Small ball probability for the condition number of random matrices

Alexander E. Litvak* Konstantin Tikhomirov†
Nicole Tomczak-Jaegermann*

Abstract

Let A be an $n \times n$ random matrix with i.i.d. entries of zero mean, unit variance and a bounded subgaussian moment. We show that the condition number $s_{\max}(A)/s_{\min}(A)$ satisfies the small ball probability estimate

$$\mathbb{P}\big\{s_{\max}(A)/s_{\min}(A) \le n/t\big\} \le 2\exp(-ct^2), \quad t \ge 1,$$

where $c > 0$ may only depend on the subgaussian moment. Although the estimate can be obtained as a combination of known results and techniques, it was not noticed in the literature before. As a key step of the proof, we apply estimates for the singular values of A, $\mathbb{P}\big\{s_{n-k+1}(A) \le ck/\sqrt{n}\big\} \le 2\exp(-ck^2)$, $\quad 1 \le k \le n$, obtained (under some additional assumptions) by Nguyen.

AMS 2010 Classification: 60B20, 15B52, 46B06, 15A18.
Keywords: Random matrices, condition number, small ball probability, invertibility, smallest singular value.

1 Introduction

We say that a random variable ξ has subgaussian moment bounded above by $K > 0$ if

$$\mathbb{P}\{|\xi| \ge t\} \le \exp\big(1 - t^2/(2K^2)\big), \quad t \ge 0.$$

Let A be an $n \times n$ random matrix with i.i.d. entries of zero mean, unit variance and subgaussian moment bounded above by K, and denote by $s_i(A)$, $1 \le i \le n$, its singular values arranged in non-increasing order. We will write $s_{\max}(A)$ and $s_{\min}(A)$ for $s_1(A)$ and $s_n(A)$, respectively. Estimating the magnitude of the condition number,

$$\kappa(A) = s_{\max}(A)/s_{\min}(A),$$

is a well studied problem, with connections to numerical analysis and computation of the limiting distribution of the matrix spectrum; we refer, in particular, to [20] for discussion.

*University of Alberta
†Georgia Institute of Technology

Since the largest singular value $s_{\max}(A)$ is strongly concentrated (see the proof of Corollary 1.2 below), estimating $\kappa(A)$ is essentially reduced to estimating $s_{\min}(A)$ from above and below.

The main result of [12] provides small ball probability estimates for $s_{\min}(A)$ of the form

$$\mathbb{P}\big\{s_{\min}(A) \le t/\sqrt{n}\big\} \le Ct + e^{-cn}, \quad t \le 1,$$

for some $C, c > 0$ depending only on the subgaussian moment. It seems natural to investigate the complementary regime — the large deviation estimates for $s_{\min}(A)$. It was shown in [13] that

$$\mathbb{P}\big\{s_{\min}(A) \ge t/\sqrt{n}\big\} \le \frac{C \ln t}{t} + e^{-cn}, \quad t \ge 2$$

(see also [21] for an extension of this result to distributions with no assumptions on moments higher than 2). The probability estimate was improved in [10] to

$$\mathbb{P}\big\{s_{\min}(A) \ge t/\sqrt{n}\big\} \le e^{-ct}, \quad t \ge 2,$$

for $c > 0$ depending only on the subgaussian moment. The existing results on the distribution of the singular values of random Gaussian matrices [4, 18] suggest that the optimal dependence on t in the exponent on the right hand side is quadratic, i.e. the variable $\sqrt{n}\, s_{\min}(A)$ is subgaussian. Specifically, it is shown in [18] that $s_{\min}(G)$ for the standard $n \times n$ Gaussian matrix G satisfies two-sided estimates

$$\exp(-Ct^2) \le \mathbb{P}\big\{s_{\min}(G) \ge t/\sqrt{n}\big\} \le \exp(-ct^2), \quad t \ge C_1,$$

where $C, C_1, c > 0$ are some universal constants. The main result of our note provides matching upper estimate for matrices with subgaussian entries:

Theorem 1.1. *Let A be an $n \times n$ random matrix with i.i.d. entries of zero mean, unit variance, and subgaussian moment bounded above by $K > 0$. Then the smallest singular value $s_{\min}(A)$ satisfies*

$$\mathbb{P}\big\{s_{\min}(A) \ge t/\sqrt{n}\big\} \le 2\exp(-ct^2), \quad t \ge 1,$$

where $c > 0$ is a constant depending only on K.

