The smallest singular value of a shifted d-regular random square matrix

Alexander E. Litvak Anna Lytova Konstantin Tikhomirov
Nicole Tomczak-Jaegermann Pierre Youssef

Abstract

We derive a lower bound on the smallest singular value of a random d-regular matrix, that is, the adjacency matrix of a random d-regular directed graph. More precisely, let $C_1 < d < c_1 n/\log^2 n$ and let $\mathcal{M}_{n,d}$ be the set of all 0/1-valued square $n \times n$ matrices such that each row and each column of a matrix $M \in \mathcal{M}_{n,d}$ has exactly d ones. Let M be uniformly distributed on $\mathcal{M}_{n,d}$. Then the smallest singular value $s_n(M)$ of M is greater than $c_2 n^{-6}$ with probability at least $1 - C_2 \log^2 d/\sqrt{d}$, where c_1, c_2, C_1, and C_2 are absolute positive constants independent of any other parameters.

AMS 2010 Classification: primary: 60B20, 15B52, 46B06, 05C80; secondary: 46B09, 60C05
Keywords: Adjacency matrices, anti-concentration, condition number, invertibility, Littlewood–Offord theory, random graphs, random matrices, regular graphs, singular probability, singularity, sparse matrices, smallest singular value

Contents

1 Introduction

The present paper concentrates on the studies of a sub-area of the random matrix theory nowadays often called *non-limiting* or *non-asymptotic* (see e.g. [22, 34]). The development of this direction of research was stimulated by some problems in statistics, compressed sensing and computer science in general, and in asymptotic geometric analysis. The object of the study is a large random matrix of a fixed size, and the typical goal is to obtain quantitative probabilistic estimates of its eigenvalues or singular values in terms of dimensions of the matrix. In this paper we avoid a detailed discussion of corresponding *limiting* results, and refer, in particular, to books [3, 14] and references therein for more information (see also [12] for interplay between limiting and non-limiting results and for applications).

The study of the non-limiting behaviour of the smallest and the largest singular values are quite important in random matrix theory. Recall that for an $n \times m$ ($m \geq n$) matrix A, the largest and the smallest singular values can be defined as

$$s_1(A) = \|A\| = \max_{\|z\|_2=1} \|Az\|_2 \quad \text{and} \quad s_n(A) = \min_{\|z\|_2=1} \|Az\|_2,$$

where $\|A\|$ denotes the operator norm of A acting from ℓ_2^n to ℓ_2^m (also called the spectral norm). In case when $m = n$ and the matrix A is invertible, we have $s_n(A) = 1/\|A^{-1}\|$. The knowledge of the magnitude of the extreme singular values gained significance in connection with asymptotic geometric analysis, the problem of approximation of covariance matrices and signal reconstruction. Moreover, for square matrices, it is also important as a crucial step in computing the limit of empirical spectral distributions as well as the condition number of a matrix. We provide a brief overview of each of the directions.

First, assume that A is a tall rectangular matrix with independent rows (satisfying certain conditions). Estimating $s_1(A)$ can be quite difficult (excluding the subgaussian case, see, for example, [25, Fact 2.4]). The lower bounds for $s_n(A)$ often requires covering arguments, estimates for small ball probabilities, anti-concentration results, and on many occasions bounds on $s_1(A)$. For bounds on $s_1(A)$ and $s_n(A)$, we refer to [1, 27, 17, 44] and references therein. We would like to notice that strong estimates for $s_n(A)$ for this model can be obtained bypassing analysis of $s_1(A)$, and under very weak conditions on the distributions of the rows [21, 30, 48, 49] (see also [16] for related yet different setting).

Another model of randomness, which is closer to the main topic of our paper, involves square random matrices or matrices with the aspect ratio m/n very close to one, with i.i.d. entries. In this setting, obtaining optimal quantitative lower bounds for $s_n(A)$ requires more delicate arguments, compared to the model considered above. We refer, in particular, to [25, 41, 32, 33, 39, 31] and references therein (see also [2] for square matrices with independent log-concave columns). In the context of numerical linear algebra, this research direction is directly related to estimating the condition number of a square matrix. Recall that the condition number of an $n \times n$ matrix A is defined as

$$\sigma(A) = s_1(A)/s_n(A) = \|A\| \, \|A^{-1}\|.$$

The condition number serves as a measure of precision of certain matrix algorithms [6, Chapter III], [37]. The study of the condition number in the random setting goes back to von Neumann and his collaborators (see [28, pp. 14, 477, 555] and [29, Section 7.8]), whose numerical experiments suggested that for a random $n \times n$ matrix A one should have $\sigma(A) \approx n$ with high probability. In a more general context, when the spectral norm $\|\cdot\|$ is replaced with an operator norm $\|\cdot\|_{X\to Y}$ for two n-dimensional Banach spaces X and Y, the quantity $\|A\|_{X\to Y}\|A^{-1}\|_{Y\to X}$ plays a crucial role in the local theory of Banach spaces and asymptotic geometric analysis through its relation to the Banach–Mazur distance [8, 45]. Estimating the condition number of a shifted matrix $A+B$ (with A random and B fixed) was put forward as an important problem by Spielman and Teng [38], in context of smooth analysis of algorithms (see, in particular, [35, 40, 42]). As a very important application, the quantitative lower bounds for $s_n(A+B)$, with B being a complex multiple of the identity, have been used to establish the circular law for the empirical spectral distribution in the i.i.d. model (see [43, 7] and references therein for the historical account of the problem). Indeed, it is known that using the Hermitization technique, one needs to show the uniform integrability of the logarithmic potential with respect to the empirical singular value distribution of the shifted matrix. Bounding the smallest singular value away from zero is therefore essential for such method to work. As the limiting distribution is not the aim of this paper and since the uniform integrability requires also a control of the remaining singular values, we leave this for a future investigation.

The model studied in this paper differs from the ones discussed above in two crucial aspects. Let us set up the framework. Let $d \le n$ be (large) integers, which we assume to be fixed throughout the paper. Consider the set $\mathcal{M}_{n,d}$ of $0/1$-valued square $n \times n$ matrices such that each row and each column of a matrix $M \in \mathcal{M}_{n,d}$ contains exactly d ones. Such matrices will be called d-regular. These are adjacency matrices of d-regular digraphs (directed graphs), where we allow loops but do not allow multiple edges. On $\mathcal{M}_{n,d}$ we take the uniform probability measure, turning $\mathcal{M}_{n,d}$ into a probability space, and consider the random matrix distributed according to this measure. The two main differences from the models mentioned in the previous paragraphs are complex dependencies between the matrix entries and (for $d \ll n$) sparsity of the matrix, i.e., large number of zero entries. Note also that by the Perron–Frobenius theorem for every $M \in \mathcal{M}_{n,d}$ one has $\|M\| = d$.

The question of estimating $s_n(M)$ (or, more generally, $s_n(M+B)$ for a fixed matrix B), where M is uniformly distributed in $\mathcal{M}_{n,d}$, can be justified in two respects. First, this is a natural model with complex dependencies between the matrix entries, which does not allow the use of standard conditioning arguments (such as fixing the span of $n-1$ rows of a random matrix and studying the conditional distribution of the distance of the remaining row to the span). Techniques developed for treating this model can potentially be adapted to more general models with dependencies. Second, as we show in this paper, unlike the Erdős–Renyi random model (see below for the definition and a more detailed comparison), the d-regularity condition guarantees strong lower bounds on $s_n(M)$ with large probability even in the case when $d \ll \log n$ when the corresponding Erdős–Renyi adjacency random matrix with the parameter $p = d/n$ is singular with large probability. This provides a better understanding as to what causes singularity of sparse random matrices ("local" obstructions to invertibility such as a zero row in the Erdős–Renyi model versus "global" obstructions when the non-trivial null vectors have many non-zero components).

Singularity of adjacency matrices of uniform random d-regular digraphs was first considered by Cook in [10]. He adapted to the case of directed graphs a conjecture of Vu from [46, Problem 8.4] which asserted that for $3 \leq d \leq n-3$ with probability going to 1 as n goes to infinity the adjacency matrix of a random d-regular undirected graph is non singular (see also 2014 ICM talks by Frieze [15, Problem 7] and Vu [47, Conjecture 5.8]). The argument in [10] was based on discrepancy properties of random digraphs studied in [9], together with some anti-concentration arguments and a sophisticated use of the simple switching operation. It established non-singularity of the adjacency matrix with a large probability for d growing to infinity faster than $\log^2 n$.

The question about singularity of adjacency matrices in the case $d \leq \log n$ remained open, moreover it was not clear whether the condition $d \gg \log n$ comes from limitations of the method used in [10] or if a random matrix uniformly distributed on $\mathcal{M}_{n,d}$ becomes singular in this regime. As we mentioned above, in the Erdős–Rényi model, a random matrix is singular with probability close to one in the case $d \ll \log n$. In [23] (see also [24]), the authors of the present paper were able to partially answer this question by showing that a random d-regular matrix is non-singular for all d bigger than a large universal constant (the probability of the singularity was estimated from above by a negative power of d). The main novelty of [23] compared to [10] rested on three new ingredients – a particular version of the covering argument which is applied to study the structure of the kernel of random matrices, on a different set of properties of random digraphs, and on a new approach to anti-concentration results.

However, both papers [10] and [23] didn't provide any quantitative estimates. Combining methods from [10] and [23] with an elaborate chaining argument, in recent papers [11] and [4], quantitative lower bounds on the smallest singular value of the adjacency matrix were proved for the uniform and permutation models, under an assumption that d is polylogarithmic in n. Moreover, considering shifted adjacency matrices, the authors of [11, 4] were able to obtain the circular law for the eigenvalue distribution (again, for d at least polylogarithmic in n). Precisely, in [11, 4] it was shown that, with some conditions on the shift W, the smallest singular value $s_n(M+W)$ of a random shifted matrix is at least $n^{-C\log_d n}$ with probability close to one. Still papers [11, 4] do not provide any bounds for s_n when d is growing slower than $\log n$ and moreover, even for d growing faster than $\log n$ but subpolynomial in n, they don't provide a polynomial in n bound for s_n.

The goal of the present paper is to provide polynomial in n lower bounds on the smallest singular value of a random matrix uniformly distributed on $\mathcal{M}_{n,d}$ for d larger than a (fixed large) absolute constant. Our approach results in better bounds not only for small d but for the entire range $C \leq d \leq cn/\log^2 n$. Our main result is the following theorem, in which we also allow shifts of random matrices for the sake of future applications (see also Remark 4.9 for more precise bounds).

Theorem 1.1. *There are universal constants $C, c > 0$ with the following property. Let $C < d < cn/((\log n)(\log\log n))$. Then for every $z \in \mathbb{C}$ with $|z| \leq d/6$ one has*

$$\mathbb{P}\Big\{M \in \mathcal{M}_{n,d} : \, s_n(M - z\mathrm{Id}) \geq n^{-6}\Big\} \geq 1 - C\log^2 d/\sqrt{d}.$$

It is natural to compare our model with the Erdős–Rényi model, i.e. matrices whose elements are i.i.d. Bernoulli 0/1 variables with the expectation d/n. Intuitively one would expect that d-regular matrices should behave in a similar way to the Erdős–Rényi model. This in turn seems to be similar (after applying a proper normalization $\sqrt{d/n}$) to random ± 1 matrices, where values 1 and -1 appears with probability 1/2. Since for the latter model one has $s_n \approx 1/\sqrt{n}$, we would expect the answer $s_n \approx \sqrt{d}/n$ for both d-regular matrices and for the Erdős–Rényi model. Indeed, the Erdős–Rényi model was recently treated in [5], where it was proved that with high probability $s_n \geq c(d/n)^{-c/\log d}\sqrt{d}/n$, provided that $c \log n \leq d \leq n - c \log n$. Note that if d is polynomial in n then this gives the expected bound $\sqrt{d}/n$. However, there is one delicate point in such a comparison. It is easy to see that for $d < \log n$ a matrix in the Erdős–Rényi model has a zero row with probability more than half, therefore at more than half of matrices in this model are singular. To the contrary, our theorem shows that in the case of d-regular matrices most matrices are non-singular. In particular, this means that the regularity prevents a matrix from being singular, in a sense reducing the randomness.

The remaining part of the introduction is devoted to a brief description of main ideas and to a short overview of the proof of Theorem 1.1. It is well understood by now that in order to estimate the smallest singular value, in other words to show that for every non-zero $x \in \mathbb{C}^n$ the ratio $\|Mx\|_2/\|x\|_2$ is bounded from below, one needs to split $\mathbb{C}^n$ and work separately with different types of vectors. The idea to split the Euclidean sphere into two parts goes back to Kashin's work [19] on an orthogonal decomposition of ℓ_1^{2n}, where the splitting was defined using the ratio of ℓ_2- and ℓ_1-norms. A similar idea was used by Schechtman [36] in the same context. In the context of the smallest singular value one usually splits $\mathbb{C}^n$ into vectors of smaller complexity (close to sparse vectors) and "spread" vectors (in particular, with a relatively small ℓ_∞-norm). Such a splitting was introduced in [25] (see also [26]) and was further formalized later in [32] into a concept of "compressible" and "incompressible" vectors in $\mathbb{C}^n$. For the spread (incompressible) vectors very good individual probability bounds are available. Using standard ε-nets argument and the union bound one can obtain good bounds for $\|Mx\|_2/\|x\|_2$ with high probability. For compressible vectors bounds on individual probabilities are not so good, but those vectors are essentially vectors of smaller dimension, so they have much smaller complexity. Therefore one can pass to ε-nets of much smaller cardinality and still apply the union bound.

In our model, due to special structure of the matrices (in particular, due to the lack of independence and sparsity) the concept of compressible and incompressible vectors is not applicable. We consider a new splitting of $\mathbb{C}^n$ into three parts. We would like to note that in [41] as well as in recent works [5, 10] the authors also had to split the sphere in three parts, however our splitting is different. To introduce it, we first define four (overlapping) classes on $\mathbb{C}^n$, which we call *steep vectors*, *sloping vectors* (that is non-steep), the *almost constant vectors*, and the *essentially non-constant vectors*. Roughly speaking, almost constant vectors are those with many coordinates almost equal to each other. The sloping vectors are vectors $x = (x_i)_i \in \mathbb{C}^n$, whose sequence $(x_i^*)_i$ (a non-increasing rearrangement of $(|x_i|)_i$) has a regular decay, i.e. has no significant jumps, where by *a jump* we mean $x_k^* \gg x_m^*$ for some $k \ll m$. The steep vectors are vectors possessing such jumps. Then we split $\mathbb{C}^n$ into essentially non-constant vectors, almost constant steep vectors, and almost constant sloping vectors and work with each part separately. The steep and sloping vectors play a similar role to the compressible and incompressible vectors respectively in the previous models, although there is no direct connection between these two splitting of the space. In fact, the compressible vectors are less complex than the steep vectors, hence in earlier papers proofs for compressible vectors were less involved than in our case. In particular, for compressible vectors it is enough to use the concentration only, while in our treatment of steep vectors we need to use the anti-concentration technique as well (see Lemma 3.9).

We now discuss steep vectors in more details. Our proof uses a new chaining argument and a very delicate construction of ε-nets. We split a vector in pieces and check if a jump occurs inside those pieces. For each particular piece with a jump, the main idea is to use the union bound, that is, to estimate the probability for an individual vector with a jump, to construct a good ε-net for such vectors, and to approximate each such vector by a vector from the ε-net. In this scheme the most important is to have the "right" balance between the size of the net and the individual probability bound. For individual probability bounds we use anti-concentration type technique together with switching argument, standard in dealing with d-regular graphs. Jumps are needed to apply anti-concentration and to show that, for a fixed vector x and a fixed index i, matrices having small inner product of row i with x belong to a certain class, to which we can apply the switching argument. For this consideration a constant jump, that is $x_k^* > 4x_m^*$ would be enough. Note that the smaller the jump and the larger the ratio m/k the better for us, since we need to have a control of the ratio x_1^*/x_m^*, which is responsible for both, for the final bound on the singular value and for the size of the net. Note also that contrary to results for matrices with i.i.d. entries we have to employ anti-concentration inequalities already for these vectors of relatively small complexity. To have a reasonable size of the net, we also work with pieces of a vector and approximate each piece separately. This delicate construction allows us to significantly decrease the size of the net (in comparison with the standard constructions). Unfortunately, the size of the net is still quite large and requires additional restrictions. First, it works only when $k \gtrsim n/d^2$ (of course, this always holds for $d^2 > n$) and $m/k \le d/\log d$. Second, in the case $n/d^2 \lesssim k \lesssim n/d^{3/2}$, to kill a large part of coordinates (in order to decrease the size of the net) we need a jump of order $d^{3/2}$. For the part of coordinates with $k \approx n/d$ and $m \approx n/\log d$ even such a big jump is not enough. But, here we intersect steep vectors with almost constant vectors. This essentially reduces the "dimension" of the vector and allows good bounds on the size of the net even with a constant jump. It remains to treat the case when a jump occurs in the first part of the vector, that is when $m \lesssim n/d^2$ (and, thus, $d^2 < n$). For this case we don't have a good balance between probability estimates and the size of the nets and therefore we have to force the bound by a large jump, so the proof in this case is more deterministic and does not require an approximation – for every "good" matrix we have a good uniform bound on vectors having a large jump. More precisely, for such vectors we use properties of d-regular graphs and their adjacency matrices, which we obtained in [23]. Using these properties, we prove that with high probability a random d-regular matrix has many rows with only one 1 in columns corresponding to the first k coordinates and no other ones till the m-th column. Thus, the inner product of such a row with x can be bounded as difference of the absolute value of one "large" coordinate and the sum of absolute values of $d-1$ "small" coordinates. Therefore, if we have a jump of order, say, $2d$, this inner product is separated from zero. In fact, we will be using a jump of order

$4d$ to "kill" shifts. This works when $m/k \lesssim \sqrt{d/\log d}$. We don't apply this scheme to the entire vector, because, first, the ratio between jump and m/k is relatively big, which affects the bound for the smallest singular value, and second, to keep this scheme for larger k, m we would have to significantly increase such a ratio.

The proof for almost constant sloping vectors is straightforward. First, since the vector is sloping, we have a control of its ℓ_2 norm in terms of x_m^* with $m \approx n/\log d$. Then, employing properties of random d-graphs again, we show that there are many rows for which most of support lies on the "almost constant" part of the vector. Therefore, since the vector is sloping, inner product of such rows with the vector is separated from zero.

