Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs – Feuille de TD 1

Éléments de solutions

19 novembre 2021

Exercice 1. Codes de Hamming sur $\mathbb{F}_q$.

Richard Hamming a introduit en 1950 l'un des premiers codes permettant de corriger de l'information. Ce code binaire a pour matrice génératrice

$$\boldsymbol{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} .$$

On peut noter que le code de Hamming a longueur 7, dimension 4 et distance minimale 3. Il permet de corriger 1 erreur efficacement. Dans cet exercice, nous allons étudier une généralisation de ce code.

Soient $P_1, \dots, P_n$ les points de l'espace projectif $\mathbb{P}^{r-1}(\mathbb{F}_q)$ où $r \geq 2$. On se fixe une représentation dans $\mathbb{F}_q^r$ des coordonnées de ces points : si $P_i = [x_0 : \dots : x_{r-1}]$, alors on choisit l'unique représentant de P_i tel que $x_1 = \cdots = x_{j-1} = 0$ et $x_j = 1$ pour un certain $j \in \{0, \dots, r-1\}$. Cette représentation est nommée représentation *standard* du point P_i.

On peut maintenant décrire le code de Hamming de la manière suivante. On définit une matrice $\boldsymbol{H} \in \mathbb{F}_q^{r \times n}$ telle que la i-ème colonne de $\boldsymbol{H}$ est constituée des coordonnées de P_i dans le système de coordonnées choisi :

$$\boldsymbol{H} = \begin{bmatrix} P_1 & P_2 & \dots & \dots & P_n \end{bmatrix} .$$

Le code de Hamming $\mathcal{H}_q(r) \subseteq \mathbb{F}_q^n$ est alors le code qui admet $\boldsymbol{H}$ pour matrice de parité. Autrement dit,

$$\mathcal{H}_q(r) = \{\boldsymbol{c} \in \mathbb{F}_q^n \mid \boldsymbol{H}\boldsymbol{c}^\top = \boldsymbol{0}\} .$$

Question 1.– Déterminer les matrices de parité $\boldsymbol{H}$ obtenues pour $(q, r) = (2, 4)$ et $(q, r) = (3, 3)$. Évidemment, les matrices sont construites à permutation des colonnes près.

Question 2.– Quelle est, en fonction de q et r, la longueur n du code de Hamming $\mathcal{H}_q(r)$? Quelle est sa dimension ?

Question 3.– En observant attentivement $\boldsymbol{H}$, donner un mot de poids 3 de $\mathcal{H}_q(r)$ pour tout q, r.

Question 4.– Existe-t-il des mots de $\mathcal{H}_q(r)$ de poids 1 ou 2 ? Conclure sur la distance minimale du code $\mathcal{H}_q(r)$. Comparer cette distance minimale aux bornes vues en cours (Singleton, Hamming, Plotkin par exemple).

Soit $\mathcal{A}_q(r) = \mathcal{H}_q(r)^\perp \subseteq \mathbb{F}_q^n$ le code orthogonal (ou code dual) au code de Hamming. Ce code a donc pour matrice génératrice $\boldsymbol{H}$.

Question 5.– Démontrer que $\mathcal{A}_q(r)$ peut être vu comme l'ensemble des vecteurs d'évaluation de toutes les formes linéaires $\phi : \mathbb{F}_q^r \to \mathbb{F}_q$, évaluées sur les représentants standards de tous les points $P_1, \dots, P_n$ de $\mathbb{P}^{r-1}(\mathbb{F}_q)$.

Question 6.– Déduire de la question précédente le poids de tout mot non-nul de $\mathcal{A}_q(r)$, puis déterminer les paramètres (longueur, dimension et distance minimale) de $\mathcal{A}_q(r)$.

Solutions de l'Exercice 1.

Solution Q1. Pour $(q,r) = (2,4)$, on a

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Pour $(q,r) = (3,3)$, on a

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{pmatrix}.$$

Solution Q2. On a $n = |\mathbb{P}^r(\mathbb{F}_q)| = \frac{q^r-1}{q-1}$ et $k = n - r$.

Solution Q3. La matrice $\boldsymbol{H}$ est constituée des colonnes

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Si l'on considère que ces colonnes sont les premières de $\boldsymbol{H}$, alors le mot $\boldsymbol{c} = (1, 1, -1, 0, \dots, 0) \in \mathcal{H}_q(r)$.