As a simple corollary of the theorem, we obtain small ball probability estimates for the condition number:

Corollary 1.2. *Let A be an $n \times n$ random matrix with i.i.d. entries of zero mean, unit variance, and subgaussian moment bounded above by $K > 0$. Then the condition number $\kappa(A)$ satisfies*

$$\mathbb{P}\big\{\kappa(A) \le n/t\big\} \le 2\exp(-ct^2), \quad t \ge 1,$$

where $c > 0$ is a constant depending only on K.

Theorem 1.1 is a consequence of the following theorem, which is of independent interest.

Theorem 1.3. *Under conditions of Theorem 1.1 one has*

$$\mathbb{P}\big\{\|A^{-1}\|_{HS} \le \min(n/t,\ \sqrt{n/t})\big\} \le 2\exp(-ct^2), \quad t \ge 0,$$

where $c > 0$ is a constant depending only on K.

The proof of Theorem 1.3 uses, as a main step, the estimates

$$\mathbb{P}\big\{s_{n-k+1}(A) \le ck/\sqrt{n}\big\} \le 2\exp(-ck^2), \qquad 1 \le k \le n,$$

for the singular values of the matrix A. These estimates, based on the restricted invertibility of matrices and certain averaging arguments, were recently obtained by Nguyen [9] under some additional assumptions (which will be discussed in the next section).

2 Preliminaries

Given a matrix A, it singular values $s_i = s_i(A)$, $i \ge 1$, are square roots of eigenvalues of AA^*. We always assume that $s_1 \ge s_2 \ge \dots$ By $\|A\|$ and $\|A\|_{HS}$ we denote the operator $\ell_2 \to \ell_2$ norm of A (also called the spectral norm) and the Hilbert–Schmidt norm respectively. Note that

$$\|A\| = s_1 \qquad \text{and} \qquad \|A\|_{HS}^2 = \sum_{i\ge 1} s_i^2.$$

The columns and rows of A are denoted by $\mathbf{C}_i(A)$ and $\mathbf{R}_i(A)$, $i \ge 1$, respectively. Given $J \subset [m]$, the coordinate projection in $\mathbb{R}^m$ onto $\mathbb{R}^J$ is denoted by P_J. For convenience, we often write A_J instead of AP_J. Given $m \ge 1$, the identity operator $\mathbb{R}^\ell \to \mathbb{R}^\ell$ we denote by I_m. Given $x, y \in \mathbb{R}^n$ by $\langle x, \cdot\rangle\, y$ we denote the operator $z \mapsto \langle x, z\rangle\, y$ (in the literature it is often denoted by $x \otimes y$ or $yx^\top$). The canonical Euclidean norm in $\mathbb{R}^m$ is denoted by $\|\cdot\|_2$ and the unit Euclidean sphere by S^{m-1}.

As the most important part of our argument, we will use the following result.

Theorem 2.1. *Let A be an $n \times n$ random matrix with i.i.d. entries of zero mean, unit variance, and subgaussian moment bounded above by $K > 0$. Then for any $1 \le k \le n$ one has*

$$\mathbb{P}\big\{s_{n-k+1}(A) \le ck/\sqrt{n}\big\} \le 2\exp(-ck^2),$$

where $c > 0$ is a constant depending only on K.

The above theorem, up to some minor modifications, was proved by Nguyen in [9]. Specifically, in the case $k \ge C\log n$, the theorem follows from [9, Theorem 1.7] (or [9, Corollary 1.8]) if one additionally assumes either that the entries of A are uniformly bounded by a constant, or that the distribution density of the entries is bounded. Removing these conditions requires a minor change of the proof in [9]. Further, in the case $k \le C\log n$, the above result (in fact, in a stronger form) is stated as formula (4) in [9, Theorem 1.4]. However, [9, Theorem 3.6], which is used to derive [9, formula (4)], provides a non-trivial probability estimate only for the event $\{s_{n-k+1}(A) \le c_\gamma k^{1-\gamma}/\sqrt{n}\}$ (for any given $\gamma \in (0,1)$ and c_γ depending on γ), see [9, formula (31)]. Again, a minor update of the argument of [9] provides the result needed for our purposes. In view of the above and for the reader's convenience, we provide a proof of Theorem 2.1 in the last section.

