Rayon de recouvrement de codes binaires non linéaires

On covering radius of non-linear binary codes

Antoine-Christophe LOBSTEIN

École Nationale Supérieure des Télécommunications, Département Systèmes et Communications, 46, rue Barrault, 75634 PARIS CEDEX 13

Étude des sacs à dos; étude du rayon de recouvrement de codes binaires, linéaires et non linéaires, notamment étude de $K(n, t)$, qui représente le nombre minimal de mots que peut contenir un code binaire de longueur n ayant t pour rayon de recouvrement [détermination de la valeur de $K(n, t)$ – ou de bornes inférieure et supérieure – pour différentes valeurs de n et t, lien entre $K(n+2, t+1)$ et $K(n, t)$...].

RÉSUMÉ

Le rayon de recouvrement d'un code est le plus petit entier t tel que tout vecteur de l'espace soit à distance au plus t du code.

Soit $K(n, t)$ le nombre minimal de mots que peut contenir un code binaire de longueur n ayant un rayon de recouvrement égal à t.

Nous démontrerons que :

$$K(7,2)=7, \qquad K(8,2)\leqq 12$$

et plus généralement :

$$K(2p+3, p)=7, \qquad K(2p+4, p)\leqq 12, \qquad \forall p\geqq 2.$$

MOTS CLÉS

Codes binaires, rayon de recouvrement.

SUMMARY

The covering radius of a code is defined as the smallest integer t such that all vectors in the space are within distance t of some codeword.

Let $K(n, t)$ be the minimum number of codewords for a binary code of length n with covering radius equal to t.

We shall prove that:

$$K(7,2)=7, \qquad K(8,2)\leqq 12$$

and more generally:

$$K(2p+3, p)=7, \qquad K(2p+4, p)\leqq 12, \qquad \forall p\geqq 2.$$

KEY WORDS

Binary codes, covering radius.

TABLE DES MATIÈRES

Introduction

F_2^n désignera l'ensemble des vecteurs binaires de longueur n.

Définition 1 : *Soit C un code binaire de longueur n. Le rayon de recouvrement t du code C peut être défini des deux manières suivantes :*

$$t = \operatorname{Max}_{x \in F_2^n} d(x, C)$$

ou

$$t = \operatorname{Min}_{r \in N} (r / \bigcup_{c \in C} S(c, r) \supseteq F_2^n).$$

La distance considérée ici est la distance de Hamming. $S(c, r)$ désigne la sphère de centre c, de rayon r.

On dira qu'un mot c de C couvre un mot x de F_2^n si $d(x, c) \leqq t$.

Un code C de longueur n, contenant K mots, sera donné par un tableau de n colonnes et K lignes constitué par les mots du code écrits les uns en dessous des autres.

Nous nous intéresserons au problème suivant : étant donnés n et t entiers, déterminer $K(n, t)$ qui désigne le nombre minimal de mots que doit contenir un code de longueur n pour avoir un rayon de recouvrement égal à t.

La complexité supposée de ce problème fait penser qu'il ne sera résolu que pour de petites valeurs de n [6].

Donnons sans démonstration quelques propriétés relatives au rayon de recouvrement :

Propriété 1 :

$$\forall n_1, n_2, t_1, t_2 \in N,$$
$$K(n_1+n_2, t_1+t_2) \leqq K(n_1, t_1) . K(n_2, t_2)$$

et ses conséquences :

$$\forall n, t \in N,$$
$$K(n+1, t) \leqq 2\, K(n, t),$$
$$K(n+1, t+1) \leqq K(n, t).$$

Propriété 2 :

$$\forall n, t \in N, \quad K(n, t) \geqq \frac{2^n}{\sum_{i=0}^{t} \binom{n}{i}}.$$

Cette borne est appelée « sphere covering bound ».

1. Codes binaires non linéaires

Ce chapitre est essentiellement consacré au calcul de $K(n, t)$ pour certaines valeurs de n et t.

Rappelons que l'on dispose d'une borne inférieure égale à :

$$\frac{2^n}{\sum_{i=0}^{t} \binom{n}{i}}.$$

Une borne supérieure peut être donnée par un code linéaire par exemple.

Notre intention est d'améliorer ces deux bornes.

Dans ce but, donnons le lemme suivant :

Lemme 1 : *Tout code de longueur $n+2$, ayant un rayon de recouvrement égal à $t+1$ et contenant strictement moins de $K(n, t)$ mots vérifie la propriété suivante :*

Deux quelconques de ses colonnes contiennent au moins une fois chacun des quatre couples 00, 01, 10 *et* 11.

Démonstration du lemme 1 : Soit C un code binaire de longueur $n+2$, de rayon de recouvrement $t+1$, et contenant K mots, avec $K < K(n, t)$.

Supposons que deux colonnes de C (on supposera que ce sont les deux premières) ne contiennent pas le couple ij ($i=0$ ou 1, $j=0$ ou 1).

Appelons C_1 (respectivement C_2) le code de longueur 2 (respectivement n) constitué des deux premières (respectivement n dernières) colonnes de C. Alors $t(C_1) \geqq 1$ car C_1 ne contient pas tous les mots de F_2^2. Et $t(C_2) \geqq t+1$ car C_2, code de longueur n, contient strictement moins de $K(n, t)$ mots.

