Complexité du décodage des codes linéaires

B. Martin

19 novembre 2001

1 Présentation du problème

1.1 Distance de Hamming

Soit $\mathbb{F}_2$ l'alphabet binaire et x, y deux mots binaires de longueur n. On définit la *distance de Hamming* entre x et $y \in (\mathbb{F}_2)^n$ comme le nombre de positions où x et y diffèrent lettre à lettre.

$$d_H(x, y) = \text{Card}\{i : x_i \neq y_i, 1 \leq i \leq n\}$$

Question 1 : Montrer que la distance de Hamming vérifie les axiomes d'une distance :

$$d_H(x, y) \geq 0; \quad d_H(x, y) = 0 \Leftrightarrow x = y; \quad d_H(x, y) = d_H(y, x); \quad d_H(x, y) \leq d_H(x, z) + d_H(z, y)$$

On définit alors le *poids* d'un mot $x \in (\mathbb{F}_2)^n$, noté $\mathbf{w}(x)$ comme la distance de x au point $\underline{0} = (0, 0, \ldots, 0)$, i.e. $\mathbf{w}(x) = d_H(\underline{0}, x)$.

1.2 Codes linéaires

Un *code linéaire* C de longueur n et de dimension k (noté (n, k)-code linéaire) sur l'alphabet binaire est un sous-espace vectoriel de dimension k de l'espace vectoriel $(\mathbb{F}_2)^n$. Autrement dit, C vérifie :

$$\forall u, v \in C, u + v \in C$$

Exemple : $C = \{000, 100, 011, 111\}$ est un code linéaire de longueur 3 et de dimension 2, sous-espace linéaire de $(\mathbb{F}_2)^3$.

Le code C peut être défini au moyen d'une matrice binaire G à k lignes et n colonnes appelée *matrice génératrice* dont les lignes forment une base de C. G définit une application linéaire de $(\mathbb{F}_2)^k \rightarrow (\mathbb{F}_2)^n$ qui associe à tout mot binaire de longueur k (le *message* vu comme un vecteur de $(\mathbb{F}_2)^k$), un mot binaire de longueur n (le *mot du code* vu comme un vecteur de $(\mathbb{F}_2)^n$). On produit ainsi à partir d'un message de k lettres un nouveau message de n lettres (plus long) qui possède une certaine redondance.

Exemple :

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

définit un $(6, 3)$ code linéaire.
A partir des vecteurs de $(\mathbb{F}_2)^k$ et de la matrice G, on peut énumérer les éléments de C :

$$C = \{a.G : a \in (\mathbb{F}_2)^k\}$$

On dira que la matrice génératrice G est sous *forme normale* si G est de la forme $(\mathrm{Id}_k \quad P)$ où Id_k représente la $k \times k$ matrice identité. Dans ce cas, les k premiers symboles d'un mot de C sont appelés les *symboles d'information* et les $n - k$ autres les *symboles de redondance*.
Il est toujours possible de mettre sous forme normale une matrice génératrice :

Théorème 1 (admis) *Deux $k \times n$ matrices G et G' engendrent des (n, k)-codes linéaires* équivalents *si on peut obtenir G à partir G' par une suite d'opérations à choisir parmi :*
– *permutation des lignes ;*
– *addition de deux lignes ;*
– *permutation des colonnes .*

Question 2 :

1. mettre sous forme normale la matrice G qui engendre le $(6, 3)$-code linéaire de l'exemple précédent ;
2. énumérer les mots du code engendrés par une matrice génératrice sous forme normale.

1.3 Distance minimale

Une notion importante est celle de *distance minimale*. Par définition, elle mesure la plus petite distance entre deux mots du code distincts.

$$d(C) = \min_{c,c' \in C, c \neq c'} \{d_H(c, c')\}$$

Question 3 : Montrer que la distance minimale peut être définie au moyen du poids :

$$d(C) = \min_{c \in C, c \neq 0} \{\mathbf{w}(c)\}$$

Question 4 : Quelle est la distance minimale du (6,3) code linéaire précédent ?
La distance minimale est le principal paramètre pour déterminer la capacité de correction du code. En effet, si on envoie $x = x_1 x_2 \dots x_n$, un mot d'un code binaire de longueur n, au travers d'un canal de communication qui peut modifier le mot sur au plus t positions, le destinataire recevra un mot binaire $y = y_1 y_2 \dots y_n$ de longueur n qui peut différer de x sur au plus t positions. Le *motif d'erreur* $e = e_1 \dots e_n$ est le vecteur de dimension n qui satisfait :

$$y = x + e$$

Le *nombre d'erreurs* est défini comme étant le poids de e, $\mathbf{w}(e)$.

