Agrégation externe de mathématiques
Épreuve de modélisation, option C : algèbre et calcul formel

(*Texte public*)

Résumé : On s'intéresse à l'étude d'une famille de codes correcteurs d'erreurs pour une métrique différente de la classique distance de Hamming. Cela mène à l'étude d'un anneau non-commutatif de polynômes pour lesquels on arrive tout de même à définir une division euclidienne qui permet un décodage rapide.

Mots clefs : corps finis, codes correcteurs, polynômes

► *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. Vous êtes laissé(e) libre d'organiser votre discussion comme vous l'entendez. Des suggestions de développement, largement indépendantes les unes des autres, vous sont proposées en fin de texte. Vous n'êtes pas tenu(e) de les suivre. Il vous est conseillé de mettre en lumière vos connaissances à partir du fil conducteur constitué par le texte. Le jury appréciera que la discussion soit accompagnée d'exemples traités sur ordinateur.*

1. Introduction

En théorie des codes correcteurs, on utilise traditionnellement la métrique de Hamming, bien adaptée à un modèle de canal de transmission qui est symétrique (le 0 et le 1 jouent des rôles symétriques) et où ce que subit un bit est indépendant de ce que subit un autre. Dans certaines situations, les mots de codes sont naturellement transmis sous forme de matrice de bits, les erreurs prenant fréquemment la forme d'effacement de certaines lignes ou colonnes. Citons par exemple le cas d'une transmission s'appuyant sur n antennes ; un mot de code sera un tableau de mn bits, la première ligne correspondant aux bits envoyés par les n antennes à l'instant 0, la seconde ligne à l'instant 1, etc. La défaillance d'une antenne se traduit par l'effacement d'une colonne, alors qu'une interférence ponctuelle dans le temps se traduit par l'effacement d'une ligne.

On s'intéresse à une métrique adaptée à ce type de transmission, et à certains codes correcteurs associés.

2. Codes en métrique rang

Dans tout le texte, le corps fini à 2 éléments sera noté $\mathbb{F}_2$. Plutôt que de raisonner avec des matrices à m lignes et n colonnes à coefficients dans $\mathbb{F}_2$, on va travailler avec des vecteurs à coefficients dans $\mathbb{F}_{2^m}$, le corps à 2^m éléments. Pour ce faire, on se donne une base de $\mathbb{F}_{2^m}$ en tant que $\mathbb{F}_2$–espace vectoriel. Ce choix de base induit un isomorphisme de $\mathbb{F}_2$–espaces vectoriels entre $\mathbb{F}_2^m$ et $\mathbb{F}_{2^m}$, puis un isomorphisme de $\mathbb{F}_2$–espaces vectoriels entre $\mathscr{M}_{m,n(\mathbb{F}_2)}$ (l'espace des matrices à m lignes et n colonnes à coefficients dans $\mathbb{F}_2$) et $\mathbb{F}_{2^m}^n$.

Exemple 1. *Pour $m = 5$ et $n = 4$, représentons $\mathbb{F}_{2^5}$ par $\mathbb{F}_2[\omega] := \mathbb{F}_2[x]/(x^5 + x^2 + 1)$. Les éléments de $\mathfrak{M}_{5,4}(\mathbb{F}_2)$ seront représentés par des vecteurs de $\mathbb{F}_{2^5}^4$. Par exemple, la matrice*

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

sera représentée par le vecteur $c = (\omega^4 + 1, \omega^3 + \omega + 1, \omega^4 + \omega^2, \omega^2 + 1) \in \mathbb{F}_{2^5}^4$.

Cet isomorphisme d'espace vectoriels entre $\mathscr{M}_{m,n}(\mathbb{F}_2)$ et $\mathbb{F}_{2^m}^n$ permet de définir la notion de *rang* d'un vecteur $\mathbb{F}_{2^m}$ comme le rang de la matrice correspondante par cet isomorphisme. Par exemple le vecteur c décrit dans l'exemple 1 est de rang 3.