Solution Q4. On remarque qu'un mot de poids 1 dans $\mathcal{H}_q(r)$ signifie que l'une des colonnes de $\boldsymbol{H}$ est nulle. C'est impossible car les colonnes sont des représentants de points de l'espace projectifs. De même, avoir un mot de poids 2 dans $\mathcal{H}_q(r)$ est équivalent à avoir deux colonnes linéairement dépendantes dans $\boldsymbol{H}$. C'est de nouveau impossible car les représentations de 2 points projectifs disctincts sont linéairement indépendantes.

On a donc $d = d_{\min}(\mathcal{H}_q(r)) = 3$.

- La borne de Singleton donne $d \le n - k + 1 = r + 1$. On a donc égalité pour $r = 2$, et un « défaut » de $r - 2$ en général.
- La borne de Plotkin est inapplicable car d est trop petit.
- La borne de Hamming donne $q^k \le \frac{q^n}{1+n(q-1)}$. Or, $n(q-1) = q^r - 1$ donc les codes de Hamming atteignent cette borne. On dit que ce sont des codes **parfaits**.

Solution Q5. Un mot du code $\mathcal{A}_q(r)$ est une combinaison linéaire des lignes de $\boldsymbol{H}$. Cette combinaison linéaire s'écrit $\boldsymbol{c} = (c_1, \dots, c_n)$ où $c_j = \sum_i \lambda_i x_i(P_j)$ et $x_i(P_j)$ est la i-ème coordonnée de P_j. Donc $\boldsymbol{c}$ est l'évaluation de la forme linéaire $\sum_i \lambda_i x_i$ sur les points P_j.

Solution Q6. Une forme linéaire non-nulle s'annule sur un hyperplan. Donc le poids de $\boldsymbol{c} \in \mathcal{A}_q(r) \setminus \{\boldsymbol{0}\}$ est $n - |\mathbb{P}^{r-2}(\mathbb{F}_q)| = \frac{q^r-1}{q-1} - \frac{q^{r-1}-1}{q-1} = q^{r-2}$.

Le code $\mathcal{A}_q(r)$ a donc pour paramètres $[\frac{q^r-1}{q-1}, r, q^{r-2}]$.

Exercice 2. Hyperovale.

On considère la conique plane $\mathcal{X}$ d'équation $Z^2 = XY$ dans le plan projectif $\mathbb{P}^2(\mathbb{F}_q)$, et l'on suppose que la caractéristique de $\mathbb{F}_q$ est égale à 2.

Question 1.– Décrire les points rationnels de la courbe $\mathcal{X}$ sur $\mathbb{F}_q$. Combien y en a t-il ?

Question 2.– Démontrer que les tangentes à $\mathcal{X}$ se coupent en même point $\mathcal{O} \in \mathbb{P}^2(\mathbb{F}_q)$ que l'on déterminera. Ce point $\mathcal{O}$ appartient-il à $\mathcal{X}(\mathbb{F}_q)$?

On note $P_1, \dots, P_{n-1}$ les points de $\mathcal{X}(\mathbb{F}_q)$ et $P_n = \mathcal{O}$. Pour chaque point $P_i = [x_i : y_i : z_i]$, on choisit comme *représentant standard* dans $\mathbb{F}_q^3$ l'unique triplet de coordonnées ayant un 1 comme premier élément non nul. Par exemple, si $P = [0 : \omega : \omega^2] \in \mathbb{P}^2(\mathbb{F}_q)$ avec $\mathbb{F}_q^\times = \langle \omega \rangle$, alors le représentant canonique de P dans $\mathbb{F}_q^3$ est $(0, 1, \omega)$.

On construit maintenant un code $\mathcal{C} \subseteq \mathbb{F}_q^n$ par sa matrice génératrice $\boldsymbol{G}$, dont les colonnes sont constituées des coordonnées de la représentation canonique des points P_i. Par exemple, le point $[0 : 1 : 0]$ est l'unique point à l'infini de $\mathcal{X}$, donc la matrice $\boldsymbol{G}$ contiendra la colonne suivante :

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Question 3.– Donner une matrice génératrice du code $\mathcal{C}$ pour $q = 8$.

Question 4.– Dans le cas général, quelle est la longueur n et la dimension k du code $\mathcal{C}$?

Question 5.– En raisonnant comme dans l'Exercice 1, démontrer que le code $\mathcal{C}$ est MDS.

