Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Bisecting and *D*-secting families for set systems

Niranjan Balachandran [a], Rogers Mathew [b], Tapas Kumar Mishra [b,*], Sudebkumar Prasant Pal [b]

[a] *Department of Mathematics, Indian Institute of Technology, Bombay 400076, India*
[b] *Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur 721302, India*

ARTICLE INFO

Article history:
Received 2 May 2016
Received in revised form 9 March 2017
Accepted 8 May 2017
Available online xxxx

Keywords:
Discrepancy
Hypergraphs
Separating family
Bisecting families

ABSTRACT

Let n be any positive integer and $\mathcal{F}$ be a family of subsets of $[n]$. A family $\mathcal{F}'$ is said to be *D-secting* for $\mathcal{F}$ if for every $A \in \mathcal{F}$, there exists a subset $A' \in \mathcal{F}'$ such that $|A \cap A'| - |A \cap ([n] \setminus A')| = i$, where $i \in D$, $D \subseteq \{-n, -n+1, \ldots, 0, \ldots, n\}$. A *D-secting* family $\mathcal{F}'$ of $\mathcal{F}$, where $D = \{-1, 0, 1\}$, is a *bisecting* family ensuring the existence of a subset $A' \in \mathcal{F}'$ such that $|A \cap A'| \in \{\lceil \frac{|A|}{2} \rceil, \lfloor \frac{|A|}{2} \rfloor\}$, for each $A \in \mathcal{F}$. In this paper, we study D-secting families for $\mathcal{F}$ with restrictions on D, and the cardinalities of $\mathcal{F}$ and the subsets of $\mathcal{F}$.

1. Introduction

Let n be any positive integer and $\mathcal{F}$ be a family of subsets of $[n]$. Another family $\mathcal{F}'$ of subsets of $[n]$ is called a *bisecting family* for $\mathcal{F}$, if for each subset $A \in \mathcal{F}$, there exists a subset $A' \in \mathcal{F}'$ such that $|A \cap A'| \in \{\lceil \frac{|A|}{2} \rceil, \lfloor \frac{|A|}{2} \rfloor\}$. What is the minimum cardinality of a bisecting family for any family $\mathcal{F}$? We pose a more general problem based on the difference between $|A \cap A'|$ and $|A \cap ([n] \setminus A')|$. We say a family $\mathcal{F}'$ is *D-secting* for $\mathcal{F}$ if for each subset $A \in \mathcal{F}$, there exists a subset $A' \in \mathcal{F}'$ such that $|A \cap A'| - |A \cap ([n] \setminus A')| = i$, where $i \in D$, $D \subseteq \{-n, -n+1, \ldots, 0, \ldots, n\}$. Let $\beta_D(\mathcal{F})$ denote the minimum cardinality of a D-secting family for $\mathcal{F}$. In particular, when $D = \{-1, 0, 1\}$, the family $\mathcal{F}'$ becomes a bisecting family for $\mathcal{F}$. We study two cases depending on D: (i) $D = \{-i, -i+1, \ldots, 0, \ldots, i\}$, and (ii) $D = \{i\}$, for some $i \in [n]$. Observe that if $D = \{i\}$, only those sets $A \in \mathcal{F}$ for which $|A| \cong i \pmod 2$ can attain a value of i for $|A \cap A'| - |A \cap ([n] \setminus A')|$. So, we consider only those sets for which $|A| \cong i \pmod 2$, when $D = \{i\}$. We define $\beta_D(n)$ as the maximum of $\beta_D(\mathcal{F})$ over all families $\mathcal{F}$ on $[n]$ and $\beta_D(n, k)$ as the maximum of $\beta_D(\mathcal{F})$ over all families $\mathcal{F} \subseteq \binom{[n]}{k}$. When $D = \{i\}$ ($D = \{-i, -i+1, \ldots, i\}$), we sometimes abuse the notation to denote $\beta_D(\mathcal{F})$ by $\beta_i(\mathcal{F})$ (resp., $\beta_{[\pm i]}(\mathcal{F})$).

Consider an example family $\mathcal{F}$ which consists of all the 4-element subsets of $\{1, \ldots, 6\}$. Note that since each subset $A \in \mathcal{F}$ has an even cardinality, $\beta_0(\mathcal{F}) = \beta_{[\pm 1]}(\mathcal{F})$. Let $\mathcal{F}' = \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 5\}\}$. It is not hard to verify that every 4-element subset $A \in \mathcal{F}$ is bisected by at least one element in $\mathcal{F}'$. So, $\beta_0(\mathcal{F}) \le 3$, for $\mathcal{F} = \binom{[6]}{4}$. In fact there is no pair of subsets of $\{1, \ldots, 6\}$ such that every 4-element subset $A \in \mathcal{F}$ is bisected by one of them, which is asserted by Proposition 21. Therefore, $\beta_0(\mathcal{F}) = 3$.

* Corresponding author.
E-mail addresses: niranj@iitb.ac.in (N. Balachandran), rogers@cse.iitkgp.ernet.in (R. Mathew), tkmishra@cse.iitkgp.ernet.in (T.K. Mishra), spp@cse.iitkgp.ernet.in (S.P. Pal).

http://dx.doi.org/10.1016/j.dam.2017.05.005
0166-218X/

Discrepancy and D-secting families

Bisecting families may also be interpreted in terms of 'discrepancy' of hypergraphs under multiple bicolorings. Let $G(V, E)$ be a hypergraph with vertex set $V = \{v_1, \ldots, v_n\}$ and hyperedge set $E = \{e_1, \ldots, e_m\}$. Given a bicoloring X, $X : V \to \{-1, +1\}$, let $\mathbb{C}_X(e) = |\sum_{v \in e} X(v)|$ denote the discrepancy of the hyperedge e under the bicoloring X. Then, the *discrepancy* of the hypergraph G, denoted by $disc(G)$, is defined as $disc(G) = \min_X \max_{e \in E} \mathbb{C}_X(e)$. For definitions, results, and extensions of discrepancy and related problems, see [2,5,9,13]. Below, we define $\beta_D(E)$ in terms of the discrepancy of a hypergraph $G(V, E)$, where $D = [\pm i]$. Let $t \in \mathbb{N}$ be the minimum number such that there exists a set of t hypergraphs $G_1, \ldots, G_t$ on vertex set $V = [n]$ with (i) $disc(G_j) \in [\pm i]$, for $1 \le j \le t$, and, (ii) $\cup_{j=1}^{t} G_j = G(V, E)$. Given an optimal D-secting family $\mathcal{F}'$ of E, it is easy to construct a set of hypergraphs $G_1, \ldots, G_{|\mathcal{F}'|}$ satisfying the above conditions. Again, given a set of t hypergraphs $G_1, \ldots, G_t$ satisfying conditions (i) and (ii) under bicolorings $X_1, \ldots, X_t$, respectively, let (A_j^{+1}, A_j^{-1}) be the bipartition of V formed by the bicoloring X_j. Then, $\mathcal{F}' = \{A_1^{+1}, \ldots, A_t^{+1}\}$ is a D-secting family for E. Thus, $\beta_{[\pm i]}(E) = t$. Moreover, the discrepancy of a hypergraph $G([n], E)$ can be defined in terms of $\beta_{[\pm i]}(E)$ as follows. The *discrepancy* of a hypergraph $G([n], E)$ is the minimum $i \in \mathbb{N}$ such that $\beta_{[\pm i]}(E) = 1$.

Separating and bisecting families

Given a family $\mathcal{F}$ of subsets of $[n]$, finding another family $\mathcal{F}'$ with certain properties has been well investigated. One of the most studied problem in this direction is the computation of *separating families*. Let $\mathcal{F}$ consist of pairs $\{i, j\}$, $i, j \in \mathbb{N}$, $i \ne j$ and $\mathcal{F}' = \{A'_1, \ldots, A'_t\}$ be another family of subsets on $[n]$ ($\mathcal{F}$ can be viewed as the edge set of a graph on vertex set $[n]$). A subset A'_l separates a pair $\{i, j\}$ if $i \in A'_l$ and $j \notin A'_l$ or vice versa, $l \in [t]$. The family $\mathcal{F}'$ is a separating family for $\mathcal{F}$ if every pair $\{i, j\} \in \mathcal{F}$ is separated by some $A' \in \mathcal{F}'$. It is easy to see that $\mathcal{F}'$ is indeed a bisecting family for $\mathcal{F}$. Let $f(n)$ denote the size of a minimum separating family $\mathcal{F}'$ for a family $\mathcal{F}$ consisting of all the $\binom{n}{2}$ pairs (edge set of a complete graph on n vertices). Rényi [17] proved that $f(n) = \lceil \log_2 n \rceil$. Observe that $f(n)$ is the minimum number of bipartite graphs needed to cover the edges of a complete graph K_n. We note the following generalization of the above statement for arbitrary graphs.

Proposition 1 (*Folklore*). *Let $\chi(G)$ denote the chromatic number of graph G. Then, $\lceil \log_2 \chi(G) \rceil$ bipartite graphs are necessary and sufficient to cover the edges of G.*

Note that $f(n)$ is equal to $\beta_0(n, 2)$, thus $\beta_0(n, 2) = \lceil \log_2 n \rceil$. In fact, when the family $\mathcal{F}$ is the edge set of a graph $G(V, E)$, where $V = [n]$, any bisecting family $\mathcal{F}'$ for $\mathcal{F}$ forms a covering of the edges of G with $|\mathcal{F}'|$ bipartite graphs. We state these observations as a corollary below.

Corollary 2. *For a graph $G(V, E)$, $\beta_0(E) = \lceil \log_2 \chi(G) \rceil$. Thus, $\beta_0(n, 2) = \lceil \log_2 n \rceil$.*

See [10,17,19] for details on separating families.

1.1. Notations and definitions

Let $[n]$ denote the set of integers $\{1, \ldots, n\}$, $\pm i$ denote the set of integers $\{-i, i\}$, and $[\pm i]$ denote the set of integers $\{-i, -i+1, \ldots, i\}$. Let $\mathcal{F}$ denote a family of subsets of $[n]$ and $\mathcal{F}'$ denote another family of subsets with some desired intersection property with elements of $\mathcal{F}$. Let $\binom{[n]}{k}$ denote the family of all the k-sized subsets of $[n]$. We use $\beta_{[\pm i]}(\mathcal{F})$ (resp., $\beta_i(\mathcal{F})$) to denote $\beta_D(\mathcal{F})$ if $D = [\pm i]$ (resp., $D = \{i\}$). We denote an n-dimensional vector $R \in \{0, 1\}^n$ (or $\{-1, +1\}^n$) as $R = (x_1, \ldots, x_n)$ where $x_j \in \{0, 1\}$ (resp., $\{-1, +1\}$). The *weight* of a vector $R = (x_1, \ldots, x_n) \in \{0, 1\}^n$ (or $\{-1, +1\}^n$) is the number of x_j's which are 1 (resp., -1), $1 \le j \le n$. Vector $R \in \{0, 1\}^n$ is even (resp., odd) if the number of 1's in R is even (resp., odd). A vector $R \in \{-1, 1\}^n$ is even (resp., odd) if the number of -1's in R is even (resp., odd). We use log to denote $\log_2$ in the rest of the paper.