After we obtain bounds for those two classes it remains to deal with essentially non-constant vectors. Since we have already found good lower bounds on the ratio $\|Mx\|_2/\|x\|_2$ in the case of other vectors we can restrict ourself only to the case of matrices having this ratio small if and only if x is an essentially non-constant vector. First, using general algebraic properties of square matrices we reduce the problem of estimating the smallest singular number to estimating distances between rows (or columns) of the matrix and certain subspaces (similar reduction was done in [32]). Then, we show that such distances for a fixed row can be estimated via the inner product of this row and a certain vector (in fact, we consider vectors at which a matrix attains its smallest singular value). Then for such a vector (note, due to our restriction this vector is essentially non-constant), we show that for most pairs of rows it is also essentially non-constant when restricted to the support of those two rows (which is almost $2d$, but still much smaller than n). This step requires two properties of random d-regular digraphs which we proved in [23]. They state that with high probability a random matrix drawn from $\mathcal{M}_{n,d}$ has no large zero minors and that the intersection of the supports of any two rows is very small. Then we choose two rows, say the first one and the second one, and split $\mathcal{M}_{n,d}$ into equivalence classes of matrices such that in every class all matrices have the same minors obtained by removing the first two rows (that is, any two matrices from a given class have the same rows starting with the third one). Thus, restricting ourself to a fixed class, we can "play" with the two first rows only (note that their sum is fixed). Then we show that on such a class the inner product of the first row with a vector can be treated as a sum of independent random variables, to which anti-concentration inequalities can be applied. At this step we use that the vector is essentially non-constant. This concluding part is in a sense similar to ideas of [23], however our technique here is more delicate, since we need to obtain quantitative estimates.

Acknowledgement. The second and the third named authors would like to thank University of Alberta for excellent working conditions in January–August 2016, when a significant part of this work has been done.

2 Preliminaries

By "universal" or "absolute" constants we always mean numbers independent of all involved parameters, in particular independent of d and n. When we say that a parameter (or a constant) is sufficiently large (resp. sufficiently small) it means that there exist an absolute positive constant such that the corresponding statement or inequality holds whenever the parameter is larger (resp. smaller) than this absolute constant. Given positive integers $\ell < k$, we denote the sets $\{1, 2, \ldots, \ell\}$ and $\{\ell, \ell+1, \ldots, k\}$ by $[\ell]$ and $[\ell, k]$, respectively. For any two real-valued functions f and g we write $f \approx g$ if there are two absolute positive constants c and C such that $cf \le g \le Cf$. By Id we denote the $n \times n$ identity matrix. For $I \subset [n]$, we denote by $I^c := [n] \setminus I$ the complement of I in $[n]$, and by P_I we denote the orthogonal projection on the coordinate subspace $\mathbb{C}^I$. For a vector $x \in \mathbb{C}^n$, we denote its coordinates by x_i, $i \le n$, its ℓ_∞-norm by $\|x\|_\infty = \max_i |x_i|$ and its ℓ_2-norm by $\|x\|_2$. The unit ball of the complex space ℓ_∞^n is denoted by B_∞^n. For a linear operator U from (complex) ℓ_2 to ℓ_2, by $\|U\|$ we denote its operator norm. Given a vector $x = (x_i)_i^n \in \mathbb{C}^n$, by $(x_i^*)_i^n$ we denote the non-increasing rearrangement of the sequence $(|x_i|)_i^n$. We don't distinguish between column and row vectors and, given $x = (x_i)_i^n \in \mathbb{C}^n$ we let $x^\dagger := (\bar{x}_i)_i^n \in \mathbb{C}^n$. For an $n \times n$ complex matrix U by $U^\dagger$ we denote the dual matrix (in the operator sense, i.e. obtained by taking transpose and complex conjugate).

We will use the following anti-concentration Littlewood–Offord type lemma ([13], see also [20]).

Proposition 2.1. *Let ξ_1, ξ_2, ..., ξ_m be independent ± 1 Bernoulli random variables and let x_1, x_2, ..., x_m be complex numbers such that $|x_i| \ge 1$, $i \le m$. Then for every $t \ge 1$ one has*

$$\sup_{a \in \mathbb{C}} \mathbb{P}\Big(\big|\sum_{i=1}^m \xi_i x_i - a\big| < t\Big) \le \frac{C_{2.1}\, t}{\sqrt{m}},$$

where $C_{2.1} > 0$ is a universal constant.

The next lemma is a "quantified" version of Claim 4.7 from [23].

Lemma 2.2. *Let $x = (x_1, x_2, \ldots, x_m) \in \mathbb{C}^m$ be a vector such that for some $\rho > 0$ and $\varepsilon \in (0,1)$ we have*

$$\forall \lambda \in \mathbb{C} \quad \big|\big\{i \le m : \, |x_i - \lambda| \ge \rho\big\}\big| \ge \varepsilon m.$$

Then there are disjoint subsets J and Q of $[m]$ such that

$$|J|, |Q| \ge \varepsilon m/4 \quad \text{and} \quad \forall i \in J, \forall j \in Q \ \ |x_i - x_j| \ge \rho/\sqrt{2}.$$

Proof. Let $y^1 := \mathrm{Re}(x)$ and $y^2 := \mathrm{Im}(x)$ be the real and imaginary part of x, respectively. First, observe that there is $k \in \{1, 2\}$ such that

$$\forall \lambda \in \mathbb{R} \quad \big|\big\{i \le m : \, |y_i^k - \lambda| \ge \rho/\sqrt{2}\big\}\big| \ge \varepsilon m/2. \tag{1}$$

Indeed, assume the opposite, i.e. there are real numbers λ_1 and λ_2 such that

$$\big|\big\{i \le m : \, |y_i^k - \lambda_k| \ge \rho/\sqrt{2}\big\}\big| < \varepsilon m/2, \quad k = 1, 2.$$

Then for $\lambda := \lambda_1 + i\,\lambda_2$ we necessarily have

$$\big|\big\{i \le m : \, |x_i - \lambda| \ge \rho\big\}\big| < \varepsilon m,$$

contradicting the assumption of the lemma.

Without loss of generality, we can assume that condition (1) holds for $k = 1$, and that the coordinates of y^1 are arranged in the non-increasing order. Denote $p := \lceil \varepsilon m/4 \rceil$. Set $J := \{1, 2, \dots, p\}$ and $Q := \{m - p + 1, \dots, m\}$. Clearly, it is enough to show that $y^1_p \geq \rho/\sqrt{2} + y^1_{m-p+1}$. Assume the opposite. Then the set $I := \{p, \dots, m - p + 1\}$ has cardinality strictly greater than $m - \varepsilon m/2$, and for $\lambda := y^1_p$ we have $|y^1_i - \lambda| < \rho/\sqrt{2}$ for all $i \in I$ contradicting (1). The result follows. $\square$

We will need the following simple combinatorial claim about relations. Let A, B be sets, and $R \subset A \times B$ be a relation. Given $a \in A$ and $b \in B$, the image of a and preimage of b are defined by

$$R(a) = \{y \in B \,:\, (a, y) \in R\} \quad \text{and} \quad R^{-1}(b) = \{x \in A \,:\, (x, b) \in R\}.$$

We also set $R(A) = \cup_{a \in A} R(a)$. We have the following standard estimate (see e.g. Claim 2.1 in [23]).

Claim 2.3. *Let $s, t > 0$. Let R be a relation between two finite sets A and B such that for every $a \in A$ and every $b \in B$ one has $|R(a)| \geq s$ and $|R^{-1}(b)| \leq t$. Then*

$$s|A| \leq t|B|.$$

We turn now to properties of d-regular matrices. Recall that $\mathcal{M}_{n,d}$ denotes the set of all $n \times n$ 0/1-valued matrices having sums of elements in every row and in every column equal to d (the set corresponds to adjacency matrices of directed d-regular graphs where we allow loops but do not allow multiple edges). For every $n \times n$ matrix U its i'th row is denoted by $R_i(U)$ and its i'th column is denoted by $X_i(U)$. Let $M = \{\mu_{ij}\} \in \mathcal{M}_{n,d}$. For $j \leq n$, we denote $\operatorname{supp} X_j(M) = \{i \leq n \,:\, \mu_{ij} = 1\}$ and for every subset $J \subset [n]$ we let

$$S_J := \{i \leq n \,:\, \operatorname{supp} R_i(M) \cap J \neq \emptyset\} = \bigcup_{j \in J} \operatorname{supp} X_j(M). \tag{2}$$

Given $k \leq n$ and $\varepsilon \in (0, 1)$, let

$$\Omega_{k,\varepsilon} = \Big\{ M \in \mathcal{M}_{n,d} \,:\, \forall J \subset [n] \;\text{ with }\; |J| = k \;\text{ one has }\; |S_J| \geq (1 - \varepsilon) dk \Big\}.$$

Clearly, if $k = 1$ then $\Omega_{k,\varepsilon} = \mathcal{M}_{n,d}$. The following theorem is essentially Theorem 2.2 of [23] (see also Theorem 3.1 there).

Theorem 2.4. *Let $e^8 < d \leq n$, $\varepsilon_0 = \sqrt{\log d/d}$, and $\varepsilon \in [\varepsilon_0, 1)$. Let $k \leq c_{2.4} \varepsilon n/d$, where $c_{2.4} \in (0, 1)$ is a sufficiently small absolute positive constant. Then*

$$\mathbb{P}(\Omega_{k,\varepsilon}) \geq 1 - \exp\left(-\frac{\varepsilon^2 dk}{8} \log\left(\frac{e\varepsilon c_{2.4} n}{kd} \right) \right),$$

in particular,

$$\mathbb{P}\left(\bigcup_{k \leq c_{2.4}\varepsilon n/d} \Omega_{k,\varepsilon} \right) \geq 1 - \left(C_{2.4} d/\varepsilon n \right)^{\varepsilon^2 d/8}.$$

We will need two more results from [23]. The following is [23, Proposition 3.3].

Proposition 2.5 (Row and columns are almost disjoint). *Let $\varepsilon \in (0,1)$ and $8 \le d \le \varepsilon n/6$. Denote*

$$\Omega_1(\varepsilon) := \Big\{ M \in \mathcal{M}_{n,d} : \forall i,j \in [n] \quad |\mathrm{supp}(R_i(M) + R_j(M))| \ge 2(1-\varepsilon)d$$
$$\text{and} \quad \big|\mathrm{supp}\big(R_i\left(M^T\right) + R_j\left(M^T\right)\big)\big| \ge 2(1-\varepsilon)d \Big\}.$$

Then

$$\mathbb{P}(\Omega_1(\varepsilon)) \ge 1 - n^2 \left(\frac{ed}{\varepsilon n}\right)^{\varepsilon d}.$$

Given $0 \le \alpha, \beta \le 1$, denote by $\Omega_0(\alpha,\beta)$ the set of matrices in $\mathcal{M}_{n,d}$ having a zero minor of size at least $\alpha n \times \beta n$, that is

$$\Omega_0(\alpha,\beta) := \{M \in \mathcal{M}_{n,d} \ : \exists I, J \subset [n] \quad \text{such that} \quad |I| \ge \alpha n,\, |J| \ge \beta n,$$
$$\text{and} \quad \forall i \in I \, \forall j \in J \quad \mu_{ij} = 0\}.$$

The next result is Theorem 3.4 from [23] (note that the condition $\beta \le 1/4$ there can be removed by adjusting absolute constants).

Proposition 2.6 (No large zero minors). *There exist absolute positive constants c, C such that the following holds. Let $2 \le d \le n/24$, $0 < \beta \le 1$, and $0 < \alpha \le \min(\beta, 1/4)$. Assume that*

$$\alpha \ge \frac{C \log(e/\beta)}{d}.$$

Then

$$\mathbb{P}\left(\Omega_0(\alpha,\beta)\right) \le \exp\left(-c\alpha\beta d n\right).$$

We now discuss another property of matrices in $\Omega_{m,\varepsilon}$. We start with the following construction. Given two disjoint sets J^ℓ, $J^r \subset [n]$ and a matrix $M \in \mathcal{M}_{n,d}$, denote

$$I^\ell = I^\ell(M) := \{i \le n : \, |\mathrm{supp} R_i \cap J^\ell| = 1 \ \text{ and } \ \mathrm{supp} R_i \cap J^r = \emptyset\},$$

and

$$I^r = I^r(M) := \{i \le n : \, \mathrm{supp} R_i \cap J^\ell = \emptyset \ \text{ and } \ |\mathrm{supp} R_i \cap J^r| = 1\}.$$

The sets J^ℓ, J^r will always be clear from the context. The upper indexes ℓ and r refer to *left* and *right*, since later, given a vector $x \in \mathbb{R}^n$ with $x_1 \ge x_2 \ge \ldots \ge x_n \ge 0$, we choose $J^\ell = [k_1]$ and $J^r = [k_2, n]$ for some $k_1 < k_2$ (this is the reason why the above formulas for $I^\ell(M), I^r(M)$ are "asymmetric").

Lemma 2.7. *Let $p \ge 2$, $m \ge 1$ be integers satisfying $pm \le c_{2.4}\varepsilon n/d$ and let $J^\ell, J^r \subset [n]$ be such that $J^\ell \cap J^r = \emptyset$, $|J^\ell| = m$, $|J^r| = (p-1)m$. Let $M \in \Omega_{pm,\varepsilon}$. Then*

$$|S_{J^\ell} \setminus S_{J^r}| \ge (1 - \varepsilon p) d |J^\ell| \quad \text{and} \quad |I^\ell| \ge (1 - 2\varepsilon p) d |J^\ell|,$$

where S_{J^ℓ}, S_{J^r} are defined by (2). In particular, if $|J^r| = |J^\ell| = m$ then

$$(1 - 4\varepsilon) dm \le \min(|I^\ell|, \, |I^r|) \le \max(|I^\ell|, \, |I^r|) \le dm.$$

Proof. Since $M \in \Omega_{pm,\varepsilon}$, we observe that $|S_{J^\ell} \cup S_{J^r}| \geq (1-\varepsilon)pd|J^\ell|$. Hence,

$$|S_{J^\ell} \setminus S_{J^r}| = |S_{J^\ell} \cup S_{J^r}| - |S_{J^r}| \geq (1-\varepsilon)pd|J^\ell| - (p-1)d|J^\ell| = (1-\varepsilon p)d|J^\ell|,$$

which proves the first estimate. To prove the second one, set

$$k := |\{i \in S_{J^\ell} \setminus S_{J^r} \,:\, |\mathrm{supp} R_i \cap J^\ell| = 1\}|.$$

Then the number of ones in the minor

$$\{\mu_{ij} \,:\, i \in S_{J^\ell} \setminus S_{J^r},\, j \in J^\ell\}$$

is at least

$$k + 2(|S_{J^\ell} \setminus S_{J^r}| - k) \geq 2(1-\varepsilon p)d|J^\ell| - k.$$

On the other hand, it cannot exceed $|J^\ell|d$. Therefore

$$k \geq 2(1-\varepsilon p)d|J^\ell| - d|J^\ell| = (1-2\varepsilon p)d|J^\ell|.$$

This completes the first part of the lemma. The second one follows by applying these estimates with $p = 2$, using that the roles of I^ℓ and I^r are interchangable and that each row contains exactly d ones. □

3 Almost constant vectors

In this section we treat almost constant vectors, which we split into almost constant sloping vectors (i.e., vectors with many coordinates almost equal to each other and without jumps) and almost constant steep vectors (i.e., almost constant vectors with jumps). First, in Theorem 3.1, we prove a bound for almost constant sloping vectors. This case is less involved. Then we turn to steep vectors. As we mentioned in the introduction, there will be three types of steep vectors, and, in fact, in two of three types we don't need to assume that vectors are almost constant. First, when a jump occurs between coordinates of $(x_i^*)_i$ indexed by k and m, with $k\sqrt{d/\log d} \approx m \lesssim n\log d/d^2$ (in fact we allow m to go up to $\approx n\sqrt{\log d}/d^{3/2}$). For such vectors, using properties of d-regular random matrices (Lemma 2.7), we show that with high probability such a matrix has many rows with only one 1 in columns corresponding to largest k coordinates and no other ones till the column corresponding to the m-th largest coordinate. We then obtain a lower bound on $\|Mx\|_2/\|x\|_2$ by using the assumption that x has the jump of order d. In this part we don't use that vector is almost constant. We will have two more classes of steep vectors – when a jump occurs between (essentially) $n\log d/d^2 \lesssim k \lesssim n\sqrt{\log d}/d^{3/2}$ and $m \approx n/d$ and between $k \approx n/d$ and $m \approx n/\log d$. For these two classes we use union bound, that is, we find a balance between the probability for individual vectors with such jumps and the size of corresponding ε-nets. Nets will be constructed in ℓ_∞ metric fixing $x_k^* = 1$ in order to control values of each coordinate indexed between k and m. Construction of nets is quite delicate, since we have rather weak control on the size of the first coordinates. For the individual probability bounds we use anti-concentration type technique together with switching argument. To control the size of the nets, we use jump of order $d^{3/2}$ for the second type of steep vectors, while in the third type of steep vectors we intersect them with almost constant vectors, reducing dimension. This allows us to use a constant jump only.

3.1 Almost constant, steep, and sloping vectors: definitions and main results

To define almost constant and steep vectors we will use the following parameters. In order to use Theorem 2.4, we fix ε_0 and a related parameter p as follows:

$$\varepsilon_0 = \sqrt{(\log d)/d}, \qquad p = \lfloor 1/(5\varepsilon_0) \rfloor = \left\lfloor \tfrac{1}{5}\sqrt{d/\log d} \right\rfloor$$

(the choice of p comes from $\varepsilon_0 p < 1$ needed in Lemma 3.7 in order to apply Lemma 2.7). We also fix three absolute positive sufficiently small constants a_1, a_2, and a_3, satisfying

$$a_3 \le a_2/28 \le a_1/28^2, \tag{3}$$

(we don't try to estimate the actual values of a_i's, the conditions on how small they are will be appearing in corresponding proofs). Set

$$n_0 := \lceil a_1 n \log d/d^2 \rceil, \quad n_2 := \lfloor a_2 n/d \rfloor, \quad \text{and} \quad n_3 := \lfloor a_3 n/\log d \rfloor.$$

If $n_0 = 1$, set $n_1 = 1$. Otherwise, fix an integer $r \ge 0$ such that $p^r < n_0 \le p^{r+1}$. Set

$$n_1 = \begin{cases} n_0, & \text{if } n_0 \le p, \\ p^{r+1}, & \text{otherwise.} \end{cases}$$

Note that

$$\frac{n_2}{n_1} \le \frac{a_2 d}{a_1 \log d}, \qquad \frac{n}{n_2} \le \frac{2d}{a_2}, \qquad \frac{n}{n_3} \le \frac{2\log d}{a_3}, \qquad \frac{n_3}{n_1} \le \frac{a_3 d^2}{a_1 \log^2 d}, \tag{4}$$

and, in the case $n_0 > 1$,

$$n_1 \le p n_0 \le a_1 \varepsilon_0 n/5d \tag{5}$$

We are ready now to describe our classes. First, given $\rho > 0$, we introduce a class of almost constant vectors by

$$\mathcal{B}(\rho) = \{x \in \mathbb{C}^n \,:\, \exists \lambda \in \mathbb{C} \text{ such that } |\{i \le n \,:\, |x_i - \lambda| \le \rho \, \|x\|_2\}| > n - n_3\}.$$

The definition of the class of steep vectors is more involved and consists of few steps at which we define sets $\mathcal{T}_0$, $\mathcal{T}_1$, and $\mathcal{T}_2$. We start with $\mathcal{T}_0$. If $n_0 = n_1 = 1$ we set $\mathcal{T}_0 = \emptyset$. If $n_0 > 1$, we denote

$$\mathcal{T}_{0,0} := \{x \in \mathbb{C}^n \,:\, x_1^* > 4d x_m^*\},$$

where $m = \min(n_0, p)$. In the case $1 < n_0 = n_1 \le p$ we set $\mathcal{T}_0 = \mathcal{T}_{0,0}$. Otherwise, if $n_0 > p$ we set for $1 \le i \le r$

$$\mathcal{T}_{0,i} := \{x \in \mathbb{C}^n \,:\, x \notin \bigcup_{j=0}^{i-1} \mathcal{T}_{0,j} \text{ and } x_{p^i}^* > 4d x_{p^{i+1}}^*\} \quad \text{and} \quad \mathcal{T}_0 = \bigcup_{i=0}^{r} \mathcal{T}_{0,i}.$$

Finally, we define two more sets of steep vectors, as

$$\mathcal{T}_1 := \{x \in \mathbb{C}^n \,:\, x \notin \mathcal{T}_0 \;\text{ and }\; x^*_{n_1} > d^{3/2}\, x^*_{n_2}\}$$

and

$$\mathcal{T}_2 := \{x \in \mathbb{C}^n \,:\, x \notin \mathcal{T}_0 \cup \mathcal{T}_1 \;\text{ and }\; x^*_{n_2} > 4\, x^*_{n_3}\}.$$

The vectors form $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ we call steep and all other vectors we call sloping.

