COMBINATORICA 6 (3) (1986) 279—285

ORTHOGONAL VECTORS IN THE n-DIMENSIONAL CUBE AND CODES WITH MISSING DISTANCES

P. FRANKL

Received 15 October 1984

For k a positive integer let $m(4k)$ denote the maximum number of ± 1-vectors of length $4k$ so that no two are orthogonal. Equivalently, $m(4k)$ is the maximal number of codewords in a code of length $4k$ over an alphabet of size two, such that no two codewords have Hamming distance $2k$. It is proved that $m(4k)=4\sum_{0\leqq i<k}\binom{4k-1}{i}$ if k is the power of an odd prime.

1. Introduction

Let C^n be the set of vertices of the standard n-dimensional cube of edge length 2 and centered at the origin. That is, the 2^n vertices of C^n are all the points of $\mathbf{R}^n$ of the form $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ where ε_i is $+1$ or -1. Two vertices $P=(\varepsilon_1, \ldots, \varepsilon_n)$ and $Q=(\delta_1, \ldots, \delta_n)$ are said to be *orthogonal* if $\sphericalangle \mathrm{POQ}=90^\circ$ holds. This is easily seen to be equivalent to $\sum_{1\leqq i\leqq n}\varepsilon_i\delta_i=0$.

In 1972 Larman and Rogers raised the following question.

Problem 1.1. Determine or estimate the maximum number, $m(n)$ of vertices of C^n so that no two are orthogonal.

In the case n is odd, the answer is trivial:

$$m(2k+1) = 2^{2k+1} \tag{1}$$

holds for all nonnegative integers k because $\sum \varepsilon_i\delta_i$ being a sum of an odd number of plus and minus one's is never zero.

A little less trivial is the following.

Proposition 1.2. *$m(4k+2)=2^{4k+1}$ holds for all nonnegative integers k.*

Therefore the really interesting case is when the dimension is divisible by 4, $n=4k$. Larman and Rogers conjectured that for k large $m(4k)$ is much smaller than $2^{4k}/k^2$. Rödl and the author (cf. [2]) proved this conjecture in a stronger form. Namely, they showed that there exists a very small but positive γ so that $m(4k)<$

AMS subject classification (1980): 05 C 35

$<(2-\gamma)^{4k}$ holds. Here we present a simpler argument which will provide the exact value of $m(4k)$, however, only in the case when k is the power of an odd prime.

Theorem 1.3. *Suppose* $k=p^\alpha$, $\alpha\geqq 1$, $p\geqq 3$, p *prime. Then*

$$m(4k) = 4\sum_{i=0}^{k-1}\binom{4k-1}{i} < 4^{4k}/3^{3k} \quad \textit{holds.} \tag{2}$$

Let us now describe Problem 1.1 from a coding theoretical viewpoint. A (binary) *code* of length n over the *alphabet* $\{a, b\}$ is a collection $\mathscr{C}$ of sequences — called *codewords* $\mathbf{x}=(x_1, x_2, \ldots, x_n)$ with $x_i\in\{a, b\}$, $i=1, \ldots n$. The *distance* of two codewords $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$d(\mathbf{x}, \mathbf{y}) = \#i\colon x_i \neq y_i.$$

Thus $d(\mathbf{x}, \mathbf{y})=0$ iff $\mathbf{x}=\mathbf{y}$, and $d(\mathbf{x}, \mathbf{y})$ is always an integer between zero and n.

Given a subset $\mathscr{C}\subset C^n$, we can consider it a code of length n over the alphabet $\{+1, -1\}$. It is easy to see that two points $\varepsilon=(\varepsilon_1, \ldots, \varepsilon_n)$, $\delta=(\delta_1, \ldots, \delta_n)$ are orthogonal if and only if $d(\varepsilon, \delta)=n/2$ holds. This shows once again that there are no orthogonal points if n is odd.

One of the principal problems of coding theory is the following.

