PUBLICATIONS MATHÉMATIQUES ET INFORMATIQUES DE RENNES

J. WOLFMANN

Évariste Galois et la planète Mars : introduction à la théorie algébrique du codage

Publications de l'Institut de recherche mathématiques de Rennes, 1985, fascicule 4 « Séminaires de mathématiques - science, histoire et société », , p. 123-147

<http://www.numdam.org/item?id=PSMIR_1985___4_123_0>

Evariste Galois et la planète Mars : Introduction à la théorie algébrique du codage.

J. WOLFMANN
G.E.C.T.
Université de Toulon
83130 LA GARDE

Cet exposé se situe dans la partie "algèbre effective" du colloque. On peut entendre par là qu'il traite d'une problèmatique ou l'essentiel n'est pas d'éclairer l'organisation d'objets mathématiques mais d'obtenir des résultats explicites, exprimables au moyen d'opérations élémentaires permettant une utilisation éventuelles pour des applications. Les mots sacrés de cette nouvelle liturgie en sont "algorithmes", "calculabilité", "complexité" etc... et les objets du culte en sont les ordinateurs. En fait il ne faut pas voir là une opposition factice entre des Mathématiques dites pures et d'autres qualifiées d'appliquées mais une nouvelle manière de poser des problèmes au Mathématiciens sous la poussée de l'informatique.
Le but de ce travail est de donner en exemple, à ce propos, la théorie algébrique des codes correcteurs, issue d'un problème technologique concernant la protection contre les erreurs lors d'une transmission de signaux ou d'information. Elle fait intervenir des outils mathématiques de plus en plus élaborés et tisse des liens avec différentes branches de l'algèbre finie et de la combinatoire.

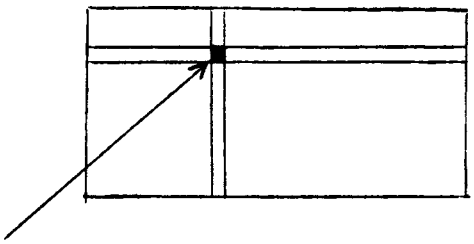
Cet exposé est un survol non exhaustif, ne contenant aucune démonstration. Il pose le problème en termes volontairement élémentaires puis décrit un certain nombre de questions algébriques qui en découlent. Il contient seulement quelques théorèmes significatifs et renvoie, pour un développement plus complet, aux ouvrages de la bibliographie.

I. LES CODES CORRECTEURS D'ERREURS

Le problème du codage correcteur d'erreurs

Le 19 Janvier 1972 la sonde spatiale "Mariner 9" transmettait une photographie d'une partie de "Grand Canyon" de la planète Mars. La très grande qualité de cette photo avait été obtenue en protégeant la transmission contre les erreurs éventuelles au moyen du "code de REED et MULLER d'ordre 1 et de longueur 32".

Pour transmettre une telle photo depuis la planète Mars on "discrétise" le problème de la manière suivante : au moyen d'un "pavage" suffisamment fin la photo est découpée en petits rectangles, chacun d'entre eux étant assimilé à un point (voir figure 4). Pour chacun de ces points on discerne une nuance de gris entre le blanc et le noir qui est caractérisé par un "niveau d'énergie". Il existe en tout 64 niveaux d'énergie.



repéré par un des 64 niveaux d'énergie

Figure 1.

On a donc besoin de 64 "messages" à transmettre représentant les différents "niveaux d'énergie". Pour des raisons technologiques, qui sont maintenant familières à tous, chaque message est représenté par une succession de 0 et de 1 (les "bits"). Dans le cas précis de cette photo chaque message contenait le même nombre de "bits" soit 32 et était obtenus de la manière suivante à une variante près : ce sont les 64 combinaisons linéaires modulo 2 des lignes de la matrice de la figure 2.
En d'autres termes ce sont les vecteurs du sous-espace de $\mathbb{F}_2^{32}$ (considéré comme espace vectoriel sur les corps de GALOIS $\mathbb{F}_2$ à deux éléments 0 et 1) engendré par les lignes de la matrice

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1
0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1
0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1

Figure 2

En indexant les éléments de $\mathbb{F}_2^{32}$ (qui peut être identifié, en tant que $\mathbb{F}_2$ - espace vectoriel, au corps de Galois $\mathbb{F}_2 32$) par les colonnes de la matrice de la figure 2 chaque vecteur du sous-espace considéré est le vecteur caractéristique d'un sous-ensemble de $\mathbb{F}_2^{32}$. Ces sous-ensembles sont l'ensemble vide $\mathbb{F}_2^{32}$, ainsi que tous les hyperplans affines de $\mathbb{F}_2^{32}$. En d'autres termes les fonctions caractéristiques de ces sous-ensembles sont les formes linéaires affines de $\mathbb{F}_2^{32}$.

Parlons maintenant du procédé qui fait que l'utilisation de ces messages binaires (nous disons dans la suite des "mots" et que leur ensemble est un "code") permet de se protéger contre les erreurs de transmissions.

Tout d'abord remarquons qu'il est, à priori, inutile d'utiliser à chaque fois 32 bits pour fabriquer 64 messages. Les 6-uples de $\mathbb{F}_2^6$ suffisent. La necessité d'utiliser plus de symboles que nécessaire est due à l'introduction de la "redondance" qui est l'un des maître-mots du codage correcteur d'erreurs. Comme dans le langage courant c'est la redondance qui permet de se protéger contre la perte d'information. C'est pourquoi les mots d'une langue sont suffisamment longs, ce qui permet qu'ils soient suffisamment différents les uns des autres, que l'on peut les reconstituer lors d'une écoute imparfaite. Impossible de confondre "feli..tations" pour félicitations avec "conged..ment" pour congédient. On enfonce même le clou en disant, par exemple, au téléphone : C comme Carole, O comme Odile, D comme Daniel, A comme Alfred, G comme Galois, E comme Eléonore, C.O.D.A.G.E. : Codage.

