ISRAEL JOURNAL OF MATHEMATICS **227** (2018), 507–544
DOI: 10.1007/s11856-018-1732-y

COVERINGS OF RANDOM ELLIPSOIDS, AND INVERTIBILITY OF MATRICES WITH I.I.D. HEAVY-TAILED ENTRIES

BY

Elizaveta Rebrova*

Department of Mathematics, University of Michigan
Ann Arbor, MI 48109-1043, USA
e-mail: erebrova@umich.edu

AND

Konstantin Tikhomirov**

Department of Mathematical and Statistical Sciences, University of Alberta
Edmonton, Alberta, Canada, T6G 2G1
Current address: Department of Mathematics, Princeton University
Princeton, NJ 08544, USA
e-mail: kt12@math.princeton.edu

* E. R. was partially supported by U.S. Air Force grant F035062.
** K. T. was partially supported by PIMS Graduate Scholarship and by Dean's Excellence Award, Faculty of Science, UofA.
Received January 25 ,2016 and in revised form February 13, 2017

ABSTRACT

Let $A = (a_{ij})$ be an $n \times n$ random matrix with i.i.d. entries such that $\mathbb{E}a_{11} = 0$ and $\mathbb{E}{a_{11}}^2 = 1$. We prove that for any $\delta > 0$ there is $L > 0$ depending only on δ, and a subset $\mathcal{N}$ of B_2^n of cardinality at most $\exp(\delta n)$ such that with probability very close to one we have

$$A(B_2^n) \subset \bigcup_{y \in A(\mathcal{N})} (y + L\sqrt{n}B_2^n).$$

In fact, a stronger statement holds true. As an application, we show that for some $L' > 0$ and $u \in [0, 1)$ depending only on the distribution law of a_{11}, the smallest singular value s_n of the matrix A satisfies

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

for all $\varepsilon > 0$. The latter result generalizes a theorem of Rudelson and Vershynin which was proved for random matrices with subgaussian entries.

1. Introduction

In this paper, we consider random matrices A satisfying

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

We are concerned with the following question: how many translates of a Euclidean ball $\sqrt{n}B_2^n$ (or its constant multiple) are needed to cover the random ellipsoid $A(B_2^n)$? Being geometrically natural, this problem, as we will see later, has an application to studying invertibility properties of the matrix A.

If the entries of A have a bounded fourth moment, then the operator norm $\|A\|_{2\to 2}$ satisfies $\|A\|_{2\to 2} \le L\sqrt{n}$ with probability close to one (see [31] and [9] for precise statements), whence

$$\mathbb{P}\{A(B_2^n) \subset L\sqrt{n}B_2^n\} \approx 1.$$

If, moreover, the entries of A are subgaussian, then for some $L > 0$ depending only on the subgaussian moment we have

$$\mathbb{P}\{A(B_2^n) \subset L\sqrt{n}B_2^n\} \ge 1 - \exp(-n).$$

On the other hand, for heavy-tailed entries the operator norm of A may have a higher order of magnitude compared to $\sqrt{n}$ with probability close to one, so the trivial argument given above is not applicable. The first main result of the paper is the following theorem:

THEOREM A: *Let $\delta \in (0, 1/4]$ and $n \geq \frac{1}{4\delta}$. Then there is a (non-random) collection $\mathcal{C}$ of parallelepipeds in $\mathbb{R}^n$ with $|\mathcal{C}| \leq \exp(13n\delta \ln \frac{2e}{\delta})$ having the following property: For any random matrix A satisfying $(*)$, with probability at least $1 - 4\exp(-\delta n/8)$ we have*

$$\forall\, x \in B_2^n\ \exists P \in \mathcal{C} \text{ such that } x \in P \text{ and } A(P) \subset Ax + \frac{C\sqrt{n}}{\delta} B_2^n.$$

Here, $C > 0$ is a universal constant.

In particular, the above theorem implies the following more elegant

COROLLARY A: *For any $\delta \in (0, 1/4]$ and $n \geq \frac{1}{4\delta}$ there exists a non-random subset $\mathcal{N} \subset B_2^n$ of cardinality at most $\exp(13n\delta \ln \frac{2e}{\delta})$ such that for any $n \times n$ matrix A satisfying $(*)$, we have*

$$\mathbb{P}\Big\{ A(B_2^n) \subset \bigcup_{y \in A(\mathcal{N})} \Big(y + \frac{C'\sqrt{n}}{\delta} B_2^n \Big) \Big\} \geq 1 - 4\exp(-\delta n/8)$$

for some universal constant $C' > 0$.

Both results have geometric interpretation in terms of covering numbers. Recall that for two subsets S and K of a vector space the **covering number** $\mathbf{N}(S, K)$ is defined as the smallest number of parallel translates of K sufficient to cover S. By Theorem A, $\mathbf{N}(A(B_2^n), \frac{C\sqrt{n}}{\delta} B_2^n) \leq \exp(13\delta n \ln \frac{2e}{\delta})$ with probability at least $1 - 4\exp(-\delta n/8)$.

Another interpretation of these results, that will be of use for us, is related to the net refinement (see Theorem A* in Section 5). Given a metric space X, an ε-**net** $\mathcal{N}$ on X is a subset of X such that any point of X is within a distance at most ε from a point of $\mathcal{N}$. It is easy to see that with probability at least $1 - 4\exp(-\delta n/8)$ the set $\mathcal{N}$ from Corollary A is a $\frac{C\sqrt{n}}{\delta}$-net on B_2^n with respect to the pseudometric $d(x, y) := \|A(x - y)\|$ $(x, y \in B_2^n)$. Here and further, $\|\cdot\|$ denotes the standard Euclidean norm in $\mathbb{R}^n$.

A crucial feature of these results is that the set $\mathcal{C}$ in the theorem is non-random. Moreover, $\mathcal{C}$ (as well as the set $\mathcal{N}$ from Corollary A) provides a "universal" covering which is independent of the distribution of the entries of A.

Finally, compared to Corollary A, the statement of Theorem A is more flexible as it enables us to choose the "anchor" points within the parallelepipeds when constructing the corresponding ε-net (this matter is covered in detail at the beginning of Section 5).

Let us briefly describe the main idea of the proof. The collection $\mathcal{C}$ of parallelepipeds is constructed using a special subset $\mathcal{D}$ of diagonal operators with diagonal elements in the interval $(0,1]$. Namely, we define $\mathcal{D}$ as the set of all diagonal operators with diagonal entries in $\{1\}\cup\{2^{-2^k}\}_{k=0}^{\infty}$ and with determinants bounded from below by $\exp(-\delta n)$. Then, for every operator D from $\mathcal{D}$, we take a covering of the ball B_2^n by appropriate translates of parallelepiped $D(L''n^{-1/2}B_\infty^n)$ (for some $L''=L''(\delta)$), and let $\mathcal{C}$ be the union of such coverings over $\mathcal{D}$. It turns out that Theorem A follows almost immediately from the following relation:

$$\begin{aligned}&\mathbb{P}\{\exists \text{ diagonal matrix } D \text{ with diagonal entries in } \{1\}\cup\{2^{-2^k}\}_{k=0}^{\infty}\\ &\quad\text{such that } \det D\geq \exp(-\delta n) \text{ and } \|AD\|_{\infty\to 2}\leq \frac{Cn}{\sqrt{\delta}}\}\\ &\qquad\geq 1-4\exp(-\delta n/8).\end{aligned} \tag{1}$$

In Section 3, we show that (1) holds true under condition $(*)$; see Theorem 3.1. Geometrically, this property means that it is possible to construct a random parallelepiped $P\subset[-1,1]^n$ with sides parallel to the standard coordinate axes, such that $\mathrm{Vol}(P)\geq\exp(-\delta n)$ and A maps P inside the Euclidean ball $\frac{Cn}{\sqrt{\delta}}B_2^n$ with probability at least $1-4\exp(-\delta n/8)$. Note that parallelepiped P will be "narrow" along directions $w\in S^{n-1}$ for which $\|Aw\|$ is large.

As we already mentioned above, Theorem A has a direct application to the problem of obtaining quantitative (non-asymptotic) estimates for the smallest singular value of A. Recall that, given an $m\times n$ ($m\geq n$) matrix M, its smallest singular value can be defined as $s_n(M)=\inf_{y\in S^{n-1}}\|My\|$. An argument based on Theorem A and results of Rudelson and Vershynin from [20, 19] yields

Theorem B: *For any $\widetilde{v}\in(0,1]$ and $\widetilde{u}\in(0,1)$ there are numbers $L>0$, $u\in(0,1)$ and $n_0\in\mathbb{N}$ depending only on $\widetilde{v}$ and $\widetilde{u}$ with the following property: Let $n\geq n_0$ and let $A=(a_{ij})$ be an $n\times n$ random matrix satisfying $(*)$, such that $\sup_{\lambda\in\mathbb{R}}\mathbb{P}\{|a_{11}-\lambda|\leq\widetilde{v}\}\leq\widetilde{u}$. Then for any $\varepsilon>0$ we have*

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

Note that any random variable α with $\mathbb{E}\alpha=0$ and $\mathbb{E}\alpha^2=1$ obviously satisfies $\sup_{\lambda\in\mathbb{R}}\mathbb{P}\{|\alpha-\lambda|\leq\widetilde{v}\}\leq\widetilde{u}$ for some $\widetilde{v}>0$ and $\widetilde{u}\in(0,1)$ determined by the law of α. Thus, the above statement does not require any additional assumptions on the matrix apart from $(*)$; by introducing the quantities $\widetilde{v}$ and $\widetilde{u}$ we make the dependence of L and u on the law of a_{11} more explicit.

Let us put Theorem B in the context of known results.

Convergence of (appropriately normalized) smallest singular values for a sequence of random rectangular matrices with i.i.d. entries and growing dimensions was established by Bai and Yin [3] (see also [27], where the result is proved under optimal moment assumptions). For non-asymptotic results in this direction, we refer the reader to papers [11, 19] for the case of i.i.d. entries (see also [28] where no moment conditions are assumed), [1, 2] for log-concave distributions of rows and [21, 12, 8, 30, 5] for more general isotropic distributions. We refer to surveys [18, 29] (see also [17]) for more information.

For random square matrices with independent standard Gaussian entries, the limiting distribution of the smallest singular value was computed by Edelman [4]; universality of this result was established in [24]. Further, for matrices with i.i.d. entries it was shown in [23] and [25] that, given any $K>0$, there are $R, L>0$ depending only on K and the law of a_{11} such that $\mathbb{P}\{s_n(A+B) \le n^{-L}\} \le Rn^{-K}$ for any non-random matrix B satisfying $\|B\|_{2\to 2} \le n^K$ (we note that analogous results were recently obtained for more general models of randomness allowing some dependence between the entries of A; see, in particular, [14] and [6]). In the case $B = \mathbf{0}$ which we study in this paper, those papers do not provide optimal estimates for $s_n(A)$. A much more precise statement was proved in [20] under the additional assumption that the entries of A are subgaussian; namely, Rudelson and Vershynin showed that $s_n(A)$ satisfies a small ball probability estimate

$$\mathbb{P}\{s_n(A) \le \varepsilon n^{-1/2}\} \le L\varepsilon + u^n, \quad \varepsilon > 0,$$

where $L > 0$ and $u \in (0,1)$ depend only on the subgaussian moment of a_{ij}'s. Note that Theorem B gives an estimate of exactly the same form, but for the matrices with heavy-tailed entries.

The idea of the proof of Theorem B can be described as follows. Denote by A' the transpose of the first $n-1$ columns of A. A principal component of the proof of [20] is an analysis of the arithmetic structure of null vectors of A', which is described with the help of the notion of the least common denominator (LCD). To show that null vectors of A' typically have an exponentially large LCD, the authors of [20] consider subsets S of the unit sphere corresponding to vectors with small LCD, and show that $\inf_{x\in S}\|A'x\| > 0$

with a large probability. For this, they use the standard ε-net argument, when the infimum is estimated by taking a Euclidean ε-net $\mathcal{N}$ on S and applying relation $\inf_{x\in S}\|A'x\| \geq \inf_{y\in\mathcal{N}}\|A'y\| - \varepsilon\|A'\|_{2\to 2}$ together with the estimate $\|A'\|_{2\to 2} \leq C\sqrt{n}$ which holds with probability very close to one under the subgaussian moment assumptions on the entries. In our setting, the principal difficulty consists in the fact that the condition $(*)$ does not guarantee a good upper bound for the operator norm $\|A'\|_{2\to 2}$. To deal with this fundamental issue, we "refine" the nets constructed in [20] by applying Theorem A. Indeed, it can be shown that Theorem A implies that, given an ε-net $\mathcal{N}$ on S, it is possible to construct a subset $\widetilde{\mathcal{N}} \subset S$ of cardinality at most $\exp(13\delta n \ln\frac{2e}{\delta})|\mathcal{N}|$ which is an $L'\varepsilon\sqrt{n}$-net on S (for some $L' = L'(\delta)$) with respect to the pseudometric $d(x,y) = \|A'(x-y)\|$ with probability at least $1 - 4\exp(-\delta n/8)$. Then, $\inf_{x\in S}\|A'x\| \geq \inf_{y\in\widetilde{\mathcal{N}}}\|A'y\| - L'\varepsilon\sqrt{n}$, so the argument does not depend any more on the value of $\|A'\|_{2\to 2}$.

The paper is organized as follows: Sections 2 and 3 are devoted to proving the main novel element of the paper—Theorem A. Then, in Section 4, we collect some results from [20], and, in Section 5, prove Theorem B.

Finally, let us discuss notation. Given a finite set S, by $|S|$ we denote its cardinality. By $e_1, e_2, \ldots, e_n$ we denote the canonical basis in $\mathbb{R}^n$. The standard inner product in $\mathbb{R}^n$ shall be denoted by $\langle\cdot,\cdot\rangle$. Given $p \in [1,\infty]$, $\|\cdot\|_p$ is the standard ℓ_p-norm. For ℓ_2, we will simply write $\|\cdot\|$. Given an $m\times n$ matrix M and $p, q \in [1,\infty]$, by $\|M\|_{p\to q}$ we shall denote the operator norm of M considered as the mapping from $(\mathbb{R}^n, \|\cdot\|_p)$ to $(\mathbb{R}^m, \|\cdot\|_q)$. Universal positive constants shall be denoted by C, c. Sometimes, to avoid confusion, we shall add a numerical subscript to the name of a constant or function defined within a statement.

