}^n \).

(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) \( d(a, c) \leq d(a, b) + d(b, c) \), with equality if and only if \( D(a, b) \cap D(b, c) = \emptyset \).

Proof. \( \square \) Let \( 1 \leq i \leq n \).

- If \( i \notin D(a, b) \cup D(b, c) \), then \( a_i = b_i = c_i \) and so \( i \notin D(a, c) \).
- If \( i \in D(a, b) \setminus D(b, c) \), then \( a_i \neq b_i = c_i \) and so \( a_i \neq c_i \) and \( i \in D(a, c) \).
- If \( i \in D(b, c) \setminus D(a, b) \), then \( a_i = b_i \neq c_i \) and so \( a_i \neq c_i \) and \( i \in D(a, c) \).
- If \( i \in D(a, b) \cap D(b, c) \), then \( a_i \neq b_i \neq c_i \). Since \( \mathbb{B} \) only has two elements this gives \( a_i = c_i \) and so \( i \notin D(a, c) \).

Hence \( i \in D(a, c) \) if and only of \( i \in (D(a, b) \setminus D(b, c)) \cup (D(b, c) \setminus D(a, b)) \) = \( D(a, b) + D(b, c) \)\( \square \):

\[
d(a, c) = |D(a, c)| = |D(a, b) + D(b, c)| = |D(a, b)| + |D(b, c)| - 2|D(a, b) \cap D(b, c)| = d(a, b) + d(b, c) - 2|D(a, b) \cap D(b, c)|
\]

\( \square \): Note that \( |D(a, b) \cap D(b, c)| \geq 0 \) with equality if and only if \( D(a, b) \cap D(b, c) = \emptyset \). So \( \square \) follows from (b). \( \square \)

Lemma V.11. Let \( a, b \in \mathbb{B}^n \). Put \( d := d(a, b) \) and let \( 0 \leq e \leq d \). Then there exists \( x \in \mathbb{B}^n \) with \( d(a, x) = e \) and \( d(b, x) = d - e \).

Proof. Since \( e \leq d = d(a, b) = |D(a, b)| \) we can choose a subset \( J \) of \( D(a, b) \) with \( |J| = e \). Define \( x \in \mathbb{B}^n \) by

\[
x_i = \begin{cases} 
 b_i & \text{if } i \in J \\
 a_i & \text{if } i \notin J.
\end{cases}
\]

Then
\[ x_i = b_i \neq a_i \quad \text{if} \ i \in J \]
\[ x_i = a_i \neq b_i \quad \text{if} \ i \in D(a, b) \setminus J \]
\[ x_i = a_i = b_i \quad \text{if} \ i \notin D(a, b) \]

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

**Definition V.12.** Let \( r, n \in \mathbb{N} \), \( x \in \mathbb{B}^n \) and \( X \subseteq \mathbb{B}^n \). Then

\[ N_r(x) = \{ y \in \mathbb{B}^n \mid d(x, y) \leq r \} \]

and

\[ N_r(X) = \{ y \in \mathbb{B}^n \mid d(x, y) \leq r \ \text{for some} \ x \in X \} = \bigcup_{x \in X} N_r(x) \]

\( N_r(x) \) is called the neighborhood of radius \( r \) around \( x \).

**Example V.13.** Compute \( N_1(0110) \).

\[ N_1(0110) = \{0110, 1110, 0010, 0100, 0111\} \]

**Lemma V.14.** Let \( C \subseteq \mathbb{B}^n \) be a code and \( \sigma \) a decision rule for \( C \). Then \( \sigma \) is \( r \)-error correcting if and only if \( \sigma(z) = a \) for all \( a \in C \) and all \( z \in N_r(a) \).

**Proof.** \( \sigma \) is \( r \)-error correcting if and only if for all \( 0 \leq k \leq r \), \( \sigma \) corrects all \( k \)-bit errors \((a, z)\) of \( C \).

This holds if and only if \( \sigma(z) = a \) for all \( a \in C \) and \( z \in \mathbb{B}^n \) with \( d(a, z) \leq r \) and so if and only if \( \sigma(z) = a \) for all \( a \in C \) and \( z \in N_r(a) \). \hfill \( \square \)

**Definition V.15.** Let \( C \subseteq \mathbb{B}^n \) be a binary code.

(a) \( C \) 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 \( a \neq b \).

(b) Let \( \sigma \) be a decision rule for \( C \). We say that \( \sigma \) is a Minimum Distance decision rule if

\[ d(\sigma(z), z) \leq d(a, z) \]

for all \( a \in C \) and all \( z \in \mathbb{B}^n \) an

**Remark V.16.** Let \( C \subseteq \mathbb{B}^n \) be a binary code. Then there exists a Minimum Distance decision rule for \( C \).
**Proof.** For each \( z \in \mathbb{B}^n \) choose \( z' \in C \) with \( d(z, z') \) minimal. Then

\[
\sigma : \mathbb{B}^n \rightarrow C, \quad z \mapsto z'
\]

is a Minimum distance decision rule for \( C \).

\[ \square \]

**Theorem V.17.** 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) \( N_r(a) \cap N_r(b) = \emptyset \) 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) \( \implies \) (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 \( C \) has minimal distance at least \( 2r + 1 \) this implies that \( a = b \). So there exists at most one codeword of distance less or equal to \( r \) from \( z \).

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

(d) \( \implies \) (e): By V.16 there exists an Minimum Distance decision rule for \( C \). So if all such rules are \( r \)-error-correcting there exists an \( r \)-error-correcting decision rule for \( C \).

(e) \( \implies \) (a): Let \( \sigma \) be an \( r \)-error-correcting decision rule for \( C \). Let \( a, b \in C \) with \( a \neq b \) and put \( d := d(a, b) \). Let \( d = 2e + \epsilon \) with \( \epsilon \in \{0, 1\} \) and \( e \in \mathbb{N} \). By V.11 there exists \( z \in \mathbb{B}^n \) with \( d(a, z) = e \) and \( d(b, z) = d - e = e + \epsilon \). Since \( a \neq b \), we have \( \sigma(a) \neq a \) or \( \sigma(z) \neq b \). So at least one of \((a, z)\) and \((b, z)\) is not corrected by \( \sigma \). Since \( d \) is \( r \)-error correcting this implies that \( d(a, z) > r \) or \( d(b, z) > r \). Thus \( e > r \) or \( e + \epsilon > r \). In either case \( e + \epsilon > r \). So \( e + 1 \geq e + \epsilon > r \) and thus \( e \geq r \). We proved that \( e \geq r \) and \( e + \epsilon > r \). Hence \( d = 2e + \epsilon = e + (e + \epsilon) > r + r = 2r \) and \( d \geq 2r + 1 \). Thus \( \delta(C) \geq 2r + 1 \) and \( C \) is \( r \)-error-correcting.
CHAPTER V. ERROR CORRECTING

V.2 The Packing Bound

Lemma V.18. 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) $|N_r(x)| = \sum_{i=0}^{r} \binom{n}{i}$.

Proof. (a) The map $y \rightarrow D(x, y)$ is bijection between $\{y \in \mathbb{B}^n \mid d(x, y) = r\}$ and the set of subsets of size $r$ of $\{1, \ldots, n\}$. Note that there are exactly $\binom{n}{r}$ subsets of size $r$ of $\{1, \ldots, n\}$ and so (a) holds.

(b) Note that $N_r(x)$ 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 V.19. Let $C \subseteq \mathbb{B}^n$ and $r \in \mathbb{N}$. Then

(a) $|N_r(C)| \leq 2^n$ with equality if and only if $|N_r(C)| = \mathbb{B}^n$.

(b) $|N_r(C)| \leq |C| \sum_{i=0}^{r} \binom{n}{i}$ with equality if and only if $C$ is an $r$-error correcting code.

Proof. (a) Just observe that $N_r(C) \subseteq \mathbb{B}^n$ and $|\mathbb{B}^n| = 2^n$.

(b) Note that $N_r(C) = \bigcup_{a \in C} N_r(a)$ and so

\begin{equation}
\left| \bigcup_{a \in C} N_r(a) \right| \leq \sum_{a \in C} |N_r(a)|
\end{equation}

with equality if and only if $N_r(a) \cap N_r(b) = \emptyset$ for all $a, b \in C$ with $a \neq b$. So by V.17 equality holds in (1) if and only if $C$ is $r$-error-correcting.

By V.18(b)

\begin{equation}
\sum_{a \in C} |N_r(a)| = \sum_{a \in C} \sum_{i=0}^{r} \binom{n}{i} = |C| \sum_{i=0}^{r} \binom{n}{i}.
\end{equation}

Combining (1) and (2) gives (b).

Theorem V.20 (The Packing Bound). Let $C \subseteq \mathbb{B}^n$ be an $r$-error-correcting code. Then

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

and

$|C| \leq \frac{2^n}{\sum_{i=0}^{r} \binom{n}{i}}$. 
V.2. THE PACKING BOUND

Proof. By V.19

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

\[
\Box
\]

Definition V.21. Let \( C \subseteq \mathbb{B}^n \) be a code. Then the information rate of \( C \) is the real number

\[
\rho(C) := \log_{|\mathbb{B}^n|} |C| = \frac{\log_2 |C|}{n}.
\]

Example V.22. Use the packing bound to find an upper bound for the information rate of a 2-error correcting code \( C \subseteq \mathbb{B}^n \) of size 100.

Note that finding a upper bound for \( \rho(C) \) is equivalent to finding lower bound for \( n \). According to the packing bound,

\[ |C| \cdot \sum_{i=0}^{r} \binom{n}{i} \leq 2^n \]

and so

\[ 100 \left( 1 + n + \binom{n}{2} \right) \leq 2^n. \]

Hence

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

and

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

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

\[ 2^m - 10m - 5 > 0 \]

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

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

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 V.23.** Let $C \subseteq B^n$. Then $C$ is called a perfect code if there exists $r \in \mathbb{N}$ such that for all $z \in B^n$ there exists a unique $a \in C$ with $d(z, a) = r$.

**Lemma V.24.** Let $C \subseteq B^n$. 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.** Let $C \subseteq B^n$. Then

1. $C$ is a perfect code
2. $\exists r \in \mathbb{N} \left( \forall z \in B^n \text{ there exists a unique } a \in C \text{ with } d(a, z) \leq r \right)$ -- def of a perfect code
3. $\exists r \in \mathbb{N} \left( \forall z \in B^n \text{ there exists at most one } a \in C \text{ with } d(a, z) \leq r \right)$ and $\forall z \in B^n$ there exists $a \in C$ with $d(a, z) \leq r$
4. $\exists r \in \mathbb{N} \left( C \text{ is } r\text{-error-correcting} \right)$ and $\forall z \in B^n$ there exists $a \in C$ with $d(a, z) \leq r$
5. $\exists r \in \mathbb{N} \left( C \text{ is } r\text{-error-correcting and } B^n = N_r(C) \right)$ -- definition of $N_r(C)$
6. $\exists r \in \mathbb{N} \left( C \text{ is } r\text{-error-correcting and } 2^n = |N_r(C)| \right)$ -- V.19(a)
7. $\exists r \in \mathbb{N} \left( C \text{ is } r\text{-error-correcting and } 2^n = |C| \sum_{i=0}^{r} \binom{n}{i} \right)$ -- V.19(b)

\[ \square \]

**Definition V.25.** Let $\delta, n \in \mathbb{N}$ with $\delta \leq n$. Then $A(n, \delta)$ is the largest possible size of code $C \subseteq B^n$ with minimum distance at least $\delta$. That is

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

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

**Lemma V.26.** Let $\delta, n \in \mathbb{N}$ with $\frac{2}{3} n < \delta \leq n$. Then $A(n, \delta) = 2$. 
Proof. The code \{00...0, 11...1\} shows that \(A(n, \delta) \geq 2\). Now let \(C \subseteq \mathbb{B}^n\) be a binary code with \(|C| \geq 3\). We will show that

\[\delta(C) \leq \frac{2}{3}n.\]

For this let \(a, b, c\) be three distinct codewords in \(C\) and assume without loss \(d(a, b) \geq \frac{2}{3}n\) and \(d(b, c) \geq \frac{2}{3}n\). Put \(I := \{1, 2, \ldots, n\}\). Then

\[d(a, c) = |D(a, c)| = |D(a, b) + D(a, c)| = |D(a, b) \setminus D(b, c)| + |D(b, c) \setminus D(a, b)| \leq |I \setminus D(b, c)| + |I \setminus D(a, b)| = (n - d(b, c)) + (n - d(a, b)) = \frac{2}{3}n + \frac{2}{3}n = \frac{2}{3}n\]

So indeed \(\delta(C) \leq \frac{2}{3}n\). \(\square\)

**Example V.27.** Consider a binary code \(C \subseteq \mathbb{B}^{10}\) with \(\delta(C) \geq 7\). Then \(\delta(C) \geq 2 \cdot 3 + 1\) and so \(C\) is an 3-error-correcting code. Hence the Packing Bound shows that

\[|C| \leq \frac{2^{10}}{1 + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}} = \frac{1024}{1 + 10 + 45 + 120} = \frac{1024}{176} = 5.\ldots\]

Thus \(|C| \leq 5\). On the other hand, \(\frac{2}{3} \cdot 10 \leq 7 \leq 10\) and so the \(V.26\) shows that \(A(10, 7) = 2\). Thus \(|C| \leq 2\). This shows that, in general, the packing bound is not the best possible bound.
Chapter VI

Linear Codes

VI.1 Introduction to linear codes

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

\[
\begin{array}{c|cc}
+ & 0 & 1 \\
\hline
0 & 0 & 1 \\
1 & 0 & 0 \\
\end{array}
\quad
\begin{array}{c|cc}
\cdot & 0 & 1 \\
\hline
0 & 0 & 0 \\
1 & 0 & 1 \\
\end{array}
\]

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

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

\[
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{pmatrix}
+ \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix} = \begin{pmatrix}
x_1 + y_1 \\
x_2 + y_2 \\
\vdots \\
x_n + y_n
\end{pmatrix}
\quad
l \begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{pmatrix} = \begin{pmatrix}
rx_1 \\
rx_2 \\
\vdots \\
rx_n
\end{pmatrix}
\]

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 $0x = \bar{0}$ for all $x \in \mathbb{F}_2^n$. Here $\bar{0} = 00\ldots0$. 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
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 VI.3.** Compute $11011 + 10110$.

We have

\[
\begin{array}{c}
11011 \\
+ \\n10110 \\
\hline
= 01101 \\
\end{array}
\]

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

(i) $\bar{0} \in C$.

(ii) $lx \in C$ for all $l \in \mathbb{F}_2, x \in \mathbb{F}_2^n$.

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

**Example VI.5.** 1. \{0000, 1101, 1011, 0111\} is not a subspace of $\mathbb{F}_2^4$. Indeed we have

\[
\begin{array}{c}
1101 \\
+ \\n1011 \\
\hline
= 0110 \\
\end{array}
\]

Since 0110 \notin C, C is not closed under addition.

2. \{011, 101, 110\} is not a subspace of $\mathbb{F}_3^3$ since 000 \notin C.

\[^1\text{Since } lx = \bar{0} \text{ or } lx = x, \text{ this condition is redundant}\]
VI.1. INTRODUCTION TO LINEAR CODES

(3) \(\{000, 011, 101, 110\}\) is a subspace of \(\mathbb{F}_2^3\).

(4) Let \(C := \{x \in \mathbb{F}_2^n \mid x_1 + x_2 + \ldots + x_n = 0\}\). Then \(C\) is subspace of \(\mathbb{F}_2^n\). Note that \(C\) consists of all \(x \in \mathbb{F}_2^n\) such that \(\{i \mid x_i = 1\}\) is even.

**Definition VI.6.** Let \(C\) be a subspace of \(\mathbb{F}_2^n\).

(a) A basis for \(C\) is a \(k\)-tuple

\((v_1, \ldots, v_k)\)

with coefficients in \(C\) such that for each \(a\) in \(C\) there exists a unique \(k\)-tuple

\((l_1, \ldots, l_k)\)

with coefficients in \(\mathbb{F}_2\) such that

\[a = l_1 v_1 + \ldots + l_k v_k.\]

(b) \(\dim C := \log_2 |C|\).

(c) The triple \((n, \dim C, \delta(C))\) is called the linear parameter of \(C\).

**Lemma VI.7.** Let \(C\) be a subspace of \(\mathbb{F}_2^n\).

(a) If \((v_1, \ldots, v_k)\) is a basis for \(C\) then \(|C| = 2^k\). In particular, \(k = \log_2 |C| = \dim C\).

(b) \(0 \leq \dim C \leq n\). Moreover, \(\dim C = 0\) if and only if \(C = \{0\}\) and \(\dim C = n\) if and only if \(C = \mathbb{F}_2^n\).

(c) \(\rho(C) = \frac{\dim C}{n}\).

**Proof.**

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

\[\mathbb{F}_2^k \to C, \quad (l_1, \ldots, l_k) \to l_1 v_1 + \ldots + l_k v_k\]

is a well-defined bijection. Thus \(|C| = \mathbb{F}_2^k| = 2^k\).

(b) Since \(1 \leq |C| \leq |\mathbb{F}_2^n| = 2^n\), we get \(0 \leq \log_2 |C| = \dim C \leq n\).

(c) \(\rho(C) = \frac{\log_2 |C|}{n} = \frac{\log_2 2^k}{n} = \frac{k}{n}\).

**Example VI.8.** Let \(C = \{000 000, 111 000, 000 111, 111 111\}\). Find a basis for \(C\). Determine \(\dim C\) and \(\rho(C)\).

\(^2\)We will prove in **VI.20** that any subspace has a basis.
Both
\[(111\ 000, 000\ 111)\]
and
\[(111\ 000, 111\ 111)\]
are bases for \(C\). Thus \(\dim C = 2\) and \(\rho(C) = \frac{2}{6} = \frac{1}{3}\).

**Definition VI.9.** Let \(x = x_1 \ldots x_n \in \mathbb{F}_2^n\). Then \(\text{wt}(x) = |\{1 \leq i \leq n \mid x_i \neq 0\}|\) is called the weight of \(x\).

**Lemma VI.10.** Let \(x, y \in \mathbb{F}_2^n\).

(a) \(x + x = \bar{0}\). In particular, \(-x = x\) and \(x + y = x - y\).

(b) \(d(x, y) = \text{wt}(x - y) = \text{wt}(x + y)\).

(c) \(\text{wt}(x) = d(x, \overline{0})\).

(d) Let \(C \subseteq \mathbb{F}_2^n\) be a binary linear code. Then \(\delta(C)\) is the minimal weight of a non-zero codeword in \(C\).

**Proof.**

(a) Since \(0 + 0 = 0\) and \(1 + 1 = 0\) in \(\mathbb{F}_2\) we have \(k + k = 0\) for all \(k \in \mathbb{F}_2\). Hence also \(x + x = 0\) for all \(x \in \mathbb{F}_2^n\).

(b) \(d(x, y) = |\{1 \leq i \leq n \mid x_i \neq y_i\}| = |\{1 \leq i \leq n \mid x_i - y_i \neq 0\}| = \text{wt}(x - y)\)

(c) \(d(x, \overline{0}) = \text{wt}(x - \overline{0}) = \text{wt}(x)\).

(d) Let \(w\) be the minimal weight of a non-zero codeword, and let \(x, y \in C\) with \(x \neq y\) and \(d(x, y) = \delta(C)\). Since \(C\) is subspace of \(\mathbb{F}_2^n\) we have \(x - y \in C\) and so \(x - y\) is a non-zero codeword. Thus \(\text{wt}(x - y) \geq w\). Hence

\[\delta(C) = d(x, y) = \text{wt}(x - y) \geq w.\]

Let \(a \in C\) be non-zero codeword with \(\text{wt}(a) = w\). Since \(\bar{0} \in C\), we have \(d(a, \bar{0}) \geq \delta(C)\). Thus

\[w = \text{wt}(a) = d(a, \bar{0}) \geq \delta(C).\]

We proved that \(\delta(C) \geq w\) and \(w \geq \delta(C)\), so \(\delta(C) = w\). \(\square\)

**Example VI.11.** Let \(C := \{000\ 000, 111\ 000, 000\ 111, 111\ 111\}\) be the code from Example VI.8. Determine the parameter of \(D\).
VI.2. CONSTRUCTION OF LINEAR CODES USING MATRICES

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

VI.2 Construction of linear codes using matrices

Definition VI.12. (a) A binary matrix is a matrix with coefficients in $\mathbb{F}_2$

(b) Let $E$ be a binary $n \times k$ matrix. Then

$$\text{Col}(E) := \{ Ey \mid y \in \mathbb{F}_2^k \}.$$ 

$\text{Col}(E)$ is called the linear code generated by $E$, and $E$ is called a generating matrix for $\text{Col}(E)$. $\text{Col}(E)$ is also called the column space of $E$.

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

$$\text{Nul}(H) := \{ x \in \mathbb{F}_2^n \mid Hx = \bar{0} \}$$

$\text{Nul}(H)$ is called the null space of $H$. $H$ is called a check matrix for $\text{Nul}(H)$.

Lemma VI.13. (a) Let $E$ be a binary $n \times k$-matrix and $y \in \mathbb{F}_2^k$. Let $e_i := \text{Col}_i(E)$ be the $i$'th column of $E$. Then

$$Ey = y_1 e_1 + \ldots + y_k e_k.$$ 

(b) Let $E$ be a binary $n \times k$ matrix. Then $\text{Col}(E)$ is a subspace of $\mathbb{F}_2^n$.

(c) Let $H$ be a binary $m \times n$-matrix. Then $\text{Nul}(H)$ is a subspace of $\mathbb{F}_2^n$.

Proof. (a) The $i$-coefficient $Ey$ is $\sum_{l=1}^k e_{il} y_l$. The $i$-coefficient of $e_i$ is $e_{il}$ and so the the $i$-coefficient of $\sum_{l=1}^k y_l e_i$ is $\sum_{l=1}^k y_l e_{il}$. Note that $\sum_{l=1}^k e_{il} y_l = \sum_{l=1}^k y_l e_{il}$ and so (a) holds.

(b) $E\bar{0} = \bar{0}$. So $\bar{0} \in \text{Col}(E)$. Let $a, b \in \text{Col}(E)$. 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 $a + b \in \text{Col}(E)$.

(c) $H\bar{0} = \bar{0}$ and so $\bar{0} \in \text{Nul}(H)$. Let $a, b \in \text{Nul}H$. The $Ha = \bar{0}$ and $Hb = \bar{0}$. Hence
and so \( a + b \in \text{Nul}H \).

**Example VI.14.** Find the minimal distance of \( \text{Col}(E) \) where

\[
E := \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1
\end{bmatrix}.
\]

Is \( \text{Col}(E) \) 1-error correcting?

We will determine the minimum weight of a non-zero codeword. So \( x \in \text{Col}(E) \) with \( x \neq \mathbf{0} \).

Then \( x = Ey \) for some \( y \in F_3^2 \) with \( y \neq \mathbf{0} \). Let \( y = abc \). Then

\[
x = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1
\end{bmatrix} \begin{bmatrix}
a \\
b \\
c \\
a + b \\
b + c \\
a + c
\end{bmatrix} = \begin{bmatrix}
a \\
b \\
c \\
a + b \\
b + c \\
a + c
\end{bmatrix}
\]

Suppose \( \text{wt}(y) = 1 \). Then exactly one of \( a, b \) and \( c \) is equal to 1. Without loss \( a = 1 \) and \( b = c = 0 \). So \( x = 100101 \) and \( \text{wt}(x) = 3 \).

Suppose \( \text{wt}(y) = 2 \). Then exactly two of \( a, b \) and \( c \) are equal to 1. Without loss \( a = b = 1 \) and \( c = 0 \). So \( x = 110011 \) and \( \text{wt}(x) = 4 \).

Suppose \( \text{wt}(y) = 3 \). Then \( a = b = c = 1 \) and \( x = 111000 \). So \( \text{wt}(x) = 3 \).

Thus \( \text{Col}(E) \) has minimum distance \( \delta(C) = 3 = 2 \cdot 1 + 1 \) and \( \text{Col}(E) \) is 1-error correcting.

### VI.3 Standard form of check matrix

**Notation VI.15.** Let \( (I_a)_{a \in A} \) and \( (J_b)_{b \in B} \) be tuples of sets. Let \( M = [M_{ab}]_{a \in A}^{b \in B} \) be an \( A \times B \) matrix such that each \( M_{ab} \) is an \( I_a \times J_b \)-matrix. Put \( I = \bigcup_{a \in A} I_a \) and \( J = \bigcup_{b \in B} J_b \). 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 \( i \in I \) and \( b \) is the unique element of \( B \) with \( j \in J \).

**Example VI.16.** Given the matrices

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

\[
X_{21} = \begin{bmatrix} 9 & 10 & 11 \end{bmatrix}, \quad \text{and} \quad X_{22} = [12].
\]

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 VI.17.** Let \( A \) be a set. Then \( I_A \) is the \( A \times A \)-matrix

\[
I_A := \left[ \delta_{ab} \right]_{a \in A \atop b \in B} \quad \text{where} \quad \delta_{ab} := \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}
\]

\( I_A \) is called the \( A \times A \) identity matrix. If \( n \in \mathbb{N} \), then

\[
I_n := I_{(1, \ldots, n)}.
\]
Example VI.18.

\[
I_4 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\] and

\[
I_{\{x,y\}} = \begin{bmatrix}
x & y \\
x & 1 \\
y & 0
\end{bmatrix}
\]

Lemma VI.19. Let \( A \) be binary \( m \times k \) matrix. Define

\[
n := m + k,
\]

\[
E := \begin{bmatrix}
I_k \\
A
\end{bmatrix},
\]

\[
H := [A I_m]
\]

and

\[
c_E : \mathbb{F}_2^k \to \mathbb{F}_2^n, \quad y \mapsto Ey.
\]

Let \( a \in \mathbb{F}_2^n \) and write \( a = vw \) with \( v \in \mathbb{F}_2^k \) and \( w \in \mathbb{F}_2^m \).  (\( v \) is called the message part of \( a \) and \( w \) the check part of \( a \))

(a) \( c_E(y) = Ey = (y, Ay) \) for all \( y \in \mathbb{F}_2^k \).

(b) \( c_E \) is a bijection from \( \mathbb{F}_2^k \) to \( \text{Col}(E) \).  In particular, \( c_E \) is a binary code for \( \mathbb{F}_2^k \) with set of codewords \( \text{Col}(E) \).

(c) The columns of \( E \) form a basis for \( \text{Col}(E) \).

(d) \( \text{Col}(E) \) is a binary linear code of length \( n \) and dimension \( k \).

(e) \( a \in \text{Col}(E) \) if and only if \( w = Av \) and if and only if \( a \in \text{Nul}(H) \).

(f) \( \text{Col}(E) = \text{Nul}(H) \).

(g) Suppose \( a \in \text{Col}(E) \).  Then \( a = c_E(v) \).

Proof. (a):

\[
c_E(y) = Ey = \begin{bmatrix}
I_k \\
A
\end{bmatrix} y = \begin{bmatrix}
y \\
Ay
\end{bmatrix} = (y, Ay)
\]

[b]: We need to show that \( c_E \) is 1-1 and \( \text{Im} c_E = \text{Col}(E) \). Let \( y, z \in \mathbb{F}_2^k \) with \( c(y) = c(z) \). Then \( (y, Ay) = c_E(y) = c_E(z) = (z, Ay) \) and so \( y = z \). Hence \( c \) is 1-1. Also

\[
\text{Col}(E) = \{E y \mid y \in \mathbb{F}_2^k\} = \{c_E(y) \mid y \in \mathbb{F}_2^k\} = \text{Im} c_E.
\]
VI.3. STANDARD FORM OF CHECK MATRIX

(e): Set \( e_i := \text{Col}_i(E) \) and let \( y \in \mathbb{F}_2^k \). Then
\[
c_E(y) = Ey = y_1e_1 + y_2e_2 + \ldots + y_ke_k.
\]
and, since \( c_E \) is a bijection from \( \mathbb{F}_2^k \) to \( \text{Col}(E) \), we conclude that for each \( a \in \text{Col}(E) \) there exists a unique \( y = (y_1, \ldots, y_k) \in \mathbb{F}_2^k \) with \( a = y_1e_1 + \ldots + y_ke_k \). Hence \((e_1, \ldots, e_k)\) is basis for \( \text{Col}(E) \).

(d): By (c) \((e_1, \ldots, e_k)\) is a basis for \( \text{Col}(E) \), so VI.7(a) shows that \( \dim \text{Col}(E) = k \). Note that \( \text{Col}(E) \subseteq \mathbb{F}_2^{m+k} = \mathbb{F}_2^n \) and so \( \text{Col}(E) \) has length \( n \).

(e): \([v, w] \in \text{Col}(E) \iff (v, w) = Ey \) for some \( y \in \mathbb{F}_2^k \)
\([v, w] = (y, Ay) \) for some \( y \in \mathbb{F}_2^k \)
\( v = y \) and \( w = Ay \) for some \( y \in \mathbb{F}_2^k \)
\( w = Av \)
\( Av - w = \tilde{0} \)
\( Av + w = \tilde{0} \)
\([A I_m] \begin{pmatrix} v \\ w \end{pmatrix} = \tilde{0} \)
\([v, w] \in \text{Nul}([A I_m]) \)

So \( a \in \text{Col}(E) \) if and only if \( w = Av \) and if and only if \( a \in \text{Nul}(H) \).

(f): By (e) \( a \in \text{Col}(E) \) if and only if \( a \in \text{Nul}(H) \), so \( \text{Col}(E) = \text{Nul}(H) \).

(g): Since \( a \in \text{Col}(E) \) (e) shows that \( w = Av \). Thus \( c_E(v) = (v, Av) = (v, w) = a \). \( \square \)

**Definition VI.20.** Let \( C \subseteq \mathbb{F}_2^n \) be a linear code.

(a) A check matrix \( H \) for \( C \) is said to be in standard form if \( H = [A I_m] \) for some \( k \times m \) 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 VI.21.** Let \( C, D \subseteq \mathbb{F}_2^n \). We say that \( D \) is a permutation of \( C \) if there exists a bijection \( \pi : \{1, \ldots, n\} \rightarrow \{1, \ldots, n\} \) with
\[
D = \{a_{\pi(1)}a_{\pi(2)}\ldots a_{\pi(n)} \mid a_1a_2\ldots a_n \in C\}.
\]
Example VI.22. Show that

\[ D = \{000, 100, 001, 101\} \]

is a permutation of

\[ C = \{000, 010, 001, 011\} \]

\( D \) is obtained from \( C \) by interchanging the first two bits, that is via the permutation:

\[ \pi : \begin{array}{ccc} 1 & 2 & 3 \\ 2 & 1 & 3 \end{array} \]

Notation VI.23. Let \( I, J \) be sets, \( i, k \in I \), \( A \) an \( I \times J \) matrix and \( x \) and \( y \) \( J \)-tuples.

