Code Theory

Manuel Gijón Agudo

September 2017 - January 2018

Índice

1. Memoryless resources **3**

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

2. Information and entropy **6**

2.1. Definitions 6

2.2. Properties of the entropy funcion 6

2.3. Shannon-Fano Code 6

2.4. Product of sources 7

2.5. Markov Chains 7

2.6. Sources with memory 8

3. Information channels **9**

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 transmision and Hamming distance 12

4. Finite fields **14**

4.1. Basic definitions 14

4.2. Propierties of finite fields 14

4.3. Factorization of polynomials 15

5. Block codes **17**

5.1. Minimun distance 17

5.2. Bounds on block codes 17

5.3. Asymptotically good codes 18

6. Linear codes **20**

6.1. Basics 20

6.2. Syndrom decoding 21

6.3. Dual code and Mc Williams identities 21

6.4. The Griesmer bound 22

7. Cyclic codes **23**

7.1. Introduction 23

7.2. Quadratic residue codes 23

7.3. BCH Codes 23

7.4. Decision problem, yes/no problem 24

8. Maximun distance separable codes **25**

8.1. Syngleton bound 25

8.2. Linear MDS codes 25

9. Alternant codes **27**

10. Low density parity check codes **28**

10.1. Bipartite graphs with the expander property 28

10.2. Low density parity check (LDPC) codes 28

10.3. Belief propagation 28

11. P-adic codes **29**

11.1. P-adic numbers 29

11.2. Polynomials over $\mathbb{Q}_p$ 29

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 $(X_1, X_2, ...)$ 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 $|T| = r$ then $\mathfrak{C}$ is a r**-ary code**.

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

insert example here

Definition 3: The **average word-lenght** 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 sequencies $u_1...u_n = v_1...v_m$ in $\mathfrak{C}$ implies $m = n$ and $u_i = v_i$ for $i = 1, ..., n$ then we say that $\mathfrak{C}$ is **uniquely decodeable**.

insert example here

insert example here

insert example here

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

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

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 $\Longleftrightarrow \mathfrak{C} \cap \mathfrak{C}_\infty = \emptyset$.

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 $l_1, l_2, .., l_q \iff$

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

proof: Insert proof here

insert example here

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

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

proof: Insert proof here

1.3. Optimal codes

Let be $\mathcal{S}$ a source with symbols $s_1, ..., s_q$ emitted with probabilities $p_1, ..., p_q$ 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 uniquel decodeable code $\mathfrak{C}$.

inset example here

insert example here

Definition 7: A code constructed in this way is called a **Hoffman 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 $s_1, ..., s_n$, emitted with $p_1, ..., p_n$ probabilities.

insert example here

2. Information and entropy

2.1. Definitions

Definition 1: the **information** coveyed by a source is a function $I : S \to [0, \infty)$ where S is a **source** [1] with the properties:

- $I(s_i)$ is a decreasing function of the propability 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 geined by two symbols is the sum of the information obtained from each where the source has symbols $s_1, ..., s_q$ emitted with probabilities $p_1, ..., p_q$.

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 coveyed by S.

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

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

Insert five examples

2.2. Properties of the entropy funcion

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

[1] 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 woed length $l_1, l_2, ..., l_1$. 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 optiml code of S^n satisfies:

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

proof: Insert proof here

---

some examples

2.5. Markov Chains

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

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

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 beeing 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 $u_0 P^r \longrightarrow u$, as $n \to \infty$.

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 $X_1, X_2, ...$ 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 $u = (u_1, ..., u_)$. Let S' be the stationary memoryless source with the same source elements as S (where s_i is emmitted 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 $X_1, X_2, ...$ where $P(X_n = a_i) = p_i$ for $a_i \in \mathcal{A}$.

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

Let $\mathcal{B}$ be a source with random variables $Y_1, Y_2, ...$ 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 recieve 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 prababilities** $r_{ij} = P(X_n = a_i, Y_n = b_j)$

insert example here

inser example here (binary eraure 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 recived $b_j \in \mathcal{B}$, $H(\mathcal{A}|Y_n = b_j) = -\sum_i q_{ij} \log(q_{ij})$.

This is relling 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 recieve 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 recieve 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 lenght L_n satisfies $\frac{L_n}{n} \longrightarrow H(\mathcal{A})$.

