ELSEVIER

Discrete Applied Mathematics 89 (1998) 269–276

DISCRETE APPLIED MATHEMATICS

Note

On the average Hamming distance for binary codes

Shutao Xia*, Fangwei Fu

Department of Mathematics, Nankai University, Tianjin, 300071, China

Received 20 September 1996; revised 29 December 1997; accepted 20 April 1998

Abstract

By using the dual distance distribution and its properties for binary code C with length n and size M, the Althöfer–Sillke inequality is improved for odd M. Let $\beta(n,M)$ denote the minimum value of average Hamming distance (AHD) of binary (n,M) codes. In this paper, $\beta(n,2^n-1)$, $\beta(n,2^{n-1}-1)$ and $\beta(n,2^{n-1}+1)$ are determined. Two recursive inequalities for $\beta(n,M)$ are derived. Furthermore, the variance of AHD of code C is studied, and lower and upper bounds are presented. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Average Hamming distance; Variance of AHD; Althöfer-Sillke inequality; Distance enumerator; MacWilliams–Delsarte identity

1. Introduction

Let $V_n=\{0,1\}^n$ be the n-dimensional vector space over the binary field $\{0,1\}$. The Hamming distance between two vectors a,b is denoted by $d_H(a,b)$. We call C a binary (n,M) code, if C is a subset of V_n with cardinality M. The average Hamming distance (AHD) of C is defined by

$$d(C)=\frac{1}{M^2}\sum_{a\in C}\sum_{b\in C}d_H(a,b). \tag{1}$$

The variance of $d(C)$ is defined by

$$\mathrm{var}(C)=\frac{1}{M^2}\sum_{a\in C}\sum_{b\in C}[d_H(a,b)-d(C)]^2. \tag{2}$$

In the efforts to solve an open problem posed by Ahlswede and Katona (see [2, pp. 10(1)]), Althöfer and Sillke proved that:

* Corresponding author. E-mail: stxia@mail.zlnet.co.cn.

0166-218X/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII: S0166-218X(98)00081-X

Theorem 1 (Althöfer and Sillke [1]).

$$d(C) \geqslant \frac{n+1}{2} - \frac{2^{n-1}}{M}, \tag{3}$$

where equality is possible only for $M = 2^n$ and for $M = 2^{n-1}$ with C being a subcube.

This inequality yields only negative values as lower bounds for $M < 2^n/(n+1)$, therefore it is only meaningful for large subsets. Moreover, Althöfer and Sillke showed that $d(C) \leqslant n/2$.

For fixed positive integers n, M, where $M \leqslant 2^n$, let

$$\beta(n,M) = \min\{d(C) \mid C \text{ is a binary } (n,M) \text{ code}\}.$$

Ahlswede and Katona [2] posed the following open problem: for every $1 \leqslant M \leqslant 2^n$, determining the exact value of $\beta(n,M)$. Theorem 1 shows that

$$\beta(n,2^n) = \frac{n}{2}, \quad \beta(n,2^{n-1}) = \frac{n-1}{2}.$$

Therefore, Althöfer and Sillke gave an answer for $M = 2^n$ or $M = 2^{n-1}$. Ahlswede and Althöfer [3] studied the asymptotic behaviour of $\beta(n,M)$. For the cases of $M \neq 2^n$ and 2^{n-1}, how to find the exact value or a good lower bound of $\beta(n,M)$ is still an open problem. In this paper, we improve Theorem 1 for odd M and give the exact values of $\beta(n, 2^n - 1)$, $\beta(n, 2^{n-1} - 1)$ and $\beta(n, 2^{n-1} + 1)$.

For the variance of $d(C)$, Fu and Shen [5] presented the following lower and upper bounds.

