TS226 - Codes correcteurs
année 2019/2020

Romain Tajan

Exercice 1 -

Soit la matrice G à coefficients dans $\mathbb{F}_2$ donnée par

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

Question 1. Vérifier que G est la matrice génératrice d'un code linéaire dont on précisera la longueur, la dimension ainsi que le rendement de code.

Question 2. Mettre le code sous forme systématique et trouver tous les mots de codes.

Question 3. Trouver une matrice de parité H, et en déduire les équations de parité du code.

Question 4. Déterminer la distance minimale de ce code ainsi que son spectre de poids.

Question 5. Combien d'erreurs peut-il détecter ?

Question 6. Combien d'erreurs peut-il corriger, par décodage au plus proche voisin ? Le code est-il MDS, parfait ?

Question 7. Après transmission dans un canal binaire symétrique, on reçoit la séquence $[1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0]$. Quelles sont les conditions sur les erreurs présentes dans la séquence, pour garantir un décodage parfait ? En supposant ces conditions satisfaites, effectuer le décodage de la séquence, à l'aide de la méthode du syndrome.

Exercice 2 - Construction de Plotkin

Soient $\mathcal{C}_1 = [n, k_1, d_1]$ et $\mathcal{C}_2 = [n, k_2, d_2]$ deux codes linéaires binaires de même longueur n. La méthode de Plotkin permet de construire un nouveau code $\mathcal{C}$ à partir de deux codes linéaires $\mathcal{C}_1$ et $\mathcal{C}_2$, donné par

$$\mathcal{C} = \{(\mathbf{c_1}, \mathbf{c_1} + \mathbf{c_2}) : \mathbf{c_1} \in \mathcal{C}_1; \mathbf{c_2} \in \mathcal{C}_2\}$$

Question 1. Montrer que le code de Plotkin est linéaire, et donner ses longueur, dimension et rendement.

Soient $\mathbf{c} = (\mathbf{c_1}, \mathbf{c_1} + \mathbf{c_2}) \in \mathcal{C}$ et $\mathbf{c}' = (\mathbf{c'_1}, \mathbf{c'_1} + \mathbf{c'_2}) \in \mathcal{C}$ soit $\mathbf{c}'' = \mathbf{c} + \mathbf{c}'$, par construction de $\mathcal{C}$, $\mathbf{c}'' = (\mathbf{c_1} + \mathbf{c'_1}, \mathbf{c_1} + \mathbf{c'_1} + \mathbf{c_2} + \mathbf{c'_2})$ donc $\mathbf{c}'' \in \mathcal{C}$ en effet, $\mathbf{c_1} + \mathbf{c'_1} \in \mathcal{C}_1$ et $\mathbf{c_2} + \mathbf{c'_2} \in \mathcal{C}_2$ car $\mathcal{C}_1$ et $\mathcal{C}_2$ sont des codes linéaires.

Question 2. En notant, G_1 et G_2 les matrices génératrices de $\mathcal{C}_1$ et $\mathcal{C}_2$, donner une matrice génératrice G de $\mathcal{C}$.

Comme G_1 et G_2 sont les matrices génératrices de $\mathcal{C}_1$ et $\mathcal{C}_2$, on a $\mathbf{c}_1 = \mathbf{u}_1 G_1$ et $\mathbf{c}_2 = \mathbf{u}_2 G_2$ où $\mathbf{u}_1$ est un message de longueur k_1 et $\mathbf{u}_2$ est un message de longueur k_2. On en déduit que $\mathbf{c} = (\mathbf{u}_1, \mathbf{u}_2) \begin{pmatrix} G_1 & G_1 \\ 0 & G_2 \end{pmatrix}$