Problem 1.4. Given a set $D\subset\{1, 2, \ldots, n\}$, determine or estimate the maximum number $m(n, D)$ of codewords in a code $\mathscr{C}$ of length n satisfying

$$d(\mathbf{x}, \mathbf{y})\in D \quad \text{for all distinct} \quad \mathbf{x}, \mathbf{y}\in\mathscr{C}.$$

A classical result of Delsarte says:

Theorem 1.5. ([1])

$$m(n, D) \leqq \sum_{i=0}^{|D|}\binom{n}{i}. \tag{3}$$

By the above $m(n)=m(n, \{1, 2, \ldots, n\}-\{n/2\})$. From (3) we obtain only $m(n)\leqq 2^n-1$ (for n even). We deduce (2) from the following.

Theorem 1.6. *Suppose* $k=p^\alpha$, p *a prime*, $p\geqq 3$, $\alpha\geqq 1$; $\mathscr{C}$ *is a code of length* l *such that* $k\nmid d(x, y)$ *holds for all distinct* $\mathbf{x}, \mathbf{y}\subset\mathscr{C}$. *Then*

$$|\mathscr{C}| \leqq \sum_{0\leqq i\leqq k-1}\binom{l}{i} \quad \textit{holds.} \tag{4}$$

In the above terminology (4) says $m(l, \{1, 2, \ldots, l\}-\{k, 2k, 3k, \ldots\})\leqq$ $\leqq \sum_{0\leqq i\leqq k-1}\binom{l}{i}$.

There is an obvious similarity between the upper bounds in (3) and (4). In fact, we prove (4) by adopting Delsarte's argument to this particular choice of D.

Conjecture 1.7. (4) *holds for all positive integers* k.

2. The exact value of $m(4k+2)$

First observe that no two points are orthogonal in $C^{4k+2}_{\text{even}}=\{(\varepsilon_1, \ldots, \varepsilon_{4k+2})\colon \varepsilon_i=+1$ holds an even number of times$\}$. This yields $m(4k+2)\geqq 2^{4k+1}$.

To prove the upper bound define for all $\varepsilon=(\varepsilon_1, \ldots, \varepsilon_{4k+2})$ a new point $\varepsilon^*=(\varepsilon_1, \ldots, \varepsilon_{2k+1}, -\varepsilon_{2k+2}, \ldots, -\varepsilon_{4k+2})$. It is clear that ε^* and ε are orthogonal, and $(\varepsilon^*)^*=\varepsilon$.

Suppose now $\mathscr{C}\subset C^{4k+2}$ and there are no two orthogonal points in $\mathscr{C}$. Then $\mathscr{C}^*=\{\varepsilon^*\colon \varepsilon\in\mathscr{C}\}$ and $\mathscr{C}$ are disjoint and $|\mathscr{C}^*|=|\mathscr{C}|$ holds. Consequently $|\mathscr{C}|\leqq (1/2)\,2^{4k+2}=2^{4k+1}$ holds.

3. The proof of theorem 1.3 using (4)

Suppose $\mathscr{C}\subset C^{4k}$ and $\mathscr{C}$ contains no pair of orthogonal points, or equivalently — in the coding terminology — $d(\mathbf{x}, \mathbf{y})\neq 2k$ holds for all $\mathbf{x}, \mathbf{y}\in\mathscr{C}$.
Define $\mathscr{C}_{\text{even}}$ by $\mathscr{C}_{\text{even}}=\{\mathbf{x}=(x_1, \ldots, x_n)\in\mathscr{C}\colon \#\, i\colon x_i=+1$ is even$\}$, $\mathscr{C}_{\text{odd}}$ is defined analogously.

It is easy to check that $d(\mathbf{x}, \mathbf{y})$ is even for all $\mathbf{x}, \mathbf{y}\in\mathscr{C}_{\text{even}}$ and also for all $\mathbf{x}, \mathbf{y}\in\mathscr{C}_{\text{odd}}$. Consequently for both these "subcodes" $d(\mathbf{x}, \mathbf{y})\neq k, 2k, 3k$ holds.

