CO349 - Information and Coding Theory

Olivier Roques

Autumn 2018

Contents

1 Coding and Decoding

Definition 1.1 (Alphabet)**.** An *alphabet* is a finite set S. Its elements are called *symbols*. *Messages* are finite sequence of symbols.

Definition 1.2. $S^0 = \varepsilon$ is the neutral element. $S_n = \{m \mid |m| = n\}$. $S^* = \bigcup_{n\in\mathbb{N}} S^n$.

Definition 1.3 (Code)**.** A *code* is an injective function $c : S \to T^*$. $c(s)$ is the *codeword* for s. $C = \{c(s) \mid s \in S\}$ is also called code.

Definition 1.4 (Uniquely Decodeable)**.** c can be extended to S^*:

$$\begin{array}{rrcl} \widetilde{c}: & S^* & \longrightarrow & T^* \\ & s_1 s_2 \dots s_n & \longmapsto & c(s_1)c(s_2)\dots c(s_n) \end{array}$$

c is *uniquely decodeable* (UD) if $\widetilde{c}$ is injective.

Definition 1.5 (Prefix-Free)**.** $c : S \to T^*$ is *prefix-free* (PF) if and only if there is no pair of codewords $q = c(s)$, $q' = c(s')$ such that $\exists r \in T^*$, $r \neq \varepsilon$, $q = qr$.

Theorem 1.1. If $c : S \to T^*$ is prefix-free then c is uniquely decodeable.

All codewords can be seen as a finite path in a binary tree (if $T = \{0, 1\} = \mathbb{B}$). If C is prefix-free, none of a codeword's descendents can be a codeword.

Definition 1.6 (Parameter, Filling rate). We define $n_i = |\{s \in S \mid |c(s)| = i\}|$ for $i \in [\![0\,, M]\!]$ where M is the maximum length of a codeword as the *parameters* of $c : S \to T^*$.
$b_i = |T|^i$ is the number of all potential codewords of length i.
$\frac{n_i}{b_i}$ is the *filling rate.*

Definition 1.7 (Kraft-McMillan Number). The *Kraft-McMillan number* is defined as $K = \sum_{i=1}^{M} \frac{n_i}{b_i}$.

Theorem 1.2. Let T be an alphabet, $|T| = b$ and $n_1, \ldots, n_M$ some parameters. If $K \leq 1$ then there exists a code $c : S \to T^*$ prefix-free with those parameters.

Definition 1.8. We define $q_r = |\{s \in S^* \mid |s| = r, |\widetilde{c}(s)| = i\}$ as the number of strings of length r in S^* encoded in a string of length i in T^*.

Definition 1.9 (Generating Functions). For a sequence of number $q(1), q(2), \ldots$, the *generating function* Q is defined as $Q(x) = q(1)x + q(2)x^2 + \ldots$.
For the sequence $q_r(1), q_r(2), \ldots, q_r(rM)$, the generating function is $Q_r(x) = q_r(1)x + q_r(2)x^2 + \cdots + q_r(rM)x^{rM}$.

Theorem 1.3 (Counting Principle). If $c : S \to T^*$ is uniquely decodeable with $\forall s \in S\ |c(s)| \leq M$ and generating function Q_r then $Q_r = Q_1^r$.

Theorem 1.4. Let $c : S \to T^*$ be an uniquely decodeable code. Then $K_c \leq 1$.

Thus we get the result:

$$\exists c \text{ UD with parameters } n_1, \ldots, n_M \iff K \leq 1 \iff \exists c \text{ PF with parameters } n_1, \ldots, n_M$$

2 Probability Theory

We consider a finite *event space* Ω with $|\Omega| = n$ and a set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ of *measurable* set in Ω. Probabilities are then assigned via a *measure i.e.* a function $P : \mathcal{B} \to \mathbb{R}$.

