Étude des codes de Hamming sur les corps finis

Andreas Pauper

1 Introduction et définitions

Dans tout le TIPE, si l'on se place sur un corps de cardinal N (où N est la puissance d'un nombre premier), le terme code fera référence à une partie de N^k, où k est un entier considéré comme la taille du code. Les éléments du codes sont en théorie des codes associés à des éléments d'un ensemble plus petit de façon à s'assurer que des messages stockés ou transmis puissent être conservés malgré les erreurs qui peuvent être rencontrées.

Le code correspond alors également à l'injection entre l'ensemble des messages de taille α (c'est-à-dire K^α) et la partie aussi appelée code $C \subset K^\beta$.

Au sein de la famille plus large des codes correcteurs, on s'intéressera ici uniquement à des codes linéaires, c'est-à-dire des codes C qui sont des sous-espaces vectoriels de K^β et où les fonctions de codages peuvent donc êtres linéaires, ce qui simplifie les calculs. Plus exactement si l'on note ϕ l'injection permettant de coder les messages de l'ensemble de départ K^α (ou K est un corps de cardinal p^k) dans l'ensemble d'arrivée C, ϕ est un code linéaire si et seulement si pour tous messages $\mathbf{m}$ et $\mathbf{m'}$ dans K^α, pour tous scalaires λ et μ dans K on a

$$\phi(\lambda\mathbf{m} + \mu\mathbf{m'}) = \lambda\phi(\mathbf{m}) + \mu\phi(\mathbf{m'})$$

Dire que le code est linéaire revient donc à dire que ϕ ainsi définie est K-linéaire. Nous utilisons ainsi dans le cadre des codes linéaires plusieurs fonctions qui sont linéaires. Ainsi si ϕ est une application linéaire de K^α dans un espace d'arrivée K^β, en représentant les mots $\mathbf{m}$ au départ et à $\mathbf{m}$ et $\mathbf{m'}$ à l'arrivée par la matrice ligne des coordonnées dans les bases canoniques des deux espaces, on peut représenter ϕ par une matrice Φ dans $\mathcal{M}_{\alpha,\beta}(K)$ avec :

$$\mathbf{m'} = \mathbf{m} \cdot \Phi^\intercal$$

Il existe de nombreux codes linéaires mais nous nous concentrerons ici sur les codes de Hamming, définis initialement sur le corps F_2. Nous étudierons ici les codes définis sur F_{2^α} pour des raisons que nous évoquerons d'ici peu.

Distance de Hamming La distance de Hamming permet de mesurer les différences entre deux mots, dans le cas où le corps d'étude est F_2, il s'agit du nombre de bits qui diffèrent. Si l'on se place sur K^α, la distance de Hamming

d entre deux messages $\mathbf{m} = [m_1 \cdots m_\alpha]$ et $\mathbf{m'} = [m'_1 \cdots m'_\alpha]$ de K^α est définie comme suit :

$$d(\mathbf{m}, \mathbf{m'}) = \mid \{0 \leq i \leq \alpha / m_i \neq m'_i\} \mid$$

Produit scalaire sur K^α Il est important pour la suite de définir un « produit scalaire » (il n'est pas défini) sur K^α, permettant entre autres de définir l'orthogonalité sur cet espace et de vérifier l'appartenance d'un mot $\mathbf{m}$ de K^α au code $C \subset K^\alpha$. En notant $\mathbf{m} = [m_1 \cdots m_n]$ et $\mathbf{m'} = [m'_1 \cdots m'_n]$ deux messages de K^α, on a :

$$\langle \mathbf{m}, \mathbf{m'} \rangle = \sum_{i=1}^{\alpha} m_i m'_i \tag{1}$$

2 Choix de p

Il a été dit plus tôt qu'un corps fini a pour cardinal une puissance d'un nombre premier p. Cependant en pratique le nombre premier 2 est toujours choisi. Cet entier premier est le seul permettant d'optimiser l'utilisation de la mémoire d'un ordinateur moderne à architecture binaire.

Démonstration. Attribuer une certaine quantité de mémoire pour représenter un élément du corps K de cardinal p^l revient à y attribuer un certains nombres de bits l'. Autrement dit, il serait souhaitable d'avoir $p^l = 2^{l'}$, c'est-à-dire qu'à chaque combinaison d'états de bits possible corresponde un unique élément de K. Alors $2|p^l$, ce qui n'est possible que si $p = 2$. □

Nous nous placerons dans le cas où $p = 2$ dans toute la suite du TIPE.

3 Représentation des erreurs

Pour qu'il y ait détection et correction d'erreurs, il faut d'abord qu'erreurs il y ait. Pour cela il convient de déterminer la probabilité qu'un bit soit erroné. Deux modèles seront étudiés ici : un premier modèle permet de représenter les erreurs affectant chaque bit individuellement et un autre permet de modéliser les cas de corruptions de plusieurs bits à la suite, les «bouffées d'erreurs», qui peuvent survenir par exemple lorsqu'on raye un CD. Nous considérons dans toute cette partie des messages de $l > 0$ bits.

3.1 Premier modèle : erreurs indépendantes

Considérons $\mathbf{m} = [m_1 \cdots m_l] \in (F_2)^l$, dans ce modèle les lettres m_i sont indépendantes deux à deux et pour $1 \leq i \leq l$, m_i suit une loi de Bernoulli de paramètre $0 < \rho < 1/2$.
En effet certains cas sont inutiles à considérer. Le cas $p = 0$ n'est pas intéressant

car il n'y aurait dans ce cas pas d'erreurs à corriger, le cas $p > 1/2$ est peu probable et pourrait se déduire en inversant tous les bits en plus de l'utilisation de codes de Hamming. Dans le cas où $p = 1/2$ il devient tout bonnement impossible de corriger les erreurs.

4 Construction et représentation de K

Si p est premier il est évident d'après le petit théorème de Fermat que $\mathbb{Z}/p\mathbb{Z}$ est un corps, cherchons maintenant à construire un corps de cardinal p^α avec α naturel non nul.