Codage linéaire

Un codage linéaire est un application linéaire g

$$g : \begin{array}{ccc} \mathcal{B}^k & \rightarrow & \mathcal{B}^n \\ m & \rightarrow & g(m) \end{array}$$

Le code $\mathcal{C} = Im(g)$ est un sous espace vectoriel de dimension k de $\mathcal{B}^n$

Le code est injectif : $Ker(g) = \{\vec{0}\}$

- $g(\vec{o}) = \vec{o}$
  Le mot nul est un mot de code
- $(\forall m \in \mathcal{B}^k)\,(\forall m' \in \mathcal{B}^k)\, g\,(m \oplus m') = g(m) \oplus g(m')$
  La somme de deux mots de codes est un mot de code

Distance d'un code linéaire

étant donne C un mot du code linéaire $\mathcal{C}$, considérons la translation t_C qui a tout mot de code c_1 associe $c_2 = c_1 \oplus C$

$$\boxed{t_C(\mathcal{C}) = \mathcal{C}}$$

- $t_C(\mathcal{C}) \subset \mathcal{C}$ puisque la somme de deux mots de codes est un mot de code
- $\mathcal{C} \subset t_C(\mathcal{C})$ en effet $(\forall c \in \mathcal{C})(c \oplus C) \oplus C = c$. Comme c et C sont deux mots de codes $c \oplus C$ est aussi un mot de code et $t_c(c \oplus C) = c$

la distance d'un code linéaire est égale au poids du mot de code de plus faible poids

cela provient du fait que d_H est invariante par translation.
soit c_1 et c_2 deux mots de codes tels que la distance du code $d = d_H(c_1, c_2)$ alors

$$\begin{aligned} d &= d_H(c_1, c_2) \\ &= w(c_1 \oplus c_2) \\ &= w((c_1 \oplus c_1) \oplus (c_2 \oplus c_1)) \\ &= d_H(c_1 \oplus c_1, c_2 \oplus c_1) \\ &= d_H(\vec{0}, c_2 \oplus c_1) \end{aligned}$$

Il existe donc un mot de code ($c = c_2 \oplus c_1$) tel que $d = d_H(\vec{0}, c)$

Matrice génératrice d'un code linéaire

soit $\mathcal{E}_k$ et $\mathcal{E}_n$ des bases respectivement de $\mathcal{B}^k$ et $\mathcal{B}^n$
soit m un mots de $\mathcal{B}^k$ et $[m]_{\mathcal{E}_k}$ sa matrice (horizontale) des coordonnées de m la base $\mathcal{E}_k$

la matrice génératrice du code linéaire g est donnée par

$$G = \begin{pmatrix} [g(e_1)]_{\mathcal{E}_n} \\ [g(e_2)]_{\mathcal{E}_n} \\ \dots \\ [g(e_k)]_{\mathcal{E}_n} \end{pmatrix}$$

Si on note $C_m = g(m)$ alors

$$[C_m]_{\mathcal{E}_n} = [m]_{\mathcal{E}_k} G$$

Matrice génératrice d'un code linéaire : forme systématique

$$g : m_1 m_2 \cdots m_k \longrightarrow C_m = \underbrace{m_1 m_2 \cdots m_k}_{k \text{ bits informatifs}} \mid \underbrace{c_1 c_2 \cdots c_{n-k}}_{n-k \text{ bits de contrôle}}$$

Relativement aux base canoniques la matrice G peut se mettre sous la forme

la matrice génératrice du code linéaire g est donnée par
$$G = \begin{pmatrix} I_{kxk} & P_{kx(n-k)} \end{pmatrix}$$

$m \in \mathcal{B}^k$ est codé par
$$C_m = mG$$
$$C_m = m|mP$$

exemple :$k = 3$, $n = 6$

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

les mots de code					
bits informatifs			bits de contrôle		
0	0	0	0	0	0
0	0	1	1	1	0
0	1	0	1	0	1
0	1	1	0	1	1
1	0	0	0	1	1
1	0	1	1	0	1
1	1	0	1	1	0
1	1	1	0	0	0

Matrice génératrice d'un code linéaire : forme systématique

Exercice :

1. si on considère $k = 3$ autrement si les message en entrée du codage canal sont de longueur, quelle doit être la longueur minimale des mots de code pour que le code corrige une erreur de façon certaine.
2. soit le code dont la matrice génératrice est