Definition 2.1 (Probability). A *probability* $P : \mathcal{B} \to \mathbb{R}$ has to fulfill:
- $P(\Omega) = 1$
- $\forall A \in \mathcal{B}$, $P(A) \in [0\,, 1]$
- $\forall A$, $B \in \mathcal{B}$ such that $A \cap B = \emptyset$, $P(A \cap B) = P(A) + P(B)$

For a finite event space Ω, we can define a probability via atoms $\omega \in \Omega$.

Definition 2.2 (Probability Distribution). A *probability distribution* is a function $p : \Omega \to [0\,, 1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$. p can be represented as a row vector in $\mathbb{R}^n$ (in bold in the following).

Definition 2.3 (Random Variable). A *random variable* is a function $X : \Omega \to \mathbb{R}$. We can represent random variables as column vectors in $\mathbb{R}^n$.

Definition 2.4 (Expectation, Variance, Standard deviation). For a random variable X and probability distribution p:
- $\mathbf{E}[X] = \sum_{\omega \in \Omega} p(\omega) X(\omega)$
- $V[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$

- $\sigma_X = \sqrt{V[X]}$

Definition 2.5 (Covariance, Correlation)**.** The *covariance* of two random variables X and Y is $Cov(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$. The *correlation coefficient* is defined as $\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$. Covariance and correlation are equal to 0 when X and Y are independent.

Theorem 2.1 (Bayes Theorem)**.** The conditional probability of event $A \in \mathcal{B}$ given that $B \in \mathcal{B}$ has happened is

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \,|\, A)P(A)}{P(B)}$$

Definition 2.6 (Probability Product)**.** Given two probability spaces (Ω_1, P_1) and (Ω_2, P_2), we can define a probability P on the cartesian product $\Omega = \Omega_1 \times \Omega_2$ via:

$$P((\omega_1, \omega_2)) = P_1(\omega_1)P_2(\omega_2)$$

If P, P_1 and P_2 have probability distributions $\mathbf{p}$, $\mathbf{p_1}$ and $\mathbf{p_2}$ then $\mathbf{p} = \mathbf{p_1} \otimes \mathbf{p_2}$. However not all distributions on $\Omega_1 \times \Omega_2$ are product.

Definition 2.7 (Tensor product)**.** Given $A \in \mathcal{M}_{n,m}(\mathbb{R})$, $B \in \mathcal{M}_{k,l}(\mathbb{R})$, we define the *tensor product* $A \otimes B$ as

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \dots & a_{1,m}B \\ \vdots & \ddots & \vdots \\ a_{n,1}B & \dots & a_{n,m}B \end{pmatrix} \in \mathcal{M}_{nk,mn}(\mathbb{R})$$

3 Representation of Information

Definition 3.1 (Source)**.** Let S be an alphabet. A *source* $(S, \mathbf{p})$ with $\mathbf{p} = \mathbf{p}^{(k)} = (p_1^{(k)}, \dots, p_n^{(k)})$ emits a stream $\sigma_1\sigma_2\dots$ of symbols with probability $P(\sigma_k = s_i) = p_i^{(k)}$.

Definition 3.2 (Memoryless Source)**.** A *memoryless source* $(S, \mathbf{p})$ emits a stream $\sigma_1\sigma_2\dots$ such that

$$\forall k,\ l \quad P(\sigma_k = s_i, \sigma_l = s_j) = P(\sigma_k = s_i)P(\sigma_l = s_j)$$

Definition 3.3 (Stationary Source)**.** A source emitting a stream $\sigma_1\sigma_2\dots$ is *stationary* if, for any positive integers k, r, l_1, l_2, $\dots$, l_r, the probabilities

$$P(\sigma_{k+l_1} = s_1, \sigma_{k+l_2} = s_2,\ \dots\ , \sigma_{k+l_r} = s_r)$$

depend only on the stream $s_1s_2\dots s_r$ and not on k. For $l_1 = 1$, $l_2 = 2$, $\dots$, $l_r = r$, we use the notation:

$$p^r(s_1s_2\dots s_r) = P(\sigma_{k+1} = s_1, \sigma_{k+2} = s_2,\ \dots\ , \sigma_{k+r} = s_r)$$

Then a memoryless source is a stationary source with $p^r(s_1s_2\dots s_r) = p^1(s_1)p^1(s_2)\dots p^1(s_r)$.

