Codes Correcteurs d'Erreurs
Cours 2 :
+ Codage-Decodage de codes linéaires en blocs
+ Borne de Hamming (codes parfaits)

Marc Chaumont

November 12, 2008

Plan

1. Codage - décodage de code linéaire en bloc
   - Tableau standard de décodage
   - Exercice : réflexion et analyse
2. Les codes parfaits
   - Inégalité de Hamming

Tableau standard

Proposition de procédure de décodage pour trouver le mot de code c le plus proche d'un mot y réçut bruité $y = c + e$.

Vecteur d'erreur

Le vecteur d'erreur $e \in \{0,1\}^n$ est produit par un BSC (canal binaire symétrique). Le BSC est un modèle des erreurs survenant sur un canal. On fait donc une hypothèse particulière en choisisant un BSC. On supposera que la probabilité d'erreur p est tel que $p < 1/2$.

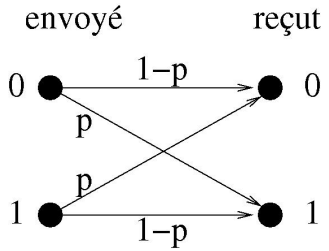

Figure: Canal binaire symétrique

Tableau standard

Tableau standard

Un tableau standard pour un code linéaire $[n, k, d_{min}]$ est une table listant tous les vecteurs y pouvant être reçu (après transmission d'un mot de code). Le tableau est construit pour pouvoir lire pour chaque vecteur y le plus proche mot de code c.

Tableau standard

Soit c_i, un des 2^k mots de code du code $[n, k, d_{min}]$.
Soit e_i, un des $2^{n-k} - 1$ **vecteurs erreur**.

erreur	*mot de code*			
	$c_0 = \underline{0}$	c_1	...	c_{2^k-1}
$e_0 = \underline{0}$	$c_0 + e_0 = \underline{0}$	$c_1 + e_0 = c_1$	...	$c_{2^k-1} + e_0 = c_{2^k-1}$
e_1	$c_0 + e_1 = e_1$	$c_1 + e_1$	...	$c_{2^k-1} + e_1$
...	...	...	...	...
$e_{2^{n-k}-1}$	$c_0 + e_{2^{n-k}-1} = e_1$	$c_1 + e_{2^{n-k}-1}$	...	$c_{2^k-1} + e_{2^{n-k}-1}$

Syndrome

Syndrome

Pour un code C, le syndrome d'un mot $y \in \mathbb{F}_2^n$ (vecteur reçu après transmission) est :

$$s = y.H^t$$

avec H la matrice de vérification de parité (matrice de contrôle) du code C.

Si un mot de code $c \in C$ est transmis sur un canal BSC et que le vecteur $y = c + e$ est reçu, le syndrome est :

$$s = yH^t = (c + e)H^t = eH^t$$

car $cH^t = 0$.

Tableau standard

		bits d'information			
		u_0	u_1	$\dots$	u_{2^k-1}
syndrome s_i	*erreur*	*mot de code*			
		$c_0 = \underline{0}$	c_1	$\dots$	c_{2^k-1}
$\underline{0}$	$e_0 = \underline{0}$	$c_0 + e_0 = \underline{0}$	$c_1 + e_0 = c_1$	$\dots$	$c_{2^k-1} + e_0 = c_{2^k-1}$
s_1	e_1	$c_0 + e_1 = e_1$	$c_1 + e_1$	$\dots$	$c_{2^k-1} + e_1$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$s_{2^{n-k}-1}$	$e_{2^{n-k}-1}$	$c_0 + e_{2^{n-k}-1} = e_1$	$c_1 + e_{2^{n-k}-1}$	$\dots$	$c_{2^k-1} + e_{2^{n-k}-1}$

Procédure de construction automatique du tableau standard

- remplir la ligne listant l'ensemble des mots d'information et caculer la ligne des mots de code associés (via la matrice G).
- remplir la première ligne correspondant à une erreur nulle (recopie de la ligne précédente).
- génération des lignes :
  - déterminer un vecteur erreur de poids le plus petit $e_i \in \mathbb{F}_2^n$ et n'appartenant pas aux lignes précédentes.
  - en déduire le syndrome associé $s_i = e_i H^t$
  - remplir la ligne pour ce vecteur erreur (première ligne + erreur e_i)
- répéter jusqu'à avoir rempli 2^{n-k} lignes.

Illustration du tableau standard pour le code linéaire [4,2,2]

Rappel : La sous-matrice de vérification de parité du code linéaire systématique [4,2,2] vaut :

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

syndrome	*bits d'information*			
	00	01	10	11
00	...	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Exercice : REMPLIR LE TABLEAU ...

Illustration du tableau standard pour le code linéaire [4,2,2]

CORRECTION

Illustration du tableau standard pour le code linéaire [4,2,2]

CORRECTION

Quelque définitions et remarques

Coset (classe) de C

Chaque ligne est ce que l'on appelle un **coset** (classe) et correspond à l'ensemble des mots de code ayant subi la même erreur : $Row_i = \{c + e_i | c \in C\}$
Un coset (classe) est également l'ensemble des mots ayant le même syndrome.

chef de coset (chef de classe)

Le leader du coset est le vecteur e_i.

Procédure de décodage

Soit $c_i \in C$ le mot de code émis et $y_i = c_i + e_i$ le mot reçu.

- calculer le syndrome $s_i = y_i H^t$,
- chercher dans le tableau standard le syndrome s_i et lire l'erreur e_i associée dans la colonne suivante,
- calculer le mot de code corrigé $\hat{c} = y_i - e_i = y_i + e_i$.

Occupation mémoire du tableau

Il n'y a besoin que de deux colonnes :

- la colonne du syndrome de taille 2^{n-k} cases de $n-k$ bits.
- la colonne des erreurs de taille 2^{n-k} cases de n bits.