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

   1. quelle est la distance du code, combien d'erreur peut-on être sûr de détecter, de corriger ?
   2. donner la matrice P de parité
   3. comment est codé le message $m = 101$? identifier les parties informative et de parité

Dual d'un code linéaire

Notation : soit $x = x_1 x_2 \cdots x_n$ et $y = y_1 y_2 \cdots y_n$ deux vecteurs de $\mathcal{B}^n$

$$< x, y >= x_1 y_1 \oplus x_2 y_2 \oplus \cdots \oplus x_n y_n$$

Soit $\mathcal{C}$ un $[n, k]$-code,
$\mathcal{C}^{\perp} = \{x \in \mathcal{B}^n tq \forall c \in \mathcal{C} < x, , c >= 0\}$
est appelé code dual de $\mathcal{C}$

d'après les résultats d'algèbre linéaire

- $\mathcal{C}^{\perp} = \{x \in \mathcal{B}^n \ tq \ Gx^t = 0\}$
- $dim(\mathcal{C}) = k \Longrightarrow dim\left(\mathcal{C}^{\perp}\right) = n - k$
- $\mathcal{C}^{\perp\perp} = \mathcal{C}$
  $\mathcal{C} = \left\{y \in \mathcal{B}^n tq \ \forall x \in \mathcal{C}^{\perp}, < y, x >= 0\right\}$
  si H est une matrice génératrice du code dual : $\mathcal{C} = \left\{y \in \mathcal{B}^n \ tq \ yH^t = 0\right\}$

Dual d'un code linéaire : Exemple

soit le [7,3]-code de matrice génératrice systématique

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

- recherche de $\mathcal{C}^{\perp}$:
  on cherche les mots $abcdefg \in \mathcal{B}^7$ tel que $G.(abcdefg)^t = 0$. On obtient le système d'équations linéaires homogènes

$$\begin{cases} a \oplus d \oplus e & = & 0 \\ b \oplus e \oplus g & = & 0 \\ c \oplus d \oplus f \oplus g & = & 0 \end{cases}$$

  ce système est de rang 3,prenons a, b, c comme inconnues principales,et d, e, f, g comme paramètres

$$\begin{cases} a & = & d \oplus e \\ b & = & e \oplus g \\ c & = & d \oplus f \oplus g \end{cases}$$

  on a donc

$$\begin{aligned} \mathcal{C}^{\perp} &= \left\{(d \oplus e)(e \oplus g)(d \oplus f \oplus g)defg | defg \in \mathcal{B}^4\right\} \\ &= \left\{d(1011000) \oplus e(1100100) \oplus f(0010010) \oplus g(0110001) | defg \in \mathcal{B}^4\right\} \\ &= Vec\left\{1011000; 1100100; 0010010; 0110001\right\} \end{aligned}$$

  une matrice génératrice du code dual est

$$H = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad H = \begin{pmatrix} P^t & I_4 \end{pmatrix}$$

Dual d'un code linéaire : Exemple

- $\mathcal{C} = \mathcal{C}^{\perp\perp}$ se traduit par la recherche des solutions du système $abcdefg \cdot H^t = 0$

$$\begin{cases} a \oplus c \oplus d & = & 0 \\ a \oplus b \oplus e & = & 0 \\ c \oplus f & = & 0 \\ b \oplus c \oplus g & = & 0 \end{cases}$$

Le code est ainsi défini comme l'intersection d'hyperplans

Matrice de contrôle de parité

Soit G une matrice génératrice d'un $(n,k]$-code linéaire $\mathcal{C}$.
On appelle matrice de contrôle de parité toute matrice H génératrice du code dual $\mathcal{C}^{\perp}$

Une matrice H de contrôle de parité d'un $[n,k]$- code est une matrice $(n-k)\times n$ définie par

$$GH^{T} = O_{k\times(n-k)}$$

Dans le cas où G est écrite sous forme systématique $G = \begin{pmatrix} I_k & P_{k\times(n-k)} \end{pmatrix}$

$$H = \begin{pmatrix} P^t_{k\times(n-k)} & I_{n-k} \end{pmatrix}$$

un mot $m \in \mathcal{B}^n$ est un mot de code si et seulement si

$$mH^t = 0_{1\times(n-k)}$$

Syndrome

Un codage linéaire est un application linéaire g

$$\begin{array}{cccc} \sigma : & \mathcal{B}^n & \to & \mathcal{B}^{n-k} \\ & & \to & \sigma(x) = x.H^t \end{array}$$

