The Annals of Probability
2024, Vol. 52, No. 5, 1992–2018
https://doi.org/10.1214/24-AOP1695

A LARGE DEVIATION INEQUALITY FOR THE RANK OF A RANDOM MATRIX

BY MARK RUDELSON[a]

Department of Mathematics, University of Michigan, [a]*rudelson@umich.edu*

Let A be an $n \times n$ random matrix with independent identically distributed nonconstant sub-Gaussian entries. Then for any $k \le c\sqrt{n}$,

$$\operatorname{rank}(A) \ge n - k$$

with probability at least $1 - \exp(-c'kn)$.

1. Introduction. Estimating the probability that an $n \times n$ random matrix with independent identically distributed (i.i.d.) entries is singular is a classical problem in probability. The first result in this direction showing that, for a matrix with Bernoulli$(1/2)$ entries, this probability is $O(n^{-1/2})$ was proved by Komlós [7] in 1967. In a breakthrough paper [6], Kahn, Komlós, and Szemerédi established the first exponential bound for Bernoulli matrices,

$$\mathbb{P}(\det(A_n) = 0) = (0.998 + o(1))^n.$$

The asymptotically optimal exponent has been recently obtained by Tikhomirov [17],

$$\mathbb{P}(\det(A_n) = 0) = \left(\frac{1}{2} + o(1)\right)^n.$$

The exponential bound for probability of singularity holds in a more general context than Bernoulli random matrices. It was proved in [14] for matrices with i.i.d. sub-Gaussian entries and extended in [11] to matrices whose entries have bounded second moment.

A natural extension of the question about the probability of singularity is estimating the probability that a random matrix has a large co-rank. More precisely, we are interested in the asymptotic of $\mathbb{P}(\operatorname{rank}(A_n) \le n - k)$, where $k < n$ is a number which can grow with n as $n \to \infty$. Such rank means that there are k columns of A_n which are linearly dependent on the other columns. Based on the fact that

$$\mathbb{P}(\operatorname{rank}(A_n) \le n - 1) = \mathbb{P}(A_n \text{ is singular}) \le \exp(-cn)$$

and the independence of the columns of A_n, one can predict that the probability that the rank of A_n does not exceed $n - k$ and is bounded by $(\exp(-cn))^k = \exp(-cnk)$. Proving such a bound amounts to obtaining a super-exponential probability estimate if $k \to \infty$ as $n \to \infty$. This makes a number of key tools in the previously mentioned papers unavailable, because these tools were intended to rule out pathological events of probability $O(\exp(-cn))$ which cannot be considered negligible in this context.

The existing results fell short of this tight bound until recently. Kahn, Komlos, and Szemeredi showed that the probability that a Bernoulli$(1/2)$ matrix has rank smaller than $n - k$ is $O(f(k)^n)$ where $f(k) \to 0$ as $k \to \infty$. The intuitive prediction above was recently confirmed by Jain, Sah and Sawney in the case when $k \in \mathbb{N}$ is a fixed number. Building on the ideas of Tikhomirov [17], they proved an optimal bound for random matrices with independent Bernoulli(p) entries. Namely, for any $p \in (0, 1/2]$, $\varepsilon > 0$, and for any $n > n_0(k, p, \varepsilon)$,

$$\mathbb{P}(\operatorname{rank}(A_n) \le n - k) \le (1 - p + \varepsilon)^{kn}.$$

Received April 2023; revised February 2024.
MSC2020 subject classifications. 60B20.
Key words and phrases. Random matrices, rank.

This completely solves the problem for Bernoulli matrices within the exponential range. However, the methods of this paper do not seem to be extendable to the case when k grows together with n, that is, to the super-exponential range of probabilities (see Section 2.2 for more details).

The main result of this paper confirms this prediction in the super-exponential range for all matrices with i.i.d. sub-Gaussian entries. A random variable ξ is called sub-Gaussian if

$$\mathbb{E}\exp\left(-\left(\frac{\xi}{K}\right)^2\right) < \infty$$

for some $K > 0$. In what follows, we regard K as a constant and allow other constants such as C, c, c', etc. depend on it. This is a rich class of random variables including, for instance, all bounded ones.

We prove the following theorem.

THEOREM 1.1. *Let $k, n \in \mathbb{N}$ be numbers such that $k \leq cn^{1/2}$. Let A be an $n \times n$ matrix with i.i.d. nonconstant sub-Gaussian entries. Then*

$$\mathbb{P}(\operatorname{rank}(A) \leq n - k) \leq \exp(-c'kn).$$

REMARK 1.2. *The bound of Theorem 1.1 should hold for $k > cn^{1/2}$ as well. The restriction on k in the theorem arises from using ν-almost orthogonal systems throughout its proof; see Definition 3.1 and Remark 3.4.*

REMARK 1.3. *Combining the technique of this paper with that of Nguyen* [10], *one can also obtain a lower bound for the singular value $s_{n-k}(A_n)$ of the same type as in* [10] *but with the additive error term* $\exp(-ckn)$ *instead of* $\exp(-cn)$. *We will not pursue this route in order to keep the presentation relatively simple.*

The importance of getting the large deviation bound of Theorem 1.1 in the regime when k grows simultaneously with n stems in particular from its application to *quantitative group testing* (QGT). This computer science problem considers a collection of n items containing k defective ones, where $k < n$ is regarded as a known number. A test consists of selecting a random pool of items choosing each one independently with probability $1/2$ and outputting the number of defective items in the pool. The aim of the QGT is to efficiently determine the defective items after a small number of tests. The question of constructing an efficient algorithm for QGT is still open. In [3], Feige and Lellouche introduced the following relaxation of the QGT: after $m > k$ tests, one has to produce a subset $S \subset [n]$ of cardinality m, containing all defective items. This means that, unlike the original QGT, the approach of Feige and Lellouche allows false positives, which makes the problem simpler and admits more efficient algorithms. Denote by A the $m \times n$ matrix whose rows are the indicator functions of the tests, and denote by $A|_S$ its submatrix with columns from the set $S \subset [n]$. Then A is a random matrix with i.i.d. Bernoulli entries. The main result of [3] asserts that if an algorithm for the relaxed problem succeeds and outputs a set $S \subset [n]$ and $\operatorname{rank}(A|_S) \geq m - O(\log n)$, then one can efficiently determine the set of defective items. Checking this criterion for a given algorithm is difficult since the set S is not known in advance. However, if we know that

$$\operatorname{rank}(A|_S) \geq m - O(\log n) \tag{1.1}$$

for all m-element sets $S \subset [n]$ at the same time, this condition would be redundant, and all algorithms for the relaxed problem could be adapted to solve the QGT. In other words, we need to estimate the minimal rank of all $m \times m$ submatrices of an $m \times n$ random matrix. We show below that Theorem 1.1 implies that the bound (1.1) holds with high probability and, moreover, that this is an optimal estimate (see Lemma 6.2).

2. Notation and the outline of the proof.

2.1. *Notation.* We denote by $[n]$ the set of natural numbers from 1 to n. Given a vector $x \in \mathbb{R}^n$, we denote by $\|x\|_2$ its standard Euclidean norm: $\|x\|_2 = (\sum_{j\in[n]} x_j^2)^{1/2}$. The unit sphere of $\mathbb{R}^n$ is denoted by S^{n-1}.

If V is an $m \times l$ matrix, we denote by $\text{Row}_i(V)$ its ith row and by $\text{Col}_j(V)$ its jth column. Its singular values will be denoted by

$$s_1(V) \geq s_2(V) \geq \cdots \geq s_m(V) \geq 0.$$

The operator norm of V is defined as

$$\|V\| = \max_{x \in S^{l-1}} \|Vx\|_2,$$

and the Hilbert–Schmidt norm as

$$\|V\|_{\text{HS}} = \left(\sum_{i=1}^{m} \sum_{j=1}^{l} v_{i,j}^2 \right)^{1/2}.$$

Note that $\|V\| = s_1(V)$ and $\|V\|_{\text{HS}} = (\sum_{j=1}^m s_j(V)^2)^{1/2}$.

Throughout the paper the letters $c, \bar{c}, C$ etc. stand for absolute constants whose values may change from line to line.

2.2. *Outline of the proof.* Let A be an $n \times n$ random matrix with i.i.d. entries. The fact that this matrix has rank at most $n - k$ means that at least k of its columns are linearly dependent on the rest. Assume that the k last columns are linearly dependent on the other. As the results of [14] show, for a typical realization of the first $n - k$ columns, the probability that a given column belongs to their linear span is $O(\exp(-cn))$. Since the last k columns are mutually independent and at the same time independent of the first $n-k$ ones, the probability that all k columns fall into the linear span of the rest is $O((\exp(-cn))^k) = O(\exp(-cnk))$, which is the content of our main theorem.

The problem with this argument, however, is in the meaning of the term "typical." It includes several requirements on the matrix with these $n - k$ rows, including that its norm is $O(\sqrt{n})$ and that its kernel contains no vector with a rigid arithmetic structure. As was shown in [14], all these requirements hold with probability at least $1 - \exp(-cn)$, which is enough to derive that the matrix is invertible with a similar probability. In our case, when we aim at bounding probability by $\exp(-ckn)$ with k which can tend to infinity with n, the events which have just exponentially small probability cannot be considered negligible any longer. In particular, we are not able to assume that the operator norm of a random matrix is bounded by $O(\sqrt{n})$. This is, however, the easiest of the arising problems, as we will be able to use a better concentrated Hilbert–Schmidt norm instead.

The problem of ruling out the arithmetic structure of the kernel turns out to be more delicate. For Bernoulli(p) random matrices with $0 < p \leq \frac{1}{2}$, Jain, Sah, and Sawney [5] overcame it by replacing the approach based on the least common denominator used in [14] with a further development of the averaging method of Tikhomirov [17]. This allowed them to prove that if k is a constant, then with probability $1 - 4^{-kn}$, the kernel of the matrix consisting of the first $n - k$ rows either consists of vectors close to sparse (compressible) or does not contain *any* vector with a problematic arithmetic structure; see [5], Proposition 2.7, whose proof follows [4], Proposition 3.7. They further derived from this fact that the probability that a random Bernoulli matrix has rank $n - k$ or smaller does not exceed

$$(1 - p + o(1))^{kn}$$

for any constant k. However, this approach is no longer feasible if k is growing at the same time with n. Indeed, the kernel of an $(n-k)\times n$ Bernoulli random matrix contains the vector $(1,\dots,1)$ with probability $(c/\sqrt{n})^n = \exp(-c'n\log n)$. It can also contain numerous other vectors of the same type with a similar probability. Hence, the kernel of such matrix contains incompressible vectors with rigid arithmetic structure for $k=\Omega(\log n)$, which includes the range important for the question of Feige and Lellouche.

Fortunately, the complete absence of vectors with a rigid arithmetic structure in the kernel is not necessary for proving a bound on the probability of a low rank. It is sufficient to rule out the situation where such vectors occupy a significant part of the kernel. More precisely, we show that if B is an $(n-k)\times n$ random matrix with i.i.d. sub-Gaussian entries, then with probability at least $1-\exp(-ckn)$, its kernel contains a $(k/2)$-dimensional subspace free of the vectors with a rigid arithmetic structure. Checking this fact is the main technical step in proving our main theorem.

We outline the argument leading to it below. We try to follow the geometric method developed in [12, 14]. However, the aim of obtaining a super-exponential probability bound forces us to work with systems of problematic vectors instead of single ones. To handle such systems, we introduce a notion of an *almost orthogonal* l-tuple of vectors in Section 3. These systems are sufficiently simple to allow efficient estimates. At the same time, we show in Lemma 3.3 that a linear subspace containing many "bad" vectors contains an almost orthogonal system of such vectors possessing an important minimality property.

Following the general scheme, we split the unit sphere of $\mathbb{R}^n$ into compressible and incompressible parts. Let us introduce the respective definitions.

DEFINITION 2.1. *Let $s\in[n]$, and let $\tau>0$. Define the set of s-sparse vectors by*

$$\operatorname{Sparse}(s)=\{x\in\mathbb{R}^n : |\operatorname{supp}(x)|\le s\}$$

and the sets of compressible and incompressible vectors by

$$\operatorname{Comp}(s,\tau)=\{x\in S^{n-1} : \operatorname{dist}(x,\operatorname{Sparse}(s))\le\tau\},$$
$$\operatorname{Incomp}(s,\tau)=S^{n-1}\setminus\operatorname{Comp}(s,\tau).$$

Note that we define the sparse vectors in $\mathbb{R}^n$ and not in S^{n-1}. This is not important but allows to shorten some future calculations.