1.2. Our contribution

We begin by addressing the problem of bounding and computing $\beta_D(n)$, where $D = [\pm i]$. We demonstrate a construction yielding an upper bound of $\lceil \frac{n}{2i} \rceil$ for $\beta_{[\pm i]}(n)$. Further, we show using a polynomial representation for the parity function that $\lceil \frac{n}{2i} \rceil$ is also a lower bound for $\beta_{[\pm i]}(n)$.

Theorem 3. $\beta_{[\pm i]}(n) = \lceil \frac{n}{2i} \rceil$, $n \in \mathbb{N}$, $i \in [n]$.

We study $\beta_{[\pm i]}(\mathcal{F})$ for a family $\mathcal{F}$ on $[n]$, in terms of i and $|\mathcal{F}|$, using Chernoff's bound.

Theorem 4. *Let $\mathcal{F}$ be a family of subsets of $[n]$ and let $m = |\mathcal{F}|$. Let $D = [\pm i]$, where $i \ge \sqrt{\frac{3n \ln(2m)}{t}}$ and $t \le \frac{1}{2} \log m$. Then, $\beta_D(\mathcal{F}) \le t$.*

In particular, if $i \geq \sqrt{4.2n+1}$ and $|\mathcal{F}| = O(n^c)$, for $c \in \mathbb{N}$, a D-secting family $\mathcal{F}'$ of cardinality $O(\log n)$ can be computed for families $\mathcal{F}$, thus improving the bound from Theorem 3 for this range of i and $|\mathcal{F}|$.

Subsequently, we study $\beta_D(n)$, where D is a singleton set, i.e., $D = \{i\}$. Note that $\beta_i(n) = \beta_{-i}(n)$. Moreover, when $D = \{-i, i\}$, note that $\beta_{\pm i}(n) \leq \beta_i(n) \leq 2\beta_{\pm i}(n)$. Therefore, we focus on establishing bounds for $\beta_i(n)$. We demonstrate a construction to show that $\beta_1(n)$ is at most $\lceil \frac{n}{2} \rceil$. We also show that $\beta_1(n)$ is at least $\lceil \frac{n}{2} \rceil$ using arguments similar to those in the proof of Theorem 3 about $\beta_{[\pm 1]}(n)$. In Section 3.2, we establish a lower bound of $\frac{n-i+1}{2}$ for arbitrary $i \in [n]$, $i \geq 2$. We demonstrate a construction establishing $\beta_i(n) \leq n - i + 1$. We have the following theorem.

Theorem 5. $\frac{n-i+1}{2} \leq \beta_i(n) \leq n - i + 1$, $n \in \mathbb{N}$, $i \in [n]$.

In Section 4, we consider families $\mathcal{F}$, $\mathcal{F} \subseteq \binom{[n]}{k}$. We study $\beta_{[\pm 1]}(n, k)$ in detail when k is even; the analysis for $\beta_i(n, k)$ for $i \in [n]$ and for the case when k is odd is analogous. We have lower bounds for $\beta_{[\pm 1]}(n, k)$ given by Theorem 6, Observation 11 (see Section 1.3), and Theorem 7 which are useful when k is a constant, k is sublinear in n, and k is linear in n, respectively. We establish the following theorem using entropy based arguments.

Theorem 6.

$$\beta_{[\pm 1]}(n, k) \geq \begin{cases} \log(n - k + 2), \textit{ when } k \textit{ is even and } \dfrac{k}{2} \textit{ is odd,} \\ \lceil (\log \lceil \dfrac{n}{\lceil \frac{k}{2} \rceil} \rceil) \rceil, \textit{ for any } k \geq 2. \end{cases}$$

When $cn < k < (1-c)n$ for a constant c, $0 < c < \frac{1}{2}$, we establish an improved lower bound for $\beta_{[\pm 1]}(n, k)$ using a vector space orthogonality argument, enabling us to apply a recent result of Keevash and Long [11].

Theorem 7. *Let c be a constant such that $0 < c < \frac{1}{2}$ and $n \in \mathbb{N}$. If $cn < k < (1-c)n$, then*

$$\max\Big\{\beta_{[\pm 1]}(n, k),\ \beta_{[\pm 1]}(n, k-1),\ \beta_{[\pm 1]}(n, k-2),\ \beta_{[\pm 1]}(n, k-3)\Big\} \geq \delta n,$$

where $\delta = \delta(c)$ is some real positive constant.

Let $\mathcal{F}$ be a family of subsets of $[n]$. The *dependency* of a subset $A \in \mathcal{F}$ denoted by $d(A, \mathcal{F})$ is the number of subsets $\widehat{A} \in \mathcal{F}$, such that (i) $|A \cap \widehat{A}| \geq 1$, and (ii) $A \neq \widehat{A}$. The *dependency of a family* $d(\mathcal{F})$ or simply d, denotes the maximum dependency of any subset A in the family $\mathcal{F}$. We study $\beta_{[\pm 1]}(\mathcal{F})$ for families $\mathcal{F}$ consisting of k-sized sets with bounded dependency and using a corollary of the Lovász local lemma from [15], we prove the following probabilistic upper bound.

Theorem 8. *For a family $\mathcal{F}$ consisting of k-sized subsets of $[n]$ and dependency d, $\beta_{[\pm 1]}(\mathcal{F}) \leq \frac{\sqrt{k}}{c}(\ln(d+1)+1)$, where $c = 0.67$.*

We also study the case when $\mathcal{F}$ consists of all the subsets of $[n]$ of cardinality more than k, $k \in [n]$ and we have the following bounds.

Theorem 9. *Let $\mathcal{F} = \binom{[n]}{k} \cup \binom{[n]}{k+1} \ldots \cup \binom{[n]}{n}$. Then, $\frac{n-k+1}{2} \leq \beta_{[\pm 1]}(\mathcal{F}) \leq \min\{\frac{n}{2}, n-k+1\}$.*

Note that when $n - k$ is a constant, Theorem 9 gives better upper bounds for $\beta_{[\pm 1]}(\mathcal{F})$.

1.3. Some quick observations

In this section, we derive a few basic results on $\beta_D(\mathcal{F})$, $\beta_D(n)$ and $\beta_D(n, k)$. $\mathcal{P}$ is a *property* for a set system if it is invariant under isomorphism.[1] It is not hard to see that for any two isomorphic families $\mathcal{F}_1$ and $\mathcal{F}_2$ on $[n]$, $\beta_D(\mathcal{F}_1) = \beta_D(\mathcal{F}_2)$. So, β_D is a property of the set system. For any two families $\mathcal{F}_1$ and $\mathcal{F}_2$, $\mathcal{F}_1 \subseteq \mathcal{F}_2$, $\beta_D(\mathcal{F}_1) \leq \beta_D(\mathcal{F}_2)$. Therefore, $\beta_D(n)$ and $\beta_D(n, k)$ are monotone with respect to n. However, $\beta_D(n, k)$ is not monotone with respect to k: $\beta_{[\pm 1]}(n, 2) = \lceil \log n \rceil$ (see Corollary 2), $\beta_{[\pm 1]}(n, \frac{n}{2}) = \Omega(\sqrt{n})$ (see Observation 11) whereas $\beta_{[\pm 1]}(n, n-2) = 3$ (see Proposition 21).

We note that for any integer t, "$\beta_D(\mathcal{F}) \leq t$" is not *hereditary*.[2] This can be demonstrated with the following example. Let $\mathcal{F} = \{\{1, 2, 4, 5\}, \{1, 3, 4, 5\}, \{2, 3, 4, 5\}\}$ be a family on $\{1, \ldots, 5\}$ and $S = \{1, 2, 3\}$. $\mathcal{F}_S = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ is the subfamily of $\mathcal{F}$ induced by S. It is easy to see that when $D = [\pm 1]$, $\beta_D(\mathcal{F}) = 1$ whereas $\beta_D(\mathcal{F}_S) = 2$.

[1] Two set systems $H = (X; E_1, E_2, \ldots, E_m)$ and $I = (Y; F_1, F_2, \ldots, F_m)$ are said to be isomorphic if they have the same number m of subsets, and if there exist a bijection $\varphi : X \to Y$ and a permutation π on $M = \{1, 2, \ldots, m\}$ such that

$$\varphi(E_i) = F_{\pi(i)} \ (i = 1, 2, \ldots, m).$$

See page 411 of [3] for related notions.

[2] For a family $\mathcal{F} = \{A_1, \ldots, A_m\}$ on $[n]$, and a set $S \subseteq [n]$, the family $\mathcal{F}_S = \{A_1^s, \ldots, A_m^s\}$ is called a family induced by S on $\mathcal{F}$ if $A_j^s = A_j \cap S$, for $1 \leq j \leq m$. A property $\mathcal{P}$ is *hereditary* if $\mathcal{F} \in \mathcal{P}$ implies $\mathcal{F}_S \in \mathcal{P}$ for every induced family $\mathcal{F}_S$ of $\mathcal{F}$, $S \subseteq [n]$.

Please cite this article in press as: N. Balachandran, et al., Bisecting and D-secting families for set systems, Discrete Applied Mathematics (2017), http://dx.doi.org/10.1016/j.dam.2017.05.005.

Observation 10. *Let $\mathcal{F}$ be a family of subsets of $[n]$ and $\mathcal{F}' = \{S_1, \ldots, S_r\}$ be a D-secting family for $\mathcal{F}$, $r \in \mathbb{N}$ and $D = [\pm i]$. Then, $\mathcal{H} = \{H_1, \ldots, H_r\}$ is also a D-secting family for $\mathcal{F}$, where $H_i \in \{[n] \setminus S_i, S_i\}$, $1 \le i \le r$.*

For the rest of the section, assume that n is even (since it does not effect the asymptotics). Note that when k is even (resp., odd), the maximum number of k-sized sets $A \in \mathcal{F}$ that can be bisected with any set $A' \subseteq [n]$ is $\binom{\frac{n}{2}}{\frac{k}{2}}^2$ (resp., $2\binom{\frac{n}{2}}{\lceil\frac{k}{2}\rceil}\binom{\frac{n}{2}}{\lfloor\frac{k}{2}\rfloor}$), $k \in [n]$. This gives a trivial lower bound for $\beta_{[\pm 1]}(n,k)$ using Stirling's approximation, i.e., $\sqrt{2\pi n}(\frac{n}{e})^n \le n! \le e\sqrt{n}(\frac{n}{e})^n$.