Exemple de correction du code linéaire [4,2,2]

Supposons que le mot de code $c = (0110)$ soit transmis et que le mot reçu soit $y = (0010)$. Le syndrome est :

$$s = yH^t = (0010)\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (10).$$

À partir du tableau standard, on trouve que le leader du coset (l'erreur) est $e = (0100)$ et nous estimons alors que le mot de code transmis est $\hat{c} = y + e = (0010) + (0100) = (0110)$.
$d_{min} = 2$, donc la capacité de correction est de 0 erreurs...
Ce code est appelé code linéaire à protection inégale.

Résumé d'un décodeur à décision dure pour les codes linéaires

- $c \in C$ un code émis,
- $y = c + e$ un code bruité reçu,
- calcul du syndrome $s = yH^t$,
- À partir du syndrome, estimation du vecteur erreur e,
- Soustraction (ou addition modulo 2 dans le cas binaire) de l'erreur au mot reçu. Le mot corrigé est alors dans le cas binaire $\hat{c} = y + e$.

Plan

1. Codage - décodage de code linéaire en bloc
   - Tableau standard de décodage
   - Exercice : réflexion et analyse
2. Les codes parfaits
   - Inégalité de Hamming

Exercice : Réflexion sur les codes

Créer un code [5,3,.] linéaire ?

- Pour cela, on donnera la matrice G telle que :
  - la matrice G soit sous forme systématique.
- Donner ensuite sa distance minimum, sa capacité de détection et de correction.
- Le mot y = (00001) est-il un mot de code ?
- Dans le cas ou 1 seul bit a été erroné lors de la transmission, quels peuvent être les mots de code qui ont mené au mot y ?
- Enfin donner un tableau de décodage (protection inégale).

Correction

Correction

Correction

Résumé

On a construit un code à bon rendement ($3/5 = 0.6$) mais il a une distance d_{min} insuffisante ($d_{min} = 2$); il a donc un pouvoir détecteur d'erreur nul.
Pour construire un code linéaire en bloc, on fait appel à l'algèbre linéaire.
Regardons maintenant les codes linéaires en blocs parfaits qui ont la propriété de couvrir équitablement l'espace $\mathbb{F}^n$ et permettent facilement de déterminer des codes à distance minimale ≥ 3.

Plan

1. Codage - décodage de code linéaire en bloc
   - Tableau standard de décodage
   - Exercice : réflexion et analyse
2. Les codes parfaits
   - Inégalité de Hamming

Région de décodage

Région de décodage

Une colonne du tableau standard contient un mot de code c_i et $2^{n-k}-1$ mots plus proche de c_i que de n'importe quel autre mot de code. Une colonne correspond à une **région de décodage** autour d'un mot de code c_i :

$$Col_j = \{c_i + e_j | e_j = l'ensemble\ des\ chefs\ de\ coset\}$$

Borne de Hamming

Pour un code binaire linéaire $C[n, k, d_{min}]$ et donc une capacité de correction du code $t = \lfloor (d_{min} - 1)/2 \rfloor$, la sphère de Hamming $S_t(c_j)$ est contenue dans la région de décodage Col_j et donc

$$S_t(c_j) \subseteq Col_j.$$

La taille (nombre de mots) d'une région de décodage Col_j est 2^{n-k}.

Borne de Hamming

- Le nombre de vecteurs erreur différents de taille n, de poids de Hamming 1 vaut la combinaison $C_n^1 = \binom{n}{1}$
- Le nombre de vecteurs erreur différents de taille n, de poids de Hamming 2 vaut $C_n^2 = \binom{n}{2}$
- ...
- Le nombre de vecteurs erreur différents de taille n, de poids de Hamming t vaut $C_n^t = \frac{n!}{t!(n-t)!} = \binom{n}{t}$

Nombre de sphères de rayon 0 jusqu'à t

Pour un code linéaire $[n, k, d_min]$, et donc une capacité de correction du code $t = \lfloor (d_{min} - 1)/2 \rfloor$, le nombre de vecteurs erreur différents de poids 0 jusqu'à t est :

$$\sum_{i=0}^{t} \binom{n}{i}$$

Borne de Hamming

Pour un code linéaire $[n, k, d_{min}]$, et donc une capacité de correction du code $t = \lfloor (d_{min} - 1)/2 \rfloor$, puisque l'ensemble des mots y résultant de la perturbation d'un mot de code c_i par des erreurs de poids inférieur ou égal à t est inclus dans la région de décodage Col_i de ce mot de code c_i, on a :

Borne de Hamming

$$\sum_{i=0}^{t} \binom{n}{i} \leq 2^{n-k}$$

Une interprétation

Le nombre de syndrome, 2^{n-k}, doit être plus grand ou égal au nombre d'erreurs corrigeables $\sum_{i=0}^{t} \binom{n}{i} \leq 2^{n-k}$

Borne de Hamming

Code parfait

Un code est dit **code parfait** lorsque $\sum_{i=0}^{t} \binom{n}{i} = 2^{n-k}$ c'est à dire lorsque les sphères de Hamming couvrent parfaitement l'espace $\mathbb{F}_2^n$.

Le code binaire à répétition [3, 1, 3] est un code parfait (cf. figure ou regarder le tableau standard ou faire le calcul).

Exercice

- Le code à répétition C[3,1, 3] est-il parfait ?
- Le code C[4,2,2] est-il parfait ?
- Le code C[7,4,3] est-il parfait ?

Correction

Remarque sur les codes parfaits

Les seuls codes parfaits **binaires** sont :

- Le code à répétition C[3,1, 3].
- Les codes de Hamming,
- Le code de Golay G_{23}.