C'est le principe du langage des aviateurs qui font danser le FOXTROT ou le TANGO au PAPA de CHARLIE utilisant ainsi un ensemble de mots, le "code", suffisamment différents les uns des autres.
Examinons maintenant ce qu'on peut faire, dans cet ordre d'idée, avec des "0" et des "1". On supposera maintenant, et dans toute la suite, que les "mots" auront toujours le même nombre de symboles, la "longueur" du mot. Commençons par un mauvais exemple, celui de la Figure 3.

Messages	Code
a	0 0 0
b	0 1 0
c	0 0 1
d	1 1 0
e	1 0 1

Figure 3

Supposons que le mot envoyé est 0 0 1 et que le mot reçu est 0 1 1 c'est à dire avec une erreur dans la deuxième position.

Il est impossible de savoir quel est le message initial, même en sachant qu'il n'y a qu'une erreur, car 0 1 1 peut très bien provenir également de 0 1 0 avec une erreur dans la troisième position. La raison de cela est évidemment que 0 0 1 est trop voisin de 0 1 0.
Au contraire un bon exemple (pédagogique) est donné par la figure 4.
On a représenté 5 mots, C_1, C_2, C_3, C_4, C_5 de longueur 6 et, en dessous de chacun d'eux, tous les mots énoncés possibles qui peuvent en provenir en faisant une erreur.

C_1 : 1 1 0 0 1 1	C_2 : 0 1 0 1 0 0	C_3 : 0 0 0 1 1 1
0 1 0 0 1 1	0 0 1 0 1 1	1 0 0 1 1 1
1 0 0 0 1 1	0 0 0 1 0 0	0 1 0 1 1 1
1 1 1 0 1 1	0 1 1 1 0 0	0 0 1 1 1 1
1 1 0 1 1 1	0 1 0 0 0 0	0 0 0 0 1 1
1 1 0 0 0 1	0 1 0 1 1 0	0 0 0 1 0 1
1 1 0 0 0 1	0 1 0 1 0 1	0 0 0 1 1 0

C_4 : 1 0 1 1 0 1	C_5 : 0 1 1 0 0 1
0 0 1 1 0 1	1 1 1 0 0 1
1 1 1 1 0 1	0 0 1 0 0 1
1 0 0 1 0 1	0 1 0 0 0 1
1 0 1 0 0 1	0 1 1 1 0 1
1 0 1 1 1 1	0 1 1 0 1 1
1 0 1 1 0 0	0 1 1 0 0 0

Figure 4

On constate que les différents ensembles de mots erronnés sont deux à-deux disjoints. Il n'y a donc pas d'ambiguité possible dans le cas d'une seule erreur et l'on peut donc "corriger" le mot reçu par simple comparaison avec les mots du code C_1, C_2, C_3, C_4, C_5. Un seul d'entre eux diffère du mot reçu en une seule position. Dans l'exemple du code la planète Mars on peut corriger jusqu'a 7 erreurs sur un mot de longueur 32 en raison de considérations analogues. Un procédé courant, utilisé notamment dans les ordinateurs, est celui du "bit de parité" qui permet dans le cas d'une seule erreur, de "détecter" l'erreur, c'est à dire de savoir qu'il y a eu une erreur sans pour autant la corriger. Le principe en est le suivant :
On fabrique d'abord des messages sous forme de mots binaire de même longueur et on ajoute une nouvelle composante : la somme modulo 2 des composantes.

Par exemple 1 0 1 1 0 devient 1 0 1 1 0 1
ou encore 1 1 0 0 0 devient 1 1 0 0 0 0
la dernière composante est le "bit de parité".

Si, dans le cas d'une seule erreur possible, à la réception la somme modulo 2 de toutes les composantes n'est pas nulle (un nombre impair de "1") c'est qu'il y a erreur.

Une autre méthode consiste à envoyer deux fois de suite le même mot. Si on suppose qu'il n'y a eu qu'une seule erreur sur l'ensemble des deux mots la détection d'une erreur sera faite en constatant que les deux mots reçus sont différents. On pourra même corriger en envoyant trois fois de suite le même mot car si l'on suppose qu'il y a une seule erreur sur l'ensemble il y aura nécessairement un mot répété deux fois à la réception et ce sera le bon. Sur ce dernier exemple on voit immédiatement apparaitre un problème important : pour corriger plusieurs erreurs et pour des mots de grande longueurs ce procédé devient rapidement très onéreux en transmission. On cherchera donc, en plus de la possibilité de correction, à chercher à ce que celle ci soit la plus économique possible. Toute la thèorie du codage correcteur d'erreurs vise à généraliser les procédés élémentaires de "bit de parité" ou de répétition, et il est très surprenant de voir que, dans son développement, elle tisse des liens avec des branches des Mathématiques aussi abstraites que celle des groupes finis simples sporadiques ou la Géométrie algébrique.

Avant de passer à la formulation mathématique du problème il y a lieu de faire quelques remarques :

- On cherche à corriger des erreurs et il serait vain de penser pouvoir corriger un nombre quelconques d'erreurs. On se restreint donc à un nombre d'erreurs raisonnables indiqué expérimentalement par la statistique des erreurs lors de l'utilisation du canal de transmission considéré. Ainsi pour la photo de Mars la probalité de dépasser 7 erreurs par mot a été trouvée comme étant très faible ce qui permet d'obtenir un

bon décodage et donc une bonne photo.

