An ensemble of high rank matrices arising from tournaments*

Niranjan Balachandran† Srimanta Bhattacharya‡

Brahadeesh Sankarnarayanan§

November 11, 2022

Abstract

Suppose $\mathbb{F}$ is a field and let $\mathbf{a} \coloneqq (a_1, a_2, \dots)$ be a sequence of non-zero elements in $\mathbb{F}$. For $\mathbf{a}_n \coloneqq (a_1, \dots, a_n)$, we consider the family $\mathcal{M}_n(\mathbf{a})$ of $n \times n$ symmetric matrices M over $\mathbb{F}$ with all diagonal entries zero and the (i, j)th element of M either a_i or a_j for $i < j$. In this short paper, we show that all matrices in a certain subclass of $\mathcal{M}_n(\mathbf{a})$—which can be naturally associated with transitive tournaments—have rank at least $\lfloor 2n/3 \rfloor - 1$. We also show that if $\mathrm{char}(\mathbb{F}) \neq 2$ and M is a matrix chosen uniformly at random from $\mathcal{M}_n(\mathbf{a})$, then with high probability $\mathrm{rank}(M) \geq \big(\frac{1}{2} - o(1)\big)n$.

2020 Mathematics Subject Classification: 15B52, 05D99, 05C20, 15A03.

Key words and phrases: rank, symmetric matrix, tournament, Talagrand's inequality, bisection closed family

1 Introduction

By $[n]$ we shall mean the set $\{1, \dots, n\}$. Suppose $\mathbb{F}$ is a field and suppose $\mathbf{a} \coloneqq (a_1, a_2, \dots)$ is a sequence of non-zero elements in $\mathbb{F}$. Write $\mathbf{a}_n \coloneqq (a_1, \dots, a_n)$. This paper concerns itself with the following

†Department of Mathematics, Indian Institute of Technology Bombay, Mumbai, India. Email: niranj@math.iitb.ac.in. ORCiD: https://orcid.org/0000-0002-4142-3857

‡Department of Computer Science and Enginnering, Indian Institute of Technology Palakkad, Palakkad, India. Email: srimanta@iitpkd.ac.in. ORCiD: https://orcid.org/0000-0002-2612-0217. This work was done while the author was at the Indian Statistical Institute, Kolkata, India.

§Corresponding author. Department of Mathematics, Indian Institute of Technology Bombay, Mumbai, India. Email: bs@math.iitb.ac.in. ORCiD: https://orcid.org/0000-0001-9191-1253. The author's research is supported by the National Board for Higher Mathematics (NBHM), Department of Atomic Energy (DAE), Govt. of India.

Problem 1. *Let $\mathcal{M}_n(\mathbf{a})$ consist of the family of all symmetric $n \times n$ matrices over $\mathbb{F}$ with all diagonal entries being zero and such that for $1 \le i < j \le n$ the (i,j)th entry is either a_i or a_j. Determine $\min_{M \in \mathcal{M}_n(\mathbf{a})} \operatorname{rank}(M)$. Here, $\operatorname{rank}(M)$ denotes the rank of M over the field $\mathbb{F}$.*

This problem first appeared in [3], though it was only stated for matrices over the reals. There it is asked whether one can find an absolute constant $c > 0$ such that $\operatorname{rank}(M) \ge cn$ for all $M \in \mathcal{M}_n(\mathbf{a})$.

The problem of determining the rank of specific matrices has been of immense interest in extremal combinatorics with applications in theoretical computer science as well—see [1, 4, 5, 7, 8, 10, 11, 13]. The question in [3] is motivated by a problem in extremal combinatorics concerning what are called *bisection closed families*: a family $\mathcal{F}$ of subsets of $[n]$ is called a bisection closed family if, for any distinct $A, B \in \mathcal{F}$, either $\frac{|A \cap B|}{|A|} = \frac{1}{2}$ or $\frac{|A \cap B|}{|B|} = \frac{1}{2}$, and one seeks to find the maximum size of a bisection closed family over $[n]$. One of the results that appears in [3] shows that any bisection closed family has size $O(n \log_2 n)$, while there are bisection closed families of size $\Omega(n)$. If the answer to the question in [3] is affirmative, then in fact it is not hard to see that any bisection closed family has size $O(n)$, and we include that simple argument here for the sake of completeness. Suppose $\mathcal{F}$ is a bisection closed family over $[n]$ of size m and let $X_{m \times n}$ be the matrix whose rows are indexed by the members of $\mathcal{F}$ and the columns by the elements of $[n]$ defined as follows: for $A \in \mathcal{F}$ and $x \in [n]$, set $X(A, x) \coloneqq 1$ if $x \in A$ and $X(A, x) \coloneqq -1$ otherwise. For two sets $A, B \in \mathcal{F}$, let $\operatorname{Tor}(A, B) \coloneqq A$ if $|A \cap B| = \frac{1}{2}|B|$ and $\operatorname{Tor}(A, B) \coloneqq B$ otherwise. Then the matrix XX^T whose rows and columns are indexed by the members of $\mathcal{F}$ satisfies

$$\begin{aligned} XX^T(A, A) &= n, \\ XX^T(A, B) &= n - 2(|A| + |B|) + 4|A \cap B| \\ &= n - 2|\operatorname{Tor}(A, B)| \end{aligned}$$