Question 5 : Si on note $\mathcal{E}_t$ l'ensemble de tous les motifs d'erreur de longueur n et de poids inférieur ou égal à t, montrer qu'un code binaire C peut permettre de corriger tous les éléments de $\mathcal{E}_t$ si et seulement si $d(C) > 2t$.

Question 6 : Pour C un code linéaire binaire de longueur n et pour une valeur de t fixée, combien $\mathcal{E}_t$, possède t'il d'éléments ? En déduire une borne sur le nombre de mots d'un (n, k)-code linéaire binaire C qui corrige toutes les erreurs de poids au plus t.

1.4 Codes duaux

Les *codes duaux* vont jouer un rôle important dans le décodage d'un message. On définit $C^\perp$ le code dual de C par $C^\perp = \{y \in (\mathbb{F}_2)^n : \forall x \in C, \langle x, y\rangle = 0\}$ où $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ est le produit scalaire. $C^\perp$ est un $(n, n-k)$-code linéaire. Si C et $C^\perp$ sont équivalents au sens du théorème 1, le code C sera dit *autodual*.
Si G est une matrice génératrice sous forme normale de C, on en déduit une matrice génératrice H du code dual $C^\perp$ par $H = ({}^tP \ \ \mathrm{Id}_{n-k})$ où ${}^tP$ dénote la transposée de la matrice P.
La matrice H permet de définir le code de manière alternative $C = \{x : xH = 0\}$.

Question 7 : Construire le code dual du $(6, 3)$ code linéaire de la question 2.

2 Décodage des codes linéaires

Dans la suite, C désignera toujours un (n, k)-code linéaire binaire.
Supposons que le mot du code x est émis au travers d'un canal de communication qui peut ajouter des erreurs. Le mot reçu est $y = y_1 y_2 \dots y_n$ et peut différer de x. On rappelle que le *motif d'erreur* e est $e = y - x = e_1 e_2 \dots e_n$.
Étant donné y, le décodeur doit décider quel mot du code x a été transmis, ou de manière équivalente, quel est le motif d'erreur. Les codes linéaires fournissent une solution élégante au problème du décodage au plus proche voisin –i.e. de manière à minimiser le poids de l'erreur–, au moyen des classes latérales.
Pour un code C et un vecteur $u \in (\mathbb{F}_2)^n$ on appelle *classe latérale* de C l'ensemble $u + C$ défini par :

$$u + C = \{u + x : x \in C\}$$

Question 8 : Montrer que, étant donnée $u + C$ une classe latérale de C, si $v \in u + C$, alors $v + C = u + C$.

Question 9 : Montrer que si C est un (n, k)-code linéaire binaire, alors

1. tout vecteur de $(\mathbb{F}_2)^n$ est dans une classe latérale de C ;
2. chaque classe latérale contient exactement 2^k vecteurs ;
3. étant données deux classes latérales, elles sont soit disjointes soit identiques.

On partitionne alors $(\mathbb{F}_2)^n = (\underline{0} + C) \cup (u_1 + C) \cup \ldots \cup (u_s + C)$ où $s = 2^{n-k} - 1$ et où les $\underline{0}, u_1, \ldots, u_s$ sont des éléments de poids minimum appelés *chefs de classe*.
On peut alors construire le *tableau standard* de C à 2^{n-k} lignes et 2^k colonnes. Il contient tous les vecteurs de $(\mathbb{F}_2)^n$. Sa première ligne correspond aux mots de C avec le vecteur 0 à gauche ; les autres lignes représentent les classes latérales $u_i + C$ avec leur chef de classe à gauche.
Les motifs d'erreurs qui pourront être corrigés sont précisément les chefs de classe, quel que soit le mot du code transmis. En choisissant des motifs d'erreur de poids minimum en tant que chefs de classe, le tableau standard assure un décodage au plus proche voisin.