- Les symboles, utilisés dans la pratique, pour fabriquer les mots sont toujours "0" et "1". Néanmoins il y a lieu de développer la théorie pour un "alphabet" quelconque car il y a de bons exemples ou l'on fabrique des mots utilisant les symboles non binaires (les codes de REED-SOLOMON ou l'alphabet est un corps de GALOIS différent de $\mathbb{F}_2$) chacun d'eux étant, pour la transmission, traduit en symboles binaires.

- Dans toute la suite les mots d'un même code auront la même longueur. De tels codes sont appelés des codes "en blocs" opposés aux codes "à longueur variable" ou "convolutionnnel" dont nous ne parlerons pas ici.

Formulation mathématique du problème du codage

1) Distance de Hamming

Soit A un ensemble fini, n un entier, $n \geq 1$. On définit une distance dans A^n (la distance de HAMMING) par :

Si $x = (x_1, x_2, \ldots, x_n) \in A^n$ et $y = (y_1, \ldots, y_n) \in A^n$

$$d(x,y) = \# \{i \in \{1, 2, \ldots, n\} | x_i \neq y_i\}$$

(nombre d'indices pour lesquelles les composantes correspondantes différent)

- Remarque : si dans le cas d'une transmission, x est le mot envoyé et y le mot reçu, $d(x,y)$ est le nombre d'erreurs.

2) Codes correcteurs

Soit C une partie de A^n

. A s'appelle l'alphabet
. C s'appelle un code de longueur n sur A
. Les éléments de C s'appellent les mots du code

Définition : Soit e un entier, $e \geq 1$

On dit qu'un code C corrige jusqu'à e erreurs si :

Pour tout x de A^n il existe au plus un mot c de C tel que $d(x,c) \leq e$

- Remarque : c'est cette condition qui permet de retrouver, sans ambiguité, un mot envoyé x à partir du mot reçu y si on a pas fait plus de e erreurs.

Proposition : (démonstration immédiate)

Les propriétés suivantes sont équivalentes

a) C corrige jusqu'à e erreurs

b) Les boules (pour la distance de Hamming) de rayon e, centrées sur les mots du code, sont deux à deux disjointes

c) Pour tout couple de mots distincts (x_1, x_2) de C^2 :

$$d(x_1, x_2) \geq 2e + 1$$

- Le plus grand nombre nombre e tel qu'un code donné C corrigé jusqu'à e erreurs est donné par la plus petite des distances entre mots distincts d'ou la :

Définition : - Un code est e-correcteur si

$$d_m = \inf_{\substack{x \in C,\ y \in C \\ x \neq y}} [d(x,y)] = 2e + 1 \text{ ou } 2e + 2$$

(ce qui est équivalent à $e = \left[\frac{d_m - 1}{2}\right]$ ou $[x]$ désigne la partie entière de x)

- Si un code est e-correcteur on dit que e est sa capacité de correction.

Les paramètres d'un code

n : longueur
N : nombre de mots
e : capacité de correction (ou d_m = distance minimum)

- Les problèmes du codage :

Trouver des codes avec le plus grand nombre de mots possibles, permettant de corriger le plus d'erreurs possibles et avec la plus petite longueur possible (! ! !).

Plus raisonnablement :

- Fabriquer des codes dont certains paramètres sont fixés en optimisant les autres.

- Quelles performances peut on atteindre dans ce domaine ?

- Comment construire les codes correspondants ?
- Problème du décodage : Soit C un code e-correcteur et y un mot de A^n dont la distance à C est au plus e. Comment trouver l'unique mot x de C tel que $d(x,y) \leq e$ autrement qu'en comparant y à tous les mots de C ? (bon algorithme)

- Exemple de problème non résolu :
  Trouver A(n,d) le plus grand nombre de mots d'un code sur A de longueur n et de distance minimum d)
  (on connait seulement des résultats partiels et des bornes)

Les codes linéaires

Dans le but de construire et d'étudier des codes il est naturel de choisir l'alphabet A comme possédant des propriétés connues, par exemple comme étant muni d'une structure algebrique. L'alphabet étant un ensemble fini on choisit pour A un corps de Galois, soit $A = \mathbb{F}_q$ avec q puissance d'un nombre premier p. Ce choix s'impose d'autant plus que l'alphabet usuel {0,1} a le bon goût d'être équipé de la structure du corps $\mathbb{F}_2$ pour les opérations modulo 2.

Poids d'un mot

- Définition : Si $x = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$ le poids de x, noté $w(x)$ est le nombre de ses composantes non nulles.

$$w(x) = \#\ \{i \in \{1, 2, \ldots, n\} \mid x_i \neq 0\}$$

- Cette définition s'impose en raison de la propriété suivante de démonstration immédiate :

(1) $\qquad d(x,y) = w(x-y)$

- Les sous-ensembles privilégiés de $\mathbb{F}_q^n$ sont les $\mathbb{F}_q$-sous espaces vectoriels. On introduit donc la définition suivante :

Définition d'un code linéaire

Soit $\mathbb{F}_q$ le corps de Galois de q éléments et n un entier, $n \geq 1$.
Une partie C est un code linéaire (n,k) sur $\mathbb{F}_q$.

Si C est un sous-espace vectoriel de dimension k de $\mathbb{F}_q^n$.

