Réseaux, information et communications (INFO-F303)
Partie Théorie de l'Information
8. Quelques familles de codes linéaires

Christophe Petit

Université libre de Bruxelles

Plan du cours

1. Notion de code
2. Source aléatoire et codes efficaces
3. Entropie et codage efficace
4. Compression sans perte
5. Canal bruité
6. Codes correcteurs d'erreurs
7. Codes linéaires
8. Quelques familles de codes linéaires

A. Rappels mathématiques (chapitre 7.1 du syllabus)

Théorie des codes

- Shannon : *"il existe un code qui transmet sans erreur sur un canal bruité, tant que le débit du code est inférieur à la capacité du canal"*
- Théorème d'existence, preuve non constructive
- En pratique, on souhaite des codes avec bons débit et capacité de correction, mais aussi des fonctions d'encodage et de décodage efficaces

Rappel : contraintes connues

- Second théorème de Shannon : $k/n \leq \mathcal{C}_p$
  (borne atteinte asymptotiquement)
- Borne de Singleton : $d \leq n - k + 1$
- Borne de Hamming : $B_s \leq q^{n-k}$ avec $s = \left\lfloor \frac{d-1}{2} \right\rfloor$
- Borne de Gilbert-Varshamov : $B_{d-1} \geq q^{n-k}$
  (pour les codes maximaux)

Codes approchant la capacité maximale

- Pour canal bruité donné, par exemple un canal symétrique de paramètres p et r, on veut un code $[n, k, d]$ avec
  - Grande capacité de détection et correction d'erreurs (asymptotiquement on veut $d \geq 2pn$)
  - Haut débit k/n (approchant la capacité du canal)
  - Algorithmes rapides de détection et correction

Codes parfaits

- Borne de Hamming : $B_s \leq q^{n-k}$ avec $s = \left\lfloor \frac{d-1}{2} \right\rfloor$
- Codes **parfaits** ssi **borne de Hamming atteinte**
- "Corrige toutes les erreurs détectables"
- Exemples : codes de Hamming, codes de Golay
- "Parfaits" ? Tous les codes de Hamming ont $d = 3$ (asymptotiquement, taux d'erreur est 1)

Codes MDS (maximal distance separable)

- Borne de Singleton : $d \leq n - k + 1$
- Codes **MDS** ssi **borne de Singleton atteinte**
- "Corrige le maximum d'erreurs pour redondance fixée"
- Exemple : codes de **Reed-Solomon** $[n, k, n - k + 1]_q$ avec $q \geq n$
  - Paramètres très généraux, mais contrainte $q \geq n$ augmente la probabilité d'erreur sur un symbole (car chaque symbole est codé sur au moins $\log q$ bits)
  - Souvent $q = n + 1$ et le code est aussi un **code BCH** (voir plus loin)

Transparents suivants

- Codes polynomiaux
- Codes cycliques
- Codes BCH
- Codes linéaires parfaits et codes de Hamming
- Brièvement
  - Codes de Reed-Muller
  - Codes de Reed-Solomon

Mots et polynômes

- Message à coder x de k caractères est vu comme le polynôme en la variable Z

$$X(Z) = \sum_{i=0}^{k-1} x_i Z^i$$

  de degré inférieur à k
- Mot à décoder y de n caractères est un polynôme $Y(Z) = \sum_{i=0}^{n-1} y_i Z^i$ de degré inférieur à n
- Syndrome σ est un polynôme $\Sigma(Z) = \sum_{i=0}^{n-k-1} \sigma_i Z^i$ de degré inférieur à $(n-k)$
- Arithmétique sur l'anneau de polynômes $F[Z]$

Codes polynomiaux

- Un **code polynomial** de dimension k et de longueur n sur un corps F est l'ensemble des multiples de degrés inférieurs à n d'un polynôme $G \in F[Z]$ de degré $(n-k)$
- Cas particulier des codes linéaires
- Polynôme $G(Z)$ est le **polynôme générateur** du code
- Pour le syndrome, on peut prendre un simple résidu

$$\Sigma(Z) = \sigma(Y) := Y(Z) \bmod G(Z)$$

On a $\sigma(Y' + Y) = \sigma(Y')$ pour tout mot du code Y
et $\sigma(Y) = 0 \Leftrightarrow Y \in K$

