Low-rank matrices, tournaments, and symmetric designs[*]

Niranjan Balachandran & Brahadeesh Sankarnarayanan[†]
Department of Mathematics, Indian Institute of Technology Bombay,
Mumbai 400076, Maharashtra, India.
{niranj,bs}@math.iitb.ac.in

29 April 2024

Abstract

Let $\mathbf{a} = (a_i)_{i \geq 1}$ be a sequence in a field $\mathbb{F}$, and $f \colon \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ be a function such that $f(a_i, a_i) \neq 0$ for all $i \geq 1$. For any tournament T over $[n]$, consider the $n \times n$ symmetric matrix M_T with zero diagonal whose (i, j)th entry (for $i < j$) is $f(a_i, a_j)$ if $i \to j$ in T, and $f(a_j, a_i)$ if $j \to i$ in T. It is known [3] that if T is a uniformly random tournament over $[n]$, then $\operatorname{rank}(M_T) \geq (\frac{1}{2} - o(1))n$ with high probability when $\operatorname{char}(\mathbb{F}) \neq 2$ and f is a linear function.

In this paper, we investigate the other extremal question: how low can the ranks of such matrices be? We work with sequences $\mathbf{a}$ that take only two distinct values, so the rank of any such $n \times n$ matrix is at least $n/2$. First, we show that the rank of any such matrix depends on whether an associated bipartite graph has certain eigenvalues of high multiplicity. Using this, we show that if f is linear, then there are real matrices $M_T(f; \mathbf{a})$ of rank at most $\frac{n}{2} + O(1)$. For rational matrices, we show that for each $\varepsilon > 0$ we can find a sequence $\mathbf{a}(\varepsilon)$ for which there are matrices $M_T(f; \mathbf{a})$ of rank at most $(\frac{1}{2} + \varepsilon)n + O(1)$. These matrices are constructed from symmetric designs, and we also use them to produce bisection-closed families of size greater than $\lfloor 3n/2 \rfloor - 2$ for $n \leq 15$, which improves the previously best known bound given in [5].

1 Introduction

Let a and b be two nonzero real numbers, and let T be a tournament over $[n]$. Consider the $n \times n$ symmetric matrix M_T with zero diagonal and $M_T(i, j) = a$ if $i \to j$ in T, and $M_T(i, j) = b$ if $j \to i$ in T, for all $i < j$.

What can be said about the rank of M_T?

1.1 Fractionally intersecting families: A motivating extremal problem

Linear algebraic methods are an important set of tools in the study of extremal problems in combinatorics; for more on this, see Babai–Frankl [1]. It is often the case that one can recast an extremal problem into the computation of certain matrix properties (its spectrum, rank, and so on).

The above question initially arose from the following extremal problem posed by Balachandran–Mathew–Mishra [5]. A θ-*intersecting family* $\mathcal{F}$ over $[n]$ is a collection of subsets of $[n]$ such that for all distinct $A, B \in \mathcal{F}$, we have $|A \cap B| = \theta|A|$ or $\theta|B|$, where $\theta \in (0,1) \cap \mathbb{Q}$. What is the maximum size of a θ-intersecting family over $[n]$?

The general problem remains open; see [2, 8, 9] for related work. A linear upper bound on the size of the θ-intersecting family $\mathcal{F} = \{A_1, \ldots, A_m\}$ over $[n]$ can be obtained[1] by showing that an associated real symmetric $m \times m$ matrix with zero diagonal, $M_{\mathcal{F}}$, has "high" rank.[2] Consequently, we turn our attention to the following ensemble of matrices.

Notation. Let $\mathbb{F}$ be a field, and $a_1, \ldots, a_n \in \mathbb{F}$. By $\mathsf{Sym}(f; a_1, \ldots, a_n)$ we denote the collection of all the $n \times n$ symmetric matrices over $\mathbb{F}$ with zero diagonal whose (i,j)th entries are either $f(a_i, a_j)$ or $f(a_j, a_i)$, for all $i < j$.

Notation. Let $\mathbb{F}$ be a field, $\mathbf{a}$ be a sequence in $\mathbb{F}$, and $f \colon \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ be a function.

1. By $\mathbf{a}_n$ we mean the tuple $(a_1, \ldots, a_n)$.

2. We say that $(f; \mathbf{a})$ is a *good pair* (over $\mathbb{F}$) if $f(a_i, a_i) \neq 0$ for all i.

3. For $\theta \in \mathbb{F}$, we write $f = f_\theta$ if $f(x, y) = x + (1 - 2\theta)y$ for all $x, y \in \mathbb{F}$.

Problem 1.1. Let $(f; \mathbf{a})$ be a good pair over $\mathbb{F}$. Is there an absolute constant $c > 0$ such that $\mathrm{rank}(M) \geq cn$ for all $M \in \mathsf{Sym}(f; \mathbf{a}_n)$?

1.2 $\mathsf{Sym}(f; \mathbf{a}_n)$: A high-rank ensemble

Problem 1.1 was initially posed in [5] for $\mathbb{F} = \mathbb{R}$, $f = f_{1/2}$ and a sequence $\mathbf{a}$ of positive reals. In particular, the $m \times m$ matrix $M_{\mathcal{F}}$ mentioned earlier is a member of $\mathsf{Sym}(f_\theta; |A_1|, \ldots, |A_m|)$.[3] However, this ensemble will in general contain many more matrices than just (those similar to) $M_{\mathcal{F}}$, and we describe a natural combinatorial interpretation for all the matrices in $\mathsf{Sym}(f; \mathbf{a}_n)$.

