1. Introduction

Le codage de l'information concerne les méthodes de représenter l'information afin de pouvoir la manipuler, la stocker ou la transmettre. La méthode de codage est choisie suivant l'application voulue:

* Le codage de source permet de faire la compression de données.
* Le codage visuel (destiné au code bare, QR code, ...)
* Le chiffrement est le codage qui permet à masquer ou brouiller une information.
* Le codage de canal permet une représentation des données de façon à résister aux erreurs de transmission.

Dans ce cours on va s'intéresser à ce dernier codage où l'échange de l'information s'effectue par transmission d'un émetteur vers un récepteur à travers un canal.
Il est aussi possible de représenter physiquement cette information par un signal.
Ce dernier codage fait naissance à la théorie des codes correcteurs qui est basée sur la redondance, en plus des corrections d'erreurs elle a des applications en cryptage, authentification, partage de secret et en stéganographie...

2. Codes Correcteurs

L'information est transformée en une grandeur mathématique: suite, vecteur, mot, polynôme, ...

Si $A=\{a_1,\dots,a_m\}$ est un alphabet (ensemble fini), un mot de longueur n est un élément de A^n, Ses composantes sont appelées lettres ou bits, quand $A=\{0,1\}$ on parle de mots binaires.

Quand le canal est soumis à des perturbations, les bits du mot envoyé peuvent être modifiés.

En théorie des codes correcteurs l'idée de base est d'allonger le mot envoyé c'est à dire lui ajouter des bits (redondance) de sorte à appartenir à un ensemble C appelé code.

La redondance est conçue pour détecter ou corriger les erreurs.

Codes en blocs

Soit n un entier naturel et A un alphabet.

Un code en blocs C de longueur n est une partie de A^n:

$$C \subset A^n.$$

Il y a des codes où les mots ne sont pas de même longueur, citons les codes en treillis, codes convolutifs....

Exemples:

a) * Soit $A=\{0,1\}$, $C=\{001, 101, 111\}$ est un code en blocs binaire de longueur 3.

** Si $A=\{a,b,c\}$, $C=\{aa, ba, ac\}$ est un code en blocs de longueur 2.

b) Pour tout alphabet A, A^n est appelé code trivial,

$C=\{\underbrace{a\dots a}_{n} : a\in A\}$ est dit code répétition de longueur n.

c) Codes contrôle de parité

Pour n fixé, aux n bits du message $m=a_1\dots a_n$ on ajoute un bit supplémentaire qui est la somme $\sum_{i=1}^{n} a_i$.

$a_i\in\mathbb{F}_2$, $m=a_1\dots a_n$ $\xrightarrow{\text{codage}}$ $m'=a_1\dots a_n(\sum a_i)\in C$.

A la reception si la somme des bits de m' n'est pas nulle cela signifie qu'il y a au moins une erreur.

Exemple:

$C=\{000, 101, 011, 110\}$ est un code controle de parité de longueur 3.

d) Code ISBN (International Standard Book Number)

est un numéro qui permet d'identifier chaque livre publié. Le code ISBN à 10 chiffres est composé de 4 segments: $A-B-C-D$ où

A identifie un grmpe de codes pour un pays ou une zone linguistique.

B identifie l'editeur de la publication.

C est un numéro d'ordre de l'ouvrage chez l'editeur.

D est un nombre pour détecter les erreurs.

Le code ISBN 10 est une suite $a_1, \dots, a_{10}$ tels que

$a_i \in \frac{\mathbb{Z}}{11\mathbb{Z}} = \{0, 1, \dots, 10 = X\}$ et $\sum_{k=1}^{10} k a_k \equiv 0 \ [11]$.

Exemple:

ISBN 1-58182-008-9 est le code de l'ouvrage:

James Reasoner, Manassas, Cumberland House (1999).

La première partie 1 concerne le groupe linguistique Anglais.