Exemple de code polynomial binaire

- Code linéaire sur $F = \mathbb{F}_2$, dimension $k = 2$, longueur $n = 5$, engendré par $G(Z) = Z^3 + Z + 1$ est

$$K = \left\{0,\ 1 + Z + Z^3,\ Z + Z^2 + Z^4,\ 1 + Z^2 + Z^3 + Z^4\right\}$$

- Matrice génératrice : colonnes sont les images des 2 vecteurs d'une base de $F[Z]_{\deg<2}$, par exemple $\left\{Z, 1\right\}$

$$G = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{array}\right]$$

Codes polynomiaux : forme canonique

- On peut choisir comme fonction d'encodage

$$K : X(Z) \mapsto Z^{n-k}\ X(Z) - \left(\left(Z^{n-k}\ X(Z)\right) \bmod G(Z)\right)$$

- Bits de redondance correspondent au polynôme

$$P_X(Z) = -\left(Z^{n-k}\ X(Z)\right) \bmod G(Z)$$

de degré inférieur à $(n-k)$

- G au transparent précédent est sous cette forme

Codes cycliques (CRC)

- Le **décalage circulaire** d'une position d'un mot

$$y = (y_0, y_1, \ldots, y_{n-1})$$

est le mot

$$y' = (y_{n-1}, y_0, y_1, \ldots, y_{n-2})$$

- Un code linéaire est **cyclique** si et seulement s'il contient tous les décalages circulaires de ses mots
- Dans une représentation polynomiale,

$$\begin{aligned} Y'(Z) &= Z \cdot Y(Z) - y_{n-1}Z^n + y_{n-1} \\ &= Z \cdot Y(Z) - y_{n-1}(Z^n - 1) \\ &= Z \cdot Y(Z) \bmod (Z^n - 1) \end{aligned}$$

Codes cycliques : polynôme générateur

- **Soit $\mathbf{G(Z)}$** mot de **degré minimal** d'un code **cyclique**. Alors **$\mathbf{G(Z})$ divise $\mathbf{Z^n - 1}$**.
- Pour tout $A(Z)$, le polynôme

$$R(Z) = A(Z)G(Z) \bmod (Z^n - 1)$$

  est un mot du code
- Supposons $G \nmid (Z^n - 1)$. Alors il existe A, R tels que

$$R(Z) = Z^n - 1 - A(Z)G(Z), \qquad \deg R < \deg G$$

- Donc R est dans le code et $\deg R < \deg G$: contradiction

Codes cycliques : polynôme générateur

- Tout code cyclique a un **polynôme générateur** $G(Z) = \sum_{i=0}^{n-k} g_i Z^i$ qui est un **diviseur de $\mathbf{Z^n - 1}$**
- Tout diviseur de $Z^n - 1$ engendre un code cyclique
- Matrice génératrice

$$G = \begin{bmatrix} g_{n-k} & 0 & & \cdots & 0 \\ g_{n-k-1} & g_{n-k} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ g_0 & \cdots & & \ddots & 0 \\ 0 & \ddots & & & g_{n-k} \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & g_0 & g_1 \\ 0 & \cdots & & 0 & g_0 \end{bmatrix}$$

Codes cyclique : polynôme de contrôle

- Polynôme de contrôle

$$H(Z) = \frac{Z^n - 1}{G(Z)}$$

- On a $P(Z) \in K \Leftrightarrow P(Z)H(Z) = 0 \bmod (Z^n - 1)$
- Matrice de contrôle associée est

$$H = \begin{bmatrix} h_k & h_{k-1} & \cdots & h_0 & 0 & \cdots & 0 \\ 0 & h_k & \ddots & & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \cdots & h_k & \cdots & h_1 & h_0 \end{bmatrix}$$

Codes BCH

- BCH = Bose-Chaudhury-Hocquenghem
- Codes polynomiaux cycliques $[q^\mu - 1, q^\mu - 1 - \rho, d]_q$ avec $\rho \leq \mu(\delta - 1)$ et $d \geq \delta$
- Contrôle de la distance minimale
- Codes MDS (Maximal Distance Separable) si $\mu = 1$: codes $[q - 1, q - d, d]_q$
- Utilisés dans les communications satellites, CD, DVD, etc

