



Constructive Regularization of the Random Matrix Norm

Elizaveta Rebrova¹ 

Received: 7 March 2019 / Revised: 11 June 2019 / Published online: 3 July 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

We study the structure of $n \times n$ random matrices with centered i.i.d. entries having only two finite moments. In the recent joint work with R. Vershynin, we have shown that the operator norm of such matrix A can be reduced to the optimal order $O(\sqrt{n})$ with high probability by zeroing out a small submatrix of A , but did not describe the structure of this “bad” submatrix nor provide a constructive way to find it. In the current paper, we give a very simple description of a small “bad” subset of entries. We show that it is enough to zero out a small fraction of the rows and columns of A with largest L_2 norms to bring the operator norm of A to the almost optimal order $O(\sqrt{n \log \log n})$, under additional assumption that the matrix entries are symmetrically distributed. As a corollary, we also obtain a constructive procedure to find a small submatrix of A that one can zero out to achieve the same norm regularization. The main component of the proof is the development of techniques extending constructive regularization approaches known for the Bernoulli matrices (from the works of Feige and Ofek, and Le, Levina and Vershynin) to the considerably broader class of heavy-tailed random matrices.

Keywords Random matrices · Operator norms · Heavy tails

Mathematics Subject Classification (2010) 60B20 · 15B52

A significant part of the work was done when Elizaveta Rebrova was a PhD student at the University of Michigan.

Work supported in part by Allen Shields Memorial Fellowship.

✉ Elizaveta Rebrova
rebrova@math.ucla.edu

¹ Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095, USA

1 Introduction

1.1 Operator Norm Regularization for Random Matrices

What should we call an *optimal* order of an operator norm of a random $n \times n$ matrix? If we consider a matrix A with independent standard Gaussian entries, then by the classical Bai–Yin law (see, for example, [16])

$$\|A\|/\sqrt{n} \rightarrow 2 \quad \text{almost surely,}$$

as the dimension $n \rightarrow \infty$. Moreover, the $2\sqrt{n}$ asymptotic holds for more general classes of matrices. By [21], if the entries of A have zero mean and bounded fourth moment, then

$$\|A\| = (2 + o(1))\sqrt{n}$$

with high probability. If we are concerned to get an explicit (non-asymptotic) probability estimate for all large enough n , an application of Bernstein’s inequality (see, for example, in [19]) gives

$$\mathbb{P}\{\|A\| \leq t\sqrt{n}\} \geq 1 - e^{-c_0 t^2 n} \quad \text{for } t \geq C_0$$

for the matrices with i.i.d. sub-Gaussian entries. Here, $c_0, C_0 > 0$ are absolute constants. The non-asymptotic extensions to more general distributions are also available, see [2,8,14,18].

Also, note that the order \sqrt{n} is the best we can generally hope for. Indeed, if the entries of A have variance C , then the typical magnitude of the Euclidean norm of a row of A is $\sim \sqrt{n}$, and the operator norm of A cannot be smaller than that. So, it is natural to assume $O(\sqrt{n})$ as the “ideal order” of the operator norm of an $n \times n$ i.i.d. random matrix.

However, if we do not assume that the matrix entries have four finite moments, we do not have ideal order $O(\sqrt{n})$: The weak fourth moment is necessary for the convergence in probability of $\|A\|/\sqrt{n}$ when n grows to infinity (see [15]). Moreover, for the matrices with the entries having two finite moments, an explicit family of examples, constructed in [11], shows that A can have $\|A\| \sim O(n^\alpha)$ for any $\alpha \leq 1$ with substantial probability.

This motivates the following questions: What are the obstructions in the structure of A that make its operator norm too large? Under what conditions and how can we regularize the matrix restoring the optimal $O(\sqrt{n})$ norm with high probability? Clearly, interesting regularization would be the one that does not change A too much, for example, that changes only a small fraction of the entries of A . We call such regularization *local*.

The first question was answered in our previous work with Vershynin [13]. We have shown that one can enforce the norm bound $\|A\| \sim \sqrt{n}$ by modifying the entries in a small submatrix of A if and only if the i.i.d. entries of A have zero moment and finite variance. The proof strategy was to construct a way to regularize $\|\cdot\|_{\infty \rightarrow 2}$ norm of A

and to apply a form of Grothendieck–Pietsch theorem (see [10, Proposition 15.11]) to claim that some additional small correction regularizes the operator norm $\|A\|$. This last step made it impossible to find the submatrix explicitly.

1.2 Main Results

In the current work, we give an (almost optimal) answer to the remaining constructiveness question, namely *when local regularization is possible, how to fix the norm of A by a small change to the optimal order?* The main result of the paper is

Theorem 1 (Constructive regularization) *Let A be a random $n \times n$ matrix with i.i.d. entries A_{ij} having symmetric distribution such that $\mathbb{E}A_{ij}^2 = 1$. Then for any $\varepsilon \in (0, 1/6]$, $r \geq 1$ with probability $1 - n^{0.1-r}$ the following holds: If we replace with zeros at most εn rows and εn columns with largest L_2 -norms (as vectors in \mathbb{R}^n), then the resulting matrix \tilde{A} will have a well-bounded operator norm*

$$\|\tilde{A}\| \leq Cr\sqrt{c_\varepsilon n \cdot \ln \ln n}. \quad (1.1)$$

Here, $c_\varepsilon = (\ln \varepsilon^{-1})/\varepsilon$ and $C > 0$ is a sufficiently large absolute constant.

Remark 1 Typically, all the rows and columns of the matrix \tilde{A} have L_2 -norms bounded by $O(\sqrt{c_\varepsilon n})$. One way to check this is via the non-constructive regularization result proved in [13]. Indeed, with probability $1 - 7 \exp(-\varepsilon n/12)$, removing some $\varepsilon n \times \varepsilon n$ submatrix of A , we get a matrix \bar{A} such that $\|\bar{A}\| \lesssim \sqrt{c_\varepsilon n}$ (see [13, Theorem 1]). It implies that all the rows and columns of \bar{A} have well-bounded L_2 -norms (of order at most $\sqrt{c_\varepsilon n}$). Since all but εn rows and εn columns of A coincide with those of \bar{A} , there can be at most εn rows and columns in A having larger L_2 -norms. Thus, regularization described in the statement of Theorem 1 zeros out them all.

Moreover, the proof Theorem 1 holds without changes if we define \tilde{A} as the result of zeroing out of all rows and columns having L_2 -norm bigger than $C\sqrt{c_\varepsilon n}$. As we just discussed, this is an even more delicate change in the matrix A with high probability.

The regularization procedures discussed above (in Theorem 1 and Remark 1) are local, as they change only a small fraction of the matrix entries. However, they still change more than $\varepsilon n \times \varepsilon n$ submatrix as promised by [13, Theorem 1.1]. As a corollary of Theorem 1, we also obtain a polynomial algorithm that regularizes the norm of A with high probability by zeroing out its *small submatrix*.

This algorithm addresses separately subsets of matrix entries having similar magnitude. We define these subsets via order statistics of i.i.d. samples A_{ij} : Let $\hat{A}_1, \dots, \hat{A}_{n^2}$ be the non-increasing rearrangement of the entries A_{ij} (in sense of absolute values, namely $|\hat{A}_1| \geq \dots \geq |\hat{A}_{n^2}|$). Then,

$$\mathcal{A}_l := \{\hat{A}_{\lceil 2^{l-1}n\varepsilon+1 \rceil}, \dots, \hat{A}_{\lceil 2^l n\varepsilon \rceil}\} \quad \text{for any } l \in \mathbb{Z}_{\geq 0}. \quad (1.2)$$

We are ready to state submatrix regularization algorithm:

Algorithm 1: Local norm regularization

Input: matrix $A = (A_{ij})_{i,j=1}^n$, constants $\varepsilon, c_\varepsilon > 0$, positive integer l_{\max} , disjoint entry subsets \mathcal{A}_l defined by (1.2) for $l \leq l_{\max}$

Output: \tilde{A} - $n \times n$ matrix, regularized version of A

1. Zero out $\lceil n\varepsilon/2 \rceil$ entries A_{ij} with the largest absolute values;
2. For $l = 0, \dots, l_{\max}$ find column index subset J_l in the following way:
 - 2a. For $j \in [n]$ define $e_j^{\text{row}}(\mathcal{A}_l) := |\{i : A_{ij} \in \mathcal{A}_l\}|$;
 - 2b. For every $i, j \in [n]$ define the weight

$$W_{ij}^l := \begin{cases} 1, & \text{if } e_j^{\text{row}}(\mathcal{A}_l) \leq c_\varepsilon n p_l \text{ or } A_{ij} \notin \mathcal{A}_l, \\ c_\varepsilon n p_l / e_j^{\text{row}}(\mathcal{A}_l), & \text{otherwise,} \end{cases}$$
 where we denoted $p_l = 2^l \varepsilon / n$;
 - 2c. Then, define $J_l := \{j : \prod_{i=1}^n W_{ij}^l \leq 0.1\}$;
3. Find subset \hat{J} of $n\varepsilon/4$ indices corresponding to the columns of A with the largest L_2 -norms, define $J := (\cup_l J_l) \cup \hat{J}$;
4. Repeat Steps 2-3 for A^T to find row subset $I := (\cup_l I_l) \cup \hat{I}$;
5. Zero out all the entries of A in the product subset $I \times J$ to get \tilde{A} .