The following result was proved in [17] as an extension of the classical Bourgain–Tzafriri restricted invertibility theorem [2]. With worse dependence on ε, the theorem was earlier proved in [22]. See also a recent paper [8] for further improvements and discussions.

Theorem 2.2 ([17])**.** *Let T be $n\times n$ matrix. Then for any $\varepsilon \in (0,1)$ there is a set $J \subset [n]$ such that*

$$\ell := |J| \geq \left\lfloor \frac{\varepsilon^2 \|T\|_{HS}^2}{\|T\|^2} \right\rfloor \qquad \textit{and} \qquad s_\ell(T_J) \geq \frac{(1-\varepsilon)\|T\|_{HS}}{\sqrt{n}}.$$

We will use two following results by Rudelson–Verhsynin. The first one was one of the key ingredients in estimating the smallest singular value of rectangular matrices. The second one is an immediate consequence of the Hanson–Wright inequality [5, 23] generalized in [15].

Theorem 2.3 ([14], Theorem 4.1)**.** *Let X be a vector in $\mathbb{R}^n$, whose coordinates are i.i.d. mean-zero, subgaussian random variables with unit variance. Let F be a random subspace in $\mathbb{R}^n$ spanned by $n-\ell$ vectors, $1 \leq \ell \leq c'n$, whose coordinates are i.i.d. mean-zero, sub-Gaussian random variables with unit variance, jointly independent with X. Then, for every $\varepsilon > 0$, one has*

$$\mathbb{P}\big\{\mathrm{dist}(X,F) \leq \varepsilon\sqrt{\ell}\big\} \leq (C\varepsilon)^\ell + \exp(-cn).$$

where $C>0$, $c, c' \in (0,1)$ are constants depending only on the subgaussian moments.

Theorem 2.4 ([15, Corollary 3.1])**.** *Let X be a vector in $\mathbb{R}^n$, whose coordinates are i.i.d. mean-zero random variables with unit variance and with subgaussian moment bounded by K. Let F be a fixed subspace in $\mathbb{R}^n$ of dimension $n-\ell$. Then, for every $t>0$, one has*

$$\mathbb{P}\big\{|\mathrm{dist}(X,F) - \sqrt{\ell}| \geq t\big\} \leq 2\exp(-ct^2/K^4).$$

where $c>0$ is an absolute constant.

We will also need the following standard claim, which can be proved by integrating the indicator functions (see e.g., [9, Claim 3.4], cf. [6, Claim 4.9]).

Claim 2.5. *Let $\alpha, p \in (0,1)$. Let $\mathcal{E}$ be an event. Let Z be a finite index set, and $\{\mathcal{E}_z\}_{z\in Z}$ be a collection of $|Z|$ events satisfying $\mathbb{P}(\mathcal{E}_z) \leq p$ for every $z \in Z$. Assume that at least $\alpha|Z|$ of events $\mathcal{E}_z$ hold whenever the event $\mathcal{E}$ occurs. Then $\mathbb{P}(\mathcal{E}) \leq p/\alpha$.*

3 Proofs of main results

Proof of Theorem 1.1. In the case $t>n$ we have

$$\mathbb{P}\big\{s_{\min}(A) \geq t/\sqrt{n}\big\} = \mathbb{P}\Big\{s_1(A^{-1}) \leq \sqrt{n}/t\Big\} \leq \mathbb{P}\Big\{\sum_{i=1}^n s_i(A^{-1})^2 \leq n^2/t^2\Big\}$$

and the result follows from Theorem 1.3.

Now we consider the case $1 \le t \le n$. Let $L \ge 1$ be a parameter which we will choose later. Then

$$\begin{aligned}
\mathbb{P}\big\{s_{\min}(A) \ge t/\sqrt{n}\big\} &= \mathbb{P}\big\{s_1(A^{-1}) \le \sqrt{n}/t\big\} \\
&\le \mathbb{P}\Big\{s_1(A^{-1})^2 \le n/t^2 \quad \text{and} \quad \sum_{i \ge \lceil t \rceil} s_i(A^{-1})^2 \ge Ln/t\Big\} \\
&\qquad + \mathbb{P}\Big\{s_1(A^{-1})^2 \le n/t^2 \quad \text{and} \quad \sum_{i \ge \lceil t \rceil} s_i(A^{-1})^2 < Ln/t\Big\} \\
&\le \mathbb{P}\Big\{\sum_{i \ge \lceil t \rceil} s_i(A^{-1})^2 \ge Ln/t\Big\} + \mathbb{P}\Big\{\sum_{i=1}^{n} s_i(A^{-1})^2 \le n/t + Ln/t\Big\}.
\end{aligned}$$