In Section 4 we show that the probability that the kernel of the matrix $B=(A_{[n-k]\times[n]})^\top$ contains an almost orthogonal system of $k/4$ compressible vectors is negligible. This is done by using a net argument, that is, by approximating vectors from our system by vectors from a certain net. The net will be a part of a scaled integer lattice, and the approximation will be performed by *random rounding*, a technique widely used in computer science and introduced in random matrix theory by Livshyts [8]. Let B be a random matrix. The general net argument relies on obtaining a uniform lower bound for $\|By\|_2$ over all points y in the net and approximating a given point x by the points of the net. In this case, one can use the triangle inequality to obtain

$$\|Bx\|_2\ge\|By\|_2-\|B\|\cdot\|x-y\|_2.$$

This approach runs into problems in the absence of a good control of $\|B\|$. However, if the net is constructed a part of a scaled integer lattice, then one can choose the approximating point y as a random vertex of the cubic cell containing x. This essentially allows to replace $\|B\|$ in the approximation above by a more stable quantity $\|B\|_{\mathrm{HS}}/\sqrt{n}$. Moreover, this replacement will be possible for a randomly chosen y with probability close to 1.

In our case we have to approximate the entire system of vectors while preserving the almost orthogonality property. This makes the situation more delicate, and we can only prove that this approximation succeeds with probability which is exponentially small in k. Fortunately, this is enough since we need just one approximation, so any positive probability is sufficient.

In Section 5, we assume that the kernel of B contains a subspace of dimension $(3/4)k$ consisting of incompressible vectors and prove that, with high probability, this subspace contains a further one of dimension $k/2$, which has no vectors with a rigid arithmetic structure. The arithmetic structure is measured in terms of the *least common denominator* (LCD), which is defined in Section 3.3. To this end we consider a minimal almost orthogonal system of $k/4$ vectors having subexponential LCDs and show that the presence of such system in the kernel is unlikely using the net argument and random rounding. This is more involved than the case of compressible vectors since the magnitude of the LCD varies from $O(\sqrt{n})$ to the exponential level and thus requires approximation on different scales. To implement it, we decompose the set of such systems according to the magnitudes of the LCDs, and then we scale each system by the sequence of its LCDs. Because of the multiplicity of scales, the approximation has to satisfy a number of conditions at once. At this step we also rely on random rounding allowing to check all the required conditions probabilistically. Verification that all of them can be satisfied simultaneously, although with an exponentially small probability performed in the proof in Lemma 5.3 is the most technical part of the argument.

Finally, in Section 6 we collect all the ingredients and finish the proof of Theorem 1.1.

3. Preliminary results.

3.1. *Almost orthogonal systems of vectors*. We will have to control the arithmetic structure of the subspace spanned by $n-k$ columns of the matrix A throughout the proof. This structure is defined by the presence of vectors which are close to the integer lattice. To be able to estimate the probability that many such vectors lie in the subspace, we will consider special configurations of almost orthogonal vectors, which are easier to analyze. This leads us to the following definition.

DEFINITION 3.1. *Let $\nu \in (0,1)$. An l-tuple of vectors $(v_1,\dots,v_l) \subset \mathbb{R}^n \setminus \{0\}$ is called ν-almost orthogonal if the $n \times l$ matrix W with columns $(\frac{v_1}{\|v_1\|_2},\dots,\frac{v_l}{\|v_l\|_2})$ satisfies*

$$1-\nu \le s_l(W) \le s_1(W) \le 1+\nu.$$

Estimating the largest and especially the smallest singular values of a general deterministic matrix is a delicate task. We employ a very crude criterion below.

LEMMA 3.2. *Let $\nu \in [0,\frac{1}{4}]$, and let $(v_1,\dots,v_l) \subset \mathbb{R}^n \setminus \{0\}$ be a an l-tuple such that*

$$\|P_{\mathrm{span}(v_1,\dots,v_j)} v_{j+1}\|_2 \le \frac{\nu}{\sqrt{l}} \|v_{j+1}\|_2 \quad \text{for all } j \in [l-1].$$

Then $(v_1,\dots,v_l) \subset \mathbb{R}^l$ is a (2ν)-almost orthogonal system. Moreover, if V is the $n \times l$ matrix with columns $v_1,\dots,v_l$, then

$$\det{}^{1/2}(V^\top V) \ge 2^{-l} \prod_{j=1}^{l} \|v_j\|_2.$$

PROOF. Construct an orthonormal system in $\mathbb{R}^n$ by setting

$$e_1 = \frac{v_1}{\|v_1\|_2}, \qquad e_{j+1} = \frac{P_{(\mathrm{span}(v_1,\dots,v_j))^\perp} v_{j+1}}{\|P_{(\mathrm{span}(v_1,\dots,v_j))^\perp} v_{j+1}\|_2} \quad \text{for all } j \in [l-1],$$

and complete it to an orthonormal basis. The $n \times l$ matrix W with columns $\mathrm{Col}_j(W) = \frac{v_j}{\|v_j\|_2}$ written in this basis has the form $W = [\begin{smallmatrix}\bar{W}\\ 0\end{smallmatrix}]$, where $\bar{W}$ is an $l \times l$ upper triangular matrix. The assumption of the lemma yields

$$\left(\sum_{i=1}^{j-1} \bar{W}_{i,j}^2\right)^{1/2} = \left\| P_{\mathrm{span}(v_1,\dots,v_{j-1})} \mathrm{Col}_j(\bar{W})\right\|_2 \le \frac{\nu}{\sqrt{l}} \quad \text{for all } j \in \{2,\dots,l\}$$

and so

$$\sqrt{1-\frac{\nu^2}{l}} \le \bar{W}_{j,j} \le 1 \quad \text{for all } j \in [l],$$

since $\|\mathrm{Col}_j(\bar{W})\|_2 = 1$. Therefore,

$$\|\bar{W} - \mathrm{diag}(\bar{W})\| \le \|\bar{W} - \mathrm{diag}(\bar{W})\|_{\mathrm{HS}} = \left(\sum_{j=1}^{l}\sum_{i<j} \bar{W}_{i,j}^2\right)^{1/2} \le \nu,$$

and thus

$$\begin{aligned}
1 - 2\nu &\le 1 - \|I_l - \mathrm{diag}(\bar{W})\| - \|\mathrm{diag}(\bar{W}) - \bar{W}\| \\
&\le s_l(\bar{W}) \le s_1(\bar{W}) \\
&\le 1 + \|I_l - \mathrm{diag}(\bar{W})\| + \|\mathrm{diag}(\bar{W}) - \bar{W}\| \\
&\le 1 + 2\nu.
\end{aligned}$$

This implies the first claim of the lemma. The second claim immediately follows from the first one. □

The next lemma shows that if $W \subset \mathbb{R}^n \setminus \{0\}$ is a closed set and $E \subset \mathbb{R}^n$ is a linear subspace, then we can find a large almost orthogonal system in $E \cap W$ having a certain minimality property or a further linear subspace $F \subset E$ of a large dimension disjoint from W. This minimality property will be a key to estimating the least common denominator below.

LEMMA 3.3 (Almost orthogonal system). *Let $W \subset \mathbb{R}^n \setminus \{0\}$ be a closed set. Let $l < k \le n$, and let $E \subset \mathbb{R}^n$ be a linear subspace of dimension k. Then at least one of the following holds*:

1. *There exist vectors $v_1, \dots, v_l \in E \cap W$ such that*:
   (a) *The l-tuple $(v_1, \dots, v_l)$ is $(\frac{1}{8})$-almost orthogonal.*
   (b) *For any $\theta \in \mathbb{R}^l$ such that $\|\theta\|_2 \le \frac{1}{20\sqrt{l}}$,*

$$\sum_{i=1}^{l} \theta_i v_i \notin W.$$

2. *There exists a subspace $F \subset E$ of dimension $k - l$ such that $F \cap W = \varnothing$.*

REMARK 3.4. *The restriction $k \le cn^{1/2}$ in the formulation of Theorem* 1.1 *stems from the condition $\|\theta\|_2 \le \frac{1}{20\sqrt{l}}$ in Lemma* 3.3 (1b), *which in turn arises from using Lemma* 3.2 *for the l-tuple $v_1, \dots, v_l$.*

PROOF. Let us construct a sequence of vectors $v_1, \dots, v_{l'}, l' \le l$ with $\|v_1\|_2 \le \|v_2\|_2 \le \dots \le \|v_{l'}\|_2$ by induction. If $E \cap W = \varnothing$, then (2) holds for any subspace F of E of dimension $k-l$, so the lemma is proved. Assume that $E \cap W \neq \varnothing$, and define v_1 as the vector of this set having the smallest norm.

Let $2 \le j \le l-1$. For convenience, denote $v_0 = 0$. Assume that $j \in [l-1]$ and the vectors $v_1, \dots, v_j$ with $\|v_1\|_2 \le \|v_2\|_2 \le \dots \le \|v_j\|_2$ and such that, for all $0 \le i \le j-1$, v_{i+1} is the vector of the smallest norm in $E \cap W$ for which the inequality

$$\|P_{\mathrm{span}(v_0,\dots,v_i)} v_{i+1}\|_2 \le \frac{1}{16\sqrt{l}} \|v_i\|_2$$

holds. Note that if $j = 1$, then the condition above is vacuous, and the vector v_1 has been already constructed. Assume that $j \ge 2$, and we have found such vectors $v_1, \dots, v_j$. Consider the set

$$H_j = \left\{ v \in E \cap W : \|P_{\mathrm{span}(v_0,\dots,v_j)} v\|_2 \le \frac{1}{16\sqrt{l}} \|v_j\|_2 \right\}.$$

If $H_j = \varnothing$, then (2) holds for any subspace of $E \cap (\mathrm{span}(v_1, \dots, v_j))^{\perp}$ of dimension $k-l$, which proves the lemma in this case. Otherwise, choose a vector $v \in H_j$ having the smallest norm, and denote it by v_{j+1}. By construction, $\|v_{j+1}\|_2 \ge \|v_j\|_2$ since otherwise it would have been chosen at one of the previous steps.

Assume that we have run this process for l steps and constructed such sequence $v_1, \dots, v_l$. Then for any $j \in [l]$,

$$\|P_{\mathrm{span}(v_1,\dots,v_{j-1})} v_j\|_2 \le \frac{1}{16\sqrt{l}} \|v_{j-1}\|_2 \le \frac{1}{16\sqrt{l}} \|v_j\|_2,$$

and Lemma 3.2 ensures that (1a) holds. Therefore, to complete the induction step, we have to check only (1b). Assume that there exists $\theta \in \mathbb{R}^{j+1}$ such that $\|\theta\|_2 \le \frac{1}{20\sqrt{l}}$ and

$$\sum_{i=1}^{j+1} \theta_i v_i \in W. \tag{3.1}$$

Let V^j be the $n \times j$ matrix with columns $v_1, \dots, v_j$. The already verified condition (1a) yields $\|V^j\| \le \frac{9}{8} \max_{i \in [j]} \|v_i\|_2 \le \frac{9}{8} \|v_j\|_2$. Since $v_{j+1} \in H_j$,

$$\begin{aligned} \left\| P_{\mathrm{span}(v_1,\dots,v_j)} \left(\sum_{i=1}^{j+1} \theta_i v_i \right) \right\|_2 &\le \left\| \sum_{i=1}^{j} \theta_i v_i \right\|_2 + |\theta_{j+1}| \cdot \|P_{\mathrm{span}(v_1,\dots,v_j)} v_{j+1}\|_2 \\ &\le \|V^j\| \cdot \|\theta\|_2 + \|\theta\|_2 \cdot \frac{1}{16\sqrt{l}} \|v_j\|_2 \\ &\le \left(\frac{9}{8} + \frac{1}{16\sqrt{l}} \right) \|\theta\|_2 \cdot \|v_j\|_2 \\ &< \frac{1}{16\sqrt{l}} \|v_j\|_2. \end{aligned}$$

The last inequality above uses that $\|\theta\|_2 \le \frac{1}{20\sqrt{l}}$. Since by the inductive construction, v_{j+1} is the vector of the smallest norm in H_j having this property, $\|\sum_{i=1}^{j+1} \theta_i v_i\|_2 \ge \|v_{j+1}\|_2$. On the other hand, by (1a) and Lemma 3.2, $\|V^{j+1}\| \le \frac{9}{8} \|v_{j+1}\|_2$, so

$$\left\| \sum_{i=1}^{j+1} \theta_i v_i \right\|_2 \le \|V^{j+1}\| \cdot \|\theta\|_2 \le \frac{9}{8} \|v_{j+1}\|_2 \cdot \|\theta\|_2 \le \frac{1}{16\sqrt{l}} \|v_{j+1}\|_2.$$

This contradiction shows that (3.1) is not satisfied, so (1b) holds. □

3.2. *Concentration and tensorization.* We will need several elementary concentration results. To formulate them, we introduce a few definitions. Denote by $\mathcal{L}(X,t)$ the Levy concentration function of a random vector $X \in \mathbb{R}^m$,

$$\mathcal{L}(X,t) = \sup_{y \in \mathbb{R}^m} \mathbb{P}(\|X - y\|_2 \le t).$$

Let $\xi \in \mathbb{R}$ be a random variable. We will call it sub-Gaussian if $\mathbb{E}\exp(\lambda|\xi|^2) < \infty$ for some $\lambda > 0$ and denote

$$\|\xi\|_{\psi_2} := \inf\left\{s > 0 : \mathbb{E}\left[\exp\left(\frac{|\xi|}{s}\right)^2\right] \le 2\right\}.$$

For technical reasons, let us restrict the class of random entries of the matrix and introduce some parameters controlling their behavior. First, without loss of generality, we may assume that the entries of A are centered, that is, $\mathbb{E}a_{i,j} = 0$. Indeed, since all entries are i.i.d., subtracting the expectation from each one results in a rank one perturbation of the matrix A, which does not affect the conclusion of Theorem 1.1. Second, since the entries are nonconstant, $\mathcal{L}(a_{i,j}, t) < 1$ for some $t > 0$. After an appropriate scaling the entries, we can assume that $t = 1$. Therefore, throughout the paper, we will assume that the entries of the matrix A are i.i.d. copies of a random variable ξ satisfying the following conditions:

$$\mathbb{E}\xi = 0, \qquad \|\xi\|_{\psi_2} \le K, \qquad \mathcal{L}(\xi, 1) \le 1 - p. \tag{3.2}$$

Without loss of generality, we may assume that $K \ge 1$.