2. Fitting a random vector into an ℓ_p^n-ball

Throughout the paper, by $\mathcal{D}_n$ we denote the set of all $n\times n$ diagonal matrices with diagonal elements belonging to the interval $(0,1]$ (we will sometimes refer to such matrices as positive diagonal contractions). Further, denote by $\mathcal{D}_n^2$ the set of all $n\times n$ positive diagonal contractions whose diagonal entries belong to the set $\{1\}\cup\{2^{-2^k}\}_{k=0}^{\infty}$. The set $\mathcal{D}_n^2$ can be regarded as a discretization of $\mathcal{D}_n$.

In this section, we consider the following problem: Let X be a random vector in $\mathbb{R}^n$ with i.i.d. coordinates. We want to find a random diagonal operator D

taking values in $\mathcal{D}_n$ such that $D(X)$ is contained in an appropriate (fixed) multiple of the ℓ_p^n-ball everywhere on the probability space and at the same time the determinant of D is typically "not too small". The statement to be proved is

Proposition 2.1: *For any $\alpha \in (0,1)$ there is a number $L = L(\alpha) > 0$ with the following property: Let $\delta \in (0,1]$, $p \in [1,\infty)$ and let $X = (x_1, x_2, \dots, x_n)$ be a random vector on $(\Omega, \Sigma, \mathbb{P})$ with i.i.d. coordinates such that $\mathbb{E}|x_i|^p < \infty$. Then there is a random positive diagonal contraction D taking values in $\mathcal{D}_n$ such that*

$$\|DX\|_p^p \le \frac{L}{\delta}\mathbb{E}\|X\|_p^p \text{ everywhere on the probability space,}$$
$$\text{and } \mathbb{E}(\det D)^{p\alpha - p} \le \exp(\delta).$$

Remark 2.2: Proposition 2.1 is a foundation block of our paper. In Section 3, we will amplify this result (the case $p = 2$) by proving its "matrix version" (Theorem 3.1). The case $p \neq 2$ in this section is considered just for completeness.

Remark 2.3: Note that a trivial definition of the diagonal operator $D = (d_{ij})$ by setting

$$d_{jj}{}^p := \min\Big(1, \frac{L}{\delta}\frac{\mathbb{E}\|X\|_p^p}{\|X\|_p^p}\Big), \quad j = 1, 2, \dots, n,$$

gives an unsatisfactory distribution of the determinant. For example, if the entries of X are $\{0,1\}$-valued with probability of taking value 1 equal to $1/n$, then $\mathbb{E}\|X\|_p^p = 1$, and for any $m \le n$ we have

$$\mathbb{P}\{\|X\|_p^p = m\} = \binom{n}{m} n^{-m}\Big(1 - \frac{1}{n}\Big)^{n-m} \ge \frac{1}{4m^m}.$$

Thus, the above definition of D would give

$$\mathbb{P}\{\det D \le 2^{-n}\} \ge \frac{1}{4}\lceil 2^p L/\delta\rceil^{-\lceil 2^p L/\delta\rceil}.$$

Our construction of the required operator is more elaborate. Let us first describe the idea informally. Assume that $p = 1$ and that X is our random vector with non-negative i.i.d. coordinates with unit expectations. We consider a sequence of non-negative numbers (levels) such that each coordinate exceeds the k-th level with probability 2^{-k}. The main observation is that X "does not fit" into the ℓ_1^n-ball $\frac{Ln}{\delta}B_1^n$ only if for some k there are much more than $2^{-k}n$ coordinates of X exceeding the level. We define the required operator D so that its restriction to the "bad" coordinates is an appropriate dilation, while on all other coordinates it acts isometrically. If there exist several "bad" levels the operator D will be defined as a product of several diagonal operators. Moreover, it will be more convenient to "replace" the vector X by a sum of independent vectors of two-valued variables, such that the sum is a majorant for X on the entire probability space. We construct the majorant in the coupling Lemma 2.5 stated below.

Given a non-negative random variable ξ with an everywhere continuous cumulative distribution function (in particular, $\mathbb{P}\{\xi = 0\} = 0$), define numbers $\tau_k(\xi)$ (**levels**) as

$$\tau_k(\xi) := \inf\{\tau \geq 0 : \mathbb{P}\{\xi \geq \tau\} = 2^{-k}\}, \quad k \geq 0.$$

Note that

$$\mathbb{E}\xi \geq \sum_{k=0}^{\infty} 2^{-k-1}\tau_k(\xi). \tag{2}$$

We will need the following standard fact:

Lemma 2.4 (see, for example, [26, Chapter 1, Theorem 3.1]): *Let ξ_1, ξ_2 be two random variables on a probability space $(\Omega, \Sigma, \mathbb{P})$, and assume that*

$$\mathbb{P}\{\xi_1 > t\} \geq \mathbb{P}\{\xi_2 > t\}$$

for all $t \in \mathbb{R}$ (that is, ξ_2 is stochastically dominated by ξ_1). Then there is a probability space $(\widetilde{\Omega}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$ and random variables $\widetilde{\xi}_1, \widetilde{\xi}_2$ on $(\widetilde{\Omega}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$ such that (1) $\widetilde{\xi}_i$ is equidistributed with ξ_i, $i = 1, 2$, and (2) $\widetilde{\xi}_1 \geq \widetilde{\xi}_2$ everywhere on $\widetilde{\Omega}$.

Lemma 2.5 (Coupling): *Let $Y = (y_1, y_2, \ldots, y_n)$ be a random vector on a probability space $(\Omega, \Sigma, \mathbb{P})$ with i.i.d. non-negative coordinates with everywhere continuous cdf and $\mathbb{E}y_i = 1$. Further, let ξ_i^k ($i \leq n$, $k = 0, 1, \ldots$) be 0-1 variables on $(\Omega, \Sigma, \mathbb{P})$ with $\mathbb{P}\{\xi_i^k = 1\} = 2^{-k}$, and such that ξ_i^k are jointly independent for all $i \leq n$ and $k \geq 0$, and set*

$$z_i := \sum_{k=0}^{\infty} \tau_{k+1}(y_i)\xi_i^k, \quad i = 1, 2, \ldots, n,$$

and $Z := (z_1, z_2, \ldots, z_n)$. Then there is a probability space $(\widetilde{\Omega}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$ and random vectors $\widetilde{Y} = (\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_n)$ and $\widetilde{Z} = (\widetilde{z}_1, \widetilde{z}_2, \ldots, \widetilde{z}_n)$ on $(\widetilde{\Omega}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$ such that

(a) *$\widetilde{Y}$ and $\widetilde{Z}$ are equidistributed with Y and Z, respectively;*
(b) *$\widetilde{z}_i \geq \widetilde{y}_i$ for all $i \leq n$ everywhere on $(\widetilde{\Omega}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$.*

Proof. Fix for a moment $i \le n$ and consider the distributions of y_i and z_i. Take any $t > 0$. If $\tau_k(y_i) \le t$ for all $k \ge 0$ then, obviously,

$$\mathbb{P}\{z_i \ge t\} \ge 0 = \mathbb{P}\{y_i \ge t\}.$$

Otherwise, let $k(t) := \max\{k \ge 0 : \tau_k(y_i) \le t\}$. Then

$$\begin{aligned}\mathbb{P}\{z_i \ge t\} \ge& \mathbb{P}\{\tau_{k(t)+1}(y_i)\xi_i^{k(t)} \ge \tau_{k(t)+1}(y_i)\}\\ =& 2^{-k(t)} = \mathbb{P}\{y_i \ge \tau_{k(t)}(y_i)\}\\ \ge& \mathbb{P}\{y_i \ge t\}.\end{aligned}$$

Thus, y_i is stochastically dominated by z_i and, by Lemma 2.4, there is a probability space $(\widetilde{\Omega}_i, \widetilde{\Sigma}_i, \widetilde{\mathbb{P}}_i)$ and variables $\widetilde{y}_i$ and $\widetilde{z}_i$ on $(\widetilde{\Omega}_i, \widetilde{\Sigma}_i, \widetilde{\mathbb{P}}_i)$ equidistributed with y_i and z_i, respectively, such that $z_i \ge y_i$ everywhere on $\widetilde{\Omega}_i$.

Finally, by taking $(\widetilde{\Omega}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$ to be the product space $\prod_i \Omega_i$ and naturally extending the variables $\widetilde{y}_i, \widetilde{z}_i$ to $(\widetilde{\Omega}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$, we obtain the random vectors $\widetilde{Y}$, $\widetilde{Z}$ satisfying the required conditions. ∎

The next lemma provides an actual construction of the required diagonal operator.

Lemma 2.6: *For any $\alpha \in (0,1)$ there is $L = L(\alpha) > 0$ with the following property. Let $(\tau_k)_{k=1}^{\infty}$ be an increasing non-negative sequence satisfying $\sum_{k=1}^{\infty} \tau_k 2^{-k} < \infty$, and let*

$$\widetilde{Z} := \sum_{k=0}^{\infty} \tau_{k+1}\xi^k,$$

where $\xi^k = (\xi_1^k, \xi_2^k, \dots, \xi_n^k)$ and ξ_i^k ($i \le n$, $k = 0, 1, \dots$) are jointly independent 0-1 random variables with $\mathbb{P}\{\xi_i^k = 1\} = 2^{-k}$. Further, let $\delta \in (0,1]$. Then there is a random positive contraction $\widetilde{D}$ taking values in $\mathcal{D}_n$ such that

$$\|\widetilde{D}\widetilde{Z}\|_1 \le \frac{L}{\delta}\mathbb{E}\|\widetilde{Z}\|_1 = \frac{Ln}{\delta}\sum_{k=0}^{\infty}\tau_{k+1}2^{-k} \text{ everywhere on the probability space,}$$

and

$$\mathbb{E}(\det \widetilde{D})^{\alpha-1} \le \exp(\delta).$$

Proof. Let $L \ge 2e$ be a number which we will determine later. Now, for each $k \ge 0$, define random variables

$$\nu_k := |\{i : \xi_i^k \neq 0\}|$$

and

$$\eta_k := \begin{cases} (\frac{\delta\nu_k}{L2^{-k}n})^{\nu_k}, & \text{if } \delta\nu_k \geq L2^{-k}n; \\ 1, & \text{otherwise.} \end{cases}$$

As building blocks of the contraction $\widetilde{D}$, let us consider random diagonal matrices $D^{(k)}$ with

$$d_{jj}^{(k)} := \begin{cases} 1, & \text{if } \xi_j^k = 0; \\ \min(1, \frac{L2^{-k}n}{\delta\nu_k}), & \text{otherwise,} \end{cases} \qquad j = 1, 2, \dots, n.$$

Then $\det D^{(k)} = {\eta_k}^{-1}$ and $\|D^{(k)}\xi^k\|_1 \leq \frac{L2^{-k}n}{\delta} = \frac{L}{\delta}\mathbb{E}\|\xi^k\|_1$ (deterministically). Note that $D^{(k)}$ acts as a dilation on the span of $\{e_i : \xi_i^k \neq 0\}$ provided that $\nu_k \geq \frac{L2^{-k}n}{\delta} = \frac{L}{\delta}\mathbb{E}\nu_k$, and as an isometry on the orthogonal complement. We construct the required contraction $\widetilde{D}$ as the product of contractions $D^{(k)}$ by setting $\widetilde{D} := \prod_{k=0}^{\infty} D^{(k)}$. Then

$$\|\widetilde{D}\widetilde{Z}\|_1 \leq \Big\| \sum_{k=0}^{\infty} \tau_{k+1} D^{(k)}\xi^k \Big\|_1 \leq \frac{Ln}{\delta} \sum_{k=0}^{\infty} \tau_{k+1} 2^{-k} = \frac{L}{\delta}\mathbb{E}\|\widetilde{Z}\|_1.$$

Note that

$$\mathbb{E}(\det \widetilde{D})^{\alpha-1} = \mathbb{E}\prod_{k=0}^{\infty} {\eta_k}^{1-\alpha} = \prod_{k=0}^{\infty} \mathbb{E}{\eta_k}^{1-\alpha}.$$

Next, for every $k \geq 0$ we have

$$\begin{aligned} \mathbb{E}{\eta_k}^{1-\alpha} \leq& 1 + \sum_{m=\lceil L2^{-k}n/\delta\rceil}^{\infty} \Big(\frac{\delta m}{L2^{-k}n}\Big)^{m-\alpha m} \mathbb{P}\{\nu_k = m\} \\ \leq& 1 + \sum_{m=\lceil L2^{-k}n/\delta\rceil}^{\infty} \Big(\frac{e\delta}{L}\Big)^{m} \Big(\frac{L2^{-k}n}{\delta m}\Big)^{\alpha m}. \end{aligned}$$

In particular, for all k such that $L2^{-k}n/\delta \geq 1$, using the relation $L \geq 2e$, we obtain

$$\mathbb{E}{\eta_k}^{1-\alpha} \leq 1 + 2\Big(\frac{e\delta}{L}\Big)^{\lceil L2^{-k}n/\delta\rceil},$$

and for all k satisfying $L2^{-k}n/\delta < 1$, we get

$$\mathbb{E}{\eta_k}^{1-\alpha} \leq 1 + 2\frac{e\delta}{L}(L2^{-k}n/\delta)^{\alpha}.$$

Now, let us choose $L = L(\alpha)$ sufficiently large so that both

$$\sum_{k:\, L2^{-k}n/\delta\geq 1} 2\Big(\frac{e\delta}{L}\Big)^{\lceil L2^{-k}n/\delta\rceil} \quad \text{and} \quad \sum_{k:\, L2^{-k}n/\delta<1} 2\frac{e\delta}{L}(L2^{-k}n/\delta)^{\alpha}$$

are less than $\delta/2$. Then, multiplying the estimates for $\mathbb{E}{\eta_k}^{1-\alpha}$, we get

$$\mathbb{E}\Big(\prod_{k=0}^{\infty}\eta_k\Big)^{1-\alpha}\leq \exp(\delta),$$

and the result follows. ∎

Proof of Proposition 2.1. Fix admissible α, δ and p. Without loss of generality, the distribution of the coordinates of the random vector X is continuous on the real line. Indeed, otherwise we can replace every coordinate x_i with $|x_i|+u_i$, where $u_1,u_2,\dots,u_n$ are jointly independent with $x_1,x_2,\dots,x_n$ and each u_i is uniformly distributed on $[0,\theta]$ for a very small parameter $\theta>0$ chosen so that $\mathbb{E}(|x_i|+u_i)^p\approx\mathbb{E}|x_i|^p$. Then the random diagonal contraction D constructed for the new vector $X':=(|x_i|+u_i)_{i=1}^n$ will also satisfy the required properties with respect to X.

