# Invertibility of Random Matrices

Yiting Wang yiting.wang@ist.ac.at

March 28, 2023

**Abstract**

The study of discrete random matrices has attracted a large amount of interest in recent years and gradually formed an area named combinatorial random matrices. Contrary to the classical study of random matrices, the approach in this area are mostly combinatorial. This report is the outcome of a rotation project supervised by Prof. László Erdös. It focuses on one particular problem: the probability that a random sign matrix is singular. We will first explain the paper of Rudelson and Vershynin [RV08] in detail, which lays the ground for how to attack this question. Then, we will sketch the breakthrough work by Tikhomirov [Tik20], which resolved this longstanding open problem.

## Contents

# 1 Introduction

Let $A$ be an $n \times n$ matrix, whose entries $A_{ij}$ are identically distributed and independent (iid.) random variables, which take value 1 with probability $1/2$ and $-1$ with probability $1/2$. What is the probability that $A$ is singular?

Let $p_n$ denote this probability. Notice that $p_n \geq 2^{-n}$, since with probability $2^{-n+1}$, the first column and the second column are the same (up to a sign), which implies that $A$ is singular. Tikhomirov shows that this is essentially tight:

**Theorem 1.1** ([Tik20]). *$p_n = (1/2 + o(1))^n$ as $n \to \infty$.*

Despite the simplicity of the statement, this problem is highly non-trivial. This was an important open problem and has a long history: Komlòs [Kom67] first proved that $p_n = o(1)$. Much later, Kahn, Komlós and Szemerédi [KKS95] showed that $p_n \leq 0.999^n$. The base is subsequently improved to $0.939, (3/4 + o(1))$ by Tao and Vu [TV08, TV09] and to $(1/\sqrt{2} + o(1))$ by Bourgain, Vu, and Wood [BVW10].

More precisely, what Tikhomirov proved was:

**Theorem 1.2.** *For every $\varepsilon > 0$, the least singular value of $A$ satisfies:*

$$\mathbb{P}(s_n(A) \leq \varepsilon n^{-1/2}) \leq (\frac{1}{2} + \varepsilon)^n + C_{1.2}\varepsilon\,,$$

*where $C_{1.2}$ is a constant depending only on $\varepsilon$.*

Tikhomirov's work builds upon the strategy the work of Rudelson and Vershynin [RV08], who proved a much more general but less precise result. They showed that if every entry is iid subgaussian random variable with bounded subgaussian moment, then the singularity probability is exponentially small, with some base $c \in (0, 1)$. See Theorem 3.1 for a precise statement.

This report aims to demonstrate the proof ideas in [RV08, Tik20]. Due to the length constraint, not all details are included. In particular, we will not prove the main technical results from both papers, but only give some explanations of them. Nonetheless, assuming these technical results, the report gives a complete description of the structure of the proof.

The report is structured as follows: In Section 2, we list out definitions and standard facts and fix our notations. In Section 3, we explain the proof for the subgaussian case, based on the paper of Rudelson and Vershynin [RV08]. At last, in Section 4, we explain Tikhomirov's proof [Tik20], focusing on the new ideas involved.

# 2 Preliminaries

**Singular Values** Let $A$ be an $n \times n$ matrix with real or complex entries. The singular values of $A$ are defined to be the eigenvalues of $\sqrt{A^*A}$, arranged in non-increasing order (i.e. $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A)$). A matrix is singular, if and only if its least singular value equals to 0. The variation characterization of singular values states that:

$$s_i(A) = \inf_{\substack{U \text{ subspace} \\ \dim(U)=n-i+1}} \sup_{\substack{x \in U \\ \|x\|_2=1}} \|Ax\|_2 = \sup_{\substack{U \text{ subspace} \\ \dim(U)=i}} \inf_{\substack{x \in U \\ \|x\|_2=1}} \|Ax\|_2.$$

In particular, we are interested in the smallest and largest singular values, which can be expressed as

$$\|A\| = s_1(A) = \sup_{x \in S^{n-1}} \|Ax\|_2, \quad s_n(A) = \inf_{x \in S^{n-1}} \|Ax\|_2.$$

**Subgaussian** A real random variable $\xi$ is called subgaussian, if for some constant $B > 0$,

$$\mathbb{P}(|\xi| > t) \leq 2\exp(-t^2/B^2) \quad \forall t > 0\,.$$

The minimal $B$ is called the subgaussian moment of $\xi$. This tail condition can be translated to a moment condition, up to constants: $\xi$ is subgaussian if there exists an absolute constant $C$ and some constant $B > 0$, such that

$$\mathbb{E}|\xi|^p \leq C(B\sqrt{p})^p \quad \forall p \geq 1\,.$$

Subgaussian random variables generalize the gaussian-like behavior to a larger class of random variables. In particular, any bounded random variable is subgaussian with some $B$.

**Notations** For convenience, we assume in this section and Section 3 that $A$ is an $n \times n$ real matrix, where $A_{ij}$ are iid. subgaussian random variables with variance at least 1 and subgaussian moment at most $B$. In Section 4, we assume $A$ is an $n \times n$ matrix, where $A_{ij} = 1$ with probability $1/2$ and $-1$ with probability $1/2$. Most of the lemmas and theorems stated below hold in more general settings, either weakening the iid assumption or the distribution assumption. All logarithms are base 2.

**Largest Singular Value** We will need the following standard fact about the largest singular value. See for instance Theorem 4.4.5 from [Ver18].

**Lemma 2.1** (Largest Singular Value)**.**

$$\mathbb{P}(\|A\| \geq C_{2.1}\sqrt{n}) \leq 2\exp(-n)\,,$$

*where $C_{2.1}$ is a constant depending only on $B$.*

**$\varepsilon$-net for $S^{n-1}$** Let $X \subset \mathbb{R}^n$. For $\varepsilon > 0$, a subset $\mathcal{N} \subseteq S$ is called an $\varepsilon$-net of $X$, if for every $x \in X$, there exists $y \in \mathcal{N}$ such that $\|x - y\|_2 \leq \varepsilon$. If $X$ is not finite, then we cannot directly union bound over it. However, we can apply union bound for the epsilon net $\mathcal{N}$ and generalize the result to all $x \in X$ by using the proximity in euclidean distance. Here is a standard result when $X = S^{n-1}$:

**Proposition 2.2** ($\varepsilon$-net for $S^{n-1}$)**.** *Let $X \subset S^{n-1}$ and $\varepsilon > 0$, then $|\mathcal{N}| \leq (\frac{3}{\varepsilon})^n$.*

See for instance Corollary 4.2.13 from [Ver18].

### Small Ball Probability

**Definition 2.3** (Small Ball Probability)**.** *Let $a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ and $\varepsilon > 0$. Let $\xi_1, \ldots, \xi_n$ be iid random variables with variance at least 1 and subgaussian moment bounded by $B$. Then, define*

$$p_\varepsilon(a) \coloneqq \sup_{t\in\mathbb{R}} \mathbb{P}(|\sum_{i=1}^n a_i\xi_i - t| \leq \varepsilon)\,.$$

The small ball probability characterizes how anti-concentrated the random sum is, because it measures the maximum probability mass a random sum falls inside an

interval of length $2\varepsilon$. The smaller is the small ball probability, the less concentrated is the random sum.

For a fixed index set $S \subseteq [n]$ and a vector $x \in \mathbb{R}^n$, let $x^S \in \mathbb{R}^{|S|}$ denote the restriction of $x$ to $S$. That is, $(x^S)_i = x_{s_i}$ for all $1 \le i \le |S|$, where $S = \{s_1 < s_2 < \ldots, s_{|S|}\}$. Moreover, let $A^S$ denote the submatrix of $A$ consisting of columns whose indices correspond to $S$. We have the following restriction lemma, which says the small ball probability can only increase, if we restrict to a smaller support.

