

Modélisation et résolutions numérique et symbolique de problèmes via les logiciels Maple et MATLAB (MODEL)

Cours n°7 : Codes correcteurs d'erreurs

Stef Graillat

Université Pierre et Marie Curie (Paris 6)



Rappel

- Examen réparti n°1 : le jeudi 12 novembre 2009, 13h30-15h30, salle 42-43/411 (pas de TD/TME cette semaine)
- Seuls les documents de cours seront autorisés
- Le programme de l'examen concerne les 5 premiers cours

- Introduction à Maple
- Rappel d'arithmétique et d'algèbre générale

Objectifs

- 1 Introduction aux codes correcteurs d'erreurs
- 2 Codes linéaires
- 3 Algorithme de codage et de décodage des codes linéaires

- 1 Généralités
- 2 Modélisation
- 3 Codes linéaires
- 4 Décodage par syndrome

Bibliographie en français

- Algèbre discrète et codes correcteurs, Odile Papini et Jacques Wolfmann, Springer, 1996
- Cours d'algèbre. Primalité. Divisibilité. Codes, Michel Demazure, Cassini, 1997
- Codage, cryptologie et applications, Bruno Martin, Presses Polytechniques et Universitaires Romandes (PPUR), 2004
- Théorie des Codes : Compression, Cryptage et Correction, J.-G. Dumas, J.-L. Roch, E. Tannier and S. Varrette, Dunod, 2007
- Codes correcteurs : théorie et applications, A. Poli et L. Huguet, Masson, 1989
- Mathématiques et Technologie, C. Rousseau et Y. Saint-Aubin, Springer, 2008
- Codes correcteurs d'erreurs, G. Cohen, J.L. Dornstetter et P. Godlewski, Masson, 1992
- Article « Code correcteur » de Wikipedia

- Applications of Abstract Algebra with Maple and MATLAB, Richard E. Klima, Neil Sigmon et Ernest Stitzinger, 2nd édition, Chapman & Hall/CRC, 2006
- An Introduction to Coding Theory, J.H. van Lint, 3e édition, Springer, 1998
- The theory of error-correcting codes, F. MacWilliams et N. Sloane, 11e édition North-Holland, 2003
- Information and Coding Theory, G. A. Jones and J. M. Jones, Springer, 2000
- Introduction to coding theory, R. M. Roth, Cambridge University Press, 2006
- Coding theory and cryptography, the essentials, 2nd édition, D.R. Hankerson, D.G Hoffman, et al, Marcel Dekker, 2000
- Fundamentals of error-correcting codes, W. C. Huffman et V. Pless, Cambridge University Press, 2003

Généralités

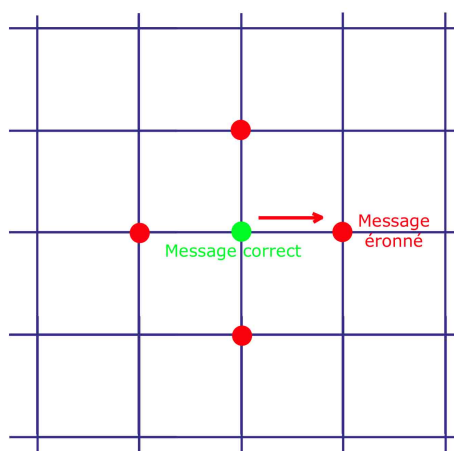
Définition 1

Un code correcteur d'erreurs est une technique de codage basée sur la redondance. Elle est destinée à corriger les erreurs de transmission d'une information (plus souvent appelée message) sur une voie de communication peu fiable.

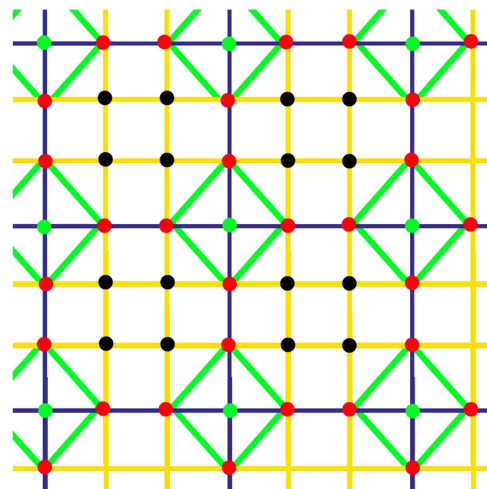
Utilisation dans

- les communications classiques (radio, câble coaxial, fibre optique)
- les supports pour le stockage (CD-ROM, RAM, etc.)