Observation 11.

$$\beta_{[\pm 1]}(n,k) \ge \frac{\binom{n}{k}}{2\binom{\frac{n}{2}}{\lceil\frac{k}{2}\rceil}\binom{\frac{n}{2}}{\lfloor\frac{k}{2}\rfloor}} = \Omega\left(\sqrt{\frac{k(n-k)}{n}}\right). \qquad (1)$$

The constant in the lower bound is $C = \frac{\sqrt{2}\pi^{2.5}}{e^4} \ge .45$. When $k = \frac{n}{2}$, this corresponds to a lower bound of $\Omega(\sqrt{n})$ for $\beta_{[\pm 1]}(n, \frac{n}{2})$. Moreover, using the monotone property, $\beta_{[\pm 1]}(n) \ge \beta_{[\pm 1]}(n, \frac{n}{2}) = \Omega(\sqrt{n})$. In what follows, we derive improved upper bounds and lower bounds for $\beta_D(n)$. We start our discussion with the case $D = [\pm i]$, $i \in [n]$, followed by the case $D = \{i\}$.

2. Bounds for $\beta_{[\pm i]}(n)$

Recall that $\beta_{[\pm i]}(n)$ is the maximum of $\beta_{[\pm i]}(\mathcal{F})$ over all families $\mathcal{F}$ on $[n]$, where $\beta_{[\pm i]}(\mathcal{F})$ denotes the minimum cardinality of a $[\pm i]$-secting family for $\mathcal{F}$.

2.1. Upper bounds

Lemma 12. $\beta_{[\pm i]}(n) \le \lceil\frac{n}{2i}\rceil$.

Proof. Let $\mathcal{F}$ denote the family consisting of all the non-empty subsets of $[n]$. In what follows, we demonstrate a construction that yields a $[\pm i]$-secting family of cardinality $\frac{n}{2i}$ for $\mathcal{F}$, assuming $2i$ divides n. Let $B_1 = \{1, 2, \ldots, \frac{n}{2}\}$. The set B_2 is obtained from B_1 by swapping the largest i elements of B_1 with the smallest i elements in $[n] \setminus B_1$. So, $B_2 = \{1, 2, \ldots, \frac{n}{2} - i, \frac{n}{2} + i, \frac{n}{2} + i - 1, \ldots, \frac{n}{2} + 1\}$ (we write the swapped elements in descending order for convenience). In general, B_{j+1} is obtained from B_j by swapping the largest i elements of $B_1 \cap B_j$ (i.e., $\{\frac{n}{2} - ij + 1, \ldots, \frac{n}{2} - ij + i\}$) with the smallest i elements of $([n] \setminus B_1) \cap ([n] \setminus B_j)$ (i.e., $\{\frac{n}{2} + ij - i + 1, \ldots, \frac{n}{2} + ij\}$). We stop the process at $B_{\frac{n}{2i}} = \{1, \ldots, i, n - i, n - (i-1), \ldots, \frac{n}{2} + 1\}$. Let $\mathcal{F}' = \{B_1, \ldots, B_{\frac{n}{2i}}\}$.

We prove that $\mathcal{F}'$ is indeed a $[\pm i]$-secting family for $\mathcal{F}$. For the sake of contradiction, we assume that there exists some $A \subseteq [n]$ such that $|A \cap B_j| - |A \cap ([n] \setminus B_j)| \notin D$, for all $B_j \in \mathcal{F}'$. Let $c_j := |A \cap B_j| - |A \cap ([n] \setminus B_j)|$, $1 \le j \le \frac{n}{2i}$. From the construction of B_{j+1} from B_j, observe that $|c_j - c_{j+1}| \le |B_j \triangle B_{j+1}| = 2i$, $1 \le j \le \frac{n}{2i} - 1$. Clearly, $c_1 = d$, for some $d \notin \{-i, \ldots, i\}$.

Claim 13. $c_{\frac{n}{2i}} \le -d + 2i$ *for* $d > 0$ *(resp.* $\ge -d - 2i$ *for* $d < 0$*).*

Proof. Let $B_{\frac{n}{2i}+1}$ be the set obtained from $B_{\frac{n}{2i}}$ by swapping the largest i elements $\{1, \ldots, i\}$ of $B_1 \cap B_{\frac{n}{2i}}$ with the smallest i elements $\{n - i + 1, \ldots, n\}$ of $([n] \setminus B_1) \cap ([n] \setminus B_{\frac{n}{2i}})$. Let $c_{\frac{n}{2i}+1} = |A \cap B_{\frac{n}{2i}+1}| - |A \cap ([n] \setminus B_{\frac{n}{2i}+1})|$. Observe that since $c_1 = d$ and $B_{\frac{n}{2i}+1}$ is $[n] \setminus B_1$, $c_{\frac{n}{2i}+1} = -d$. Moreover, $|c_{\frac{n}{2i}} - c_{\frac{n}{2i}+1}| \le 2i$. So, $c_{\frac{n}{2i}}$ is at most $-d + 2i$. The proof for the case of $d < 0$ is similar. □

We now have these exhaustive cases.

1. $d \ge 2i$ (or $d \le -2i$): Note that $D = \{-i, \ldots, +i\}$ and $|c_j - c_{j+1}| \le 2i$, for all $1 \le j \le \frac{n}{2i} - 1$. Using Claim 13, $c_{\frac{n}{2i}} \le 0$ (resp., $c_{\frac{n}{2i}} \ge 0$). Therefore, there exists at least one index l, $1 \le l \le \frac{n}{2i} - 1$, such that $c_l \cdot c_{l+1} \le 0$. Observe that either of c_l or c_{l+1}, or both lie in $\{-i, \ldots, +i\}$. This is a contradiction to our assumption that A is not D-sected by $\mathcal{F}'$.
2. $i < d < 2i$: From Claim 13, it is clear that $c_{\frac{n}{2i}} < i$. So, if there exists an index l, $1 \le l \le \frac{n}{2i} - 1$, such that $c_l \cdot c_{l+1} \le 0$, either c_l or c_{l+1} or both lie in $\{-i, \ldots, +i\}$. Otherwise, $c_{\frac{n}{2i}} \in \{0, \ldots, i-1\} \subset D$ as desired.
3. $-2i < d < -i$: Similar to the previous case.

This establishes that $\beta_{[\pm i]}(n)$ is at most $\frac{n}{2i}$, when $2i$ divides n. Note that when n is not divisible by $2i$, we can construct $\mathcal{F}'$ of cardinality $\lceil\frac{n}{2i}\rceil$ with the same procedure, where $B_{\lceil\frac{n}{2i}\rceil} = \{1, \ldots, p, n - p, n - (p-1), \ldots, \frac{n}{2} + 1\}$, $p = n \bmod 2i$. This completes the proof of Lemma 12. □

2.2. Lower bounds

To obtain a lower bound for $\beta_D(n)$, it is natural to remove one or two points from $[n]$ and to proceed with induction. However, we note that, even when $D = \{-1, 0, 1\}$, such a direct induction only yields a lower bound of $\log n$, which is not useful (since we already have a lower bound of $\Omega(\sqrt{n})$ from Section 1.3). In order to derive a tight lower bound for $\beta_D(n)$, we use the vector representations of sets and a polynomial representation of Boolean functions.

For any subset $A \subseteq [n]$, let (i) $X_A = (x_1, \dots, x_n) \in \{0,1\}^n$ be the incidence vector such that $x_i = 1$ if and only if $i \in A$; and, (ii)$R_A = (r_1, \dots, r_n) \in \{-1,1\}^n$ be the incidence vector such that $r_i = 1$ if and only if $i \in A$. Observe that for any two subsets A and A' of $[n]$, the dot product of $X_A = (x_1, \dots, x_n)$ with $R_{A'} = (r_1, \dots, r_n)$, denoted by $\langle X_A, R_{A'}\rangle$, is equivalent to $|A \cap A'| - |A \cap ([n] \setminus A')|$. For an even (resp., odd) cardinality subset $A \in \mathcal{F}$, note that the corresponding incidence vector $X_A = (x_1, \dots, x_n)$ is even (resp., odd). Let $\mathcal{F}$ be a family of subsets of $[n]$. Observe that for any even subset $A_e \in \mathcal{F}$ and any arbitrary subset $A' \subseteq [n]$, $\langle X_{A_e}, R_{A'}\rangle \equiv 0 \bmod 2$, i.e., $\langle X_{A_e}, R_{A'}\rangle \in \{0, \pm 2, \pm 4, \dots\}$. Moreover, for any odd subset $A_o \in \mathcal{F}$, $\langle X_{A_o}, R_{A'}\rangle \equiv 1 \bmod 2$, i.e., $\langle X_{A_o}, R_{A'}\rangle \in \{\pm 1, \pm 3, \pm 5, \dots\}$.

We demonstrate that the polynomial representation of Boolean functions [16,18] is useful to establish lower bounds for $\beta_D(n)$. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function on n variables,say $y_1, \dots, y_n$. For instance, the *parity* function on n variables is simply equal to the monomial $\prod_{j=1}^{n} y_j$. Let *sign* : $\mathbb{R} \setminus \{0\} \to \{0,1\}$ be a function defined as (i) $sign(\alpha) = 1$ if $\alpha > 0$, and (ii) $sign(\alpha) = 0$, otherwise, for $\alpha \in \mathbb{R} \setminus \{0\}$. A multilinear polynomial $P(y_1, \dots, y_n)$ *weakly represents* f if P is nonzero and for every $Y = (y_1, \dots, y_n)$ where $P(Y)$ is nonzero, $sign(f(Y)) = sign(P(Y))$. The *weak degree* of a function f is the degree of the lowest degree polynomial which weakly represents f. We have the following result that follows from Lemma 2.29 of [18] originally proved by Minsky and Papert in [14].

Lemma 14. *The weak degree of the parity function on n variables is n.*

In what follows, we use the notion of weak degree of the parity function to establish Theorem 3.

Lemma 15. $\beta_{[\pm i]}(n) \geq \lceil \frac{n}{2i} \rceil$.

