

RESEARCH STATEMENT

ELIZAVETA REBROVA

INTRODUCTION

My Ph.D. research is mostly devoted to the **non-asymptotic random matrix theory**.

Random matrix theory is a beautiful area where the tools of probability, functional analysis, linear algebra, and combinatorics can work together to study the structure and properties of the matrices taken from some probability distribution. Classical and powerful way to understand a structure of the random matrix is to look at the **matrix spectrum**: eigenvalues, eigenvectors, or singular values and vectors. The crucial observation that makes results of the theory possible is that the spectrum stabilizes as the size of the matrices (taken from some distribution over all the matrices) grows to infinity, so there are limit laws for the distribution of the spectrum. Among the classical laws are, for example, classical Wigner semicircular law for the limiting eigenvalue empirical measure distribution of random symmetric matrices, Tracy–Widom law for the limiting distribution of the largest eigenvalue, Marchenko–Pastur law for the sample covariance matrices (see, for example, [29, 2]).

But what if we would like to get some explicit high probability estimate for a large, but finite size matrix? This is what non-asymptotic branch of the theory is concerned with. One powerful tool that allows us to obtain high probability results is the **concentration of measure phenomenon**. It is a geometrical property of high dimensional measure spaces, which states – in its simplest form – that the majority of the volume of the unit sphere in \mathbb{R}^n (when n is large) lies in a thin strip near the equator. In more general versions, it implies enough similarity of the large objects to claim that almost all of them behave “like the expectation”. This idea bridges the study of random matrices with the **geometry of high dimensional spaces**, which is also important for my research, both as a tool and as a source of further interesting problems.

Geometric methods and explicit probability estimates for large finite matrices make this area useful in **application dealing with large high dimensional data objects**. Among the examples are graph and network analysis, statistics, compressed sensing and more. I am also interested in identifying and building such connections between random matrix theory and data applications.

1. INVERTIBILITY OF HEAVY-TAILED RANDOM MATRICES.

My research deals with the matrices having random elements in \mathbb{R} and their real spectrum, i.e. singular values and vectors. Singular values $s_{max} = s_1 \geq s_2 \geq \dots \geq s_n = s_{min}$ of a real $n \times n$ matrix A are the eigenvalues of the matrix $\sqrt{X^T X}$. The largest and the smallest singular values determine the basic geometric properties of A as a linear operator, namely, the norm of A and its inverse:

$$(1) \quad s_{max}(A) = \sup_{\|x\|_2=1} \|Ax\|_2 = \|A\|, \quad s_{min}(A) = \inf_{\|x\|_2=1} \|Ax\|_2 = 1/\|A^{-1}\|.$$

So, the study of the smallest singular value of the random matrix A can answer the questions like what is the probability that A is singular? If this probability is small (and thus A is likely to be invertible), how large is a typical operator norm of the inverse?

The invertibility problem was extensively studied from several angles, including the methods of geometric functional analysis, additive combinatorics, and mathematical physics.

For the square matrices with independent standard Gaussian entries, limiting distribution of the smallest singular value was computed in [7] and [27]: it was proved that for any $\varepsilon \in (0, 1)$

$$\mathbb{P}\{s_n(A) \leq \varepsilon n^{-1/2}\} \sim \varepsilon.$$

This result was later extended for more general distributions and improved in a number of papers, including [22, 31] (random sign matrices), [23] (finite fourth moment entries), [30, 32] (finite second moment entries and non-zero mean). In the paper [23] the authors also got much more precise probability estimate for the narrower class of subgaussian matrices. A random variable ξ is called *subgaussian*, if it has at least gaussian tail decay, i.e., there exists a number $K > 0$ such that

$$\mathbb{P}\{|\xi| > t\} \leq 2 \exp(-t^2/K^2), \quad \text{for all } t > 0.$$

In particular, it implies that all the moments of ξ are finite, $\mathbb{E}\xi^p \lesssim \sqrt{p}$ (see also [35]). In [23], Rudelson and Vershynin have shown that an $n \times n$ random matrix A with centered subgaussian random entries satisfies for every $\varepsilon > 0$