Donc $t(C) \geqq t+2$ et on aboutit à une contradiction.

Donnons enfin une définition :

Définition 2 : *Un code* C *contenant* K *mots est dit équilibré si et seulement si :*

– lorsque K *est pair chacune de ses colonnes contient* K/2 *fois la valeur* 0 *et* K/2 *fois la valeur* 1;

– lorsque K *est impair chacune de ses colonnes contient* (K+1)/2 *fois la valeur* 0 *et* (K−1)/2 *fois la valeur* 1, *ou l'inverse.*

1.1. Le cas $t=1$

Il a été étudié par R. G. Stanton, J. G. Kalbfleisch [9, 10], N. J. A. Sloane (communication personnelle), et moi-même.

Il est notamment établi que $K(5,1)=7$ et $K(6,1)=12$.

Par exemple, les codes suivants :

$$C_{5,7,1}=\begin{matrix}00000\\00001\\01110\\10000\\11101\\11011\\10111\end{matrix},$$

$$C_{6,12,1}=\begin{matrix}000000\\101000\\011000\\110100\\001110\\110010\\110001\\111111\\010111\\100111\\001011\\001101\end{matrix},$$

ont un rayon de recouvrement égal à 1, et ont respectivement 7 et 12 mots.

1.2. Le cas $t=2$

Nos deux principaux résultats sont les suivants :

Théorème 1 :

$$\boxed{K(7,2)=7.}$$

Théorème 2 :

$$\boxed{9\leqq K(8,2)\leqq 12.}$$

Démonstration du théorème 1 : On sait que $5\leqq K(7,2)\leqq 8$.

Le code de longueur 7 suivant :

$$C_{7,7,2}=\begin{matrix}00000\ 00\\00001\ 00\\01110\ 00\\10000\ 11\\11101\ 11\\11011\ 11\\10111\ 11\end{matrix},$$

contient 7 mots et a 2 pour rayon de recouvrement, ce qui montre que $K(7,2)\leqq 7$.

Montrons maintenant que $K(7,2)>6$. Pour cela supposons que C est un code binaire de longueur 7, contenant 6 mots, et ayant un rayon de recouvrement égal à 2.

Première étape : Montrons que C est équilibré (toutes ses colonnes contiennent 3 fois la valeur 0 et 3 fois la valeur 1).

Comme $K(5,1)=7$ on peut appliquer le lemme 1 à C. Il en découle que toute colonne de C contient au moins 2 fois la valeur 0 et 2 fois la valeur 1.

(*a*) Supposons que C admette deux colonnes déséquilibrées, les deux premières, qui contiennent 2 fois la valeur 0 et 4 fois la valeur 1 (ceci sans perte de généralité). Les cinq autres colonnes sont équilibrées ou non. D'après le lemme 1, on a forcément (aux permutations de lignes près) la configuration suivante sur les deux premières colonnes de C :

$$\begin{matrix} & 1&2&3&4&5&6&7\\ c_1 & 0&0&.&.&.&.&.\\ c_2 & 0&1\\ c_3 & 1&0\\ c_4 & 1&1\\ c_5 & 1&1\\ c_6 & 1&1 \end{matrix}$$

$$c_i\in F_2^7 \quad \text{pour} \quad i=1,2,\ldots,6.$$

Soit S l'ensemble des mots de F_2^7 dont les deux premières composantes sont 0. $|S|=2^5=32$.

Or c_1 peut couvrir au maximum :

$$1+5+\binom{5}{2}=16 \text{ mots de S.}$$

c_2 et c_3 peuvent couvrir au maximum :

$$2(1+5)=12 \text{ mots de S.}$$

c_4, c_5 et c_6 peuvent couvrir au maximum :

$$3.1=3 \text{ mots de S.}$$

Donc au maximum 31 mots de S peuvent être couverts par C, ce qui montre qu'il est impossible que C ait deux colonnes déséquilibrées.

(*b*) Supposons que C admette une seule colonne déséquilibrée, la deuxième, qui contient 2 fois 0 et 4 fois 1 (ceci sans perte de généralité). Les six autres colonnes sont équilibrées. D'après le lemme 1, on a forcément (aux permutations de lignes près) la configuration suivante sur les deux premières colonnes de C :

1 2 3 4 5 6 7
0 0
1 0
0 1
1 1
0 1
1 1

On peut poser sans perte de généralité $c_1 = 0000000$ (ce que nous noterons par la suite $c_1 = \bar{0}$). Comme d'après le lemme 1 les colonnes 2-3, 2-4, ..., 2-7 doivent contenir 01 on en déduit que $c_2 = 1011111$:

1 2 3 4 5 6 7
0 0 0 0 0 0 0
1 0 1 1 1 1 1
0 1
1 1
0 1
1 1

Comme les colonnes 3 à 7 doivent être équilibrées, elles doivent être complétées avec les six quadruplets
1 1 1 0 0 0
1 0 0 1 1 0
0' 1' 0' 1' 0' 1'
0 0 1 0 1 1

