

Sharp Bound for t -Distance Sets of Diameter at Most $n/2$

Minghui Ouyang*

Abstract

For integers $n, t \geq 1$, we prove that every binary t -code $\mathcal{C} \subseteq \{0, 1\}^n$ with distance set $D \subseteq \{1, 2, \dots, \lfloor n/2 \rfloor\}$ satisfies $|\mathcal{C}| \leq \sum_{\substack{0 \leq j \leq t \\ j \equiv t \pmod{2}}} \binom{n}{j}$, which is sharp for $D = \{2, 4, \dots, 2t\}$. The proof uses the annihilating polynomial associated with D and shows that its coefficients in the Krawtchouk basis follow an alternating sign pattern. The result then follows from a rank bound of Barg and Musin.

1 Introduction

Let $\mathcal{H}_n = \{0, 1\}^n$ be the Hamming cube endowed with the distance $d_H(\mathbf{x}, \mathbf{y}) := |\{i \in [n] : x_i \neq y_i\}|$. For $\mathcal{C} \subseteq \mathcal{H}_n$, write $D(\mathcal{C}) := \{d_H(\mathbf{x}, \mathbf{y}) : \mathbf{x}, \mathbf{y} \in \mathcal{C}, \mathbf{x} \neq \mathbf{y}\}$. A family $\mathcal{C}$ is called a t -distance set or a binary t -code if $|D(\mathcal{C})| = t$. Let $A(\mathcal{H}_n, D)$ denote the maximum cardinality of a family $\mathcal{C} \subseteq \mathcal{H}_n$ with distance set D .

Kleitman [9] first studied this problem for $D = \{1, 2, \dots, d\}$ and established the discrete isodiametric inequality. Delsarte [3] subsequently proved $|\mathcal{C}| \leq \sum_{j=0}^t \binom{n}{j}$ for general t -distance sets. Barg and Musin [2], Barg, Glazyrin, Kao, Lai, Tseng and Yu [1], and Musin and Nozaki [11] later obtained more precise estimates for families with few distances. For uniform set families¹, Ray-Chaudhuri and Wilson [13] proved that $|\mathcal{C}| \leq \binom{n}{t}$ when exactly t distances occur. Delsarte's bound has the correct order of magnitude² as $n \rightarrow \infty$ for fixed t . However, the exact maximum size remains unknown in general.

For a fixed distance set D , Huang, Klurman and Pohoata [8] studied the case where D is an interval of integers, obtaining bounds of the correct order and sharp asymptotics in special cases. Dong, Gao, Liu, Ouyang and Zhou [5] later determined the asymptotic maximum family size for homogeneous arithmetic progressions. By analyzing intersections with a fixed set, their approach translates distance constraints into intersection constraints and applies classical results on set families [4, 6], allowing $\max D = O(\log n)$ for fixed t with a sufficiently small implied constant.

In this paper, we build on the work of Barg and Musin [2] to obtain a sharp upper bound on the size of t -distance sets under the condition $\max D \leq n/2$. We use the Krawtchouk expansion of the annihilating polynomial associated with D , together with a rank bound that retains only terms corresponding to positive coefficients in this expansion [1, 2].

Theorem 1.1. *For integers $n, t \geq 1$, let $D \subseteq \{1, 2, \dots, \lfloor n/2 \rfloor\}$ with $|D| = t$. If $\mathcal{C} \subseteq \mathcal{H}_n$ satisfies $d_H(\mathbf{x}, \mathbf{y}) \in D$ for all distinct $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, then $|\mathcal{C}| \leq \sum_{\substack{0 \leq j \leq t \\ j \equiv t \pmod{2}}} \binom{n}{j}$.*

Moreover, this bound is attained by the family $\mathcal{F}_t := \{A \subseteq [n] : |A| \leq t, |A| \equiv t \pmod{2}\}$, whose distance set is $\{2, 4, \dots, 2t\}$.

*School of Mathematical Sciences, Peking University, Beijing 100871, China. ouyangminghui1998@gmail.com.

¹A set family with constant weight.

