

# Code Theory

Manuel Gijón Agudo

September 2017 - January 2018

# Índice

<b>1. Memoryless resources</b>	<b>3</b>
1.1. Sources and average word length . . . . .	3
1.2. Uniquely decodeable codes . . . . .	3
1.3. Optimal codes . . . . .	4
1.4. Extension of sources . . . . .	5
<b>2. Information and entropy</b>	<b>6</b>
2.1. Definitions . . . . .	6
2.2. Properties of the entropy function . . . . .	6
2.3. Shannon-Fano Code . . . . .	6
2.4. Product of sources . . . . .	7
2.5. Markov Chains . . . . .	7
2.6. Sources with memory . . . . .	8
<b>3. Information channels</b>	<b>9</b>
3.1. Channel matrix . . . . .	9
3.2. System Entropies and mutual information . . . . .	9
3.3. Extension of noiseless coding theorem to information channels . . . . .	10
3.4. Decision rules . . . . .	11
3.5. Improving reliability . . . . .	11
3.6. Rates of transmission and Hamming distance . . . . .	11
<b>4. Finite fields</b>	<b>12</b>
4.1. Basic definitions . . . . .	12
4.2. Properties of finite fields . . . . .	12
4.3. Factorization of polynomials . . . . .	12
<b>5. Block codes</b>	<b>13</b>
5.1. Minimum distance . . . . .	13
5.2. Bounds on block codes . . . . .	13
5.3. Asymptotically good codes . . . . .	13

<b>6. Linear codes</b>	<b>14</b>
6.1. Basics . . . . .	14
6.2. Syndrom decoding . . . . .	14
6.3. Dual code and Mc Williams identities . . . . .	14
6.4. The Griesmer bound . . . . .	14
<b>7. Cyclic codes</b>	<b>15</b>
7.1. Introduction . . . . .	15
7.2. Quadratic residue codes . . . . .	15
7.3. BCH Codes . . . . .	15
<b>8. Maximum distance separable codes</b>	<b>16</b>
8.1. Singleton bound . . . . .	16
8.2. Linear MDS codes . . . . .	16
<b>9. Alternant codes</b>	<b>17</b>
<b>10. Low density parity check codes</b>	<b>18</b>
10.1. Bipartite graphs with the expander property . . . . .	18
10.2. Low density parity check (LDPC) codes . . . . .	18
10.3. Belief propagation . . . . .	18
<b>11. P-adic codes</b>	<b>19</b>
11.1. P-adic numbers . . . . .	19
11.2. Polynomials over $\mathbb{Q}_p$ . . . . .	19

## 1. Memoryless resources

### 1.1. Sources and average word length

**Definition 1:** a **source** is a finite set  $\mathcal{S}$  together with a set of random variables  $(X_1, X_2, \dots)$  whose range is  $\mathcal{S}$ .

If  $P(X_n = \mathcal{S}_i)$  only depends on  $i$  and not on  $n$  then we say the source is **stationary** and if the  $X_n$  are independent then it's **memoryless**.

Insert example here

**Definition 2:** Let  $\mathcal{T}$  be a finite set called **alphabet**. A map  $\mathfrak{C} : \mathbb{S} \longrightarrow \mathbb{U}_{n \geq 1} T^n$  is called a **code**.

If  $|\mathcal{T}| = r$  then  $\mathfrak{C}$  is a  **$r$ -ary code**.

A code extends from  $\mathbb{S}$  to  $T \cup T^2 \cup \dots$  to  $\mathbb{S} \cup \mathbb{S}^2 \cup \dots$  to  $T \cup T^2 \cup \dots$  in obvious way.

insert example here

**Definition 3:** The **average word-length** of a code  $\mathfrak{C}$  is  $L(\mathfrak{C}) := \sum_{i=1}^n p_i l_i$  where  $l_i$  is the length of the image of the symbol of  $\mathbb{S}$ , which is emitted with probability  $p_i$ .

For now, we write  $\mathfrak{C}$  to be the image of  $\mathfrak{C}$ .

### 1.2. Uniquely decodeable codes

**Definition 4:** If for any sequences  $u_1 \dots u_n = v_1 \dots v_m$  in  $\mathfrak{C}$  implies  $m = n$  and  $u_i = v_i$  for  $i = 1, \dots, n$  then we say that  $\mathfrak{C}$  is **uniquely decodeable**.

insert example here

insert example here

insert example here

Let  $\mathfrak{C}_0 = \mathfrak{C}$ :

- $\mathfrak{C}_n := \{\omega \in T \cup T^2 \cup \dots \mid u\omega = v \text{ for some } u \in \mathfrak{C}_{n-1}, v \in \mathfrak{C} \text{ or } u\omega = v \text{ for some } u \in \mathfrak{C}, v \in \mathfrak{C}_{n-1}\}$
- $\mathfrak{C}_\infty := \bigcup_{k \geq 1} \mathfrak{C}_k$

Since everythig is finite either  $\mathfrak{C}_m = \emptyset$  for some  $m$  and then  $\mathfrak{C}_n = \emptyset$  for  $n \geq m$  or it will be periodic and start repeating.