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

(b) \( R_{i,k} x y A = R_i x (R_k y A) \). (So \( R_{i,k} x y A \) is the matrix obtained from \( A \) by replacing \( \text{Row}_k \) by \( y \) and then replacing \( \text{Row}_i \) by \( x \).)

Definition VI.24. Let \( I \) and \( J \) be sets and \( i, k \in I \) with \( i \neq k \). Let \( a^i := \text{Row}_i(A) \). Then \( R_i \leftrightarrow R_k \) and \( R_i + R_k \rightarrow R_k \) are the functions with domain the binary \( I \times J \)-matrix defined as follows:

(a) \( (R_i \leftrightarrow R_k)(A) = R_{i,k}a^ka^iA \). So \( R_i \leftrightarrow R_k \) interchanges row \( i \) and row \( k \) of \( A \)

(b) \( (R_i + R_k \rightarrow R_k)(A) = R_{i,k}(a^i + a^k)A \). So \( (R_i + R_k \rightarrow R_k) \) adds row \( i \) to row \( k \) of \( A \).

An elementary row operation is one of the function \( R_i \leftrightarrow R_j \) and \( R_i + R_k \rightarrow R_k \).

Elementary column operations are defined similarly, using the symbol \( C \) in place of \( R \) and using columns rather than rows.

Lemma VI.25. (a) Let \( H \) and \( G \) be binary \( m \times n \)-matrices and suppose \( G \) is obtain from \( H \) by sequence of elementary row operation. Then \( \text{Nul}(H) = \text{Nul}(G) \).

(b) Let \( E \) and \( F \) be binary \( m \times n \)-matrices and suppose \( F \) is obtain from \( E \) by a sequence of elementary column operation. Then \( \text{Col}(E) = \text{Col}(F) \).

Proof. (a) Let \( h^i := \text{Row}_i(H) \). Note that \( x \in \text{Nul}(H) \) if and only \( Hx = 0 \) and if and only of if \( h^i x = 0 \) for all \( 1 \leq i \leq m \).

Suppose that \( G = (R_i \leftrightarrow R_k)(H) \). Then \( H \) and \( G \) have the same rows (just in a different order) and so \( x \in \text{Nul}(H) \) if and only if \( x \in \text{Nul}(G) \).

Suppose next that \( G = (R_i + R_k \rightarrow R_k)(H) \). Let \( x \in \text{Nul}(H) \). \( 1 \leq l \leq n \) with \( l \neq k \), 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 \( x \in \text{Nul}(G) \) and thus \( \text{Nul}(H) \subseteq \text{Nul}(G) \).
Note that $h^k = h^i + (h^i + h^k) = g^i + g^k$. Hence $(R_i + R_k \rightarrow R_k)(G) = H$ so Nul($G$) \subseteq Nul($H$) by the result of the previous paragraph, applied with $G$ and $H$ interchanged. Thus Nul($G$) = Nul($H$) and (a) holds.

(b): Put $e_j = \text{Col}_j(E)$. Then $x \in \text{Col}(E)$ if and only if $x = Ey$ for some $y \in \mathbb{F}_2^k$ and so if and only if

$$x = \sum_{i=1}^{n} y_i e_j$$

for some $(y_1, \ldots, y_k) \in \mathbb{F}_2^k$.

Suppose that $F = (C_j \leftrightarrow C_k)(E)$. Then $E$ and $F$ have the same columns (just in a different order) and $x \in \text{Col}(E)$ if and only if $x \in \text{Col}(F)$.

Suppose next that $F = (C_j + C_k \rightarrow C_k)(E)$ and let $x \in \text{Col}(E)$. Then, for some $y \in \mathbb{F}_2^k$,

$$x = \sum_{l \neq j} y_l e_l = y_j e_j + y_k e_k + \sum_{1 \leq l \leq n \text{ or } j \neq k} y_l e_l = (y_j + y_k) e_j + y_k e_k + \sum_{1 \leq l \leq n \text{ or } j \neq k} y_l f_l$$

and so $x \in \text{Col}(F)$.

Note that $E = (C_j + C_k \rightarrow C_k)(F)$ and so the preceding result, applied with $E$ and $F$ interchanged, gives $\text{Col}(F) \subseteq \text{Col}(E)$. Thus $\text{Col}(F) = \text{Col}(E)$ and (b) holds.

**Theorem VI.26.** Let $C$ be a subspace of $\mathbb{F}_2^n$. Then there exists a permutation $D$ of $C$ such that $D$ has a generating matrix and a check matrix in standard form. In particular, $C$ has a basis.

**Proof.** We will first show that there exists a permutation of $D$ with generating matrix in standard form. For this we will construct a finite sequence of subspaces $C_k \subseteq \mathbb{F}_2^n$ and matrices $E_k$, $k = 0, 1, 2, \ldots$, such that

(i) $C_k$ is a permutation of $C$,

(ii) $E_k$ is a generating matrix for $C_k$, and

(iii) $E_k$ is of the form

$$E_k = \begin{bmatrix} I_k & 0 \\ * & * \end{bmatrix}.$$
CHAPTER VI. LINEAR CODES

Put $C_0 := C$ and let $E_0$ be any generating matrix for $C_0$, for example one can choose $E_0$ to be a matrix whose columns consists of all the codewords of $C$ (in some order). Suppose now that $C_k$ and $E_k$ have already been constructed. Put $l := k + 1$.

**Step 1.** Let $A_l$ be the matrix obtained from $E_k$ by deleting the zero columns of $E_k$.

Observe that $A_l$ is still a generating matrix for $C_k$ and, since none of the first $k$-columns of $E_k$ have been deleted, $A_l$ still is of the form $[I_k \ 0 \ \ast \ \ast \ \ast]$. Let $m$ be the number of columns of $A_l$.

If $m = k$, then $A_l$ is in standard form and we can choose $D = C_k$ and we are done.

So suppose $m > k$. Note that all columns of $A$ are non-zero and so have an entry equal to 1.

**Step 2.** Choose $1 \leq i \leq n$ with $a_{li} = 1$. Let $C_l$ by the code obtain from $C_k$ via the permutation $l \leftrightarrow i$. Put $B_l = (R_l \leftrightarrow R_i)(A_l)$.

Since $A_l$ is of shape $[I_k \ 0 \ \ast \ \ast \ \ast]$ we have $a_{lj} = 0$ for $1 \leq j \leq k$. Thus $l \leq i \leq n$. So the first $k$ rows of $A_l$ are the same as the first $k$ rows of $A_l$. Also $b_{ii} = a_{ii} = 1$. Thus $B_l$ has shape

$$B_l = \begin{bmatrix}
I_k & 0 & 0 \\
* & 1 & * \\
* & * & *
\end{bmatrix}.$$ 

Since $A_l$ is a generating matrix for $C_k$, $B_l$ is a generating matrix for $C_l$.

**Step 3.** Let $E_l$ be the matrix obtained from $B_l$ by adding Column $l$ of $B_l$ to Column $j$ of $B_l$ for each $1 \leq j \leq m$ such that $j \neq l$ and the $jl$-coefficient of $B_l$ is equal to 1.

Since column operations do not change the columns space, $E_l$ is still a generating matrix for $D_l$. Also $E_l$ has shape

$$E_l = \begin{bmatrix}
I_k & 0 & 0 \\
0 & 1 & 0 \\
* & * & *
\end{bmatrix} = \begin{bmatrix}
I_l & 0 \\
* & * 
\end{bmatrix}.$$ 

Observe that the number of columns of $E_l$ is the number of non-zero columns of $E_k$. So the numbers of columns of $E_l$ is less or equal to the number of columns of $E_0$. Thus the above algorithm will terminate in a find number of iterations.

This completes the proof that there exists a permutation $D$ of $C$ with a generating matrix $E = \left[ \begin{array}{c} I_k \\ A \end{array} \right]$ in standard form.

Put $m := n - k$ and $H = [A I_m]$. By VI.19 Nul$(H) = \text{Col}(E) = D$ and so $H$ is check matrix in standard form for $D$. Moreover, the columns of $E$ form a basis for $\text{Col}(E) = D$. Hence $D$ has a basis and so also $C$ has a basis. \qed
Example VI.27. Let

\[ C := \{0000, 0011, 1110, 1101\} \]

Find a permutation \( D \) of \( C \), a generating matrix \( E \) for \( D \) in standard form, and check matrix \( H \) of \( D \) in standard form.

<table>
<thead>
<tr>
<th>( l )</th>
<th>( A_l )</th>
<th>( l \leftrightarrow i )</th>
<th>( B_l )</th>
<th>( E_l )</th>
<th>( C_l )</th>
</tr>
</thead>
</table>
| 0 | \[
\begin{bmatrix}
0 & 1 & 1 \\
0 & 1 & 1 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
\end{bmatrix}
\] | 1 \( \leftrightarrow \) 3 | \[
\begin{bmatrix}
1 & 1 & 0 \\
0 & 1 & 1 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
\end{bmatrix}
\] | \[
\begin{bmatrix}
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 \\
\end{bmatrix}
\] | \[
\begin{bmatrix}
0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 \\
\end{bmatrix}
\] |
| 1 | \[
\begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & 1 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
\end{bmatrix}
\] | | \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
1 & 1 & 0 \\
\end{bmatrix}
\] | \( E_1 \) | \( C_1 \) |
| 2 | \[
\begin{bmatrix}
1 & 0 \\
0 & 1 \\
0 & 1 \\
1 & 1 \\
\end{bmatrix}
\] | \( 2 \leftrightarrow 2 \) | \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
1 & 1 & 0 \\
\end{bmatrix}
\] | \( E_1 \) |

Put

\[
E := A_3 = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
0 & 1 \\
1 & 1 \\
\end{bmatrix} \quad H := \begin{bmatrix}
0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 \\
\end{bmatrix} \quad \text{and} \quad D := C_2 = \begin{bmatrix}
0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 \\
\end{bmatrix}
\]

Then \( D \) is a permutation of \( C \) with generating matrix \( E \) and check matrix \( H \).

Moreover, \((1001, 0111)\) (the columns of \( E \)) is a basis for \( D \) and since \( C \) is obtained from \( D \) via the permutation \( 1 \leftrightarrow 3 \) we conclude that

\((0011, 1101)\)
VI.4 Constructing 1-error-correcting linear codes

**Definition VI.28.** Let \( v_1, \ldots, v_k \in \mathbb{F}_2^n \). Then \((v_1, \ldots, v_k)\) is called linearly dependent if there exists a non-zero \((l_1, \ldots, l_k) \in \mathbb{F}_2^k\) with

\[
l_1v_1 + l_2v_2 + \ldots + l_kv_k = \bar{0}.
\]

\((v_1, \ldots, v_k)\) is called linearly independent, if it is not linearly dependent.

**Remark VI.29.** A \( k \)-tuple \((v_1, \ldots, v_k)\) with \( v_i \in \mathbb{F}_2^n \) is linearly independent if for all \((l_1, \ldots, l_k) \in \mathbb{F}_2^k\):

\[
l_1v_1 + l_2v_2 + \ldots + l_kv_k = \bar{0} \iff l_1 = 0, l_2 = 0, \ldots, l_k = 0
\]

**Lemma VI.30.** Let \( H \) be a check matrix for the binary linear code \( C \).

(a) The minimum distance of \( C \) is the minimal number of columns of \( H \) whose sum is equal to \( \bar{0} \). It is also equal to the minimal number of linearly dependent columns of \( H \).

(b) Let \( r \in \mathbb{Z}^+ \). Then \( C \) is \( r \)-error-correcting if and only if no (non-empty) sum of \( 2r \) or less columns of \( H \) is equal to \( \bar{0} \).

**Proof.** (a): Recall first from [VI.10](#) that \( \delta(C) \) is the minimum weight of a non-zero codeword. Let \( x = x_1x_2 \ldots x_n \in \mathbb{F}_2^n \) with \( x \neq \bar{0} \) and let \( h_i \) be the \( i \)-th column of \( H \). Then

\[
Hx = x_1h_1 + x_2h_2 + \ldots + x_nh_n.
\]

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 \( d = \text{wt}(x) \). Since \( H \) is a check matrix for \( C \) we have \( x \in C \) if only if \( Hx = \bar{0} \). Hence there exists a codeword 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 in (a). 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

\[
(*) \quad l_1h_{k_1} + l_2h_{k_2} + \ldots + l_eh_{k_e} = \bar{0}.
\]
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

$$h_{k_1} + h_{k_2} + \ldots + h_{k_e} = 0.$$ 

Thus the minimal number of linear dependent columns is also the minimal number of columns whose sum is equal to 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 VI.31.** Let $H$ be a check matrix for the binary linear code $C$. Then $C$ is 1-error-correcting if and only if the columns of $H$ are non-zero and pairwise distinct.

**Proof.** We apply VI.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 0. Since $x + y = 0$ 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. □

**Example VI.32.** Let $C$ be the binary linear code with check matrix

$$H := \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Is $C$ 1-error correcting? Is $C$ 2-error-correcting?

The columns of $H$ are non-zero and pairwise distinct. So by VI.31 $C$ is 1-error correcting. The sums of the first, fourth, fifth and sixth column is 0. So a sum 2 $\cdot$ 2 columns of $H$ is 0 and thus by VI.30(b) $C$ is not 2-error correcting. Observe that the sums of these four columns is 0 is equivalent to the fact to 100111 is in $C$. So $C$ has minimum weight at most 4.

**Lemma VI.33.** (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

$$k \leq n - \lfloor \log_2(n + 1) \rfloor.$$ 

(b) The maximal information rate of a 1-error correcting, binary linear code of length $n$ is

$$1 - \frac{\lceil \log_2(n + 1) \rceil}{n}.$$
Proof. (a) Suppose first that there exists a 1-error-correcting code $C$ of length $n$ and dimension $k$. Then the Packing bound shows that

$$2^k (1 + n) = |C| \sum_{i=0}^{n} \binom{n}{i} \leq 2^n$$

and so $k + \log_2(n + 1) \leq n$. Thus $k \leq n - \log_2(n + 1)$. Since $k$ and $n$ are integers this gives $k \leq n - \lceil \log_2(n + 1) \rceil$

Suppose next that $k \leq n - \lceil \log_2(n + 1) \rceil$. Then $\log_2(1 + n) \leq n - k$. Put $m := n - k$. $\log_2(1 + n) \leq m$ and so $1 + n \leq 2^m$. Hence

$$2^m - 1 - m \geq (1 + n) - 1 - m = n - m = k.$$ 

Put $e_i = \text{Col}_i(I_m)$. Then

$$|\mathbb{F}_2^m \setminus \{0, e_1, \ldots, e_m\}| = 2^m - m - 1 \geq k.$$ 

So there exists $k$ pairwise distinct vectors $a_1, \ldots, a_k \in \mathbb{F}_2^m \setminus \{0, e_1, \ldots, e_m\}$. Let $A = [a_1 a_2 \ldots a_k]$ be the $m \times k$ matrix with $\text{Col}_i(A) = a_i$ and put

$$H := [A \ I_m] = [a_1 a_2 \ldots a_k e_1 e_2 \ldots e_m].$$ 

Note that $k + m = n$. Thus $H$ is a binary $m \times n$ matrix with pairwise distinct non-zero columns. Hence VI.31 shows that $\text{Nul}(H)$ is 1-error-correcting code of length $n$. As $H$ is in standard form VI.19 shows that $\dim \text{Nul}(H) = k$.

(b): By VI.7(c) we have $\rho(C) = \frac{\dim C}{n}$. So (b) follows from (a). \qed

Corollary VI.34. Let $\epsilon > 0$. Then there exists $N \in \mathbb{Z}^+$ such that for all $n \in \mathbb{Z}^+$ with $n \geq N$ there exists a binary, linear, 1-error-correcting code of length $n$ and information rate at least $1 - \epsilon$.

Proof. Observe that

$$\lim_{n \rightarrow \infty} \frac{1 - \lceil \log_2(n + 1) \rceil}{n} = 1$$

and so there exists $N \in \mathbb{N}$ such that

$$1 - \frac{\lceil \log_2(n + 1) \rceil}{n} \geq 1 - \epsilon$$

for all $n \geq N$. Let $n \in \mathbb{N}$ with $n \geq N$. By VI.33(b) there exists a binary, linear 1-error-correcting code $C$ of length $n$ with $\rho(C) = 1 - \frac{\lceil \log_2(n + 1) \rceil}{n}$. Thus $\rho(C) \geq 1 - \epsilon$ and the corollary holds. \qed

Definition VI.35. A Hamming code is a perfect, 1-error correcting, binary, linear code.
Theorem VI.36. Let $C \subseteq \mathbb{F}_2^n$ be a linear code with an $m \times n$ check matrix $H$ in standard form. Then $C$ is a Hamming code if and only if $n = 2^m - 1$ and the columns of $H$ are the non-zero vectors of $\mathbb{F}_2^m$ (in some order).

Proof. Since $C$ is binary and linear, $C$ is a Hamming code if and only if $C$ is 1-error-correcting and perfect. So by V.24

(*) $C$ is a Hamming code if and only if $C$ is 1-error correcting and $|C|(1+n) = 2^n$.

Since $H$ is in standard form, we know that $\dim C = n - m$, see VI.19(d). So $|C| = 2^{n-m}$ and

$$|C|(1+n) = 2^n \iff 2^{n-m}(1+n) = 2^n \iff 1+n = 2^m \iff n = 2^m - 1$$

By VI.31, $C$ is 1-error correcting if and only if the columns of $H$ are non-zero and pairwise distinct.

Thus

(**) $C$ is a Hamming code if and only if and only if $n = 2^m - 1$ and the columns of $H$ are non-zero and pairwise distinct.

Since $\mathbb{F}_2^m$ has exactly $2^m - 1$ non-zero vectors, the theorem follows from (**). \qed

Corollary VI.37. 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. Suppose $C$ is a Hamming code of length $n$. By VI.26 there exists a permutation $D$ of $C$ with an $m \times n$ check matrix $H$ in standard form. Then VI.36 shows that $n = 2^m - 1$ and the columns of $H$ are the non-zero vectors of $\mathbb{F}_2^m$ (in some order). This determines $H$ up to a permutation of the columns. Thus $D$ and so also $C$ is unique up to permutation.

Suppose now that $n = 2^m - 1$ for some $m \in \mathbb{N}$ and let $H$ be binary $m \times n$-matrix check matrix in standard form, whose columns are the non-zero vectors of $\mathbb{F}_2^m$ (in some order). Now VI.36 implies that $	ext{Nul}(H)$ is a Hamming code of length $n$. \qed

Example VI.38. Find a Hamming code of length 7.

We have $7 = 2^3 - 1$. So $m = 3$ and we choose the check matrix

$$H := \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1
\end{bmatrix}$$
Observe that $H$ is in standard form and thus by VI.19

$$E = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1
\end{bmatrix}$$

is a generating matrix for $C$. We obtain all codewords by computing the sums of any $i$ of the columns of $E$ for $0 \leq i \leq 4$.

\begin{align*}
i = 0: & \quad 000000 \\
i = 1: & \quad 100111 \ 0100110 \ 0010101 \ 0001011 \\
i = 2: & \quad 110001 \ 1010010 \ 1001100 \ 0110011 \ 0101101 \ 0011110 \\
i = 3: & \quad 1110100 \ 1101010 \ 1101010 \ 0111001 \\
i = 4: & \quad 1111111
\end{align*}

**Lemma VI.39.** Let $C \subseteq F_2^n$ be a binary linear code with check matrix $H$. Let $z \in F_2^n$ and $1 \leq i \leq n$. Define $a \in F_2^n$ by $a_j = z_j$ for $j \neq i$ and $a_i \neq z_i$. Then $(a, z)$ is a 1-bit error of $C$ if and only if $a \in C$ and if and only if $Hz = \text{Col}_i(H)$.

**Proof.** By definition, $(a, z)$ is a 1-bit error of $C$ if and only if $a \in C, z \in F_2^n$ and $d(a, z) = 1$. As $z \in F_2^n$ and $d(a, z) = 1$ this holds if and only if $a \in C$. Put $e_i := \text{Col}_i(I_n)$ and note that $a = z + e_i = z - e_i$. Then

$$a \in C \iff Ha = \bar{0} \quad \text{---} \quad -H \text{ is a check matrix for } C$$

$$a \in C \iff H(z - e_i) = \bar{0} \quad \text{---} \quad a = z - e_i$$

$$a \in C \iff Hz - He_i = \bar{0} \quad \text{---} \quad Hz = He_i$$

$$a \in C \iff Hz = \text{Col}_i(H)$$

\[\Box\]
Example VI.40. Let $C$ be the binary linear code with check matrix

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

(a) Is $C$ 1-error correcting?

(b) For $z = 101100$ and $z = 101110$, does there exist $a \in C$ such that $(a, z)$ is a 1-bit error? If yes, find $a$.

(a) The columns of $H$ are non-zero and pairwise distinct. So $C$ is 1-error correcting.

(b) Consider first that case $z = 101100$. We compute

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

Thus $Hz$ is not a column of $H$ and so cannot by the result of 1-bit error.

Now consider $z = 101110$. Then

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

Thus $Hz$ is the third column of $H$. Put $a := z + e_3 = 100110$. Then by VI.39 $(a, z)$ is a 1-bit error. To confirm that $a$ is codeword we compute:

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

So $Ha = 0$ and $a = 100110$ is indeed a codeword.
VI.5 Decoding linear codes

Definition VI.41. 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) $z + C := \{ z + a \mid a \in C \}$ is called the coset of $C$ containing $z$.

Remark VI.42. Let $H$ be a check matrix for the linear code $C \subseteq \mathbb{F}_2^n$ and $z \in \mathbb{F}_2^n$.

(a) $z \in C$ if and only if the syndrome of $z$ with respect to $H$ is $\tilde{0}$.

(b) $z \in z + C$.

Proof. (a) Since $H$ is a check matrix $C = \text{Nul}H$ and so $z \in C$ if and only if $Hz = \tilde{0}$.

(b) Since $C$ is a linear code, $\tilde{0} \in C$. Thus $z = z + \tilde{0} \in \{ z + c \mid c \in C \} = z + C$.  

Lemma VI.43. Let $H$ be a check matrix for the linear code $C \subseteq \mathbb{F}_2^n$. Let $y, z \in \mathbb{F}_2^n$. Then the following statements are equivalent:

(a) $y$ and $z$ have the same syndrome with respect to $H$.

(b) $y + z \in C$.

(c) $z = y + a$ for some codeword $a \in C$.

(d) $z \in y + C$.

(e) $(y + C) \cap (z + C) \neq \varnothing$.

(f) $y + C = z + C$.

Proof. We have

\[ Hy = Hz \]
\[ \iff \]
\[ Hz + Hy = \tilde{0} \]  \hfill (a)
\[ \iff \]
\[ H(z + y) = \tilde{0} \]
\[ \iff \]
\[ z + y \in C \]  \hfill (b)
\[ \iff \]
\[ z + y = a \text{ for some } a \in C \]  \hfill (c)
\[ \iff \]
\[ z = y + a \text{ for some } a \in C \]  \hfill (d)
\[ \iff \]
\[ z \in y + C \]  \hfill (d)
So the first four statements are equivalent.

\[ (d) \implies (e): \] Note that \( z = z + \tilde{0} \in z + C \). So if \( z \in y + C \), then \( z \in (z + C) \cap (y + C) \) and \( (z + C) \cap (y + C) \neq \emptyset \) and (e) holds.

\[ (e) \implies (f): \] Let \( u \in (y + C) \cap (z + C) \). Then \( u \in y + C \). Since (d) implies (a), we conclude that \( H u = H y \). Similarly as \( u \in z + C \) we have \( H u = H z \). So

\[ H y = H z. \]

Let \( v \in \mathbb{F}_2^n \). Since (a) and (d) are equivalent, we get

\[ v \in y + C \iff H v = H y \iff H v = H z \iff v \in z + C. \]

So \( y + C = z + C \).

\[ (f) \implies (d): \] If \( y + C = z + C \), then \( z = z + \tilde{0} \in z + C = y + C \).

**Corollary VI.44.** 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.** Let \( y \in \mathbb{F}_2^n \). Then by VI.43 \( z \in y + C \) if and only if \( y + C = z + C \). So indeed \( z + C \) is the unique coset of \( C \) containing \( z \).

**Example VI.45.** Let \( C \) be the binary linear 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 \( C \), see VI.19. Computing the sum of each set of columns of \( E \) we obtain

\[ C = \{0000, 1010, 0101, 1111\} \].
We compute

<table>
<thead>
<tr>
<th></th>
<th>0000 + C</th>
<th>1000 + C</th>
<th>0100 + C</th>
<th>1100 + C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>1000</td>
<td>0100</td>
<td>1100</td>
<td></td>
</tr>
<tr>
<td>1010</td>
<td>0010</td>
<td>1110</td>
<td>0110</td>
<td></td>
</tr>
<tr>
<td>0101</td>
<td>1101</td>
<td>0001</td>
<td>1001</td>
<td></td>
</tr>
<tr>
<td>1111</td>
<td>0111</td>
<td>1011</td>
<td>0011</td>
<td></td>
</tr>
<tr>
<td>00</td>
<td>10</td>
<td>01</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

Here in last row we list the common syndrome of the elements in the coset.

Which of the cosets contains 1101?

\[
\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
\]

So 1101 lies in the coset with syndrome 10, that is 1000 + C.

**Lemma VI.46.** Let \( C \) be a binary linear code of length \( n \) and dimension \( k \). Let \( H \) be an \( m \times n \)-check matrix for \( C \).

(a) The set of syndromes of \( H \) is \( \text{Col}(H) \).

(b) The function

\[
\alpha_H : \{ z + C \mid z \in \mathbb{F}_2^n \} \to \text{Col}(H), \quad z + C \mapsto Hz
\]

is well defined bijection.

(c) The numbers of syndromes for \( C \) with respect to \( H \) is \( 2^{n-k} \). In particular, \( |\text{Col}(H)| = 2^{n-k} \).

(d) \( n - k \leq m \).

(e) If \( m = n - k \) (for example if \( H \) is in standard form), every element of \( \mathbb{F}_2^n \) 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 \( s \in \text{Col}H \).

(b) Let \( y, z \in \mathbb{F}_2^n \). By VI.43.
VI.5. DECODING LINEAR CODES

\[ y + C = z + C \iff H y = H z \]

The forward direction shows that the function \( \alpha_H \) is well-defined. The backward direction shows that \( \alpha_H \) is 1-1. By definition, \( \text{Col}(H) = \left\{ H z \mid z \in \mathbb{F}_2^n \right\} = \text{Im} \alpha_H \) and so \( \alpha_H \) is onto.

(c) Let \( u \) be the numbers of cosets of \( C \) in \( \mathbb{F}_2^n \). Note that for \( z \in \mathbb{F}_2^n \) the function \( C \to z + C, a \to z + a \) is a bijections. Thus \( |z + C| = |C| = 2^k \). Thus each cosets of \( C \) contains \( 2^k \) elements. Since there are \( u \) coset of \( C \) and each of the \( 2^n \) elements of \( \mathbb{F}_2^n \) lies in a unique coset of \( C \) elements we conclude that

\[ u \cdot 2^k = 2^n. \]

Thus \( u = 2^{n-k} \). By (b) \( u = |\text{Col}(H)| \) and by (a) \( \text{Col}(H) \) is the set of syndromes of \( H \). Thus the number of syndrome of \( H \) is \( 2^{n-k} \) and (c) is proved.

(d) Note that \( \text{Col}(H) \subseteq \mathbb{F}_2^n \) and so

\[ 2^{n-k} \leq |\text{Col}(H)| \leq |\mathbb{F}_2^n| = 2^m. \]

Thus \( n - k \leq m \).

(e) If \( m = n - k \), then (d) shows that \( C \) has \( 2^m \) syndromes with respect to \( H \). As \( |\mathbb{F}_2^m| = 2^m \) this implies that every element of \( \mathbb{F}_2^m \) is a syndrome of \( C \) with respect to \( H \).

Remark VI.47. Let \( H \) be binary \( m \times n \)-matrix. Then \( \dim \text{Nul} H + \dim \text{Col} H = n. \)

Proof. Put \( k := \dim \text{Nul}(H) \). By VI.46 \( |\text{Col}(H)| = 2^{n-k} \) Thus

\[ \dim \text{Col}(H) = \log_2 |\text{Col}(H)| = n - k = n - \dim \text{Nul} H. \]

Definition VI.48. Let \( C \) be binary linear code with an \( m \times n \) check matrix \( H \). A syndrome look-up table for \( H \) is a function

\[ \tau : \text{Col}(H) \to \mathbb{F}_2^n \]

such that for each syndrome \( s \) of \( H \):

(i) \( \tau(s) \) has syndrome \( s \) with respect to \( H \), i.e. \( H \tau(s) = s \), and

(ii) \( \tau(s) \) is a vector of minimal weight in \( \tau(s) + C \), i.e. \( \text{wt} \left( \tau(s) \right) \leq \text{wt}(z) \) for all \( z \in \tau(s) + C \).

Remark VI.49. Let \( C \) be linear code of length \( n \) with check matrix \( H \) and \( \tau \) a syndrome look-up table for \( H \). Then
CHAPTER VI. LINEAR CODES

(a) Let $s$ be a syndrome of $H$. Then $\tau(s) + C$ is the set of vectors in $\mathbb{F}_2^n$ which have syndrome $s$ with respect to $H$.

(b) $z + \tau(Hz) \in C$ for all $z \in \mathbb{F}_2^n$.

Proof. (a) Let $z \in \mathbb{F}_2^n$. By definition of a syndrome look-up-table, $\tau(s)$ has syndrome $s$. Thus $z$ has syndrome $s$ if and only if $z$ and $\tau(s)$ have the same syndrome, and so by VI.43 if and only if $z \in \tau(s) + C$.

(b) Note that $z$ and $\tau(Hz)$ both have syndrome $Hz$. Thus by VI.43 $z + \tau(Hz) \in C$.

**Definition VI.50.** Let $C$ be binary linear code of length $n$ with check matrix $H$ and $\tau$ a syndrome look-up table for $H$. Then the function

$$\sigma : \mathbb{F}_2^n \to C, \quad z \mapsto z + \tau(Hz)$$

is called the decision rule for $C$ with respect to $H$ and $\tau$.

Note that by the preceding remark $z + \tau(Hz) \in C$, so $\sigma$ is well-defined.

