THE SMALLEST SINGULAR VALUE OF INHOMOGENEOUS SQUARE RANDOM MATRICES

BY GALYNA V. LIVSHYTS[1,*], KONSTANTIN TIKHOMIROV[1,†] AND ROMAN VERSHYNIN[2]

[1]*Department of Mathematics, Georgia Institute of Technology,* *glivshyts6@math.gatech.edu;*
†*konstantin.tikhomirov@math.gatech.edu*

[2]*Department of Mathematics, University of California, Irvine, vershyn@uci.edu*

We show that, for an $n \times n$ random matrix A with independent uniformly anticoncentrated entries such that $\mathbb{E}\|A\|_{\mathrm{HS}}^2 \leq Kn^2$, the smallest singular value $\sigma_n(A)$ of A satisfies

$$\mathbb{P}\left\{\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right\} \leq C\varepsilon + 2e^{-cn}, \quad \varepsilon \geq 0.$$

This extends earlier results (*Adv. Math.* **218** (2008) 600–633; *Israel J. Math.* **227** (2018) 507–544) by removing the assumption of mean zero and identical distribution of the entries across the matrix as well as the recent result (Livshyts (2018)) where the matrix was required to have i.i.d. rows. Our model covers inhomogeneous matrices allowing different variances of the entries as long as the sum of the second moments is of order $O(n^2)$.

In the past advances, the assumption of i.i.d. rows was required due to lack of Littlewood–Offord-type inequalities for weighted sums of non-i.i.d. random variables. Here, we overcome this problem by introducing *the Randomized Least Common Denominator* (RLCD) which allows to study anticoncentration properties of weighted sums of independent but not identically distributed variables. We construct efficient nets on the sphere with *lattice structure* and show that the lattice points typically have large RLCD. This allows us to derive strong anticoncentration properties for the distance between a fixed column of A and the linear span of the remaining columns and prove the main result.

1. Introduction. Given a random matrix A, the question of fundamental interest is, *"How likely is A to be invertible* and, more quantitatively, *well conditioned*?" These questions can be expressed in terms of the singular values $\sigma_1(A) \geq \cdots \geq \sigma_n(A) \geq 0$ which are defined as the square roots of the eigenvalues of $A^\mathsf{T}A$. The extreme singular values are especially interesting. They can be expressed as

$$(1) \qquad \sigma_1(A) = \max_{x \in \mathbb{S}^{n-1}} |Ax| \quad \text{and} \quad \sigma_n(A) = \min_{x \in \mathbb{S}^{n-1}} |Ax|,$$

where $\mathbb{S}^{n-1}$ is the unit Euclidean sphere in $\mathbb{R}^n$. In this paper we will be concerned with the smallest singular value $\sigma_n(A)$. Its value is nonzero if and only if A is invertible, and the magnitude of $\sigma_n(A)$ provides us with a quantitative measure of invertibility.

The behavior of the smallest singular values of random matrices have been extensively studied [2–4, 8, 12–15, 17–20, 25–34]. For Gaussian random matrices with i.i.d. $N(0,1)$ entries, the magnitude of $\sigma_n(A)$ is of order $1/\sqrt{n}$ with high probability. This observation goes back to von Neumann and Goldstine [35], and it was rigorously verified, with precise tail bounds, by Edelman [5] and Szarek [24]. Extending this result beyond the Gaussian distribution is nontrivial due to the absence of rotation invariance. After the initial progress

by Tao and Vu [28] and Rudelson [18], the following lower bound on $\sigma_n(A)$ was proved by Rudelson and Vershynin [19] for matrices with sub-Gaussian, mean zero, unit variance and i.i.d. entries:

$$(2) \qquad \mathbb{P}\left\{\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right\} \leq C\varepsilon + 2e^{-cn}, \quad \varepsilon \geq 0.$$

This result is optimal up to positive constants C and c (depending only on the sub-Gaussian moment). It has been further extended and sharpened in various ways [14, 17, 20, 29, 34]. In particular, Rebrova and Tikhomirov [17] relaxed the sub-Gaussian assumption on the distribution of the entries to just having unit variance.

It has remained unclear, however, if one can completely drop the assumption of the identical distribution of the entries of A. The identical distribution seemed to be crucial in the existing versions of the Littlewood–Offord theory [11] which allowed to handle arithmetic structures that arise in the invertibility problem for random matrices. A partial result was obtained recently by Livshyts [14] who proved (2) under the assumption that the rows of A are identically distributed (the entries must be still independent but not necessarily i.i.d.). In the present paper we remove the latter requirement, as well, and thus prove (2) without any identical distribution assumptions whatsoever.

We only assume the following about the entries of A: (a) they are independent; (b) the sum of their second moments is $O(n^2)$ which is weaker than assuming that each entry has unit second moment; (c) their distributions are uniformly anticoncentrated, that is, not concentrated around any single value. The latter assumption is convenient to state in terms of the Lévy concentration function which for a random variable Z is defined as

$$\mathcal{L}(Z, t) := \sup_{u \in \mathbb{R}} \mathbb{P}\{|Z - u| < t\}, \quad t \geq 0.$$

The following is our main result.

THEOREM 1.1 (Main).   *Let A be an $n \times n$ random matrix whose entries A_{ij} are independent and satisfy $\sum_{i,j=1}^{n} \mathbb{E}A_{ij}^2 \leq Kn^2$ for some $K > 0$ and $\max_{i,j} \mathcal{L}(A_{ij}, 1) \leq b$ for some $b \in (0, 1)$. Then,*

$$\mathbb{P}\left\{\sigma_n(A) \leq \frac{\varepsilon}{\sqrt{n}}\right\} \leq C\varepsilon + 2e^{-cn}, \quad \varepsilon \geq 0.$$

Here, $C, c > 0$ depend only on K and b.

We would like to emphasize that prior to this paper even the problem of *singularity* of inhomogeneous random matrices was not resolved in the literature. In particular, it was not known if, for an $n \times n$ random matrix B with independent discrete entries (say, uniformly bounded and with variances separated from zero), the singularity probability is *exponentially small* in dimension. (Theorem 1 of [14] only implied a polynomial bound on the singularity probability without the assumption of i.i.d. rows.)

The following theorem is the primary tool in proving the main result of the paper.

THEOREM 1.2 (Distances).   *For any $K > 0$ and $b \in (0, 1)$, there are $r, C, c > 0$, depending only on K and b with the following property. Let A be a random $n \times n$ matrix as in Theorem 1.1. Denote the columns of A by $A_1, \ldots, A_n$, and define*

$$H_j = \text{span}\{A_i : i \neq j, i = 1, \ldots, n\}, \quad j \leq n.$$

Take any $j \leq n$ such that $\mathbb{E}|A_j|^2 \leq rn^2$, and let v_j be a random unit vector orthogonal to H_j and measurable with respect to the sigma–field generated by H_j. Then,

$$\mathcal{L}(\langle v_j, A_j \rangle, \varepsilon) \leq C\varepsilon + 2e^{-cn}, \quad \varepsilon \geq 0.$$

In particular, for every such j we have

$$\mathbb{P}\{\mathrm{dist}(A_j, H_j) \le \varepsilon\} \le C\varepsilon + 2e^{-cn}, \quad \varepsilon \ge 0.$$

Let us outline how Theorem 1.1 can be deduced from Theorem 1.2. The first step follows the argument in [19] which is to decompose the sphere into compressible and incompressible vectors. Fix some parameters $\rho, \delta \in (0, 1)$ which, for simplicity, can be thought of as small constants. The set of compressible vectors $\mathrm{Comp}(\delta, \rho)$ consists of all vectors on the unit sphere $\mathbb{S}^{n-1}$ that are within Euclidean distance ρ to δn-sparse vectors (those that have at most δn nonzero coordinates). The remaining unit vectors are called incompressible, and we have the decomposition of the sphere,

$$\mathbb{S}^{n-1} = \mathrm{Comp}(\delta, \rho) \cup \mathrm{Incomp}(\delta, \rho).$$

By the characterization (1) of the smallest singular value, the invertibility problem reduces to finding a uniform lower bound over the sets of compressible and incompressible vectors,

$$(3) \quad \mathbb{P}\left\{\sigma_n(A) \le \frac{\varepsilon}{\sqrt{n}}\right\} \le \mathbb{P}\left\{\inf_{x \in \mathrm{Comp}(\delta, \rho)} |Ax| \le \frac{\varepsilon}{\sqrt{n}}\right\} + \mathbb{P}\left\{\inf_{x \in \mathrm{Incomp}(\delta, \rho)} |Ax| \le \frac{\varepsilon}{\sqrt{n}}\right\}.$$

For the compressible vectors, Lemma 5.3 from [14] gives the upper bound $2e^{-cn}$ on the corresponding probability in (3). For the incompressible vectors we use a version of the "invertibility via distance" bound from [19], which holds, for any $n \times n$ random matrix A (regardless of the distribution),

$$(4) \quad \mathbb{P}\left\{\inf_{x \in \mathrm{Incomp}(\delta, \rho)} |Ax| \le \frac{\varepsilon\rho}{\sqrt{n}}\right\} \le \frac{4}{\delta n} \inf_{J} \sum_{j \in J} \mathbb{P}\{\mathrm{dist}(A_j, H_j) \le \varepsilon\},$$

where the infimum is over all subsets $J \subset [n]$ of cardinality at least $n - \delta n/2$. To handle the distances, we apply Theorem 1.2. Due to our assumption $\sum_{i,j=1}^{n} \mathbb{E}A_{ij}^2 = \sum_{j=1}^{n} \mathbb{E}|A_j|^2 \le Kn^2$, all, except at most K/r terms, satisfy $\mathbb{E}|A_j|^2 \le rn^2$. Denoting the set of these terms by J and applying Theorem 1.2, we get

$$\mathbb{P}\{\mathrm{dist}(A_j, H_j) \le \varepsilon\} \le C\varepsilon + 2e^{-cn} \quad \text{for all } j \in J.$$

Since the cardinality of J is, at least, $n - K/r \ge n - \delta n/2$ for large n, we can substitute this bound into (4) and conclude that the last term in (3) is bounded by $\lesssim \varepsilon + e^{-cn}$ (recall that δ is a constant and we suppress it here). Putting all together, the probability in (3) gets bounded by $\lesssim \varepsilon + e^{-cn}$, as claimed in Theorem 1.1.

REMARK 1.3. Given Theorem 1.1, the second assertion of Theorem 1.2 can be formally strengthened as follows. Since the matrix A is shown to be singular with probability at most $2e^{-cn}$, we have that for any $j \le n$ and any random unit vector v_j orthogonal to H_j, $|\langle v_j, A_j \rangle| = \mathrm{dist}(A_j, H_j)$ with probability at least $1 - 2e^{-cn}$. Hence, the assertion of Theorem 1.2 can be replaced with

$$\mathcal{L}(\mathrm{dist}(A_j, H_j), \varepsilon) \le C\varepsilon + 2e^{-cn}, \quad \varepsilon \ge 0, \text{ whenever } \mathbb{E}|A_j|^2 \le rn^2,$$

for some $r, c, C > 0$, depending only on K, b.

An earlier version of Theorem 1.2, under the assumption that the coordinates of A_i are i.i.d., was obtained by Rudelson and Vershynin [19]. They discovered an arithmetic-combinatorial invariant of a vector (in this case, a normal vector of H_i), which they called an essential least common denominator (LCD). The authors of [19] proved a strong Littewood–Offord-type inequality for linear combinations of i.i.d. random variables in terms of the LCD

of the coefficient vector, and thus were able to estimate $\mathcal{L}(\text{dist}(A_i, H_i), \varepsilon)$. However, in the case when A_i do not have i.i.d. coordinates, the essential LCD is no longer applicable. Moreover, none of the existing Littlewood–Offord-type results could be used even to show that the distance $\text{dist}(A_i, H_i)$ is zero with an exponentially small probability (which would allow to conclude that the singularity probability for the inhomogeneous random matrix is exponentially small in dimension).