$$\mathbb{P}\{s_n(A) \leq \varepsilon n^{-1/2}\} \leq L\varepsilon + u^n,$$

where $L > 0$ and $u \in (0, 1)$ are absolute constants. Note that this implies linear probability decay with ε (like in the Gaussian case) until ε reaches an exponentially small probability level. The latter is necessary since singularity probability for the discrete matrices is non-zero and at least 0.5^n (for example, in the case of i.i.d. Bernoulli entries with $\mathbb{P}\{A_{ij} = 0\} \sim e^{-n}$).

My work with K. Tikhomirov [20] shows that exactly the same probability estimate holds for much more general class of *heavy-tailed matrices*, such that

$$(2) \quad A \text{ is } n \times n; \text{ the entries of } A \text{ are i.i.d., with } \mathbb{E}A_{ij} = 0, \quad \mathbb{E}A_{ij}^2 = 1.$$

Namely, we proved the following

Theorem 1.1. *Let $n \geq n_0$ and let A be a heavy-tailed matrix satisfying (2). Then for any $\varepsilon > 0$*

$$(3) \quad \mathbb{P}\{s_n(A) \leq \varepsilon n^{-1/2}\} \leq L\varepsilon + u^n,$$

where L and u are constants (independent from A and n , $L > 1$ and $u \in (0, 1)$).

1.1. Related questions and further work. Several natural questions related to the invertibility of random matrices arise from Theorem 1.1.

• **Optimal moment assumptions.** A crucial step of our proof (construction of the universal covering for ellipsoids $\{\|Ax\| : \|x\| = 1, A \text{ satisfies (2)}\}$) requires finite second moment assumption. But it is an open question to identify an optimal moment condition for the invertibility estimate (3) for square random matrices having i.i.d. elements.

For example, in the case of random rectangular matrices prior work [17, 24] was extended by K. Tikhomirov ([33]) to the matrices with no assumptions on moments at all. However, this argument works only for the matrices of the size $m \times n$, such that $m \leq \delta n$ for some constant $\delta < 1$.

• **Non-identically distributed entries.** We use the i.i.d. entries assumption in Theorem 1.1 several times, in particular, adopting from [23] the small ball probability estimate via LCD (least common denominator is a notion that measures the additive structure of the vector). One approach I am pursuing to relax identical distribution assumption is to generalize the notion of LCD of a vector to the several not identically distributed vectors, and prove a new small ball probability estimate with it. For example, it can be done in a simple partial case of several Bernoulli vectors.

Another approach I was recently looking at is employing Lindeberg's technique ([12]), which was recently used by Tropp and Oymak [18] to bound a restricted smallest singular value (minimum of $\|Ax\|$ over the subset of a unit sphere) of the matrices with no identical distribution assumed.

The lower bound for s_{min} for the matrices with the non-identically distributed entries will help, in particular, to improve probability estimates for the sparse random matrices (in continuation of the work [14]).

2. REGULARIZATION OF THE NORM.

Another branch of my work is connected with the study of the maximal singular value of a random matrix, which is, as we noted in (1), its operator norm. If we consider matrix A with the independent standard Gaussian entries, then by the classical Bai-Yin law (see, for example, [29])

$$s_1(A) \rightarrow 2\sqrt{n} \text{ almost surely,}$$

as the dimension $n \rightarrow \infty$. Moreover, the $2\sqrt{n}$ asymptotics holds for more general classes of matrices. By [4], if the entries of A have zero mean and bounded fourth moment, then

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

with high probability. In the non-asymptotic regime, an application of Bernstein's inequality (see, for example, in [35]) gives

$$\mathbb{P}\{s_1(A) \leq t\sqrt{n}\} \geq 1 - e^{-c_0 t^2 n} \quad \text{for } t \geq C_0$$

for the matrices with i.i.d. subgaussian entries. Here, $c_0, C_0 > 0$ are universal constants. The non-asymptotic extensions to more general distributions are also available, see [25, 13, 5, 9].