For a tournament T over $[n]$, define $M_T \in \mathsf{Sym}(f; \mathbf{a}_n)$ by

$$M_T(i, j) = \begin{cases} f(a_i, a_j), & i \to j \text{ in } T; \\ f(a_j, a_i), & j \to i \text{ in } T, \end{cases}$$

for all $i < j$. Conversely, given a matrix $M \in \mathsf{Sym}(f; \mathbf{a}_n)$, define a tournament T_M over $[n]$ by setting $i \to j$ in T_M if $M(i, j) = f(a_i, a_j)$, and $j \to i$ in T if $M(i, j) = f(a_j, a_i)$, for all $i < j$. Note that the correspondence $M \to T_M$ is not one-one since we could have $f(a_i, a_j) = f(a_j, a_i)$ for some $i < j$.

While there are good reasons to believe that $M_{\mathcal{F}}$ has high rank,[4] is there evidence to suggest that

---

[1] The argument is implicit in [5] and appears explicitly in [11].

[2] To be precise, if $\mathrm{rank}(M_{\mathcal{F}}) \geq cm$, then $|\mathcal{F}| \leq c^{-1}(n + 1)$.

[3] Specifically, it is the matrix whose (i, j)th entry is $f(|A_i|, |A_j|)$ if $|A_i \cap A_j| = \theta|A_j|$, and $f(|A_j|, |A_i|)$ otherwise.

[4] Indeed, we do not even know of any θ-intersecting family over $[n]$ of size at least $2n$.

all the matrices in the ensemble $\mathsf{Sym}(f; \mathbf{a}_n)$ are of high rank? In [12], Tao and Vu study the rank of a random symmetric matrix $M_n = ((\xi_{ij}))$, where ξ_{ij} are all jointly independent (for $i < j$) and also independent of ξ_{ii} (which are also independent) with the additional property that for all $i < j$ and all real x, $\mathbb{P}(\xi_{ij} = x) \leq 1 - \mu$ for some fixed constant μ. They show that the spectrum of M_n is simple with high probability (*whp*); so, a random matrix in $\mathsf{Sym}(f; \mathbf{a}_n)$ has rank at least $n - 1$ with high probability in the case when all the a_i are distinct reals.

To see why we require $(f; \mathbf{a})$ to be a good pair, consider the following matrix

$$
M_0 = \begin{bmatrix}
0 & 1 & 4 & \cdots & (n-1)^2 \\
1 & 0 & 1 & \cdots & (n-2)^2 \\
4 & 1 & 0 & \cdots & (n-3)^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(n-1)^2 & (n-2)^2 & (n-3)^2 & \cdots & 0
\end{bmatrix},
$$

whose (i, j)th entry is $(i - j)^2$. It is easy to see that $\mathrm{rank}(M_0) \leq 3$ for all $n \geq 3$, since M_0 is the sum of three rank one matrices, M_1, M_2, and M_3, given by $M_1(i, j) = i^2$, $M_2(i, j) = -2ij$, $M_3(i, j) = j^2$. Note that $M_0 \in \mathsf{Sym}(f; \mathbf{a}_n)$, where $f(x, y) = (x - y)^2$ and $a_i = i$, but the same argument works for any sequence $\mathbf{a}$ since $f(x, x) = 0$ for all $x \in \mathbb{F}$. Hence, $\mathrm{rank}(M) \leq 3$ for all $M \in \mathsf{Sym}(f; \mathbf{a}_n)$ for $f(x, y) = (x - y)^2$.

Another reason is that in [3] the authors showed that for a rank r function f, say $f(x, y) = \sum_{i=1}^{r} g_i(x) h_i(y)$, the rank of a matrix in $\mathsf{Sym}(f; \mathbf{a}_n)$ arising from a uniformly random tournament is at least $(\frac{1}{2r} - o(1))n$ *whp*, provided that $(f; \mathbf{a})$ is a good pair over a field of characteristic different from 2.

1.3 Low rank matrices in $\mathsf{Sym}(f; \mathbf{a}_n)$

The discussion in the previous section shows why a typical member of $\mathsf{Sym}(f; \mathbf{a}_n)$ has high rank. However, we do not know of many concrete examples where the rank can be computed exactly. In [4], it is shown that the matrices associated to transitive tournaments all have high rank (specifically, at least $n - 1$), when $f = f_{1/2}$. Furthermore, the probabilistic results mentioned in the previous section cannot rule out the existence of a vanishingly small proportion of matrices in $\mathsf{Sym}(f; \mathbf{a}_n)$ whose ranks drop arbitrarily low. So, it is interesting to see what can be said about matrices of *low* rank in $\mathsf{Sym}(f; \mathbf{a}_n)$.

The best examples that we know of are those associated to maximal bisection-closed families (i.e., $\frac{1}{2}$-intersecting families). The family $\mathcal{F}_{\text{sunflower}}$ over $[n]$ is constructed as a union of two sunflowers, containing $n - 1$ sets of size 2 and $\frac{n}{2} - 1$ sets of size 4. The family $\mathcal{F}_{\text{Hadamard}}$ over $[n]$ is constructed from a Hadamard matrix of order n, and contains $n - 1$ sets of size $n/2$ and $\frac{n}{2} - 1$ sets of size $n/4$. Since these maximal families over $[n]$ have size $3n/2 - 2$, the associated matrices have rank at most $\frac{2n}{3}$.