We introduce the following functions h_i, $0 \le i \le r+1$,

$$h_{r+1} := \begin{cases} \sqrt{3p}\, n_1^{2+\alpha_d} & \text{if } n_0 > p, \\ 2d^{3/2}/\sqrt{\log d} & \text{if } 1 < n_0 \le p, \\ \sqrt{n} & \text{if } n_0 = 1, \end{cases} \qquad h_i := \begin{cases} \sqrt{n} & \text{if } i = 0, \\ \sqrt{n} + \sqrt{2p}\, p^{i(2+\alpha_d)} & \text{if } 1 \le i \le r, \end{cases}$$

where $\alpha_d = \log 4d/\log p - 2$ (note $2\log\log d/\log d \le \alpha_d \le 4\log\log d/\log d$ for large d). We also denote

$$b_{\mathcal{T}} := \begin{cases} 4d^{3/2} h_{r+1} & \text{if } \; n_0 > 1, \\ d\sqrt{n} & \text{if } \; n_0 = 1. \end{cases}$$

In this section we prove two following theorems. The first one treats almost constant sloping vectors, the second one treats almost constant sleep vectors (in fact, a slightly larger class).

Theorem 3.1. *Let $0 < \rho \le 1/(5\, b_{\mathcal{T}})$ and $x \in \mathcal{B}(\rho) \setminus (\mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2)$. Then for every $M \in \mathcal{M}_{n,d}$ and for every $z \in \mathbb{C}$ with $|z| \le d/6$ one has*

$$\|(M - z\mathrm{Id})x\|_2 \ge \frac{d\sqrt{3n}}{20\, b_{\mathcal{T}}}\, \|x\|_2.$$

Theorem 3.2. *There are absolute constants $C > 1 > c, c_1 > 0$ such that the following holds. Let $C < d < c_1 n$ and $0 < \rho \le 1/(d^{3/2}\, b_{\mathcal{T}})$. Let $z \in \mathbb{C}$ be such that $|z| \le d$. Denote*

$$\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1 \cup (\mathcal{T}_2 \cap \mathcal{B}(\rho))$$

and

$$\mathcal{E}_{steep} := \Big\{ M \in \mathcal{M}_{n,d} \,:\, \exists\, x \in \mathcal{T} \quad \textit{such that} \quad \|(M - z\mathrm{Id})x\|_2 < \frac{\sqrt{n_2 d}}{18 b_{\mathcal{T}}}\, \|x\|_2 \Big\}.$$

Then

$$\mathbb{P}(\mathcal{E}_{steep}) \le n^{-c\min(\log n, \sqrt{d \log d})}.$$

Remark 3.3. *In Section 4 we will use these two theorems in the following way. Let $\rho = 1/(d^{3/2}\, b_{\mathcal{T}})$, $|z| \le d/6$,*

$$\mathcal{B}_0(\rho) := \mathcal{B}(\rho) \cap \{x \in \mathbb{C}^n : \|x\|_2 = 1\},$$

and

$$\mathcal{E} := \Big\{ M \in \mathcal{M}_{n,d} \,:\, \exists\, x \in \mathcal{B}_0(\rho) \quad \textit{such that} \quad \|(M - z\mathrm{Id})x\|_2 < \rho^2/16 \Big\}.$$

Then Theorems 3.1 and 3.2 imply that

$$\mathbb{P}(\mathcal{E}) \le n^{-c\min(\log n, \sqrt{d \log d})} \le 1/2n^2.$$

Remark 3.4. *Note that*

$$\frac{d\sqrt{3n}}{20\, b_{\mathcal{T}}} \geq \frac{\sqrt{n_2 d}}{18 d^{3/2} h_{r+1}}.$$

In the proof of Theorem 3.2 we show also that

$$\frac{\sqrt{n_2 d}}{18 d^{3/2} h_{r+1}} \geq h(d,n),$$

where

$$h(d,n) = \begin{cases} c d^{-3/2} & \text{if } n_1 = 1 \text{ (that is, if } a_1 n \leq \frac{d^2}{\log d}\text{)}, \\ c\sqrt{n}\, d^{-3} (\log d)^{-1/2} & \text{if } 1 < n_1 \leq p \text{ (that is, if } \frac{d^2}{\log d} < a_1 n \leq \frac{d^{5/2}}{5 \log^{3/2} d}\text{)}, \\ c d^{5/4} (\log d)^2 n^{-3/2-\alpha_d} & \text{if } n_1 > p \text{ (that is, if } a_1 n > \frac{d^{5/2}}{5\log^{3/2} d}\text{)}. \end{cases}$$

In the proof of both theorems we will use the comparison of ℓ_2-norm of a given vector with a fixed coordinate. The next lemma provides such a bound in terms of the functions h_i. Moreover, we also estimate the ℓ_∞-norm. Note that we clearly have $\|x\|_2 \leq \sqrt{n}\, x_1^*$ for every $x \in \mathbb{C}^n$.

Lemma 3.5. *Let $d \leq n$ be large enough and $x \in \mathbb{C}^n$, $x \neq 0$. If $x \in \mathcal{T}_{0,i}$ for some $0 \leq i \leq r$, then*

$$\|x\|_2 \leq h_i\, x_{p^i}^*.$$

Moreover,

$$\|x\|_2 \leq \begin{cases} h_{r+1}\, x_{n_1}^* & \text{if } x \notin \mathcal{T}_0, \\ (b_{\mathcal{T}}/4)\, x_{n_2}^* & \text{if } x \notin \mathcal{T}_0 \cup \mathcal{T}_1, \\ b_{\mathcal{T}}\, x_{n_3}^* & \text{if } x \notin \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2. \end{cases}$$

Proof. The case $x \in \mathcal{T}_{0,0}$ is trivial.

If $1 < n_0 = n_1 \leq p$ then $\mathcal{T}_0 = \mathcal{T}_{0,0}$ and thus for $x \notin \mathcal{T}_0$ we observe

$$\|x\|_2^2 = \sum_{i=1}^{n_1-1} (x_i^*)^2 + \sum_{i=n_1}^{n} (x_i^*)^2 \leq 16 d^2 n_1 (x_{n_1}^*)^2 + n (x_{n_1}^*)^2 \leq (16 d^2 p + n)(x_{n_1}^*)^2.$$

The result follows since $n_0 \leq p$ implies $a_1 n \leq d^2 p/\log d$ and because d is large enough.

We now assume that $n_0 > p$. Let $x \in \mathcal{T}_{0,i}$ for some $1 \leq i \leq r$ or let $x \notin \mathcal{T}_0$ in which case we set $i = r+1$. Then for every $j < i$, one has $x \notin \mathcal{T}_{0,j}$, hence, assuming without loss of generality that $x_{p^i}^* = 1$,

$$x_1^* \leq (4d) x_p^* \leq (4d)^2 x_{p^2}^* \leq \ldots \leq (4d)^i x_{p^i}^* = (4d)^i = p^{i \log 4d/\log p}.$$

This implies

$$\|x\|_2^2 = ((x_1^*)^2 + \ldots (x_p^*)^2) + ((x_{p+1}^*)^2 + \ldots (x_{p^2}^*)^2) + \ldots \leq p(4d)^{2i} + p^2 (4d)^{2(i-1)} + \ldots + p^i (4d)^2 + n$$

$$= \frac{p(4d)^2((4d)^{2i} - p^i)}{(4d)^2 - p} + n \leq 2p(4d)^{2i} + n = 2p\, p^{2i\log 4d/\log p} + n,$$

which implies the result for $i \leq r$. In the case $i = r+1$, that is, if $x \notin \mathcal{T}_0$, this gives $\|x\|_2^2 \leq 2p\, n_1^{4+2\alpha_d} + n$. Note that we are in the case, $n_0 > p$, hence $n_1 \geq p^2$. Using the definition of n_0, we observe that $a_1 n \geq d^2 p/\log d$ and therefore

$$n_1^4 \geq p^6 n_1 \geq \frac{d^3}{(6\log d)^3} \frac{a_1 n \log d}{d^2} \geq \frac{a_1 n\, d}{\log d}$$

which implies for sufficiently large d that $\|x\|_2 \leq \sqrt{3p}\, n_1^{2+\alpha_d}$.

If $x \notin \mathcal{T}_0 \cup \mathcal{T}_1$ then clearly $x^*_{n_1} \leq d^{3/2} x_{n_2}$, and, if additionally $n_0 = 1$, then

$$\|x\|_2^2 = \sum_{i=1}^{n_2-1} (x_i^*)^2 + \sum_{i=n_2}^{n} (x_i^*)^2 \leq d^3 n_2 (x^*_{n_2})^2 + n(x^*_{n_2})^2 \leq (a_2 d^2 n + n)(x^*_{n_2})^2 \leq d^2 n (x^*_{n_2})^2/16,$$

provided $a_2 < 1/20$ and d is large enough. The case $x \notin \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2$ follows as well, since in this case $x^*_{n_3} \leq 4x^*_{n_2}$. This completes the proof. $\square$

3.2 Proof of Theorem 3.1

We will use the following simple claim.

Claim 3.6. *Let $J \subset [n]$, $k = |J|$, and $A > 1$. Let $M \in \mathcal{M}_{n,d}$. Then*

$$|\{i \leq n \,:\, |\mathrm{supp} R_i(M) \cap J| \geq Akd/n\}| \leq n/A.$$

Proof. The number of ones in the minor $[n] \times J$ is kd. Thus

$$|\{i \leq n \,:\, |\mathrm{supp} R_i(M) \cap J| \geq Akd/n\}| \cdot Akd/n \leq kd,$$

which implies the result. $\square$

Proof of Theorem 3.1. Clearly, we may assume $x \neq 0$. Fix a permutation $\sigma = \sigma_x$ of $[n]$ such that $x_i^* = |x_{\sigma(i)}|$ for $i \leq n$. Note that since $x \notin \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2 \cup \{0\}$ we have $|x^*_{n_3}| \neq 0$.

Fix $\lambda_0 = \lambda_0(x) \in \mathbb{C}$ such that the cardinality of

$$J_1 := \{i \leq n \,:\, |x_i - \lambda_0| \leq \rho\, \|x\|_2\}$$

is at least $n - n_3 + 1$. Then $\sigma(n_3) \in J_1$. By Lemma 3.5,

$$\|x\|_2 \leq b_{\mathcal{T}}\, x^*_{n_3} = b_{\mathcal{T}}\, |x_{\sigma(n_3)}|,$$

and, thus

$$|\lambda_0| - x^*_{n_3}/5 \leq |\lambda_0| - \rho\, \|x\|_2 \leq x^*_{n_3} = |x_{\sigma(n_3)}| \leq |\lambda_0| + \rho\, \|x\|_2 \leq |\lambda_0| + x^*_{n_3}/5,$$

where we also used that $\rho \leq 1/5b_{\mathcal{T}}$. This implies

$$(5/6)|\lambda_0| \leq x^*_{n_3} \leq (5/4)\, |\lambda_0|$$

(in particular, $|\lambda_0| \neq 0$) and, using that $\rho \leq 1/5b_{\mathcal{T}}$,

$$\rho \, \|x\|_2 \leq x^*_{n_3}/5 \leq |\lambda_0|/4.$$

Set

$$J_2 = \sigma([n_2]) \setminus J_1, \quad J_3 = \sigma([n_3]) \setminus (J_1 \cup J_2), \quad \text{and} \quad J_4 = [n] \setminus (J_1 \cup \sigma([n_3])).$$

Then $|J_3|, |J_4| \leq n_3$, $[n] = J_1 \cup J_2 \cup J_3 \cup J_4$ and

$$\forall j \in J_4 \;\; |x_j| \leq x^*_{n_3} \leq 5\lambda_0/4 \qquad \text{and} \qquad \forall j \in J_3 \;\; |x_j| \leq x^*_{n_2} \leq 4x^*_{n_3} \leq 5\lambda_0. \tag{6}$$

Now, given a matrix $M \in \mathcal{M}_{n,d}$, consider

$$I_2 = \{i \leq n \, : \, \mathrm{supp} R_i(M) \cap J_2 \neq \emptyset\} \;\text{ and }\; I_\ell = \{i \leq n \, : \, |\mathrm{supp} R_i(M) \cap J_\ell| \geq 16 n_3 d/n\},$$

for $\ell = 3, 4$. Since $M \in \mathcal{M}_{n,d}$ and by Claim 3.6, we have for small enough a_2,

$$|I_2| \leq d\, n_2 \leq n/16 \quad \text{and} \quad |I_\ell| \leq n/16 \;\text{ for }\; \ell = 3, 4.$$

Set $I := [n] \setminus (I_2 \cup I_3 \cup I_4 \cup \sigma([n_3]))$. Then

$$|I| \geq n - 3n/16 - n_3 \geq 3n/4 \quad \text{and} \quad \forall i \in I \;\; |x_i| \leq x^*_{n_3} \leq (5/4)|\lambda_0|.$$

Moreover, for every $i \in I$, denote $\overline{J_\ell} = J_\ell \cap \mathrm{supp} R_i(M)$ for $1 \leq \ell \leq 4$, and note that $\overline{J_2} = \emptyset$ since $i \notin I_2$. Using the triangular inequality, we observe

$$|\left\langle R_i(M - z\mathrm{Id}), x^\dagger \right\rangle| \geq \Big| \sum_{j \in \overline{J_1}} x_j \Big| - \sum_{j \in \overline{J_3}} |x_j| - \sum_{j \in \overline{J_4}} |x_j| - |z x_i|.$$

We estimate each of the previous terms separately. Using the triangle inequality and the definition of J_1, we have

$$\Big| \sum_{j \in \overline{J_1}} x_j \Big| \geq |\lambda_0| \, |\overline{J_1}| - \sum_{j \in \overline{J_1}} \Big| x_j - \lambda_0 \Big| \geq |\overline{J_1}| \, (|\lambda_0| - \rho \|x\|_2) \geq (d - 32 n_3 d/n) \, (|\lambda_0| - \rho \|x\|_2),$$

where for the last inequality we used that for $i \notin I_3 \cup I_4$ one has $|\overline{J_1}| = d - |\overline{J_2}| - |\overline{J_3}| - |\overline{J_4}| \geq d - 32 n_3 d/n$. Using (6),

$$\Big| \sum_{j \in \overline{J_3}} x_j \Big| + \Big| \sum_{j \in \overline{J_4}} x_j \Big| \leq |\overline{J_3}| \, x^*_{n_2} + |\overline{J_4}| \, x^*_{n_3} \leq 100 |\lambda_0| n_3 d/n.$$

Putting together the above estimates, we obtain for large enough d

$$\begin{aligned} |\left\langle R_i(M - z\mathrm{Id}), x^\dagger \right\rangle| &\geq (d - 32 n_3 d/n)(|\lambda_0| - \rho \|x\|_2) - 100 |\lambda_0| n_3 d/n - (5/4)|\lambda_0||z| \\ &\geq |\lambda_0| d/2, \end{aligned}$$

where we used $|\lambda_0| - \rho \|x\|_2 \geq (3/4)|\lambda_0|$, $n_3/n \leq c/\log d$, and $|z| \leq d/6$. This implies

$$\|(M - z\mathrm{Id})x\|_2 \geq (|\lambda_0| d/2) \sqrt{3n/4},$$

which completes the proof. $\square$

3.3 Lower bounds on $\|Mx\|_2$ for vectors from $\mathcal{T}_0$

Here we provide lower bounds on the ratio $\|Mx\|_2/\|x\|_2$ for vectors x from $\mathcal{T}_0$. Recall that given ε and k the set $\Omega_{k,\varepsilon}$ was introduced before Theorem 2.4.