**Proposition 2.4** (Restriction). *For any $a \in \mathbb{R}^n$ and any $S \subseteq [n]$ and any $\varepsilon \ge 0$, we have*

$$p_\varepsilon(a) \le p_\varepsilon(a^S)\,.$$

This follows immediately after we condition on $\xi_i$ for $i \in [n]\backslash S$.

**Tensorization Lemma** For a random vector with independent entries, if we have an anti-concentration estimate for each entry separately, then we have an estimate of its $\ell_2$ norm. The name comes from that we tensorize a lower dimensional estimate to a higher dimensional one.

**Lemma 2.5** (Tensorization). *Let $\xi_1, \ldots, \xi_n$ be independent non-negative random variables. Let $K, \varepsilon_0 \ge 0$. Assume for each $k$, we have $\mathbb{P}(\xi_k < \varepsilon) \le K\varepsilon$ for all $\varepsilon \ge \varepsilon_0$, then there exists some absolute constant $C$ such that*

$$\mathbb{P}(\sum_{k=1}^{n} \xi_k^2 < \varepsilon^2 n) \le (CK\varepsilon)^n \quad \text{for all } \varepsilon \ge \varepsilon_0\,.$$

# 3 Subgaussian Case

In this section, we will prove the following theorem [RV08]:

**Theorem 3.1** (Invertibility: Subgaussian Case). *Let $A$ be an $n \times n$ real matrix, where $A_{ij}$ are iid subgaussian random variables with variance at least 1 and subgaussian moment at most $B$. For every $\varepsilon > 0$, we have*

$$\mathbb{P}(s_n(A) \le \varepsilon n^{-1/2}) \le C_{3.1}\varepsilon + c_{3.1}^n\,,$$

*where $C_{3.1} \ge 1, c_{3.1} \in (0,1)$ are constants depending only on $B$.*

Here are several remarks:

1. If we let $\varepsilon \to 0$, we conclude that the singular probability of such random matrix $A$ is exponentially small. This in particular implies the singularity probability of a random sign matrix is exponentially small. However, this proof can only give $c_{3.1}$ up to constant factors, so we cannot deduce the exact base $(1/2 + \varepsilon)$ immediately from this proof.
2. In general, both the linear term $C_{3.1}\varepsilon$ and the exponential term $c_{3.1}^n$ are needed. On one hand, it is known that for random gaussian matrices, this upper bound should be $C\varepsilon$ [Ede88]. On the other hand, for random sign matrices, this upper bound should be $c^n$ [KKS95].

## 3.1 General Strategy

Recall that $s_n(A) = \inf_{x\in\mathbb{R}^n:\|x\|_2=1} \|Ax\|_2$. Thus, it suffices to prove a lower bound of $\|Ax\|_2$ for all $x \in S^{n-1}$. The first step is to divide the set of unit vectors into compressible/incompressible vectors based on their structures. Recall that the support of a vector $v \in \mathbb{R}^n$ is defined as $\mathrm{supp}(v) = \{i \in [n] : v_i \neq 0\}$.

**Definition 3.2** ($\delta-$sparse Vectors)**.** *Let $\delta \in (0,1)$. A vector $v \in \mathbb{R}^n$ is called $\delta$-sparse if $|\mathrm{supp}(v)| \leq \delta n$. We denote the set of $\delta$-sparse vectors as* $\mathrm{Sparse}(\delta)$.

**Definition 3.3** ($(\delta,\rho)$-Compressible/Incompressible Vectors)**.** *Let $\rho, \delta \in (0,1)$. A unit vector $v \in S^{n-1}$ is $(\delta,\rho)-$compressible, if there exists a $\delta$-sparse vector $w$ such that $\|v-w\|_2 \leq \rho$. Otherwise, it is called $(\delta,\rho)-$incompressible. The set of $(\delta,\rho)$ compressible vectors is denoted as* $\mathrm{Comp}(\delta,\rho)$ *and incompressible vectors as* $\mathrm{Incomp}(\delta,\rho)$.

By definition, an incompressible vector is not close to any sparse vector. In other word, the mass of it tends to spread out instead of concentrating on a few coordinates. This intuition can be made precise as follows:

**Proposition 3.4** (Incompressible implies spread)**.** *Let $a \in \mathrm{Incomp}(\delta,\rho)$. Then, there exists $S \subset [n]$ of cardinality at least $\frac{\rho^2\delta}{2}n$ such that*

$$\frac{\rho}{\sqrt{2}}\frac{1}{\sqrt{n}} \leq |a_i| \leq \frac{1}{\sqrt{\delta}}\frac{1}{\sqrt{n}} \quad \forall i \in S\,.$$

*Proof.* Let $S_1 = \{i \in [n] : |a_i| < \frac{\rho}{\sqrt{2}}\frac{1}{\sqrt{n}}\}$ and $S_2 = \{i \in [n] : |a_i| > \frac{1}{\sqrt{\delta}}\frac{1}{\sqrt{n}}\}$. The set $S$ in the statement equals $[n]\backslash(S_1 \cup S_2)$. Since $\|a\|_2 = 1$, we know $|S_2| \leq \delta n$. Using the definition of incompressiblity, we know that

$$\rho^2 \leq \|a^{[n]\backslash S_2}\|_2^2 \leq n \cdot \frac{\rho^2}{2n} + |S|\frac{1}{\delta n}\,,$$

which implies $|S| \geq \frac{\rho^2\delta}{2}n$. □

We call $S$ the *spread part* of $a$. The proposition tells us each incompressible vector has a large spread part. Moreover, we define $\widehat{a} := (\sqrt{n}a)^S$.

By definition, $S^{n-1} = \mathrm{Comp}(\delta,\rho) \sqcup \mathrm{Incomp}(\delta,\rho)$. Thus,

$$\begin{aligned}\mathbb{P}(s_n(A) \leq \varepsilon n^{-1/2}) &= \mathbb{P}(\inf_{x\in S^{n-1}} \|Ax\|_2 \leq \varepsilon n^{-1/2}) \\ &\leq \mathbb{P}(\inf_{x\in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \leq \varepsilon n^{-1/2}) + \mathbb{P}(\inf_{x\in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 \leq \varepsilon n^{-1/2}).\end{aligned} \tag{1}$$

It remains to deal with compressible and incompressible vectors separately. For compressible vectors, we use the set of sparse vectors as a $\rho$-net and union bound. For incompressible vectors, we reduce the problem to a Littlewood-Offord type question and then use a strong small ball probability estimate (that takes into account of the structure of the vector).

By Lemma 2.1, we fix some constant $K \geq C_{2.1}$, so that

$$\mathbb{P}(\|A\| \leq K\sqrt{n}) \geq 1 - 2\exp(-n)\,.$$

We may assume the condition $\|A\| \leq K\sqrt{n}$ holds, as $\exp(-n)$ cannot affect Theorem 3.1. The subgaussian moment $B$ and the upper bound on the largest singular value $K$ will be treated as fixed parameters below.