Proof. Let $\mathcal{F}$ denote the $2^n - 1$ non-empty subsets of $[n]$. Let $\mathcal{F}'$ be a minimum cardinality $[\pm i]$-secting family for $\mathcal{F}$. Let $\mathcal{R}$ be set of incidence vectors of sets in $\mathcal{F}'$, where each vector R in $\mathcal{R}$ is an element of $\{-1,+1\}^n$. We start the analysis assuming i is even and $i > 0$, and then extend to odd i. For every odd set $A_o \in \mathcal{F}$, there exists a vector $R \in \mathcal{R}$ such that $\langle X_{A_o}, R\rangle - d = 0$, for some $d \in \{-i+1, -i+3, \dots, i-1\}$. Let $X = (x_1, \dots, x_n) \in \{0,1\}^n$. We use X to denote the incidence vector of any arbitrary set in $\mathcal{F}$. Consider the polynomial M on $X = (x_1, \dots, x_n)$ as

$$M(X) = \left(\prod_{R \in \mathcal{R}} \left((\langle X, R\rangle)^2 - 1^2\right) \prod_{R \in \mathcal{R}} \left((\langle X, R\rangle)^2 - 3^2\right) \dots \prod_{R \in \mathcal{R}} \left((\langle X, R\rangle)^2 - (i-1)^2\right) \right)^2. \quad (2)$$

From the definitions of $\mathcal{R}$ and M, it is clear that $M(X)$ is (i) zero when $X = X_{A_o}$ for all odd subsets $A_o \in \mathcal{F}$; and (ii) positive when $X = X_{A_e}$ for all even subsets $A_e \in \mathcal{F}$.

Domain conversion and multilinearization

Recall that a vector $T \in \{0,1\}^n$ is even if the number of 1's in T is even and a vector $T \in \{-1,1\}^n$ is even if the number of -1's in T is even. Consider the polynomial N on $Y = (y_1, \dots, y_n)$, where each $y_i = \pm 1$.

$$N(y_1, \dots, y_n) = M(x_1, \dots, x_n), \quad (3)$$

where $x_j = \frac{1-y_j}{2}$, $1 \leq j \leq n$. Note that if $y_i = -1$ (resp. 1), then $\frac{1-y_i}{2}$ becomes 1 (resp. 0). So, if some vector $Y = (y_1, \dots, y_n)$ includes an even number of -1's, then the vector $(\frac{1-y_1}{2}, \dots, \frac{1-y_n}{2})$ has an even number of 1's, i.e., the reduction of the vector $(y_1, \dots, y_n)$ from the $\{-1,1\}^n$ domain to $(\frac{1-y_1}{2}, \dots, \frac{1-y_n}{2})$ in the $\{0,1\}^n$ domain preserves the definition of *evenness*. Note that (i) $N(Y)$ evaluates to zero, when $Y = Y_{A_o} \in \{-1,1\}^n$ for all odd subsets $A_o \in \mathcal{F}$; (ii) $sign(N(Y)) = sign(parity(Y))$, when $Y = Y_{A_e} \in \{-1,1\}^n$ for all even subsets $A_e \in \mathcal{F}$. Let $N'(Y = (y_1, \dots, y_n))$ be the multilinear polynomial obtained from $N(Y = (y_1, \dots, y_n))$ by repeatedly replacing each y_i^2 in the monomials by 1. $deg(N'(Y)) \leq deg(N(Y))$ and $N'(Y) = N(Y)$, for vectors $Y \in \{-1,1\}^n$.

Clearly, $N'(Y)$ weakly represents the parity function. Each term $(\prod_{R \in \mathcal{R}}((\langle X, R\rangle)^2 - j^2))^2, j \in \{1, \dots, (i-1)\}$, contributes a degree of $4|\mathcal{R}|$ to the degree of $M(X)$, and, there are $\frac{i}{2}$ such terms. Therefore, the degree of $M(X)$ is $2|\mathcal{R}|i$. Moreover, from Eq. (3), $deg(N'(Y)) \leq deg(N(Y)) = deg(M(X))$. However, from Lemma 14, $deg(N'(Y)) \geq n$, which implies $\beta_{[\pm i]}(n) = |\mathcal{R}| \geq \frac{n}{2i}$.

If $i > 1$ is odd, $M(X)$ is defined as

$$\prod_{R \in \mathcal{R}} \left((\langle X, R\rangle)^2\right) \left(\prod_{R \in \mathcal{R}} \left((\langle X, R\rangle)^2 - 2^2\right) \prod_{R \in \mathcal{R}} \left((\langle X, R\rangle)^2 - 4^2\right) \dots \prod_{R \in \mathcal{R}} \left((\langle X, R\rangle)^2 - (i-1)^2\right) \right)^2.$$

Observe that $M(X)$ vanishes for all even vectors and is positive for all odd vectors. The polynomial N on $Y = (y_1, \dots, y_n)$, where each $y_i = \pm 1$, is now defined as

$$N(y_1, \dots, y_n) = -M(x_1, \dots, x_n). \quad (4)$$

Note that degree of $M(X)$ is $2|\mathcal{R}| + 4|\mathcal{R}|\frac{i-1}{2} = 2|\mathcal{R}|i$ and the rest of the arguments are same as the previous case.

We are only left with the cases when $i = 0$ and $i = 1$. Observe that $\beta_D(n)$ for the case of $D = \{0\}$ and $D = \{-1, 0, 1\}$ is same: any bisecting family for a family $\mathcal{F}_1$ consisting of only the $2^{n-1} - 1$ non-empty even subsets of $[n]$ must bisect all the $2^n - 1$ subsets of $[n]$. In this case, take $M(X) = \prod_{R\in\mathcal{R}} \left((\langle X, R\rangle)^2\right)$ and proceed as before to get $\beta_{[\pm 1]}(n) \geq \frac{n}{2}$. □

From Lemmas 12 and 15, Theorem 3 follows, which is restated below.

Statement. $\beta_{[\pm i]}(n) = \lceil \frac{n}{2i} \rceil$, $n \in \mathbb{N}$, $i \in [n]$.

Let $\mathcal{F}$ consist of $2^n - 1$ non-empty subsets of $[n]$. Then, Theorem 3 asserts that the construction of $[\pm i]$-secting family of cardinality $\lceil \frac{n}{2i} \rceil$ in Section 2.1 is indeed optimal. Moreover, Theorem 3 implies that if we allow the imbalances of intersections up to $\sqrt{n}$, i.e., $D = [\pm\sqrt{n}]$, then a family $\mathcal{F}'$ of cardinality $\frac{\sqrt{n}}{2}$ is necessary and sufficient for $\mathcal{F}$.

Corollary 16. *For $D = [\pm\sqrt{n}]$, $n \in \mathbb{N}$, $\beta_D(n) = \lceil \frac{\sqrt{n}}{2} \rceil$.*

In what follows, we demonstrate that D-secting families of cardinality much smaller than $\frac{\sqrt{n}}{2}$ can be computed when $|\mathcal{F}|$ is small.

2.3. Computing $\beta_{[\pm i]}(\mathcal{F})$ for arbitrary families

In Section 1, we discussed about the discrepancy interpretation of the bisection problems. Probabilistic method is a useful tool in computing low discrepancy colorings. The following Chernoff's bound is used extensively to establish upper bounds on the discrepancy of hypergraphs.

Lemma 17 ([5]). *If $X = \sum_{i=1}^{n} X_i$ is the sum of n independent random variables distributed uniformly over $\{-1, 1\}$, then for any $\Delta > 0$,*

$$P[|X| > \Delta] \leq 2e^{-\frac{\Delta^2}{2n}}.$$

In what follows, we obtain an upper bound on $\beta_{[\pm i]}(\mathcal{F})$, when $\mathcal{F}$ is a family of arbitrary sized subsets, with a simple application of Lemma 17.

Proof of Theorem 4

Statement. *Let $\mathcal{F}$ be a family of subsets of $[n]$ and let $|\mathcal{F}| = m$. Let $D = [\pm i]$, where $i = \sqrt{\frac{3n\ln(2m)}{t}}$ and $t \leq \frac{1}{2}\log m$. Then, $\beta_D(\mathcal{F}) \leq t$.*

Proof. We pick a set $\mathcal{F}'$ of t random subsets $\{A'_1, \ldots, A'_t\}$ of $[n]$, where for each j, $1 \leq j \leq t$, a point $a \in [n]$ is chosen independently and uniformly at random into A'_j. Let $R_{A'_j} = (r_1, \ldots, r_n) \in \{-1, 1\}^n$ be the incidence vector corresponding to A'_j: r_i is 1 if and only if $i \in A'_j$. For any subset $A \in \mathcal{F}$, $|A\cap A'_j| - |A\cap([n]\setminus A'_j)|$ can be viewed as sum of $|A|$ random variables distributed uniformly over $\{-1, 1\}$. We say a subset $A \in \mathcal{F}$ is bad with respect to subset $A'_j \in \mathcal{F}'$ if $\||A\cap A'_j| - |A\cap([n]\setminus A'_j)\| > \sqrt{\frac{3|A|\ln(2m)}{t}}$. Using Chernoff's bound, the probability that a subset $A \in \mathcal{F}$ is bad with respect to a random subset $A'_j \in \mathcal{F}'$ is

$$P\left[\||A\cap A'_j| - |A\cap([n]\setminus A'_j)\| > \sqrt{\frac{3|A|\ln(2m)}{t}}\right] \leq 2e^{-\frac{3|A|\ln(2m)}{2t|A|}} = 2\left(\frac{1}{2m}\right)^{\frac{3}{2t}}.$$

Any subset A is bad with respect to $\mathcal{F}'$ if $\||A\cap A'_j| - |A\cap([n]\setminus A'_j)\| > \sqrt{\frac{3|A|\ln(2m)}{t}}$, for all $A'_j \in \mathcal{F}'$. So, A is bad with respect to $\mathcal{F}'$ with probability at most $2^t(\frac{1}{2m})^{\frac{3t}{2t}} = \frac{2^{t-1.5}}{m^{1.5}}$. Using union bound, the probability that some subset in $\mathcal{F}$ is bad with respect to $\mathcal{F}'$ is at most $m\frac{2^{t-1.5}}{m^{1.5}}$. So, if $2^t \leq \sqrt{m}$ (i.e., $t \leq \frac{1}{2}\log m$), the probability that any subset in $\mathcal{F}$ is bad with respect to $\mathcal{F}'$ is at most $\frac{1}{2\sqrt{2}}$. Since the failure probability is less than $\frac{1}{2}$, in expected two iterations, we can obtain a family $\mathcal{F}'$ of t subsets such that for every $A \in \mathcal{F}$, there is an $A'_j \in \mathcal{F}'$ with $\||A\cap A'_j| - |A\cap([n]\setminus A'_j)\| \leq \sqrt{\frac{3n\ln(2m)}{t}}$. □

Note that if $i \geq \sqrt{4.2n+1}$ and $|\mathcal{F}| = O(n^c)$, $c \in \mathbb{N}$, a D-secting family for $\mathcal{F}$ of cardinality $O(\log n)$ can be computed as discussed above. Note that this yields D-secting families of size much smaller than that guaranteed by Corollary 16 for $\mathcal{F}$ provided $|\mathcal{F}|$ is polynomial in n.