Cependant le quadruplet $\begin{matrix}0\\1\\0\\1\end{matrix}$ ne peut pas être utilisé pour compléter une colonne, car cette colonne serait alors identique à la première, et en conséquence, ces deux colonnes ne contiendraient pas le couple 10 (ou 01), et C ne vérifierait pas le lemme 1. De même on ne peut pas utiliser le même quadruplet pour compléter deux des cinq dernières colonnes. On est donc obligé d'utiliser cinq quadruplets, au plus une fois, pour ces cinq colonnes, c'est-à-dire que chacun sera utilisé une et une seule fois. Aux translations et permutations de lignes et de colonnes près, on obtient donc forcément :

$$C = \begin{matrix}0000000\\1011111\\0111100\\1110010'\\0101011\\1100101\end{matrix}$$

Or ce code a un rayon de recouvrement supérieur ou égal à 3 : le mot 0000111 en est à distance 3. On a donc établi que nécessairement C est un code équilibré.

Deuxième étape : Nous allons montrer que C ne peut pas contenir sur quatre quelconques de ses colonnes et trois quelconques de ses lignes la configuration
0000
1111. Pour cela, supposons que C présente cette confi-
1111
guration sur les quatre premières colonnes de ses trois premiers mots :

1 2 3 4 5 6 7
0 0 0 0
1 1 1 1
1 1 1 1

Pour être équilibrées, les quatre premières colonnes de C doivent être alors complétées par les trois triplets
1 0 0
0, 1, 0.
0 0 1
Mais un triplet ne pouvant être utilisé 2 fois pour compléter deux colonnes, cela est impossible. De même C ne peut contenir sur quatre quelconques de ses colonnes et trois quelconques de ses lignes la
1111
configuration 0000.
0000
Il en résulte des conséquences intéressantes, si l'on pose que C contient $\bar{0}$:

si C contient 2 mots de poids 5, ces 2 mots ne peuvent être à distance 2 l'un de l'autre;

si C contient 1 mot de poids 6, il ne contient aucun mot de poids 5, et aucun autre mot de poids 6;

si C contient 1 mot de poids 2, il ne contient aucun mot de poids 4 (en effet, d'après la deuxième étape, si C contient le mot $\bar{0}$, 1 mot de poids 2 et 1 mot de poids 4, alors au moins 2 colonnes de C présentent
0 0
la configuration 0 0.
0 0
On ne peut plus alors compléter ces deux colonnes pour qu'à la fois elles soient équilibrées et vérifient le lemme 1.)

Troisième étape : Soit $A_0, A_1, \ldots, A_7$ le nombre de mots de poids 0, 1, ..., 7 figurant dans C.

$A_1, A_2, \ldots, A_7$ doivent nécessairement vérifier les inégalités suivantes (on peut prendre sans perte de généralité $A_0 = 1$, ce qui assure que les mots de poids 0, 1 et 2 sont couverts) :

$$\begin{gathered}15A_1+5A_2+13A_3+4A_4+10A_5 \geqq 35,\\ 10A_2+4A_3+13A_4+5A_5+15A_6 \geqq 35,\\ 6A_3+3A_4+11A_5+6A_6+21A_7 \geqq 21,\\ 3A_4+2A_5+7A_6+7A_7 \geqq 7,\\ A_5+A_6+A_7 \geqq 1.\end{gathered}$$

Ces contraintes expriment la couverture des mots de F_2^7 de poids i par les mots de C de poids $i-2$, $i-1$, i, $i+1$ et $i+2$ (pour $i = 3, 4, 5, 6, 7$), de la manière la plus simple possible, sans tenir compte d'éventuelles intersections.

Comme le code C contient 6 mots et est équilibré, on a de plus :

$$A_1+A_2+A_3+A_4+A_5+A_6+A_7=5,$$
$$A_1+2A_2+3A_3+4A_4+5A_5+6A_6+7A_7=21$$

et enfin :

$$0\leqq A_i\leqq 5 \qquad (i=1, 2, 3, 4, 5, 6),$$
$$0\leqq A_7\leqq 1.$$

Ce système, simple à programmer, donne très rapidement les sept octuplets de valeurs suivants pour A_0, A_1, ..., A_7 :

	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
1....	1	0	0	1	2	2	0	0
2....	1	0	0	2	1	1	1	0
3....	1	0	0	3	0	0	2	0
4....	1	0	1	0	1	3	0	0
5....	1	0	1	1	0	2	1	0
6....	1	1	0	0	1	2	1	0
7....	1	1	0	1	0	1	2	0

Remarquons que le même système, calculé sans la contrainte $\sum_{i=1}^{7} iA_i=21$, imposée par le fait que C est équilibré, donnerait 28 solutions au lieu de 7. La quatrième étape consiste maintenant à éliminer ces 7 cas.

Quatrième étape : Les cas n^{os} 2, 3, 4, 5, 6 et 7 sont éliminés en vertu des conséquences de la deuxième étape.