Throughout the paper we consider random matrices whose entries are independent copies of a random variable ξ satisfying (3.2). The constants c, C, C' etc., appearing in various formulas below, may depend on p and $\|\xi\|_{\psi_2}$.

LEMMA 3.5 (Operator norm). *Let $m \le n$, and let Q be an $m \times n$ matrix with centered independent entries $q_{i,j}$ such that $|q_{i,j}| \le 1$. Then*

$$\mathbb{P}(\|Q\| \ge C_{3.5}\sqrt{n}) \le \exp(-c_{3.5}n).$$

Lemma 3.5 follows from a general norm estimate for a random matrix with centered sub-Gaussian entries; see, for example, [14]. It is easy to see that the statement of the lemma is optimal up to constants C, c. Note that the event that $\|Q\| \ge C\sqrt{n}$ has probability which is exponentially small in n. Such bound is sufficient for the application we have in mind but is not strong enough to bound the operator norm of A. Indeed, as our aim is to prove the bound $\exp(-ckn)$ for the probability that the rank of A is smaller than $n - k$ and k can be large, we cannot exclude events of probability $\exp(-cn)$. This forces us to consider another matrix norm which enjoys stronger concentration properties.

For a matrix with sub-Gaussian entries, we prove a stronger bound for the Hilbert–Schmidt norm.

LEMMA 3.6 (Hilbert–Schmidt norm). *Let $m \le n$, and let A be an $m \times n$ matrix whose entries are independent copies of a random variable ξ satisfying* (3.2). *Then*

$$\mathbb{P}(\|A\|_{\mathrm{HS}} \ge 2Kn) \le \exp(-cn^2).$$

PROOF. Since $\mathbb{E}\xi^2 \le \|\xi\|_{\psi_2} < \infty$,

$$\mathbb{E}\exp\left(\frac{|\xi^2 - \mathbb{E}\xi^2|}{\|\xi\|_{\psi_2}^2}\right) \le \mathbb{E}\exp\left(\frac{\xi^2}{\|\xi\|_{\psi_2}^2} + 1\right) \le 2e,$$

which shows that $Y = \xi^2 - \mathbb{E}\xi^2$ is a centered subexponential random variable. Taking into account that

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{E}\, a_{i,j}^2 \le K^2 n^2$$

and Bernstein's inequality [19], we obtain

$$\mathbb{P}(\|A\|_{\mathrm{HS}} \ge 2Kn) = \mathbb{P}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{i,j}^2 - \mathbb{E}\, a_{i,j}^2) \ge 3K^2n^2\right) \le \exp(-cn^2),$$

as required. □

We will also need a tensorization lemma for the small ball probability similar to Lemma 2.2 [14].

LEMMA 3.7 (Tenzorization). *Let $m, M > 0$, and let $Y_1, \dots, Y_n \ge 0$ be independent random variables such that $\mathbb{P}(Y_j \le s) \le (Ms)^m$ for all $s \ge s_0$. Then*

$$\mathbb{P}\left(\sum_{j=1}^{n} Y_j \le nt\right) \le (CMt)^{mn} \quad \textit{for all } t \ge s_0.$$

PROOF. Let $t \ge s_0$. By Markov's inequality

$$\begin{aligned}\mathbb{P}\left(\sum_{j=1}^{n} Y_j \le nt\right) &\le \mathbb{E}\left[\exp\left(mn - \frac{m}{t}\sum_{j=1}^{n} Y_j\right)\right] \\ &= e^{mn}\prod_{j=1}^{n}\mathbb{E}\exp\left(-\frac{m}{t}Y_j\right),\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}\exp\left(-\frac{m}{t}Y_j\right) &= \int_0^1 \mathbb{P}\left[\exp\left(-\frac{m}{t}Y_j\right) > s\right]ds = \int_0^\infty e^{-u}\mathbb{P}\left[Y_j < \frac{t}{m}u\right]du \\ &\le \int_0^m e^{-u}\mathbb{P}\left[Y_j < \frac{t}{m}\right]du + \int_m^\infty e^{-u}\mathbb{P}\left[Y_j < \frac{t}{m}u\right]du \\ &\le (Mt)^m + \int_m^\infty e^{-u}\left(\frac{Mt}{m}u\right)^m du \\ &\le \left(1 + \frac{\Gamma(m+1)}{m^m}\right)\cdot (Mt)^m \le (CMt)^m.\end{aligned}$$

Here we used that $\mathbb{P}[Y_j < \frac{t}{m}u] \le \mathbb{P}[Y_j < \frac{t}{m}]$ for $u \in (0,1)$ in the first inequality and the Stirling formula in the last one. Combining the two inequalities above completes the proof. □

3.3. *Least common denominators and the small ball probability.* The least common denominator (LCD) of a sequence of real numbers originally introduced in [14] turned out to be a useful tool to gauge the behavior of the Levy concentration function of a linear combination of independent random variables with constant coefficients. Its various versions played a crucial role in proving quantitative estimates of invertibility of random matrices; see, for example, [13] and the references therein as well as more recent works, including [1, 2, 9, 18], and numerous other papers. In what follows, we use the extension of the LCD to matrices introduced in [16].

DEFINITION 3.8. *Let V be an $m \times n$ matrix, and let $L > 0$, $\alpha \in (0, 1]$. Define the* least common denominator (*LCD*) *of V by*

$$D_{L,\alpha}(V) = \inf\left(\|\theta\|_2 : \theta \in \mathbb{R}^m,\ \operatorname{dist}(V^\top\theta, \mathbb{Z}^n) < L\sqrt{\log_+ \frac{\alpha\|V^\top\theta\|_2}{L}}\right).$$

If $E \subset \mathbb{R}^n$ is a linear subspace, we can adapt this definition to the orthogonal projection P_E on E setting

$$D_{L,\alpha}(E) = D_{L,\alpha}(P_E) = \inf\left(\|y\|_2 : y \in E,\ \operatorname{dist}(y, \mathbb{Z}^n) < L\sqrt{\log_+ \frac{\alpha\|y\|_2}{L}}\right).$$

This is a modification of [18], Definition 6.1, and [16], Definition 7.1, where the same notion was introduced with $\alpha = 1$.

We will use the following concentration function estimate in terms of the LCD and its corollary proved in [16].

THEOREM 3.9 (Small ball probabilities via LCD). *Consider a random vector $\xi = (\xi_1, \dots, \xi_n)$, where ξ_k are i.i.d. copies of a real-valued random variable ξ satisfying* (3.2). *Consider a matrix $V \in \mathbb{R}^{n\times m}$. Then for every $L \geq \sqrt{m/p}$, we have*

$$\mathcal{L}(V^\top\xi, t\sqrt{m}) \leq \frac{(CL/(\alpha\sqrt{m}))^m}{\det(VV^\top)^{1/2}}\left(t + \frac{\sqrt{m}}{D_{L,\alpha}(V^\top)}\right)^m, \quad t \geq 0. \tag{3.3}$$

Theorem 3.9 with $\alpha = 1$ is [16], Theorem 7.5. We notice that exactly the same proof with the modified definition of the LCD yields Theorem 3.9 with a general α.

COROLLARY 3.10 (Small ball probabilities for projections). *Consider a random vector $\xi = (\xi_1, \dots, \xi_N)$, where ξ_k are i.i.d. copies of a real-valued random variable ξ satisfying* (3.2). *Let E be a subspace of $\mathbb{R}^N$ with $\dim(E) = m$, and let P_E denote the orthogonal projection onto E. Then for every $L \geq \sqrt{m/p}$, we have*

$$\mathcal{L}(P_E\xi, t\sqrt{m}) \leq \left(\frac{CL}{\alpha\sqrt{m}}\right)^m \left(t + \frac{\sqrt{m}}{D_{L,\alpha}(E)}\right)^m, \quad t \geq 0. \tag{3.4}$$

We will need a lemma which essentially generalizes the fact that the LCD of an incompressible vector is $\Omega(\sqrt{n})$. We will formulate it in a somewhat more technical way required for the future applications.

LEMMA 3.11. *Let $s, \alpha \in (0, 1)$. Let U be an $n \times l$ matrix such that $U\mathbb{R}^l \cap S^{n-1} \subset$* Incomp$(sn, \alpha)$ *Then any $\theta \in \mathbb{R}^l$ with $\|U\theta\|_2 \leq \sqrt{sn}/2$ satisfies*

$$\operatorname{dist}(U\theta, \mathbb{Z}^n) \geq L\sqrt{\log_+ \frac{\alpha\|U\theta\|_2}{L}}.$$

PROOF. Take any $\theta \in \mathbb{R}^l$ such that $\|U\theta\|_2 \leq \sqrt{sn}/2$. Let $x \in \mathbb{Z}^l$ be such that

$$\|U\theta - x\|_2 = \operatorname{dist}(U\theta, \mathbb{Z}^n) \leq \|U\theta\|_2.$$

Then by the triangle inequality, $\|x\|_2 \leq \sqrt{sn}$. Since the coordinates of x are integer, this implies that $|\operatorname{supp}(x)| \leq sn$. Therefore,

$$\left\|\frac{U\theta}{\|U\theta\|_2} - \frac{x}{\|U\theta\|_2}\right\|_2 \geq \alpha,$$

since $\frac{x}{\|U\theta\|_2} \in \text{Sparse}(sn)$. Combining the two previous inequalities, we see that

$$\alpha\|U\theta\|_2 \le \|U\theta - x\|_2 = \text{dist}(U\theta, \mathbb{Z}^n).$$

The desired inequality follows now from an elementary estimate $t > \sqrt{\log_+ t}$ valid for all $t > 0$, which is applied with $t = \frac{\alpha\|U\theta\|_2}{L}$. □

3.4. *Integer points inside a ball.* We will need a simple lemma estimating the number of integer points inside a ball in $\mathbb{R}^n$. Denote the Euclidean ball of radius R centered at 0 by $B(0, R)$ and the cardinality of a set F by $|F|$.

LEMMA 3.12. *For any $R > 0$,*

$$|\mathbb{Z}^n \cap B(0, R)| \le \left(2 + \frac{CR}{\sqrt{n}}\right)^n.$$

The proof immediately follows by covering $B(0, R)$ by unit cubes and estimating the volume of their union.

4. Compressible vectors. The aim of this section is to prove that it is unlikely that the kernel of a rectangular matrix with i.i.d. entries satisfying (3.2) contains a large, almost orthogonal, system of compressible vectors. More precisely, we prove that the probability of such event does not exceed $\exp(-clm)$, where m is the number of rows of B and l is the number of vectors in the system. The compressibility parameters will be selected in the process of the proof and after that, fixed for the rest of the paper. In Section 6 we will apply this statement with $m = n - k$ and $l = k/4$ in which case the probability of existence of such almost orthogonal system becomes negligible for our purposes.