Set $Y:=(|x_1|^p,|x_2|^p,\dots,|x_n|^p)$ and let $\widetilde{Y},\widetilde{Z}$ be random vectors on a space $(\widetilde{\Omega},\widetilde{\Sigma},\widetilde{\mathbb{P}})$ constructed in Lemma 2.5 with respect to Y. By Lemma 2.6 and in view of relation (2), we can find a random positive contraction $\widetilde{D}$ on $\widetilde{\Omega}$ taking values in $\mathcal{D}_n$ such that for some $L=L(\alpha)>0$ we have

$$\|\widetilde{D}\widetilde{Y}\|_1\leq\|\widetilde{D}\widetilde{Z}\|_1\leq\frac{L}{\delta}\mathbb{E}\|\widetilde{Z}\|_1\leq\frac{4L}{\delta}\mathbb{E}\|\widetilde{Y}\|_1 \text{ everywhere on } \widetilde{\Omega}$$

and

$$\mathbb{E}(\det\widetilde{D})^{\alpha-1}\leq\exp(\delta).$$

In general, the operator $\widetilde{D}$ is not a function of $\widetilde{Y}$, which creates (purely technical) issues in defining a corresponding operator on the original space $(\Omega,\Sigma,\mathbb{P})$. For completeness, let us describe an elementrary discretization argument resolving the problem:

Let $\{B_z\}$ be a partition of $\mathbb{R}^n_+$ into Borel subsets, indexed over

$$z=(z_1,z_2,\dots,z_n)\in(\mathbb{Z}\cup\{-\infty\})^n$$

and defined by

$$B_z:=\{W\in\mathbb{R}^n_+:\,W_i\in(2^{z_i-1},2^{z_i}]\text{ for all }i=1,2,\dots,n\}$$

(we set $W_i=0$ for $z_i=-\infty$). Further, for every z we let

$$\widetilde{\Omega}_z:=\{\widetilde{\omega}\in\widetilde{\Omega}:\,\widetilde{Y}(\widetilde{\omega})\in B_z\}$$

and

$$Q_z:=\widetilde{D}(\widetilde{\Omega}_z)=\{M\in\mathcal{D}_n:\,M=\widetilde{D}(\widetilde{\omega})\text{ for some }\widetilde{\omega}\in\widetilde{\Omega}_z\}.$$

For each $z \in (\mathbb{Z} \cup \{-\infty\})^n$ such that $\widetilde{\Omega}_z$ is non-empty, choose an operator D_z from the closure of Q_z such that $\det D_z \geq \det M$ for all $M \in Q_z$ (of course, the choice of D_z does not have to be unique). Otherwise, if $\widetilde{\Omega}_z$ is empty then we set $D_z := \min(1, \frac{4L}{\delta \sum_{i=1}^n 2^{z_i}} \mathbb{E}\|\widetilde{Y}\|_1)\mathrm{Id}_n$. Finally, define a function $h : \mathbb{R}_+^n \to \mathcal{D}_n$ by setting $h(W) := D_z$ for all $W \in B_z$ and $z \in (\mathbb{Z} \cup \{-\infty\})^n$. Observe that h is Borel. Further, by the choice of D_z's, we have $\det h(\widetilde{Y}) \geq \det \widetilde{D}$ everywhere on $\widetilde{\Omega}$, whence $\mathbb{E}(\det h(\widetilde{Y}))^{\alpha-1} \leq \exp(\delta)$. Next, by the choice of sets B_z, we have $\|M(W)\|_1 \leq 2\|M'(W')\|_1$ for any two couples $(M, W), (M', W') \in Q_z \times B_z$. Together with the conditions on $\widetilde{D}$ and the definition of D_z's, this implies $\|D_z(W)\|_1 \leq \frac{8L}{\delta}\mathbb{E}\|\widetilde{Y}\|_1$ for all $W \in B_z$, whence

$$\|h(W)\, W\|_1 \leq \frac{8L}{\delta}\mathbb{E}\|\widetilde{Y}\|_1 \text{ everywhere on } \mathbb{R}_+^n.$$

Now, taking $T := h(Y)$, we obtain a random diagonal contraction on $(\Omega, \Sigma, \mathbb{P})$ such that

$$\|T^{1/p}X\|_p^p = \|TY\|_1 \leq \frac{8L}{\delta}\mathbb{E}\|X\|_p^p \text{ everywhere on } \Omega$$

and $\mathbb{E}(\det T)^{\alpha-1} \leq \exp(\delta)$. Finally, setting $D := T^{1/p}$, we get the required operator. ∎

The above statement can be "tensorized". In what follows, we are interested only in the case $p = 2$ and $\alpha = 1/2$.

Proposition 2.7: *There is a universal constant $C > 0$ with the following property. Let $A = (a_{ij})$ be an $n \times n$ random matrix satisfying $(*)$, and let $\delta \in (0, 1]$. Then there is a random positive contraction D taking values in $\mathcal{D}_n$ such that the Euclidean norms of the rows of AD are uniformly bounded by $\frac{C}{\sqrt{\delta}}\sqrt{n}$ everywhere on the probability space, and*

$$\mathbb{E}\det D^{-1} \leq \exp(\delta n).$$

Proof. Indeed, for any $i = 1, 2, \ldots, n$, let D_i be the positive contraction defined with respect to the i-th row of A using Proposition 2.1 (with parameters $\alpha = 1/2, p = 2$), so that $D_1, D_2, \ldots, D_n$ are jointly independent. Then the product of these contractions $D := \prod_{i=1}^n D_i$ satisfies the required conditions. ∎

Remark 2.8: It is not difficult to see that for any positive contraction $M \in \mathcal{D}_n$ there is an element $\widetilde{M} \in \mathcal{D}_n^2$ such that $\widetilde{M} \leq \sqrt{2}M$ and $\det \widetilde{M}^{-1} \leq \det M^{-2}$. Indeed, this follows easily from the fact that for any number $t \in (0, 1]$ there is $\widetilde{t} \in \{1\} \cup \{2^{-2^k}\}_{k=0}^{\infty}$ with $t^2 \leq \widetilde{t} \leq \sqrt{2}t$ (the constant $\sqrt{2}$ on the right-hand side is achieved for $t = \sqrt{2}/2 - o(1)$). Hence, the above statement implies that, given a matrix A satisfying $(*)$ and a number $\delta > 0$, one can construct a random contraction $\widetilde{D}$ taking values in $\mathcal{D}_n^2$ such that each row of $A\widetilde{D}$ has Euclidean norm at most $\frac{C}{\sqrt{\delta}}\sqrt{n}$ (for some universal constant $C > 0$), and $\mathbb{E}\det\widetilde{D}^{-1/2} \le \exp(\delta n)$.

3. Coverings of random ellipsoids

The main result of the section is

Theorem 3.1: *Let $\delta \in (0,1]$ and let $A = (a_{ij})$ be an $n \times n$ random matrix satisfying $(*)$. Then*

$$\mathbb{P}\Big\{\exists D \in \mathcal{D}_n^2 : \det D \ge \exp(-\delta n) \text{ and } \|AD\|_{\infty\to 2} \le \frac{C_{3.1}}{\sqrt{\delta}} n\Big\} \ge 1 - 4\exp(-\delta n/8),$$

where $C_{3.1} > 0$ is a universal constant.

Remark 3.2: The above theorem can be seen as a way to "regularize" the random matrix A by reducing its norm while preserving its "structure". In this connection, let us mention work [10] where a very general problem of regularizing random matrices was discussed (see [10, Section 5.4]).

As we have mentioned in the introduction, Theorem A follows almost immediately from the above statement; we give the proof of Theorem A at the very end of the section. The section is organized as follows. First, we use $\widetilde{D}$ constructed in Remark 2.8 to verify Theorem 3.1 under an additional assumption that the entries of A are symmetrically distributed (see Proposition 3.6). Then, we will apply a symmetrization procedure to prove Theorem 3.1 in full generality.

A random variable ξ is **subgaussian** if there exists a number $K > 0$ such that

$$\mathbb{P}\{|\xi| > t\} \le 2\exp(-t^2/K^2), \quad t > 0. \tag{3}$$

To put an emphasis on the value of K, we will sometimes call ξ K-subgaussian. We note that the smallest value of K satisfying (3) is equivalent to the **subgaussian norm** of ξ (see, for example, [29, Lemma 5.5]); however, the latter notion is less convenient for us and will not be used in this paper.

The next lemma is equivalent to a standard Khintchine-type inequality (see, for example, [7]).

Lemma 3.3: *Let $r_1, r_2, \dots, r_n$ be independent Rademacher random variables. Then for any vector $y \in S^{n-1}$ the random variable $\sum_{i=1}^n y_i r_i$ is $C_{3.3}$-subgaussian, where $C_{3.3} > 0$ is a universal constant.*

The sum of squares of subgaussian variables has good concentration properties; the bound below follows from a standard "Laplace transform" argument (see, for example, [29, Corollary 5.17]):

Lemma 3.4: *For any $T > 0$ there is $L_{3.4} > 0$ depending on T with the following property: Let $\xi_1, \xi_2, \dots, \xi_n$ be independent centered 1-subgaussian random variables. Then*

$$\mathbb{P}\Big\{ \sum_{i=1}^n \xi_i^2 > L_{3.4} n \Big\} \le \exp(-Tn).$$

The next proposition implies that for a random matrix A satisfying $(*)$ with symmetrically distributed entries and the operator $\widetilde{D}$ from Remark 2.8, the norm $\|A\widetilde{D}\|_{\infty\to 2}$ can be efficiently bounded from above as long as $\widetilde{D}$ is a Borel function of $|A|$ (here and further in the text, given a matrix $B = (b_{ij})$, by $|B|$ we shall denote the matrix $(|b_{ij}|)$).

Proposition 3.5: *Let $K > 0$ and let A be an $n \times n$ random matrix satisfying $(*)$, with symmetrically distributed entries. Further, let $\mathcal{F} \subset \mathcal{D}_n$ be any countable subset. Denote by $\mathcal{E}$ the event*

$$\mathcal{E} := \{\exists D \in \mathcal{F} : \text{ all rows of } AD \text{ have Euclidean norms at most } K\sqrt{n}\}.$$

Then

$$\mathbb{P}\{\exists D \in \mathcal{F} : \|AD\|_{\infty\to 2} \le CKn\} \ge \mathbb{P}(\mathcal{E}) - \exp(-n),$$

where $C > 0$ is a universal constant.

Proof. Fix any admissible K and $\mathcal{F}$. Clearly, for any $n \times n$ matrix B and a diagonal matrix D, the Euclidean norms of rows of BD and $|B|D$ are the same. Hence, we may assume that there is a Borel function $f : \mathbb{R}_+^{n\times n} \to \mathcal{F}$ such that

$$\mathcal{E} = \{\text{all rows of } A\, f(|A|) \text{ have norms at most } K\sqrt{n}\}.$$

For any $D \in \mathcal{F}$, let

$$\mathcal{E}_D := \mathcal{E} \cap \{f(|A|) = D\}.$$

Without loss of generality, $\mathbb{P}(\mathcal{E}_D) > 0$ for any $D \in \mathcal{F}$.

Next, as the unit cube $[-1,1]^n$ is the convex hull of its vertices $V=\{-1,1\}^n$, we have

$$\|Af(|A|)\|_{\infty\to 2}=\sup_{y\in B_\infty^n}\|Af(|A|)y\|=\sup_{v\in V}\|Af(|A|)v\|. \tag{4}$$

Note that, given event $\mathcal{E}_D$, the entries of $Af(|A|)=AD$ are symmetrically distributed, so the distribution of ADv given $\mathcal{E}_D$ is the same for any vertex $v\in V$. Fix a vertex v.

Observe that for any $t>0$ we have

$$\mathbb{P}_{\mathcal{E}_D}\{\|ADv\|>t\}\le\sup_B\mathbb{P}\{\|\widetilde{B}Dv\|>t\}, \tag{5}$$

where by $\mathbb{P}_{\mathcal{E}_D}$ we denote the conditional probability given $\mathcal{E}_D$ and the supremum is taken over all matrices $B=(b_{ij})$ such that the rows of BD have Euclidean norms at most $K\sqrt{n}$, and $\widetilde{B}=(r_{ij}b_{ij})$, with r_{ij} being jointly independent Rademacher (±1) variables. Fix any admissible $B=(b_{ij})$.