Taking a cue from these constructions, we restrict our attention to matrices in $\mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$, where $\alpha, \beta \in \mathbb{R}$ satisfy $f(\alpha, \alpha) \neq 0$ and $f(\beta, \beta) \neq 0$, and $(\alpha^{(m)}, \beta^{(n)})$ is shorthand for the sequence $(\alpha, \ldots, \alpha, \beta, \ldots, \beta)$ containing m occurrences of α and n occurrences of β. Naturally, every matrix in $\mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$ has rank at least $r := \max\{m, n\}$. How close can one get to r?

3

Our first result says that the rank of a matrix $M \in \mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$ depends on whether a scalar $\mu(f; \alpha, \beta) \in \mathbb{C}$ such that

$$\big(\mu(f; \alpha, \beta)\big)^2 = \frac{f(\alpha, \alpha) f(\beta, \beta)}{(f(\alpha, \beta) - f(\beta, \alpha))^2}. \tag{1}$$

is an eigenvalue of high multiplicity for a bipartite graph G_M associated to M, which we define in Section 2.

Theorem 1.1. *Let $M \in \mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$, where $f(\alpha, \alpha)$, $f(\beta, \beta)$, and $f(\alpha, \beta) - f(\beta, \alpha)$ are all nonzero. Let $\mu(f; \alpha, \beta) \in \mathbb{C}$ be as in (1). There exists a bipartite graph G_M with part sizes m and n such that $m + n - 2 - \nu \le \mathrm{rank}(M) \le m + n + 2 - \nu$, where ν is the multiplicity of $\mu(f; \alpha, \beta)$ as an eigenvalue of G_M. In particular, if $f(\alpha, \alpha) f(\beta, \beta) < 0$, then $\mathrm{rank}(M) > m + n - 2$.*

Note that when $f = f_\theta$ (for $\theta \in (0, 1)$), the hypotheses of Theorem 1.1 are satisfied so long as α and β are distinct and nonzero. Next, we show that for any $\theta \in (0, 1)$ there exist $\alpha, \beta \in \mathbb{R}$ for which there are matrices in $\mathsf{Sym}(f_\theta; \alpha^{(n)}, \beta^{(n)})$ having rank at most $n + 3$.

Theorem 1.2. *Let $\theta \in (0, 1)$. Let $\beta \in \mathbb{R}$ be a root of the quadratic equation $x^2 - (2 + (\theta^{-1} - 1)^2)x + 1 = 0$. For every $n \in \mathbb{N}$, there is a matrix $M \in \mathsf{Sym}(f_\theta; 1^{(n)}, \beta^{(n)})$ with $\mathrm{rank}(M) \le n + 3$.*

When $\theta = 1/2$, the matrices that we find in the proof of Theorem 1.2 have irrational entries. This raises the question whether there exist rational matrices in $\mathsf{Sym}(f_{1/2}; \alpha^{(n)}, \beta^{(n)})$ of rank close to $n/2$. While we are unable to settle the question completely, we show in the next theorem that we can get arbitrarily close to $n/2$.

Theorem 1.3. *For each $\varepsilon > 0$, there exists $c_\varepsilon \in (\frac{1}{2}, \frac{1}{2} + \varepsilon)$ and $\alpha_\varepsilon, \beta_\varepsilon \in \mathbb{Q}$ such that there is a sequence of matrices $M_n \in \mathsf{Sym}(f_{1/2}; \alpha_\varepsilon^{(n)}, \beta_\varepsilon^{(n)})$ for which $\mathrm{rank}(M_n) \le c_\varepsilon n + O(1)$.*

We conclude with some combinatorial consequences of these results. In particular, we show that in some small cases (i.e., for $n \le 15$), we can use these constructions to find bisection-closed families of size greater than $\lfloor 3n/2 \rfloor - 2$, which was the previously best known lower bound given in [5].

2 Preliminaries

For the rest of the paper, we will be working with real matrices in $\mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$, where $f(\alpha, \alpha) \ne 0$ and $f(\beta, \beta) \ne 0$. The latter conditions just say that $(f; \mathbf{a})$ is a good pair for any sequence $\mathbf{a}$ taking values in $\{\alpha, \beta\}$. Let I_n denote the $n \times n$ identity matrix, let $J_{m \times n}$ denote the $m \times n$ all-ones matrix, and let $J_n := J_{n \times n}$, for all $m, n \in \mathbb{N}$. We sometimes denote $\mu(f; \alpha, \beta)$ simply by μ when the context is clear. Recall that by f_θ we mean the function given by $(x, y) \mapsto x + (1 - 2\theta)y$.