In particular, if $\mathcal{F} = \{A_1, \ldots, A_m\}$ and we let J denote the $m \times m$ matrix consisting entirely of ones, then $\frac{1}{2}(nJ - XX^T) \in \mathcal{M}_n(\mathbf{a})$ for the sequence $\mathbf{a}_n = (|A_1|, \ldots, |A_m|)$. Hence, if the conjecture holds, then $\operatorname{rank}(XX^T) \ge cm$. But, since $\operatorname{rank}(XX^T) \le \operatorname{rank}(X) \le n$, it follows that $m \le (n+1)/c$, which establishes an asymptotically tight bound on the size of the bisection closed family.

In this short paper, we make some steps towards settling this problem in the affirmative. In order to describe our results, we note that to each $M \in \mathcal{M}_n(\mathbf{a})$ there corresponds a tournament on the vertex set $[n]$ in the following natural manner: for $i < j$ we direct the edge ij as $i \to j$ if $M(i,j) = a_i$, and the edge is directed in the reverse direction if $M(i,j) = a_j$. Conversely, for a tournament T on $[n]$, we can associate the matrix $M_T(\mathbf{a}) \in \mathcal{M}_n(\mathbf{a})$ in exactly the same way, namely, for $i < j$, set $M_T(i,j) = a_i$ iff $i \to j$. Note that this correspondence is not necessarily one-to-one, since the a_i need not be distinct.

Our first result gives a lower bound in the case where the underlying tournament is transitive, i.e., when there is a total order $\prec$ on $[n]$ such that $i \to j$ whenever $i \prec j$.

Theorem 2. *If T is transitive, then* $\operatorname{rank}(M_T(\mathbf{a})) \geq \lfloor \frac{2n}{3} \rfloor - 1$.

In [3], there are constructions of bisection closed families $\mathcal{F}$ of size $\frac{3}{2}n - 2$ which admit a uniform subfamily of size $n-1$. So, by the remarks above, the corresponding matrix M has a principal submatrix of rank at least $2n/3 + 1$, so the constant $2/3$ in the theorem is best possible over *all* tournaments. It is instructive to compare this with a result of de Caen's [6] on the rank of tournament matrices (where the entries are only 0 and 1). In [6], among other things, it is shown that the rank of any $n \times n$ tournament matrix is at least $\frac{n-1}{2}$ over any field and at least $n-1$ over the reals. Our result is in a similar spirit, since it answers a question in a more general setup, but is also in contrast with de Caen's since we consider symmetric matrices.

Our second result shows that almost all the matrices in $\mathcal{M}_n(\mathbf{a})$ have high rank. More precisely, we show that for a *random tournament*—a tournament with the edges being directed in either direction with probability 1/2 each and independently—then *with high probability (whp)* the rank is at least $(1/2 - o(1))n$. Here, the phrase "with high probability" means that the probability that the said event occurs asymptotically tends to 1 as $n \to \infty$.

Theorem 3. *Suppose* $\operatorname{char}(\mathbb{F}) \neq 2$ *and* $\mathbf{a}$ *is a sequence of non-zero elements of* $\mathbb{F}$. *If T is a uniformly random tournament, then* whp $\operatorname{rank}(M_T(\mathbf{a})) \geq \frac{n}{2} - 21\sqrt{n \log n}$.

To give some perspective on this result vis-à-vis the existing literature on similar problems, the behavior of random symmetric matrices is an immensely active area of research and there are several papers that consider various random models (see [9, 14] and the references therein) and Theorem 3 may be regarded as another addition to that list, though there is a fundamental difference between our result and all the others. For one, as we have pointed out earlier, the matrix that arises from a bisection closed family has rank at most $2n/3 + O(1)$, so in that sense our result is somewhat qualitatively different from those that appear in several of those papers. It must be pointed out that the main result in [14] considers random symmetric matrices $M_n = ((\xi_{ij}))$ 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, and their result shows that *whp* the spectrum is simple. This does establish (in a strong form) Theorem 3 over the reals, *in the special case where a_i are all pairwise distinct.* But otherwise the best bound this suggests is of the order $\Omega(\sqrt{n})$. Secondly, our result holds over all fields $\mathbb{F}$ with $\operatorname{char}(\mathbb{F}) \neq 2$ whereas most other results usually work specifically with $\mathbb{R}$ or $\mathbb{C}$ (though they have stronger results). To also contrast the results of Theorems 2 and 3, note that Theorem 2 holds over all fields whereas for Theorem 3 we need $\operatorname{char}(\mathbb{F}) \neq 2$.