Principe des codes correcteurs

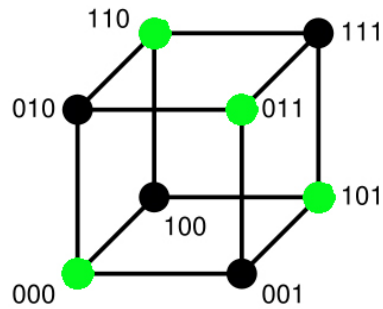


Code sans redondance



Code correcteur

Sur un paquet de k bits, on ajoute un bit de parité de sorte que la somme des $k + 1$ bits envoyés soit paire. On détecte une erreur sur $k + 1$ bits envoyés.



Modélisation

L'information à transmettre peut être vue comme une suite de symboles pris dans un ensemble fini.

- un **alphabet** \mathcal{A} est un ensemble fini de symboles (typiquement $\mathcal{A} = \mathbb{F}_2$ ou \mathcal{A} corps fini).
- Un **message** ou un **mot** est une suite à valeur dans un alphabet, il correspond à une suite de lettres.
- possibilité de modification (pas d'effacement) de symboles lors de la transmission.
- on traite l'information bloc par bloc (code en bloc).

Distance de Hamming

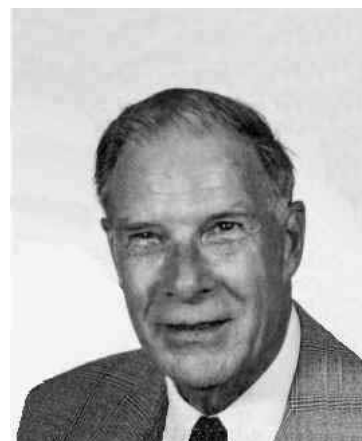
Définition 2

Soit \mathcal{A} un alphabet fini de symboles ($\mathcal{A} = \mathbb{F}_2$ ou \mathcal{A} un corps fini).

- La **distance de Hamming** de deux mots $x = x_1 \cdots x_n$ et $y = y_1 \cdots y_n$ de \mathcal{A}^n est le nombre de lettres qui diffèrent entre x et y , i.e.

$$d_H(x, y) = \text{card}\{i \in \{1, 2, \dots, n\} \mid x_i \neq y_i\}$$

- Le **poids de Hamming** de x est $w_H(x) = d_H(x, \mathbf{0}_n)$.



Richard Wesley
Hamming
1915 - 1998

La distance de Hamming mesure le nombre d'erreurs entre un vecteur u envoyé et un vecteur v réceptionné.

Propriété 1

La distance de Hamming est une distance, c'est-à-dire,

- $d(x, y) = 0$ si et seulement si $x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$

Boule de Hamming

Définition 3

La *boule (fermée)* de centre $c \in \mathcal{A}^n$ et de rayon r est définie par

$$B_H(c, r) = \{u \in \mathcal{A}^n, d_H(c, u) \leq r\}.$$

On note :

$$V_r(c) = \text{card}(B_H(c, r)).$$

Propriété 2

Si $q = \text{card}(\mathcal{A})$ alors

$$V_r(c) = \sum_{i=0}^r C_n^i (q-1)^i.$$

Définition 4

Soit \mathcal{A} un alphabet fini de symboles ($\mathcal{A} = \mathbb{F}_2$ ou \mathcal{A} un corps fini).

- Un code (correcteur d'erreur) en bloc de longueur n est un sous-espace métrique non vide de \mathcal{A}^n .
- La distance d'un mot $u \in \mathcal{A}^n$ à un code \mathcal{C} est

$$d_H(u, \mathcal{C}) = \min_{c \in \mathcal{C}} d_H(u, c).$$

- La *distance minimale* d'un code \mathcal{C} est

$$d_H(\mathcal{C}) = \min_{u, v \in \mathcal{C}, u \neq v} d_H(u, v).$$

Décodage

Définition 5

Un *algorithme de décodage* d'un code \mathcal{C} est une procédure qui à tout élément de \mathcal{A}^n associe un mot de \mathcal{C} ou échoue (symbole ∞),

$$\begin{aligned} \phi : \mathcal{A}^n &\rightarrow \mathcal{C} \cup \infty \\ c &\mapsto \phi(c) \end{aligned}$$

Définition 6

Un algorithme de décodage de \mathcal{C} est dit à *vraisemblance maximale* si pour tout $y \in \mathcal{A}^n$, le mot $x = \phi(y) \in \mathcal{C}$ est dans et réalise le maximum de la probabilité $P(x \text{ "émis"} \mid y \text{ "reçu"})$.