Lemma 3.7. *Let $C \le d \le n$, where C is an absolute positive constant and $x \in \mathcal{T}_0$. Let $z \in \mathbb{C}$ be such that $|z| \le d$. If $1 < n_0 = n_1 \le p$ and $M \in \Omega_{n_0,\varepsilon_0}$ then $\|(M - zId)x\|_2 \ge \sqrt{d/8n}\,\|x\|_2$. If $n_0 > p$ and*

$$M \in \bigcap_{j=1}^{r+1} \Omega_{p^j,\varepsilon_0}$$

then

$$\|(M - z\mathrm{Id})x\|_2 \ge \min\left\{\sqrt{d/8n},\ \frac{p\sqrt{n_1 d}}{8h_{r+1}}\right\}\|x\|_2.$$

Proof. We prove the case $n_0 \ge p$, the other case is similar. Fix $x \in \mathcal{T}_0$ and fix $0 \le i \le r$ such that $x \in \mathcal{T}_{0,i}$ and denote $m = p^i$. Fix a permutation $\sigma = \sigma_x$ of $[n]$ such that $x_i^* = |x_{\sigma(i)}|$ for $i \le n$. Then $x_m^* > 4dx_{pm}^*$. Let

$$J^\ell = \sigma([m]), \quad J^r = \sigma([pm] \setminus [m]), \quad \text{and} \quad J_3 := (J^\ell \cup J^r)^c.$$

Then, for sufficiently small a_1,

$$|J^\ell \cup J^r| = pm \le pn_0 \le c_{2.4}\varepsilon_0 n/d \qquad \text{and} \qquad |J^r| = (p-1)|J^\ell| = (p-1)m.$$

Denote by I_0 the set of rows having exactly one 1 in J^ℓ and no 1's in J^r. Lemma 2.7 implies that

$$|I_0| \ge (1 - 2p\varepsilon_0)md \ge 3md/5.$$

Let $I = I_0 \setminus (J^\ell \cup J^r)$ (so that the minor $I \times (J^\ell \cup J^r)$ does not intersect the main diagonal). Then $|I| \ge 3md/5 - pm \ge md/2$ provided that d is large enough. By definition, for every $s \in I$ there exists $j(s) \in J^\ell$ such that

$$\mathrm{supp} R_s \cap J^\ell = \{j(s)\}, \quad \mathrm{supp} R_s \cap J^r = \emptyset, \quad \text{and} \quad \max_{i \in J_3} |x_i| \le x_{mp}^*.$$

Using Lemma 3.5, the fact that $s \notin J^\ell \cup J^r$ (which implies $x_s^* \le x_{pm}^*$), and that $j(s) \in J^\ell$ (which implies $|x_{j(s)}| \ge x_m^* > 4dx_{mp}^*$), we obtain

$$\begin{aligned}|\langle R_s(M - z\mathrm{Id}),\, x^\dagger\rangle| &= \Big|x_{j(s)} + \sum_{j \in J_3 \cap \mathrm{supp} R_s} x_j - zx_s\Big| \\ &\ge |x_{j(s)}| - (d-1)\,x_{mp}^* - |z|\,x_{mp}^* \ \ge x_m^*/2 \ge \|x\|_2/2h_i.\end{aligned}$$

Since the number of such rows is $|I| \ge md/2 = p^i d/2$ we obtain

$$\|(M - z\mathrm{Id})x\|_2 \ge \sqrt{p^i d}\,\|x\|_2/(2\sqrt{2}h_i).$$

If $i = 0$ then $p^{i/2}/h_i = 1/\sqrt{n}$. If $i \ge 1$ and $\sqrt{n} \ge \sqrt{2p}\,p^{i\,(2+\alpha_d)}$, then $h_i \le 2\sqrt{n}$ and $p^{i/2}/h_i \ge p^{i/2}/(2\sqrt{n}) \ge 1/\sqrt{n}$ provided d is large enough. If $\sqrt{n} \le \sqrt{2p}\,p^{i\,(2+\alpha_d)}$ then $h_i \le 2\sqrt{2p}\,p^{i\,(2+\alpha_d)}$. Using this and that $p^i \le p^r = n_1/p$, we get

$$\frac{p^{i/2}}{h_i} \ge \frac{p^{i/2}}{2\sqrt{2p}\,p^{i\,(2+\alpha_d)}} \ge \frac{p^{r/2}}{2\sqrt{2p}\,p^{r\,(2+\alpha_d)}} \ge \frac{p}{2\sqrt{2}}\,\frac{\sqrt{n_1}}{n_1^{2+\alpha_d}},$$

which implies the result. $\square$

3.4 Nets for steep vectors from $\mathcal{T}_1 \cup \mathcal{T}_2$

For the rest of steep vectors (i.e., for vectors from $\mathcal{T}_1 \cup \mathcal{T}_2$) we will use the union bound together with a covering argument. We first construct nets for "normalized" versions of the sets $\mathcal{T}_i$ and then provide individual probability bounds for elements of the nets. The natural normalization would be $x^*_{n_1} = 1$, which we use for $\mathcal{T}_1$. However, for individual probability bounds below and to have the same level of approximation, it is more convenient to use a slightly different normalization for $\mathcal{T}_2$. Moreover, since $\mathcal{T}_2$ has a constant jump, we can't just ignore the tail of the sequence as we will do for vectors in $\mathcal{T}_1$. To overcome this difficulty, and to have a better control on the size of a net, we intersect this set with the set of almost constant vectors. We set

$$\mathcal{T}_1' = \{x \in \mathcal{T}_1 \,:\, x^*_{n_1} = 1\} \quad \text{and} \quad \mathcal{T}_2' = \mathcal{T}_2'(\rho) = \{x \in \mathcal{T}_2 \,:\, x^*_{n_2} = 1\} \cap \mathcal{B}(\rho),$$

where $0 < \rho \leq 1/(d^{3/2}\, b_{\mathcal{T}})$ (the intersection with $\mathcal{B}(\rho)$ here is needed to better control the size of ε-nets constructed below for such vectors).

Lemma 3.8 (Cardinalities of nets)**.** *Let $d \leq n$ be large enough and $0 < \rho \leq 1/(d^{3/2}\, b_{\mathcal{T}})$. Then, for each $i = 1, 2$, there exists a $d^{-3/2}$-net $\mathcal{N}_i$ in $\mathbb{C}^n$ for $\mathcal{T}_i'$ in ℓ_∞-metric with*

$$|\mathcal{N}_i| \leq \exp\left(dn_i/4\right),$$

and for every $y \in \mathcal{N}_i$ one has $y^*_j \leq 1/4 + 1/d^{3/2}$ for all $j \geq n_{i+1}$.

Proof. The constructions for $i = 1$ and $i = 2$ are quite similar, and we carry out the argument simultaneously for both cases, making adjustments where necessary. For every $x \in \mathcal{T}_i'$ ($i = 1, 2$) fix a permutation $\sigma = \sigma_x$ of $[n]$ such that $x^*_j = |x_{\sigma(j)}|$ for $j \leq n$.

The main idea is to split a given vector from $\mathcal{T}_i'$ into three parts according to the behaviour of its coordinates (essentially, parts corresponding to the largest coordinates, middle sized coordinates, and the smallest coordinates with small adjustment in the case $i = 2$) and approximate each part separately. Then we construct nets for vectors with the same splitting and take the union over all nets. To be more precise, for each $x \in \mathcal{T}_i'$ ($i = 1, 2$) we consider a partition of $[n]$ into three sets $B_1(x)$, $B_2(x)$, $B_3(x)$ corresponding to x, as follows. If $n_1 = 1$ (i.e., if $d^2/\log d \geq a_1 n$) we set $B_1(x) = \emptyset$. Otherwise, if $n_1 > 1$, we set $B_1(x) = \sigma_x([n_1])$. Further, let us define sets $B_2(x), B_3(x)$ (this definition will depend on i). For $i = 1$ we set

$$B_2(x) = \sigma_x([n_2]) \setminus B_1(x) \quad \text{and} \quad B_3(x) = \sigma_x([n] \setminus [n_2]).$$

If $i = 2$ then since $x \in \mathcal{B}(\rho)$ there exists $\lambda_0(x)$ such that the cardinality of the set

$$B_0(x) := \{j \leq n \,:\, |x_j - \lambda_0(x)| \leq \rho \|x\|_2\}$$

is larger than $n - n_3$. This in particular implies that $\sigma_x(n_3) \in B_0(x)$ and hence, by Lemma 3.5,

$$|\lambda_0(x)| \leq x^*_{n_3} + \rho\|x\|_2 < x^*_{n_2}/4 + (\rho b_{\mathcal{T}}/4) x^*_{n_2} \leq 1/3.$$

So, for $i = 2$ we choose an arbitrary subset $B_3(x) \subset B_0(x)$ of cardinality $n - n_3$ and fix it, and we let $B_2(x) = [n] \setminus (B_1(x) \cup B_3(x))$.

Note that if $n_1 > 1$ then for every $x \in \mathcal{T}_i'$ ($i = 1$ or $i = 2$) we always have

$$|B_1(x)| = n_1, \quad |B_2(x)| = n_{i+1} - n_1, \quad \text{and} \quad |B_3(x)| = n - n_{i+1}.$$

Thus, given a partition of $[n]$ into three sets B_1, B_2, B_3 with cardinalities $|B_1| = n_1$, $|B_2| = n_{i+1} - n_1$, $|B_3| = n - n_{i+1}$, it is enough to construct a net for vectors $x \in \mathcal{T}_i'$ with $B_1(x) = B_1$, $B_2(x) = B_2$, $B_3(x) = B_3$ and then take the union of nets over all such partitions $\{B_1, B_2, B_3\}$ of $[n]$. In what follows, we skip the case $n_1 = 1$ (and $B_1 = \emptyset$) as the simplest one, and assume that $n_1 > 1$.

Now we describe our construction. Note that for $x \in \mathcal{T}_i'$ ($i = 1, 2$) one has $x^*_{n_1} \leq d^{(3/2)(i-1)}$ and

$$x_1^* \leq (4d) x_p^* \leq (4d)^2 x_{p^2}^* \leq \ldots \leq (4d)^{r+1} x_{p^{r+1}}^* = (4d)^{r+1} x_{n_1}^* \leq d^{(3/2)(i-1)} \, (4d)^{r+1} \tag{7}$$

(with corresponding adjustment for the case $n_1 < p$). Recall that we deal with the case $n_1 > 1$ (otherwise, $B_1(x) = \emptyset$ and we skip the first part). Fix $I_0 \subset [n]$ with $|I_0| = n_1$ (which will play the role of B_1). We construct a $d^{-3/2}$-net $\mathcal{N}_{I_0}$ in the set

$$\mathcal{T}_{I_0} := \big\{ x \in (\mathcal{T}_0)^c : \, \sigma_x([n_1]) = B_1(x) = I_0, \; x^*_{n_1} = d^{(3/2)(i-1)}, \; x^*_{n_1+1} = 0 \big\}.$$

Clearly, the nets $\mathcal{N}_{I_0}$ for various I_0's can be related by appropriate permutations, so without loss of generality we can assume that $I_0 = [n_1]$. First, let us construct a partition of I_0. If $n_1 = n_0 \leq p$, let $I_1 = [n_1]$. Otherwise, recall that $n_1 = p^{r+1}$ and let

$$I_1 = [p], \;\; I_2 = [p^2] \setminus [p], \;\; I_3 = [p^3] \setminus [p^2], \; \ldots, \;\; I_{r+1} = [p^{r+1}] \setminus [p^r].$$

Then the sets $I_1, \ldots, I_{r+1}$ form a partition of $I_0 = [n_1]$. Now, consider the set

$$\mathcal{T}^* := \big\{ x \in \mathcal{T}_{[n_1]} : \, \sigma_x(I_j) = I_j, \;\; j = 1, 2, \ldots, r+1 \big\}$$

and construct a $d^{-3/2}$-net $\mathcal{N}^*$ in $\mathcal{T}^*$ in the following way. Below we provide the proof for the case $n_1 > p$ (i.e., when we have at least two sets in the partition), the other case is simpler. By (7), for every $x \in \mathcal{T}^*$, one has $\|P_{I_j} x\|_\infty \leq b := d^{(3/2)(i-1)} \, (4d)^{r+2-j}$ for every $j \leq r+1$ (where P_I denotes the coordinate projection onto $\mathbb{C}^I$). Set

$$\mathcal{N}^* := \mathcal{N}_1 \oplus \mathcal{N}_2 \oplus \cdots \oplus \mathcal{N}_{r+1},$$

where $\mathcal{N}_j$ is a $d^{-3/2}$-net (in the ℓ_∞-metric) of cardinality at most

$$(3bd^{3/2})^{2|I_j|} \leq (4d)^{2(r+5-j)p^j}$$

in the coordinate projection of the complex cube $P_{I_j}(bB_\infty^n)$. Since d is large enough and $n_1 = p^{r+1}$, we observe

$$\sum_{j=1}^{r+1} 2(r+5-j)p^j = 2p^{r+1} \sum_{m=0}^{r} (m+4)p^{-m} \leq 10p^{r+1} = 10n_1,$$

which implies

$$|\mathcal{N}^*| \le \prod_{j=1}^{r+1} |\mathcal{N}_j| \le \exp(10 n_1 \log(4d)).$$

To pass from the net for $\mathcal{T}^*$ to the net for $\mathcal{T}_{[n_1]}$, let $\mathcal{N}_{[n_1]}$ be the union of nets constructed as $\mathcal{N}^*$ but for arbitrary partition I_1', ..., I_{r+1}' of $[n_1]$ with $|I_j'| = |I_j|$. Using that $p = \lfloor (1/5)\sqrt{d/\log d} \rfloor$, we observe that

$$\sum_{j=1}^{r} p^j \log(ep) \le \frac{p^{r+1}}{p-1} \log(ep) \le n_1 \log^2 d/\sqrt{d}.$$

Therefore, for large enough d,

$$|\mathcal{N}_{[n_1]}| \le |\mathcal{N}^*| \prod_{j=1}^{r} \binom{p^{j+1}}{p^j} \le |\mathcal{N}^*| \prod_{j=1}^{r} (ep)^{p^j} \le \exp(11 n_1 \log d).$$

Now we construct a net for the second part of the vector. Fix $J_0 \subset [n]$ with $|J_0| = n_{i+1} - n_1$ (which will play the role of B_2). We construct a $d^{-3/2}$-net in the set

$$\mathcal{T}_{J_0} := \{P_{B_2(x)} x \, : \, x \in \mathcal{T}_i', \ B_2(x) = J_0, \ x^*_{n_{i+1}} = 0\}.$$

Since $x^*_{n_1} \le d^{(3/2)(i-1)}$ for $x \in \mathcal{T}_i'$, it is enough to take $d^{-3/2}$-net $\mathcal{K}_{J_0}$ of cardinality at most

$$(3 d^{3/2} d^{(3/2)(i-1)})^{2|J_0|} \le (2d)^{6 n_{i+1}}$$

in the coordinate projection of the complex cube $P_{J_0}(d^{(3/2)(i-1)} B_\infty^n)$.

It remains to construct a net for the third part of the vector, corresponding to coordinates in B_3. Fix B of cardinality $n - n_{i+1}$ and consider the set

$$\mathcal{T}_B := \{P_{B_3(x)} x \, : \, x \in \mathcal{T}_i', \ B_3(x) = B\}.$$

If $i = 1$ then, by definitions, $\|y\|_\infty < d^{-3/2}$ for every $y \in \mathcal{T}_B$, therefore our net, $\mathcal{O}_B$, will consist of 0 only. In the case $i = 2$, for $x \in \mathcal{T}_2'$ and $j \in B$, using Lemma 3.5 and the condition on ρ, we have that

$$|x_j - \lambda_0(x)| \le \rho \|x\|_2 \le (\rho b_{\mathcal{T}}/4) x^*_{n_2} \le 1/(4d^{3/2}) \quad \text{and} \quad |\lambda_0(x)| \le 1/3.$$

Take a $3/(4d^{3/2})$-net $\mathcal{O}$ in the set $\{\lambda \in \mathbb{C} \, : \, |\lambda| \le 1/3\}$ of cardinality at most $2d^3$ and let

$$\mathcal{O}_B := \{y \in \mathbb{C}^B \, : \, \exists \lambda \in \mathcal{O} \text{ such that } \forall j \in B \text{ one has } y_j = \lambda\}.$$

Clearly, $\mathcal{O}_B$ is a $d^{-3/2}$-net for $\mathcal{T}_B$.

Finally consider the net

$$\mathcal{N} := \bigcup \{y = y_1 + y_2 + y_3 \, : \, y_1 \in \mathcal{N}_{I_0}, \, y_2 \in \mathcal{K}_{J_0}, \, y_3 \in \mathcal{O}_B\},$$

where the union is taken over all partitions of $[n]$ into I_0, J_0, B with $|I_0| = n_1$, $|J_0| = n_{i+1} - n_1$, and $|B| = n - n_i$. Clearly, $\mathcal{N}$ is a $d^{-3/2}$-net for $\mathcal{T}_i'$ and, using (4) and (3), we obtain for large enough d,

$$|\mathcal{N}| \leq \binom{n}{n_{i+1}} \binom{n_{i+1}}{n_1} |\mathcal{N}_{I_0}| \, |\mathcal{K}_{J_0}| \, |\mathcal{O}_B| \leq \left(\frac{en}{n_{i+1}}\right)^{n_{i+1}} \left(\frac{en_{i+1}}{n_1}\right)^{n_1} (2d)^{11n_1+6n_{i+1}+3}$$

$$\leq \exp\left(7n_{i+1}\log d\right) \leq \exp\left(7(a_{i+1}/a_i)dn_i\right) \leq \exp\left(dn_i/4\right).$$

Without loss of generality (by removing unnecessary vectors from $\mathcal{N}$), we may assume that every $y \in \mathcal{N}$ approximates some $x \in \mathcal{T}_i'$. This implies that for every $y \in \mathcal{N}$ one has $y_j^* \leq 1/4 + 1/d^{3/2}$ for all $j \geq n_{i+1}$, completing the proof. □

3.5 Individual probability bounds

To obtain the lower bounds on $\|(M+W)x\|_2$, where W is a fixed matrix, for vectors x from our nets, we investigate the behavior of coordinates of $(M+W)x$, that is of the inner products $\langle R_i(M+W), x^\dagger \rangle$. One of the tools that we use is Theorem 2.4 together with Lemma 2.7 applied to the $2m$ columns of M corresponding to the m biggest and m smallest (in the absolute value) coordinates of x with properly chosen m. Then, using jumps, we show that the inner product of some row $R_i(M+W)$ with the first part of the vector and with the second part of the vector cannot be simultaneously large. This will reduce the set of matrices under consideration to a much smaller set, where it is easier to obtain a good probability bound. To make our scheme work we will use the following subdivision of $\mathcal{M}_{n,d}$.

Given $J \subset [n]$ and $M \in \mathcal{M}_{n,d}$ we denote

$$I(J, M) = \{i \leq n \, : \, |\mathrm{supp} R_i(M) \cap J| = 1\}$$

(*cf.*, the definition of $I^\ell(M)$, $I^r(M)$ before Lemma 2.7, clearly, if we split J into J^ℓ and J^r, then $I(J,M) = I^\ell(M) \cup I^r(M)$).

Fix $J \subset [n]$ and denote $J_0 = J^c$. Given a subset I of $[n]$ and $V = \{v_{ij}\} \in \mathcal{M}_{n,d}$, consider the class

$$\mathcal{F}(I, V) = \{M \in \mathcal{M}_{n,d} \, : \quad I(J,M) = I \quad \text{and} \quad \forall i \leq n \, \forall j \in J_0 \;\; \mu_{ij} = v_{ij}\}$$

(depending on the choice of I such a class can be empty). In words, we first fix the columns indexed by J_0 and then fix a set of indices I such that the rows indexed by I have only one 1 in columns indexed by J. Clearly, $\mathcal{M}_{n,d}$ splits into disjoint union of classes $\mathcal{F}(I,V)$ over some subset of matrices V in $\mathcal{M}_{n,d}$ and all $I \subset [n]$.

Lemma 3.9 (individual probability)**.** *There exist absolute constants $C > 1 > \varepsilon > 0$ such that the following holds. Let $C < d < n$, $i = 1, 2$, and W be a complex $n \times n$ matrix. Assume $x \in \mathbb{C}^n$ satisfies*

$$x_{n_i}^* \geq 1/2 + x_j^* \quad \textit{for every} \quad j \geq n_{i+1}.$$

Denote $E(x) := \left\{M \in \mathcal{M}_{n,d} \, : \, \|(M+W)x\|_2 \leq \sqrt{n_i d}/13\right\}$. Then

$$\mathbb{P}(E \cap \Omega_{2n_i,\varepsilon}) \leq \exp(-n_i d/2).$$

Proof. Fix x satisfying the condition of the lemma. Let σ be a permutation of $[n]$ such that $x_j^* = |x_{\sigma(j)}|$ for all $j \leq n$. Denote $m = n_i$. Let

$$J^\ell = \sigma([n_i]) \quad \text{ and } \quad J^r = \sigma([n - n_i + 1, n]).$$

Denote $J = J^\ell \cup J^r$ and $J_0 = J^c$. Fix $\varepsilon > 0$ small enough. We assume that $a_2 < c_{2.4}\varepsilon/2$. Then $m = n_i \leq n_2 \leq c_{2.4}\varepsilon n/2d$.