Codes BCH : préliminaires

- Soit q une puissance de premier et μ un entier
- Soit $\mathbb{F}_{q^\mu}$ corps fini à q^μ éléments et soit $n := q^\mu - 1$
- Le groupe multiplicatif associé $\mathbb{F}_{q^\mu}^*$ est un groupe cyclique à n éléments, toutes les racines $n$èmes de l'unité
- Soit α un générateur de $\mathbb{F}_{q^\mu}^*$. On a

$$Z^n - 1 = \prod_{i=1}^{n} (Z - \alpha^i)$$

- La plupart des α^i sont définies sur $\mathbb{F}_{q^\mu}$, mais certains sont définis dans le sous-corps $\mathbb{F}_q$ (par exemple $1 = \alpha^n$)

Codes BCH : préliminaires

- Soit $I_i = \{\alpha^i, \alpha^{iq}, \alpha^{iq^2}, \ldots\}$
  (au plus μ éléments car $\alpha^{q^\mu} = \alpha$)
- Tous les $\alpha \in I_i$ ont le même polynôme minimal sur $\mathbb{F}_q$ égal à

$$M_i(Z) = \prod_{\alpha \in I_i} (Z - \alpha)$$

  (polynôme invariant par conjugaison de Galois i.e. si coefficients élevés à la puissance q)
- Polynôme $Z^n - 1$ ne se factorise pas nécessairement en facteurs linéaires sur $\mathbb{F}_q$; c'est le produit de tous les polynômes minimaux $M_i(Z)$

Exemple

- Soit $q = 3$ et $\mu = 2$ et $n = q^\mu - 1 = 8$
- Soit $\alpha \in \mathbb{F}_9^*$ générateur (donc $\alpha^4 = -1$)
- On a $I_0 = \{1\}$, $I_1 = I_3$, $I_2 = I_6$, $I_4 = \{-1\}$, $I_5 = I_7$
- Sur $\mathbb{F}_3$ on a la factorisation irréductible

$$Z^8 - 1 = (Z-1)(Z+1)(Z^2+1)(Z^2-Z-1)(Z^2+Z-1)$$

  - Les deux premiers facteurs correspondent aux racines $+1$ et -1 déjà dans le corps
  - Le suivant à une extension par une racine 4^e de 1

$$\iota^2 + 1 = 0 \Leftrightarrow \iota^4 = 1 \text{ et } \iota^2 = -1 \neq 1$$

  - Les deux derniers correspondent aux racines 8^e primitives regroupées deux par deux en racines *conjuguées*

Codes BCH : construction

- Soit q premier et μ entier
- Soit $n = q^\mu - 1$
- Soit α un générateur de $\mathbb{F}_{q^\mu}^*$
- Soit $\delta \geq 1$ **borne inférieure sur distance minimale**
- Code BCH est le code cyclique généré par

$$G(Z) = \text{lcm}\left\{ M_i(Z) \,\middle|\, i = 1 \ldots (\delta - 1) \right\}$$

- On a $k = n - \deg G$ et $\deg G \leq \mu(\delta - 1)$
- On va prouver $d \geq \delta$

Codes BCH : distance minimale

- Soit $Y(Z) = \sum_{i=0}^{n-1} y_i Z^i \in K$ avec $w = w_H(Y) < \delta$
- Soit $\{i_1, \ldots, i_w\} = \{i \mid y_i \neq 0\}$
- On a $Y(Z) = A(Z)G(Z) + B(Z)(Z^n - 1)$ donc $Y(\alpha^j) = 0$ pour $j = 1, \ldots, \delta - 1$ donc

$$\begin{pmatrix} \alpha^{i_1} & \alpha^{i_2} & \cdots & \alpha^{i_w} \\ \alpha^{2i_1} & \alpha^{2i_2} & \cdots & \alpha^{2i_w} \\ \vdots & \vdots & & \vdots \\ \alpha^{wi_1} & \alpha^{wi_2} & \cdots & \alpha^{wi_w} \end{pmatrix} \begin{pmatrix} y_{i_1} \\ y_{i_2} \\ \vdots \\ y_{i_w} \end{pmatrix} = 0$$

- Déterminant vaut $\prod_{j=1}^{w} \alpha^{i_j} \prod_{1 \leq j < k \leq w} (\alpha^{i_j} - \alpha^{i_k})$
- Déterminant nul ssi deux racines égales (contradiction)