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.

We do not have ideal $O(\sqrt{n})$ -order in the heavy-tailed regime (2). It is suggested by the fact that weak fourth moment is necessary for the convergence in probability of $\|A\|/\sqrt{n}$ when n grows to infinity (see [28]). Moreover, an explicit family of examples, constructed in [15], shows heavy-tailed matrices A that have $\|A\| \sim O(n^\alpha)$ for any $\alpha \leq 1$ with substantial probability.

A natural question is: what are the obstructions in the structure of A that cause too large norm? Can we regularize the matrix restoring the “ideal norm”? Clearly, interesting regularization would be the one that does not change A too much.

So, the first question to answer is when we can enforce the “ideal” $\|A\|$ with high probability by modifying just the small fraction of the entries? Let us assume nothing about the distribution of the i.i.d. entries of A . Under what moment conditions the regularization of the matrix norm is a *local problem*?

We answered these questions in our work with R. Vershynin [21]. We have shown that local regularization (on a small submatrix of the matrix) is possible if and only if the entries of A have zero mean and unit variance:

Theorem 2.1 (Local problem). *Consider an $n \times n$ random matrix A satisfying (2), and let $\varepsilon \in (0, 1/6]$. Then, with probability at least $1 - 7 \exp(-\varepsilon n/12)$, there exists an $\varepsilon n \times \varepsilon n$ submatrix of A such that replacing all of its entries with zero leads to a well-bounded matrix \tilde{A} :*

$$\|\tilde{A}\| \leq \frac{C \ln \varepsilon^{-1}}{\sqrt{\varepsilon}} \cdot \sqrt{n},$$

where C is a sufficiently large absolute constant.

Theorem 2.2 (Global problem). *Consider an $n \times n$ random matrix A_n whose entries are i.i.d. copies of a random variable that has either nonzero mean or infinite second moment, and let $\varepsilon \in (0, 1)$. Then*

$$\min \frac{\|\tilde{A}_n\|}{\sqrt{n}} \rightarrow \infty \quad \text{as } n \rightarrow \infty$$

almost surely. Here the minimum is with respect to the matrices \tilde{A}_n obtained by any modification of any $\varepsilon n \times \varepsilon n$ submatrix of A_n .

Our proof utilizes the cut norm and Grothendieck-Pietsch factorization for matrices ([19, 1]), and it combines the methods developed in [20] and [16].

Returning to the question about the obstructions: if the local regularization is possible (that is, under the conditions of Theorem 2.1), what exactly causes the norm of a centered random matrix A to be too large? A natural guess is that the only troublemakers are a few large entries of A , and so we can obtain a result like Theorem 2.1 simply by zeroing them out.

This intuition turns out to be misleading. Only in the case when A_{ij} have *more* than two finite moments the truncation idea works and the following result ([21, Proposition 1.4]) can be quickly deduced from known bounds on random matrices such as [9, 25, 3]:

Proposition 2.3 ($2 + \varepsilon$ moments). *Let $\varepsilon \in (0, 1]$ and $n > n_0(\varepsilon)$. Consider any $n \times n$ random matrix A with i.i.d. mean zero entries which satisfy $\mathbb{E}|A_{ij}|^{2+\varepsilon} \leq 1$. With the probability at least $1 - 2\exp(-n^{\varepsilon/5})$, zeroing out at most $n^{1-\varepsilon/9}$ largest entries of A leads to the matrix \tilde{A} such that*

$$\|\tilde{A}\| \leq 9\sqrt{n}.$$

The heavy-tailed model (2) is qualitatively harder: to find a small submatrix, deletion of which regularizes the norm, one has to account for the mutual positions of the large elements of A .

2.1. Related questions and further work. There are several connected questions that are of special interest for me right now:

- **Constructive regularization.** Theorem 2.1 does not provide an answer how to find the small submatrix to be deleted – it is rather an existential result. It would be nice to have an explicit description of a submatrix to be removed.