Then the variables $\langle\widetilde{B}Dv,e_i\rangle$, $i=1,2,\dots,n$, are jointly independent and, in view of Lemma 3.3 and the choice of B, each variable $K^{-1}n^{-1/2}\langle\widetilde{B}Dv,e_i\rangle$ is $C_{3.3}$-subgaussian. By Lemma 3.4, there is a universal constant $C>0$ such that

$$\mathbb{P}\{\|\widetilde{B}Dv\|>CKn\}=\mathbb{P}\Big\{\frac{1}{n}\sum_{i=1}^n\langle\widetilde{B}Dv,e_i\rangle^2>(CK)^2n\Big\}\le\exp(-(1+\ln 2)n).$$

Then, taking a union bound over 2^n vertices of the unit cube and using (5) and (4), we get an estimate

$$\mathbb{P}_{\mathcal{E}_D}\{\|AD\|_{\infty\to2}>CKn\}\le 2^n\cdot\sup_B\mathbb{P}\{\|\widetilde{B}Dv\|>CKn\}\le\exp(-n).$$

Finally, clearly

$$\begin{aligned}\mathbb{P}\{\|AD\|_{\infty\to2}>CKn\}\le&\mathbb{P}(\mathcal{E}^c)+\sum_D\mathbb{P}_{\mathcal{E}_D}\{\|AD\|_{\infty\to2}>CKn\}\mathbb{P}(\mathcal{E}_D)\\ \le&\mathbb{P}(\mathcal{E}^c)+\exp(-n),\end{aligned}$$

and the result follows. ∎

Proposition 3.6: *Let $\delta\in(0,1]$ and let $A=(a_{ij})$ be an $n\times n$ random matrix satisfying $(*)$, with symmetrically distributed entries. Then*

$$\mathbb{P}\Big\{\exists D\in\mathcal{D}_n^2:\det D\ge\exp(-\delta n)\ \text{and}\ \|AD\|_{\infty\to2}\le\frac{C_{3.6}}{\sqrt{\delta}}n\Big\}\ge1-2\exp(-\delta n/4).$$

Proof. Fix any $\delta \in (0,1]$. In view of Remark 2.8, there is a random contraction D taking values in $\mathcal{D}_n^2$ such that each row of AD has the Euclidean norm at most $\frac{C}{\sqrt{\delta}}\sqrt{n}$ and $\mathbb{E}\det D^{-1/2} \le \exp(\delta n/4)$. Denote by $\mathcal{E}$ the event

$$\mathcal{E} := \{\det D \ge \exp(-\delta n)\}.$$

In view of the conditions on D and Markov's inequality, we have

$$\mathbb{P}(\mathcal{E}) \ge 1 - \exp(-\delta n/4).$$

Hence, by Proposition 3.5, taking $\mathcal{F}$ to be the set of all contractions from $\mathcal{D}_n^2$ having determinant at least $\exp(-\delta n)$, we obtain

$$\mathbb{P}\Big\{\exists D \in \mathcal{D}_n^2 : \det D \ge \exp(-\delta n) \text{ and } \|AD\|_{\infty\to 2} \le \frac{C_{3.6}}{\sqrt{\delta}} n\Big\} \ge 1 - \exp(-\delta n/4) - \exp(-n)$$

for a universal constant $C_{3.6} > 0$. ∎

For the next lemma we will need the following definition (essentially taken from [13]). Let S be a finite set and d be a pseudometric on S. We say that (S,d) is **of length at most** ℓ (for some $\ell > 0$) if there is $n \in \mathbb{N}$, positive numbers $b_1, b_2, \dots, b_n$ with $\|(b_1, b_2, \dots, b_n)\| \le \ell$ and a sequence $(S_k)_{k=0}^n$ of partitions of S such that

(1) $S_0 = \{S\}$;
(2) $S_n = \{\{s\}\}_{s\in S}$;
(3) S_k is a refinement of S_{k-1} for all $k = 1, 2, \dots, n$;
(4) for each $k \in \{1, 2, \dots, n\}$ and any $Q, Q' \in S_k$ such that $Q \cup Q'$ is a subset of an element of S_{k-1}, there is a one-to-one mapping $\phi : Q \to Q'$ such that $d(s, \phi(s)) \le b_k$ for all $s \in Q$.

In particular, the above conditions on S_k imply that all elements of S_k have the same cardinality.

THEOREM 3.7 (see [13, Theorem 7.8]): *Let (S,d) be a finite pseudometric space of length at most ℓ and let μ be the normalized counting measure on S. Then for any function $f : S \to \mathbb{R}$ satisfying $|f(s) - f(s')| \le d(s,s')$ $(s, s' \in S)$ and all $t > 0$ we have*

$$\mu\Big\{\Big|f - \int f\,d\mu\Big| \ge t\Big\} \le 2\exp\Big(-\frac{t^2}{4\ell^2}\Big).$$

Remark 3.8: In [13], the above theorem is formulated for metric spaces. It is easy to see that passing to pseudometrics does not change the picture.

Denote by Π_n the set of permutations of $[n] := \{1, 2, \dots, n\}$.

Lemma 3.9: *Let $y = (y_1, y_2, \dots, y_n)$ be a non-zero vector and $v = (v_1, v_2, \dots, v_n)$ be a vertex of the cube $[-1, 1]^n$. Further, let μ be the normalized counting measure on Π_n. Define a function $f : \Pi_n \to \mathbb{R}$ as*

$$f(p) := \sum_{j=1}^{n} v_{p(j)} y_j, \quad p \in \Pi_n.$$

Then

$$\mu\Big\{\Big|f - \int f\, d\mu\Big| \geq t\Big\} \leq 2 \exp\Big(-\frac{t^2}{64\|y\|^2}\Big), \quad t > 0.$$

Proof. Without loss of generality, we can assume that $|y_j| \geq |y_{j+1}|$ $(j = 1, 2, \dots, n-1)$. Define a pseudometric d on Π_n: for any $p, q \in \Pi_n$ let

$$d(p, q) := |f(p) - f(q)|.$$

Further, we define a sequence of partitions $(\Pi_{n,k})_{k=0}^{n}$ of Π_n: let $\Pi_{n,0} := \{\Pi_n\}$ and for each $k = 1, 2, \dots, n$, let $\Pi_{n,k}$ consist of all subsets of Π_n of the form

$$\{p \in \Pi_n : \ p(1) = i_1, p(2) = i_2, \dots, p(k) = i_k\}$$

for all $\{i_1, i_2, \dots, i_k\} \subset [n]$.

Now, let $k \in \{1, 2, \dots, n\}$ and let $Q, Q' \in \Pi_{n,k}$ be such that $Q \cup Q'$ is a subset of an element of $\Pi_{n,k-1}$. Note that there are numbers $i_1, i_2, \dots, i_k$, i'_k such that $p(j) = i_j$ for all $j < k$ and $p \in Q \cup Q'$; $p(k) = i_k$ for all $p \in Q$ and $p(k) = i'_k$ for all $p \in Q'$. Define a one-to-one mapping $\phi : Q \to Q'$ by

$$\phi(p)(j) := p(j) \text{ for } j \neq k, p^{-1}(i'_k); \quad \phi(p)(k) := i'_k; \quad \phi(p)(p^{-1}(i'_k)) := i_k.$$

For any $p \in Q$, we have

$$d(p, \phi(p)) \leq 2|y_k| + 2|y_{p^{-1}(i'_k)}| \leq 4|y_k|,$$

with the last inequality due to the fact that $p^{-1}(i'_k) \geq k$. Thus, the space (Π_n, d) is of length at most $4\|y\|$. Applying Theorem 3.7, we get the result. ∎

The next statement shall be used in a symmetrization argument within the proof of Theorem 3.1; we think it may be of interest in itself.

Proposition 3.10: *Let $B = (b_{ij})$ be a non-random $n \times n$ matrix such that the Euclidean norm of every row is at most $\sqrt{n}$ and such that*

$$\Big|\sum_{j=1}^{n} b_{ij}\Big| \le \sqrt{n}, \quad i = 1, 2, \dots, n.$$

Further, let π_i ($i = 1, 2, \dots, n$) be independent random permutations uniformly distributed on Π_n, and denote by $\widetilde{B} = (\widetilde{b}_{ij})$ the random $n \times n$ matrix with entries defined by

$$\widetilde{b}_{ij} := b_{i,\pi_i(j)}.$$

Then

$$\mathbb{P}\{\|\widetilde{B}\|_{\infty\to 2} \le C_{3.10} n\} \ge 1 - \exp(-n)$$

for a universal constant $C_{3.10} > 0$.

Proof. We will show that for any $v \in \{-1, 1\}^n$ we have

$$\mathbb{P}\{\|\widetilde{B}v\| > C_{3.10} n\} \le \exp(-n - n\ln 2)$$

for a sufficiently large universal constant $C_{3.10}$ and then take the union bound over the vertices of the cube.

Fix any $v = (v_1, v_2, \dots, v_n) \in \{-1, 1\}^n$ and let m be the number of ones in $(v_1, \dots, v_n)$. Clearly, the random variables $\langle \widetilde{B}v, e_i\rangle$ ($i = 1, 2, \dots, n$) are independent. Next, for a fixed i, the distribution of $\langle \widetilde{B}v, e_i\rangle$ coincides with that of the variable $\xi_i := \sum_{j=1}^{n} v_{\pi_i(j)} b_{ij}$. By Lemma 3.9 and in view of the condition on the rows of B, we have

$$\mathbb{P}\{|\xi_i - \mathbb{E}\xi_i| > \tau\} \le 2\exp\Big(-\frac{\tau^2}{64n}\Big), \quad \tau > 0.$$

Hence, the variables $n^{-1/2}(\xi_i - \mathbb{E}\xi_i)$ ($i = 1, 2, \dots, n$) are C-subgaussian for a universal constant $C > 0$. In view of Lemma 3.4, we get that

$$\mathbb{P}\Big\{\sum_{i=1}^{n} (\xi_i - \mathbb{E}\xi_i)^2 > \widetilde{C} n^2\Big\} \le \exp(-n - n\ln 2) \tag{6}$$

for some constant $\widetilde{C} > 0$. Finally, observe that

$$\sum_{i=1}^{n} \xi_i^2 \le 2\sum_{i=1}^{n} (\xi_i - \mathbb{E}\xi_i)^2 + 2\sum_{i=1}^{n} (\mathbb{E}\xi_i)^2 \quad \text{(deterministically)},$$

so, applying the estimate

$$|\mathbb{E}\xi_i| = \left|\frac{2m-n}{n}\sum_{j=1}^n b_{ij}\right| \le \sqrt{n}$$

and (6), we obtain

$$\mathbb{P}\{\|\widetilde{B}v\|^2 > (2\widetilde{C}+2)n^2\} = \mathbb{P}\bigg\{\sum_{i=1}^n \xi_i^2 > (2\widetilde{C}+2)n^2\bigg\} \le \exp(-n-n\ln 2). \qquad \blacksquare$$

Proof of Theorem 3.1. Let $\widetilde{A}$ be an independent copy of A. Obviously

$$\mathbb{E}\bigg(\sum_{j=1}^n \widetilde{a}_{ij}\bigg)^2 = \mathbb{E}\sum_{j=1}^n \widetilde{a}_{ij}^2 = n$$

for every $i = 1, 2, \dots, n$. Then, in view of Markov's inequality, each row of $\widetilde{A}$ satisfies

$$\bigg|\sum_{j=1}^n \widetilde{a}_{ij}\bigg| \le \sqrt{\frac{32n}{\delta}} \quad \text{and} \quad \sum_{j=1}^n \widetilde{a}_{ij}^2 \le \frac{32n}{\delta}$$

with probability at least $1-\delta/16 > \exp(-\delta/8)$. Denote by $\widetilde{\mathcal{E}}$ the event

$$\widetilde{\mathcal{E}} := \bigg\{\bigg|\sum_{j=1}^n \widetilde{a}_{ij}\bigg| \le \sqrt{\frac{32n}{\delta}} \text{ and } \sum_{j=1}^n \widetilde{a}_{ij}^2 \le \frac{32n}{\delta} \text{ for all } i = 1, 2, \dots, n\bigg\}.$$

In view of the above, $\mathbb{P}(\widetilde{\mathcal{E}}) \ge \exp(-\delta n/8)$. Let $\pi_1, \pi_2, \dots, \pi_n$ be random permutations uniformly distributed on Π_n and jointly independent with $\widetilde{A}$, and denote by $\widetilde{B} = (\widetilde{b}_{ij})$ the random matrix with entries $\widetilde{b}_{ij} := \widetilde{a}_{i,\pi_i(j)}$ $(i, j \le n)$. Then Proposition 3.10 yields

$$\mathbb{P}\bigg\{\|\widetilde{B}\|_{\infty\to 2} \le C_{3.10}\sqrt{n}\max_{i\le n}\bigg(\sum_{j=1}^n \widetilde{a}_{ij}^2\bigg)^{1/2} \mid \widetilde{A}\bigg\} \ge 1-\exp(-n),$$

whence, in particular,

$$\mathbb{P}\{\|\widetilde{B}\|_{\infty\to 2} \le C_{3.10}\sqrt{32/\delta}\, n \mid \widetilde{\mathcal{E}}\} \ge 1-\exp(-n).$$

But $\widetilde{B}$ is equidistributed with $\widetilde{A}$ given $\widetilde{\mathcal{E}}$, so that

$$\mathbb{P}\{\|\widetilde{A}\|_{\infty\to 2} \le C_{3.10}\sqrt{32/\delta}\, n \mid \widetilde{\mathcal{E}}\} \ge 1-\exp(-n).$$

Clearly, $\|\widetilde{A}D\|_{\infty\to 2} \le \|\widetilde{A}\|_{\infty\to 2}$ for any contraction $D \in \mathcal{D}_n$ (deterministically), so we obtain for the event $\mathcal{E}_1 := \{\|\widetilde{A}D\|_{\infty\to 2} \le C_{3.10}\sqrt{32/\delta}\, n \text{ for all } D \in \mathcal{D}_n\}$

$$\mathbb{P}(\mathcal{E}_1) \ge (1-\exp(-n))\mathbb{P}(\widetilde{\mathcal{E}}) \ge \frac{1}{2}\exp(-\delta n/8).$$

Next, the matrix $2^{-1/2}(A-\widetilde{A})$ has symmetrically distributed entries, and satisfies conditions of Proposition 3.6. Hence,

$$\begin{aligned}\mathbb{P}\{\|(A-\widetilde{A})D\|_{\infty\to 2}\le C_{3.6}\sqrt{2/\delta}\,n \text{ for some } D\in\mathcal{D}_n^2 \text{ with } \det D\ge\exp(-\delta n)\}\\ \ge 1-2\exp(-\delta n/4).\end{aligned}$$

