I - Codes correcteurs

Résumé

Quelques notions sur les codes correcteurs d'erreurs. Ces codes permettent de détecter les erreurs de transmission et éventuellement de réaliser une correction automatique. Ils sont utilisés en téléphonie, sur les disques optiques, etc.

1 Codage d'un message

1.1 Erreurs de transmission

Lors de la transmission informatique d'un message des erreurs de transmission surviennent. Un message m étant envoyé ($m \in \mathbf{B}^n$) chaque bit de m est susceptible d'être reçu erroné. Un 0 peut être transformé en 1 et vice-versa. À la réception du message, se pose deux questions :
– Le message reçu est-il le bon (détection d'erreurs éventuelles) ?
– Le cas échéant est-il possible de corriger ces erreurs ?
On essaiera d'apporter des réponses à ces deux questions.

1.2 Distance de Hamming

Définition 1 *Soit $m \in \mathbf{B}^n$, on appelle poids de m le nombre de bits de m égaux à 1.*

$$\omega(m) = \sharp\{m_i | m_i = 1, i = 1 \cdots n\}$$

Exemple 1

Définition 2 *Soient a et b deux mots de $\mathbf{B}^n$, on appelle somme modulo 2 de a et de b le mot suivant :*
$\overline{a_1 + b_1} \cdot \overline{a_2 + b_2} \cdot \cdots \cdot \overline{a_n + b_n}$ *où les sommes ont lieu dans $\mathbb{Z}/2\mathbb{Z}$.*
Il sera noté $a \oplus b$.

Exemple 2

Définition 3 *On appelle distance de Hamming entre a et b l'entier $\omega(a \oplus b)$ noté $d(a; b)$*

Exemple 3

Proposition 1 – *d est une distance et vérifie les propriétés suivantes :*
  – $d(a; b) \geq 0$
  – $d(a; b) = d(b; a)$
  – $d(a; b) = 0 \iff a = b$
  – $d(a; b) + d(b; c) \geq d(a; c)$
– $d(a \oplus x; b \oplus x) = d(a; b)$

La distance de Hamming entre a et b est le nombre d'arêtes qu'il faut parcourir pour se rendre de a à b par le chemin le plus court possible.

1.3 Vecteur erreur

On envoie un message m et l'on reçoit le message M. Le mot $e = m \oplus M$ est appelé **vecteur erreur**. Le poids de e est le nombre de bits mal transmis.

Remarque 1 Si l'on connaît le vecteur erreur, on peut retrouver le message original. $m = M \oplus e$.

Dans la suite on supposera que le canal de transmission est sans mémoire, c'est à dire que la probabilité qu'un bit soit envoyé de manière erronée ne dépend pas des bits précédemment envoyé.
On suppose de plus que la probabilité qu'un bit soit mal transmis soit égale à p.

Théorème 1 *Lorsqu'on envoie un message de longueur n sur un canal sans mémoire.*

- *La probabilité que le vecteur erreur soit e est $p^{\omega(e)}(1-p)^{n-\omega(e)}$*
- *La probabilité que le nombre d'erreur soit k est $\begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$*

1.4 Codage par blocs

On appelle codage par bloc la technique de codage qui consiste à découper le message en bloc de taille fixe n et de transmettre un message de taille $n+k$.
Un codage par bloc consistant à adjoindre un mot de taille k au bloc de taille n est appelé codage systématique.
Les k bits ajoutés sont appelés les bits de contrôle.
Dans ce cas la vérification est facile, il suffit de prendre les n premiers bits, de calculer les bits de contrôles et de vérifier qu'ils correspondent avec ceux envoyés.

Exemple 4 Exemple fondamentaux

Exemple 5 Exemple de codage non systématique

Le codage est une application de $\mathbf{B}^n$ dans $\mathbf{B}^{n+k}$ qui a chaque bloc de taille n associe un bloc de taille $k+n$.