Question 10 : Donner un algorithme qui permette de construire le tableau standard.

Le décodeur va utiliser le tableau standard de la façon suivante : lorsque le mot y est reçu, on recherche sa position dans le tableau standard. Le décodeur décide alors que le motif d'erreur e correspond au chef de classe qui est situé dans la première colonne de la même ligne et peut décoder y comme $x = y - e$ en choisissant le mot du code de la première ligne sur la même colonne que y.

Question 11 : Quelle est la complexité de cet algorithme de décodage ?

2.1 Application

On applique le décodage par les classes latérales au $(6, 3)$-code linéaire de la question 2.

Question 12 : Construire le tableau standard de ce code.

Question 13 : Décodez le message 110011.

3 La NP -complétude du décodage linéaire

3.1 Le problème du décodage des codes linéaires

Si le mot reçu y a subi des altérations, le *syndrome* $s = yH$ du mot reçu est non nul et le mot du code x lui correspondant vérifie $x = y + e$ où e est la solution de poids minimum (au sens de la définition de la section 1.1) de l'équation $s = zH$. Pour trouver cette solution de poids minimum, il semble nécessaire de rechercher les solutions parmi les 2^k solutions possibles. C'est un problème de *minimisation*.

3.2 Le problème du poids des classes latérales

On s'intéresse tout d'abord au problème de décision suivant :
POIDS DES CLASSES LATÉRALES
Instance : A une matrice binaire, y un vecteur binaire, $w \in \mathbb{N}$

Question : Existe-t-il un vecteur binaire x tel que le poids de x, $\mathbf{w}(x) \leq w$ et $xA = y$?

Question 14 : Donnez une relation entre le problème de minimisation du décodage des codes linéaires et le problème de décision du poids des classes latérales.

Nous allons montrer la NP -complétude du problème du poids des classes latérales.

Question 15 Montrer que le problème du poids des classes latérales est dans NP.

La NP-difficulté du problème du poids des classes latérales est obtenue par réduction de 3DM (réputé NP-complet par réduction de 3-SAT) dont on rappelle l'énoncé :

COUPLAGE À TROIS (3DM)
Instance : Un ensemble fini T et un sous-ensemble $U \subseteq T \times T \times T$.

Question : Existe-t'il $W \subseteq U$ tel que $|W| = |T|$ et dont toutes les coordonnées sont 2 à 2 distinctes ?

3DM est une généralisation du problème de couplage «classique» : étant donnés n hommes et n femmes et une liste de couples hommes/femmes qui souhaitent se marier, peut-on trouver n mariages en évitant la polygamie et la polyandrie et que chacun ait un époux convenable ?
De manière similaire, 3DM est la donnée de trois sexes distincts et chaque triplet correspond à un mariage à trois possible. On cherche à satisfaire l'ensemble des participants en évitant les alliances multiples.
Voici un exemple de couplage à trois. soit $T = \{1, 2, 3, 4\}$ et $|U| = 6$ défini comme suit :

$1)$	$(1, 2, 1)$	$(1, 2, 1)$
$2)$	$(1, 3, 2)$	$(1, 3, 2)$
$3)$	$(2, 1, 4)$	$(2, 1, 4)$
$4)$	$(2, 2, 3)$	$(2, 2, 3)$
$5)$	$(3, 1, 1)$	$(3, 2, 1)$
$6)$	$(4, 4, 4)$	$(4, 4, 4)$

Le premier ensemble de triplets (colonne de gauche) vérifie la propriété du couplage à trois pour les lignes $2), 4), 5)$ et $6)$ mais pas le second.

Question 16 : Montrer la NP-difficulté du problème du poids des classes latérales par réduction de 3DM et en déduire la NP -completude du problème du poids des classes.
Indication : Il est suggéré de coder U, l'ensemble de triplets sous la forme d'une $|U| \times 3|T|$ matrice d'incidence binaire

Question 17 : Trouvez un $(6, 3)$-code linéaire équivalent à celui de la question 2 qui code une instance de 3DM. Décrivez cette instance et dites si elle a une solution.