Conditioning on $\mathcal{E}_1$, we get

$$\begin{aligned}\mathbb{P}\{\|(A-\widetilde{A})D\|_{\infty\to 2}\le C_{3.6}\sqrt{2/\delta}\,n \text{ for some } D\in\mathcal{D}_n^2 \text{ with } \det D\ge\exp(-\delta n)\,|\,\mathcal{E}_1\}\\ \ge 1-\frac{2\exp(-\delta n/4)}{\mathbb{P}(\mathcal{E}_1)}\\ \ge 1-4\exp(-\delta n/8).\end{aligned}$$

Note that, given $\mathcal{E}_1$, we have $\|AD\|_{\infty\to 2}\le\|(A-\widetilde{A})D\|_{\infty\to 2}+C_{3.10}\sqrt{32/\delta}\,n$ for all contractions $D\in\mathcal{D}_n$. Combining this with the last formula, we obtain

$$\begin{aligned}&\mathbb{P}\{\|AD\|_{\infty\to 2}\le C_{3.6}\sqrt{2/\delta}\,n+C_{3.10}\sqrt{32/\delta}\,n\\ &\quad\text{for some } D\in\mathcal{D}_n^2 \text{ with } \det D\ge\exp(-\delta n)\,|\,\mathcal{E}_1\}\ge 1-4\exp(-\delta n/8).\end{aligned}$$

Finally, since A is independent of $\mathcal{E}_1$, the conditioning in the last estimate can be dropped, and we obtain the statement. ∎

To complete the proof of Theorem A, we will need two more technical lemmas:

Lemma 3.11: *For any $\delta\in(0,1/2]$ and all $n\in\mathbb{N}$ we have*

$$|\{D\in\mathcal{D}_n^2\,:\,\det D\ge\exp(-\delta n)\}|\le\Big(\frac{2e}{\delta}\Big)^{4\delta n}.$$

Proof. Denote $\mathcal{S}:=\{D\in\mathcal{D}_n^2\,:\,\det D\ge\exp(-\delta n)\}$. Note that for any matrix $D\in\mathcal{S}$ and for any $k\ge 0$, the number of diagonal elements of D equal to 2^{-2^k} is less than $2^{-k+1}\delta n$. Hence, the cardinality of $\mathcal{S}$ can be estimated as

$$\begin{aligned}|\mathcal{S}|\le&\prod_{k=0}^{\infty}\binom{n}{\lfloor 2^{-k+1}\delta n\rfloor}\le\prod_{k=0}^{\infty}\Big(\frac{e}{\delta}\Big)^{2^{-k+1}\delta n}2^{k2^{-k+1}\delta n}\\ =&\Big(\frac{e}{\delta}\Big)^{4\delta n}2^{4\delta n}=\Big(\frac{2e}{\delta}\Big)^{4\delta n}.\end{aligned}$$ ∎

Lemma 3.12: *For any $n\in\mathbb{N}$ and $K\in[2,2\sqrt{n}]$, the unit Euclidean ball B_2^n can be covered by at most $(2eK^2)^{8n/K^2}$ translates of the cube $\frac{K}{\sqrt{n}}B_\infty^n$.*

Proof. First, note that for any $y \in B_2^n$ we have

$$\Big|\Big\{i \leq n :\, |y_i| \geq \frac{K}{2\sqrt{n}}\Big\}\Big| \leq \frac{4n}{K^2}.$$

Hence, it is sufficient to show that the set $|\{y \in B_2^n : |\mathrm{supp}(y)| \leq \frac{4n}{K^2}\}|$ can be covered by at most $(2eK^2)^{8n/K^2}$ translates of $\frac{K}{2\sqrt{n}} B_\infty^n$. A simple volumetric argument, together with an estimate $\mathrm{Vol}(B_2^m) \leq (\frac{2\pi e}{m})^{m/2}$, implies that B_2^m can be covered by at most 7^m translates of $\frac{1}{\sqrt{m}} B_\infty^m$ (for any $m \in \mathbb{N}$). As a consequence, we obtain a covering of $B_2^{\lceil 4n/K^2 \rceil}$ by at most $7^{\lceil 4n/K^2 \rceil}$ translates of $\frac{K}{2\sqrt{n}} B_\infty^n$. Finally, the cardinality of the optimal covering of $|\{y \in B_2^n : |\mathrm{supp}(y)| \leq \frac{4n}{K^2}\}|$ can be estimated from above by

$$\binom{n}{\lceil 4n/K^2 \rceil} 7^{\lceil 4n/K^2 \rceil} \leq (2eK^2)^{8n/K^2}. \qquad ∎$$

Proof of Theorem A. Let $\delta \in (0, 1/4]$ and $n \geq \frac{1}{4\delta}$. First, applying Lemma 3.12 with $K = 1/\sqrt{\delta}$, we see that B_2^n can be covered by $(2e/\delta)^{8n\delta}$ translates of the dilated cube $\frac{1}{\sqrt{n\delta}} B_\infty^n$. Let

$$\mathcal{Q} = \{D \in \mathcal{D}_n^2 :\, \det D \geq \exp(-\delta n)\}.$$

Then, in view of Lemma 3.11, we get that B_∞^n can be covered by at most $(2e/\delta)^{4\delta n} \exp(\delta n)$ parallelepipeds in such a way that for any $y \in B_\infty^n$ and $D \in \mathcal{Q}$, y is covered by a translate of $D(B_\infty^n)$. Combining the two coverings, we get a collection $\mathcal{C}$ of parallelepipeds covering B_2^n such that

$$|\mathcal{C}| \leq (2e/\delta)^{4\delta n} \exp(\delta n) \cdot (2e/\delta)^{8n\delta} = \exp\Big(\delta n + 12 n\delta \ln \frac{2e}{\delta}\Big),$$

and for any $y \in B_2^n$ and $D \in \mathcal{Q}$, the set $\mathcal{C}$ contains a translate of $D(\frac{1}{\sqrt{n\delta}} B_\infty^n)$ covering y. Finally, applying Theorem 3.1, we get that with probability at least $1 - 4\exp(-\delta n/8)$ for some $D \in \mathcal{Q}$ we have $AD(B_\infty^n) \subset \frac{C_{3.1} n}{\sqrt{\delta}} B_2^n$, implying

$$\mathbb{P}\Big\{\forall\, x \in B_2^n \;\; \exists\, P \in \mathcal{C} \text{ such that } x \in P \text{ and } A(P) \subset Ax + \frac{2 \cdot C_{3.1}\sqrt{n}}{\delta} B_2^n\Big\} \geq 1 - 4\exp(-\delta n/8)$$

(the multiple “2” in the last formula appears because the translation $-Ax + A(P)$ is not origin-symmetric in general). ∎

Proof of Corollary A. Fix n and δ, and let $\mathcal{C}$ be the collection of parallelepipeds defined in Theorem A. For each $P \in \mathcal{C}$, choose a point $y_p \in P \cap B_2^n$, and let $\mathcal{N} := \{y_P : P \in \mathcal{C}\}$. Then, clearly,

$$|\mathcal{N}| = |\mathcal{C}| \leq \exp\Big(\delta n + 12 n\delta \ln \frac{2e}{\delta}\Big),$$

and with probability at least $1 - 4\exp(-\delta n/8)$ for every $x \in B_2^n$ there is $y = y(x) \in \mathcal{N}$ with $-Ax + Ay \in \frac{C\sqrt{n}}{\delta} B_2^n$. In short,

$$\mathbb{P}\Big\{A(B_2^n) \subset \bigcup_{y\in A(\mathcal{N})} \Big(y + \frac{C\sqrt{n}}{\delta} B_2^n\Big)\Big\} \geq 1 - 4\exp(-\delta n/8). \qquad \blacksquare$$

4. The smallest singular value—Preliminaries

As we already mentioned in the introduction, the proof of Theorem B heavily relies on results obtained by Rudelson and Vershynin in papers [20] and [19]. In this section, we will state several intermediate results from those papers that we will need in Section 5 to complete our proof.

A crucial step in the proof of [20, Theorem 1.2] is a decomposition of the unit sphere into sets of "compressible" and "incompressible" vectors.

Definition 4.1 (Sparse, compressible and incompressible vectors): Fix parameters $\theta, \rho \in (0,1)$. A vector $x \in \mathbb{R}^n$ is called θn-**sparse** if $|\mathrm{supp}(x)| \leq \theta n$. A vector $x \in S^{n-1}$ is called **compressible** if x is within Euclidean distance ρ from the set of all θn-sparse vectors. Otherwise, x will be called **incompressible**. The set of all compressible unit vectors will be denoted by $\mathrm{Comp}_n(\theta,\rho)$, and the set of incompressible vectors by $\mathrm{Incomp}_n(\theta,\rho)$. Sometimes, when the dimension n or the parameters θ, ρ are clear from the context, we will simply write Comp, Incomp to denote the sets.

Remark 4.2: A similar decomposition of the unit sphere was already introduced in an earlier paper [11] for the purpose of bounding the smallest singular value of rectangular matrices.

Obviously, for any $\varepsilon > 0$ we have

$$\mathbb{P}\{s_n(A) < \varepsilon n^{-1/2}\} \leq \mathbb{P}\{\inf_{y\in\mathrm{Comp}} \|Ay\| < \varepsilon n^{-1/2}\} + \mathbb{P}\{\inf_{y\in\mathrm{Incomp}} \|Ay\| < \varepsilon n^{-1/2}\}.$$

Treatment of the compressible vectors is simpler due to the fact that the set Comp is "small"; we will deal with this set in the first part of Section 5. Let us remark that, unlike in the subgaussian result of [20], where an estimate for compressible vectors follows almost directly from an analogue of Lemma 4.9 (see below) together with a standard covering argument, in our case we will still need to use additional results (proved in Section 3) as the norm $\|A\|_{2\to 2}$ may be “too large”. We will need the following simple lemma:

Lemma 4.3: *For any $\theta, \rho \in (0,1]$ the set* Comp $=$ $\mathrm{Comp}_n(\theta,\rho)$ *admits a Euclidean 3ρ-net $\mathcal{N} \subset$ Comp of cardinality $|\mathcal{N}| \le (e/\theta)^{\theta n}(\frac{5}{\rho})^{\theta n}$.*

Proof. Note that the definition of Comp implies that for any $y \in$ Comp there is $y' \in S^{n-1}$ such that $|\mathrm{supp}(y')| \le \theta n$ and $\|y - y'\| \le 2\rho$. Hence, it is enough to show that one can find a Euclidean ρ-net $\mathcal{N}$ on the set of θn-sparse unit vectors, with the required estimate on $|\mathcal{N}|$. This follows from a standard estimate on the cardinality of an optimal ρ-net on $S^{\lfloor \theta n \rfloor - 1}$, together with a bound for the binomial coefficient $\binom{n}{\lfloor \theta \rfloor}$. ∎

Incompressible vectors have the important property that a significant portion of their coordinates are of order $n^{-1/2}$. In paper [20], this property was referred to as “incompressible vectors are spread”. For the reader’s convenience, we provide a proof of this fact below (let us note once again that analogous concepts were already considered in [11]).

Lemma 4.4 ([20, Lemma 3.4]): *For any $\theta, \rho \in (0,1)$ and for any vector $x \in \mathrm{Incomp}_n(\theta,\rho)$ there is a subset of indices $\sigma(x) \subset \{1,2,\ldots,n\}$ of cardinality at least $\frac{1}{2}\rho^2\theta n$ such that for all $i \in \sigma(x)$ we have*

$$\frac{\rho}{\sqrt{2n}} \le x_i \le \frac{1}{\sqrt{\theta n}}.$$

Proof. For every subset $I \subset \{1,2,\ldots,n\}$, let P_I be the coordinate projection onto the span of $\{e_i : i \in I\}$. Let $\sigma = \sigma(x) := \sigma_1 \cap \sigma_2$, where

$$\sigma_1 = \Big\{i \le n : |x_i| \le \frac{1}{\sqrt{\theta n}}\Big\}, \quad \sigma_2 = \Big\{i \le n : |x_i| \ge \frac{\rho}{\sqrt{2n}}\Big\}.$$

Since $\|x\| = 1$, we have $|\sigma_1^c| \le \theta n$, and $P_{\sigma_1^c}(x)$ is a θn-sparse vector. Then the condition that x is incompressible implies

$$\|P_{\sigma_1}(x)\| = \|x - P_{\sigma_1^c}(x)\| > \rho.$$

Hence,

$$\|P_\sigma(x)\|^2 \ge \|P_{\sigma_1}(x)\|^2 - \|P_{\sigma_2^c}(x)\|^2 \ge \rho^2 - n\cdot\|P_{\sigma_2^c}(x)\|_\infty^2 \ge \rho^2/2. \tag{7}$$

On the other hand, in view of the inclusion $\sigma(x) \subset \sigma_1$, we get

$$\|P_\sigma(x)\|^2 \leq \|P_\sigma(x)\|_\infty^2 \cdot |\sigma| \leq \frac{1}{\theta n} \cdot |\sigma|. \tag{8}$$

Together, (7) and (8) imply that $|\sigma| \geq \frac{1}{2}\rho^2\theta n$. ∎

For incompressible vectors we will need the following basic estimate from [20].