If the matrix A is taken from the same model as in Theorem 1, the regularization provided by Algorithm 1 finds $\varepsilon n \times \varepsilon n$ submatrix $I \times J$ that one can replace with zeros to get a matrix \tilde{A} with well-bounded norm. This is proved in the following

Corollary 1 (Constructive regularization, submatrix version) *Let A be a random $n \times n$ matrix with i.i.d. entries A_{ij} having symmetric distribution such that $\mathbb{E}A_{ij}^2 = 1$. Let $\varepsilon \in (0, 1/6]$, $l_{\max} = \lfloor \log_2(\ln n / \ln \varepsilon^{-4}) \rfloor$ and $c_\varepsilon = (\ln \varepsilon^{-1})/\varepsilon$. Subsets \mathcal{A}_l are defined by (1.2) for $l = 0, \dots, l_{\max}$.*

Suppose that matrix \tilde{A} is constructed from A by Algorithm 1. Then, with probability $1 - n^{0.1-r}$ matrix, \tilde{A} differs from A on at most $\varepsilon n \times \varepsilon n$ submatrix, and

$$\|\tilde{A}\| \leq Cr^{3/2} \sqrt{c_\varepsilon n \cdot \ln \ln n},$$

where $C > 0$ is a sufficiently large absolute constant.

1.3 Proof Strategy

General idea of the proof of Theorem 1 is to split the entries of A into subsets with similar absolute values and bound each of these “levels” by the properly scaled 0–1 Bernoulli variables. Then, we can use known regularization results that hold for Bernoulli random matrices (see Sect. 2.1 for their short exposition) for each “level” separately, and sum the norms over the “levels”. However, this intuitive approach conceals several difficulties to be resolved on the way.

First, we cannot directly substitute the entries of A by their absolute values: The norm of $|A|$ might have global obstructions to the regularization (consider an example of ± 1 symmetric Rademacher entries). To approach the level splitting process more delicately, we strengthen the following ideas (initially proposed by Friedman, Kahn and Szemerédi for the regular graphs in [6] and modified by Feige and Ofek for more

general Bernoulli case in [5]). In the standard approximation of the operator norm $\|A\|$ by the supremum of a quadratic form on a discrete ε -net \mathcal{N} on the unit sphere

$$\sup_{u,v \in \mathcal{N} \subset S^{n-1}} \left| \sum_{i,j=1}^n A_{ij} u_i v_j \right|,$$

for each fixed pair of unit vectors $(u, v) = ((u_i)_{i=1}^n, (v_j)_{j=1}^n)$, one can split their indices $(i, j) \in [n] \times [n]$ into *light* and *heavy* couples. *Light* couples correspond to the sum members $u_i v_j$ that are bounded well enough for the classical concentration results to work, and *heavy* couples make a smaller set for which regularization does not need mean zero assumption any more (see Sect. 2.2 for more details). In the current work, we define similar notions of the *light* and *heavy* members that additionally take into account non-unit absolute values of A_{ij} (see Sect. 3).

The second difficulty is that a simple (based on L_2 norms) regularization of the matrix A does not imply proper regularization of all the “levels.” Even if we know that the rows and columns of the matrix A are well bounded, some of the “levels” might have too large row or column norms, if the others were small enough. To address this issue without making regularization procedure more complex, we employ an additional structural decomposition for Bernoulli random matrices, first shown in the work of Le et al. [9]. Essentially, we split the entries on each “level” into three parts: a part with bounded rows and columns, and two exceptional *sparse* parts. The sparsity comes in handy for their separate norm estimation.

Finally, we would like to make the number of “levels” as small as possible (as their quantity creates an additional factor in the resulting norm). We were able to keep this number as small as double logarithm of the matrix size, using the matrix entries truncations, similar to ones used in the preceding work [13].

The fact that the submatrix regularization Algorithm 1 is more involved than the one presented in Theorem 1 is somewhat natural. Zeroing out a small submatrix must still bring the L_2 -norms of all rows and columns to the order $O(\sqrt{n})$. Since the majority of rows and columns stays untouched in such regularization, essentially, one needs to find the most “dense” part of the matrix.

The procedure of assigning weights to the matrix entries row-wise, multiplying them to set column weights and then thresholding columns with the low weights is a delicate way to do so. This weight construction was originally used in [12, 13] for the matrices with i.i.d. scaled Bernoulli entries. Here, we employ the same construction to regularize the entries at every “level” independently. (Here, k th “level” contains the entries of A that belong to 2^{-k} -quantile of the distribution of A_{ij}^2 .) Additionally, to make the algorithm distribution oblivious, we estimate quantiles by order statistics of the matrix entries (since a random matrix naturally contains n^2 samples of the distribution $\xi \sim A_{ij}$). The idea to estimate quantiles of some distribution by the order statistics of a set of samples is both natural and well known in the statistics literature (see, e.g., [4, 22]).

1.4 Notations and Structure of the Paper

We use the following standard notations throughout the paper. Positive absolute constants are denoted C, C_1, c, c_1 , etc. Their values may be different from line to line. We often write $a \lesssim b$ to indicate that $a \leq Cb$ for some absolute constant C .

The discrete interval $\{1, 2, \dots, n\}$ is denoted by $[n]$. Given a finite set S , by $|S|$, we denote its cardinality. The standard inner product in \mathbb{R}^n shall be denoted by $\langle \cdot, \cdot \rangle$. Given $p \in [1, \infty]$, $\|\cdot\|_p$ is the standard ℓ_p^n -norm in \mathbb{R}^n . Given a matrix A , $\|\cdot\|$ denotes the operator $l_2 \rightarrow l_2$ norm of the matrix:

$$\|A\| := \max_{x \in S^{n-1}} \|Ax\|_2.$$

We write $\text{row}_1(A), \dots, \text{row}_n(A) \in \mathbb{R}^m$ to denote the rows of any $m \times n$ matrix A and $\text{col}_1(A), \dots, \text{col}_m(A) \in \mathbb{R}^n$ to denote its columns. We are going to use sparsity of the matrices in the proof. We denote by $e_i^{\text{row}}(A)$ the number of nonzero entries in the i th row of the matrix A , and also $e_i^{\text{col}}(A)$ denotes the number of nonzero entries in the i th column of A .

The rest of the paper is structured as follows. In Sect. 2, we list auxiliary results (from the works of Feige and Ofek [5], and Le et al. [9]) specific to the Bernoulli matrices. In Sect. 3, we show how to extend the Bernoulli techniques to more general class of matrices and prove central Proposition 2. In Sect. 4, we combine these techniques to conclude the proof of Theorem 1. In Sect. 5, we prove Corollary 1, and the last Sect. 6 contains discussion of the results and related open questions.

2 Auxiliary Results for Bernoulli Random Matrices

In this section, we review several useful results related to the regularization of the norms of Bernoulli matrices.

2.1 Regularization of the Norms of Bernoulli Random Matrices

Consider a $n \times n$ Bernoulli matrix B with independent 0–1 entries such that $\mathbb{P}\{B_{ij} = 1\} = p$. Since the second moment of its entries $\mathbb{E}B_{ij}^2 \sim p$, from the facts discussed in the beginning of Sect. 1, one would expect an ideal operator norm $\|B\| \sim \sqrt{np}$.

