Agrégation externe de mathématiques, session 2005
Épreuve de modélisation, option calcul scientifique : méthodes numériques et symboliques

---

(*527*) CODES CORRECTEURS D'ERREURS

Résumé : Ce texte décrit le code de Shannon en le présentant comme un tour de prestidigitation : trouver en sept questions un nombre compris entre 0 et 15, en permettant de mentir à l'une des questions.
Il présente aussi les définitions de base des codes (distance, etc.).

Thème applicatif, mots clefs : Codes correcteurs, code de Shannon.

► *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. Vous êtes laissé(e) libre d'organiser votre discussion comme vous l'entendez. Des suggestions de développement, largement indépendantes les unes des autres, sont proposées en fin de texte. Vous n'êtes pas tenu(e) de les suivre. Il vous est conseillé de mettre en lumière vos connaissances à partir du fil conducteur constitué par le texte. Le jury demande que la discussion soit accompagnée d'exemples traités sur ordinateur. Il est souhaitable que vous organisiez votre présentation comme si le jury n'avait pas connaissance du texte. Le jury aura néanmoins le texte sous les yeux pendant votre exposé.*

CODES CORRECTEURS D'ERREURS

La problématique des *codes correcteurs d'erreurs* est la suivante : un expéditeur A envoie un message m à B ; durant la transmission de ce message, des erreurs se produisent éventuellement, et B reçoit un message m' qui comporte peut-être des erreurs. Il s'agit de trouver comment faire pour que B,

1) d'une part détecte l'existence d'erreurs,

2) d'autre part, si elles ne sont pas trop nombreuses, sache les corriger.

Dans certains cas, lorsqu'il est rapide de réexpédier le message, 1) suffit. Néanmoins, dans d'autres cas, 2) s'avère indispensable : par exemple, si A est une fusée intersidérale qui envoie une photographie d'une météorite, et s'il faut par exemple une semaine pour que cette photographie atteigne B, à savoir la Terre, on ne pourra pas demander à la fusée de reprendre la même photographie de la météorite - qui sera bien loin de la fusée quinze jours après.

Le même cas de figure se produit pour un disque compact commercial : s'il possède quelques défauts, il est plus rentable d'avoir un procédé pour corriger les erreurs que de devoir le racheter.

L'explosion des technologies dites de l'information a rendu nécessaire l'élaboration de tels codes correcteurs performants. Le plus simple est celui qui consiste à affecter à un message m écrit en binaire, par exemple un octet, donc un élément $m = (m_1, \dots, m_8)$ de $\mathbf{F}_2^8$, où $\mathbf{F}_2$ désigne le corps fini $\mathbf{Z}/2\mathbf{Z}$, un neuvième bit, dit *bit de parité*, égal à $\sum m_i$. Il vaut donc 0 si le nombre des m_i égaux à 1 est pair, et 1 sinon.

L'expéditeur A envoie donc à B non pas le message m, mais un message $\overline{m}$ formé de m et d'informations supplémentaires. Dans le cas du bit de parité, $\overline{m} = (m, \varepsilon(m))$, donc $\overline{m}$ est un élément de l'hyperplan $\sum_{i=1}^{9} x_i = 0$ de $\mathbf{F}_2^9$.

Si le message m' reçu par B n'est pas dans cet hyperplan, il y a eu une erreur (au moins). Mais B est dans l'incapacité de corriger cette erreur ; en plus, il est tout-à-fait possible que l'erreur de transmission ait eu lieu sur le neuvième bit (le bit de parité), et qu'en fait les huit premiers bits aient été correctement transmis !

Nous décrivons dans le paragraphe suivant un tour de prestidigitation, qui illustre bien les différentes problématiques de la théorie des codes correcteurs d'erreurs.

1. Un tour de prestidigitation

Le prestidigitateur (B) demande à un spectateur (A) de choisir un nombre entre 0 et 15, et il lui pose sept questions. Le prestidigitateur pose sept questions à A, qui peut répondre par oui (1) ou non (0), et peut mentir au plus une fois.

Le prestidigitateur dit alors au spectateur s'il a menti, et dans ce cas à quelle question, et enfin annonce le nombre que le spectateur avait choisi :

PENSEZ À UN NOMBRE ENTRE 0 ET 15.	
Répondez aux questions suivantes :	(**Vous pouvez mentir une fois**.)
1) Est-il ≥ 8 ?	2) Est-il dans $\{4,5,6,7,12,13,14,15\}$?
3) Est-il dans $\{2,3,6,7,10,11,14,15\}$?	4) Est-il impair ?
5) Est-il dans $\{1,2,4,7,9,10,12,15\}$?	6) Est-il dans $\{1,2,5,6,8,11,12,15\}$?
7) Est-il dans $\{1,3,4,6,8,10,13,15\}$?	

Comment trouver le nombre

Notons $m_i = 1$ ou 0 suivant que la réponse à la i-ème question est oui ou non, et soit $m = (m_1,\ldots,m_7)$ le vecteur de $\mathbf{F}_2^7$ correspondant.