The same source of inconstructiveness – Grothendieck-Pietsch factorization for matrices – appears earlier in [16]. There, for the case Bernoulli random matrices, it eventually does not disrupt the constructive nature of that regularization algorithm. My attempts to employ similar methods to the case of heavy-tailed matrices gave me only partial results so far ($\sqrt{\ln n}$ times weaker). My work in progress is to combine some ideas from [20] and [10] in the direction of the following

Conjecture 2.4. *Consider an $n \times n$ random matrix A satisfying (2). Let $R_i(A)$ and $T_i(A)$ denote the i -th row and column vector of the matrix A respectively. Let \tilde{A} be the matrix that obtained from A by zeroing out all rows and columns such that*

$$(4) \quad \|R_i(A)\|_1 \leq 10\mathbb{E}\|R_i(A)\|_1, \quad \|T_i(A)\|_1 \leq 10\mathbb{E}\|T_i(A)\|_1.$$

Then with probability $1 - o(1)$ the operator norm satisfies $\|\tilde{A}\| \leq C'\sqrt{n}$.

This would mean that to regularize the matrix, it is enough to delete “heavy” rows and columns. On the other hand, in [20] we construct a way to cut a small fraction of columns only to achieve (4). Thus, by a little more work, (4) can be obtained by a local modification in the small submatrix.

- **Dependence on ε .** The dependence on ε in Theorem 2.1 is best possible up to the $\ln \varepsilon^{-1}$ factor (to see this, one can consider $\frac{\sqrt{n}}{\sqrt{\varepsilon}}$ -Bernoulli example, or $\varepsilon \leq n^{-1}$ and any matrix A such that $\|A\| \sim n$). However, if the entries of A are $O(\sqrt{n})$ -bounded almost surely, then the dependence on ε is potentially much better:

Conjecture 2.5. *If A is the matrix satisfying (2) and $|A_{ij}| \leq \alpha\sqrt{n}$ a.s. Then, with probability $1 - o(1)$, there exists an $\varepsilon n \times \varepsilon n$ submatrix of A such that replacing all of its entries with zero leads to a well-bounded matrix \tilde{A} :*

$$\|\tilde{A}\| \leq C(\alpha) \ln(\varepsilon^{-1})\sqrt{n}.$$

A motivating case when Hypothesis 2.5 holds true is $n \times n$ Bernoulli zero-one matrices such that $np = \text{const}$, where $p := \mathbb{P}\{A_{ij} = 1\}$.

- **Intermediate singular values.** As it is discussed above, my current work focuses on the study of the smallest and largest singular values of heavy-tailed random matrices. In particular, it answers the question if s_{\min} and s_{\max} typically have the same order as in “ideal” gaussian case:

for $s_{min} = s_n$ the answer is “yes,” and for $s_{max} = s_1$ the answer is “no” (but local regularization is possible).

It would be interesting to know the answer for the other s_k , $1 < k < n$. The distribution of all s_k for “ideal” square matrices is known: $s_k \sim \frac{n+1-k}{\sqrt{n}}$ (see [26] for the Gaussian case, [24] and [37] imply the same estimate for subgaussian under an additional mild assumption). Is there a threshold k_0 , such that for $k > k_0$ the order of s_k stays the same for the heavy-tailed matrices, and for $k < k_0$ it is not? How does the size of the submatrix to be deleted depend on $k > 1$? I am planning to try this problem combining refined net and regularization techniques developed in [21, 20] with the projection techniques in the flavor of [37].

3. APPLICATIONS TO RANDOM GRAPHS.

One of my smaller projects is about regularization of the random graphs adjacency matrices. It is closely related to the work of Le, Levina and Vershynin [16] on inhomogeneous Erdős-Rényi random graphs¹. Often (for example, in community detection problems), the question of interest is to estimate some features of the probability matrix (p_{ij}) from random graphs drawn from $G(n, p_{ij})$. Concentration of the adjacency matrix A around its expectation matrix $\mathbb{E}A$, when it holds, guarantees that such features can be recovered.