This is exactly what happens with high probability when success probability p is large enough ($p \gtrsim \sqrt{\ln n}$, see, e.g., [5]). If $p \ll \sqrt{\ln n}$, the norm can stay larger than optimal [7]. However, it is known that the regularization procedure described in Theorem 1 works in the case of Bernoulli matrices, and moreover, results in the optimal norm order $\|\tilde{B}\| \sim \sqrt{np}$. Since all nonzero entries in B have the same size, the L_2 norm description can be simplified in terms of the number of nonzero entries in each row or column.

Namely, Feige and Ofek proved in [5] that if we zero out all rows and columns that contain more than Cnp nonzero entries, then the resulting matrix satisfies $\|\tilde{B}\| \sim \sqrt{np}$. This result was improved by Le, Levina and Vershynin. In [9], the authors demonstrate

that it is enough to zero out any part of the rows and columns with too many nonzeros or reweigh them in any way, to satisfy $e_i^{\text{row}} \leq Cnp$ and $e_i^{\text{col}} \leq Cnp$ for any $i \in [n]$ to obtain the resulting matrix $\|\tilde{B}\| \sim \sqrt{np}$.

2.2 Regularization and the Quadratic Form

So, zeroing out all rows and columns that contain more than Cnp nonzero entries regularizes the quadratic form

$$\|B\| = \sup_{u, v \in S^{n-1}} \left| \sum_{ij} B_{ij} u_i v_j \right| \quad (2.1)$$

to the optimal order \sqrt{np} .

We will need the following Lemma 1, addressing the part of the sum (2.1) (over the indices i, j such that $\{|u_i v_j| \geq \sqrt{p/n}\}$). It was first proved by Feige and Ofek in [5] and later appears in [3]. For completeness, we give a short sketch of its proof below. Let us also emphasize that even though for the regularization procedure introduced in [5] it is crucial to zero out a *product* subset of the entries, in the framework of Lemma 1 it is possible to zero out *any* subset of the entries.

Lemma 1 *Let B be a $n \times n$ Bernoulli matrix with independent 0–1 entries such that $\mathbb{P}\{B_{ij} = 1\} = p$. Let $r \geq 1$. Let $\mathcal{B} \subset [n] \times [n]$ be an index subset, such that if we zero out all B_{ij} with $(i, j) \notin \mathcal{B}$, then every row and column of B has at most $C_0 r p n$ nonzero entries. Then, with probability $1 - n^{-r}$*

$$\sup_{u, v \in S^{n-1}} \sum_{\substack{(i, j) \in \mathcal{B}: \\ |u_i v_j| \geq \sqrt{p/n}}} B_{ij} |u_i v_j| \leq Cr \sqrt{np},$$

where C is a large enough absolute constant.

The proof of Lemma 1 is based on a technical Lemma 2 stated below. The proof of Lemma 2 is completely deterministic and can be found in [3,5].

Lemma 2 [3, Lemma 21] *Let B be an $n \times n$ matrix with 0–1 elements. Let $p > 0$, such that every row and column of B contains at most $C_0 np$ ones. For index subsets $S, T \subset [n]$ define*

$$e(S, T) := \sum_{i \in S, j \in T} B_{ij}$$

(i.e., number of nonzero elements in the submatrix spanned by $S \times T$). Suppose that for any $S, T \subset [n]$ one of the following conditions holds:

- (A) $e(S, T) \leq C_1 |S| |T| p$, or
- (B) $e(S, T) \cdot \log \left(\frac{e(S, T)}{|S| |T| p} \right) \leq C_2 |T| \log \left(\frac{n}{|T|} \right)$,

with some constants C_1 and C_2 independent from S, T and n . Then, for any $u, v \in S^{n-1}$

$$\sum_{i,j:|u_i v_j| \geq \sqrt{p/n}} B_{ij} |u_i v_j| \leq C \sqrt{np},$$

where the constant $C = \max\{16, 3C_0, 32C_1, 32C_3\}$.

Proof (of Lemma 1) In view of Lemma 2, it is enough to show that with probability $1 - n^{-r}$ for every $S, T \subset [n]$ and

$$e_{\mathcal{B}}(S, T) = \sum_{(i,j) \in \mathcal{B} \cap (S \times T)} B_{ij}$$

one of the conditions (A) and (B) holds. Without loss of generality, let us assume that $|T| \geq |S|$.

If $|T| \geq n/e$, we have

$$e_{\mathcal{B}}(S, T) \leq |S| \cdot C_0 r p n \leq C_0 r p e |T| |S|.$$

Hence, condition (A) holds with $C_1 = C_0 r e$.

If both $|S|, |T| < n/e$, then $\mathbb{P}\{\mathcal{E}\} \leq n^{-r}$ for the event

$$\mathcal{E} := \{\exists S, T : |S|, |T| < n/e, e_{\mathcal{B}}(S, T) > l_{S,T} p |S| |T|\}$$

and $l_{S,T}$ being a number such that

$$l_{S,T} \ln l_{S,T} := \ln \frac{n}{|T|} \cdot \frac{3(r+6)|T|}{p|S||T|}.$$

Indeed, the probability estimate for \mathcal{E} follows from the Chernoff’s inequality applied to the sum of independent variables $e(S, T) \geq e_{\mathcal{B}}(S, T)$, combined with the Stirling formula estimating the number of proper sets S and T , and the fact that the function $f(x) = (x/n)^x$ is monotonically increasing on $[1, n/e]$ (see [5, Section 2.2.5] for the computation details).

Thus, with probability $1 - n^{-r}$, for any S, T , such that $|S|, |T| < n/e$, condition (B) holds:

$$\begin{aligned} e_{\mathcal{B}}(S, T) \cdot \ln \left(\frac{e_{\mathcal{B}}(S, T)}{|S||T|p} \right) &\leq l_{S,T} p |S| |T| \cdot \ln l_{S,T} \\ &\leq 3(r+6)|T| \ln \frac{n}{|T|}. \end{aligned}$$

This concludes Lemma 1 with $C = \max\{32C_0 r e, 100(r+6)\}$.

2.3 Decomposition of Bernoulli Matrices

As we mentioned above, the idea is to apply an approach developed for Bernoulli matrices for the truncations of the entries of A , having absolute values on the same “level,” and then to sum over these “levels.”

However, even if we know that the rows and columns of the general matrix A are well bounded, there is no direct way to see which “levels” are well bounded: Some of them might be too large if the others are small enough. To address this issue, we are going to use additional structural decomposition for Bernoulli random matrices. The next proposition is a direct corollary of [9, Theorem 2.6]:

Proposition 1 (Decomposition lemma) *Let B be a $n \times n$ Bernoulli matrix with independent 0–1 entries such that $\mathbb{P}\{B_{ij} = 1\} = p$. Then, for any $n \geq 4$, $p \geq 0$ and $r \geq 1$, with probability at least $1 - 3n^{-r}$, all entries of B can be divided into three disjoint classes $[n] \times [n] = \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \mathcal{B}_3$, such that*

1. $e_i^{\text{row}}(\mathcal{B}_1) \leq C_1 r^3 np$ and $e_i^{\text{col}}(\mathcal{B}_1) \leq C_1 r^3 np$
2. $e_i^{\text{row}}(\mathcal{B}_2) \leq C_2 r$
3. $e_i^{\text{col}}(\mathcal{B}_3) \leq C_2 r$,

where $e_i^{\text{col/row}}(\mathcal{B})$ is the number of nonzero elements in i th row or column of B belonging to the class \mathcal{B} and C_1, C_2 are absolute constants.

Remark 2 Following the same methods as was employed in the proof of [9, Theorem 2.6], one can check that Proposition 1 actually holds with linear (instead of cubic) dependence on r , namely $e_i^{\text{row}}(\mathcal{B}_1) \leq C_1 r np$ and $e_i^{\text{col}}(\mathcal{B}_1) \leq C_1 r np$.