## 3.2 Compressible Vectors

The goal of this section is to show

**Lemma 3.5** (Invertibility: Compressible Vectors)**.** *There exists $\delta, \rho, C_{3.5}, C'_{3.5}$ that depend only on $B, K$ such that*

$$\mathbb{P}(\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \le C_{3.5}\sqrt{n}) \le \exp(-C'_{3.5}n)\,.$$

Notice that this result is much stronger than required for Theorem 3.1. For any $\varepsilon > 0$, if $n$ is sufficiently large we can ensure

$$\mathbb{P}(\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \le \varepsilon n^{-1/2}) \le \mathbb{P}(\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \le C_{3.5}\sqrt{n}) \le \exp(-C'_{3.5}n)\,.$$

Before proving the theorem, we need an auxiliary result that says the least singular value of a rectangular matrix is large:

**Lemma 3.6** (Least Singular Value of Rectangular Matrices [LPRTJ05])**.** *Let $G$ be an $n \times k$ submatrix of $A$. Then, there exist $C_{3.6}, C'_{3.6}$ and $\delta_{3.6} \in (0,1)$ depending only on $B, K$ such that if $k \le \delta_{3.6} n$, then*

$$\mathbb{P}(\inf_{x \in S^{k-1}} \|Gx\|_2 \le C_{3.6}\sqrt{n}) \le \exp(-C'_{3.6}n)\,. \tag{2}$$

*Proof of Lemma 3.5.* The proof is a standard $\varepsilon$-net argument. We first prove the result for sparse vectors. Let $\delta \le \delta_{3.6}$ to be fixed and let $k = \lceil \delta n \rceil$. Slightly abusing notations, we define $\binom{[n]}{k}$ to be all size $k$ subsets of $[n]$.

$$\begin{aligned}
\mathbb{P}(\inf_{y \in \mathrm{Sparse}(\delta) \cap S^{n-1}} \|Ay\|_2 \le C_{3.6}\sqrt{n}) &\le \mathbb{P}(\exists S \in \binom{[n]}{k} : \inf_{y \in S^{n-1}} \|A^S y^S\|_2 \le C_{3.6}\sqrt{n}) \\
&\le \mathbb{P}(\exists S \in \binom{[n]}{k} : \inf_{z \in S^{k-1}} \|A^S z\|_2 \le C_{3.6}\sqrt{n}) \\
&\le \binom{n}{k} \exp(-C'_{3.6}n) \le \exp(-C'_{3.5}n) \quad \text{(using (2))}\,,
\end{aligned}$$

if we select a small enough $\delta$, depending on $C'_{3.6}$ and $\delta_{3.6}$.

By definition, we know that $\mathrm{Sparse}(\delta)$ forms a $\rho$-net for $\mathrm{Comp}(\delta,\rho)$. Thus, for every $x \in \mathrm{Comp}(\delta,\rho)$, we can find $y_x \in \mathrm{Sparse}(\delta)$ and $z_x \in \mathbb{R}^n$ such that $x = y_x + z_x$ and $\|z_x\|_2 \le \rho$. Using triangle inequality and the definition of the $\|A\|$, we conclude $\|y_x\|_2 \ge 1 - \rho$ and

$$\|Ay_x\|_2 = \|Ax - Az_x\|_2 \le \|Ax\|_2 + \|A\|\|z_x\|_2 \le \|Ax\|_2 + K\rho\sqrt{n}.$$

Thus, if we select appropriate $\rho$ and $C_{3.5}$ depending only on $C_{3.6}$, then

$$\begin{aligned}
\mathbb{P}(\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \le C_{3.5}\sqrt{n}) &\le \mathbb{P}(\inf_{y_x \in \mathrm{Sparse}(\delta)} \|Ay_x\|_2 \le (C_{3.5} + K\rho)\sqrt{n}) \\
&\le \mathbb{P}(\inf_{y'_x \in \mathrm{Sparse}(\delta) \cap S^{n-1}} \|Ay'_x\|_2 \le \frac{1}{(1-\rho)}(C_{3.5} + K\rho)\sqrt{n}) \\
&\le \mathbb{P}(\inf_{y'_x \in \mathrm{Sparse}(\delta) \cap S^{n-1}} \|Ay'_x\|_2 \le C_{3.6}\sqrt{n}) \le \exp(-C'_{3.5}n)\,,
\end{aligned}$$

which finishes the proof. □

## 3.3 Incompressible Vectors

The goal of this section is to show

**Lemma 3.7** (Invertibility: Incompressible Vectors)**.** *Let $\delta, \rho, \varepsilon > 0$. There exists $C_{3.7}, C'_{3.7}$ that depends only on $B, K$ such that*

$$\mathbb{P}(\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 \le \varepsilon n^{-1/2}) \le C_{3.7}\varepsilon + \exp(-C'_{3.7} n)\,.$$

Assuming this is true, we can prove Theorem 3.1:

*Proof of Theorem 3.1.* Fix $K \ge C_{2.1}$ that depends only on $B$ such that $\mathbb{P}(\|A\| \ge K\sqrt{n}) \le 2\exp(-n)$. Fix arbitrary $\varepsilon > 0$. Lemma 3.5 shows there exist $\delta, \rho, C_{3.5}, C'_{3.5}$ depending only on $B$ such that

$$\mathbb{P}(\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \le \varepsilon n^{-1/2}) \le \exp(-C'_{3.5} n)\,.$$

Lemma 3.7 shows there exist $C_{3.7}, C'_{3.7}$ that depend only on $B$ such that

$$\mathbb{P}(\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 \le \varepsilon n^{-1/2}) \le C_{3.7}\varepsilon + \exp(-C'_{3.7} n)\,.$$

Putting them together, we conclude

$$\mathbb{P}(s_n(A) \le \varepsilon n^{-1/2}) \le C_{3.7}\varepsilon + \exp(-C'_{3.7} n) + \exp(-C'_{3.5} n) + 2\exp(-n) \le C_{3.1}\varepsilon + c_{3.1}^n\,,$$

for appropriately chosen $C_{3.1}, c_{3.1}$ that depend only on $B$ and sufficiently large $n$. □

### 3.3.1 Reduction to Littlewood-Offord Problem

We want to reduce the problem of estimating $\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2$ in Lemma 3.7 to a problem of estimating the lower bound of $|\langle X, Y\rangle|$, where both $X, Y$ are random vectors satisfying certain properties.

Let $X_1, \dots, X_n$ denote the columns of $A$ and let $H_k = \mathrm{span}(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$. We define the distance

$$\mathrm{dist}(X_k, H_k) := \inf_{x \in H_k} \|X_k - x\|_2 = \sup_{x \in H_k^\perp, \|x\|_2 = 1} |\langle X_k, x\rangle|\,.$$

**Lemma 3.8** (Reduction to Distance Problem)**.** *For every $\delta, \rho \in (0,1)$ and $\varepsilon > 0$, we have*

$$\mathbb{P}(\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 < \varepsilon n^{-1/2}) \le \frac{1}{\delta}\mathbb{P}\Big(\mathrm{dist}(X_n, H_n) \le \frac{\varepsilon}{\rho}\Big)\,.$$

*Proof.* Let $S = \{i \in [n] : |x_i| \ge \rho n^{-1/2}\}$. From the definition of the incompressible vectors, we conclude that $|S| \ge \delta n$, since otherwise $x^S$ is a sparse vector and $\|x^{[n]\setminus S}\|_2 < \rho$, a contradiction.

