Réseaux, information et communications (INFO-F303)
Partie Théorie de l'Information
7. Codes linéaires

Christophe Petit

Université libre de Bruxelles

Plan du cours

Motivation

- Second théorème de Shannon : il existe famille de codes de débit tendant vers l'optimal (capacité du canal) et taux d'erreur tendant vers zéro
- Résultat asymptotique ; nécessite de grands codes
- Pour les grands codes : problèmes d'efficacité pour
  - stocker les mots du code
  - vérifier ses propriétés (distance minimale, etc)
  - calculer mot de code le plus proche du message reçu
- Idée : choix de codes “structurés”

Codes linéaires

- But : familles de codes avec de bonnes propriétés (et une représentation compacte)
- Codes linéaires : définition et propriétés
- Spécialisation des bornes de Hamming et Singleton
- Matrice génératrice, syndrome et matrice de contrôle
- Forme canonique et matrice de parité
- Décodage par syndrome

Codes linéaires

- Espace des messages codés (mots du code et erronés) est l'espace vectoriel $\mathbb{F}_q^n$ de dimension n sur $\mathbb{F}_q$ ($\mathbb{F}_q =$ corps fini à $q = p^\nu$ éléments, où p est un nombre premier, caractéristique du corps)
- Espace des symboles originels est l'espace vectoriel $\mathbb{F}_q^k$ de dimension k sur $\mathbb{F}_q$
- La fonction de codage est une **application linéaire**
- Code est un **sous-espace vectoriel** $K \subset \mathbb{F}_q^n$ de dimension k

Matrice génératrice

- La fonction de codage est une application linéaire
- **Matrice génératrice** G, de n lignes et k colonnes

$$
\begin{aligned}
G(K) \colon \mathbb{F}_q^k &\to \mathbb{F}_q^n \\
\boxed{x} &\mapsto \boxed{y} = \boxed{G} \cdot \boxed{x}
\end{aligned}
$$

Exemple : code à répétition

$$K: C \to C^n : c \mapsto \overbrace{cc\ldots c}^{n}$$

.

- On a

$$G = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \in \mathbb{F}_q^{n \times 1}$$

Exemple : code à somme de contrôle

$$K\colon C^{n-1} \to C^n : c_1 \dots c_{n-1} \mapsto c_1 \dots c_{n-1}(\sum_i c_i)$$

.

- On a

$$G = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} I_{n-1} \\ \mathbf{1} \end{bmatrix} \in \mathbb{F}_q^{n\times(n-1)}$$

Exemple : code de Hamming binaire

- Famille de codes linéaires **parfaits** ($B_s = B_c = 2^{n-k}$) avec $n = 2^m - 1$, $k = 2^m - 1 - m$ et $d = 3$
- Un mot $c = c_1 \dots c_n \in \{0,1\}^n$ du code de Hamming binaire est tel que les bits x_i dont l'indice i est une puissance de deux sont des bits de contrôle, les autres sont des bits de données
- Le bit de contrôle d'indice c_i pour $i = 2^\ell$ est la somme modulo 2 de tous les bits de données x_j dont l'indice j écrit en base 2 a le $(\ell+1)^{\text{ème}}$ bit à 1

Exemple : code de Hamming binaire $(7, 4)$

- Pour $n = 7$, un mot $c = c_1 \dots c_7$ du code de Hamming est tel que
  - les bits c_1, c_2, c_4 sont des bits de contrôle
  - les bits c_3, c_5, c_6, c_7 sont des bits de données
- On a

$$\begin{aligned}
c_1 &= c_{110} + c_{101} + c_{111} = c_3 + c_5 + c_7 \mod 2 \\
c_2 &= c_{110} + c_{011} + c_{111} = c_3 + c_6 + c_7 \mod 2 \\
c_4 &= c_{101} + c_{011} + c_{111} = c_5 + c_6 + c_7 \mod 2
\end{aligned}$$

Code de Hamming est linéaire

- Matrice génératrice G dans $\mathbb{F}_2^{7\times 4}$:

$$c = y = Gx = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 + x_4 \\ x_1 + x_3 + x_4 \\ x_1 \\ x_2 + x_3 + x_4 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