Proposition 4.5 ([20, Lemma 3.5]): *Let M be a random $n \times n$ matrix with column vectors X^1, $X^2, \ldots, X^n$, and let H_j $(j = 1, 2, \ldots, n)$ be the span of all column vectors except the j-th. Then for every $\varepsilon > 0$ we have*

$$\mathbb{P}\{\inf_{y \in \mathrm{Incomp}(\theta,\rho)} \|My\| < \varepsilon\rho n^{-1/2}\} \leq \frac{1}{\theta n}\sum_{j=1}^{n} \mathbb{P}\{\mathrm{dist}(X^j, H_j) < \varepsilon\}.$$

In view of independence and equi-measurability of the columns of A in our model, the above proposition yields for any $\varepsilon > 0$

$$\mathbb{P}\{\inf_{y \in \mathrm{Incomp}(\theta,\rho)} \|Ay\| < \varepsilon\rho n^{-1/2}\} \leq \frac{1}{\theta}\mathbb{P}\bigg\{\bigg|\sum_{i=1}^{n} X_i^* a_{in}\bigg| < \varepsilon\bigg\},$$

where $X^* = (X_1^*, X_2^*, \ldots, X_n^*)$ denotes a random normal unit vector to the span of the first $n-1$ columns of A. Obtaining small ball probability estimates for $|\sum_{i=1}^{n} X_i^* a_{in}|$ was a crucial ingredient of [20].

Given a real-valued random variable ξ, define its **Levy concentration function** as

$$\mathcal{L}(\xi, z) := \sup_{\lambda \in \mathbb{R}} \mathbb{P}\{|\xi - \lambda| \leq z\}, \quad z \geq 0.$$

First, let us look at some well known estimates of $\mathcal{L}(\xi, v)$ and then state a stronger bound from [20].

Theorem 4.6 (Rogozin, [16]): *Let $n \in \mathbb{N}$, let $\xi_1, \xi_2, \ldots, \xi_n$ be jointly independent random variables and let $t_1, t_2, \ldots, t_n$ be some positive real numbers. Then for any $t \geq \max_j t_j$ we have*

$$\mathcal{L}\bigg(\sum_{j=1}^{n} \xi_j, t\bigg) \leq C_{4.6}\, t \left(\sum_{j=1}^{n} (1 - \mathcal{L}(\xi_j)) {t_j}^2\right)^{-1/2},$$

where $C_{4.6} > 0$ is a universal constant.

Obviously, if ξ is essentially non-constant, there are $v > 0$ and $u \in (0,1)$ such that $\mathcal{L}(\xi, v) \leq u$. The following lemma is an elementary consequence of

Theorem 4.6 (see [11, Lemma 3.6] and [20, Lemma 2.6] for similar statements proved under additional moment assumptions on the variable).

Lemma 4.7: *Let ξ be a random variable with $\mathcal{L}(\xi,\widetilde{v}) \le \widetilde{u}$ for some $\widetilde{v} > 0$ and $\widetilde{u} \in (0,1)$. Then there are $v' > 0$ and $u' \in (0,1)$ depending only on $\widetilde{u},\widetilde{v}$ with the following property: Let $\xi_1,\xi_2,\dots,\xi_n$ be independent copies of ξ. Then for any vector $y \in S^{n-1}$ we have*

$$\mathcal{L}\Big(\sum_{j=1}^n y_j\xi_j, v'\Big) \le u'.$$

Proof. By Theorem 4.6, for any $y \in S^{n-1}$ and any $h \ge \max_j |y_j|\widetilde{v}$, we have

$$\mathcal{L}\Big(\sum_{j=1}^n y_j\xi_j, h\Big) \le \frac{C_{4.6}h}{\widetilde{v}\sqrt{1-\widetilde{u}}}.$$

Define $v' := \frac{\widetilde{v}\sqrt{1-\widetilde{u}}}{2C_{4.6}}$ and consider two cases.

(1) For every $j = 1,\dots,n$ we have $|y_j| \le \frac{\sqrt{1-\widetilde{u}}}{2C_{4.6}}$. Then $v' \ge \max_j |y_j|\widetilde{v}$, and we obtain from the above relation

$$\mathcal{L}\Big(\sum_{j=1}^n y_j\xi_j, v'\Big) \le \frac{1}{2}.$$

(2) There is j_0 such that $|y_{j_0}| > \frac{\sqrt{1-\widetilde{u}}}{2C_{4.6}}$. Then we get

$$\mathcal{L}\Big(\sum_{j=1}^n y_j\xi_j, v'\Big) \le \mathcal{L}(y_{j_0}\xi_{j_0}, v') \le \widetilde{u}.$$

Thus, we can take $u' := \max(1/2,\widetilde{u})$. ∎

Lemma 4.8 ("Tensorization lemma", [20, Lemma 2.2]): *Let $\alpha_1,\alpha_2,\dots,\alpha_n$ be i.i.d. random variables, and let $\varepsilon_0 > 0$.*

- *Assume that*

$$\mathcal{L}(\alpha_1,\varepsilon) \le L\varepsilon \quad \text{for some } L > 0 \text{ and for all } \varepsilon \ge \varepsilon_0.$$

  Then

$$\mathbb{P}\Big\{\sum_{j=1}^n {\alpha_j}^2 \le \varepsilon^2 n\Big\} \le (CL\varepsilon)^n \quad \text{for all } \varepsilon \ge \varepsilon_0,$$

  where $C > 0$ is a universal constant.

- *Assume that $\mathcal{L}(\alpha_1, v') \le u'$ for some $v' > 0$ and $u' \in (0,1)$. Then there are $v > 0$ and $u \in (0,1)$ depending only on u', v' such that*

$$\mathbb{P}\Big\{\sum_{j=1}^{n} {\alpha_j}^2 \le vn\Big\} \le u^n.$$

As a consequence of Lemmas 4.7 and 4.8, we get

LEMMA 4.9: *Let α be a random variable with $\mathcal{L}(\alpha, \widetilde{v}) \le \widetilde{u}$ for some $\widetilde{v} > 0$ and $\widetilde{u} \in (0,1)$. Then there are $v > 0$ and $u \in (0,1)$ depending only on $\widetilde{u}, \widetilde{v}$ with the following property: Let A be an $n \times n$ random matrix with i.i.d. entries equidistributed with α. Then for any $y \in S^{n-1}$ we have*

$$\mathbb{P}\{\|Ay\| \le v\sqrt{n}\} \le u^n.$$

Remark 4.10: Lemma 4.9 can be compared with [11, Proposition 3.4] and [20, Corollary 2.7]; however, those statements were proved with additional assumptions on the entries of A.

To get a stronger estimate than the one obtained in Lemma 4.7, the following notion was developed in [20] and [19] (see also preceding work [22] by Tao and Vu).

Definition 4.11 (Essential least common denominator): For parameters $r \in (0,1)$ and $h > 0$ and any non-zero vector $x \in \mathbb{R}^n$, define

$$\mathrm{LCD}_{h,r}(x) := \inf\{t > 0 : \mathrm{dist}(tx, \mathbb{Z}^n) < \min(r\|tx\|, h)\}.$$

We note that later we shall choose r sufficiently small and h to be a small multiple of $\sqrt{n}$. Thus, most of the coordinates of $\mathrm{LCD}_{h,r}(x) \cdot x$ are within a small distance to integers. For a detailed discussion of the above notion, we refer to [17].

The next statement is proved in [19].

THEOREM 4.12 ([19, Theorem 3.4]): *Let $\xi_1, \xi_2, \dots, \xi_n$ be independent copies of a centered random variable such that $\mathcal{L}(\xi_i, v) \le u$ for some $v > 0$ and $u \in (0,1)$. Further, let $x = (x_1, x_2, \dots, x_n) \in S^{n-1}$ be a fixed vector. Then for every $h > 0$, $r \in (0,1)$ and for every*

$$\varepsilon \ge \frac{1}{\mathrm{LCD}_{h,r}(x)},$$

we have

$$\mathcal{L}\Big(\sum_{i=1}^{n} x_i\xi_i, \varepsilon v\Big) \le \frac{C_{4.12}\varepsilon}{r\sqrt{1-u}} + C_{4.12}\exp(-2(1-u)h^2),$$

where $C_{4.12}$ *is a universal constant.*

Thus, in order to get a satisfactory small ball probability estimate for the infimum over incompressible vectors, it is sufficient to show that the random normal X^* has exponentially large LCD with probability close to one. This will be done in the second part of Section 5. As for the set Comp, our treatment of the random normal will be based on results of Section 3.

5. The smallest singular value—proof of Theorem B

In this section we give a proof of Theorem B stated in the introduction. Let us start with a version of Theorem A more convenient for us:

THEOREM A*: *Let* $\delta \in (0,1/4]$, $n \ge \frac{1}{4\delta}$, $\varepsilon \in (0,1/2]$, $S \subset S^{n-1}$, *and let* $\mathcal{N} \subset S$ *be a Euclidean* ε*-net on* S*. Then there exists a (deterministic) subset* $\widetilde{\mathcal{N}} \subset S$ *with* $|\widetilde{\mathcal{N}}| \le \exp(13\delta n \ln \frac{2e}{\delta})|\mathcal{N}|$ *such that for any* $n\times n$ *random matrix* A *satisfying* $(*)$, *with probability at least* $1-4\exp(-\delta n/8)$ *the set* $\widetilde{\mathcal{N}}$ *is a* $(\frac{\varepsilon C_\star}{\delta}\sqrt{n})$*-net on* S *with respect to the pseudometric* $d(x,y) := \|A(x-y)\|$ $(x,y\in S^{n-1})$*, where* $C_\star > 0$ *is a universal constant.*

Proof. Fix parameters n and δ, and let $\mathcal{C}$ be the collection of parallelepipeds from Theorem A covering B_2^n. Define a set

$$\widetilde{\mathcal{C}} := \{\varepsilon P + y : \ P \in \mathcal{C},\ y\in\mathcal{N},\ S\cap(\varepsilon P + y) \neq \emptyset\}$$

and for every $\widetilde{P}\in\mathcal{C}$ let $y_{\widetilde{P}}$ be a point in the intersection $S\cap\widetilde{P}$. Finally, set

$$\widetilde{\mathcal{N}} := \{y_{\widetilde{P}} : \ \widetilde{P}\in\widetilde{\mathcal{C}}\}.$$

Informally speaking, $\widetilde{\mathcal{C}}$ is a "product" of the rescaled collection $\varepsilon\cdot\mathcal{C}$ and the net $\mathcal{N}$. For each parallelepiped in $\widetilde{\mathcal{C}}$ having a non-empty intersection with S, we take one (arbitrary) point from this intersection to construct the refined net $\widetilde{\mathcal{N}}$. What remains is to check that with high probability $\widetilde{\mathcal{N}}$ is indeed a $(\frac{\varepsilon C}{\delta}\sqrt{n})$-net on S with respect to the pseudometric $d(x,y) := \|A(x-y)\|$.

Observe that

$$|\widetilde{\mathcal{N}}| = |\widetilde{\mathcal{C}}| \le |\mathcal{C}|\cdot|\mathcal{N}| \le \exp\Big(13\delta n \ln\frac{2e}{\delta}\Big)|\mathcal{N}|.$$

Next, let A be an $n \times n$ random matrix satisfying $(*)$, and define event $\mathcal{E}$ as

$$\mathcal{E} := \Big\{\forall\, x \in B_2^n\ \exists P \in \mathcal{C} \text{ such that } x \in P \text{ and } A(P) \subset Ax + \frac{C\sqrt{n}}{\delta} B_2^n\Big\}.$$

By Theorem A, we have $\mathbb{P}(\mathcal{E}) \geq 1 - 4\exp(-\delta n/8)$.

Fix any point $x \in S$. By the definition of $\mathcal{N}$, there is a vector $y \in \mathcal{N}$ such that $\varepsilon^{-1}(x-y) \in B_2^n$. Hence, for any point $\omega \in \mathcal{E}$ on the probability space, there is a parallelepiped $P = P(\omega) \in \mathcal{C}$ such that $\varepsilon^{-1}(x-y) \in P$ and

$$A_\omega(P) \subset A_\omega(\varepsilon^{-1}(x-y)) + \frac{C\sqrt{n}}{\delta} B_2^n.$$

Note that $S \cap (\varepsilon P + y) \supset \{x\} \neq \emptyset$, whence $\widetilde{P} := \varepsilon P + y \in \widetilde{\mathcal{C}}$, and, from the above relation,

$$A_\omega(\widetilde{P}) \subset A_\omega x + \frac{\varepsilon C\sqrt{n}}{\delta} B_2^n,$$

whence

$$A_\omega y_{\widetilde{P}} - A_\omega x \subset \frac{\varepsilon C\sqrt{n}}{\delta} B_2^n,$$

where $y_{\widetilde{P}} \in \widetilde{\mathcal{N}}$. We have shown that

$$\mathcal{E} \subset \Big\{\forall x \in S\ \exists y = y(x) \in \widetilde{\mathcal{N}} \text{ such that } \|A(x-y)\| \leq \frac{\varepsilon C\sqrt{n}}{\delta}\Big\},$$

and the result follows. ∎

Remark 5.1: Let us note that a weaker version of Theorem A*, with condition $\widetilde{\mathcal{N}} \subset S$ dropped, can be proved by applying Corollary A instead of Theorem A.

At this point, a significant part of our argument follows the same scheme as in [20]. In the first part of this section, we are dealing with compressible vectors.

Proposition 5.2 (Compressible vectors): *Let α be a centered random variable with unit variance such that $\mathcal{L}(\alpha, \widetilde{v}) \leq \widetilde{u}$ for some $\widetilde{v} > 0$ and $\widetilde{u} \in (0,1)$. Then there are numbers $\theta_{5.2}, v_{5.2} > 0$ and $u_{5.2} \in (0,1)$ depending only on $\widetilde{v}, \widetilde{u}$ with the following property: Let $n \in \mathbb{N}$ and let A be an $n \times n$ random matrix with i.i.d. entries equidistributed with α. Then for* $\mathrm{Comp} = \mathrm{Comp}_n(\theta_{5.2}, \theta_{5.2})$ *we have*

$$\mathbb{P}\{\inf_{y \in \mathrm{Comp}} \|Ay\| < v_{5.2}\sqrt{n}\} \leq 5\, {u_{5.2}}^n.$$

Proof. Without loss of generality, we can assume that n is large. First, note that by Lemma 4.9 we have a strong probability estimate for any fixed unit vector: there are $v > 0$ and $u \in (0,1)$ depending on $\widetilde{v}, \widetilde{u}$ such that for any $y \in S^{n-1}$ we get

$$\mathbb{P}\{\|Ay\| < v\sqrt{n}\} \le u^n. \tag{9}$$

In order to obtain a uniform estimate over a set $S = \mathrm{Comp}_n(\theta, \theta)$ for some small parameter θ, we will take a net $\mathcal{N} \subset S$ constructed in Lemma 4.3 and refine it with the help of Theorem A* to get a net $\widetilde{\mathcal{N}}$ with respect to pseudometric $\|A(x-y)\|$. We will apply Theorem A* with parameter δ defined as the largest number in $(0, 1/4]$ so that $\exp(13\delta n \ln \frac{2e}{\delta}) \le u^{-n/3}$. Let us describe the procedure in more detail.

