Problems Session 1

Basic Exercises on Codes

In what follows, $|\cdot|$ will denote the Hamming weight, namely

$$\forall \mathbf{x} \in \mathbb{F}_q^k, \quad |\mathbf{x}| \overset{\text{def}}{=} \sharp \left\{ i \in [\![1,n]\!],\ x_i \neq 0 \right\}.$$

Exercise 1. Give the dimension of the following linear codes:

1. $\{(f(x_1),\dots,f(x_n)) : f \in \mathbb{F}_q[X] \text{ and } \deg(f) < k\}$ where the x_i's are distinct elements of $\mathbb{F}_q$,
2. $\{(\mathbf{u},\mathbf{u}+\mathbf{v}) : \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}$ where U (*resp.* V) is an $[n,k_U]_q$-code (*resp.* $[n,k_V]_q$-code).

Exercise 2. *Let* $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ *be a generator matrix of some code* $\mathcal{C}$*. Let* $\mathbf{H} \in \mathbb{F}_q^{(n-k)\times n}$ *of rank* $n-k$ *such that* $\mathbf{G}\mathbf{H}^\top = \mathbf{0}$*. Show that* $\mathbf{H}$ *is a parity-check matrix of* $\mathcal{C}$*.*

Exercise 3. Give the minimum distance of the following linear codes:

1. $\{(f(x_1),\dots,f(x_n)) : f \in \mathbb{F}_q[X] \text{ and } \deg(f) < k\}$ where the x_i's are distinct elements of $\mathbb{F}_q$.
2. $\{(\mathbf{u},\mathbf{u}+\mathbf{v}) : \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}$ where U (*resp.* V) is a code of length n over $\mathbb{F}_q$ and minimum distance d_U (*resp.* d_V).
3. The Hamming code of length $2^r - 1$, namely the code which admits a parity check matrix $\mathbf{H} \in \mathbb{F}_q^{r\times(2^r-1)} \overset{\text{def}}{=} \left(\mathbf{x}^\top\right)_{\mathbf{x}\in\mathbb{F}_2^r\setminus\{\mathbf{0}\}}$.

   Hint: *A code has minimum distance d if and only if for some parity-check matrix* $\mathbf{H}$ *every* $(d-1)$*-tuple of columns are linearly independent and there is at least one linearly linked* d*–tuple of columns.*

Decoding Generalized Reed-Solomon Codes

The purpose of this Section is to describe a decoding algorithm for Generalized Reed-Solomon (GRS) codes. In the last Practice Session, we will see that this family of codes is not suitable for instanciating the McEliece Cryptosystem, but it is nevertheless the starting point to NIST candidate Classic McEliece which uses a family of codes known as *Goppa Codes*, derived from Reed-Solomon codes.

Reed-Solomon Codes

Let $k \leq n$ be integers. Let $\mathbf{x} \overset{\text{def}}{=} (x_i)_{1\leq i\leq n}$ be a tuple of *pairwise distinct* elements of a finite field $\mathbb{F}_q$, and let $\mathbf{z} \overset{\text{def}}{=} (z_i)_{1\leq i\leq n} \in \left(\mathbb{F}_q^\star\right)^n$ (the z_i may be equal). Recall that the Generalized Reed-Solomon code $\mathrm{RS}_k(\mathbf{x},\mathbf{z})$ with *evaluation points* $\mathbf{x}$ and *multipliers* $\mathbf{z}$ is the code over $\mathbb{F}_q$ defined by

$$\mathrm{RS}_k(\mathbf{x},\mathbf{z}) \overset{\text{def}}{=} \{z_1 f(x_1), \dots, z_n f(x_n) \colon f \in \mathbb{F}_q[X]_{<k}\}\,.$$

Suppose we are given a noisy codeword

$$\mathbf{y} \overset{\text{def}}{=} \mathbf{c} + \mathbf{e} \tag{1}$$

where $\mathbf{c} = (z_i f(x_i))_{1\leq i\leq n} \in \mathrm{RS}_k(\mathbf{x},\mathbf{z})$, and $\mathbf{e} \in \mathbb{F}_q^n$ has Hamming weight t. Our goal is to recover $\mathbf{c}$, or equivalently the polynomial $f \in \mathbb{F}_q[X]_{<k}$.

1. Explain why we can suppose without loss of generality that all the z_i are equal to 1.

2. From now on, we assume that $z_i = 1$ for all $i \in \{1,\dots,n\}$. Let us introduce the following polynomial

   $$\Lambda(X) = \prod_{i\colon e_i \neq 0} (X - x_i),$$

   known as *error locator polynomial.*

   Show that

   $$\forall i \in \{1,\dots,n\}, \quad y_i\Lambda(x_i) = f(x_i)\Lambda(x_i) \tag{2}$$

3. Remind that the y_i's and x_i's are known.

   (a) Deduce from eq. (2) a system $\mathcal{S}_1$ of equations satisfied by the coefficients of f.

   (b) How many equations and unknowns does this system have?

   (c) Can you recover f?