Propriétés :

a) Pour un code linéaire la distance minimum est le plus petit poids de ses mots non nul

$$d_m = \inf_{x \in C,\ x \neq 0} w(x)$$

b) Le nombre de mots d'un code linéaire (n,k) sur $\mathbb{F}_q$ est q^n.

- La démonstration de a) résulte immédiatement de la relation (1) et b) vient du fait que le code est isomorphe à $\mathbb{F}_q^k$

Exemple (pédagogique) donné par la figure 5

			poids
Base	1 0 0 1 1 0	a	3
	0 1 0 1 0 1	b	3
	0 0 1 1 1 1	c	4
	1 1 0 0 1 1	a+b	4
	1 0 1 0 0 1	a+c	3
	0 1 1 0 1 0	b+c	3
	1 1 1 1 0 0	a+b+c	4
	0 0 0 0 0 0		

Figure 5

Ce code est 1-correcteur car $d_m = \inf_{x \in C,\ x \neq 0} w(x) = 3$ (=2e+1 pour e=1)

Les différentes descriptions d'un code linéaire

Si C est un code (n,k) sur $\mathbb{F}_q$ il s'agit de réaliser effectivement une injection de $\mathbb{F}_q^k$ dans $\mathbb{F}_q^n$.

Pour ce faire on peut distinguer les différents objets suivants :
(les exemples se refèrent au code décrit ci-dessus)

Base $C_1, C_2, \ldots, C_n$ base de C.

les mots sont les $\sum_{i=1}^{k} \lambda_i C_i$

exemple : $C_1 = (1\ 0\ 0\ 1\ 1\ 0)$ $C_2 = (0\ 1\ 0\ 1\ 0\ 1)$ $C_3 = (0\ 0\ 1\ 1\ 1\ 1)$

Matrice génératrice

C'est une matrice G dont les lignes forment une base de C. Les mots sont alors de la forme :

$(a_1, a_2, \ldots, a_k) \times G$ avec $a_i \in \mathbb{F}_q$

Ceci revient à décrire C comme image de $\mathbb{F}_q^{\,k}$ dans $\mathbb{F}_q^{\,n}$ par l'application linéaire injective de matrice G^t par rapport aux bases naturelles.

Exemple :

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

Formes coordonnées

$L = \{l_1, l_2, \ldots, l_n\}$ ensemble de n formes linéaires de $\mathbb{F}_q^{\,k}$, de rang k tels que les mots sont de la forme

$m_a = (l_1(a), l_2(a), \ldots, l_n(a))$, $a \in \mathbb{F}_q^{\,k}$

Exemple : $l_1(a_1, a_2, a_3) = a_1$, $l_2(a_1, a_2, a_3) = a_2$, $l_3(a_1, a_2, a_3) = a_3$

$l_4(a_1, a_2, a_3) = a_1+a_2+a_3$, $l_5(a_1, a_2, a_3) = a_1+a_3$, $l_6(a_1, a_2, a_3) = a_2+a_3$

Sous ensemble de rang k de $\mathbb{F}_q^{\,k}$

Au moyen du produit scalaire usuel de $\mathbb{F}_q^{\,k}$ les formes coordonnés peuvent être identifiées aux colonnes d'une matrice génératrice G, les mots étant de la forme :

$m_a = (a\cdot\omega_1, a\cdot\omega_2, \ldots, a\cdot\omega_n)$ ou $a \in \mathbb{F}_q^{\,k}$ et

$\Omega = (\omega_1, \ldots, \omega_n)$ est l'ensemble des colonnes de G.

Avec cette description le code sera désigné par $C = C(\Omega)$

exemple : $\omega_1 = (1, 0, 0)$, $\omega_2 = (0, 1, 0)$, $\omega_3 = (0, 0, 1)$, $\omega_4 = (1, 1, 1)$

$\omega_5 = (1, 0, 1)$, $\omega_6 = (0, 1, 1)$.

Matrice de controle :

C peut être considéré comme le noyau d'une application linéaire de $\mathbb{F}_q^n$ dans $\mathbb{F}_q^k$ dont la matrice, par rapport aux bases naturelles est une matrice génératrice H de l'orthogonal de C pour le produit scalaire usuel. Une telle matrice est appelée matrice de controle de C et :

$$(x_1, x_2, \ldots, x_n) \in C \iff (x_1, \ldots, x_n) \times H^t = 0$$

Exemple : $H = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}$

Codage au moyen d'un code linéaire

Il est réalisé au moyen du schéma suivant :

$$\text{message} \longmapsto (a_1, a_2, \ldots, a_k) \longmapsto (a_1, a_2, \ldots, a_k) \times G = (x_1, \ldots, x_n)$$

codage non correcteur — codage correcteur — matrice génératrice — mot du code

On code tout d'abord chaque message a transmettre par un k-uple sur $\mathbb{F}_q$. L'entier k est déterminé de façon que q^k (nombre total de k-uples sur $\mathbb{F}_q$) soit au moins égal au nombre de message à transmettre. Ce codage n'est pas correcteur puisque le poids minimum du code constitué par $\mathbb{F}_q^k$ tout entier est 1.

En choisissant convenablement un code e-correcteur ou e est le nombre d'erreurs que l'on souhaite corriger ou réalise une injection de $\mathbb{F}_q^k$ dans $\mathbb{F}_q^n$ au moyen d'une matrice génératrice G comme indiqué précédemment.