**Lemma VI.51.** Let $C$ be linear code of length $n$ with check matrix $H$, $\tau$ a syndrome look-up table for $H$ and $\sigma$ the corresponding decision rule. Then $\sigma$ is a Minimum Distance decision rule.

Proof. Let $z \in \mathbb{F}_2^n$ and $a \in C$. Put $s := Hz$. By definition of syndrome look-up table $\tau(s)$ has syndrome $s$. Since also $z$ has syndrome $s$, we conclude from VI.43 that $z + C = \tau(s) + C$. Thus $z + a \in \tau(s) + C$. By definition of syndrome look-up table, $\tau(s)$ is of minimal weight in $\tau(s) + C$. So

$$(*) \quad \text{wt}(z + a) \leq \text{wt}(\tau(s)).$$

We compute

$$d(\sigma(z), z) = d(z + \tau(Hz), z) \overset{VI.10}{=} d(z + \tau(s), z) = \text{wt}(z + \tau(s) + z) \overset{(*)}{\geq} \text{wt}(z + a) \overset{VI.10}{=} d(a, z).$$

Thus $\sigma$ is Minimum-Distance decision rule.

**Algorithm VI.52.** Let $C$ be binary 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 $\text{wt}(z_i) \leq \text{wt}(z_j)$ for all $1 \leq i < j \leq 2^n$. Define sets $S_l$ and functions $\tau_l : S_l \to \mathbb{F}_2^n$ inductively as follows:

For $l = 0$ let $S_0 = \emptyset$ and $\tau_0 = \emptyset$.

Suppose $l > 0$ and that $S_{l-1}$ and $\tau_{l-1}$ have been defined. Compute $s_l := Hz_l$. 

VI.5. DECODING LINEAR CODES

- If \( s_l \in S_{l-1} \), put \( S_l := S_{l-1} \) and \( \tau_l := \tau_{l-1} \).

- If \( s_l \notin S_l \), put \( S_l := S_{l-1} \cup \{ s_l \} \) and extend \( \tau_{l-1} \) to a function \( \tau_l \) on \( S_l \) by \( \tau_l(s_l) := z_l \).

Put \( d := \dim \text{Col}(H) \). Stop the algorithm when \( |S_l| = 2^d \), that is when \( S_l = \text{Col}(H) \). Then the last \( \tau_l \) is a syndrome look-up table for \( H \).

**Example VI.53.** Let \( C \) be the binary linear code with check matrix

\[
H := \begin{bmatrix}
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\]

Construct a syndrome look-up table \( \tau \) for \( H \) and compute \( \sigma(11001) \), where \( \sigma \) is the decision rule for \( C \) with respect to \( H \) and \( \tau \).

| \( z_l \) | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
|\hline
| \( s_l \) | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
|\hline
| \( |S_l| \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | = | = | 8 |

\[
\begin{pmatrix}
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 \\
0 \\
1 \\
0
\end{pmatrix} = \begin{pmatrix}
1 \\
0 \\
1 \\
0
\end{pmatrix}.
\]

\( \tau(100) = 00100 \)

and so

\[
\begin{align*}
\sigma(11001) &= 11001 \\
&+ \ 00100 \\
&= \ 11101
\end{align*}
\]
To double check

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
1 \\
0 \\
1
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}.
\]
Chapter VII

Algebraic Coding Theory

VII.1 Classification and properties of cyclic codes

In this chapter we will denote a string $a \in \mathbb{F}_2^n$ by $a_0 \ldots a_{n-1}$ rather than $a_1 \ldots a_n$, since we will consider the associated polynomial $a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}$.

**Definition VII.1.** A binary code $C \subseteq \mathbb{F}_2^n$ is called cyclic if

(i) $C$ is linear, and

(ii) $a_{n-1}a_0 \ldots a_{n-2} \in C$ for all $a = a_0 \ldots a_{n-1} \in C$.

**Example VII.2.** Which of following codes are cyclic?

(1) \{000, 100, 111, 011\}.

(2) \{000, 100, 010, 001\}

(3) \{000, 1010, 0101, 1111\}

(1): Not cyclic, since 100 is in the code, but 010 is not.

(2): Not cyclic, since its not linear: 100 and 010 are in the code, but $100 + 010 = 110$ is not.

(3): Is cyclic.

**Definition VII.3.** An ideal in a ring $R$ is a subset $S$ of $R$ such that

(i) $0 \in S$.

(ii) If $s, t \in S$ then $s + t \in S$.

(iii) If $a \in R$ and $s \in S$, then $as \in S$ and $sa \in S$. 
Example VII.4. The set of even integers is an ideal in the ring of integers.

Indeed, 0 is even. Sums of even integers are even and any multiple of an even integer is even.

Definition VII.5. Let $R$ be a ring with identity. $R[x]$ is the ring defined as follows:

(i) The elements of $R[x]$ are the infinite sequences $f = (f_i)_{i=0}^\infty$ with coefficients in $R$ such that there exists $n \in \mathbb{N}$ with $f_i = 0$ for all $i \in \mathbb{N}$ with $i > 0$.

(ii) $f + g := (f_i + g_i)_{i=0}^\infty$ for all $f, g \in R[x]$.

(iii) $fg = \left(\sum_{k=0}^i f_k g_{i-k}\right)_{i=0}^\infty$ for all $f, g \in R[x]$.

$R[x]$ is called the polynomial ring over $R$. The elements of $R[x]$ are called polynomial. $x$ denotes the polynomial $x := (0, 1, 0, 0, \ldots)$.

$\deg f := \max\{i \in \mathbb{N} \mid f_i \neq 0\}$ with $\deg f = -\infty$ if $f = 0$.

Remark VII.6. Let $\mathbb{F}$ be a field and $f, g \in \mathbb{F}[x]$.

(a) $f = \sum_{i=0}^{\deg f} f_i x^i$.

(b) $\deg(f + g) \leq \max(\deg f, \deg g)$.

(c) If $\deg g < \deg f$, then $\deg(f + g) = \deg f$.

(d) $\deg(fg) = \deg f + \deg g$. Here we define $a + (-\infty) = (\infty) + q = \infty$ for all $a \in \mathbb{N} \cup \{-\infty\}$.

(e) $\deg(a f) = \deg f$ for all $a \in \mathbb{F}$ with $a \neq 0$.

(f) $\deg(-f) = \deg f$.

Lemma VII.7. Let $\mathbb{F}$ be a field and $f, g \in \mathbb{F}[x]$ with $g \neq 0$. Then there exist uniquely determined $q, r \in \mathbb{F}[x]$ with

$$f = qh + r$$

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

\[ \tilde{f} = \tilde{q}h + \tilde{r} \quad \text{and} \quad \deg \tilde{r} < \deg h \]

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 that

\[ f = qh + r = q^*h + r^*, \quad \deg r < \deg h \quad \text{and} \quad \deg r^* < \deg h \]

for some \(q, q^*, r, r^* \in \mathbb{F}[x]\). Then

\[ (q - q^*)h = r^* - r. \]

By A.8(a), \(\deg(r^* - r) \leq \max(\deg r, \deg r^*) < \deg h\) and so \(\deg(q - q^*)h < \deg h\). If \(q - q^* \neq 0\), then A.8(c) implies that \(\deg(q - q^*)h \geq \deg h\), a contradiction. Thus \(q - q^* = 0\) and so also \(r^* - r = (q - q^*)h = 0h = 0\).

Hence \(q = q^*\) and \(r = r^*\). So \(q\) and \(r\) are unique. \(\square\)

**Example VII.8.** Consider the polynomials \(f = x^4 + x\) and \(g = x^2 + 1\) in \(\mathbb{F}_2[x]\). Find \(q, r \in \mathbb{F}_2[x]\) with \(f = qg + r\) and \(\deg r > \deg g\).

\[
\begin{array}{c|cc}
  & x^2 & + 1 \\
\hline
x^2 + 1 & x^4 & + x \\
  & x^4 & + x^2 \\
  & x^2 & + x \\
  & x^2 & + 1 \\
  & x & + 1 \\
\end{array}
\]

Thus

\[ x^4 + x = (x^2 + 1) \cdot (x^2 + 1) + (x + 1). \]
Definition VII.9. 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

$$f^{\odot 0} = \overline{1}$$ and $$f^{\odot (n+1)} = f^{\odot n} \odot f.$$  

Define

$$\mathbb{F}^h[x] = \{ f \in \mathbb{F}[x] \mid \deg f < \deg h \}$$

$$(\mathbb{F}^h[x], \oplus, \odot)$$ is called the ring of polynomials modulo $h$ with coefficients in $\mathbb{F}$.

**Example VII.10.** Determine the addition and multiplication in $\mathbb{R}^{x^2 + 1}[x]$.

Since $\deg x^2 + 1 = 2$ we have $\mathbb{R}^{x^2 + 1}[x] = \{ a + bx \mid a, b \in \mathbb{R} \}$. We compute

$$\begin{align*}
(a + bx) \oplus (c + dx) &= \overline{(a + bx) + (c + dx)} \\
&= \overline{(a + c) + (b + d)x} \\
&= (a + c) + (b + d)x
\end{align*}$$

and

$$\begin{align*}
(a + bx) \odot (c + dx) &= \overline{(a + bx) \cdot (c + dx)} \\
&= \overline{ac + adx + bcx + bdx^2} \\
&= \overline{(ac - bd) + (ad + bc)x + bd(x^2 + 1)} \\
&= (ac - bd) + (ad + bc)x.
\end{align*}$$

Note that this is the same addition and multiplication as for the ring of complex numbers $\mathbb{C} = \{ a + bi \mid a, b \in \mathbb{R} \}$:

$$\begin{align*}
(a + bi) + (c + di) &= (a + c) + (b + d)i \\
\text{and} \\
(a + bi)(c + di) &= (ac - bd) + (ad + bc)i.
\end{align*}$$
Lemma VII.11. Let $F$ be a field and $h \in F[x]$ with $h \neq 0$. Let $f, g \in F[x]$.

(a) $\overline{f} \in \overline{Fh}[x]$.

(b) $f = \overline{f}$ if and only if $\deg f < \deg h$ and if and only if $f \in \overline{Fh}[x]$.

(c) $\overline{Fh}[x] = \{\overline{f} \mid f \in F[x]\}$.

(d) Suppose $f, g \in \overline{Fh}[x]$. Then $f \oplus g = f + g$

(e) Suppose $f \in \overline{Fh}[x]$. Then $a \odot f = af$.

Proof. (a) By definition, $\deg f > \deg h$. So $\overline{f} \in \overline{Fh}[x]$.

(b) Suppose $f = \overline{f}$. By (a), $\overline{f} \in \overline{Fh}[x]$, so $f \in \overline{Fh}[x]$.

(c) If $f \in \overline{Fh}[x]$, then (b) gives $f = \overline{f} \in \{\overline{f} \mid f \in F[x]\}$. Also by (b), $\overline{f} \in \overline{Fh}[x]$ for all $f \in F[x]$.

(d) Observe that $\deg(f + g) \leq \max(\deg f, \deg g) < \deg h$ and so by (b) $\overline{f + g} = f + g$.

(e) Since $a \in F$, $\deg a \leq 0$. So $\deg af = \deg a + \deg f \leq \deg f < \deg h$. Thus (b) gives $a \overline{f} = \overline{af}$.

Remark VII.12. Let $F$ be a field and $h \in F[x]$ with $h \neq 0$. Then $\overline{Fh}[x]$ is vector space over $F$.

Definition VII.13. 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 VII.14. Let $F$ be a field and $h \in F[x]$ with $h \neq 0$. Let $f, g \in F[x], a \in F$ and $n \in \mathbb{N}$. Then

(a) $0 \oplus f = 1 \odot f = \overline{f} = f \odot 1 = f \oplus 0$.

(b) $f \oplus g = \overline{f + g} = \overline{f} \oplus \overline{g} = \overline{f} \oplus g = f \oplus g$.

(c) $f \odot g = \overline{fg} = \overline{f} \odot \overline{g} = \overline{f} \odot g = f \odot g$.

(d) $\overline{f}^n = f^\odot n = (\overline{f})^\odot n = (\overline{f})^n$.

(e) $-\overline{f} = -\overline{f}$.

(f) $\overline{f} = 0$ if and only if $h$ divides $f$.

(g) $\overline{f} = \overline{g}$ if and only if $h$ divides $g - f$. 
(h) \( a \odot f = af = a f \).

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) \( 1 \odot f = \overline{1} \odot f = 0 \odot f = f = f \odot 0 = f \oplus 0 \). Since \( f \in \mathbb{F}[h] \), VII.11(b) gives \( f = f \).

(b) We compute

\[
\begin{align*}
\overline{f} + \overline{g} &= (ph + \overline{f}) + (qh + \overline{g}) = (p + q)h + (\overline{f} + \overline{g}).
\end{align*}
\]

Note that

\[
\deg (\overline{f} + \overline{g}) \leq \max(\deg \overline{f}, \deg \overline{g}) < \deg h
\]

and so \( \overline{f} + \overline{g} \) is the remainder of \( f + g \) when divided by \( h \). Hence \( f \oplus g = \overline{f} + \overline{g} = \overline{f + g} \).

By VII.11(d), \( f \oplus g = f \ominus g \). The remaining parts of (b) now follow from \( \overline{f} = f \) and \( \overline{g} = g \).

(c) By definition of \( f \odot g \) we know that

\[
\begin{align*}
f \odot g &= \overline{fg}.
\end{align*}
\]

By definition of \( \overline{fg} \) we have \( \overline{fg} = th + \overline{fg} \) for some \( t \in \mathbb{F}[x] \) and \( \deg \overline{fg} < \deg h \). We compute

\[
\begin{align*}
fg &= (ph + \overline{f})(qh + \overline{g}) = (phq + \overline{f}q + p\overline{g})h + \overline{fg} = (phq + \overline{f}q + p\overline{g} + t)h + \overline{fg}
\end{align*}
\]

So \( \overline{fg} \) is the remainder of \( fg \) when divided by \( h \), that is

\[
\begin{align*}
\overline{fg} &= \overline{fg}.
\end{align*}
\]

By definition of \( \overline{f} \odot g \) we get

\[
\begin{align*}
\overline{f} \odot g &= \overline{fg}.
\end{align*}
\]

By definition of \( \overline{f} \odot g \) we have \( \overline{f} \odot g = \overline{fg} \). By (*) applied with \( \overline{f} \) in place of \( f \), we know that \( \overline{fg} = \overline{fg} = \overline{fg} \) and so

\[
\begin{align*}
\overline{f} \odot g &= \overline{fg}.
\end{align*}
\]

Similarly \( f \odot g = \overline{fg} \) and (c) is proved.

(d) For \( n = 0 \) all four expression are equal to \( 1 \).

Suppose (d) holds for \( n \). Then

\[
\begin{align*}
\overline{f}^{(n+1)} &\overset{\text{Def}}{=} \overline{f}^n \odot f \overset{\text{Ind}}{=} f \odot f \overset{\text{Def}}{=} f^{(n+1)},
\end{align*}
\]

\[
\begin{align*}
\overline{f}^{(n+1)} &\overset{\text{Def}}{=} \overline{f}^n \odot f \overset{\text{Ind}}{=} f \odot f \overset{\text{Def}}{=} f^{(n+1)},
\end{align*}
\]

\[
\begin{align*}
\overline{f}^{(n+1)} &\overset{\text{Def}}{=} \overline{f}^n \odot f \overset{\text{Ind}}{=} f \odot f \overset{\text{Def}}{=} f^{(n+1)},
\end{align*}
\]

\[
\begin{align*}
\overline{f}^{(n+1)} &\overset{\text{Def}}{=} \overline{f}^n \odot f \overset{\text{Ind}}{=} f \odot f \overset{\text{Def}}{=} f^{(n+1)},
\end{align*}
\]

\[
\begin{align*}
\overline{f}^{(n+1)} &\overset{\text{Def}}{=} \overline{f}^n \odot f \overset{\text{Ind}}{=} f \odot f \overset{\text{Def}}{=} f^{(n+1)},
\end{align*}
\]

\[
\begin{align*}
\overline{f}^{(n+1)} &\overset{\text{Def}}{=} \overline{f}^n \odot f \overset{\text{Ind}}{=} f \odot f \overset{\text{Def}}{=} f^{(n+1)},
\end{align*}
\]
\[ (\overline{f})^{(n+1)} \overset{\text{Def}}{=} (\overline{f})^{\oplus n} \otimes \overline{f} \overset{\text{Ind.}}{=} \overline{f^n} \otimes f \overset{(**)}{=} f^{\oplus (n+1)} \]

and

\[ (\overline{f})^{(n+1)} \overset{\text{Def}}{=} (\overline{f})^{\oplus n} \overset{\text{Ind.}}{=} \overline{f^n} \otimes f \overset{\text{Def}}{=} f^{\oplus (n+1)} \]

\[ -f = (-p)h + (-\overline{f}) \text{ and } \deg(-f) = \deg f < \deg h, \text{ so } -\overline{f} \text{ is the remainder of } f \text{ when divided by } h. \]

\[ h \text{ divides } f \text{ if and only if } f = sh \text{ for some } s \in F[x] \text{ and so if and only if } f = sh + 0 \text{ for some } s \in F[x]. \] Since \( \deg 0 < \deg h \), this holds if and only if 0 is the remainder of \( f \) when divided by 0.

\[ \overline{f} = \overline{g} \text{ if and only if } \overline{f} - \overline{g} = 0 \text{ and if and only if } \overline{f} - \overline{g} = 0. \] By (c) the latter holds if and only if \( h \) divides \( f - g \).

\[ a \otimes f = a\overline{f} = a \otimes \overline{f} \text{ and by VII.11(c) } a \otimes \overline{f} = a\overline{f}. \]

**Lemma VII.15.** Let \( F \) be a field, \( 0 \neq h \in F[x] \) and \( I \subseteq F^h[x] \). Then \( I \) is an ideal in \( F^h[x] \) if and only if \( I \) is an \( F \)-subspace of \( F^h[x] \) and \( x \otimes f \in I \) for all \( f \in I \).

**Proof.** Let \( f, g \in I \) and \( a \in F \).

\( \implies \): Suppose that \( I \) is an ideal in \( F^h[x] \). Then by definition of an ideal, \( 0 \in I \) and \( f + g = f \oplus g \in I \). Since \( f \in I \) and \( I \) is an ideal we have \( af = a \otimes f \in I \). Thus \( I \) is an \( F \)-subspace of \( F^h[x] \). Also \( x \otimes f = x \otimes f \in I \) and so the forward direction is proved.

\( \impliedby \): Suppose \( I \) is an \( F \)-subspace of \( F^h[x] \) and \( x \otimes f \in I \) for all \( f \in I \). Then by definition of a subspace \( 0 \in I \) and \( f \oplus g = f + g \in I \).

We claim that \( x^i \otimes f \in I \) for all \( i \in \mathbb{N} \). We have \( x^0 \otimes f = 1 \otimes f = f \) and so the claim holds for \( i = 0 \). Suppose inductively that \( x^i \otimes f \in I \). Then by assumption also \( x \otimes (x^i \otimes f) \in I \) and hence

\[ x^{i+1} \otimes f = x^{i+1}f = x(x^i f) = x \otimes x^i f = x \otimes (x^i \otimes f) \in I. \]

So indeed \( x^i \otimes f \in I \) for all \( i \in \mathbb{N} \) and \( f \in I \).

Let \( g \in F^h[x] \). Then \( g = \sum_{i=0}^{n-1} a_i x^i \) for some \( a_i \in F \) and so

\[ g \otimes f = \left( \sum_{i=0}^{n-1} a_i x^i \right) \otimes f = \sum_{i=0}^{n-1} \left( (a_i x^i) \otimes f \right) = \sum_{i=0}^{n-1} a_i \left( x^i \otimes f \right). \]

Since each \( x^i \otimes f \in I \) and \( I \) is an \( F \)-subspace of \( F^h[x] \) we conclude that \( g \otimes f \in I \). Thus \( I \) is an ideal in \( F^h[x] \). \( \Box \)

**Definition VII.16.** Let \( n \in \mathbb{Z}^+ \). Then \( V^n[x] := F_2^{x^{n-1}}[x] \).
Lemma VII.17. Let $\mathbb{F}$ be a field, $n \in \mathbb{Z}^+$ and for $f \in \mathbb{F}[x]$ let $\overline{f}$ be remainder of $f$ when divided by $x^n - 1$.

(a) Let $i, j \in \mathbb{N}$ with $0 \leq j < n$. Then $\overline{x^{ni+j}} = x^j$.

(b) Let $f \in \mathbb{F}[x]$ with

$$f = \sum_{i=0}^{r} \sum_{j=0}^{n-1} a_{ij}x^{ni+j}$$

for some $a_{ij} \in \mathbb{F}$. Then

$$\overline{f} = \sum_{j=0}^{n-1} \left( \sum_{i=0}^{r} a_{ij} \right) x^j$$

Proof. (a): Since $x^n = (x^n - 1) + 1$ and $\deg 1 = 0 < n = \deg (x^n - 1)$ we have $\overline{x^n} = 1$. Thus

$$\overline{x^{ni+j}} = \overline{(x^n)^i} \overline{x^j} = \overline{x^i} \overline{x^j} = \overline{x^j} = x^j.$$

So (a) holds.

(b): We compute

$$\sum_{i=0}^{r} \sum_{j=0}^{n-1} a_{ij}x^{ni+j} = \sum_{i=0}^{r} \sum_{j=0}^{n-1} a_{ij} \overline{x^{ni+j}} = \sum_{j=0}^{n-1} \left( \sum_{i=0}^{r} a_{ij} \right) x^j$$

and so (b) holds. \qed

Example VII.18. Compute the remainder of $f = 1 + x + x^4 + x^5 + x^7 + x^{11} + x^{17} + x^{28}$ when divided by $x^6 + 1$ in $\mathbb{F}_2[x]$.

$$f = 1 + x + x^4 + x^5 + x^{6+1} + x^{6+5} + x^{12+5} + x^{24+4}$$

and so

$$\overline{f} = 1 + x + x^4 + x^5 + x + x^5 + x^5 + x^4 = 1 + 2x + 2x^4 + 3x^5 = 1 + x^5.$$

Definition VII.19. Let $a = a_0 \ldots a_{n-1} \in \mathbb{F}_2^n$ and $C \subseteq \mathbb{F}_2^n$.

(a) For $0 \leq i < n$ define

$$a^{(i)} = a_{n-i} \ldots a_{n-1}a_0a_1 \ldots a_{n-1-i},$$

where $a_0, a_1, \ldots, a_{n-1}$ are the coefficients of $a$.
VII.1. CLASSIFICATION AND PROPERTIES OF CYCLIC CODES

that is

\[ a = a^{(0)} = a_0 \ a_1 \ a_2 \ldots a_{n-2} \ a_{n-1} \]
\[ a^{(1)} = a_{n-1} \ a_0 \ a_1 \ldots a_{n-3} \ a_{n-2} \]
\[ a^{(2)} = a_{n-2} \ a_{n-1} \ a_0 \ldots a_{n-4} \ a_{n-3} \]
\[ \vdots \]
\[ a^{(n-1)} = a_1 \ a_2 \ a_3 \ldots a_{n-1} \ a_0 \]

\( a^{(i)} \) is called the cyclic \( i \)-shift of \( a \).

(b) \( \langle a \rangle \) is the subspace of \( \mathbb{F}_2^n \) spanned by \( a^{(0)}, a^{(1)}, \ldots, a^{(n-1)} \). \( \langle a \rangle \) is called the binary cyclic code generated by \( a \).

(c) \( a(x) := \sum_{i=0}^{n-1} a_i x^i = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} \in V^n[x] \).

(d) \( C(x) := \{ a(x) \mid a \in C \} \).

Example VII.20. Compute the binary cyclic code generated by 1100.

The cyclic shifts of 1100 are

1100, 0110, 0011, 1001

The compute the subspace of \( \mathbb{F}_2^n \) spanned by these four codewords by computing the sum of any subset:

\{0000, 1100, 0110, 0011, 1001, 1010, 1111, 0101\}

Lemma VII.21. Let \( a \in \mathbb{F}_2^n \) and \( C \subseteq \mathbb{F}_2^n \).

(a) The function \( \delta : \mathbb{F}_2^n \to V^n[x], x \mapsto a(x) \) is an isomorphism of \( \mathbb{F}_2 \)-vector spaces.

(b) \( a \in C \) if and only if \( a(x) \in C(x) \).

(c) Let \( a \in \mathbb{F}_2^n \) and \( 0 \leq i < n \). Then \( a^{(i)}(x) = x^i \odot a(x) \).

(d) Let \( C \subseteq \mathbb{F}_2^n \). Then \( C \) is a cyclic code if and only if \( C(x) \) is an ideal in \( V^n[x] \).

Proof. (a): Any \( a \in \mathbb{F}_2^n \) can be uniquely written as \( a_0 \ldots a_{n-1} \) with \( a_i \in \mathbb{F}_2 \). Any \( f \in V^n[x] \) can be uniquely written as \( a_0 + a_1 x + \ldots a_{n-1} x^{n-1} \). So \( \delta \) is a bijections. Let \( a, b \in \mathbb{F}_2^n \). Then

\[ \delta(a + b) = (a + b)(x) = (a_0 + b_0) + (a_1 + b_1)x + \ldots (a_{n-1} + b_{n-1}) x^{n-1} = a(x) + b(x) = \delta(a) + \delta(b). \]

So \( \delta \) is also \( \mathbb{F}_2 \)-linear. Thus \( \delta \) is an isomorphism.
VII.15

(a) By (a) we know that \( \delta \) is a bijection. So (b) holds.

(c): Let \( a = a_0 \ldots a_{n-1} \) with \( a_j \in \mathbb{F}_2 \). Then

\[
x^i \odot a(x) = x^i(a_0 + a_1 x + \ldots + a_{n-1} x^{n-1})
\]

\[
= a_0 x^i + a_1 x^{i+1} + \ldots + a_{n-1-i} x^{n-1} + a_{n-i} x^n + \ldots + a_{n-1-x^{n-i}}
\]

\[
= a_0 x^i + a_1 x^{i+1} + \ldots + a^n - i x^{n-1} + a_{n-i} + \ldots + a_{n-1} x^{i-1}
\]

\[
= a_{n-i} + \ldots + a_{n-1} x^{i-1} + a_0 x^i + \ldots + a_{n-1} x^{n-1}
\]

\[
= a^{(i)}(x)
\]

(d): Note that \( 0 \in C \) if and only if \( 0 = \tilde{a}(x) \in C(x) \). Let \( a, b \in C \). Then \( a + b \in C \) if and only if \( (a + b)(x) \in C(x) \) and so if and only if \( a(x) + b(x) \in C(x) \). Thus

(*) \( C \) is a subspace of \( \mathbb{F}_2^n \) if and only if \( C(x) \) is a subspace of \( V^n[x] \).

By (c) \( a^{(1)}(x) = x \odot a(x) \) and so (b) implies:

(**) \( a^{(1)} \in C \) if and only if \( x \odot a(x) \in C(x) \).

By definition, \( C \) is cyclic if and only if \( C \) is a subspace of \( \mathbb{F}_2^n \) and \( a^{(1)} \in C \) for all \( a \in C \). By VII.15 \( C(x) \) is an ideal in \( V^n[x] \) if and only if \( C(x) \) is a subspace of \( V^n(x) \) and \( x \odot a(x) \in C(x) \) for all \( a \in C \). Together with (*) and (**) this proves the lemma.

Definition VII.22. Let \( R \) be a commutative ring with identity and \( a \in R \). Define

\[
\{a\} := \{ra \mid r \in R\}.
\]

\( \{a\} \) is called the ideal in \( R \) generated by \( a \).

Lemma VII.23. Let \( R \) be a commutative ring with identity and \( a \in R \). Then \( \{a\} \) is the smallest ideal of \( R \) containing \( a \), that is

(a) \( a \in \{a\} \),

(b) \( \{a\} \) is an ideal in \( R \), and

(c) \( \{a\} \subseteq I \) for any ideal \( I \) of \( R \) with \( a \in I \).

Proof. \( (a) \) \( a = 1a \in \{a\} \).

(b) We have \( 0 = 0a \in \{a\} \). Also for any \( r, s \in R \), \( ra + sa = (r + s)a \in \{a\} \) and \( (ra)s = s(ra) = (sr)a \in \{a\} \).

(c) By definition of an ideal, \( ra \in I \) for all \( r \in R \) and so \( \{a\} \subseteq I \).
Lemma VII.24. Let \( \mathbb{F} \) be a field, \( h \in \mathbb{F}[x] \) with \( h \neq 0 \) and \( f \in \mathbb{F}^h[x] \). Let \( (f) \) be the ideal in \( \mathbb{F}^h[x] \) generated by \( f \). Then

(a) \( (f) \) 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) \( (f) = \{ g \odot f \mid g \in \mathbb{F}[x] \} \).

Proof. (a) Let \( g = \sum_{i=0}^{n-1} a_i x^i \in \mathbb{F}[x] \). 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 \( (f) \) are exactly the \( \mathbb{F} \)-linear combinations of \( 1 \odot f, x \odot f, \ldots, x^{n-1} \odot f \).

(b) \( (f) = \{ g \odot f \mid g \in \mathbb{F}^h[x] \} = \{ g \odot f \mid g \in \mathbb{F}[x] \} = \{ g \odot f \mid g \in \mathbb{F}[x] \} \).

Example VII.25. Find the ideal in \( \mathbb{V}^3[x] \) generated by \( f = 1 + x^2 \) and determined the corresponding cyclic code.

\[
1 \odot (1 + x^2) = 1 + x^2
\]
\[
x \odot (1 + x^2) = x + x^3 = 1 + x
\]
\[
x^2 \odot (1 + x^2) = x^2 + x^4 = x + x^2
\]

Note that \( (1 + x^2) + (1 + x) = x + x^2 \). So

\[
(1 + x^2) = \{0, 1 + x^2, 1 + x, x + x^2\}.
\]

The corresponding cyclic code is

\[
\{000, 101, 110, 011\}
\]

Lemma VII.26. Let \( \mathbb{F} \) be a field and \( f, g, t \in \mathbb{F}[x] \) with \( t \neq 0 \).

(a) \( ft = gt \) if and only if \( f = g \).

(b) \( f \mid g \) if and only if \( ft \mid gt \).
Proof. (a): $\Rightarrow$: Suppose $ft = gt$. Then $(f - g)t = ft - gt = 0$ and so
\[
\deg(f - g) + \deg t = \deg((f - g)t) = \deg 0 = -\infty.
\]
Since $t \neq 0$, we have $\deg t \in \mathbb{N}$. As $\deg(f - g) + \deg t = -\infty$ we conclude that $\deg(f - g) = -\infty$.