Définition 4 *Le rapport $\tau = \frac{n}{n+k}$ est appelé rendement du code.*

2 Correction et détection des erreurs de transmission

On considère un codage systématique $\mathbf{B}^n \longrightarrow \mathbf{B}^m$ $(m \geq n)$.
n est la dimension du code, m est la longueur du code.
Lorsque que l'on reçoit un message et qu'une erreur est détectée, on choisit de corriger l'erreur par le mot de code le plus proche du message.
En pratique, si l'on reçoit m et qu'une erreur est détectée, on choisit le vecteur e erreur de poids le plus faible possible et l'on corrige par $m \oplus e$.

Définition 5 *On appelle distance minimale d'un code la plus petite distance séparant deux mots de code distinct.*

Théorème 2 *On considère un code de distance minimale d.*

- *Le codage détecte de façon certaine tous les messages faux dont le nombre d'erreurs N est strictement plus petit que d. ($0 < N < d$).*
- *La méthode de correction corrige correctement tous les messages dont le nombre d'erreurs est strictement plus petit que $\frac{d}{2}$ ($0 \leq N < \frac{d}{2}$.)*

On pose $t = \mathrm{E}(\frac{d}{-}12)$, t est le nombre d'erreur corrigées par le code.

Remarque 2 Les trois nombres n, m et d décrivent les caractéristiques d'un code, on parlera de code $[m, n, d]$.

3 Code linéaire

3.1 Principes du codage linéaire

3.2 Matrice génératrice normalisée

On considère un code linéaire systématique f de $\mathbf{B}^n$ vers $\mathbf{B}^m$. Pour un code linéaire, il suffit de s'intéresser au mot de code associé aux vecteurs de la base canonique. Soit $e_1, e_2, \cdots, e_n$ la base canonique de $\mathbf{B}^n$. Ces mots de code sont $f(e_1), f(e_2), \cdots, f(e_n)$; le codage étant systématique :

$$\begin{aligned} f(e_1) &= f(10\cdots0) = 10\cdots0\cdot k_1 \\ f(e_2) &= f(01\cdots0) = 01\cdots0\cdot k_2 \\ \vdots & \qquad\qquad \vdots \\ f(e_n) &= f(00\cdots1) = 00\cdots1\cdot k_n \end{aligned}$$

Où chaque k_i est un élément de $\mathbf{B}^{m-n}$.
Que l'on peut réécrire matriciellement :

$$G_{n,m} = \begin{pmatrix} & I_n & \\ & & \\ k_1 & k_2 \quad \cdots & k_n \end{pmatrix} \qquad\qquad f(e_i) = G_{n,m} \times e_i$$

Définition 6 *La matrice $G_{n,m}$ est appelée matrice génératrice du code.*

$$G_{n,m} = \begin{pmatrix} & I_n & \\ & & \\ - & - & - \\ & K_{m-n;n} & \\ & & \end{pmatrix}$$ où $K_{m-n;n}$ est appelée matrice des clefs.

Proposition 2 *Soit c un message de $\mathbf{B}^n$ le mot de code associé est $G_{n,m} \times c$*

3.3 Fonction syndrome

3.3.1 Matrice de contrôle normalisée

Soit $\mathcal{C}$ un code de $\mathbf{B}^n \longrightarrow \mathbf{B}^m$ de matrice génératrice normalisé $G = G_{n,m}$. On pose $G = \begin{pmatrix} & I_n & \\ & & \\ - & - & - \\ & K & \end{pmatrix}$

$c = (i, q)$ appartient à $\mathcal{C}$ si et seulement si $Ki = q$. Condition que l'on peut réécrire $Ki + q = 0$.