3 From Bernoulli to General Matrices

The goal of this section is to prove Proposition 2 that provides a way to generalize the regularization results known for Bernoulli random matrices to the general case:

Proposition 2 *Suppose A is a random $n \times n$ matrix with i.i.d. symmetric entries A_{ij} with $\mathbb{E}A_{ij}^2 = 1$. Let \tilde{A} be the resulting matrix after we have zeroed out row and column subsets of A in any way, such that*

$$\|\text{row}_i(\tilde{A})\|_2^2 \leq c_\varepsilon n \text{ and } \|\text{col}_i(\tilde{A})\|_2^2 \leq c_\varepsilon n \quad (3.1)$$

for all $i = 1, \dots, n$. Let

$$\tilde{M} := \tilde{A} \cdot \mathbb{1}_{\{|\tilde{A}_{ij}| \in (2^{k_0}, 2^{k_1}]\}}, \text{ and } k_1 - k_0 =: \kappa.$$

Then, with probability at least $1 - 10\kappa n^{-r}$, we have

$$\|\tilde{M}\| \leq Cr\sqrt{c_\varepsilon n \kappa}.$$

Here, c_ε is any positive ε -dependence and $C > 0$ is an absolute constant.

Let us first collect several auxiliary lemmas that will be used in the proof of Proposition 2.

Lemma 3 Consider a $n \times n$ random matrix M with independent symmetric entries and $\mathbb{E}M_{ij}^2 \leq 1$. Consider two vectors $u = (u_i)_{i=1}^n$ and $v = (v_j)_{j=1}^n$ such that $u, v \in S^{n-1}$. Denote the event

$$\mathcal{M}_{ij}^{\text{light}} = \{|M_{ij}| |u_i v_j| \leq 2/\sqrt{n}\}$$

and let $Q \subset [n] \times [n]$ be an index subset. Then, for any constant $C \geq 3$

$$|\sum_{ij} u_i M_{ij} \mathbb{1}_{\{(i,j) \in Q\}} \mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}} v_j| \leq C\sqrt{n}$$

with probability at least $1 - 2 \exp(-Cn/2)$.

Proof Let $R_{ij} := M_{ij} \mathbb{1}_{\{(i,j) \in Q\}} \mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}}$. Note that R_{ij} are centered due to the symmetric distribution of M_{ij} , and they are independent as M_{ij} are. So we can apply Bernstein’s inequality for bounded distributions (see, for example, [20, Theorem 2.8.4]) to bound the sum:

$$\mathbb{P}\{|\sum_{ij} u_i R_{ij} v_j| \geq t\} \leq 2 \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right),$$

where

$$K = \max_{i,j} |u_i R_{ij} v_j| \leq 2/\sqrt{n} \quad \text{and} \quad \sigma^2 = \sum_{ij} \mathbb{E}(u_i R_{ij} v_j)^2.$$

Note that $\mathbb{E}R_{ij}^2 \leq \mathbb{E}M_{ij}^2$, as $R_{ij}^2 \leq M_{ij}^2$ almost surely, and $\mathbb{E}M_{ij}^2 \leq 1$. So,

$$\sigma^2 = \sum_{ij} u_i^2 \mathbb{E}R_{ij}^2 v_j^2 \leq \sum_{i,j} u_i^2 v_j^2 = 1,$$

as $\sum_i u_i^2 = \sum_j v_j^2 = 1$. So, taking $t = C\sqrt{n}$, we obtain

$$\mathbb{P}\{|\sum_{(i,j)} u_i M_{ij} \mathbb{1}_{\{(i,j) \in Q\}} \mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}} v_j| \geq C\sqrt{n}\} \leq 2 \exp(-Cn/2)$$

for any $C \geq 3$. This concludes the statement of the lemma. □

The following lemma is a version of [9, Lemma 3.3].

Lemma 4 For any matrix Q and vectors $u, v \in S^{n-1}$, we have

$$\sum_{ij} Q_{ij} u_i v_j \leq \max_j \|\text{col}_j(Q)\|_2 \cdot (\max_i e_i^{\text{row}}(Q))^{1/2}.$$

Proof Indeed,

$$\begin{aligned}
 \sum_{ij} Q_{ij} u_i v_j &\leq \sum_j v_j \left(\sum_{i:Q_{ij}\neq 0} Q_{ij} u_i \right) \\
 &\leq \sum_j v_j \left(\sum_{i:Q_{ij}\neq 0} Q_{ij}^2 \right)^{1/2} \left(\sum_{i:Q_{ij}\neq 0} u_i^2 \right)^{1/2} \quad (*) \\
 &\leq \max_j \|\text{col}_j(Q)\|_2 \sum_j v_j \left(\sum_{i:Q_{ij}\neq 0} u_i^2 \right)^{1/2} \\
 &\leq \max_j \|\text{col}_j(Q)\|_2 \left(\sum_j v_j^2 \right)^{1/2} \left(\sum_j \sum_{i:Q_{ij}\neq 0} u_i^2 \right)^{1/2} \quad (*) \\
 &\leq \max_j \|\text{col}_j(Q)\|_2 \cdot 1 \cdot \left(\sum_i u_i^2 \sum_{j:Q_{ij}\neq 0} 1 \right)^{1/2} \quad (\text{since } \|v\|_2 = 1) \\
 &\leq \max_j \|\text{col}_j(Q)\|_2 \left(\sum_i u_i^2 e_i^{\text{row}}(Q) \right)^{1/2} \\
 &\leq \max_j \|\text{col}_j(Q)\|_2 \cdot \left(\max_i e_i^{\text{row}}(Q) \right)^{1/2}. \quad (\text{since } \|u\|_2 = 1)
 \end{aligned}$$

Steps (*) hold by the Cauchy–Schwarz inequality. Lemma 4 is proved. \square

In the proof of Proposition 2, we are going to use a standard splitting “by size” of a nonnegative random variable. Let $X = 0$ or $X \in (2^{k_0}, 2^{k_1}]$ almost surely. Then, clearly,

$$X \leq \sum_{k=k_0+1}^{k_1} 2^k \mathbb{1}_{\{X \in (2^{k-1}, 2^k]\}}.$$

If additionally $\mathbb{E}X^2 \leq 1$ and we denote

$$p_k := \mathbb{P}\{|M_{ij}| \in (2^{k-1}, 2^k]\}, \quad (3.2)$$

the following estimates hold for p_k . First, the sum

$$\sum p_k 2^{2k} \leq 4 \sum p_k 2^{2(k-1)} \leq 4\mathbb{E}X^2 \leq 4. \quad (3.3)$$

Now, we are ready to prove Proposition 2.

3.1 Proof of Proposition 2

Step 1. Net approximation

Let \mathcal{N} be a $1/2$ -net on S^{n-1} with cardinality $|\mathcal{N}| \leq 5^n$. (The existence of such net is a standard fact that can be found, e.g., in [20].) We will use a simple net approximation of the norm (see, e.g., [20, Lemma 4.4.1]), namely

$$\|\tilde{M}\| \leq 4 \max_{u,v \in \mathcal{N}} \langle \tilde{M}u, v \rangle = 4 \max_{u,v \in \mathcal{N}} \left| \sum_{ij} \tilde{M}_{ij} u_i v_j \right|.$$

We will split the sum into two parts and bound each of them separately, based on the absolute value of the element.

Let $M := A \cdot \mathbb{1}_{\{|A_{ij}| \in (2^{k_0}, 2^{k_1}]\}}$. For any fixed $u, v \in \mathcal{N}$ and every $i, j \in [n]$, we define an event

$$\mathcal{M}_{ij}^{\text{light}} = \{|M_{ij}| |u_i v_j| \leq 2/\sqrt{n}\}.$$

A random set of indices

$$\mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}} = \{(i, j) : |M_{ij}| |u_i v_j| \leq 2/\sqrt{n}\}$$

is an analogue of the non-random set of “light couples” (i, j) used by Feige and Ofek in [5]. The difference appears since in the non-Bernoulli setting we have to take into account additional random scaling by absolute values of the corresponding matrix entries $|M_{ij}|$.