Theorem 2 (Fu and Shen [5]).

$$\frac{n-1}{4} + \frac{2^{n-1}}{M} - \frac{2^{2n-2}}{M^2} \leqslant \operatorname{var}(C) \leqslant \frac{n-2}{4} + \frac{2^{n-1}}{M}, \tag{4}$$

and the lower bound of var(C) *is achieved for $M = 2^n$ and $M = 2^{n-1}$ with C being a subcube.*

For fixed positive integers n, M, where $M \leqslant 2^n$, let

$$\alpha(n,M) = \min\{\operatorname{var}(C) \mid C \text{ is a binary } (n,M) \text{ code}\}.$$

Theorem 2 implies that

$$\alpha(n,2^n) = \frac{n}{4}, \quad \alpha(n,2^{n-1}) = \frac{n-1}{4}.$$

2. Preliminary

The Hamming weight of $x \in V_n$ is the number of non-zero coordinates, and is denoted by $w_H(x)$. Let $\langle \cdot, \cdot \rangle$ be the scalar product of two vectors. The distance distribution of code C is defined by

$$D_i = \frac{1}{M^2} |\{(a,b) \mid a, b \in C,\ d_H(a,b) = i\}|, \quad i = 0, 1, \ldots, n.$$

The dual distance distribution of code C is defined by

$$\hat{D}_i = \frac{1}{M^2} \sum_{\substack{u \in V_n \\ w_H(u) = i}} \left[\sum_{c \in C} (-1)^{\langle u, c \rangle} \right]^2, \quad i = 0, 1, \ldots, n. \tag{5}$$

Lemma 1 (MacWilliam and Sloane [4]). $\hat{D}_i \geqslant 0$, $\hat{D}_0 = 1$, $\sum_{i=0}^{n} \hat{D}_i = 2^n / M$.

The distance enumerator of code C is defined as $f(s) = \sum_{i=0}^{n} D_i s^i$. The dual distance enumerator of code C is defined as $g(s) = \sum_{i=0}^{n} \hat{D}_i s^i$. The MacWilliams–Delsarte identity gives the relationship between $f(s)$ and $g(s)$.

Lemma 2 (MacWilliam and Sloane [4]) (*MacWilliams–Delsarte identity*).

$$g(s) = (1 + s)^n f\left(\frac{1 - s}{1 + s}\right), \tag{6}$$

$$f(s) = \frac{1}{2^n} (1 + s)^n g\left(\frac{1 - s}{1 + s}\right). \tag{7}$$

It is easy to see from the MacWilliams–Delsarte identity or the Pless identity for the moments of distance distribution) that:

Lemma 3.

$$d(C) = \frac{n}{2} - \frac{\hat{D}_1}{2}, \tag{8}$$

$$\mathrm{var}(C) = \frac{n}{4} - \frac{\hat{D}_1^2}{4} + \frac{\hat{D}_2}{2}. \tag{9}$$

Lemma 4 (Best et al. [6]). *If* $1 \leqslant M \leqslant 2^n$ *and* M *is odd, then for every* $i = 1, 2, \ldots, n$,

$$\hat{D}_i \geqslant \frac{1}{M^2} \binom{n}{i}.$$

The equality holds for a fixed $1 \leqslant i \leqslant n$ *if and only if for every* $u \in V_n$ *with* $w_H(u) = i$,

$$\sum_{a \in C} (-1)^{\langle a, u \rangle} = 1 \text{ or } -1.$$

3. Improvements of Theorems 1 and 2

By Lemmas 1 and 3, we have

$$d(C)=\frac{n}{2}-\frac{1}{2}\left(\frac{2^n}{M}-1-\hat{D}_2-\cdots-\hat{D}_n\right)$$
$$=\frac{n+1}{2}-\frac{2^{n-1}}{M}+\frac{1}{2}(\hat{D}_2+\cdots+\hat{D}_n).$$

Hence, Lemma 1 implies the inequality (3) in Theorem 1, and the equality holds if and only if $\hat{D}_2=\hat{D}_3=\cdots=\hat{D}_n=0$, i.e.

$$\sum_{a\in C}(-1)^{\langle a,u\rangle}=0 \quad \text{for every } u\in V_n \text{ with } w_H(u)\geqslant 2. \tag{10}$$

Comparing with Theorem 1, we know that (10) holds if and only if C is V_n or its subcube with cardinality 2^{n-1}. By Lemma 4, we know that for odd M,

$$\hat{D}_2+\cdots+\hat{D}_n \geqslant \frac{1}{M^2}\left[\binom{n}{2}+\cdots+\binom{n}{n}\right]$$
$$=\frac{2^n-n-1}{M^2}.$$

Therefore, we have the following result which improves Theorem 1.