In the present paper we develop a *randomized* version of the least common denominator and show how it can handle the non-i.i.d. coordinates. Given a random vector X in $\mathbb{R}^n$ and a (deterministic) vector v in $\mathbb{R}^n$ as well as parameters $L > 0$, $u \in (0, 1)$, the Randomized Least Common Denominator of $v = (v_1, \ldots, v_n)$ (with respect to the distribution of $X = (X_1, \ldots, X_n)$) is

$$\text{RLCD}_{L,u}^X(v) = \inf\{\theta > 0 : \mathbb{E}\,\text{dist}^2\big(\theta(v_1\bar{X}_1, \ldots, v_n\bar{X}_n), \mathbb{Z}^n\big) < \min(u|\theta v|^2, L^2)\},$$

where $\bar{X}_i$ denotes a symmetrization of X_i defined as $\bar{X}_i := X_i - X_i'$, with X_i' being an independent copy of X_i, $i = 1, 2, \ldots, n$ (for the sake of comparison, let us recall that the essential least common denominator for random vectors with i.i.d. components was defined in [20] as $\text{LCD}(v) := \inf\{\theta > 0 : \text{dist}(\theta(v_1, \ldots, v_n), \mathbb{Z}^n) < \min(u|\theta v|, L)\})$. In this paper we establish a few key properties of the RLCD, in particular, its relation to anticoncentration as well as stability under perturbations of a vector. Other essential elements of the proof of Theorem 1.2 are a discretization argument based on the concept of random rounding and a double-counting procedure for estimating cardinalities of ε–nets. Those were, in a rather different form, used in [14] and [32].

In Section 2 we discuss some preliminaries and introduce our main tool, the RLCD. In Section 3 we outline the discretization procedure, based on the idea of random rounding. In Section 4 we outline the key result which, informally, states that "lattice vectors are usually nice," and is based on the idea of double counting. In Section 5 we combine the results of Sections 3 and 4 and prove Theorem 1.2. In Section 6 we conclude by formally deriving Theorem 1.1 from Theorem 1.2.

REMARK 1.4.   The main results of this paper are stated here for real random matrices and can be extended to random matrices with complex entries. This was recently done in the preprint [7] following the approach we presented in the present paper.

2. Preliminaries.   The inner product in $\mathbb{R}^n$ is denoted $\langle \cdot, \cdot \rangle$, the Euclidean norm is denoted $|\cdot|$ and the sup norm is denoted $\|x\|_\infty = \max_i |x_i|$. The Euclidean unit ball and sphere in $\mathbb{R}^n$ are denoted B_2^n and $\mathbb{S}^{n-1}$, respectively. The unit cube and the cross-polytope in $\mathbb{R}^n$ are denoted

$$B_\infty^n = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}, \qquad B_1^n = \left\{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i| \leq 1\right\}.$$

The integer part of a real number a (i.e., the largest integer which is smaller or equal to a) is denoted by $\lfloor a \rfloor$, and the fractional part by $\{a\} = a - \lfloor a \rfloor$. The cardinality of a finite set I is denoted by $\sharp I$.

Columns of an $N \times n$ matrix M will be denoted by M_j, for $j = 1, \ldots, n$, and the rows will be denoted M^i, with $i = 1, \ldots, N$.

For a random variable X we denote by $\overline{X}$ the symmetrization of X defined as $\overline{X} = X - X'$, where X' is an independent copy of X. Note that

$$(5) \qquad\qquad\qquad\qquad \mathbb{E}|\overline{X}|^2 = 2\,\text{Var}(X),$$

where we defined the variance of a random vector X as the covariance of X with itself, that is, $\text{Var}(X) = \text{Cov}(X, X) = \mathbb{E}|X - \mathbb{E}X|^2$.

2.1. *Decomposition of the sphere.* We shall follow the scheme developed by Rudelson and Vershynin in [19], the first step of which is to decompose the sphere to the set of compressible and incompressible vectors. Such decomposition in some form goes back to earlier works, in particular, that of Litvak, Pajor, Rudelson and Tomczak-Jaegermann [12], and it was used in many papers since then [17, 20, 29, 30].

Fix some parameters $\delta, \rho \in (0, 1)$ whose values will be chosen later and define the sets of sparse, compressible and incompressible vectors as follows:

$$\mathrm{Sparse}(\delta) := \{u \in \mathbb{S}^{n-1} : \#\mathrm{supp}(u) \leq \delta n\},$$

$$\mathrm{Comp}(\delta, \rho) := \{u \in \mathbb{S}^{n-1} : \mathrm{dist}(u, \mathrm{Sparse}(\delta)) \leq \rho\},$$

$$\mathrm{Incomp}(\delta, \rho) := \mathbb{S}^{n-1} \setminus \mathrm{Comp}(\delta, \rho).$$

We will use a result of [14] which gives a good uniform lower bound for $|Ax|$ on the set of compressible vectors.

LEMMA 2.1 (Lemma 5.3, [14]). *Let A be an $N \times n$ random matrix with $N \geq n$, whose entries A_{ij} are independent and satisfy $\sum_{i=1}^{N} \sum_{j=1}^{n} \mathbb{E} A_{ij}^2 \leq KNn$ for some $K > 0$ and $\max_{i,j} \mathcal{L}(A_{ij}, 1) \leq b$ for some $b \in (0, 1)$. Then,*

$$\mathbb{P}\left\{\inf_{x \in \mathrm{Comp}(\delta, \rho)} |Ax| \leq c\sqrt{N}\right\} \leq 2e^{-cN}.$$

Here, $\rho, \delta \in (0, 1)$ and $c > 0$ depend only on K and b.

The rest of our argument will be about incompressible vectors.

2.2. *Randomized least common denominator.* We will need the following lemma due to Esseen (see Esseen [6] or, e.g., Rudelson–Vershynin [19]):

LEMMA 2.2 (Esseen). *Given a variable ξ with the characteristic function $\varphi(\cdot) = \mathbb{E}\exp(2\pi \mathbf{i}\xi\cdot)$,*

$$\mathcal{L}(\xi, t) \leq C \int_{-1}^{1} \left|\varphi\left(\frac{s}{t}\right)\right| ds, \quad t > 0,$$

where $C > 0$ is an absolute constant.

Rudelson and Vershynin [19, 20] specialized Esseen's lemma for weighted sums of independent random variables $\langle X, v \rangle = \sum_{i=1}^{n} v_i X_i$.

LEMMA 2.3. *Let $X = (X_1, \ldots, X_n)$ be a random vector with independent coordinates. Then, for every vector $v \in \mathbb{R}^n$ and any $t > 0$, we have*[1]

$$\mathcal{L}(\langle X, v \rangle, t) \leq C_{2.3} \int_{-1}^{1} \exp\left(-c_{2.3}\mathbb{E}\left(\sum_{i=1}^{n}\left[1 - \cos\left(\frac{2\pi s \overline{X}_i v_i}{t}\right)\right]\right)\right) ds.$$

The constants $C_{2.3}, c_{2.3} > 0$ are absolute.

---

[1]Recall that $\overline{X}_i$ denotes the symmetrization of X_i, which we defined in the beginning of Section 2.

For completeness we outline the argument here.

PROOF.   Let φ be the characteristic function of $\langle X, v \rangle$, and φ_i be the characteristic function of X_i. By independence we have

$$\varphi(s) = \prod_{i=1}^{n} \varphi_i(sv_i), \quad s \in \mathbb{R}.$$

By definition of $\overline{X}_i$, we have, for each $i \leq n$,

$$|\varphi_i(sv_i)| = \sqrt{\mathbb{E} \cos(2\pi s v_i \overline{X}_i)} \leq \exp\left(-\frac{1}{2}(1 - \mathbb{E}\cos(2\pi s v_i \overline{X}_i))\right), \quad s \in \mathbb{R},$$

where the last step uses the inequality $|a| \leq \exp(-\frac{1}{2}(1 - a^2))$ valid for all $a \in \mathbb{R}$. To finish the proof, it remains to use Lemma 2.2.   □

In analogy with the notion of the essential least common denominator (LCD) developed by Rudelson and Vershynin [19–21], we define a randomized version of LCD which will be instrumental in controlling the sums nonidentically distributed random variables.

DEFINITION 2.4.   For a random vector X in $\mathbb{R}^n$, a (deterministic) vector v in $\mathbb{R}^n$ and parameters $L > 0$, $u \in (0, 1)$, define

$$\mathrm{RLCD}_{L,u}^X(v) := \inf\{\theta > 0 : \mathbb{E}\,\mathrm{dist}^2(\theta v \star \overline{X}, \mathbb{Z}^n) < \min(u|\theta v|^2, L^2)\}.$$

Here, by $\star$ we denote the Schur product

$$v \star X := (v_1 X_1, \ldots, v_n X_n).$$

The usefulness of RLCD is demonstrated in the following lemma which shows how RLCD controls the concentration function of a sum of independent random variables.

LEMMA 2.5.   Let $X = (X_1, \ldots, X_n)$ be a random vector with independent coordinates. Let $c_0 > 0$, $L > 0$ and $u \in (0, 1)$. Then, for any vector $v \in \mathbb{R}^n$ with $|v| \geq c_0$ and any $\varepsilon \geq 0$, we have

$$\mathcal{L}(\langle X, v \rangle, \varepsilon) \leq C\varepsilon + C\exp(-\widetilde{c}L^2) + \frac{C}{\mathrm{RLCD}_{L,u}^X(v)}.$$

Here, $C > 0$, $\widetilde{c} > 0$ may only depend on c_0, u.

PROOF.   Take any $\varepsilon \geq 1/\mathrm{RLCD}_{L,u}^X(v)$. By Lemma 2.3 we have

$$\mathcal{L}(\langle X, v \rangle, \varepsilon) \leq C_{2.3} \int_{-1}^{1} \exp\left(-c_{2.3}\mathbb{E}\left(\sum_{i=1}^{n}\left[1 - \cos\left(\frac{2\pi s \overline{X}_i v_i}{\varepsilon}\right)\right]\right)\right) ds.$$

For each $s \in [-1, 1]$ and $i \leq n$, we have

$$\mathbb{E}\left[1 - \cos\left(\frac{2\pi s \overline{X}_i v_i}{\varepsilon}\right)\right] \geq \widetilde{c}\,\mathbb{E}\,\mathrm{dist}^2(s\overline{X}_i v_i/\varepsilon, \mathbb{Z})$$

for some universal constant $\widetilde{c} > 0$. Hence,

$$\mathcal{L}(\langle X, v \rangle, \varepsilon) \leq C_{2.3} \int_{-1}^{1} \exp(-c_{2.3}\widetilde{c}\mathbb{E}\,\mathrm{dist}^2(s\overline{X} \star v/\varepsilon, \mathbb{Z}^n))\,ds$$

$$= C_{2.3}\varepsilon \int_{-1/\varepsilon}^{1/\varepsilon} \exp(-c_{2.3}\widetilde{c}\mathbb{E}\,\mathrm{dist}^2(s\overline{X} \star v, \mathbb{Z}^n))\,ds$$

$$\leq C_{2.3}\varepsilon \int_{-1/\varepsilon}^{1/\varepsilon} \exp(-c_{2.3}\widetilde{c}\min(u|sv|^2, L^2))\,ds,$$

where at the last step we used the definition of RLCD and the assumption on ε. A simple computation finishes the proof. $\square$

We shall also need the notion of the randomized LCD for matrices.

DEFINITION 2.6.   For an $m \times n$ matrix M with rows $M^1, \ldots, M^m$ and a vector $v \in \mathbb{R}^n$, define

$$\mathrm{RLCD}^M_{L,u}(v) := \min_{i=1,\ldots,m} \mathrm{RLCD}^{M^i}_{L,u}(v).$$

Recall the following "tensorization" lemma of Rudelson and Vershynin [19]:

LEMMA 2.7 (Tensorization lemma, Rudelson–Vershynin [19]).   *Suppose that $\varepsilon_0 \in (0, 1)$, $K \geq 1$, and let $Y_1, \ldots, Y_m$ be independent random variables such that each Y_i satisfies*

$$\mathbb{P}\{|Y_i| \leq \varepsilon\} \leq K\varepsilon \quad \text{for all } \varepsilon \geq \varepsilon_0.$$

Then,

$$\mathbb{P}\left\{\sum_{i=1}^m Y_i^2 \leq \varepsilon^2 m\right\} \leq (CK\varepsilon)^m, \quad \varepsilon \geq \varepsilon_0,$$

where $C > 0$ is a universal constant.