$\mathcal{A} \longrightarrow \mathcal{B}$, 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 $\frac{L_n}{n} \longrightarrow H(\mathcal{A}|\mathcal{B})$.

proof: Insert proof here

3.4. Decision rules

$$\mathcal{A} \overset{\Gamma}{\longrightarrow} \mathcal{B}$$

Where $\mathcal{A}$ is the **input**, $\mathcal{B}$ is the **output** and Γ is the **channel**.

The channel is given by a matrix (p_{ij}), $p_{ij} = P(Y_n = b_j|X_n = a_i)$. We defined $r_{ij} = P(X_n = a_i|X_n = b_j)$.

So if we recive b_J we should "decode" b_j as a_{j*} where $r_{j*j} \geq r_{ij}$ for all i.

Definition 7: We would define our decision $\Delta : \mathcal{B} \longrightarrow \mathcal{A}$ as $\Delta(b_j) := a_{j*}$, this is called the **ideal observer rule**.

Howecer, most likely we only know p_{ij}'s.

Definition 8: In **maximun likelihood decoding** we use the decision rule $\Delta(b_j) := a_{j*}$, where $p_{j*j} \geq p_{ij}$ for all i.

Definition 9: The **average probability of a correct decoding** is:

$$P_{cor} := \sum_j q_j q_{j*j} - \sum_j r_{j*j}$$

Remind $q_{ij} = P(X_n = a_i|Y_n = b_j)$. Given that we recived b_j if we dcode it as a_{j*} then the probability we have decoded correctly is $P(X_n = a_{j*}|Y_n = b_j) = q_{j*j}$

3.5. Improving reliability

Suposse Γ is the binary symmetrical channel $\begin{pmatrix} \phi & 1-\phi \\ 1-\phi & \phi \end{pmatrix}$ (and assume $\phi > \frac{1}{2}$).

If we extends the source $\mathcal{A} = \{0,1\}$ to $\{000, 001\}$ then the outpout source if $\{000, 001, 010, 100, 110, 101, 011, 111\}$. Now we have the channel matrix:

$$\begin{pmatrix} \phi^3 & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & (1-\phi)^3 \\ (1-\phi)^3 & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^2(1-\phi) & \phi^3 \end{pmatrix}$$

if we decode $\Delta(000) = \Delta(001) = \Delta(010) = \Delta(100) = 0$ and $\Delta(111) = \Delta(110) = \Delta(101) = \Delta(011) = 1$.

effectively we have the channel:

$$\begin{pmatrix} \phi^3 + 3\phi^2(1-\phi) & 3\phi^2(1-\phi) + (1-\phi)^3 \\ 3\phi^2(1-\phi) + (1-\phi)^3 & \phi^3 + 3\phi^2(1-\phi) \end{pmatrix}$$

since $\phi > 1 - \phi$ we have $\phi^3 + 3\phi^2(1-\phi) > \phi$.

So we have proved the reliability of the channel, because $P_{cor} = \sum_j r_{j*j} = p(\phi^3 + 3\phi^2(1-\phi)) + (1-p)(\phi^3 + 3\phi^2(1-\phi) = \phi^3 + 3\phi^2(1-\phi)$.

Observe if we do not extend the sorce $P_{cor} = \phi$.

3.6. Rates of transmision and Hamming distance

noindent Suppose $\mathcal{A}$ is a source with r symbols. By extending the source, consider $\mathcal{C}$ to be a subset of $\mathcal{A}^n$.

Definition 10: The **(transmision) rate of** $\mathcal{C}$ is:

$$R := \frac{\log_r |\mathcal{C}|}{n}$$

By increasing n in the previous exercise we can make $P_{cor} \longrightarrow 1$. However $R \longrightarrow 0$ since $|\mathcal{C} = \frac{\log_2 2}{n} \longrightarrow 0$.

Definition 11: The **capacity of a channel** Γ is:

$$\Lambda = \max_{\mathcal{A},\mathcal{B}} I(\mathcal{A},\mathcal{B})$$

Maximising over $\mathcal{A},\mathcal{B}$ means we can vary p_i's and q_j's.

Since $\mathcal{C}$ is a subset of $\mathcal{A}^n$ the rate tell us how many bits od information we can send in n bits (it is Rn).