To each matrix $M \in \mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$, we associate a subgraph G_M of the complete bipartite graph $K_{m,n}$ as follows: for each $i \in [m]$ and $j \in [n]$, there is an edge ij in G_M if and only if $M(i, j) = f(\alpha, \beta)$. Conversely, given a subgraph G of $K_{m,n}$, the associated matrix $M_G \in \mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$ is defined as follows: if $B_{m \times n}$ is the $\{0, 1\}$-incidence matrix of G, let $C = f(\alpha, \beta)B + f(\beta, \alpha)(J_{m \times n} - B)$, and define M_G to be the $(m + n) \times (m + n)$ matrix

$$\begin{bmatrix} f(\alpha, \alpha)(J_m - I_m) & C \\ C^T & f(\beta, \beta)(J_n - I_n) \end{bmatrix}.$$

These correspondences are natural inverses of each other.

4

If $f(\alpha, \beta) = f(\beta, \alpha)$, then there is only one matrix in the family $\mathsf{Sym}(f; \alpha^{(m)}, \beta^{(n)})$, and it has rank at least $m + n - 2$. So, to avoid this trivial scenario, we henceforth assume that $f(\alpha, \beta) \neq f(\beta, \alpha)$.

The following lemma describes some properties of $\mu(f_\theta; \alpha, \beta)$, which we will refer to in Section 4.

Lemma 2.1. *If α, β are distinct nonzero real numbers, then*

$$\left(\mu(f_\theta; \alpha, \beta)\right)^2 = \frac{(1-\theta)^2 \alpha\beta}{\theta^2 (\alpha - \beta)^2}.$$

In particular, $\mu(f_\theta; \alpha, \beta) = \mu(f_\theta; \beta, \alpha)$ and $\mu(f_\theta; \alpha, \beta) = \mu(f_\theta; k\alpha, k\beta)$ for all nonzero real numbers k. Thus, $\mu(f_\theta; \alpha, \beta) = \mu(f_\theta; 1, \frac{\alpha}{\beta}) = \mu(f_\theta; 1, \frac{\beta}{\alpha})$.

Proof. We have

$$f_\theta(\alpha, \alpha) = 2(1-\theta)\alpha,$$
$$f_\theta(\beta, \beta) = 2(1-\theta)\beta,$$
$$f_\theta(\alpha, \beta) - f_\theta(\beta, \alpha) = 2\theta(\alpha - \beta).$$

Hence,

$$\left(\mu(f_\theta; \alpha, \beta)\right)^2 = \frac{f_\theta(\alpha, \alpha)f_\theta(\beta, \beta)}{(f_\theta(\alpha, \beta) - f_\theta(\beta, \alpha))^2} = \frac{(1-\theta)^2 \alpha\beta}{\theta^2 (\alpha - \beta)^2}.$$

Since the RHS is symmetric and homogeneous of degree 0 in α and β, we have $\mu(f_\theta; \alpha, \beta) = \mu(f_\theta; \beta, \alpha)$ and $\mu(f_\theta; \alpha, \beta) = \mu(f_\theta; k\alpha, k\beta)$ for all nonzero real numbers k. Taking $k = 1/\alpha$, we get $\mu(f_\theta; \alpha, \beta) = \mu(f_\theta; 1, \frac{\alpha}{\beta})$. By symmetry, this equals $\mu(f_\theta; \frac{\alpha}{\beta}, 1)$. Now, taking $k = \frac{\beta}{\alpha}$, we see that this also equals $\mu(f_\theta; 1, \frac{\beta}{\alpha})$. $\square$

We will refer to the following result due to Rowlinson in Section 4.1, which concerns bipartite graphs having an eigenvalue of high multiplicity.

Theorem 2.2 (Rowlinson [10, Theorems 3.4 and 4.2]). *Let G be a connected bipartite graph of order $n > 5$, with $\mu \notin \{-1, 0\}$ as an eigenvalue of multiplicity $\nu > 1$.*

(a) *If d is the maximum degree in G, then $\nu \leq n - 1 - d$.*

(b) *If equality holds in (a), then $\nu \leq d - 1$.*

(c) *If equality holds in (b), then G is the bipolar cone over a graph G_0, where G_0 is either the incidence graph of a symmetric 2-design, or a 2-balanced bipartite graph.*

3 Main results

Proof of Theorem 1.1

Without loss of generality, assume that $m \geq n$. The matrix

$$\Gamma = \begin{bmatrix} f(\alpha, \alpha)J_m & f(\beta, \alpha)J_{m \times n} \\ f(\beta, \alpha)J_{n \times m} & f(\beta, \beta)J_n \end{bmatrix}$$

has rank at most 2, so $\text{rank}(M - \Gamma) - 2 \leq \text{rank}(M) \leq \text{rank}(M - \Gamma) + 2$. Since $\text{rank}(M - \Gamma) = m + n - \text{nullity}(M - \Gamma)$, we shall find the multiplicity of 0 as a root of the characteristic polynomial $\mathsf{CharPol}(x)$ of $M - \Gamma$:

$$\mathsf{CharPol}(x) = \det \begin{bmatrix} (x + f(\alpha, \alpha))I_m & (f(\beta, \alpha) - f(\alpha, \beta))B \\ (f(\beta, \alpha) - f(\alpha, \beta))B^T & (x + f(\beta, \beta))I_n \end{bmatrix}.$$

For a block matrix $X = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$, if $\det(P) \neq 0$, then $\det(X) = \det(P)\det(S - RP^{-1}Q)$. So, $\mathsf{CharPol}(x)$ equals

$$(f(\beta, \alpha) - f(\alpha, \beta))^{2n}(x + f(\alpha, \alpha))^{m-n} \det\left(\frac{(x + f(\alpha, \alpha))(x + f(\beta, \beta))}{(f(\beta, \alpha) - f(\alpha, \beta))^2}I_n - B^T B \right).$$

Let A be the adjacency matrix of the bipartite graph G_M, so $A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}$. Note that:

- the spectrum of A is real and symmetric about 0;

- $A^2 = \begin{bmatrix} BB^T & 0 \\ 0 & B^T B \end{bmatrix}$; and

- the spectra of BB^T and $B^T B$ are identical (other than the $m - n$ many additional zeros in the spectrum of BB^T).