Definition 3.4 (Stochastic Process)**.** A discrete time *stochastic process* on S is a collection of S-valued random variables X_i with $i \in \mathbb{N}$ or $\mathbb{Z}$.

Definition 3.5 (Markov Chain)**.** A discrete time *Markov chain* is a stochastic process such that

$$\forall N, \quad P(X_N = s_{i_N} \,|\, X_0 = s_{i_0}, \ldots, X_{N-1} = s_{i_{N-1}}) = P(X_N = s_{i_N} \,|\, X_{N-1} = s_{i_{N-1}})$$

A Markov chain can be represented by a *stochastic matrice*: $P = (P_{i,j})$ with $P_{i,j} = P(X_N = s_i \,|\, X_{N-1} = s_j)$.

We thus have:

- Remember nothing: Memoryless process
- Remeber last state: Markov chain
- Remember everything: Stochastic Process

Definition 3.6 (Average Word Length)**.** The *average word length* L of a code $c : S \to T^*$ for a source $(S, \mathbf{p})$ is defined by $L = \sum_{i=1}^{m} p_i |c(s_i)|$.

Definition 3.7 (Optimal Code)**.** A uniquely decodeable code $c : S \to T*$ is *optimal* if there is no other code with smaller average word-length.

Definition 3.8 (Entropy)**.** Given a distribution $\mathbf{p} = (p_1, p_2, \ldots, p_m)$, the *entropy* of $\mathbf{p}$ (in base b) is given by

$$H_b(p) = \sum_{i=1}^{m} p_i \log_b(\frac{1}{p_i})$$

For $p_i = 0$, we set $p_i \log_b(\frac{1}{p_i}) = 0$. Entropy can be seen as the average number of bits needed to encode an information.

Property 3.1 (Entropy)**.**

- $H(\mathbf{p} \otimes \mathbf{q}) = H(\mathbf{p}) + H(\mathbf{q} \,|\, \mathbf{p})$
- $H((p_1, \ldots, p_n, 0)) = H((p_1, \ldots, p_n))$

Theorem 3.1 (Comparison Theorem)**.** Given probability distributions $\mathbf{p}$ and $\mathbf{q}$ then

$$H_b(\mathbf{p}) = \sum_{i=1}^{m} p_i \log_b(\frac{1}{p_i}) \le \sum_{i=1}^{m} p_i \log_b(\frac{1}{q_i})$$

Theorem 3.2. $H_b(\mathbf{p})$ is maximal for uniform distribution $\mathbf{p}$ *i.e.* $H_b(\mathbf{p}) \le \log_g(m)$ and there is equality if and only if $p_i = \frac{1}{m}$.

Theorem 3.3 (Fundamental Theorem)**.** The average word-length L of any uniquely decodeable code $c : S \to T^*$ with $|T| = b$ for the source $(S, \mathbf{p})$ satisfies $L \ge H_b(\mathbf{p})$.

Theorem 3.4 (Shannon-Fano Rule)**.** There exists a prefix-free code $c : S \to T^*$ for the source $(S, \mathbf{p})$ which satisfies $L < H_b(\mathbf{p}) + 1$. To build it, select for the word length $y_i = |c(s_i)|$ the least positive integer such that $b^{y_i} \ge \frac{1}{p_i}$.

We thus have $H \le L_{opt} \le L_{SF} \le H + 1$

Property 3.2. An optimal prefix-free code $c : S \to \mathbb{B}^*$ for a source $(S, \mathbf{p})$ has the following properties:

- If $c(s') \ge c(s)$ then $p_{s'} \le p_s$

- Among the codewords of maximal length there are two of the form $w0$ and $w1$ for some $w \in \mathbb{B}^*$

Theorem 3.5 (Huffman's Rule)**.** Let $(S, \mathbf{p})$ be a source. To construct an optimal code:

1. If s' and s'' have the smallest probability, construct a new source $(S^*, \mathbf{p}^*)$ by replacing s' and s'' with a new symbol s^* with probability $p_{s^*} = p_{s'} + p_{s''}$.
2. If we have a prefix-free binary code h^* for $(S^*, \mathbf{p}^*)$ with $h^*(s^*) = w$, then define a binary code h for $(S, \mathbf{p})$ with $h(s') = w0$ and $h(s'') = w1$.