3. Bounds for $\beta_i(n)$

In Section 2, we established tight bounds for $\beta_D(n)$ when $D = [\pm i]$. In this section, we study $\beta_D(n)$, when D is a singleton set, i.e., $D = \{i\}$.

3.1. Tight bounds for $\beta_1(n)$

Theorem 18. $\beta_1(n) = \lceil \frac{n}{2} \rceil$, $n \in \mathbb{N}$.

Proof. As mentioned in Section 1, when $D = \{1\}$, the family $\mathcal{F}$ should consist of all the odd subsets of $[n]$. Let $\mathcal{R}$ be a minimum sized set of $\{-1, +1\}^n$ vectors such that for every odd set $A_o \in \mathcal{F}$, there exists a vector $R \in \mathcal{R}$ such that $\langle A_o, R \rangle - 1 = 0$. Consider the polynomial M on $X = (x_1, \ldots, x_n)$.

$$M(X) = \prod_{R \in \mathcal{R}} (\langle X, R \rangle - 1)^2. \quad (5)$$

Note that if $N'(Y)$ is obtained from $M(X)$ after domain conversion and multilinearization, N' weakly represents the parity function. Using Lemma 14, $deg(M(X)) = 2|\mathcal{R}| \geq deg(N'(Y)) \geq n$ and therefore $|\mathcal{R}| \geq \lceil \frac{n}{2} \rceil$. In what follows, we demonstrate a construction of a family $\mathcal{F}'$ of cardinality $\lceil \frac{n}{2} \rceil$ such that for every odd subset $A \in \mathcal{F}$, there exists some $A' \in \mathcal{F}'$ with $|A \cap A'| - |A \cap ([n] \setminus A')| = 1$.

Consider the family $\mathcal{F}$ consisting of all the odd subsets of $[n]$. Consider the case when n is even; the odd case is similar except the ceilings in the final expression. Note that if $n \leq 2$, we can choose $\mathcal{F}' = \{\{1, 2\}\}$ to get the desired intersection property. So, we consider the case when $n \geq 4$. Let $B_1 = \{1, 2, \ldots, \frac{n}{2} + 1\}$. B_2 is obtained from B_1 by swapping $\{\frac{n}{2} + 1\}$ with $\{\frac{n}{2} + 2\}$, i.e., $B_2 = \{1, 2, \ldots, \frac{n}{2}, \frac{n}{2} + 2\}$. In general, B_{j+1} is obtained from B_j by replacing the point $\frac{n}{2} - j + 2$ with $\frac{n}{2} + j + 1$. We stop the process at $B_{\frac{n}{2}} = \{1, 2, n, n-1, \ldots, \frac{n}{2} + 2\}$. Let $\mathcal{F}' = \{B_1, \ldots, B_{\frac{n}{2}}\}$.

Claim 19. (i) *For any odd subset $A_o \subseteq \{3, \ldots, n\}$, there exist some B_j and B_l in $\mathcal{F}'$ such that $|A \cap B_j| = \lceil \frac{|A|}{2} \rceil$, and $|A \cap B_l| = \lfloor \frac{|A|}{2} \rfloor$, and* (ii) *For any even subset $A_e \subseteq \{3, \ldots, n\}$, there exists some B_j in $\mathcal{F}'$ such that $|A \cap B_j| = \frac{|A|}{2}$.*

To see the correctness of the claim, consider an arbitrary set A, $A \subseteq \{3, \ldots, n\}$, such that $|A \cap B_1| - |A \cap ([n] \setminus B_1)| = d$, for some $d \in \mathbb{N} \setminus 0$. Then, it follows from the construction that $|A \cap B_{\frac{n}{2}}| - |A \cap ([n] \setminus B_{\frac{n}{2}})| = -d$. Observe that for any j, $1 \leq j \leq \frac{n}{2} - 1$, the difference between $|A \cap B_{j+1}| - |A \cap ([n] \setminus B_{j+1})|$ and $|A \cap B_j| - |A \cap ([n] \setminus B_j)|$ is either -2, 0 or 2. So, the claim follows.

Now, to complete the proof, we need to consider the following exhaustive case for an odd subset A_o.

1. $A_o \subseteq \{3, \ldots, n\}$: A_o has the desired intersection property using Claim 19.
2. $|A_o \cap \{3, \ldots, n\}| = |A_o| - 1$: Using Claim 19, there exists some B_j in $\mathcal{F}'$ such that the even subset $A_o \cap \{3, \ldots, n\}$ is bisected by B_j. Clearly, $|A_o \cap B_j| = \lceil \frac{|A_o|}{2} \rceil$.
3. $|A_o \cap \{3, \ldots, n\}| = |A_o| - 2$: In this case, $\{1, 2\} \subset A_o$. From Claim 19, there exists some B_j in $\mathcal{F}'$ such that $|A'_o \cap B_j| = \lfloor \frac{|A'_o|}{2} \rfloor$, where $A'_o = A_o \cap \{3, \ldots, n\}$. Then, $|A_o \cap B_j| = \lceil \frac{|A_o|}{2} \rceil$.

This establishes that $\beta_1(n)$ is at most $\lceil \frac{n}{2} \rceil$ and completes the proof of Theorem 18. □

3.2. Bounds for $\beta_i(n)$, $i \geq 2$

In the following section, we extend the notion of $\beta_1(n)$ to arbitrary values of i. Note that when $i = 0$, $\beta_0(n) = \beta_{[\pm 1]}(n) = \lceil \frac{n}{2} \rceil$ (see Theorem 3). The case when $i = 1$ is resolved by Theorem 18. We assume that $i \geq 2$ in the remainder of the section.

3.2.1. Proof of Theorem 5

Statement. $\frac{n-i+1}{2} \leq \beta_i(n) \leq n - i + 1$, $n \in \mathbb{N}$, $i \in [n]$.

Proof. Let $\mathcal{F}$ consist of all subsets of $[n]$ such that $A \in \mathcal{F}$ if and only if $|A| \cong i \bmod 2$ and $|A| \geq i$. Let $\mathcal{F}' = \{B_1 = [i], B_2 = B_1 \cup \{i+1\}, \ldots, B_{n-i+1} = B_{n-i} \cup \{n\}\}$. Observe that $\mathcal{F}'$ is indeed an i-secting family for $\mathcal{F}$. Therefore, $\beta_i(n) \leq n - i + 1$. In what follows, we prove the lower bound for $\beta_i(n)$ assuming i to be an even integer greater than 1. The case for odd i can be treated analogously.

We invoke the notion of weak representation of the parity function to establish a lower bound. Let $\mathcal{F}$ denote the $2^n - 1$ non-empty subsets of $[n]$. Let $\mathcal{F}'$ be a minimum cardinality $[\pm i]$-secting family for $\mathcal{F}$. Let $\mathcal{R}$ be the set of incidence vectors of sets in $\mathcal{F}'$, where each vector R in $\mathcal{R}$ is an element of $\{-1, +1\}^n$. So, for any even subset $A_e \subseteq [n]$ with $|A_e| \geq i$, there exists a vector $R \in \mathcal{R}$ such that $\langle X_{A_e}, R \rangle - i = 0$, where X_{A_e} is the 0–1 incidence vector of A_e. We define the polynomials P, M and F on $X = (x_1, \ldots, x_n)$ as follows.

$$M(X) = \prod_{R \in \mathcal{R}} (\langle X, R \rangle - i)^2. \quad (6)$$

$$F(X) = \sum_{S \in \binom{[n]}{i-1}} \prod_{j \in S} x_j.$$

$$P(X) = M(X)F(X). \quad (7)$$

Observe that (i) $P(X)$ evaluates to zero when $X = X_A$, for all subsets A of size at most $i-2$ (since $F(X)$ vanishes for these subsets), (ii) $P(X)$ evaluates to zero when $X = X_{A_e}$, for all even subsets A_e of size at least i (since $M(X)$ vanishes for these subsets), and, (iii) $P(X)$ is strictly positive when $X = X_{A_o}$, for all odd subsets A_o of size at least $i-1$. Consider the polynomial Q on $Y = (y_1, \ldots, y_n)$, where each $y_j \in [\pm 1]$.

$$Q\,(y_1, \ldots, y_n) = -P\,(x_1, \ldots, x_n) \tag{8}$$

where $x_j = \frac{1-y_j}{2}$, $1 \le j \le n$. Let $Q'(Y)$ be the multilinear polynomial obtained from $Q(Y)$ by replacing each occurrence of a y_j^2 by 1, repeatedly. Note that (i) $Q'(Y)$ evaluates to zero for even subsets of $[n]$, and (ii) if $Q'(Y)$ is non-zero on some odd subset Y, then $sign(Q'(Y)) = sign(parity(Y))$. Therefore, $Q'(Y)$ weakly represents the parity function. From Lemma 14, $Q'(Y)$ has degree at least n, and $deg(P(X)) = (i-1) + 2|\mathcal{R}| \ge deg(Q'(Y)) \ge n$. So, $|\mathcal{R}| \ge \frac{n-i+1}{2}$. □

4. Bisecting k-uniform families

In this section, we discuss the problem of bisection for k-uniform families. We focus on establishing bounds for $\beta_D(n, k)$ when $D = [\pm 1]$.