Éliminons maintenant le cas n° 1 :

$$A_0=1, \quad A_1=0, \quad A_2=0, \quad A_3=1,$$
$$A_4=2, \quad A_5=2, \quad A_6=0, \quad A_7=0.$$

D'après la deuxième étape, les 2 mots de poids 5 sont à distance 4 l'un de l'autre. On peut donc poser, sans perte de généralité :

1 2 3 4 5 6 7
0 0 0 0 0 0 0
1 1 1 1 1 0 0
1 1 1 0 0 1 1

Pour être équilibrées, les trois premières colonnes doivent être complétées par les trois triplets
1 0 0
0, 1 et 0.
0 0 1
Chacun ne pouvant être utilisé qu'une fois, ils le sont tous une et une seule fois. On a donc :

1 2 3 4 5 6 7
0 0 0 0 0 0 0
1 1 1 1 1 0 0
1 1 1 0 0 1 1
1 0 0
0 1 0
0 0 1

Les mots c_4, c_5 et c_6 étant interchangeables pour le moment, et les colonnes 6-7 devant être complétées avec les couples 11,01 et 10 pour être équilibrées et vérifier le lemme 1, on peut prendre sans perte de généralité :

1 2 3 4 5 6 7
0 0 0 0 0 0 0
1 1 1 1 1 0 0
1 1 1 0 0 1 1
1 0 0 1 0
0 1 0 0 1
0 0 1 1 1

De même il reste à mettre en colonnes 4-5 les couples 01, 10 et 11. c_6 ne pouvant être de poids 5, on peut lui attribuer 10 en colonnes 4-5, puis (c_4 et c_5 étant encore interchangeables) 11 pour c_4 et 01 pour c_5.

Finalement, on a de manière unique (aux translations et permutations de lignes et colonnes près) :

1 2 3 4 5 6 7
0 0 0 0 0 0 0 c_1
1 1 1 1 1 0 0 c_2
C=1 1 1 0 0 1 1 c_3
1 0 0 1 1 1 0 c_4
0 1 0 0 1 0 1 c_5
0 0 1 1 0 1 1 c_6

Or ce code a un rayon de recouvrement supérieur ou égal à 3 : le mot 0010110 en est à distance 3.

Conclusion : $K(7,2)>6$ et le théorème 1 est démontré.

Ce théorème s'inscrit dans le cadre d'un résultat plus général :

$$\forall p\geqq 1, \quad K(2p+3, p)=7 \quad (\textit{voir} \text{ th. 5}).$$

Démonstration du théorème 2 : On sait que $7\leqq K(8,2)\leqq 16$.

Le code de longueur 8 suivant :

$$C_{8,12,2}=\begin{matrix}000000\ 00\\101000\ 00\\011000\ 00\\110100\ 00\\110010\ 00\\001110\ 00\\110001\ 11\\111111\ 11\\010111\ 11\\100111\ 11\\001011\ 11\\001101\ 11\end{matrix},$$

contient 12 mots et a 2 pour rayon de recouvrement, ce qui montre que $K(8,2)\leqq 12$.

Ce résultat s'inscrit dans le cadre d'un résultat plus général :

$$\forall p \geqq 1, \quad K(2p+4, p) \leqq 12 \quad (\textit{voir} \text{ th. } 6).$$

Les inégalités que doivent vérifier A_1, A_2, ..., A_8 (on prendra à nouveau $A_0 = 1$) s'écrivent :

$$\begin{gathered}
21A_1 + 6A_2 + 16A_3 + 4A_4 + 10A_5 \geqq 56, \\
15A_2 + 5A_3 + 17A_4 + 5A_5 + 15A_6 \geqq 70, \\
10A_3 + 4A_4 + 16A_5 + 6A_6 + 21A_7 \geqq 56, \\
6A_4 + 3A_5 + 13A_6 + 7A_7 + 28A_8 \geqq 28, \\
3A_5 + 2A_6 + 8A_7 + 8A_8 \geqq 8, \\
A_6 + A_7 + A_8 \geqq 1.
\end{gathered}$$

Si on y ajoute :

$$\begin{gathered}
A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 = 7, \\
0 \leqq A_i \leqq 7 \qquad (i = 1, 2, ..., 7), \\
0 \leqq A_8 \leqq 1,
\end{gathered}$$

l'ensemble des solutions est vide. On a donc $K(8,2) > 8$.

Si on y ajoute :

$$\begin{gathered}
A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 = 8, \\
0 \leqq A_i \leqq 8 \qquad (i = 1, 2, ..., 7), \\
0 \leqq A_8 \leqq 1,
\end{gathered}$$

on trouve 86 solutions : $K(8,2) = 9$ est possible.

Le théorème 2 est ainsi démontré.

1.3. Le cas $t = 3$

Nos deux principaux résultats sont les suivants :

Théorème 3 :

$$\boxed{K(9,3) = 7.}$$

Théorème 4 :

$$\boxed{7 \leqq K(10,3) \leqq 12.}$$

Démonstration du théorème 3 : On sait que $4 \leqq K(9,3) \leqq 8$.

Le code de longueur 9 suivant :

$$C_{9,7,3} = \begin{matrix} 00000\ 0000 \\ 00001\ 0000 \\ 01110\ 0000 \\ 10000\ 1111, \\ 11101\ 1111 \\ 11011\ 1111 \\ 10111\ 1111 \end{matrix}$$

contient 7 mots et a 3 pour rayon de recouvrement, ce qui montre que $K(9,3) \leqq 7$.