If the code h^* is optimal for $(S^*, \mathbf{p}^*)$, then h is optimal for $(S, \mathbf{p})$.

Definition 3.9 (Products and Dstribution)**.** Let $(S = S' \times S'', \mathbf{p})$ be a source defined on the cartesian product of alphabets S' and S''. The *marginal distributions* $\mathbf{p}'$ on S' and $\mathbf{p}''$ on S'' are given by:

$$p'_i = \sum_{j=1}^{n} p_{ij} \quad \text{and} \quad p''_j = \sum_{i=1}^{m} p_{ij}$$

$\mathbf{p}'$ and $\mathbf{p}''$ are independent if and only if $p_{ij} = p'_i p''_j$ or $\mathbf{p} = \mathbf{p}' \otimes \mathbf{p}''$.

Theorem 3.6 (Entropy of Product)**.** For a distribution $\mathbf{p}$ on $S' \times S''$ and its marginal distributions $\mathbf{p}'$ and $\mathbf{p}''$ we have $H(\mathbf{p}) \leq H(\mathbf{p}') + H(\mathbf{p}'')$. Equality holds if and only if $\mathbf{p}'$ and $\mathbf{p}''$ are independent.

Definition 3.10 (Entropy of a Stationary Source)**.** The entropy H of a stationary source with probability distribution $\mathbf{p^r}$ is defined as

$$H = \inf_{r \in \mathbb{N}^*} \frac{H(\mathbf{p}^1)}{r}$$

In particular for a memoryless stationary source, $H = H(\mathbf{p^1})$.

Theorem 3.7. For a stationary source on S and entropy H, given $\varepsilon > 0$ there exists $n \in \mathbb{N}^*$ and a prefix-free binary code $(S^n, \mathbf{p^n})$ such that $\frac{L_n}{n} < H + \varepsilon$.

Definition 3.11 (Dictionary)**.** A *dictionary* D based on an alphabet S is a finite sequence of distinct words in S^*: $D = (d_1, d_2, \ldots, d_N)$. Dictionaries are used to record the encoding of symbols and blocks: keys (indexes) are codewords, values are symbols or blocks.

Theorem 3.8 (LZW Compression)**.** $X = x_1 x_2 \ldots x_n$ is a message in the alphabet $S = \{s_1, \ldots, s_m\}$, $D_0 = (d_1, d_2, \ldots, d_m)$ is a dictionary with $d_i = s_i$. The *LZW coding* constructs $c(X) = c_1 c_2 c_3 \ldots$ as follows:

1. The first symbol x_1 is encoded as $c_1 = p$ where p is taken such that $x_1 = s_p = d_p$.
2. We define $D_1 = (D_0, d_{m+1})$ where $d_{m+1} = x_1 x_2$.
3. Find the longest string starting with x_2 present in D_1 and repeat the first two steps.
4. Repeat until X is completely encoded.

The *LZW* code constructed as above is uniquely decodeable.

4 Transmission of Information

Definition 4.1 (Channel)**.** A *channel* Γ with input set $I = \{s_1, \ldots, s_m\}$ and output $J = \{r_1, \ldots, r_m\}$ is a stochastic matrix where $\Gamma_{ij} = P(r_j \mid s_i)$. If there exist $i \neq j$ such that $\Gamma_{ij} \neq 0$ then the channel is *noisy*.

Definition 4.2 (Binary Symmetric Chanel)**.** A *binary symmetric chanel* (BSC) corresponds to the channel matrix of the form:

$$\Gamma = \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix}$$

with *error* $e > 0$.