Remarque. Cette construction de code est particulièrement exceptionnelle. En effet, depuis des dizaines années il est conjecturé que, sauf quelques exceptions déjà connues, tout code MDS a une longueur $n \leq q+1$. Les exceptions sont pour les paramètres de codes suivants : $[n, 2, n-1]_q$ pour tout n, $[2^\ell + 2, 3, 2^\ell]_{2^\ell}$ ainsi que les duaux de ces codes.

Solutions de l'Exercice 2.

Solution Q1. Soit $[x : y : z] \in \mathbb{P}^2(\mathbb{F}_q)$ une solution de $xy = z^2$.

- Pour $x = 0$, on obtient un unique point projectif, c'est le point $[0 : 1 : 0]$.
- Si $x = 1$, alors l'ensemble des solutions est :

$$\{[1 : z^2 : z], z \in \mathbb{F}_q\}.$$

La conique $\mathcal{X}$ a donc $q+1$ points sur $\mathbb{F}_q$ (cela se confirme par le fait que c'est une courbe de genre 0).

Solution Q2. Le polynôme qui définit la courbe $\mathcal{X}$ est $F(X, Y, Z) = XY - Z^2$. En caractéristique 2, les tangentes à $\mathcal{X}$ en un point $P = [x : y : z]$ ont donc pour équation

$$\begin{aligned} \frac{\partial F}{\partial X}(x,y,z)(X-x) + \frac{\partial F}{\partial Y}(x,y,z)(Y-y) + \frac{\partial F}{\partial Z}(x,y,z)(Z-Z) \\ &= y(X-x) + x(Y-y) + (-2z) \times (Z-z) \\ &= y(X-x) + x(Y-y) \\ &= xY + yX = 0. \end{aligned}$$

Ces tangentes se coupent donc toutes en $\mathcal{O} = [0 : 0 : 1]$. Le point $\mathcal{O}$ n'appartient pas à $\mathcal{X}$.

Solution Q3. Si l'on note ω un générateur de $\mathbb{F}_8^\times$, alors la matrice génératrice est :

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & \omega^2 & \omega^4 & \omega^6 & \omega & \omega^3 & \omega^5 & 1 & 0 \\ 0 & 1 & \omega & \omega^2 & \omega^3 & \omega^4 & \omega^5 & \omega^6 & 0 & 1 \end{pmatrix}$$

Solution Q4. Le code $\mathcal{C}$ a pour longueur $n = q + 2$ et pour dimension 3.

Solution Q5. Un mot non-nul du code $\mathcal{C}$ correspond à l'évaluation d'une forme linéaire non-nulle sur $\mathcal{P} = \{\mathcal{O}, P_1, \dots, P_{n-1}, P_n = \mathcal{O}\} \subset \mathbb{P}^2(\mathbb{F}_q)$. Pour déterminer le poids des mots du code, on cherche donc le nombre de points de $\mathcal{P}$ surlesquels les mots du code s'annulents.

Généralement, l'ensemble des zéros d'une forme linéaire est un hyperplan projectif (ici, donc, une droite projective de $\mathbb{P}^2(\mathbb{F}_q)$). Montrons que toute droite projective D intersecte $\mathcal{P}$ en moins de 2 points :

— si D passe par $\mathcal{O}$ ($q+1$ droites possibles), alors D est tangente à la conique, donc l'intersecte en 1 point,

— sinon, D coupe la conique en au plus 2 points.

Le nombre de zéros de la forme linéaire est donc au plus 2, par conséquent le poids d'un mot non-nul est au moins $n - 2$. On a bien un code MDS $[n, 3, n-2]$ avec $n = q + 2$.

Exercice 3. Algorithme de décodage de codes de Reed-Solomon.

Dans cet exercice, on étudie l'algorithme de Gao [1] qui permet de décoder les codes de Reed-Solomon jusqu'au rayon de décodage unique.