Theorem 3. *If M is odd, then*

$$d(C)\geqslant\frac{n+1}{2}-\frac{2^{n-1}}{M}+\frac{2^n-n-1}{2M^2}, \tag{11}$$

where the equality holds if and only if $\hat{D}_i=(1/M^2)\binom{n}{i}$, $i=2,3,\ldots,n$, i.e. for every $u\in V_n$ with $w_H(u)\geqslant 2$, $\sum_{a\in C}(-1)^{\langle a,u\rangle}=1$ or -1.

Remark. The inequality is meaningful only for $M\geqslant 2^n/(n+1)-1$.

Next, we will determine several exact values of $\beta(n,M)$ by Theorem 3.

- Let C be V_n, fix $a_0\in C$, remove a_0 from C, we get a binary code C_0 with size 2^n-1.
- Let C be a subcube of V_n with size 2^{n-1}, fix $a_0\in C$, remove a_0 from C, we get a binary code C_1 with size $2^{n-1}-1$.
- Let C be a subcube of V_n with size 2^{n-1}, fix $a_0\notin C$, add a_0 to C, we get a binary code C_2 with size $2^{n-1}+1$.

It is easy to see from (10) that for every $u\in V_n$ with $w_H(u)\geqslant 2$,

$$\sum_{a\in C_i}(-1)^{\langle a,u\rangle}=1 \quad \text{or} \ -1, \quad i=1,2,3.$$

Therefore, the lower bound in Theorem 3 is achieved for C_0, C_1 and C_2. By substituting M with 2^n-1, $2^{n-1}-1$ and $2^{n-1}+1$ into the right-hand side of (11) separately, we obtain the following results.

Corollary 1.

$$\begin{aligned}
\beta(n,2^n-1) &= \frac{n}{2}-\frac{n}{2(2^n-1)^2},\\
\beta(n,2^{n-1}-1) &= \frac{n-1}{2}-\frac{n-1}{2(2^{n-1}-1)^2},\\
\beta(n,2^{n-1}+1) &= \frac{n-1}{2}+\frac{2^{n+1}-n+1}{2(2^{n-1}+1)^2}.
\end{aligned}$$

By Lemmas 3 and 4, we know that for odd M,

$$d(C)=\frac{n}{2}-\frac{\hat{D}_1}{2}\leqslant\frac{n}{2}-\frac{n}{2M^2}.$$

This fact was first observed by Ahlswede and Katona (see [2, pp. 10]).

Below we improve Theorem 2 for odd M by using the same argument. From Lemma 3 and Lemma 4, we know that for odd M,

$$\begin{aligned}
\operatorname{var}(C) &= \frac{n}{4}-\frac{\hat{D}_1^2}{4}+\frac{1}{2}\left(\frac{2^n}{M}-1-\hat{D}_1-\hat{D}_3-\cdots-\hat{D}_n\right)\\
&\leqslant \frac{n}{4}-\frac{1}{4}\left[\frac{1}{M^2}\binom{n}{1}\right]^2+\frac{1}{2}\left[\frac{2^n}{M}-1-\frac{1}{M^2}\binom{n}{1}-\frac{1}{M^2}\binom{n}{3}-\cdots-\frac{1}{M^2}\binom{n}{n}\right]\\
&= \frac{n-2}{4}+\frac{2^{n-1}}{M}-\Delta_1,
\end{aligned}$$