In order to apply (4) we define

$$\mathscr{C}^+_{\text{even}} = \{(x_1, \ldots, x_{n-1})\colon (x_1, \ldots, x_n)\in\mathscr{C}_{\text{even}}, x_n=+1\},$$

and analogously $\mathscr{C}^-_{\text{even}}, \mathscr{C}^+_{\text{odd}}, \mathscr{C}^-_{\text{odd}}$. Then all four are codes of length $4k-1$ in which the distance of two distinct codewords is never divisible by k.

Now (4) implies:

$$|\mathscr{C}| \leqq 4 \sum_{0\leqq i<k} \binom{4k-1}{i},$$

which is the upper bound part of (2).

For the lower bound part k can be an arbitrary positive integer. Define

$$\mathscr{D} = \{(\varepsilon_1, \ldots, \varepsilon_{4k})\in C^{4k}\colon \big|\{i\colon 1\leqq i\leqq 4k-1,\ \varepsilon_i=1\}\big|$$

is either less than k or more than $3k\}$.

It is easy to check that $\mathscr{D}$ has the desired size and it contains no two orthogonal vectors, concluding the proof of (2).

4. The proof of theorem 1.6

For each $\mathbf{x}\in\mathscr{C}$ define $S(\mathbf{x})=\{i\colon x_i=+1\}$. For $0\leqq j<k$ we define $M(j)$ a $|\mathscr{C}|$ by $\binom{l}{j}$ matrix whose rows are indexed by the codewords, the columns by the j-element subsets of $\{1, 2, \ldots, l\}$ and the entry in the position $(\mathbf{x}, A)$ is $(-1)^{|S(\mathbf{x})\cap A|}$. Thus $M(j)$ is a ± 1-matrix. Clearly the rank of $M(j)$ does not exceed

the number of its columns, i.e.,

$$\text{rank } M(j) \leqq \binom{l}{j} \text{ holds.} \tag{5}$$

Let us compute the general entry of the matrix $N(j)=M(j)M(j)^T$. It is a $|\mathscr{C}|$ by $|\mathscr{C}|$ matrix whose $(\mathbf{x}, \mathbf{y})$-entry $n_j(\mathbf{x}, \mathbf{y})$ is the scalar product of rows $\mathbf{x}$ and $\mathbf{y}$ of $M(j)$.

Proposition 4.1.

$$n_j(\mathbf{x}, \mathbf{y}) = \sum_{0 \leqq i \leqq j} (-1)^i \binom{d(\mathbf{x}, \mathbf{y})}{i} \binom{l-d(\mathbf{x}, \mathbf{y})}{j-i}.$$

Proof. By definition

$$n_j(\mathbf{x}, \mathbf{y}) = \sum_{\substack{A \subset \{1, 2, \ldots, l\} \\ |A|=j}} (-1)^{|S(\mathbf{x}) \cap A|} (-1)^{|S(\mathbf{y}) \cap A|} \tag{6}$$

$$= \sum_{\substack{A \subset \{1, 2, \ldots, l\} \\ |A|=j}} (-1)^{|(S(\mathbf{x}) \Delta S(\mathbf{y})) \cap A|}$$

where $S(\mathbf{x})\Delta S(\mathbf{y})$ is the symmetric difference of $S(\mathbf{x})$ and $S(\mathbf{y})$. Thus $|S(\mathbf{x})\Delta S(\mathbf{y})| = d(\mathbf{x}, \mathbf{y})$ holds. To evaluate (6) note that given $S(\mathbf{x})\Delta S(\mathbf{y})$ there are exactly

$$\binom{d(\mathbf{x}, \mathbf{y})}{i} \binom{l-d(\mathbf{x}, \mathbf{y})}{j-i}$$

j-element sets A which intersect $S(\mathbf{x})\Delta S(\mathbf{y})$ in i elements. We infer

$$n_j(\mathbf{x}, \mathbf{y}) = \sum_{0 \leqq i \leqq j} (-1)^i \binom{d(\mathbf{x}, \mathbf{y})}{i} \binom{l-d(\mathbf{x}, \mathbf{y})}{j-i}. \tag{7}$$