Theorem 4.1. Let Γ be a channel matrix and $(I, \mathbf{p})$, $(J, \mathbf{q})$ be the sources associated to the input and output respectively. Then $\mathbf{q} = \mathbf{p}\Gamma$.

Theorem 4.2 (Conditional Entropy)**.** Consider the probability distribution $\mathbf{t}$ on $I \times J$ defined as $t_{ij} = P(s_i \cap r_j) = p_i\Gamma_{ij}$. The *conditional entropy* is defined as $H(\mathbf{p} \,|\, \mathbf{q}) = H(\mathbf{t}) - H(\mathbf{q}) = H(\Gamma; \mathbf{p})$. It can be seen as the measure of incertitude on $\mathbf{p}$ once we observe $\mathbf{q}$.

Theorem 4.3. The conditional entropy $H(\mathbf{q} \,|\, \mathbf{p})$ can also be calculated as

$$H(\mathbf{q} \,|\, \mathbf{p}) = \sum_i p_i H(\mathbf{q} \,|\, i)$$

where $H(\mathbf{q} \,|\, i) = \sum_j \Gamma_{ij} \log(\frac{1}{\Gamma_{ij}})$

Theorem 4.4 (Conditional Entropy for BSC)**.** Let Γ be a BSC with bit-error probability $\mathbf{e} = (1-e, e)$ and $\mathbf{p}$ the source distribution. Then $H(\Gamma; \mathbf{p}) = H(\mathbf{p}) + H(\mathbf{e}) - H(\mathbf{q})$.

Theorem 4.5. Let Γ be a channel and $\mathbf{p}$ an input source distribution. Then $H(\Gamma; \mathbf{p}) \leq H(\mathbf{p})$. Equality holds if and only if $\mathbf{p}$ and $\mathbf{q} = \mathbf{p}\Gamma$ are independent.

Definition 4.3 (Capacity)**.** The *capacity* γ of a channel Γ is defined as:

$$\gamma(\Gamma) = \max_{\mathbf{p}}(H(\mathbf{p}) - H(\Gamma; \mathbf{p}))$$

We can interpret $H(\mathbf{p}) - H(\Gamma; \mathbf{p})$ as the *mutual information* between $\mathbf{p}$ and $\mathbf{q} = \mathbf{p}\Gamma$ *i.e.* the information we get on $\mathbf{p}$ when we have observed $\mathbf{q}$ and conversely. The capacity maximizes that mutual information: it's the fundamental limit of bits we can transmit over a channel reliably.

Theorem 4.6 (Capacity of a BSC)**.** Let Γ be a BSC with bit-error probability $\mathbf{e}$ with $0 \leq e \leq \frac{1}{2}$. Then $\gamma(\Gamma) = 1 - H(\mathbf{e})$.

Definition 4.4 (Decision rule, Mistake)**.** Let $C \subset \mathbb{B}^n$ be a set of binary words. A *decision rule* for C is a function $\sigma : \mathbb{B}^n \to C$ which assigns to each $z \in \mathbb{B}^n$ a codeword in C. We say that a mistake occurs if a codeword in the final stream is different from the codeword in the encoded stream.

The transmission of information follows these steps:

1. Original Stream $\to$ Encoded Stream via coding $c : S \to C \subseteq \mathbb{B}^n$.
2. Encoded Stream $\to$ Received Stream. Channel $\Gamma : \mathbb{B}^m \to \mathbb{B}^n$ introduces errors.
3. Received Stream $\to$ Final Stream via decision rules $\sigma : \mathbb{B}^n \to C$.
4. Final Stream decoded to retrieve Original Stream.

Definition 4.5 (Extended Channel)**.** Given a channel Γ with input alphabet I and output alphabet J. The *extended channel* is defined on words of length n as Γ^n (for the tensor product).

Definition 4.6 (Hamming Distance)**.** Given two binary words x, $y \in \mathbb{B}^n$, the *Hamming distance* $d(x, y)$ is the number of places where x and y differ.