For the first summand in the last expression, we apply Theorem 2.1. Since $\sum_{i=\lfloor t \rfloor}^{\infty} \frac{1}{i^2} \le \frac{2}{t}$, we obtain

$$\begin{aligned}
\mathbb{P}\Big\{\sum_{i=\lceil t \rceil}^{n} s_i(A^{-1})^2 \ge Ln/t\Big\} &\le \sum_{i=\lceil t \rceil}^{n} \mathbb{P}\big\{s_i(A^{-1})^2 \ge Ln/(2i^2)\big\} \\
&= \sum_{i=\lceil t \rceil}^{n} \mathbb{P}\big\{s_{n-i+1}(A) \le \sqrt{2}i/\sqrt{Ln}\big\}.
\end{aligned}$$

Choosing L so that $\sqrt{2/L}$ is equal to the constant from Theorem 2.1, we get

$$\sum_{i=\lfloor t \rfloor}^{n} \mathbb{P}\big\{s_{n-i+1}(A) \le \sqrt{2}i/\sqrt{Ln}\big\} \le 2 \sum_{i=\lfloor t \rfloor}^{n} \exp(-ci^2) \le 3\exp(-c't^2)$$

for some $c' > 0$ depending only on K. The bound on the second summand follows from Theorem 1.3 applied with $t/(L+1)$ instead of t. This completes the proof. □

Proof of Corollary 1.2. Theorem 2.4 implies that there exists an absolute constant $c_1 > 0$ depending only on K such that for every $i \le n$

$$\mathbb{P}(\|\mathbf{C}_i(A)\|_2 \le \sqrt{n}/2) \le \exp(-c_1 n)$$

(this can be shown by direct calculations as well, see e.g. Fact 2.5 in [7]). Since the entries of A are independent, we obtain

$$\mathbb{P}(\|A\| \le \sqrt{n}/2) \le \prod_{i=1}^{n} \mathbb{P}(\|\mathbf{C}_i(A)\|_2 \le \sqrt{n}/2) \le \exp(-c_1 n^2).$$

Note that if $\|A\| \ge \sqrt{n}/2$ and $\kappa(A) \le n/2t$ then $s_n(A) = \|A\|/\kappa(A) \ge t/\sqrt{n}$. Therefore, by Theorem 1.1,

$$\mathbb{P}\{\kappa(A) \le n/2t\} \le 2\exp(-ct^2) + \exp(-c_1 n^2).$$

By adjusting constants, this implies the conclusion for $t \le n$. Since $\kappa(A) \ge 1$, the case $t > n$ is trivial. □

Proof of Theorem 1.3. Adjusting the constant in the exponent if needed, without loss of generality, we assume that $t \geq C_0$, where $C_0 > 0$ is a large enough constant depending only on K. Denote

$$\mathcal{E}_0 := \Big\{ \sum_{i=1}^n s_i(A^{-1})^2 \leq n/t \Big\}.$$

We first consider the case $t \leq n$. Applying the negative second moment identity (see e.g. Exercise 2.7.3 in [19]),

$$\sum_{i=1}^n s_i(A^{-1})^2 = \sum_{i=1}^n \mathrm{dist}\big(\mathbf{C}_i(A), \mathrm{span}\{\mathbf{C}_j(A),\ j \neq i\}\big)^{-2},$$

we observe that on the event $\mathcal{E}_0$,

$$\big|\big\{i \leq n:\ \mathrm{dist}\big(\mathbf{C}_i(A), \mathrm{span}\{\mathbf{C}_j(A),\ j \neq i\}\big) \geq \sqrt{t/2}\big\}\big| \geq n/2.$$

For each subset $I \subset [n]$ of cardinality $k \leq n/2$ (the actual value of k will be defined later), let $\mathbf{1}_I$ be the indicator of the event

$$\big\{\mathrm{dist}\big(\mathbf{C}_i(A), \mathrm{span}\{\mathbf{C}_j(A),\ j \in [n]\setminus I\}\big) \geq \sqrt{t/2} \text{ for all } i \in I\big\}.$$