Forme systèmatique

Si la matrice G est de la forme

$G = (I_k, M)$ ou I_k est la matrice identité d'ordre k alors l'injection de $\mathbb{F}_q^k$ dans $\mathbb{F}_q^n$ est réalisé par :

$$(a_1, \ldots, a_k) \longmapsto \underbrace{(a_1, \ldots, a_k}_{\text{Information}}, \underbrace{x_{k+1}, \ldots, x_n)}_{\text{redondance (ou controle)}}$$

On retrouve ici le procédé d'adjonction d'une redondance décrit dans le début du chapitre à l'image de "U comme Ursule".
On peut toujours se ramener à cette situation moyennant une permutation des colonnes d'une matrice génératrice ce qui ne change pas les propriétés de poids des mots donc la capacité de correction (code équivalent).

Mesures de l'efficacité relative d'un code linéaire

Pour un code linéaire (n,k) de poids minimum d le "cout" del'adjonction de redondance permettant de corriger jusqu'à e erreurs (d = 2e + 1 ou 2e + 2) est mesuré par les nombres :

$\frac{k}{n}$: rendement

$\frac{d}{n}$: taux de correction

On prouve aisément que :

$$\frac{k}{n} + \frac{d}{n} \leq 1 + \frac{1}{n} \quad \text{(Borne de SINGLETON)}$$

Dans le cas de l'égalité (d = n-k+1) on dit que le code est "Maximum distance séparable" (M.D.S)

Les problèmes sur les codes linéaires

D'une manière non exhaustive on peut distinguer les problèmes suivants :

1) Optimisation des paramètres

Deux des paramètres (longueur, dimension, poids minimum) étant fixés quel est la meilleure valeur possible pour le troisième ? (plus petite longueur, plus grande dimension, plus grand poids minimum).

On ne connait de résultats exacts pour des petites valeurs des paramètres et dans le cas général on de dispose que de bornes qui ne sont pas toujours très fines, et pour lesquels l'existence et la construction de codes correspondants est souvent un problème ouvert.

2) Les classes de "bons codes"

Une classe de "bons codes" est une classe $(C_i)_{i\in\mathbb{N}}$ de codes linéaires (n_i, k_i) de poids minimum d_i tels que :

$$\lim n_i = \infty,\ \lim \frac{k_i}{n_i} > 0 \text{ et } \lim \frac{d_i}{n_i} > 0$$

(les suites des rendements et des taux de correction ne tendent pas vers zéro).

La théorie de l'information de SHANNON prédit de manière non constructive, l'existence de tels classes mais on n'en connait que très peu actuellement.

3) Codes M.D.S.

Problème de l'existence et de la construction des codes M.D.S. sur $\mathbb{F}_q$

- Pour chaque q il existe des codes M.D.S. triviaux
  (codes (n,1) de poids n ou (n, n-1) de poids 2 ou (n, n) de poids 1)
- Pour q = 2 les codes M.D.S. sont triviaux
- Dans le cas général il existe une conjoncture sur les paramètres
  (voir [2] chap. 11)

4) Le problème du décodage

Si y est un mot reçu provenant du mot x d'un code C e-correcteur, et si il n'y a pas plus de e erreurs alors la seule boule de rayon e centrée sur un mot du code et contenant y est celle de centre x. Le décodage consiste donc à identifier le centre de la boule de rayon e, centrée sur un mot du code, et contenant y. On peut tout simplement calculer la distance de y à tous les mots du code mais ce procédé direct est rapidement onéreux lorsque la longueur et la dimension du code augmentent.

Le problème du décodage consiste à trouver des algorithmes de meilleure complexité que celui de cette comparaison exhaustive.

Les codes cycliques

Définition : Un code cyclique est un code linéaire C possédant la propriété suivante :

Si $(a_0, a_1, \ldots, a_{n-1}) \in C$ alors $(a_{n-1}, a_0, \ldots, a_{n-2}) \in C$ (invariance du code par permutation circulaire).

Représentation polynomiale

Soit θ l'application de $\mathbb{F}_q^n$ dans $\mathbb{F}_q[x]/x^n-1$

telle que :

$$\theta : a = (a_0, a_1, \ldots, a_{n-1}) \longmapsto a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} = a(x)$$

On obtient facilement :

- θ est un isomorphisme vectoriel
- C est un code cyclique si et seulement si $\theta(C)$ est un idéal de $\mathbb{F}_q[x]/x^n-1$

On en déduit également :

- Pour chaque code cyclique de longueur n sur $\mathbb{F}_q$ il existe un diviseur (sur $\mathbb{F}_q$) de x^n-1 et un seul, soit $g(x)$, tel que les éléments de $\theta(C)$ sont les multiples de $g(x)$ ($g(x)$ s'appelle le générateur de C).

Dans le cas ou n et q sont premier entre eux l'algèbre $\mathbb{F}_q[x]/x^n-1$ est semi-simple et on dispose de la théorie des idempotents primitifs pour décrire et utiliser les code cycliques.

Les codes B.C.H. (BOSE, CHAUDURI, HOCQUEMGHEM)

Le théorème suivant permet de fabriquer des codes cycliques corrigeant un nombre d'erreurs donné.

Théorème

Soit $g(x)$ le générateur d'un code cyclique de longueur n sur $\mathbb{F}_q$ et α une racine primitive du corps des racines n-ièmes de l'unité sur $\mathbb{F}_q$.

Si $g(x)$ admet pour racines

$\alpha^r, \alpha^{r+1}, \alpha^{r+2}, \ldots, \alpha^{r+\delta-2}$ avec r,δ entiers alors le poids minimum de C est au moins δ (voir une démonstration par exemple dans [2])

Les codes auto-duaux binaires

Un code $C(n,k)$ sur $\mathbb{F}_q$ est auto-dual sir $C = C^{\perp}$ (orthogonal de C pour le produit scalaire usuel de $\mathbb{F}_q^n$.