Theorem 4.7. Given two binary words x, $y \in \mathbb{B}^n$, the entry $(\Gamma)_{xy}$ in the channel matrix of the extended BSC with bit-error e is given by $(\Gamma)_{xy} = e^d(1-e)^{n-d}$ where d is the Hamming distance between x and y.

Definition 4.7 (Ideal Observer Rule)**.** The *ideal observer rule* is given by $\sigma(z) = c$ if the probability that z was sent given that c was received is maximal *i.e.* $P(c \,|\, z) = \max_{c'} P(c' \,|\, z)$.

Definition 4.8 (Maximal Likelihood Rule)**.** The *maximal likelihood rule* is given by $\sigma(z) = c$ if $P(z \,|\, c) = \max_{c'} P(z' \,|\, c')$

Definition 4.9 (Minimum Distance Rule)**.** The *minimum distance rule* (MD) is given by $\sigma(z) = c$ such that $d(z, c) = \min_{c'} d(z, c')$.

Theorem 4.8. For an extended BSC channel Γ^n with bit-error $e \leq \frac{1}{2}$, the maximal likelihood rule is equivalent to the minimum distance rule.

Definition 4.10 (Minimum Distance)**.** Given a set of binary words $C \subseteq \mathbb{B}^n$, the *minimum distance* of C is defined as $\delta = \min_{c \neq c'} d(c, c')$.

Theorem 4.9. Let $C \subseteq \mathbb{B}^n$ be a set of binary words with $\delta \geq 2r + 1$ used as input and applying the MD rule. If less than r bit-errors are made during transmission, then there will be no mistakes.

Theorem 4.10 (Packing Bound)**.** Let $C \subseteq \mathbb{B}^n$ be a code with $\delta \geq 2r + 1$. Then:

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

Definition 4.11 (Information Rate)**.** Given a code $C \subseteq \mathbb{B}^n$, its *information rate* is given by $\rho = \frac{\log_2 |C|}{n}$. It is the ratio between the amount of bits needed to encode all codewords and the number of bits available.

Definition 4.12 (Probability of Mistake)**.** The probability of a mistake when the encoded stream comes from a source $(S, \mathbf{p})$ is given by:

$$M(C, \mathbf{p}) = M_\sigma(C, \mathbf{p}) = \sum_{c \in C} p_c M_{c,\sigma}$$

where:

$$M_{c,\sigma} = \sum_{z \in F_\sigma(c)} P(z \,|\, c) \quad \text{and} \quad F_\sigma(c) = \{z \in \mathbb{B}^n \mid \sigma(z) \neq c\}$$

Theorem 4.11 (Capacity of Extended BSC)**.** If the capacity of a BSC Γ is $\gamma(\Gamma) = 1 - H(\mathbf{e})$ then the capacity of the extended BSC Γ^n is $\gamma(\Gamma^n) = n\gamma(\Gamma)$.

Theorem 4.12 (Rate vs. Capacity)**.** Let $C \subseteq \mathbb{B}^n$ be a code with information rate ρ and $\mathbf{p}^*$ be the uniform probability distribution on C. Consider source $(C, \mathbf{p}^*)$ through an extended BSC Γ^n with capacity $\gamma(\Gamma) = \gamma$. Then:

$$H(\Gamma^n; \mathbf{p}^*) \geq n(\rho - \gamma)$$

Theorem 4.13 (Fano's Inequality)**.** Given a code $C \subseteq \mathbb{B}^n$ and $M = M(C, \mathbf{p})$ the probability of a mistake for the source $(C, \mathbf{p})$ being transmitted through the extended BSC Γ^n using the MD rule, then we have:

$$H(\Gamma; \mathbf{p}) \leq H(\mathbf{M}) + M \log(|C| - 1)$$

Theorem 4.14 (Shannon's Theorem)**.** For $\rho < \gamma$ it is possible to construct a sequence of codes $C_n \subseteq \mathbb{B}^n$ such that:

$$|C_n| \geq 2^{\rho n} \quad \text{and} \quad \lim_{n \to \infty} M(C_n, \mathbf{p}) = 0$$

In other words, any transmission link can be made reliable provided the capacity is large enough.

