MODÉLISATION MATHÉMATIQUE

Dossier : Codes correcteurs.

Introduction

Il est impossible d'envisager la transmission d'information sans prendre en compte la possibilité de corruption des données au cours de la chaine de transmission (appelé bruit). Il est donc obligatoire d'anticiper cette perte d'information dans la manière dont on transmet l'information si l'on veut être efficace. Il s'agit du but des codes détecteur et correcteur : coder l'information de telle manière qu'il soit possible à la reception de deviner s'il y a eu un problème de communication (et donc de redemander l'information), voir d'être capable de corriger le message sans requêtes supplémentaires.

Notation et vocabulaire

On suppose que l'information que l'on veut transmettre a déjà été numérisée, on veut donc transmettre un ensemble de bits.

Definition 1 : Un bloc de k bits sera appele bloc, mot ou vecteur. L'ensemble des mots de k bits sera vu avec une structure d'espace vectoriel sur un corps fini : $\{0,1\}^k = (\mathbb{Z}/2\mathbb{Z})^k = {\mathbb{F}_2}^k$. On parlera indifféremment de bits ou de lettres ; et un mot m formé des k bits $m_1, \ldots, m_k$ sera note $m_1 m_2 ... m_k$, ou éventuellement $\begin{pmatrix} m_1 & ... & m_k \end{pmatrix}$

Exercice 1 : Combien de mots de k bits existent-il ?

On va alors transformer cet ensemble de bits (de manière à toujours pouvoir retrouver le message d'origine), souvent pour rajouter une propriété de robustesse vis-à-vis de la transmission (détection ou correction d'erreurs)

Definition 2 : Un code est une application injective $\phi : {\mathbb{F}_2}^k \to {\mathbb{F}_2}^n$; le parametre k est appelé la dimension du code ϕ et le paramètre n est appele la longueur du code : on dit que ϕ est un code de paramètres (k,n). Si de plus pour tout mot m de ${\mathbb{F}_2}^k$, m est un préfixe de ϕ(m) (c'est a dire si l'application de ϕ consiste seulement a rajouter des bits ,dits de controle, pour assister à la communication efficace), on dira que ϕ est un code systematique.

L'ensemble $C = \{\phi(m), m \in {\mathbb{F}_2}^k\}$ est appelé l'image du code ϕ. Les éléments de C sont appelés les mots de code de ϕ (en opposition aux éléments originels de ${\mathbb{F}_2}^k$ qui sont appelés mots de source). Deux codes ayant la même image sont dits équivalents.

Exercice 2 : Si on a un code $\phi : {\mathbb{F}_2}^k \to {\mathbb{F}_2}^n$, comment peut-on s'en servir pour coder un long message $m \in {\mathbb{F}_2}^r$ (r très grand devant k) ?

1 Initiation au sujet : Code détecteurs

1.1 La répétition

La première idée qu'on pourrait avoir pour transmettre plus efficacement l'information, serait de doubler chacun des bits. Ainsi, (0) serait codé par (00), et $\begin{pmatrix}0 & 1 & 0 & 1 & 1\end{pmatrix}$ serait codé par $\begin{pmatrix}0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1\end{pmatrix}$
En supposant que dans un (court) message, au maximum une erreur de transmission a été commise, alors on pourrais le repérer et même dire où elle a été commise.

Exercice 3 : Comment serait codé le message $\begin{pmatrix}0 & 0 & 1 & 1 & 0\end{pmatrix}$?
Si vous recevez les mots de codes suivant, quels étaits les mots de sources (ou bien où est l'erreur) ? $\begin{pmatrix}0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1\end{pmatrix}$, $\begin{pmatrix}0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1\end{pmatrix}$, et $\begin{pmatrix}1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\end{pmatrix}$.
Quels sont les avantages et inconvénients de ce code ?
Ce code ne permet pas de corriger les erreurs. Pourquoi ? Comment le modifier pour que la correction devienne possible ?