Hence $f - g = 0$ and $f = g$.

$\Leftarrow$: If $f = g$, then clearly $ft = gt$.

(b):
\[
\divides ft \iff \exists l \in F[x] \quad \text{such that} \quad gt = l(ft)
\]
\[
\divides gt \iff \exists l \in F[x] \quad \text{such that} \quad g = (lf)t
\]
\[
\divides g \iff \exists l \in F[x] \quad \text{such that} \quad f = g
\]

Theorem VII.27. Let $F$ be a field, $h \in F[x]$ with $h \neq 0$ and $I$ an ideal in $F[h][x]$. Observe that $h = 0 \in I$ and choose $g \in F[x]$ of minimal degree subject to $\overline{g} \in I$ and $g \neq 0$.

(a) Let $f \in F[x]$. Then $\overline{f} \in I$ if and only if $g$ divides $f$ in $F[x]$.

(b) $I = (\overline{g})$ is the ideal in $F[h][x]$ generated by $\overline{g}$.

(c) $g$ divides $h$ in $F[x]$.

According to (c) choose $t \in F[x]$ with $h = tg$. Put $k := \deg t$.

(d) $t \neq 0$ and $k \in \mathbb{N}$.

(e) Let $f, f^* \in F[x]$. Then $f \odot g = f^* \odot g$ if and only if $t$ divides $f^* - f$.

(f) $I = \{sg \mid s \in F[x], \deg s < k\}$. Moreover, for each $f \in I$ there exist a unique $s \in F[x]$ with $f = sg$ and $\deg s < k$.

(g) $(g, xg, x^2g, \ldots, x^{k-1}g)$ is an $F$-basis for $I$.

(h) $\dim_F I = k$.

(i) Let $f \in F[x]$. Then $\overline{f} \in I$ if and only if $f \odot t = 0$.

(j) Suppose that $h = x^n - 1$ and let $f \in F[h][x]$. Then $f \in I$ if and only if the coefficient of $x^i$ in $ft$ is 0 for all $k \leq i < n$. 
VII.1. CLASSIFICATION AND PROPERTIES OF CYCLIC CODES

Proof. (a): \(\implies\): Suppose \(\overline{f} \in I\). Let \(f = qg + r\) with \(q, r \in \mathbb{F}[x]\) and \(\deg r < \deg g\). Then

\[
\overline{f} = f + (-q)g = \overline{f} \oplus \overline{-q} \circ \overline{g}.
\]

Since both \(\overline{f}\) and \(\overline{g}\) are in \(I\) and \(I\) is an ideal, this gives \(\overline{r} \in I\). As \(\deg r < \deg g\), the minimal choice of \(\deg g\) shows that \(r = 0\) and so \(g \mid f\).

\(?\): Suppose that \(g \mid f\). Then \(f = qg\) for some \(q \in \mathbb{F}[x]\) and so \(\overline{f} = \overline{qg} = \overline{q} \circ \overline{g}\). Recall that \(\overline{q} \in \mathbb{F}^{h}[x]\), \(\overline{g} \in I\) and \(I\) is an ideal in \(\mathbb{F}^{h}[x]\). Thus \(\overline{f} = \overline{q} \circ \overline{g} \in I\).

\(b\): Let \(f \in I\). Then \(f \in \mathbb{F}^{h}[x]\) and so \(\overline{f} = f \in I\). Hence \(\text{(a)}\) shows that \(g\) divides \(f\) and so \(f = eg\) for some \(e \in \mathbb{F}[x]\). Therefore \(f = \overline{f} = \overline{e} \circ \overline{g}\) and so \(f \in (\overline{g})\). Hence \(I \subseteq (\overline{g})\).

Since \(\overline{g} \in I\) and \(I\) is an ideal, \(\text{VII.23(c)}\) shows \((\overline{g}) \subseteq I\). Thus \(I = (\overline{g})\).

\(c\): Note that \(\overline{h} = 0 \in I\) and so \(g \mid h\) by \(\text{(a)}\).

\(d\): Since \(h \neq 0\) and \(h = tg\) we get \(t \neq 0\). As \(t \neq 0\) we have \(k = \deg(t) \in \mathbb{N}\).

\(e\): We have

\[
\begin{align*}
f \circ g &= f^* \circ g \\
\iff\quad \overline{fg} &= \overline{f^*g} & \text{definition of } \circ \\
\iff\quad h \text{ divides } fg - f^*g & \text{ VII.14(g)} \\
\iff\quad tg \text{ divides } (f - f^*)g & \text{ VII.26} \\
\iff\quad t \text{ divides } f - f^* & \text{ VII.26}
\end{align*}
\]

\(f\): Let \(f \in \mathbb{F}^{h}[x]\) and let \(s\) be the remainder of \(f\) when divided by \(t\). Then \(\deg s : \det t\) and \(t\) divides \(f - s\) in \(\mathbb{F}[x]\). Thus \(\text{(e)}\) gives \(f \circ g = s \circ g\). Note that

\[
\deg sg = \deg s + \deg g < \deg t + \deg g = \deg tg = \deg h,
\]

and so \(\overline{sg} = sg\). Hence

\[
f \circ \overline{g} = f \circ g = s \circ g = \overline{sg} = sg.
\]

Note also that \(\deg s < \deg t = k\). Thus

\[
I = (\overline{g}) = \{f \circ \overline{g} \mid f \in \mathbb{F}^{h}[x]\} = \{sg \mid s \in \mathbb{F}[x], \deg s < k\}.
\]

So for each \(f \in I\) there exists \(s \in \mathbb{F}[x]\) with \(\deg s < k\) and \(f = sg\). Suppose that \(\tilde{s} \in \mathbb{F}[x]\) with \(\deg \tilde{s} < k\) and \(f = \tilde{sg}\). Then \(sg = \tilde{sg}\). It follows that \((s - \tilde{s})g = 0\) and since \(g \neq 0\) we get \(s - \tilde{s} = 0\) and \(s = \tilde{s}\). Hence \(s\) is unique.
Let \( f \in I \). By \( (g) \) there exists a unique \( s \in \mathbb{F}[x] \) with \( \deg s < k \) and \( f = sg \). Note that \( s = \sum_{i=0}^{k-1} a_i x^i \) for unique \( a_i \in \mathbb{F} \). Then

\[
(*) \quad f = sg = \sum_{i=0}^{k-1} a_i x^i g.
\]

So for each \( f \in I \) there exists a unique \((a_0, \ldots, a_{k-1}) \in \mathbb{F}^k \) with \( f = \sum_{i=0}^{k-1} a_i x^i g \). Thus \( (g) \) holds.

Note that \( \dim_{\mathbb{F}} I \) is the size of any \( \mathbb{F} \)-basis of \( I \). So \( (h) \) follows from \( (g) \).

Let \( f \in \mathbb{F}^h[x] \) and let \( f = qg + r \), where \( q, r \in \mathbb{F}[x] \) with \( \deg r < \deg g \).

By \( (a) \),

\[
(**) \quad f = \overline{f} \in I \quad \iff \quad g \mid f \quad \iff \quad gt \mid ft \quad \iff \quad h \mid ft \quad \iff \quad ft = 0 \quad \iff \quad f \odot t = 0
\]

We claim that \( \deg(qg) \leq \deg f \). Indeed, if \( q = 0 \), then also \( qg = 0 \) and so \( \deg(qg) = -\infty \leq \deg f \). If \( q \neq 0 \), then \( \deg(qg) = \deg(q) + \deg(g) \geq \deg(q) > \deg r \) and so by \( (VII.6) \) \( \deg f = \deg(qg+r) = \deg(qg) \). We proved that \( \deg(qg) \leq \deg f \). Thus

\[
\deg q + \deg g = \deg(qg) \leq \deg f < \deg h = \deg(tg) = \deg t + \deg g = k + \deg g
\]

and so

\[
(***) \quad \deg qg < k.
\]

Also

\[
\deg rt = \deg r + \deg t < \deg g + \deg t = \deg gt = \deg h = \deg(x^n - 1) = n.
\]

and so
VII.1. CLASSIFICATION AND PROPERTIES OF CYCLIC CODES

deg rt < n.

If r ≠ 0, then deg rt ≥ deg t = k. If r = 0, then deg(rt) = deg 0 = −∞ < k. So r = 0 if and only if deg rt < k. Since deg rt < n, this gives

(+) \quad r = 0 \quad \iff \quad \text{the coefficient of } x^i \text{ in } rt \text{ is 0 for all } k \leq i < n

We compute

\begin{align*}
ft = (qg + r)t &= qgt + rt = qh + rt = q \cdot (x^n - 1) + rt = qx^n - q + rt
\end{align*}

Let k ≤ i < n. Since i < n, the coefficient of x^i in qx^n is 0. Since i ≥ k and, by (**) \deg q < k, the coefficient of x^i in q is 0. Thus

(++) \quad \text{For all } k \leq i < n, \text{ the coefficient of } x^i \text{ in } ft \text{ is the same as the coefficient of } x^i \text{ in } rt.

From (**) \quad (++) \quad \text{and (++) we see that } f \in I \text{ if and only if the coefficient of } x^i \text{ in } ft \text{ is 0 for all } k \leq i < n. \text{ Hence } (j) \text{ holds.}

\textbf{Definition VII.28.} Let \mathbb{F} be a field.

(a) Let \( f \in \mathbb{F}[x] \). If \( f = \sum_{i=0}^{m} a_i x^i \) with \( a_m \neq 0 \), define \( \text{lead}(f) := a_m \). If \( f = 0 \) define \( \text{lead}(f) = 0 \). Then \( \text{lead}(f) \) is called the leading coefficient of \( f \).

(b) A monic polynomial is a polynomial with leading coefficient 1.

(c) Let \( h \in \mathbb{F}[x] \) with \( h \neq 0 \) and let \( I \) be an ideal in \( \mathbb{F}[x] \). Let \( g \in \mathbb{F}[x] \) be a monic polynomial of minimal degree with \( g \in I \). Then \( g \) is called the canonical generator for \( I \).

\textbf{Example VII.29.} Determine the cyclic code generated by 0101 and find the canonical generator for the corresponding ideal in \( \mathbb{V}^4[x] \).

The cyclic shifts of 0101 are

\begin{align*}
0101, \quad 1010.
\end{align*}

Thus

\begin{align*}
\langle 1010 \rangle &= \{0000, 0101, 1010, 1111\}.
\end{align*}

The non-zero codeword with the most trailing zeros is 1010. So the canonical generator is \( 1 + x^2 \).

\textbf{Lemma VII.30.} Let \( \mathbb{F} \) be a field and \( h \in \mathbb{F}[x] \) with \( h \neq 0 \).

(a) Let \( I \) an ideal in \( \mathbb{F}[x] \). Then there exists a unique canonical generator \( g \) of \( I \).
(b) $I = (\overline{g})$ in $\mathbb{F}^h[x]$ and $g$ divides $h$ in $\mathbb{F}[x]$.

(c) If $I \neq \{0\}$, then $\deg g < \deg h$ and $0 \neq g = \overline{g} \in I$. If $I = \{0\}$, then $\deg g = \deg h$, $\overline{g} = 0$ and $g = (\text{lead } h)^{-1}h$.

(d) Suppose $I \neq 0$. Then $g$ is a non-zero polynomial of minimal degree in $I$. Moreover, $g$ is the unique monic polynomial of minimal degree in $I$.

Proof. Choose $g \in F[x]$ of minimal degree subject to $g \neq 0$ and $\overline{g} \in I$. Since $I$ is an ideal we get $(\text{lead } g)^{-1}g = (\text{lead } g)^{-1}\overline{g} \in I$. So replacing $g$ by $(\text{lead } g)^{-1}g$ we may assume that $g$ is monic.

(a) To show that $g$ is a canonical generator for $I$, let $f$ be a monic polynomial with $\overline{f} \in I$. Then $f \neq 0$ and so $\deg f \geq \deg g$, by choice of $g$. Hence $g$ is a canonical generator.

Let $f$ be any canonical generator for $I$. Then the minimal choice of the degree of a canonical generator shows that $\deg g \leq \deg f$ and vice versa. Thus $\deg g = \deg f$. Since both $g$ and $f$ are monic we conclude that $\deg(g - f) < \deg g$. Also $\overline{g - f} = \overline{g} - \overline{f} \in I$ and the minimal choice of $\deg g$ gives $g - f = 0$. Thus $f = g$ and $g$ is the unique canonical generator for $I$.

(b) See VII.27(b),(c).

(c) Suppose that $I \neq 0$ and let $f \in I$ with $f \neq 0$. Then $\deg f < \deg h$ and so $\overline{f} = f \in I$. The minimal choice of $\deg g$ implies that $\deg f \leq \deg g$. Thus $\deg g < \deg h$. Hence $g = \overline{g} \in I$.

Suppose that $I = 0$. Then $\overline{g} = 0$. Thus $h|g$ and so $\deg h \leq \deg g$. By VII.27 we know that $h = tg$ for some $t \in \mathbb{F}[x]$ with $t \neq 0$. Then

$$\deg g \leq \deg(t) + \deg(g) = \deg(tg) = \deg h \leq \deg g.$$ 

It follows that $\deg t = 0$. So $t \in \mathbb{F}$. As $h = tg$ we get lead$(h) = t \text{ lead}(g) = t1 = t$. So $t = \text{lead}(h)$ and $h = tg = \text{lead}(h)g$. Thus $g = (\text{lead } h)^{-1}h$.

(d) Let $f \in I$ with $f \neq 0$. Then $f \in \mathbb{F}^h[x]$ and so $\overline{f} = f \in I$. The minimal choice of $\deg g$ now implies that $\deg f \geq \deg g$. By (c) $g \in I$ and so (d) holds.

Lemma VII.31. Let $F$ be a field, $h \in \mathbb{F}[x]$ with $h \neq 0$ and let $f \in \mathbb{F}[x]$ be monic. Then $f$ is the canonical generator for the ideal $(\overline{f})$ in $\mathbb{F}^h[x]$ if and only if $f \mid h$.

Proof. If $f$ is the canonical generator for $(\overline{f})$, then VII.27(c) shows that $f \mid h$.

Conversely suppose that $f \mid h$ and let $e$ be the canonical generator for $(\overline{f})$. Since $\overline{e} \in (\overline{f})$,

$$\overline{e} = d \circ \overline{f} = df$$

for some $d \in \mathbb{F}^h[x]$. Thus VII.14 shows that $h$ divides $e - df$. Hence $e - df = lh$ for some $l \in \mathbb{F}[x]$. So $e = df + lh$. Since $f \mid h$ this shows that $f \mid e$. Hence $\deg e \geq \deg f$. Note that $f$ is monic polynomial with $\overline{f} \in (\overline{f})$. So the definition of the canonical generator $e$ implies $\deg e \leq \deg f$. Hence $\deg f = \deg e$ and so also $f$ is a canonical generator of $(\overline{f})$. \qed
Corollary VII.32. Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$ with $h \neq 0$. Then the function

$$g \mapsto \langle \overline{g} \rangle$$

is a 1-1 correspondence between monic divisors of $h$ in $F[x]$ and the ideals of $\mathbb{F}^h[x]$, with inverse

$$I \mapsto \text{canonical divisor of } I.$$

Proof. Let $\alpha$ be the first function and $\beta$ the second. We just need to show that the two functions are inverse to each other.

Let $g$ be a monic divisor of $h$. Then $\alpha(g) = \langle \overline{g} \rangle$. By VII.31 $g$ is a canonical generator of $\langle \overline{g} \rangle$. Thus

$$g = \beta(\langle \overline{g} \rangle) = \beta(\alpha(g)).$$

Now let $I$ be an ideal in $\mathbb{F}^h[x]$. Let $g$ be the canonical generator of $I$, so $g = \beta(I)$. By VII.27(c) $g$ is a divisor of $h$ and by VII.27(b) we have $I = \langle \overline{g} \rangle$. Thus

$$I = \langle \overline{g} \rangle = \alpha(g) = \alpha(\beta(I)).$$

Theorem VII.33. Let $C \subseteq \mathbb{F}_2^n$ be a binary cyclic code. Let $g$ be the canonical generator of the ideal $C(x)$ in $V^n[x]$. Let $t \in \mathbb{F}_2[x]$ with $gt = x^n - 1$. Put $k := \deg t$ and $m := n - k = \deg g$. Let

$$g = c_0 + c_1 x + \ldots + c_m x^m \quad \text{and} \quad t = h_0 + h_1 x + \ldots h_k x^k,$$

with $c_i, h_i \in \mathbb{F}_2$.

(a) The $n \times k$ matrix
CHAPTER VII. ALGEBRAIC CODING THEORY

\[ E = \begin{bmatrix}
    c_0 & 0 & 0 & \ldots & 0 & 0 \\
    c_1 & c_0 & 0 & \ldots & 0 & 0 \\
    \vdots & c_1 & c_0 & \ldots & \vdots & \vdots \\
    \vdots & \vdots & c_1 & \ldots & 0 & \vdots \\
    c_{m-1} & \vdots & \vdots & \ldots & c_0 & 0 \\
    c_m & c_{m-1} & \vdots & \ldots & c_1 & c_0 \\
    0 & c_m & c_{m-1} & \vdots & \vdots & \vdots \\
    0 & 0 & c_m & \vdots & \vdots & \vdots \\
    \vdots & 0 & 0 & \ldots & c_{m-1} & \vdots \\
    \vdots & \vdots & \vdots & \ldots & c_m & c_{m-1} \\
    0 & 0 & 0 & \ldots & 0 & c_m
\end{bmatrix} \]

is a generating matrix or \( C \). (Note here that for \( g = x^n - 1 \), one has \( m = n, k = 0, E = [] \) and \( C = \{0\} \).

(b) The \( m \times n \) matrix

\[ H = \begin{bmatrix}
    h_k & h_{k-1} & h_{k-2} & \ldots & h_1 & h_0 & 0 & 0 & \ldots & 0 \\
    0 & h_k & h_{k-1} & \ldots & h_2 & h_1 & h_0 & 0 & \ldots & 0 \\
    0 & 0 & h_k & \ldots & h_3 & h_2 & h_1 & h_0 & \ldots & 0 \\
    \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\
    0 & 0 & 0 & \ldots & h_k & \ldots & h_1 & h_0 & 0 \\
    0 & 0 & 0 & \ldots & 0 & h_k & \ldots & h_2 & h_1 & h_0
\end{bmatrix} \]

is a check matrix for \( C \).

Proof. (a) Let \( c \in \mathbb{F}_2^n \) with \( c(x) = g \). So \( c = c_0 \ldots c_m 0 \ldots 0 \) and \( c \) is the first column of \( E \). By VII.27 \( C(x) \) is spanned by the polynomials \( x^i g, 0 \leq i < k \). Thus \( C \) is spanned by the cyclic shifts \( c^{(i)}, 0 \leq i < k \). Since the columns of \( E \) are the \( c^{(i)} \), (a) holds.

(b) Let \( d = d_0 d_1 \ldots d_{n-1} \in \mathbb{F}_2^n \). By VII.27 \( d(x) \in C(x) \) if and only if the coefficient \( a_s \) of \( x^s \) in \( t \cdot d(x) \) is equal to 0 for all \( k \leq s < n \). Since \( \deg t = k \) we have \( h_l = 0 \) for \( l > k \). Thus
\[ a_s = \sum_{l=0}^{s} h_l d_{s-l} = \sum_{l=0}^{k} h_l d_{s-l} = h_k d_{s-k} + h_{k-1} d_{s-k+1} + \ldots + h_1 d_{s-1} + h_0 d_s. \]

Hence

\[
\begin{align*}
a_k &= h_k d_0 + h_{k-1} d_1 + h_{k-2} d_2 + \ldots + h_1 d_{k-1} + h_0 d_k \\
a_{k+1} &= h_k d_1 + h_{k-1} d_2 + \ldots + h_2 d_{k-1} + h_1 d_k + h_0 d_{k+1} \\
a_{k+2} &= h_k d_2 + \ldots + h_3 d_{k-1} + h_2 d_k + h_1 d_{k+1} + h_0 d_{k+2} \\
\vdots & \vdots \quad \ddots \quad \ddots \quad \ddots \quad \ddots \\
a_{n-2} &= h_k d_{n-k-2} + \ldots + \ldots + h_1 d_{n-3} + h_0 d_{n-2} \\
a_{n-1} &= h_k d_{n-k-1} + \ldots + h_2 d_{n-3} + h_1 d_{n-2} + h_0 d_{n-1}
\end{align*}
\]

and so

\[
\begin{pmatrix}
  a_k \\
a_{k+1} \\
\vdots \\
a_{n-2} \\
a_{n-1}
\end{pmatrix} =
H \cdot 
\begin{pmatrix}
  d_0 \\
d_1 \\
\vdots \\
d_{n-1} \\
d_n
\end{pmatrix}
\]

It follows that \( a_s = 0 \) for \( k \leq s < n \) if and only if \( Hd = 0 \). Thus \( H \) is a check matrix for \( C \).

**Example VII.34.** Find a generating and a check matrix for the cyclic code \( C \) of length 7 with canonical generator \( g = 1 + x^2 + x^3 + x^4 \). Is 1001011 in the code?

Let \( E \) be the generating matrix from VII.33(a). Since \( g = 1 + x^2 + x^3 + x^4 \) and \( n = 7 \), the first column of \( E \) is 1011100. So

\[
E = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 1 \\
1 & 1 & 0 \\
1 & 1 & 1 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{bmatrix}
\]
Let $H$ be the check matrix from VII.33(b). We first compute $t := \frac{x^7+1}{g}$:

\[
\begin{array}{c}
1101 \\
1110110000001 \\
11101101001 \\
111011101 \\
0
\end{array}
\]

Hence $t = x^3 + x^2 + 1$. Also $n = 7$ and so the first row of $H$ is 1101000. Thus

\[
H = \begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 1
\end{bmatrix}
\]

To check whether $d = 1001011$ is in the code we compute $Hd$:

\[
Hd = \begin{pmatrix}
1 & 0 & 0 & 1 & 0 & 1 & 1
\end{pmatrix}
\begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 1
\end{bmatrix} = \begin{pmatrix}
1+1+0+0 \\
0+0+0+0 \\
0+1+1+0 \\
0+1+0+1
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
0 \\
0
\end{pmatrix}
\]

So $Hd = \tilde{0}$ and $d \in C$.

We could also have observed that $d$ is the sum of the first and last column of $E$ and so $d \in C$.

### VII.2 Definition of a family of BCH codes

**Irreducible polynomials**

**Definition VII.35.** Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$. Then $h$ is called irreducible provided that
VII.2. DEFINITION OF A FAMILY OF BCH CODES

(i) $\deg h > 0$, and

(ii) if $h = fg$ for some $f, g \in \mathbb{F}[x]$, then $\deg f = 0$ or $\deg g = 0$.

Remark VII.36. Let $\mathbb{F}$ be a field and $h \in \mathbb{F}[x]$. Then $h$ is irreducible if and only

(I) $\deg h > 0$, and

(II) if $f \mid h$ for some $f \in \mathbb{F}[x]$, then $\deg f = 0$ or $\deg f = \deg h$.

Proof. We may assume that $\deg h > 0$.

$\implies$: Suppose first that $h$ is irreducible and let $f \in \mathbb{F}[x]$ with $f \mid h$. Then $h = fg$ for some $g \in \mathbb{F}[x]$. Since $h$ is irreducible, $\deg f = 0$ or $\deg g = 0$. If $\deg g = 0$, then $\deg f = \deg h - \deg g = \deg f$.

$\Longleftarrow$: Suppose next that (I) 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. $\square$

Lemma VII.37. 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. $\implies$: Suppose $sr = 1$ for some $s \in R$. Let $t \in R$. Then $t = t1 = tsr = (ts)r \in \langle r \rangle$ and so $R = \langle r \rangle$.

$\Longleftarrow$: Suppose that $R = \langle r \rangle$. Then $1 \in \langle r \rangle$ and so $1 = sr$ for some $s \in R$. $\square$

Lemma VII.38. 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) $\mathbb{E}$ together with the addition $\oplus$ and multiplication $\odot$ is a field.

(c) Let $h = \sum_{i=0}^{n} h_i x^i$ with $h_i \in \mathbb{F}$. Then $x$ is a root in $\mathbb{E}$ of the polynomial $\sum_{i=0}^{n} h_i y^i$ in $\mathbb{E}[y]$.

Proof. (a) 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 know that $a, b \in \mathbb{F}^h[x]$, $a \oplus b = a + b$ and $a \odot b = ab$. So $\mathbb{F}$ is a subfield of $\mathbb{F}^h[x]$.

(b) Let $0 \neq f \in \mathbb{F}^h[x]$ and put $I := \langle f \rangle$. Let $g$ be the canonical generator for $I$. As $I \neq 0$ we conclude from VII.30(c) that $\deg g < \deg h$ and $g = \overline{g} \in I$. Moreover, VII.30(b) $g| h$. Since $h$ is irreducible, we get $\deg g = 0$, see VII.36. As $g$ is monic, this gives $g = 1$. So $1 = 1g$ and VII.37 implies that $\langle g \rangle = \mathbb{F}^h[x]$. Hence $\langle f \rangle = I = \langle g \rangle = \mathbb{F}^h[x]$ and another application of VII.37 shows that $s \odot f = 1$ for some $s \in \mathbb{F}^h[x]$. Hence $\mathbb{F}^h[x]$ is a field.

(c) Let $e \in \mathbb{E}$. In $\mathbb{E}$ we have to use the operations $\oplus$ and $\odot$. Hence $e$ is a root of $\sum_{i=0}^{n} h_i y^i$ in $\mathbb{E}$ if and only if
\[ h_0 \oplus h_1 \oplus e \oplus h_2 \oplus e^{\circ 2} \oplus \ldots \oplus h_n \oplus e^{\circ n} = 0, \]

and if and only if

\[ \sum_{i=0}^{n} h_i e^i = 0. \]

So \( x \) is a root of \( \sum_{i=0}^{n} h_i y^i \) if and only if

\[ \sum_{i=0}^{n} h_i x^i = 0. \]

But this just say \( \overline{h} = 0 \), which is true.

**Example VII.39.** 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 \( \alpha \) for \( x \) to indicate that our computation are in the field \( \mathbb{E} \) (rather than in \( \mathbb{F}[x] \)). Then

\[ 1 + \alpha + \alpha^3 = \overline{1 + x + x^3} = \overline{h} = 0. \]

Thus

\[ \alpha^3 = -(1 + \alpha) = 1 + \alpha. \]

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 \( \alpha \).

\[
\begin{align*}
\alpha^0 &= 1 \\
\alpha^1 &= \alpha \\
\alpha^2 &= \alpha^2 \\
\alpha^3 &= 1 + \alpha \\
\alpha^4 &= \alpha \alpha^3 = \alpha (1 + \alpha) = \alpha + \alpha^2 \\
\alpha^5 &= \alpha \alpha^4 = \alpha (\alpha + \alpha^2) = \alpha^2 + \alpha^3 = \alpha^2 + (1 + \alpha) = 1 + \alpha + \alpha^2 \\
\alpha^6 &= \alpha \alpha^5 = \alpha (1 + \alpha + \alpha^2) = \alpha + \alpha^2 + \alpha^3 = \alpha + \alpha^2 + (1 + \alpha) = 1 + \alpha^2 \\
\alpha^7 &= \alpha \alpha^6 = \alpha (1 + \alpha^2) = \alpha + \alpha^3 = \alpha + 1 + \alpha = 1.
\end{align*}
\]
We will verify this by direct computation:

\[ h(\alpha^2) = 1 + \alpha^2 + (\alpha^2)^3 = 1 + \alpha^2 + \alpha^6 = 1 + \alpha^2 + (1 + \alpha^2) = 0 \]

and

\[ h(\alpha^4) = 1 + \alpha^4 + (\alpha^4)^3 = 1 + \alpha^4 + \alpha^{12} = 1 + \alpha^4 + \alpha^5 = 1 + (\alpha + \alpha^2) + (1 + \alpha + \alpha^2) = 0. \]

Thus \( \alpha, \alpha^2, \alpha^4 \) are the roots of \( 1 + x + x^3 \).

From \( \alpha^7 = 1 \) we have \((\alpha^i)^7 = (\alpha^7)^i = 1 \). So if \( 0 \neq \epsilon \in \mathbb{E} \), then \( \epsilon^7 = 1 \) and \( \epsilon \) is a root of \( x^7 - 1 \).

Note that \((1 + x + x^3)(1 + x) = 1 + x + x^3 + x^2 + x^4 = 1 + x^2 + x^3 + x^4 \). In Example VII.34 we computed that

\[ \frac{x^7-1}{1+x+x^2+x^3} = 1 + x^2 + x^3. \]

Thus

\[ x^7 - 1 = (1 + x)(1 + x + x^3)(1 + x^2 + x^3). \]

1 is a root of \( 1 + x \) and \( \alpha, \alpha^2, \alpha^4 \) are the roots of \( 1 + x + x^3 \). So \( \alpha^3, \alpha^4 \) and \( \alpha^6 \) 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 \( \alpha^6 \) and \( \alpha^5 \) are roots of \( 1 + x^2 + x^3 \).

**Lemma VII.40.** Let \( n \in \mathbb{Z}^+ \) and write

\[ x^n - 1 = f_0 f_1 \cdots f_l \]

where \( f_0, f_1, \ldots, f_l \) are irreducible polynomials in \( \mathbb{F}_2[x] \). Let \( C \subseteq \mathbb{F}_2^n \) be a cyclic code and let \( g \) the canonic generator for the ideal \( C(x) \) in \( V^n[x] \) corresponding to \( C \). Then there exist \( \epsilon_i \in \{0, 1\}, 0 \leq i \leq l \) with

\[ g = f_0^{\epsilon_0} \cdots f_l^{\epsilon_l} \]

Moreover, \( x^n - 1 = gt \) where

\[ t = f_0^{\delta_0} \cdots f_l^{\delta_l} \]

and \( \delta_i = 1 - \epsilon_i \).

**Proof.** By VII.27 \( g \) divides \( x^n - 1 \) in \( \mathbb{F}_2[x] \). The lemma now follows from A.13.\qed
Example VII.41. 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 binary cyclic codes of length 7.