Thus, 0 is an eigenvalue of $M - \Gamma$ if and only if $\mu(f; \alpha, \beta)$ is an eigenvalue of A (i.e., of G_M). $\qquad \square$

Proof of Theorem 1.2

By Lemma 2.1, there is no loss of generality in setting $\alpha = 1$, so we state the result by specifying the required real number β. It is easy to see that $\text{Spec}(K_n) = \{n-1, (-1)^{n-1}\}$. Let G_n be the bipartite cover of K_n, so that G_n is the complete bipartite graph $K_{n,n}$ minus a perfect matching. Note that if v is an eigenvector of K_n with eigenvalue λ, then (v, v) and $(v, -v)$ are eigenvectors of G_n with eigenvalues λ and $-\lambda$, respectively. Consequently, $\text{Spec}(G_n) = \{n-1, (1)^{n-1}, (-1)^{n-1}, -n+1\}$. Given a nonzero β, let $M(\beta)$ be the matrix in $\mathsf{Sym}(f_\theta; 1^{(n)}, \beta^{(n)})$ associated to G_n. From Lemma 2.1, we have

$$\left(\mu(f_\theta; 1, \beta)\right)^2 = \frac{(1-\theta)^2\beta}{\theta^2(1-\beta)^2}.$$

So, we set $\mu(f_\theta; 1, \beta) = 1$ and by rearranging terms we see that β must be a solution to the quadratic equation $x^2 - (2 + (\theta^{-1} - 1)^2)x + 1 = 0$. In particular, we have

$$\beta^2 - (2 + (\theta^{-1} - 1)^2)\beta + 1 = 0 \iff \beta_\pm = \frac{2 + (\theta^{-1} - 1)^2 \pm (\theta^{-1} - 1)\sqrt{(\theta^{-1} - 1)^2 + 4}}{2}.$$

Note that $\beta_\pm \in \mathbb{R}$, so $M(\beta_\pm)$ is in fact a real matrix, and furthermore $\text{rank}(M(\beta_\pm)) \leq n + n + 2 - (n-1) = n + 3$. $\qquad \square$

Note that $\beta \in \mathbb{Q}$ for certain values of θ in Theorem 1.2; for example, if $\theta = \frac{2}{5}$, then $\beta_+ = 4$, and if $\theta = \frac{3}{11}$, then $\beta_+ = 9$, and so on. However, when $\theta = \frac{1}{2}$, we have $\beta_\pm = \frac{3\pm\sqrt{5}}{2}$. Nevertheless, can one find rational (or integral) matrices $M \in \mathsf{Sym}(f_{1/2}; \alpha^{(n)}, \beta^{(n)})$ having rank close to n?

6

To answer this question in the form of Theorem 1.3, we turn to symmetric designs. For Δ a symmetric 2-(v, k, λ) design, let G_Δ be its point-block bipartite incidence graph, and $M_\Delta \in \mathsf{Sym}(f; (\alpha^{(n)}, \beta^{(n)}))$ be the matrix associated to G_Δ.[5] It is well-known that the spectrum of G_Δ is given by $\{v, (\sqrt{k-\lambda})^{(v-1)}, (-\sqrt{k-\lambda})^{(v-1)}, -v\}$. So, Theorem 1.1 implies that either $\mathrm{rank}(M_\Delta) \geq 2v - 3$ or $\mathrm{rank}(M_\Delta) \leq v + 3$. We will use the matrices M_Δ associated to Hadamard designs Δ in order to prove Theorem 1.3. Recall that a Hadamard design is a symmetric 2-(v, k, λ) design with parameters $v = 4n - 1$, $k = 2n - 1$, and $\lambda = n - 1$, for n a positive integer. A Hadamard design with these parameters exists if and only if there is a Hadamard matrix of order $4n$.

Proof of Theorem 1.3

By Lemma 2.1, there is no loss of generality in restricting to integral α and β. The Paley construction [6, §V.6.1.6–7] gives a skew-Hadamard matrix of order $q + 1$ for every prime power $q \equiv 3$ (mod 4). Also, Williamson [13] shows that there is a Hadamard matrix of order $\beta^2 - \beta$ whenever there is a skew-Hadamard matrix of order β.

So, given $\varepsilon > 0$, fix $N \in \mathbb{N}$ such that $c_\varepsilon := \frac{4N+1}{8N-2} \in (\frac{1}{2}, \frac{1}{2} + \varepsilon)$. Choose β_ε to be a prime power congruent to 3 modulo 4 that is sufficiently large so that $n_0 := \beta_\varepsilon^2 - \beta_\varepsilon \geq N$, and let $\alpha_\varepsilon := \beta_\varepsilon - 1$. Let H be a normalized Hadamard matrix of order $4n_0$ obtained from the Kronecker product of any two Hadamard matrices of orders n_0 and 4, respectively. Let Δ be the Hadamard design induced by H, and $M_\Delta \in \mathsf{Sym}(f_{1/2}; \alpha_\varepsilon^{(4n_0-1)}, \beta_\varepsilon^{(4n_0-1)})$ be the associated matrix.