**Theorem 1:**  $\mathfrak{C}$  is uniquely decodeable  $\iff \mathfrak{C} \cap \mathfrak{C}_\infty = \emptyset$ .

*proof:* Insert proof here

insert example here

insert example here

insert example here

**Definition 5:** A code is a **prefix-code** if no codeword is prefix of another (ie.  $\mathfrak{C}_1 = \emptyset$ ).

A prefix code is uniquely decodeable.

**Theorem 2: (Kraft's inequality)**  $\exists r$ -ary prefix code with word lengths  $l_1, l_2, \dots, l_q \iff$

$$\sum_{i=1}^q r^{-l_i} \leq 1$$

*proof:* Insert proof here

insert example here

**Theorem 3: (McMillan's inequality)**  $\exists r$ -ary uniquely decodeable code with word lengths  $l_1, l_2, \dots, l_q \iff$

$$\sum_{i=1}^q r^{-l_i} \leq 1$$

*proof:* Insert proof here

### 1.3. Optimal codes

Let be  $\mathcal{S}$  a source with symbols  $s_1, \dots, s_q$  emitted with probabilities  $p_1, \dots, p_q$  and  $\mathfrak{C}$  is a code which encodes  $s_i$  with a codeword length  $l_i$ . Recall  $L(\mathfrak{C}) = \sum_{i=1}^q p_i l_i$ .

**Definition 6:** An **optimal code** for  $\mathcal{S}$  is an uniquely decodeable code  $\mathfrak{D}$  such that  $L(\mathfrak{C}) \geq L(\mathfrak{D})$  for all unique decodeable code  $\mathfrak{C}$ .

inset example here

insert example here

**Definition 7:** A code constructed in this way is called a **Huffman code**.

insert example here

Construct the  $r$ -arg Huffman code we sum together (at each step) the  $r$  smallest probabilities.

For this to work we need  $q \equiv 1(r-1)$ . Recall  $q$  is the number of symbols in the source. If not, then we add symbols with probabilities zero so that it is.

insert example here

**Lemma 1:** Every source  $\mathcal{S}$  has an optimal binary code  $\mathfrak{D}$  in which two of the longest codewords are **siblings**, ie.  $\exists x$  (a string) such that  $x_0, x_1 \in \mathfrak{D}$ .

*proof:* Insert proof here

**Theorem 4:** The Huffman code is an optimal code.

*proof:* Insert proof here

## 1.4. Extension of sources

Given a source  $\mathcal{S}$  we define  $\mathcal{S}^n$  the source with  $|\mathcal{S}|^n$  symbols, typically  $s_1, \dots, s_n$ , emitted with  $p_1, \dots, p_n$  probabilities.

insert example here

## 2. Information and entropy

### 2.1. Definitions

**Definition 1:** the **information** conveyed by a source is a function  $I : S \rightarrow [0, \infty)$  where  $S$  is a **source**<sup>1</sup> with the properties:

- $I(s_i)$  is a decreasing function of the probability  $p_i$ , with  $I(s_i) = 0$  if  $p_i = 1$ .
- $I(s_i s_j) = I(s_i) + I(s_j)$ , ie. the information gained by two symbols is the sum of the information obtained from each where the source has symbols  $s_1, \dots, s_q$  emitted with probabilities  $p_1, \dots, p_q$ .

**Lemma 1:**  $I(s_i) = -\log_r p_i$  for some  $r$ .

*proof:* Insert proof here

---

**Definition 2:** The  $r$ -ary **entropy**  $H_r(S)$  of a source  $S$  is the average information conveyed by  $S$ .

$$H_r(S) := - \sum_{i=1}^q p_i \log_r p_i$$

, by convention  $x \log_r x$  evaluated at 0 is 0.

Insert five examples

### 2.2. Properties of the entropy function

**Theorem 1:**  $H_r(S) \leq \log_r q$  with equality if and only iff  $S$  is the source where each symbol is emitted with probability  $1/q$ .

*proof:* Insert proof here

---

**Theorem 2:**  $H_r(S) \leq L(C)$  for unique decodeable code  $C$ .

*proof:* Insert proof here

---

### 2.3. Shannon-Fano Code

Let  $S$  be the source with symbols  $s_i$  and probabilities  $p_i$ . Let  $l_i := \lceil \log_r 1/p_i \rceil$ .