The tensorization lemma is useful when one wants to control the anticoncentration of $|Mx|$, where M is an $m \times n$ random matrix with independent rows M^i and x is a fixed vector. Indeed, in this case $|Mx|^2 = \sum_{i=1}^m \langle M^i, x \rangle^2$, and one can use Lemma 2.7 for $Y_i := \langle M^i, x \rangle$. Furthermore, one can use Lemma 2.5 to control the concentration function of each Y_i. This gives

LEMMA 2.8.   *Let M be an $m \times n$ random matrix with independent entries M_{ij}. Let $L > 0$, $c_0 > 0$ and $u \in (0, 1)$. Then, for any $x \in \mathbb{R}^n$ with $|x| \geq c_0$ and any $\varepsilon \geq C_{2.8} \exp(-\tilde{c}_{2.8} L^2) + C_{2.8}/\mathrm{RLCD}^M_{L,u}(x)$, we have*

$$\mathbb{P}\{|Mx| \leq \varepsilon\sqrt{m}\} \leq (C_{2.8}\varepsilon)^m.$$

Here, $C_{2.8}, \tilde{c}_{2.8} > 0$ may only depend on c_0 and u.

A crucial property of the RLCD, which will enable us to discretize the range of possible realizations of random unit normals, is *stability of RLCD with respect to small perturbations*.

LEMMA 2.9 (Stability of RLCD).   *Consider a random vector X in $\mathbb{R}^n$ with uncorrelated coordinates, a (deterministic) vector x in $\mathbb{R}^n$ and parameters $L, u > 0$. Fix any tolerance level $r > 0$ that satisfies*

(6) $$r^2 \mathrm{Var}(X) \leq \frac{1}{8}\min\left(u|x|^2, \frac{L^2}{D^2}\right),$$

where $D = \mathrm{RLCD}^X_{L,u}(x)$. Then, for any $y \in \mathbb{R}^n$ with $\|x - y\|_\infty < r$, we have

$$\mathrm{RLCD}^X_{2L,4u}(y) \leq \mathrm{RLCD}^X_{L,u}(x) \leq \mathrm{RLCD}^X_{L/2,u/4}(y).$$

PROOF.   Note that

$$\mathbb{E}|x \star \overline{X} - y \star \overline{X}|^2 = \mathbb{E} \sum_{i=1}^n \overline{X}_i^2 (x_i - y_i)^2 < r^2 \mathbb{E}|\overline{X}|^2 = 2r^2 \operatorname{Var}(X),$$

where the last identity is (5). Since $\operatorname{RLCD}_{L,u}^X(x) = D$, the definition of RLCD yields

$$\mathbb{E}\operatorname{dist}^2(Dx \star \overline{X}, \mathbb{Z}^n) = \min(uD^2|x|^2, L^2).$$

By the inequality $(a + b)^2 \le 2a^2 + 2b^2$, we get

$$\mathbb{E}\operatorname{dist}^2(Dy \star \overline{X}, \mathbb{Z}^n) \le 2\mathbb{E}\operatorname{dist}^2(Dx \star \overline{X}, \mathbb{Z}^n) + 2\mathbb{E}|Dx \star \overline{X} - Dy \star \overline{X}|^2$$

$$< 2\min(uD^2|x|^2, L^2) + 4D^2r^2 \operatorname{Var}(X) \le 4\min(uD^2|x|^2, L^2),$$

where the last step follows from our assumptions (6) on r. By definition of RLCD, this immediately gives

$$\operatorname{RLCD}_{2L,4u}^X(y) \le D$$

which proves the first conclusion of the lemma.

The second conclusion can be derived similarly. For any $\theta < D$, the definition of RLCD yields

$$\mathbb{E}\operatorname{dist}^2(\theta x \star \overline{X}, \mathbb{Z}^n) \ge \min(u\theta^2|x|^2, L^2).$$

By the inequality $(a + b)^2 \ge a^2/2 - b^2$, we get

$$\mathbb{E}\operatorname{dist}^2(\theta y \star \overline{X}, \mathbb{Z}^n) \ge \frac{1}{2}\mathbb{E}\operatorname{dist}^2(\theta x \star \overline{X}, \mathbb{Z}^n) - \mathbb{E}|\theta x \star \overline{X} - \theta y \star \overline{X}|^2$$

$$\ge \frac{1}{2}\min(u\theta^2|x|^2, L^2) - 2\theta^2 r^2 \operatorname{Var}(X) \ge \frac{1}{4}\min(u\theta^2|x|^2, L^2),$$

where in the last step we used the bound $\theta < D$ and our assumptions (6) on r. Thus,

$$\mathbb{E}\operatorname{dist}^2(\theta y \star \overline{X}, \mathbb{Z}^n) \ge \min(u\theta^2|x|^2/4, L^2/4) \quad \text{for all } \theta \in (0, D),$$

and, by the definition of RLCD, this immediately gives

$$\operatorname{RLCD}_{L/2,u/4}^X(y) \ge D$$

which proves the second conclusion of the lemma. $\square$

The following result is a version of [20], Lemma 3.6.

LEMMA 2.10 (Incompressible vectors have large RLCD).   *For any $b, \delta, \rho \in (0, 1)$, there are $n_0 = n_0(b, \delta, \rho), h_{2.10} = h_{2.10}(b, \delta, \rho) \in (0, 1)$ and $u_{2.10} = u_{2.10}(b, \delta, \rho) \in (0, 1/4)$ with the following property. Let $n \ge n_0$, let $x \in \operatorname{Incomp}_n(\delta, \rho)$, and assume that a random vector $X = (X_1, \ldots, X_n)$ with independent components satisfies $\mathcal{L}(X_i, 1) \le b, i \le n$, and $\operatorname{Var}|X| \le T$, for some fixed parameter $T \ge n$. Then, for any $L > 0$, we have $\operatorname{RLCD}_{L,u_{2.10}}^X(x) \ge h_{2.10} \cdot \frac{n}{\sqrt{T}}$.*

PROOF.   For clarity of the argument, we shall often hide the parameters $b, \delta, \rho, h_{2.10}$ and $u_{2.10}$ in the notation such as $\lesssim, \gtrsim$; the reader will find it easy to fill in the details.

By definition of RLCD and since x is a unit vector, it suffices to show that

$$\mathbb{E}\operatorname{dist}^2(\theta x \star \overline{X}, \mathbb{Z}^n) \gtrsim \theta^2 \quad \forall \theta \in \left(0, h_{2.10} \cdot \frac{n}{\sqrt{T}}\right).$$

Suppose that

$$\mathbb{E}\operatorname{dist}^2(\theta x \star \overline{X}, \mathbb{Z}^n) \ll \theta^2$$

for some $\theta > 0$; we want to show that in this case $\theta \gtrsim \frac{n}{\sqrt{T}}$. Let $p \in \mathbb{Z}^n$ denote a closest integer vector to $\theta x \star \overline{X}$; note that p is a random vector. Then, $\mathbb{E}\left|\theta x \star \overline{X} - p\right|^2 \ll \theta^2$, and Markov's inequality yields that $\left|\theta x \star \overline{X} - p\right| \ll \theta$ with high probability. Dividing both sides by θ gives $\left|x \star \overline{X} - p/\theta\right| \ll 1$, so another application of Markov's inequality shows that

$$\left|x_i \overline{X}_i - \frac{p_i}{\theta}\right| \ll \frac{1}{\sqrt{n}} \quad \text{for } n - o(n) \text{ coordinates } i.$$

Moreover, $\mathbb{E}\left|\overline{X}\right|^2 = 2\operatorname{Var}|X| \le 2T$ by (5). So, a similar double application of Markov's inequality shows that, with high probability,

$$\left|\overline{X}_i\right| \lesssim \sqrt{\frac{T}{n}} \quad \text{for } n - o(n) \text{ coordinates } i.$$

Furthermore, incompressible vectors are "spread" in the sense that

$$I := \left\{i : |x_i| \asymp \frac{1}{\sqrt{n}}\right\} \quad \text{satisfies } |I| \gtrsim n.$$

This fact is easy to check; a formal proof can be found in [19], Lemma 3.4.

Finally, the assumption on the concentration function shows that $\mathbb{P}\{\left|\overline{X}_i\right| \ge 1\} \ge b$. By the independence of $\overline{X}_i$'s this implies that, with high probability,

$$\left|\overline{X}_i\right| \ge 1 \quad \text{for } b\,|I|/2 \gtrsim n \text{ coordinates } i \in I$$

(this conclusion follows by considering the sum of independent indicator variables $\mathbf{1}_{\{|\overline{X}_i| \ge 1\}}$, $i \in I$).

Taking the intersection of these events and sets of coordinates, we see that, with high probability, there must exist a coordinate i for which we have simultaneously the following three bounds:

$$\left|x_i \overline{X}_i - \frac{p_i}{\theta}\right| \ll \frac{1}{\sqrt{n}}, \quad 1 \le \left|\overline{X}_i\right| \lesssim \sqrt{\frac{T}{n}}, \quad |x_i| \asymp \frac{1}{\sqrt{n}}.$$

Then, using the triangle inequality, we get

$$\left|\frac{p_i}{\theta}\right| \ge \left|x_i \overline{X}_i\right| - o\left(\frac{1}{\sqrt{n}}\right) \ge \frac{c}{\sqrt{n}} \cdot 1 - o\left(\frac{1}{\sqrt{n}}\right) > 0.$$

Thus, $p_i \ne 0$, and, since p_i is an integer, we necessarily have $|p_i| \ge 1$.

On the other hand, a similar application of the triangle inequality gives

$$\left|\frac{p_i}{\theta}\right| \le \left|x_i \overline{X}_i\right| + o\left(\frac{1}{\sqrt{n}}\right) \lesssim \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{T}{n}} + o\left(\frac{1}{\sqrt{n}}\right) \lesssim \frac{\sqrt{T}}{n}.$$

This yields that $\theta \gtrsim |p_i| \cdot \frac{n}{\sqrt{T}} \ge \frac{n}{\sqrt{T}}$, as claimed. $\quad\square$

3. Discretization.   In this section we outline the required discretization results. They essentially follow from the results in Section 3 of [14]; however, they are not stated there in the form we need, and, thus, we repeat certain arguments here.

DEFINITION 3.1 (Discretization, part 1).   Given a vector of weights $\alpha \in \mathbb{R}^n$ and a resolution parameter $\varepsilon > 0$, we consider the set of approximately unit vectors whose coordinates are quantized at scales $\alpha_i \varepsilon / \sqrt{n}$. Precisely, we define

$$\Lambda_\alpha(\varepsilon) := \left(\frac{3}{2} B_2^n \setminus \frac{1}{2} B_2^n \right) \cap \left(\frac{\alpha_1 \varepsilon}{\sqrt{n}} \mathbb{Z} \times \cdots \times \frac{\alpha_n \varepsilon}{\sqrt{n}} \mathbb{Z} \right).$$

LEMMA 3.2 (Rounding).   *Fix any accuracy* $\varepsilon \in (0, 1/2)$, *a weight vector* $\alpha \in [0, 1]^n$ *and any* (*deterministic*) $N \times n$ *matrix* A *whose columns we denote* A_i. *Then, for any* $x \in \mathbb{S}^{n-1}$, *one can find* $y \in \Lambda_\alpha(\varepsilon)$ *such that*

$$\|x - y\|_\infty \leq \frac{\varepsilon}{\sqrt{n}} \quad \text{and} \quad |A(x - y)| \leq \frac{\varepsilon}{\sqrt{n}} \left(\sum_{j=1}^{n} \alpha_j^2 |A_j|^2 \right)^{1/2}.$$

PROOF.   Our construction of y is probabilistic and amounts to *random rounding* of x. The technique of random rounding has been used in computer science (see the survey by Srinivasan [23], papers [1, 9]), asymptotic convex geometry [10] and random matrix theory [14, 30].

A random rounding of $x \in \mathbb{S}^{n-1}$ is a random vector y with independent coordinates that takes values in the $\Lambda_\alpha(\varepsilon)$ and satisfies $\mathbb{E}y = x$ and

$$(7) \qquad\qquad |x_j - y_j| \leq \frac{\alpha_j \varepsilon}{\sqrt{n}}, \quad j = 1, \ldots, n, \text{ for any realization of } y.$$

One can construct such a distribution of y by rounding each coordinate of x up or down, at random, to a neighboring point in the lattice $(\alpha_j \varepsilon / \sqrt{n})\mathbb{Z}$. The identity $\mathbb{E}y = x$ can be enforced by choosing the probabilities of rounding up and down accordingly.[2]

To check that y indeed takes values in $\Lambda_\alpha(\varepsilon)$, note that the bound in (7) and the assumption that $\alpha_i \in [0, 1]$ imply

$$(8) \qquad\qquad \|x - y\|_\infty \leq \frac{\varepsilon}{\sqrt{n}} \quad \text{for any realization of } y.$$

It follows that $\|x - y\|_2 \leq \varepsilon < 1/2$, and, since $\|x\|_2 = 1$, this implies by triangle inequality that $1/2 < \|y\|_2 < 3/2$. This verifies that the random vector y takes values in $\Lambda_\alpha(\varepsilon)$, as we claimed.