Lemme 1. *Le rang d'un vecteur $c \in \mathbb{F}_{2^m}^n$ ne dépend pas du choix de la base de $\mathbb{F}_{2^m}$ comme $\mathbb{F}_2$–espace vectoriel.*

Démonstration. Si pour un choix de base fixé $c \in \mathbb{F}_{2^m}^n$ correspond à $M \in \mathscr{M}_{m,n}(\mathbb{F}_2)$ alors, pour une autre base, c correspond à $P^{-1} \cdot M$ où P est la matrice de changement de base. Le rang est donc inchangé. □

Lemme 2. *Le rang de $c = (c_1, \ldots, c_n) \in \mathbb{F}_{2^m}^n$ est égal à la dimension du sous–$\mathbb{F}_2$–espace vectoriel de $\mathbb{F}_{2^m}$ engendré par $c_1, \ldots, c_n$.*

Définition 1. *Soient m et n deux entiers strictement positifs. Sur l'ensemble $\mathbb{F}_{2^m}^n$, on définit la distance $d_r(x, y)$ entre deux vecteurs x et y comme le rang de $x - y$.*

On vérifie que cela définit bien une distance au sens usuel.

Définition 2. *Un code $\mathscr{C}$ est un sous–$\mathbb{F}_{2^m}$–espace vectoriel de $\mathbb{F}_{2^m}^n$. La distance minimale de $\mathscr{C}$ est la valeur minimale de $d_r(c_1, c_2)$ lorsque c_1 et c_2 sont deux éléments distincts de $\mathscr{C}$. Pour des raisons de linéarité c'est aussi le rang minimal d'un élément $c \in \mathscr{C} \setminus \{0\}$.*

3. Une famille de codes

Définition 3. *Un polynôme $\mathbb{F}_2$–linéaire sur $\mathbb{F}_{2^m}$ est un élément de $\mathbb{F}_{2^m}[x]$ dont tous les coefficients correspondant à un degré qui n'est pas une puissance de 2 sont nuls.*

À tout polynôme $\mathbb{F}_2$–linéaire P, on peut associer sa fonction polynomiale de $\mathbb{F}_{2^m}$ dans lui-même $x \mapsto P(x)$. Comme tous les monômes de P sont de degré une puissance de 2, cette fonction est un endomorphisme du $\mathbb{F}_2$–espace vectoriel $\mathbb{F}_{2^m}$.

Lemme 3. *Les racines dans $\mathbb{F}_{2^m}$ d'un polynôme $\mathbb{F}_2$–linéaire P de degré 2^k sur $\mathbb{F}_{2^m}$ sont les éléments du noyau de l'endomorphisme associé. Elles forment donc un $\mathbb{F}_2$–espace vectoriel dont la dimension est au plus k.*

Soit $\mathbf{g} = (g_1, \ldots, g_n)$ un vecteur de n éléments de $\mathbb{F}_{2^m}$ linéairement indépendants sur $\mathbb{F}_2$ (donc $n \leqslant m$), et soit k un entier strictement positif avec $k < n$. On définit le code $\mathrm{Gab}_k(\mathbf{g})$ comme l'ensemble des vecteurs de la forme

$$\big(P(g_1), P(g_2), \ldots, P(g_n)\big),$$

où P décrit l'ensemble des polynômes $\mathbb{F}_2$–linéaires sur $\mathbb{F}_{2^m}$ de degré strictement inférieur à 2^k.