Soit k l'entier dont la représentation binaire est $\overline{cba}$, avec

$$a = m_1+m_3+m_5+m_7,\ \ b = m_2+m_3+m_6+m_7,\ \ c = m_4+m_5+m_6+m_7 \bmod 2.$$

Si $k = 0$, il n'y a pas eu de mensonge.

Sinon, la k-ième réponse est fausse. On corrige m en changeant la k-ième réponse, et la représentation en base 2 du nombre cherché est donnée par les quatre premières composantes de m.

Remarques.- 1) Il fallait au moins sept questions :

Six questions donnent en effet 2^6 choix, alors qu'on a $16\times(6+1) = 112$ possibilités ; comme $112 > 64$, six questions ne suffisent donc pas.

Par contre, le même calcul avec 7 questions donne $2^7 = 16\times(7+1)$: non seulement on a pu calculer la réponse correcte, mais il n'y a aucune information superflue !

2) Soit

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

Le vecteur $w = Hm$ est le k-ième vecteur colonne, qui n'est autre que k écrit en base 2.

Il est clair que ce jeu est un exemple de code correcteur d'erreurs : le spectateur envoie un message $(m_1,\dots,m_7)$ (formé du nombre compris entre 0 et 15, dont l'écriture en base deux est (m_1,m_2,m_3,m_4), à savoir les quatre premières questions, et de trois informations (m_5,m_6,m_7) supplémentaires ; le fait qu'il puisse mentir au plus une fois peut s'interpréter par le fait qu'il y a au plus une erreur, et dans ce cas B peut retrouver le nombre initial.

Pour montrer pourquoi ce tour fonctionne, nous allons d'abord donner quelques définitions générales sur les codes correcteurs.

Quelques définitions.

Soit k un corps fini (par exemple $\mathbf{F}_2=\mathbf{Z}/2\mathbf{Z}$), n un entier, et E le k-espace vectoriel k^n. Pour tout vecteur x de k^n, le *poids* $w(x)$ de x est le nombre de composantes non nulles de x.

La *distance de Hamming* $d : E\times E\to\mathbf{N}$ est définie par $d(x,y)=w(x-y)$. On vérifie aisément que c'est bien une distance.

Un *code* de longueur n à coefficients dans k est une partie C de E. Par définition, la *distance* de C est égale à $d(C)=\inf_{m,m'} d(m,m')$, où m et m' parcourent les couples d'éléments distincts de C.

Il est alors clair que, pour tout entier t tel que $2t+1\leq d(C)$, et pour tout élément $x\in E$, il existe au plus un élément $m\in C$ tel que $d(x,m)\leq t$. Pour un tel t, on dit que C est t correcteur-parfait si E est l'union disjointe des boules de rayon t centrées en les éléments de C.

Par ailleurs, si C est un sous-espace vectoriel de E, on dit que C est un *code linéaire*.

Codage de messages

Le lien de toutes ces définitions abstraites avec le problème pratique décrit dans l'introduction est le suivant :

Les messages m qu'envoie l'expéditeur A sont des éléments de C. Si C a comme distance d, B pourra donc corriger au plus $(d-1)/2$ erreurs. Le but de la théorie des codes correcteurs est donc de trouver des codes ayant simultanément une grande distance (afin de corriger le maximum d'erreurs) et une grande dimension (afin que le nombre de mots que l'on peut coder soit le plus grand possible).

Le tour de prestidigitation

Dans le cas du tour de prestidigitation, le code C est le sous-espace de dimension quatre de $\mathbf{F}_2^7$ noyau de l'application linéaire définie par la matrice H. On vérifie facilement qu'il est de distance 3, et qu'il est 1-correcteur parfait.

Suggestions pour le développement

▶ *Soulignons qu'il s'agit d'un menu à la carte et que vous pouvez choisir d'étudier certains points, pas tous, pas nécessairement dans l'ordre, et de façon plus ou moins fouillée. Vous pouvez aussi vous poser d'autres questions que celles indiquées plus bas. Il est très vivement souhaité que vos investigations comportent une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats.*

- Le candidat pourra expliquer le "tour de prestidigitation" décrit au paragraphe 1, notamment la méthode pour trouver le nombre. (Pourquoi donne-t-elle le nombre choisi ?)
- Il pourra également l'illustrer sur ordinateur (par exemple en faisant jouer le jury).
- Au sujet des définitions sur les codes, pourquoi la distance de Hamming est-elle effectivement une distance ? Quelles sont les distances possibles des codes de dimension 1, de codimension 1 ?
- On pourra montrer que, si $C \subset k^n$ est un code t-correcteur, on a
$$\mathrm{Card}\, B(0,t) \times \mathrm{Card} C \leq \mathrm{Card}\, k^n,$$
avec égalité si de plus C est t-correcteur parfait.
- On pourra prouver que, si C est t-correcteur parfait, sa distance est égale à $2t+1$.
- On pourra établir l'inégalité $\mathrm{Card} C \leq \mathrm{Card}\, k^{n-d+1}$, où d est la distance du code C. Cette inégalité peut-elle être une égalité dans le cas où C est un code linéaire de dimension et codimension ≥ 2 ?
- On pourra, pour un code de distance ≥ 3, indiquer un algorithme permettant de corriger une erreur, et le programmer.