Then:  $\sum_{i=1}^q r^{-l_i} \leq \sum r^{-\log_r 1/p_i} = \sum p_i = 1$

---

<sup>1</sup>A **source** is a finite set  $S$  together with a sequence of random variables  $X_i$  whose range is  $S$

**Definition 3:** by Kraft exists a prefix code with word length  $l_1, l_2, \dots, l_1$ . This code is called **Shannon-Fano code**.

Inert example here

**Lemma 2:** For the Shannon-Fano code  $C$ :  $H_r(S) \leq L(C) < H_r(S) + 1$ .

*proof:* Insert proof here

---

## 2.4. Product of sources

Let  $S$  and  $T$  be two memoryless sources,  $S$  with symbols  $s_i$  and probabilities  $p_i$  and  $T$  with symbols  $t_j$  and probabilities  $q_j$ .

**Definition 4:** The **product source**  $S \times T$  is a source with symbols  $s_i t_j$  and probabilities  $p_i q_j$ .

**Theorem 3:**  $H_r(S \times T) = H_r(S) + H_r(T)$ .

*proof:* Insert proof here

---

**Corollary 1:**  $H_r(S^n) = nH_r(S)$ .

**Theorem 4: Noiseless Coding** The average word length  $L_n$  of an optimal code of  $S^n$  satisfies:

$$\frac{L_n}{n} \rightarrow H_r(S), n \rightarrow \infty$$

*proof:* Insert proof here

---

some examples

## 2.5. Markov Chains

**Definition 4:** A **Markov Chain** is a sequence of random variables where  $X_{n+1}$  depends only for  $X_n$ .

$$P(X_{n+1} = s_j | X_n = s_i) = p_{i,j}$$

This can be represented in a direct graph and also by a matrix  $P := (p)_{i,j}$ .

Suppose  $u_0$  is the vector which describes the initial distribution, ie. the  $i$ -th coordinate of  $u_0$  is probability we start at  $s_i$ . Probability of being in the  $i$ -th state after  $r$  steps is the  $i$ -th coordinate of  $u_0 P^r$ .

**Theorem 5:** if  $\exists r \in \mathbb{N}$  such that  $P^r$  has no zero entries, then  $u_0 P^r \rightarrow u$ , as  $n \rightarrow \infty$ .



**Definition 5:** This vector  $u$  is called the **stationary distribution**. It is normalised eigenvector of  $P^t$  with eigenvalue 1, ie.  $u_j = \sum_i p_{i,j} u_i$  and  $\sum_j u_j = 1$ .

**Definition 6:** If  $P$  is the matrix of a Markov Chain and  $\exists r$  such that  $P^r$  has non zero entries then we say that the Markov Chain is **regular**.

## 2.6. Sources with memory

Suppose  $S$  is a Markov Chain source with random variables  $X_1, X_2, \dots$  such that

$$P(X_{n+1} = s_j | X_n = s_j) = p_{i,j}$$

**Definition 7:**  $S$  is **not memoryless**, but it is stationary.

**Theorem 6:** suppose  $S$  is a regular Markov Chain source with stationary distribution  $u = (u_1, \dots, u_n)$ . Let  $S'$  be the stationary memoryless source with the same source elements as  $S$  (where  $s_i$  is emitted with probability  $w_i$ ). Then:

$$H_r(S) \leq H_r(S')$$

*proof:* Insert proof here

---

### 3. Information channels

#### 3.1. Channel matrix

Let  $\mathcal{A}$  be a stationary memoryless source with random variables  $X_1, X_2, \dots$  where  $P(X_n = a_i) = p_i$  for  $a_i \in \mathcal{A}$ .

Suppose we transmit  $\mathcal{A}$  through a channel  $\Gamma$ .

Let  $\mathcal{B}$  be a source with random variables  $Y_1, Y_2, \dots$  where  $P(Y_n = b_j) = q_j$

For  $b_j$  emerging from the channel:

$$\mathcal{A} \xrightarrow{\Gamma} \mathcal{B}$$

**Definition 1:** The **channel** is defined by a matrix  $(p_{ij})$  where  $p_{ij} = P(X_n = b_j | X_n = a_i)$  the probability we receive  $b_j$  given that  $a_i$  was sent,  $p_{ij}$ -**forward probabilities**. The **backwards probabilities** are  $q_{ij} = P(X_n = a_i | Y_n = b_j)$  and **joint probabilities**  $r_{ij} = P(X_n = a_i, Y_n = b_j)$

insert example here

inser example here (binary erasure channel)

#### 3.2. System Entropies and mutual information

**Definition 2:** We define the **input entropy** as:

$$H(\mathcal{A}) := - \sum_i p_i \log(p_i)$$

**Definition 3:** We define the **output entropy** as:

$$H(\mathcal{B}) := - \sum_j q_j \log(q_j)$$

We suppress the  $r$  (base) in the  $\log_r$  but it's always the same for every one.