We start with bounding the probability of presence of a fixed almost orthogonal system in the kernel of B. This bound relies on a corollary of the Hanson–Wright inequality; see [15], Corollary 2.4. Note that this result applies to any almost orthogonal system, not only to a compressible one.

LEMMA 4.1. *Let $m \le n$, and let let B be an $m \times n$ matrix whose entries are i.i.d. random variables satisfying (3.2). Let $l \le n$, and let $v_1, \dots, v_l \in S^{n-1}$ be an l-tuple of $(\frac{1}{2})$-almost orthogonal vectors. Then*

$$\mathbb{P}(\|Bv_j\|_2 \le C_{4.1}\sqrt{m} \text{ for all } j \in [l]) \le \exp(-c_{4.1}lm).$$

PROOF. Let $V = (v_1, \dots, v_l)$ be the $n \times l$ matrix formed by columns $v_1, \dots, v_l$. The assumption of the lemma implies that $\|V\| \le 2\max_{j\in[l]}\|v_j\|_2 = 2$. On the other hand,

$$\sum_{j=1}^{l} s_j^2(V) = \|V\|_{\text{HS}}^2 = \sum_{j=1}^{l}\|v_j\|_2^2 = l.$$

Note that if ξ is a random variable satisfying (3.2), then $\mathbb{E}\xi^2 \ge \mathbb{P}(|\xi| \ge 1) \ge p$. Let $\eta \in \mathbb{R}^n$ be a random vector with i.i.d. coordinates satisfying (3.2). By the Hanson–Wright inequality,

$$\mathbb{P}\left(\|V^\top\eta\|_2^2 \le \frac{p}{2}l\right) \le \mathbb{P}\left(\|V^\top\eta\|_2^2 \le \frac{p}{2}\|V\|_{\text{HS}}^2\right) \le \exp(-cl).$$

Let $\eta_1, \dots, \eta_m$ be i.i.d. copies of η. Then

$$\mathbb{P}\left(\sum_{i=1}^{m}\|V^\top\eta_i\|_2^2 \le \frac{p}{4}ln\right) \le \exp\left(-\frac{c}{2}lm\right).$$

Indeed, the condition $\sum_{i=1}^{m} \|V\eta_i\|_2^2 \le \frac{p}{4} lm$ implies that $\|V\eta_i\|_2^2 \le \frac{p}{2} l$ for at least $m/2$ indexes $i \in [m]$. These events are independent, and the probability of each one does not exceed $\exp(-cl)$.

Applying the inequality above with $\eta_j = (\mathrm{Row}_j(B))^\top$, we obtain

$$\mathbb{P}\left(\|Bv_j\|_2^2 \le \frac{p}{4} m \text{ for } j \in [l]\right) \le \mathbb{P}\left(\sum_{j=1}^{l} \|Bv_j\|_2^2 \le \frac{p}{4} m \cdot l\right)$$

$$= \mathbb{P}\left(\sum_{i=1}^{m} \|V^\top \eta_i\|_2^2 \le \frac{p}{4} lm\right) \le \exp\left(-\frac{c}{2} lm\right).$$

The proof is complete. □

The next statement, Proposition 4.2, contains the main result of this section. We will extend the bound of Lemma 4.1 from the presence of a fixed, almost orthogonal, system of compressible vectors in the kernel of a random matrix to the presence of any such system.

The proof of Proposition 4.2 follows the general roadmap of the geometric method. We start with constructing a special net for the set of compressible vectors in S^{n-1}. This net will consist of the vectors from a scaled copy of the integer lattice in $\mathbb{R}^n$. The vectors of an almost orthogonal system will be then approximated by the vectors from this net using the procedure of *random rounding*. This procedure, whose use in random matrix theory was pioneered by Livshyts [8], has now numerous applications in the problems related to invertibility. One of its advantages is that it allows to bound the approximation error in terms of a highly concentrated Hilbert–Schmidt norm instead of the operator norm of the matrix. In our case this approximation presents two new special challenges. First, we have to approximate all vectors $x_1, \dots, x_l$ forming an almost orthogonal system at the same time and in a way that preserves almost orthogonality. Second, the vectors of the approximating system have to retain some sparsity properties of the original vectors. We will show below that all these requirements can be satisfied simultaneously for a randomly chosen approximation. The probability of that will be exponentially small in l yet positive, which is sufficient since we need only to show the existence of such approximation.

PROPOSITION 4.2. *Let $k, n \in \mathbb{N}$ be such that $k \le n/2$, and let B be an $(n-k) \times n$ matrix whose entries are i.i.d. random variables satisfying* (3.2). *There exists $\tau > 0$ such that the probability that there exists a $(\frac{1}{4})$-almost orthogonal l-tuple $x_1, \dots, x_l \in \mathrm{Comp}(\tau^2 n, \tau^4)$ with $l \le \tau^3 n$ and*

$$\|Bx_j\|_2 \le \tau\sqrt{n} \quad \text{for all } j \in [l]$$

is less than $\exp(-cln)$.

PROOF. Let $\tau \in (0, 1/2)$ be a number to be chosen later, and set

$$T = \left\{v \in \frac{\tau}{\sqrt{n}} \mathbb{Z}^n : \|v\|_2 \in \left[\frac{1}{2}, 2\right]\right\}.$$

Then Lemma 3.12, applied to $R = \frac{\sqrt{n}}{\tau}$, yields

$$|T \cap \mathrm{Sparse}(4\tau^2 n)| \le \binom{n}{4\tau^2 n} \cdot \left(2 + \frac{C}{\tau}\right)^{4\tau^2 n} \le \left(\frac{C'}{\tau^3}\right)^{4\tau^2 n}. \tag{4.1}$$

Denote the coordinates of a vector $x \in \mathbb{R}^n$ by $x(1), \dots, x(n)$. Consider a $(\frac{1}{4})$-almost orthogonal l-tuple $x_1, \dots, x_l \in \mathrm{Comp}(\tau^2 n, \tau^4)$. Since $x_j \in \mathrm{Comp}(\tau^2 n, \tau^4)$, there is a set $I_1(j) \subset [n]$ with $|I_1(j)| \le \tau^2 n$ such that

$$\sum_{i \in [n] \setminus I_1(j)} x_j^2(i) \le \tau^8.$$

Using an elementary counting argument, we conclude that there exists $I_2(j) \supset I_1(j)$ with $|I_2(j)| \le 2\tau^2 n$ such that

$$|x_j(i)| \le \frac{\tau^3}{\sqrt{n}} \quad \text{for any } i \in [n] \setminus I_2(j).$$

For $j \in [l]$, define the vector $w_j = (w_j(1), \dots, w_j(n))$ by

$$w_j(i) = \frac{\tau}{\sqrt{n}} \cdot \left\lfloor \frac{\sqrt{n}}{\tau} |x_j(i)| \right\rfloor \mathrm{sign}(x_j(i)).$$

This form of rounding is chosen to approximate small in the absolute value coordinates of x_j by zeros.

Define independent random variables $\varepsilon_{i,j}$ such that

$$\mathbb{P}(\varepsilon_{i,j} = w_j(i) - x_j(i)) = 1 - \frac{\sqrt{n}}{\tau} |x_j(i) - w_j(i)|,$$

and

$$\mathbb{P}\left(\varepsilon_{i,j} = w_j(i) - x_j(i) + \frac{\tau}{\sqrt{n}} \mathrm{sign}(x_j(i))\right) = \frac{\sqrt{n}}{\tau} |x_j(i) - w_j(i)|.$$

Set

$$v_j = x_j + \sum_{i=1}^{n} \varepsilon_{i,j} e_i.$$

Then v_j is a random vector such that $\mathbb{E}\, v_j = x_j$. Moreover, $\|v_j - x_j\|_2 \le \tau < 1/2$, and so $\|v_j\|_2 \in [\frac{1}{2}, 2]$, which implies that $v_j \in T$ for all $j \in [l]$.

The definition of v_j above means that for any $i \in [n] \setminus I_2(j)$, $w_j(i) = 0$, and so $\mathbb{P}(v_j(i) \neq 0) \le \tau^2$ for these i. Set $I_3(j) = I_2(j) \cup \{i \in [n] \setminus I_2(j) v_j(i) \neq 0\}$. Since the events $v_j(i) \neq 0$ are independent for all $i \in [n] \setminus I_2(j)$ for a given $j \in [l]$, Chernoff's inequality in combination with the union bound over j yield

$$\mathbb{P}(\forall j \in [l] |I_3(j)| \le 4\tau^2 n) \ge 1 - l \exp(-c\tau^2 n). \tag{4.2}$$

Note that if $|I_3(j)| < 4\tau^2 n$ for all $j \in [l]$, then all the vectors $v_1, \dots, v_l$ belong to $T \cap \mathrm{Sparse}(4\tau^2 n)$.

Let us form the $n \times l$ matrices X and V with columns $x_1, \dots, x_l$ and $v_1, \dots, v_l$, respectively. Then the matrix $V - X$ has independent centered entries $\varepsilon_{i,j}$ whose absolute values are bounded by $\frac{\tau}{\sqrt{n}}$. This means that the random variables $\frac{\sqrt{n}}{\tau} \varepsilon_{i,j}$ satisfy the assumptions of Lemma 3.5. In view of this lemma,

$$\mathbb{P}(\|V - X\| \le C_{3.5} \tau) \ge 1 - \exp(-c_{3.5} n).$$

Define the diagonal matrix $D_V = \mathrm{diag}(\|v_1\|_2, \dots, \|v_l\|_2)$. Recall that

$$1 - \tau \le \|v_j\|_2 \le 1 + \tau$$

for all $j \in [l]$, and $s_l(X) \geq \frac{3}{4}$ since the vectors $x_1, \ldots, x_l$ are $(\frac{1}{4})$-almost orthogonal. Hence, if the event $\|V - X\| \leq C_{3.5}\tau$ occurs, then

$$\begin{aligned} s_l(V D_V^{-1}) &\geq s_l(X D_V^{-1}) - \|X - V\| \cdot \|D_V^{-1}\| \\ &\geq s_l(X) \cdot s_l(D_V^{-1}) - \|X - V\| \cdot \|D_V^{-1}\| \\ &\geq \frac{3}{4}(1+\tau)^{-1} - C_{3.5}\tau \cdot (1-\tau)^{-1} \\ &\geq \frac{1}{2}, \end{aligned}$$

where the last inequality holds if

$$\tau \leq \tau_0$$

for some $\tau_0 > 0$.

Similarly, we can show that $s_1(V D_V^{-1}) \leq \frac{3}{2}$, thus proving that the vectors $v_1, \ldots, v_l$ are $(\frac{1}{2})$-almost orthogonal. This shows that

$$\mathbb{P}\left(v_1, \ldots, v_l \text{ are } \left(\frac{1}{2}\right)\text{-almost orthogonal}\right) \geq 1 - \exp(-c_{3.5}n). \tag{4.3}$$

Let $\mathcal{E}_{\mathrm{HS}}$ be the event that $\|B\|_{\mathrm{HS}} \leq 2Kn$. Lemma 3.6 yields that

$$\mathbb{P}(\mathcal{E}_{\mathrm{HS}}) \geq 1 - \exp(-cn^2).$$

Condition on a realization of the matrix B such that $\mathcal{E}_{\mathrm{HS}}$ occurs. Since the random variables $\varepsilon_{i,j}$ are independent,

$$\begin{aligned} \mathbb{E}\|B(x_j - v_j)\|_2^2 &= \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_{i,j} B e_i\right\|_2^2 = \sum_{i=1}^n \mathbb{E}\,\varepsilon_{i,j}^2 \|B e_i\|_2^2 \\ &\leq \left(\frac{\tau}{\sqrt{n}}\right)^2 \|B\|_{\mathrm{HS}}^2 \leq 4K^2\tau^2 n. \end{aligned}$$

Hence, by Chebyshev's inequality

$$\mathbb{P}\left[\|B(x_j - v_j)\|_2 \leq 3K\tau\sqrt{n} \,|\, \mathcal{E}_{\mathrm{HS}}\right] \geq \frac{1}{2}.$$

In view of the independence of these events for different j,

$$\mathbb{P}\left[\forall j \in [l] \, \|B(x_j - v_j)\|_2 \leq 3K\tau\sqrt{n} \,|\, \mathcal{E}_{\mathrm{HS}}\right] \geq 2^{-l}. \tag{4.4}$$

Let us summarize (4.2), (4.3), and (4.4). Recall that $l \leq \tau^3 n$. If τ is sufficiently small, that is, $\tau \leq \tau_1$ for some $\tau_1 > 0$, then

$$1 - \exp(-c_{3.5}n) - l\exp(-c\tau^2 n) + 2^{-l} > 1.$$

This means that, conditionally on B for which the event $\mathcal{E}_{\mathrm{HS}}$ occurs, we can find a realization of random variables $\varepsilon_{i,j}$, $i \in [n]$, $j \in [l]$ such that:

- the vectors $v_1, \ldots, v_l$ are $(\frac{1}{2})$-almost orthogonal;
- $v_1, \ldots, v_l \in T \cap \mathrm{Sparse}(4\tau^2 n)$, and
- $\|B(x_j - v_j)\|_2 \leq 3K\tau\sqrt{n}$ for all $j \in [l]$.