Let $T = \{i \in [n] : \mathrm{dist}(X_i, H_i) \ge \frac{\varepsilon}{\rho}\}$. If $|T| > (1-\delta)n$, then by pigeonhole principle, there exists $i \in S \cap T$, meaning that $|x_i| \ge \rho n^{-1/2}$ and $\mathrm{dist}(X_i, H_i) \ge \frac{\varepsilon}{\rho}$. However,

$$\|Ax\|_2 = \sup_{y \in \mathbb{R}^n : \|y\|_2 = 1} |\langle Ax, y\rangle| \ge \max_{i \in [n]} \mathrm{dist}(Ax, H_i) = \max_{i \in [n]} |x_i| \mathrm{dist}(X_i, H_i) \ge \varepsilon n^{-1/2}\,.$$

This contradicts the condition $\inf_{x\in\mathrm{Incomp}(\delta,\rho)}\|Ax\|_2<\varepsilon n^{-1/2})$. In other word,

$$\inf_{x\in\mathrm{Incomp}(\delta,\rho)}\|Ax\|_2<\varepsilon n^{-1/2})\implies |T|\le(1-\delta)n\implies\sum_{i=1}^{n}\mathbb{1}(\mathrm{dist}(X_i,H_i)\le\frac{\varepsilon}{\rho})\ge\delta n\,.$$

Apply Markov's inequality and use the property that $\mathbb{P}(\mathrm{dist}(X_i,H_i)\le\frac{\varepsilon}{\rho})$ is the same for all $i$, the proof is finished. □

Note that $\mathrm{dist}(X_n,H_n)=\sup_{x\in H_n^{\perp},\|x\|_2=1}|\langle X_n,x\rangle|\ge|\langle X_n,X^*\rangle|$ for any unit vector $X^*\in H_n^{\perp}$. Fix arbitrary such $X^*$ and call it a random normal. Our problem reduces to prove a small ball probability estimate for a random normal:

$$\mathbb{P}(\inf_{x\in\mathrm{Incomp}(\delta,\rho)}\|Ax\|_2<\varepsilon n^{-1/2})\le\frac{1}{\delta}\mathbb{P}(|\langle X_n,X^*\rangle|\le\frac{\varepsilon}{\rho})=\frac{1}{\delta}p_{\varepsilon/\rho}(X^*)\,,$$

where $p_{\varepsilon/\rho}(X^*)$ was defined in Definition 2.3.

### 3.3.2 Small Ball Probability Estimate

A weak small ball probability estimate can be derived immediately from the Berry-Esséen central limit theorem (c.f. [RV08] Corollary 2.9). However, this only gives polynomial dependency on $n$ instead of exponential dependency in Theorem 3.1.

In order to prove strong small ball probability estimate, Rudelson and Vershynin [RV08] introduced the notion of essential least common denominator (essential LCD). Informally speaking, essential LCD measures the arithmetic structuredness of a vector.

**Definition 3.9** (Essential LCD)**.** *Let $\alpha\in(0,1),\kappa\in(0,n)$. For a vector $a\in\mathbb{R}^n$, its essential LCD, denoted by, $D_{\alpha,\kappa}(a)$ is defined to be the infimum of $t>0$ such that all except $\kappa$ coordinates of $ta$ are of distance at most $\alpha$ to nonzero integers. That is,*

$$D_{\alpha,\kappa}(a):=\inf\{t>0:|\{i\in[n]:[(ta)_i]>\alpha\}|=\kappa\}\,,$$

*where $[x]:=\min\{x-\lfloor x\rfloor,1-x+\lfloor x\rfloor\}$.*

It turns out that essential LCD efficiently controls the small ball probability for vectors with bounded coefficients:

**Theorem 3.10** (Small Ball Probability)**.** *Let $\xi_1,\cdots,\xi_n$ be iid. subgaussian random variables with variance at least 1 and subgaussian moment at most $B$. Let $a\in\mathbb{R}^n$ be a vector satisfying $K_1\le|a_i|\le K_2$ for some constants $K_1,K_2$. Then, for every $\alpha\in(0,1),\kappa\in(0,n)$ and $\varepsilon>0$, we have*

$$p_\varepsilon(a)\le\frac{C_{3.10}}{\sqrt{\kappa}}(\varepsilon+\frac{1}{D_{\alpha,\kappa}(a)})+C_{3.10}\exp(-c_{3.10}\alpha^2\kappa)\,,$$

*where $C_{3.10},c_{3.10}$ are constants depending only on $B,K_1,K_2$.*

This is the main technical contribution of [RV08]. It gives fairly good estimate when coordinates are bounded by constants. The proof reduces the problem to studying the Lebesgue measure of the level set of a certain function and showing that the essential LCD efficiently controls it. The proof is outside the range of this report. Interested readers should refer to the Section 4 of [RV08].

We now prove a special case of Theorem 3.10 where $a$ is a random normal. For this purpose, we first show a random normal is very likely to be incompressible and then restrict it to its spread part.

**Proposition 3.11** (Random normal is incompressible). *There exists constant $C_{3.11}$ depeneding only on $B, K$ such that*

$$\mathbb{P}(X^* \in \mathrm{Comp}(\delta, \rho)) \leq \exp(-C_{3.11} n)\,.$$

*Proof.* By definition, $X^*$ is orthogonal to $X_1, \ldots, X_{n-1}$. We can therefore reuse the proof of Lemma 3.5, with $n$ replaced by $n-1$. □

**Corollary 3.12** (Small Ball Probability for Random Normal). *There exists $\delta, \rho, C_{3.11}$ that depend only on $B, K$ such that for any $\alpha \in (0,1), \kappa \in (0,n)$ and $\varepsilon > 0$, with probability at least $1 - \exp(-C_{3.11} n)$, we have*

$$p_\varepsilon(X^*) \leq \frac{C_{3.10}}{\sqrt{\kappa}}\left(\sqrt{n}\varepsilon + \frac{1}{D_{\alpha,\kappa}(\widehat{X^*})}\right) + C_{3.10}\exp(-c_{3.10}\alpha^2\kappa)\,,$$

*where $C_{3.10}, c_{3.10}$ depend only on $B, K$.*

*Proof.* Proposition 3.11 ensures the random normal $X^*$ is spread with probability $1-\exp(-C_{3.11}n)$ and Proposition 3.4 ensures it has a large spread part. The restriction proposition (Proposition 2.4) tells us

$$p_\varepsilon(X^*) = p_{\sqrt{n}\varepsilon}(\sqrt{n}X^*) \leq p_{\sqrt{n}\varepsilon}(\widehat{X^*})\,,$$

where $\widehat{X^*}$ is the restriction of $\sqrt{n}X^*$ to only its spread part. At last, we use small ball probability estimate (Theorem 3.10), with $K_1 = \frac{\rho}{\sqrt{2}}$ and $K_2 = \frac{1}{\sqrt{\delta}}$. □

### 3.3.3 Proof of Lemma 3.7

The remaining step is to show that the essential LCD of a random normal $X^*$ is exponentially large with very high probability, which gives a strong enough small ball probability estimate to finish the proof. To prove this, we will dyadically partition $\mathrm{Incomp}(\delta, \rho)$ based on the essential LCD and show that $X^*$ does not belong to any subset of $\mathrm{Incomp}(\delta, \rho)$ whose essential LCD is below an exponential order.

**Lemma 3.13** (Random Normal has Large Essential LCD). *There exist $\alpha \in (0, 1/2), \kappa \in (0, n), c_{3.13}, c'_{3.13}$ depending only on $B$ and $K$ such that*

$$\mathbb{P}(D_{\alpha,\kappa}(\widehat{X^*}) \leq e^{c_{3.13}n}) \leq \exp(-c'_{3.13} n)\,.$$

We will choose $\alpha \in (0, 1/2)$ later. Let $K_1 = \frac{\rho}{\sqrt{2}}, K_2 = \frac{1}{\sqrt{\delta}}$ and $n_0 = \frac{\rho^2\delta n}{2}$. Note that the definition of the essential LCD implies $D_{\alpha,n_0/2}(\widehat{X^*}) \geq \frac{1-\alpha}{K_2} > \frac{1}{2K_2}$. Let $D_0 := \frac{1}{2K_2}$ denote this lower bound.