- σ est linéaire
- $Ker(\sigma) = \mathcal{C}$

Soit $m = i|p$ un mot en sortie de canal .
Son syndrome est

$$\begin{aligned} \sigma(m) &= i|p \cdot H^T \\ &= i|p \begin{pmatrix} P \\ I \end{pmatrix} \\ &= iP \oplus p \end{aligned}$$

iP est la partie de contrôle du mot de code associé à la partie informative i, p est la partie de contrôle reçue.
Le syndrome de m est donc le vecteur d'erreur de la partie contrôle lorsqu'on suppose la partie informative exacte
Exemple :

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

1. combien y -a-t-il de syndromes ?
2. donner la liste des syndromes des mots de poids 1 .
3. pour les autres syndromes déterminer les mots de plus faible poids dont ils sont le syndrome

Syndrome

syndromes			mots associés				
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1
0	1	0	0	0	0	1	0
0	1	1	0	1	0	0	0
1	0	0	,0	0	1	0	0
1	0	1	1	0	0	0	0
1	1	0	1	1	0	0	0
1	1	1	1	0	0	1	0

Relation d'équivalence sur $\mathcal{B}^n$

Etant donné deux mots m_1 et m_2 de $\mathcal{B}^n$ on définit la relation $\mathcal{R}$ par
$m_1 \mathcal{R} m_2 \iff \sigma(m_1) = \sigma(m_2)$

comme

$$\sigma(m_1) = \sigma(m_2) \iff m_1 \oplus m_2 \in Ker(\sigma)$$

on en déduit

$$m_1 \mathcal{R} m_2 \iff m_2 \in m_1 \oplus \mathcal{C}$$

- $\mathcal{R}$ est une relation d'équivalence
- Les classes d'équivalences peuvent être étiquetées par leur syndrome

$$\overline{\sigma(m)} = m \oplus \mathcal{C}$$

- Il y a 2^{n-k} classes d'équivalences de cardinal 2^k

Décodage par tableau standard

pour tout $m \in \mathcal{B}^n$
$m \oplus \overline{\sigma(m)} = \mathcal{C}$

pour corrigé un message reçu en sortie de canal suivant le maximum de vraisemblance on cherchera parmi $\overline{\sigma(m)}$ s'il existe un unique de plus faible poids mot m' .
On corrigera m par $c = m \oplus m'$
Construction du tableau standard

1. la première ligne est composée des mots des k mots de code.
2. pour composé la ligne j, on sélectionne un mot m_j de plus faible poids n'étant pas dans les lignes précédentes. La ligne j est composée des mots de $m_j \oplus \mathcal{C}$

$\mathcal{C}$	0	c_1	c_2	$\cdots$	c_{2^k}
$m_1 \oplus \mathcal{C}$	m_1	$m_1 \oplus c_1$	$m_1 \oplus c_2$	$\cdots$	$m_1 \oplus c_{2^k}$
$m_2 \oplus \mathcal{C}$	m_2	$m_2 \oplus c_1$	$m_2 \oplus c_2$	$\cdots$	$m_2 \oplus c_{2^k}$
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$m_{2^{n-k}} \oplus \mathcal{C}$	$m_{2^{n-k}}$	$m_{2^{n-k}} \oplus c_1$	$m_{2^{n-k}} \oplus c_2$	$\cdots$	$m_{2^{n-k}} \oplus c_{2^k}$

Pour corriger le mot m reçu on le repère dans le tableau standard. S'il est dans la ligne j et que m_j est l'unique mot de plus faible poids de cette ligne, on le corrige par $c = m \oplus m_j$

0	$\cdots$	$\cdots$	c	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\uparrow$	$\cdots$
m_j	$\longleftarrow$		m	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Décodage par tableau standard :exemple

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

La distance du code est $d = 3$: pon est sûr de pouvoir corriger une erreur $k = 2$ il ya donc $2^2 = 4$ mots de codes, $n = 5$ il y a donc 2^3 syndromes.
Le tableau comporte 8 colonnes et 4 colonnes

00000	01101	10011	11110
00001	01100	10010	11111
00010	01111	10001	11100
00100	01001	10111	11010
01000	00101	11011	10110
10000	11101	00011	01110
00110	01011	10101	11000
01010	00111	11001	10100

- Si on reçois $m = 10001$ le vecteur d'erreur de plus faible poids est 00010 . On corrige par le mot $c = m \oplus 00010 = 10011$
- Si on reçois $m = 10001$ il y a deux vecteurs d'erreurs de plus faible poids : 00110 et 11000. Le mot ne peut pas être corrigé.