Question 3. Montrer que $d_{min}(\mathcal{C}) \leq \min\{2d_1, d_2\}$

Les mots de codes de la forme $(\mathbf{c}_1, \mathbf{c}_1)$ sont dans $\mathcal{C}$, en effet ils sont obtenus en considérant $\mathbf{c}_1 \in \mathcal{C}_1$ et $\mathbf{0} \in \mathcal{C}_2$. Comme $\mathcal{C}_1$ est de distance minimale d_1, on en déduit que $d \leq 2d_1$. Les mots de codes de la forme $(\mathbf{0}, \mathbf{c}_2)$ sont dans $\mathcal{C}$, en effet ils sont obtenus en considérant $\mathbf{0} \in \mathcal{C}_1$ et $\mathbf{c}_2 \in \mathcal{C}_2$. Comme $\mathcal{C}_2$ est de distance minimale d_2, on en déduit que $d \leq d_2$. En couplant ces deux résultats, on obtient finalement

$$d \leq \min(2d_1, d_2)$$

Question 4. Montrer que pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{F}_2^n$,

$$w_H(\mathbf{x} + \mathbf{y}) = w_H(\mathbf{x}) + w_H(\mathbf{y}) - 2w_H(\mathbf{x} \odot \mathbf{y}),$$

où $\odot$ représente le produit de Hadamard (produit terme à terme).

Soient $\mathbf{x}, \mathbf{y} \in \mathbb{F}_2^n$, $w_H(\mathbf{x} + \mathbf{y})$ représente le nombre d'éléments non nuls dans $\mathbf{x} + \mathbf{y}$. De plus $x_i + y_i = 1$ si et seulement si ($x_i = 1$ et $y_i = 0$) ou ($x_i = 0$ et $y_i = 1$). Donc on a

$$w_H(\mathbf{x} + \mathbf{y}) = \underbrace{w_H(\mathbf{x}) + w_H(\mathbf{y})}_{\substack{\text{Ici on a compté en trop} \\ \text{les cas où } x_i = y_i = 1 \text{ (donc } x_i + y_i = 0) \\ \text{de plus on les a comptés 2 fois}}} \quad -2w_H(\underbrace{\mathbf{x} \odot \mathbf{y}}_{x_i \cdot y_i = 1 \Leftrightarrow x_i = y_i = 1})$$

Question 5. Déduire de la question précédente que pour $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_1 + \mathbf{c}_2) \in \mathcal{C}$, $w_H(\mathbf{c}) \geq w_H(\mathbf{c}_2)$.

Soit $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_1 + \mathbf{c}_2) \in \mathcal{C}$,

$$\begin{aligned} w_H(\mathbf{c}) &= w_H(\mathbf{c}_1) + w_H(\mathbf{c}_1 + \mathbf{c}_2) \\ &= 2w_H(\mathbf{c}_1) + w_H(\mathbf{c}_2) - 2w_H(\mathbf{c}_1 \odot \mathbf{c}_2) \\ &= 2(w_H(\mathbf{c}_1) - w_H(\mathbf{c}_1 \odot \mathbf{c}_2)) + w_H(\mathbf{c}_2) \end{aligned}$$

Il suffit maintenant d'observer qu'il ne peut pas y avoir plus de 1 dans $\mathbf{c}_1 \odot \mathbf{c}_2$ que dans $\mathbf{c}_1$, en effet pour qu'un élément de $\mathbf{c}_1 \odot \mathbf{c}_2$ vaille 1, il est nécessaire (mais pas suffisant) que l'élément de $\mathbf{c}_1$ vaille aussi 1. On en déduit que $w_H(\mathbf{c}_1) - w_H(\mathbf{c}_1 \odot \mathbf{c}_2) \geq 0$ et que $w_H(\mathbf{c}) \geq w_H(\mathbf{c}_2) \geq d_2$.

Question 6. Déduire finalement que $d_{min}(\mathcal{C}) = \min\{2d_1, d_2\}$.

Des questions précédentes, nous avons déjà l'encadrement suivant

$$d_2 \leq d_{min}(\mathcal{C}) \leq \min\{2d_1, d_2\}$$

Il y a maintenant deux cas de figure :

— Si $d_2 \leq 2d_1$ alors $\min\{2d_1, d_2\} = d_2$ et $d_2 \leq d_{min}(\mathcal{C}) \leq d_2$

— Si $2d_1 \leq d_2$ alors $\min\{2d_1, d_2\} = 2d_1$ et $2d_1 \leq d_2 \leq d_{min}(\mathcal{C}) \leq 2d_1$ d'où $d_{min}(\mathcal{C}) = 2d_1$

On en conclut le résultat demandé $d_{min}(\mathcal{C}) = \min\{2d_1, d_2\}$.