²For $n \geq 2t$, the family of all t -element subsets of $[n]$ has $\binom{n}{t}$ members and distance set $\{2, 4, \dots, 2t\}$.

For comparison, Barg and Musin [2] proved that $A(\mathcal{H}_n, D) \leq \binom{n}{t} + \sum_{j=0}^{t-2} \binom{n}{j}$ for $t \geq 2$ whenever $\sum_{d \in D} d \leq tn/2$. Under the stronger assumption $\max D \leq n/2$, Theorem 1.1 improves this bound by removing the extra terms $\binom{n}{t-3}, \binom{n}{t-5}, \dots$. Hegedüs [7] also obtained the same upper bound under the different assumption that $D(\mathcal{C})$ is invariant under $d \mapsto n - d$.

Throughout the paper, $[n] := \{1, \dots, n\}$ and $\binom{[n]}{j}$ denotes the collection of j -element subsets of $[n]$. We write $B_t(n)$ for the upper bound in Theorem 1.1. We set $\binom{n}{j} = 0$ if $j < 0$ or $j > n$, and $A(\mathcal{H}_n, D) = 0$ if no family in $\mathcal{H}_n$ has distance set D . A sum over an empty index set is zero. All matrices and polynomial coefficients are real. The notation $\mathbf{K} \succeq 0$ means that $\mathbf{K}$ is positive semidefinite.

2 Proofs

For $\mathbf{x} \in \mathcal{H}_n$, let $\mathbf{u}_{\mathbf{x}} := ((-1)^{x_1}, \dots, (-1)^{x_n}) \in \{-1, 1\}^n$ be its associated sign vector. Then $\langle \mathbf{u}_{\mathbf{x}}, \mathbf{u}_{\mathbf{y}} \rangle = n - 2d_H(\mathbf{x}, \mathbf{y})$. For $S \subseteq [n]$, write $\chi_S(\mathbf{u}) := \prod_{i \in S} u_i$, with $\chi_{\emptyset}(\mathbf{u}) = 1$.

We define the Krawtchouk polynomials [2, 10] as follows.

Definition 2.1. Fix an integer $n \geq 1$. For $0 \leq j \leq n$, the *Krawtchouk polynomial* in d is

$$K_j(d; n) := \sum_{\ell=0}^j (-1)^\ell \binom{d}{\ell} \binom{n-d}{j-\ell}.$$

Here the binomial coefficients are interpreted as polynomials: $\binom{a}{0} = 1$ and $\binom{a}{k} = a(a-1) \cdots (a-k+1)/k!$ for $a \in \mathbb{R}$ and integers $k \geq 1$. Moreover, we define

$$p_j(z) := K_j\left(\frac{n-z}{2}; n\right) = \sum_{\ell=0}^j (-1)^\ell \binom{(n-z)/2}{\ell} \binom{(n+z)/2}{j-\ell}.$$

The dependence of p_j on n is left implicit.

With the convention $p_{-1}(z) = 0$, these polynomials satisfy $p_0(z) = 1$ and the three-term recurrence

$$zp_j(z) = (j+1)p_{j+1}(z) + (n-j+1)p_{j-1}(z), \quad 0 \leq j < n. \quad (1)$$

The recurrence implies that p_j has degree j , leading coefficient $1/j!$, and parity $p_j(-z) = (-1)^j p_j(z)$. For $n \geq 3$, the first few polynomials are $p_1(z) = z$, $p_2(z) = (z^2 - n)/2$, and $p_3(z) = (z^3 - (3n-2)z)/6$. In particular, $p_0, \dots, p_t$ form a basis for the polynomials of degree at most t , whenever $t \leq n$.

We use the following identity [2] for Krawtchouk polynomials and give a direct proof.