The expression on the right hand side is a polynomial in $d(\mathbf{x}, \mathbf{y})$ of degree j: the coefficient of $d(\mathbf{x}, \mathbf{y})^j$ is

$$\sum_{0 \leqq i \leqq j} \frac{(-1)^i}{i!} \frac{(-1)^{j-i}}{(j-i)!} = \frac{(-1)^j}{j!} \sum_{0 \leqq i \leqq j} \binom{j}{i} = \frac{(-2)^j}{j!}.$$

This polynomial $K(l, j, z)$ is called the Krawtchouk polynomial and it plays an important role in coding theory.

Consider the polynomial

$$\binom{z-1}{k-1} = \frac{(z-1)(z-2)\ldots(z-k+1)}{(k-1)!}.$$

As a polynomial of degree $k-1$, $p(z)$ can be uniquely written as a linear combination of $K(l, j, z)$ for $0 \leqq j \leqq k-1$: $p(z)=\sum \alpha_j K(l, j, z)$. Consider the matrix $N = \sum_{0 \leqq j < k} \alpha_j N(j)$, where addition is entrywise. In view of (6) the $|\mathscr{C}|$ by $|\mathscr{C}|$ matrix N has the general entry

$$\sum_{0 \leqq j < k} K(l, j, d(\mathbf{x}, \mathbf{y})) = \binom{d(\mathbf{x}, \mathbf{y})-1}{k-1}.$$

In view of $N(j)=M(j)M(j)^T$, rank $N(j)\leqq$rank $M(j)$ holds. Consequently (5) implies

$$\text{rank } N \leqq \sum_{0\leqq j\leqq k-1} \text{rank } M(j) \leqq \sum_{0\leqq j\leqq k-1} \binom{l}{j}.$$

Thus the proof of (4) will be complete as soon as we show that N has full rank, i.e., rank $N=|\mathscr{C}|$ holds.

We need an easy lemma of number-theoretic character.

Proposition 4.2. *Suppose* $k=p^\alpha$, $\alpha\geqq 1$, p *prime,* d *an integer. Then*

$$\binom{d-1}{k-1} \equiv \begin{cases} 1 \pmod p & \text{if } k|d \\ 0 \pmod p & \text{if } k\nmid d. \end{cases}$$

Proof. Suppose first $b=ak-1$ for some integer a. Then

$$\binom{b}{k-1} = \prod_{1\leqq i\leqq k-1} \frac{ak-i}{k-i}.$$

As $k=p^\alpha$ and $1\leqq i\leqq k-1$, $ak-i$ and $k-i$ are divisible by the same power of p, say $p^{\alpha(i)}$. Moreover $\alpha(i)<\alpha$ and consequently

$$((ak-i)/p^{\alpha(ii)})/((k-i)/p^{\alpha(i)}) \equiv 1 \pmod p$$

holds. This yields $\binom{b}{k-1}\equiv 1 \pmod p$.

Suppose next $b\not\equiv -1 \pmod k$, i.e., $p^\alpha\nmid(b+1)$. Consider the binomial coefficient

$$\binom{b+1}{k} = \frac{b+1}{k}\binom{b}{k-1}.$$

It is an integer, thus p^α divides $(b+1)\binom{b}{k-1}$. We infer $p\Big|\binom{b}{k-1}$. ∎

Now we can conclude the proof of Theorem 1.6. The diagonal terms of N are equal to $\binom{-1}{k-1}\equiv 1 \pmod p$. However $k\nmid d(\mathbf{x},\mathbf{y})$ implies that all the off diagonal terms are divisible by p. Thus $\det N\equiv 1 \pmod p$, in particular $\det N\neq 0$ holds. ∎

5. Hadamard graphs

The vertices of the Hadamard graph $H(k)$ are the vertices of C^{4k}; with two vertices forming an edge if and only if they are orthogonal. The reader is referred for lot of interesting information on $H(k)$ to the recent paper of Ito [5]. Let us mention only a few properties of $H(k)$. It has two isomorphic connected components. Denote by $H_0(k)$ the component containing $(1, 1, \ldots, 1)$. It consists of all vertices which have an even number of 1's.