Assume that there exists a $(\frac{1}{4})$-almost orthogonal l-tuple $x_1, \dots, x_l \in \mathrm{Comp}(\tau^2 n, \tau^4)$ such that $\|Bx_j\|_2 \le \tau\sqrt{n}$ for all $j \in [l]$. Then the above argument shows that, conditionally on B such that $\mathcal{E}_{\mathrm{HS}}$ occurs, we can find vectors $v_1, \dots, v_l \in T \cap \mathrm{Sparse}(4\tau^2 n)$, which are $(\frac{1}{2})$-almost orthogonal such that $\|Bv_j\|_2 \le 4K\tau\sqrt{n}$ for all $j \in [l]$ since $K \ge 1$. Therefore,

$$\begin{aligned}
&\mathbb{P}(\exists x_1, \dots, x_l \in \mathrm{Comp}(\tau^2 n, \tau^4) : \|Bx_j\|_2 \le \tau\sqrt{n} \text{ for all } j \in [l] \text{ and } \mathcal{E}_{\mathrm{HS}}) \\
&\quad \le \mathbb{P}\biggl[\exists v_1, \dots, v_l \in T \cap \mathrm{Sparse}(4\tau^2 n) : v_1, \dots, v_l \text{ are } \left(\frac{1}{2}\right)\text{-almost orthogonal} \\
&\qquad \text{and } \|Bv_j\|_2 \le 4K\tau\sqrt{n} \text{ for all } j \in [l] | \mathcal{E}_{\mathrm{HS}}\biggr] \cdot \mathbb{P}(\mathcal{E}_{\mathrm{HS}}) \\
&\quad = \mathbb{P}\biggl(\exists v_1, \dots, v_l \in T \cap \mathrm{Sparse}(4\tau^2 n) : v_1, \dots, v_l \text{ are } \left(\frac{1}{2}\right)\text{-almost orthogonal} \\
&\qquad \text{and } \|Bv_j\|_2 \le 4K\tau\sqrt{n} \text{ for all } j \in [l] \text{ and } \mathcal{E}_{\mathrm{HS}}\biggr).
\end{aligned}$$

Let us show that the latter probability is small. Assume that

$$\tau \le \min\left(\tau_0, \tau_1, \frac{C_{4.1}}{4K}\right).$$

Recall that the number of rows of B satisfies $n - k \ge n/2$. In view of Lemma 4.1 and (4.1),

$$\begin{aligned}
&\mathbb{P}\biggl(\exists v_1, \dots, v_l \in T \cap \mathrm{Sparse}(4\tau^2 n) : (v_1, \dots, v_l) \text{ is } \left(\frac{1}{2}\right)\text{-almost orthogonal} \\
&\qquad \text{and } \|Bv_j\|_2 \le 4K\tau\sqrt{n} \text{ for all } j \in [l]\biggr) \\
&\quad \le \left(\frac{C}{\tau^3}\right)^{4\tau^2 n \cdot l} \cdot \exp\left(-\frac{c_{4.1}}{2} ln\right) \le \exp\left(-\left[\frac{c_{4.1}}{2} - 4\tau^2 \log\left(\frac{C}{\tau^3}\right)\right] ln\right) \\
&\quad \le \exp\left(-\frac{c_{4.1}}{4} ln\right),
\end{aligned}$$

where the last inequality holds if we choose τ sufficiently small.

The previous proof shows that

$$\begin{aligned}
&\mathbb{P}(\exists x_1, \dots, x_l \in \mathrm{Comp}(\tau^2 n, \tau^4) \|Bx_j\|_2 \le \tau\sqrt{n} \ \text{ for all } j \in [l] \text{ and } \mathcal{E}_{\mathrm{HS}}) \\
&\quad \le \exp\left(-\frac{c_{4.1}}{4} ln\right).
\end{aligned}$$

In combination with the inequality $\mathbb{P}(\mathcal{E}_{\mathrm{HS}}^c) \le \exp(-cn^2)$, this completes the proof. □

We will fix the value of τ for which Proposition 4.2 holds for the rest of the paper.

5. Incompressible vectors. The main statement of this section, Proposition 5.1, bounds the probability that the kernel of a rectangular matrix B with i.i.d. entries satisfying assumptions (3.2) contains an almost orthogonal system of incompressible vectors with subexponential common denominators. In what follows, B will be the $(n-k) \times n$ matrix whose rows are $\mathrm{Col}_1(A)^\top, \dots, \mathrm{Col}_{n-k}(A)^\top$, and the required probability should be of order $\exp(-ckn)$ to fit Theorem 1.1. The need to achieve such a tight probability estimate requires considering the event that l vectors in the kernel of B have subexponential least common denominator. The number l here is proportional to k. Recall that a vector has a relatively small least common denominator if, after being scaled by a moderate factor, it becomes close to the integer lattice.

Since we have to consider l such vectors at once and the norms of these scaled copies vary significantly, it is more convenient to consider these copies and not the original unit vectors, as we did in Proposition 4.2. Moreover, to bound the probability, we have to consider all vectors with a moderate least common denominator in the linear span of the original system of l vectors in the kernel of B. To make the analysis of such linear span more manageable, we will restrict our attention to the almost orthogonal systems. This restriction will be later justified by using Lemma 3.3.

Throughout the paper we set

$$L = \sqrt{k/p}, \qquad \alpha = \frac{\tau^4}{4}, \tag{5.1}$$

where k appears in Theorem 1.1, p is a parameter from (3.2), and τ was chosen at the end of Section 4.

PROPOSITION 5.1. *Let $\rho \in (0, \rho_0)$, where $\rho_0 = \rho_0(\tau)$ is some positive number. Assume that $l \le k \le \frac{\rho}{2}\sqrt{n}$.*

Let B be an $(n-k) \times n$ matrix with i.i.d. entries satisfying (3.2). Consider the event $\mathcal{E}_l$ that there exist vectors $v_1, \dots, v_l \in \ker(B)$ having the following properties:

1. $\frac{\tau}{8}\sqrt{n} \le \|v_j\|_2 \le \exp(\frac{\rho^2 n}{4L^2})$ *for all* $j \in [l]$.
2. $\operatorname{span}(v_1, \dots, v_l) \cap S^{n-1} \subset \operatorname{Incomp}(\tau^2 n, \tau^4)$.
3. *The vectors $v_1, \dots, v_l$ are $(\frac{1}{8})$-almost orthogonal.*
4. $\operatorname{dist}(v_j, \mathbb{Z}^n) \le \rho\sqrt{n}$ *for* $j \in [l]$.
5. *The $n \times l$ matrix V with columns $v_1, \dots, v_l$ satisfies*

$$\operatorname{dist}(V\theta, \mathbb{Z}^n) > \rho\sqrt{n}$$

for all $\theta \in \mathbb{R}^l$ such that $\|\theta\|_2 \le \frac{1}{20\sqrt{l}}$ and $\|V\theta\|_2 \ge \frac{\tau}{8}\sqrt{n}$.

Then

$$\mathbb{P}(\mathcal{E}_l) \le \exp(-ln).$$

Conditions (1)–(4) mean that $v_1, \dots, v_l$ is a (1/8)-almost orthogonal system of incompressible vectors close to the integer lattice, and condition 5 is a certain minimality property of this system.

To simplify the analysis, we will tighten condition (1) of Proposition 5.1 restricting the magnitudes of the norms of v_j to some dyadic intervals. Denote for shortness

$$r = \frac{\tau}{16}, \qquad R = \exp\left(\frac{\rho^2 n}{4L^2}\right). \tag{5.2}$$

Consider a vector $\mathbf{d} = (d_1, \dots, d_l) \in [r\sqrt{n}, R]^l$, and define the set $W_{\mathbf{d}}$ be the set of l-tuples of vectors $v_1, \dots, v_l \in \mathbb{R}^n$, satisfying

$$\|v_j\|_2 \in [d_j, 2d_j] \quad \text{for all } j \in [l]$$

and conditions (2)–(5) of Proposition 5.1. We will prove the proposition for vectors $v_1, \dots, v_l$ with such restricted norms first and derive the general statement by taking the union bound over $d_1, \dots, d_l$ being dyadic integers.

We begin the proof of Proposition 5.1 with constructing a special net for the set $W_{\mathbf{d}}$. This will follow by proving that the l-tuples from the net approximate any point of $W_{\mathbf{d}}$ in a number of senses. After that we will prove the individual small ball probability estimate for *some* l-tuples from the net. These will be exactly those tuples that appear as a result of approximation of points from $W_{\mathbf{d}}$.

To make the construction of the net simpler, we introduce another parameter. Given ρ, as in Proposition 5.1, we will chose $\delta > 0$ such that

$$\delta \le \rho \quad \text{and} \quad \delta^{-1} \in \mathbb{N}. \tag{5.3}$$

The parameter δ will be adjusted several times throughout the proof, but its value will remain independent of n.

LEMMA 5.2 (Net cardinality). *Let $\mathbf{d} = (d_1, \dots, d_l)$ be a vector such that $d_j \in [r\sqrt{n}, R]$ for all $j \in [l]$. Let δ be as in (5.3), and $\mathcal{N}_{\mathbf{d}} \subset (\delta\mathbb{Z}^n)^l$ be the set of all l-tuples of vectors $u_1, \dots, u_l$ such that*

$$\|u_j\|_2 \in \left[\frac{1}{2} d_j, 4d_j\right] \quad \textit{for all } j \in [l]$$

and

$$\operatorname{dist}(u_j, \mathbb{Z}^n) \le 2\rho\sqrt{n}.$$

Then

$$|\mathcal{N}_{\mathbf{d}}| \le \left(\frac{C\rho}{r\delta}\right)^{ln} \left(\prod_{j=1}^{l} \frac{d_j}{\sqrt{n}}\right)^n.$$

PROOF. Let $\mathcal{M}_j = \mathbb{Z}^n \cap 2d_j B_2^n$. Taking into account that $d_j \ge r\sqrt{n}$, we use Lemma 3.12 to conclude that

$$|\mathcal{M}_j| \le \left(2 + \frac{Cd_j}{\sqrt{n}}\right)^n \le \left(\frac{C'}{r}\right)^n \cdot \left(\frac{d_j}{\sqrt{n}}\right)^n.$$

Define the set $\mathcal{M}$ by $\mathcal{M} = \delta\mathbb{Z}^n \cap 2\rho\sqrt{n} B_2^n$. Similarly, Lemma 3.12 yields

$$|\mathcal{M}| \le \left(\frac{C\rho}{\delta}\right)^n.$$

Set $\mathcal{N}_j = \mathcal{M}_j + \mathcal{M} \subset \delta\mathbb{Z}^n$. Here we used the assumption that $\delta^{-1} \in \mathbb{N}$. Then by construction

$$|\mathcal{N}_j| \le \left(\frac{C\rho}{r\delta}\right)^n \cdot \left(\frac{d_j}{\sqrt{n}}\right)^n.$$

Set $\mathcal{N}_{\mathbf{d}} = \prod_{j=1}^{l} \mathcal{N}_j$. Multiplying the previous estimates, we obtain

$$|\mathcal{N}_{\mathbf{d}}| \le \left(\frac{C\rho}{r\delta}\right)^{ln} \left(\prod_{j=1}^{l} \frac{d_j}{\sqrt{n}}\right)^n,$$

as required. □

The next step is the central technical part of this section. Our next task is to show that, for any $(v_1, \dots, v_l) \in W_{\mathbf{d}}$, there exists a sequence $(u_1, \dots, u_l) \in \mathcal{N}_{\mathbf{d}}$ which approximates it in various ways. As some of these approximations hold only for a randomly chosen point of $\mathcal{N}_{\mathbf{d}}$ and we need all of them to hold simultaneously, we have to establish all of them at the same time. This will be done by using random rounding, as in the proof of Proposition 4.2. The implementation of this method here is somewhat different since we have to control the least common denominator of the matrix U formed by the vectors $u_1, \dots, u_l$.

We will prove the following lemma.