Propriétés du code de Hamming

- On a $d = 3$: tout changement d'un bit de donnée $c_{\sum_{j=0}^{m-1} e_j 2^j}$ impacte tous les bits de contrôle c_{2^j} pour $e_j = 1$ (c'est-à-dire toujours au moins 2 bits de contrôle, et exactement 2 pour tout e de poids binaire 2)
- Le code est 1-correcteur. En effet, considérons la somme e des indices des bits de contrôle erronés. S'il n'y a qu'une seule erreur, elle ne peut provenir que du bit d'indice e

Code $[n, k, d]_q$

- Un code $[n, k, d]_q$ est un code linéaire de paramètres
  - n est la **longueur** du code, celle des mots encodés
  - k est la **dimension** du code, le nombre de caractères des symboles originels
  - d est la **distance minimale** du code
  - q est le **nombre primaire de caractères** ($q = p^\nu$ pour p premier)

Propriétés des codes linéaires

- Débit d'un code linéaire

$$R = \frac{k}{n}$$

- Distance minimale = poids minimal du code
- Borne de Singleton pour les codes linéaires

$$d \leq n - k + 1$$

- Borne de Hamming pour les codes linéaires

$$B_s \leq q^{n-k}$$

Matrice de contrôle

- Code linéaire défini par **matrice génératrice** $G \in \mathbb{F}_q^{n\times k}$

$$K = \mathrm{Im}\ G = \{Gx \mid x \in \mathbb{F}_q^k\}$$

$$K = \{\text{combinaisons linéaires des colonnes de } G\}$$

- Alternative : **matrice de contrôle** $H \in \mathbb{F}_q^{(n-k)\times n}$ telle que

$$K = \mathrm{Ker}\ H = \{y \in \mathbb{F}_q^n \mid Hy = 0\}$$

- On a

$$\boxed{\quad H \quad} \cdot \boxed{\begin{matrix} \\ G \\ \\ \end{matrix}} = \boxed{\ 0\ }$$

Exemple : code à répétition

$$K : C \to C^n : c \mapsto \overbrace{cc\dots c}^{n}$$

.

- On a
$$G = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^t \in \mathbb{F}_q^{n\times 1}$$
- On peut prendre
$$H = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ -1 & 0 & 0 & & 1 & 0 \\ -1 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \in \mathbb{F}_q^{(n-1)\times n}$$

Exemple : code à somme de contrôle

$$K: C^{n-1} \to C^n : c_1 \dots c_{n-1} \mapsto c_1 \dots c_{n-1} \left(\sum_i c_i\right)$$

.

- On a

$$G = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} I_{n-1} \\ \mathbf{1} \end{bmatrix} \in \mathbb{F}_q^{n \times (n-1)}$$

- On a

$$H = \begin{bmatrix} -1 & \dots & -1 & 1 \end{bmatrix} \in \mathbb{F}_q^{1 \times n}$$

Exemple : code de Hamming

- Matrice génératrice G dans $\mathbb{F}_2^{7\times 4}$:

$$c = y = Gx = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 + x_4 \\ x_1 + x_3 + x_4 \\ x_1 \\ x_2 + x_3 + x_4 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

- On peut prendre $H \in \mathbb{F}_2^{3\times 7}$ *sous forme itérative*

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(contient tous les vecteurs non nuls de $\mathbb{F}_2^3$, dans l'ordre)

Code de Hamming : démonstration $HG = 0$

- $H_{ij} = 1 \Leftrightarrow i$th bit de j vaut 1
- $H_i G$ est la somme des lignes de G d'indices j tels que le ith bit de j vaut 1
- Ces indices j incluent
  - un bit de contrôle ($j = 2^i$)
  - $2^{n-1} - 1$ bits de données
- Par construction, le bit de contrôle vaut la somme des bits de données, donc la somme est nulle

Matrice de contrôle et distance minimal

- Distance minimale = poids minimal = nombre minimal de colonnes dépendantes de H