First, define parameter $\theta \in (0, 1/6]$ as the largest number satisfying the inequalities

$$\Big(\frac{5e}{\theta^2}\Big)^{\theta n} \le u^{-n/3} \quad \text{and} \quad \frac{3\theta C_\star}{\delta} \le \frac{v}{2}.$$

Let S be as above. By Lemma 4.3, there is a 3θ-net $\mathcal{N} \subset S$ on S (with respect to the usual Euclidean metric) of cardinality $|\mathcal{N}| \le (\frac{5e}{\theta^2})^{\theta n}$. Now, by Theorem A*, there is a deterministic subset $\widetilde{\mathcal{N}} \subset S$ having the following properties:

- $|\widetilde{\mathcal{N}}| \le \exp(13\delta n \ln \frac{2e}{\delta}) \cdot |\mathcal{N}| \le u^{-n/3} \cdot (\frac{5e}{\theta^2})^{\theta n} \le u^{-2n/3}$;
- with probability at least $1 - 4\exp(-\delta n/8)$ for every $y \in S$ there exists $x(y) \in \widetilde{\mathcal{N}}$ such that

$$\|A(x-y)\| \le \frac{3\theta \cdot C_\star}{\delta}\sqrt{n} \le \frac{v}{2}\sqrt{n}.$$

Applying the union bound over $\widetilde{\mathcal{N}}$ to relation (9), we get

$$\mathbb{P}\{\|Ay'\| < v\sqrt{n} \text{ for some } y' \in \widetilde{\mathcal{N}}\} \le |\widetilde{\mathcal{N}}| u^n \le u^{n/3}.$$

On the other hand, the second property of $\widetilde{\mathcal{N}}$ implies that

$$\mathbb{P}\Big\{\inf_{y \in S} \|Ay\| < \inf_{y \in \widetilde{\mathcal{N}}} \|Ay\| - \frac{v\sqrt{n}}{2}\Big\} \le 4\exp(-\delta n/8).$$

Combining the two estimates, we get

$$\mathbb{P}\{\|Ay\| < v\sqrt{n}/2 \text{ for some } y \in S\} \le u^{n/3} + 4\exp(-\delta n/8),$$

and the result follows with $u_{5.2} := \max\{u^{1/3}, \exp(-\delta/8)\}$. ∎

Remark 5.3: It is not difficult to see that Proposition 5.2 can be stated and proved in the same way for A which is not square, but instead is an $n-1 \times n$ matrix with i.i.d. entries equidistributed with α. Indeed, for n large enough we can assume that $\gamma \cdot n < (n-1) < n$ for γ as close to one as we want (the values of $\theta_{5.2}$, $u_{5.2}$ and $v_{5.2}$ may differ in that case). This will be important for us later.

Remark 5.4: Proposition 5.2 could be proved by a completely different argument based on [27, Proposition 13] and not using results of Section 3 at all. However, we prefer to have a "uniform" treatment of both compressible and incompressible vectors.

Let us turn to estimating the infimum over incompressible vectors. As we already discussed in Section 4, it suffices to show that the random unit normal vector to the span of the first $n-1$ columns of A has exponentially large LCD with probability very close to one. This property is verified in Theorem 5.9 below. We start with some auxiliary statements. First, note that Theorem 4.12 together with Lemma 4.8 imply that anti-concentration probability for a single vector can be estimated in terms of the LCD of the vector. Namely, the bigger $\mathrm{LCD}(x)$ is, the less is the probability that the image Ax concentrates in a small ball:

LEMMA 5.5 (Small ball probability for a single vector; see [20, Lemma 5.5]): *Let $h > 0$, $r \in (0,1)$ and let α be a random variable satisfying $\mathcal{L}(\alpha, \widetilde{v}) \le \widetilde{u}$ for some $\widetilde{v} > 0$ and $\widetilde{u} \in (0,1)$. Then there is $L_{5.5} \ge 1$ depending only on $\widetilde{v}, \widetilde{u}$ with the following property: Let A' be an $n-1 \times n$ random matrix with i.i.d. elements equidistributed with α. Then for any vector $x \in S^{n-1}$ and any*

$$\varepsilon \ge \widetilde{v} \cdot \max\Big(\frac{1}{\mathrm{LCD}_{h,r}(x)}, \exp(-2(1-\widetilde{u})h^2)\Big)$$

we have

$$\mathbb{P}\{\|A'x\| < \varepsilon\sqrt{n}\} \le (L_{5.5}\varepsilon/r)^{n-1}.$$

Proof. Fix any vector $x \in S^{n-1}$ and denote $Y = (Y_1, Y_2, \dots, Y_{n-1}) := A'x$. Note that, in view of Theorem 4.12, we have

$$\mathcal{L}(Y_i, \varepsilon) \le \frac{C_{4.12}\varepsilon}{r\sqrt{1-\widetilde{u}}} + C_{4.12}\exp(-2(1-\widetilde{u})h^2) \le \frac{C_{4.12}(1+\widetilde{v}^{-1})\varepsilon}{r\sqrt{1-\widetilde{u}}}, \quad i \le n,$$

for any ε satisfying conditions of the lemma. Hence, by Lemma 4.8,

$$\mathbb{P}\Big\{\sum_{i=1}^{n-1} Y_i^2 \le \varepsilon^2(n-1)\Big\} \le \Big(\frac{C'(1+\widetilde{v}^{-1})\varepsilon}{r\sqrt{1-\widetilde{u}}}\Big)^{n-1}. \qquad \blacksquare$$

The above statement is useful for incompressible vectors: the following Lemma 5.6 shows that incompressible vectors have LCD at least of order $\sqrt{n}$. The lemma is taken from papers [20, 19], and its proof is included for completeness.

Lemma 5.6 (see [19, Lemma 3.6]): *For every $\theta, \rho \in (0,1)$ there are $q_{5.6} = q_{5.6}(\theta,\rho) > 0$ and $r_{5.6} = r_{5.6}(\theta,\rho) > 0$ such that for every $h > 0$ any vector $x \in \mathrm{Incomp}_n(\theta,\rho)$ satisfies*

$$\mathrm{LCD}_{h,r_{5.6}}(x) \geq q_{5.6}\sqrt{n}.$$

Proof. Set $a := \frac{1}{2}\rho^2\theta$ and $b := \rho/\sqrt{2}$. We choose $r = r_{5.6} := b\sqrt{\frac{a}{2}} = \frac{1}{2}\rho^2\sqrt{\theta}$ and $q = q_{5.6} := (1/\sqrt{\theta} + \frac{2r}{a})^{-1} = \sqrt{\theta}/3$.

Let $x \in \mathrm{Incomp}_n(\theta,\rho)$, $h > 0$ and assume that $\mathrm{LCD}_{h,r}(x) < q\sqrt{n}$. Then, by definition of the least common denominator, there exist $p \in \mathbb{Z}^n$ and $\lambda \in (0, q\sqrt{n})$ such that

$$\|\lambda x - p\| < r\lambda < rq\sqrt{n} = \frac{1}{6}\rho^2\theta\sqrt{n} = \frac{1}{3}a\sqrt{n}. \tag{10}$$

It is easy to check that for a vector with such norm the set

$$\widetilde{\sigma}(x) := \{i \leq n : |\lambda x_i - p_i| < 2/3\}$$

has a cardinality at least $(1 - \frac{a^2}{4})n$. Further, by Lemma 4.4, the set of "spread" coordinates $\sigma(x)$ has cardinality at least an. Hence, the set $I(x) := \sigma(x) \cap \widetilde{\sigma}(x)$ is non-empty, and $|I(x)| > \frac{a}{2}n$. For any $i \in I(x)$ we have

$$|p_i| < \lambda|x_i| + \frac{2}{3} < \frac{q}{\sqrt{\theta}} + \frac{2rq}{a} = 1$$

(in the last step we used our definition of q). Since $p \in \mathbb{Z}^n$, we get that $p_i = 0$ for all $i \in I(x)$.

Finally, due to the definition of $I(x)$ and our choice of r, denoting by P_J the coordinate projection on a span $\{i \in J : e_i\}$, we obtain

$$\|\lambda x - p\|^2 \geq \|\lambda P_I(x)\|^2 > \lambda^2 |I(x)| \frac{\rho^2}{2n} = \lambda^2 \frac{\rho^2 a}{4} = (r\lambda)^2,$$

which contradicts (10) and, hence, the assumption that $\mathrm{LCD}_{h,r}(x) < q\sqrt{n}$. ∎

Let $n \in \mathbb{N}$, $h > 0$, $\theta, \rho \in (0,1)$, and let $q_{5.6}$ and $r_{5.6}$ be as in the above statement. Following [20], we consider the "level sets" S_k of $\mathrm{Incomp}_n(\theta,\rho)$ defined as

$$S_k = S_k(\theta,\rho,h) := \{x \in \mathrm{Incomp}_n(\theta,\rho) \,:\, k \leq \mathrm{LCD}_{h,r_{5.6}}(x) < 2k\}, \quad k \geq 0.$$

In the proof of the theorem below we will partition $\mathrm{Incomp}_n(\theta,\rho)$ into subsets of vectors having LCD's of the same order:

$$\mathrm{Incomp}_n(\theta,\rho) = \bigsqcup_{k=2^i, i\geq i_0} S_k, \tag{11}$$

where, using Lemma 5.6, we introduce the lower bound $i_0 := \log_2(q_{5.6}\sqrt{n}/2)$ (we have $S_k = \emptyset$ for all $k < q_{5.6}\sqrt{n}/2$). Following [20], we are going to combine estimates for individual sets S_k.

A principal observation made in [19] and [20] is that the sets S_k admit Euclidean ε-nets of relatively small cardinality. We give both the formal statement and its proof from [19] below for the sake of completeness:

Lemma 5.7 ([19, Lemma 4.8]): *For any $\theta,\rho \in (0,1)$ there is $L = L(\theta,\rho) > 0$ such that for every $h \geq 1$ and $k > 0$ the set S_k admits a Euclidean $(4h/k)$-net of cardinality at most $(kL/\sqrt{n})^n$.*

Proof. In view of Lemma 5.6, we can assume that $k \geq q_{5.6}\sqrt{n}/2$. Further, without loss of generality $\frac{4h}{k} < 2$; otherwise a one-point net works.

Fix for a moment a point $x \in S_k$. Then, by definition of the "level sets", $k \leq \mathrm{LCD}_{h,r_{5.6}}(x) < 2k$. By definition of LCD, there exists $p = p(x) \in \mathbb{Z}^n$ such that

$$\|\mathrm{LCD}_{h,r_{5.6}}(x)\cdot x - p\| \leq h.$$

Hence,

$$\Big\| x - \frac{p}{\mathrm{LCD}_{h,r_{5.6}}(x)}\Big\| \leq \frac{h}{k} < \frac{1}{2}.$$

It is a simple planimetric observation that if we normalize the vector $p/\mathrm{LCD}_{h,r_{5.6}}(x)$, the distance to the unit vector x cannot increase more than twice:

$$\Big\| x - \frac{p}{\|p\|}\Big\| \leq \frac{2h}{k}.$$

Thus, the set

$$\mathcal{N}_{int} := \Big\{ \frac{p}{\|p\|} \,:\, p = p(x) \text{ for some } x \in S_k \Big\}$$

is a $2h/k$-net for S_k. How many different $p \in \mathbb{Z}^n$ do we have to consider? Note that for any $x \in S_k$, the norm of $p(x)$ cannot be too large: since $\|x\| = 1$, $\mathrm{LCD}_{h,r_{5.6}}(x) < 2k$ and $4h/k < 2$, we get

$$\|p(x)\| \leq \mathrm{LCD}_{h,r_{5.6}}(x) + h < 3k.$$

Hence, all vectors $p \in \mathbb{Z}^n$ in the definition of $\mathcal{N}_{int}$ belong to the Euclidean ball of radius $3k$ centered at the origin. A standard volumetric argument shows that there are at most $(1+Ck/\sqrt{n})^n$ integer points in this ball for a sufficiently large constant $C > 0$. Recall that $k \geq q_{5.6}\sqrt{n}/2$, whence

$$|\mathcal{N}_{int}| \leq \Big(1 + \frac{Ck}{\sqrt{n}}\Big)^n \leq \Big(\frac{kL}{\sqrt{n}}\Big)^n$$

for an appropriate number $L = L(\theta, \rho) > 0$. The net $\mathcal{N}_{int}$ does not have to be contained in S_k. But, by a standard argument, we can "replace" $\mathcal{N}_{int}$ with a $4h/k$-net of the same cardinality, and with elements from the set S_k. ∎

Together with Theorem A*, the above lemma gives

LEMMA 5.8: *For any $\theta, \rho \in (0,1)$ there is $L_{5.8} = L_{5.8}(\theta, \rho) \geq 1$ such that for every $h \geq 1$ and $k > 0$ there is a finite subset $\mathcal{N} \subset S_k$ of cardinality at most $(kL_{5.8}/\sqrt{n})^n$ with the following property. The event*

$$\{\text{For every } y \in S_k \text{ there is } y' = y'(y) \in \mathcal{N} \text{ such that } \|A(y - y')\| \leq hL_{5.8}\sqrt{n}/k\}$$

has probability at least $1 - 4\exp(-n/32)$.