We prove Theorems 2 and 3 in the next section. The final section includes some concluding remarks and poses some further questions.

2 Proofs of Theorems 2 and 3

Proof of Theorem 2. It suffices to prove that whenever 3 divides n we have $\operatorname{rank}(M_T(\mathbf{a})) \geq 2n/3$, since we may then interpolate to those n such that 3 does not divide n to show that $\operatorname{rank}(M_T(\mathbf{a})) \geq \lfloor 2n/3 \rfloor - 1$ for all $n \geq 3$. We shall prove this by induction on $n \geq 3$ such that 3 divides n. Without loss of generality we may assume that the ordering of the elements coincides with the natural ranking order on $[n]$, i.e., $i \prec j$ iff $i > j$. Also, we denote the matrix corresponding to the transitive tournament on $[n]$ by $D_n(a_1, \ldots, a_n)$. When the a_i are clear from the context, we will simply call this matrix D_n. Then, for the base case, i.e., for $n = 3$, it is easy to see that

$$D_3(a_1, a_2, a_3) = \begin{pmatrix} 0 & a_2 & a_3 \\ a_2 & 0 & a_3 \\ a_3 & a_3 & 0 \end{pmatrix}$$

has rank at least $2 = \frac{2 \cdot 3}{3}$ for any non-zero values of a_2 and a_3. Next, we assume the assertion to be true for n and prove it for $n + 3$. Then, up to relabeling of the indices, we can write $D_{n+3}(a_1, \ldots, a_{n+3})$ as

$$D_{n+3} = \begin{pmatrix} D_3 & B \\ B^T & D_n \end{pmatrix}, \tag{1}$$

where $D_3 = D_3(a_1, a_2, a_3)$, $D_n = D_n(a_4, \ldots, a_{n+3})$, and B is the matrix

$$\begin{pmatrix} a_4 & a_5 & \cdots & a_{n+3} \\ a_4 & a_5 & \cdots & a_{n+3} \\ a_4 & a_5 & \cdots & a_{n+3} \end{pmatrix}$$

and by the induction hypothesis D_n has rank $2n/3$. Now, assume that D_{n+3} has rank $\leq \frac{2n}{3} + 1$. Let $\mathcal{L}$ be a basis of D_{n+3}. So, by our assumption $|\mathcal{L}| \leq \frac{2n}{3} + 1$. Next, we will arrive at a contradiction by considering the following cases.

For ease of presentation, let us write $D_{n+3} = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$, $D_3 = \begin{pmatrix} \mathbf{a}'_1 & \mathbf{a}'_2 & \mathbf{a}'_3 \end{pmatrix}$, and $B^T = \begin{pmatrix} \mathbf{a}''_1 & \mathbf{a}''_2 & \mathbf{a}''_3 \end{pmatrix}$. Let $\mathcal{D} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

Case 1: $|\mathcal{L} \cap \mathcal{D}| = 0$. In this case, $\mathcal{L} \subseteq \{\mathbf{a}_4, \ldots, \mathbf{a}_n\}$. But, this is not possible since the column space of B, which is of rank one, cannot include the column space of D_3, which is of rank three. In fact, no column of D_3 belongs to the column space of matrix B. More concretely, for $\mathbf{a}_i \in \mathcal{D}, 1 \leq i \leq 3$, let $\mathbf{a}_i = \sum \gamma_j \mathbf{a}_j$, where $\gamma_j \in \mathbb{F}$ and $\mathbf{a}_j \in \mathcal{L}$ with $j \in \{4, \ldots, n+3\}$. Then, it follows that $\mathbf{a}'_i = \sum \gamma_j \mathbf{a}'_j$, where $\mathbf{a}'_j = \begin{pmatrix} a_j & a_j & a_j \end{pmatrix}^T$. But, this is a contradiction since $0 = \sum_j \gamma_j a_j = a_3 \neq 0$.

Case 2: $|\mathcal{L} \cap \mathcal{D}| = 1$. This is similar to Case 1. More precisely and without loss of generality, let $\mathcal{L}$ contain $\mathbf{a}_3$. Now, let $\mathbf{a}_2 = \gamma_3 \mathbf{a}_3 + \sum_j \gamma_j \mathbf{a}_j$, where $\gamma_j \in \mathbb{F}$ and $\mathbf{a}_j \in \mathcal{L}$ with $j \in \{4, \ldots, n+3\}$. This implies $\mathbf{a}'_2 = \gamma_3 \mathbf{a}'_3 + \sum_j \gamma_j \mathbf{a}'_j$. Then, following the considerations of Claim 1 we have $\gamma_3 \neq 0$. But then we have that $0 = \gamma_3 a_3 + \sum_j \gamma_j a_j = a_2 \neq 0$ for $j \in \{4, \ldots, n+3\}$, which is a contradiction. Similar arguments hold for the cases when $\mathbf{a}_1 \in \mathcal{L}$ or $\mathbf{a}_2 \in \mathcal{L}$.