$\sum k a_k = 1.1 + 2.5 + 3.8 + 4.1 + 5.8 + 6.2 + 7.0 + 8.0 + 9.8 + 10.9 = 253 \equiv 0 \ [11]$.

Vu l'augmentation du nombre de publications électroniques, la capacité de numérotation du système ISBN 10 est devenue insuffisante, pour cela en Janvier 2007 la longueur a été étendue à 13 chiffres.

Distance de Hamming.

La distance de Hamming est utilisée en télécommunications pour compter le nombre de bits altérés dans la transmission d'un message de longueur fixée.

Définition:

Soit A un alphabet.

Si $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ sont deux éléments de A^n, la distance de Hamming entre x et y correspond au nombre de composantes pour lequel ces deux vecteurs diffèrent:

$$d(x,y) = \text{card}\{ i : x_i \neq y_i , 1 \leq i \leq n \}.$$

Distance minimale:

Soit $C \subset A^n$ un code de longueur n, sa distance minimale de Hamming d_{min} est définie par

$$d_{min} = \min \{ d(x,y) : x, y \in C \text{ et } x \neq y \}.$$

Exemple:

Soit $C = \{ a = 1001 , b = 0101 , c = 0111 \}$.

On a $d(a,b) = 2$, $d(a,c) = 3$, $d(b,c) = 1$

donc $d_{min}(C) = 1$.

Poids d'un vecteur

Soit $x \in A^n$, $x = (x_1, \ldots, x_n)$, son poids $w(x)$ est le nombre de ses composantes non nulles:

$$w(x) = \text{card}\{ i : x_i \neq 0 \}.$$

Décodage

Soit un message $m = (x_1, \ldots, x_k)$ de longueur k.
Soit C un code de longueur n, $(n > k)$, pour coder m on le plonge dans C (par une application injective).
m codé devient $c = c_1 \ldots c_n$ et il est envoyé par le canal, à la réception le mot reçu $\acute{c} = \acute{c}_1 \ldots \acute{c}_n$ est de longueur n, le nombre d'erreurs est exactement $d(c, \acute{c})$.

Pour décoder et retrouver m on cherche le mot de C, le plus proche de $\acute{c}$, ceci est possible si le nombre d'erreurs ne dépasse pas la capacité de correction de C.

Définition:

On dit que $C \subset A^n$ corrige t erreurs si pour tout $x \in A^n$ il existe au plus $c \in C$ tel que

$$d(x,c) \leq t.$$

On dit que C est de capacité e si C corrige e erreurs mais il ne corrige pas e+1 erreurs.

On montre que:

Si C est un code de distance minimale d, il detecte d-1 erreurs et sa capacité de correction est $e = \left[\frac{d-1}{2}\right]$ où $[x]$ est la partie entière de x.

Des détails supplémentaires seront donnés dans les prochaines sections.

3. Codes linéaires

Soit K un corps fini.

Définition :

on appelle code linéaire de longueur n tout sous-espace vectoriel C de K^n.

Si $\dim C = k$ et C est de distance minimale d on dit que C est un $[n, k, d]$ code sur K.

Donc si $|K| = q$ on note K par $\mathbb{F}_q$ et on a

$$|C| = |\mathbb{F}_q|^{\dim C} = q^k.$$

Exemples :

a) Sur $\mathbb{F}_3 = \{0, 1, 2\}$, $C = \{00, 11, 22\}$ est $[2, 1, 2]$ code, il est engendré par 11 :

$C = \{0 \cdot 11 = 00,\ 1 \cdot 11 = 11,\ 2 \cdot 11 = 22\}$

b) Le code contrôle de parité binaire de longueur n est défini par :

$$C = \left\{ (x_1, \dots, x_n) \in \mathbb{F}_2^n : \sum_{i=1}^{n} x_i = 0 \right\}.$$

on a $x_n = \sum_{i=1}^{n-1} x_i$ et $\{e_1, \dots, e_{n-1}\}$ avec $e_i = (0, \dots, \underset{i}{1}, 0 \dots 0 1)$

est une base de C, d'où $\dim C = n - 1$.