On considère donc $\mathcal{C} = \mathrm{RS}(\boldsymbol{x}, k) \subseteq \mathbb{F}_q^n$ le code de Reed-Solomon évalué sur les points $\boldsymbol{x} = (x_1, \dots, x_n)$ où les $x_i \in \mathbb{F}_q$ sont deux à deux distincts. **On suppose que n et k sont pairs pour éviter la manipulation de parties entières.** La méthode de Gao pour décoder dans $\mathcal{C}$ est décrite dans l'Algorithme 2. Celui-ci prend en entrée un mot transmis $\boldsymbol{y} \in \mathbb{F}_q^n$, tel que $\boldsymbol{y} = \boldsymbol{c} + \boldsymbol{e}$, où $\boldsymbol{c}$ est un mot de $\mathcal{C}$ et $\boldsymbol{e}$ est une erreur de poids $\mathrm{wt}(\boldsymbol{e}) \le t := \frac{n-k}{2}$. Le but de l'exercice est de démontrer que l'algorithme retourne le message associé à $\boldsymbol{c}$, c'est-à-dire les coefficients du polynôme $C \in \mathbb{F}_q[X]_{<k}$ tel que $c_i = C(x_i)$ pour tout $i = 1, \dots, n$.

Dans tout l'exercice, on note $Y(X)$ le polynôme interpolateur des points (x_i, y_i), c'est-à-dire le polynôme $Y(X) \in \mathbb{F}_q[X]$ de plus petit degré tel que $y_i = Y(x_i)$. On note également $C(X) \in \mathbb{F}_q[X]$ le polynôme interpolateur des (x_i, c_i). Enfin, on note $A(X) = (X - x_1) \dots (X - x_n)$ et $E(X) = (X - x_{j_1}) \dots (X - x_{j_w})$ où $\{j_1, \dots, j_w\}$ sont les indices des erreurs présentes dans $\boldsymbol{e}$. Autrement dit $E(x_j) = 0$ si et seulement si $e_j \neq 0$.

L'algorithme de Gao (Algorithme 2) utilise une variante tronquée de l'algorithme d'Euclide étendu, décrite dans l'Algorithme 1.

Algorithme 1 : Algorithme d'Euclide tronqué, aussi noté `partial_xgcd`

Entrée : deux polynômes $A, B \in \mathbb{F}_q[X]$ tels que $\deg A = n$ et $\deg B \le n - 1$, et un entier $s \le n - 1$

Sortie : trois polynômes $R, U, V \in \mathbb{F}_q[X]$ tels que $UA + BV = R$ et $\deg R < s$

Initialiser $R_0 = A$, $R_1 = B$, $U_0 = 1$, $U_1 = 0$, $V_0 = 0$ et $V_1 = 1$.

Tant que $\deg R_1 \ge s$ **faire**

    Calculer le quotient Q de la division euclidienne de R_0 par R_1.

    Mettre à jour $(R_0, R_1) \leftarrow (R_1, R_0 - QR_1)$.

    Mettre à jour $(U_0, U_1) \leftarrow (U_1, U_0 - QU_1)$.

    Mettre à jour $(V_0, V_1) \leftarrow (V_1, V_0 - QV_1)$.

Retourner $R = R_1$, $U = U_1$ et $V = V_1$.

Question 1.– Donner des bornes (ou la valeur exacte lorsque c'est possible) sur les degrés des polynômes $A(X)$, $C(X)$, $E(X)$ et $Y(X)$.

Question 2.– Démontrer que $Y(X)E(X) \equiv C(X)E(X) \mod A(X)$. En déduire qu'il existe un couple de polynômes $(N, E) \in \mathbb{F}_q[X]$ tel que

$$Y(X)E(X) \equiv N(X) \mod A(X), \quad \deg E \le \tfrac{n-k}{2}, \quad \deg N \le k - 1 + \tfrac{n-k}{2}\,. \tag{1}$$

Question 3.– Démontrer que si (N, E) et (N', E') sont deux couples de solutions de (1), alors on a

$$\frac{N}{E} = \frac{N'}{E'} = C(X)\,.$$

Algorithme 2 : Algorithme de Gao [1] pour le décodage des codes de Reed-Solomon

Entrée : un mot $\boldsymbol{y} \in \mathbb{F}_q^n$ tel que $\boldsymbol{y} = \boldsymbol{c} + \boldsymbol{e}$, $\boldsymbol{c} \in \mathcal{C}$ et $\mathrm{wt}(\boldsymbol{e}) \leq \frac{n-k}{2}$
Sortie : le message $\boldsymbol{m}$ initial

Calculer $A(X) = (X - x_1)(X - x_2) \cdots (X - x_n)$.
Calculer $Y(X)$ le polynôme interpolateur des points $\{(x_i, y_i)\}_{1 \leq i \leq n}$.
Calculer $s = (k+n)/2$.
Effectuer $(R, U, V) \leftarrow \texttt{partial\_xgcd}(A, Y, s)$.
Calculer $C(X) = \frac{R(X)}{V(X)}$.
Retourner les coefficients de C.