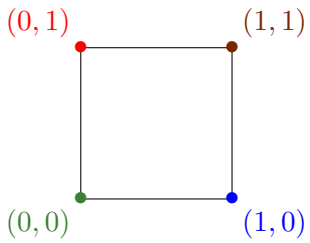
1.2 le bit de parité

Pour être plus efficace, on pourrait se dire qu'on n'a besoin que d'un bit pour marquer l'erreur. En effet, on peut rajouter au début du message un bit de telle manière qu'il y ait dans le message un nombre pair de bits ayant la valeur 1.

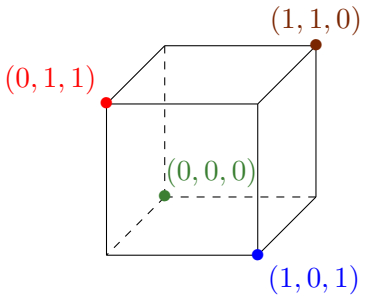
Avec ce codage, $\begin{pmatrix}0 & 0 & 1 & 1\end{pmatrix}$ est codé par $\begin{pmatrix}0 & 0 & 0 & 1 & 1\end{pmatrix}$, et $\begin{pmatrix}1 & 0 & 1 & 0 & 1\end{pmatrix}$ est codé par $\begin{pmatrix}1 & 1 & 0 & 1 & 0 & 1\end{pmatrix}$.

Alors comme la propriété "il y a un nombre pair de 1" n'est pas conservé s'il y a une erreur de transmission, on détectera encore une fois l'erreur, mais sans savoir la corriger.

La figure ci-dessous illustre bien cette propriété. Si un seul bit du code est altéré, on se déplace d'un sommet sur le cube en suivant une arête, et on obtient alors 3 bits qui ne correspondent pas à un message. On détecte donc facilement l'erreur. Mais si on a un message erroné, comme par exemple $\begin{pmatrix}1 & 1 & 1\end{pmatrix}$, on ne peut pas retrouver le message d'origine (sur la figure, un sommet qui ne correspond pas à un code d'un message a trois sommets voisins qui correspondent à trois messages differents).



Les messages à coder



Leurs codes

Exercice 4 : Comment serait codé le message $\begin{pmatrix}1 & 0 & 0 & 1 & 1 & 0\end{pmatrix}$?
Si vous recevez les mots de codes suivant, il y a t il une erreur ? $\begin{pmatrix}0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1\end{pmatrix}$, $\begin{pmatrix}0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1\end{pmatrix}$, et $\begin{pmatrix}1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\end{pmatrix}$.

Quels sont les avantages et inconvénients de ce code ?

2 Un exemple de code correcteur : Le code de Hamming C(7,4)

Le code de Hamming C(7,4) est un code permettant de transmettre un mot de 4 lettres $(u_1, u_2, u_3, u_4) \in \mathbb{Z}/2\mathbb{Z})^4$ avec une longueur de code 7. Encore une fois, ce code fait l'hypothèse que lors de la transmission, au plus une erreur sera produite.

Codage : Pour transmettre le mot (u_1, u_2, u_3, u_4), on va ajouter 3 bits de contrôles et envoyer le mot $(v_1, v_2, v_3, v_4, v_5, v_6, v_7)$ définit comme (avec addition dans $\mathbb{Z}/2\mathbb{Z}$) :

$$\begin{aligned} v_1 &= u_1, \\ v_2 &= u_2, \\ v_3 &= u_3, \\ v_4 &= u_4, \\ v_5 &= u_1 + u_2 + u_4, \\ v_6 &= u_1 + u_3 + u_4, \\ v_7 &= u_2 + u_3 + u_4. \end{aligned}$$

Ainsi, par exemple, $\begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix}$ est codé par $\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}$

Decodage : Le récepteur reçoit le mot à 7 lettres $(w_1, w_2, w_3, w_4, w_5, w_6, w_7)$, et calcule alors sa propre version des bits de contrôle (W_5, W_6, W_7) comme suit :

$$\begin{aligned} W_5 &= w_1 + w_2 + w_4, \\ W_6 &= w_1 + w_3 + w_4, \\ W_7 &= w_2 + w_3 + w_4. \end{aligned}$$

Ensuite, il se réfère à la liste qui suit :