Let \( C \) be a four dimensional binary cyclic code of length 7. Let \( g \) be the canonical generator of \( C(x) \) and let \( t \in \mathbb{F}_2[x] \) with \( gt = x^7 - 1 \). Then by VII.27(h) \( \deg t = \dim C = 4 \) and so \( \deg g = 7 - 4 = 3 \). By VII.40

\[
g = (1 + x)\epsilon_1(1 + x + x^3)\epsilon_2(1 + x^2 + x^3)\epsilon_3
\]

for some \( \epsilon_1, \epsilon_2, \epsilon_3 \in \{0, 1\} \). We have \( 3 = \deg g = \epsilon_1 + 3\epsilon_2 + 3\epsilon_3 \). It follows that \( 3 \mid \epsilon_1 \). Thus \( \epsilon_1 = 0 \) and \( 3 = 3\epsilon_2 + 3\epsilon_3 \). Hence \( 1 = \epsilon_2 + \epsilon_3 \). It follows that either \( \epsilon_2 = 1 \) and \( \epsilon_3 = 0 \) or \( \epsilon_2 = 0 \) and \( \epsilon_3 = 1 \). Therefore,

\[
g = 1 + x + x^3 \quad \text{and} \quad t = (1 + x)(1 + x^2 + x^3) = (1 + x^2 + x^3) + (x + x^3 + x^4) = x^4 + x^2 + x + 1
\]

or

\[
g = 1 + x^2 + x^3 \quad \text{and} \quad t = (1 + x)(1 + x + x^3) = (1 + x + x^3) + (x + x^2 + x^4) = x^4 + x^3 + x^2 + 1
\]

Thus

\[
E = \begin{bmatrix}
1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
1 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

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
\[
E = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\quad \text{and} \quad
H = \begin{bmatrix}
1 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 1
\end{bmatrix}
\]

Observe that for both codes the columns of \( H \) are the non-zero elements of \( \mathbb{F}_2^3 \). Hence, by VI.36 both codes are Hamming codes.

**Lemma VII.42.** Let \( \mathbb{F} \) be a field and \( 0 \neq f \in \mathbb{F}[x] \). Let \( a = \text{lead}(f) \) and \( n = \text{deg}f \). Then there exists a field \( \mathbb{E} \) containing \( \mathbb{F} \) and elements \( \alpha_1, \alpha_2 \ldots, \alpha_n \) in \( \mathbb{E} \) such that
\[
f = a(x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_n)
\]

Moreover, if \( \mathbb{F} \) is finite we can choose \( \mathbb{E} \) to be finite.

Such a field \( \mathbb{E} \) is called a splitting field for \( f \) over \( \mathbb{F} \).

**Proof.** The proof is by induction on \( \text{deg}f \). If \( \text{deg}f = 0 \), then \( f = a \) and the lemma holds with \( \mathbb{E} = \mathbb{F} \). So suppose \( \text{deg}f > 0 \). Then \( f = gh \) with \( g, h \in \mathbb{F}[x] \) and \( h \) irreducible. Put \( \mathbb{K} := \mathbb{F}^h[x] \). Then by VII.38 \( \mathbb{K} \) is a field and there exists a root \( \alpha \) of \( h \) in \( \mathbb{K} \). Moreover, \( \mathbb{K} \) is finite if \( \mathbb{F} \) is finite. Since \( \alpha \) is a root of \( h \) in \( \mathbb{K} \) we know that \( x - \alpha \) divides \( h \) in \( \mathbb{K}[x] \). As \( f = gh \) we conclude that \( x - \alpha \) divides \( f \) and so \( f = d \cdot (x - \alpha) \) for some \( d \in \mathbb{K}[x] \). Observe that \( \text{deg}d = n - 1 \) and \( \text{lead}(d) = \text{lead}(f) = a \). By induction there exist a field \( \mathbb{E} \) containing \( \mathbb{K} \) and elements \( \alpha_1, \alpha_2 \ldots, \alpha_{n-1} \) in \( \mathbb{E} \) such that
\[
d = a(x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_{n-1})
\]
and \( \mathbb{E} \) is finite if \( \mathbb{K} \) is finite. Hence
\[
f = d \cdot (x - \alpha) = a(x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_{n-1})(x - \alpha_n),
\]
where \( \alpha_n = \alpha \).

**Lemma VII.43.** Let \( \mathbb{F} \) and \( \mathbb{E} \) be a fields with \( \mathbb{F} \subseteq \mathbb{E} \). Let \( \alpha \in \mathbb{E} \) and suppose that \( \alpha \) is a root of some non-zero polynomial in \( \mathbb{F}[x] \). Let \( m \in \mathbb{F}[x] \) be a monic polynomial of minimal degree with respect to \( m(\alpha) = 0 \). Put \( \mathbb{F}[\alpha] := \{ f(\alpha) \mid f \in \mathbb{F}[x] \} \). Let \( f, g \in \mathbb{F}[x] \) and let \( r \) and \( s \) be the remainders of \( f \) and \( g \), respectively, when divided by \( m \). Then
(a) \( f(\alpha) = r(\alpha) \).

(b) \( f(\alpha) = 0 \) if and only if \( r = 0 \) and if and only if \( m \) divides \( f \) in \( \mathbb{F}[x] \).

(c) 
\[
 f(\alpha) = g(\alpha) \iff r = s \iff m \mid f - g.
\]

(d) For each \( e \in \mathbb{F}[\alpha] \) there exists a unique \( r \in \mathbb{F}^m[x] \) with \( e = r(\alpha) \).

(e) Let \( n = \deg m \). Then \( (1, \alpha, \alpha^2, \ldots, \alpha^{n-1}) \) is an \( \mathbb{F} \)-basis for \( \mathbb{F}[\alpha] \). In particular, \( \dim_{\mathbb{F}} \mathbb{F}[\alpha] = n \).

(f) \( m \) is the unique monic irreducible polynomial in \( \mathbb{F}[x] \) with \( m(\alpha) = 0 \).

(g) The function
\[
\Phi : \mathbb{F}^m[x] \to \mathbb{F}[\alpha], \quad f \mapsto f(\alpha)
\]
is an isomorphism of rings. In particular, \( \mathbb{F}[\alpha] \) is a subfield of \( \mathbb{K} \) isomorphic to \( \mathbb{F}^m[x] \).

\( m \) is called the minimal polynomial of \( \alpha \) over \( \mathbb{F} \) and is denoted by \( m_\alpha \) or \( m_{\alpha} \).

**Proof.**

(a) Let \( f = qm + r \) with \( q, r \in \mathbb{F}[x] \) and \( \deg r < \deg m \). Then \( f(\alpha) = q(\alpha)m(\alpha) + r(\alpha) = q(\alpha)0 + r(\alpha) = r(\alpha) \). So (a) holds.

(b) We have \( f(\alpha) = 0 \) if and only if \( r(\alpha) = 0 \). Since \( \deg r < \deg m \), the minimality of \( \deg m \) shows that \( r(\alpha) = 0 \) if and only if \( r = 0 \). By (d) \( r = 0 \) if and only if \( m \mid f \). So (b) holds.

(c) Observe that \( r - s \) is the remainder of \( f - g \) when divided by \( m \). Hence by (b) applied to \( f - g \) in place of \( f \):
\[
(f - g)(\alpha) = 0 \iff r - s = 0 \iff m \mid f - g.
\]

Observe that \( (f - g)(\alpha) = 0 \) if and only if \( f(\alpha) = g(\alpha) \). Also \( r - s = 0 \) if and only if \( r = s \). So (c) is proved.

(d) Let \( e \in \mathbb{F}[\alpha] \). Then \( e = f(\alpha) \) for some \( f \in \mathbb{F}[x] \). By (a) we have \( f(\alpha) = r(\alpha) \). Since \( \deg r < \deg m \) we know that \( r \in \mathbb{F}^m[x] \). This shows that existence of \( r \). Suppose \( e = g(\alpha) \) for some \( g \in \mathbb{F}^m[x] \). Then \( g(\alpha) = e = f(\alpha) \). So (c) implies that \( r = s \). As \( g \in \mathbb{F}^m[x] \) we have \( g = s \). Hence \( g = r \) and \( r \) is unique.

(e) Let \( e \in \mathbb{F}[\alpha] \).

By (d) there exists a unique \( r \in \mathbb{F}^m[x] \) with \( e = r(\alpha) \). Then \( r = \sum_{i=0}^{n-1} a_i x^n \) for unique \( a_0, \ldots, a_{n-1} \in \mathbb{F} \).

This shows that there exists unique \( a_0, \ldots, a_{n-1} \in F \) with \( e = a_0 + a_1 \alpha + \ldots + a_{n-1} \alpha^{n-1} \). Thus \( (1, \alpha, \ldots, \alpha^{n-1}) \) is an \( \mathbb{F} \)-basis for \( \mathbb{F}[\alpha] \).
VII.2. DEFINITION OF A FAMILY OF BCH CODES

[1] Suppose that $m = gh$ for some $g, h ∈ \mathbb{F}[x]$. Then $g(α)h(α) = m(α) = 0$ and since $\mathbb{E}$ is a field, $g(α) = 0$ or $h(α) = 0$. Without loss $g(α) = 0$. 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 $g(α) = 0$. By [1], $m|g$ and since $g$ is irreducible, [A.11] implies $m = g$. Thus $m$ is unique.

[2] By (d) $\Phi$ is a bijection. Also

$$\begin{align*}
\Phi(f + g)(α) &= f(α) + g(α) = \Phi(f) + \Phi(g), \\
\Phi(f ⋅ g)(α) &= f(α) ⋅ g(α) = \Phi(f) ⋅ \Phi(g).
\end{align*}$$

So $\Phi$ is an isomorphism.

Definition VII.44. Let $\mathbb{E}$ be a field containing $\mathbb{F}_2$, let $C ⊆ \mathbb{F}_2^n$ be a binary linear code and let $H$ be an $m × n$-matrix with coefficients in $\mathbb{E}$. We say that $H$ is a check matrix for $C$ over $\mathbb{E}$ if

$$C = \{c ∈ \mathbb{F}_2^n \mid Hc = 0\}.$$  

Lemma VII.45. Let $\mathbb{F}$ be a field and $α_i, 1 ≤ i ≤ d$ be elements in $\mathbb{F}$. Let $H$ be the $d × n$ matrix

$$H = [α^j]_{0 ≤ j ≤ n} = \begin{bmatrix}
1 & α_1 & α_1^2 & \cdots & α_1^{n-1} \\
1 & α_2 & α_2^2 & \cdots & α_2^{n-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & α_d & α_d^2 & \cdots & α_d^{n-1}
\end{bmatrix}.$$  

and let $c = c_0c_1 \ldots c_{n-1} ∈ \mathbb{F}^n$. Then the $i$-th coefficient of $Hc$ is

$$c(α_i) := c_0 + c_1α_i + c_2α_i^2 + \ldots + c_{n-1}α_i^{n-1} = \sum_{j=0}^{n-1} c_jα_i^j.$$  

In particular, $Hc = 0$ if and only if $α_1, α_2, \ldots, α_d$ all are roots of $c(x)$.

Proof. The $i$'th coefficient of $Hc$ is

$$\sum_{j=0}^{n-1} α_i^j c_j = \sum_{j=0}^{n-1} c_jα_i^j = c(α_i).$$

□

Lemma VII.46. Let $C ⊆ \mathbb{F}_2^n$ be a binary cyclic code with canonical generator $g ∈ \mathbb{F}_2[x]$. Suppose that $g = m_1 \ldots m_s$, where $m_1, \ldots, m_s$ are pairwise distinct, irreducible, monic polynomials in $\mathbb{F}_2[x]$. Let $\mathbb{E}$ be a field containing $\mathbb{F}_2$ and let $α_1, \ldots, α_d$ be pairwise distinct elements in $\mathbb{E}$. Suppose that
(i) for each $1 \leq i \leq d$ there exists $1 \leq j \leq s$ with $m_j(\alpha_i) = 0$, and

(ii) for each $1 \leq j \leq s$ there exists $1 \leq i \leq d$ with $m_j(\alpha_i) = 0$.

Put

$$H := [\alpha_i^j]_{1 \leq i \leq d}.$$

Then $H$ is a check matrix for $C$ over $E$ and

$$C = \{ c \in \mathbb{F}_2^n | c(\alpha_i) = 0 \text{ for all } 1 \leq i \leq d \}.$$

**Proof.** Let $c \in \mathbb{F}_2^n$. We will first show that:

$$(*) \quad c \in C \quad \text{if and only if} \quad g \text{ divides } c(x).$$

Note first that $c \in C$ if and only if $c(x) \in C(x)$. Since $g$ is a canonical generator for $C(x)$, this is the case if and only if $g$ divides $c(x)$, see VII.27[a].

Next we shows

$$(***) \quad g \text{ divides } c(x) \quad \text{if and only if} \quad c(\alpha_i) = 0 \text{ for all } 1 \leq i \leq d.$$

Suppose $g$ divides $c(x)$. Let $1 \leq i \leq d$. By (i) there exists $1 \leq j \leq s$ with $m_j(\alpha_i) = 0$. Since $m_j$ divides $g$, $g(\alpha_i) = 0$ and since $g$ divides $c(x)$ we get $c(\alpha_i) = 0$.

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 $m_j(\alpha_i) = 0$. So by VII.43 $m_j$ divides $c(x)$. Since $g = m_1 \ldots m_s$ and the $m_j$’s are pairwise distinct monic irreducible polynomials we conclude from A.13 that $g$ divides $c(x)$.

From $(*)$ and $(***)$ we get

$$c \in C \quad \text{if and only if} \quad c(\alpha_i) = 0 \text{ for all } 1 \leq i \leq d.$$

and so

$$C = \{ c \in \mathbb{F}_2^n | c(\alpha_i) = 0 \text{ for all } 1 \leq i \leq d \}.$$

Hence VII.45 implies that $C = \{ c \in \mathbb{F}_2^n | Hc = \bar{0} \}$. Thus $H$ is check matrix for $C$ over $E$. 

### VII.3 Properties of BCH codes

**Definition VII.47.** Let $E$ be a finite field and $\alpha \in E$. Put $n := |E| - 1$. Then $\alpha$ is called a primitive element for $E$ if

$$\alpha^n = 1 \quad \text{and} \quad E \setminus \{0\} = \{ \alpha^i \mid 0 \leq i \leq n - 1 \}.$$
Lemma VII.48. Every finite field has a primitive element.

Proof. For a proof see A.18 in the appendix.

Definition VII.49. Let \( E \) be a finite field containing \( \mathbb{F}_2 \). Put \( n := |E| - 1 \) and let \( \alpha \) be a primitive element for \( E \). Let \( 1 \leq d \leq n - 1 \) and put

\[
g = \text{lcm}(m_{\alpha^i} | 1 \leq i \leq d).
\]

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 \( d + 1 \) with respect to \( \alpha \).

Lemma VII.50. Let \( F \) be a field and \( \beta_j, 1 \leq j \leq e \) be pairwise distinct non-zero elements in \( F \). Let \( 1 \leq d \leq e \) and let \( A \) be the \( d \times e \) matrix

\[
A := [\beta_j^i]_{1 \leq i \leq d} = \begin{bmatrix}
\beta_1 & \beta_2 & \ldots & \beta_e \\
\beta_1^2 & \beta_2^2 & \ldots & \beta_e^2 \\
\vdots & \vdots & \ddots & \vdots \\
\beta_1^d & \beta_2^d & \ldots & \beta_e^d
\end{bmatrix}
\]

Then any \( d \) columns of \( A \) are linearly independent over \( F \).

Proof. Replacing \( A \) by \( d \) of its columns we may assume that \( e = d \). Put

\[
B := [\beta_j^i]_{0 \leq i \leq d-1} = \begin{bmatrix}
1 & 1 & \ldots & 1 \\
\beta_1 & \beta_2 & \ldots & \beta_d \\
\vdots & \vdots & \ddots & \vdots \\
\beta_1^{d-1} & \beta_2^{d-1} & \ldots & \beta_{d-1}^{d-1}
\end{bmatrix}.
\]

Then \( \beta_j \cdot \text{Col}_j(B) = \text{Col}_j(A) \). Since \( \beta_j \neq 0 \) for all \( 1 \leq j \leq d \), the columns of \( B \) are linearly independent if and only if columns 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 \( B^T d = 0 \). Note that

\[
B^T = [\beta_j^i]_{0 \leq i \leq d-1}.
\]

By VII.45, \( B^T c = 0 \) if and only if \( \beta_1, \beta_2, \ldots, \beta_d \) are roots of \( c(x) \). Observe that a non-zero polynomial of degree less or equal to \( d - 1 \) has at most \( d - 1 \) roots. Since \( c(x) \) is a polynomial of degree at most \( d - 1 \) and since \( \beta_1, \beta_2, \ldots, \beta_d \) are \( d \) distinct 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 \( c = 0 \). \( \square \)
Theorem VII.51. Let \( \mathbb{E} \) be a finite field containing \( \mathbb{F}_2 \). Put \( n := |\mathbb{E}| - 1 \) and let \( 1 \leq d \leq n - 1 \). Let \( C \) be the BCH-code of length \( n \) and designated distance \( d + 1 \) with respect to the primitive element \( \alpha \) in \( \mathbb{E} \). Let
\[
\{ m_{\alpha_i} | 1 \leq i \leq d \} = \{ m_1, \ldots, m_s \}.
\]
where the \( m_i \)'s, \( 1 \leq i \leq s \) are pairwise distinct.

(a) Let \( g \) be the canonical generator of \( C \), then \( g = m_1 m_2 \ldots m_s \).

(b) Put \( H := [\alpha^j]_{0 \leq j \leq n-1} \). Then \( H \) is a check matrix for \( C \) over \( \mathbb{E} \) and
\[
C = \{ c \in \mathbb{F}_2^n | c(\alpha^i) = 0 \text{ for all } 1 \leq i \leq d \}.
\]

(c) For \( 1 \leq i \leq s \) let \( \alpha_i \) be a root of \( m_i \) in \( \mathbb{E} \). Put \( \tilde{H} := [\alpha_i^j]_{0 \leq j \leq n-1} \). Then \( \tilde{H} \) is a check matrix for \( C \) and
\[
C = \{ c \in \mathbb{F}_2^n | c(\alpha_i) = 0 \text{ for all } 1 \leq i \leq s \}.
\]

(d) \( C \) has minimum distance at least \( d + 1 \).

(e) \( \dim C \geq n - \left\lfloor \frac{d}{2} \right\rfloor \log_2(n+1) \).

(f) Suppose \( d \) is even and let \( d = 2r \) with \( r \in \mathbb{Z} \). Then \( C \) is an \( r \)-error correcting code and \( \dim C \geq n - r \log_2(n+1) \).

Proof. (a) By definition of a BCH-code we have \( g = \text{lcm}(m_{\alpha_i} | 1 \leq i \leq d) \). Since
\[
\{ m_{\alpha_i} | 1 \leq i \leq d \} = \{ m_1, \ldots, m_s \}
\]
we get \( g = \text{lcm}(m_i | 1 \leq i \leq s) \). As the \( m_i \) are pairwise distinct monic irreducible polynomials we conclude from \( \text{A.13} \) that \( g = m_1 m_2 \ldots m_s \).

(b) We will verify conditions \( \text{VII.46} \) and \( \text{(ii)} \) are fulfilled for \( \alpha_i = \alpha^i \):

(i): Let \( 1 \leq i \leq d \). Choose \( 1 \leq j \leq s \) with \( m_j = m_{\alpha^i} \). As \( \alpha^i \) is a root of \( m_{\alpha^i} \) we conclude that \( m_j(\alpha^i) = 0 \).

(ii): Let \( 1 \leq j \leq s \). Then \( m_j = m_{\alpha^i} \) for some \( 1 \leq i \leq d \) and so \( m_j(\alpha^i) = 0 \).

Thus we can apply \( \text{VII.46} \) and so \( \text{(ii)} \) holds.

(c) Note that \( m_i(\alpha_i) = 0 \) for all \( 1 \leq i \leq s \). So both conditions in \( \text{VII.46} \) are fulfilled (with \( d = s \)) and thus \( \text{(c)} \) holds.

(d) By \( \text{VII.50} \) applied with \( \beta_j = \alpha^{j-1} \), \( 1 \leq j \leq n \), any \( d \)-columns of \( H \) are linearly independent over \( \mathbb{E} \). Hence the sum of any \( d \)-columns of \( H \) is non-zero and so \( Hc \neq 0 \) for any \( 0 \neq c \in \mathbb{F}_2^n \) with \( \text{wt}(c) \leq d \). Thus \( C \) has minimum weight at least \( d + 1 \).
Put \( l := \log_2(n + 1) \). Then \(|E| = n + 1 = 2^l\). Put \( t_i := \deg m_i\). By \text{(VI.43)} we have \( \dim_{F_2} \mathbb{F}_2[\alpha_i] = t_i \) and so \(|\mathbb{F}_2[\alpha_i]| = 2^{t_i}\). Since \(|\mathbb{F}_2[\alpha_i]| \leq |E| = 2^l\) we have \( t_i \leq l \). Thus

\[
\deg g = \deg(m_1 m_2 \ldots m_s) = \sum_{i=1}^{s} \deg m_i = \sum_{i=1}^{s} t_i \leq s l = s l.
\]

By \text{(VI.27)} we have \( \dim C = n - \deg g \geq n - sl \).

Since \( \alpha^j \) is a root of \( m_{\alpha_j} \) and \( \alpha^{2j} = (\alpha^j)^2 \), also \( \alpha^{2j} \) is a root of \( m_{\alpha_j} \) (see Exercise 5 on Homework 6). Hence \( m_{\alpha_j} \) is an irreducible monic polynomials with root \( \alpha^{2j} \). By \text{(VI.43)} \( m_{\alpha_j} \) is the unique such polynomial, so \( m_{\alpha_j} = m_{\alpha^j} \). If \( i = 2k + j \) with \( j \) odd we conclude that \( m_{\alpha^j} = m_{\alpha^i} \). Therefore,

\[
\{m_{\alpha^i} \mid 1 \leq i \leq d\} = \{m_{\alpha^j} \mid 1 \leq j \leq d, j \text{ odd}\}
\]

If \( d \) is even, the number of odd integers \( j \) with \( 1 \leq j \leq d \) is \( \frac{d}{2} \) and if \( d \) is odd the number of such integers is \( \frac{d+1}{2} \). So in either case the number of such integers is \( \left\lceil \frac{d}{2} \right\rceil \). Hence \( s \leq \left\lceil \frac{d}{2} \right\rceil \). Thus

\[
\dim C \geq n - sl \geq n - \left( \left\lceil \frac{d}{2} \right\rceil \right) \log_2(n + 1).
\]

Suppose that \( d = 2r \). By \text{(d)} \( C \) has minimum distance at least \( d = 2r + 1 \) and so \( C \) is \( r \)-error-correcting. Also \( \left\lceil \frac{d}{2} \right\rceil = r \) and so by \text{(c)} \( \dim C \geq n - r \log_2(n + 1) \).

**Example VII.52.** Put \( \mathbb{E} := \mathbb{F}_2^{x^4 + x + 1}[x] \) and \( \alpha := x \in \mathbb{E} \). Let \( C \) be the BCH-code of length 15 and designated distance 7 with respect to \( \alpha \). Verify that \( \mathbb{E} \) is a field and \( \alpha \) 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 for \( C \) over \( \mathbb{E} \).

To show that \( \mathbb{E} \) is a field it suffices to show that \( x^4 + x + 1 \) is irreducible in \( \mathbb{F}_2[x] \), (see Lemma \text{[VII.38]}.)

Suppose that \( x^4 + x + 1 \) is reducible. Then \( x^4 + x + 1 = pq \) where \( p, q \) are non-constant polynomials in \( \mathbb{F}_2[x] \). Note that \( \deg p + \deg q = \deg pq = 4 \). Hence \( \deg p \leq 2 \) or \( \deg q \leq 2 \). Without loss \( \deg p \leq 2 \). 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 \( x^2 \). It follows that \( p = x^2 + x + 1 \). Hence \( \deg q = 2 \) and by symmetry, \( q = x^2 + x + 1 \). Thus

\[
x^4 + x + 1 = pq = (x^2 + x + 1)^2 = x^4 + x^2 + 1,
\]
a contradiction. So \( x^4 + x + 1 \) is irreducible.

In order to shows that \( \alpha \) is a primitive element for \( \mathbb{E} \), we will compute the powers of \( \alpha \). Recall first from \text{[VII.38]} that \( \alpha \) is a root of \( x^4 + x + 1 \). So \( \alpha^4 = -(1 + \alpha) = 1 + \alpha \).
\[\begin{align*}
\alpha^0 &= 1 \\
\alpha^1 &= \alpha \\
\alpha^2 &= \alpha^2 \\
\alpha^3 &= \alpha^3 \\
\alpha^4 &= 1 + \alpha \\
\alpha^5 &= \alpha + \alpha^2 \\
\alpha^6 &= \alpha^2 + \alpha^3 \\
\alpha^7 &= 1 + \alpha + \alpha^3 \\
\alpha^8 &= 1 + \alpha^2 \\
\alpha^9 &= \alpha + \alpha^3 \\
\alpha^{10} &= 1 + \alpha + \alpha^2 \\
\alpha^{11} &= 1 + \alpha^2 + \alpha^3 \\
\alpha^{12} &= 1 + \alpha + \alpha^2 + \alpha^3 \\
\alpha^{13} &= 1 + \alpha^2 + \alpha^3 \\
\alpha^{14} &= 1 + \alpha^3 \\
\alpha^{15} &= 1
\end{align*}\]

Hence \(\alpha\) is a primitive element. Since \(C\) has designated distance 7, \(d = 6\). To find the canonical generator \(g\) we need to determine \(m_{\alpha^i}\) for \(1 \leq i \leq 6\).

\(\alpha, \alpha^2\) and \(\alpha^4\) are roots of \(1 + x + x^4\) and so \(m_{\alpha} = m_{\alpha^2} = m_{\alpha^4}\).

We now introduce a method to compute \(m_{\beta}\) for \(\beta \in \mathbb{F}_2\). Note that \(\beta = b_0 + b_1\alpha + b_2\alpha^2 + b_3\alpha^3\) for some \(b_0, b_1, b_2, b_3 \in \mathbb{F}_2\). We call \(b_0b_1b_2b_3\) the string associated to \(\beta\) and write \(\beta \leftrightarrow b_0b_1b_2b_3\).

Let \(s_i\) be the string associated to \(\beta^i\). Let \(l \in \mathbb{N}\) be minimal such that

\[a_0s_0 + a_1s_2 + \ldots + a_\ell s_\ell = 0\]

for some \(a_0, \ldots, a_\ell \in \mathbb{F}_2\) with \(a_\ell = 1\). Then \(l\) is also minimal in \(\mathbb{N}\) with respect to

\[a_0 + a_1\beta + \ldots + a_\ell \beta^\ell = 0\]

for some \(a_0, \ldots, a_\ell \in \mathbb{F}_2\) with \(a_\ell = 1\), so \(m_{\beta} = a_0 + a_1x + \ldots + a_\ell x^\ell\).

Hence we can compute \(m_{\beta}\) by computing the strings \(s_0, s_1, s_2, \ldots\) associated to \(1, \beta, \beta^2, \ldots\) until we reach \(l \in \mathbb{N}\) such that \(s_0, s_2, \ldots, s_l\) are linearly dependent.

For \(\beta = \alpha^3\) we have

\[\begin{align*}
(\alpha^3)^0 &= 1 \quad \leftrightarrow \quad 1000 \\
(\alpha^3)^1 &= \alpha^3 \quad \leftrightarrow \quad 0001 \\
(\alpha^3)^2 &= \alpha^6 \quad = \quad \alpha^2 + \alpha^3 \quad \leftrightarrow \quad 0011 \\
(\alpha^3)^3 &= \alpha^9 \quad = \quad \alpha \quad + \alpha^3 \quad \leftrightarrow \quad 0101 \\
(\alpha^3)^4 &= \alpha^{12} \quad = \quad 1 + \alpha + \alpha^2 + \alpha^3 \quad \leftrightarrow \quad 1111
\end{align*}\]

The first four strings are linearly independent. The sum of all five is zero.

So
VII.3. PROPERTIES OF BCH CODES

\[ m_{\alpha^3} = 1 + x + x^2 + x^3 + x^4 \quad \text{and} \quad m_{\alpha^6} = m_{\alpha^3} = 1 + x + x^2 + x^3 + x^4. \]

For \( \beta = \alpha^5 \) we get

\[
\begin{align*}
(\alpha^5)^0 &= 1 \leftrightarrow 1000 \\
(\alpha^5)^1 &= \alpha + \alpha^2 \leftrightarrow 0110 \\
(\alpha^5)^2 &= \alpha^{10} = 1 + \alpha + \alpha^2 \leftrightarrow 1110
\end{align*}
\]

The first two strings are linearly independent and the sum of all three is zero. So

\[ m_{\alpha^5} = 1 + x + x^2. \]

Thus we can choose \( m_1 = 1 + x + x^4 \), \( m_2 = 1 + x + x^2 + x^3 + x^4 \) and \( m_3 = 1 + x + x^2 \). Hence

\[ g = m_1m_2m_3 = (1 + x + x^4)(1 + x + x^2 + x^3 + x^4)(1 + x + x^2) \]

We compute

\[
\begin{array}{cccccccccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 + x + x^2 + x^3 + x^4 & \leftrightarrow & 1 & 1 & 1 & 1 & 1 & \cdot 1 \\
& & 1 & 1 & 1 & 1 & 1 & \cdot x \\
& & 1 & 1 & 1 & 1 & 1 & \cdot x^4 \\
& & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & \cdot 1 \\
& & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & \cdot x \\
& & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & \cdot x^2 \\
g & \leftrightarrow & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1
\end{array}
\]

Thus \( g = 1 + x + x^2 + x^4 + x^5 + x^8 + x^{10} \). Since \( \deg g = 10 \), \( \dim C = 15 - 10 = 5 \). Observe that the string corresponding to \( g \) is a codeword and has weight 7. So \( C \) has minimum distance at most 7. By \([\text{VII.51}]\) \( C \) has minimum weight at least 7 and so \( \delta(C) = 7 \).

Since \( \alpha, \alpha^3 \) and \( \alpha^5 \) are roots of \( m_1, m_2 \) and \( m_3 \) respectively we obtain the following check matrix for \( C \) over \( \mathbb{E} \):
Lemma VII.53. Let \( C \subseteq \mathbb{F}_2^n \) be a BCH code with respect to the primitive element \( \alpha \). Let \( (a, z) \) be a 1-bit error for \( C \). Then \( z(\alpha) \neq 0 \). Moreover, if \( i \in \mathbb{N} \) is the unique element with \( 0 \leq i < n \) and \( z(\alpha) = \alpha^i \), then \( a_i \neq z_i \).