Given that we have received  $b_j \in \mathcal{B}$ ,  $H(\mathcal{A} | Y_n = b_j) = - \sum_i q_{ij} \log(q_{ij})$ .

This is telling us the average information of  $\mathcal{A}$  knowing that  $Y_n = b_j$ .

If  $H(\mathcal{A} | Y_n = b_j) = 0$  then  $\exists m$  such that  $q_{ij} = 0$  for all  $i \neq m$  and  $q_{ij} = 1$  if  $i = m$ , ie.  $P(X_n = a_m | Y_n = b_j) = 1$ , ie. if we receive  $b_j$  then we know that  $a_m$  was sent.

If  $H(\mathcal{A} | Y_n = b_j) = H(\mathcal{A})$  then we learn nothing about  $\mathcal{A}$  when we receive  $b_j$  and this occurs when  $q_{ij} = P(X_n = a_i | Y_n = b_j) = P(X_n = a_i) = p_i$ .

**Definition 4:** Averaging over  $b_j \in \mathcal{B}$  we get the **condicional entropy**:

$$H(\mathcal{A} | \mathcal{B}) := - \sum_j P(Y_n = b_j) H(\mathcal{A} | Y_n = b_j) = - \sum_{i,j} q_j q_{ij} \log q_{ij}$$

Similary:

$$H(\mathcal{B}|\mathcal{A}) := - \sum_{i,j} p_i p_{ij} \log p_{ij}$$

**Definition 5:** The **joint entropy**:

$$H(\mathcal{A}, \mathcal{B}) := - \sum_{i,j} r_{ij} \log r_{ij}$$

insert example here

**Theorem 1:** For sources  $\mathcal{A}$  and  $\mathcal{B}$ :

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}|\mathcal{B}) + H(\mathcal{B}) = H(\mathcal{B}|\mathcal{A}) + H(\mathcal{A})$$

*proof:* Insert proof here

**Definition 6:** We define the **mutual information** as the amount of information about  $\mathcal{A}$  we have learnt from  $\mathcal{B}$  and vice-versa:

$$I(\mathcal{A}, \mathcal{B}) := H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}) = H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B})$$

If  $H(\mathcal{A}) = H(\mathcal{A}|\mathcal{B})$  then  $\mathcal{B}$  tells us nothing about  $\mathcal{A}$ , so  $I(\mathcal{A}, \mathcal{B}) = 0$ . This is an unrialiable channel and useless as a mean of communication.

If  $H(\mathcal{A}|\mathcal{B}) = 0$  then knowing  $\mathcal{B}$  we know everythin about  $\mathcal{A}$ , so  $I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A})$ . This is the perfect situation because when we recive something, we know exactly what was sent.

insert example here

### 3.3. Extension of noiseless coding theorem to information channels

We have proved that given a source  $\mathcal{A}$  we can find an encoding of  $\mathcal{A}^n$  such that the average word lenglht  $L_n$  satisfies  $\frac{L_n}{n} \rightarrow H(\mathcal{A})$ .

$\mathcal{A} \rightarrow \mathcal{B}$ , imagine we know  $\mathcal{B}$ .

**Lemma 1:**  $H(\mathcal{A}^n|\mathcal{B}^n) = nH(\mathcal{A}|\mathcal{B})$

*proof:* EXERCISE

**Theorem 2:** if  $\mathcal{B}$  is know then we can find encodings of  $\mathcal{A}^n$  such that the average word length  $L_n$  satisfies  $\frac{L_n}{n} \rightarrow H(\mathcal{A}|\mathcal{B})$ .

*proof:* Insert proof here

**3.4. Decision rules**

**3.5. Improving reliability**

**3.6. Rates of transmission and Hamming distance**

## 4. Finite fields

### 4.1. Basic definitions

### 4.2. Properties of finite fields

### 4.3. Factorization of polynomials

## 5. Block codes

### 5.1. Minimum distance

### 5.2. Bounds on block codes

### 5.3. Asymptotically good codes

## **6. Linear codes**

### **6.1. Basics**

### **6.2. Syndrom decoding**

### **6.3. Dual code and Mc Williams identities**

### **6.4. The Griesmer bound**

## 7. Cyclic codes

### 7.1. Introduction

### 7.2. Quadratic residue codes

### 7.3. BCH Codes

Decision problem, yes/no problem



## 8. Maximun distance separable codes

### 8.1. Syngleton bound

### 8.2. Linear MDS codes

## 9. Alternant codes

## 10. Low density parity check codes

### 10.1. Bipartite graphs with the expander property

### 10.2. Low density parity check (LDPC) codes

### 10.3. Belief propagation

## 11. P-adic codes

Breve comenterio

### 11.1. P-adic numbers

### 11.2. Polynomials over $\mathbb{Q}_p$