Dans le cas $q = 2$ (codes binaires) les codes auto-duaux offrent beaucoup d'intérêt pour, en particulier, les raisons suivantes :

- Il existe une-classe de bons codes linéaires binaires auto-duaux.

- Il existe une classe de bons codes linéaires binaires auto-duaux dont tous les poids sont des multiples de quatre.

- On peut associer à un éventuel plan projectif d'ordre 10 (voir dans la suite la partie "codes et combinatoire") un code auto-dual à poids multiples de quatre.

- Certains groupes finis simples sporadiques sont des groupes d'automorphismes (voir dans la suite "codes et groupes") de codes auto-duaux interessant.

- Les codes auto-duaux binaire à poids multiples de quatre sont les sous-espaces totalement singuliers maximaux d'une certaine forme quadratique sur $\mathbb{F}_2^k$.

. Il existe des bornes pour le poids minimum des codes auto-duaux binaires :

poids pairs : $d_{min} \leq 2\left[\frac{n}{8}\right] + 2$

poids multiples de quatre : $d_{min} \leq 4\left[\frac{n}{24}\right] + 4$

($[x]$ partie entière de x)

Si l'égalité est atteinte on dit que le code est "extremal".

Les problèmes :

- Construire effectivement une classe de bons codes auto-duaux, (dont seule l'existence est connue).

- Existence et construction de codes extrémaux. En particulier existe-t-il un code auto-dual à poids multiples de quatre sur $\mathbb{F}_2$ (72, 36) de poids minimum 16 ?

- Etude du code associé à l'éventuel plan projectif d'ordre 10.

- Approfondir le lien entre codes auto-duaux et groupes finis simples.

II. CODES ET GEOMETRIE DE GALOIS

La géométrie de Galois est celle qui est fabriqué classiquement sur un espace vectoriel sur un corps de Galois.

Différentes questions, évoquées dans ce qui précède, peuvent se formuler de la manière suivante :

Soit Ω une partie de $\mathbb{F}_q^k$

. Pour les poids :

- Soit $s \in \mathbb{N}$. Trouver Ω pour que s vecteurs quelconques distincts de Ω soient linéairement indépendants

- Déterminer la cardinal de l'intersection de Ω avec chaque hyperplan de $\mathbb{F}_q^k$

. Pour le groupe d'automorphisme : (voir la suite)

- Déterminer effectivement le stabilisateur de Ω dans le groupe linéaire GL (k,q).

. Pour le décodage "par permutations" (voir [2] chap 16)

- Soit $e \in \mathbb{N}$ et $I \subset \Omega$. Trouver un sous-ensemble $\mathcal{E}$ du stabilisateur de Ω tel que :

Pour tout sous-ensemble E de Ω de cardinal e il existe un élément de $\mathcal{E}$ qui envoie E à l'extérieur de I.

. Pour le rayon de recouvrement du code de Reed et Muller d'ordre 1

Le rayon de recouvrement d'un code C de longueur n sur A est le plus petit entier r tel que les boules (de Hamming) de rayon r centrées sur les mots du code recouvrent A^n.

Le code de Reed et Mullerd'ordre 1 de longueur 2^k, soit $R_1(k)$, est l'ensemble des mots binaires dont les fonctions caractéristiques associées sont les formes linéaire de $\mathbb{F}_2^k$ sur $\mathbb{F}_2$.

Le problème (non résolu pour k impair) de la détermination du rayon de recouvrement de $R_1(k)$ est équivalent à :

- Quel est le maximum du cardinal d'une partie Ω de $\mathbb{F}_2^k$ qui contient au plus la moitié des points de chaque hyperplan affine ?

III. CODES ET GROUPES

Si C est un code linéaire (n,k) sur $\mathbb{F}_q$ on introduit tout naturellement la notion suivante :

Définition : Une isométrie de C est un automorphisme vectoriel de C qui conserve les poids.

- On notera $\mathcal{I}(C)$ kes groupes des isométries de C

Exemples

- Le plongement habituel du groupe symétrique S_n, dans GL(n,q) détermine des isométries. Si $\sigma \in S_n$ et $\hat{\sigma}(x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ alors $\hat{\sigma}$ est une isométrie de C si $x \in C$ implique $\hat{\sigma}(x) \in C$

- Soit $u = (u_1, u_2, \ldots, u_n) \in \mathbb{F}_q^n$ avec u_i non nul pour chaque i et δ_u : $(x_1, x_2, \ldots, x_n) \longrightarrow (u_1x_1, u_2x_2, \ldots, u_nx_n)$ alors δ_u est une isométrie de C si $x \in C$ implique $\delta_u(x) \in C$.

Ces deux types d'isométries permettent de trouver toutes les autres :

Théorème (Mac WILLIAMS)

Toute isométrie de C est la restriction à C d'une application de la forme $\delta_u \circ \hat{\sigma}$

Groupe d'automorphisme d'un code linéaire

Définition : le groupe d'automorphisme du code linéaire C est le sous-groupe de S_n formé des σ tels que $\hat{\sigma}$ est une isométrie de C (on le note A(C)).

Indexation d'un code

Il peut être utile d'indexer les composantes de mots de longueur n, non pas par les entiers de 1 à n, mais par les éléments d'un ensemble E de cardinal n au moyen d'une bijection de $\{1, 2, \ldots, n\}$ dans E.

Définition

Une permutation Θ de E est un "automorphisme de C indéxé par E" si et seulement si

$\sigma = \varphi^{-1} \circ \Theta \circ \varphi$ est un automorphisme de C.