4.1. Some observations for $\beta_{[\pm 1]}(n, k)$

Observation 20. *Let n be an even integer and $\mathcal{F}'$ be an optimal bisecting family for a family $\mathcal{F} = \binom{[n]}{k}$ such that each subset $A' \in \mathcal{F}'$ has cardinality $\frac{n}{2}$. Then, $\beta_{[\pm 1]}(n, n-k) \le \beta_{[\pm 1]}(n, k)$.*

Proof. It is not hard to see that the bisecting family $\mathcal{F}'$ for $\mathcal{F}$ is also a bisecting family for $\overline{\mathcal{F}} = \binom{[n]}{n-k}$ when n is even and each subset in $\mathcal{F}'$ is a part of an equal-sized bipartition of n. □

From Corollary 2, we know that $\beta_{[\pm 1]}(n, 2) = \lceil \log n \rceil$. Moreover, when n is of the form 2^t, for some $t \in \mathbb{N}$, we can obtain a bisecting family $\mathcal{F}' = \{A_1, \ldots, A_{\log n}\}$ for the family $\mathcal{F} = \binom{[n]}{2}$ in the following way. (i) For $j \in [n]$, obtain the $\log n$ bit binary code equivalent to $j-1$ and assign it to j. (ii) Elements with lth bit as 1 form the set A_l. Using Corollary 2, $\mathcal{F}'$ is an optimal bisecting family for $\mathcal{F}$, and $|A_l| = \frac{n}{2}$, for all $A_l \in \mathcal{F}'$. Using Observation 20, it follows that $\beta_{[\pm 1]}(n, n-2) \le \log n$, when n is a power of 2. However, when the difference between n and k is a small constant, we can achieve much better bounds for $\beta_{[\pm 1]}(n, k)$ as follows.

Proof of Theorem 9

Statement. *Let $\mathcal{F} = \binom{[n]}{k} \cup \binom{[n]}{k+1} \ldots \cup \binom{[n]}{n}$. Then, $\frac{n-k+1}{2} \le \beta_{[\pm 1]}(\mathcal{F}) \le \min\{\frac{n}{2}, n-k+1\}$.*

Proof. The upper bound of $\frac{n}{2}$ follows from Lemma 12. Let $x = n-k$. We obtain a bisecting family for $\mathcal{F}$ of cardinality $x+1$ in the following way. Let S and T denote two disjoint $\lceil \frac{k}{2} \rceil$ and $\lfloor \frac{k}{2} \rfloor$ elements subset of $[n]$, respectively. Let $c_1, \ldots, c_x$ denote the remaining elements of $[n]$. Let $S_0 = S$, and for any $j \in [x]$, $S_j = S_{j-1} \cup \{c_j\}$. Let $\mathcal{F}' = \{S_0, \ldots, S_x\}$. We claim that $\mathcal{F}'$ is a bisecting family for a $\mathcal{F}$. For any set A of cardinality k', $k \le k' \le n$, that is not bisected by S_0, $|A \cap S_0| < \frac{k'}{2}$ and $|A \cap S_x| \ge \frac{k'}{2}$. The upper bound follows from the observation that $|A \cap S_{j+1}|$ differs from $|A \cap S_j|$ by at most 1.

The proof of the lower bound $\frac{n-k+1}{2}$ for $\beta_{[\pm 1]}(\mathcal{F})$ is in the same spirit as the proof of the lower bound of Theorem 5; we give the proof for completeness. We assume that $k \ge 2$ and is even; the case when k is odd is analogous. Let $\mathcal{F}'$ be a minimum cardinality $[\pm 1]$-secting family for $\mathcal{F}$. Let $\mathcal{R}$ be the set of incidence vectors of sets in $\mathcal{F}'$, where each vector R in $\mathcal{R}$ is an element of $\{-1, +1\}^n$. We define the polynomials P, M and F on $X = (x_1, \ldots, x_n)$ as follows.

$$M(X) = \prod_{R \in \mathcal{R}} (\langle X, R \rangle)^2 \text{ (note the difference from Eq. (6)).} \tag{9}$$

$$F(X) = \sum_{S \in \binom{[n]}{k-1}} \prod_{j \in S} x_j. \tag{10}$$

$$P(X) = M(X)F(X). \tag{11}$$

Observe that (i) $P(X)$ evaluates to zero when $X = X_A$, for all subsets A of size at most $k-2$ (since $F(X)$ vanishes for these subsets), (ii) $P(X)$ evaluates to zero when $X = X_{A_e}$, for all even subsets A_e of size at least k (since $M(X)$ vanishes for these subsets), and, (iii) $P(X)$ is strictly positive when $X = X_{A_o}$, for all odd subsets A_o of size at least $k-1$. Note that if $Q'(Y)$ is obtained from $P(X)$ after domain conversion and multilinearization, $Q'(Y)$ weakly represents the parity function. From Lemma 14, $Q'(Y)$ has degree at least n, and $deg(P(X)) = (k-1) + 2|\mathcal{R}| \ge deg(Q'(Y)) \ge n$. So, $|\mathcal{R}| \ge \frac{n-k+1}{2}$. □

Note that using Theorem 9 for $k = n-2$, we get, $\beta_{[\pm 1]}(n, n-2) \le 3$. This is surprising since (i) $\mathcal{F} = \binom{[n]}{n-2}$ has the same number of subsets as $\overline{\mathcal{F}} = \binom{[n]}{2}$, (ii) the maximum number of sets of $\mathcal{F}$ and $\overline{\mathcal{F}}$ that can be bisected by a single set $A' \in \mathcal{F}'$ is $(\frac{n}{2})^2$, and (iii) $\beta_0(n, 2) = \lceil \log n \rceil$.

Proposition 21. *$\beta_{[\pm 1]}(n, n-2) = 3$, for every even integer n greater than 4.*

Proof. We only need to show that $\beta_{[\pm 1]}(n, n-2) > 2$. Note that since the hyperedges are of cardinality $n-2$, every set in an optimal bisecting family $\mathcal{F}'$ is of cardinality $\frac{n}{2}-1$, $\frac{n}{2}$, or $\frac{n}{2}+1$. Consider an optimal bisecting family $\mathcal{F}' = \{A_1, A_2\}$ of cardinality 2 for $\mathcal{F} = \binom{[n]}{n-2}$. We know that $\beta_{[\pm 1]}(n, n-2) \le 3$. For the sake of contradiction, assume that there exists an optimal bisecting family $\mathcal{F}'$ for $\mathcal{F}$ consisting of sets of size only $\frac{n}{2}$. Using Observation 20, $\mathcal{F}'$ is a bisecting family of cardinality less than $\log n$ for $\binom{[n]}{2}$, a contradiction to Corollary 2. Without loss of generality, assume that $|A_1| \ne \frac{n}{2}$. Using Observation 10, we can also assume that $|A_1| = \frac{n}{2} - 1$. The rest of the proof is an exhaustive case analysis based on the cardinality of A_2. Let $A^1 = A_1 \cap A_2$ and $A^2 = A_1 \setminus A_2$.

1. $|A_2| = \frac{n}{2}$. At least one of A^1 or A^2 is of size at least two. The $(n-2)$-sized subset missing two elements of $[n]$ both from either A^1 or A^2 is not bisected by $\mathcal{F}'$.
2. $|A_2| = \frac{n}{2}+1$. If $|A^2| \ge 2$, the $(n-2)$-sized subset missing two elements both from A^2 is not bisected by $\mathcal{F}'$. So, $|A^2| \le 1$. If $A^2 = \{y\}$, then an $(n-2)$-sized subset missing y and one element from A^1 is not bisected by $\mathcal{F}'$. If $A^2 = \emptyset$, then any $(n-2)$-sized subset missing one element each from A_1 and $[n] \setminus A_2$ is not bisected by $\mathcal{F}'$.
3. $|A_2| = \frac{n}{2} - 1$. Using Observation 10, this case is identical to Case 2. □

4.2. Proof of Theorem 6

Note that the lower bound of $\Omega(\sqrt{\frac{k(n-k)}{n}})$ for $\beta_{[\pm 1]}(n, k)$ is given by Observation 11. However, when k is a constant, Observation 11 asserts only a $\Omega(\sqrt{k})$ lower bound on $\beta_{[\pm 1]}(n, k)$. An improved lower bound on $\beta_{[\pm 1]}(n, k)$ for constant k given by Theorem 6 is proven below.

Statement.

$$\beta_{[\pm 1]}(n, k) \ge \begin{cases} \log(n-k+2), \text{ when } k \text{ is even and } \dfrac{k}{2} \text{ is odd,} \\ \lceil(\log\lceil \dfrac{n}{\lceil \frac{k}{2} \rceil}\rceil)\rceil, \text{ for any } k \ge 2. \end{cases}$$

Proof. We prove the first lower bound given in Theorem 6 under the assumption that k is even and $\frac{k}{2}$ is odd. Let $\mathcal{F}' = \{A'_1, \ldots, A'_t\}$ be a bisecting family for the family $\mathcal{F} = \binom{[n]}{k}$. For every $A'_j \in \mathcal{F}'$, let $\mathcal{F}_j$ be the collection of k-sized sets that are bisected by A'_j. We estimate a lower bound for t. We associate a graph $G(\mathcal{F})$ with the collection $\mathcal{F}$ of k-sized sets in the following way:

$$V(G(\mathcal{F})) = \{S \in \binom{[n]}{\frac{k}{2}} : S \subseteq A, A \in \mathcal{F}\}$$
$$E(G(\mathcal{F})) = \{\{S_1, S_2\} : S_1 \cap S_2 = \emptyset, S_1, S_2 \in V(G(\mathcal{F}))\}.$$

Observe that $G(\mathcal{F})$ is the Kneser graph $KG(n, \frac{k}{2})$ (for definitions and results related to Kneser graphs, see [1,4]). For every k-sized subset $A \in \mathcal{F}$, there are $\binom{k}{\frac{k}{2}}$ edges in $E(G(\mathcal{F}))$: an edge between any two disjoint $\frac{k}{2}$ sets. From the definition of $\mathcal{F}_1, \ldots, \mathcal{F}_t$, $\cup_{j=1}^{t} G(\mathcal{F}_j) = G(\mathcal{F})$.

Claim 22. *Each $G(\mathcal{F}_j)$ is a bipartite graph.*

Let $A \in \mathcal{F}_j$. Consider a fixed $\frac{k}{2}$ sized subset S of A. If $|S \cap A'_j| > \lfloor \frac{k}{4} \rfloor$, S is placed in the first partite set of $G(\mathcal{F}_j)$; otherwise S is placed in the second partite set of $G(\mathcal{F}_j)$. Note that since $\frac{k}{2}$ is odd, $|S \cap A'_j|$ can never be equal to $|S \cap ([n] \setminus A'_j)|$. It is now easy to see that there is no edge inside the first or second partite set of $G(\mathcal{F}_j)$.

$G(\mathcal{F}_1), \ldots, G(\mathcal{F}_t)$ are bipartite graphs whose union covers $G(\mathcal{F})$. Since $G(\mathcal{F})$ is the Kneser graph $KG(n, \frac{k}{2})$, its chromatic number is $n-k+2$ (see [1,12]). So, using Proposition 1, we get, $t \ge \lceil \log(n-k+2) \rceil$.[3] That is, $\beta_{[\pm 1]}(n, k) \ge \lceil \log(n-k+2) \rceil$, when k is even and $\frac{k}{2}$ is odd. This concludes the proof of the first lower bound given by Theorem 6.