Then, in view of the above, everywhere on the event $\mathcal{E}_0$ we have

$$\sum_{I \subset [n],\ |I|=k} \mathbf{1}_I \geq \binom{\lceil n/2 \rceil}{k} \geq \Big(\frac{n}{2k}\Big)^k \geq (2e)^{-k}\binom{n}{k}.$$

Hence, by Markov's inequality and permutation invariance of the matrix distribution,

$$\mathbb{P}(\mathcal{E}_0) \leq (2e)^k\, \mathbb{E}\, \mathbf{1}_{[k]}.$$

As the last step of the proof, we estimate the expectation of $\mathbf{1}_{[k]}$ (with a suitable choice of k). In view of independence and equidistribution of the matrix columns, we have

$$\mathbb{E}\, \mathbf{1}_{[k]} = \Big(\mathbb{P}\big\{\mathrm{dist}\big(\mathbf{C}_1(A), \mathrm{span}\{\mathbf{C}_j(A),\ j \in [n]\setminus[k]\}\big) \geq \sqrt{t/2}\big\}\Big)^k.$$

Choose $k := \lfloor t/4 \rfloor \leq n/2$ and denote

$$D := \mathrm{dist}\big(\mathbf{C}_1(A), \mathrm{span}\{\mathbf{C}_j(A),\ j \in [n]\setminus[k]\}\big).$$

Using independence of columns of the matrix A and applying Theorem 2.4 with $\ell = k$ and $F = \mathrm{span}\{\mathbf{C}_j(A),\ j \in [n]\setminus[k]\}$, we obtain

$$\mathbb{P}\Big\{D \geq \sqrt{t/2}\Big\} \leq \mathbb{P}\Big\{D - \sqrt{k} \geq (\sqrt{2}-1)\,\sqrt{t/4}\Big\} \leq 2\exp(-\bar{c}\,t)$$

for some $\bar{c} > 0$ depending only on K. Hence,

$$\mathbb{P}(\mathcal{E}_0) \leq (2e)^k\, 2^k\, \exp(-\bar{c}\,t\,k) \leq \exp(-\bar{c}t^2/16),$$

provided that t is larger than a certain constant depending only on K. This implies the desired result for $t \le n$.

In the case $t > n$ we essentially repeat the argument along the same lines. Define

$$\mathcal{E}_0' := \Big\{ \sum_{i=1}^n s_i(A^{-1})^2 \le n^2/t^2 \Big\}.$$

Observe that on the event $\mathcal{E}_0'$,

$$\big|\big\{ i \le n : \ \mathrm{dist}\big(\mathbf{C}_i(A), \mathrm{span}\{\mathbf{C}_j(A), \ j \ne i\}\big) \ge t/\sqrt{2n} \big\}\big| \ge n/2.$$

Repeating the above computations with the same notation and with $k = \lfloor n/4 \rfloor$ we obtain

$$\mathbb{P}\Big\{ D \ge t/\sqrt{2n} \Big\} \le \mathbb{P}\Big\{ D - \sqrt{k} \ge t/(5\sqrt{n}) \Big\} \le 2\exp(-\bar{c}\, t^2/n),$$

which leads to

$$\mathbb{P}(\mathcal{E}_0') \le (2e)^k \, 2^k \, \exp(-\bar{c}\, kt^2/n) \le \exp(-\bar{c}t^2/16),$$

provided that $t > Cn$ for large enough C depending only on K. For $n < t \le Cn$ the result follows by adjusting the absolute constants. □

4 Small ball estimates for singular values

The goal of this section is to prove Theorem 2.1. As we have noted, the argument essentially reproduces that of [9]. An important part of the proof is the use of restricted invertibility (see also [3] and [11] for some recent applications of restricted invertibility in the context of random matrices).

We will use a construction from [9]. Given an integer k and an $n \times n$ matrix A define a $k \times n$ matrix $Z = Z(A, k)$ in the following way. Consider singular value decomposition $A = \sum_{i=1}^n s_i \langle v_i, \cdot \rangle \, w_i$, where $s_i = s_i(A)$ are singular values of A (arranged in non-increasing order) and $\{v_i\}_i$, $\{w_i\}_i$ are two orthonormal systems in $\mathbb{R}^n$. For $i \le k$ denote $z_i = v_{n-i+1}$. Let Z be the matrix whose rows are $\mathbf{R}_i(Z) = z_i$. Clearly, the rows of Z are orthonormal and for every $i \le k$,

$$\|Az_i\|_2 = s_{n-i+1} \le s_{n-k+1}. \tag{1}$$

Moreover,

$$\|Z\| = 1 \qquad \text{and} \qquad \|Z\|_{HS} = \sqrt{k}.$$

The matrix Z is not uniquely defined when some of the k smallest singular values of A have non-trivial multiplicity; we will however assume that for each realization of A, a single admissible Z is chosen in such a way that Z is a (measurable) random matrix.