Then,

$$\begin{aligned} & \max_{u,v \in \mathcal{N}} \left| \sum_{i,j} \tilde{M}_{ij} u_i v_j \right| \\ & \leq \max_{u,v \in \mathcal{N}} \left| \sum_{i,j} \tilde{M}_{ij} (\mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}} + \mathbb{1}_{(\mathcal{M}_{ij}^{\text{light}})^c}) u_i v_j \right| \\ & \leq \max_{u,v \in \mathcal{N}} \left| \sum_{i,j} \tilde{M}_{ij} \mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}} u_i v_j \right| + \max_{u,v \in \mathcal{N}} \left| \sum_{i,j} \tilde{M}_{ij} \mathbb{1}_{(\mathcal{M}_{ij}^{\text{light}})^c} u_i v_j \right|. \end{aligned}$$

Step 2. Light members

By Lemma 3, for any fixed $u, v \in S^{n-1}$ and a fixed subset of indices Q (assuming that Q^c is a set of rows and columns to delete),

$$\left| \sum_{i,j} u_i M_{ij} \mathbb{1}_{\{(i,j) \in Q\}} \mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}} v_j \right| \leq 12\sqrt{n} \tag{3.4}$$

with probability at most $2 \exp(-6n)$. Now, taking union bound over 5^n choices for $u \in \mathcal{N}$, as many choices for $v \in \mathcal{N}$, and 2^{2n} choices for the row and column subset Q^c , we obtain that

$$\mathbb{P}\left\{\sum_{i,j} u_i \tilde{M}_{ij} \mathbb{1}_{\mathcal{M}_{ij}^{\text{light}}} v_j \leq 12\sqrt{n}\right\} \geq 1 - 2 \exp(-n). \tag{3.5}$$

Step 3. Other members

The second sum can be roughly bounded by the sum of absolute values of its members:

$$\begin{aligned} \left| \sum_{i,j} \tilde{M}_{ij} \mathbb{1}_{(\mathcal{M}_{ij}^{\text{light}})^c} u_i v_j \right| &\leq \sum_{i,j} |\tilde{M}_{ij}| \mathbb{1}_{(\mathcal{M}_{ij}^{\text{light}})^c} |u_i v_j| \\ &\leq \sum_{i,j} \left[\sum_{k=k_0+1}^{k_1} 2^k \mathbb{1}_{\{|\tilde{M}_{ij}| \in (2^{k-1}, 2^k]\}} \right] \mathbb{1}_{\{|M_{ij}| |u_i v_j| \geq 2/\sqrt{n}\}} |u_i v_j| \end{aligned}$$

Note that as long as $\mathbb{1}_{\{|\tilde{M}_{ij}| \in (2^{k-1}, 2^k]\}} = 1$, we also have that $|M_{ij}| \leq 2^k$. Indeed, $|M_{ij}| > 2^k$ implies either $|\tilde{M}_{ij}| > 2^k$ or $|\tilde{M}_{ij}| = 0$. In any case, $|\tilde{M}_{ij}| \notin (2^{k-1}, 2^k]$. So, the last expression is bounded above by

$$\sum_{i,j} \sum_{k=k_0+1}^{k_1} 2^k \mathbb{1}_{\{|\tilde{M}_{ij}| \in (2^{k-1}, 2^k]\}} \mathbb{1}_{\{2^k |u_i v_j| \geq 2/\sqrt{n}\}} |u_i v_j|$$

Since $\mathbb{E}M_{ij}^2 \leq \mathbb{E}A_{ij}^2 = 1$, from (3.3), we have $2^{1-k} \geq \sqrt{p_k}$ for any k , where p_k is probability of the k th level (defined in (3.2)). As a result, in Step 3, we got

$$\begin{aligned} \left| \sum_{i,j} \tilde{M}_{ij} \mathbb{1}_{(\mathcal{M}_{ij}^{\text{light}})^c} u_i v_j \right| &\leq \sum_{k=k_0+1}^{k_1} 2^k \sum_{i,j: |u_i v_j| \geq \sqrt{p_k/n}} \mathbb{1}_{\{|\tilde{M}_{ij}| \in (2^{k-1}, 2^k]\}} |u_i v_j|. \tag{3.6} \end{aligned}$$

Step 4. Bernoulli matrices

For each “size” $k = k_0 + 1, \dots, k_1$, let us define a $n \times n$ matrix B^k with independent Bernoulli entries

$$B_{ij}^k := \mathbb{1}_{\{|M_{ij}| \in (2^{k-1}, 2^k]\}}, \quad \mathbb{E}B_{ij}^k = p_k.$$

By Decomposition Lemma 1 (and Remark 2), with probability $1 - 3n^{-r}$, the entries of every B^k can be assigned to one of three disjoint classes: \mathcal{B}_1^k , where all rows and columns have at most $C_1 r p_k n$ nonzero entries; \mathcal{B}_2^k , where all the columns have at most $C_2 r$ nonzero entries; and \mathcal{B}_3^k , where all the rows have at most $C_2 r$ nonzero entries. We are going to further split the sum (3.6) into three sums containing elements of these three classes and estimate each of them separately.

- \mathcal{B}_1 : The part with the entries from \mathcal{B}_1^k satisfies the conditions of Lemma 1. For any $k = k_0 + 1, \dots, k_1$

$$\sum_{\substack{(i,j) \in \mathcal{B}_1^k: \\ |u_i v_j| \geq \sqrt{p_k/n}}} \mathbb{1}_{\{|\tilde{M}_{ij}| \in (2^{k-1}, 2^k]\}} |u_i v_j| \leq \sum_{\substack{(i,j) \in \mathcal{B}_1^k: \\ |u_i v_j| \geq \sqrt{p_k/n}}} B_{ij}^k |u_i v_j|.$$

By Lemma 1, this sum is bounded by $Cr\sqrt{p_k n}$ with probability at least $1 - n^{-r}$. So, for any $u, v \in S^{n-1}$

$$\sum_{k=k_0+1}^{k_1} 2^k \sum_{\substack{(i,j) \in \mathcal{B}_1^k: \\ |u_i v_j| \geq \sqrt{p_k/n}}} B_{ij}^k |u_i v_j| \leq Cr\sqrt{n} \sum_{k=k_0+1}^{k_1} 2^k \sqrt{p_k}.$$

Then, by the Cauchy–Schwarz inequality and estimate (3.3),

$$\sum_{k=k_0+1}^{k_1} 2^k \sqrt{p_k} \leq \sqrt{\sum_{k=k_0+1}^{k_1} 2^{2k} p_k} \sqrt{\sum_{k=k_0+1}^{k_1} 1} \leq 2\sqrt{\kappa}.$$

- \mathcal{B}_2 : The part with the entries from \mathcal{B}_2^k can be estimated by Lemma 4. We have that

$$\sum_{k=k_0+1}^{k_1} 2^k \sum_{\substack{(i,j) \in \mathcal{B}_2^k: \\ |u_i v_j| \geq \sqrt{p_k/n}}} \mathbb{1}_{\{|\tilde{M}_{ij}| \in (2^{k-1}, 2^k]\}} |u_i v_j| \leq \sum_{i,j} Q_{ij} |u_i v_j|,$$

where

$$Q_{ij} := \sum_{k=k_0+1}^{k_1} 2^k \mathbb{1}_{\{(i,j) \in \mathcal{B}_2\}} \mathbb{1}_{\{|\tilde{M}_{ij}| \in (2^{k-1}, 2^k]\}}$$

Note that for every fixed $i = 1, \dots, n$, number of nonzero entries Q_{ij} in the row i is at most $C_2 r \kappa$. Also, $|Q_{ij}| \leq 2|\tilde{M}_{ij}| \leq 2|\tilde{A}_{ij}|$ almost surely, so maximum L_2 -norm of the column of Q is $\sqrt{c_\varepsilon n}$. By Lemma 4, this implies that for any $u, v \in S^{n-1}$

$$\sum_{i,j} Q_{ij} |u_i v_j| \leq \sqrt{C_2 r \kappa c_\varepsilon n}$$

- \mathcal{B}_3 : The part with the entries from \mathcal{B}_3^k can be estimated in the same way as \mathcal{B}_2^k , repeating the argument for A^T .

Step 5. Conclusion

Now, we can combine the estimates obtained for the light members (3.5) and all three parts of the non-light sum, to get that

$$\|\tilde{M}\| \leq 12\sqrt{n} + Cr\sqrt{\kappa n} + 2\sqrt{C_2 r \kappa c_\varepsilon n} \lesssim r\sqrt{c_\varepsilon \kappa n}$$

with probability at least $1 - 2e^{-n} - 3\kappa n^{-r} - 6n^{-r} \geq 1 - 10n^{-r}\kappa$ for all n large enough. Proposition 2 is proved.