Finally, we have

$$\mathbb{E}|A(x - y)|^2 = \mathbb{E}\left| \sum_{j=1}^{n} (x_j - y_j) A_j \right|^2 = \sum_{i=1}^{n} \mathbb{E}(x_j - y_j)^2 \cdot |A_j|^2 \quad \left(\text{since } \mathbb{E}(x_j - y_j) = 0\right)$$

$$\leq \frac{\varepsilon^2}{n} \sum_{j=1}^{n} \alpha_j^2 |A_j|^2 \quad \text{(using the bound in (7)).}$$

Combining this with (8), we conclude that there exists a realization of the random vector y that satisfies the conclusion of the lemma. $\square$

---

[2]Precisely, if $x_j = (\alpha_j \varepsilon / \sqrt{n})(k_j + p_j)$ for some $k_j \in \mathbb{Z}$ and $p_j \in [0, 1)$, we let y_j take value $(\alpha_j \varepsilon / \sqrt{n})k_j$ with probability $1 - p_j$ and value $(\alpha_j \varepsilon / \sqrt{n})(k_j + 1)$ with probability p_j. Clearly, this yields $\mathbb{E}y = x$.

LEMMA 3.3. *Let $M \geq 1$. There exists a subset $\Xi \subset \mathbb{R}_+^n$ of cardinality at most $(CM)^n$ and such that the following holds. For every vector $x \in \mathbb{R}_+^n$ with $\|x\|_1 \leq Mn$, there exists $y \in \Xi$ such that $\|y\|_1 \leq (M+1)n$ and $y \geq x$ coordinatewise.*

PROOF. Define $y := \lceil x \rceil$ where the ceiling function is applied coordinatewise. Then, $\|y\|_1 \leq \|x\|_1 + n \leq (M+1)n$, as claimed. In particular, there are as many vectors y as there are integer points in the ℓ_1-ball $\{z \in \mathbb{R}^n : \|z\|_1 \leq (M+1)n\}$. According to classical results (see [16], Exercise 29; [22]), the number of integer points in this ball is bounded by $(CM)^n$ (see also [10] for a similar covering argument). The lemma is proved. $\square$

Fix $\kappa > e$, and consider the set

$$(9) \qquad \Omega_\kappa := \left\{ \alpha \in [0,1]^n : \prod_{j=1}^n \alpha_j \geq \kappa^{-n} \right\}.$$

The following result is a corollary of [14], Lemma 3.11.

LEMMA 3.4. *For any $\kappa > e$, there exists a subset $\mathcal{F} \subset \Omega_{e\kappa}$ of cardinality, at most, $(C \log \kappa)^n$ and such that the following holds. For every vector $\beta \in \Omega_\kappa$, there exists $\alpha \in \mathcal{F}$ such that $\alpha \leq \beta$ coordinatewise.*

PROOF. Apply Lemma 3.3 for $x = -\log \beta$, $y = -\log \alpha$ (defined coordinatewise) and $M = \log \kappa$. $\square$

DEFINITION 3.5 (Discretization—part 2). Assuming the dimension n fixed, for the parameters $\kappa > e$ and $\varepsilon > 0$, we shall use the notation

$$(10) \qquad \Lambda^\kappa(\varepsilon) := \bigcup_{\alpha \in \mathcal{F}} \Lambda_\alpha(\varepsilon),$$

with $\mathcal{F}$ being the set whose existence is guaranteed by Lemma 3.4.

REMARK 3.6. It is immediate from the above definition that, for any $\kappa > e$, there is $C_\kappa > 0$ depending only on κ such that $\sharp \Lambda^\kappa(\varepsilon) \leq \sum_{\alpha \in \mathcal{F}} \sharp \Lambda_\alpha(\varepsilon) \leq (C_\kappa / \varepsilon)^n$ for every $\varepsilon \in (0,1]$.

The following notion from [14] will help us to control the norms of the columns A_j of an $N \times n$ matrix A in the absence of any distributional assumptions on A_j:

$$\mathcal{B}_\kappa(A) := \min \left\{ \sum_{j=1}^n \alpha_j^2 |A_j|^2 : \alpha \in \Omega_\kappa \right\}.$$

THEOREM 3.7. *Fix $\varepsilon \in (0, 1/2)$, $\kappa > e$ and any (deterministic) $N \times n$ matrix A. Then, for every $x \in \mathbb{S}^{n-1}$, one can find $y \in \Lambda^\kappa(\varepsilon)$ so that*

$$\|x - y\|_\infty \leq \frac{\varepsilon}{\sqrt{n}} \quad \text{and} \quad |A(x - y)| \leq \frac{\varepsilon}{\sqrt{n}} \sqrt{\mathcal{B}_\kappa(A)}.$$

PROOF. By Lemma 3.2, for any $x \in \mathbb{S}^{n-1}$ we can find $y \in \Lambda^\kappa(\varepsilon)$ that approximates x in the ℓ_∞ norm, as required, and such that

$$|A(x - y)| \leq \frac{\varepsilon}{\sqrt{n}} \left(\min_{\alpha \in \mathcal{F}} \sum_{j=1}^n \alpha_j^2 |A_j|^2 \right)^{1/2} \leq \frac{\varepsilon}{\sqrt{n}} \left(\min_{\beta \in \Omega_\kappa} \sum_{j=1}^n \beta_j^2 |A_j|^2 \right)^{1/2} \qquad \text{(by Lemma 3.4)}$$

$$= \frac{\varepsilon}{\sqrt{n}} \sqrt{\mathcal{B}_\kappa(A)}.$$

The proof is complete. $\square$

Lastly, we recall the important property concerning the large deviation behavior of $\mathcal{B}_\kappa$; here, Lemma 3.11 from [14] is quoted with a specific choice of parameters.

LEMMA 3.8 (Lemma 3.11 from [14]).   *Let A be a random matrix with independent columns. Then, for any $\kappa > e$, we have*

$$\mathbb{P}\{\mathcal{B}_\kappa(A) \geq 2\mathbb{E}\|A\|_{\mathrm{HS}}^2\} \leq \left(\frac{\kappa}{\sqrt{2}}\right)^{-2n}.$$

Finally, we are ready to state the main result of this section which will follow as a corollary of Lemma 2.9, Theorem 3.7 and Lemma 3.8. Given $\gamma > 0$, $\omega \in (0, 1)$, $D > 0$ and a distribution of a random matrix M, we shall use the notation

$$S_{\omega,\gamma}^M(D) := \left\{ x \in \frac{3}{2}B_2^n \setminus \frac{1}{2}B_2^n : \mathrm{RLCD}_{\gamma\sqrt{n},\omega}^M(x) \in [D, 2D] \right\},$$

$$\tilde{S}_{\omega,\gamma}^M(D) := \left\{ x \in \frac{3}{2}B_2^n \setminus \frac{1}{2}B_2^n : \mathrm{RLCD}_{2\gamma\sqrt{n},4\omega}^M(x) \leq 2D, \mathrm{RLCD}_{0.5\gamma\sqrt{n},0.25\omega}^M(x) \geq D \right\}$$

for the level sets of the RLCD.

THEOREM 3.9 (Approximation).   *Fix any $\varepsilon \in (0, 0.1)$, $\kappa > e$, $\gamma > 0$, $\omega \in (0, 1)$, $K > 0$. Let M be an $m \times n$ random matrix with independent columns and whose rows M^i satisfy*

$$(11) \qquad \qquad \varepsilon^2 \mathrm{Var}(M^i) \leq \frac{1}{8} \min\left(\omega n, \frac{\gamma^2 n^2}{D^2}\right), \quad i = 1, \ldots, m.$$

Then, with probability, at least, $1 - (\kappa/\sqrt{2})^{-2n}$, for every $x \in \mathbb{S}^{n-1} \cap S_{\omega,\gamma}^M(D)$, there exists $y \in \Lambda^\kappa(\varepsilon) \cap \tilde{S}_{\omega,\gamma}^M(D)$ such that

$$(12) \qquad \qquad \|x - y\|_\infty \leq \frac{\varepsilon}{\sqrt{n}}, \qquad |M(x - y)| \leq \frac{\sqrt{2}\varepsilon}{\sqrt{n}}\left(\mathbb{E}\|M\|_{\mathrm{HS}}^2\right)^{1/2}.$$

PROOF.   Lemma 3.8 says that the event

$$\mathcal{E} := \{\mathcal{B}_\kappa(M) \leq 2\mathbb{E}\|M\|_{\mathrm{HS}}^2\}$$

occurs with probability, at least, $1 - (\kappa/\sqrt{2})^{-2n}$. Fix any realization of the random matrix M for which this event happens.

Let y be the approximation of x given by Theorem 3.7. Then, (12) follows from the conclusion of Theorem 3.7 and the definition of our event. The fact that $y \in \tilde{S}_{\omega,L}^M(D)$ follows from Lemma 2.9 (applied with $r = \epsilon/\sqrt{n}$) together with the assertion of Theorem 3.7 (applied with $A = M$): indeed, the assumption (11) allows us to appeal to Lemma 2.9. $\square$

4. Anticoncentration on lattice points.

The goal of this section is to study anticoncentration properties of random sums with coefficients taken from sets of the form

$$(13) \qquad \Lambda := \left(\frac{3}{2}B_2^n \cap \left\{ x \in \mathbb{R}^n : \sharp\left\{ i : |x_i| \geq \frac{\rho}{\sqrt{n}} \right\} \geq \delta n \right\}\right) \cap \left(\frac{\lambda_1}{\sqrt{n}}\mathbb{Z} \times \cdots \times \frac{\lambda_n}{\sqrt{n}}\mathbb{Z}\right).$$

The main result of this section is the following.

THEOREM 4.1 (Most lattice points are unstructured).   *For any $U \geq 1$, $b \in (0, 1)$ and $\delta, \rho \in (0, 1/2]$, there exist $n_0 = n_0(U, b, \delta, \rho)$, $\gamma = \gamma(U, b, \delta, \rho) \in (0, 1)$ and $u =$*

$u(b, \delta, \rho) \in (0, 1/4)$ *such that the following holds. Let* $n \geq n_0$. *Consider a random vector* X *in* $\mathbb{R}^n$ *with independent components* X_i *that satisfies*

$$\mathrm{Var}(X) \leq \frac{1}{8}(1 - b)\delta\gamma^2 n^2 \quad and \quad \max_i \mathcal{L}(X_i, 1) \leq b.$$

Fix numbers $\lambda_1, \ldots, \lambda_n$ *satisfying* $6^{-n} \leq \lambda_i \leq 0.01$, *and let* W *be a vector uniformly distributed on the set* Λ *defined in* (13). *Then,*

$$\mathbb{P}_W \left\{ \mathrm{RLCD}^X_{\gamma\sqrt{n}, u}(W) < \min_i 1/\lambda_i \right\} \leq U^{-n}.$$

The above theorem will be used to control the cardinality of ε-nets on the set of "typical" realizations of unit normal vectors to the spans of columns of our random matrix and forms a crucial step in the proof of Theorem 1.2. The idea of using double counting to verify structural properties of random normals was applied earlier in [32].

We start with an observation that will allow us to reduce the Euclidean ball $\frac{3}{2}B_2^n$ by a parallelotope in the definition of Λ.

LEMMA 4.2. *There is a universal constant* $C_0 > 0$ *with the following property. For any* $n \geq 1$, *there is a collection of parallelotopes* $\mathcal{P} = \{P_i\}$ *in* $\mathbb{R}^n$ *of cardinality at most* $2^{C_0 n}$, *such that*:

- *Each* P_i *is centered at the origin with the edges parallel to the coordinate axes;*
- *Each edge of* P_i *is of length at least* $2/\sqrt{n}$;
- $\frac{3}{2}B_2^n \subset \bigcup_i P_i \subset 3B_2^n$.