4.1 Proof of Theorem 2.1, the case $k \geq \ln n$

Let C, c, c' be constants from Theorem 2.3. Let $\gamma = \sqrt{c'}$. Note that C, c, c', γ depend only on K. Let $Z = Z(A, k)$ be the $k \times n$ matrix constructed above. Applying Theorem 2.2 to Z (one can add zero rows to make it an $n \times n$ matrix), there exists $J \subset [n]$ such that

$$|J| = \ell := \lfloor \gamma^2 k \rfloor \leq c'k \qquad \text{and} \qquad s_\ell(Z_J) \geq (1-\gamma)\sqrt{k/n}.$$

Fix a (small enough, depending on K) constant $c_0 > 0$. Define the event

$$\mathcal{E}_k := \left\{ s_{n-k+1}(A) \leq c_0 k/\sqrt{n} \right\}.$$

Consider the $n \times k$ matrix $B = AZ^\top$. Using property (1), on the event $\mathcal{E}_k$, we have for every $i \leq k$,

$$\|\mathbf{C}_i(B)\|_2 = \|Az_i\|_2 \leq c_0 k/\sqrt{n},$$

hence $\|B\|_{HS} \leq c_0 k^{3/2}/\sqrt{n}$. Now, since $s_\ell(Z_J) > 0$, there exists a $k \times \ell$ matrix M such that $Z_J^\top M = I_\ell$. Then

$$\|M\| = 1/s_\ell(Z) \leq (1-\gamma)^{-1}\sqrt{n/k}.$$

Therefore,

$$\|BM\|_{HS} \leq \|B\|_{HS}\,\|M\| \leq c_0(1-\gamma)^{-1}k.$$

Writing $B = A_J(Z_J)^\top + A_{J^c}(Z_{J^c})^\top$, we also have $BM = A_J + A_{J^c}(Z_{J^c})^\top M$. Next denote

$$F = F(A, J) := \operatorname{span}\{\mathbf{C}_i(A_{J^c})\}_{i \in J^c},$$

and let P be the orthogonal projection on $F^\perp$. Then, on the event $\mathcal{E}_k$,

$$c_0^2(1-\gamma)^{-2}k^2 \geq \|PBM\|_{HS}^2 \geq \|PA_J\|_{HS}^2 = \sum_{i \in J} \|P\,\mathbf{C}_i(A_J)\|_2^2 = \sum_{i \in J} \operatorname{dist}^2(\mathbf{C}_i(A), F).$$

Therefore, for at least $\ell/2$ indices $i \in J$, one has

$$\operatorname{dist}(\mathbf{C}_i(A), F) \leq \sqrt{2}c_0(1-\gamma)^{-1}k/\sqrt{\ell} \leq 2c_0\sqrt{\ell}/((1-\gamma)\gamma^2).$$

Note that the subspace F is spanned by $n-\ell$ random vectors, it is independent of columns $\mathbf{C}_i(A)$, $i \in J$, and that columns of A are independent. Therefore, by Theorem 2.3 and the union bound we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_k) &\leq \sum_{\substack{J \subset [n] \\ |J| = \ell}} \sum_{\substack{J_1 \subset J \\ |J| = \lceil \ell/2 \rceil}} \mathbb{P}\Big\{ \forall i \in J_1 \;\; \operatorname{dist}(\mathbf{C}_i(A), F) \leq 2c_0\sqrt{\ell}/((1-\gamma)\gamma^2) \Big\} \\
&\leq \binom{n}{\ell} 2^\ell \left((2Cc_0/((1-\gamma)\gamma^2))^\ell + \exp(-cn) \right)^{\ell/2} \\
&\leq \left(\frac{4en}{\ell} \max\left\{ \left(\frac{\sqrt{2Cc_0}}{\gamma\sqrt{1-\gamma}} \right)^\ell, \exp(-cn/2) \right\} \right)^\ell
\end{aligned}$$