Lemma 2: The capacity of a binary symetric channel $\begin{pmatrix} \phi & 1-\phi \\ 1-\phi & \phi \end{pmatrix}$ is $\Lambda = 1 + \phi \log_2 \phi + (1-\phi) \log_2 (1-\phi)$.

proof: Insert proof here

Definition 12: For any $u, v \in \mathcal{A}^n$, the **Hamming distance** is $d(u,v) :=$ number of cordinates where u and v differ.

Lemma 3: The Hamming distande satisfies the triangle inequality $d(u,v) \leq d(u,w) + d(w,v)$

proof: Insert proof here

Lemma 4: Fot the binary symmetric channerl, maximun likelihood decoding is $\Delta(v) = u$, where u is the closest element of $\mathcal{C}$ with respect the Hamming distance.

proof: Insert proof here

Definition 13: in general this decoding is called **nearest neighbour decoding**.

Lemma 5: For $0 < \lambda < \frac{1}{2}$:

$$\sum_{i=0}^{\lambda n} \binom{n}{i} \leq 2^{n(-\lambda \log(\lambda) - (1-\lambda)\log(1-\lambda))}$$

proof: Insert proof here

Theorem 2: (**Shannon**) Let $\delta, \varepsilon > 0$. For all sufficiently large n there is a code of length n and rate R satisfying $\Lambda - \varepsilon < R < \Lambda$ together with a decision rule Δ such that $P_{cor} \longrightarrow 1 - \delta$.

proof: Insert proof here (ONLY FOR BINARY SYMETRIC CHANNEL)

Lemma 6: For an input source $\mathcal{A}$ and output source $\mathcal{B}$ and decision rule $\Delta(b_j) = a_{j*}$.

$$H(\mathcal{A}|\mathcal{B}) \leq -P_{cor}\log(P_{cor}) - (1-P_{cor})\log(1-P_{cor}) + (1-P_{cor})(\log|\mathcal{C}| - 1)$$

where $\mathcal{C}$ is the set of input source elements emitted with non zero probability.

Theorem 3: If $\Lambda^{‘} > \Lambda$ and we fix the input probability distribution is uniform then ther is no sequence of codes C_n of rate R satisfying $\Lambda^{‘} - \varepsilon < R < \Lambda^{‘}$ such that $P_{cor} \longrightarrow 1$ as $n \to \infty$.

proof: Insert proof here

4. Finite fields

4.1. Basic definitions

Definition 1: A **field** is a commutable ring in which every non-zero element has a multiplicative inverse.

insert example here

inse example here

Notation 1: We denote as (f) with $f \in \mathbb{F}_p[X]$, the **ideal consisiting of all multples of f**.

Theorem 1: if f is an irreducible polynomial of degree h then $\mathbb{F}_p[X]/(f)$ is a finite field with p^h elements.

proof: Insert proof here

insert examples here

Exercise : construct a field wih 9 elements.

Let $\mathbb{F}$ be a finite field. Ler n minimal such that adding 1 n times gives 0.

Since $\overbrace{(1 + ... + 1)}^{n} = \overbrace{(1 + ... + 1)}^{r}\overbrace{(1 + ... + 1)}^{n/r} = 0$ minimaliy implies that $n = p$ is prime.

Definition 2: In this situation, we say that $\mathbb{F}$ has **characteristic** p. If no such p exits then wa sat that $\mathbb{F}$ has **characteristic zero**, in which case $\mathbb{F} \supset \mathbb{Z}$ and so $\mathbb{F} \supseteq \mathbb{Q}$.

insert exercise here

4.2. Propierties of finite fields

Theorem 2: Ler $\mathbb{F}$ be a field with q elements. For all $x \in \mathbb{F}$. $x^q = x$.

proof: Insert proof here

The finite field with q elements is unique since it is the splitting field of the polynomial $x^t - x \in \mathbb{F}_p[X]$.

Considerer the map $x \longmapsto x^p$ in $\mathbb{F}$ ($q =$ ‘p^h).

$$(x + y)^p = \sum_{j=0}^{p} \binom{p}{j} x^j y^{p-j} = x^p + y^p$$

Observe that $\binom{p}{j} = 0$ (modulo p) for $j = 1, ..., p - 1$.

$$(x * y)^p = x^p y^p$$

So this map os catiallu an automorphism of $\mathbb{F}_p$ since ot preserve addiction and multiplication.