For any $D_0 \leq D \leq \exp(c_{3.13}n)$, define the level set $S_D := \{x \in \mathrm{Incomp}(\delta, \rho) : D \leq D_{\alpha,n_0/2}(\widehat{X^*}) \leq 2D\}$. We now show that on $S_D$, there is a $4\alpha/D$ net of "small" size. The saving on the exponent turns out to be crucial for the theorem.

**Lemma 3.14** (Small net for $S_D$). *There exist $\alpha_0 \in (0,1), c_{3.14}, C_{3.14}$ depending only on $B$ and $K$ such that for $\alpha \leq \alpha_0$ and $D \geq D_0$, there exists a $(4\alpha/D)$-net for $S_D$, with cardinality at most $(\frac{C_{3.14}D}{\alpha^{1-c_{3.14}}})^n$.*

*Proof.* Fix arbitrary $x \in S_D$. Recall that $x \in S_D$ means $D \le D_{\alpha,n_0/2}(\hat{x}) \le 2D$. Combining with the definition of LCD (Definition 3.9), we conclude that there exists an integer vector $p \in Z^{\mathrm{supp}(\hat{x})}$ such that $|p_i - D_{\alpha,n_0/2}(\hat{x})\hat{x}_i| \le \alpha$ for $n_0 - n_0/2 = n_0/2$ coordinates.

For other coordinates, choose the closest one from $\alpha\mathbb{Z}$. By construction, this new vector $q$ (as an extension of $p$), satisfies $|q_i - \sqrt{n}D_{\alpha,n_0/2}(\hat{x})x_i| \le \alpha$ for all $i \in [n]$ and

$$q \in \mathcal{Q} \coloneqq \bigcup_{\substack{S\subseteq[n]\\|S|=\frac{n_0}{2}}} \mathbb{Z}^S \oplus \alpha\mathbb{Z}^{[n]\setminus S}\,.$$

Then, by summing up coordinates, we know $\|\sqrt{n}D(\hat{x})x - q\|_2 \le \alpha\sqrt{n}$, which futher implies

$$\|x - \frac{q}{\sqrt{n}D(\hat{x})}\|_2 \le \frac{\alpha}{D_{\alpha,n_0/2}(\hat{x})} \le \frac{\alpha}{D} \le \frac{\alpha_0}{D_0} \le 1/4\,,$$

if we choose $\alpha_0 \le D_0/4$.

By elementary calculations, we conclude $\|x - \frac{q}{\|q\|_2}\|_2 \le 2\alpha/D$ and $\|q\|_2 \le 3\sqrt{n}D$. Define $B_2^n \coloneqq \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}$. Then, we have a $(2\alpha/D)$-net of the form:

$$\mathcal{N} = \{\frac{q}{\|q\|_2} : q \in \mathcal{Q} \cap 3\sqrt{n}\cdot B_2^n\}\,.$$

It is not necessarily true that $\mathcal{N} \subseteq S_D$. However, this can be easily fixed at the cost of making it a $(4\alpha/D)$-net via a standard trick.

We now calculate the size of the net.

$$\begin{aligned}
|\mathcal{N}| &\le |\mathcal{Q} \cap 3\sqrt{n}D \cdot B_2^n| \\
&\le \binom{n}{n_0/2} |\mathbb{Z}^{n_0/2} \cap 3\sqrt{n}D \cdot B_2^{n_0/2}| \cdot |\alpha\mathbb{Z}^{n-n_0/2} \cap 3\sqrt{n}D \cdot B_2^{n-n_0/2}| \\
&\le 2^n (3C\cdot D)^{n_0/2} (\frac{3C\cdot D}{\alpha})^{n-n_0/2} \le (\frac{C_{3.14}D}{\alpha^{1-c_{3.14}}})^n\,,
\end{aligned}$$

where $C$ is an absolute constant. □

*Proof of Lemma 3.13.* Fix some constant $c > 0$ and any $D_0 \le D \le e^{cn}$. We will argue that with very high probability, $X^* \notin S_D$. Equivalently, no vector $x \in S_D$ satisfies $\langle x, X_i\rangle = 0$ for all $i \in [n-1]$.

Fix arbitrary $x \in S_D$. Define $y \in \mathbb{R}^{n-1}$, where $y_i \coloneqq \langle X_i, x\rangle$ for $1 \le i \le n-1$.

By Corollary 3.12, we conclude

$$\mathbb{P}(|y_i| < \varepsilon) \le p_\varepsilon(x) \le \frac{C_{3.10}}{\sqrt{\kappa}}(\sqrt{n}\varepsilon + \frac{1}{D_{\alpha,\kappa}(\widehat{x})}) + C_{3.10}\exp(-c_{3.10}\alpha^2\kappa) \le C\varepsilon + C\frac{1}{\sqrt{n}D}\,,$$

for some large enough $C$ depending only on $B, K$. By tensorization lemma (Lemma 2.5),

$$\mathbb{P}(\|y\|_2 \le \varepsilon\sqrt{n}) \le (C\varepsilon + \frac{C}{\sqrt{n}D})^{n-1}\,.$$

If we choose $\varepsilon = \frac{5K\alpha}{D}$, then the first term dominates, so

$$\mathbb{P}(\|y\|_2 < \frac{5K\alpha}{D}\sqrt{n}) \le (C'\frac{5K\alpha}{D})^{n-1}\,.$$

Taking an union bound over the net in Lemma 3.14, we conclude

$$\mathbb{P}(\exists x \in \mathcal{N} : \|Ax\|_2 \le \frac{5K\alpha}{D}\sqrt{n}) \le (\frac{C_{3.14}D}{\alpha^{1-c_{3.14}}})^n \cdot (C'\frac{5K\alpha}{D})^{n-1} \le e^{-n}\,,$$

if we choose $\alpha$ sufficiently small in the range of $(n^{-1/2}, \alpha_0)$, depending on $B, K$.

Now, we transfer the net result to the whole sphere:

$$\inf_{x\in S_D} \|Ax\|_2 \ge \inf_{y\in\mathcal{N}}\{\|Ay\|_2 - \|x-y\|_2\|A\|\} \ge \frac{K\alpha}{D}\sqrt{n}\,.$$

Thus, with probability at least $1-\exp(-n)$, we have $\inf_{x\in S_D}\|Ax\|_2 \ge \frac{K\alpha}{D}\sqrt{n}$.

It remains to dyadically divide the range of $D$ and union bound over all $D$. Since $D_0 \le D \le e^{cn}$, there are only polynomially many events. Because each happens with an exponential probability, the proof is finished. □

*Proof of Lemma 3.7.* This is an immediate corollary of Corollary 3.12 and Lemma 3.13. □

# 4 Random Sign Case

The goal of the section is to prove Theorem 1.2. From Theorem 3.1, we already know the correct answer is $c^n$ for some $c \in (0,1)$, so the focus now is to get the asymptotically exact base $(1/2+\varepsilon)$. For convenience, we redefine the small ball probability of a vector $a \in \mathbb{R}^n$ to be,

$$p_\varepsilon(a) = \sup_{t\in\mathbb{R}} \mathbb{P}(|\sum_{i=1}^n a_i\xi_i - t| \le \varepsilon)\,,$$

where $\xi_i$ are distributed as iid Ber(1/2).