Let $M \in \Omega_{2m,\varepsilon}$. Let the sets $I^\ell(M)$ and $I^r(M)$ be defined as before Lemma 2.7. Since $|J| = 2m \leq c_{2.4}\varepsilon n/d$, this lemma implies that $|I^\ell(M)|, |I^r(M)| \in [(1-4\varepsilon)md,\, md]$, in particular $I = I^\ell(M) \cup I^r(M)$ satisfy

$$|I| \in [2(1-4\varepsilon)md,\, 2md]. \tag{8}$$

Now we split $\mathcal{M}_{n,d}$ into disjoint union of classes $\mathcal{F}(I,V)$ defined at the beginning of this subsection and note that $\Omega_{2m,\varepsilon} \cap \mathcal{F}(I,V) \neq \emptyset$ implies that I satisfies (8). Thus, to prove our lemma it is enough to prove uniform upper bound for such classes, indeed,

$$\mathbb{P}(E(x) \cap \Omega_{2m,\varepsilon}) \leq \max \mathbb{P}(E(x) \cap \Omega_{2m,\varepsilon} \,|\, \mathcal{F}(I,V)) \leq \max \mathbb{P}(E(x) |\, \mathcal{F}(I,V)),$$

where the first maximum is taken over all classes $\mathcal{F}(I,V)$ with $\Omega_{2m,\varepsilon} \cap \mathcal{F}(I,V) \neq \emptyset$ and the second maximum is taking over $\mathcal{F}(I,V)$ with I's satisfying (8).

Fix such a class $\mathcal{F}(I,V)$ for some $I \subset [n]$ with $t_1 := |I| \in [2(1-4\varepsilon)md,\, 2md]$ and denote the uniform probability on it just by $\mathbb{P}_{\mathcal{F}}$, that is

$$\mathbb{P}_{\mathcal{F}}(\cdot) = \mathbb{P}(\cdot \,|\, \mathcal{F}(I,V)).$$

Without loss of generality we assume that $I = [t_1]$.

By definition, for matrices $M \in E(x)$ we have

$$\|(M+W)x\|_2^2 = \sum_{i=1}^n |\langle R_i(M+W),\, x^\dagger\rangle|^2 \leq md/169.$$

Therefore there are at most $t_0 := 4md/169$ rows $R_i = R_i(M+W)$ with $|\langle R_i,\, x^\dagger\rangle| \geq 1/4$. Hence,

$$|\{i \in I \,:\, |\langle R_i,\, x^\dagger\rangle| < 1/4\}| \geq t_1 - t_0.$$

Denote $t := \lceil t_1 - t_0 \rceil$. The above bound implies that for every $M \in E(x)$ there is a set of indices $B(M) \subset I$ such that $|B(M)| = t$ and for every $i \in B(M)$ one has $|\langle R_i,\, x^\dagger\rangle| < 1/4$. Thus, denoting

$$\Omega_i := \{M \in \mathcal{F}(I,V) : \, |\langle R_i,\, x^\dagger\rangle| < 1/4\},$$

we obtain

$$\mathbb{P}_{\mathcal{F}}(E(x)) \leq \sum_{\substack{B\subset I\\ |B|=t}} \mathbb{P}_{\mathcal{F}}\Big(\bigcap_{i\in B} \Omega_i\Big) = \binom{t_1}{t} \max_{\substack{B\subset I\\ |B|=t}} \mathbb{P}_{\mathcal{F}}\Big(\bigcap_{i\in B} \Omega_i\Big) \leq \Big(\frac{et_1}{t_0}\Big)^{t_0} \max_{\substack{B\subset I\\ |B|=t}} \mathbb{P}_{\mathcal{F}}\Big(\bigcap_{i\in B} \Omega_i\Big). \tag{9}$$

Next for every $i \in I$ by $\mathcal{F}_I^\ell(i)$ and $\mathcal{F}_I^r(i)$ denote the sets

$$\{M \in \mathcal{F}(I,V) :\, i \in I^\ell(M)\} = \{M \in \mathcal{F}(I,V) :\, |\mathrm{supp} R_i(M) \cap J^\ell| = 1,\, \mathrm{supp} R_i(M) \cap J^r = \emptyset\}$$

and

$$\{M \in \mathcal{F}(I,V) :\, i \in I^r(M)\} = \{M \in \mathcal{F}(I,V) :\, |\mathrm{supp} R_i(M) \cap J^r| = 1,\, \mathrm{supp} R_i(M) \cap J^\ell = \emptyset\}.$$

We show that for every $i \in I$ either $\Omega_i \subset \mathcal{F}_I^\ell(i)$ or $\Omega_i \subset \mathcal{F}_I^r(i)$. Indeed, assume that $M_1 \in \mathcal{F}_I^\ell(i)$ and $M_2 \in \mathcal{F}_I^r(i)$. By the definition of our sets and by the conditions on x, we have

$$J_1 := \mathrm{supp} R_i(M_1) \setminus J = \mathrm{supp} R_i(M_2) \setminus J,$$

and there exist $j_\ell \in J^\ell$, $j_r \in J^r$ such that

$$\langle R_i(M_1),\, x^\dagger\rangle = x_{j_\ell} + \sum_{j \in J_1} x_j \quad \text{ and } \quad \langle R_i(M_2),\, x^\dagger\rangle = x_{j_r} + \sum_{j \in J_1} x_j.$$

Hence,

$$|\langle R_i(M_1+W),\, x^\dagger\rangle| + |\langle R_i(M_2+W),\, x^\dagger\rangle| \geq |\langle R_i(M_1+W),\, x^\dagger\rangle - \langle R_i(M_2+W),\, x^\dagger\rangle|$$
$$= |x_{j_\ell} - x_{j_r}| \geq x^*_{n_i} - |x_{j_r}| \geq 1/2.$$

Thus, it is impossible to simultaneously have both

$$|\langle R_i(M_1+W),\, x^\dagger\rangle| < 1/4 \quad \text{ and } \quad |\langle R_i(M_2+W),\, x^\dagger\rangle| < 1/4$$

and therefore either $\Omega_i \subset \mathcal{F}_I^\ell(i)$ or $\Omega_i \subset \mathcal{F}_I^r(i)$. This implies for every $B \subset I$ with $|B| = t$,

$$\mathbb{P}\Big(\bigcap_{i \in B} \Omega_i\Big) \leq \max_{B_0 \subset B} \mathbb{P}_{\mathcal{F}_I}\Big(\bigcap_{i \in B_0} \mathcal{F}_I^\ell(i) \bigcap \bigcap_{i \in B \setminus B_0} \mathcal{F}_I^r(i)\Big) = \max_{B_0 \subset [t]} \mathbb{P}_{\mathcal{F}_I}\Big(\bigcap_{i \in B_0} \mathcal{F}_I^\ell(i) \bigcap \bigcap_{i \in [t] \setminus B_0} \mathcal{F}_I^r(i)\Big),$$

where in the last equality we used permutation invariance.

Claim 3.10. *If d is large enough and ε is small enough then for every $B_0 \subset [t]$ one has*

$$\mathbb{P}_{\mathcal{F}}\Big(\bigcap_{i \in B_0} \mathcal{F}_I^\ell(i) \bigcap \bigcap_{i \in [t] \setminus B_0} \mathcal{F}_I^r(i)\Big) \leq e^{-t/3}.$$

Recall that $t_1 \in [2(1-4\varepsilon)md,\, 2md]$, $t_0 = 4md/169$, and $t = \lceil t_1 - t_0 \rceil$, so that

$$t/3 - t_0 \log(et_1/t_0) \geq md\,((2 - 8\varepsilon - 4/169)/3 - (4/169)\log(169e/2))) \geq md/2,$$

provided that ε is small enough. Therefore Claim 3.10 and (9) imply the desired result. $\square$

Proof of Claim 3.10. Fix $B_0 \subset [t]$. Denote $\ell_0 := |B_0|$ and without loss of generality assume that $\ell_0 \geq t/2$. Let $q = \lfloor \ell_0/2 \rfloor$. To compare the cardinalities of

$$A := \bigcap_{i \in B_0} \mathcal{F}_I^\ell(i) \bigcap \bigcap_{i \in [t] \setminus B_0} \mathcal{F}_I^r(i)$$

and $\mathcal{F}(I,V)$ we construct a relation R between them as follows. Let $M \in A$. We say that $(M, M') \in R$ if $M' \in \mathcal{F}(I,V)$ can be obtained from M in the following way. Choose a subset $B_1 \subset B_0$ of cardinality q. There are

$$\binom{\ell_0}{q} \geq \frac{2^{\ell_0}}{2\sqrt{\ell_0}}$$

such choices. Let $i_1 < i_2 < \ldots < i_q$ be the elements of B_1. Recall that $M \in \mathcal{F}_I^\ell(i_s)$ for every $s \leq q$. Let $j_1, \ldots, j_q$ be elements of J^ℓ such that M has ones on positions (i_s, j_s) for $s \leq q$. Choose a subset $B_2 \subset I^r(M)$ of cardinality q. There are

$$\binom{|I^r(M)|}{q} \geq \binom{\lceil (1-4\varepsilon)md \rceil}{q}$$

such choices. Let $v_1 < v_2 < \ldots < v_q$ be elements of B_2. Let $w_1, \ldots, w_q$ be elements of J^r such that M has ones on positions (v_s, w_s) for $s \leq q$. Let $M' \in \mathcal{F}(I,V)$ be obtained from M by substituting ones with zeros on places (i_s, j_s) and (v_s, w_s) and substituting zeros with ones on places (i_s, w_s) and (v_s, j_s) for all $s \leq q$. By construction we have

$$|R(A)| \geq \frac{2^{\ell_0}}{2\sqrt{\ell_0}} \binom{\lceil (1-4\varepsilon)md \rceil}{q}.$$

Now we estimate the cardinalities of preimages. Let $M' \in R(A)$. Then the set $B_3 = B_0 \cap I^r(M')$ must have cardinality q. Write $B_3 = \{i_1, i_2, \ldots i_q\}$ with $i_1 < i_2 < \ldots < i_q$. Let $w_1, \ldots, w_q$ be elements of J^r such that M' has ones on positions (i_s, w_s) for $s \leq q$. If $(M, M') \in R$, M has to have zeros on those positions. We now compute how many such matrices $M \in \mathcal{F}(I,V)$ can be constructed, that is, how many possibilities to have ones in rows i_s, $s \leq q$, exist. Since $M' \in R(A)$, we have

$$|I^\ell(M') \setminus B_0| = |I^\ell(M')| - (|B_0| - q) \leq md.$$

Choose $B_4 \subset I^\ell(M') \setminus B_0$ of cardinality q. Write $B_4 = \{v_1, v_2, \ldots, v_q\}$ with $v_1 < v_2 < \ldots < v_q$. Let $j_1, \ldots, j_q$ be elements of J^r such that M' has ones on positions (v_s, j_s) for $s \leq q$. Then M is obtained from M' by substituting zeros with ones on places (i_s, j_s) and (v_s, w_s) and substituting ones with zeros on places (i_s, w_s) and (v_s, j_s) for all $s \leq q$. Thus, $|R^{-1}|$ is bounded above by the number of choices for the set B_4, that is $|R^{-1}(A)| \leq \binom{md}{q}$. Using that for every integers N and s with $N - s > q$ one has

$$\frac{\binom{N}{q}}{\binom{N-s}{q}} = \frac{N \ldots (N-s+1)}{(N-s)\ldots(N-s-q+1)} \leq \left(\frac{N-s+1}{N-s-q+1}\right)^s \leq \exp\left(\frac{sq}{N-s-q+1}\right),$$

that $q = \lfloor \ell_0/2 \rfloor \le t/2$, $t \le t_1 - t_0 \le (2 - 4/169)md$, and Claim 2.3 we observe that

$$\frac{|A|}{|\mathcal{F}(I,V)|} \le \frac{2\sqrt{\ell_0}}{2^{\ell_0}} \exp\left(\frac{q\, 4\varepsilon md}{(1-4\varepsilon)md - q + 1}\right) \le \frac{\sqrt{2t}}{2^{t/2}} \exp\left(\frac{2\varepsilon tmd}{(1-4\varepsilon)md - t/2}\right)$$

$$\le \frac{\sqrt{2t}}{2^{t/2}} \exp\left(\frac{85\varepsilon t}{1 - 169\varepsilon}\right) \le e^{-t/3},$$

provided that ε is small enough and d (hence t) is large enough. $\square$

3.6 Proof of Theorem 3.2

We are ready to complete the proof.

Proof of Theorem 3.2. Recall that d is large enough, $\varepsilon_0 = \sqrt{(\log d)/d}$, $p = \lfloor 1/5\varepsilon_0 \rfloor$, and let ε be a small positive constant from Lemma 3.9. In most formulas below we assume that $n_0 > 1$, otherwise $T_0 = \emptyset$ and the proof is easier. We make corresponding remarks in the text. Below we deal with matrices from

$$\Omega_0 = \bigcap_{j=2}^{r+1} \Omega_{p^j, \varepsilon_0} \cap \Omega_{k_1, \varepsilon_0} \cap \bigcap_{i=1}^{2} \Omega_{2n_i, \varepsilon},$$

where $k_1 = \min\{n_0, p\}$ and where we do not have the first intersection if $n_1 = n_0 \le p$ and we do not have the second term if $n_1 = n_0 = 1$.

If $x \in \mathcal{T}_0$ and $M \in \Omega_0$ then Lemma 3.7 implies that

$$\|(M - z\mathrm{Id})x\|_2 \ge \min\left\{\sqrt{d/8n},\ \frac{p\sqrt{n_1 d}}{8h_{r+1}}\right\} \|x\|_2.$$

We turn now to the case $x \in \mathcal{T}_i$ for $i = 1, 2$. Let

$$\mathcal{E}_i := \left\{ M \in \mathcal{M}_{n,d} \,:\, \exists\, x \in \mathcal{T}_i \text{ such that } \|(M - z\mathrm{Id})x\|_2 \le \frac{\sqrt{n_i d}}{18\, b_i} \|x\|_2 \right\},$$

where $b_1 = h_{r+1}$ and $b_2 = d^{3/2} h_{r+1}$ in the case $n_0 > 1$ and $b_2 = d\sqrt{n}$ in the case $n_0 = 1$. By Lemma 3.5 for $x \in \mathcal{T}_i$ one has $\|x\|_2 \le b_i x^*_{n_i}$. Thus, for $M \in \mathcal{E}_i$ there exists $x = x(M) \in \mathcal{T}_i$ with

$$\|(M - z\mathrm{Id})x\|_2 \le \frac{\sqrt{n_i d}}{18}\, x^*_{n_i}.$$

Normalizing $x \in \mathcal{T}_i$, so that $x^*_{n_i} = 1$ (that is, $x \in \mathcal{T}_i'$), we observe that there exists $y = y(x)$ from the net constructed in Lemma 3.8 with $y^*_{n_i} \ge 1 - d^{-3/2} > 3/4$, and $y^*_j \le 1/4$ for $j > n_{i+1}$ and such that

$$\|x - y\|_2 \le \sqrt{n}\, \|x - y\|_\infty \le d^{-3/2}\sqrt{n} \le \frac{1}{108}\sqrt{n_i/d}.$$

Therefore, using that $\|M\| = d$ and $|z| \le d$, we have

$$\|(M - z\mathrm{Id})y\|_2 \le \|(M - z\mathrm{Id})x\|_2 + (\|M\| + |z|)\|x - y\|_2 \le \sqrt{n_j d}/13.$$

Now we use the union bound over vectors in the net together with individual probability bounds. Lemmas 3.9 and 3.8 imply for $i = 1, 2$,

$$\mathbb{P}(\mathcal{E}_i \cap \Omega_0) \le \exp(-n_i d/4).$$

Combining all cases we obtain that for $x \in \mathcal{T}$ one has $\|(M - z\mathrm{Id})x\|_2 \le A\|x\|$, where

$$A := \min\left(\frac{\sqrt{d}}{2\sqrt{2n}}, \frac{\sqrt{n_1 d}}{18b_1}, \frac{\sqrt{n_2 d}}{18b_2}\right),$$

with probability at most $p_0 := \mathbb{P}(\Omega_0^c) + \exp(-n_1 d/4) + \exp(-n_2 d/4)$.

We first estimate A. If $n_0 = n_1 = 1$ then $\mathcal{T}_0 = \emptyset$, $d^2 > n$, and $h_{r+1} = \sqrt{n}$. Therefore

$$\frac{\sqrt{n_1 d}}{18b_1} = \frac{\sqrt{d}}{18\sqrt{n}} \quad \text{and} \quad \frac{\sqrt{n_2 d}}{18b_2} \ge \frac{\sqrt{a_2}}{20d},$$

which implies that $A \ge c/d$ in this case. If $n_1 > 1$ then $n_1 d \ge a_1 n \log d/d \ge a_2 n/d \approx n_2$. Therefore, in the case $1 < n_0 = n_1 \le p$, one has

$$A \ge \sqrt{a_2 n}/(20d^{3/2}h_{r+1}) \ge \sqrt{a_2 n}/(40d^3\sqrt{\log d}),$$

while in the case $n_0 > p$, using that by (5), $n_1 \le a_1\sqrt{\log d}\, n/5d^{3/2}$,

$$A \ge \frac{\sqrt{a_2 n}}{20d^{3/2}h_{r+1}} \ge \frac{\sqrt{a_2 n}}{20\sqrt{3p}\, d^{3/2} n_1^{2+\alpha_d}} \ge \frac{\sqrt{a_2 n}\, d^{3+3\alpha d/2}}{3\sqrt{p}\, d^{3/2} a_1^3 n^{2+\alpha_d}} \ge \frac{\sqrt{a_2}\, d^{5/4}\log^2 d}{3a_1^3 n^{3/2+\alpha_d}}.$$

We now estimate the probability p_0 using Theorem 2.4. Recall that c_1, c_2, ... always denote (sufficiently small) positive absolute constants. First note that Theorem 2.4 implies

$$p_1 := \sum_{i=1}^{2}\left(\mathbb{P}\left(\Omega_{2n_i,\varepsilon}^c\right) + \exp(-n_i d/4)\right)$$

$$\le \sum_{i=1}^{2}\left(\exp\left(-\frac{\varepsilon^2 d n_i}{4}\log\left(\frac{ec_{2.4}\varepsilon n}{2dn_i}\right)\right) + \exp\left(-\frac{n_i d}{4}\right)\right) \le \exp(-c_1 n_1 d).$$

In the case $n_1 = n_0 = 1$ we have $a_1 n \le d^2/\log d$ and hence $p_1 \le \exp(-c_2\sqrt{n})$. In the case $n_1 > 1$ we have $a_1 n \log d \ge d^2$, hence

$$n_1 d \ge n_0 d \ge (a_1 n \log d)/d \ge \sqrt{a_1 n \log d},$$

thus again $p_1 \le \exp(-c_2\sqrt{n})$.