LEMMA 5.3 (Approximation). *Let $k \le cn$. Let $\mathbf{d} = (d_1, \dots, d_l) \in [r\sqrt{n}, R]^l$. Let $\delta > 0$ be a sufficiently small constant satisfying* (5.3). *Let B be an $(n-k) \times n$ matrix such that $\|B\|_{HS} \le 2Kn$. For any sequence $(v_1, \dots, v_l) \in W_{\mathbf{d}} \cap \mathrm{Ker}(B)$, there exists a sequence $(u_1, \dots, u_l) \in \mathcal{N}_{\mathbf{d}}$ with the following properties*:

1. $\|u_j - v_j\|_\infty \le \delta$ *for all* $j \in [l]$.
2. *Let U and V be $n \times l$ matrices with columns $u_1, \dots, u_l$ and $v_1, \dots, v_l$ respectively. Then*
$$\|U - V\| \le C\delta\sqrt{n}.$$
3. *The system $(u_1, \dots, u_l)$ is* $(1/4)$*-orthogonal.*
4. $\mathrm{span}(u_1, \dots, u_l) \cap S^{n-1} \subset \mathrm{Incomp}(\tau^2, \tau^4/2)$.
5. $\mathrm{dist}(u_j, \mathbb{Z}^n) \le 2\rho\sqrt{n}$ *for all* $j \in [n]$.
6. *Let U be as in* (2). *Then*
$$\mathrm{dist}(U\theta, \mathbb{Z}^n) > \frac{\rho}{2}\sqrt{n}$$
for any $\theta \in \mathbb{R}^n$ satisfying
$$\|\theta\|_2 \le \frac{1}{20\sqrt{l}} \quad \textit{and} \quad \|U\theta\|_2 \ge 8r\sqrt{n}.$$
7. $\|Bu_j\|_2 \le 2K\delta n$ *for all* $j \in [l]$.

The conditions (3)–(6) are the same as (2)–(5) of Proposition 5.1 up to a relaxation of some parameters.

PROOF. Let $(v_1, \dots, v_l) \in W_{\mathbf{d}}$. Choose $(v_1', \dots, v_l') \in \delta\mathbb{Z}^n$ be such that
$$v_j \in v_j' + \delta[0,1]^n \quad \text{for all } j \in [l].$$
Define independent random variables $\varepsilon_{i,j}$, $i \in [n]$, $j \in [l]$ by setting
$$\mathbb{P}(\varepsilon_{i,j} = v_j'(i) - v_j(i)) = 1 - \frac{v_j(i) - v_j'(i)}{\delta}$$
and
$$\mathbb{P}(\varepsilon_{i,j} = v_j'(i) - v_j(i) + \delta) = \frac{v_j(i) - v_j'(i)}{\delta}.$$
Then $|\varepsilon_{i,j}| \le \delta$ and $\mathbb{E}\,\varepsilon_{i,j} = 0$. Consider a random point
$$u_j = v_j + \sum_{i=1}^n \varepsilon_{i,j} e_i \in \delta\mathbb{Z}^n.$$
Then $\mathbb{E}\,u_j = v_j$ and $\|u_j - v_j\|_\infty \le \delta$ for all $j \in [l]$, as in (1). Let us check that $(u_1, \dots, u_l) \in \mathcal{N}_{\mathbf{d}}$ for any choice of $\varepsilon_{i,j}$. Indeed, for any $j \in [l]$,
$$\|u_j - v_j\|_2 \le \delta\sqrt{n} \quad \text{and} \quad \left(1 - \frac{\delta}{r}\right)\|v_j\|_2 \le \|u_j\|_2 \le \left(1 + \frac{\delta}{r}\right)\|v_j\|_2,$$
as $\|v_j\|_2 \ge r\sqrt{n}$ for all $j \in [l]$. This, in particular, implies that $\|u_j\|_2 \in [\frac{1}{2}d_j, 4d_j]$ for all $j \in [l]$ and any values of $\varepsilon_{i,j}$.

Let U and V be the $n \times l$ matrices with columns $u_1, \dots, u_l$ and $v_1, \dots, v_l$, respectively. Then the matrix $U - V$ has independent entries $\varepsilon_{i,j}$, $i \in [n]$, $j \in [l]$, which are centered and bounded by δ in the absolute value. By Lemma 3.5
$$\mathbb{P}(\|U - V\| \ge C_{3.5}\delta\sqrt{n}) \le \exp(-c_{3.5}n),$$
and so condition (2) holds with probability at least $1 - \exp(-c_{3.5}n)$.

Let us check that condition (3) follows from (2). Let D_U be the diagonal matrix $D_U = \operatorname{diag}(\|u_1\|_2, \dots, \|u_l\|_2)$, and define D_V in a similar way. If $\|U - V\| \le C_{3.5}\delta\sqrt{n}$, then by the $(\frac{1}{8})$-almost orthogonality of $(v_1, \dots, v_l)$, we get

$$\begin{aligned}\|U D_U^{-1}\| &\le \|U D_V^{-1}\| \cdot \|D_V D_U^{-1}\| \\ &\le [\|V D_V^{-1}\| + \|U - V\| \cdot \|D_V^{-1}\|] \cdot \|D_V D_U^{-1}\| \\ &\le \left[\frac{9}{8} + C_{3.5}\delta\sqrt{n} \cdot \frac{1}{r\sqrt{n}}\right] \cdot \left(1 - \frac{\delta}{r}\right)^{-1} \le \frac{5}{4}\end{aligned}$$

if $\delta \le cr$ for an appropriately small constant $c > 0$. Similarly,

$$\begin{aligned}s_l(U D_U^{-1}) &\ge s_l(U D_V^{-1}) \|D_U D_V^{-1}\|^{-1} \\ &\ge [s_l(V D_V^{-1}) - \|U - V\| \cdot \|D_V^{-1}\|] \cdot \|D_U D_V^{-1}\|^{-1} \\ &\ge \left[\frac{7}{8} - C_{3.5}\frac{\delta}{r}\right] \cdot \left(1 + \frac{\delta}{r}\right)^{-1} \ge \frac{3}{4}\end{aligned}$$

confirming our claim. The last inequality above follows again by choosing $\delta < cr$ with a sufficiently small $c > 0$.

Let us check that condition (4) follows from (2) and (3). Indeed, let $\theta \in \mathbb{R}^l$ be such that $\|U\theta\|_2 = 1$. Since $\min(\|u_1\|_2, \dots, \|u_l\|_2) \ge r\sqrt{n}$ and the system $u_1, \dots, u_l$ is $(\frac{1}{4})$-almost orthogonal, we have

$$\|\theta\|_2 \le (s_l(U))^{-1}\|U\theta\|_2 \le \frac{4}{3} \cdot \frac{1}{r\sqrt{n}}.$$

At the same time,

$$\|V\theta\|_2 \ge \|U\theta\|_2 - \|U - V\| \cdot \|\theta\|_2 \ge 1 - C_{3.5}\delta\sqrt{n} \cdot \frac{4}{3} \cdot \frac{1}{r\sqrt{n}} = 1 - C_{3.5}\frac{4\delta}{3r}.$$

Take any $y \in \operatorname{Sparse}(\tau^2 n)$. Then

$$\|V\theta - y\|_2 \ge \left(1 - C_{3.5}\frac{4\delta}{3r}\right) \cdot \left\|\frac{V\theta}{\|V\theta\|_2} - \frac{y}{\|V\theta\|_2}\right\|_2 \ge \left(1 - C_{3.5}\frac{4\delta}{3r}\right) \cdot \tau^4$$

since $\frac{V\theta}{\|V\theta\|_2} \in \operatorname{Incomp}(\tau^2 n, \tau^4)$. Therefore,

$$\|U\theta - y\|_2 \ge \|V\theta - y\|_2 - \|U - V\| \cdot \|\theta\|_2 \ge \left(1 - C_{3.5}\frac{4\delta}{3r}\right) \cdot \tau^4 - C_{3.5}\frac{4\delta}{3r} \ge \frac{1}{2}\tau^4,$$

where the last inequality holds if δ is appropriately adjusted depending on r and τ. This proves that if $\|U\theta\|_2 = 1$, then $\operatorname{dist}(U\theta, \operatorname{Sparse}(\tau^2 n)) \ge \tau^4/2$, that is, $U\theta \in \operatorname{Incomp}(\tau^n, \tau^4/2)$. Thus, (4) is verified.

Condition (5) immediately follows from (1) and the triangle inequality,

$$\operatorname{dist}(u_j, \mathbb{Z}^n) \le \operatorname{dist}(v_j, \mathbb{Z}^n) + \delta\sqrt{n} \le 2\rho\sqrt{n}$$

since $\delta \le \rho$.

Condition (6) follows from (2) and (3). Indeed, let θ be as in (6), and assume that $\|U - V\| \le C_{3.5}\delta\sqrt{n}$. Since both $(v_1, \dots, v_l)$ and $(u_1, \dots, u_l)$ are $(\frac{1}{4})$-almost orthogonal and $\|u_j\|_2 \ge \frac{1}{2}\|v_j\|_2$,

$$\|V\theta\|_2^2 \ge \frac{1}{4}\sum_{j=1}^{l} \theta_j^2 \|v_j\|_2^2 \ge \frac{1}{16}\sum_{j=1}^{l} \theta_j^2 \|u_j\|_2^2 \ge \frac{1}{64}\|U\theta\|_2^2 \ge r^2 n.$$

As $(v_1, \dots, v_l) \in W_{\mathbf{d}}$, this implies that

$$\operatorname{dist}(V\theta, \mathbb{Z}^n) > \rho\sqrt{n}.$$

Therefore,

$$\begin{aligned}\operatorname{dist}(U\theta, \mathbb{Z}^n) &\geq \operatorname{dist}(V\theta, \mathbb{Z}^n) - \|(U-V)^\top\theta\|_2 \\ &> \rho\sqrt{n} - \|U-V\| \cdot \|\theta\|_2 \geq \rho\sqrt{n} - C_{3.5}\delta\sqrt{n} \cdot \frac{1}{20\sqrt{l}} \\ &\geq \frac{\rho}{2}\sqrt{n},\end{aligned}$$

where we adjust δ again, if necessary. Thus, $(u_1, \dots, u_l)$ satisfy (2)–(6) with probability at least $1 - \exp(-c_{3.5}n)$.

It remains to show that we can choose $(u_1, \dots, u_l)$ satisfying (7) at the same time. For any $j \in [l]$, we have

$$\mathbb{E}\|B(u_j - v_j)\|_2^2 = \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_{i,j} Be_i\right\|_2^2 = \sum_{i=1}^n \mathbb{E}\,\varepsilon_{i,j}^2 \|Be_i\|_2^2 \leq \frac{\delta^2}{4}\|B\|_{\mathrm{HS}}^2 \leq K^2\delta^2 n^2.$$

By Chebyshev's inequality

$$\mathbb{P}(\|B(u_j - v_j)\|_2 \leq 2K\delta n) \geq \frac{1}{2}.$$

In view of independence of these events for different j,

$$\mathbb{P}(\forall j \in [l] \|B(u_j - v_j)\|_2 \leq 2K\delta n) \geq 2^{-l}.$$

As

$$1 - \exp(-c_{3.5}n) + 2^{-l} > 1,$$

there is a realization $(u_1, \dots, u_l) \in \mathcal{N}_{\mathbf{d}}$, satisfying (2)–(6), for which

$$\|Bu_j\|_2 = \|B(v_j - w_j)\|_2 \leq 2K\delta n$$

holds for all $j \in [l]$ simultaneously. This finishes the proof of the lemma. □

Fix the value of δ satisfying (5.3) such that Lemma 5.3 holds for the rest of the proof.

We will now use the small ball probability estimate of Theorem 3.9 to show that the event $W_{\mathbf{d}} \cap \operatorname{Ker}(B) \neq \varnothing$ is unlikely.

LEMMA 5.4. *Let* $\mathbf{d} = (d_1, \dots, d_l) \in [r\sqrt{n}, R]^l$ *where* R, r *are defined above. Let* $k \leq \frac{\delta}{20}\sqrt{n}$ *and* $\frac{k}{10} \leq l \leq k$. *Then*

$$\mathbb{P}(W_{\mathbf{d}} \cap \operatorname{Ker}(B) \neq \varnothing) \leq \exp(-2ln).$$

PROOF. Let $\mathcal{N}_{\mathbf{d}}$ be the net constructed in Lemma 5.2. Let $\tilde{\mathcal{N}_{\mathbf{d}}}$ be the set of all $(u_1, \dots, u_l) \in \mathcal{N}_{\mathbf{d}}$ which satisfy conditions (3)–(6) of Lemma 5.3. Consider an l-tuple $u_1, \dots, u_l \in \tilde{\mathcal{N}_{\mathbf{d}}}$. Let U be the $n \times l$ matrix with columns $u_1, \dots, u_l$.