- Si $W_5 = w_5$, $W_6 = w_6$, et $W_7 = w_7$ alors la transmission n'a eu aucune erreur,
- Si $W_5 \neq w_5$, $W_6 \neq w_6$, et $W_7 = w_7$, alors w_1 a été modifié par la transmission,
- Si $W_5 = w_5$, $W_6 \neq w_6$, et $W_7 \neq w_7$, alors w_3 a été modifié par la transmission,
- Si $W_5 \neq w_5$, $W_6 = w_6$, et $W_7 \neq w_7$, alors w_2 a été modifié par la transmission,
- Si $W_5 \neq w_5$, $W_6 \neq w_6$, et $W_7 \neq w_7$, alors w_4 a été modifié par la transmission,
- Si $W_5 \neq w_5$, $W_6 = w_6$, et $W_7 = w_7$, alors w_5 a été modifié par la transmission,
- Si $W_5 = w_5$, $W_6 \neq w_6$, et $W_7 = w_7$, alors w_6 a été modifié par la transmission,
- Si $W_5 = w_5$, $W_6 = w_6$, et $W_7 \neq w_7$, alors w_7 a été modifié par la transmission,

ce qui lui permet de retrouver le message d'origine.
Ainsi par exemple le message $\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$ se decode $\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}$

Exercice 5 : Comment serait codé le message $\begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix}$?

Si vous recevez les mots de codes suivant, quel était le message d'origine ? $\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$, et $\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$.

Ecrire matriciellement l'encodage.
(Plus compliqué) Ecrire une matrice $H \in M_{3,7}(\mathbb{Z}/2\mathbb{Z})$ telle que si $\overline{w}$ est un message, $H\overline{w} = 0$ s'il n'y a pas eu d'erreur de transmission, et si le bit en position numéro i a été modifié, alors $H\overline{w}$ est égal à la i^{me} colonne de H.
Quels sont les avantages et inconvénients de ce code ?

3 Travail demandé

On vous demande de traiter les points suivants. Lorsqu'il y a marqué "pour aller plus loin", ce qui suit est optionnel, mais on vous demande de traiter au moins un des points optionnel.
Les exercices dans le sujet sont là pour aider à la familiarisation avec ces concepts nouveaux.

- Vous approfondirez les notions présentées ici à l'aide des documents joints et de ce que vous pourrez trouver par vous-même, en particulier les définitions théoriques et propriété principales.
- Vous implémenterez le codage, insertion de bruit aléatoire et décodage (si possible) du bit de parité et du code de Hamming C(7,4). Pour aller plus loin, vous pourrez faire de même pour le code de Hamming C($2^k - 1$, $2^k - k - 1$), ou encore le code de Reed-Salomon C($2^m - 1$,k), voir [1].
- Pour aller plus loin, on pourra s'intéresser à la définition de la distance de Hamming ainsi que celle du poids de Hamming (qui ont une importance théorique). En particulier, on pourra ecrire un programme qui calcule le poids d'un mot, et un programme qui donne la distance de Hamming entre deux mots, voir [2]
- Pour aller plus loin, on pourra s'intéresser aux efficacités des divers codage en fonction de la probabilité qu'un bit soit transmis erroné. On étudiera en particulier le cas où la modification de deux bits different est indépendante et de même probabilité p, très petite ($p \leq \frac{1}{1000}$) et une étude numérique frequentielle pourra soutenir et illustrer un calcul théorique (par exemple celui proposé pour l'exercice 8 de [1]).
- Pour aller plus loin, on pourra s'intéresser aux conditions sur le cardinal de l'alphabet pour pouvoir généraliser les algorithmes vu précédemment (Peut-on au lieu d'utiliser que des 0 et des 1, utiliser tous les nombres (de 0 à 9) ? Tout les nombres et une lettre ? ...).
- Vous rédigerez un rapport présentant votre travail et préparerez une presentation orale de celui-ci.

Références

[1] Christiane Rousseau, Yvan Saint-Aubin, **Mathématiques et Technologie**, Springer, 2000, chapitre 6. (*version papier disponible au CDI, version pdf sur demande auprès des enseignants)*

[2] Pour les détails sur la distance de Hamming : https://fr.wikipedia.org/wiki/Distance_de_Hamming