Proof. Since \( (a, z) \) is 1-bit error, \( a_j \neq z_j \) for a unique \( 0 \leq j < n \). Thus \( z(x) = c(x) + x^j \). By VII.46 \( c(\alpha) = 0 \) for all \( c \in C \) and so

\[
z(\alpha) = c(\alpha) + \alpha^j = \alpha^j.
\]

Since \( \alpha \) 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 VII.54. Let \( C \) be the code from example VII.52. Does there exist a 1-bit error \( (a, z) \) with

\[
z = 0111111100000000?
\]

We have \( z(\alpha) = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 + \alpha^6 + \alpha^7 \). We calculate the remainder of the corresponding polynomial when divided by \( x^4 + x + 1 \).

\[
\begin{array}{l}
111 \\
1011111110 \\
10011 \\
1100110 \\
10011 \\
101010 \\
10011 \\
1100 \\
\end{array}
\]

So the remainder is \( x^3 + x^2 \). Thus \( z(\alpha) = \alpha^3 + \alpha^2 = \alpha^2(1 + \alpha) = \alpha^2 \alpha^4 = \alpha^6 \). Hence, if \( (a, z) \) is a 1-bit error, then the error occurred at \( z_6 \). So

\[
a = 0111110100000000.
\]

We compute \( a(\alpha^3) \) and \( a(\alpha^5) \) to check whether \( a \) is in the code:

\[
a(\alpha^5) = 0 + \alpha^5 + \alpha^{10} + 1 + \alpha^5 + \alpha^{10} + 0 + \alpha^5 = 1 + \alpha^5 \neq 0
\]

So \( a \) is not in the code. Thus \( z \) cannot be the result of a 1-bit-error.
Chapter VIII

The RSA cryptosystem

VIII.1 Public-key cryptosystems

Definition VIII.1. A cryptosystem $\Omega$ 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 $D_k^* \circ E_k = \text{id}_{\mathcal{M}}$.

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 $D_k$ is called a decryption function. If $k, k^* \in \mathcal{K}$ with $D_k^* \circ E_k = \text{id}_{\mathcal{M}}$, then $k^*$ is called an inverse key for $k$.

Example VIII.2. Let $\mathcal{M} = \mathcal{C} = \mathcal{A} = \{A, B, C, D, \ldots, Z, \}$, $\mathcal{K} = \{0, 1, \ldots, 25\}$ 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.21. Note that $D_{26-k}$ is the inverse of $E_k$, so this is indeed a cryptosystem.

Definition VIII.3. A public-key cryptosystem is pair $(\Omega, \xi)$ where $\Omega$ is a cryptosystem and $\xi$ is a function

$$\xi : A \to \mathcal{K} \times \mathcal{K}, \ a \to (k_a, k_a^*)$$

such that, for all $a \in A$, $k_a^*$ is an inverse key for $k_a$. $k_a$ is called a public key and $k_a^*$ a private key.

In public key cryptography, all the ingredients except the private 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 $\xi$. 

139
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.

VIII.2 The Euclidean Algorithm

Lemma VIII.4. 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 \( a \) and \( b \) and so \( e \leq d \). Hence \( e = d \). \( \square \)

Theorem VIII.5 (Euclidean Algorithm). Let \( a \) and \( b \) be integers and let \( E_{-1} \) and \( E_0 \) be the equations

\[
E_{-1} : a = 1a + 0b \\
E_0 : b = 0a + 1b
\]

and suppose inductively we defined equation \( E_k, -1 \leq k \leq i \) of the form

\[
E_k : r_k = x_k a + y_k b
\]

If \( r_i \neq 0 \), let \( E_{i+1} \) be equation obtained by subtracting \( q_{i+1} \) times equation \( E_i \) from \( E_{i-1} \) where \( q_{i+1} \) is the integer quotient of \( r_{i-1} \) when divided by \( r_i \) (so \( q_{i+1} = \lfloor r_{i-1} / r_i \rfloor \)). Let \( m \in \mathbb{N} \) be minimal with \( r_m = 0 \) and put \( d = r_{m-1}, x = x_{m-1} \) and \( y = y_{m-1} \). Then

(a) \( \gcd(a, b) = |d| \)

(b) \( x, y \in \mathbb{Z} \) and \( d = xa + yb \).

Proof. Observe that \( r_{i+1} = r_{i-1} - q_{i+1} r_i, x_{i+1} = x_{i-1} - q_{i+1} x_i \) and \( y_{i+1} = y_{i-1} - q_{i+1} x_i \). So inductively \( r_{i+1}, x_{i+1}, y_{i+1} \) are integers and \( r_{i+1} \) is the remainder of \( r_{i-1} \) when divided by \( r_i \). So \( r_{i+1} < |r_i| \) and the algorithm will terminate in finitely many steps.

From \( r_{i-1} = q_{i+1} r_i + r_{i+1} \) and VIII.4 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 \( E_{m-1} \). \( \square \)
Example VIII.6. Let \( a = 1492 \) and \( b = 1066 \). Then

\[
\begin{align*}
1492 &= 1 \cdot 1492 + 0 \cdot 1066 \\
1066 &= 0 \cdot 1492 + 1 \cdot 1066 & q_1 = 1 \\
426 &= 1 \cdot 1492 - 1 \cdot 1066 & q_2 = 2 \\
214 &= -2 \cdot 1492 + 3 \cdot 1066 & q_3 = 1 \\
212 &= 3 \cdot 1492 - 4 \cdot 1066 & q_4 = 1 \\
2 &= -5 \cdot 1492 + 7 \cdot 1066 & q_5 = 106 \\
0 &= & &
\end{align*}
\]

So \( \gcd(1492, 1066) = 2 \) and \( 2 = -5 \cdot 1492 + 7 \cdot 1066 \).

Definition VIII.7. 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 ‘\( \equiv \) (mod \( n \))’ on \( \mathbb{Z} \) is defined by

\[
a \equiv b \pmod{n} \iff n \mid b - a
\]

Lemma VIII.8. Let \( a, b, a', b', n \in \mathbb{Z} \) with \( n > 0 \). If

\[
a \equiv a' \pmod{n} \quad \text{and} \quad b \equiv b' \pmod{n}
\]

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

\[
a'b' - ab = (a + kn)(b + ln) - ab = (al + kb + kl)n
\]

So \( n \) divides \( a'b' - ab \) the lemma holds.
Lemma VIII.9. Let \( n \in \mathbb{Z}^+ \). Let \( d \in \mathbb{Z} \) with \( \gcd(d, n) = 1 \). Then there exists \( e \in \mathbb{Z}_n^* \) with \( de \equiv 1 \pmod{n} \).

Proof. By the Euclidean algorithm there exist \( r, s \in \mathbb{Z} \) with

\[
1 = \gcd(d, n) = rd + sn.
\]

Hence \( rd \equiv 1 \pmod{n} \). Put \( e = [r]_n \). Then \( 0 \leq e < n \) and so \( e \in \mathbb{Z}_n \). Note that \( e \equiv r \pmod{n} \) and so by \textbf{VIII.8} \( ed \equiv rd \equiv 1 \pmod{n} \).

In particular, \( n \mid ed - 1 \) and any divisor of \( n \) and \( e \) will divide 1. Thus \( \gcd(n, e) = 1 \) and so \( e \in \mathbb{Z}_n^* \).

Lemma VIII.10. 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 \textbf{VIII.4}.

Lemma VIII.11. Let \( n \in \mathbb{Z}^+ \), \( x, y \in \mathbb{Z} \), \( a, b \in \mathbb{Z}_n \) and \( d \in \mathbb{Z} \) with \( \gcd(d, n) = 1 \).

(a) If \( a \equiv b \pmod{n} \), then \( a = b \).

(b) If \( xd \equiv yd \pmod{n} \), then \( x \equiv y \pmod{b} \)

(c) 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 \textbf{VIII.9} there exists \( e \in \mathbb{Z}_n^* \) with \( de \equiv 1 \pmod{n} \). Since \( xd \equiv yd \pmod{n} \) have \( exd \equiv eyd \pmod{n} \) and so also \( (de)x \equiv (de)y \pmod{n} \). As \( de \equiv 1 \pmod{n} \) this gives Thus \( x \equiv y \pmod{n} \).

(c) From \( [ad]_n = [bd]_n \) we get \( ad \equiv bd \pmod{n} \). Hence \( c \) shows that \( a \equiv b \pmod{n} \) and then \( a \) gives \( a = b \).

Lemma VIII.12. Let \( n, m \in \mathbb{Z}^+ \) with \( \gcd(n, m) = 1 \). Then \( \phi(nm) = \phi(n)\phi(m) \).

Proof. Consider the map

\[
\alpha : \mathbb{Z}_{nm} \rightarrow \mathbb{Z}_n \times \mathbb{Z}_m, a \rightarrow ([a]_n, [a]_m)
\]

We claim that \( \alpha \) 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 \textbf{VIII.11}. So \( \alpha \) is 1-1. Since \( |\mathbb{Z}_{nm}| = nm = |\mathbb{Z}_n \times \mathbb{Z}_m| \), \( \alpha \) is also onto.

Since \( \gcd(n, m) = 1 \), \( \gcd(a, n) = \gcd(a, m) \). Hence \( \gcd(a, nm) = 1 \) if and only if \( \gcd(a, n) = 1 \) and \( \gcd(a, m) \) is 1. By \textbf{VIII.10} this holds if and only if \( \gcd([a]_n, n) = 1 \) and
VIII.2. THE EUCLIDEAN ALGORITHM

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 \( \alpha \) induces a bijection

\[ \alpha^*: \mathbb{Z}_{nm}^* \rightarrow \mathbb{Z}_n^* \times \mathbb{Z}_m^*, \quad a \rightarrow ([a]_n, [a]_m). \]

So

\[ \phi(nm) = |\mathbb{Z}_{nm}^*| = |\mathbb{Z}_n^* \times \mathbb{Z}_m^*| = \phi(n)\phi(m). \]

**Corollary VIII.13.** Let \( p \) and \( q \) be distinct primes. Then

\[ \phi(p) = p - 1 \quad \text{and} \quad \phi(pq) = (p - 1)(q - 1) = pq + 1 - (p + q). \]

**Proof.** Note that \( \mathbb{Z}_p^* = \{1, 2, \ldots, p - 1\} \) and so \( \phi(p) = p - 1 \). By VII.12 \( \phi(pq) = \phi(p)\phi(q) = (p - 1)(q - 1) = pq + 1 - (p + q) \).

**Lemma VIII.14.** 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 \( \phi(p) = p - 1 \). By VII.10 also \( \gcd(\overline{ab}, n) = 1 \). Thus \( \overline{ab} \in \mathbb{Z}_n^* \) and we obtain a well-defined function

\[ \alpha: \mathbb{Z}_n^* \rightarrow \mathbb{Z}_n^*, \quad b \mapsto \overline{ab}. \]

Let \( b, c \in \mathbb{Z}_n^* \) with \( \alpha(b) = \alpha(c) \). Then \( \overline{ab} = \overline{ac} \) and VIII.11 shows that \( b = c \). Thus \( \alpha \) is 1-1.

As \( \alpha \) is function from a finite set to itself, this shows that \( \alpha \) is bijection.

It follows that

\[ \prod_{b \in \mathbb{Z}_n^*} b = \prod_{b \in \mathbb{Z}_n^*} \overline{ab}. \]

Since \( \overline{ab} \equiv ab \pmod{n} \) we conclude from VIII.8 that

\[ \prod_{b \in \mathbb{Z}_n^*} b \equiv \prod_{b \in \mathbb{Z}_n^*} \overline{ab} \equiv \prod_{b \in \mathbb{Z}_n^*} ab = \prod_{b \in \mathbb{Z}_n^*} a \prod_{b \in \mathbb{Z}_n^*} b \equiv a^{\phi(n)} \prod_{b \in \mathbb{Z}_n^*} b \pmod{n}. \]

Put \( e := \prod_{b \in \mathbb{Z}_n^*} b \). Then the above equation reads

\[ 1e \equiv e \equiv a^{\phi(n)}e \pmod{n}. \]

Observe that \( \gcd(e, n) = 1 \) and so VIII.11 implies that

\[ 1 \equiv a^{\phi(n)} \pmod{n}. \]
Chapter VIII. The RSA Cryptosystem

VIII.3 Definition of the RSA public-key cryptosystem

Definition VIII.15. Let

\( \mathcal{M} \) be an alphabet (plaintext messages)

Let \( c \in \mathbb{Z}^+ \). Put

\[
\mathcal{C} := \{ n \in \mathbb{Z}^+ | n \leq c \},
\]

\[
N := \{ n \in \mathcal{C} | \phi(n) \geq |\mathcal{M}| \}
\]

and

\[
\mathcal{K} := \{ (n, d) | n \in N, d \in \mathbb{Z}^*_\phi(n) \}.
\]

For \( n \in N \) let

\( \alpha_n : \mathcal{M} \to \mathbb{Z}^*_n \) be a code and \( \beta_n : \mathbb{Z}^*_n \to \mathcal{M} \) be a function

with

\[
\beta_n \circ \alpha_n = \text{id}_\mathcal{M}.
\]

Given a key \((n, d) \in \mathcal{K}\). Define

\( E_{n,d} : \mathcal{M} \to \mathcal{C}, \ m \mapsto \left[ \alpha_n(m)^d \right]_n \) (encryption functions)

\( D_{n,d} : \mathcal{M} \to \mathcal{C}, \ z \mapsto \beta_n\left( \left[ z^d \right]_n \right) \) (decryption functions)

Let \( \pi \) be a set of primes such that

\[
|\mathcal{M}| \leq (p - 1)^2 \leq c
\]

for all \( p \in \pi \). Define

\[
A := \{ (p, q, d, e) | p, q \in \pi, \ p \neq q, \ d, e \in \mathbb{Z}^*_\phi(pq), \ de \equiv 1 \pmod{\phi(pq)} \}
\]

and

\[
\xi : A \to \mathcal{K} \times \mathcal{K}, \quad (p, q, d, e) \mapsto ((pq, d), (pq, e))
\]

Then \((\mathcal{M}, \mathcal{C}, (E_{n,d})_{(n,d) \in \mathcal{K}}, (D_{n,d})_{(n,d) \in \mathcal{K}}, \xi)\) is called an RSA public-key cryptosystem.

Theorem VIII.16. Any RSA public-key cryptosystem is a public-key cryptosystem.
VIII.3. DEFINITION OF THE RSA PUBLIC-KEY CRYPTOSYSTEM

Proof. Let \((n, d) \in \mathcal{K}\). By definition of an RSA public-key cryptosystem we have \(d \in \mathbb{Z}_{\phi(n)}^*\) and so by VIII.9 there exists \(e \in \mathbb{Z}_{\phi(n)}^*\) with

\[ de \equiv 1 \pmod{\phi(n)}. \]

We will show \((n, e)\) is an inverse key for \((n, d)\), that is \(D_{n,e} \circ E_{n,d} = \text{id}_M\). For this let \(m \in \mathcal{M}\) and put \(w := \alpha_n(m)\). By definition of RSA public-key cryptosystem we have \(w \in \mathbb{Z}_n^*\) and \(E_{n,d}(m) = [w^d]_n\). Note that \(de = 1 + q\phi(n)\) for some \(q \in \mathbb{Z}\). Moreover, by VIII.14 \(w^{\phi(n)} \equiv 1 \pmod{n}\). Hence

\[ w^de \equiv w^{1+q\phi(n)} \equiv w \cdot (w^{\phi(n)})^q \equiv w1^q \equiv w \pmod{n}. \]

Thus

\[ \left([w^d]_n\right)^e \equiv \left[w^{de}\right]_n = w \]

and so

\[ D_{n,e}(E_{n,d}(m)) = \beta_n\left([E_{n,d}(m)^e]_n\right) = \beta_n\left([\left[w^d\right]_n^e]\right) = \beta_n(w) = \beta_n(\alpha_n(m)) = m. \]

So \((n, e)\) is an inverse key for \((n, d)\). This shows that any RSA public-key cryptosystem is cryptosystem.

Let \((p, q, d, e) \in A\) and put \(n := pq\). By definition of an RSA public-key cryptosystem we have \(\xi(p, q, d, e) = ((n, d), (n, e))\), \(d, e \in \mathbb{Z}_{\phi(n)}^*\) and \(de \equiv 1 \pmod{\phi(n)}\). As just seen, this means that \((n, e)\) is an inverse key to \((n, d)\). Hence any RSA public-key cryptosystem is a public-key cryptosystem.

Example VIII.17. Let

\[ \mathcal{M} = \mathcal{A} = \{\cup, A, \ldots, Z\} = \{l_0, l_1, \ldots, l_{26}\} \]

and \(c = 1000\). Let \(N = \{1, 2, \ldots, 1000\}\). For \(n \in N\) let

\[ \mathbb{Z}_n^* = \{a_{1,n}, \ldots, a_{\phi(n),n}\} \]

with \(a_{i,n} < a_{i+1,n}\) for \(1 \leq i < \phi(n)\). Define

\[ \alpha_n : \mathcal{M} \to \mathbb{Z}_n^*, \quad l_i \mapsto a_{i+1,n} \]

and

\[ \beta_n : \mathbb{Z}_n^* \to \mathcal{M}, \quad j \mapsto l_{[i-1]_{27}}, \quad \text{where} \quad 1 \leq i \leq \phi(n) \text{ with } j = a_{i,n} \]

Compute \(u := E_{(667,5)}(K)\). Find the inverse key \((667, e)\) for \((667, 5)\) and verify that \(D_{667,e}(u) = K\).
Recall first that $E_{667,5}(K) = [\alpha_{667}(K)^5]_{667}$. Note that $K = l_{11}$ and so $\alpha_{667} = a_{12,667}$ is the twelve positive integer coprime to 667. As 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 all are coprime to 667. Thus

$$\alpha_{667}(K) = 12$$

To compute $[12^5]_{667}$ we determined $12^5$ modulo 667

$$12^2 = 144$$
$$12^3 = \frac{1440}{1728}$$
$$12^4 \equiv 1728 - 3 \cdot 667 = -(2001 - 1728) = -273$$
$$12^4 \equiv -273 \cdot 12 = -\frac{2730}{3286}$$
$$12^4 \equiv -3286 + 5 \cdot 667 = 3335 - 3286 = 59$$
$$12^5 \equiv 59 \cdot 12 = 720 - 12 = 708$$
$$12^5 \equiv 708 - 667 = 41$$

Hence

$$E_{667,5}(K) = E_{667,5}(K) = [\alpha_{667}(K)^5]_{667} = [12^5]_{667} = 41.$$  

Recall from the proof of VIII.16 that $(667, e)$ will be an inverse key for $(667, 5)$ provided that $e \in \mathbb{Z}_{\phi(667)}$ with $5 \cdot e \equiv 1 \pmod{667}$. To compute $\phi(667)$ we need to factorize 667 as a product of primes. None of 3, 5, 11 divides 667, $667 - 7 = 660$, $667 - 13 = 680$, $667 - 17 = 650$, $667 + 23 = 690$. So $667 = 23 \cdot 29$. Thus $\phi(667) = 667 - (23 + 29) + 1 = 668 - 52 = 616$. Note that

$$1 = 616 - 123 \cdot 5$$

So $e = [-123]_{616} = 616 - 123 = 493$. A long calculation by hand or a quick computer calculation shows that $41^{493} \equiv 12 \pmod{667}$. Hence

$$D_{667,493}(41) = \beta_{667}([41^{493}]_{667}) = \beta_{667}(12) = l_{[12-1]}_{27} = l_{11} = K.$$
Chapter IX

Noisy Channels

IX.1 The definition of a channel

Definition IX.1. Let $I$ and $J$ be alphabets. An $I \times J$-channel is an $I \times J$-matrix $\Gamma = [\Gamma_{ij}]_{i \in I, j \in J}$ 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 $\Gamma$ and $J$ the output alphabet.

We interpret $\Gamma_{ij}$ as the probability that the symbol $j$ is received when the symbol $i$ is send through the channel $\Gamma$.

<table>
<thead>
<tr>
<th>$\Gamma$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Example IX.2. $b$ is a channel with input alphabet $\{a,b,c,d\}$ and output alphabet $\{a,b,c,d\}$.

<table>
<thead>
<tr>
<th>$\Gamma$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>$c$</td>
<td>0.7</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>$d$</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Lemma IX.3. Let $I$ and $J$ be alphabets and $\Gamma$ an $I \times J$-matrix with coefficients in $\mathbb{R}$. Then $\Gamma$ is a channel if and only if each row of $\Gamma$ 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 $i \in I$. 

147
Definition IX.4.  (a) The transpose of an $I \times J$-matrix $M = [m_{ij}]_{i \in I, j \in J}$ is the $J \times I$ matrix $M^{Tr} = [m_{ij}]_{j \in J, i \in I}$.

(b) An $I \times J$-channel $\Gamma$ is called symmetric if $\frac{|J|}{|I|} \Gamma^{Tr}$ is a channel. Note that this is the case if and only if $\sum_{j \in J} \Gamma_{ij} = |J|$ for all $j \in J$ and if and only if all columns of $\Gamma$ have the same sum.

(c) A binary symmetric channel BSC is a symmetric channel with input and output alphabet $\mathbb{B}$.

(d) Let $\Gamma$ be a binary symmetric channel. Then $e = \Gamma_{01}$ is called the bit error of $\Gamma$.

Lemma IX.5. Let $e$ be the bit error of a binary symmetric channel $\Gamma$. Then

$$
\Gamma = \begin{bmatrix}
0 & 1 \\
0 & 1-e & e \\
1 & e & 1-e 
\end{bmatrix}
$$

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

Notation IX.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

$$
\text{BSC}(e) = \begin{bmatrix}
1-e & e \\
e & 1-e
\end{bmatrix}
$$

Example IX.7. Consider the simplified keypad

```
  A  B  C
  D  E  F
```
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

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$D$</th>
<th>$E$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0.8</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$B$</td>
<td>0.1</td>
<td>0.7</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$C$</td>
<td>0</td>
<td>0.1</td>
<td>0.8</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$D$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>$E$</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.7</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$F$</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**IX.2 Transferring a source through a channel**

**Definition IX.8.** Let $I$ and $J$ be alphabets, let $\Gamma$ 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 $\Gamma$ if $q = p\Gamma$, that is $q_j = \sum_{i \in I} p_i \Gamma_{ij}$.

(b) $\text{Diag}(p)$ 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 IX.9.** Let $I$ and $J$ be alphabets, let $\Gamma$ 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

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

(a) $p$ is the marginal tuple of $t$ on $I$ if and only if $\Gamma_i$ is a probability distribution 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 $\Gamma$ is a channel.

(c) $q$ is linked to $p$ via $\Gamma$ if and only if $q$ is the marginal tuple of $t$ on $J$. 

Proof. (a) \[ p \text{ is the marginal distribution of } t \text{ on } I \]
\[ \iff \sum_{j \in J} t_{ij} = p_i \text{ for all } i \in I \]
\[ \iff \sum_{j \in J} p_i \Gamma_{ij} = p_i \text{ for all } i \in I \]
\[ \iff \sum_{j \in J} \Gamma_{ij} = 1 \text{ for all } i \in I \text{ with } p_i \neq 0 \]
\[ \iff \Gamma_i \text{ is a probability distribution on } J \text{ for all } i \in I \text{ with } p_i \neq 0 \]
So (a) holds.

(b) Follows from (a).

(c) \[ q \text{ is linked to } p \text{ via } \Gamma \]
\[ \iff q = p \Gamma \]
\[ \iff q_j = \sum_{i \in I} p_i \Gamma_{ij} \text{ for all } j \in J \]
\[ \iff q_j = \sum_{i \in I} t_{ij} \text{ for all } j \in J \]
\[ \iff q \text{ is the marginal distribution of } t \text{ on } J \]

Definition IX.10. Let I and J be alphabets.

(a) I \times J-channel system is a tuple \((\Gamma, t, p, q)\) such that \(\Gamma\) is a I \times J-channel, t, p and q are probability distribution on I \times J, I and J respectively, \(t = \text{Diag}(p)\Gamma\) and \(q = p\Gamma\).

(b) Let \(\Gamma\) be a I \times J-channel and p a probability distribution on T. Then \(t = \text{Diag}(p)\Gamma\) is called the joint distribution for \(\Gamma\) and p. \((\Gamma, t, p, p\Gamma)\) is called the Channel system for \(\Gamma\) 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 \(\Gamma\) an I \times J-channel, \(\Gamma\) is called a channel associated to t (and \((\Gamma, t, p, q)\) is called a channel system associated to t) if \(t = \text{Diag}(p)\Gamma\).

(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 \(\Pr^\Sigma(i, j)\).
- \(p_i\) is called probability that \(i\) is send, and is denoted by \(\Pr^\Sigma(i, \ast)\).
- \(q_j\) is called the probability that \(j\) is received, and is denoted by \(\Pr^\Sigma(\ast, j)\).
IX.2. TRANSMITTING A SOURCE THROUGH A CHANNEL

- \( \Gamma_{ij} \) is called the probability that \( j \) is received when \( i \) is send and is denoted by \( \text{Pr}^\Sigma(j|i) \).

Assuming that there is no doubt underlying channel system, we will usually drop the superscript \( \Sigma \).

**Example IX.11.** Compute the channel system for the channel BSC(\( e \)) and the probability distribution \( (p, 1-p) \).

\[
t = \text{Diag}(p, 1-p)) \text{BSC}(e)
= \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix} \begin{bmatrix} 1-e & e \\ e & 1-e \end{bmatrix}
= \begin{bmatrix} p(1-e) & pe \\ (1-p)e & (1-p)(1-e) \end{bmatrix}
= \begin{bmatrix} p - pe & pe \\ e - pe & 1 - p - e + pe \end{bmatrix}
\]

Since \( q \) is the column sum of \( t \):

\[
q = (p(1-e) + (1-p)e, pe + (1-p)(1-e)) = (p + e - 2pe, 1 + 2pe - p - e)
\]

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 IX.12. Let $I$ and $J$ be alphabets and let $t$ be a probability distribution, $p$ the marginal distribution of $t$ on $I$ and $\Gamma$ an $I \times J$-matrix.

(a) $\Gamma$ 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) $\Gamma_i$ 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 $\text{Diag}(p)^{-1} \Gamma$ is the unique channel associated to $t$.

Proof. (a) Suppose first that $\Gamma$ is a channel associated to $t$. Then $t = \text{Diag}(p) \Gamma$ and so $t_i = p_i \Gamma_i$. Thus (a:i) holds. Since $\Gamma$ 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 $t = \text{Diag}(p) \Gamma$. From IX.9 we conclude that $\Gamma_i$ is a probability distribution on $J$ for all $i \in I$ with $p_i \neq 0$. Together with (a:ii) this shows that $\Gamma$ is a channel.

(b) Let $\Gamma$ be the $I \times J$-matrix such that $\Gamma_i = \frac{1}{p_i} t_i$ if $p_i \neq 0$ and $\Gamma_i$ is the equal probability distribution on $J$ if $p_i = 0$. Then by (a) $\Gamma$ is a channel associated to $t$.

(c) Follows immediately from (a).

IX.3 Conditional Entropy

Definition IX.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 IX.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 $\Gamma$ be a $I \times J$-channel and $E$ a probability distribution on $J$. We say that $\Gamma$ is additive with row distribution $E$ if each row of $\Gamma$ is a permutation of $E$. 
Example IX.15. \((0,1,0,3,0,4,0,2)\) is a permutation of \((0.4,0.2,0.3,0.1)\).

Example IX.16. BSC\((e)\) is additive with row distribution \(E = (e, 1 - e)\).

Lemma IX.17. (a) Let \(p\) and \(p'\) be probability distributions. If \(p\) is a permutation of \(p'\), then \(H(p) = H(p')\).

(b) Let \((\Gamma, t, p, q)\) be a channel system. Suppose that \(\Gamma\) is additive with row distribution \(E\). Then \(t\) is a permutation of \(p \otimes E\) and

\[
\begin{align*}
H(t) &= H(p) + H(E) \quad H(p|q) = H(p) + H(E) - H(q) \\
H(q|p) &= H(E) \quad I(p,q) = H(q) - H(E)
\end{align*}
\]

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 \(p_i E\). Since \(\Gamma_i\) is a permutation of \(E\), also \(p_i \Gamma_i\) is a permutation of \(p_i E\). So \(t\) is a permutation of \(p \otimes E\). Thus by (a) and IV.11

\[
H(t) = H(p \otimes E) = H(p) + H(E)
\]

Hence

\[
H(p|q) = H(t) - h(q) = H(p) + H(E) - H(q),
\]

\[
H(q|p) = H(t) - h(p) = H(E)
\]

and

\[
I(p,q) = H(p) + H(q) - H(t) = H(q) - H(E)
\]

\(\square\)

Corollary IX.18. Let \((\text{BSC}(e), t, p, q)\) be a channel system. Then

\[
H(p|q) = H(p) + H((e, 1-e)) - H(q)
\]

IX.4 Capacity of a channel

Theorem IX.19. Let \((\Gamma, t, p, q)\) be a channel system. Then

\[
I(p, q) \geq 0
\]

with equality 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 \[ IV.11 \]

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 IX.20.** Let $\Gamma$ 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

$$\gamma(\Gamma) = \max_{p \in \mathcal{P}(I)} f_\Gamma(p)$$

Then $\gamma(\Gamma)$ is called the capacity of the channel $\Gamma$.

In little less precise notation

$$\gamma(\Gamma) = \max_p I(p, q)$$

**Theorem IX.21.** Let $\Gamma$ 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 $(\Gamma, t, p, q)$ the channel system for $\Gamma$ and $p$. So $q = p\Gamma$ and $t = \text{Diag}(p)\Gamma$. By \[ IX.17 \]

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

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. \[ \square \]

**Corollary IX.22.** Let $\Gamma$ be an symmetric, additive $I \times J$ channel with row distribution $E$. Then

$$\gamma(\Gamma) = \log |J| - H(E)$$
IX.4. CAPACITY OF A CHANNEL

Proof. Let \( p = \left( \frac{1}{|I|} \right)_{i \in I} \) be the equal probability distribution on \( I \). Let \( j \in J \). We compute

\[
q_j = \sum_{i \in I} p_i \Gamma_{ij} = \frac{1}{|I|} \sum_{i \in I} \Gamma_{ij} = \frac{1}{|I||J|} \frac{|I|}{|J|} = \frac{1}{|J|}
\]

where the second equality holds since \( \Gamma \) is symmetric. So \( q \) is the equal probability distribution on \( J \). Since \( \Gamma \) is additive, the Corollary now follows from IX.21.

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

Proof. Since BSC(\( e \)) is a symmetric, additive channel with row distribution \((e, 1-e)\) and output alphabet of size 2, this follows immediately from IX.22.