Cette définition permet d'énoncer le résultat suivant :

Théorème : Soit C un code linéaire (n,k) sur $\mathbb{F}_q$.

Si $C = C(\Omega)$ alors le groupe des automorphismes de C indexé par Ω est l'ensemble des restrictions des automorphismes vectoriels de $\mathbb{F}_q^k$ qui conservent Ω.

Remarques

- Le groupe des automorphismes de C indexé par Ω est une représentation linéaire de A(C) sur F_q de degré k

- La détermination de A(C) est équivalente à celle du stabilisateur, dans GL (k,q) d'une partie de $\mathbb{F}_q^k$.

Exemples de groupes finis égaux à (ou contenus dans) des groupes d'automorphismes de codes importants

- Groupe cyclique d'ordre fini
- GL(n,k)
- Groupes de MATHIEU M_{24}, M_{23}, M_{11}, M_{12}
- $PSL_2(p)$
- Groupe affine d'un corps de Galois
- Groupe orthogonal d'une forme quadratique sur un corps de Galois.

Remarque

Il existe un lien très interessant entre les codes, les réseaux (sous $\mathbb{Z}$ modules libres de R^m) et certain groupes finis simples sporadiques (Mathieu, Conway, Suzuki, Higman-Sims etc...)

On peut, à ce propos, consulter [8].

Les problèmes sur les groupes d'automorphismes de codes

- Quel est le groupe d'automorphisme d'un code donné ? (difficile)

- Trouver un code dont le groupe d'automorphisme est (ou contient) un groupe donné (pas facile si on cherche un code interessant).

Utilisations des groupes d'automorphismes

- Analyse des propriétés d'un code
- Décodage
- Interprétation des groupes finis (en particulier simples sporadiques)
- Représentation linéaires de groupes finis sur des corps de Galois.

Idéaux des algèbres de groupes

Les codes cycliques de longueur n sur $\mathbb{F}_q$, comme on l'a vu précédemment, peuvent se représenter comme des idéaux de $\mathbb{F}_q[x]/x^n-1$ qui s'identifie avec l'algèbre de groupe $\mathbb{F}_q[G]$ ou G est le groupe des entiers modulo n.

D'une manière analogue on peut chercher des codes interessants comme idéaux d'une algèbre $\mathbb{F}_q[G]$ ou G est un groupe abelien fini. Des résultats interessants, en particulier pour la construction de codes auto-duaux, ont été trouvés dans le cas ou $G = (\mathbb{F}_2 k, +)$ groupe additif d'un corps de Galois en caractéristique 2 et q puissance de 2 (voir par exemple [11]).

IV. CODES ET COMBINATOIRE

La combinatoire dont il s'agit ici est celle des configurations.

t-configurations (t-design)

Problème ouvert :

Existe-t-il des configurations pour $t \geq 6$?

- Les dernières 5-configurations trouvées proviennent de codes.

Plan projectif (fini) Un plan projectif est une 1-configuration particulière.

Définition : Un plan projectif est un ensemble P fini dont les éléments sont appellés "points" contenant des sous ensembles $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N$ appellés "droites" tel que :

1) Deux points distincts sont dans une droite et une seule.

2) Deux droites distinctes se coupent en un point.

3) P contient au moins 4 points 3 à 3 non dans une même droite.

Propriétés :

Nombre de points sur chaque droite : $n + 1$

Nombre de droites contenant un point : $n + 1$

Nombre de points : $n^2 + n + 1$

Nombre de droites : $n^2 + n + 1$

l'entier n s'appelle l'ordre du plan.

Matrice génératrice

C'est la matrice (a_{ij}) sur $\mathbb{F}_2$ dont les lignes sont indexée par les droites $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N$ et les colonnes par les points $P_1, P_2, \ldots, P_N$ telle que :

$a_{ij} = 1$ si $P_j \in \mathcal{D}_i$

$a_{ij} = 0$ si $P_j \notin \mathcal{D}_i$

Remarque

Les espaces projectifs de dimension 2 sur des corps de Galois $\mathbb{F}_q$ donnent des exemples de plans projectifs. L'ordre est alors q c'est à dire une puissance d'un nombre premier. On ne sait pas si il existe des plans projectifs d'ordre n non puissance d'un premier.

On a démontré l'inexistence pour n = 6 et le premier entier pour lequel le problème est ouvert est n = 10.

Le problème de l'existence d'un plan projectif d'ordre 10

A propos de l'existence d'un tel plan projectif on connait entre autres les résultats suivants :

- le groupe d'automorphisme (permutation conservant l'ensemble des droites) est réduit à l'identité (1981).
- Le sous-espace vectoriel de $\mathbb{F}_2^{111}$ engendré par les lignes de la matrice d'incidence est un code C linéaire (111, 56) sur $\mathbb{F}_2$ de poids minimum 11.
- Soit $\hat{C}$ le code étendu de C :

$$(a_1, a_2, \ldots, a_N, a_{N+1}) \in \hat{C} \iff \begin{cases} (a_1, a_2, \ldots, a_N) \in C \\ a_{N+1} = \sum_{i=1}^{N} a_i \end{cases}$$

$\hat{C}$ est un code auto-dual, de longueur 112, de dimension 56, de poids minimum 12 et à poids multiples de quatre.

Remarque : le fait que le groupe d'automorphisme du plan soit trivial fait dire à M. HALL (celui des groupes) que le seul objet, associé à un éventuel plan d'ordre 10, que l'on peut utiliser est le code qui lui est associé.