4 Conclusions: Proof of Theorem 1 and Further Directions

In this section, we conclude the proof of Theorem 1. As we have seen in the previous section, splitting the entries of the matrix A into κ “levels” with the similar absolute value produces an extra $\sqrt{\kappa}$ factor in the norm estimate. Hence, we care to make the number of levels as small as possible. We are going to show that this number can be as small as $C \ln \ln n$, where n is the size of the matrix. The reason is that we need only to consider the “average” entries of the matrix, those with the absolute values between $O(\sqrt{n/\ln n})$ and $O(\sqrt{n})$. The “large” entries will be all replaced by zeros in regularization, and restriction to the “small” entries creates a matrix with the optimal norm (no regularization is needed). One way to check this is by applying the following result of Bandeira and Van Handel:

Theorem 2 [2, Lemma 21] *Let X be an $n \times n$ matrix whose entries X_{ij} are independent centered random variables. For any $\varepsilon \in (0, 1/2]$, there exists a constant c_ε such that for every $t \geq 0$,*

$$\mathbb{P}\{\|X\| \geq (1 + \varepsilon)6\sigma + t\} \leq n \exp(-t^2/c_\varepsilon \sigma_*^2),$$

where σ is a maximum expected row and column norm:

$$\sigma^2 := \max(\sigma_1^2, \sigma_2^2), \text{ where } \sigma_1^2 = \max_i \sum_j \mathbb{E}(X_{ij}^2), \quad \sigma_2^2 = \max_j \sum_i \mathbb{E}(X_{ij}^2);$$

and σ_* is a maximum absolute value of an entry:

$$\sigma_* := \max_{ij} \|X_{ij}\|_\infty.$$

Lemma 5 *Suppose S is a random $n \times n$ matrix with i.i.d. mean zero entries S_{ij} , such that $\mathbb{E}S_{ij}^2 \leq 1$ and $|S_{ij}| < \bar{c}\sqrt{n}/\sqrt{\ln n}$. Let $r \geq 1$. If $\bar{c} < c$, then with probability at least $1 - n^{-r}$*

$$\|S\| \leq 13r\sqrt{n}.$$

Here, c is a small enough absolute constant.

Proof The proof follows directly from Theorem 2 with $t = r\sqrt{n}$ and $\varepsilon = 1$. It is enough to take $c = 1/11c_1$, where c_1 is a constant from the statement of Theorem 2. \square

Now, we are ready to prove Theorem 1.

4.1 Proof of Theorem 1

Let us decompose A into a sum of three $n \times n$ matrices

$$A := S + M + L, \tag{4.1}$$

where S contains the entries of A that satisfy $|A_{ij}| \leq 2^{k_0}$, the matrix M contains the entries for which $2^{k_0} < |A_{ij}| \leq 2^{k_1}$, and L contains large entries, satisfying $|A_{ij}| > 2^{k_1}$. Here,

$$k_0 := \left\lceil \frac{1}{2} \log_2 \frac{c_1 n}{\ln n} \right\rceil, \quad k_1 := \left\lceil \frac{1}{2} \log_2 (C_2 c_\varepsilon n) \right\rceil, \quad \text{where } c_\varepsilon = (\ln \varepsilon^{-1})/\varepsilon$$

and $c_1, C_2 > 0$ are absolute constants.

Note that S, M and L inherit essential properties of the matrix A . First, they also have i.i.d. entries (since they are obtained by independent individual truncations from the i.i.d. elements A_{ij}). Due to the symmetric distribution of the entries of A , the entries of S, M and L have mean zero. Also, their second moment is bounded from above by $\mathbb{E}A_{ij}^2 = 1$.

Note that all the entries in S satisfy $|A_{ij}| \leq \sqrt{c_1 n / \ln n}$. Thus, as long as we choose constant c_1 small enough to satisfy the condition in Lemma 5, the norm of S can be estimated as

$$\mathbb{P}\{\|S\| > 13\sqrt{n}\} < n^{-r}. \tag{4.2}$$

Clearly, replacing by zeros some rows and column subset can only decrease the norm of S .

By Remark 1, with probability at least $1 - 7 \exp(\varepsilon n/12)$, all rows and columns of \tilde{A} have bounded norms: for $i = 1, \dots, n$

$$\|\text{row}_i(\tilde{A})\|_2 \leq C\sqrt{c_\varepsilon n} \text{ and } \|\text{col}_i(\tilde{A})\|_2 \leq C\sqrt{c_\varepsilon n}.$$

In particular, it implies that all the entries of \tilde{A} have absolute values bound by $C\sqrt{c_\varepsilon n}$. So, by taking constant $C_2 \geq C^2$, we achieve that L will be empty after the regularization.

Proposition 2 estimates the norm of M after the regularization (zeroing out row and columns with large norms):

$$\mathbb{P}\{\|\tilde{A} \cdot \mathbf{1}_{\{|A_{ij}| \in (2^{k_0}, 2^{k_1})\}}\| > Cc_\varepsilon r \sqrt{n\kappa}\} \leq 10n^{-r} \kappa. \tag{4.3}$$

By definition,

$$\kappa := k_1 - k_0 \leq \frac{1}{2} \left[\log_2(C_2 c_\varepsilon n) - \log_2 \frac{c_1 n}{\ln n} + 2 \right] \leq \log_2 \ln n$$

for all large enough n .

Using triangle inequality to combine norm estimates (4.2) and (4.3), we get $\|\tilde{A}\| \lesssim c_\varepsilon r \sqrt{n} \cdot \ln \ln n$ with probability at least

$$1 - n^{-r} - 7e^{-\varepsilon n/12} - 10n^{-r} \ln \ln n \geq 1 - n^{-r+0.1}$$

for all n large enough. This concludes the proof of Theorem 1.

Remark 3 Note that the only place in the argument where it matters what entry subset we replace by zeros is Step 2 of the proof of Proposition 2. To have a manageable union bound estimate, we need to be sure that the number of options for the potential subset to be deleted is of order $\exp(n)$. (So, $\exp(\ln 2 \cdot n^2)$ options for a general entry subset would be too much.)

Hence, also recalling Remark 1, we emphasize that the norm estimate (1.1) holds with the probability $1 - n^{0.1-r}$ as long as we achieve L_2 -norm of all rows and columns bounded by $C\sqrt{c_\varepsilon n}$ by zeroing out any product subset of the entries of A .

5 Proof of Corollary 1

General idea of the proof of Corollary 1 is to show that after the regularization procedure, all rows and columns of the matrix have well-bounded L_2 -norms and then apply Theorem 1.

Originally, the core part of the algorithm (Step 2) was presented for Bernoulli random matrices in [12,13]. We will use the following version of [13, Lemma 5.1] (based on the ideas developed in [12, Proposition 2.1]):

Lemma 6 *Let B be a $n \times n$ matrix with independent 0–1 valued entries, $\mathbb{E}B_{ij} \leq p$. Then, for any $L \geq 10$ with probability $1 - \exp(-n \exp(-Lpn))$, the following holds. If we define*

$$W_{ij} := \begin{cases} 1, & \text{if } e_j^{\text{row}}(B) \leq Lnp \text{ or } B_{ij} = 0, \\ Lnp/e_j^{\text{row}}(B), & \text{otherwise.} \end{cases}$$

and $V_j := \prod_{i=1}^n W_{ij}$, and $J := \{j : V_j < 0.1\}$, then

$$|J| \leq n \exp(-Lnp) \quad \text{and} \quad \sum_{j \in J^c} B_{ij} \leq 10Lnp, \text{ for any } i \in [n].$$

In order to pass from Bernoulli case to the general distribution case, we are going to use some version of “level truncation” idea once again. Note that here we need

the probabilities of the levels p_l to be both not too large (for the joint cardinality estimate) and not too small (for the probability estimate union bound). This motivates the idea to group “similar size” entries A_{ij} not by absolute value $|A_{ij}|$, but by common 2^{-l} -quantiles of the distribution of $\xi \sim A_{ij}^2$.

Remark 4 A version of Corollary 1 can be proved, when one would define the sets \mathcal{A}_l to contain all A_{ij} such that $A_{ij}^2 \in (q_{k-1}, q_k]$, where q_l is 2^l th quantile of the distribution of A_{ij}^2 , namely

$$q_l := \inf\{t : \mathbb{P}\{A_{ij}^2 > t\} \leq 2^{-l}\}. \tag{5.1}$$

The proof of this version is actually almost identical to the one presented below and gives smaller absolute constants. However, an additional requirement to know quantiles of the distribution of the entries in order to regularize the norm of the matrix seems undesirable. So, we are going to prove the distribution oblivious version as presented by Algorithm 1.

The next lemma shows that the order statistics used in the statement of Corollary 1 approximate quantiles of the distribution of A_{ij}^2 .