4. Let N be the polynomial defined by $\Lambda \cdot f$.

   (a) Write a *linear* system $\mathcal{S}_2$ satisfied by the coefficients of N.

   (b) How many unknowns and equations does $\mathcal{S}_2$ have?

5. Show that *any* solution of $\mathcal{S}_1$ is indeed a solution of $\mathcal{S}_2$. In other words, $\mathcal{S}_2$ is *a priori*, more general than $\mathcal{S}_1$. It turns out that they are in fact equivalent.

6. Let (Λ_1, N_1) and (Λ_2, N_2) be two pairs of *non zero* solutions of $\mathcal{S}_2$ (in particular, $\Lambda_1, \Lambda_2 \neq 0$). Show that

   $$\frac{N_1}{\Lambda_1} = \frac{N_2}{\Lambda_2} = f.^{1}$$

[1] In other words, the solutions of $\mathcal{S}_2$ form an $\mathbb{F}_q[X]-$module of rank 1.

7. Conclude:

   (a) Deduce an algorithm which, given a noisy codeword $\mathbf{y} = \mathbf{c} + \mathbf{e}$, the *evaluation vector* $\mathbf{x}$, recovers the codeword $\mathbf{c}$.

   (b) What is its time complexity?

About the Average Decoding Problem

Our goal in this section is to show that the average decoding problem can be stated, without incidence on its average hardness, via the parity-check or generator point of view. To this aim we will rely on the so-called statistical distance.

Recall that the statistical distance, also called the total variational distance, is a distance between probability distributions. In in the case where X and Y are two random variables taking their values in a same finite space $\mathcal{E}$, it is defined as

$$\Delta(X,Y) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{a \in \mathcal{E}} |\mathbb{P}(X=a) - \mathbb{P}(Y=a)|$$

The statistical distance enjoys many interesting properties. Among others, it cannot increase by applying any probabilistic algorithm as you have to show in the following question.

1. Let $X, Y : \Omega \to \mathcal{D}$ be two random variables. Let $\mathcal{A}$ be an algorithm taking as input $(x,r) \in \mathcal{D} \times \{0,1\}^{\ell}$ (the input r denotes the internal randomness of $\mathcal{A}$). Assume that the distribution R of the internal randomness of $\mathcal{R}$ is independent from X and Y. Show that,

$$\Delta\Big(\mathcal{A}\left(X,R\right), \mathcal{A}\left(Y,R\right)\Big) \leq \Delta(X,Y)$$

2. Let $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ (resp. $\mathbf{H} \in \mathbb{F}_q^{(n-k)\times n}$) be a uniformly random matrix and $\mathbf{G}_k \in \mathbb{F}_q^{k \times n}$ (resp. $\mathbf{H}_{n-k} \in \mathbb{F}_q^{(n-k)\times n}$) be a uniformly random matrix of rank k (resp. $n-k$). Show that

$$\Delta\left(\mathbf{G}, \mathbf{G}_k\right) = O\left(q^{-(n-k)}\right) \quad \left(\textit{resp. } \Delta\left(\mathbf{H}, \mathbf{H}_{n-k}\right) = O\left(q^{-k}\right)\right).$$

   Hint: You can admit that the density of rank k matrices among all the $\ell \times n$ matrices with entries in $\mathbb{F}_q$ is equal to $1 - O(q^{n-\ell})$

Let us recall the following variant of DP which was introduced during the lecture: $\mathsf{DP}'(n, q, R, \tau)$. Let $k \overset{\text{def}}{=} \lfloor Rn \rfloor$ and $t \overset{\text{def}}{=} \lfloor \tau n \rfloor$.

Input: $(\mathbf{G}, \mathbf{y} \overset{\text{def}}{=} \mathbf{sG} + \mathbf{x})$ where $\mathbf{G}, \mathbf{s}$ and $\mathbf{x}$ are uniformly distributed over $\mathbb{F}_q^{k \times n}$, $\mathbb{F}_q^k$ and words of Hamming weight t in $\mathbb{F}_q^n$.

Output: an error $\mathbf{e} \in \mathbb{F}_q^n$ of Hamming weight t such that $\mathbf{y} - \mathbf{e} = \mathbf{mG}$ for some $\mathbf{m} \in \mathbb{F}_q^k$.

Show that for any algorithm $\mathcal{A}$ solving this problem with probability ε and time T, there exists an algorithm $\mathcal{B}$ which solves $\mathsf{DP}(n, q, R, \tau)$ in time $O\left(n^3 + T\right)$ with probability $\geq \varepsilon - O\left(q^{-\min(k,n-k)}\right)$. Show that we can exchange DP' by DP in the previous question.

3. You are now ready to show the following statements.

   (a) For any algorithm $\mathcal{A}$ solving DP', with probability ε and time T, there exists an algorithm $\mathcal{B}$ which solves $\mathsf{DP}(n, q, R, \tau)$ in time $O\left(n^3 + T\right)$ with probability $\geq \varepsilon - O\left(q^{-\min(k,n-k)}\right)$.

   (b) Show that we can exchange DP' by DP in the previous question.