Lemma 2.2. For $\mathbf{u}, \mathbf{v} \in \{-1, 1\}^n$ and $0 \leq j \leq n$, we have

$$p_j(\langle \mathbf{u}, \mathbf{v} \rangle) = \sum_{S \in \binom{[n]}{j}} \chi_S(\mathbf{u}) \chi_S(\mathbf{v}). \quad (2)$$

Consequently, if $\mathcal{C} \subseteq \mathcal{H}_n$, then the matrix $\mathbf{E}_j := (p_j(\langle \mathbf{u}_{\mathbf{x}}, \mathbf{u}_{\mathbf{y}} \rangle))_{\mathbf{x}, \mathbf{y} \in \mathcal{C}}$ is positive semidefinite and satisfies $\text{rank } \mathbf{E}_j \leq \binom{n}{j}$.

Proof. Put $w_i = u_i v_i$, $\mathbf{w} = (w_1, \dots, w_n)$, and $w_S := \prod_{i \in S} w_i$. Then $w_i^2 = 1$. Define $e_j(\mathbf{w}) := \sum_{S \in \binom{[n]}{j}} w_S$,

with $e_{-1}(\mathbf{w}) = 0$ and $e_0(\mathbf{w}) = 1$. For $0 \leq j < n$, splitting the sum over $i \in [n]$ according to whether $i \in S$ gives

$$\left(\sum_{i=1}^n w_i \right) e_j(\mathbf{w}) = \sum_{S \in \binom{[n]}{j}} \sum_{i \notin S} w_i w_S + \sum_{S \in \binom{[n]}{j}} \sum_{i \in S} w_i w_S.$$

If $i \notin S$, then $w_i w_S = w_{S \cup \{i\}}$. Each fixed set T of size $j+1$ occurs for exactly $j+1$ pairs (S, i) , namely $S = T \setminus \{i\}$ with $i \in T$. If $i \in S$, then $w_i w_S = w_i^2 w_{S \setminus \{i\}} = w_{S \setminus \{i\}}$. For $j \geq 1$, each fixed set T of size $j-1$ occurs for exactly $n-j+1$ pairs, namely $S = T \cup \{i\}$ with $i \in [n] \setminus T$. For $j = 0$, the second sum is empty. Hence

$$\sum_{S \in \binom{[n]}{j}} \sum_{i \notin S} w_i w_S = (j+1) e_{j+1}(\mathbf{w}), \quad \sum_{S \in \binom{[n]}{j}} \sum_{i \in S} w_i w_S = (n-j+1) e_{j-1}(\mathbf{w}).$$

Adding these expressions yields

$$\left(\sum_{i=1}^n w_i \right) e_j(\mathbf{w}) = (j+1) e_{j+1}(\mathbf{w}) + (n-j+1) e_{j-1}(\mathbf{w}), \quad 0 \leq j < n.$$

Since $\sum_i w_i = \langle \mathbf{u}, \mathbf{v} \rangle$, this is the recurrence (1) with the same initial values. Induction therefore gives $e_j(\mathbf{w}) = p_j(\langle \mathbf{u}, \mathbf{v} \rangle)$ for $0 \leq j \leq n$. Since $w_S = \chi_S(\mathbf{u}) \chi_S(\mathbf{v})$, this proves (2).

Now put $M = |\mathcal{C}|$. For $S \subseteq [n]$, let $\mathbf{h}_S := (\chi_S(\mathbf{u}_x))_{x \in \mathcal{C}} \in \mathbb{R}^M$. The identity gives the matrix decomposition

$$\mathbf{E}_j = \sum_{S \in \binom{[n]}{j}} \mathbf{h}_S \mathbf{h}_S^\top.$$

For every $\mathbf{a} \in \mathbb{R}^M$,

$$\mathbf{a}^\top \mathbf{E}_j \mathbf{a} = \sum_{S \in \binom{[n]}{j}} (\mathbf{a}^\top \mathbf{h}_S)^2 \geq 0,$$

so $\mathbf{E}_j$ is positive semidefinite. Each summand has rank at most one, and rank subadditivity gives

$$\text{rank } \mathbf{E}_j \leq \sum_{S \in \binom{[n]}{j}} \text{rank}(\mathbf{h}_S \mathbf{h}_S^\top) \leq \binom{n}{j}. \quad \square$$

The next lemma bounds $|\mathcal{C}|$ by a sum of binomial coefficients, retaining only terms corresponding to positive Krawtchouk coefficients. This bound is stated in [1] and follows from the general argument of [2]. Its spherical version is given by Nozaki and Shinohara [12]. We state it here for any annihilating polynomial F satisfying $F(n) > 0$ and include a short proof using a rank argument for completeness.