To prove the second lower bound of Theorem 6, consider a bisecting family $\mathcal{F}' = \{A'_1, \ldots, A'_t\}$ of $\mathcal{F} = \binom{[n]}{k}$. Observe that for every $\lceil \frac{k}{2} \rceil + 1$-sized set $S \subseteq [n]$, there exists an $A'_j \in \mathcal{F}'$ such that $S \cap A'_j \ne \emptyset$ and $S \cap ([n] \setminus A'_j) \ne \emptyset$. For every $A'_j \in \mathcal{F}'$, let $\mathcal{F}_j$ be the collection of $\lceil \frac{k}{2} \rceil + 1$-sized sets that has a non-empty intersection with both A'_j and $[n] \setminus A'_j$. Observe that

$$\bigcup_{j=1}^{t} \mathcal{F}_j = \binom{[n]}{\lceil \frac{k}{2} \rceil + 1}. \tag{12}$$

[3] Note that Proposition 1 does not guarantee equality since the $\lceil \log(n-k+2) \rceil$ bipartite graphs that cover $G(\mathcal{F})$ as per Proposition 1 may not correspond to valid $\mathcal{F}_j$'s.

Construct hypergraphs $G_1, \ldots, G_t$, where $V(G_j) = [n]$ and $E(G_j) = \mathcal{F}_j$. To each point $v \in [n]$, assign an t length 0–1 bit vector: jth bit is 1 if and only if $v \in A_j$. Color the points in $[n]$ with the decimal equivalent of its bit vector. Let $f : [n] \to \{0, 1, \ldots, 2^t - 1\}$ denote this coloring. We show that none of the $\binom{[n]}{\lceil \frac{k}{2} \rceil + 1}$ sets remain monochromatic under f. Assume for the sake of contradiction that $S \in \binom{[n]}{\lceil \frac{k}{2} \rceil + 1}$ is monochromatic under f. From Eq. (12), there exists an $\mathcal{F}_j$ such that $S \in \mathcal{F}_j$. From the definition of $\mathcal{F}_j$, S has non-empty intersection with both A'_j and $[n] \setminus A'_j$. Therefore, the jth bits of the t length 0–1 bit vectors of all the points in S cannot be the same. Therefore, S contains at least two points of different color under f, i.e., S is not monochromatic. It is well known that the chromatic number of $\binom{[n]}{\lceil \frac{k}{2} \rceil + 1}$, $\chi(\binom{[n]}{\lceil \frac{k}{2} \rceil + 1})$, is $\lceil \frac{n}{\lceil \frac{k}{2} \rceil} \rceil$. Since f uses 2^t colors, we have, $2^t \geq \lceil \frac{n}{\lceil \frac{k}{2} \rceil} \rceil$. Therefore, $\beta_{[\pm 1]}(n, k) = |\mathcal{F}'| = t \geq \lceil (\log \lceil \frac{n}{\lceil \frac{k}{2} \rceil} \rceil) \rceil$.

This completes the proof of Theorem 6. □

4.3. Proof of Theorem 7

We know that $\beta_{[\pm 1]}(n) = \lceil \frac{n}{2} \rceil$ (see Theorem 3). The number of $\frac{n}{2}$-sized subsets of $[n]$ that can be bisected by a single subset $A' \subseteq [n]$ is at most $2(\binom{\frac{n}{2}}{\frac{n}{4}})^2$. This gives a trivial lower bound of $\Omega(\sqrt{n})$ for $\beta_{[\pm 1]}(n, \frac{n}{2})$. In this section, we prove a stronger result using a theorem of Keevash and Long [11] which is an improvement over a theorem of Frankl and Rödl [8]. Given $q \in \mathbb{N}$, a set $\mathcal{C}$ is called a q-ary code if $\mathcal{C} \subseteq [q]^n$, for $q \geq 2$. For any $x, y \in [q]^n$, the *Hamming distance* between x and y, where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, denoted by $d_H(x, y)$, is $|\{i \in [n] : x_i \neq y_i\}|$. For any code $\mathcal{C}$, let $d(\mathcal{C})$ be the set of all the Hamming distances allowed for any $x, y \in \mathcal{C}$. A code is called *d-avoiding* if $d \notin d(\mathcal{C})$. We have the following upper bound on the cardinality of a *d-avoiding* code $\mathcal{C}$ as given in [11].

Theorem 23 ([11]). *Let $\mathcal{C} \subseteq [q]^n$ and let ϵ satisfy $0 < \epsilon < \frac{1}{2}$. Suppose that $\epsilon n < d < (1 - \epsilon)n$ and d is even if $q = 2$. If $d \notin d(\mathcal{C})$, then $|\mathcal{C}| \leq q^{(1-\delta)n}$, for some positive constant $\delta = \delta(\epsilon)$.*

In what follows, we prove Theorem 7.

Statement. *Let c be a constant such that $0 < c < \frac{1}{2}$ and $n \in \mathbb{N}$. If $cn < k < (1 - c)n$, then*

$$\max\left\{\beta_{[\pm 1]}(n, k), \beta_{[\pm 1]}(n, k-1), \beta_{[\pm 1]}(n, k-2), \beta_{[\pm 1]}(n, k-3)\right\} \geq \delta n,$$

where $\delta = \delta(c)$ is some real positive constant.

Proof. Consider a bisecting family $\mathcal{F}' = \{A'_1, \ldots, A'_m\}$ of minimum cardinality for $\binom{[n]}{l}$, where $cn < l < (1 - c)n$ is even and $\frac{l}{2}$ is odd, for some constant c, $0 < c < \frac{1}{2}$. Let X_A denote the 0–1 incidence vector corresponding to a set $A \subseteq [n]$. Let V denote the vector space generated by the incidence vectors of $\mathcal{F}'$ over $\mathbb{F}_2$. Observe that for any $A \in \binom{[n]}{l}$, there exists an $A' \in \mathcal{F}'$ such that $|A \cap A'| = \frac{l}{2}$. Since $\frac{l}{2}$ is odd, $\langle X_A, X_{A'} \rangle = 1$, i.e., $X_A \notin V^\perp$, where $V^\perp$ is the subspace of the vector space $\{0, 1\}^n$ over $\mathbb{F}_2$ which contains all the vectors perpendicular to V. So, $V^\perp$ is a subspace containing no vector of weight l. For any $X_B, X_C \in V^\perp$, $X_B + X_C$ has weight $|B \triangle C| \neq l$. Moreover, l is even. Since $cn < l < (1 - c)n$, using Theorem 23, there exists a positive constant $\delta = \delta(c)$ such that $|V^\perp| \leq 2^{n(1-\delta)}$. So, $dim(V^\perp) \leq n - \lfloor \delta n \rfloor$. It follows that $dim(V) \geq \lfloor \delta n \rfloor$. To complete the proof of the theorem, note that for any k, there exists an $l \in \{k, k-1, k-2, k-3\}$ such that l is even and $\frac{l}{2}$ is odd. □

4.4. $\beta_0(n, k)$ and computation of bisecting families

An important probabilistic tool used in this section is the Lovász local lemma [6]. Let $\mathcal{F}$ be a family of subsets of $[n]$. The *dependency* of a set $A \in \mathcal{F}$ denoted by $d(A, \mathcal{F})$ is the number of subsets $\widehat{A} \in \mathcal{F}$, such that (i) $|A \cap \widehat{A}| \geq 1$, and (ii) $A \neq \widehat{A}$. The *dependency* of a family $\mathcal{F}$, denoted by $d(\mathcal{F})$ or simply d, is the maximum dependency of any subset A in the family $\mathcal{F}$. We have the following corollary of the Lovász local lemma from [15].

Lemma 24 ([15]). *Let $\mathcal{P}$ be a finite set of mutually independent random variables in a probability space. Let $\mathcal{A}$ be a finite set of events determined by these variables, where $m = |\mathcal{A}|$. For any $A \in \mathcal{A}$, let $\Gamma(A)$ denote the set of all the events in $\mathcal{A}$ that depend on A. Let $d = \max_{A \in \mathcal{A}} |\Gamma(A)|$. If $\forall A \in \mathcal{A} : P[A] \leq p$ and $ep(d + 1) \leq 1$, then an assignment of the variables not violating any of the events in $\mathcal{A}$ can be computed using expected $\frac{1}{d}$ resamplings per event and expected $\frac{m}{d}$ resamplings in total.*

Proof of Theorem 8

Statement. *For a family $\mathcal{F}$ consisting of k-sized subsets of $[n]$ and dependency d, $\beta_{[\pm 1]}(\mathcal{F}) \leq \frac{\sqrt{k}}{c}(\ln(d+1)+1)$, where $c = 0.67$.*

Please cite this article in press as: N. Balachandran, et al., Bisecting and D-secting families for set systems, Discrete Applied Mathematics (2017), http://dx.doi.org/10.1016/j.dam.2017.05.005.

Proof. Let $\mathcal{F}$ be a family of k-sized subsets of $[n]$, $\mathcal{F} \subseteq \binom{[n]}{k}$, with dependency d. Assume that k is even. Consider a family $\mathcal{F}' = \{A'_1, \ldots, A'_t\}$: each $A'_j \in \mathcal{F}'$ is a random subset of $[n]$ where each point $x \in [n]$ is chosen into A'_j independently with probability $\frac{1}{2}$. Let p be the probability that a fixed subset $A \in \mathcal{F}$ is bisected by some $A'_j \in \mathcal{F}'$.

$$p = \frac{\binom{k}{\frac{k}{2}}}{\binom{k}{0} + \binom{k}{1} + \cdots + \binom{k}{k}} \geq \frac{c}{\sqrt{k}}, \text{ where } c = 0.67.$$

So, the failure probability that A is not bisected by A'_j is $1 - p$ which is at most $1 - \frac{c}{\sqrt{k}}$. Therefore, the failure probability that A is not bisected by any $A'_j \in \mathcal{F}'$ is $(1-p)^t$ which is at most $(1 - \frac{c}{\sqrt{k}})^t \leq e^{-\frac{ct}{\sqrt{k}}}$. Using Lemma 24, we get $t \geq \frac{\sqrt{k}}{c}(\ln(d+1)+1)$. This implies that there exists a bisecting family for any family $\mathcal{F}$ of k-sized sets of size $\frac{\sqrt{k}}{c}(\ln(d+1)+1)$, where d denotes the dependency of family $\mathcal{F}$.