Montrons maintenant que $K(9,3) > 6$. Pour cela supposons que C est un code binaire de longueur 9, contenant 6 mots, et ayant un rayon de recouvrement égal à 3.

Première étape : Montrons que C est équilibré.

Comme $K(7,2) = 7$, on peut appliquer le lemme 1 à C. En particulier, toute colonne de C contient au moins 2 fois la valeur 0 et 2 fois la valeur 1.

(*a*) Supposons que C admette deux colonnes déséquilibrées, les deux premières, qui contiennent 2 fois la valeur 0 et 4 fois la valeur 1 (ceci sans perte de généralité). Les sept autres colonnes sont équilibrées ou non. D'après le lemme 1, on a forcément (aux permutations de lignes près) la configuration suivante sur les deux premières colonnes de C :

$$\begin{matrix} 1\ 2\ \ldots\ldots\ 9 \\ 0\ 0 \\ 0\ 1 \\ 1\ 0 \\ 1\ 1 \\ 1\ 1 \\ 1\ 1 \end{matrix}$$

Soit S l'ensemble des mots de F_2^9 dont les deux premières composantes sont 0. $|S| = 2^7 = 128$.

Or c_1 peut couvrir au maximum :

$$1 + 7 + \binom{7}{2} + \binom{7}{3} = 64 \text{ mots de S.}$$

c_2 et c_3 peuvent couvrir au maximum :

$$2\left(1 + 7 + \binom{7}{2}\right) = 58 \text{ mots de S.}$$

c_4, c_5 et c_6 peuvent couvrir au maximum :

$$3(1+7) = 24 \text{ mots de S,}$$

soit au total 146 mots. Donc au plus 18 mots de S peuvent être couverts par 2 mots différents de C.

Posons (sans perte de généralité) $c_1 = \underline{0}$.

Comme 01 doit apparaître en colonnes 1-3, 1-4, ..., 1-9, on a :

$$c_2 = 011111111.$$

De même 01 devant apparaître en colonnes 2-3, 2-4, ..., 2-9, on a :

$$c_3 = 101111111.$$

Alors c_2 et c_3 recouvrent 2 fois 29 mots de S, ce qui montre que C ne peut avoir deux colonnes déséquilibrées.

(*b*) Supposons que C admette une seule colonne déséquilibrée, la deuxième, qui contient (sans perte de généralité) 2 fois la valeur 0 et 4 fois la valeur 1. Les huit autres colonnes sont équilibrées. D'après le lemme 1, on a forcément (aux permutations de lignes

près) la configuration suivante sur les deux premières colonnes de C :

$$\begin{array}{c} 1\ 2\ \ldots\ldots\ 9 \\ 0\ 0 \\ 1\ 0 \\ 0\ 1 \\ 1\ 1 \\ 0\ 1 \\ 1\ 1 \end{array}$$

On peut poser sans perte de généralité $c_1 = \underline{0}$. Alors d'après le lemme 1, $c_2 = 101111111$:

$$\begin{array}{c} 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\ 1\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1 \\ 0\ 1 \\ 1\ 1 \\ 0\ 1 \\ 1\ 1 \end{array}$$

Comme les colonnes 3 à 9 doivent être équilibrées, elles doivent être complétées avec les six quadruplets

1 1 1 0 0 0
1 0 0 1 1 0
0' 1' 0' 1' 0' 1'
0 0 1 0 1 1

Cela est impossible, chaque quadruplet ne pouvant être pris qu'une fois, toujours d'après le lemme 1. On a donc établi que C est un code équilibré.

Deuxième étape : Supposons maintenant (sans perte de généralité) que C contient le vecteur nul. Pour être équilibrées, les 9 colonnes de C doivent être complétées avec les 10 quintuplets :

1 1 1 1 1 1 0 0 0 0
1 1 1 0 0 0 1 1 1 0
1, 0, 0, 1, 1, 0, 1, 1, 0 et 1.
0 1 0 1 0 1 1 0 1 1
0 0 1 0 1 1 0 1 1 1

Chacun ne pouvant être pris qu'une fois, on peut supposer, sans perte de généralité, que $\begin{matrix}0\\0\\1\\1\end{matrix}$ est le seul à n'être pas pris.

On a donc de manière unique (aux translations et permutations de lignes et colonnes près) :

$$C = \begin{array}{c} 000000000 \\ 111111000 \\ 111000111 \\ 100110110' \\ 010101101 \\ 001011011 \end{array}$$

Or ce code a un rayon de recouvrement égal ou supérieur à 4 : le mot 001101110 en est à distance 4.

Conclusion : $K(9,3) > 6$ et le théorème 3 est démontré.

Ce théorème s'inscrit dans le cadre d'un résultat plus général :

$$\forall p \geqq 1, \quad K(2p+3, p) = 7 \quad (\textit{voir} \text{ th. 5}).$$

Démonstration du théorème 4 : On sait que $6 \leqq K(10,3) \leqq 16$.