Lemma 7 (Order statistics as approximate quantiles) *Let $\hat{A}_1, \dots, \hat{A}_{n^2}$ be all the entries of $n \times n$ random matrix A in an non-increasing order, and q_k be 2^{-k} quantiles of the distribution of A_{ij}^2 defined by (5.1). Then, with probability at least $1 - 4 \exp(-n^2 2^{-k_1 - 2})$ for all $k = 1, \dots, k_1$*

$$q_{k-2} \leq \hat{A}_{\lceil n^2 2^{1-k} \rceil} \leq q_k. \tag{5.2}$$

Proof A direct application of Chernoff’s inequality shows that for any k

$$\mathbb{P}\{\nu_1 > \lceil n^2 2^{1-k} \rceil\} \leq \exp(-n^2 2^{-k} / 4),$$

where ν_1 is a number of entries A_{ij} such that $A_{ij}^2 > q_k$. Another application of Chernoff’s inequality lower bound shows that

$$\mathbb{P}\{\nu_2 < \lceil n^2 2^{1-k} \rceil\} \leq \exp(-n^2 2^{-k} / 4),$$

where ν_2 is a number of entries A_{ij} such that $A_{ij}^2 > q_{k-2}$.

Then, with probability at least $1 - 2 \exp(-n^2 2^{-k})$, the order statistics $\hat{A}_{\lceil n^2 2^{2-k} \rceil}$ is at least q_{k-2} and at most q_k . Taking union bound, Eq. (5.2) holds with probability $1 - 4 \exp(-n^2 2^{-k_1} / 4)$ for all $k = 1, \dots, k_1$. \square

Remark 5 We will use $k_1 = \lceil \log_2(8n/\varepsilon) \rceil$. An easy computation using (5.2) shows that

$$q_{k_1-l-3} \leq \hat{A}_{\lceil 2^{4+l-k_1} n^2 \rceil} \leq \hat{A}_{\lceil 2^l n \varepsilon \rceil} \leq \hat{A}_{\lceil 2^{3+l-k_1} n^2 \rceil} \leq q_{k_1-l-2}$$

for all $l = 0, \dots, l_{\max}$ with probability $1 - 4 \exp(-n\varepsilon/4)$. Then, for \mathcal{A}_l as defined in (1.2) and all $l \leq l_{\max}$,

$$\mathbb{P}\{A_{ij} \in \mathcal{A}_l\} \leq 2^{3+l-k_1} \leq 2^l \varepsilon / n. \tag{5.3}$$

Now, we are ready to prove Corollary 1.

5.1 Proof of Corollary 1

Let $k_1 = \lceil \log_2(8n/\varepsilon) \rceil$, and q_{k_1} is a corresponding 2^{-k_1} quantile of the distribution of A_{ij}^2 (as defined by (5.1)). It is easy to check by Chernoff’s inequality that the total number of entries in A such that $A_{ij}^2 \geq q_{k_1}$ is at most $\varepsilon n/2$ with probability at least $1 - e^{-\varepsilon n/4}$. So, all these “large” entries will be replaced by zeros at Step 1 of regularization Algorithm 1.

To prove Corollary 1, it is enough to show that:

1. Algorithm 1 makes all rows and columns of the truncated matrix $A \cdot \mathbb{1}_{\{A_{ij}^2 < q_{k_1}\}}$ have norms bounded by $C\sqrt{c_\varepsilon n}$. Then, in view of Remark 3, we can apply Theorem 1 to conclude the desired norm estimate.
2. the cardinalities of the exceptional index subsets $|I|, |J| \leq \varepsilon n/2$ with high probability. Then, the regularization procedure is indeed *local*.

The matrix $\bar{A} := A \cdot \mathbb{1}_{\{A_{ij}^2 < q_{k_1}\}}$ is naturally decomposed into the union of l_{\max} “levels” with the entries coming from sets \mathcal{A}_l and “the leftover” part that contains A_{ij} such that $A_{ij}^2 < \hat{A}_{\lceil n\varepsilon 2^{l_{\max}} \rceil}$. So,

$$\bar{A}_{ij} = A_{ij} \mathbb{1}_{\{A_{ij} \in \cup \mathcal{A}_l\}} + A_{ij} \mathbb{1}_{\{|A_{ij}| < \hat{A}_{\lceil n\varepsilon 2^{l_{\max}} \rceil}\}} =: A^{\text{Large}} + A^{\text{Small}}.$$

All the rows and columns of A^{Small} have L_2 -norms at most $C\sqrt{nr c_\varepsilon}$ with probability at least $1 - n^{-r}$ (without any regularization). This follows from an application of Bernstein’s inequality (e.g., [20, Theorem 2.8.4]) for a sum of independent centered entries bounded by $\sqrt{Cn/\ln n}$. Indeed, we just need to check the boundedness condition. Recall that $l_{\max} = \lfloor \log_2(\ln n / \ln \varepsilon^{-4}) \rfloor$. By definition of quantiles q_k and Markov’s inequality

$$\mathbb{P}\{A_{ij}^2 \geq q_{k_1 - 2 - l_{\max}}\} \geq 2^{-k_1 + 1} \frac{\ln n}{\ln \varepsilon^{-4}} \geq \mathbb{P}\left\{A_{ij}^2 \geq \frac{32c_\varepsilon n}{\ln n}\right\}.$$

Hence, the entries of A^{Small} can be estimated from above by

$$\hat{A}_{\lceil n\varepsilon 2^{l_{\max}} \rceil} \leq q_{k_1 - 2 - l_{\max}} \lesssim \frac{32c_\varepsilon n}{\ln n}.$$

For the part A^{Large} , we use Lemma 6 applied to $n \times n$ matrices with i.i.d. entries $B_{ij}^l = \mathbb{1}_{\{A_{ij} \in \mathcal{A}_l\}}$ for each $l = 0, \dots, l_{\max}$ with $L = 2c_\varepsilon$ and $p_l = 2^l \varepsilon/n$ (which is a valid choice by Remark 5). From the union bound estimate, we can conclude that the statement of Lemma 6 holds for all $l \leq l_{\max}$ with high probability

$$1 - \sum_{l \leq l_{\max}} \exp(-n \exp(-2c_\varepsilon n 2^l \varepsilon/n)) \geq 1 - \exp(-n^{0.5}).$$

Recall that $\bar{J} = \cup_l J_l$ is the union of all exceptional column index subsets found for all matrices $A \cdot \mathbb{1}_{\{A_{ij} \in \mathcal{A}_l\}}$ with $l = 0, \dots, l_{\max}$. Note that by the definition of quantiles and second moment condition,

$$\sum_{s=0}^{\infty} q_s 2^{-s-1} \leq \mathbb{E}A_{ij}^2 \leq 1. \tag{5.4}$$

By Lemma 6, we can estimate for every $i \in [n]$

$$\begin{aligned} \|\text{row}_i(A_{[n] \times \bar{J}^c}^{\text{Large}})\|_2^2 &\leq \sum_{l \leq l_{\max}} q_{k_1-l-2} 20c_\varepsilon n p_l \\ &\leq \sum_{l \leq l_{\max}} q_{k_1-l-2} 20c_\varepsilon n \frac{2^{l-k_1} \varepsilon}{n} 2^{k_1} \leq 160c_\varepsilon n, \end{aligned}$$

as $2^{k_1} \leq 16n/\varepsilon$, we used (5.4) with $s = k_1 - l - 2$ in the last step.