Lemma 2.3. *Let $\mathcal{C} \subseteq \mathcal{H}_n$ and $0 \leq s \leq n$. Suppose that $F(z) = \sum_{j=0}^s f_j p_j(z)$ is a real polynomial*

satisfying $F(n) > 0$ and $F(n - 2d_H(\mathbf{x}, \mathbf{y})) = 0$ for all distinct $\mathbf{x}, \mathbf{y} \in \mathcal{C}$. Then $|\mathcal{C}| \leq \sum_{\substack{0 \leq j \leq s \\ f_j > 0}} \binom{n}{j}$.

Proof. We may assume $M := |\mathcal{C}| > 0$. By Lemma 2.2 and the expansion of F , the matrix $\sum_{j=0}^s f_j \mathbf{E}_j$ has diagonal entries $F(n)$ and off-diagonal entries $F(n - 2d_H(\mathbf{x}, \mathbf{y})) = 0$ by assumption. Hence $F(n) \mathbf{I}_M = \sum_{j=0}^s f_j \mathbf{E}_j$. Separating the terms according to the signs of f_j , we obtain

$$\sum_{\substack{0 \leq j \leq s \\ f_j > 0}} f_j \mathbf{E}_j = F(n) \mathbf{I}_M + \sum_{\substack{0 \leq j \leq s \\ f_j < 0}} (-f_j) \mathbf{E}_j.$$

Since $F(n) > 0$ and every $\mathbf{E}_j$ is positive semidefinite, the matrix on the right is positive definite. Hence $\sum_{f_j > 0} f_j \mathbf{E}_j$ has full rank M . On the other hand, subadditivity of rank and the rank estimate in Lemma 2.2 imply

$$M \leq \sum_{\substack{0 \leq j \leq s \\ f_j > 0}} \text{rank } \mathbf{E}_j \leq \sum_{\substack{0 \leq j \leq s \\ f_j > 0}} \binom{n}{j}. \quad \square$$

For the annihilating polynomial $P_D(z) = \prod_{d \in D} (z - (n - 2d))$, we have $P_D(n) = \prod_{d \in D} 2d > 0$.

Ignoring the coefficient signs in Lemma 2.3 recovers Delsarte's bound $|\mathcal{C}| \leq \sum_{j=0}^t \binom{n}{j}$. To improve this bound to $B_t(n)$, it suffices to show that the coefficients of P_D in degrees of parity opposite to t are nonpositive. The next two lemmas establish the required signs for all degrees simultaneously.

We first prove an elementary consequence of the recurrence (1).

Lemma 2.4. *For each $0 \leq k \leq n$, we can write $z^k = \sum_{\substack{0 \leq j \leq k \\ j \equiv k \pmod{2}}} a_{k,j} p_j(z)$ with nonnegative coefficients $a_{k,j}$. Moreover, $a_{k,k} = k!$.*

Proof. For $k = 0$, we have $z^0 = 1 = p_0(z)$, so $a_{0,0} = 1$. Suppose the assertion holds for some $0 \leq k < n$, and write

$$z^k = \sum_{j=0}^k a_{k,j} p_j(z), \quad a_{k,j} \geq 0.$$

We set $a_{k,j} = 0$ whenever $j < 0$, $j > k$, or $j \not\equiv k \pmod{2}$. Multiplying by z and using (1), we obtain

$$\begin{aligned} z^{k+1} &= \sum_{j=0}^k a_{k,j} z p_j(z) \\ &= \sum_{j=0}^k a_{k,j} ((j+1)p_{j+1}(z) + (n-j+1)p_{j-1}(z)) \\ &= \sum_{j=0}^{k+1} (j a_{k,j-1} + (n-j) a_{k,j+1}) p_j(z). \end{aligned}$$

Thus the coefficients in the new expansion are

$$a_{k+1,j} = j a_{k,j-1} + (n-j) a_{k,j+1} \geq 0, \quad 0 \leq j \leq k+1.$$

The inequality follows because $j \geq 0$ and $n-j \geq 0$ for $j \leq k+1 \leq n$. If $j \not\equiv k+1 \pmod{2}$, then both $j-1$ and $j+1$ have parity opposite to k , so $a_{k,j-1} = a_{k,j+1} = 0$ and hence $a_{k+1,j} = 0$. This proves the required expansion by induction.