5 Representation of codes

Using closest codeword decoder (*i.e.* using minimum distance rule), any code with minimum distance d can *correct* up to $\lfloor \frac{d-1}{2} \rfloor$ errors. Any code with minimum distance d can also *detect* up to $d - 1$ errors.

Definition 5.1 (Linear Code)**.** A *linear code block* C of length n is a sub-vector space of $\mathbb{F}_2^n$. We use the notation $[n, k, d]_q$ to describe a linear code C:

- n is the code length
- k is the dimension of C
- d is the minimum distance
- q is the size of the source alphabet

Definition 5.2 (Minimal Distance of a Linear Code)**.** The minimal distance of a linear code is

$$d = \min_{\mathbf{c} \in C, \mathbf{c} \neq \mathbf{0}} w_H(\mathbf{c})$$

where w_H is the Hamming weight function.

Definition 5.3 (Generator Matrix)**.** Let $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k)$ be a base of a linear code C. Then $\mathbf{c} \in C$ (row vector) can be written as $\mathbf{c} = \sum_{i=1}^{k} d_i \mathbf{e}_i$ or $\mathbf{c} = \mathbf{d}G$ where $G \in \mathcal{M}_{k,n}(\mathbb{F}_2)$ is the *generator matrix* for the code C:

$$G = \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k \end{pmatrix}$$

Definition 5.4 (Systematic Encoder)**.** Given a code $C = [n, k, d]$, a *systematic generator matrix* (or systematic encoder) of C is a matrix obtained by applying the Gaussian elimination algorithm to any generator matrix of C. A systematic generator matrix G_s can be written as:

$$G_s = \begin{bmatrix} I_k & P \end{bmatrix}$$

where P is the *checksum* matrix. Every linear code has a systematic encoder.

Definition 5.5 (MDS Code)**.** A *maximum distance separable* code C is a code where any combination of k columns of a generator matrix of C are linearly independent. The minimum distance of a MDS code is $d = n - k + 1$.

Definition 5.6 (Dual Code)**.** The *dual code* $C^{\perp}$ of a linear code $C = [n, k, d]$ is defined as $C^{\perp} = \{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{c} = \sum_{i=1}^{n} x_i c_i = 0 \ \forall \mathbf{c} \in C\}$. The dual code of a MDS code is also a MDS code.

Definition 5.7 (Parity Check Matrix)**.** Let $C = [n, k, d]$ be a linear code. A *parity check matrix* of C is any generator matrix of $C^{\perp}$ (of $n - k$ rows and n columns). A matrix is a parity check matrix if and only if $G \cdot H^T = \mathbf{0}$ and we have $\mathbf{c} \in C \iff \mathbf{c} \cdot H^T = \mathbf{0}$.

Definition 5.8 (Systematic Parity Check Matrix)**.** A *systematic parity check matrix* can be written as $\begin{bmatrix} -P^T & I_{n-k} \end{bmatrix}$.

Theorem 5.1 (Minimal Distance)**.** For a linear code $C = [n, k, d]$:

- $d \leq n - k + 1$
- d is equal to the minimum number of linearly dependent columns of H.

Definition 5.9 (Syndrome)**.** A *syndrome* is a vector $\mathbf{s} = \mathbf{y} \cdot H^T$ for any vector $\mathbf{y}$ of length n.

Definition 5.10 (Syndrome Decoding)**.** Given a received message $\mathbf{y} = \mathbf{c} + \mathbf{e}$ where $\mathbf{e}$ is an error vector, we want to find the original codeword $\mathbf{c}$. The *syndrome decoding* algorithm follows these steps:

1. Compute the syndrome $\mathbf{s} = \mathbf{y} \cdot H^T$.
2. If $\mathbf{s} = \mathbf{0}$, then $\mathbf{c} = \mathbf{y}$ and the algorithm stops.
3. Check if $\mathbf{s}^T$ is a column of H. If $\mathbf{s}^T = \mathbf{h_i}$, $\mathbf{c} = (y_1, \cdots, 1 - y_i, \cdots, y_n)$ and the algorithm stops.
4. Check if $\mathbf{s}^T$ is the sum of two columns of H. If $\mathbf{s}^T = \mathbf{h}_i + \mathbf{h}_j$, $\mathbf{c} = (y_1, \cdots, 1 - y_i, \cdots, 1 - y_j, \cdots, y_n)$ and the algorithm stops. If there is more than one choice for i and j, choose one pair at random.
5. Repeat until $\mathbf{c}$ is found.