where

$$\Delta_1=\frac{2^{n+1}-n(n-1)-2}{4M^2}+\frac{n^2}{4M^4}.$$

On the other hand,

$$\begin{aligned}
\hat{D}_1 &= \frac{2^n}{M}-1-\hat{D}_2-\cdots-\hat{D}_n\\
&\leqslant \frac{2^n}{M}-1-\frac{1}{M^2}\left[\binom{n}{2}+\binom{n}{3}+\cdots+\binom{n}{n}\right]\\
&= \frac{2^n}{M}-1-\frac{1}{M^2}(2^n-1-n).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\operatorname{var}(C) &\geqslant \frac{n}{4}-\frac{1}{4}\left[\frac{2^n}{M}-1-\frac{1}{M^2}(2^n-1-n)\right]^2+\frac{1}{2M^2}\binom{n}{2}\\
&= \frac{n-1}{4}+\frac{2^{n-1}}{M}-\frac{2^{2n-2}}{M^2}+\Delta_2,
\end{aligned}$$

where

$$\Delta_2 = \frac{n(n-1)}{4M^2} + \frac{1}{2M^2}(2^n - n - 1)\left[\frac{2^n}{M} - 1 - \frac{1}{2M^2}(2^n - 1 - n)\right].$$

The lower bound is achieved only when $\hat{D}_i = (1/M^2)\binom{n}{i}$, $i = 2, 3, \ldots, n$. Hence, Theorem 2 can be improved as follows.

Theorem 4. *If M is odd, then*

$$\frac{n-1}{4} + \frac{2^{n-1}}{M} - \frac{2^{2n-2}}{M^2} + \Delta_2 \leqslant \mathrm{var}(C) \leqslant \frac{n-2}{4} + \frac{2^{n-1}}{M} - \Delta_1.$$

The lower bound is achieved only when $\hat{D}_i = (1/M^2)\binom{n}{i}$, $i = 2, 3, \ldots, n$.

Similar to Corollary 1, we can obtain the following results by using the lower bound of Theorem 4.

Corollary 2.

$$\alpha(n, 2^n - 1) = \frac{n}{4} + \frac{n(n-1)}{4(2^n - 1)^2} - \frac{n^2}{4(2^n - 1)^4},$$

$$\alpha(n, 2^{n-1} - 1) = \frac{n-1}{4} + \frac{(n-1)(n-2)}{4(2^{n-1} - 1)^2} - \frac{(n-1)^2}{4(2^{n-1} - 1)^4},$$

$$\alpha(n, 2^{n-1} + 1) = \frac{n-1}{4} + \frac{n^2 + n + 2}{4(2^{n-1} + 1)^2} + \frac{2^{2n} + 2^n + n + 1}{4(2^{n-1} + 1)^4}.$$

Remark. (1) The values of $\beta(n, 2^n - 1)$ and $\alpha(n, 2^n - 1)$ can also be obtained directly from the properties of Hamming distance.

(2) We can also improve Theorems 1 and 2 for $M \equiv 2 \pmod 4$ by using Theorems 7 and 8 in [6].

4. Recursive inequalities of $\beta(n, M)$

Let C be a binary code with length n and size M. Let A be the binary $M \times n$ matrix, where the row vectors consist of all of the codewords of code C. Let $h_1, h_2, \ldots, h_n$ be the column vectors of A. It is not hard to see from (5) that

$$\hat{D}_1 = \frac{1}{M^2} \sum_{i=1}^{n} [M - 2w_H(h_i)]^2. \tag{12}$$

Let

$$W(C) = \frac{1}{M} \sum_{c \in C} w_H(c) \tag{13}$$

be the average Hamming weight of code C. By Lemma 3, (12) and the Cauchy inequality,

$$\begin{aligned} d(C) &= \frac{n}{2} - \frac{1}{2M^2}\sum_{i=1}^{n}[M - 2w_H(h_i)]^2 \\ &= 2\frac{1}{M}\sum_{i=1}^{n} w_H(h_i) - 2\frac{1}{M^2}\sum_{i=1}^{n} w_H^2(h_i) \\ &\leqslant 2W(C) - 2\frac{1}{M^2}\frac{[\sum_{i=1}^{n} w_H(h_i)]^2}{n} \\ &= 2W(C) - \frac{2}{n}W^2(C). \end{aligned} \tag{14}$$