$K(10,3) \leqq 12$ est un cas particulier du résultat plus général :

$$\forall p \geqq 1, \quad K(2p+4, p) \leqq 12 \quad (\textit{voir} \text{ th. 6}).$$

Les inégalités que doivent vérifier $A_1, A_2, \ldots, A_{10}$ (on prend $A_0 = 1$) s'écrivent :

$$\begin{aligned}
&84A_1 + 28A_2 + 70A_3 + 25A_4 + 55A_5 \\
&\qquad + 15A_6 + 35A_7 \geqq 210, \\
&56A_2 + 21A_3 + 66A_4 + 26A_5 + 66A_6 \\
&\qquad + 21A_7 + 56A_8 \geqq 252, \\
&35A_3 + 15A_4 + 55A_5 + 25A_6 + 70A_7 \\
&\qquad + 28A_8 + 84A_9 \geqq 210, \\
&20A_4 + 10A_5 + 40A_6 + 22A_7 + 64A_8 \\
&\qquad + 36A_9 + 120A_{10} \geqq 120, \\
&10A_5 + 6A_6 + 24A_7 + 17A_8 + 45A_9 + 45A_{10} \geqq 45, \\
&10A_5 + 6A_6 + 24A_7 + 17A_8 + 45A_9 + 45A_{10} \geqq 45, \\
&4A_6 + 3A_7 + 10A_8 + 10A_9 + 10A_{10} \geqq 10, \\
&A_7 + A_8 + A_9 + A_{10} \geqq 1.
\end{aligned}$$

Si on y ajoute :

$$A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9 + A_{10} = 5$$

$$0 \leqq A_i \leqq 5 \quad (i = 1, 2, \ldots, 9),$$
$$0 \leqq A_{10} \leqq 1,$$

l'ensemble des solutions est vide. On a donc $K(10,3) > 6$.

Si on y ajoute :

$$A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9 + A_{10} = 6$$

$$0 \leqq A_i \leqq 6 \quad (i = 1, 2, \ldots, 9),$$
$$0 \leqq A_{10} \leqq 1,$$

on trouve quatre solutions que nous n'avons pu éliminer : $K(10,3) = 7$ est possible.

Le théorème 4 sera donc démontré dès que le théorème 6 l'aura été.

1.4. Résultats généraux

Nous avons notamment obtenu les deux résultats suivants :

Théorème 5 :

$$\boxed{\forall p \geqq 1, \quad K(2p+3, p) = 7.}$$

Théorème 6 :

$$\boxed{\forall p \geqq 1, \quad 7 \leqq K(2p+4, p) \leqq 12.}$$

Avant de démontrer ces deux théorèmes, donnons le lemme suivant :

Lemme 2 : *Soit* C *un code binaire de longueur* n *possédant les propriétés suivantes :* $t(C) \geqq 1$, *et il existe deux sous-codes* C_0 *et* C_1 *de* C *tels que :*

– $C = C_0 \cup C_1$ et $C_0 \cap C_1 = \varnothing$;

– C_0 (*respectivement* C_1) *est à distance au plus* $t(C)+1$ *de tous les mots de* F_2^n, *sauf éventuellement des mots de* C_1 (*respectivement* C_0) *qui peuvent être à distance* $t(C)+2$.

Alors on peut construire un code C′ *de longueur* $n+2p$ ($p \geqq 1$) *contenant autant de mots que* C *et dont le rayon de recouvrement vaut au plus* $t(C)+p$.

Démonstration du lemme 2 :

$$C = \begin{array}{|c|}\hline C_0 \\ \hline C_1 \\ \hline \end{array} \quad \underbrace{}_{n}$$

Pour $p \geqq 1$, on construit le code C′ de la manière suivante, par exemple :

$$C' = \begin{array}{|c|l|}\hline C_0 & 000\ldots\ldots\ldots 0 \\ & \vdots \\ & 0\ldots\ldots\ldots 0 \\ \hline C_1 & 111\ldots\ldots\ldots 1 \\ & \vdots \\ & 1\ldots\ldots\ldots 1 \\ \hline \end{array}$$

n $\qquad$ $2p$

C′ est un code de longueur $n+2p$, contenant autant de mots que C. Montrons que $t(C') \leqq t(C)+p$.

Soit $v \in F_2^{n+2p}$. Sans perte de généralité, écrivons :

$$v = (v_1 \underbrace{00\ldots\ldots}_{k_1} \underbrace{011\ldots\ldots\ldots 1}_{2p-k_1}),$$

avec $v_1 \in F_2^n$ et $0 \leqq k_1 \leqq p$.

Pour $c_0 \in C_0$ posons :

$$\bar{c}_0 = (c_0 \underbrace{00\ldots\ldots 0}_{2p}) : \bar{c}_0 \in C'.$$

Pour $c_1 \in C_1$ posons :

$$\bar{c}_1 = (c_1 \underbrace{11\ldots\ldots 1}_{2p}) : \bar{c}_1 \in C'.$$

Premier cas : $k_1 = p$:

$$\exists c \in C \quad \text{tel que} \quad d(c, v_1) \leqq t(C).$$

Alors soit :

$$c \in C_0 \quad \text{et} \quad d(\bar{c}, v) \leqq t(C)+p,$$

soit :

$$c \in C_1 \quad \text{et} \quad d(\bar{c}, v) \leqq t(C)+p.$$

Dans les deux cas,

$$d(C', v) \leqq t(C)+p.$$

Deuxième cas : $k_1 = p-1$.