However, this concentration holds only in the case when the graph is *dense* (i.e. maximal expected degree of the vertices $d := \max_{ij} np_{ij} > \log n$). For the *sparse* graphs, especially when expected degree is constant, the regularization question appears. Can we modify the graph on a small subset of its edges such that $\|A - \mathbb{E}A\|$ become well-bounded, and so concentration will be restored? And which exactly modification can help us?

An answer was given in the Feige and Ofek work [10]: with high probability, it is enough to delete all the edges adjacent to the “heavy” vertices (n/d vertices with the largest degrees). Le, Levina and Vershynin presented in [16] more general way to regularize these edges, as well as a new approach to the proof. Using similar techniques, I checked that it is actually enough to modify much smaller subgraph of the graph and also described the structure of this “bad” subgraph:

Proposition 3.1. *Let $G(n, p_{ij})$ be an inhomogeneous Erdős-Rényi random graph and $d := \max_{ij} np_{ij}$ be its maximal expected degree, $n > n_0$. Then with high probability we can numerate the vertices of G such that the adjacency matrix A has the following properties:*

- if \tilde{A} is obtained from A by zeroing out the entries of top left $s \times s$ submatrix A_0 , then

$$\|\tilde{A} - \mathbb{E}A\| \leq C\sqrt{d},$$

- A_0 is a very small square part of A , as its size $s \lesssim ne^{-d} \ll n/d$,
- A_0 has no more than 64 ones in every column above its main diagonal.

The last condition means that we can direct the edges inside the “bad” (too dense) subgraph such that every vertex will have a finite number of the outgoing edges.

3.1. Further questions. I would like to explore the connections of non-asymptotic random matrix theory with the networks analysis further (in flavor of works [10, 16, 6]). One question I was thinking about is how to prove the results similar to [16] in terms of $d_{ave} = \max_i \sum_j p_{ij}$ instead of d . This would allow to apply the concentration results to the networks having certain patterns (some edges are always present whereas some others are unlikely).

REFERENCES

- [1] N. Alon, A. Naor, *Approximating the cut-norm via Grothendieck’s inequality*. SIAM Journal on Computing 35, no. 4 (2006): 787–803.

¹Inhomogeneous Erdős-Rényi random graph $G(n, p_{ij})$ is a graph with n vertices, such that the edge between i -th and j -th vertices is present with probability p_{ij}