Now, we can prove

THEOREM 5.9: *Let α be a centered random variable of unit variance such that $\mathcal{L}(\alpha, \widetilde{v}) \leq \widetilde{u}$ for some $\widetilde{v} > 0$ and $\widetilde{u} \in (0,1)$. Then there exist $q, s, w, r > 0$ depending only on $\widetilde{v}, \widetilde{u}$ with the following property: let $X^1, X^2 \dots, X^{n-1}$ be random n-dimensional vectors whose coordinates are jointly independent copies of α. Consider any random unit vector X^* orthogonal to $\{X^1, X^2, \dots, X^{n-1}\}$. Then*

$$\mathbb{P}\{\mathrm{LCD}_{s\sqrt{n},r}(X^*) < \exp(qn)\} \leq 2\exp(-wn).$$

Proof. Without loss of generality, we can assume that n is a large number and that $\widetilde{v} \leq 1$. Denote by A' the $n-1 \times n$ matrix with rows $X^1, X^2, \dots, X^{n-1}$. Then, by the definition of X^*, we have $A'X^* = 0$ almost surely. Let $\theta_{5.2}$ and $u_{5.2}$ be defined as in Remark 5.3 (with A' replacing A). Then, by Proposition 5.2 and Remark 5.3, we have

$$\mathbb{P}\{X^* \in \mathrm{Comp}_n(\theta_{5.2}, \theta_{5.2})\} \leq 5{u_{5.2}}^n \leq \exp(-wn)$$

for $w > 0$ such that, say, $\exp(-2w) > u_{5.2}$, and provided that n is large. Thus, it is enough to prove that

$$\mathbb{P}\{\mathrm{LCD}_{s\sqrt{n},r}(X^*) < \exp(qn), X^* \in \mathrm{Incomp}_n(\theta_{5.2}, \theta_{5.2})\} \le \exp(-wn)$$

for small enough r, w, s, q depending only on $\widetilde{v}, \widetilde{u}$. We start by defining

$$r := r_{5.6}(\theta_{5.2}, \theta_{5.2}).$$

Note that, by Lemma 5.6, we have

$$\mathrm{Incomp}_n(\theta_{5.2}, \theta_{5.2}) \subset \{x \in S^{n-1} : \mathrm{LCD}_{s\sqrt{n},r}(x) \ge q_{5.6}\sqrt{n}\}$$

for any $s > 0$ and, in particular, for s defined by

$$s := \frac{\widetilde{v}r}{4L_{5.8}^2 L_{5.5}},$$

where $L_{5.8} = L_{5.8}(\theta_{5.2}, \theta_{5.2})$ and $L_{5.5}$ are taken from Lemmas 5.8 and 5.5, respectively, and $q_{5.6} = q_{5.6}(\theta_{5.2}, \theta_{5.2})$. Let us emphasize that no vicious cycle is created here in regard to interdependence between s and r. Finally, we let $q := 2s^2(1 - \widetilde{u})$ (w will be defined at the very end of the proof).

We will make use of representation (11) of the set $\mathrm{Incomp}_n(\theta_{5.2}, \theta_{5.2})$. Denote

$$\mathcal{K} := \{2^i : i \in [\log_2(q_{5.6}\sqrt{n}) - 1, qn/\ln 2] \cap \mathbb{N}\}.$$

Then, in view of Lemma 5.6, we have

$$\{x \in \mathrm{Incomp}_n(\theta_{5.2}, \theta_{5.2}) : \mathrm{LCD}_{s\sqrt{n},r}(x) < \exp(qn)\} \subset \bigsqcup_{k \in \mathcal{K}} S_k.$$

It is sufficient to prove that

$$\mathbb{P}\{X^* \in S_k\} \le 5\exp(-n/32) \quad \text{for all } k \in \mathcal{K}. \tag{12}$$

Indeed, since $|\mathcal{K}| < qn$, the union bound over $\mathcal{K}$ will conclude the theorem.

In turn, (12) will follow as long as we show that

$$\mathbb{P}\{A'x = 0 \text{ for some } x \in S_k\} \le 5\exp(-n/32) \quad \text{for all } k \in \mathcal{K}.$$

Fix for a moment any $k \in \mathcal{K}$ and let $\mathcal{N}_k$ be the subset of S_k of cardinality at most $(kL_{5.8}/\sqrt{n})^n$, constructed in Lemma 5.8 (with $h := s\sqrt{n}$). Further, take

$$\varepsilon := \frac{\widetilde{v}r\sqrt{n}}{2kL_{5.8}L_{5.5}}.$$

Note that, in view of the definition of q and $\mathcal{K}$, we have $k \le \exp(2s^2(1-\widetilde{u})n)$. Hence, for n large enough, ε satisfies the condition of Lemma 5.5:

$$\varepsilon \ge \widetilde{v} \cdot \max\Big(\frac{1}{k}, \exp(-2s^2(1-\widetilde{u})n)\Big) \ge \widetilde{v} \cdot \max\Big(\frac{1}{\mathrm{LCD}_{h,r}(x)}, \exp(-2(1-\widetilde{u})h^2)\Big).$$

Hence,

$$\begin{aligned}\mathbb{P}\{\|A'y\| \ge \varepsilon\sqrt{n} \text{ for all } y \in \mathcal{N}_k\} &\ge 1 - |\mathcal{N}_k|(L_{5.5}\varepsilon/r)^{n-1}\\ &\ge 1 - \Big(\frac{kL_{5.8}}{\sqrt{n}}\Big)^n\Big(\frac{L_{5.5}\varepsilon}{r}\Big)^{n-1}\\ &\ge 1 - \frac{kL_{5.8}}{\sqrt{n}} \cdot \Big(\frac{\widetilde{v}}{2}\Big)^{n-1}\\ &\ge 1 - 2^{-n}\exp(2s^2(1-\widetilde{u})n),\end{aligned}$$

where the last relation follows by the assumption $\widetilde{v} \le 1$. Finally, note that, since $s \le 1/4$, the last quantity is bounded from below by $1 - 2^{-n/2}$. Applying the definition of $\mathcal{N}_k$ in Lemma 5.8 and, noticing that $hL_{5.8}\sqrt{n}/k \le \varepsilon\sqrt{n}/2$, we get

$$\begin{aligned}\mathbb{P}\{\|A'y\| \ge \varepsilon\sqrt{n}/2 \text{ for all } y \in S_k\} \ge& 1 - 4\exp(-n/32) - 2^{-n/2}\\ \ge& 1 - 5\exp(-n/32).\end{aligned}$$

This proves (12) and implies the result. ∎

Proof of Theorem B. Without loss of generality, the dimension n is large. Let $A = (a_{ij})$ be an $n \times n$ random matrix with i.i.d. centered entries with unit variance such that for some $\widetilde{v} > 0$ and $\widetilde{u} \in (0,1)$ we have $\mathcal{L}(a_{ij}, \widetilde{v}) \le \widetilde{u}$. We define $\theta := \theta_{5.2}(\widetilde{v}, \widetilde{u})$ and $v := v_{5.2}(\widetilde{v}, \widetilde{u})$, where $\theta_{5.2}, v_{5.2}$ are taken from Proposition 5.2, and let q, s, w, r be as in Theorem 5.9 (with respect to $\widetilde{v}, \widetilde{u}$). We will prove a small ball probability bound for $s_n(A)$.

It is sufficient to consider the parameter domain $\varepsilon \in (\theta\widetilde{v}\exp(-qn), 1]$. We have

$$\begin{aligned}\mathbb{P}\{s_n(A) < \varepsilon n^{-1/2}\} \le& \mathbb{P}\{\inf_{y \in \mathrm{Comp}_n(\theta,\theta)} \|Ay\| < v\sqrt{n}\}\\ &+ \mathbb{P}\{\inf_{y \in \mathrm{Incomp}_n(\theta,\theta)} \|Ay\| < \varepsilon n^{-1/2}\}\\ \le& 5{u_{5.2}}^n + \mathbb{P}\{\inf_{y \in \mathrm{Incomp}_n(\theta,\theta)} \|Ay\| < \varepsilon n^{-1/2}\},\end{aligned}$$

where we have applied Proposition 5.2. Further, by Proposition 4.5, we have

$$\mathbb{P}\{\inf_{y\in \mathrm{Incomp}_n(\theta,\theta)} \|Ay\| < \varepsilon n^{-1/2}\} \le \frac{1}{\theta}\mathbb{P}\left\{\left|\sum_{i=1}^{n} X_i^* a_{in}\right| < \frac{\varepsilon}{\theta}\right\},$$

where X^* denotes a random unit normal vector to the span of the first $n-1$ columns of A. In view of Theorem 4.12, this last relation implies

$$\begin{aligned}\mathbb{P}\{\inf_{y\in \mathrm{Incomp}_n(\theta,\theta)} \|Ay\| < \varepsilon n^{-1/2}\} \le& \theta^{-1}\mathbb{P}\{\mathrm{LCD}_{s\sqrt{n},r}(X^*) < \theta\widetilde{v}\varepsilon^{-1}\}\\ &+ \frac{C_{4.12}\varepsilon}{\theta\widetilde{v}r\sqrt{1-\widetilde{\widetilde{u}}}} + C_{4.12}\exp(-2s^2(1-\widetilde{u})n).\end{aligned}$$

Finally, noticing that $\theta\widetilde{v}\varepsilon^{-1} \le \exp(qn)$ and applying Theorem 5.9, we get

$$\begin{aligned}\mathbb{P}\{\inf_{y\in \mathrm{Incomp}_n(\theta,\theta)} \|Ay\| < \varepsilon n^{-1/2}\} \le& 2\theta^{-1}\exp(-wn)\\ &+ \frac{C_{4.12}\varepsilon}{\theta\widetilde{v}r\sqrt{1-\widetilde{\widetilde{u}}}} + C_{4.12}\exp(-2s^2(1-\widetilde{u})n).\end{aligned}$$

Together with an estimate for the compressible vectors, this implies the result. ∎

Acknowledgements. We would like to thank N. Tomczak-Jaegermann and R. Vershynin for valuable suggestions that helped improve the presentation of the work. The second named author is grateful to A. Litvak for inspiring discussions. Both authors are indebted to the referee for very useful remarks and suggestions.

References

[1] R. Adamczak, A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in Log-concave Ensembles*, Journal of the American Mathematical Society **234** (2010), 535–561.

[2] R. Adamczak, A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Sharp bounds on the rate of convergence of empirical covariance matrix*, Comptes Rendus Mathématique. Académie des Sciences. Paris **349** (2011), 195–200.

[3] Z. D. Bai and Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Annals of Probability **21** (1993), 1275–1294.

[4] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM Journal on Matrix Analysis and Applications **9** (1988), 543–560.

[5] O. Guedon, A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *On the interval of fluctuation of the singular values of random matrices*, Journal of the European Mathematical Society **19** (2017), 1469–1505.

[6] F. Götze, A. Naumov and A. Tikhomirov, *On minimal singular values of random matrices with correlated entries*, Random Matrices: Theory and Applications **4** (2015), 1550006.

[7] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963), 13–30.

[8] V. Koltchinskii and S. Mendelson, *Bounding the smallest singular value of a random matrix without concentration*, International Mathematics Research Notices **23** (2015), 12991–13008.

[9] R. Latala, *Some estimates of norms of random matrices*, Proceedings of the American Mathematical Society **133** (2005), 1273–1282.

[10] C. Le, E. Levina and R. Vershynin, *Concentration and regularization of random graphs*, Random Structures & Algorithms **51** (2017), 538–561.

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

[12] S. Mendelson and G. Paouris, *On the singular values of random matrices*, Journal of the European Mathematical Society **16** (2014), 823–834.

[13] V. D. Milman and G. Schechtman, *Asymptotic Theory of Finite-dimensional Normed Spaces*, Lecture Notes in Mathematics, Vol. 1200, Springer, Berlin, 1986.

[14] H. H. Nguyen and S. O'Rourke, *The elliptic law*, International Mathematics Research Notices **17** (2015), 7620–7689.

[15] G. Rempala and J. Wesolowski, *Asymptotics for products of independent sums with an application to Wishart determinants*, Statistics & Probability Letters **74** (2005), 129–138.

[16] B. A. Rogozin, *On the increase of dispersion of sums of independent random variables*, (Russian) Teorija Verojatnosteĭ i ee Primenenija **6** (1961), 106–108.

[17] M. Rudelson, *Lecture notes on non-aymptotic random matrix theory*, Notes from the AMS Short Course on Random Matrices, 2013.

[18] M. Rudelson and R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, in *Proceedings of the International Congress of Mathematicians. Volume III*, Hindustan Book Agency, New Delhi, 2010, pp. 1576–1602.

[19] M. Rudelson and R. Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics **62** (2009), 1707–1739.

[20] M. Rudelson and R. Vershynin, *The Littlewood–Offord problem and invertibility of random matrices*, Advances in Mathematics **218** (2008), 600–633.

[21] N. Srivastava and R. Vershynin, *Covariance estimation for distributions with* $2+\varepsilon$ *moments*, Annals of Probability **41** (2013), 3081–3111.

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

[23] T. Tao and V. Vu, *Random matrices: the circular law*, Communications in Contemporary Mathematics **10** (2008), 261–307.

[24] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values*, Geometric and Functional Analysis **20** (2010), 260–297.

[25] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, Mathematics of Computation **79** (2010), 2333–2352.

[26] H. Thorisson, *Coupling, Stationarity, and Regeneration*, Probability and its Applications (New York), Springer, New York, 2000.

[27] K. Tikhomirov, *The limit of the smallest singular value of random matrices with i.i.d. entries*, Advances in Mathematics **284** (2015), 1–20.

[28] K. E. Tikhomirov, *The smallest singular value of random rectangular matrices with no moment assumptions on entries*, Israel Journal of Mathematics **212** (2016), 289–314.

[29] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, in *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, 2012, pp. 210–268.

[30] P. Yaskov, *Lower bounds on the smallest eigenvalue of a sample covariance matrix*, Electronic Communications in Probability **19** (2014), 10pp.

[31] Y. Q. Yin, Z. D. Bai and P. R. Krishnaiah, *On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix*, Probability Theory and Related Fields **78** (1988), 509–521.