PROOF. First, standard volumetric estimates imply that there is a covering of $\frac{3}{2}B_2^n$ by parallel translates of the cube $\frac{1}{2\sqrt{n}}B_\infty^n$, of cardinality, at most, $2^{C_0 n}$ for a universal constant $C_0 > 0$. Let $\{x_i\}_{i \in I}$ be a collection of, at most, $2^{C_0 n}$ points in $\frac{3}{2}B_2^n$ such that each of the cubes from the covering contains at least one point x_i from the collection. Now, define $\mathcal{P} = \{P_i\}_{i \in I}$ by taking, for each $i \in I$, $P_i := \widetilde{P}_i + \frac{1}{\sqrt{n}}B_\infty^n$, where $\widetilde{P}_i$ is the unique parallelotope centered at the origin and with x_i being one of its vertices. It is elementary to check that the collection satisfies the required properties. □

LEMMA 4.3. *For any* $b \in (0, 1)$ *and* $\delta, \rho \in (0, 1/2]$, *there exists* $n_0 = n_0(b, \delta, \rho)$ *such that the following holds. Let* $n \geq n_0$ *and* $\gamma \in (0, 1)$. *Fix any subset* $J \subset [n]$, *and consider a fixed (deterministic) vector* $x \in \mathbb{R}^n$ *satisfying*

(14) $$|x|^2 \leq \frac{1}{4}(1 - b)\delta\gamma^2 n^2 \quad and \quad \sharp\{i \in J : |x_i| \geq 1\} \geq \frac{1}{2}(1 - b)\delta n.$$

Furthermore, fix numbers $\lambda_1, \ldots, \lambda_n$ *satisfying* $6^{-n} \leq \lambda_i \leq 0.01$ *and a vector* $a = (a_1, \ldots, a_n)$ *satisfying* $|a| \leq 3$ *and* $\min a_i \geq 1/\sqrt{n}$. *Consider the parallelotope* $P := \prod_{i=1}^n [-a_i, a_i]$, *and define*

$$\Lambda' := \left\{ w \in P : |w_i| \geq \frac{\rho}{\sqrt{n}} \ \forall i \in J \right\} \cap \left(\frac{\lambda_1}{\sqrt{n}}\mathbb{Z} \times \cdots \times \frac{\lambda_n}{\sqrt{n}}\mathbb{Z} \right).$$

Let W *be a random vector uniformly distributed on* Λ'. *Then, for* $D := \min_i 1/\lambda_i$, *we have*

(15) $$\mathbb{P}\left\{ \min_{\theta \in (0, D)} \mathrm{dist}(\theta W \star x, \mathbb{Z}^n)^2 < \min(c|\theta W|^2/2, 16\gamma^2 n) \right\} \leq (C\gamma)^{cn},$$

where $C, c > 0$ *depending only on* b, δ, ρ.

PROOF.   *Step 1. Halving the set I*. The assumptions on x imply that the set

$$I := \{i \in J : 1 \le |x_i| \le \gamma\sqrt{n}\} \quad \text{satisfies } \sharp I \ge \frac{1}{4}(1-b)\delta n.$$

Next, let $\mu = \mu(x)$ be a median of the set $\{a_i |x_i| : i \in I\}$. Thus, each of the subsets

$$I' := \{i \in I : a_i |x_i| \le \mu\} \quad \text{and} \quad I'' := \{i \in I : a_i |x_i| \ge \mu\}$$

contains at least a half of the elements of I,

(16) $$\min(\sharp I', \sharp I'') \ge \frac{1}{2}\sharp I \ge \frac{1}{8}(1-b)\delta n \ge cn,$$

where $c > 0$ depends only on b and δ. Take $\theta \in (0, D)$, and consider two cases.

Step 2. Ruling out small multipliers θ. We claim that the range for θ in (15) can automatically be narrowed to $(\frac{1}{2\mu}, D)$. To check this, it suffices to show that, for any $\theta \in (0, \frac{1}{2\mu}]$, the bound

(17) $$\text{dist}(\theta W \star x, \mathbb{Z}^n)^2 \ge c|\theta W|^2/2$$

holds deterministically, that is, for any realization of the random vector W.

By construction the coordinates W_i of W for $i \in I$ are uniformly distributed in lattice intervals, namely,

(18) $$W_i \sim \text{Unif}\left(\left[\frac{\rho}{\sqrt{n}}, a_i\right] \cap \frac{\lambda_i}{\sqrt{n}}\mathbb{Z}\right), \quad i \in I.$$

This means in particular that the coordinates of $\theta W \star x$ for $i \in I'$ satisfy

$$\theta |W_i x_i| \le \theta a_i |x_i| \le \theta\mu \le \frac{1}{2},$$

where we used the definition of I' and the smallness of θ. This bound in turn yields

$$\text{dist}(\theta |W_i x_i|, \mathbb{Z}) = \theta |W_i x_i| \ge \theta \cdot \frac{\rho}{\sqrt{n}} \cdot 1,$$

where in the last step we used the range of W_i from (18) and the definition of I. Square both sides of this bound, and sum over $i \in I'$ to get

$$\text{dist}(\theta W \star x, \mathbb{Z}^n)^2 \ge \frac{\theta^2 \rho^2}{n}\sharp I' \ge c\theta^2 \rho^2 \ge c_0\theta^2 |W|^2/2,$$

where we used (16), suppressed ρ into c_0 and noted that $|W|^2 \le |a|^2 \le 9$ by definition of W and assumption on a. We have proved (17).

Step 3. Handling a fixed multiplier θ. Due to the previous step, our remaining task is to show that

$$\mathbb{P}\left\{\min_{\theta \in (1/2\mu, D)} \text{dist}(\theta W \star x, \mathbb{Z}^n)^2 < 16\gamma^2 n\right\} \le (C\gamma)^{cn}.$$

To do this, let us first estimate the probability that $\text{dist}(\theta W \star x, \mathbb{Z}^n)^2 < 49\gamma^2 n$ for a *fixed* multiplier[3] $\theta \in (1/2\mu, D+1)$.

Let $i \in I''$. Recall from (18) that the random variable $|W_i|$ is uniformly distributed in a lattice interval whose diameter is, at least,

$$a_i - \frac{\rho}{\sqrt{n}} - \frac{2\lambda_i}{\sqrt{n}} \ge \frac{a_i}{3};$$

---

[3]Extending the range by *one* will be help us in the next step to unfix θ; increasing the constant factor 16 to 49 will help us run a net approximation argument in Step 4.

here, we used the assumptions $a_i \geq 1/\sqrt{n}$, $\rho \leq 1/2$ and $\lambda_i \leq 0.01$. Thus, the random variable $\theta \, |W_i x_i|$, that is, the absolute value of a coordinate of $\theta W \star x$ is distributed in a lattice interval of diameter, at least,

$$\frac{a_i}{3}\theta \, |x_i| \geq \frac{\theta \mu}{3} \geq \frac{1}{6};$$

here, we used the definition of I'' and the largeness of θ. Moreover, the step of that lattice interval (the distance between any adjacent points) is

$$\frac{\lambda_i}{\sqrt{n}}\theta \, |x_i| \leq \lambda_i \theta \gamma \leq \lambda_i (D+1)\gamma \leq 2\gamma;$$

here, we used the definition of I, the range of θ, the definition of D and the assumption that $\lambda_i \leq 0.01$.

The random variable $\theta \, |W_i x_i|$ that is uniformly distributed on a lattice interval of diameter, at least, $1/6$ and with step at most 2γ satisfies

$$\mathbb{P}\{\text{dist}(\theta \, |W_i x_i|, \mathbb{Z}) < \varepsilon\} \leq C\varepsilon \quad \text{for any } \varepsilon \geq 4\gamma,$$

where C is an absolute constant. Squaring the distances, summing them over $i \in I''$ and using tensorization Lemma 2.7, we conclude that

$$\mathbb{P}\{\text{dist}(\theta W \star x, \mathbb{Z}^n)^2 < \varepsilon^2 \sharp I''\} \leq (C'\varepsilon)^{\sharp I''} \quad \text{for any } \varepsilon \geq 4\gamma.$$

Recall from (16) that $\sharp I'' \geq cn$. Hence, substituting $\varepsilon = C_0 \gamma$ with sufficiently large C_0 (depending on c and thus, ultimately, on b and δ), we get

$$\mathbb{P}\{\text{dist}(\theta W \star x, \mathbb{Z}^n)^2 < 49\gamma^2 n\} \leq (C''\gamma)^{cn}.$$

Step 4. Unfixing the multiplier θ. It remains to make the distance bound hold simultaneously for all θ in the range $(1/2\mu, D)$. To this end, we use a union bound combined with a discretization argument. To discretize the range of θ, consider the lattice interval

$$\Theta := \left(\frac{1}{2\mu}, D\right) \cap \frac{1}{\sqrt{n}}\mathbb{Z}.$$

For sufficiently large n, its cardinality can be bounded as follows:

$$\sharp \Theta \leq (D+1)\sqrt{n} + 1 \leq (6^n + 1)\sqrt{n} + 1 \leq 7^n;$$

here, we used that $D = \min_i(1/\lambda_i)$ by definition and $\lambda_i \geq 6^{-n}$ by assumption. The construction of Θ shows that any $\theta \in (1/2\mu, D)$ can be approximated by some $\theta_0 \in \Theta$ in the sense that

$$\theta \leq \theta_0 \leq \theta + \frac{1}{\sqrt{n}}.$$

Note, in particular, that θ_0 falls in the range $(1/2\mu, D+1)$, which we handled in the previous step of the proof.

Recall that we need to bound the probability of the event

$$\mathcal{E} := \left\{\min_{\theta \in (1/2\mu, D)} \text{dist}(\theta W \star x, \mathbb{Z}^n) < 4\gamma \sqrt{n}\right\}.$$

Suppose this event occurs. Let θ be the multiplier that realizes the minimum and consider an approximation $\theta_0 \in \Theta$ as above. By triangle inequality it satisfies

$$\text{dist}(\theta_0 W \star x, \mathbb{Z}^n) < 4\gamma \sqrt{n} + |\theta_0 - \theta| \, |W \star x|.$$

By construction, we have $|\theta_0 - \theta| \leq 1/\sqrt{n}$ and

$$|W \star x| \leq \|W\|_\infty |x| \leq 3\gamma n;$$

here, we used that $\|W\|_\infty \leq \|a\|_\infty \leq |a| \leq 3$ by definition of W and assumptions on a as well as $|x| \leq \gamma n$ by assumption on x. Thus,

$$\mathrm{dist}(\theta_0 W \star x, \mathbb{Z}^n) \leq 7\gamma n.$$

For each fixed θ_0, the result of the previous step of the proof shows that the probability of this event is, at most, $(C''\gamma)^{cn}$.

As we know, the number of possible choices of θ is, at most, $\sharp\Theta \leq 7^n$. Thus, the union bound gives

$$\mathbb{P}(\mathcal{E}) \leq 7^n(C''\gamma)^{cn} \leq (C\gamma)^{cn}.$$

This completes the proof of the lemma. $\square$

REMARK 4.4. Note that with our choice of parameters, Λ' is nonempty, and, therefore, W is well defined in the lemma above.

From Lemma 4.3 we deduce the following.

LEMMA 4.5. *For any $U \geq 1$, $b \in (0, 1)$ and $\delta, \rho \in (0, 1/2]$, there exist $n_0 = n_0(U, b, \delta, \rho)$, $\gamma = \gamma(U, b, \delta, \rho) \in (0, 1)$ and $u = u(b, \delta, \rho) \in (0, 1/4)$ such that the following holds. Let $n \geq n_0$, and let J be a fixed subset of $[n]$ of cardinality at least δn. Further, consider a random vector X in $\mathbb{R}^n$ with independent components X_i that satisfies*

$$\mathbb{E}|X|^2 \leq \frac{1}{8}(1 - b)\delta\gamma^2 n^2 \quad and \quad \max_i \mathcal{L}(X_i, 1) \leq b.$$

Consider a set Λ' described in Lemma 4.3 and a random vector W uniformly distributed on Λ'. Then,

$$\mathbb{P}_W\left\{\mathrm{RLCD}^X_{\gamma\sqrt{n},u}(W) < \min_i 1/\lambda_i\right\} \leq U^{-n}.$$

PROOF. We apply a simple argument based on change of integration order or a "double-counting" trick. Without any loss of generality, we can assume that the random vector X is uniformly distributed on a finite set $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ so that, for any $x \in \mathcal{X}$, we have

$$\mathbb{P}\{X = x\} = \frac{1}{\sharp\mathcal{X}}.$$

Indeed, this follows from a simple fact that any multidimensional distribution $\zeta = (\zeta_1, \ldots, \zeta_n)$ with independent components can be approximated by a discrete distribution $\tau = (\tau_1, \ldots, \tau_n)$ of the above form so that

$$\sup_{\theta\in[0,6^n]} \sup_{v\in S^{n-1}} \left|\mathbb{E}\,\mathrm{dist}^2(\theta(v_1\bar{\zeta}_1, \ldots, v_n\bar{\zeta}_n), \mathbb{Z}^n) - \mathbb{E}\,\mathrm{dist}^2(\theta(v_1\bar{\tau}_1, \ldots, v_n\bar{\tau}_n), \mathbb{Z}^n)\right|$$

is arbitrarily small. Then, the definition of RLCD would imply that proving the required assertion for τ implies corresponding assertion for ζ, perhaps with a different choice of γ, u, n_0.