Si $v_1 \in C_0$ alors :

$$d(C', v) \leqq 2p-k_1 = p+1 \leqq t(C)+p.$$

Si $v_1 \notin C_0$ alors :

$$\exists c_1 \in C_1 \quad \text{tel que} \quad d(c_1, v_1) \leqq t(C)+1,$$

et donc :

$$d(\bar{c}_1, v) \leqq t(C)+1+k_1 = t(C)+p.$$

Troisième cas : $k_1 \leqq p-2$:

$$\exists c_1 \in C_1 \quad \text{tel que} \quad d(v_1, c_1) \leqq t(C)+2,$$

et donc :

$$d(v, \bar{c}_1) \leqq t(C)+2+k_1 \leqq t(C)+p.$$

Dans tous les cas, on a donc $d(C', v) \leqq t(C)+p$, ce qui montre que $t(C') \leqq t(C)+p$ et achève la démonstration du lemme 2.

Démonstration du théorème 5 :

Première étape :

$$\forall p \geqq 1, \quad K(2p+3, p) \leqq 7.$$

Rappelons que :

$$C_{5,7,1} = \begin{matrix} 00000 \\ 00001 \\ 01110 \\ 10000 \\ 11101 \\ 11011 \\ 10111 \end{matrix}$$

a un rayon de recouvrement égal à 1. Posons :

$$C_0 = \begin{matrix} 00000 \\ 00001, \\ 01110 \end{matrix}$$

$$C_1 = \begin{matrix} 10000 \\ 11101 \\ 11011 \cdot \\ 10111 \end{matrix}$$

Tous les mots de F_2^5 sont à distance au plus 2 de C_0, sauf les mots 11101, 11011 et 10111 qui appartiennent à C_1 et qui sont à distance 3 de C_0.

Tous les mots de F_2^5 sont à distance 2 au plus de C_1, sauf le mot 01110 qui appartient à C_0 et qui est à distance 3 de C_1.

$C_{5,7,1}$ remplit donc les conditions du lemme 2.
Donc $\forall p \geqq 1$ on peut construire un code C′ de longueur $5+2p$, contenant 7 mots, et de rayon de recouvrement inférieur ou égal à $p+1$, de la manière exposée dans la démonstration du lemme 2. Ceci montre que :

$$\forall p \geqq 1, \quad K(5+2p, p+1) \leqq 7.$$

Comme $K(5,1)=7$, on a donc :

$$\forall p \geqq 1, \quad K(3+2p, p) \leqq 7.$$

Les codes $C_{7,7,2}$ et $C_{9,7,3}$ donnés plus haut (*voir* th. 1 et 3) ont été construits ainsi.

Deuxième étape :

$$\forall p \geqq 1, \quad K(2p+3, p) > 6.$$

La démonstration se fait par récurrence.
Les cas $p=1$, $p=2$ et $p=3$ ont été étudiés plus haut (*voir* [1 . 1.], et théorèmes 1 et 3). Supposons que cette propriété soit vraie à l'ordre p_0 :

$$K(2p_0+3, p_0) > 6$$

[et donc $K(2p_0+3, p_0)=7$] pour $p_0 \geqq 3$.
Soit C un code de longueur $n=2(p_0+1)+3$ ($n \geqq 11$), contenant 6 mots et ayant un rayon de recouvrement égal à p_0+1.
Une première étape consiste à montrer que C est équilibré. D'après l'hypothèse de récurrence, on peut appliquer le lemme 1 à C. En particulier, toute colonne de C contient au moins 2 fois la valeur 0 et 2 fois la valeur 1.
(*a*) Supposons que C admette deux colonnes déséquilibrées, les deux premières, qui contiennent 2 fois la valeur 0 et 4 fois la valeur 1 (ceci sans perte de généralité). Les $(2p_0+3)$ autres colonnes sont équilibrées ou non. D'après le lemme 1, on a forcément (aux permutations de lignes près) la configuration suivante sur les deux premières colonnes de C :

1 2 3 . . . . $2p_0+5$
0 0
0 1
1 0
1 1
1 1
1 1

On peut poser sans perte de généralité $c_1=\underline{0}$. Alors d'après le lemme 1,

$$c_2 = 01111.....1 \quad \text{et} \quad c_3 = 10111.....1.$$

Les colonnes 3 à $2p_0+5$ doivent maintenant être complétées avec les huit triplets :

0 0 0 1 1 1 0 1
0, 0, 1, 0, 1, 0, 1, 1.
0 1 0 0 0 1 1 1

Mais chaque triplet pouvant être pris au maximum une fois, et plus de neuf colonnes devant être ainsi complétées, cela montre qu'il est impossible que C admette deux colonnes déséquilibrées.
(*b*) Supposons que C admette une seule colonne déséquilibrée, la deuxième, qui contient 2 fois la valeur 0 et 4 fois la valeur 1 (ceci sans perte de généralité). Les $(2p_0+4)$ autres colonnes sont équilibrées. D'après le lemme 1, on a forcément (aux permutations de lignes près) la configuration suivante sur les deux premières colonnes de C :