Propriétés :

a) Si $x, y \in K^n$, on a $d(x,y) = w(x-y)$

2) $w(x) = 0 \iff x = 0$

3) $w(x+y) \leq w(x) + w(y)$

4) $w(\lambda x) = w(x)$ pour $\lambda \in K - \{0\}$

b) Si C est un code linéaire, $d_{min}(C) = w_{min}(C) = \min\{w(x) : x \in C - \{0\}\}$

c) Inégalité de Singletm :

Soit C un $[n, k, d]$ code alors $d \leq n - k + 1$.

Remarque :

quand $d = n - k + 1$ le code C est dit MDS (Maximum Distance Separable).

Description des codes par les matrices génératrices

En algèbre linéaire, il est connu que l'image d'une application linéaire est un sous-espace vectoriel de l'espace d'arrivée et son noyau est un sous-espace vectoriel de l'espace de départ, ceci permet d'associer à tout code linéaire C deux matrices : une matrice génératrice G et une matrice de contrôle H.

Définition :

Soit C un code linéaire de longueur n, de dimention k sur K. Une matrice génératrice G de C est une matrice $k \times n$ à éléments dans K dont les lignes forment une base de C.

Alors $C = \{ (x_1, \ldots, x_k) G : (x_1, \ldots, x_k) \in K^k \}$.

Donc un $[n,k]$ code est complètement déterminé par une matrice $k \times n$ de rang k sur K.

Exemple:

Considérons le code contrôle de parité de longueur 3:

$C = \{000, 101, 011, 110\}$.

comme $\{101, 110\}$ est une base de C, $G = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ est une matrice génératrice de C d'où

$C = \{ (x_1, x_2) G = (x_1 + x_2, x_2, x_1) : x_i \in \mathbb{F}_2 \}$.

A une autre base de C: $\{011, 110\}$ on associe une autre matrice génératrice $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, donc une matrice génératrice n'est pas unique.

Définition:

Un $[n,k]$ code est dit systématique s'il admet une matrice génératrice de la forme $G = [I_k | A]$ où I_k est la matrice d'identité d'ordre k et $A \in M_{k,n-k}(K)$.

Dans ce cas G est dite normalisée et une partie du mot codé coincide avec le message.

Exemple: Soient $G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ et $m = (x_1, x_2)$

alors $mG = (\underbrace{x_1, x_2}_{m}, x_1 + x_2)$.

Description des codes par les matrices de contrôle

Soit C un $[n,k]$ code linéaire sur K.
on appelle matrice de contrôle de C, toute matrice H de type $(n-k)\times n$, de rang $n-k$ telle que

$$C = \{ x \in K^n : x \cdot {}^tH = 0 \} \quad \text{où } {}^tH \text{ est la transposée de } H.$$

Donc la matrice H contrôle l'appartenance d'un élément x au code C : $\quad x \in C \iff x \cdot {}^tH = 0$

Remarque :

a) Soit C un code de matrice génératrice G et $H \in M_{n-k,n}(K)$, on montre que

H est une matrice de contrôle de $C \iff$ 1) $\text{rang}(H) = n-k$ et 2) $G \cdot {}^tH = 0$

b) Pour un code systématique C de matrice normalisée $G = (I_k \mid A)$ il est possible de calculer une matrice de contrôle par $H = ({}^tA \mid -I_{n-k})$.

Exemple :

Soit le code binaire de matrice génératrice

$$G = \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{array}\right) \quad \text{(normalisée)}.$$

Une matrice de contrôle H est donnée par $\left(\begin{array}{ccc|cc} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{array}\right)$.