## 4.1 General Strategy

The starting move of Tikhomirov's proof is exactly the same as Rudelson and Vershynin's: divide the sphere into compressible and incompresible vectors and deal with them separately. By taking $\delta, \rho$ small enough, we can prove $(1/2+\varepsilon)^n$ bound for compressible vectors using the exact same proof as Lemma 3.5. Next, following the same reduction steps, the invertibility for incompressible vectors reduces to estimate the small ball probability for a random normal. This is where the approach differs. However, broadly speaking, structure of the proof for this part is still similar. We find a parameter (which was essential LCD) that characterizes the small ball probability, dyadically partition the set of incompressible vectors based on it, and then show that the random normal does not belong to weak parameter parts, which leads to a strong small ball probability estimate. Now we describe the differences in detail.

The parameter chosen in [Tik20] is called threshold:

**Definition 4.1** (Threshold)**.** *Let $x \in S^{n-1}$ and $L > 0$. The threshold*

$$\mathcal{T}(x, L) := \sup\{t \in (0,1] : p_t(x) > Lt\}\,.$$

The connection with the small ball probability is as follows: Fix $x$ and $L$. Note $p_t(x)$ is an increasing function in $t$. If $t < \mathcal{T}(x,L)$, then $p_t(x) \le p_{\mathcal{T}(x,L)}(x) \le L\mathcal{T}(x,L)$.

If $t > \mathcal{T}(x, L)$, then by definition, we know $p_t(x) \le Lt$. Thus, we have a small ball probability estimate $p_t(x) = L\max(\mathcal{T}(x, L), t)$. Threshold can be viewed as a lower bound of the range of $t$ where the small ball probability estimate is good. The goal now is to prove the following, which directly implies Theorem 1.2:

**Lemma 4.2** (Random Normal has Small Threshold)**.** *There exists $L > 0$ such that with probability at least $1 - 2^{-n}$, the threshold of random normal $\mathcal{T}(X^*, L) \le (1/2 + o(1))^n$.*

Let $\delta, \rho$ be parameters in Lemma 3.5. Fix $L > 0, \varepsilon > 0$ and $T \le (1/2 + \varepsilon)^n$. Define

$$R(T) \coloneqq \{x \in \mathrm{Incomp}(\delta, \rho) : T \le \mathcal{T}(x, L) \le 2T\}.$$

In order to show $X^* \notin R(T)$ for any $T \le (1/2 + o(1))^n$, Tikhomirov constructed a subset of the integer lattice $\mathcal{N}_T$ such that for every (properly scaled) $x \in R(T)$, there is an integer vector $y \in \mathcal{N}_T$ which closely resembles $x$ entrywise and in terms of threshold using an idea called "randomized rounding". Then by union bound,

$$\mathbb{P}(x \in R(T)) \le |\mathcal{N}_T| \max_{y \in \mathcal{N}_T} \mathbb{P}(y \text{ is almost orthogonal to } X_1, \cdots, X_{n-1}).$$

As comparison, recall that an $\varepsilon$-net is used instead of an intger lattice in the old approach. The probability is estimated using Tensorization lemma (Lemma 2.5) as before. The size of the net uses a new idea named "inversion of randomness". Instead of estimating $|\mathcal{N}_T|$ directly, we first find a larger lattice subset $\mathcal{M}_T \supseteq \mathcal{N}_T$ whose size is easy to compute. Then, Tikhomirov showed that if we sample a vector uniformly at random from $\mathcal{M}_T$ (random for each coordinate), the threshold of this vector is not in the range of $(T, 2T]$, with probability *superexponentially* close to 1. Thus, $|\mathcal{N}_T| \le |\mathcal{M}_T| \exp(-\omega(n)) \ll 2^{-n}$ for sufficiently large $n$. It remains to union bound over different $T$ (only polynomially many of them) to finish the proof.

## 4.2 Random Averaging Over $\ell_1$ norm

The superexponential probability guarantee is a special case of a much general phenomenon. In order to describe it, we first introduce the concept of an admissible set, which is a structured subset of $\mathbb{Z}^n$.

**Definition 4.3** (Admissible Set)**.** *Let $N, n \ge 1$ be integers, $\delta \in (0, 1]$, and $K \ge 1$ be some real number. A subset $\mathcal{A} \subset \mathbb{Z}^n$ is $(N, n, K, \delta)$-admissible if*

- *$\mathcal{A} = A_1 \times \cdots \times A_n$ where every $A_i$ is an origin-symmetric subset of $\mathbb{Z}$.*
- *For every $i \le \delta n$, $A_i \cap [-N, N] = \emptyset$ and $A_i$ is a union of two integer intervals of total cardinality at least $2N$.*
- *For every $i > \delta n$, $A_i$ is an integer interval of cardinality at least $2N + 1$.*
- *$|A_1| \cdots |A_n| \le (KN)^n$.*
- *For all $1 \le i \le n$, $\max A_i < nN$.*

Then, on an admissible set $\mathcal{A}$, we can define an averaged version of $f$ as follows:

**Definition 4.4** (Averaged function)**.** *Let $f : \mathbb{Z} \to \mathbb{R}$ be an arbitrary function and let $\mathcal{A}$ be an $(N, n, K, \delta)$-admissible set for some $N, n, K, \delta$. For $1 \le \ell \le n$, the averaged version of $f$ is defined as:*

$$f_{\mathcal{A},\ell}(t) \coloneqq \sum_{v \in \{0,1\}^\ell} 2^{-\ell} f(t + v_1 X_1 + \cdots + v_\ell X_\ell)$$

*where each $X_i$ is drawn uniformly from $A_i$ uniformly at random and independently.*

Note that $f_{\mathcal{A},\ell}(\cdot)$ is a random function, whose randomness comes from $X_i$s.

Intuitively, averaging should smooth out the spikes of the function, so we expect that averaged function should have much smaller $\ell_\infty$ norm. It turns out this is indeed the case, subject to some conditions on the function $f$:

**Theorem 4.5** ($\ell_\infty$ Estimate)**.** *Fix any $\delta \in (0,1], \varepsilon \in (0,p), K, M \geq 1$, there exist $n_{4.5} = n_{4.5}(\delta,\varepsilon,K,M) \geq 1, \eta_{4.5} = \eta_{4.5}(\delta,\varepsilon,K,M) \in (0,1]$ and $L_{4.5} = L_{4.5}(\delta,\varepsilon,K) > 0$ such that if $n \geq n_{4.5}, 1 \leq N \leq (1/2+\varepsilon)^{-n}$ and $\mathcal{A}$ is an $(N,n,K,\delta)$-admissible set, and $f : \mathbb{Z} \to \mathbb{R}$ is a non-negative function with $\|f\|_1 = 1$ and $\log_2 f$ is $\eta_{4.5}$-Lipschitz. Then,*

$$\mathbb{P}_{X_1,\dots,X_n}(\|f_{\mathcal{A},n}\|_\infty > L_{4.5}(N\sqrt{n})^{-1}) \leq \exp(-Mn)\,.$$

The crucial point of this theorem is the parameter controlling the $\ell_\infty$ of the average ($L_{4.5}$) is decoupled from the one that controls the probability $M$. It means after fixing $\delta, \varepsilon$ and $K$, the $L_{4.5}$ is fixed, whereas we can take $M$ arbitrarily large, at the cost of increasing $n_{4.5}$ and decreasing $\eta_{4.5}$.

Although intuitively clear, this phenomenon is highly non-trivial and interested reader should refer to Section 4 in [Tik20] for the proof. Instead, we demonstrate why it is useful for the invertibility problem.