From the above quadratic inequality, it is easy to obtain that

$$\frac{n}{2}\left[1 - \sqrt{1 - \frac{2}{n}d(C)}\right] \leqslant W(C) \leqslant \frac{n}{2}\left[1 + \sqrt{1 - \frac{2}{n}d(C)}\right]. \tag{15}$$

For a fixed codeword $c_0 \in C$, let $C = \{c_0\} \cup C^*$ and $c_0 + C^* = \{c_0 + a \mid a \in C^*\}$, we have

$$\begin{aligned} d(C) &= \frac{1}{M^2}\left[\sum_{a,b\in C^*} d_H(a,b) + 2\sum_{a\in C^*} d_H(c_0,a)\right] \\ &= \frac{1}{M^2}\left[(M-1)^2 d(C^*) + 2\sum_{a\in C^*} w_H(c_0 + a)\right] \\ &= \frac{1}{M^2}[(M-1)^2 d(C^*) + 2(M-1)W(c_0 + C^*)]. \end{aligned} \tag{16}$$

Note that $d(C^*) = d(c_0 + C^*)$. It follows from (15) and (16) that

$$d(C) \geqslant \frac{(M-1)^2}{M^2} d(C^*) + \frac{(M-1)n}{M^2}\left[1 - \sqrt{1 - \frac{2}{n}d(C^*)}\right], \tag{17}$$

$$d(C) \leqslant \frac{(M-1)^2}{M^2} d(C^*) + \frac{(M-1)n}{M^2}\left[1 + \sqrt{1 - \frac{2}{n}d(C^*)}\right]. \tag{18}$$

Let C be the binary (n, M) code such that $d(C) = \beta(n, M)$. Since $d(C^*) \geqslant \beta(n, M-1)$, it is easy to see from (17) that

$$\beta(n,M) \geqslant \frac{(M-1)^2}{M^2}\beta(n, M-1) + \frac{(M-1)n}{M^2}\left[1 - \sqrt{1 - \frac{2}{n}\beta(n, M-1)}\right]. \tag{19}$$

Let C^* be the binary $(n,M-1)$ code such that $d(C^*)=\beta(n,M-1)$. Since $d(C)\geqslant \beta(n,M)$, it is easy to see from (18) that

$$\beta(n,M)\leqslant \frac{(M-1)^2}{M^2}\beta(n,M-1)+\frac{(M-1)n}{M^2}\left[1+\sqrt{1-\frac{2}{n}\beta(n,M-1)}\right]. \qquad (20)$$

Theorem 5. *$\beta(n,M)$ satisfies the recursive inequalities* (19) *and* (20).

Acknowledgements

The authors wish to thank the anonymous referees and editors for their comments and suggestions that helped to improve the paper.

References

[1] I. Althöfer, T. Sillke, An "average distance" inequality for large subsets of the cube, J. Combin. Theory Ser. B 56 (1992) 296–301.
[2] R. Ahlswede, G. Katona, Contributions to the geometry of Hamming spaces, Discrete Math. 17 (1977) 1–22.
[3] R. Ahlswede, I. Althöfer, The asymptotic behaviour of diameters in the average, J. Combin. Theory Ser. B 61 (1994) 167–177.
[4] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland, New York, 1985.
[5] Fang-Wei Fu, Shi-Yi Shen, On the expectation and variance of Hamming distance between two binary i.i.d. random vectors, Acta Math. Appl. Sinica 13 (1997) 243–250.
[6] M.R. Best, A.E. Brouwer, F.J. MacWilliams, A.W. Odlyzko, N.J.A. Sloane, Bounds for binary codes of length less than 25, IEEE Trans. Inform. Theory IT-24 (1981) 81–93.