Case 3: $|\mathcal{L} \cap \mathcal{D}| \in \{2, 3\}$ for any choice of $\mathcal{L}$. This implies that $\operatorname{rank}\left(\begin{pmatrix} B^T & D_n \end{pmatrix}^T\right) \leq \frac{2n}{3} - 1$, which is a contradiction since $\operatorname{rank}(D_n)$ is already at least $2n/3$ by our assumption. □

Before we get to the proof of Theorem 3, we state a few results that we shall use. We state the versions as they appear in [12].

Theorem 4 (Chernoff bound). *If X is distributed as the binomial random variable $B(n, p)$, then for any $0 \leq t \leq np$*

$$\mathbb{P}(|X - np| > t) < 2 \exp\left(-\frac{t^2}{3np}\right).$$

The other main technical tool is Talagrand's inequality. By a *trial* we shall simply mean a Bernoulli event.

Theorem 5 (Talagrand's inequality). *Suppose X is a non-negative random variable, not identically zero, which is determined by n independent trials $T_1, \ldots, T_n$, and satisfying the following for some $c, r > 0$:*

1. *(c-**Lipschitz**) Changing the outcome of any one trial T_i changes X by at most c,*
2. *(r-**certifiability**) For any $s \geq 0$, if $X \geq s$ then there is a set of at most rs trials whose outcomes certify that $X \geq s$, i.e., there is a set $I \subset [n]$ of size at most rs and a set of outcomes of the trials T_i for $i \in I$ such that fixing the outcomes of T_i for $i \in I$ ensures $X \geq s$* irrespective of the outcomes of *T_j for $j \notin I$.*

If $\operatorname{Med}(X)$ denotes the median of X and $0 \leq t \leq \operatorname{Med}(X)$, we have

$$\mathbb{P}(|X - \operatorname{Med}(X)| > t) \leq 4 \exp\left(-\frac{t^2}{8c^2 r \operatorname{Med}(X)}\right).$$

Proof of Theorem 3. We begin with a couple of simple observations.

1. Fix a pair (i, j) with $i < j$, and let T be a tournament on $[n]$. If T' is the tournament obtained from T by changing the orientation of only the edge ij, then $M_{T'}(\mathbf{a}) = M_T(\mathbf{a}) + D$ for a matrix D comprising of zeros everywhere except at the (i, j) and (j, i) positions. Consequently, $|\operatorname{rank}(M_T(\mathbf{a})) - \operatorname{rank}(M_{T'}(\mathbf{a}))| \leq 2$. Also, observe that if $\mathbf{a}'$ is the sequence with a_i replaced by some $z \in \mathbb{F}$, then for any tournament T,

$M_T(\mathbf{a})$ differs from $M_T(\mathbf{a}')$ only in the entries of the ith row and column, so again in particular, $|\text{rank}(M_T(\mathbf{a})) - \text{rank}(M_T(\mathbf{a}'))| \leq 2$.

2. For a tournament T, let T_R denote the *reverse* tournament, i.e., if $i \to j$ in T then $j \to i$ in T_R. Then $M_T(\mathbf{a}) + M_{T_R}(\mathbf{a}) = M$ where $M(i,i) = 0$ and $M(i,j) = a_i + a_j$. In particular, $M = DJ + JD - 2D$ where D is the diagonal matrix $\text{diag}(a_1, \ldots, a_n)$ and J represents, as before, the all-ones matrix. In particular, since $a_i \neq 0$ and $\text{char}(\mathbb{F}) \neq 2$, it follows that $\text{rank}(M) \geq n-2$. Consequently, at least one of $\text{rank}(M_T)$ and $\text{rank}(M_{T_R})$ is at least $n/2 - 1$.

First, suppose that $\text{char}(\mathbb{F})$ does not divide $n-1$. A uniformly random tournament T is completely determined by the trials T_e for the pairs $e = (i,j)$ with $i < j$ with T_e distributed as $\text{Ber}(1/2)$. Fix some $z \neq 0$ in $\mathbb{F}$ and consider the more general ensemble $\mathcal{M}_n(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)$ is the random sequence with $x_i = z$ with probability $1/\sqrt{n}$ and $x_i = a_i$ with probability $1 - \frac{1}{\sqrt{n}}$. (We can view $\mathbf{x}$ as arising from n flips of a biased coin where heads occurs with probability $1/\sqrt{n}$, and $x_i = z$ if a head occurs on the ith toss and $x_i = a_i$ if a tail occurs on the ith toss.) Let $X = \text{rank}(M_T(\mathbf{x}))$. By the first observation, it follows that X is 2-Lipschitz. Also, for any $s \geq 1$, if we fix $x_1 = \cdots = x_s = x_{s+1} = z$, then, irrespective of T, either the principal $s \times s$ matrix or the principal $(s+1) \times (s+1)$ submatrix of $M_T(\mathbf{x})$ has non-zero determinant in $\mathbb{F}$, so this establishes that X is 2-certifiable. Note that if T is a transitive tournament, then so is T_R, so it follows by Theorem 2 and the second observation above that $\text{Med}(X) \geq \frac{n-2}{2}$. Hence, by Talagrand's inequality,