Proposition 1. *Le code* $\mathrm{Gab}_k(\mathbf{g})$ *est un code linéaire : ses éléments forment un* $\mathbb{F}_{2^m}$*–espace vectoriel de dimension* k*. Sa distance minimale est* $d = n - k + 1$*. Une* $\mathbb{F}_{2^m}$*–base de* $\mathrm{Gab}_k(\mathbf{g})$ *est donnée par les lignes de la matrice suivante, à coefficients dans* $\mathbb{F}_{2^m}$ *:*

$$G = \left(\begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & P_1(g_{k+1}) & \cdots & P_1(g_n) \\ 0 & 1 & \cdots & 0 & P_2(g_{k+1}) & \cdots & P_2(g_n) \\ \vdots & & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 & P_k(g_{k+1}) & \cdots & P_k(g_n) \end{array}\right),$$

où P_i *est l'unique polynôme* $\mathbb{F}_2$*–linéaire de degré au plus* 2^{k-1} *vérifiant pour tout* $j \leqslant k$*,* $P_i(g_j) = \delta_{ij}$*.*

Démonstration. L'unicité des P_i est une conséquence du Lemme 3 et leur existence sera prouvée en § 4.

Montrons maintenant les assertions sur la dimension et la distance minimale. Soit P un polynôme $\mathbb{F}_2$–linéaire non nul de degré $< 2^k$. D'après le lemme 3 et le théorème du rang, la dimension de l'espace engendré par $P(g_1), \ldots, P(g_n)$ est supérieure ou égale à $n - k + 1$. On en déduit grâce au lemme 2 que $d \geqslant n - k + 1$. On peut ensuite vérifier que les lignes de la matrice G sont de rang au plus $n - k + 1$ ce qui entraîne $d = n - k + 1$.

On a en particulier montré que si P est un polynôme $\mathbb{F}_2$–linéaire on nul de degré $< 2^k$, alors $(P(g_1), \ldots, P(g_n))$ est non nul. L'application $P \mapsto (P(g_1), \ldots, P(g_n))$ est donc injective, ce qui nous fournit la dimension du code. □

4. Arithmétique des polynômes $\mathbb{F}_2$–linéaires

L'ensemble des polynômes $\mathbb{F}_2$–linéaires est un $\mathbb{F}_{2^m}$–espace vectoriel ; de plus, si P et Q sont deux polynômes $\mathbb{F}_2$–linéaires, leur composition $P \circ Q$ en est aussi un.

Proposition 2. *L'ensemble des polynômes* $\mathbb{F}_2$*–linéaires muni des lois* $+$ *et* $\circ$ *est un anneau.*

Attention! Il s'agit d'un anneau non-commutatif, la composition n'étant pas commutative.

Proposition 3. *L'anneau des polynômes* $\mathbb{F}_2$*–linéaires sur* $\mathbb{F}_{2^m}$ *est commutatif si et seulement si* $m = 1$*.*

En pratique, l'algorithme naïf pour composer deux polynômes de degré 2^k et 2^l va coûter de l'ordre de kl multiplications dans $\mathbb{F}_{2^m}$ ainsi que des additions et des élévations au carré que l'on négligera par la suite. En effet, si l'on prend une représentation adaptée du corps

$\mathbb{F}_{2^m}$ à l'aide d'un polynôme ayant peu de coefficients non nuls, le coût de ces deux dernières opérations est presque linéaire en m. C'est par exemple le cas si on représente $\mathbb{F}_{2^{127}}$ comme $\mathbb{F}_2[x]/(x^{127}+x+1)$.

Le lemme 3 donne la structure des racines d'un polynôme $\mathbb{F}_2$–linéaire comme espace vectoriel. Inversement, étant donnés k éléments $\mathbb{F}_2$–linéairement indépendants $e_1,\dots,e_k$ de $\mathbb{F}_{2^m}$, on peut trouver un polynôme ayant le sous-espace engendré par les e_i comme racines grâce à la construction suivante :

— $P_1(x) = x^2 - e_1 x$;

— Pour tout $i \in [2,k]$, $P_i(x) = (x^2 - P_{i-1}(e_i)x) \circ P_{i-1}(x)$.

Cette construction peut s'effectuer en $O(k^2)$ multiplications dans $\mathbb{F}_{2^m}$.