Finally, taking $j = k+1$ in the coefficient recurrence gives

$$a_{k+1,k+1} = (k+1) a_{k,k} + (n-k-1) a_{k,k+2} = (k+1) a_{k,k}.$$

Since $a_{0,0} = 1$, it follows that $a_{k,k} = k!$ for all $0 \leq k \leq n$. $\square$

The next lemma then determines the signs of the Krawtchouk coefficients.

Lemma 2.5. *Let $1 \leq t \leq n$ and $r_1, \dots, r_t \geq 0$. Write $P(z) = \prod_{i=1}^t (z - r_i) = \sum_{j=0}^t c_j p_j(z)$. Then $(-1)^{t-j} c_j \geq 0$ for $0 \leq j \leq t$.*

Proof. For $0 \leq q \leq t$, let $e_q = e_q(r_1, \dots, r_t)$ be the elementary symmetric polynomial of degree q , with $e_0 = 1$. All e_q are nonnegative since all $r_i \geq 0$. Expanding the product gives $P(z) = \sum_{k=0}^t (-1)^{t-k} e_{t-k} z^k$. Now apply Lemma 2.4. For a fixed index j , only powers z^k with $k \equiv j \pmod{2}$ contribute to the coefficient of p_j . Their signs agree, because $(-1)^{t-k} = (-1)^{t-j}$. More explicitly,

$$c_j = (-1)^{t-j} \sum_{\substack{j \leq k \leq t \\ k \equiv j \pmod{2}}} e_{t-k} a_{k,j}.$$

After factoring out $(-1)^{t-j}$, every contribution to c_j is nonnegative. This proves the claimed sign pattern for all degrees. $\square$

We now complete the proof of Theorem 1.1. Applying Lemma 2.5 to P_D shows that the terms in degrees of parity opposite to t can be removed from the rank bound in Lemma 2.3. This yields the desired upper bound.

Proof of Theorem 1.1. Write $D = \{d_1, \dots, d_t\}$ and put $r_i := n - 2d_i$ and $P_D(z) := \prod_{i=1}^t (z - r_i)$. The condition $d_i \leq n/2$ gives $r_i \geq 0$ for every i .

If $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ are distinct, then $d_H(\mathbf{x}, \mathbf{y}) \in D$, so the identity $\langle \mathbf{u}_x, \mathbf{u}_y \rangle = n - 2d_H(\mathbf{x}, \mathbf{y})$ yields $P_D(\langle \mathbf{u}_x, \mathbf{u}_y \rangle) = 0$. We also have $P_D(n) = \prod_{i=1}^t (n - r_i) = \prod_{i=1}^t 2d_i > 0$.

Write $P_D(z) = \sum_{j=0}^t c_j p_j(z)$. By Lemma 2.5, $c_j \leq 0$ whenever $j \not\equiv t \pmod{2}$. Applying Lemma 2.3, we obtain

$$|\mathcal{C}| \leq \sum_{\substack{0 \leq j \leq t \\ c_j > 0}} \binom{n}{j} \leq \sum_{\substack{0 \leq j \leq t \\ j \equiv t \pmod{2}}} \binom{n}{j} = B_t(n).$$

Finally, for $D = \{2, 4, \dots, 2t\}$ with $2t \leq n/2$, we have $|\mathcal{F}_t| = \sum_{\substack{0 \leq j \leq t \\ j \equiv t \pmod{2}}} \binom{n}{j} = B_t(n)$. Distances

within $\mathcal{F}_t$ are even and at most $2t$, proving the sharpness. $\square$

3 Concluding remarks

Lemma 2.5 applies to every monic polynomial of degree at most n with nonnegative roots. By Lemma 2.4, the Krawtchouk expansion of each monomial has nonnegative coefficients and contains only degrees of the same parity. These two facts make all contributions to a fixed Krawtchouk coefficient have the same sign. For the annihilating polynomial P_D , the condition $\max D \leq n/2$ ensures that every root $n - 2d$ is nonnegative. The resulting upper bound holds for all admissible distance sets and remains valid when t and D vary with n .