$$\mathbb{P}\left(X > \frac{n}{2} - 16\sqrt{n \log n}\right) < \frac{4}{n^4}.$$

Now, let Y denote the number of elements of $\mathbf{x}$ that are equal to z (that is, the number of occurrences of heads in the sequence of coin flips). By the Chernoff bound, it follows that $\mathbb{P}(Y > 2\sqrt{n}) < 2e^{-\sqrt{n}/3}$, so with probability at least $1 - \frac{4}{n^4} - 2e^{-\sqrt{n}/3}$ both the events, $X > n/2 - 16\sqrt{n \log n}$ and $Y < 2\sqrt{n}$ hold simultaneously. Hence, if we only consider the submatrix M' of $M_T(\mathbf{x})$ indexed by those rows and columns of the sequence $\mathbf{x}$ that do not comprise of any z, then $\text{rank}(M') \geq n/2 - 16\sqrt{n \log n} - 4\sqrt{n}$. Now finally, if $\text{char}(\mathbb{F})$ divides $n-1$, then we restrict ourselves to the principal submatrix of M_T of order $n-1$, and the same argument as above gives us $\text{rank}(M_T(\mathbf{a})) \geq n/2 - 21\sqrt{n \log n}$ *whp* with room to spare. □

3 Concluding remarks

- In the statement of Theorem 3, the randomness is over the orientations of the tournament edges, and *not* over the elements a_i since they come from the given sequence $\mathbf{a}$. However, the proof uses a randomization of the sequence in order to be able to use Talagrand's inequality, and it does not seem straightforward to stay within the confines of the given sequence to be able to prove the same statement.

- Our bound of $n/2$ is constricted by our estimate of $\mathrm{Med}(\mathrm{rank}(M_T(\mathbf{a}))$. If one can get a better bound, then the same proof gives a better rank bound as well. Interestingly, this is an instance where using Talagrand's inequality for concentration around the median gives a decidedly better bound than the analogous version for concentration around the mean.

- Our error probability of $O(n^{-3})$ is easily improved to $O\big(e^{-n^{1/3}}\big)$ if we take $t = n^{2/3}$ for instance in the proof of Theorem 3.

- As remarked after the statement of Theorem 2, while the lower bound in Theorem 2 does achieve the bound that is best possible for all tournaments, we believe that the bound must be substantially better when restricted to transitive tournaments. In fact, we believe that for transitive tournaments, $\mathrm{rank}(M_T(\mathbf{a})) \geq n - o(n)$ must hold as well though we are unable to prove this even over the reals.[1]

- As indicated in the remarks in the introduction, it must be possible to improve upon the results obtained here when we restrict ourselves to the fields $\mathbb{R}$ or $\mathbb{C}$, or if the elements a_i themselves satisfy other constraints. For instance, when the sequence $\mathbf{a}$ is the constant sequence $(a, a, \dots)$, $\mathrm{rank}(M) \geq n-1$ for all $M \in \mathcal{M}_n(\mathbf{a})$ and over any field. This also shows that if $\mathbb{F}$ is any finite field, then $\mathrm{rank}(M) \geq \frac{n}{|\mathbb{F}|-1} - 1$ for all $M \in \mathcal{M}_n(\mathbf{a})$ for *any* sequence $\mathbf{a}$ in $\mathbb{F}$. On the other hand, it is not even clear whether the results of this paper extend to the case when $\mathrm{char}(\mathbb{F}) = 2$, or when infinitely many of the a_i are distinct.

- A more general setup is the following: Suppose $f\colon \mathbb{F}^2 \to \mathbb{F}$ and let $\mathbf{a} = (a_1, a_2, \dots)$ as before and consider the central problem of this paper over the more general family $\mathcal{M}_n^{(f)}(\mathbf{a})$ which is defined as follows. For any tournament T on the vertex set $[n]$ the matrix $M_T^{(f)}(\mathbf{a}) \in \mathcal{M}_n^{(f)}(\mathbf{a})$ consists of zeros on the diagonal, and for $i < j$ the (i,j) entry of $M_T^{(f)}(\mathbf{a})$ equals $f(a_i, a_j)$ if $i \to j$ in T, and equals $f(a_j, a_i)$ otherwise. Our proof of Theorem 3 is easily modified to show that for a random tournament T, $\mathrm{rank}\big(M_T^{(f)}(\mathbf{a})\big) \geq \big(\frac{1}{2} - o(1)\big)n$ for any function $f(x,y) = \alpha x + \beta y$ with the property $\alpha + \beta \neq 0$. The only difference in the proof is that we use Talagrand's inequality with concentration about the mean instead of the median. If $f(x,y) = \alpha x + \beta y$ with $\alpha + \beta \neq 0$, then the same argument as in the proof of Theorem 3 also shows that $\mathrm{rank}\big(M_T^{(f)}(\mathbf{a})\big) + \mathrm{rank}\big(M_{T_R}^{(f)}(\mathbf{a})\big) \geq n - 2$, so again for a random tournament T the expected rank of $M_T^{(f)}(\mathbf{a})$ is at least $n/2 - 1$. We omit the details.[2]