Lemma IX.24. Let \((\Gamma, t, p, q)\) 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

\[
p_i = \sum_{j \in J} t_{ij}
\]

Also

\[
\frac{p_i}{t_{ij}} = \frac{p_i}{p_i \Gamma_{ij}} = \frac{1}{\Gamma_{ij}}.
\]
We compute

\[ H(q|p) = H(t) - H(p) \]
\[ = \sum_{(i,j) \in I \times J} t_{ij} \log \left( \frac{1}{t_{ij}} \right) - \sum_{i \in I} p_i \log \left( \frac{1}{p_i} \right) \]

\[ = \sum_{(i,j) \in I \times J} t_{ij} \log \left( \frac{1}{t_{ij}} \right) - \sum_{i \in I} \left( \sum_{j \in J} t_{ij} \right) \log \left( \frac{1}{p_i} \right) \]

\[ = \sum_{(i,j) \in I \times J} t_{ij} \log \left( \frac{p_i}{t_{ij}} \right) \]

\[ = \sum_{(i,j) \in I \times J} t_{ij} \log \left( \frac{1}{\Gamma_{ij}} \right) \]

\[ = \sum_{i \in I} p_i \Gamma_{ij} \log \left( \frac{1}{\Gamma_{ij}} \right) \]

\[ = \sum_{i \in I} p_i \left( \sum_{j \in J} \Gamma_{ij} \log \left( \frac{1}{\Gamma_{ij}} \right) \right) \]

\[ = \sum_{i \in I} p_i H(\Gamma_i) \]
Chapter X

The noisy coding theorems

X.1 The probability of a mistake

Definition X.1. Let $I$ and $J$ be alphabets.

(a) An $I \times J$-decision rule is a function $\sigma : J \rightarrow I$.

(b) Let $i \in I$ and $j \in J$. Then $(i, j)$ is called a mistake for $\sigma$ if $i \neq \sigma(j)$.

(c) An $I \times J$-decision system is a tuple $(\Gamma, t, p, q, \sigma)$, where $(\Gamma, t, p, q)$ is an $I \times J$-channel system and $\sigma$ an $I \times J$-decision rule.

(d) Let $\Gamma$ be an $I \times J$-channel, $p$ a probability distribution of $I$ and $\sigma$ and $I \times J$ decision rule.

Let $(\Gamma, t, p, q)$ be the channel system for $\Gamma$ and $p$. Then $(\Gamma, t, p, q, \sigma)$ is called the decision system for $\Gamma, p$ and $\sigma$ and is denoted by $\Sigma(\Gamma, p, \sigma)$.

Definition X.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) $M^\Sigma = \sum_{i \in I} p_i M^\Sigma(i)$. $M^\Sigma$ is called the probability of a mistake.

Of course we will often drop the superscripts.

Definition X.3 (Ideal Observer Rule). Let $\Sigma = (\Gamma, t, p, q, \sigma)$ be an $I \times J$-decision system. We say that $\sigma$ is an Ideal observer rule with respect to $\Sigma$ 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 X.4.** Find the Ideal Observer Rule for the channel BSC(0.3) and probability distribution $(0.2, 0.8)$. What is the probability of a mistake?

We have
\[
\Gamma = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \quad \text{and} \quad t = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.14 & 0.06 \\ 0.24 & 0.56 \end{bmatrix}
\]

Let $\sigma$ 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 X.5 (Maximum Likelihood Rule).** Let $\Sigma = (\Gamma, t, p, q, \sigma)$ be a $I \times J$-decision system. We say that $\sigma$ is a Maximum Likelihood Rule with respect to $\Gamma$ 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 X.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 \( \sigma \) be a Maximum Likelihood rule. The largest entry in the first column of \( \Gamma \) 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
\[ M = p_0M(0) + p_1M(1) = 0.2 \cdot 0.3 + 0.8 \cdot 0.3 = 0.3 \]

### X.2 Fano’s inequality

**Lemma X.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
\[ (I \times J) \setminus K = \{(\sigma(j), j) \mid j \in J\}, \]
\[ M = \sum_{(i,j) \in K} t_{ij} \quad \text{and} \quad 1 - M = \sum_{j \in J} t_{\sigma(j)j} \]

**Proof.** Note that \( K = \{(i, j) \in I \times J \mid \sigma(j) \neq i\} \). So
\[ (I \times J) \setminus K = \{(i, j) \in I \times J \mid i = \sigma(j)\} = \{(\sigma(j), j) \mid j \in J\} \]
and the first statement is proved.
We compute
\[
\sum_{(i,j) \in K} t_{ij} = \sum_{(i,j) \in I \times J} t_{ij} = \sum_{i \in I} \left( \sum_{j \in F(i)} t_{ij} \right) \\
= \sum_{i \in I} \left( \sum_{j \in F(i)} p_i \Gamma_{ij} \right) = \sum_{i \in I} p_i \left( \sum_{j \in F(i)} \Gamma_{ij} \right) \\
= \sum_{i \in I} p_i M(i) = M
\]
Thus the second statement holds. Since \( 1 = \sum_{(i,j) \in I \times J} t_{ij} \), the first two statement imply the third. \( \square \)

**Theorem X.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

\[
H(\Gamma; p) \leq H((M, 1 - M)) + M \log(|I| - 1)
\]

**Proof.** Let \(K\) be the sets of mistakes. By \(\text{X.7}\) \(M = \sum_{(i,j) \in K} t_{ij}\) and \(1 - M = \sum_{j \in J} t_{\sigma(j)j}\). Thus

\[
H((M, 1 - M)) = M \log \left( \frac{1}{M} \right) + (1 - M) \log \left( \frac{1}{1 - M} \right)
\]

Also by \(\text{X.7}\)

\[
I \times J = K \cup \{(\sigma(j), j) \mid j \in J\}.
\]

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

\[
H(\Gamma; p) = H(p \mid q) = \sum_{j \in J} q_j H(\Delta_j)
\]

\[
= \sum_{j \in J} \sum_{i \in I} q_j \Delta_{ji} \log \left( \frac{1}{\Delta_{ji}} \right) = \sum_{j \in J} \sum_{i \in I} \frac{t_{ij}}{q_j} \log \left( \frac{1}{\frac{t_{ij}}{q_j}} \right)
\]

\[
= \sum_{(i,j) \in I \times J} t_{ij} \log \left( \frac{q_j}{t_{ij}} \right) = \sum_{(i,j) \in K} t_{ij} \log \left( \frac{q_j}{t_{ij}} \right) + \sum_{j \in J} t_{\sigma(j)j} \log \left( \frac{q_j}{t_{\sigma(j)j}} \right)
\]

Put

\[
S_1 = \sum_{(i,j) \in K} t_{ij} \log \left( \frac{q_j}{t_{ij}} \right) - \sum_{(i,j) \in K} t_{ij} \log \left( \frac{1}{M} \right) = \sum_{(i,j) \in K} t_{ij} \log \left( \frac{M q_j}{t_{ij}} \right)
\]

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

\[
H(\Gamma; p) - H((M, 1 - M)) = S_1 + S_2.
\]

So it suffices to show \(S_1 \leq M \log(|I| - 1)\) and \(S_2 \leq 0\).
For \((i, j) \in K\) put
\[
v_{ij} = \frac{t_{ij}}{M} \quad \text{and} \quad w_{ij} = \frac{q_j}{|I| - 1}.
\]
Since \(\sum_{(i,j) \in K} t_{ij} = M\), \((v_{ij})_{(i,j) \in K}\) is a probability distribution on \(K\). Also
\[
\sum_{(i,j) \in K} v_{ij} = \sum_{j \in J} \sum_{i \in I, i \neq \sigma(j)} q_j = (|I| - 1) \sum_{j \in J} q_j = |I| - 1,
\]
and so also \((w_{ij})_{(i,j) \in K}\) is a probability distribution on \(K\). Thus by the Comparison Theorem IV.1
\[
0 \geq \sum_{(i,j) \in K} v_{ij} \log \left( \frac{1}{v_{ij}} \right) - \sum_{(i,j) \in K} v_{ij} \log \left( \frac{1}{w_{ij}} \right)
= \sum_{(i,j) \in K} v_{ij} \log \left( \frac{w_{ij}}{v_{ij}} \right)
= \sum_{(i,j) \in K} \frac{t_{ij}}{M} \log \left( \frac{q_j M}{t_{ij} (|I| - 1)} \right)
= \sum_{(i,j) \in K} \frac{t_{ij}}{M} \log \left( \frac{q_j M}{t_{ij}} \right) - \sum_{(i,j) \in K} \frac{t_{ij}}{M} \log \left( |I| - 1 \right)
= \frac{1}{M} S_1 + \log(|I| - 1)
\]
Thus indeed \(S_1 \leq M \log(|I| - 1)\).
For \(j \in J\), put
\[
u_j = \frac{t_{\sigma(j)j}}{1 - M}.
\]
Since \(\sum_{j \in J} t_{\sigma(j)j} = 1 - M\) both \(q\) and \((u_j)_{j \in J}\) are probability distributions on \(J\). So by the Comparison Theorem IV.1
\[
0 \geq \sum_{j \in J} u_j \log \left( \frac{1}{u_j} \right) - \sum_{j \in J} u_j \log \left( \frac{1}{q_j} \right)
= \sum_{j \in J} u_j \log \left( \frac{q_j}{u_j} \right)
= \sum_{j \in J} \frac{t_{\sigma(j)j}}{1 - M} \log \left( \frac{q_j (1 - M)}{t_{\sigma(j)j}} \right)
= \frac{1}{1 - M} S_2
\]
and so indeed \(S_2 \leq 0\).

**X.3  A lower bound for the probability of a mistake**

**Theorem X.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 \ref{x.8}

\[(*) \quad H(\Gamma;p) \leq H((M,1-M)) + M \log(|I| - 1) \leq \log 2 + M \log(|I| - 1) < 1 + M \log(|I|)\]

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 \ref{iv.2} \( H(p) = \log(|I|) \) and so

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

\( \Box \)

\section*{X.4 Extended Channels}

Let \( \Gamma \) be an \( I \times J \) channel and \( \Gamma' \) 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 \( \Gamma \) and \( i' \) via \( \Gamma' \). Then the probability that the pair of symbols \( jj' \) is received is \( \Gamma_{ij} \Gamma'_{ij'} \). So the combined channel \( \Gamma'' \) has input \( I \times I' \), output \( J \times J' \) and
\[ \Gamma_{ii',jj'}'' = \Gamma_{ij} \Gamma_{ij'}' \]

This leads to the following definitions:

**Definition X.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

\[ m''_{ii',jj'} = m_{ij} m'_{ij'} \]

\( M'' \) 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

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

**Example X.11.** Compute

\[ \begin{array}{cccc|cc}
  a & b & c & v & w \\
  d & 0 & 1 & 2 & x & 4 & 5 \\
  e & 3 & -1 & 0 & y & -1 & 3 \\
\end{array} \]

and

\[ \begin{bmatrix} 0.2 & 0.8 \\ 0.3 & 0.7 \end{bmatrix}^{\otimes 2} \]

<table>
<thead>
<tr>
<th>av</th>
<th>aw</th>
<th>bv</th>
<th>bw</th>
<th>cv</th>
<th>cw</th>
</tr>
</thead>
<tbody>
<tr>
<td>dx</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>dy</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>3</td>
<td>-2</td>
</tr>
<tr>
<td>ex</td>
<td>12</td>
<td>15</td>
<td>-4</td>
<td>-5</td>
<td>0</td>
</tr>
<tr>
<td>ey</td>
<td>-3</td>
<td>9</td>
<td>1</td>
<td>-3</td>
<td>0</td>
</tr>
</tbody>
</table>

and

\[ \begin{bmatrix} 0.04 & 0.16 & 0.16 & 0.64 \\ 0.06 & 0.14 & 0.24 & 0.56 \\ 0.06 & 0.24 & 0.14 & 0.56 \\ 0.09 & 0.21 & 0.21 & 0.49 \end{bmatrix} \]

**Lemma X.12.** Let \( \Gamma \) be an \( I \times J \)-channel and \( \Gamma' \) 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 \( i, i' \in I' \).

(b) \( \Gamma'' \) 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 \( \Gamma \).

(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. \( (a) \)
\[
(\Gamma''_{ii'})_{jj'} = \Gamma''_{ii'}^{jj'} = \Gamma'_{ij} \Gamma''_{ij} = (\Gamma_i)_{jj'} = \Gamma_i \otimes \Gamma_{ij}^{jj'}
\]

\( (b) \) Let \( i \in I \) and \( i' \in I' \). Since \( \Gamma \) and \( \Gamma' \) are channels, \( \Gamma_i \) is a probability distribution on \( J \) and \( \Gamma'_{ij} \) is a probability distribution on \( J' \). Thus by \((IV.10)\) \( \Gamma_i \otimes \Gamma_{ij}^{jj'} \) is a probability distribution on \( J'' = J \times J' \). So by \((a)\) all rows of \( \Gamma'' \) are probability distributions and hence \( \Gamma'' \) is a channel.

\( (c) \) Follows from \((b)\) and induction on \( n \).

\( (d) \) Can be proved using an easy induction argument. \( \square \)

**Lemma X.13.** Let \( \Gamma \) be an \( I \times J \)-channel and \( \Gamma' \) 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 \( H(\Gamma''_{ii'}) = H(\Gamma_i) + H(\Gamma'_{ii}) \).

\( (d) \) \( H(q''|p'') = H(q|p) + H(q'|p') \)

\( (e) \) \( \gamma(\Gamma'') = \gamma(\Gamma) + \gamma(\Gamma') \).

**Proof.** Since \( \Sigma'' \) is a channel system

\[
(1) \quad t''_{ii', jj'} = p''_{ii'} \Gamma''_{ii', jj'} = p''_{ii'} \Gamma'_{ij} \Gamma''_{ij}
\]

Since \( \Gamma' \) is a channel, \( \Sigma_{i \in I} \Gamma'_{ij} = 1 \). Also \( t \) and \( p \) are marginal distributions of \( t'' \) and \( p'' \).

So summing (1) over all \( i' \in I' \), \( j' \in J' \) gives

\[
t_{ij} = \sum_{i' \in I, j' \in J'} t''_{ii', jj'} = \left( \sum_{i' \in I'} \left( p''_{ii'} \sum_{j' \in J'} \Gamma'_{ij} \right) \right) \Gamma_{ij} = \left( \sum_{i' \in I'} p''_{ii'} \right) \Gamma_{ij} = p_i \Gamma_{ij}
\]

So \( t \) is the joint distribution of \( p \) and \( \Gamma \).

Since \( \Sigma'' \) 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 \( \Sigma \) is a channel system. By symmetry also \( \Sigma' \) is a channel system.

\( (b) \) Suppose that \( p \) and \( p' \) are independent with respect to \( p'' \). Then \( p''_{ii'} = p_i p'_{i'} \) and
q_{jj'} = \sum_{i \in I, i' \in I'} p_{ii'}^{ii'} \Gamma_{ii',jj'}^{ii'} = \sum_{i \in I, i' \in I'} p_{ii'} p_{i'i'}^i \Gamma_{ij}^i \Gamma_{i'j'}^{i'} = \left( \sum_{i \in I} p_i \Gamma_{ij}^i \right) \left( \sum_{i' \in I'} p_{i'} \Gamma_{i'j'}^{i'} \right) = q_j q_{j'}

Hence \( q'' = q \otimes q' \) and \( q \) and \( q' \) are independent with respect to \( q'' \).

(c) Let \( i \in I \) and \( i' \in I' \). By X.12(a),

\[ \Gamma_{ii'} = \Gamma_i \otimes \Gamma_{i'} \]

and so (c) follows from IV.11.

(d) By IX.24

\[ H(q''|p'') = \sum_{i \in I, i' \in I'} p_{ii'}^{ii'} H(\Gamma_{ii'}) \]

and so by (c)

\[ H(q''|p'') = \sum_{i \in I} \left( \sum_{i' \in I'} p_{ii'}^{ii'} \right) H(\Gamma_i) + \sum_{i' \in I'} \left( \sum_{i \in I} p_{ii'}^{ii'} \right) H(\Gamma_{i'}) = \sum_{i \in I} p_i H(\Gamma_i) + \sum_{i' \in I'} p_{i'} H(\Gamma_{i'}) \]

Two more applications of IX.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

\[ f_{\Gamma''}(p'') = I(p'', q'') = H(p'') + H(q'') - H(t'') \]

Since \( q \) and \( q' \) are the marginal distributions of \( q'' \) IV.11 gives

\[ f_{\Gamma''}(p'') = I(p'', q'') = H(p'') + H(q'') - H(t'') = H(q'') - H(q''|p''). \]

(2) \[ f_{\Gamma''}(p'') = I(p'', q'') = H(p'') + H(q'') - H(t'') = H(q'') - H(q''|p''). \]

Since \( q \) and \( q' \) are the marginal distributions of \( q'' \) IV.11 gives

\[ H(q'') \leq H(q) + H(q') \]

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
Proof. \((a)\) This clearly holds for \(n = 1\). Suppose its true for \(n\). Then

\[
\gamma(\Gamma^{\otimes(n+1)}) = \gamma(\Gamma^{\otimes n} \otimes \Gamma) = \gamma(\Gamma^{\otimes n}) + \gamma(\Gamma) = n\gamma(\Gamma) + \gamma(\Gamma) = (n + 1)\gamma(\Gamma)
\]

Thus \((a)\) also holds for \(n + 1\) and thus for all \(n\), \((b)\) follows from \((a)\). \(\Box\)
X.5  Coding at a given rate

Definition X.15. Let $I$ and $J$ be alphabets with $|J| > 1$. Then the information rate of $I$ relative to $J$ is $\log_2|I|/\log_2|J|$. 

Note that $\log_2|I|/\log_2|J| = \log_2|J|/\log_2|I| = \log_2|J|/\log_2|I|$.

Theorem X.16 (Noisy Coding Theorem I). Let $\rho > 0$ be a real number and let $\Gamma$ 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 $\sigma_i$.

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 $\sigma_i$. Suppose that

(i) $\log_2|C_i| \geq n_i \rho$ for all $i \in \mathbb{Z}^+$, and

(ii) $\lim_{i \to \infty} M_i = 0$.

Then $\rho \leq \gamma(\Gamma)$.

Proof. Let $\Gamma_i$ be the channel $\Gamma^{\otimes n_i} |_{C_i \times J^{n_i}}$. By X.1

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

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

(2) $\gamma(\Gamma_i) \leq n_i \gamma(\Gamma)$.

By (i)

(3) $\log_2|C_i| \geq n_i \rho$.

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 $\rho \leq \gamma(\Gamma)$. □
Chapter X. The Noisy Coding Theorems

Theorem X.17 (Noisy Coding Theorem II). Let $\rho > 0$ and $\Gamma$ 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 $C_i \subseteq I^m$ such that

(i) $\log_C \frac{|C_i|}{n_i} \geq \rho$ 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 $C_i \times J^n$ decision rules $\sigma_i$ 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 $\Gamma^\otimes n_i |_{C_i \times J^n}$, the probability distribution $p_i$ and the decision rule $\sigma_i$.

Proof. Beyond the scope of these lecture notes.

X.6 Minimum Distance Decision Rule

Lemma X.18. Let $\Gamma = \text{BSC}(e)$, $n \in \mathbb{Z}^+$ and $x, y \in \mathbb{B}^n$. Put $d = d(x, y)$. Then

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

Proof. We have

$$\Gamma^\otimes n_{xy} = \prod_{k=1}^n \Gamma_{x_ky_k}$$

Observe that $\Gamma_{x_ky_k} = e$ if $x_k \neq y_k$ and $\Gamma_{x_ky_k} = 1 - e$ if $x_k = y_k$. Note that there are $d$'s with $x_k \neq y_k$ and $n-d$'s with $x_k = y_k$. So $\Gamma^\otimes n_{xy} = e^d(1 - e)^{n-d}$.

Lemma X.19. Let $0 < e < \frac{1}{2}$, $C \subseteq \mathbb{B}^n$ and $\sigma$ a decision rule for $C$. Then $\sigma$ is a minimum distance rule if and only if $\sigma$ is a maximum likelihood rule with respect to $\text{BSC}^\otimes n(e)$.

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

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

It follow that

$$\Gamma_{az} \leq \Gamma_{a'z} \iff d \geq d'$$

So $\Gamma_{a'z}$ is maximal if and only if $d'$ is minimal.
X.6. MINIMUM DISTANCE DECISION RULE

Lemma X.20. Let \( C \subseteq \mathbb{B}^n \). Let \( \Sigma \) be a decision system with channel \( \text{BSC}^\otimes n(e) \) 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) \( M \leq (\sum_{aeC} p_a |F(a)|) e^{r+1} \).

Proof. (a) Let \( a \in C \) and \( z \in \mathbb{B}^n \) with \( d(a, z) \leq r \). Since \( \sigma \) 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) \( M = \sum_{aeC} p_a M_a \leq \sum_{aeC} p_a |F(a)|e^{r+1} = \left( \sum_{aeC} p_a |F(a)| \right) e^{r+1} \).

Example X.21. Suppose \( C = \{000, 111\} \). Determine a minimal distance rule \( \sigma \) for \( C \). Compute \( F(e) \), \( M_e \) and \( M \) for the decision system determined by \( \text{BSC}(e), (p, 1-p) \) and \( \sigma \).

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

\[
F_{000} = \{011, 101, 110, 111\} \quad \text{and} \quad F_{111} = \{000, 001, 010, 100\}.
\]

So

\[
M_{000} = \Gamma_{000,011} + \Gamma_{000,101} + \Gamma_{000,110} + \Gamma_{000,111} \\
= e^2(1-e) + e^2(1-e) + e^2(1-e) + e^3 \\
= e^2(3(1-e) + e) \\
= e^2(3 - 2e)
\]

By symmetry, also \( M_{111} = e^2(3 - 2e) \) and so

\[
M = pe^2(3 - 2e) + (1-p)e^2(3 - 2e) = e^2(3 - 2e).
\]

Note that each of the four summand in \( M_{000} \) is at most \( e^2 \). So \( M_{000} \leq 4e^2 \) and also \( M \leq 4e^2 \).
Chapter XI

Cryptography in theory and practice

XI.1 Encryption in terms of a channel

Definition XI.1. Let $(\mathcal{M}, \mathcal{C}, (E_k)_{k \in \mathcal{K}}, (D_l)_{l \in \mathcal{K}})$ be a cryptosystem and $p$ and $r$ probability distributions on $\mathcal{M}$ and $\mathcal{K}$, respectively. Define

$$u : \mathcal{M} \times \mathcal{K} \times \mathcal{C} \to [0, 1]$$

$$(m, k, c) \to \begin{cases} p_m r_k & \text{if } c = E_k(m) \\ 0 & \text{if } c \neq E_k(m) \end{cases}$$

$t, q, s$ are defined to be the marginal distribution of $u$ on $\mathcal{M} \times \mathcal{C}, \mathcal{K} \times \mathcal{C}$ respectively. Define the $\mathcal{M} \times \mathcal{C}$-matrix $\Gamma$ by

$$\Gamma_{mc} = \sum_{k \in \mathcal{K}} r_k$$

$\Gamma$ is called the encryption channel for the cryptosystem with respect to $r$.

We interpreted $u_{mkc}$ as the probability $\text{Prob}(m, k, c)$ that plain text message $m$ was encrypted via the key $k$ and the cipher text $c$ was obtained.

Lemma XI.2. With the notation as in XI.1

(a) $u$ is a probability distribution.

(b) $p, r$ and $p \otimes r$ are the marginal distribution of $u$ on $\mathcal{M}, \mathcal{K}$ and $\mathcal{M} \times \mathcal{K}$, respectively.

(c) $\Gamma$ is a channel associated to the distribution $t$ on $\mathcal{M} \times \mathcal{C}$.

(d) $E_k$ is 1-1 for all $k \in \mathcal{K}$. 

171
(e) \[ q_c = \sum_{(m,k) \in \mathcal{M} \times \mathcal{K}} p_m r_k. \]

(f) \( s_{kc} = 0 \) if \( c \notin E_k(\mathcal{M}) \) and \( s_{kc} = p_m r_k \) 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.7] that also \( u \) is a probability 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 \( \Gamma \) is a channel: Let \( m \in \mathcal{M} \), then

\[ \sum_{c \in \mathcal{C}} \Gamma_{mc} = \sum_{c \in \mathcal{C}} \left( \sum_{k \in \mathcal{K}} \frac{r_k}{E_k(m)=c} \right) = \sum_{k \in \mathcal{K}} r_k = 1 \]

So \( \Gamma \) is indeed a channel.

Since \( u_{mkc} = 0 \) for \( c \neq E_k(m) \) we have

\[ t_{mc} = \sum_{k \in \mathcal{K}} u_{mkc} = \sum_{k \in \mathcal{K}} \frac{p_m r_k}{E_k(m)=c} = \sum_{k \in \mathcal{K}} p_m r_k = p_m \sum_{k \in \mathcal{K}} r_k = p_m \Gamma_{mc} \]

Thus \( t = \text{Diag}(p) \Gamma \) and so (see Definition [IX.10(c)]) \( \Gamma \) 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 \( D_{k^*} \circ E_k = \text{id}_\mathcal{M} \). 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),
XI.1. ENCRYPTION IN TERMS OF A CHANNEL

\[ q_c = \sum_{m \in \mathcal{M}} p_m \Gamma_{mc} = \sum_{m \in \mathcal{M}} p_m \left( \sum_{k \in \mathcal{K}} r_k \right) = \sum_{(m,k) \in \mathcal{M} \times \mathcal{K}} E_k(m) = c p_m r_k. \]

(f) Since \( s \) is the marginal distribution of \( u \) on \( \mathcal{K} \times \mathcal{C} \), \( s_{kc} = \sum_{m \in \mathcal{M}} u_{pm} = \sum_{m \in \mathcal{M}} \sum_{E_k(m) = c} p_m r_k \).

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.11) gives

\[ H(p \otimes r) = H(p) + H(r) \]

Consider the function:

\[ \pi : \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{K} \times \mathcal{C}, \quad (m, k) \rightarrow (k, E_k(m)) \]

Since each \( E_k \) is 1-1, also \( \pi \) is 1-1 and so \( \pi \) is a bijection from \( \mathcal{M} \times \mathcal{K} \) to \( \text{Im} \, \pi \). 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 \( \pi \) is a bijection from \( \mathcal{M} \times \mathcal{K} \) to \( \text{Im} \, \pi \), (IX.17) gives

\[ H(p \otimes r) = H(s|_{\text{Im} \, \pi}) \]

Also \( s_{kc} = 0 \) if \( (k,c) \notin \text{Im} \, \pi \). So

\[ H(s|_{\text{Im} \, \pi}) = H(s) \]

and (g) holds.

(h) Using the definition of \( H(r \mid q) \) and (g),

\[ H(r \mid q) = H(s) - H(q) = H(p) + H(r) - H(q). \]

\[ \square \]

**Definition XI.3.**  
(a) The key equivocation of a cryptosystem with respect to the probability distribution \( p \) and \( r \) on \( \mathcal{M} \) and \( \mathcal{K} \), respectively, is

\[ H(r \mid q) \]

(b) A cryptosystem 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 \( \Gamma, r, s \) and \( q \) are as defined in (XI.1).
$H(r|q)$ 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.

### XI.2 Perfect secrecy

**Theorem XI.4.** Given a cryptosystem and a probability distribution $r$ on $K$. 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 $M$.

(c) There exists a positive probability distribution $p$ on $M$ such that $p$ and $q$ are independent.

(d) There exists a positive probability distribution $p$ on $M$ such that $\Gamma_{mc} = q_c$ for all $m \in M$ and all $c \in C$.

(e) Each column of $\Gamma$ is constant, that is there exist a $C$-tuple $(l_c)_{c \in C}$ with $\Gamma_{mc} = l_c$ for all $m \in M, c \in C$.

(f) $\Gamma_{mc} = q_c$ for all probability distributions $p$ on $M$, all $m \in M$ and all $c \in C$.

**Proof.** (a) $\iff$ (b): By IX.19 $I(p, q) = 0$ if and only if $p$ and $q$ are independent.

(b) $\implies$ (c): Just choose $p$ to be the equal probability distribution on $M$.

(c) $\implies$ (d): Since $p$ and $q$ independent we have $t_{mc} = p_m q_c$ for all $m \in M$ and $c \in C$. Since $\Gamma$ is a channel associated to $t$, $t_{mc} = p_m \Gamma_{mc}$. Thus $p_m \Gamma_{mc} = p_m q_c$ for all $m \in M$ and all $c \in C$. Since $p$ is positive $p_m \neq 0$ and so $\Gamma_{mc} = q_c$.

(d) $\implies$ (e): All entries in column $c$ of $\Gamma$ are equal to $q_c$.

(e) $\implies$ (f): Let $c \in C$. Since column $c$ of $\Gamma$ is constant there exists $l_c \in [0, 1]$ with $\Gamma_{mc} = l_c$ for all $m \in M$. Let $p$ be any probability distribution on $M$. Then

$$q_c = \sum_{m \in M} p_m \Gamma_{mc} = \sum_{m \in M} p_m l_c = \left(\sum_{m \in M} p_m\right) l_c = 1 l_c = l_c$$

and so (f) holds.

(f) $\implies$ (b): Let $p$ be a probability distribution on $M$. 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 XI.5.** If a cryptosystem has perfect secrecy (with respect to some probability distribution on $K$), then the numbers of keys is greater or equal to the number of plaintext messages.
XI.3. THE ONE-TIME PAD

Proof. Fix \( c \in \mathcal{C} \) with \( q_c \neq 0 \). Then by XI.4 \( \Gamma_{mc} = q_c > 0 \) for all \( m \in \mathcal{M} \). By XI.2

\[
\sum_{k \in \mathcal{K}} r_k = \Gamma_{mc} > 0
\]

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 \( |\mathcal{M}| \leq |\mathcal{K}| \).

XI.3 The one-time pad

Definition XI.6. 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 \( \Omega(\mathbb{F}^n) = (\mathbb{F}^n, \mathbb{F}^n, (E_k)_{k \in \mathbb{F}^n}, (D_k)_{k \in \mathbb{F}^n}) \) is called the one-time pad determined by \( \mathbb{F}^n \).

Lemma XI.7 (One-Time Pad). Any one-time pad is a cryptosystem and has perfect secrecy with respect to the equal-probability distribution \( r \) on the set of keys.

Proof. Given a one-time pad \( \Omega(\mathbb{F}^n) \). Since \( (m + k) + (-k) = m + (k + (-k)) = m + 0 = m \), \( D_{-k} \circ E_k = \text{id}_{\mathbb{F}^n} \). Thus the one-time pad is a cryptosystem.

Put \( e = \frac{1}{|\mathbb{F}|} \). Then \( r_k = e \) for all \( k \in \mathcal{K} \).