On peut résumer ces calculs sur les polynômes $\mathbb{F}_2$–linéaires dans le lemme suivant :

Lemme 4. *On suppose que l'on peut négliger les carrés et les additions dans $\mathbb{F}_{2^m}$.*

— *Les compositions $P \circ Q$ et $Q \circ P$ de deux polynômes $\mathbb{F}_2$–linéaires de degrés 2^k et 2^l peuvent se calculer en $O(kl)$ opérations dans $\mathbb{F}_{2^m}$.*

— *Un polynôme $\mathbb{F}_2$–linéaire de degré 2^k s'annulant en k éléments distincts se calcule en $O(k^2)$ opérations dans $\mathbb{F}_{2^m}$.*

La construction d'un polynôme à partir de ses racines fournit une preuve constructive de l'existence de la matrice G de la proposition 1.

Plus surprenant, on peut aussi définir une division euclidienne à droite et une division euclidienne à gauche :

Proposition 4. *Soient A et B deux polynômes $\mathbb{F}_2$–linéaire sur $\mathbb{F}_{2^m}$ avec B non nul. Alors il existe deux couples uniques de polynômes $\mathbb{F}_2$–linéaires (q_1,r_1) et (q_2,r_2) tels que*

$$\begin{array}{rcll} A & = & B \circ q_1 + r_1 & \textit{avec} \quad \deg(r_1) < \deg(B) \\ A & = & q_2 \circ B + r_2 & \textit{avec} \quad \deg(r_2) < \deg(B) \end{array}$$

La preuve de cette proposition est constructive : l'algorithme classique de division euclidienne de polynômes s'étend à ces divisions euclidiennes à droite et à gauche. Comme la composition de polynômes a pour effet d'élever les coefficients à certaines puissances de 2, il est nécessaire de le prendre en compte et le compenser lors de la division euclidienne, mais cela ne perturbe pas la stratégie générale habituelle.

5. Décodage

Nous nous intéressons désormais au problème du décodage, c'est-à-dire de retrouver le mot de code le plus proche d'un mot reçu pour la métrique considérée.

Soit P un polynôme $\mathbb{F}_2$–linéraire de degré inférieur à 2^k, et soit $\mathbf{c} = (P(g_1),\dots,P(g_n))$ le mot de code associé dans $\mathrm{Gab}_k(\mathbf{g})$. Le message est bruité par une erreur $\mathbf{e} = (e_1,\dots,e_n)$ de petit $\mathbb{F}_2$–rang t, si bien que le message reçu est

$$\mathbf{y} = \mathbf{c} + \mathbf{e}. \tag{1}$$

On souhaite néanmoins retrouver P. On va tâcher de le retrouver en traduisant le problème en termes de polynômes $\mathbb{F}_2$–linéaires.

Lemme 5. *Il existe un polynôme $\mathbb{F}_2$–linéaire V de degré au plus 2^t tel que pour tout i, on ait $V(y_i) = V \circ P(g_i)$.*

Indication : Prendre V dont les racines sont le sous–$\mathbb{F}_2$–espace de $\mathbb{F}_{2^m}$ engendré par les e_i.

Le polynôme P n'est pas connu (c'est lui que l'on cherche), si bien que trouver V tel que dans le lemme ne peut pas se faire directement. On va en fait chercher à résoudre un problème plus vaste dont les solutions contiendront les solutions du problème initial : soit N le produit $V \circ P$ qui est un polynôme $\mathbb{F}_2$–linéaire de degré au plus 2^{t+k-1}. On cherche les couples de polynômes $\mathbb{F}_2$–linéaires (V, N) tels que $\deg(V) \leqslant 2^t$ et $\deg(N) \leqslant 2^{k+t-1}$ et

$$\forall i \in \{1, \ldots, n\}, \quad V(y_i) = N(g_i). \tag{2}$$

Ce système linéaire a n équations et $2t + k + 1$ inconnues (les coefficients de V et de N) dont on peut calculer l'espace vectoriel des solutions.

Si l'espace des solutions est de dimension 0, on en déduit qu'il n'y a pas de mot de code à distance inférieure ou égale à t. Sinon, pour chaque solution non nulle (V, N), on calcule le quotient P dans la division euclidienne à gauche de N par V, et on teste la validité de cette solution. Si le rang de l'erreur est suffisamment petit, la solution sera toujours valide, comme l'affirme l'énoncé suivant.