V. CODES ET GEOMETRIE ALGEBRIQUE

La géométrie algébrique est la dernière en date des branches des Mathématiques, introduite dans la théorie des codes. On utilise des courbes algébriques sur des corps de Galois pour définir les codes de GOPPA (généralisés) (voir [9] et l'exposé de MICHON).

Soit $\mathcal{C}$ une courbe algébrique projective sur $\mathbb{F}_q$ de genre g.

- On considère deux diviseurs sur $\mathcal{C}$:

$D = \Sigma\, m_i\, P_i$, chaque P_i étant rationnel

$G = \Sigma\, m_Q\, Q$ avec :

. $m_Q > 0$

. supp (G) $\cap$ supp (D) = $\emptyset$

G invariant par le groupe de Galois de $\overline{\mathbb{F}}$ (cloture algébrique) sur $\mathbb{F}$

- On définit les deux applications suivantes :

a) Φ_Ω de $\Omega(G-D)$ dans $\mathbb{F}_q^n$ telle que :

$\Phi_\Omega : \omega \longrightarrow (\mathrm{Res}_{P_1}(\omega) \ldots \mathrm{Res}_{P_n}(\omega))$

ou $\Omega(G-D)$ est l'espace des différentielles ω sur $\mathcal{C}$ telles que $\omega = 0$ ou $(\omega) \geq D$

et $\mathrm{Res}_{P_i}(\omega)$ est le résidu de ω en P_i

b) Φ_L de $L(G)$ dans $\mathbb{F}_q^n$ telle que :

$\Phi_L : f \longrightarrow (f(P_1) \ldots f(P_n))$

ou $L(G)$ est l'espace des fonctions rationnelles sur $\mathcal{C}$ telle que $f = 0$ ou $(f) \geq -D$

- Pour une courbe suffisamment régulière (voir [10]) on obtient les résultats suivants en utilisant le théorème de RIEMANN-ROCH.

Théorème : (codes géométriques de GOPPA)

1) Si $2g - 2 < \deg G \leq n + 2g - 2$ alors $C_\Omega = I_m(\Phi_\Omega)$

est un code linéaire (n,k) de poids minimum d avec :

$d \geq \deg G - 2g + 2$

$k \geq n - \deg G - 1 + g$ avec égalité si $\deg G < n$

2) Si $0 \leq \deg G < n$ alors $C_L = I_m(\Phi_L)$

est un code linéaire (n,k) de poids minimum d avec :

$d \geq n - \deg G$

$k \geq \deg G - g + 1$ avec égalité si $\deg G > 2g - 2$

3) Si $2g - 2 < \deg G < n$ alros C_L et C_Ω sont orthogonaux.

Remarques

- Ce théorème est important en particulier parce qu'il donne une borne inférieure du poids minimum des codes obtenus.
- C'est au moyen de tels codes qu'on a récemment construit, à partir de courbes modulaires, une classe de bon codes dépassant la borne asymtotique de

VARSHAMOV-GILBERT qui était, jusqu'alors, la meilleure borne connue.

- On déduit du théorème :

  $n - k + 1 - g \leq d$

  ce qui montre que, pour $g = 0$, les codes sont M.D.S.

Les problèmes de la construction des codes de GOPPA

- Trouver des points rationnels sur une courbe (difficile).
- Trouver effectivement une base de $\Omega(G-D)$ et une base de $L(G)$ au moyen d'un bon algorithme.
- Quel est le poids minimum exact des codes de GOPPA ?
- Trouver un algorithme de décodage.
- Le cas particulier $q = 2$.

BIBLIOGRAPHIE

Sur la théorie du codage

[1] BLAKE, I.F. and MULLIN, R.C.
"The Mathematical Theory of Coding"
ACADEMIC PRESS, NEW YORK (1975)

[2] Mac WILLIAMS, F.J. and SLOANE, N.J.A.
"The theory of error correcting codes"
NORTH HOLLAND. AMSTERDAM (1977)

[3] Mc ELIECE, R.J.
"The Theory of Information and Coding"
ENCYCLOPEDIA OF MATH. AND ITS APPL. 3
ADDISON-WESLEY, READING (1977)

[4] VANLINT, J.H.
"Coding Theory"
LECTURE NOTES IN MATH. Vol 201
SPRINGER VERLAG, BERLIN (1971)

[5] VAN LINT, J.H.
"Introduction to Coding Theory"
GRADUATE TEXTS IN MATH.
SPRINGER VERLAG, BERLIN (1982)

Sur les Codes et la Combinatoire

[6] BRIDGES, W.G., HALL, M.Jr, HAYDEN, J.L.
"Codes and Designs"
JOURNAL OF COMB. THEORY, Series A 31 (1981) 155-174

[7] Mac WILLIAMS, F.J., SLOANE, N.J.A., THOMPSON, J.G.
"On the existence of a projective plane of order 10"
J. COMB. THEORY Ser. A 14(1973) 66-78

Sur les codes et les groupes simples

[8] THOMPSON, J.G.
"From error correcting codes through sphere packing to simple groups"
THE CARUS MATHEMATICAL MONOGRAPH. MAS XIV WASHINGTON (1983)

Sur les codes et la géométrie algébrique

[9] GOPPA, V.D.
"Codes and Information"
RUSSIAN MATH. SURVEYS 39 : 1 (1984)

[10] LACHAUD, G.
"Les codes géométriques de GOPPA"
SEMINAIRE BOURBAKI, exposé n°641 (1985)

Sur la construction de codes auto-duaux

[11] WOLFMANN, J.
"A class of doubly even self dual binary codes"
A paraitre dans DISCRETE MATH.