In fact, if $\mathcal{F}$ is $\binom{[n]}{k}$ and we choose the subsets $A'_j \in \mathcal{F}'$ of cardinality exactly $\frac{n}{2}$ uniformly and independently at random from $\binom{[n]}{\frac{n}{2}}$, then $p = \frac{\binom{\frac{n}{2}}{\frac{k}{2}}^2}{\binom{n}{k}} \geq c_1\sqrt{\frac{n}{(n-k)k}}$ ($c_1 \geq 0.53$). Therefore, the failure probability that A is not bisected by any $A'_j \in \mathcal{F}'$ is $(1-p)^t$. Using Lemma 24, we can compute a bisecting family for $\binom{[n]}{k}$ of size $\frac{1}{c_1}\sqrt{\frac{k(n-k)}{n}}(\ln(d+1)+1)$. Therefore, using Observation 11, $\beta_{[\pm 1]}(n,k)$ is $O((\ln(d+1)+1))$-approximable.

The proof for the case when k is odd is similar to the above proof. In fact, we get a small constant factor improvement over the bound given in Theorem 8. □

Let $m = |\mathcal{F}|$. Since, $d+1 \leq m \leq \binom{n}{k} < (\frac{en}{k})^k$, we get, $\beta_{[\pm 1]}(n,k) \leq \frac{1}{c_1}\sqrt{\frac{k(n-k)}{n}}(\ln m + 1) \leq \frac{k}{c_1}\sqrt{\frac{k(n-k)}{n}} \ln(\frac{en}{k})$.

5. Discussion and open problems

The discrepancy interpretation of bisecting families leads us to the investigation of $\beta_{[\pm 1]}(\mathcal{F})$ for recursive Hadamard set systems.

Bisecting families for Hadamard set systems

Definition 25. *A Hadamard matrix H is a $n \times n$ matrix with* (i) *each entry being either $+1$ or -1, and* (ii) *any two distinct columns being orthogonal, i.e., $H^T H = nI$, where I is the $n \times n$ identity matrix.*

By convention, the first row and first column of H are all ones. By a recursive construction, $H(k)$ of size $2^k \times 2^k$ can be obtained from $H(k-1)$ of size $2^{k-1} \times 2^{k-1}$ as follows:

$$H(k) = \begin{bmatrix} H(k-1) & H(k-1) \\ H(k-1) & -H(k-1) \end{bmatrix},$$

where $H(0) = 1$. Note that except the first row, every other row of the Hadamard matrix $H(k)$ must contain equal number of 1's and -1's, since the columns are orthogonal and $H(k)$ is symmetric. Let $A = \frac{1}{2}(H(k) + J(k))$, where J is the $2^k \times 2^k$ matrix whose every entry is $+1$. The matrix A corresponds to the Hadamard set system $HF(k)$, where $HF(k) = \{A_1, \ldots, A_{2^k}\}$, and, $j \in A_i$ if and only if the (i,j) entry of A is one. So, from construction, every subset $A_j \in HF(k)$ except A_1 is of cardinality exactly 2^{k-1}. It is a well known fact that a Hadamard set system HF of order $n \times n$ has a discrepancy at least $\frac{\sqrt{n-1}}{2}$ [13, p. 106]. Therefore, $\beta_{[\pm 1]}(HF(k)) \geq 2$. In what follows, we show that $\beta_{[\pm 1]}(HF(k)) \leq 2$ for all Hadamard set systems obtained from the recursively constructed Hadamard matrix $H(k)$, $k > 1$. Consider the Hadamard set system $HF(k)$, which is represented by the incidence matrix A. Let $B_1 = \{1, \ldots, 2^{k-1}\}$. Observe that A_1 through $A_{2^{k-1}}$ of $HF(k)$ are bisected by B_1 due to the recursive construction. $A_{2^{k-1}+1}$ represented by the $2^{k-1}+1$th row of A is not bisected by B_1. In fact, $|A_{2^{k-1}+1} \cap B_1| - |A_{2^{k-1}+1} \cap ([2^k] \setminus B_1)| = 2^{k-1}$. The subsets $A_{2^{k-1}+2}$ through A_{2^k} of $HF(k)$ are bisected by B_1 since every row, except the first row, of $H(k-1)$ and $-H(k-1)$ contains equal number of 1's and -1's. $A_{2^{k-1}+1}$ represented by the $2^{k-1}+1$th row of A can be bisected by a second subset $B_2 = \{1, \ldots, 2^{k-2}\}$. So, this establishes $\beta_{[\pm 1]}(HF(k)) = 2$, $k > 1$.

From the above discussion, it is clear that discrepancy of a set system $\mathcal{F}$ can be arbitrarily large as compared to $\beta_{[\pm 1]}(\mathcal{F})$. On the other extreme, we know that discrepancy of a family of 2-sized subsets $\mathcal{F}$ of $[n]$ cannot exceed 2, whereas $\beta_{[\pm 1]}(\mathcal{F})$ can be as large as $\log n$. Thus, there exist families $\mathcal{F}$ and $\mathcal{G}$ where $\beta_{[\pm 1]}(\mathcal{F})$ and $disc(\mathcal{G})$ are constants whereas $disc(\mathcal{F})$ and $\beta_{[\pm 1]}(\mathcal{G})$ are arbitrarily large. However, this does not rule out a possible relationship between these two parameters and other hypergraph parameters. One possibility of making progress in this direction is obtaining tight upper and lower bounds for $\beta_{[\pm 1]}(\mathcal{F})$. Recall that the discrepancy of a family $\mathcal{F}$ is the minimum $i \in \mathbb{N}$ such that $\beta_{[\pm i]}(\mathcal{F}) \leq 1$. Below, we demonstrate the usage of such tight bounds where $\mathcal{F} = 2^{[n]}$ and n is a power of 2. From Theorem 3, we have, $\frac{n}{2} \geq \beta_{[\pm 1]}(n) \geq 2\beta_{[\pm 2]}(n) \geq \cdots \geq 2^j\beta_{[\pm 2^j]}(n)$. So, when $j = \log(\frac{n}{2})$, we get, $\beta_{[\pm 2^j]}(n) \leq 1$. This gives a known trivial upper bound for $disc(\mathcal{F})$.

As mentioned in the introduction, $\beta_{[\pm 1]}(E)$ is $\lceil \log \chi(G) \rceil$ for a graph $G(V,E)$. We know that it is impossible to approximate the chromatic number of graphs on n vertices within a factor of $n^{1-\epsilon}$ for any fixed $\epsilon > 0$, unless $NP \subseteq ZPP$ (see Feige and

Killian [7]). Therefore, it is not difficult to see that under the assumption $NP \not\subseteq ZPP$, no polynomial time algorithm can approximate $\beta_{[\pm 1]}(E)$ for an n-vertex graph $G(V, E)$ within an additive approximation factor of $(1-\epsilon)\log n - 1$, for any fixed $\epsilon > 0$.

In Section 1.3, we have seen that $\beta_D(n, k)$ is not monotone with k in general. However, it is possible that $\beta_D(n, k)$ is monotone with k in certain ranges, say when $k \leq \frac{n}{2}$. In Section 3.2, we established the lower bound of $\frac{n-i+1}{2}$ for $\beta_i(n)$. However, the best upper bound we have for this case is just $n - i + 1$. So, there is a gap between the lower and upper bounds for $\beta_i(n)$.

Acknowledgments

The research of the third author is supported by the doctoral fellowship program of Ministry of Human Resources and Development, Govt. of India. The research of the first author is supported by grant 12IRCCSG016, IRCC, IIT Bombay.

References

[1] Martin Aigner, Günter M. Ziegler, Alfio Quarteroni, Proofs from the Book, Springer, 2010.
[2] József Beck, Vera T. Sós, in: R.L. Graham, M. Grötschel, L. Lovász (Eds.), Handbook of Combinatorics, Vol. 2, MIT Press, Cambridge, MA, USA, 1995, pp. 1405–1446.
[3] Claude Berge, Graphs and Hypergraphs, North-Holland publishing company Amsterdam, 1973.
[4] Béla Bollobás, Extremal Graph Theory, Dover Publications, INC, Mineola, New York, 2004, pp. 259–261.
[5] Bernard Chazelle, The Discrepancy Method, Cambridge University Press, 2000, Cambridge Books Online.
[6] P. Erdős, L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, Colloq. Math. Soc. János Bolya 10 (1975) 609–627.
[7] Uriel Feige, Joe Kilian, Zero knowledge and the chromatic number, J. Comput. System Sci. 57 (2) (1998) 187–199.
[8] Peter Frankl, Vojtěch Rödl, Forbidden intersections, Trans. Amer. Math. Soc. 300 (1) (1987) 259–286.
[9] Nils Hebbinghaus, Anand Srivastav, Multicolor discrepancy of arithmetic structures, in: A Panorama of Discrepancy Theory, Springer, 2014, pp. 319–424.
[10] G. Katona, On separating systems of a finite set, J. Combin. Theory I (1966) 174–194.
[11] Peter Keevash, Eoin Long, Frankl-rödl-type theorems for codes and permutations, Trans. Amer. Math. Soc. 369 (2) (2017) 1147–1162.
[12] L. Lovász, Kneser's conjecture, chromatic number, and homotopy, J. Combin. Theory Ser. A 25 (3) (1978) 319–324.
[13] Jiří Matoušek, Combinatorial discrepancy, in: Geometric Discrepancy, in: Algorithms and Combinatorics, vol. 18, Springer Berlin Heidelberg, 1999, pp. 101–135.
[14] Marvin Minsky, Seymour Papert, Perceptron: an introduction to computational geometry, The MIT Press, Cambridge, Expanded Edition, 1969, p. 2.
[15] Robin A. Moser, Gábor Tardos, A constructive proof of the general Lovász local lemma, J. ACM 57 (2) (2010) 11:1–11:15.
[16] R. Paturi, M.E. Saks, On threshold circuits for parity, in: 31st Annual Symposium on Foundations of Computer Science, 1990, Proceedings, vol. 1, 1990, pp. 397–404.
[17] A. Rényi, On random generating elements of a finite boolean algebra, Acta Sci. Math. (Szeged) 22 (1–2) (1961) 75–81.
[18] Michael E. Saks, Slicing the hypercube, in: Surveys in Combinatorics, 1993, Cambridge University Press, 1993, pp. 211–256.
[19] I. Wegener, On separating systems whose elements are sets of at most k elements, Discrete Math. 28 (1979) 219–222.

Please cite this article in press as: N. Balachandran, et al., Bisecting and D-secting families for set systems, Discrete Applied Mathematics (2017), http://dx.doi.org/10.1016/j.dam.2017.05.005.