Proposition 3 $c = (i, q) \in \mathcal{C} \Longleftrightarrow Ki + q = 0$

3.3.2 Syndrome

Soit $S : \mathbf{B}^n \longrightarrow \mathbf{B}^{m-n}$ qui à $c = (i, q)$ associe $Ki + q$

La matrice associée est $C_{m-n,m} = \left(\begin{array}{c|c} & \\ K_{m-n;n} & I_{m-n} \\ & \end{array}\right)$

D'après ce qui précède $c = (i, q)$ appartient à $\mathcal{C}$ si et seulement si $C_{m-n,m}c = 0$

Définition 7 *Soit $\mathcal{C}$ un code, une application linéaire telle que $\mathcal{C}$ soit la pré-image de 0 est appelée fonction de contrôle ou fonction syndrome*

Exemple 6

3.3.3 Détection des erreurs

On envoie M et l'on reçoit $\tilde{M} = M \oplus e$

Proposition 4 *Soit e un vecteur erreur de poids non nul, l'erreur est détectée si et seulement si $S(e) \neq 0$*

Preuve $S(\tilde{M}) = S(M \oplus e) = S(M) \oplus S(e) = 0 \oplus S(e) = S(e)$

CQFD.

3.4 Tableau standard

L'idée est de ranger les mots de code suivant leur syndrome dans un tableau. Ce tableau sera appelé tableau standard de correction.

3.4.1 Construction

Deux mots de code sont dit équivalent pour S si $S(u) = S(v)$. Cette relation d'équivalence induit une partition de $\mathbf{B}^m$ en différente classe d'équivalence. La classe d'équivalence de 0 étant égale à $\mathcal{C}$.
Notons S_j le vecteur de $\mathbf{B}^{m-n}$ dont la suite des composantes représente le nombre j en base 2.

- On commence le tableau en rangeant sur la première ligne tous les syndromes ; S_0 à gauche, puis S_1 etc.
- On range tous les mots de $\mathbf{B}^n$, suivant leurs syndromes en commençant par ceux de poids le plus faible.

Remarque 3 La première colonne contient tous les mots du code.

3.4.2 Utilisation du tableau standard pour corriger un message

On envoie M et l'on reçoit $\tilde{M} = M \oplus e$, on calcul le syndrome de $\tilde{M}$ et l'on corrige par $\tilde{M} + e$ où e est un vecteur de poids minimum trouvé dans la colonne étiquetée par $S(\tilde{M})$.

Remarque 4 Il peut y avoir plusieurs vecteur erreur de poids minimum.

Exemple 7 Voici le tableau standard du code linéaire de parité $[4, 3, 2]$.

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

$S_0 = 0$	$S_1 = 1$
0000	0001
1001	1000
0101	0100
0011	0010
1100	1101
0110	0111
1010	1011
1110	1111

3.5 Tableau standard réduit

La table de correction peut être grande pour de grande valeur de m.

3.5.1 Construction

On pourra alors préférer le tableau standard réduit qui se construit comme le tableau précédent, mais en ne faisant figurer qu'un seul vecteur de poids minimum par colonne.

Exemple 8 En reprenant l'exemple précédent :

$S_0 = 0$	$S_1 = 1$
0000	0001

Remarque 5 Lorsqu'il y a plusieurs vecteurs erreurs de poids minimum, il y a un choix à effectuer, cela signifie également que les capacités de correction du code ne sont pas parfaite.

Exemple 9 Si l'on envoie $M = 1001$ que l'on reçoit $\tilde{M} = 0001$ de syndrome 1, la correction automatique à l'aide du tableau standard réduit donne $0001 \oplus 0001 = 0000$!

3.5.2 Utilisation du tableau standard pour corriger un message

3.5.3 En pratique

- La liste des syndromes est placée en première ligne
- Le représentant de la classe de 0 est 0_m
- Chaque vecteur de poids 1 est rangé en dessous de son syndrome, si la place n'est pas déjà occupée
- Si la deuxième ligne n'est pas complète on continue avec les vecteurs de poids 2, 3 etc. Jusqu'à ce que la deuxième ligne soit complète.