Set $\mathcal{X}' := \{x \in \mathcal{X} : x \text{ satisfies } (14)\}$. In view of our assumptions on X (and assuming that n is sufficiently large), we have

$$\mathbb{P}\{X \in \mathcal{X}'\} \geq 1/4,$$

while, in view of the assertion of Lemma 4.3 and summing over $x \in \mathcal{X}'$, we get

(19)
$$\sharp\Big\{(x, w) \in \mathcal{X}' \times \Lambda' : \min_{\theta \in (0, D)} \text{dist}(\theta w \star x, \mathbb{Z}^n)^2 \geq \min(c|\theta w|^2/2, 16\gamma^2 n)\Big\}$$
$$\geq (1 - (C\gamma)^{cn})\sharp\mathcal{X}'\sharp\Lambda',$$

where $D = \min_i 1/\lambda_i$. This implies

$$\sharp\Big\{w \in \Lambda' : \sharp\Big\{x \in \mathcal{X}' : \min_{\theta \in (0, D)} \text{dist}(\theta w \star x, \mathbb{Z}^n)^2 \geq \min(c|\theta w|^2/2, 16\gamma^2 n)\Big\} \geq \sharp\mathcal{X}'/4\Big\}$$
$$\geq (1 - 2(C\gamma)^{cn})\sharp\Lambda'$$

(indeed, if the last assertion were not true, we would get that the cardinality of the set in (19) was bounded above by $(1 - 2(C\gamma)^{cn})\sharp\Lambda' \cdot \sharp\mathcal{X}' + 2(C\gamma)^{cn}\sharp\Lambda' \cdot \sharp\mathcal{X}'/4 \leq (1 - 3(C\gamma)^{cn}/2)\sharp\mathcal{X}'\sharp\Lambda')$. Back from counting to probabilities, we get from the last bound and the estimate $\sharp\mathcal{X}'/4 \geq \sharp X/16$,

$$\sharp\Big\{w \in \Lambda' : \min_{\theta \in (0, D)} \mathbb{E}_X \text{dist}(\theta w \star X, \mathbb{Z}^n)^2 \geq \min(c|\theta w|^2/32, \gamma^2 n)\Big\} \geq (1 - 2(C\gamma)^{cn})\sharp\Lambda'.$$

This can be equivalently rewritten with $u := c/32$ as

$$\sharp\big\{w \in \Lambda' : \text{RLCD}^X_{\gamma\sqrt{n}, u}(w) > D\big\} \geq (1 - 2(C\gamma)^{cn})\sharp\Lambda',$$

and the result follows by taking any $\gamma \in (0, 1)$ satisfying $2(C\gamma)^{cn} \leq U^{-n}$. $\square$

PROOF OF THEOREM 4.1.   Without loss of generality, $\mathbb{E}X = 0$, so that $\text{Var}(X) = \mathbb{E}|X|^2$. We obtain the results as a combination of Lemmas 4.2 and 4.5. To do so, note that Λ can be covered by $2^{C_1 n}$ sets of the type Λ' (one for each paralellotope and a support set J). Then, the probability measures on Λ and a given Λ' are within $2^{C_1 n}$ from each other. Thus, the probability in the conclusion of Theorem 4.1 is bounded by $2^{C_1 n} U^{-n} \leq (cU)^{-n}$. It remains to redefine $U \to cU$ to get the result. $\square$

5. Proof of Theorem 1.2.   In this section we split the Euclidean unit sphere S^{n-1} into *level sets* collecting (incompressible) unit vectors having comparable RLCD. To show that with a high probability the normal vector does not belong to a level set with a small RLCD, we consider a discrete approximating set whose cardinality is well controlled from above by using a combination of Theorem 3.9 and Theorem 4.1. In view of the stability property of RLCD, the event that the normal vector has a small RLCD is contained within the event that one of the vectors in the approximating set has a small RLCD. We then apply the small ball probability estimates for individual vectors, combined with the union bound, to show that the latter event has probability close to zero.

For any $D \geq 1$, $\gamma, u \in (0, 1)$ and an $m \times n$ random matrix M define, as before,

$$S_D(M, \gamma, u) := \{v \in \mathbb{S}^{n-1} : \text{RLCD}^M_{\gamma\sqrt{n}, u} \in [D, 2D]\}.$$

As the first step we combine the approximation Theorem 3.9 with Theorem 4.1 to obtain

PROPOSITION 5.1.   *For arbitrary* $b, \rho, \delta \in (0, 1)$, $U \geq 1$ *and* $K \geq 1$ *there exist* $n_{5.1} = n_{5.1}(b, \delta, \rho, U, K)$, $u_{5.1} = u_{5.1}(b, \delta, \rho) \in (0, u_{2.10}(b, \delta, \rho))$, $\gamma_{5.1} = \gamma_{5.1}(b, \delta, \rho, U, K) \in (0, 1/2)$ *with the following property. Let* $D \geq 1$ *and* $0 < \varepsilon \leq 1/D$. *Let* $n \geq n_{5.1}$, $m \geq 1$, *and let* M *be an* $m \times n$ *matrix with independent entries* M_{ij} *such that* $\mathcal{L}(M_{ij}, 1) \leq b$, *for all* i, j,

$$\text{Var}(M^\top e_i) \leq \frac{1}{8}\min\big((1 - b)\delta\gamma_{5.1}^2 n^2, \varepsilon^{-2}u_{5.1}n\big)$$

for every $i \leq m$ and

$$\mathbb{E}\|M\|_{\mathrm{HS}}^2 \leq Kn^2.$$

Then, there is a nonrandom set $\Lambda \subset \mathbb{R}^n$ of cardinality at most $(\varepsilon U)^{-n}$ having the following properties:

- *For any $y \in \Lambda$, we have $3/2 \geq |y| \geq 1/2$;*
- *For any $y \in \Lambda$, $\mathrm{RLCD}^M_{\gamma_{5.1}\sqrt{n}/2, u_{5.1}/4}(y) \geq D$ and $\mathrm{RLCD}^M_{2\gamma_{5.1}\sqrt{n}, 4u_{5.1}}(y) \leq 2D$;*
- *With probability at least $1 - e^{-n}$, for any $x \in S_D(M, \gamma_{5.1}, u_{5.1}) \cap \mathrm{Incomp}(\delta, \rho)$, there is $y \in \Lambda$ with $\|x - y\|_\infty \leq \varepsilon/\sqrt{n}$ and $|M(x - y)| \leq \varepsilon\sqrt{n}$.*

PROOF.   Set $\kappa := 5$, and let $C_\kappa > 0$ be the constant from Remark 3.6. Let $U \geq 1$, $U' := 100\sqrt{2K} U C_\kappa / \rho$, and set

$$n_{5.1} := n_0(U', b, \delta, \rho/2), \qquad \gamma = \gamma_{5.1} := \gamma(U', b, \delta, \rho/2),$$

$$u = u_{5.1} := u(b, \delta, \rho/2) \in \left(0, \frac{1}{4}\right),$$

where the functions $n_0(\cdot)$, $\gamma(\cdot)$, $u(\cdot)$ are taken from Theorem 4.1. Finally, set

$$\varepsilon' := \frac{\rho\varepsilon}{100\sqrt{2}\max(K, 1)} \in (0, 0.01),$$

and let $\Lambda^\kappa(\varepsilon')$ be as in Definition 3.5.

Let Λ be a subset of all vectors $y \in \Lambda^\kappa(\varepsilon')$ such that

$$\mathrm{RLCD}^M_{\gamma\sqrt{n}/2, u/4}(y) \geq D \quad \text{and} \quad \mathrm{RLCD}^M_{2\gamma\sqrt{n}, 4u}(y) \leq 2D,$$

and such that the ℓ_∞–distance of y to $\mathrm{Incomp}(\delta, \rho)$ is, at most, $\varepsilon'/\sqrt{n}$. Note that the last condition implies that for any $y \in \Lambda$, $\sharp\{i \leq n : |y_i| \geq \rho/(2\sqrt{n})\} \geq \delta n$; see the argument in Lemma 3.4 from [19].

By our choice of ε' and the condition on the matrix, we have

$$(\varepsilon')^2 \mathrm{Var}(M^\top e_i) \leq \frac{1}{8}\frac{\gamma^2 n^2}{D^2}; \qquad (\varepsilon')^2 \mathrm{Var}(M^\top e_i) \leq \frac{1}{8}un.$$

Then, according to Theorem 3.9, with probability at least $1 - (5/\sqrt{2})^{-2n}$ for any incompressible vector $x \in S_D(M, \gamma, u)$ there is a vector $y \in \Lambda$ such that $\|x - y\|_\infty \leq \varepsilon'/\sqrt{n}$ and $|M(x - y)| \leq \sqrt{2}\varepsilon'\sqrt{K}\sqrt{n} \leq \varepsilon\sqrt{n}$.

It remains to estimate the cardinality of Λ. We recall that

$$\Lambda^\kappa(\varepsilon') = \bigcup_{\alpha \in \mathcal{F}} \Lambda_\alpha(\varepsilon'),$$

where the collection $\mathcal{F}$ of parameters $(\alpha_1, \ldots, \alpha_n) \in (0, 1]^n$ is given by Lemma 3.4. Fix for a moment any $(\alpha_1, \ldots, \alpha_n) \in \mathcal{F}$, and set $\lambda_i := \alpha_i \varepsilon' \in (0, 0.01]$, $i \leq n$. Observe that $1/\lambda_i \geq 1/\varepsilon' > 2/\varepsilon \geq 2D$, $i \leq n$. Hence, we can apply Theorem 4.1 to obtain

$$\sharp(\Lambda \cap \Lambda_\alpha(\varepsilon')) \leq \sharp\Lambda_\alpha(\varepsilon')(U')^{-n}.$$

Taking the union over all $(\alpha_1, \ldots, \alpha_n) \in \mathcal{F}$, we then get

$$\sharp\Lambda \leq (U')^{-n} \sum_{\alpha \in \mathcal{F}} \sharp\Lambda_\alpha(\varepsilon') \leq (\varepsilon U)^{-n},$$

where at the last step we used our definition of U'. $\square$

Next, we combine the discrete approximation set introduced above with the small ball probability of Lemma 2.8.

PROPOSITION 5.2.   *For any $b, \rho, \delta \in (0, 1)$ and $K \geq 1$, there are $n_{5.2} = n_{5.2}(b, \delta, \rho, K)$, $u_{5.2} = u_{5.2}(b, \delta, \rho) \in (0, u_{2.10}(b, \delta, \rho))$, $\gamma_{5.2} = \gamma_{5.2}(b, \delta, \rho, K) \in (0, 1/2)$ and $\gamma'_{5.2} = \gamma'_{5.2}(b, \delta, \rho, K)$ with the following property. Let $n \geq n_{5.2}$, $e^2 \leq D \leq D_0 \leq e^{\gamma'_{5.2}n}$, $0 \leq k \leq n/\ln D_0$, $m := n - k$, and let M be an $m \times n$ random matrix with independent entries M_{ij} such that $\mathcal{L}(M_{ij}, 1) \leq b$, for all i, j,*

$$\text{(20)} \qquad \text{Var}(M^i) \leq \frac{1}{64} \min((1 - b)\delta\gamma_{5.2}^2 n^2, D_0^2 u_{5.2} n)$$

for every $i \leq m$, and

$$\mathbb{E}\|M\|_{\text{HS}}^2 \leq Kn^2.$$

Let $M^{(1)}$ be the matrix obtained from M by removing the first row. Then,

$$\mathbb{P}\{\exists x \in \text{Incomp}(\delta, \rho) \cap S_D(M, \gamma_{5.2}, u_{5.2}) \text{ s.t. } \text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{M^{(1)}}(x) \geq D_0, M^{(1)}x = 0\} \leq 2e^{-n}.$$

PROOF.   First, we should carefully define the parameters. We choose $u := u_{5.1}(b, \delta, \rho)$. Next, set $U := 2e^3 C_{2.8}^2$, where $C_{2.8}$ is taken from Lemma 2.8 with parameters $c_0 := 1/2$ and $u/4$, and we assume without loss of generality that $C_{2.8} \geq 1$. Finally, take $\gamma := \gamma_{5.1}(b, \delta, \rho, U, K)$, $\gamma' := \tilde{c}_{2.8}\gamma^2/4 \leq 1$.