Definition 3: This is called the **Frobenious automorphism**.

$$x \longmapsto x^p \longmapsto x^{p^2} \longmapsto x^{p^3} \longmapsto ... \longmapsto x^{p^{h-1}} \longmapsto x$$

4.3. Factorization of polynomials

Let $\mathbb{F}_p$ denote the unique finite field with q elements ($q = p^h$).

Lemma 1: The polynomial $x^{q-1} - 1$ factories into distinct linear factors in $\mathbb{F}_q[X]$.

proof: Insert proof here

Lemma 2: The polynomial $x^q - 1$ factories into distinct irreducible factors whose degre divides h.

proof: Insert proof here

insert example here

insert example here

Observation 1: if q is odd $x^{q-1} - 1 = (x^{\frac{q-1}{2}} - 1)(x^{\frac{q-1}{2}} + 1)$ the zeros of the first dactor anr on the non-zeros squares in $\mathbb{F}_q$ and vice-versa ($x = y^2$ then $x^{\frac{q-1}{2}} = y^{q-1} = 1$).

Observation 2: if $q^1 = q^r$ then $x^n - 1 = (x^{n/q'} - 1)^{q`}$ so if we want to factorise $x^n - 1$ in $\mathbb{F}_p[x]$ we can assume $(n, p) = 1$.

To factorise $x^n - 1$ in $\mathbb{F}_q[X]$, we find and extension field in $\mathbb{F}_q$ which contains n-th roots of 1, ie. find h such that n divides $q^h - 1$ since then $x^{q^n - 1} - 1$ is divisible by $x^n - 1$, ie. $q^n = 1 \pmod 1$, ie. h is the multiplicative order of q in $\mathbb{Z}/n\mathbb{Z}$.

If we let ε be a primitive n-th root of 1 in $\mathbb{F}_{q^n}$ then $(x - \varepsilon)(x - \varepsilon^q)(x - \varepsilon^{q^2})...(x - \varepsilon^{q^{h-1}})$ is a polynomial whose coefficients are in $\mathbb{F}_q$ since $(x - \varepsilon)(x - \varepsilon^q)(x - \varepsilon^{q^2})...(x - \varepsilon^{q^h})$.

insert example here

insert exercise here

insert example here

5. Block codes

5.1. Minimun distance

Let $\mathcal{A}$ be a finite set (an alphabet).

Definition 1: A **block code** $\mathfrak{C}$ of length n is a subset of $\mathbb{A}^n$.

Definition 2: The **minimun distance of** $\mathfrak{C}$ is the minumun Hamming distance berween any 2 codewords (elements of $\mathfrak{C}$).

We are goning to use nearest neightbour decoding so we want d as larde as possible. We also cant $|\mathfrak{C}|$ to be as large as posible.

Lemma 1: A block code of minimun distance d can correct up to do $\lceil \frac{d-1}{2} \rceil$ errors using nearest neightbour decoding.

proof: Insert proof here

insert example here

insert example here

Definition 3: Let $\mathfrak{C}$ be a binary code of lenght n. The **extended code** $\overline{\mathfrak{C}}$ is the code of length $n+1$ defined by:

$$\overline{\mathfrak{C}} := \{(u_1, ..., u_{n+1}) : u \in \mathfrak{C} \text{ where } u_{n+1} = u_1 + ... + u_n (\text{mod } 2)$$

Theorem 1: if the minimun distande d of a binary code is odd then the minimun distance of $\overbrace{\mathfrak{C}}$ is $d+1$.

proof: Insert proof here

5.2. Bounds on block codes

Let $\mathcal{A}_r(n, d)$ denote the maximun $|\mathfrak{C}|$, such that exits a block code $\mathfrak{C}$ of length n, minimun distance d over an alphabeth with r-elements.

Theorem 1: (**Gilbert-Varshamov Bound**)

$$\mathcal{A}_r(n, d)\Big(1 + \binom{n}{1}(r-1) + ... + \binom{n}{d}(r-1)^d\Big) \geq r^n$$

proof: Insert proof here

Recall 1: we defined the binary entropy function as $h(p) = -p \log p - (1-p) \log (1-p)$.