Let \( m \in \mathcal{M} \) and \( c \in \mathcal{C} \). By definition \( \Gamma \) and \( E_k \):

\[
\Gamma_{mc} = \sum_{k \in \mathcal{K}} r_k = \sum_{k \in \mathcal{K}} r_k
\]

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 \( \Gamma \) are constant and so by XI.4 the one-time pad has perfect secrecy.

XI.4 Iterative methods

Definition XI.8. 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 \( k = (k_1, \ldots, k_r) \in \mathcal{K} \) define \( X_{i+1}, 1 \leq i \leq r \) inductively by

\[
X_{i+1} = X_{i-1} + F(k_i, X_i)
\]
Lemma XI.9. Any Feistel system is a cryptosystem.

Proof. Let \( E \) the Feistel system determined by \( P \). We will show that also \( D \) holds for all \( 0 \leq i \leq r \) inductively by

\[
Y_{i+1} = Y_{i-1} - F(k_i, Y_i), 1 \leq i \leq r
\]

Define

\[
D_k : \mathcal{C} \to M, (Y_0, Y_1) \to (Y_{r+1}, Y_r).
\]

Put \( \Omega(\mathbb{F}^n, K, r, F) = (\mathbb{F}^n \times \mathbb{F}^n, \mathbb{F}^n \times \mathbb{F}^n, (E_k)_{k \in K^r}, (D_k)_{k \in K^r}) \). Then \( \Omega(\mathbb{F}^n, K, r, F) \) is called the Feistel system determined by \( \mathbb{F}^n, K, r \) and \( F \).

Define

\[
E_k : M \to \mathcal{C}, (X_0, X_1) \to (X_{r+1}, X_r).
\]

For \( (Y_0, Y_1) \in \mathcal{C} \) and \( k = (k_1, \ldots, k_r) \in K \) define \( Y_{i+1}, 1 \leq i \leq r \) inductively by

\[
Y_{i+1} = Y_{i-1} - F(k_i, Y_i), 1 \leq i \leq r
\]

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:

\[
Y_{i+1} = Y_{i-1} - F(k_i, Y_i) = X_{(r+1)-(i+1)} = X_{r+1-i} - F(k_{r+1-i}, X_{r+1-i})
\]

Hence \( P(i) \) holds for all \( 0 \leq i \leq r+1 \) and so

\[
D_k(\mathcal{C}) = (Y_{r+1}, Y_r) = (X_0, X_1)
\]

Thus \( D_k \circ E_k = \text{id}_M \) and the Feistel system is indeed a cryptosystem. □

Example XI.10. Consider the Feistel system with \( \mathbb{F}^n = \mathbb{F}_2^3 \), \( K = \mathbb{F}_2^3 \), \( r = 3 \) and

\[
F : \mathbb{F}_2^3 \times \mathbb{F}_2^3 \to \mathbb{F}_2^3, (\alpha \beta \gamma, xyz) \to (\alpha x + yz, \beta y + xz, \gamma z + xy)
\]

Compute \( E_k(m) \) for \( k = (100, 101, 001) \) and \( m = (101, 110) \). Verify that \( D_k \circ (E_k(m)) = m \).
XI.5. THE DOUBLE-LOCKING PROCEDURE

<table>
<thead>
<tr>
<th>i</th>
<th>$k_i$</th>
<th>$X_i$</th>
<th>$F(k_i, X_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>101</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>110</td>
<td>(1 · 1 + 1 · 0, 0 · 1 + 1 · 0, 0 · 0 + 1 · 1) = 101</td>
</tr>
<tr>
<td>2</td>
<td>101</td>
<td>000</td>
<td>000</td>
</tr>
<tr>
<td>3</td>
<td>001</td>
<td>110</td>
<td>(0 · 1 + 1 · 0, 0 · 0 + 1 · 0, 1 · 1 + 0 · 1) = 001</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>001</td>
<td>101</td>
</tr>
</tbody>
</table>

So $E_k(m) = (001, 110)$. To decrypt (001, 110) we use the key $k^* = (001, 101, 100)$.

<table>
<thead>
<tr>
<th>i</th>
<th>$k_i^*$</th>
<th>$Y_i$</th>
<th>$F(k_i^*, Y_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>001</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>001</td>
<td>110</td>
<td>(0 · 1 + 1 · 0, 0 · 1 + 1 · 0, 1 · 0 + 1 · 1) = 001</td>
</tr>
<tr>
<td>2</td>
<td>101</td>
<td>000</td>
<td>000</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>110</td>
<td>(1 · 1 + 1 · 0, 0 · 0 + 1 · 0, 0 · 0 + 1 · 1) = 101</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>101</td>
<td></td>
</tr>
</tbody>
</table>

So $D_{k^*}(E_k(m)) = (101, 110) = m.$

XI.5 The Double-Locking Procedure

Two cryptosystems $\Omega$ and $\tilde{\Omega}$ are called compatible if $\mathcal{M} = \mathcal{C} = \tilde{\mathcal{M}} = \tilde{\mathcal{C}}$.

Given compatible cryptosystems $\Omega$ and $\tilde{\Omega}$, keys $k, k^*$ in $\Omega$ and keys $\tilde{k}, \tilde{k}^*$ in $\tilde{\Omega}$ with $D_{k^*} \circ E_k = \text{id}_\mathcal{M}$ and $\tilde{D}_{\tilde{k}^*} \circ \tilde{E}_{\tilde{k}} = \text{id}_{\tilde{\mathcal{M}}}$.

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 $m_2 = \tilde{E}_k(m_1)$ and sends $m_2$ to $X$.  
- $X$ computes $m_3 = D_{k^*}(\tilde{m})$ and sends $\tilde{X}$.  
- $\tilde{X}$ computes $m_4 = \tilde{D}_{\tilde{k}^*}(m_3)$.  

- $X$ computes $m_1 = E_k(m_0)$ and sends $m_1$ to $\tilde{X}$.  
- $\tilde{X}$ computes $m_2 = \tilde{E}_k(m_1)$ and sends $m_2$ to $X$.  
- $X$ computes $m_3 = D_{k^*}(\tilde{m})$ and sends $\tilde{X}$.  
- $\tilde{X}$ computes $m_4 = \tilde{D}_{\tilde{k}^*}(m_3)$.
CHAPTER XI. CRYPTOGRAPHY IN THEORY AND PRACTICE

\[ \begin{array}{c}
m_0 \xrightarrow{E_k} m_1 \xrightarrow{\tilde{E}_k} m_2 \xrightarrow{D_{k^*}} m_3 \xrightarrow{\tilde{D}_{k^*}} m_4 \\
\end{array} \]

Is \( m_4 = m_0 \)?

Consider the following example \( \mathcal{M} = \mathcal{C} = \{1, 2, 3\} \),

\[
\begin{array}{c|c|c|c}
E_k = D_{k^*} & 1 & 2 & 3 \\
\hline
1 & 3 & 2 \\
\end{array}
\quad
\begin{array}{c|c|c|c}
\tilde{E}_k = \tilde{D}_{k^*} & 1 & 2 & 3 \\
\hline
2 & 1 & 3 \\
\end{array}
\]

and \( m_0 = 1 \)

\[ \begin{array}{c}
1 \xrightarrow{E_k} 1 \xrightarrow{\tilde{E}_k} 2 \xrightarrow{D_{k^*}} 3 \xrightarrow{\tilde{D}_{k^*}} 3 \\
\end{array} \]

So \( m_4 \neq m_0 \).

In general

\[
m_4 = (\tilde{D}_{k^*} \circ D_{k^*} \circ \tilde{E}_k \circ E_k)(m_0)
\]

If \( D_{k^*} \) commutes with \( \tilde{E}_k \), that is \( D_{k^*} \circ \tilde{E}_k = \tilde{E}_k \circ D_{k^*} \) then

\[
\tilde{D}_{k^*} \circ D_{k^*} \circ \tilde{E}_k \circ E_k = (\tilde{D}_{k^*} \circ \tilde{E}_k) \circ (D_{k^*} \circ E_k) = \text{id}_M \circ \text{id}_M = \text{id}_M
\]

and procedure works.

Since addition in finite field is commutative, one-time pads provide examples where \( D_{k^*} \) commutes with \( \tilde{E}_k \).

**Example XI.11.** Suppose \( \Omega \) and \( \tilde{\Omega} \) both are the one-time pad determined by \( \mathbb{F}_2^4 \). Given the following public information:

\[ \begin{array}{c}
m_0 \xrightarrow{E_k} 1101 \xrightarrow{\tilde{E}_k} 0110 \xrightarrow{D_{k^*}} 1100 \xrightarrow{\tilde{D}_{k^*}} m_4 \\
\end{array} \]

What is \( m_0 \)?

Since \( 0110 + k^* = 1100 \) and so \( k^* = -\frac{1100}{0110} = 1010 \). So \( m_0 = D_{k^*}(m_1) = +\frac{1101}{1010} = 0111 \).

So one-time pads should not be used for the double-locking procedure.

**Lemma XI.12.** Let \( \Omega \) and \( \tilde{\Omega} \) be compatible cryptosystem. Let \( \beta \) be an encryption function in \( \tilde{\Omega} \) and let \( \gamma \) and \( \gamma' \) be decryption functions in \( \Omega \) which commute with \( \beta \), that is

\[
\gamma \circ \beta = \beta \circ \gamma, \quad \text{and} \quad \gamma' \circ \beta = \beta \circ \gamma'
\]

Let \( m_1 \in \mathcal{M} \) and put \( m_2 = \beta(m_1) \). Then

\[
\gamma(m_2) = \gamma'(m_2) \implies \gamma(m_1) = \gamma'(m_2)
\]
XI.5. THE DOUBLE-LOCKING PROCEDURE

Proof.

\[ \beta(\gamma(m_1)) = (\beta \circ \gamma)(m_1) = (\gamma \circ \beta)(m_1) = \gamma(\beta(m_1)) = \gamma(m_2) \]

By symmetry, \( \beta(\gamma'(m_1)) = \gamma'(m_2) \). So if \( \gamma(m_2) = \gamma'(m_2) \) we conclude that

\[ \beta(\gamma(m_1)) = \beta(\gamma'(m_1)) \]

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 \( \Omega \) with \( D_l(m_2) = m_3 \) can compute \( m_0 \), namely \( m_0 = D_l(m_1) \).
Appendix A

Rings and Field

A.1 Basic Properties of Rings and Fields

Definition A.1. 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 \(a, b \in R\); \([\text{closure for addition}]\)

(Ax 2) \((a + b) + c = (a + c) + b\) for all \(a, b, c \in R\); \([\text{associative addition}]\)

(Ax 3) \(a + b = b + a\) for all \(a, b \in R\). \([\text{commutative addition}]\)

(Ax 4) there exists an element in \(R\), denoted by \(0_R\) and called ‘zero \(R\)’, \([\text{additive identity}]\)

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 \(-a\) \([\text{additive inverses}]\)

and called ‘negative \(a\)’, such that \(a + (-a) = 0_R\);

(Ax 6) \(ab \in R\) for all \(a, b \in R\); \([\text{closure for multiplication}]\)

(Ax 7) \((ab)c = (a(bc))\) for all \(a, b, c \in R\); \([\text{associative multiplication}]\)
APPENDIX A. RINGS AND FIELD

(Ax 8) \( a(b + c) = ab + ac \) and \( (a + b)c = ac + bc \) for all \( a, b, c \in R \). \( \text{[distributive laws]} \)

**Definition A.2.** A ring \((R, +, \cdot)\) is called commutative if \( \text{(Ax 9) } ab = ba \text{ for all } a, b \in R. \) \( \text{[commutative multiplication]} \)

**Definition A.3.** An element \( 1_R \) in a ring \((R, +, \cdot)\) is called an (multiplicative) identity if \( \text{(Ax 10) } 1_R \cdot a = a = a \cdot 1_R \text{ for all } a \in R. \) \( \text{[multiplicative identity]} \)

**Definition A.4.** A field is a commutative ring \((F, +, \cdot)\) with identity \( 1_F \neq 0_F \) such that \( \text{(Ax 11) } \) for each \( a \in R \) with \( a \neq 0_F \) there exists an element in \( R \), denoted by \( a^{-1} \) \( \text{[multiplicative inverses]} \)

and called ‘a inverse’, such that \( a \cdot a^{-1} = 1_R = a^{-1} \cdot a; \)

If \((R, +, \cdot)\) is a ring, we will often just say that \( R \) is a ring, assuming that there is no confusion about the underlying addition and multiplication. Also we will usually write 0 for \( 0_R \) and 1 for \( 1_R \).

With respect to the usual addition and multiplication:
The real number and the rational numbers are fields. The integers are a commutative ring but not a field. \( \mathbb{F}_2 \) is a field.

**Lemma A.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, \( (-1)a = -a. \)
A.2. Polynomials

**Definition A.6.** Let \( R \) be a 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

\[
f + g = \sum_{i=0}^{\min(n,m)} (a_i + b_i) x^i,
\]

where \( l = \max(n,m) \), \( a_i = 0 \) for \( i > n \) and \( b_i = 0 \) for \( i > m \); and

\[
f g = \sum_{k=0}^{n+m} \left( \sum_{i=0}^{k} a_i b_{k-i} \right) x^i
\]

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

\[
\text{if } a \in \mathbb{F}, \text{ then } (af)g = a(fg)
\]
(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 \( \odot \):

\[
\begin{align*}
(\ast) & \quad f \oplus g = \overline{f + g} \text{ and } f \odot g = \overline{fg}.
\end{align*}
\]

By VII.11[19]

\[
(\ast\ast) & \quad \overline{f} = f \text{ for all } f \in \mathbb{F}^h[x].
\]

By VII.14

\[
(\ast\ast\ast) & \quad \overline{f + g} = \overline{f} \oplus \overline{g} = f \oplus g = \overline{f + g} \text{ and } \overline{f} \odot \overline{g} = \overline{f \odot g} = f \odot g = \overline{fg}.
\]

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

\[
e \oplus (f \oplus g) = e \oplus \overline{f + g} = e + (f + g) = (e + f) + g
\]

and

\[
(e \oplus f) \oplus g = \overline{e + f} \oplus \overline{g} = (e + f) + g.
\]

Thus condition (ii) holds.

We have \( 0 \oplus f = \overline{0 + f} = \overline{f} \) and so for \( f \in \mathbb{F}^h[x] \), \( 0 \oplus f = f \). Hence condition (iii) holds.

\[
f \oplus -f = \overline{f + (-f)} = \overline{0} = 0
\]

and condition (iv) is proved.
A.3. **Irreducible Polynomials**

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
\[ (e \odot f) \oplus (e \odot g) = \overline{ef \oplus eg} = \overline{ef + eg} \]
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
\[ f \odot g = \overline{fg} = \overline{g}f = g \odot f \]
and condition (ix) holds.

\[ \square \]

**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 \( h \mid fg \). Then \( h \mid f \) or \( h \mid g \).

*Proof.* Since \( h \mid fg \), 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 VII.38, \( \mathbb{F}^h[x] \) 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, \( \deg f > 0 \), \( f \mid g \) and \( g \) is irreducible. Then \( f = g \).

*Proof.* Since \( f \mid g \), \( g = fh \) 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 \( g, f_1, \ldots, f_r \) be irreducible monic polynomials in \( \mathbb{F}[x] \). If \( g \) divides \( af_1 \ldots f_r \) in \( \mathbb{F}[x] \), 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 \( h = a f_1 \ldots f_{r-1} \). 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 \( 1 \leq i \leq r - 1 \). \( \square \)

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 \( f_1, f_2, \ldots, f_r \in \mathbb{F}[x] \) with

\[
f = af_1f_2\ldots f_r
\]

Moreover, the \( f_1, f_2, \ldots, f_r \) 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

\[
g = b f_1^{\epsilon_1} f_2^{\epsilon_2} \ldots f_r^{\epsilon_r}
\]

Moreover, if \( g \) is of this form, then \( f = gh \), where

\[
h = c f_1^{\delta_1} f_2^{\delta_2} \ldots f_r^{\delta_r}
\]

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 \( f_1, \ldots, f_r \). 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

\[
g = bg_1g_2\ldots g_s \text{ and } h = ch_1\ldots h_t
\]

where \( b = \text{lead}(g) \), \( c = \text{lead}(h) \) and \( g_1, \ldots, g_s, h_1, \ldots, h_t \) are monic irreducible polynomials. Since \( a = bc \) we can choose \( r = s + t \) and

\[
f_1 = g_1, \ldots, f_s = g_s, f_{s+1} = h_1, \ldots, f_{s+t} = h_t
\]

To prove the existence suppose that

\[
f = af_1\ldots f_r = ag_1\ldots g_s
\]

for some monic irreducible polynomials \( f_1, \ldots, f_r, g_1, \ldots, g_s \).

Then \( f | f = ag_1\ldots g_s \) and so A.12 show that \( f_i = g_i \) for some \( 1 \leq i \leq r \). Reordering the \( g_i \)'s we may assume that \( f_1 = g_1 \). Hence also
The induction assumptions implies that \( r = s \) and after reordering \( f_2 = g_2, \ldots, f_r = g_r \).

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 \( f = gh = t \tilde{g} \), \( t \) divides \( f \) and so by A.12, \( t = f_i \) for some \( 1 \leq i \leq t \).

Without loss \( i = 1 \). Then

\[
\tilde{g}h = af_2 \ldots f_n
\]

Note that lead(\( \tilde{g} \)) = lead(\( g \)) = \( b \). By induction

\[
\tilde{g} = bf_2^{\epsilon_2} \ldots f_r^{\epsilon_r} \quad \text{and} \quad h = f_2^{\delta_2} \ldots f_n^{\delta_n}
\]

where \( c = \frac{a}{b} \), \( \epsilon_i \in \{0, 1\} \) and \( \delta_i = 1 - \epsilon_i \) for \( 2 \leq i \leq r \). Thus \( \box{b} \) holds with \( \epsilon_1 = 1 \) and \( \delta_1 = 0 \).

\[
A.4 \quad \text{Primitive elements in finite field}
\]

**Lemma A.14.** Let \( \mathbb{E} \) be a finite field and put \( t = |\mathbb{E}| - 1 \). Let \( e \in \mathbb{E}^\ast \). Then

(a) There exists positive integer \( m \) with \( e^m = 1 \). The smallest such positive integer is called the order of \( e \) in \( \mathbb{E}^\ast \) 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:

(*) 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 \( |\mathbb{E}| = t \), the elements \( e^i, 0 \leq i \leq t \) cannot be pairwise distinct. So by (*) 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 (a).

(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), $e^m = e^r$ 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 $\mathbb{E}_1$, 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 $\mathbb{E}_1$ lies in exactly one equivalence class and since each equivalence class has $|e|$ elements, $|\mathbb{E}_1| = d|e|$. 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) $D_d(n) = \{ed \mid e \in \mathbb{Z}_d^*\}$.

(b) $|D_d(n)| = \varphi(\frac{n}{d})$.

(c) $n = \sum_{d \mid n} \varphi(d)$.

(d) $\varphi(n) \geq 1$.

**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 $e \in \mathbb{Z}_d^*$.

Conversely, if $e \in \mathbb{Z}_d^*$, 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

$$n = |\{m \mid 0 \leq m < n\}| = \sum_{f \mid n} |D_f(n)| = \sum_{f \mid n} |D_f(n)| = \sum_{f \mid n} \varphi \left(\frac{n}{f}\right) = \sum_{d \mid n} \varphi(d)$$

(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 $x^n - 1$. 

APPENDIX A. RINGS AND FIELD
Note that if \( \alpha \) 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 \( e \in \mathbb{F}^\# \) 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 \( \mathbb{F}^\# \) has exactly \( \phi(d) \) elements of order \( d \), namely the elements \( e^i, i \in D_\frac{n}{d}(n) \).

**Proof.**

(a) Let \( 0 \leq i < n \). Then \((e^i)^n = (e^n)^i = 1\) and so \( e^i \) is a root of \( x^n - 1 \). Since the \( e^i, 0 \leq i < n \) 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 \( \frac{n}{d} \mid l \). Thus \( e^m \) has order \( \frac{n}{d} \).

(c) Let \( a \in \mathbb{F}^\# \). 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 \( 0 \leq i < n \). 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 \( |D_\frac{n}{d}(n)| = \phi\left(\frac{n}{d}\right) = \phi(d) \), (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
\[
\mathbb{E}^t = \{ \alpha^i \mid 0 \leq i < t \}
\]
Such an \( \beta \) 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 \( \mathbb{E}^\# \). If \( e \in \mathbb{E}^\# \) then \( |e| \) is a divisor of \( t \) and so
\[
(\ast) \quad t = |\mathbb{E}^\#| = | \bigcup_{n \mid t} A_n | = \sum_{n \mid t} |A_n |.
\]

Let \( n \mid t \). Suppose \( A_n \neq \emptyset \). Then \( \mathbb{E}^t \) 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
\[
(\ast\ast) \quad \sum_{n \mid t} |A_n | \leq \sum_{n \mid t} \phi(n) = t.
\]

Together with (\ast) we conclude that equality must holds everywhere in (\ast) and (\ast\ast). 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 \( \beta \) of order \( t \). Then \( \{ \beta^i \mid 0 \leq i < t \} \) are \( t \)-pairwise distinct elements in \( \mathbb{E}^t \) and the theorem is proved. \( \square \)
Lemma A.19. Let $n$ be a positive integer. Let $n = 2^k m$ where $m, k \in \mathbb{N}$ with $m$ odd. Let $E$ be a splitting field for $x^m - 1$ over $\mathbb{F}_2$ and let $\alpha_1, \alpha_2, \ldots, \alpha_m \in E$ 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 $1 \leq i < k \leq m$. Put $\alpha = \alpha_i$. Then

$$x^n - 1 = (x - \alpha)^2 g$$

for some $g \in E[x]$. Taking derivatives gives

$$mx^{n-1} = 2(x - \alpha)g + (x - \alpha)^2 g'.$$

Observe that in $E$, $2 = 0$ and, since $m$ is odd, $m = 1$. Therefore,

$$x^{n-1} = (x - \alpha)^2 g'.$$

Hence $\alpha$ is a root of $x^{n-1}$ and so $\alpha = 0$, a contradiction to $\alpha^m = 1$. \hfill \Box

Lemma A.20. Let $n$ be a positive odd integers and let $E$ be a finite field containing $\mathbb{F}_2$. Let $t = |E| - 1$ and let $\beta$ be a primitive root for $E$.

(a) $E$ is a splitting field for $x^n - 1$ if and only if $n$ divides $t$.

(b) Suppose $n$ divides $t$ and put $\alpha = \beta^{t/n}$. Then $\alpha$ is a primitive root of $x^n - 1$.

Proof. Let $d = \gcd(t, n)$ and put $s = \frac{t}{d}$. Then $\beta^m$ 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 $E$ are

$$\beta^{is}, \quad 0 \leq i < d.$$

Therefore $E$ contains exactly $d$ roots of $x^n - 1$.

From A.19, $E$ is a splitting field for $x^n - 1$ if and only if $E$ 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 $\beta^s = \beta^{t/n} = \alpha$ and so the roots of $x^n - 1$ are $\alpha^i, 0 \leq i < n$. \hfill \Box
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 $I \times J$ (via $I \times J \times K = (I \times J) \times K$) and let $f_I$ the marginal tuple of $f$ on $I$ (via $I \times J \times K = I \times (J \times K)$). 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

$$g(i) = \sum_{j \in J} f_{I \times J}(i, j) = \sum_{j \in J} \left( \sum_{k \in K} f(i, j, k) \right) = \sum_{(j, k) \in J \times K} f(i, j, k) = f_I(i) \quad \square$$

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 $(l_1, \ldots, l_r)$ of positive integers and all $t \in \mathbb{N}$

$$p^{(l_1, \ldots, l_r)} = p^{(l_1 + t, \ldots, l_r + t)}$$

(b) $(S, P)$ is stationary.

(c) For all $r \in \mathbb{Z}^+$, $p^r = p^{(2, \ldots, r+1)}$.

Proof. Suppose (a) holds. Choosing $(l_1, \ldots, l_r) = (1, \ldots, r)$ we see that

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

191
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 \((l_1, \ldots, l_r)\) 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 \(u = l, v = u - l\) and let \(k = (k_1, \ldots, k_v)\) be the increasing \(v\)-tuple of positive integers with \(\{1, 2, \ldots, u\} = \{l_1, \ldots, l_r\} \cup \{k_1, \ldots, k_u\}\). Let \(l = (l_1, \ldots, l_r)\). Identifying \(S^u\) with \(S^r \times S^v\) via \(s \mapsto (s_1, s_k)\) we see that \(P^1\) is the marginal distribution of \(P^u\) on \(S^r\).

Let \(\bar{k} = (k_1 + 1, \ldots, k_v + 1)\) and \(\bar{l} = (l_1 + 1, \ldots, l_r + 1)\)

Identifying \(S^{u+n}\) with \(S^r \times S^v \times S\) via \(s \mapsto (s_1, s_{\bar{k}}, s_1)\) we see that \(P^{\bar{I}}\) is the marginal distribution of \(P^{u+1}\) on \(S^r\). Also \(P^{(2, \ldots, u+1)}\) is the marginal distribution of \(P^{u+1}\) on \(S^r \times S^v\). Thus B.1 shows that \(P^{\bar{I}}\) is the marginal distribution of \(P^{(2, \ldots, u+1)}\) on \(S^r\).

Since (c) holds, \(P^u = P^{(2, \ldots, u+1)}\). Hence also the marginal distribution \(P^1\) and \(P^{\bar{I}}\) of these distribution on \(S^r\) are equal. \(\square\)

### B.2 A stationary source which is not memory less

**Example B.3.** An example of a stationary source which is not memory less.

For \(z = z_1 \ldots z_n \in B^+\) 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 \(P(0) = P(1) = \frac{1}{2} \frac{3^n}{4^{n-1}} = \frac{1}{2}\). Also \(u(zs) = u(z) + 1\) if \(s = s_n\) and \(u(zs) = u(s)\) if \(z_n \neq s\). Hence for \(z \in B^n\) with \(n \geq 1\) and \(s \in B\):

\[
P(zs) = \begin{cases} \frac{3}{4} P(z) & \text{if } s = z_n \\ \frac{1}{4} P(z) & \text{if } s \neq z_n \end{cases}
\]

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

\[
p^n(z) = P(z) = P(0z) + P(1z) = P^{(2, \ldots, n+1)}(z)
\]

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

\[
(\ast) \quad t := \sum_{i \in I} f_i = \sum_{j \in J} g_j
\]

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 $f$.

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 $\ast$

\[
f_a + f_b = k = g_c + g_d \quad \text{and so} \quad g_d - f_a = f_b - g_c
\]

So we can define $h$ as follows

<table>
<thead>
<tr>
<th></th>
<th>$c$</th>
<th>$d$</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$u$</td>
<td>$f_a - u$</td>
<td>$f_a$</td>
</tr>
<tr>
<td>$b$</td>
<td>$g_c - u$</td>
<td>$g_d - f_a + u = f_b - g_c + u$</td>
<td>$f_b$</td>
</tr>
<tr>
<td>$g$</td>
<td>$g_c$</td>
<td>$g_d$</td>
<td>$t$</td>
</tr>
</tbody>
</table>

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 $\bar{J} = J \setminus \{v\}$. Define a $\bar{J}$-tuple $\bar{g}$ on $\bar{J}$ by

\[
\bar{g}_j = \begin{cases} 
  g_j & \text{if } j \neq u \\
  g_u + g_v & \text{if } j = u
\end{cases}
\]

Then

\[
\sum_{j \in \bar{J}} \bar{g}_j = \bar{g}_u + \sum_{j \in \bar{J} \setminus u} \bar{g}_j = g_u + \sum_{j \in J \setminus u} g_j = \sum_{j \in J} g_j = t = \sum_{i \in I} f_i
\]

Inductively we may assume that we found all possible $I \times \bar{J}$-matrices $\bar{h}$ with coefficients in $\mathbb{R}^\geq 0$ and marginal distribution $f$ and $\bar{g}$.
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 coefficients in \( \mathbb{R}^{\geq 0} \) and marginal distribution \( \hat{f} \) and \( \hat{g} \). Define the \( I \times J \)-matrix \( h \) by

\[
h_{ij} = \begin{cases} 
\tilde{h}_{ij} & \text{if } j \in J \setminus \hat{J} \\
\hat{h}_{ij} & \text{if } j \in \hat{J}
\end{cases}
\]

So columns \( u \) and \( v \) of \( h \) come from \( \hat{h} \), while the remaining columns come from \( \tilde{h} \).
Appendix C

More On channels

C.1 Sub channels

Lemma C.1. Let $\Gamma : I \times J \rightarrow [0, 1]$ be a channel. Let $K \subseteq I$ and let $\Xi$ be the restriction of $\Gamma$ to $K \times J$. (So $\Xi$ is the function from $K \times I \rightarrow [0, 1]$ with $\Xi_{kj} = \Gamma_{kj}$ for all $k \in K, \ j \in J$.) Let $p$ be a probability distribution on $K$, and define $\hat{p} : I \rightarrow [0, 1]$ 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) $\Xi$ 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 $Row_k(\Xi) = Row_k(\Gamma)$ and so $Row_k(\Xi)$ 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) $H_{\Xi}(q \mid k) = H(\text{Row}_k(\Xi)) = H(\text{Row}_k(\Gamma)) = H_{\Gamma}(q \mid k)$.

(e) Using IX.24 twice we have
\[ H^\Xi(q \mid p) = \sum_{i \in K} p_i H^\Xi(q \mid i) = \sum_{i \in K} p_i H^\Gamma(q \mid i) = \sum_{i \in \hat{K}} p_i H^\Gamma(q \mid i) = H^\Gamma(q \mid \hat{p}). \]

Let \( \mathcal{P}(I) \) and \( \mathcal{P}(K) \) be the set of probability distribution on \( I \) and \( K \) respectively. Then using (c) and (e),

\[ \gamma(\Xi) = \max_{p \in \mathcal{P}(K)} H(q) - H^\Xi(q \mid p) \leq \max_{p \in \mathcal{P}(K)} H(q) - H^\Gamma(q \mid \hat{p}) = \max_{p \in \mathcal{P}(I)} H(q) - H^\Gamma(q \mid p) = \gamma(\Gamma). \]

\( \square \)
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.

\[
\begin{pmatrix}
\bar{0} & \bar{0} & \bar{0} \\
J & P^2 & P \\
P & J & P^2 \\
P^2 & P & J \\
I & I + P & I + P^2 \\
I + P^2 & I & I + P \\
I + P & I + P^2 & I \\
\bar{1} & \bar{1} & \bar{1}
\end{pmatrix}
\]

Here

\[
\bar{0} = 000, \quad \bar{1} = 111, \quad I = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}
\]

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