  This more general ensemble of matrices may pose yet more interesting difficulties even for relatively simple functions. For instance, even for $f(x,y) = \frac{x}{y}$, the problem

[1]This has recently been settled in the affirmative in a strong form; see [2].

[2]Note that resolving the problem for the family $\mathcal{M}_n^{(f)}(\mathbf{a})$ also gives a linear upper bound on the size of any θ-intersecting family (cf. [3]), in a similar manner.

is already quite non-trivial.

- For a given sequence $\mathbf{a}$ in $\mathbb{F}$ and a matrix $M = M_T(\mathbf{a}_n) \in \mathcal{M}_n(\mathbf{a})$ arising from a self-dual tournament T, the matrix $M_{T_R}(\mathbf{a}_n)$ arising from the reverse tournament T_R can also be viewed as $M_T(\sigma\mathbf{a}_n)$ for a permutation σ of $\mathbf{a}_n = (a_1, \ldots, a_n)$. We have shown that at least one of M_T and M_{T_R} has rank at least $n/2 - 1$ for any tournament T on $[n]$. An interesting question is whether, for a fixed tournament T, the matrices $M_T(\sigma\mathbf{a}_n)$ have the same rank for all permutations σ of $\mathbf{a}_n$. A positive answer to this question will tell us, in particular, that matrices arising from self-dual tournaments (such as Paley tournaments) all have high rank.

References

[1] N. Alon, *Approximating sparse binary matrices in the cut-norm*, Linear Algebra Appl. **486** (2015), 409–418, DOI 10.1016/j.laa.2015.08.024. MR 3401770, Zbl 1327.15044 ↑2

[2] N. Balachandran, S. Bhattacharya, and B. Sankarnarayanan, *Almost full rank matrices arising from transitive tournaments*, Linear Multilinear Algebra (2023), DOI 10.1080/03081087.2022.2158168. ↑7

[3] N. Balachandran, R. Mathew, and T. K. Mishra, *Fractional L-intersecting families*, Electron. J. Combin. **26** (2019), no. 2, article no. P2.40, 12 pp., DOI 10.37236/7846, available at arXiv:1803.03954 [math.CO]. MR 3982269, Zbl 1416.05275 ↑2, 3, 7

[4] B. Barak, Z. Dvir, A. Wigderson, and A. Yehudayoff, *Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes*, Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing: San Jose, California, June, 6–8, 2011 (S. Vadhan, ed.), Proc. Annual ACM STOC, ACM, N. Y., 2011, pp. 519–528, DOI 10.1145/1993636.1993705, available at arXiv:1009.4375 [math.CO]. Extended abstract. Full version available at https://eccc.weizmann.ac.il/report/2010/149/. MR 2932002, Zbl 1288.05153 ↑2

[5] B. Bukh, *Ranks of matrices with few distinct entries*, Israel J. Math. **222** (2017), no. 1, 165–200, DOI 10.1007/s11856-017-1586-8, available at arXiv:1508.00145 [math.AC]. MR 3736503, Zbl 1380.15005 ↑2

[6] D. de Caen, *The ranks of tournament matrices*, Amer. Math. Monthly **98** (1991), no. 9, 829–831, DOI 10.2307/2324270. MR 1132999, Zbl 0749.05035 ↑3

[7] A. Coja-Oghlan, A. A. Ergür, P. Gao, S. Hetterich, and M. Rolvien, *The rank of sparse random matrices*, Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms: Salt Lake City, UT, January 5–8, 2020 (S. Chawla, ed.), Proc. Annual ACM-SIAM SODA, ACM, N. Y., 2020, pp. 579–591, DOI 10.1137/1.9781611975994.35, available at arXiv:1906.05757 [math.CO]. MR 4141217, Zbl 07304058 ↑2

[8] S. M. Fallat and L. Hogben, *The minimum rank of symmetric matrices described by a graph: A survey*, Linear Algebra Appl. **426** (2007), no. 2–3, 558–582, DOI 10.1016/j.laa.2007.05.036. MR 2350678, Zbl 1122.05057 ↑2