- [2] G. W. Anderson, A. Guionnet, O. Zeitouni, *An introduction to random matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press (2010).
- [3] A. Auffinger, S. Tang, *Extreme eigenvalues of sparse, heavy tailed random matrices*. Stochastic Processes and their Applications 126, no. 11 (2016): 3310–3330.
- [4] Z. D. Bai, P. R. Krishnaiah, Y. Q. Yin, *On limit of the largest eigenvalue of the large dimensional sample covariance matrix*. Probability Theory and Related Fields 78 (1988): 509–521.
- [5] A. S. Bandeira, R. van Handel, *Sharp nonasymptotic bounds on the norm of random matrices with independent entries*. The Annals of Probability 44, no. 4 (2016): 2479–2506.
- [6] P. Chin, A. Rao, V. Vu, *Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery*. In Conference on Learning Theory (2015): 391–423.
- [7] A. Edelman, *Eigenvalues and condition numbers of random matrices*. SIAM Journal on Matrix Analysis and Applications 9, no. 4 (1988): 543–560.
- [8] L. Erdos, B. Schlein, H. T. Yau, *Wegner estimate and level repulsion for Wigner random matrices*. Int. Math. Res. Not. 3 (2010): 436–479.
- [9] R. van Handel, *On the spectral norm of Gaussian random matrices*. Transactions of the American Mathematical Society 369, no. 11 (2017): 8161–8178.
- [10] U. Feige, E. Ofek, *Spectral techniques applied to sparse random graphs*. Random Structures and Algorithms 27, no. 2 (2005): 251–275.
- [11] M. Krivelevich, B. Sudakov, *The largest eigenvalue of sparse random graphs*. Combinatorics, Probability and Computing 12, no. 1 (2003): 61–72.
- [12] J. W. Lindeberg, *Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung*. Math. Z. 15 (1922): 211–225.
- [13] R. Latała, *Some estimates of norms of random matrices*. Proc. Am. Math. Soc. 133, no. 5 (2005): 1273–1282.
- [14] A. E. Litvak, O. Rivasplata, *Smallest singular value of sparse random matrices*. arXiv:1106.0938 (2011).
- [15] A. E. Litvak, S. Spektor, *Quantitative version of a Silverstein’s result*. Geometric Aspects of Functional Analysis, Springer International Publishing (2014): 335–340.
- [16] C. Le, E. Levina, R. Vershynin, *Concentration and regularization of random graphs*. Random Structures and Algorithms, to appear.
- [17] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*. Advances in Mathematics 195, no. 2 (2005): 491–523.
- [18] S. Oymak, J. A. Tropp, *Universality laws for randomized dimension reduction, with applications*. arXiv:1511.09433 (2015).
- [19] A. Pietsch, *Operator Ideals*. Deutscher Verlag der Wissenschaften (1978).
- [20] E. Rebrova, K. Tikhomirov, *Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries*. Israel Journal of Mathematics, to appear.
- [21] E. Rebrova, R. Vershynin, *Norms of random matrices: local and global problems*. Adv. Math., to appear.
- [22] M. Rudelson, *Invertibility of random matrices: norm of the inverse*. Annals of Mathematics (2008): 575–600.
- [23] M. Rudelson, R. Vershynin, *The Littlewood–Offord problem and invertibility of random matrices*. Advances in Mathematics 218, no. 2 (2008): 600–633.
- [24] M. Rudelson, R. Vershynin, *Smallest singular value of a random rectangular matrix*. Communications on Pure and Applied Mathematics 62 (2009): 1707–1739.
- [25] Y. Seginer, *The expected norm of random matrices*. Comb., Prob. and Comp. 9, no. 2 (2000): 149–166.
- [26] S. J. Szarek, *Spaces with large distance to l_∞^n and random matrices*. Am. J. of Math., 112 (1990): 899–942.
- [27] S. J. Szarek. *Condition numbers of random matrices*. J. Complexity 7(2) (1991): 131–149.
- [28] J. W. Silverstein, *On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix*. Journal of Multivariate Analysis 30, no. 2 (1989): 307–311.
- [29] T. Tao, *Topics in random matrix theory*. Vol. 132. Providence, RI: American Mathematical Society, 2012.
- [30] T. Tao, V. Vu, *Smooth analysis of the condition number and the least singular value*. Mathematics of Computation 79 (2010): 2333–2352.
- [31] T. Tao, V. Vu, *Inverse Littlewood–Offord theorems and the condition number of random discrete matrices*. Annals of Mathematics (2009): 595–632.
- [32] T. Tao, V. Vu, *Random matrices: the circular law*. Comm. in Cont. Math. 10, no. 02 (2008): 261–307.
- [33] K. E. Tikhomirov, *The smallest singular value of random rectangular matrices with no moment assumptions on entries*. Israel Journal of Mathematics 212, no. 1 (2016): 289–314.
- [34] R. Vershynin, *Invertibility of symmetric random matrices*. Rand. Struct. and Alg. 44 (2014): 135–182.
- [35] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*. Compressed sensing, Cambridge University Press (2012): 210–268.
- [36] R. Vershynin, *High-Dimensional Probability*. Cambridge University Press, to appear.
- [37] W. Feng, *Upper bound for intermediate singular values of random matrices*. Journal of Mathematical Analysis and Applications 445, no. 2 (2017): 1530–1547.