By Theorem 1.1, $\mathrm{rank}(M_\Delta) \leq 4n_0 + 1$ if and only if $\left(\mu(f_{1/2}; \alpha_\varepsilon, \beta_\varepsilon)\right)^2 = k - \lambda$, where $k = 2n_0 - 1$ and $\lambda = n_0 - 1$. On the one hand, $\left(\mu(f_{1/2}; \alpha_\varepsilon, \beta_\varepsilon)\right)^2 = (\beta_\varepsilon - 1)\beta_\varepsilon$. On the other hand, by construction we have $(\beta_\varepsilon - 1)\beta_\varepsilon = n_0$, which equals $k - \lambda$. Thus, by Theorem 1.1, $\mathrm{rank}(M_\Delta) \leq 4n_0 + 1$, and $\frac{4n_0+1}{8n_0-2} \leq \frac{4N+1}{8N-2} = c_\varepsilon$. Lastly, if G_Δ is the bipartite graph associated to M_Δ, then we take arbitrarily many disjoint copies of G_Δ to get a sequence of matrices $M_n \in \mathsf{Sym}(f_{1/2}; (\beta_\varepsilon - 1)^{(n)}, \beta_\varepsilon^{(n)})$ of rank at most $c_\varepsilon n + O(1)$, as required. $\qquad\square$

4   Combinatorial consequences and concluding remarks

4.1   Bipartite graphs with high multiplicity eigenvalues

In order to apply Theorem 1.1 to find matrices of low rank in $\mathsf{Sym}(f; \alpha, \beta)$, we need to construct bipartite graphs that have some eigenvalue of high multiplicity (in particular, one equal to $\mu(f; \alpha, \beta)$). The maximum multiplicity of any eigenvalue that is greater than -2 for bipartite graphs (in fact, more generally for triangle-free graphs) was investigated by Rowlinson [10], and his results that we quoted in Section 2 as Theorem 2.2 show that if there is a bipartite graph having an eigenvalue of maximum possible multiplicity, then it arises as a bipolar cone either of the incidence graph of a symmetric design, or of a 2-balanced bipartite graph. Rowlinson further notes in [10] that there are no known examples of the latter kind, so it is really the symmetric designs that are of interest when trying to construct bipartite graphs having an eigenvalue of high multiplicity.

---

[5]As an aside, the graph G_{M_Δ} as defined in Section 2 is precisely the graph G_Δ in this setting.

4.2 Coincidences in the set sizes of $\mathcal{F}_{\mathrm{sunflower}}$ and $\mathcal{F}_{\mathrm{Hadamard}}$

As noted in Section 1.3, the maximal bisection-closed families $\mathcal{F}_{\mathrm{sunflower}}$ and $\mathcal{F}_{\mathrm{Hadamard}}$ have sets of only two sizes (2 and 4 in $\mathcal{F}_{\mathrm{sunflower}}$, and $n/2$ and $n/4$ in $\mathcal{F}_{\mathrm{Hadamard}}$); moreover, the number of sets of size i in $\mathcal{F}_{\mathrm{sunflower}}$ and of size n/i in $\mathcal{F}_{\mathrm{Hadamard}}$ are equal, for $i = 2, 4$. Lemma 2.1 explains why this is more than just a coincidence. We ignore small additive constants in this discussion for simplicity.

Suppose that $\mathcal{F}_1$ and $\mathcal{F}_2$ are two maximal θ-intersecting families over $[N]$, such that each $\mathcal{F}_i$ only contains sets of two distinct sizes, α_i and β_i. Suppose that these families have the same size, so that the matrices $M_{\mathcal{F}_i} \in \mathsf{Sym}(f_\theta; \alpha_i^{(m_i)}, \beta_i^{(n_i)})$ have the same size. As mentioned in Section 1.1, these matrices will also have the same rank. Theorem 1.1 says that this is because $\mu(f_\theta; \alpha_i, \beta_i)$ is an eigenvalue of the same multiplicity for the associated graphs G_i.

Now, one way for this to happen is that the graphs G_i are isomorphic to each other, and $\mu(f_\theta; \alpha_1, \beta_1) = \mu(f_\theta; \alpha_2, \beta_2)$. As it turns out, the intersection patterns of the families $\mathcal{F}_{\mathrm{sunflower}}$ and $\mathcal{F}_{\mathrm{Hadamard}}$, which are encoded by the associated bipartite graphs, are *almost identical*.[6] But what about the multiplicity of the eigenvalues? Lemma 2.1 assures that $\mu(f_\theta; \alpha_1, \beta_1) = \mu(f_\theta; \alpha_2, \beta_2)$ if $\alpha_1/\beta_1 = (\alpha_2/\beta_2)^{\pm 1}$. And, we do indeed have $2/4 = (n/4)/(n/2)$ in comparing the sunflower and Hadamard families.