**Corollary 4.6.** *Let $\delta, \varepsilon \in (0,1], K, M \geq 1$. There exists $n_{4.6} = n_{4.6}(\delta,\varepsilon,K,M) \geq 1, L_{4.6} = L_{4.6}(\delta,\varepsilon,K) > 0$ such that if $n \geq n_{4.6}, 1 \leq N \leq (1/2+\varepsilon)^{-n}$ and $\mathcal{A}$ is an $(N,n,K,\delta)$-admissible set, then*

$$\mathbb{P}_{X_1,\dots,X_n}(p_{\sqrt{n}}((X_1,\dots,X_n)) \geq \frac{L_{4.6}}{N}) \leq \exp(-Mn)\,.$$

*Proof.* Consider the function $f(t) \coloneqq \frac{1}{m} 2^{-\frac{|t|}{\sqrt{n}}}$, where $m = \sum_{t\in\mathbb{Z}} 2^{-|t|/\sqrt{n}}$. By definition, $\|f\|_1 = 1$ and $\log_2 f$ is $n^{-1/2}$ Lipschitz. Moreover, it is easy to verify for $-\sqrt{n} \leq t \leq \sqrt{n}$, for some absolute constant $c$, we have $f(t) \geq \frac{c}{\sqrt{n}} \mathbb{1}\{-\sqrt{n} \leq t \leq \sqrt{n}\}$.

Now apply Theorem 4.5, we get

$$\mathbb{P}_{X_1,\dots,X_n}(\|f_{\mathcal{A},n}\|_\infty \geq L_{4.5}(N\sqrt{n})^{-1}) \leq \exp(-Mn)\,.$$

Condition on a fixed realization of $(X_1,\dots,X_n) \in \mathcal{A}$, we have

$$\begin{aligned} p_{\sqrt{n}}((X_1,\dots,X_n)) \geq \frac{L_{4.6}}{N} &\implies \sup_\lambda \mathbb{P}_\xi\{-\sqrt{n} \leq \sum_{i=1}^n X_i\xi_i - \lambda \leq \sqrt{n}\} \geq \frac{L_{4.6}}{N} \\ &\implies \sup_\lambda \mathbb{P}_\xi(f(\sum_{i=1}^n X_i\xi_i - \lambda)) \geq \frac{cL_{4.6}}{N\sqrt{n}} \\ &\implies \|f_{\mathcal{A},n}\|_\infty \geq L_{4.5}(N\sqrt{n})^{-1}\,. \end{aligned}$$

Thus, this is true also for unconditioned $(X_1,\dots,X_n)$, which finishes the proof. □

## 4.3 Randomized Rounding

We need to show that for every vector $x \in \mathbb{R}^n$, there exists an integer vector $y \in \mathbb{Z}^n$ that almost preserves its anti-concentration property and is close to it.

**Lemma 4.7** (Randomized Rounding)**.** *Let $x \in \mathbb{R}^n$ and $L > 0$. Let $\lambda$ be a number such that $\mathbb{P}(|\sum_{i=1}^n b_i x_i - \lambda| \leq t) \leq Lt$, where each $b_i$ are iid $\mathrm{Ber}(1/2)$. Then, there exists an integer vector $y \in \mathbb{Z}^n$ satisfying:*

1. $\|x - y\|_\infty \le 1$,
2. $\mathbb{P}(|\sum_{i=1}^n b_i y_i - \lambda| \le t) \le C_{4.7} L t$ *for all* $t \ge \sqrt{n}$,
3. $p_{\sqrt{n}}(y) \ge c_{4.7} p_{\sqrt{n}}(x)$,
4. $|\sum_{i=1}^n x_i - \sum_{i=1}^n y_i| \le C_{4.7}\sqrt{n}$,

*where* $C_{4.7}, c_{4.7}$ *are absolute constants.*

*Proof Sketch.* The method of finding this vector $y$ is called randomized rounding. The idea is to set $y_i = \lfloor x_i \rfloor$ with probability $p_i$ and $\lfloor x_i \rfloor + 1$ with probability $1 - p_i$ and select $p_i$ appropriately so that $\mathbb{E} y_i = x_i$. Then, it remains to argue that the failure probability for each property is small enough, so that their sum is strictly less than 1. See Lemma 5.3 in [Tik20] for details. □

Apply this lemma to a $\frac{\sqrt{n}}{\mathcal{T}(x,L)} x$ and use the definition of $\mathcal{T}(x, L)$, we get an integer vector $y \in \mathbb{Z}^n$ satisfying:

(a) $\|\frac{\sqrt{n}}{\mathcal{T}(x,L)} x - y\|_\infty \le 1$,

(b) $\mathbb{P}(|\sum_{i=1}^n b_i y_i - \frac{\sqrt{n}}{\mathcal{T}(x,L)} \sum_{i=1}^n x_i| \le t) \le \frac{C_{4.7} L \mathcal{T}(x,L)}{\sqrt{n}} t$ for all $t \ge \sqrt{n}$,

(c) $p_{\sqrt{n}}(y) \ge c_{4.7} L \mathcal{T}(x, L)$,

(d) $|\frac{\sqrt{n}}{\mathcal{T}(x,L)} \sum_{i=1}^n x_i - \sum_{i=1}^n y_i| \le C_{4.7}\sqrt{n}$.

After possibly permuting the entries, the vector $y$ has some structure. More importantly, the set of permutations that reveals structure for any such $y$ has a cardinality only exponentially large, which is ignorable when compared to the superexponentially small probability.

**Lemma 4.8** (Structure of Integer Approximation)**.** *Let* $y \in \mathbb{Z}^n$ *be an integer approximation for some* $\frac{\sqrt{n}}{\mathcal{T}(x,L)} x$*. There exists a set of permutations* $\Pi$ *of size at most* $C_{4.8}^n$ *for some absolute constant* $C_{4.8} > 0$ *satisfying the following: for every such* $y$*, there exists* $\pi \in \Pi$ *such that* $\tilde{y} = y_\pi$ *satisfies*

- $|\tilde{y}_i| > \frac{\nu}{\mathcal{T}(x,L)} - 1$ *for all* $i \le \delta n$,
- $|\tilde{y}_i| \le \frac{2^{(j+1)/2}}{\sqrt{\delta}\mathcal{T}(x,L)} + 1$ *for* $i \ge 2^{-j}\delta n$ *and* $0 \le j \le \log(\delta n)$.

*Proof Sketch.* $\Pi$ consists of all chains of subsets $[n] \supset I_0 \supset I_1 \supset \cdots \supset I_{\log(\delta n)}$, where $|I_i| = \delta n 2^{-i}$. Thus,

$$|\Pi| \le \binom{n}{\delta n}\binom{\delta n}{\frac{1}{2}\delta n} \cdots \binom{2}{1} \le 2^{2\delta n - 1} \le 4^n .$$

Given a vector $y$, we sort its entries decreasingly according to the absolute values. By definition, there exists $\pi \in \Pi$ that agrees with this order for the first $\delta n$ entries (ignoring the order within each subset). From the definition of incompressibility, it is not hard to verify that this $\pi$ satisfies these two properties, using a simple argument similar to Proposition 3.4. □