In the case $1 < n_0 = n_1 \le p$ we have $k_1 = n_1$, $a_1 n \log d \ge d^2$, and $a_1 n \log^{3/2} d \le d^{2.5}$. Therefore, by Theorem 2.4,

$$p_2 := \mathbb{P}\left(\Omega_{k_1,\varepsilon_0}^c\right) \le \exp\left(-\frac{n_1\log d}{8}\log\left(\frac{ec_{2.4}n\log d}{d^{3/2}n_1}\right)\right) \le \exp\left(-c_3\log^2 n\right).$$

Recall that in the definition of Ω_0 we do not have the first intersection if $n_1 = n_0 \le p$ and we do not have the second term if $n_1 = n_0 = 1$. This implies that in the case $n_1 \le p$ we have $p_0 \le p_1 + p_2 \le \exp\left(-c_4 \log^2 n\right)$.

Finally, in the case $n_1 > p$, we have $k_1 = p$, $r \ge 1$, and, $c_4 n \ge d^{5/2}/\log^{3/2} d$. Therefore, by Theorem 2.4,

$$p_3 := \sum_{i=2}^{r+1} \mathbb{P}\left(\Omega^c_{p^j,\varepsilon_0}\right) + \mathbb{P}\left(\Omega^c_{k_1,\varepsilon_0}\right) \le \sum_{i=1}^{r+1} \exp\left(-\frac{p^i \log d}{8} \log\left(\frac{ec_{2.4}\varepsilon_0 n}{dp^i}\right)\right)$$

$$\le \exp\left(-\frac{p\log d}{9} \log\left(\frac{ec_{2.4}\varepsilon_0 n}{dp}\right)\right) \le \exp\left(-c_5 \sqrt{d \log d}\, \log n\right).$$

Since $p_0 \le p_1 + p_2 + p_3$, the desired estimate follows. $\square$

4 Bounds for essentially non-constant vectors and completing the proof of the main theorem

In this section, we complete our proof of the lower bound for the smallest singular value of a random matrix uniformly distributed in $\mathcal{M}_{n,d}$, shifted by $z\,\mathrm{Id}$ for a fixed $z \in \mathbb{C}$. To better separate various techniques we use in this paper, we prefer to give an "autonomous" proof of the result, conditioned on a rather general assumption about the structure of the kernel of our random matrix. This assumption, for a specific choice of parameters, is actually proved in Section 3 (see Remark 3.3), so the argument presented here implies the main result of the paper regarding the magnitude of s_n. We provide the details in Section 4.4.

We start by introducing notations. Fix an $n \times n$ (complex) matrix W. Further, take positive parameters $\kappa, \rho \in (0,1)$, and $\delta \in (1/n, 1)$ (the parameters may and in fact *will* depend on n and d). Define the subset $S(\rho,\delta)$ of the unit sphere in $\mathbb{C}^n$ by

$$S(\rho,\delta) := \Big\{x \in \mathbb{C}^n :\ \|x\|_2 = 1 \text{ and } \forall \lambda \in \mathbb{C} \quad |\{i \le n :\ |x_i - \lambda| \le \rho\}| \le \delta n\Big\}$$

(note that $S(\rho,\delta) = (\mathbb{C}^n \setminus \mathcal{B}(\rho)) \cap \{x \in \mathbb{C}^n :\ \|x\|_2 = 1\}$ and that $S(\rho,\delta) = \emptyset$ for $\delta < 1/n$). Further, define two events

$$\mathcal{E}_4(W,\kappa,\rho,\delta) := \big\{M \in \mathcal{M}_{n,d} :\ \forall\, x \in \mathbb{C}^n \text{ with } \|x\|_2 = 1 \text{ and }$$
$$\min(\|(M+W)x\|_2, \|x^\dagger(M+W)\|_2) \le \kappa \text{ one has } x \in S(\rho,\delta)\big\},$$

and

$$\mathcal{E}_{4.1}(W,\kappa) := \big\{M \in \mathcal{M}_{n,d} :\ s_n(M+W) \le \kappa\big\}.$$

The parameters W, ρ, δ, κ are usually clear from the context, and we will simply write $\mathcal{E}_4$ and $\mathcal{E}_{4.1}$ to denote the respective events.

Theorem 4.1. *There exist positive absolute constants c, C_0, and C with the following property. Let $\delta \in (1/n, 1)$, $\rho \in (0,1)$, $\kappa := \rho^2/16$, and*

$$C \le d \le \frac{c(1-\delta)}{\log(e/(1-\delta))}\, n.$$

Further, assume that W is a complex matrix such that the event $\mathcal{E}_4 = \mathcal{E}_4(W, \kappa, \rho, \delta)$ has probability at least $1 - 1/n^2$. Then

$$\mathbb{P}(\mathcal{E}_{4.1}) \leq \frac{C_0\sqrt{\log(e/(1-\delta))}}{(1-\delta)^{3/2}} \, \frac{1}{\sqrt{d}}.$$

One can describe the structure of the above theorem as follows: provided that for a random matrix M uniformly distributed in $\mathcal{M}_{n,d}$, vectors "close" to the kernel of M are unstructured (i.e., not almost constant), the smallest singular value of M is at least κ with large probability (later we choose κ to be a (negative) constant power of n). Theorem 4.1 should be compared with recent results of [11, 4] which were already discussed in the introduction. Note that the high-level structure of the theorem is similar to [11, Lemma 6.2], where invertibility properties of the random matrix are also derived conditioned on a "good" event which encapsulates properties of "almost null" vectors of the matrix. At the same time, technical details of both proofs are in many respects different.

4.1 Some relations for random square matrices

In this subsection we present two lemmas – one probabilistic and the other linear algebraic – which work for a wide class of square matrices. The next lemma is analogous to [32, Lemma 3.5]. The proof follows the same lines, and we include it for the sake of completeness.

Lemma 4.2. *Fix parameters $\rho, \delta, \delta_0, \varepsilon > 0$, and assume that $1/n \leq \delta < \delta_0 \leq 1$. Further, let*

$$K_0 \subset \{(i,j) \, : \, 1 \leq i \neq j \leq n\}$$

be such that $|K_0| \geq \delta_0 n(n-1)$. Let A be an $n \times n$ random matrix on some probability space such that $\sum_{i=1}^n R_i(A) = v$ a.s. for a fixed vector $v \in \mathbb{C}^n$. Then

$$\mathbb{P}\big\{ \inf_{x \in S(\rho,\delta)} \|x^\dagger A\|_2 \leq \varepsilon\rho \big\} \\ \leq \frac{1}{n^2(\delta_0 - \delta)} \sum_{(i,j)\in K_0} \mathbb{P}\big\{ \mathrm{dist}\big(R_i(A), \mathrm{span}\,\big\{\{R_k(A)\}_{k\neq i,j}, \, R_i(A) + R_j(A)\big\}\big) < \varepsilon \big\}.$$

Proof. In this proof we denote $R_i(A)$ just by R_i ($i \leq n$), and set $K := \{(i,j) \, : \, 1 \leq i \neq j \leq n\}$. Without loss of generality we assume that $\sum_{i=1}^n R_i = v$ everywhere on the probability space. For each pair $(i,j) \in K$, set

$$d_{ij} = d_{ij}(A) := \mathrm{dist}\big(R_i, \mathrm{span}\,\big\{\{R_k\}_{k\neq i,j}, \, R_i + R_j\big\}\big).$$

Note that $d_{ij} = \mathrm{dist}\big(R_i, \mathrm{span}\,\big\{\{R_k\}_{k\neq i,j}, \, v\big\}\big)$. Since $x^\dagger A = \sum_{k=1}^n \bar{x}_k R_k$, for every $(i,j) \in K$ we have

$$\|x^\dagger A\|_2 = \Big\| (\bar{x}_i - \bar{x}_j) R_i + \bar{x}_j v + \sum_{k\neq i,j} (\bar{x}_k - \bar{x}_j) R_k \Big\|_2 \geq |\bar{x}_i - \bar{x}_j| d_{ij}.$$

The above relation is the principal point of the proof. Now, if "many" distances d_{ij} are "large", then, since the vector x is essentially non-constant, we can find a pair (i,j) such that both $|\bar{x}_i - \bar{x}_j|$ and d_{ij} are large, and we get a lower bound $\|x^\dagger A\|_2 > \varepsilon\rho$. Thus, we can estimate the probability of the considered event in terms of probability that "not so many" distances d_{ij} are large which is in turn done via Markov's inequality. Below is a rigorous argument.

Let $K_1 := \{(i,j) \in K_0 : d_{ij} \geq \varepsilon\}$. Denote by $\mathcal{E}$ the event that $|K_1| > \delta n^2 - n$. Note that if $M \in \mathcal{E}^c$, we have

$$|\{(i,j) \in K_0 : d_{ij} < \varepsilon\}| \geq \delta_0 |K| - \delta n^2 + n = (\delta_0 - \delta)n^2 + n(1-\delta_0) \geq (\delta_0 - \delta)n^2.$$

Therefore, using Markov's inequality,

$$\mathbb{P}(\mathcal{E}^c) \leq \frac{\mathbb{E}(|\{(i,j) \in K_0 : d_{ij} < \varepsilon\}|)}{n^2(\delta_0 - \delta)} = \frac{1}{n^2(\delta_0 - \delta)} \sum_{(i,j) \in K_0} \mathbb{P}\{d_{ij} < \varepsilon\}.$$

Now, we condition on the event $\mathcal{E}$. Fix a vector $x \in S(\rho,\delta)$. By the definition of $S(\rho,\delta)$,

$$|\{(i,j) \in [n] \times [n] : |\bar{x}_i - \bar{x}_j| \leq \rho\}| \leq \delta n^2,$$

therefore the set $K_2 = K_2(x) := \{(i,j) \in [n] \times [n] : |\bar{x}_i - \bar{x}_j| > \rho\}$ contains at least $n^2(1-\delta)$ elements. Clearly, $K_2 \subset K$. Thus we have $K_1 \cup K_2 \subset K$ and $|K_1| + |K_2| > \delta n^2 - n + n^2(1-\delta) = n(n-1) = |K|$. Hence $K_1 \cap K_2 \neq \emptyset$. Choose $(i_0, j_0) \in K_1 \cap K_2$. Then

$$\|x^\dagger A\|_2 \geq |\bar{x}_{i_0} - \bar{x}_{j_0}| d_{i_0 j_0} > \rho\,\varepsilon.$$

Summarizing, we have shown that

$$\mathbb{P}\big\{\inf_{x \in S(\rho,\delta)} \|x^\dagger A\|_2 \leq \varepsilon\rho\big\} \leq \mathbb{P}(\mathcal{E}^c) \leq \frac{1}{n^2(\delta_0 - \delta)} \sum_{(i,j) \in K_0} \mathbb{P}\{d_{ij} < \varepsilon\}.$$

□

The above lemma will be used to reduce the question of bounding the smallest singular value to estimating distances between rows or columns of our random matrix and certain linear subspaces of $\mathbb{C}^n$. In order to estimate the distance between the first row R_1 and span $\{R_1 + R_2, R_3, R_4, \ldots, R_n\}$ of a random matrix, we will need the following lemma.

Lemma 4.3. *Let A be an $n \times n$ matrix (either deterministic or random) and denote $R_i := R_i(A)$, $i \leq n$. Further, let $A^{1,2}$ be the $(n-2) \times n$ matrix obtained by removing the first two rows of A, and let Y be the linear span (in $\mathbb{C}^n$) of $R_1 + R_2$, R_3, $R_4, \ldots, R_n$. Then for every unit complex vector $v \in \mathbb{C}^n$ we have*

$$\operatorname{dist}\big(R_1, Y\big) \geq \frac{s_n(A)\,|\langle R_1^\dagger, v\rangle|}{s_n(A) + \|A^{1,2}v\|_2 + |\langle R_1^\dagger + R_2^\dagger, v\rangle|}.$$

In particular, if a unit complex vector $v \in \mathbb{C}^n$ satisfies

$$\|A^{1,2}v\|_2 \leq s_n(A) \quad \text{and} \quad |\langle R_1^\dagger + R_2^\dagger, v\rangle| \leq 2 s_n(A)$$

then

$$\operatorname{dist}\big(R_1, Y\big) \geq |\langle R_1^\dagger, v\rangle|/4.$$

Proof. Let x be an arbitrary vector from Y, i.e. $x = b(R_1 + R_2) + \sum_{i=3}^{n} a_i R_i$ for some $b, a_3, a_4, \ldots, a_n \in \mathbb{C}$. Fix a unit vector $v \in \mathbb{C}^n$ and let $\bar{v}$ be the vector of complex conjugates for the coordinates of v. We clearly have

$$\|R_1 - x\|_2 \geq |\langle R_1 - x, \bar{v}\rangle| \geq |\langle R_1, \bar{v}\rangle| - |\langle x, \bar{v}\rangle|. \tag{10}$$

Consider the vector $y := (1 - b, -b, -a_3, \ldots, -a_n)$. Then, up to transposition, $R_1 - x = A^T y$, whence

$$\|R_1 - x\|_2 \geq s_n(A^T)\|y\|_2 = s_n(A)\|y\|_2.$$

Therefore, using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}|\langle x, \bar{v}\rangle| &\leq |b||\langle R_1 + R_2, \bar{v}\rangle| + \Big(\sum_{i=3}^{n} |a_i|^2\Big)^{\frac{1}{2}} \Big(\sum_{i=3}^{n} |\langle R_i, \bar{v}\rangle|^2\Big)^{\frac{1}{2}} \\ &\leq \|y\|_2 \left(|\langle R_1 + R_2, \bar{v}\rangle| + \|A^{1,2} v\|_2\right) \\ &\leq \frac{1}{s_n(A)} \|R_1 - x\|_2 \left(|\langle R_1 + R_2, \bar{v}\rangle| + \|A^{1,2} v\|_2\right).\end{aligned}$$

This, together with (10), implies that

$$\|R_1 - x\|_2 \geq \frac{s_n(A)\, |\langle R_1, \bar{v}\rangle|}{s_n(A) + |\langle R_1 + R_2, \bar{v}\rangle| + \|A^{1,2} v\|_2}.$$

The lemma follows by noting that $|\langle R_1, \bar{v}\rangle| = |\langle R_1^\dagger, v\rangle|$, $|\langle R_1 + R_2, \bar{v}\rangle| = |\langle R_1^\dagger + R_2^\dagger, v\rangle|$, and by taking the infimum over $x \in Y$. $\square$

Observe that for a unit vector v_0 orthogonal to the span of $R_1 + R_2$, R_3, $R_4, \ldots$, R_n, we have $A^{1,2}\bar{v}_0 = 0$ and $\langle R_1 + R_2, v_0\rangle = 0$ so the lemma applied to $\bar{v}_0$ gives a trivial bound $\operatorname{dist}\big(R_1, Y\big) \geq |\langle R_1, v\rangle|$. Thus the above statement can be viewed as a "continuous" version of this trivial estimate.

4.2 Proof of Theorem 4.1

For the rest of the section, we fix a function f on the set of $n \times n$ complex matrices, which associates with every matrix A a complex vector $f(A)$ such that $\|Af(A)\|_2 = s_n(A)$. Note that in general the corresponding singular vector is not uniquely defined. For those matrices, we fix *some* vector $f(A)$ satisfying the above condition. Since we work with shifted matrices, we also adopt another notation: given a (fixed) complex matrix W, by f_W we denote the function on the set of $n \times n$ matrices defined by $f_W(A) := f(A + W)$.

Fix parameters $\kappa, \rho > 0$, $\delta \in (1/n, 1 - 1/\sqrt{d})$ and a complex matrix W (note that the bound for probability in Theorem 4.1 for $\delta \geq 1 - \sqrt{d}$ becomes greater than one, hence the theorem holds automatically). For the rest of the section, we assume that the parameters are given, and will specify each time what restrictions on the numbers κ, ρ, δ, d and the matrix W we impose. Further, define

$$\varepsilon_1 = \varepsilon_1(\delta) := (1 - \delta)\left(C_1 \log(2e/(1 - \delta))\right)^{-1},$$

where C_1 is a sufficiently large absolute constant (it is enough to take the constant from Proposition 2.6 multiplied by 9). Set $\alpha := (1-\delta)/(9\varepsilon_1 d)$ and $\beta := (1-\delta)/2$. Note that with such a choice of α, β we have $\alpha \geq (C\log(e/\beta))/d$ and, using that $\delta < 1-1/\sqrt{d}$ and that d is large enough, we also have $\alpha \leq \min(\beta, 1/4)$. In other words the conditions of Proposition 2.6 are satisfied. Let $\Omega_0 = \Omega_0(\alpha,\beta)$ and $\Omega_1(\varepsilon_1)$ be the events defined in and after Proposition 2.5. Define the event

$$\mathcal{E}_0 = \mathcal{E}_0(W,\kappa,\rho,\delta) := \Omega_0^c \cap \Omega_1(\varepsilon_1) \cap \mathcal{E}_4.$$

In words, $\mathcal{E}_0$ corresponds to the set of matrices in $\mathcal{M}_{n,d}$ without large zero minors, with almost no overlap between supports of any two rows or columns, and with the structural assumption on vectors "close" to the kernel of the respective shifted matrix. Note that under assumptions of Theorem 4.1, together with Propositions 2.5 and 2.6, we have

$$\mathbb{P}(\mathcal{E}_0^c) \leq 2n^{-2} \tag{11}$$

(the assumption on d in Theorem 4.1 comes from $d \leq \varepsilon_1 n/6$ needed in Propositions 2.5).

The next lemma shows, roughly speaking, that there are relatively few matrices $M \in \mathcal{M}_{n,d}$ such that the corresponding singular vector $f_W(M)$ is "almost constant" when restricted to supports of a large number of rows of M.