Note. L'équation (1) est parfois appelée « équation-clé » pour le décodage. Observons qu'on peut l'exprimer comme un système d'équations linéaires à $(\deg N + 1) + \deg E \leq n - 1$ inconnues (le polynôme E est unitaire) et n équations. On peut résoudre ce système par une élimination Gaussienne en temps $O(n^3)$. Dans cet exercice, on va chercher à résoudre l'équation-clé par des divisions euclidiennes de polynômes.

Question 4.– Dans l'algorithme d'Euclide étendu, on a $U_i A + V_i B = R_i$ à chaque étape de l'algorithme. Démontrer la relation suivante sur les degrés de certains de ces polynômes :

$$\deg V_i + \deg R_{i-1} = \deg A\,.$$

Question 5.– Démontrer que pour $s = \frac{n+k}{2}$, le triplet de sortie (R, U, V) de la procédure $\texttt{partial\_xgcd}(A, Y, s)$ est tel que $(N = R, E = V)$ vérifie l'équation (1).

Question 6.– Conclure sur la correction de l'algorithme et donner une borne sur sa complexité en fonction de n.

Solutions de l'Exercice 3.

Solution Q1. On a $\deg A = n$, $\deg C \leq k - 1$, $\deg E \leq \frac{n-k}{2}$ et $\deg Y \leq n - 1$.

Solution Q2. On observe que $e_i E(x_i) = 0$ pour tout $i \in \{1, \ldots, n\}$. Donc, $y_i E(x_i) = c_i E(x_i) + e_i E(x_i) = c_i E(x_i)$. Autrement dit, $Y(x_i)E(x_i) = C(x_i)E(x_i)$ ce qui se transcrit comme $Y(X)E(X) \equiv C(X)E(X) \mod A(X)$. En posant $N = CE$, on a le résultat escompté.

Solution Q3. Si (N, E) et (N', E') sont des solutions, alors on pose $F = NE' - N'E$. On a $F \equiv YE'E - YEE' \equiv 0 \mod A$. De plus, $\deg F = (k - 1 + \deg n - k2) + \frac{n-k}{2} \leq n - 1$ et $\deg A = n$ donc $F = 0$. Ainsi, $\frac{N}{E} = \frac{N'}{E'} = C$ car le couple (CE, E) est un solution particulière.

Solution Q4. Par définition on a $R_{i+1} = R_{i-1} - Q_i R_i$ avec $\deg R_{i+1} < \deg R_i$. Ceci implique que $\deg R_{i-1} = \deg Q_i R_i = \deg Q_i + \deg R_i$.

D'autre part, $V_{i+1} = V_{i-1} - Q_i V_i$ et $\deg Q_i > 0$. Par conséquent, $\deg V_{i+1}$ est strictement croissante et $\deg V_{i+1} = \deg Q_i V_i = \deg Q_i + \deg V_i$.

On en déduit que $\deg V_{i+1} + \deg R_i = \deg V_i + \deg R_{i-1}$, puis par induction, $\deg V_i + \deg R_{i-1} = \deg V_1 + \deg R_0 = \deg A$.

Solution Q5.

En sortie de $\texttt{partial\_xgcd}(A, Y, s)$, on obtient R_i, U_i, V_i tels que $R_i = AU_i + YV_i$. En posant $E = V_i$ et $N = R_i$, on a donc $YE \equiv N \mod A$. Il reste a vérifier les constraintes sur les degrés.

La condition d'arrêt de la boucle du pgcd tronqué est $\deg R_i < s$. On a donc $\deg R_{i-1} \geq s$. Puis,

$$\deg N = \deg R_i \leq s - 1 = \frac{n+k}{2} - 1 = k - 1 + \frac{n-k}{2} \quad \text{et} \quad \deg V_i = \deg A - \deg R_{i-1} \leq n - s = \frac{n-k}{2}\,.$$

Solution Q6. L'algorithme termine en calculant $\frac{R}{V} = \frac{N}{E} = C$ d'après la **Question 3.**

Références

[1] Shuhong Gao. A New Algorithm for Decoding Reed-Solomon Codes. In Vijay K. Bhargava, H. Vincent Poor, Vahid Tarokh, and Seokho Yoon, editors, *Communications, Information and Network Security*, pages 55–68. Springer US, 2003.