Soit \mathcal{C} un code de distance minimale d .

- Deux boules de rayon $(d - 1)/2$ centrées en deux mots de code distincts sont disjointes.
⇒ un code de distance minimale d peut corriger $\lfloor (d - 1)/2 \rfloor$ erreurs
- Toute boule de rayon $d - 1$ centrée en un mot de code ne contient aucun autre mot de code.
⇒ un code de distance minimale d peut détecter $d - 1$ erreurs

On dit que le code est $\lfloor (d - 1)/2 \rfloor$ -correcteur et $d - 1$ -détecteur.

Capacité de correction

Définition 7

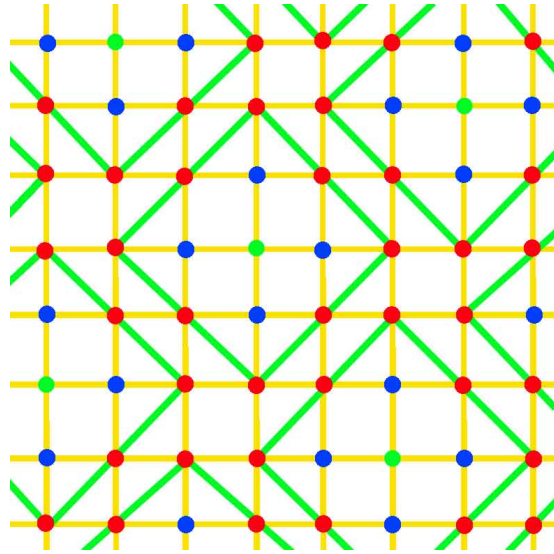
Soit \mathcal{C} un code sur \mathcal{A} (\mathcal{A} de cardinal q). La *capacité de correction* de \mathcal{C} est le plus grand entier t pour lequel toutes les boules fermées de rayon t centrées sur les mots de code sont disjointes.

Nous avons la majoration :

$$\sum_{c \in \mathcal{C}} V_t(c) \leq q^n$$

Le code est parfait si nous avons l'égalité, i.e. :

$$\sum_{c \in \mathcal{C}} V_t(c) = q^n$$



Codes linéaires

Définition 8

L'alphabet est un corps fini $\mathcal{A} = \mathbb{F}_q$. L'espace de Hamming \mathcal{A}^n est un espace vectoriel.

- Un *code linéaire* \mathcal{C} de longueur n sur \mathbb{F}_q est un sous-espace vectoriel de \mathbb{F}_q^n
- La dimension d'un code \mathcal{C} est sa dimension en tant que sous espace vectoriel de \mathbb{F}_q^n (\mathbb{F}_q^n est un espace vectoriel)
- Si k est la dimension de \mathcal{C} , son *taux d'information* est k/n

Nous parlerons de code $[n; k]_q$ si le code est de dimension k et de code $[n; k; d]_q$ si sa distance minimale est d .

Matrice génératrice

Définition 9

Soit \mathcal{C} un code $[n; k]_q$. Il existe une matrice $G \in \mathbb{F}_q^{k \times n}$ dont les k lignes sont les vecteurs d'une base de \mathcal{C} sur \mathbb{F}_q . Ainsi :

$$\mathcal{C} = \{u \cdot G \mid u \in \mathbb{F}_q^k\}$$

La matrice G est appelée *matrice génératrice* de \mathcal{C} .

La matrice génératrice est sous forme systématique si

$$G = [I_k \mid M],$$

avec I_k est l'identité de $\mathbb{F}_q^{k \times k}$ et $M \in \mathbb{F}_q^{k \times (n-k)}$

Le codage est l'opération suivante :

$$\begin{aligned} \mathbb{F}_q^k &\rightarrow \mathcal{C} \subset \mathbb{F}_q^n \\ u &\mapsto u \cdot G \in \mathcal{C} \end{aligned}$$

Exemple du bit de parité :

Soit $\mathbb{K} = \mathbb{F}_2$. On transmet 3 bits d'information avec 1 bit de parité paire. C'est un code linéaire de matrice génératrice :

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

Matrice de contrôle

Définition 10 (Code dual)

Soit \mathcal{C} un code $[n; k]_q$. Le *code dual* \mathcal{C}^\perp de \mathcal{C} est l'ensemble des vecteurs $y = (y_1, \dots, y_n) \in \mathbb{F}_q^n$ tels que pour tout $x = (x_1, \dots, x_n) \in \mathcal{C}$

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = 0.$$

Le code dual \mathcal{C}^\perp a pour dimension $\dim(\mathcal{C}^\perp) = n - \dim(\mathcal{C})$.