Proposition 5. *Si $t \leqslant (n-k)/2$, et si $\mathbf{e}$ est de rang $\leqslant t$, alors pour toute solution $(V, N) \neq (0, 0)$ du système (2) on a : $N = V \circ P$.*

Démonstration. Soit (V, N) une solution non nulle de (2). D'après (1), on a

$$\forall i \in \{1, \ldots, n\}, \ (V \circ P - N)(g_i) = -V(e_i). \tag{3}$$

Supposons que $V \circ P - N \neq 0$, alors ce polynôme est de degré au plus 2^{k+t-1} et, d'après la proposition 1, on en déduit que

$$\mathrm{Rang}(V \circ P(g_1), \ldots, V \circ P(g_n)) \geqslant n - (k + t) + 1. \tag{4}$$

D'un autre côté, $\mathrm{Rang}(V(e_1), \ldots, V(e_n)) \leqslant t$. De cette affirmation et de (3) et (4), on déduit

$$t > \frac{n-k}{2}, \tag{5}$$

ce qui contredit l'hypothèse de l'énoncé. Donc, $V \circ P - N = 0$. □

La conséquence de cette dernière proposition est qu'il suffit de prendre un couple (V, N) au hasard parmi les solutions non nulles du système (2) et de calculer la division euclidienne à gauche de N par V.

Le coût principal dans cet algorithme de décodage est celui de la résolution d'un système d'équations linéaires.

Suggestions pour le développement

► *Soulignons qu'il s'agit d'un menu à la carte et que vous pouvez choisir d'étudier certains points, pas tous, pas nécessairement dans l'ordre, et de façon plus ou moins fouillée. Vous pouvez aussi vous poser d'autres questions que celles indiquées plus bas. Il est très vivement souhaité que vos investigations comportent une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats.*

— Vérifier que d_r est bien une distance.
— Détailler les preuves des résultats énoncés dans le texte.
— Proposer un exemple de code $\mathrm{Gab}_k(\mathbf{g})$, et illustrer comment un message peut-être codé à l'aide de sa matrice génératrice.
— Détailler le coût d'un carré dans $\mathbb{F}_{2^m}$ et pourquoi on le néglige.
— Expliciter l'algorithme de division euclidienne dans le contexte des polynômes $\mathbb{F}_2$–linéaires. Par exemple, dans le corps $\mathbb{F}_{2^6} = \mathbb{F}_2[w]$ avec $w^6 + w^4 + w^3 + w + 1 = 0$, calculer la division Eulidienne à gauche de $X^8 + (w+1)X^4 + (w+1)X^2$ par $X^4 + wX^2 + X$.
— Considérons de nouveau le corps $\mathbb{F}_{2^6} = \mathbb{F}_2[w]$ avec $w^6 + w^4 + w^3 + w + 1 = 0$ et la base $\mathbf{g} = (1, w, w^2, w^3, w^4, w^5)$. Les mots suivants appartiennent-ils à $Gab_3(\mathbf{g})$? Si non, l'algorithme proposé en section 5 permet-il de trouver le mot de code le plus proche.

$\left(0, w^4 + w^2, w^5 + w^2 + w + 1, w^5 + w^4 + w, w^5 + w^2, w^5 + w^4 + w^2 + w\right)$

$\left(0, w^2 + 1, w^4 + 1, w^4 + w^3 + w, w^5 + w^4 + w^2 + w, w^5 + w^4\right)$

$\left(w, w^4 + w^3 + w, w^4 + w + 1, w^4 + w^2 + w + 1, w^5 + w^4 + w^2 + w, w^4 + w^3 + w^2 + w\right)$

— Du point de vue du décodage, la borne sur le rang de l'erreur de la proposition 5 vous semble-t-elle optimale ? Quel type de difficultés pourrait-on rencontrer dans le cas d'une erreur de rang $> (n-k)/2$?