1 2 3 . . . . $2p_0+5$
0 0
1 0
0 1
1 1
0 1
1 1

Posons $c_1=\underline{0}$. D'après le lemme 1, $c_2=10111.....1$.
Les colonnes 3 à $2p_0+5$ étant équilibrées, elles doivent être complétées avec les six quadruplets :

1 1 1 0 0 0
1 0 0 1 1 0
0' 1' 0' 1' 0' 1'
0 0 1 0 1 1

Cela est impossible, chaque quadruplet pouvant être pris au plus une fois. On en déduit que C est équilibré.
Supposons maintenant (sans perte de généralité) que C contient le vecteur nul. Pour être équilibrées, les $(2p_0+5)$ colonnes de C doivent être complétées avec les 10 quintuplets :

1 1 1 1 1 1 0 0 0 0
1 1 1 0 0 0 1 1 1 0
1, 0, 0, 1, 1, 0, 1, 1, 0, 1.
0 1 0 1 0 1 1 0 1 1
0 0 1 0 1 1 0 1 1 1

Chacun ne pouvant être pris qu'une fois, et C ayant 11 colonnes au moins, cela est impossible. On en déduit que $K(2p_0+5, p_0+1) > 6$, ce qui achève la démonstration du théorème 5.

Démonstration du théorème 6 : L'inégalité $7 \leqq K(2p+4, p)$ vient du fait que $K(2p+3, p)=7$. Montrons maintenant que :

$$\forall p \geqq 1, \quad K(2p+4, p) \leqq 12.$$

Rappelons que :

$$C_{6,12,1} = \begin{matrix} 000000 \\ 101000 \\ 011000 \\ 110100 \\ 001110 \\ 110010 \\ 110001 \\ 111111 \end{matrix}$$

$$\begin{matrix}010111\\100111\\001011\\001101\end{matrix}$$

a un rayon de recouvrement égal à 1.
Soit :

$$C_0 = \begin{matrix}000000\\101000\\011000\\110100\\001110\\110010\end{matrix}, \qquad C_1 = \begin{matrix}110001\\111111\\010111\\100111\\001011\\001101\end{matrix}.$$

Tous les mots de F_2^6 sont à distance au plus 2 de C_0 sauf les mots 100111, 010111 et 111111 qui appartiennent à C_1 et qui sont à distance 3 de C_0.
Tous les mots de F_2^6 sont à distance 2 au plus de C_1 sauf les mots 000000, 101000 et 011000 qui appartiennent à C_0 et qui sont à distance 3 de C_1.
$C_{6,12,1}$ vérifie donc les conditions du lemme 2.
Donc $\forall p \geqq 1$ on peut construire un code C′ de longueur $6+2p$, contenant 12 mots et de rayon de recouvrement inférieur ou égal à $p+1$. Ceci montre que :

$$\forall p \geqq 1, \quad K(2p+6, p+1) \leqq 12.$$

Comme $K(6,1) = 12$, on a donc :

$$\forall p \geqq 1, \quad K(2p+4, p) \leqq 12.$$

Le code $C_{8,12,2}$ donné plus haut (*voir* th. 2) a été construit ainsi.
Le cas particulier $p=3$ a été donné plus haut sans démonstration (*voir* th. 4).
Le théorème 6 est donc démontré.

Conclusion

Nous avons donc établi de nouveaux résultats sur $K(n, t)$ (jusqu'ici, seul le cas $t=1$ avait été l'objet d'études), établissant la valeur de $K(2p+3, p)$ $(p \geqq 2)$ et une borne supérieure pour $K(2p+4, p)$ $(p \geqq 2)$.

BIBLIOGRAPHIE

[1] G. D. Cohen, M. R. Karpovsky, H. F. Mattson Jr et J. R. Schatz, Covering Radius: Survey and Recent Results, *IEEE Transactions on Information Theory* (à paraître).

[2] S. Even, *Combinatorics*, MacMillan Company, New York.

[3] M. R. Garey et D. S. Johnson, *Computers and Intractability: a Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, San Francisco.

[4] A. Hartman, W. H. Mills et R. C. Mullin, *Covering Triples by Quadruples: an Asymptotic Solution.*

[5] A. Hartman, R. C. Mullin et D. R. Stinson, Exact Covering Configurations and Steiner Systems, *J. London Math. Soc.*, (2), 25, 1982.

[6] A. M. MacLoughlin, *The Complexity of Computing the Covering Radius of a Code*, Preprint.

[7] F. J. MacWilliams et N. J. A. Sloane, *The Theory of Error Correcting Codes* (I).

[8] *The Theory of Error Correcting Codes* (II), North Holland Mathematical Library.

[9] R. G. Stanton et J. G. Kalbfleisch, Covering Problems for Dichotomized Matchings, *Aequat. Math.*, I, 1968.

[10] R. G. Stanton et J. G. Kalbfleisch, Intersection Inequalities for the Covering Problem, *SIAM J. Appl. Math.*, 17, n° 6, November 1969.