Definition 5.11 (Hamming Code)**.** A *Hamming Code* is a code $C = [2^m - 1, 2^m - m - 1, 3]$ where $m \geq 3$. The parity check matrix H of a Hamming code of parameter m is such that its columns enumerate all non-zero m-bit strings.

6 Algebraic Codes

Definition 6.1 (Reed-Solomon)**.** A *Reed-Solomon* code over $GF(q)$ is defined by a set of distinct evaluation points $\alpha_1, \ldots, \alpha_n \in GF(q)$ and by a dimension k. The codewords of a Reed-Solomon code is the set of evaluation vectors of all polynomials of degree $< k$ *i.e.* :

$$C_{RS} = \left\{ (f(\alpha_1), \ldots, f(\alpha_n)) \;\middle|\; f : x \mapsto \sum_{i=0}^{k-1} c_i x^i \quad \forall \ c_0, \ldots, c_{k-1} \in GF(q) \right\}$$

Therefore we need $n < q$.

Property 6.1 (Reed-Solomon Code)**.** For a Reed-Solomon code C_{RS} of dimension k:

- C_{RS} is linear, $C_{R,S} = [n, k, d]$.
- $|C_{RS}| = q^k$.
- A natural generator matrix for C_{RS} is the *Vandermonde matrix*: $G_{RS} = \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_1 \\ \vdots & \cdots & \vdots \\ \alpha_n^{k-1} & \cdots & \alpha_n^{k-1} \end{pmatrix}$.

- The rank of the Vandermonde matrix is k if $\alpha_1, \dots, \alpha_n$ are distinct.
- The minimal distance is $d_{RS} = n - k + 1$, therefore C_{RS} is a MDS code.
- Because C_{RS} is MDS, if C_{RS} has a rate R it can correct up to $\lfloor \frac{1-R}{2} n \rfloor$ errors.

Definition 6.2 (Reed-Muller)**.** Codewords of a *Reed-Muller* code $RM(r, m)$ are evaluation vectors of polynomials of m variables of degree at most r (the order) at all points of $(GF(q))^m$.

Property 6.2 (Reed-Muller Code)**.** For a Reed-Muller code $RM(r, m)$:
- Because we consider all points of $(GF(q))^m$, $n = q^m$.
- Dimension is the number of monomials of a polynomial $f(x_1, \dots, x_m)$ of degree at most r: $k = \binom{m+r}{m} = \binom{m+r}{r}$ (only valid if $q > r$).
- Any polynomial (possibly multivariate) of degree at most r is zero on at most $\frac{r}{q}$ fraction of all possible points. Therefore $d \geq q^m - rq^{m-1} = n(1 - \frac{r}{q})$.

Definition 6.3 (Hadamard Code)**.** A *Hadamard code* is a special case of a Reed-Muller code where $q = 2$, $r = 1$. A Hadamard code verifies:
- $n = 2^m$.
- $k = m + 1$.
- $d = 2^{m-1}$.

Theorem 6.1 (Bounds)**.** The rate R_q of a code $[n, k, d]_q$ where δ is a given "relative distance" verifies:
- *Singleton bound*: $k \leq n - d + 1 \implies R_q(\delta) \leq 1 - \delta$.
- *Hamming bound*: $R_q(\delta) \leq 1 - h_q\left(\frac{\delta}{2}\right)$.
- *Gilbert-Varshamov bound*: $R_q(\delta) \geq 1 - h_q(\delta)$.