Choosing small enough c_0 and using $k \geq \ln n$, we obtain $\mathbb{P}(\mathcal{E}_k) \leq \exp(-c_3\ell^2)$, where $c_3 > 0$ depends only on K. By adjusting constants this proves the desired result for $k \geq \ln n$. $\square$

4.2 Proof of Theorem 2.1, the case $k \le \ln n$

Let A be as in Theorem 2.1. It is well known (see e.g. Fact 2.4 in [7]) that there is an absolute constant $C_1 > 0$ such that

$$\mathbb{P}\big\{\|A\| \le C_1 K\sqrt{n}\big\} \ge 1 - e^{-n}. \tag{2}$$

Let $\mathcal{E}_{bd}$ denote the event from this equation. Further, from [16, Theorem 1.5] one infers that for any $\gamma > 0$ there are $\gamma_1, \gamma_2, \gamma_3 > 0$ depending only on γ and K such that, denoting

$$\mathcal{E}_{inc}(\gamma) := \big\{\forall x \in S^{n-1} \text{ with } \|Ax\|_2 \le \gamma_1\sqrt{n},\ \forall I \subset [n] \\ \text{with } |I| \ge \gamma n \text{ one has } \|P_I x\|_2 \ge \gamma_2\big\},$$

the event satisfies

$$\mathbb{P}(\mathcal{E}_{inc}(\gamma)) \ge 1 - 2e^{-\gamma_3 n}. \tag{3}$$

The following statement was proved by Nguyen ([9, Corollary 3.8]).

Proposition 4.1. *For any $K > 0$ there are $C, c_1, c_2, \gamma > 0$ depending only on K with the following property. Let A be an $n \times n$ random matrix with i.i.d. entries of zero mean, unit variance, and subgaussian moment bounded above by K. Let $2 \le k \le n/(C\ln n)$, and let the random $k \times n$ matrix $Z = Z(A,k)$ be defined as above. Then everywhere on the event $\big\{s_{n-k+1}(A) \le c_1 k/\sqrt{n}\big\} \cap \mathcal{E}_{inc}(\gamma) \cap \mathcal{E}_{bd}$ one has*

$$\Big|\big\{J \subset [n] :\ |J| = \lfloor k/2\rfloor,\ s_{\lfloor k/2\rfloor}(Z_J) \ge c_1\sqrt{k/n}\big\}\Big| \ge c_2^{k\ln k}\, n^{\lfloor k/2\rfloor}.$$

Now assume that $k \le \ln n$. Without loss of generality we may also assume that k is bounded below by a large constant. Let C, c, c' be constants from Theorem 2.3 and c_1, c_2, γ from Proposition 4.1. Fix for a moment any realization of A from the event $\big\{s_{n-k+1}(A) \le c_0 k/\sqrt{n}\big\} \cap \mathcal{E}_{inc}(\gamma) \cap \mathcal{E}_{bd}$, where $c_0 \in (0, c_1]$ will be chosen later. Let $\ell := \lfloor k/2\rfloor$ and

$$\mathcal{J} := \big\{J \subset [n] :\ |J| = \lfloor k/2\rfloor,\ s_{\lfloor k/2\rfloor}(Z_J) \ge c_1\sqrt{k/n}\big\}.$$

Fix $J \in \mathcal{J}$ and repeat the procedure used in Subsection 4.1 with J and ℓ. We obtain that for at least $\ell/2$ indices $i \in J$, one has

$$\operatorname{dist}(\mathbf{C}_i(A), F) \le \sqrt{2}c_0 k/(c_1\sqrt{\ell}) \le 4c_0\sqrt{\ell}/c_1, \tag{4}$$

where $F = \operatorname{span}\{\mathbf{C}_i(A_{J^c})\}_{i\in J^c}$. For any fixed subset $J \subset [n]$ of cardinality ℓ consider the event

$$\mathcal{E}_J := \big\{\text{for at least } \ell/2 \text{ indices } i \in J \text{ inequality (4) holds}\big\}.$$

Applying Theorem 2.3 and the union bound we observe

$$\mathbb{P}(\mathcal{E}_J) \le 2^{\ell}\,\big((4c_0C/c_1)^{\ell} + \exp(-cn)\big)^{\ell/2} \le \Big(4\,\max\Big\{(4c_0C/c_1)^{\ell},\, \exp(-cn)\Big\}\Big)^{\ell/2}.$$