To apply the Levy concentration estimate of Theorem 3.9, we have to bound the LCD of $U^\top$ from below. Let us show that

$$D_{L,\alpha}(U^\top) \geq \frac{1}{20\sqrt{l}}. \tag{5.4}$$

Take $\theta \in \mathbb{R}^l$ such that $\|\theta\|_2 \le \frac{1}{20\sqrt{l}}$. Assume first that

$$\|U\theta\|_2 \le 8r\sqrt{n} = \sqrt{\tau^2 n}/2.$$

Recall that L and α are defined as in (5.1). Since the columns of U satisfy (4) of Lemma 5.3, applying Lemma 3.11 yields

$$\operatorname{dist}(U\theta, \mathbb{Z}^n) \ge L\sqrt{\log_+ \frac{\alpha\|U\theta\|_2}{L}}.$$

Assume now that $\|U\theta\|_2 > 8r\sqrt{n}$. By definition of the set $\mathcal{N}_{\mathbf{d}}$ in Lemma 5.2,

$$\|U\|_{\mathrm{HS}} \le \sqrt{l} \max_{j\in[l]} \|u_j\|_2 \le \sqrt{l}R.$$

Hence,

$$L\sqrt{\log_+ \frac{\alpha\|U\theta\|_2}{L}} \le L\sqrt{\log_+ \frac{\|U\|_{\mathrm{HS}}}{L}} \le L\sqrt{\log_+ R} \le L\sqrt{\frac{\rho^2}{4}\cdot\frac{n}{L^2}} \le \frac{\rho}{2}\sqrt{n}.$$

By condition (6) of Lemma 5.3,

$$\operatorname{dist}(U\theta, \mathbb{Z}^n) > \frac{\rho}{2}\sqrt{n}$$

whenever $\theta \in \mathbb{R}^n$ satisfies

$$\|\theta\|_2 \le \frac{1}{20\sqrt{l}} \quad \text{and} \quad \|U\theta\|_2 \ge 8r\sqrt{n}.$$

Combining these two cases, we see that any vector $\theta \in \mathbb{R}^l$ with $\|\theta\|_2 \le \frac{1}{20\sqrt{l}}$ satisfies

$$\operatorname{dist}(U\theta, \mathbb{Z}^n) \ge L\sqrt{\log_+ \frac{\alpha\|U\theta\|_2}{L}},$$

which proves (5.4).

Using condition (3) of Lemma 5.3 and Lemma 3.2, we infer

$$\det(U^\top U)^{1/2} \ge 4^{-l} \prod_{j=1}^{l} \|u_j\|_2 \ge 8^{-l} \prod_{j=1}^{l} d_j.$$

Let $i \in [n]$. Recall that $\mathrm{Row}_i(B) \in \mathbb{R}^n$ is a vector with i.i.d. random coordinates satisfying (3.2) and that $l \le k \le \frac{\delta}{20\sqrt{n}}$. Combining this with (3.3) used with

$$t \ge \delta\sqrt{n} \ge 20l \ge \frac{\sqrt{l}}{D_{L,\alpha}(U^\top)}$$

and recalling that $L = O(\sqrt{k})$ by (5.1) and $l \ge k/10$, we obtain

$$\mathbb{P}(\|U^\top(\mathrm{Row}_i(B))^\top\|_2 \le t\sqrt{l}) \le \frac{(CL/\sqrt{l})^l}{\det(U^\top U)^{1/2}}\left(t + \frac{\sqrt{l}}{D_{L,\alpha}(U^\top)}\right)^l \le \frac{C^l}{\prod_{j=1}^l d_j} t^l.$$

Denote

$$Y_i = \frac{1}{l}\|U^\top(\mathrm{Row}_i(B))^\top\|_2^2, \qquad M = \frac{C^2}{(\prod_{j=1}^l d_j)^{2/l}}.$$

Then we can rewrite the last inequality as

$$\mathbb{P}(Y_i \le s) \le (Ms)^{l/2} \quad \text{for } s \ge s_0 = \delta^2 n.$$

In view of Lemma 3.7 applied with $m = l/2$ and $t = 4K^2 s_0$ with K from (3.2), this yields

$$\begin{aligned}\mathbb{P}\big(\|Bu_j\|_2 \le 2K\delta n \text{ for all } j \in [l]\big) &\le \mathbb{P}\Bigg(\sum_{j=1}^{l} \|Bu_j\|_2^2 \le 4K^2\delta^2 l n^2\Bigg)\\ &= \mathbb{P}\Bigg(\sum_{i=1}^{n-k} \|U^\top (\mathrm{Row}_i(B))^\top\|_2^2 \le 4K^2\delta^2 l n^2\Bigg)\\ &= \mathbb{P}\Bigg(\sum_{i=1}^{n-k} Y_i \le n \cdot 4K^2\delta^2 n\Bigg) \le \big(C' M\delta^2 n\big)^{(n-k)l/2}\\ &= (C''\delta)^{l(n-k)} \cdot \Bigg(\prod_{j=1}^{l} \frac{\sqrt{n}}{d_j}\Bigg)^{n-k}.\end{aligned}$$

Since $\tilde{\mathcal{N}_{\mathbf{d}}} \subset \mathcal{N}_{\mathbf{d}}$, a combination of the small ball probability estimate above and Lemma 5.2 gives

$$\begin{aligned}&\mathbb{P}\big(\exists (u_1,\dots,u_l) \in \tilde{\mathcal{N}_{\mathbf{d}}} : \|Bu_j\|_2 \le \delta n,\ j \in [l]\big)\\ &\quad\le |\mathcal{N}_{\mathbf{d}}| \cdot (C''\delta)^{l(n-k)} \cdot \Bigg(\prod_{j=1}^{l} \frac{\sqrt{n}}{d_j}\Bigg)^{n-k}\\ &\quad\le \bigg(\frac{C\rho}{r\delta}\bigg)^{ln} \Bigg(\prod_{j=1}^{l} \frac{d_j}{\sqrt{n}}\Bigg)^{n} \cdot (C''\delta)^{l(n-k)} \cdot \Bigg(\prod_{j=1}^{l} \frac{\sqrt{n}}{d_j}\Bigg)^{n-k}\\ &\quad= \bigg(\frac{C'\rho}{r}\bigg)^{ln} \cdot \delta^{-lk} \Bigg(\prod_{j=1}^{l} \frac{d_j}{\sqrt{n}}\Bigg)^{k}.\end{aligned}$$

Recall that, by (5.2),

$$d_j \le R = \exp\bigg(\frac{\rho^2 n}{4L^2}\bigg) \le \exp\bigg(\frac{C\rho^2 n}{k}\bigg) \quad \text{for all } j \in [l],$$

where the last inequality follows from (5.1). Therefore,

$$\begin{aligned}\mathbb{P}\big(\|Bu_j\|_2 \le 2K\delta n \text{ for all } j \in [l]\big) &\le \bigg(\frac{C'\rho}{r}\bigg)^{ln} \cdot \bigg(\frac{R}{\delta\sqrt{n}}\bigg)^{lk} \le \bigg(\frac{\tilde{C}\rho}{r} \exp\big(C\rho^2\big)\bigg)^{ln}\\ &\le \exp(-2\,\mathrm{ln})\end{aligned}$$

if $\rho < cr$ for a sufficiently small constant $c > 0$.

Notice that

$$\begin{aligned}\mathbb{P}\big(W_{\mathbf{d}} \cap \mathrm{Ker}(B) \neq \varnothing\big) &\le \mathbb{P}\big(W_{\mathbf{d}} \cap \mathrm{Ker}(B) \neq \varnothing \text{ and } \|B\|_{\mathrm{HS}} \le 2Kn\big)\\ &\quad + \mathbb{P}\big(\|B\|_{\mathrm{HS}} \ge 2Kn\big).\end{aligned}$$

In view of Lemma 3.6, the second term is smaller than $\exp(-cn^2)$, which means that we have to concentrate on the first one.

Assume that the events $W_{\mathbf{d}} \cap \mathrm{Ker}(B) \neq \varnothing$ and $\|B\|_{\mathrm{HS}} \le 2Kn$ occur, and pick an l-tuple $(v_1,\dots,v_l) \in W_{\mathbf{d}} \cap \mathrm{Ker}(B)$. Choose an approximating l-tuple $(u_1,\dots,u_l) \in \mathcal{N}_{\mathbf{d}}$ as in Lemma 5.3. Then $(u_1,\dots,u_l) \in \tilde{\mathcal{N}_{\mathbf{d}}}$ and $\|Bu_j\|_2 \le 2K\delta n$ per condition (7) of this lemma. The argument above shows that the probability of the event that such a tuple $(u_1,\dots,u_l) \in \tilde{\mathcal{N}_{\mathbf{d}}}$ exists is at most $\exp(-2\,\mathrm{ln})$. The lemma is proved. □

Proposition 5.1 follows from Lemma 5.4 by taking the union bound over dyadic values of the coordinates of $\mathbf{d}$.

PROOF OF PROPOSITION 5.1. Let $\mathcal{E}_{\mathbf{d}}$ be the event that $W_{\mathbf{d}} \cap \operatorname{Ker}(B) \neq \varnothing$. Then

$$\mathcal{E}_l = \cup \mathcal{E}_{\mathbf{d}},$$

where the union is taken over all vectors $\mathbf{d}$ with dyadic coordinates: $d_j = 2^{s_j}$, $s_j \in \mathbb{N}$ such that $2^{s_j} \in [r\sqrt{n}, R]$. Since there are at most

$$\left[\log\left(\frac{2R}{r\sqrt{n}}\right)\right]^l \leq \left(\frac{C\rho^2 n}{L^2}\right)^l$$

terms in the union, Lemma 5.4 yields

$$\mathbb{P}(\mathcal{E}_{\mathbf{d}}) \leq \left(\frac{C\rho^2 n}{L^2}\right)^l \exp(-2ln) \leq \exp(-ln),$$

where we took into account that $L > 1$. This finishes the proof of the proposition. □

6. Rank of a random matrix. We will complete the proof of Theorem 1.1 using the probability estimates of Propositions 4.2 and 5.1. These propositions show that the linear subspace, orthogonal to the span of the first $n-k$ columns of the matrix A, is unlikely to contain a large, almost orthogonal, system of vectors with a small or moderate least common denominator. Applying Lemma 3.3, we will show that, with high probability, this subspace contains a further subspace of a dimension proportional to k, which has no vectors with a subexponential least common denominator. The next lemma shows that, in such a typical situation, it is unlikely that the rank of the matrix A is $n-k$ or smaller.

LEMMA 6.1. *Let A be an $n \times n$ random matrix whose entries are independent copies of a random variable ξ satisfying* (3.2). *For $k < \sqrt{n}$, define*

$$\Omega_k = \Omega_k\big(\operatorname{Col}_1(A), \dots, \operatorname{Col}_{n-k}(A)\big)$$

as the event that there exists a linear subspace $E \subset (\operatorname{span}(\operatorname{Col}_1(A), \dots, \operatorname{Col}_{n-k}(A)))^{\perp}$ such that $\dim(E) \geq k/2$ *and*

$$D_{L,\alpha}(E) \geq \exp\left(C\frac{n}{k}\right).$$

Then

$$\mathbb{P}\big(\operatorname{Col}_j(A) \in \operatorname{span}\big(\operatorname{Col}_i(A), i \in [n-k]\big) \textit{ for } j = n-k+1, \dots, n \textit{ and } \Omega_k\big) \\ \leq \exp(-c'nk).$$

PROOF. Assume that Ω_k occurs. The subspace E can be selected in a measurable way with respect to the sigma-algebra generated by $\operatorname{Col}_1(A), \dots, \operatorname{Col}_{n-k}(A)$. Therefore, conditioning on $\operatorname{Col}_1(A), \dots, \operatorname{Col}_{n-k}(A)$ fixes this subspace. Denote the orthogonal projection on the space E by P_E. Since E is independent of $\operatorname{Col}_{n-k+1}(A), \dots, \operatorname{Col}_n(A)$ and these columns are mutually independent as well, it is enough to prove that

$$\begin{aligned}
&\mathbb{P}\big(\operatorname{Col}_j(A) \in \operatorname{span}\big(\operatorname{Col}_i(A), i \in [n-k]\big) \text{ for } j = n-k+1, \dots, n | E\big) \\
&\quad \leq \mathbb{P}\big(\operatorname{Col}_j(A) \in E^{\perp} \text{ for } j = n-k+1, \dots, n | E\big) \\
&\quad = \big(\mathbb{P}\big(P_E \operatorname{Col}_n(A) = 0 | E\big)\big)^k \\
&\quad \leq \exp(-cnk),
\end{aligned}$$

or

$$\mathbb{P}\big(P_E \operatorname{Col}_n(A) = 0 | E\big) \le \exp(-cn).$$

Using Corollary 3.10 with $m = k/2$ and $t = 0$, we obtain

$$\mathbb{P}\big(P_E \operatorname{Col}_n(A) = 0 | E\big) \le C^m \left(\sqrt{m} \exp\left(-C\frac{n}{k}\right)\right)^m \le \exp(-cn),$$

as required. □

With all ingredients in place, we are now ready to prove the main theorem.