Exercice 3 - Extension du code de parité

Considérons un message à transmettre de 4 bits $\mathbf{u} = (u_0, \ldots, u_3)$. On ajoute au message 5 bits $\mathbf{p} = (p_0, \ldots, p_4)$ pour former le mot de code $\mathbf{c} = (p_0, \ldots, p_4, u_0, \ldots, u_3)$ tels que la matrice

$$M = \begin{bmatrix} u_0 & u_2 & p_0 \\ u_1 & u_3 & p_1 \\ p_2 & p_3 & p_4 \end{bmatrix}$$

ait des lignes et colonnes de somme nulle.

Question 1. Vérifier que le code défini est bien linéaire. Quelles sont sa longueur n et sa dimension k ?

Soit $\mathbf{c} \in \mathcal{C}$, les bits p_i sont ajoutés de sorte à avoir les sommes des lignes et des colonnes nulles. On en déduit les équations suivantes :

- $p_0 = u_0 + u_2$
- $p_1 = u_1 + u_3$
- $p_2 = u_0 + u_1$
- $p_3 = u_2 + u_3$
- $p_4 = u_0 + u_1 + u_2 + u_3$

Ce code est donc décrit par l'équation de parité suivante

$$(p_0, p_1, p_2, p_3, p_4, u_0, u_1, u_2, u_3) \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} = (0, 0, 0, 0, 0)$$

Ce qui est de la forme $\mathbf{c}H^T = \mathbf{0}$ avec $H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$ On en déduit que $\mathcal{C}$ est un code linéaire.

Question 2. Donner une matrice génératrice de ce code.

La matrice H présentée dans la question précédente est sous forme systématique donc

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

Question 3. Quelle est la distance minimale ? Combien d'erreur peut-il corriger/détecter ?

Les lignes de G possèdent toutes 4 valeurs à 1, on en déduit que $d_{min} \leq 4$. On cherche maintenant à savoir s'il existe un mot de code $\mathbf{c} \neq \mathbf{0}$ ayant moins de 4 composantes à 1.

On commence par remarquer que si $\mathbf{c} \neq \mathbf{0}$ alors $\mathbf{u} \neq \mathbf{0}$. Les lignes de G représentant les mots de codes obtenus pour les messages de poids de Hamming 1. Il nous reste à observer les mots de codes obtenus pour les poids 2 et 3.

Prenons le cas où $u_0 = u_1 = 1$ et $u_2 = u_3 = 0$, on remarque que $p_0 = p_1 = 1$ et $p_2 = p_3 = p_4 = 0$ donnant $w_H(\mathbf{c}) = \mathbf{4}$. Par construction du code, ce résultat restera inchangé si on échange de deux colonnes de la matrice M ou si on transpose M. Il nous reste donc le cas $u_0 = u_3 = 1$ et $u_1 = u_2 = 0$ qui donne $p_0 = p_1 = p2 = p3 = 1$ et $p_4 = 0$ donnant $w_H(\mathbf{c}) = \mathbf{6}$. Ce résultat étant inchangé si on transpose la matrice M, on peut en déduire qu'il n'y a pas de mot de code de poids inférieur à 4 obtenus par encodage d'un message contenant 2 bits.

Le cas $w_h(\mathbf{u}) = 3$ est plus simple. En effet, si $w_h(\mathbf{u}) = 3$, alors $p_4 = u_0 + u_1 + u_2 + u_3 = 1$ donc $w_h(\mathbf{c}) \geq 4$

On en conclut donc que $d_{min}(\mathcal{C}) = 4$ et que ce code peut corriger 1 erreur et détecter 3 erreurs.

Exercice 4 - Sur le code à répétitions et le code de parité

Soit $\mathcal{C}$ le code à n répétitions encodant des messages de $k = 1$ bit.

Question 1. Énumérer tous les mots de codes possibles.
Les mots de codes pour le code à n répétitions sont $\mathbf{c}_0 = (0, 0, \ldots, 0)$ et $\mathbf{c}_1 = (1, 1, \ldots, 1)$.

Question 2. Quelle est la distance minimale de ce code, son spectre des poids.
La distance minimale de ce code est n, il n'y a qu'un seul mot de ce poids là donc $A_n = 1$.

Question 3. Ce code est-il MDS ? Un code est MDS s'il vérifie la borne de Singleton : $d_{min}(\mathcal{C}) = n - k + 1$ où n est la longueur du code et k sa dimension. Ici $k = 1$ et $d_{min}(\mathcal{C}) = n$, ce code est MDS.