Corollary 1: in the case $r = 2$:

$$\frac{1}{n} \log_2 \mathcal{A}_2(n.d) \geq 1 - h(\delta) \text{ , where } \delta = \frac{d}{n}$$

Definition 4: $\delta = \frac{d}{n}$ is called **relative minimun distance**.

proof: Insert proof here

Theorem 2: (**Sphere packing bound**)

$$\mathcal{A}_r(n,d)\Big(1+\binom{n}{d}(r-1)+...+\binom{n}{t}(r-1)^t\Big) \leq r^n \text{ where } t = \left\lfloor \frac{d-1}{2} \right\rfloor$$

proof: Insert proof here

Definition 5: A code meeting the Spheree-packing bound is called **perfect code**.

Observation 1: the parameteres (n,t,r) must be such that:

$$1+\binom{n}{d}(r-1)+...+\binom{n}{t}(r-1)^t \text{ is a power of } r$$

insert example and exercise here

Lemma 2: (**Plotking Lemma**) An r-ary code $\mathfrak{C}$ of length n and minimun distance d satisfies $|\mathfrak{C}|$ $(d+\frac{n}{r}-n) \leq d$.

proof: Insert proof here

insert exercise here

Theorem 3: (**Plotkin-Bound**) if $\mathfrak{C}$ is a binary code of length n, minimun distance $d < \frac{n}{2}$. then:

$$|\mathfrak{C}| \leq d2^{n-2d+2}$$

proof: Insert proof here

5.3. Asymptotically good codes

We will construct and use short length codes which we can encode and decoode quickly, this is very useful in manyaplications.

insert short examples here

However, in many cases we will have a lot of data and if we chop n bits into $\frac{n}{n_0}$ chunks which we can send with $P_{cor} = P$ close to 1.

$$P^{\frac{n}{n_0}} \longrightarrow 0$$

Let's suppose we have a binary code of length n and rate R (so $|\mathfrak{C}| \approx 2^{nR}$).

In the proof of the Shannon's Theorem, we wed to the fact that the expected number of errors (using the binary symmetric channel) was $(1-\phi)n$, so if we are going to use the nearest neighbour decoding we need that d is also linear in n (as n gets very large), so we want $\delta = \frac{d}{n} > 0$.

Definition 5: We call the sequency codes of length n, where $n \to \infty$ and $\delta > 0$. $R > 0$. **asumptotically good**.

inset exercise here

Theorem 4: (**Sprieve packing bound**) Asymptotically (for n large):

$$R \leq 1 - h\Big(\frac{\delta}{2}\Big)$$

proof: Insert proof here

Theorem 5: (**Plotkin**) if $\delta \leq \frac{1}{2}$ then $R \leq 1 - 2\delta$.

proof: Insert proof here

Definition 6: Let $\mathcal{A}(n, d, \omega)$, **The maximun size** of a binary code of length n with minimun distance d in which all the codewords have weight ω.

(For any tuple $v \in \mathcal{A}^n$ where $0 \in \mathcal{A}$, the **weight** $wt(v) :=$ { number of non-zero coordinates that it has}).

Lemma 3:

$$\mathcal{A}(n, d, \omega) \leq \frac{nd}{2\omega^2 - 2n\omega + dn}$$

proof: Insert proof here

CONJETURE: there's no perfect constant (apart from the trivial bounds) weight codes.

Theorem 6: Let R be the rate of a sequence of asymptotucally good binary codes if $\delta < \frac{1}{2}$ then:

$$R < 1 - h\Big(\frac{1}{2}\big(1 - \sqrt{1 - 2\delta}\big)\Big)$$

where $h(p) = -p \log_2(p) - (1-p)\log_2(1-p)$

6. Linear codes

6.1. Basics

Definition 1: Let $\mathcal{A} = \mathbb{F}_q$. If $\mathcal{C}$ is a subspace of $\mathfrak{F}_q^n$ then we say $\mathcal{C}$ is a **linear code**.

Id $\mathcal{C}$ is a k-dimensional subspace the $|\mathcal{C}| = q^k$.

Definition 2: For $v \in \mathbb{F}_q^n$, $wt(v) :=$ {number of non-zero coordinates that it has}.