Lemma 4.4. *Assume that d is large enough. For every pair of indices $\ell \neq i$ define the event*

$$\begin{aligned}\mathcal{E}_{4.4}^{\ell,i} :=\big\{&M \in \mathcal{E}_{4.1}\cap\mathcal{E}_0 \,:\, \exists\lambda\in\mathbb{C} \text{ such that}\\ &\big|\{j\in \operatorname{supp}(R_\ell(M)+R_i(M)) \,:\, |(f_W(M))_j-\lambda|\leq \rho/4\}\big| > 2(1-2\varepsilon_1)d\big\}.\end{aligned}$$

Then for every (fixed) $\ell \leq n$ one has

$$\sum_{i:\, i\neq\ell} |\mathcal{E}_{4.4}^{\ell,i}| \leq \frac{(1-\delta)n}{9\varepsilon_1 d}\,|\mathcal{M}_{n,d}|.$$

Proof. Without loss of generality, we can assume that $\ell = 1$. Let $\mathcal{E}$ denote the event

$$\begin{aligned}\big\{&M\in\mathcal{E}_{4.1}\cap\mathcal{E}_0 \,:\, \exists\lambda\in\mathbb{C} \text{ with}\\ &\big|\{j\in\operatorname{supp}R_1(M) : |(f_W(M))_j-\lambda|\leq\rho/2\}\big| > (1-4\varepsilon_1)d\big\}.\end{aligned}$$

Note that $\mathcal{E}_{4.4}^{1,i} \subset \mathcal{E}$ for every $i \geq 2$. Indeed if $M \in \mathcal{E}_{4.4}^{1,i}$ for some $i \geq 2$, then there exists $\lambda \in \mathbb{C}$ such that

$$\big|\{j\in\operatorname{supp}(R_1(M)+R_i(M)) \,:\, |(f_W(M))_j-\lambda|\leq\rho/4\}\big| > 2(1-2\varepsilon_1)d.$$

Therefore

$$\begin{aligned}\big|\{j\in\operatorname{supp}R_1(M) : |(f_W(M))_j-\lambda|\leq\rho/2\}\big| &\geq \big|\{j\in\operatorname{supp}R_1(M) : |(f_W(M))_j-\lambda|\leq\rho/4\}\big|\\ &\geq \big|\{j\in\operatorname{supp}(R_1(M)+R_i(M)) \,:\, |(f_W(M))_j-\lambda|\leq\rho/4\}\big| - \big|\operatorname{supp}(R_i(M)\big|\\ &\geq (1-4\varepsilon_1)d,\end{aligned}$$

which means that M belongs to $\mathcal{E}$. For every $M\in\mathcal{E}$, fix a number $\lambda_0=\lambda_0(M)\in\mathbb{C}$ such that

$$\big|\big\{j\in \mathrm{supp} R_1(M):\, |(f_W(M))_j-\lambda_0|\le \rho/2\big\}\big|>(1-4\varepsilon_1)d. \tag{12}$$

Now, take any $M\in\mathcal{E}$ and let

$$J_M:=\{j\le n:\, |(f_W(M))_j-\lambda_0|\le\rho\}.$$

Since $M\in\mathcal{E}_0\subset\mathcal{E}_4$ (i.e., all vectors "close" to the kernel of $M+W$ are essentially non-constant) and $\|(M+W)f_W(M)\|_2\le\kappa$, we have $|J_M|\le\delta n$. Let also

$$I_M:=\{i\le n:\, |\mathrm{supp} R_i(M)\cap J_M|\ge(1-4\varepsilon_1)d\}.$$

We first show that $|I_M|\le\frac{(1-\delta)n}{9\varepsilon_1 d}$. Assume the opposite. Choose a set $\widetilde{I}\subset I_M$ with $|\widetilde{I}|=\left\lceil\frac{(1-\delta)n}{9\varepsilon_1 d}\right\rceil$. Clearly,

$$\forall (i,j)\in\widetilde{I}\times\big(\cup_{i\in\widetilde{I}}\mathrm{supp} R_i(M)\big)^c\quad\text{one has}\quad \mu_{ij}=0,$$

and

$$\begin{aligned}\big|\big(\cup_{i\in\widetilde{I}}\mathrm{supp} R_i(M)\big)^c\big| &\ge n-|J_M|-\big|\cup_{i\in\widetilde{I}}\mathrm{supp} R_i(M)\setminus J_M\big|\\ &\ge n-\delta n-4\varepsilon_1 d|\widetilde{I}|\\ &\ge (1-\delta)n/2.\end{aligned}$$

This contradicts the assumption $M\in\Omega_0^c$ (no large zero-minors).

By the definition of I_M, for every $i\in(I_M)^c$,

$$|\{j\in\mathrm{supp} R_i(M)\ :\ |(f_W(M))_j-\lambda_0|>\rho\}|\ge 4\varepsilon_1 d.$$

This implies for every λ satisfying $|\lambda-\lambda_0|\le 3\rho/4$,

$$|\{j\in\mathrm{supp}(R_1(M)+R_i(M))\ :\ |(f_W(M))_j-\lambda|>\rho/4\}|\ge 4\varepsilon_1 d.$$

Using the triangular inequality together with (12), we also observe that for every λ satisfying $|\lambda-\lambda_0|>3\rho/4$,

$$\begin{aligned}&|\{j\in\mathrm{supp}(R_1(M)+R_i(M))\ :\ |(f_W(M))_j-\lambda|>\rho/4\}|\\ &\qquad\ge|\{j\in\mathrm{supp}(R_1(M)+R_i(M))\ :\ |(f_W(M))_j-\lambda_0|>\rho/2\}|\\ &\qquad\ge|\{j\in\mathrm{supp} R_1(M)\ :\ |(f_W(M))_j-\lambda_0|>\rho/2\}|\\ &\qquad\ge(1-4\varepsilon_1)d\ge 4\varepsilon_1 d.\end{aligned}$$

Thus for every $i\in I_M^c$ and every $\lambda\in\mathbb{C}$ we obtain

$$\big|\big\{j\in\mathrm{supp}(R_1(M)+R_i(M))\ :\ |(f_W(M))_j-\lambda|\le\rho/4\big\}\big|\le|\mathrm{supp}(R_1(M)+R_i(M))|-4\varepsilon_1 d.$$

This proves that for every $M\in\mathcal{E}$ and $i\in I_M^c$ one has $M\in\mathcal{E}\setminus\mathcal{E}_{4.4}^{1,i}$. Therefore,

$$\sum_{i=2}^n|\mathcal{E}_{4.4}^{1,i}|=\sum_{i=2}^n\sum_{M\in\mathcal{E}}\chi_{\{M\in\mathcal{E}_{4.4}^{1,i}\}}=\sum_{M\in\mathcal{E}}\sum_{i=2}^n\chi_{\{M\in\mathcal{E}_{4.4}^{1,i}\}}\le\sum_{M\in\mathcal{E}}|I_M|\le\frac{(1-\delta)n\,|\mathcal{E}|}{9\varepsilon_1 d}.$$

□

Remark 4.5. *Note that for every $i \neq \ell$ and every matrix $M \in (\mathcal{E}_{4.1} \cap \mathcal{E}_0) \setminus \mathcal{E}_{4.4}^{\ell,i}$, one has*

$$\begin{aligned}\forall \lambda \in \mathbb{C} \quad & \big|\{j \in \mathrm{supp} R_\ell(M) \triangle \mathrm{supp} R_i(M) \,:\, |(f_W(M))_j - \lambda| > \rho/4\}\big| \\ & > |\mathrm{supp} R_i(M) \triangle \mathrm{supp} R_\ell(M)| - 2d + 4\varepsilon_1 d \\ & \geq 2\varepsilon_1 d,\end{aligned}$$

where $\triangle$ denotes the symmetric difference of sets.

The next observation is a direct consequence of Lemma 4.2 and Lemma 4.4.

Corollary 4.6. *Assume that $1/n \leq \delta < 1$, and that d satisfies the assumptions of Lemma 4.4. Then there exists a pair $(\ell, j) \in [n] \times [n]$ with $\ell \neq j$ such that*

$$|\mathcal{E}_{4.4}^{\ell,j}| \leq \frac{1}{4\varepsilon_1 d}\, |\mathcal{M}_{n,d}| \tag{13}$$

and, setting $R_i^W := R_i(M+W)$, $i \leq n$, for an (implicitly) given matrix $M \in \mathcal{M}_{n,d}$, we have for any $\varepsilon > 0$:

$$\begin{aligned}&\Big|\big\{M \in \mathcal{E}_{4.1} \cap \mathcal{E}_0 \,:\, \inf_{x \in S(\rho,\delta)} \|x^\dagger (M+W)\|_2 \leq \varepsilon\rho\big\}\Big| \\ &\quad \leq \frac{2}{1-\delta}\, \Big|\big\{M \in \mathcal{E}_{4.1} \cap \mathcal{E}_0 \,:\, \mathrm{dist}\big(R_\ell^W, \mathrm{span}\,\big\{\{R_k^W\}_{k \neq \ell, j},\, R_\ell^W + R_j^W\big\}\big) < \varepsilon\big\}\Big|.\end{aligned}$$

Proof. Denote $K := \{(\ell, j) \,:\, 1 \leq \ell \neq j \leq n\}$. Set $\delta_0 = (1+\delta)/2$. Lemma 4.4 implies that for every fixed $\ell \leq n$ there are at least $\delta_0(n-1)$ choices of $j \neq \ell$ satisfying (13). Therefore, the subset

$$K_0 := \big\{(\ell, j) \in K \,:\, (\ell, j) \;\text{ satisfies }\; (13)\big\}$$

has cardinality at least $\delta_0 n(n-1) = \delta_0 |K|$. Choosing a pair $(\ell, j) \in K_0$ with maximal

$$\Big|\big\{M \in \mathcal{E}_{4.1} \cap \mathcal{E}_0 \,:\, \mathrm{dist}\big(R_\ell^W, \mathrm{span}\,\big\{\{R_k^W\}_{k \neq \ell, j},\, R_\ell^W + R_j^W\big\}\big) < \varepsilon\big\}\Big|$$

and applying Lemma 4.2 to the random matrix $A = M + W$, where M is uniformly distributed in $\mathcal{E}_{4.1} \cap \mathcal{E}_0$, we obtain the desired result. □

Corollary 4.6 reduces the question of bounding the infimum over "non-constant" vectors to calculating the distance between a particular matrix row and corresponding linear span, and additionally makes sure that the singular vector $f_W(M)$ is essentially non-constant when restricted to the union of the supports of j-th and ℓ-th rows. The latter allows to apply Littlewood–Offord–type anti-concentration statements. Note that, instead of bounding the cardinality of the event $\mathcal{E}_{4.1}$ directly, we will bound the cardinality of the intersection of $\mathcal{E}_{4.1}$ with a "good" event $\mathcal{E}_0$, and then use the fact that $\mathcal{E}_0^c$ is small (under the assumptions of the theorem).

We are now ready to describe a partition of the event $\Omega_0^c \cap \Omega_1(\varepsilon_1)$, which will be used in the proof of Theorem 4.1. Fix d, parameters ρ, δ and complex matrix W. Let κ be defined as in Theorem 4.1 and assume that all the conditions of the theorem (including assumptions on the parameters) are satisfied. Let the pair (ℓ, j) be given by Corollary 4.6. From now on, to simplify notation, we will assume that $(\ell, j) = (1, 2)$. We would like to emphasize that the proof below can be carried for any admissible pair (ℓ, j) by simply adjusting indices.

Consider a set of $(n-2) \times n$ matrices

$$\mathcal{H} := \{M^{1,2} : M \in \Omega_0^c \cap \Omega_1(\varepsilon_1)\}.$$

For every $H \in \mathcal{H}$, let C_H be the equivalence class of matrices sharing the same minor, that is

$$C_H := \{M \in \Omega_0^c \cap \Omega_1(\varepsilon_1) : M^{1,2} = H\}.$$

Note that for $M_1, M_2 \in C_H$ one has $R_1(M_1) + R_2(M_1) = R_1(M_2) + R_2(M_2)$, that is the intersection and the union of the supports of the first two rows is the same for all matrices in the class:

$$S_1 = S_1(H) := \mathrm{supp} R_1(M_1) \cap \mathrm{supp} R_2(M_1) = \mathrm{supp} R_1(M_2) \cap \mathrm{supp} R_2(M_2)$$

and

$$S_2 = S_2(H) := \mathrm{supp} R_1(M_1) \cup \mathrm{supp} R_2(M_1) = \mathrm{supp} R_1(M_2) \cup \mathrm{supp} R_2(M_2).$$

In particular, $|C_H| = \binom{2m}{m}$, where $m = m(H) = |S_2 \setminus S_1|$ is the cardinality of the symmetric difference of the first two rows for any matrix in C_H. Observe that, because our matrices belong to $\Omega_1(\varepsilon_1)$, we have $m(H) \geq 2(1-\varepsilon_1)d$. In every class C_H, fix a subset $\widetilde{C}_H \subset C_H$ of matrices satisfying

$$\forall \widetilde{M} \in \widetilde{C}_H \ \forall M \in C_H \setminus \widetilde{C}_H : \quad s_n(\widetilde{M} + W) \leq s_n(M + W) \ \text{ and } \ \frac{1}{2\sqrt{\varepsilon_1 d}} \leq \frac{|\widetilde{C}_H|}{|C_H|} \leq \frac{1}{\sqrt{\varepsilon_1 d}}.$$

Thus, $\widetilde{C}_H$ is the set of matrices $\widetilde{M}$ delivering a "small" minimal singular value of $\widetilde{M} + W$, compared to other matrices in C_H. Denote $\mathcal{E}_{4.4} := \mathcal{E}_{4.4}^{1,2}$ and define

$$\mathcal{H}_1 := \{H \in \mathcal{H} : \widetilde{C}_H \cap \mathcal{E}_{4.1}^c \neq \emptyset\}, \ \ \mathcal{H}_2 := \{H \in \mathcal{H}_1^c : \widetilde{C}_H \subset \mathcal{E}_4^c \cup \mathcal{E}_{4.4}\}, \ \ \mathcal{H}_3 := \mathcal{H}_1^c \setminus \mathcal{H}_2.$$

Roughly speaking, the set $\mathcal{H}_1$ is the collection of all $(n-2) \times n$ minors such that a vast majority of the corresponding shifted matrices have "large" smallest singular value. The set $\mathcal{H}_2$ is the set of all minors not in $\mathcal{H}_1$ such that the corresponding shifted matrices have "bad" characteristics in regard to their "almost null" vectors as well as the vectors delivering the smallest singular value. Finally, $\mathcal{H}_3$ is all the remaining minors. It is the third category which is the most interesting for us and which will require Littlewood–Offord type anti-concentration arguments.

Consider the partition

$$\Omega_0^c \cap \Omega_1(\varepsilon_1) = \bigcup_{H \in \mathcal{H}_1} C_H \cup \bigcup_{H \in \mathcal{H}_2} C_H \cup \bigcup_{H \in \mathcal{H}_3} C_H. \tag{14}$$

We will analyze separately each of the sets $\bigcup_{H \in \mathcal{H}_i} C_H$, $i \leq 3$. First we show that for $i = 1, 2$ the respective unions have a small cardinality.

By the definition of $\mathcal{E}_{4.1}$, for every $H \in \mathcal{H}_1$ there exists a matrix $M \in \widetilde{C}_H$ with $s_n(M+W) > \kappa$. Hence, by the definition of $\widetilde{C}_H$,

$$|\{M \in C_H : s_n(M+W) \le \kappa\}| \le |\widetilde{C}_H| \le \frac{1}{\sqrt{\varepsilon_1 d}}\,|C_H|,$$

which implies

$$\Big|\bigcup_{H\in\mathcal{H}_1} C_H \cap \mathcal{E}_{4.1}\Big| \le \frac{1}{\sqrt{\varepsilon_1 d}}\,|\mathcal{M}_{n,d}|. \tag{15}$$

Further, by the definitions of $\widetilde{C}_H$ and $\mathcal{H}_2$, the assumptions of Theorem 4.1, and Corollary 4.6, we have

$$\Big|\bigcup_{H\in\mathcal{H}_2} C_H\Big| \le 2\sqrt{\varepsilon_1 d}\sum_{H\in\mathcal{H}_2}|\widetilde{C}_H| \le 2\sqrt{\varepsilon_1 d}\,|\mathcal{E}_4^c \cup \mathcal{E}_{4.4}| \le 2\sqrt{\varepsilon_1 d}\Big(n^{-2}+\frac{1}{4\varepsilon_1 d}\Big)|\mathcal{M}_{n,d}|. \tag{16}$$

Regarding the set $\mathcal{H}_3$, we prove the following lemma.

Lemma 4.7. *Denoting $R_i^W := R_i(M+W)$, $i \le n$, we have*

$$\Big|\Big\{M \in \bigcup_{H\in\mathcal{H}_3} C_H : \operatorname{dist}\big(R_1^W, \operatorname{span}\big\{\{R_k^W\}_{k>2},\, R_1^W + R_2^W\big\}\big) < \rho/16\Big\}\Big| \le C(\varepsilon_1 d)^{-1/2}\,|\mathcal{M}_{n,d}|,$$

where $C>0$ is a universal constant.

Proof. The set $\mathcal{H}_3$ can be equivalently written as

$$\{H \in \mathcal{H} : \widetilde{C}_H \subset \mathcal{E}_{4.1} \quad\text{and}\quad \widetilde{C}_H \cap \mathcal{E}_4 \cap \mathcal{E}_{4.4}^c \neq \emptyset\}.$$

Fix any $H \in \mathcal{H}_3$ and a matrix $\widetilde{M} \in \widetilde{C}_H \cap \mathcal{E}_4 \cap \mathcal{E}_{4.4}^c$. For every $M \in C_H \setminus \widetilde{C}_H$ we have

$$\|(M+W)^{1,2} f_W(\widetilde{M})\|_2 \le s_n(\widetilde{M}+W) \le s_n(M+W)$$

and

$$|\langle (R_1(\widetilde{M}+W))^\dagger + (R_2(\widetilde{M}+W))^\dagger,\, f_W(\widetilde{M})\rangle| \le 2s_n(\widetilde{M}+W) \le 2s_n(M+W).$$

This and Lemma 4.3 applied to the matrix $M+W$ imply that for at least

$$|C_H| - |\widetilde{C}_H| \ge \Big(1-\frac{1}{\sqrt{\varepsilon_1 d}}\Big)|C_H|$$

matrices $M \in C_H$, one has

$$\operatorname{dist}\big(R_1^W, \operatorname{span}\big\{\{R_k^W\}_{k>2},\, R_1^W + R_2^W\big\}\big) \ge |\langle (R_1(M+W))^\dagger,\, f_W(\widetilde{M})\rangle|/4.$$

The following claim, whose proof we postpone, completes the proof. $\square$

Claim 4.8. *With the above notation, for every $H \in \mathcal{H}_3$ and $\widetilde{M} \in \widetilde{C}_H \cap \mathcal{E}_4 \cap \mathcal{E}_{4.4}^c$ we have*

$$|\{M \in C_H : |\langle (R_1(M+W))^\dagger,\, f_W(\widetilde{M})\rangle| < \rho/4\}| \le c(\varepsilon_1 d)^{-1/2}\,|C_H|$$

for some universal constant $c>0$.