Let $n \geq 2, \delta \in [1/n, 1/2], \nu \in (0,1]$ and $T \in (0,1]$ such that $\nu/T \geq 2$. Define

$$A_i := \begin{cases} \mathbb{Z} \cap [-\frac{2\sqrt{n}}{T} - 1, \frac{2\sqrt{n}}{T} + 1] \backslash [1 - \frac{\nu}{T}, \frac{\nu}{T} - 1] & \text{if } i = 1 \\ \mathbb{Z} \cap [-\frac{2^{(j+3)/2}}{\sqrt{\delta}T} - 1, \frac{2^{(j+3)/2}}{\sqrt{\delta}T} + 1] & \text{if } 1 \leq j \leq \log(\delta n) \text{ and } 2^{-j}\delta n \leq i \leq 2^{-j+1}\delta n \\ \mathbb{Z} \cap [-\frac{\sqrt{8}}{\sqrt{\delta}T} - 1, \frac{\sqrt{8}}{\sqrt{\delta}T} + 1] & \text{if } i > \delta n\,. \end{cases}$$

and take $\mathcal{A}(n, \delta, \nu, T) = A_1 \times \cdots \times A_n \subset \mathbb{Z}^n$. From the definition, it is straightforward to vertify that for each vector $x \in R(T)$, there exists $y \in \mathbb{Z}^n$ that is an approximation for $\frac{\sqrt{n}}{\mathcal{T}(x,L)}$ and $\pi \in \Pi$ satisfying $y_\pi \in \mathcal{A}(n, \delta, \nu, T)$. Moreover,

**Proposition 4.9** ($\mathcal{A}(n, \delta, \nu, T)$ is Admissible)**.** *For any $\delta \in (0, 1/2]$ and $\nu \in (0, 1]$, there exists $n_{4.9} = n_{4.9}(\delta, \nu) \geq 1$, $K_{4.9} = K_{4.9}(\delta, \nu) \geq 1$ such that for any $n \geq n_{4.9}$ and $T \in (0, \nu/2]$ and set $N = \lfloor \frac{\nu}{T} \rfloor - 1$ such that $\mathcal{A}(n, \delta, \nu, T)$ is $(N, n, K_{4.9}, \delta)$-admissible.*

## 4.4 Proof of Lemma 4.2

Before proving Lemma 4.2, we first state a simple upper bound on the threshold, which determines the range of thresholds we need to consider (for the same purpose, recall we had a lower bound $D_0$ for essential LCD, see the text below Lemma 3.13).

**Proposition 4.10** (Upper Bounds on Threhsold)**.** *For every $\delta, \nu \in (0, 1]$, there exist $K_{4.10} = K_{4.10}(\delta, \nu) > 0$ and $L_{4.10} = L_{4.10}(\delta, \nu) \geq 1$ such that for $n \geq 2$ and $L \geq L_{4.10}$ and $x \in \mathrm{Incomp}(\delta, \nu)$, we have $\mathcal{T}(x, L) \leq K_{4.10} \cdot n^{-1/2}$.*

*Proof Sketch.* Use the restriction restriction (Proposition 2.4) on the spread part of $x$ and then use standard results on the small ball probability for binomial random variables. □

*Proof of Lemma 4.2.* Fix $L \geq L_{4.10}$ so that $\mathcal{T}(x, L) \leq K_{4.10}/\sqrt{n}$. Recall that $\delta, \nu$ was fixed by Lemma 3.5 and we defined $R(T) = \{x \in \mathrm{Incomp}(\delta, \nu) : T \leq \mathcal{T}(x, L) \leq 2T\}$. Dyadically partition the thresholds into $T_j = 2^{-j} K_{4.10}/\sqrt{n}$ for $j = 0, 1, \ldots$. It suffices to focus our attention to $\mathcal{T}(X^*, L) \in R(T_j)$ for some $j \leq \lfloor -n \log(\frac{1}{2} + \varepsilon) \rfloor$. Fix an arbitrary $j$ in this range and let $T = T_j$. Our goal is to upper bound $\mathbb{P}(\mathcal{T}(X^*, L) \in R(T))$. Let

$$\mathcal{N}_T = \{y : y \text{ is approximation of } \frac{\sqrt{n}}{\mathcal{T}(x, L)} x \text{ for some } x \in R(T)\} \subset \mathbb{Z}^n\,.$$

Recall that $X_1, \ldots, X_n$ are columns of $A$. If $X^* \in R(T)$, then we know $\langle \frac{\sqrt{n}}{\mathcal{T}(x,L)} x, X_i \rangle = 0$ for all $1 \leq i \leq n-1$. From Lemma 4.7, we know that there exists $y \in \mathcal{N}_T$ that approximates $X^*$ well. The properties of $y$ allows us to deduce that

$$\|(\langle X_1, y \rangle, \ldots, \langle X_{n-1}, y \rangle)\|_2 \leq Cn\,,$$

for some constant $C > 1$. By union bound, we conclude that

$$\mathbb{P}(X^* \in R(T)) \leq |\mathcal{N}_T| \mathbb{P}(\|(\langle X_1, y \rangle, \ldots, \langle X_{n-1}, y \rangle)\|_2 \leq Cn)\,.$$

The probability estimate follows from an application of the tensorization Lemma 2.5, same as the previous proof. To estimate the cardinality of the net, we use Lemma 4.8. Define $\mathcal{N}'_T = \{y_\pi : y \in \mathcal{N}_T, \pi \in \Pi\}$. It is immediate that $|\mathcal{N}_T| \le |\Pi| \cdot |\mathcal{N}'_T| \le C_{4.8}^n |\mathcal{N}'_T|$. From Proposition 4.9, we know that $\mathcal{N}'_T \subseteq \mathcal{A}$ and that $\mathcal{A}$ is an $(N, n, K_{4.9}, \delta)$-admissible set. Note that permuting entries does not change the small ball probability estimate of $y$. In particular, vector $y' \in \mathcal{N}'_T$ has a lower bound of $\frac{L}{N}$ on the small ball probability that comes from $y \in \mathcal{N}_T$ and further from the vector $\frac{\sqrt{n}}{\mathcal{T}(x,L)}x$.

Now we can use the definition of admissible set and Corollary 3.12 to conclude that $|\mathcal{N}'_T| \le \exp(-Mn)|\mathcal{A}| \le \exp(-Mn)(KN)^n$. For large enough $M$, the superexponentially low probability clearly dominates over the exponential terms. Thus,

$$\mathbb{P}(X^* \in R(T)) \le C_{4.8}^n (KN)^n \exp(-Mn) = o(2^{-n})$$

for large enough $n$. It remains to union bound over polynomially many different $T$. $\square$

# References


[BVW10] Jean Bourgain, Van H Vu, and Philip Matchett Wood. On the singularity probability of discrete random matrices. *Journal of Functional Analysis*, 258(2):559–603, 2010. 2

[Ede88] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM journal on matrix analysis and applications*, 9(4):543–560, 1988. 4

[KKS95] Jeff Kahn, János Komlós, and Endre Szemerédi. On the probability that a random±1-matrix is singular. *Journal of the American Mathematical Society*, 8(1):223–240, 1995. 2, 4

[Kom67] János Komlós. On the determinant of (0-1) matrices. *Studia Scientiarum Mathematicarum Hungarica*, 2:7–21, 1967. 2

[LPRTJ05] Alexander E Litvak, Alain Pajor, Mark Rudelson, and Nicole Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Advances in Mathematics*, 195(2):491–523, 2005. 6

[RV08] Mark Rudelson and Roman Vershynin. The littlewood–offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(2):600–633, 2008. 1, 2, 4, 8

[Tik20] Konstantin Tikhomirov. Singularity of random bernoulli matrices. *Annals of Mathematics*, 191(2):593–634, 2020. 1, 2, 11, 13, 14

[TV08] Terence Tao and Van Vu. Random matrices: the circular law. *Communications in Contemporary Mathematics*, 10(02):261–307, 2008. 2

[TV09] Terence Tao and Van H Vu. Inverse littlewood-offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, pages 595–632, 2009. 2

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018. 3