Lemma 1: (**Minimun Weight Lemma**) the minimun distance of a linear code $\mathcal{C}$ is equal to te minimun non-zero weight of the vector in $\mathcal{C}$.

proof: Insert proof here

Definition 3: We can describe $\mathcal{C}$ ny a basis and if $\mathfrak{G}$ os a kxn matrix whose rows are a basis for $\mathcal{C}$ then we say that $\mathfrak{G}$ is a **generator matrix** for $\mathcal{C}$.

$$\mathcal{C} := \{u\mathfrak{G} : u \in \mathbb{F}_q^n\}$$

Linear codes encode q^k multiple mensajes by simply multiplying by a matriz:

$$u \longmapsto u\mathfrak{G}$$

$$\text{message} \longmapsto \text{codeword}$$

insert exercise here

Observation 1: The rate od a k-dimensional linnear code is:

$$R = \frac{\log |\mathcal{C}|}{n} = \frac{k}{n}$$

Definition 4: a **check matrix** for a linear code is an mxn matrix $\mathfrak{H}$ such that:

$$\mathcal{C} := \{u \in \mathbb{F}_q^n : u\mathfrak{H}^t = 0\}$$

insert example here

insert exercise here

Lemma 2: if $\mathfrak{G}$ is a check matrix for $\mathcal{C}$ and $\mathfrak{H}$ its check matrix then $\mathfrak{G}\mathfrak{H}^t = 0$.

proof: Insert proof here

insert example here

6.2. Syndrom decoding

Definition 5: Let $\mathcal{C}$ be a linear code with check matrix $\mathfrak{H}$. The **syndrome of a vector** $v \in \mathbb{F}_q^n$ is $s(v) := v\mathfrak{H}^t$, observe that $v \in \mathcal{C} \iff s(v) = 0$.

Suppose that $t = \left\lfloor \frac{d-1}{2} \right\rfloor$ and we correctly up to t errors to use syndrome decoding we calculate $s(e)$ for all vectors $e \in \mathbb{F}_q^n$ such that $wt(e) \leq t$.

Then if we recieve $v \in \mathbb{F}_q^n$ we look for e such that $s(v) = s(e)$ necaise this implies $s(v-e) = 0 \Rightarrow v - e \in \mathcal{C}$ and we have found the codeword.

insert 5 examples here

insert exercise here

6.3. Dual code and Mc Williams identities

Definition 6: Let $\mathcal{C}$ be a k-dimensional linear code of length n (ie. k-dimensional subspace of $\mathfrak{F}_2^n$). We denote by:

$$\mathcal{C}^\perp := \{v \in \mathbb{F}_q^h : uv = 0 \forall u \in \mathcal{C}\}$$

$\mathcal{C}^\perp$ is a $(n-k)$-dimensional code of length n.

$\mathcal{C}^\perp$ is the **dual code**.

Lemma 3: if $\mathfrak{H}$ is an $(nxk)xn$ check matrix for $\mathcal{C}$ then $\mathfrak{H}$ is a generator matrix for $\mathcal{C}^\perp$ likewise if $\mathfrak{G}$ is a (kxn) generator matrix for $\mathcal{C}$ then it is a check matrix for $\mathcal{C}^\perp$.

Definition 7: if $\mathcal{C} = \mathcal{C}^\perp$ then we say $\mathcal{C}$ is **self-dual**.

Observation 2: in a self-dual binary code the weight of a codeword is evens since $\overline{u}u = u = wt(u)$ must be zero.

Definition 8: Let $\mathcal{A}_i$ denote the number of codewords of eright of weight i. The **weight enumerator polynomial** is:

$$\mathcal{A}(t) := \sum_{i=0}^{n} \mathcal{A}_i t^i = \sum_{u \in \mathcal{C}} t^{wt(u)}$$

Theorem 1: Let $\mathcal{A}^\perp(t)$ be the weight enumerator for $\mathcal{C}^\perp$:

$$\mathcal{A}^\perp(t) = q^{-k}\big(1 + (q-1)t\big)^n \mathcal{A}\Big(\frac{1-t}{1+(q-1)t}\Big)$$

insert example here

insert example here

6.4. The Griesmer bound

Lemma 4: Let $\mathcal{S}$ be a set of columns if a kxn generator matrix $\mathfrak{G}$ for a linear code $\mathcal{C}$. $\mathcal{S}$ is a set of n vectors in $\mathfrak{F}_q^k$ with property that any hyperplane of $\mathfrak{F}_q^k$ contains at most $n-d$ vectors of $\mathcal{S}$.