Remarque : espace mémoire occupé (16g=Go pour un [32, 6]-code !)

Décodage par syndrome

chaque ligne du tableau est une classe d'équivalence . On établit une liste des syndrome en affectant à chaque syndrome comme représentant un mot se plus faible de poids. La correction est possible code est correcteur dans la mesure ou ce représentant est unique

$\mathcal{C}$	0	c_1	c_2	$\cdots$	c_k	$\sigma(0)$
$m_1 \oplus \mathcal{C}$	m_1	$m_1 \oplus c_1$	$m_1 \oplus c_2$	$\cdots$	$m_1 \oplus c_k$	$\sigma(m_1)$
$m_2 \oplus \mathcal{C}$	m_2	$m_2 \oplus c_1$	$m_2 \oplus c_2$	$\cdots$	$m_2 \oplus c_k$	$\sigma(m_2)$
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
$m_{n-k} \oplus \mathcal{C}$	m_{n-k}	$m_{n-k} \oplus c_1$	$m_{n-k} \oplus c_2$	$\cdots$	$m_{n-k} \oplus c_k$	$\sigma(m_{n-k})$

Le calcul du syndrome du message m reçu désigne le vecteur d'erreur m_j. Le décodage obtenu est $c = m \oplus m_j$

Décodage par syndrome : exemple

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

$n = 6$, $k = 3$, la distance du code vaut $d = 3$. On est sûr de pouvoir corriger une erreur.
Les syndromes des mots de poids 1 sont les ligne de H^t

	syndrome				mots de poids 1 associé					
	1	1	0		1	0	0	0	0	0
	1	0	1		0	1	0	0	0	0
H^t	0	1	1		0	0	1	0	0	0
	1	0	0		0	0	0	1	0	0
	0	1	0		0	0	0	0	1	0
	0	0	1		0	0	0	0	0	1

Si un le syndrome d'un m se trouve dans la transposée de la matrice de contrôle, son numéro de ligne est le numéro du bit à corriger (en partant de la gauche)
Il a $2^3 - = 7$ syndrômes non nuls. Dans notre tableau il en manque 1 : 111.
Les éléments de la classe d'équivalence $\overline{111}$ sont tous au moins de poids 2

	syndrome				mots de poids 2 associés					
					1	0	0	0	0	1
	1	1	1		0	1	0	0	1	0
					0	0	1	1	0	0

Dans cette exemple si le syndrome du message reçu m

- est 000 alors m est un mot de code,
- appartient aux lignes de H^t on corrige en modifiant le bit de rang le numéro de la ligne
- est 111 on ne peut pas le corriger

Code :contrôle de parité C(3,2)

$k = 2$, aux deux bits informatif on adjoint un bit de parité égale à leur somme

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

$$P = \begin{matrix} 1 \\ 1 \end{matrix}$$

$$H = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

les mots de code sont les mots de $\mathcal{B}^3$, $c_1 c_2 c_3$ tels que $c_1 \oplus c_2 \oplus c_3 = 0$

Code de Hamming

- On choisit un $[n, k]$-code tel les classes d'équivalences des syndromes non nuls admettent toutes comme représentant un mot de poids 1 (et un seul)
  il y a n mots de poids 1 , $2^{n-k} - 1$ syndromes non nuls.

Pour code de Hamming :

$$n \;=\; 2n - k \;-\; 1$$

Exemple $n = 7$, $k = 4$

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

Extrait sujet 2020

On considère le code bloc linéaire de matrice génératrice :

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

1. Donner la taille n des mots de code, le nombre k de bits d'information
2. Dire de combien de mots le code est composé et écrire l'ensemble des mots de code
3. Calculer la distance minimale du code et en déduire sa capacité de correction d'erreur
4. Ecrire la matrice de contrôle de parité H
5. On note c le mot de code en entrée de canal. Le canal est supposé binaire symétrique de probabilité de transition p. On note $\mathbf{y}$ la séquence associée en sortie du canal.
    1. En supposant qu'il s'est produit une seule erreur, dire quelles sont les valeurs possibles du syndrome et donner un algorithme de correction d'erreur (autre que la recherche exhaustive)
    2. le syndrome vaut 0101, que peut-on dire ?
    3. Peut-on construire un code à répétition de même rendement et de même distance minimale ? Justifier en cas de réponse négative ou donner sa matrice génératrice dans le cas positif