[9] A. Ferber and V. Jain, *Singularity of random symmetric matrices: A combinatorial approach to improved bounds*, Forum Math. Sigma **7** (2019), article no. e22, 29 pp., DOI 10.1017/fms.2019.21, available at arXiv:1809.04718 [math.PR]. MR 3993806, Zbl 1423.60016 ↑3

[10] X. Hu, C. R. Johnson, C. E. Davis, and Y. Zhang, *Ranks of permutative matrices*, Spec. Matrices **4** (2016), no. 1, 233–246, DOI 10.1515/spma-2016-0022. MR 3507976, Zbl 1338.05030 ↑2

[11] N. Kishore Kumar and J. Schneider, *Literature survey on low rank approximation of matrices*, Linear Multilinear Algebra **65** (2017), no. 11, 2212–2244, DOI 10.1080/03081087.2016.1267104, available at arXiv:1606.06511 [math.NA]. MR 3740692, Zbl 1387.65039 ↑2

[12] M. Molloy and B. Reed, *Graph Colouring and the Probabilistic Method*, with 19 figures, Algorithms Combin., vol. 23, Springer, Berl., 2002, DOI 10.1007/978-3-642-04016-0. MR 1869439, Zbl 0987.05002 ↑5

[13] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev. **52** (2010), no. 3, 471–501, DOI 10.1137/070697835, available at arXiv:0706.4138 [math.OC]. MR 2680543, Zbl 1198.90321 ↑2

[14] T. Tao and V. Vu, *Random matrices have simple spectrum*, Combinatorica **37** (2017), no. 3, 539–553, DOI 10.1007/s00493-016-3363-4, available at arXiv:1412.1438 [math.PR]. MR 3666791, Zbl 1399.60008 ↑3

ADDENDUM

We fix a small error in the proof of Theorem 3. The statement of the theorem remains unchanged.

The proof of Theorem 3

Our application of Talagrand's inequality (Theorem 5) in the proof of Theorem 3 is incorrect: it is true that if we fix $x_1 = \cdots = x_s = x_{s+1} = z$ gives $X \geq s$, but this is not sufficient to establish that X is 2-certifiable. Instead, we need a version of McDiarmid's inequality for concentration bounds on product measure spaces, and we use the one stated in [A2, Lemma 1.2]:

Theorem 6 (Independent Bounded Differences Inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables, with X_k taking values in a set Ω_k for each k. Suppose that the measurable function $f\colon \prod_k \Omega_k \to \mathbb{R}$ satisfies, for each k,*

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq c_k$$

whenever the vectors $\mathbf{x}$ and $\mathbf{x}'$ differ only in the kth coordinate. Let Y be the random variable $f(X_1, \dots, X_n)$. Then, for any $t > 0$,

$$\mathbb{P}(|Y - \mathbf{E}(Y)| > t) \leq 2\exp\bigl(-2t^2/\textstyle\sum_k c_k^2\bigr).$$

Proof of Theorem 3. The notation $[m, n]$ denotes the set of integers i such that $m \leq i \leq n$. Let $\Omega_k = \{0,1\}^{[k+1,n]}$ for $k = 1, \dots, n-1$. View each vector $\mathbf{x}_k = (x_k^{(k+1)}, \dots, x_k^{(n)}) \in \Omega_k$ as a win-loss record for player k against the players $k+1, \dots, n$ in that order. Thus, any $(n-1)$-tuple $(\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$, where $\mathbf{x}_k \in \Omega_k$ for each k, determines a unique tournament T on $[n]$, and each tournament on $[n]$ arises from some point in $\prod_k \Omega_k$.

Define $f\colon \prod_k \Omega_k \to \mathbb{R}$ by $f(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}) = \operatorname{rank}(M_T(\mathbf{a}))$, where T is the tournament uniquely determined by $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$. As observed previously, changing the orientation of any edge ij to get a tournament T' changes the rank by at most 2, since the matrix $M_{T'}(\mathbf{a})$ is obtained from $M_T(\mathbf{a})$ by adding a matrix of rank at most 2. In fact, for any fixed i, flipping the orientations of any subcollection of the edges ij, where $j \geq i$, changes the rank by at most 2, since this again corresponds to adding a matrix of rank at most 2 to $M_T(\mathbf{a})$. Hence, $|f(\mathbf{x}) - f(\mathbf{x}')| \leq 2$ whenever $\mathbf{x}$ and $\mathbf{x}'$ differ only in the kth coordinate.

Now, let T be a uniformly random tournament on $[n]$, i.e., one for which the orientation of each edge ij is chosen by a fair coin toss. Then, T gives rise to random variables X_k taking values in Ω_k for each $1 \leq k \leq n-1$, and $X_1, \dots, X_{n-1}$ are independent. Define

Date: July 13, 2023

$Y \coloneqq f(X_1, \ldots, X_{n-1})$. As observed previously, if T_R denotes the reverse tournament of T, then $\operatorname{rank}(M_T(\mathbf{a}) + M_{T_R}(\mathbf{a})) \geq n-2$, so $\mathbf{E}(Y) \geq \frac{n}{2} - 1$.