proof: Insert proof here

Observation 3: Since there is a codeword of weight d threre is a hyperlplane of $\mathfrak{F}_q^k$ containing exactly $n-d$ vectors of $\mathcal{S}$.

proof: Insert proof here

Theorem 2: (**The Griesmer bound**) If there is a k-dimensioanl linear code over $\mathfrak{F}_q$ of length n and minimun distance d then:

$$n \geq \sum_{j=0}^{k-1} \left\lceil \frac{d}{q^j} \right\rceil$$

proof: Insert proof here

insert 4 examples here

7. Cyclic codes

7.1. Introduction

Definition 1: A linear code $\mathcal{C}$ is **cyclic** if $(c_0, c_1, ..., c_{n-1}) \in \mathcal{C} \Rightarrow (c_{n-1}, c_1, c_2, ..., c_{n-2}) \in \mathcal{C}$.

Observation 1: There is a $1-1$ correspondence between codewords in $\mathcal{C}$ and polynomials in $\mathfrak{F}_q/(x^n-1)$:

$$(c_0, c_1, ..., c_{n-1}) \longleftrightarrow c_0 + c_1 x + ... + c_{n-1}x^{n-1}$$

Lemma 1: A cyclic code corresponds an ideal in $\mathfrak{F}_q[X]/(x^n-1)$.

proof: Insert proof here

Lemma 2: The cyclic code $\mathcal{C} =< g >$, for some polynomial g divuding $x^n - 1$ and has dimension ar least n =degre(g).

proof: Insert proof here

Definition 2: For any polunomial h, h **reverse** is $\overleftarrow{h} := x^{\text{degre}(h)} h(x^{-1})$.

The following theorem implies $\dim(\mathcal{C}) = n-$ degre(g).

Theorem 1: The dual code of $< g >$ is ¡$\overleftarrow{h}$ $>$ where $g(x)h(x) = x^n - 1$.

(This implies $\dim(\mathcal{C}^\perp) \geq n - \deg(h) = \deg(g) \Rightarrow \dim(\mathcal{C}) = n-\dim(\mathcal{C}^\perp) \leq n - \deg(g)$).

proof: Insert proof here

insert example here

7.2. Quadratic residue codes

Let$n$ be a prime and q (a prime too) a square in $\mathbb{F}_n$.

insert example here.

Let α be a primitive n-root of 1 in some extension of $\mathbb{F}_q$. Let $g(x) = \prod_{r \text{ squares in } \mathbb{F}_n} (x - \alpha^r)$. g divides $x^n - 1$, so $< g >$ defines a cyclic code of length n over $\mathbb{F}_q$.

Definition 3: This g is called **quadratic residue code**.

insert example here

insert exercise here

7.3. BCH Codes

Let α be a primitive n-th root of 1 in some extension of $\mathbb{F}_q$ and suppose that $g \in \mathbb{F}_q[X]$ is the minimun gegree polynomial suche that $g(\alpha^j) = 0$ for $j = 1, ..., d_0$.

Definition 4: Then $< g >$ is a **BCH code**.

Theorem 2: The minimun distance of $< g >$ is a least d_0.

proof: Insert proof here

insert exercise here

insert 3 examples here

insert exercise here

Theorem 3: There is no infinite sequence of k-dimensional linear BHC codes of length n and minimun distance d with both $\frac{k}{n}$ and $\frac{d}{n}$ bounded away from zero.

7.4. Decision problem, yes/no problem

P

Decision problems which can be resolve with polynomial time algorithms. Polynomial in the input (which is given as a part of the problem).

Example: is n prime? Size of input is the number of bits $\approx \log_2 n$.

NP

Gives a positive answer to the problem we can checj with a polynomial time algorithm.

Example: does a graph have a Hamiltonian cycle?

Observation 2: Recall in syndrom,e decoding we calculate $s(v) = v\mathfrak{H}^t$ where v is the receved vector and we look for e whose $wt(e) \leq t$ $s(v) = s(e)$.

Does such a vector e exits is a decision problem, ie. given $\mathfrak{H}$ and s, does $\exists e$ of $wt(e) \leq t$ and $e\mathfrak{H}^t = s$.