While the implications run in the opposite direction to the one we would ideally like, it puts the coincidences remarked upon above into a different light; at the very least, it shows why there might exist these relationships between the sizes of the uniform subfamilies of $\mathcal{F}_{\mathrm{sunflower}}$ and $\mathcal{F}_{\mathrm{Hadamard}}$ as well as the sizes of the sets in them.

4.3 Ruling out candidates for low rank matrices in $\mathsf{Sym}(f_{1/2}; \alpha^{(m)}, \beta^{(n)})$

Theorem 1.3 shows that we can get a sequence M_n of integral matrices in $\mathsf{Sym}(f_{1/2}; \alpha_n^{(n)}, \beta_n^{(n)})$ whose ranks asymptotically tend to $n/2$, using Hadamard designs. We show that many of the other known families of symmetric designs cannot give us such matrices.

Proposition 4.1. *Let Δ be a symmetric 2-(v, k, λ) design with $k - \lambda = p^m$ for some prime p and integer $m \geq 1$. Consider $M_\Delta \in \mathsf{Sym}(f_{1/2}; \alpha^{(v)}, \beta^{(v)})$. If $\mathrm{rank}(M_\Delta) \leq v + 3$, then $p^m = 2$.*

Proof. We may scale the matrix M_Δ by $1/\gcd(\alpha, \beta)$ without changing its rank, so we assume that $\gcd(\alpha, \beta) = 1$. If $\mathrm{rank}(M_\Delta) \leq v + 3$, then $k - \lambda = \left(\mu(f_{1/2}; \alpha, \beta)\right)^2 = \frac{\alpha\beta}{(\alpha - \beta)^2}$. Now, it is easy to see that if there is a prime factor of $(\alpha - \beta)^2$ that is also a factor of $\alpha\beta$, then p must divide both α and β. Since $\gcd(\alpha, \beta) = 1$, we must have $\alpha = \beta \pm 1$, and $p^m = k - \lambda = (\beta \pm 1)\beta$. Without loss of generality, let $p^m = (\beta - 1)\beta$. But, $\gcd(\beta - 1, \beta) = 1$, so $\beta - 1 = 1$, that is, $\beta = 2 = k - \lambda$. In particular, $p^m = 2$. □

Proposition 4.1 rules out most of the known examples of symmetric designs as viable candidates for inducing matrices of low rank. Indeed, a quick glance through [6, §II.6.8] shows that none of Families 1, 7, 8, 9, 13, or 14 are viable, except for the Fano plane $PG(2, 2)$ from Family 1, which

---

[6]To be precise, there is a set $A \in \mathcal{F}_{\mathrm{Hadamard}}$ such that the graphs associated to the families $\mathcal{F}_1 = \mathcal{F}_{\mathrm{Hadamard}} \setminus \{A\}$ and $\mathcal{F}_2 = \mathcal{F}_{\mathrm{sunflower}} \setminus \{12\}$ are isomorphic.

we discuss in further detail below. A similar analysis also rules out the Menon designs (Family 6 in [6, §II.6.8]) which have parameters $v = 4t^2$, $k = 2t^2 - 1$, $\lambda = t^2 - t$.

4.4 θ-intersecting families with sets of only two distinct sizes

As mentioned in Section 1.1, a positive answer to Problem 1.1 would prove that any θ-intersecting family $\mathcal{F}$ over $[n]$ has size at most linear in n, whereas the best known upper bound [5] is $\mathcal{F} \leq O(n \log n)$. When we restrict $\mathcal{F}$ to have sets of only two distinct sizes, as in the known maximal families $\mathcal{F}_{\text{sunflower}}$ and $\mathcal{F}_{\text{Hadamard}}$, we have the trivial upper bound $|\mathcal{F}| \leq 2n$, whereas $|\mathcal{F}_{\text{sunflower}}| = \lfloor 3n/2 \rfloor - 2 = |\mathcal{F}_{\text{Hadamard}}|$. So, can the results in this work improve the trivial upper bound in this scenario?

Theorem 1.2 says we cannot hope to do so by considering ensembles of *real* matrices $\mathsf{Sym}(f_\theta; \alpha^{(m)}, \beta^{(n)})$. Theorem 1.3 says that even restricting to rational or integral ensembles will not refine the $2n$ upper bound. Then, a further natural restriction would to demand that the two set sizes be in $1 : 2$ proportion, just as in $\mathcal{F}_{\text{sunflower}}$ and $\mathcal{F}_{\text{Hadamard}}$. This means that we want to find matrices of low rank in $\mathsf{Sym}(f_{1/2}; 1^{(m)}, 2^{(n)})$.

The Fano plane

The construction by Hadamard designs is no longer helpful here. In fact, the following argument due to Gordon Royle [7] shows that the only symmetric 2-(v, k, λ) designs that give matrices of low rank in $\mathsf{Sym}(f_{1/2}; 1^{(n)}, 2^{(n)})$ are those arising from the Fano plane and its complement. By Theorem 1.1, we require $k - \lambda = \left(\mu(f_{1/2}; 1, 2) \right)^2 = 2$. Any 2-$(v, k, \lambda)$ design Δ satisfies $(v - 1)\lambda = k(k - 1)$, so we must have $(v - 1)\lambda = (\lambda + 2)(\lambda + 1)$. That is, $v = \lambda + 4 + \frac{2}{\lambda}$. Since $v, \lambda \in \mathbb{N}$, we must have $\lambda \mid 2$.