Proof of Theorem 4.1. Recall that $\mathcal{E}_0 = \Omega_0^c \cap \Omega_1(\varepsilon_1) \cap \mathcal{E}_4$ and that $\kappa = \rho^2/16$. By (11) we have

$$|\mathcal{E}_{4.1}| \le |\mathcal{E}_{4.1} \cap \mathcal{E}_0| + |\mathcal{E}_0^c| \le |\mathcal{E}_{4.1} \cap \mathcal{E}_0| + 2n^{-2}|\mathcal{M}_{n,d}|.$$

Next, using the definitions of the events $\mathcal{E}_{4.1}$, $\mathcal{E}_4$, and $\mathcal{E}_0$, we observe that

$$\begin{aligned}|\mathcal{E}_{4.1} \cap \mathcal{E}_0| &= \big|\{M \in \mathcal{E}_{4.1} \cap \mathcal{E}_0 : \inf_{\|x\|_2=1} \|x^\dagger(M+W)\|_2 \le \rho^2/16\}\big| \\ &= \big|\{M \in \mathcal{E}_{4.1} \cap \mathcal{E}_0 : \inf_{x \in S(\rho,\delta)} \|x^\dagger(M+W)\|_2 \le \rho^2/16\}\big|.\end{aligned}$$

Recall that we agreed to assume that the pair of indices $(1,2)$ satisfies the conditions in Corollary 4.6. In particular, this implies for $R_i^W := R_i(M+W)$, $i \le n$,

$$|\mathcal{E}_{4.1} \cap \mathcal{E}_0| \le \frac{2}{1-\delta}\,\big|\big\{M \in \mathcal{E}_{4.1} \cap \mathcal{E}_0 : \operatorname{dist}(R_1^W, \operatorname{span}\big\{\{R_k^W\}_{k\ge 2},\, R_1^W + R_2^W\big\}) < \rho/16\big\}\big|.$$

Finally estimates (14)–(16) and Lemma 4.7 imply that

$$\big|\big\{M \in \mathcal{E}_{4.1} \cap \mathcal{E}_0 : \operatorname{dist}(R_1^W, \operatorname{span}\big\{\{R_k^W\}_{k\ge 2},\, R_1^W + R_2^W\big\}) \le \rho/16\big\}\big| \le \frac{C'\,|\mathcal{M}_{n,d}|}{(1-\delta)\sqrt{\varepsilon_1 d}}$$

for a universal constant $C' > 0$. Since $\varepsilon_1 = (1-\delta)\,(C_1 \log(e/(1-\delta))^{-1}$, this implies the desired result. $\square$

4.3 Proof of Claim 4.8

We will use the notations from Lemma 4.7 of the previous subsection. Recall that $\widetilde{M} \in \widetilde{C}_H \cap \mathcal{E}_4 \cap \mathcal{E}_{4.4}^c \subset \mathcal{E}_0 \cap \mathcal{E}_{4.1} \cap \mathcal{E}_{4.4}^c$ and that

$$S_1(H) = \operatorname{supp} R_1(M) \cap \operatorname{supp} R_2(M), \quad S_2(H) = \operatorname{supp} R_1(M) \cup \operatorname{supp} R_2(M)$$

do not depend on the choice of $M \in C_H$. Also, we set

$$S_3 := S_2 \setminus S_1 = \operatorname{supp} R_1(M) \,\triangle\, \operatorname{supp} R_2(M).$$

Take $y := f_W(\widetilde{M})$. Using Remark 4.5 and applying Lemma 2.2 to the vector $\{y_j\}_{j\in S_3}$ we find two disjoint sets $A_1, A_2 \subset S_3$ with cardinalities $|A_1|, |A_2| \ge \ell := \lceil \varepsilon_1 d/2 \rceil$ and such that for all $i \in A_1$ and $j \in A_2$ one has $|y_i - y_j| \ge \rho/(4\sqrt{2})$. For the rest of the proof, we fix ℓ couples of distinct indices $(i_1, j_1), (i_2, j_2), \ldots, (i_\ell, j_\ell) \in A_1 \times A_2$. Next, we define auxiliary subsets of C_H as follows: for any subset $I \subset [\ell]$ and any $S \subset S_3 \setminus \bigcup_{k\in I}\{i_k, j_k\}$ we set

$$\begin{aligned}\operatorname{cpl}(I,S) := \Big\{M \in C_H \,:\, &\{k \,:\, |\operatorname{supp} R_1(M) \cap \{i_k, j_k\}| = 1\} = I \quad \text{and} \\ &\operatorname{supp} R_1(M) \setminus \Big(S_1 \cup \bigcup_{k\in I}\{i_k, j_k\}\Big) = S\Big\}.\end{aligned}$$

Roughly speaking, each subclass $\mathrm{cpl}(I,S)$ is obtained by picking a subset of the couples (i_k, j_k) on which the first row of a matrix is "allowed to vary" while fixing all other coordinates of R_1. Note that $\mathrm{cpl}(I,S)$ can be empty for some I, S. Observe that

$$\Big|\bigcup_{|I|\le \ell/4,\,S}\mathrm{cpl}(I,S)\Big| \le \frac{1}{d^2}\,|C_H|, \tag{17}$$

where the union is taken over all subsets $I \subset [\ell]$ of cardinality at most $\ell/4$ and all admissible sets S, and where d is large enough. Indeed, recall that the class C_H can be identified via a natural bijection with the collection of all m-element subsets of $[2m]$, where $m := |S_3|/2$. With such an identification and by choosing an appropriate permutation of $[2m]$, the set of matrices on the left hand side of (17) corresponds to the collection of m-element subsets B of $[2m]$ such that $|\{k \le \ell \,:\, |B \cap \{k, k+\ell\}| = 1\}| \le \ell/4$, where, in view of the definition of ε_1 and the condition on d, we have $\ell \ge \sqrt{d}$. Then a direct calculation shows that for large d the number of such subsets B is much less than $(\varepsilon_1 d)^{-2}\binom{2m}{m}$.

As the final step in the proof of the claim, we fix a non-empty subclass $\mathrm{cpl}(I,S)$ with $|I| > \ell/4$ and observe that $|\mathrm{cpl}(I,S)| = 2^{|I|}$. In fact, each matrix M in $\mathrm{cpl}(I,S)$ can be uniquely determined by picking either i_k or j_k for every $k \in I$ and then defining the support of the first row of M as the union of the chosen indices, the set S and the intersection part S_1. Moreover, for each $M \in \mathrm{cpl}(I,S)$ the inner product $\langle (R_1(M+W))^\dagger,\, y\rangle$ can be written as

$$\langle (R_1(M+W))^\dagger,\, y\rangle = \langle R_1^\dagger(M),\, y\rangle + \langle R_1^\dagger(W),\, y\rangle = U + \sum_{k\in I}\xi_k(M)(\bar{y}_{i_k} - \bar{y}_{j_k}),$$

where U is a complex number which is *the same* for all $M \in \mathrm{cpl}(I,S)$, and $\xi_k(M)$, $k \in I$, are 0/1-valued functions of M defined as $\xi_k(M) := |\mathrm{supp} R_1(M) \cap \{i_k\}|$. In other words, $\xi_k(M)$ is the indicator of the event that the support of the first row of M contains i_k and not j_k. It is not difficult to see that the functions $\xi_k(M)$, $k \in I$, considered as random variables uniformly distributed on $\mathrm{cpl}(I,S)$, are jointly independent; and that for each $k \in I$ one has

$$|\{M \in \mathrm{cpl}(I,S) \,:\, \xi_k(M) = 1\}| = \tfrac{1}{2}\,|\mathrm{cpl}(I,S)| = 2^{|I|-1}.$$

Further, by our choice of the pairs (i_k, j_k), we have $|\bar{y}_{i_k} - \bar{y}_{j_k}| = |y_{i_k} - y_{j_k}| \ge \rho/(4\sqrt{2})$ for all $k \in I$. Note that $\eta_k = 2\xi_k(M) - 1$, $k \in I$, are independent ± 1 Bernoulli random variables and that for every $v \in \mathbb{C}^I$,

$$\sum_{k\in I}\xi_k(M)v_k = \sum_{k\in I}\eta_k(M)v_k/2 + \sum_{k\in I}v_k/2.$$

Therefore, applying Proposition 2.1, we obtain

$$\big|\big\{M \in \mathrm{cpl}(I,\,S) \,:\, |\langle (R_1(M+W))^\dagger,\, y\rangle| < \rho/4\big\}\big| \le \frac{c\,|\mathrm{cpl}(I,S)|}{\sqrt{|I|}}$$

for some universal constant $c > 0$. Taking the union over all $|I| > \ell/4$, we get

$$\Big|\Big\{M \in \bigcup_{|I|>\ell/4,\,S}\mathrm{cpl}(I,S) \,:\, |\langle (R_1(M+W))^\dagger,\, y\rangle| < \rho/4\Big\}\Big| \le 2c\big|C_H\big|/\sqrt{\ell}.$$

Together with (17), this proves the claim.

4.4 Proof of the main theorem

Here we explain how Theorems 3.1, 3.2, and 4.1 imply our main result, Theorem 1.1. Fix $\rho = 1/(d^{3/2}b_{\mathcal{T}})$, $\kappa = \rho^2/16$, and $\delta = (n - n_3)/n \le 1 - a_3/\log d$. Then the condition on d means $d \le cn/((\log d)(\log\log d))$. Fix $z \in \mathbb{C}$ with $|z| \le d/6$ and $W = -z\mathrm{Id}$. Recall that

$$S(\rho, \delta) = (\mathbb{C}^n \setminus \mathcal{B}(\rho)) \cap \{x \in \mathbb{C}^n : \|x\|_2 = 1\} = S(\rho, \delta).$$

As is mentioned in Remark 3.3, Theorems 3.1 and 3.2 (applied twice for matrices and for their conjugates) imply that $\mathbb{P}(\mathcal{E}_4) \ge 1 - 1/n^2$. Thus, applying Theorem 4.1, we obtain

$$\mathbb{P}(\mathcal{E}_{4.1}) \le \frac{C_1 \log^{3/2} d\, \sqrt{\log\log d}}{\sqrt{d}},$$

which implies the probability bound. Next,

$$\kappa = \rho^2/16 = 1/(16 d^3 {b_{\mathcal{T}}}^2),$$

where $b_{\mathcal{T}} = 4d^{3/2}h_{r+1}^2$ in the case $n_0 > 1$ and $b_{\mathcal{T}} = d\sqrt{n}$ if $n_0 = 1$. This implies

$$s_n \ge \begin{cases} c/(pd^6 n_1^{4+2\alpha_d}) & \text{if} \quad n_0 > p, \\ (c\log d)/(d^9) & \text{if} \quad 1 < n_0 \le p, \\ c/(d^5 n) & \text{if} \quad n_0 = 1, \end{cases}$$

If $1 < n_0 \le p$, then $d^2 \le a_1 n \log d$ and $d^{2.5} \ge a_1 n \log^{1.5} d$, therefore

$$d^9/\log d \le C_1 n^{4.5} \log^{3.5} n.$$

If $n_0 > p$, then, using the definition of α_d, we observe $s_n \ge d^{3/2}\log^{4.5} d/C_2 n^{4+2\alpha_d}$. This implies the estimate in Theorem 1.1. □

Remark 4.9. *In fact we proved that there exists absolute positive constants c, C_1, and C_2 such that*

$$s_n \ge \begin{cases} cd^{3/2}\log^{4.5} d\; n^{-4-2\alpha_d} & \text{if} \quad d < c_1 n^{2/5}\log^{3/5} n, \\ cn^{-4.5}\log^{-3.5} n & \text{if} \quad c_1 n^{2/5}\log^{3/5} n \le d < c_2\sqrt{n\log n}, \\ c/(d^5 n) & \text{if} \quad c_2\sqrt{n\log n} \le d \le \frac{cn}{(\log n)(\log\log n)} \end{cases}$$

with probability at least $(C_1 \log^{3/2} d\, \sqrt{\log\log d})/\sqrt{d}$.

References

[1] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave Ensembles,* J. Amer. Math. Soc. **23** (2010), 535–561.

[2] R. Adamczak, O. Guedon, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Condition number of a square matrix with i.i.d. columns drawn from a convex body,* Proc. Amer. Math. Soc., **140** (2012), 987–998.

[3] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, second edition, Springer Series in Statistics, Springer, New York, 2010. MR2567175

[4] A. Basak, N. Cook, O. Zeitouni, *Circular law for the sum of random permutation matrices*, preprint, https://arxiv.org/abs/1705.09053.

[5] A. Basak, M. Rudelson, *Invertibility of sparse non-hermitian matrices*, Advances in Math., to appear.

[6] D. Bau, L. Trefethen, *Numerical linear algebra.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[7] C. Bordenave and D. Chafaï, *Around the circular law*, Probab. Surv., **9** (2012), 1–89.

[8] S. Brazitikos, A. Giannopoulos, P. Valettas, B. Vritsiou, *Geometry of isotropic convex bodies.* Mathematical Surveys and Monographs, 196. American Mathematical Society, Providence, RI, 2014.

[9] N. A. Cook, Discrepancy properties for random regular digraphs, Random Structures Algorithms **50** (2017), no. 1, 23–58. MR3583025

[10] N. A. Cook, On the singularity of adjacency matrices for random regular digraphs, Probab. Theory Related Fields **167** (2017), no. 1-2, 143–200. MR3602844

[11] N. Cook, *The circular law for random regular digraphs*, ArXiv:1703.05839.

[12] K.R. Davidson, S.J. Szarek, *Local operator theory, random matrices and Banach spaces.* Handbook of the geometry of Banach spaces, Vol. I, 317366, North-Holland, Amsterdam, 2001.

[13] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902.

[14] L. Erdös, H.-T. Yau, A Dynamical Approach to Random Matrix Theory, to appear.

[15] A. Frieze, *Random structures and algorithms*, Proceedings ICM, Vol. 1, 2014, 311–340.

[16] S. van de Geer, A. Muro, *On higher order isotropy conditions and lower bounds for sparse quadratic forms*, Electron. J. Statist., **8** (2014), 3031–3061.

[17] O. Guedon, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann, On the interval of fluctuation of the singular values of random matrices, J. Eur. Math. Soc. (JEMS) **19** (2017), no. 5, 1469–1505. MR3635358

[18] W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables* Journal of the American Statistical Association Vol. 58, No. 301 (1963), 13–30

[19] B. S. Kashin, *Diameters of some finite-dimensional sets and classes of smooth functions*, Izv. Akad. Nauk SSSR, Ser. Mat. **41** (1977), 334–351.

[20] D.J. Kleitman, *On a Lemma of Littlewood and Offord on the Distributions of Linear Combinations of Vectors*, Adv. Math., **5** (1970), 155–157.

[21] V. Koltchinskii, S. Mendelson, *Bounding the smallest singular value of a random matrix without concentration*, Int. Math. Res. Notices. **23** (2015), 12991–13008.

[22] M. Ledoux, M. Talagrand, *Probability in Banach spaces. Isoperimetry and processes*, Springer-Verlag, Berlin, 1991.

[23] A.E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef, *Adjacency matrices of random digraphs: singularity and anti-concentration*, J. Math. Anal. Appl., **445** (2017), 1447–1491.

[24] A.E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef, *Anti-concentration property for random digraphs and invertibility of their adjacency matrices*, C.R. Math. Acad. Sci. Paris, **354** (2016), 121–124.

[25] A.E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes,* Adv. Math., **195** (2005), 491–523.

[26] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R. Vershynin, *Random Euclidean embeddings in spaces of bounded volume ratio,* C.R. Acad. Sci. Paris, Ser 1, Math., **339** (2004), 33–38.

[27] S. Mendelson, G. Paouris, *On singular values of matrices*, Journal of EMS, 16 (2014), 823–834.

[28] J. von Neumann. *Collected works. Vol. V: Design of computers, theory of automata and numerical analysis*, General editor: A. H. Taub. A Pergamon Press Book The Macmillan Co., New York 1963.

[29] J. von Neumann, H.H. Goldstine. *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc. **53** (1947), 10211099.

[30] R.I. Oliveira, *The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties*, arXiv:1312.2903.

[31] E. Rebrova, K. Tikhomirov, Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries, Israel J. Math., to appear. arXiv:1508.06690.

[32] M. Rudelson, R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices,* Adv. Math. **218** (2008), 600–633.

[33] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix, Comm. Pure Appl. Math. **62** (2009), no. 12, 1707–1739. MR2569075

[34] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings ICM, 2010.

[35] A. Sankar, D. A. Spielman and S.-H. Teng, Smoothed analysis of the condition numbers and growth factors of matrices, SIAM J. Matrix Anal. Appl. **28** (2006), no. 2, 446–476 (electronic). MR2255338

[36] G. Schechtman, *Special orthogonal splittings of* L_1^{2k}, Isr. J. Math. **139** (2004), 337–347.

[37] S. Smale, *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc. (N.S.) **13** (1985), 87–121.

[38] D. A. Spielman and S.-H. Teng, Smoothed analysis of algorithms, in *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, 597–606, Higher Ed. Press, Beijing. MR1989210

[39] N. Srivastava and R. Vershynin, Covariance estimation for distributions with $2+\varepsilon$ moments, Ann. Probab. **41** (2013), no. 5, 3081–3111. MR3127875

[40] T. Tao and V. Vu, The condition number of a randomly perturbed matrix, in *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 248–255, ACM, New York. MR2402448

[41] T. Tao, V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Annals of Math. **169** (2009), 595–632.

[42] T. Tao and V. Vu, Smooth analysis of the condition number and the least singular value, Math. Comp. **79** (2010), no. 272, 2333–2352. MR2684367

[43] T. Tao and V. Vu, *Random matrices: universality of ESDs and the circular law,* Ann. Probab. **38(5)** (2010), 2023-2065. With an appendix by Manjunath Krishnapur.

[44] K. Tikhomirov, Sample covariance matrices of heavy-tailed distributions, Int. Math. Res. Notes, to appear. arXiv:1606.03557

[45] N. Tomczak-Jaegermann, *Banach-Mazur distances and finite-dimensional operator ideals.* Pitman Monographs and Surveys in Pure and Applied Mathematics, 38. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1989.

[46] V. Vu, *Random discrete matrices*, Horizons of combinatorics, Bolyai Soc. Math. Stud., **17**, 257–280, Springer, Berlin, 2008.

[47] V.H. Vu, *Combinatorial problems in random matrix theory*, Proceedings ICM, Vol. 4, 2014, 489–508.

[48] P. Yaskov, *Lower bounds on the smallest eigenvalue of a sample covariance matrix*, Electron. Commun. Probab. 19 (2014), 1–10.

[49] P. Yaskov, Sharp lower bounds on the least singular value of a random matrix without the fourth moment condition, Electron. Commun. Probab. **20** (2015), no. 44, 9 pp. MR3358966

Alexander E. Litvak and Nicole Tomczak-Jaegermann,
Dept. of Math. and Stat. Sciences,
University of Alberta,
Edmonton, AB, Canada, T6G 2G1.
e-mails: aelitvak@gmail.com and nicole.tomczak@ualberta.ca

Anna Lytova,
Faculty of Math., Physics, and Comp. Science,
University of Opole,
plac Kopernika 11A, 45-040,
Opole, Poland.
e-mail: alytova@math.uni.opole.pl

Konstantin Tikhomirov,
Dept. of Math., Princeton University,
Fine Hall, Washington road,
Princeton, NJ 08544.
e-mail: ktikhomi@ualberta.ca

Pierre Youssef,
Université Paris Diderot,
Laboratoire de Probabilités et de modèles aléatoires,
75013 Paris, France.
e-mail: youssef@math.univ-paris-diderot.fr