Theorem 4: The decision problem of whether we can decode a linear code using syndrome decoding is NP.

proof: Insert proof here

8. Maximun distance separable codes

8.1. Syngleton bound

Theorem 1: If $\mathcal{C}$ is a r-ary code of length n and minimun distance d, then $|\mathcal{C}| \leq r^{n-d+1}$.

proof: Insert proof here

Definition 1: A **maximun distance separable (MDS) code** is a code reaching the Singleton bound.

Example:

Theorem 2: If $\mathcal{C}$ is a k-dimensional linear code of length n and minimun distance d over $\mathbb{F}_q$ then $k \leq n - d + 1$.

proof: Insert proof here

A k-dimensionl linear MDS code of length n has minumun distance $n - k + 1$.

Example: (Reed-Solomon code)

Theorem 3: There is a (fast) polynomial time algorithm for decoding Reed-Solomon codes.

proof: Insert proof here

8.2. Linear MDS codes

Theorem 4: Let $\mathcal{S}$ be the set of columns of a generator matriz $\mathfrak{G}$ of a k-dimensional linear MDS code over $\mathbb{F}_q$. Then every k-subset of $\mathcal{S}$ is a basis of $\mathbb{F}_q^k$ (which is called an **arc**).

proof: Insert proof here

Theorem 5: if $k \geq q$ then a k-dimensionla linear MDS code has minimun distance at most 2.

proof: Insert proof here

Suppose we want to make a biunary code from a linear MDS code. We can write out every element of $\mathbb{F}_q$ as a tuple of $\{0, 1\}$ of length $\approx \log_2 q$.

Suppose MDS code has length N, then the binary code has length $n \approx N \log_2 q$ and q^k codewords. The binary code hase rate:

$$R = \frac{\log_2 |\mathcal{C}|}{n} \approx \frac{k \log_2 q}{N \log_2 q} = \frac{k}{N}$$

We can make R-S codes of any rate we like so this is good.

The **relative minimun distance** is $\approx \frac{d}{N \log_2 q} \left(=: \delta \right) = \frac{N-k+1}{n} \approx \frac{1}{\log_2 q} - \frac{k-1}{N \log_2 q} = \frac{1-k/n}{\log_2 q}$.

So for a fiz rate if we have an infinite sequence of codes of length n then $q \longrightarrow \infty$ as well, so $\delta \longrightarrow 0$. So this will not give us asymptotically good codes.

A long burst of errors provohes onlu a few errors in the MDS code. Our theorem frem before says we should concentrate on $k \leq q - 1$.

The Reed-Solomon code gives MDS code of length $n = q+1$ for $k-1 \leq q-2$ (since $x^{q-1} = 1$ for $x \in \mathbb{F}_q - \{0\}$) are there any better codes.

If $k = 3$ and q even, then:

$$\mathfrak{G} = \begin{pmatrix} 1 & ... & 1 & 0 \\ x_1 & ... & x_p & 0 \\ x_1^2 & ... & x_p^2 & 1 \end{pmatrix}$$

we can extend the R-S code.

We have to check that:

$$\begin{vmatrix} 1 & 1 & 0 \\ x & y & 1 \\ x^2 & y^2 & 0 \end{vmatrix} = 0 \Rightarrow x = y$$

$$x^2 + y^2 = (x+y) = 0 \Rightarrow x = y (\text{q is even!!})$$

MDS CONJETURE: if $4 \leq k \leq q-2$ then a linear MDS code has length $q+1$.

insert exercise here

$\Rightarrow$ if dimension$(\mathcal{C}) = k$ then $\dim(\mathcal{C}^{\perp}) = n-k$ so we need to prove that minumun distance of $\mathcal{C}^{\perp}$ is $n-(n-k)+1 = k+1$.

The example with $k = 3$ of length $q+2$ gives us an example with $k = q-1$ of length $q+2$.

Lemma 3: A k-dimensional linear MDS code over $\mathbb{F}_q$ has length at most $q+k-1$.

proof: Insert proof here

Theorem 6: if q is prime then MDS conjeture is true. In fact $(q = p^n)$ if $k \leq p$ then $n \leq q+1$.

Theorem 7: if $k \leq p$ then a linear MDS code of length $q+1$ is a R-S code.

Example:

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 comentario

11.1. P-adic numbers

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