Formes systématiques (ou canoniques)

- Matrices génératrice et de contrôle pas uniques
- Par combinaisons linéaires inversibles des colonnes (permutations, etc), on peut ramener G à la forme
$$G = \left[\begin{array}{c} I_k \\ \hline P \end{array}\right]$$
avec **matrice de parité** $P \in K^{(n-k)\times k}$
- Px est la **redondance**
$$y = \left[\begin{array}{c} I_k \\ \hline P \end{array}\right] x = \left[\begin{array}{c} x \\ \hline Px \end{array}\right]$$
- On peut aussi choisir H de la forme
$$H = \left[\begin{array}{c|c} -P & I_k \end{array}\right]$$

Forme systématique : codes à répétition et checksum

- Les matrices données ci-dessus sont déjà sous forme canonique

Forme systématique : Hamming binaire $(7,4)$

- Matrice génératrice canonique G dans $\mathbb{F}_2^{7\times 4}$

$$y = Gx = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_4 \\ x_2 + x_3 + x_4 \mod 2 \\ x_1 + x_3 + x_4 \mod 2 \\ x_1 + x_2 + x_4 \mod 2 \end{bmatrix}$$

- Matrice de contrôle canonique dans $\mathbb{F}_2^{3\times 7}$

$$H = \left[\begin{array}{cccc|ccc} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

Syndrome

- L'image d'un mot par H est appelée **syndrome** du mot

$$
\begin{array}{rcl}
\sigma : \mathbb{F}_q^n & \to & \mathbb{F}_q^{n-k} \\
\boxed{y} & \mapsto & \boxed{\sigma} = \sigma(y) := \boxed{\quad H \quad} \cdot \boxed{y}
\end{array}
$$

- On a $y \in K \iff \sigma(y) = 0$ (permet la détection d'erreur)

Syndrome (2)

- Si y encode x avec des erreurs $\varepsilon = y - Gx$, alors

$$\sigma(y) = Hy = H(Gx + \varepsilon) = HGx + H\varepsilon = H\varepsilon$$

  le syndrome ne dépend que de ces erreurs

- A chaque syndrome correspond au plus un vecteur dans la boule $\mathcal{B}(0, t)$ si $t < \frac{d}{2}$

Codage et décodage

- **Codage** : multiplication de la matrice génératrice par le mot $x \in \mathbb{F}_q^k$ à encoder

$$y = Gx$$

- **Décodage** : table comprenant (au plus) q^{n-k} entrées, qui à chaque syndrome $\sigma(y)$ associe l'unique erreur ε de poids au plus $s = \lfloor (d-1)/2 \rfloor$ telle que $\sigma(y) = H\varepsilon$

Efficacité du décodage

- Décodage par syndrome plus efficace que tester tous les mots si
$$q^k > q^{n-k} \quad \Leftrightarrow \quad k > n/2$$

- Stockage de la table de syndromes limite les paramètres (même si petites optimisations possibles)
- Solution : familles de codes correcteur avec un **algorithme** de décodage efficace

Exemple : code de Hamming binaire $(7, 4)$

- On a
$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$
- Supposons qu'on reçoive $y = [0\ 1\ 1\ 1\ 0\ 1\ 0]^T$
- On calcule $Hy = [1\ 1\ 0]^T$
- On déduit $Hy = H\varepsilon$ avec $\varepsilon = [0\ 0\ 1\ 0\ 0\ 0\ 0]^T$ (erreur sur le troisième bit de données)

Questions ?

?

Crédits et remerciements

- Mes transparents suivent fortement les notes de cours développées par le Professeur Yves Roggeman pour le cours INFO-F303 à l'Université libre de Bruxelles
- Une partie des transparents et des exercices ont été repris ou adaptés des transparents développés par le Professeur Jean Cardinal pour ce même cours
- Je remercie chaleureusement Yves et Jean pour la mise à disposition de ce matériel pédagogique, et de manière plus large pour toute l'aide apportée pour la reprise de ce cours
- Les typos et erreurs sont exclusivement miennes (merci de les signaler !)