A clique in $H(k)$ is a collection of pairwise orthogonal vertices of C^{4k}. One of the outstanding open problems of combinatorics is to decide whether for every k, $H(k)$ contains a clique of size $4k$, i.e., there exists a Hadamard matrix of order $4k$.

For a graph G let $\alpha(G)$ denote the independence number of G, i.e., the maximum number of vertices so that no two of them form an edge. Clearly $\alpha(H(k))=2\alpha(H_0(k))$ and $\alpha(k)=m(4k)$ hold. Thus Theorem 1.3 determines the independence number of $H(k)$ and of $H_0(k)$ if k is an odd prime power.

For a graph G let $\chi(G)$ denote the chromatic number of G.

Theorem 5.1. *If k is an odd prime power then*

$$\left(\frac{27}{16}\right)^k < \frac{2^{4k}}{\alpha(H(k))} \leqq \chi(H(k)) = \chi(H_0(k)) < \frac{3k2^{4k}}{\alpha(H(k))} < \left(\frac{27}{16}+o(1)\right)^k,$$

with the upper bound holding for all $k\geqq 1$.

Proof. The first two inequalities are direct consequences of Theorem 1.3. The last inequality follows from the lower bound part of the same theorem—noting that, that is valid for all integers. To prove the upper bound for $\chi(H(k))$ we apply the probabilistic method.

First of all if $\mathbf{x}=(x_1, \ldots, x_n)$ and $\mathbf{y}=(y_1, \ldots, y_n)$ then one defines their Hadamard product by $\mathbf{xy}=(x_1y_1, \ldots, x_ny_n)$.

If A is an independent set in C^{4k}, $\mathbf{y}\in C^{4k}$ then $\mathbf{y}A=\{\mathbf{yx}: \mathbf{x}\in A\}$ is an independent set also. Suppose $|A|=\alpha(H(k))$ and let us choose independently at random s elements of C^{4k}: $y_1, \ldots, y_s$; $s=3k\cdot 2^{4k}/\alpha(H(k))$. We claim that with positive probability $C^{4k}=y_1A\cup\ldots\cup y_sA$ holds. Clearly this will conclude the proof of the theorem.

For $\mathbf{z}\in C^{4k}$ and $1\leqq i\leqq s$, the probability of $\mathbf{z}\in y_iA$ is $\alpha(H(k))/2^{4k}$. By the independent choice of y_i the probability $p(\mathbf{z})$ of the event that $\mathbf{z}$ is not contained in any of y_iA, $1\leqq i\leqq s$ is given by $p(\mathbf{z})=(1-\alpha(H(k))/2^{4k})^s<e^{-3k}<2^{-4k}$. Since there are only 2^{4k} choices for $\mathbf{z}$, the claim follows. ∎

The Hadamard multiplication — defined in the proof — makes $H(k)$ into a group, isomorphic to Z_2^{4k}. $H_0(k)$ is a subgroup of index 2.

Let $\alpha_g(H(k))$ denote the maximum size of a subgroup of $H(k)$ which is an independent set.

Conjecture 5.2. (Ito [5]) $\alpha_g(H(k))=2^{2k}$.

To see that $\alpha_g(H(k))\geqq 2^{2k}$ consider first the cylinder $G=\{\mathbf{x}\in C^{4k}: x_i=0 \text{ for } i>2k\}$. The subgroup $\langle G, (1, \ldots, 1)\rangle$ is larger but still independent.