Question 4. Donner une matrice génératrice G pour $\mathcal{C}$ sous la forme systématique. Pour un code de répétition, la matrice génératrice G est donnée par

$$G = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

Question 5. Déduire de la question précédente une matrice de parité H pour $\mathcal{C}$.
Il convient de remarquer que la matrice G donnée précédemment est de la forme (P, I_1) où I_1 est la matrice identité de taille 1. Donc la matrice de parité s'obtient simplement comme $H = (I_{n-1}, P^T)$ d'où

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Question 6. Montrer que le code de parité est le code dual du code de répétition. Le code dual du code de répétition possède H comme matrice génératrice. Soit $\mathbf{u}'$ un message de taille $n - 1$, soit $\mathbf{c}' = H\mathbf{u}'$, par construction de H, on a $\mathbf{c}' = (\mathbf{u}', \sum_k u'_k)$ ce qui correspond bien à un code de parité.

Exercice 5 - Borne de Plotkin

Soit $\mathcal{C} = [n, k, d]$ de matrice génératrice G.

Question 1. Dans le cas d'un message $\mathbf{u} = (u_0, \ldots, u_{k-1}) \in \mathbb{F}_2^k$, l'équation vérifiée par le j-ème symbole c_j du mot de code $c = (c_0, \ldots, c_{n-1})$ associé, est donnée par

$$c_j = \sum_{i=0}^{k-1} u_i G_{i,j}$$

En supposant la j-ième colonne de G non nulle, montrer qu'il y a exactement 2^{k-1} mots de code de $\mathcal{C}$ dont la j-ème composante est non nulle.

Il y a deux cas possible, soit la colonne de G possède un nombre impair de 1 soit un nombre pair. Supposons le cas où G possède un nombre impair de 1. On remarque qu'il y a 2^k messages $\mathbf{u}$ possibles et que si $\mathbf{u}$ donne c_j alors $\mathbf{u}' = \mathbf{1} + \mathbf{u}$ donne $c'_j = \sum_{i=0}^{k-1} u_i G_{i,j} + \sum_{i=0}^{k-1} G_{i,j} = 1 + c_j$ il y aura donc 2^{k-1} messages qui donneront $c_j = 0$ et autant qui donneront $c_j = 1$.

Supposons le cas où G possède un nombre pair de 1 sur sa j-ème colonne, il y aura donc au moins 2 valeurs à 1 sur cette colonne. Notons i_0 tel que $G_{i_0,j} = 1$ alors $\mathbf{u}' = \mathbf{1}+\mathbf{u}+\mathbf{e}_{i_0}$ ($\mathbf{e}_{i_0}$ est le vecteur possédant un unique 1 à la position i_0) donnera $c'_j = \sum_{i=0}^{k-1} u_i G_{i,j} + \sum_{i=0}^{k-1} G_{i,j} + 1 = 1 + c_j$

Question 2. Déduire une borne supérieure sur la valeur de la somme des poids de tous les mots du code $\mathcal{C}$, i.e. un majorant de $\sum_{\mathbf{c}\in\mathcal{C}} w_H(\mathbf{c})$. Une borne supérieure peut être obtenue avec le raisonnement suivant :

$$\begin{aligned}
\sum_{\mathbf{c}\in\mathcal{C}} w_H(\mathbf{c}) &= \sum_{\mathbf{c}\in\mathcal{C}} \sum_{j=0}^{n-1} w_H(c_j) \\
&= \sum_{j=0}^{n-1} \sum_{\mathbf{c}\in\mathcal{C}} w_H(c_j) \\
&= \sum_{j=0}^{n-1} 2^{k-1} 1(\sum_i G_{i,j} > 0) \\
&\leq n2^{k-1}
\end{aligned}$$

Question 3. Déduire alors la borne de Plotkin :

$$d \leq \frac{n2^{k-1}}{2^k - 1}$$

Il y a $2^k - 1$ mots de code non nuls donc par définition de la distance minimale d, on a $\sum_{\mathbf{c}\in\mathcal{C}} w_H(\mathbf{c}) \geq (2^k - 1)d$. En combinant cette inégalité et la précédente, on obtient

$$d \leq \frac{n2^{k-1}}{2^k - 1}$$