Now, we apply McDiarmid's inequality to get:

$$\mathbb{P}\left(Y < \frac{n}{2} - 1 - 4\sqrt{n \log n}\right) \leq 2e^{-\frac{32n \log n}{4(n-1)}} < \frac{2}{n^8},$$

which proves the result. $\square$

The above proof is simpler than the original attempt in that we do not randomize the sequence $\mathbf{a}$ in addition to picking the tournament T on $[n]$ at random. Also, the above proof uses concentration around the mean, and not around the median as in Talagrand's inequality. Consequently, our first two comments in Section 3 are no longer relevant.

A question on general ensembles of the form $\mathcal{M}_n^{(f)}(\mathbf{a})$

In Section 3, we defined a more general ensemble $\mathcal{M}_n^{(f)}(\mathbf{a})$ as consisting of those symmetric matrices $M_T^{(f)}(\mathbf{a})$ with zero diagonal such that the (i,j)th entry is either $f(a_i, a_j)$ or $f(a_j, a_i)$ depending on the orientation of the edge ij in the tournament T. We raised the problem of finding a lower bound on the rank of the matrices coming from this general ensemble, and observed that the same methods work for linear functions of the form $f(x,y) = \alpha x + \beta y$ for which $\alpha + \beta \neq 0$, and when $\operatorname{char}(\mathbb{F}) \neq 2$, and that the problem is non-trivial for other (even relatively simple) functions, such as $f(x,y) = x/y$.

As it turns out, this specific example does not illustrate the non-triviality of this question, since the same methods suffice to prove a linear lower bound (*whp*) for functions of finite rank. More precisely, let $\operatorname{char}(\mathbb{F}) \neq 2$, and suppose that $f\colon \mathbb{F}^2 \to \mathbb{F}$ is a function such that $f(x,x) \neq 0$ for all x. Suppose that there exist functions $g_i, h_i\colon \mathbb{F} \to \mathbb{F}$, $1 \leq i \leq k$, such that $f(x,y) = \sum_{i=1}^k g_i(x) h_i(y)$. For $X \in \mathbb{F}^n$, define $G_i(X) \coloneqq (g_1(X), \ldots, g_k(X))$ and $H_i(X) \coloneqq (h_1(X), \ldots, h_k(X))$ for all $1 \leq i \leq k$. Then, for any tournament T on $[n]$ and any sequence $\mathbf{a}$ in $\mathbb{F}^*$, we have $M_T^{(f)}(\mathbf{a}) + M_{T_R}^{(f)}(\mathbf{a}) = \sum_{i=1}^k \big(G_i(\mathbf{a}_n)^T H_i(\mathbf{a}_n) + H_i(\mathbf{a}_n)^T G_i(\mathbf{a}_n)\big) - 2\operatorname{diag}\big(f(a_1,a_1), \ldots, f(a_n,a_n)\big)$. The RHS is the sum of a diagonal matrix—of full rank—and $2k$ matrices of rank one. Hence, at least one of $M_T^{(f)}$ or $M_{T_R}^{(f)}$ has rank at least $(n-2k)/2$. Hence, the proof of Theorem 3 above shows that, even in this case, we have *whp* $\operatorname{rank}\big(M_T^{(f)}(\mathbf{a})\big) \geq \frac{n}{2} - o(n)$ for a uniformly random tournament T on $[n]$.

It is worth noting that the condition $f(x,x) \neq 0$ for all x is crucial. For example, let $\mathbb{F} = \mathbb{Q}$, $f(x,y) = (x-y)^2$ and $\mathbf{a}_n = (1, 2, \ldots, n)$, then $\operatorname{rank}(M) = 3$ for all $M \in \mathcal{M}_n^{(f)}(\mathbf{a})$, $n \geq 3$ ([A1, Theorem 2.4]). Thus, we refine our original question and ask: for what functions $f\colon \mathbb{F}^2 \to \mathbb{F}$ does there exist a constant $c > 0$ such that $\operatorname{rank}(M) \geq cn$ for all $M \in \mathcal{M}_n^{(f)}(\mathbf{a})$?

References

[A1] C. Grood, J. Harmse, L. Hogben, T. J. Hunter, B. Jacob, A. Klimas, and S. McCathern, *Minimum rank with zero diagonal*, Electron. J. Linear Algebra **27** (2014), 458–477, DOI 10.13001/1081-3810.1630. MR 3240026, Zbl 1320.05075 ↑11

[A2] C. McDiarmid, *On the method of bounded differences*, Surveys in Combinatorics, 1989 (J. Siemens, ed.), London Math. Soc. Lecture Not. Ser., vol. 141, Camb. Univ. Press, Camb., 1989, pp. 148–188, DOI 10.1017/CBO9781107359949.008. MR 1036755, Zbl 0712.05012 ↑10