If an allowed distance d exceeds $n/2$, the corresponding root $n - 2d$ of the annihilating polynomial is negative. The elementary symmetric polynomials in the proof of Lemma 2.5 are then no longer guaranteed to be nonnegative. Consequently, coefficients c_j with $j \not\equiv t \pmod{2}$ may be positive, and their terms $\binom{n}{j}$ need further sign information before they can be removed. Lemma 2.3 still

applies: P_D vanishes at the inner products of distinct codewords and $P_D(n) = \prod_{i=1}^t 2d_i > 0$. We believe that for each fixed $t \geq 2$ and all sufficiently large³ n , the bound $|\mathcal{C}| \leq B_t(n)$ remains valid even when $\max D > n/2$. A different choice of annihilating polynomial may be helpful.

³As noted by Barg et al. [1, Section 6.1], the dual Golay code of length 23 is a three-distance set with 2048

References

- [1] A. Barg, A. Glazyrin, W.-J. Kao, C.-Y. Lai, P.-C. Tseng and W.-H. Yu, On the size of maximal binary codes with 2, 3, and 4 distances, *Combin. Theory* **4** (2024), no. 1, Paper No. 7, 28 pp.
- [2] A. Barg and O. R. Musin, Bounds on sets with few distances, *J. Combin. Theory Ser. A* **118** (2011), no. 4, 1465–1474.
- [3] P. Delsarte, Four fundamental parameters of a code and their combinatorial significance, *Information and Control* **23** (1973), no. 5, 407–438.
- [4] M. Deza, P. Erdős and P. Frankl, Intersection properties of systems of finite sets, *Proc. London Math. Soc.* (3) **36** (1978), no. 2, 369–384.
- [5] Z. Dong, J. Gao, H. Liu, M. Ouyang and Q. Zhou, Set families: restricted distances via restricted intersections, *Math. Proc. Cambridge Philos. Soc.* **181** (2026), no. 2, 1001–1020.
- [6] P. Frankl and R. M. Wilson, Intersection theorems with geometric consequences, *Combinatorica* **1** (1981), no. 4, 357–368.
- [7] G. Hegedüs, Upper bounds for the size of set systems with a symmetric set of Hamming distances, *Acta Math. Hungar.* **171** (2023), no. 1, 176–182.
- [8] H. Huang, O. Klurman and C. Pohoata, On subsets of the hypercube with prescribed Hamming distances, *J. Combin. Theory Ser. A* **171** (2020), Article 105156, 21 pp.
- [9] D. J. Kleitman, On a combinatorial conjecture of Erdős, *J. Combin. Theory* **1** (1966), no. 2, 209–214.
- [10] M. Krawtchouk, Sur une généralisation des polynômes d’Hermite, *C. R. Acad. Sci. Paris* **189** (1929), 620–622.
- [11] O. R. Musin and H. Nozaki, Bounds on three- and higher-distance sets, *European J. Combin.* **32** (2011), no. 8, 1182–1190.
- [12] H. Nozaki and M. Shinohara, On a generalization of distance sets, *J. Combin. Theory Ser. A* **117** (2010), no. 7, 810–826.
- [13] D. K. Ray-Chaudhuri and R. M. Wilson, On t -designs, *Osaka J. Math.* **12** (1975), no. 3, 737–744.

members, exceeding $B_3(23) = \binom{23}{3} + 23 = 1794$. Thus, we require n to be sufficiently large in terms of t . However, the bound in Theorem 1.1 holds for all n and t under $\max D \leq n/2$.