Nous pouvons donc définir une matrice $H \in \mathbb{F}_q^{(n-k) \times n}$ dont les $n - k$ lignes sont les vecteurs d'une base de \mathcal{C}^\perp sur \mathbb{F}_q^n

$$\mathcal{C}^\perp = \{u \cdot H \mid u \in \mathbb{F}_q^{n-k}\}$$

La matrice H est appelée **matrice de contrôle** de \mathcal{C} .

Théorème 1

Soit \mathcal{C} un code $[n; k]_q$, G une matrice génératrice et H une matrice de contrôle. Alors on a

$$GH^T = 0.$$

Les lignes de H forment une base du noyau de G .

Syndrome

Définition 11

Soit \mathcal{C} un code $[n; k]_q$ et $H \in \mathbb{F}_q^{(n-k) \times n}$ sa matrice de contrôle. Le *syndrome* d'un mot $x \in \mathbb{F}_q^n$ est le mot de longueur $n - k$

$$s = x \cdot H^T.$$

Nous avons

$$\mathcal{C} = \{x \in \mathbb{F}_q^n \mid x \cdot H^T = 0\}$$

Les mots du code sont donc les mots dont le syndrome est nul.

Propriété 3

Soit \mathcal{C} un code $[n; k; d]_q$. Nous avons

-

$$\min_{\{u,v \in \mathcal{C}: u \neq v\}} (d_H(u, v)) = \min_{\{u \in \mathcal{C}: u \neq 0\}} (w_H(u))$$

- Soit H la matrice de contrôle de \mathcal{C}

$$d = \min\{s \in \mathbb{N}^* : \exists s \text{ colonnes de } H \text{ linéairement dépendantes}\}$$

- *Borne de Singleton* : $d \leq n - k + 1$.

Forme systématique

Définition 12

Un code est sous forme systématique si les bits d'information sont au début des mots.

Propriété 4

Soit \mathcal{C} un code $[n; k]_q$ sous forme systématique. La matrice génératrice G s'écrit alors sous la forme

$$G = [I_k \mid M],$$

avec I_k l'identité sur $\mathbb{F}_q^{k \times k}$ et $M \in \mathbb{F}_q^{k \times (n-k)}$.

Théorème 2

Soit \mathcal{C} un code sous forme systématique de matrice génératrice $G = [I_k \mid M]$. Alors on peut choisir une matrice de contrôle H sous la forme

$$H = [-M^T \mid I_{n-k}]$$

Définition 13

Deux codes \mathcal{C} et \mathcal{C}' de longueur n sont équivalents si on obtient \mathcal{C}' à partir de \mathcal{C} en réordonnant les vecteurs de base de \mathbb{F}_q^n .

Théorème 3

Tout code est équivalent à un code sous forme systématique.

Décodage par syndrome

Soit \mathcal{C} un code $[n; k; d]_q$, et $H \in \mathbb{F}_q^{(n-k) \times k}$ une matrice de contrôle. On transmet $c = u + e$, avec $u \in \mathcal{C}$ et une erreur $e \in \mathbb{F}_q^n$

- On calcule

$$s = c \cdot H^T = u \cdot H^T + e \cdot H^T = e \cdot H^T.$$

- On cherche ensuite $e' \in \mathbb{F}_q^n$ de poids minimum tel que :

$$s = e' \cdot H^T.$$

- On décode $c - e' \in \mathcal{C}$. Nous avons $c - e' = u$, si $w_H(e) \leq t$, avec t la capacité de correction du code.

- Calculer les syndromes s_i des mots e_i de poids inférieurs à la capacité de correction t du code.
- On les stocke dans une table d'association $[(s_i, e_i)]_{i=1}^{i=t}$, les syndromes corrigibles.
- Si on reçoit un mot $c \in \mathbb{F}_q^n$ de syndrome $s = c \cdot H^T$ non nul, il faut regarder si s est corrigible.
- soit $s \in \mathbb{F}_q^{n-k}$ un syndrome corrigible et $e \in \mathbb{F}_q^n$ le mot de poids minimum produisant s . Décodeur $c = u - e$.

- Nous avons vu les codes linéaires
- Une famille importante de codes linéaires est la famille des codes cycliques