Then, by the L_2 -norm triangle inequality applied to the rows of $A_{[n] \times \bar{J}^c}^{\text{Large}}$ and A^{Small} , we have the row boundedness condition satisfied for $\tilde{A}_{[n] \times \bar{J}^c}$. Next, at Step 3, we add the set \hat{J} of columns with largest L_2 -norms. The same argument as in Remark 1 shows that with probability at least $1 - n^{-r}$, there are no columns with the norm larger than $C\sqrt{c_\varepsilon n}$ outside the set \hat{J} . So, matrix $\tilde{A}_{[n] \times J^c}$ has all rows and columns norms well bounded (recall that $J := \bar{J} \cup \hat{J}$). Then, by Theorem 1, with high probability $1 - C(\ln \ln n)n^{-r}$

$$\|\tilde{A}_{[n] \times J^c}\| \lesssim r^{3/2} \sqrt{c_\varepsilon n \ln \ln n}. \tag{5.5}$$

Repeating the same argument for the transpose, we have that

$$\|\tilde{A}_{J^c \times [n]}\| \leq \|\tilde{A}_{[n] \times J^c}\| \lesssim r^{3/2} \sqrt{c_\varepsilon n \ln \ln n}. \tag{5.6}$$

Now, we can combine (5.5) and (5.6) by triangle inequality for the operator norm to conclude the desired norm estimate for \tilde{A} on the intersection of good events, namely with probability

$$1 - \exp^{-\varepsilon n/4} - n^{-r} - \exp(-n^{0.5}) - 2C(\ln \ln n)n^{-r} \geq 1 - n^{0.1-r}.$$

Finally, let us check that the regularization is local. Again by Lemma 6, the total number of exceptional columns

$$|\bar{J}| = \left| \bigcup_l J_l \right| \leq \sum_{l \leq l_{\max}} n \exp(-2c_\varepsilon n 2^l \varepsilon/n) \leq n\varepsilon/4,$$

since we are summing a geometric progression and $l \geq 0$. Since the same argument holds for the cardinality of \bar{I} , we can conclude that with high probability, Algorithm 1 makes changes only in a $n\varepsilon \times n\varepsilon$ submatrix of A . This concludes the proof of Corollary 1.

6 Discussion

6.1 Regularization by the Individual Corrections of Entries

Do we actually need to look at the rows and columns of A ? A simpler and very intuitive idea would be to regularize the norm of A just by zeroing out a few large entries of A . However, this approach does not work for the case when the entries have only two finite moments: For the efficient local regularization, one has to account for the mutual positions of the entries in the matrix, not only for their absolute values.

Only in the case when A_{ij} have *more* than two finite moments, the truncation idea works and it is not hard to derive the following result from known bounds on random matrices such as [1, 14, 18] (see also the discussion in [13, Section 1.4]).

In the two moments case, individual correction of the entries can guarantee a bound with bigger additional factor $\ln n$ in the norm. It can be derived from known general bounds on random matrices, such as the matrix Bernstein's inequality [17]. One would apply the matrix Bernstein's inequality for the entries truncated at level \sqrt{n} to get that $\|\hat{A}\| \leq \varepsilon^{-1/2} \sqrt{n} \cdot \ln n$.

We consider Theorem 1 more advantageous with respect to the individual corrections approach not only because we are able to bound the norm closer to the optimal order \sqrt{n} , but also due to the fact that it gives more adequate information about the obstructions to the ideal norm bound. Namely, they are not only in the entries that are too large, but also in the rows and columns that accumulate too many entries (all of which are, potentially, of average size).

6.2 Symmetry Assumption

An assumption that the entries of A has to have symmetric distribution does not look natural and potentially can be avoided. We need it in the current argument to keep zero mean after various truncations by absolute value (in (4.1) and also in (3.4)). The standard symmetrization techniques (see [10, Lemma 6.3]) would not work in this case since we combine convex norm function with truncation (zeroing out of a product subset), which is not convex.

6.3 Dependence on n

Another potential improvement is an extra $\sqrt{\ln \ln n}$ factor on the optimal n -order $\|\hat{A}\| \sim \sqrt{n}$. The reason for its appearance in our proof is that we consider restrictions of A to the discretization “levels” independently, and independently estimate their norms. The second moment assumption gives us that $\sum 2^{2k} p_k \sim 1$. However, the best we can hope for a norm of one “level” (after proper regularization) is $2^k \sqrt{n p_k}$ (since this is an expected L_2 -norm of a restricted row). Thus, we end up summing square roots of the converging series, $\sum (2^{2k} p_k)^{1/2}$, which for some distributions is as large as square root of the number of summands ($\ln \ln n$ in our case).

It would be desirable to remove extra $\sqrt{\ln \ln n}$ term and symmetric distribution assumption, proving something like the following

Conjecture 1 Consider an $n \times n$ random matrix A with i.i.d. mean zero entries such that $\mathbb{E}A_{ij}^2 = 1$. Let \tilde{A} be the matrix that obtained from A by zeroing out all rows and columns such that

$$\|\text{row}_i(A)\|_m \geq C\mathbb{E}\|\text{row}_i(A)\|_m, \quad \|\text{col}_i(A)\|_m \geq C\mathbb{E}\|\text{col}_i(A)\|_m \quad (6.1)$$

for some L_m -norm to be specified (e.g., $m = 2$). Then with probability $1 - o(1)$, the operator norm satisfies $\|\tilde{A}\| \leq C'\sqrt{n}$.

Note that this result would be somewhat similar to the estimate proved by Seginer [14]: in expectation, the norm of the matrix with i.i.d. elements is bounded by the largest norm of its row or column. However, note that after cutting “heavy” rows and columns, we lose independence of the entries in the resulting matrix. And in general, the question of the norm regularization is not equivalent to another interesting question about the sufficient conditions on the distribution of the entries that ensure optimal order of the operator norm.

Acknowledgements The author is grateful to Roman Vershynin for the suggestion to look at the work of Feige and Ofek, helpful and encouraging discussions, as well as comments related to the presentation of the paper. The author is also grateful to Konstantin Tikhomirov for mentioning the idea to estimate quantiles of the entries distribution from their order statistics, which made Algorithm 1 more elegant.

References

1. Auffinger, A., Tang, S.: Extreme eigenvalues of sparse, heavy tailed random matrices. *Stoch. Process. Appl.* **126**(11), 3310–3330 (2016)
2. Bandeira, A.S., van Handel, R.: Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Prob.* **44**(4), 2479–2506 (2016)
3. Chin, P., Rao, A., Vu, V.: Stochastic block model and community detection in sparse graphs: a spectral algorithm with optimal rate of recovery. In: *Conference on Learning Theory*, pp. 391–423 (2015)
4. David, H.A., Nagaraja, H.N.: Order statistics. *Encycl. Stat. Sci.* **9**, 22 (2004)
5. Feige, U., Ofek, E.: Spectral techniques applied to sparse random graphs. *Random Struct. Algorithms* **27**(2), 251–275 (2005)
6. Friedman, J., Kahn, J., Szemerédi, E.: On the second eigenvalue of random regular graphs. In: *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, pp. 587–598 (1989)
7. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. *Comb. Prob. Comput.* **12**(1), 61–72 (2003)
8. Latała, R.: Some estimates of norms of random matrices. *Am. Math. Soc.* **133**(5), 1273–1282 (2005)
9. Le, C.M., Levina, E., Vershynin, R.: Concentration and regularization of random graphs. *Random Struct. Algorithms* **51**(3), 538–561 (2017)
10. Ledoux, M., Talagrand, M.: *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin (2013)
11. Litvak, A.E., Spector, S.: Quantitative version of a Silverstein’s result. In: Klartag, B., Milman, E. (eds.) *Geometric Aspects of Functional Analysis. Lecture Notes in Mathematics*, vol. 2116, pp. 335–340. Springer, Cham (2014)
12. Rebrova, E., Tikhomirov, K.: Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries. *Israel J. Math.* **227**(2), 507–544 (2018)
13. Rebrova, E., Vershynin, R.: Norms of random matrices: local and global problems. *Adv. Math.* **324**, 40–83 (2018)
14. Seginer, Y.: The expected norm of random matrices. *Comb. Prob. Comput.* **9**(2), 149–166 (2000)
15. Silverstein, J.W.: On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivar. Anal.* **30**(2), 307–311 (1989)

16. Tao, T.: Topics in Random Matrix Theory, vol. 132. American Mathematical Society, London (2012)
17. Tropp, J.A.: An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8**(1–2), 1–230 (2015)
18. van Handel, R.: On the spectral norm of Gaussian random matrices. *Trans. Am. Math. Soc.* **369**(11), 8161–8178 (2017)
19. Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices. In: *Compressed Sensing Theory and Applications*, pp. 210–268 (2012)
20. Vershynin, R.: *High-Dimensional Probability: An Introduction with Applications in Data Science*, vol. 47. Cambridge University Press, Cambridge (2018)
21. Yin, Y.-Q., Bai, Z.-D., Krishnaiah, P.R.: On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Prob. Theory Related Fields* **78**(4), 509–521 (1988)
22. Zieliński, R., Zieliński, W.: Best exact nonparametric confidence intervals for quantiles. *Statistics* **39**(1), 67–71 (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.