Choosing c_0 to be small enough we obtain that $\mathbb{P}(\mathcal{E}_J) \le \exp(-c_4k^2)$, where $c_4 > 0$ depends only on K. Combining this with Claim 2.5 and Proposition 4.1 we obtain

$$\mathbb{P}\big(\big\{s_{n-k+1}(A) \le c_0k/\sqrt{n}\big\} \cap \mathcal{E}_{inc}(\gamma) \cap \mathcal{E}_{bd}\big) \le c_2^{-k\ln k}\exp(-c_4k^2) \le \exp(-c_5k^2)$$

provided that $k \geq C_2$, where $C_2 \geq 1 \geq c_5 > 0$ are constants depending on on K only. By (2) and (3) this completes the proof in the case $k \leq \ln n$. $\square$

Acknowledgment. The authors are grateful to the anonymous referee for careful reading and valuable suggestions that have helped to improve the presentation. The second named author would like to thank the Department of Mathematical and Statistical Sciences, University of Alberta for ideal working conditions.

References

[1] J. D. Batson, D. A. Spielman and N. Srivastava, *Twice-Ramanujan sparsifiers*, in *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, 255–262, ACM, New York. MR2780071

[2] J. Bourgain and L. Tzafriri, *Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis*, Israel J. Math. **57** (1987), no. 2, 137–224. MR0890420

[3] N. Cook, *Lower bounds for the smallest singular value of structured random matrices*, Ann. of Prob., to appear, arXiv:1608.07347

[4] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. **9** (1988), no. 4, 543–560. MR0964668

[5] D. L. Hanson and F. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, Ann. Math. Statist. **42** (1971), 1079–1083. MR0279864

[6] A.E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, P. Youssef, *Adjacency matrices of random digraphs: singularity and anti-concentration*, J. Math. Anal. Appl. **445** (2017), no. 2, 1447–1491. MR3545253

[7] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), no. 2, 491–523. MR2146352

[8] A. Naor, P. Youssef, *Restricted invertibility revisited*, A journey through discrete mathematics, 657–691, Springer, Cham, 2017. MR3726618

[9] H. H. Nguyen, *Random matrices: overcrowding estimates for the spectrum*, J. Funct. Anal. **275** (2018), no. 8, 2197–2224. MR3841540

[10] H. H. Nguyen and V. H. Vu, *Normal vector of a random hyperplane*, Int. Math. Res. Not. **2018**, no. 6, 1754–1778. MR3800634

[11] M. Rudelson, K. Tikhomirov, *The sparse circular law under minimal assumptions*, submitted.

[12] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices,* Adv. Math. **218** (2008), no. 2, 600–633. MR2407948

[13] M. Rudelson and R. Vershynin, *The least singular value of a random square matrix is $O(n^{-1/2})$,* C. R. Math. Acad. Sci. Paris **346** (2008), no. 15-16, 893–896. MR2441928

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

[15] M. Rudelson and R. Vershynin, *Hanson-Wright inequality and sub-Gaussian concentration,* Electron. Commun. Probab. **18** (2013), no. 82, 9 pp. MR3125258

[16] M. Rudelson and R. Vershynin, No-gaps delocalization for general random matrices, Geom. Funct. Anal. **26** (2016), no. 6, 1716–1776. MR3579707

[17] D. A. Spielman and N. Srivastava, *An elementary proof of the restricted invertibility theorem*, Israel J. Math. **190** (2012), 83–91. MR2956233

[18] S. J. Szarek, *Condition numbers of random matrices,* J. Complexity **7** (1991), no. 2, 131–149. MR1108773

[19] T. Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics **132**, Providence, RI: American Mathematical Society, 2012.

[20] T. Tao and V. H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices,* Ann. of Math. (2) **169** (2009), no. 2, 595–632. MR2480613

[21] K. Tatarko, *An upper bound on the smallest singular value of a square random matrix,* J. Complexity **48** (2018), 119–128. MR3828841

[22] R. Vershynin, *John's decompositions: selecting a large part*, Israel J. Math. **122** (2001), 253–277. MR1826503

[23] F. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric*, Ann. Probability **1** (1973), no. 6, 1068–1070. MR0353419

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`

Konstantin Tikhomirov,
School of Mathematics,
Georgia Institute of Technology,
686 Cherry street,
Atlanta, GA 30332, USA.
`e-mail: ktikhomirov6@gatech.edu`