MPRI **Year 2020–21**

Module 2.13.2 : Error correcting codes and applications to cryptography

Mid-term exam, November 26

You have 2h30 Including time for scanning/taking pictures.
You can answer either in French or in English.

Exercise 1. True or false. You should justify your answers.

1. A linear code has a unique generator matrix.

2. A linear code has a unique parity–check matrix.

3. Given a linear code with a generator matrix $\mathbf{G}$, multiplying $\mathbf{G}$ on the right by a non-singular matrix does not change the code.

4. Given a linear code with a generator matrix $\mathbf{G}$, multiplying $\mathbf{G}$ on the left by a non-singular matrix does not change the code.

5. There is no linear code with parameters $[n, k, n-k+2]$.

6. There is no linear code whose parameters exceed the Gilbert-Varshamov bound.

7. Asymptotic Plotkin bound is always sharper than asymptotic Singleton bound.

8. The weight distribution of a linear code of length n and dimension $n-3$ can be computed in polynomial time in n.

9. For any linear code, decoding up to half the minimum distance can be done in polynomial time.

10. For any linear code of length n and dimension k, computing a codeword of weight $\leqslant n-k+1$ can be done in polynomial time.

Exercise 2. A *Boolean function* in m variables is an m-variable polynomial which is a sum of monomials $X_1^{i_1} \cdots X_m^{i_m}$ where $i_1, \dots, i_m \in \{0, 1\}$. The *degree* of a monomial $X_1^{i_1} \cdots X_m^{i_m}$ is the sum $i_1 + \cdots + i_m$. The *degree* of a Boolean function is the maximum degree of its monomials. For instance, the Boolean functions :

$$F(X_1, X_2, X_3) = X_1 + X_1X_2 + X_2X_3 \quad \text{and} \quad G(X_1, X_2, X_3, X_4) = X_1X_3X_4 + X_2X_3 + 1$$

have respective degrees 2 and 3. By convention, the degree of the function 0 is set to $-\infty$. The space of Boolean functions of degree $\leqslant r$ in m variables is denoted by $\mathcal{B}_r(m)$ and the whole space of Boolean functions in m variables is denoted by $\mathcal{B}(m)$.

Question 1. Give the full list of the elements of the sets $\mathcal{B}_0(2)$ and $\mathcal{B}_1(2)$.

Question 2. Prove that

(a) for any $0 \leqslant r < m$, we have $\dim_{\mathbb{F}_2} \mathcal{B}_r(m) = \sum_{j=0}^{r} \binom{m}{j}$.

(b) $\dim_{\mathbb{F}_2} \mathcal{B}(m) = 2^m$.

Fix integers $m > 0$ and $r > 0$, then the Reed–Muller code $\mathcal{R}(r, m)$ is defined as

$$\mathcal{R}(r, m) := \left\{(P(x_1, \ldots, x_m))_{(x_1,\ldots,x_m)\in\mathbb{F}_2^m} \mid P \in \mathcal{B}_r(m)\right\}.$$

where the elements of $\mathbb{F}_2^m$ are sorted in the lexicographic order. For instance for $m = 3$, the elements of $\mathbb{F}_2^3$ are sorted as :

$$(000) \prec (100) \prec (010) \prec (110) \prec (001) \prec (101) \prec (011) \prec (111).$$

Question 3. Prove that for any $m \geqslant 0$, the code $\mathcal{R}(0, m)$ is the repetition code of length 2^m.

Question 4. Give a generator matrix of the code $\mathcal{R}(1, 3)$.

We focus on the encoding of the code $\mathcal{R}(1, m)$, which is given by the map

$$\mathrm{Enc}_m : \left\{ \begin{array}{ccc} \mathcal{B}_1(m) & \longrightarrow & \mathbb{F}_2^{2^m} \\ F = a_0 + a_1X_1 + \cdots + a_mX_m & \longmapsto & (F(x_1, \ldots, x_m))_{(x_1,\ldots,x_m)\in\mathbb{F}_2^m}. \end{array} \right.$$

Question 5. Prove that a naive encoding of the code $\mathcal{R}(1, m)$ has a complexity of $O(n \log n)$, where $n = 2^m$ denotes the code length.

Question 6. This encoding may be improved using the following principle :

(a) Prove that, given $F_{m-1} = a_0 + a_1X_1 + \cdots + a_{m-1}X_{m-1}$ and $F = F_{m-1} + a_mX_m$, then

$$\mathrm{Enc}_m(F) = (\underbrace{\mathrm{Enc}_{m-1}(F_{m-1})}_{\text{length } 2^{m-1}} \mid \underbrace{\mathrm{Enc}_{m-1}(F_{m-1})}_{\text{length } 2^{m-1}} + \underbrace{(a_m, \ldots, a_m)}_{\text{length } 2^{m-1}}), \tag{1}$$

where the "|" stands for the concatenation of codewords.

(b) Deduce from the previous question a faster encoding algorithm.

(c) Prove that this faster encoding has complexity $O(n)$, where $n = 2^m$ denotes the code length.

We now focus on the decoding of these codes. Indexing words of $\mathbb{F}_2^{2^m}$ with elements of $\mathbb{F}_2^m$ sorted in the lexicographic order, for any $\mathbf{c} \in \mathbb{F}_2^{2^m}$, one defines

$$\Delta_\alpha(\mathbf{c}) := \left(\mathbf{c}_{(x_1+\alpha_1,\ldots,x_m+\alpha_m)} + \mathbf{c}_{(x_1,\ldots,x_m)}\right)_{(x_1,\ldots,x_m)\in\mathbb{F}_2^m}$$

Question 8. Let $\mathbf{c} = (1\ 1\ 0\ 1\ 0\ 0\ 0\ 1) \in \mathbb{F}_2^8$ and $\alpha = (1\ 0\ 1) \in \mathbb{F}_2^3$. Compute $\Delta_\alpha(\mathbf{c})$.

Question 9. If $\mathbf{c} \in \mathbb{F}_2^{2^m}$ has weight t,

(a) prove that for any $\alpha \in \mathbb{F}_2^m$ the vector $\Delta_\alpha(\mathbf{c})$ has weight less than or equal to $2t$;

(b) Give an example where this $2t$ is reached (*Hint. choose a small value of m to design your example*)

Question 10. Let $\mathbf{b}_1, \ldots, \mathbf{b}_m$ be the canonical basis of $\mathbb{F}_2^m$. Let $F = a_0 + a_1X_1 + \cdots + a_mX_m \in \mathcal{B}_1(m)$ and $\mathbf{c} = \mathrm{Enc}_m(F) \in \mathcal{R}(1, m)$. Prove that for any $i \in \{1, \ldots, m\}$, we have

$$\Delta_{\mathbf{b}_i}(\mathbf{c}) = (a_i, \ldots, a_i).$$

Question 11. Suppose you received a corrupted codeword $\mathbf{y} = \mathbf{c} + \mathbf{e}$ with $\mathbf{c} = \mathrm{Enc}_m(F) \in \mathcal{R}(1, m)$ and $\mathbf{e}$ has Hamming weight $w_{\mathrm{H}}(\mathbf{e}) \leqslant 2^{m-2} - 1$.

(a) Explain how to recover $a_1, \ldots, a_m$ using derivations and detail why the bound on the weight of $\mathbf{e}$ asserts that these coefficients are uniquely recovered.

(b) Once $a_1, \ldots, a_m$ are known, explain how to find a_0.

(c) Give the complexity of this decoding algorithm.

(d) Looking at Theorem 10.8 of these notes, what can you say about the decoding radius (*i.e.* the amount of errors it corrects) of this algorithm ?