PROOF OF THEOREM 1.1. Recall that it is enough to prove Theorem 1.1 under the condition that the entries of A are i.i.d. copies of a random variable satisfying (3.2).

Assume that $\operatorname{rank}(A) \le n - k$. Then there exists a set $J \subset [n]$, $|J| = n - k$ such that $\operatorname{Col}_j(A) \in \operatorname{span}(\operatorname{Col}_i(A), i \in J)$ for all $j \in [n] \setminus J$. Since the number of such sets is

$$\binom{n}{k} \le \exp\left(k \log\left(\frac{en}{k}\right)\right) \ll \exp(ckn),$$

it is enough to show that

$$\mathbb{P}\big(\operatorname{Col}_j(A) \in \operatorname{span}\big(\operatorname{Col}_i(A), i \in J\big) \text{ for all } J \in [n] \setminus J\big) \le \exp(-ckn)$$

for a single set J. As the probability above is the same for all such sets J, without loss of generality assume that $J = [n - k]$.

Consider the $(n - k) \times n$ matrix B with rows $\operatorname{Row}_j(B) = (\operatorname{Col}_j(A))^\top$ for $j \in [n - k]$. Let $E_0 = \operatorname{Ker}(B)$, and denote by P_{E_0} the orthogonal projection onto E_0. Then the condition $\operatorname{Col}_j(A) \in \operatorname{span}(\operatorname{Col}_i(A), i \in [n - k])$ reads $P_{E_0} \operatorname{Col}_j(A) = 0$.

Let τ be the constant appearing in Proposition 4.2, and denote

$$W_0 = \operatorname{Comp}(\tau^2 n, \tau^4).$$

Set $l = k/4$. Lemma 3.3 asserts that at least one of the events described in (1) and (2) of this lemma occurs. Denote these events $\mathcal{E}_{3.3}^{(1)}$ and $\mathcal{E}_{3.3}^{(2)}$, respectively. In view of Proposition 4.2,

$$\mathbb{P}\big(\mathcal{E}_{3.3}^{(1)}\big) \le \exp\left(-c\frac{k}{4}n\right).$$

Here we used only condition (1a) in Lemma 3.3 ignoring condition (1b).

Assume now that $\mathcal{E}_{3.3}^{(2)}$ occurs, and consider the subspace $F \subset E_0$, $\dim(F) = \frac{3}{4}k$ such that $F \cap \operatorname{Comp}(\tau^2 n, \tau^4) = \varnothing$. Let ρ be the constant appearing in Proposition 5.1, and let L be as in (5.1). Set

$$W_1 = \left\{v \in F : \frac{\tau}{8}\sqrt{n} \le \|v\|_2 \le \exp\left(\frac{\rho^2 n}{4L^2}\right) \text{ and } \operatorname{dist}\big(v, \mathbb{Z}^n\big) \le \rho\sqrt{n}\right\}.$$

Applying Lemma 3.3 to W_1 and $l = \frac{k}{4}$, we again conclude that one of the following events occurs:

1. there exist vectors $v_1, \dots, v_{k/4} \in F \cap W_1$ such that
   (a) the $(k/4)$-tuple $(v_1, \dots, v_{k/4})$ is $(\frac{1}{8})$-almost orthogonal, and
   (b) for any $\theta \in \mathbb{R}^{k/4}$ with
   $$\|\theta\|_2 \le \frac{1}{20\sqrt{k/4}},$$
   $\sum_{i=1}^{k/4} \theta_i v_i \notin W_1$

   or
2. there is a subspace $\tilde{F} \subset F$ with $\dim(\tilde{F}) = \frac{k}{2}$ such that $\tilde{F} \cap W_1 = \varnothing$.

Denote these events $\mathcal{V}_{3.3}^{(1)}$ and $\mathcal{V}_{3.3}^{(2)}$, respectively. In view of Proposition 5.1,

$$\mathbb{P}(\mathcal{V}_{3.3}^{(1)}) \le \exp\left(-\frac{k}{4}n\right).$$

Assume now that the event $\mathcal{V}_{3.3}^{(2)}$ occurs. We claim that, in this case,

$$D_{L,\alpha}(\tilde{F}) \ge R := \exp\left(\frac{\rho^2 n}{4L^2}\right).$$

The proof is similar to the argument used in the proof of Lemma 5.4. Let $S : \mathbb{R}^{k/2} \to \mathbb{R}^n$ be an isometric embedding such that $S\mathbb{R}^{k/2} = \tilde{F}$. Then $D_{L,\alpha}(\tilde{F}) = D_{L,\alpha}(S^\top)$. Let $\theta \in \mathbb{R}^{k/2}$ be a vector such that

$$\operatorname{dist}(S\theta, \mathbb{Z}^n) < L\sqrt{\log_+ \frac{\alpha\|\theta\|_2}{L}}.$$

Since

$$S\mathbb{R}^{k/2} \cap S^{n-1} \subset F \cap S^{n-1} \subset \operatorname{Incomp}(\tau^2 n, \tau^4).$$

Lemma 3.11 applied with $U = S$ and $s = \tau^2$ yields

$$\|\theta\|_2 \ge \tau\sqrt{n}/2.$$

On the other hand, if $\|\theta\|_2 \le R$, then

$$L\sqrt{\log_+ \frac{\alpha\|\theta\|_2}{L}} \le \rho\sqrt{n},$$

and, therefore, $\operatorname{dist}(S\theta, \mathbb{Z}^n) < \rho\sqrt{n}$. Since $\tilde{F} = S\mathbb{R}^{k/2} \cap W_1 = \varnothing$, this implies that

$$\|\theta\|_2 = \|S\theta\|_2 > R = \exp\left(\frac{\rho^2 n}{4L^2}\right),$$

thus proving our claim and checking the assumption of Lemma 6.1.

Finally,

$$\begin{aligned}
&\mathbb{P}\big(\operatorname{Col}_j(A) \in \operatorname{span}\big(\operatorname{Col}_i(A), i \in [n-k]\big) \text{ for } j = n-k+1, \dots, n\big) \\
&\quad \le 2\exp\left(-\frac{k}{4}n\right) \\
&\qquad + \mathbb{P}\big(\operatorname{Col}_j(A) \in \operatorname{span}\big(\operatorname{Col}_i(A), i \in [n-k]\big) \text{ for } j = n-k+1, \dots, n \text{ and } \mathcal{V}_{3.3}^{(2)}\big).
\end{aligned}$$

Lemma 6.1 shows that the last probability does not exceed $\exp(-c'(k/2)n)$. The proof is complete. □

After the theorem is proved, we can derive an application to the question of Feige and Lellouche.

LEMMA 6.2. *Let $q \in (0,1)$, and $m, n \in \mathbb{N}$ be numbers such that*

$$m \le n \le \exp(C_q'\sqrt{m}).$$

Let A be an $m \times n$ matrix with independent Bernoulli(q) *entries. Then with probability at least $1 - \exp(-m\log n)$, all $m \times m$ submatrices of A have rank greater than $m - C_q \log n$.*

Furthermore, if $n \ge m^2$, then with probability at least $1 - \exp(-cm)$, there exists an $m \times m$ submatrix $A|_S$ of A with $|S| = m$ such that

$$\operatorname{rank}(A|_S) \le m - c_q \log n.$$

The constants $C_q > c_q > 0$ above can depend on q.

PROOF. The entries of A are i.i.d. sub-Gaussian random variables, so Theorem 1.1 applies to an $m \times m$ submatrix of A as long as $k \le c\sqrt{m}$. In view of the assumption of the lemma, the last inequality holds if we take $k = C_q \log n$. Combining Theorem 1.1 with the union bound, we obtain

$$\begin{aligned}
&\mathbb{P}\big(\exists S \subset [n] : |S| = m \text{ and } \operatorname{rank}(A|_S) \le n - C_q \log n\big) \\
&\quad \le \binom{n}{m} \exp(-c'm \cdot C_q \log n) \le \exp\Big(m \log\Big(\frac{en}{m}\Big) - c'm \cdot C_q \log n\Big) \\
&\quad \le \exp(-m \log n)
\end{aligned}$$

if C_q is chosen sufficiently large.

To prove the second part of the lemma, take $k < m$, and define a random subset $J \subset [n]$ by

$$J = \{j \in [n] : a_{1,j} = \cdots = a_{k,j} = 1\}.$$

Then for any $j \in [n]$,

$$\mathbb{P}(j \in J) = q^k,$$

and these events are independent for different $j \in [n]$. Take $k = c_q \log n$, and choose c_q so that $nq^k \ge 10m$. Using Chernoff's inequality, we obtain

$$\mathbb{P}\big(|J| \ge m\big) = 1 - \exp(-cm).$$

On the other hand, $\operatorname{rank}(A|_J) \le n - k$ since this matrix contains k identical rows. The lemma is proved. □

Acknowledgments. This work was performed when the author held an Erna and Jakob Michael Visiting Professorship at the Department of Mathematics at the Weizmann Institute of Science. The author thanks the Weizmann Institute for its hospitality. He is especially grateful to Ofer Zeitouni for bringing this problem to his attention and numerous helpful discussions.

Funding. Research supported in part by NSF Grant DMS 2054408 and a fellowship from the Simons Foundation.

REFERENCES

[1] BASAK, A. and RUDELSON, M. (2017). Invertibility of sparse non-Hermitian matrices. *Adv. Math.* **310** 426–483. MR3620692 https://doi.org/10.1016/j.aim.2017.02.009

[2] CAMPOS, M., JENSSEN, M., MICHELEN, M. and SAHASRABUDHE, J. (2024). The least singular value of a random symmetric matrix. *Forum Math. Pi* **12** Paper No. e3, 69. MR4695501 https://doi.org/10.1017/fmp.2023.29

[3] FEIGE, U. and LELLOUCHE, A. Quantitative group testing and the rank of random matrices. Available at arXiv:2006.09074.

[4] JAIN, V., SAH, A. and SAWHNEY, M. (2021). Singularity of discrete random matrices. *Geom. Funct. Anal.* **31** 1160–1218. MR4356701 https://doi.org/10.1007/s00039-021-00580-6

[5] JAIN, V., SAH, A. and SAWHNEY, M. (2022). Rank deficiency of random matrices. *Electron. Commun. Probab.* **27** Paper No. 14, 9. MR4386039 https://doi.org/10.15388/namc.2022.27.25203

[6] 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

[7] KOMLÓS, J. (1967). On the determinant of (0, 1) matrices. *Studia Sci. Math. Hungar.* **2** 7–21. MR0221962

[8] LIVSHYTS, G. V. (2021). The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding. *J. Anal. Math.* **145** 257–306. MR4361906 https://doi.org/10.1007/s11854-021-0183-2

[9] LIVSHYTS, G. V., TIKHOMIROV, K. and VERSHYNIN, R. (2021). The smallest singular value of inhomogeneous square random matrices. *Ann. Probab.* **49** 1286–1309. MR4255145 https://doi.org/10.1214/20-aop1481

[10] NGUYEN, H. H. (2018). Random matrices: Overcrowding estimates for the spectrum. *J. Funct. Anal.* **275** 2197–2224. MR3841540 https://doi.org/10.1016/j.jfa.2018.06.010

[11] 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

[12] 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

[13] RUDELSON, M. (2014). Recent developments in non-asymptotic theory of random matrices. In *Modern Aspects of Random Matrix Theory. Proc. Sympos. Appl. Math.* **72** 83–120. Amer. Math. Soc., Providence, RI. MR3288229 https://doi.org/10.1090/psapm/072/00616

[14] 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

[15] RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. MR3125258 https://doi.org/10.1214/ECP.v18-2865

[16] RUDELSON, M. and VERSHYNIN, R. (2016). No-gaps delocalization for general random matrices. *Geom. Funct. Anal.* **26** 1716–1776. MR3579707 https://doi.org/10.1007/s00039-016-0389-0

[17] 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

[18] VERSHYNIN, R. (2014). Invertibility of symmetric random matrices. *Random Structures Algorithms* **44** 135–182. MR3158627 https://doi.org/10.1002/rsa.20429

[19] VERSHYNIN, R. (2018). *High-Dimensional Probability*: *An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596