Note that Conjecture 5.2 can also be reformulated in coding terms: if $\mathscr{C}$ is a linear code of length $4k$ of dimension greater than $2k$ over GF(2) then $\mathscr{C}$ contains a word of weight $2k$ (or equivalently two words at distance $2k$).

6. Concluding remarks

For two points $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ their Euclidean distance is defined by

$$d_e(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{1 \leq i \leq n} (x_i - y_i)^2}.$$

It is easily seen that $\mathbf{x}, \mathbf{y} \in C^n$ are orthogonal if and only if d_e, their distance is equal to $\sqrt{2n}$.

Suppose now that $\mathbf{R}^n$ is partitioned into $X_1 \cup \ldots \cup X_m$, $n=4k$, k is a power of an odd prime and $m < 3^{3k}/2^{4k}$. Defining $Y_i = X_i \cap C^n$, obviously $Y_1 \cup \ldots \cup Y_m = C^n$. Thus for some i, $|Y_i| > 4^{4k}/3^{3k}$ holds. Therefore Theorem 1.3 implies the existence of two points $\mathbf{x}, \mathbf{y} \in Y_i$ whose Euclidean distance is exactly $\sqrt{2n}$.

Taking a homothetical image of C^n one can replace $\sqrt{2n}$ by an arbitrary positive number, e.g. by 1.

Since $\mathbf{R}^n$ is contained in $\mathbf{R}^{n'}$ for $n < n'$, we obtain

Theorem 6.1. *Suppose $n \geq 4k$, k the power of an odd prime and $\mathbf{R}^n = X_1 \cup \ldots \cup X_m$ with $m < 3^{3k}/2^{4k}$. Then there exists i, $1 \leq i \leq m$ and two points in X_i whose Euclidean distance is exactly* 1.

Let us denote by $\chi(n)$ the first value of m for which the statement of Theorem 6.1 fails. It is easy to see that $\chi(1)=2$ but the exact value of $\chi(n)$ is unknown for all $n \geq 2$. This problem goes back to Hadwiger [4]. Larman and Rogers [6] prove $\chi(n) < 3^n$ for n sufficiently large. Theorem 6.1 implies $\chi(n) > 1.13^n$ for n sufficiently large. In [3] $\chi(n) > 1.2^n$ was shown for $n > n_0$.

Recently Rödl and the author [2] showed that given $d \geq 2$, there exists a positive $\varepsilon = \varepsilon(d)$ such that whenever $\mathbf{R}^n = X_1 \cup \ldots \cup X_m$ with $m < (1+\varepsilon)^n$, then some X_i contains the vertices of a regular simplex of dimension d and of edgelength 1, i.e., $d+1$ points whose pairwise distance is exactly one.

Added in proof: For a proof of conjecture 5.2, cf. [7].

References

[1] Ph. Delsarte, On the four principal parameters of a code, *Information and Control* **23** (1973), 407—438.

[2] P. Frankl and V. Rödl, Forbidden intersections, *Transactions AMS, to appear.*

[3] P. Frankl and R. M. Wilson, Intersection theorems with geometric consequences, *Combinatorica* **1** (1981), 357—368.

[4] H. Hadwiger, Überdeckungssätze für den Euklidischen Raum, *Portugaliae Math.* **4** (1944), 140—144.

[5] N. Ito, Hadamard graphs, *Graphs and Combinatorics* **1** (1985), 57—64.

[6] D. G. Larman and C. A. Rogers, The realization of distances within sets in euclidean space, *Mathematika* **19** (1972), 1—24.

[7] H. Enomoto, P. Frankl, N. Ito, and K. Nomura, Bounds on the size of code with given distances, *Graphs and Combinatorics, to appear.*

P. Frankl

U.E.R. de Mathématiques
Université Paris VII
2, Place Jussieu
75221 Paris Cedex 05
France