Let $e^2 \leq D \leq D_0 \leq e^{\gamma'n}$, and let random matrix M satisfy the assumptions of the proposition. Let Λ be the set defined in Proposition 5.1 with $\varepsilon := 1/D_0$. Set

$$\mathcal{E}_D := \{\exists x \in \text{Incomp}(\delta, \rho) \cap S_D(M, \gamma, u) \text{ s.t. } \text{RLCD}_{\gamma\sqrt{n}, u}^{M^{(1)}}(x) \geq D_0, M^{(1)}x = 0\}.$$

Note that whenever x and y are two vectors in $\mathbb{R}^n$ with $\text{RLCD}_{\gamma\sqrt{n}, u}^{M^{(1)}}(x) \geq D_0$ and $\|x - y\|_\infty \leq \frac{1}{D_0\sqrt{n}}$, then necessarily $\text{RLCD}_{\gamma\sqrt{n}/2, u/4}^{M^{(1)}}(y) \geq D_0$ (as follows from Lemma 2.9).

Hence, applying Proposition 5.1, we get

$$\mathbb{P}(\mathcal{E}_D) \leq e^{-n} + \mathbb{P}\{\text{There is } y \in \Lambda \text{ with } |M^{(1)}y| \leq \sqrt{n}/D_0 \text{ and } \text{RLCD}_{\gamma\sqrt{n}/2, u/4}^{M^{(1)}}(y) \geq D_0\}$$

$$\leq e^{-n} + \sharp\Lambda \sup_y \mathbb{P}\{|M^{(1)}y| \leq \sqrt{n}/D_0\}$$

$$\leq e^{-n} + (D_0/U)^n \sup_y \mathbb{P}\{|M^{(1)}y| \leq \sqrt{n}/D_0\},$$

where the supremum is taken over all vectors $y \in \frac{3}{2}B_2^n \setminus \frac{1}{2}B_2^n$ with $\text{RLCD}_{\gamma\sqrt{n}/2, u/4}^{M^{(1)}}(y) \geq D_0$.

Fix any y satisfying the above conditions. Set $\tilde{\varepsilon} := 2C_{2.8}/D_0$, and observe that, by our conditions on D_0,

$$\tilde{\varepsilon} \geq C_{2.8} \exp(-\tilde{c}_{2.8}\gamma^2 n/4) + C_{2.8}/\text{RLCD}_{\gamma\sqrt{n}/2, u/4}^{M^{(1)}}(y).$$

Applying Lemma 2.8, we then obtain

$$\mathbb{P}\{|M^{(1)}y| \leq \sqrt{n}/D_0\} \leq \mathbb{P}\{|M^{(1)}y| \leq 2\sqrt{m-1}/D_0\}$$

$$\leq \mathbb{P}\{|M^{(1)}y| \leq \tilde{\varepsilon}\sqrt{m-1}\} \leq (C_{2.8}\tilde{\varepsilon})^{m-1}.$$

Taking the supremum over all admissible y, we then get

$$\mathbb{P}(\mathcal{E}_D) \leq e^{-n} + (D_0/U)^n(C_{2.8}\tilde{\varepsilon})^{m-1} \leq e^{-n} + D_0^{n-m+1}U^{-n}(2C_{2.8}^2)^n.$$

The result follows by the choice of U and the condition on m. $\square$

Our proof of Theorem 1.2, in the case $\text{Var}(A_j) = \Theta(n)$, $j = 1, 2, \ldots, n$, is a straightforward application of Proposition 5.2 (taking a dyadic sequence of level sets), together with results of [14] on invertibility over compressible vectors. The fact that in our model some columns may have variances much greater than n adds some complexity to the proof because the relation (20) for such columns may hold true only for "large enough" D_0 leaving a gap in the treatment of small values of the parameter. We deal with this issue in the statement below by carefully splitting the event in question into subevents and invoking Lemma 2.10 that allows to deterministically bound RLCD in terms of the variance.

PROPOSITION 5.3.   *Let $b, \delta, \rho \in (0, 1)$ and $K \geq 1$ be parameters, and let $u_{5.2}$, $\gamma_{5.2}$ be taken from Proposition 5.2. Then, there are $n_{5.3}(b, \delta, \rho, K)$ and $\gamma'_{5.3}(b, \delta, \rho, K)$ with the following property. Let $n \geq n_{5.3}$, let $n \times n$ matrix A be as in the statement of Theorem 1.2 and let $j \leq n$ be such that*

$$\text{Var}(A_j) \leq \min\left(h_{2.10}^2 e^{-4} n^2, \frac{1}{64}(1-b)\delta\gamma_{5.2}^2 n^2\right),$$

where $h_{2.10}$ is taken from Lemma 2.10. Then,

$$\mathbb{P}\left\{\exists x \in \text{Incomp}(\delta, \rho) \text{ orth. to } A_i, i \neq j, \text{ with } \text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{A_j}(x) \leq e^{\gamma'_{5.3}n}\right\} \leq 2^{-n/2}.$$

PROOF.   We will assume that n is large and that $\gamma' > 0$ is a small parameter whose value can be recovered from the proof below. Without loss of generality, $j = 1$. Let A' be the submatrix of A composed of all columns A_i satisfying

$$\text{Var}(A_i) \leq \min\left(h_{2.10}^2 e^{-4} n^2, \frac{1}{64}(1-b)\delta\gamma_{5.2}^2 n^2\right).$$

We note that the number of columns of A' is, at least, $n - K / \min(h_{2.10}^2 e^{-4}, \frac{1}{64}(1-b)\delta\gamma_{5.2}^2)$. Further, let M be the transpose of A', and denote by W the submatrix of $M^{(1)}$ formed by removing rows with variances, at least, $n^{9/8}$.

The proof of the statement is reduced to estimating probability of the event

$$\mathcal{E}' := \left\{\exists x \in \text{Incomp}(\delta, \rho) \text{ with } M^{(1)}x = 0 \text{ and } \text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{A_1}(x) \leq e^{\gamma'n}\right\}.$$

We can write

$$\mathbb{P}(\mathcal{E}') \leq \sum_{\log_2 n - 1 \leq \ell \leq \gamma'n\log_2 e} \mathbb{P}\left\{\exists x \in \text{Incomp}(\delta, \rho) \cap S_{2^\ell}(M, \gamma_{5.2}, u_{5.2}) \text{ with } M^{(1)}x = 0\right\}$$

$$+ \mathbb{P}\left\{\exists x \in \text{Incomp}(\delta, \rho) \text{ with } M^{(1)}x = 0 \text{ and } \text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{M}(x) < n\right\}.$$

The first sum can be estimated directly by applying Proposition 5.2 with $D_0 := D := 2^\ell$, $\log_2 n - 1 \leq \ell \leq \gamma'n\log_2 e$ (note that the relation (20) is fulfilled for such D for all rows of M and that the proposition can be applied as long as $K / \min(h_{2.10}^2 e^{-4}, \frac{1}{64}(1-b)\delta\gamma_{5.2}^2) \leq 1/\gamma'$). Further, the condition that $\text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{M}(x) < n$ implies that either $\text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{W}(x) < n$ or $\text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{W}(x) \geq n$ and $\text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{M^q}(x) < n$ for some row M^q of M. Hence, we get

$$\mathbb{P}(\mathcal{E}') \leq 2n \cdot 2e^{-n} + \sum_q \mathbb{P}\left\{\exists x \in \text{Incomp}(\delta, \rho) \text{ with } Wx = 0 \text{ and } \text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{W}(x) \geq n\right.$$

$$\left. \text{and } \text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{M^q}(x) < n\right\}$$

$$+ \mathbb{P}\left\{\exists x \in \text{Incomp}(\delta, \rho) \text{ with } Wx = 0 \text{ and } \text{RLCD}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}^{W}(x) < n\right\}.$$

To estimate the sum, we apply Lemma 2.10 which, together with our restrictions on the variances, allows to deterministically bound the RLCD with respect to M^q by e^2. Thus, we get

$$\mathbb{P}\{\exists x \in \text{Incomp}(\delta, \rho) \text{ with } Wx = 0 \text{ and } \text{RLCD}^W_{\gamma_{5.2}\sqrt{n}, u_{5.2}}(x) \geq n$$

$$\text{and } \text{RLCD}^{M^q}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}(x) < n\}$$

$$= \mathbb{P}\{\exists x \in \text{Incomp}(\delta, \rho) \text{ with } Wx = 0 \text{ and } \text{RLCD}^W_{\gamma_{5.2}\sqrt{n}, u_{5.2}}(x) \geq n$$

$$\text{and } e^2 \leq \text{RLCD}^{M^q}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}(x) < n\}.$$

Splitting the interval $[e^2, n]$ into dyadic subintervals and applying Proposition 5.2 with $D_0 := n$ and for the matrix formed by concatenating W and M^q, we get an upper bound $2e^{-n}\log_2 n$ for the probability.

In order to estimate probability of the event

$$\{\exists x \in \text{Incomp}(\delta, \rho) \text{ with } Wx = 0 \text{ and } \text{RLCD}^W_{\gamma_{5.2}\sqrt{n}, u_{5.2}}(x) < n\},$$

we apply Lemma 2.10; this time the definition of W implies that RLCD with respect to each row is deterministically bounded from below by $n^{3/8}$ for a sufficiently large n. Again, splitting of the interval $[n^{3/8}, n]$ into dyadic subintervals reduces the question to estimating events of the form

$$\{\exists x \in \text{Incomp}(\delta, \rho) \cap S_D(W, \gamma_{5.2}, u_{5.2}) \text{ with } Wx = 0\}$$

for some $D \in [n^{3/8}, n]$. Taking $D_0 := D$, one can see that the condition (20) is fulfilled for all rows of W and that the difference between the number of columns and rows of W is clearly less than $n/\ln D_0$. Thus, Proposition 5.2 is applicable.

Summarizing, we get $\mathbb{P}(\mathcal{E}') \leq C'ne^{-n}\ln n$ for a universal constant $C' > 0$. The result follows for all sufficiently large n. $\square$

Now, we are in position to prove Theorem 1.2.

PROOF OF THEOREM 1.2. We will assume that n is large. We start by recording a property of A which follows immediately from Lemma 2.1 (i.e., [14], Lemma 5.3). For any $j \leq n$, with probability, at least, $1 - e^{-c_1 n}$ any unit vector orthogonal to $\{A_i, i \neq j\}$ is (δ, ρ)-incompressible for some $\delta, \rho \in (0, 1)$ depending only on b, K (here, $c_1 \in (0, 1)$ depends only on b, K). Indeed, let $j \leq n$, let B be the $n \times (n-1)$ matrix formed from A by removing A_j and define $M := B^\top$. Then,

$$\mathbb{P}\{\exists x \in \text{Comp}(\delta, \rho) \text{ orthogonal to } H_j\} \leq \mathbb{P}\Big\{\inf_{x \in \text{Comp}(\delta, \rho)} |Mx| = 0\Big\} \leq e^{-c_1 n},$$

where in the last passage Lemma 2.1 (i.e., [14], Lemma 5.3) was used.