1. If $\lambda = 1$, then $v = 7$ and $k = 3$. There is a unique design with these parameters, which is the Fano plane, and its incidence graph is the Heawood graph.

2. If $\lambda = 2$, then $v = 7$, and $k = 4$. Again, there exists a unique design with these parameters, which is the design whose blocks are the complements of the blocks of the Fano plane, and its incidence graph is the bipartite complement of the Heawood graph.

The above constructions give matrices $M_n \in \mathsf{Sym}(f_{1/2}; 1^{(7n)}, 2^{(7n)})$ of rank at most $\frac{4}{7}n$, which improves upon the upper bound $\frac{2}{3}n$ given by the $\mathcal{F}_{\text{sunflower}}$ and $\mathcal{F}_{\text{Hadamard}}$ families. Furthermore, for $n = 1$, consider the family $\mathcal{F}_{\text{Fano}}$ over $[8]$:

$$\mathcal{F}_{\text{Fano}} = \mathcal{F}_{\text{sunflower}} \cup \{1357, 1368, 1458, 1467\}$$
$$= \{12, 13, 14, 15, 16, 17, 18, 1234, 1256, 1278, 1357, 1368, 1458, 1467\}.$$

$\mathcal{F}_{\text{Fano}}$ is a bisection-closed family of size 14 over $[8]$, and it is the symmetric 2-$(7, 3, 1)$ design, which induces the Heawood graph as its incidence graph G_{Fano}.

Modifications to $\mathcal{F}_{\text{sunflower}}$ as above using the sets 1357, 1368, 1458, and 1467 can produce bisection-closed families over $[n]$ of size more than $\lfloor 3n/2 \rfloor - 2$ up to $n \leq 15$. Also, it does not appear that there are any bisection-closed families arising from M_n for $n > 1$. As such, the original conjecture that any bisection-closed family over $[n]$ has size at most $\lfloor 3n/2 \rfloor - 2$ remains open for large n.

9

References

[1] L. Babai, and P. Frankl, *Linear Algebra Methods in Combinatorics*, Univ. of Chicago, Dept. Computer Sci., Oct. 2022, unpublished, preliminary version 2.2, available at
https://people.cs.uchicago.edu/~laci/babai-frankl-book2022.pdf

[2] N. Balachandran, S. Bhattacharya, K. V. Kher, R. Mathew, and B. Sankarnarayanan, *On hierarchically closed fractional intersecting families*, Electron. J. Combin. **30** (Dec. 2023), no. 4, #P4.37, doi:10.37236/11651. MR4672585, Zbl 07806264

[3] N. Balachandran, S. Bhattacharya, and B. Sankarnarayanan, *An ensemble of high rank matrices arising from tournaments*, Linear Algebra Appl. **658** (Feb. 2023), 310–318, doi:10.1016/j.laa.2022.11.004, addendum available at arXiv:2108.10871v3 [math.CO]. MR4513895, Zbl 1506.15034

[4] N. Balachandran, S. Bhattacharya, and B. Sankarnarayanan, *Almost full rank matrices arising from transitive tournaments*, Linear Multilinear Algebra **72** (2024), no. 2, 153–161, doi:10.1080/03081087.2022.2158168. MR4703565, Zbl 07808500

[5] N. Balachandran, R. Mathew, and T. K. Mishra, *Fractional L-intersecting families*, Electron J. Combin. **26** (June 2019), no. 2, #P2.40, doi:10.37236/7846. MR3982269, Zbl 1416.05275

[6] C. J. Colbourn and J. H. Dinitz (eds.), *CRC Handbook of Combinatorial Designs*, 2nd ed., Discrete Math. Appl., Chapman & Hall/CRC, Boca Raton, 2007. Zbl 1101.05001

[7] Gordon Royle (userID:1492), *Is there a bipartite graph with $\sqrt{2}$ as an eigenvalue with high multiplicity, specifically more than in the Heawood graph?*, MathOverflow answer (ID:456725), version:2023-10-19 (revision 1).

[8] R. Mathew, T. K. Mishra, R. Ray, and S. Srivastava, *Modular and fractional L-intersecting families of vector spaces*, Electron J. Combin. **29** (Mar. 2022), no. 1, #P1.45, doi:10.37236/10358. MR4395249, Zbl 1486.05302

[9] T. K. Mishra, *On size of a r-wise fractional L-intersecting family*, preprint, available at arXiv:1909.13217v3 [cs.DM].

[10] P. Rowlinson, *Eigenvalue multiplicity in triangle-free graphs*, Linear Algebra Appl. **493** (Mar. 2016), 484–493, doi:10.1016/j.laa.2015.12.012. MR3452752, Zbl 1329.05199

[11] B. Sankarnarayanan, *Some problems in combinatorics: Excursions in graph colorings and extremal set theory*, Ph.D. Thesis, Department of Mathematics, Indian Institute of Technology Bombay, 2024.

[12] T. Tao and V. Vu, *Random matrices have simple spectrum*, Combinatorica **37** (June 2017), no. 3, 539–553, doi:10.1007/s00493-016-3363-4. MR3666791, Zbl 1399.60008

[13] J. Williamson, *Hadamard's determinant theorem and the sum of four squares*, Duke Math. J. **11** (1944), 65–82, doi:10.1215/S0012-7094-44-01108-7. MR0009590, Zbl 0060.03202