Codes BCH : cas $\mu = 1$

- $n = q - 1$
- α un générateur de $\mathbb{F}_q^*$
- $Z^n - 1 = \prod_{i=1}^{n}(Z - \alpha^i)$ sur $\mathbb{F}_q$
- $G(Z) = \prod_{j=1}^{\delta-1}(Z - \alpha^j)$
- deg $G = \delta - 1$ donc $k = n - \delta + 1$
- $\delta \leq d \leq n - k + 1 = \deg G + 1 = \delta$ donc $d = \delta$
- Code MDS $[q - 1, q - d, d]_q$

Codes BCH : décodage

- Equation clé

$$\Lambda_P(Z) \cdot S_y(Z) + \Xi_y(Z) \cdot Z^{\delta-1} = \Omega_y(Z)$$

avec

  - $S_y(Z) = \sum_{i=1}^{\delta-1} Y(\alpha^i) Z^{i-1}$ syndrome dans l'extension
  - $\Lambda_P(Z) = \prod_{j=1}^{t} (1 - \alpha^{p_j}.Z)$ polynôme de localisation des erreurs (en position p_j)
  - $\Omega_y(Z)$ polynôme d'évaluation des erreurs

- Algorithme
  - Calcul de S_y à partir de Y
  - Λ_P, $\Xi_y(Z)$ et Ω_y par l'algorithme d'Euclide étendu
  - Positions des erreurs via les racines de Λ_P
  - Valeur des erreurs via $\Omega_y(Z)$

Codes parfaits

- Rayons d'empilement et de recouvrement égaux

$$t = c = s \quad \text{et} \quad d = 2t + 1$$

- Borne de Hamming atteinte

$$|B_s| = |B_t| = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i = q^{n-k}$$

Codes parfaits

$$|B_s| = |B_t| = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i = q^{n-k}$$

- $t = 0$: on a $d = 1$ et $K = \mathbb{F}_q^n$ trivial
- $t = n$: on a $K = \{0\}$ trivial
- $t = 1$: on a $d = 3$ et $n = \frac{q^{n-k}-1}{q-1}$ codes de Hamming
- Code à répétitions sur $\mathbb{F}_2$ avec $n = 2t + 1$
- Codes de Golay
- Exemples non linéaires avec $t = 1$

Codes de Hamming

- 1950 : code de Hamming binaire $(7, 4)$: encodage demi-octet avec taux d'erreur élevé
- Codes de Hamming : famille de codes parfaits avec $t = 1$, $d = 3$ et $n = \frac{q^{n-k}-1}{q-1}$
- Codes dérivés :
  - Codes de Hamming étendu : bit additionel de parité (donc $d = 4$; corrige une erreur, mais en détecte 2)
  - Code simplexe : **dual** du code de Hamming ($G \leftrightarrow H$)

Code de Hamming binaire

- Famille de codes parfaits avec $n = 2^m - 1$, $k = 2^m - 1 - m$, $d = 3$
- Un mot $c = c_1 \dots c_n \in \{0,1\}^n$ du code de Hamming binaire est tel que les bits c_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 c_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

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

Exemple : code de Hamming binaire

- 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 $e_j = 1$
- 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

Codes de Reed-Muller

- Codes linéaires avec décodage rapide par majorité
- Codes $[2^{\mu},\ \sum_{i=0}^{\delta} \binom{\mu}{i},\ 2^{\mu-\delta}]_2$
- Mots du codes $\sim$ fonctions de μ variables et degré $\leq \delta$, évaluées en chaque valeur possible des μ variables
- Utilisés pour communications satellites dans les années 60
- Cfr syllabus d'Yves Roggeman

Codes de Reed-Solomon

- Codes linéaires $[n, k, n-k+1]_q$
- Codes MDS (borne de Singleton atteinte)
- Mots du code : évaluation d'un polynôme de degré $< k$ sur $\mathbb{F}_q$ en n valeurs distinctes
- Intuition : polynôme déterminé par évaluation en k points
- Codes BCH si $q = n + 1$
- Décodage par l'algorithme Peterson–Gorenstein–Zierler, ou techniques semblables aux codes BCH

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 !)