Set

$$r := \min\Big(h_{2.10}^2 e^{-4}, \frac{1}{64}(1-b)\delta\gamma_{5.2}^2\Big),$$

where $h_{2.10}$ and $\gamma_{5.2}$ are defined in respective lemmas with the parameters b, K, δ, ρ. Pick any index $j \leq n$ such that $\text{Var}(A_j) \leq rn^2$, and let v be a random unit vector orthogonal to H_j and measurable with respect to the sigma-field generated by H_j. Applying Proposition 5.3 together with the above observation, we get

$$v \text{ is } (\delta, \rho)\text{-incompressible and } \text{RLCD}^{A_j}_{\gamma_{5.2}\sqrt{n}, u_{5.2}}(v) \geq e^{\gamma'_{5.3}n}$$

with probability, at least, $1 - e^{c_1 n} - 2^{-n/2}$. Application of Lemma 2.5 finishes the proof. $\square$

REMARK 5.4.   In our proof, the randomized least common denominator acts like a mediator in the relationship between anticoncentration properties of matrix-vector products and cardinalities of corresponding discretizations (nets), following the ideas developed in [19]. A crucial element of our argument is the fact that RLCD is stable with respect to small perturbations of the vector, which we quantify in Lemma 2.9.

An alternative approach recently considered in [32] is based on directly estimating the concentration function for "typical" points on a multidimensional lattice. The argument of [32] uses, as an important step, certain stability properties of the Lévy concentration function and of small ball probability estimates for linear combinations of Bernoulli random variables. However, in the general (non-Bernoulli) setting and with different distributions of entries of the matrix, obtaining satisfactory stability properties similar to those in [32] seems to be a very nontrivial problem, in the situation when the approximation is done by a random vector. We note here that in our net construction the approximating vector is, indeed, random and depends on the realization of the matrix.

On a technical level, since RLCD is a structural (geometric) property, its stability follows from relatively simple computations, while the Lévy concentration function is much more difficult to control; in particular, the Esseen lemma provides only an upper bound for the concentration function, hence cannot be relied on when studying its stability.

6. Proof of the Theorem 1.1.   In this section we formally derive Theorem 1.1 from Theorem 1.2, using a modification of the "invertibility via distance" lemma from [19].

LEMMA 6.1 (Invertibility via distance).   *Let A be any $n \times n$ random matrix. Fix a pair of parameters $\delta, \rho \in (0, \frac{1}{2})$, and assume that $n \geq 4/\delta$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\left\{\inf_{x \in \mathrm{Incomp}(\delta, \rho)} |Ax| \leq \varepsilon \frac{\rho}{\sqrt{n}}\right\} \leq \frac{4}{\delta n} \inf_{\substack{I \subset [n], \\ \sharp I = n - \lfloor \delta n/2 \rfloor}} \sum_{j \in I} \mathbb{P}\{\mathrm{dist}(A_j, H_j) \leq \varepsilon\},$$

where H_j denotes the subspace spanned by all the columns of A, except for A_j.

PROOF.   Fix any $I \subset [n]$ with $\sharp I = n - \lfloor \delta n/2 \rfloor$, and consider event

$$\mathcal{E} := \left\{\inf_{x \in \mathrm{Incomp}(\delta, \rho)} |Ax| \leq \varepsilon \frac{\rho}{\sqrt{n}}\right\}.$$

Fix any realization of the matrix A such that the event holds, that is, there exists a vector $x \in \mathrm{Incomp}(\delta, \rho)$ with $|Ax| \leq \varepsilon \frac{\rho}{\sqrt{n}}$. In view of the definition of the set $\mathrm{Incomp}(\delta, \rho)$, there is a subset $J_x \subset [n]$ of cardinality $\lfloor \delta n \rfloor$ such that $|x_i| \geq \rho/\sqrt{n}$ for all $i \in J_x$, whence

$$\mathrm{dist}(A_i, H_i) \leq |x_i|^{-1} |Ax| \leq \varepsilon, \quad i \in J_x.$$

Note that $J_x \cap I$ has cardinality, at least, $\lfloor \delta n \rfloor - \lfloor \delta n/2 \rfloor \geq \delta n/4$. Thus,

$$\mathcal{E} \subset \{\sharp\{i \in I : \mathrm{dist}(A_i, H_i) \leq \varepsilon\} \geq \delta n/4\}.$$

It remains to note that

$$\mathbb{P}\{\sharp\{i \in I : \mathrm{dist}(A_i, H_i) \leq \varepsilon\} \geq \delta n/4\} \leq \frac{4}{\delta n} \mathbb{E}\sharp\{i \in I : \mathrm{dist}(A_i, H_i) \leq \varepsilon\}. \qquad \square$$

PROOF OF THEOREM 1.1.   The theorem follows from Lemma 2.1 (i.e., Lemma 5.3 from [14]), Lemma 6.1 and Theorem 1.2, by taking $I_0 := \{i \in [n] : \mathbb{E}|A_i|^2 \leq rn^2\}$ and noting that, in view of the assumption $\mathbb{E}\|A\|_{\mathrm{HS}}^2 \leq Kn^2$, we have $\sharp I_0 = n - K/r \geq n - \lfloor \delta n/2 \rfloor$ for all sufficiently large n so that, for all large enough n,

$$\mathbb{P}\left\{\inf_{x \in \mathrm{Incomp}(\delta, \rho)} |Ax| \leq \varepsilon \frac{\rho}{\sqrt{n}}\right\} \leq \frac{4}{\delta n} \sum_{j \in I_0} \mathbb{P}\{\mathrm{dist}(A_j, H_j) \leq \varepsilon\}. \qquad \square$$

REFERENCES

[1] ALON, N. and KLARTAG, B. (2017). Optimal compression of approximate inner products and dimension reduction. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 639–650. IEEE Comput. Soc., Los Alamitos, CA. MR3734268 https://doi.org/10.1109/FOCS.2017.65

[2] BAI, Z. D. and YIN, Y. Q. (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* **16** 1729–1741. MR0958213

[3] BOURGAIN, J., VU, V. H. and WOOD, P. M. (2010). On the singularity probability of discrete random matrices. *J. Funct. Anal.* **258** 559–603. MR2557947 https://doi.org/10.1016/j.jfa.2009.04.016

[4] COOK, N. (2018). Lower bounds for the smallest singular value of structured random matrices. *Ann. Probab.* **46** 3442–3500. MR3857860 https://doi.org/10.1214/17-AOP1251

[5] EDELMAN, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** 543–560. MR0964668 https://doi.org/10.1137/0609045

[6] ESSEEN, C. G. (1966). On the Kolmogorov–Rogozin inequality for the concentration function. *Z. Wahrsch. Verw. Gebiete* **5** 210–216. MR0205297 https://doi.org/10.1007/BF00533057

[7] JAIN, V. and SILWAL, S. (2020). A note on the universality of ESDs of inhomogeneous random matrices. Preprint. Available at arXiv:2006.05418.

[8] KAHN, J., KOMLÓS, J. and SZEMERÉDI, E. (1995). On the probability that a random ±1-matrix is singular. *J. Amer. Math. Soc.* **8** 223–240. MR1260107 https://doi.org/10.2307/2152887

[9] KANNAN, R. and VEMPALA, S. (1999). Sampling lattice points. In *Proc.* 29*th ACM Symposium on the Theory of Computing* (*STOC*'97) (*El Paso*, *TX*) 696–700. ACM, New York. MR1753411

[10] KLARTAG, B. and LIVSHYTS, G. V. (2020). The lower bound for Koldobsky's slicing inequality via random rounding. In *Geometric Aspects of Functional Analysis* 43–63. Springer, Berlin.

[11] LITTLEWOOD, J. E. and OFFORD, A. C. (1943). On the number of real roots of a random algebraic equation. III. *Math. Sci.* **12** 277–286. MR0009656

[12] LITVAK, A. E., PAJOR, A., RUDELSON, M. and TOMCZAK-JAEGERMANN, N. (2005). Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* **195** 491–523. MR2146352 https://doi.org/10.1016/j.aim.2004.08.004

[13] LITVAK, A. E. and RIVASPLATA, O. (2012). Smallest singular value of sparse random matrices. *Studia Math.* **212** 195–218. MR3009072 https://doi.org/10.4064/sm212-3-1

[14] LIVSHYTS, G. V. (2018). The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding. Preprint.

[15] MENDELSON, S. and PAOURIS, G. (2014). On the singular values of random matrices. *J. Eur. Math. Soc.* (*JEMS*) **16** 823–834. MR3191978 https://doi.org/10.4171/JEMS/448

[16] PÓLYA, G. and SZEGÖ, G. (1964). *Aufgaben und Lehrsätze aus der Analysis. Band I: Reihen. Integralrechnung. Funktionentheorie. Dritte Berichtigte Auflage. Die Grundlehren der Mathematischen Wissenschaften* **19**. Springer, Berlin. MR0170985

[17] REBROVA, E. and TIKHOMIROV, K. (2018). Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries. *Israel J. Math.* **227** 507–544. MR3846333 https://doi.org/10.1007/s11856-018-1732-y

[18] RUDELSON, M. (2008). Invertibility of random matrices: Norm of the inverse. *Ann. of Math.* (2) **168** 575–600. MR2434885 https://doi.org/10.4007/annals.2008.168.575

[19] RUDELSON, M. and VERSHYNIN, R. (2008). The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600–633. MR2407948 https://doi.org/10.1016/j.aim.2008.01.010

[20] RUDELSON, M. and VERSHYNIN, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707–1739. MR2569075 https://doi.org/10.1002/cpa.20294

[21] RUDELSON, M. and VERSHYNIN, R. (2015). Delocalization of eigenvectors of random matrices with independent entries. *Duke Math. J.* **164** 2507–2538. MR3405592 https://doi.org/10.1215/00127094-3129809

[22] SCHÜTT, C. (1984). Entropy numbers of diagonal operators between symmetric Banach spaces. *J. Approx. Theory* **40** 121–128. MR0732693 https://doi.org/10.1016/0021-9045(84)90021-2

[23] SRINIVASAN, A. (1999). Approximation algorithms via randomized rounding: A survey. In *Lectures on Approximation and Randomized Algorithms. Series in Advanced Topics in Mathematics* 9–71. Polish Scientific Publishers PWN, Warsaw.

[24] SZAREK, S. J. (1991). Condition numbers of random matrices. *J. Complexity* **7** 131–149. MR1108773 https://doi.org/10.1016/0885-064X(91)90002-F

[25] TAO, T. and VU, V. (2006). On random ±1 matrices: Singularity and determinant. *Random Structures Algorithms* **28** 1–23. MR2187480 https://doi.org/10.1002/rsa.20109

[26] TAO, T. and VU, V. (2007). On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.* **20** 603–628. MR2291914 https://doi.org/10.1090/S0894-0347-07-00555-3

[27] TAO, T. and VU, V. (2010). Random matrices: The distribution of the smallest singular values. *Geom. Funct. Anal.* **20** 260–297. MR2647142 https://doi.org/10.1007/s00039-010-0057-8

[28] TAO, T. and VU, V. H. (2009). Inverse Littlewood–Offord theorems and the condition number of random discrete matrices. *Ann. of Math.* (2) **169** 595–632. MR2480613 https://doi.org/10.4007/annals.2009.169.595

[29] TATARKO, K. (2018). An upper bound on the smallest singular value of a square random matrix. *J. Complexity* **48** 119–128. MR3828841 https://doi.org/10.1016/j.jco.2018.06.002

[30] TIKHOMIROV, K. (2015). The limit of the smallest singular value of random matrices with i.i.d. entries. *Adv. Math.* **284** 1–20. MR3391069 https://doi.org/10.1016/j.aim.2015.07.020

[31] TIKHOMIROV, K. (2020). Invertibility via distance for noncentered random matrices with continuous distributions. *Random Structures Algorithms* **57** 526–562. MR4129729 https://doi.org/10.1002/rsa.20920

[32] TIKHOMIROV, K. (2020). Singularity of random Bernoulli matrices. *Ann. of Math.* (2) **191** 593–634. MR4076632 https://doi.org/10.4007/annals.2020.191.2.6

[33] TIKHOMIROV, K. E. (2016). The smallest singular value of random rectangular matrices with no moment assumptions on entries. *Israel J. Math.* **212** 289–314. MR3504328 https://doi.org/10.1007/s11856-016-1287-8

[34] VERSHYNIN, R. (2011). Spectral norm of products of random and deterministic matrices. *Probab. Theory Related Fields* **150** 471–509. MR2824864 https://doi.org/10.1007/s00440-010-0281-z

[35] VON NEUMANN, J. and GOLDSTINE, H. H. (1947). Numerical inverting of matrices of high order. *Bull. Amer. Math. Soc.* **53** 1021–1099. MR0024235 https://doi.org/10.1090/S0002-9904-1947-08909-6