Pour expliciter les éléments de C on calcule

$(x_1, x_2, x_3)\, G = (x_1, x_2, x_3, x_1+x_2, x_1+x_3)$ avec $x_i \in \mathbb{F}_2$.

d'où $C = \{ 00000, 10011, 01010, 00101, 11011,$

$01111, 10110, 11100 \}$.

on peut retrouver les éléments de C en utilisant H :

$$(x_1, x_2, x_3, x_4, x_5) \in C \iff (x_1, x_2, x_3, x_4, x_5) \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (0,0)$$

en résolvant le système

$$\begin{cases} x_1 + x_2 + x_4 = 0 \\ x_1 + x_3 + x_5 = 0 \end{cases}$$

on obtient $x_1 = x_3 + x_5$ et $x_2 = x_3 + x_4 + x_5$

et $C = \{ (x_3+x_5, x_3+x_4+x_5, x_3, x_4, x_5) \,;\, x_3, x_4, x_5 \in \mathbb{F}_2 \}$.

Master CS

Département d'informatique

TD 1 : Théorie des codes linéaires

EX:1

a) Combien y a-t-il de codes binaires de longueur 3 ?

b) Combien y a-t-il de codes sur $\mathbb{F}_q$ et de longueur n ?

EX:2

Montrer que la distance de Hamming vérifie :

a) $d(x,y) = d(y,x)$

b) $x = y \iff d(x,y) = 0$

c) $d(x,y) \leq d(x,z) + d(z,y)$

pour $x, y, z \in K^n$.

EX:3

Calculer la distance minimale d pour les codes suivants :

a) $C = A^n$, (A étant un alphabet).

b) C code répétition sur A de longueur n.

EX:4

Soit $K = \mathbb{F}_q$ un corps fini de cardinal q.

a) Combien y-a-t-il de mots de K^n de poids m, $m \leq n$?

distinguer les deux cas : $q = 2$, $q \neq 2$.

b) Expliciter ces nombres pour $q = 2$, $n = 2k$, $m = 2$,

les nombres obtenus sont appelés nombres hexagonaux.

EX; 5

On considère les codes binaires suivants:

$C_1 = \{011, 110, 101, 000\}$

$C_2 = \{1100, 1010, 1001\}$

$C_3 = \{00000, 10101, 01111, 11010\}$

dire dans chaque cas si le code est linéaire ?

EX; 6

Soit A un alphabet de cardinal q et a un élément fixe de A.

Posons $C = \{c = x \underbrace{\dots}_{d} x \underbrace{a \dots a}_{n-d} : x \in A\}$

a) Calculer la distance minimale de C et $|C|$.

b) Si $A = \mathbb{F}_q$ est un corps fini, C est-t-il linéaire ?

EX; 7 Soit la matrice binaire

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

1) Déterminer $C = \{(x_1, x_2, x_3) G : x_i \in \mathbb{F}_2\}$

et vérifier que C est linéaire.

2) Déterminer sa dimension, son cardinal et sa distance minimale.

EX; 8

Considérons C le code ISBN (10).

a) Montrer que C est linéaire et préciser sa matrice de contrôle.

b) Déterminer ses paramètres : n, k, d.

EX : 9

Soit C un code contrôle de parité de longueur n :

$$C = \left\{ (x_1, \dots, x_n) \in \mathbb{F}_2^n : \sum_{i=1}^{n} x_i = 0 \right\}$$

a) Déterminer une matrice de contrôle de C.

b) Calculer les paramètres de C : n, k, d.

EX : 10

Soit C un code linéaire de longueur n sur $\mathbb{F}_q$.

1) Vérifier que si $x, y \in C$, $d(x,y) = w(x-y)$.

2) Montrer $d_{min} = w_{min}$.

3) Comparer le nombre de possibilités dans le calcul de la distance minimale par les deux méthodes :

$$d_{min} \quad , \quad w_{min}$$

EX : 11

Soit C un code linéaire sur $\mathbb{F}_3$ de matrice génératrice

$$G = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 2 & 1 & 0 & 2 \end{pmatrix}$$

1) Écrire les éléments de C.

2) C est-il MDS ?

3) Peut-on calculer une matrice de contrôle de C.