An upper bound on the smallest singular value of dense random combinatorial matrices

Dongbin Li, Alexander E. Litvak, and Tingzhou Yu

Abstract

Let M be an $n \times n$ random matrix with entries in $\{0,1\}$, where each row is independently and uniformly sampled from the set of all vectors in $\{0,1\}^n$ containing exactly d ones, with $d = pn$ for some fixed constant $p \in (0, 1/2]$. A recent result of Tran states that the smallest singular value $s_n(M)$ is bounded below by $c_p n^{-1/2}$ with high probability. In this note, we establish a complementary upper bound for $s_n(M)$, proving that

$$\forall \varepsilon > 0 \qquad \mathbb{P}\left(s_n(M) \leq \frac{\sqrt{d}}{\varepsilon^2 n}\right) \geq 1 - C_p\left(\varepsilon + \frac{1}{\sqrt{d}}\right),$$

where C_p is a positive constant depending only on p. This result confirms that the least singular value $s_n(M)$ of dense random combinatorial matrices is typically of the order $n^{-1/2}$.

AMS 2010 Classification: primary: 60B20, 15B52, 46B06; secondary: 60C05, 05C80, 46B09.
Keywords: condition number, invertibility of random matrices, random graphs, random matrices, regular graphs, singular probability, smallest singular value.

1 Introduction

Let M be an $n \times n$ matrix with real entries, recall that the smallest singular value of M, denoted by $s_n(M)$, can be defined as

$$s_n(M) := \inf_{x \in \mathbb{S}^{n-1}} \|Mx\|_2,$$

where $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ is the unit sphere in $\mathbb{R}^n$. Note that $s_n(M) > 0$ if and only if M is invertible. When M is invertible, $s_n(M) = 1/\|M^{-1}\|$, where $\|M\| = \sup_{\|x\|_2=1} \|Mx\|_2$ is the operator norm of $M : \ell_2^n \to \ell_2^n$. The study of the smallest singular value is of fundamental importance due to its crucial role in theoretical questions and practical applications. In numerical analysis, the smallest singular value is closely related to the condition number, which measures the worst-case precision loss in computational problems [34, 42]. Beyond numerical stability, the smallest singular value is essential in a variety of contexts, including the analysis of matrix singularity probabilities (for 0/1 matrices with independent entries see, for example, [3, 14, 15, 23, 41]), and the study of the empirical spectral distribution of random matrices (see [39, 4, 30] and references therein for the historical account of the problem).

The quantitative study of the smallest singular value of random matrices dates back to von Neumann and his collaborators, who conjectured that the smallest singular value is of order $n^{-1/2}$ with high probability (see [44, 45]). This estimate was later confirmed by Edelman [8] and Szarek [35] for random matrices with independent identically distributed (i.i.d.) standard Gaussian entries. In [31, 32], Rudelson and Vershynin established the sharp bound $\Theta(n^{-1/2})$ for square matrices with i.i.d. subgaussian entries, that is, $cn^{-1/2} \leq s_n(M) \leq Cn^{-1/2}$. A similar behaviour for square matrices with i.i.d. entries possessing a finite second moment is obtained by combining the results by Rebrova and Tikhomirov [29] and Tatarko [40]. For other results and bounds on the smallest singular value of square random matrices with independent entries, we refer the reader to [17, 25, 37, 38]. The study of the corresponding bounds for the rectangular matrices with i.i.d. subgaussian variables was started in [22] and further developed in [33, 24, 26].

In contrast, estimating the smallest singular value becomes considerably more challenging when dependencies exist between the matrix entries, and progress has been comparatively much slower. For example, only very recently a sharp lower tail bound for the least singular value of $n \times n$ symmetric matrices with i.i.d. subgaussian entries has been obtained by Campos, Jenssen, Michelen and Sahasrabudhe [5], despite the analogous result for non-symmetric matrices being established nearly 17 years earlier by Rudelson and Vershynin [32]. In recent years, considerable attention has also been given to models derived from combinatorics and graph theory, beyond symmetric random matrices. In particular, estimates of the least singular value of such models have been studied extensively (see, for example, [6, 13, 16, 20, 21, 43] for some recent developments in this area). Our work focuses on one of such models.

Let $d \leq n$ be positive integers. We consider the set $\mathcal{M}_{n,d}$ of all $n \times n$ matrices with entries in $\{0,1\}$, where each row is chosen independently and uniformly from the set of $\{0,1\}^n$ vectors with exactly d ones. An element $M \in \mathcal{M}_{n,d}$ can be interpreted as the adjacency matrix of a directed graph (digraph) on n labeled vertices, where each vertex has exactly d outgoing edges. Alternatively, M can be viewed as the adjacency matrix of a bipartite graph between two disjoint sets of n vertices each, with each vertex on the left having degree d. Note that while the out-degree (row sum) of each vertex is fixed at d, the in-degree (column sums) are random with mean d. This suggests that theoretically our model can produce matrices with zero columns, especially when d is small. In fact, it is not difficult to verify that for any (fixed) $\varepsilon > 0$ and $d \leq (1-\varepsilon)\log n$, the random matrix M asymptotically almost surely contains a zero column [2, 19]. Moreover, the threshold $d = \log n$ is sharp in the sense that for any given $\varepsilon > 0$, if $\min(d, n-d) \geq (1+\varepsilon)\log n$, then a matrix M chosen uniformly at random from $\mathcal{M}_{n,d}$ is asymptotically almost surely invertible, a result established by Ferber, Kwan, and Sauermann [10, Theorem 1.2]. It is noteworthy that this sharp threshold for singularity also arises in the context of random Bernoulli matrices (see, e.g., [1, 3, 23]).

The behavior of the least singular value of this model in the dense regime, specifically for degree $d = n/2$ was first studied in [27, 28]. Nguyen and Vu showed that for any $C > 0$, there exists $D > 0$ such that

$$\mathbb{P}(s_n(M) < n^{-D}) \le n^{-C}.$$

Building on the earlier work of Ferber, Jain, Luh, and Samotij [9], Jain [13] improved this bound, showing that

$$\mathbb{P}(s_n(M) < \varepsilon n^{-2}) \le C\varepsilon + 2e^{-n^{0.0001}}.$$

More recently, Tran [43] showed that there exist constants $C, c > 0$ such that for all $\varepsilon \ge 0$,

$$\mathbb{P}(s_n(M) \le \varepsilon n^{-1/2}) \le C\varepsilon + 2e^{-cn}.$$

Further refinements to the probability bounds for $s_n(M)$ when $d = n/2$ were obtained in [14]. It was shown that for every $\varepsilon > 0$, there exists $C = C(\varepsilon)$ such that for all $t \ge 0$,

$$\mathbb{P}(s_n(M) \le tn^{-1/2}) \le Ct + (1/2 + \varepsilon)^n.$$

The result also confirms that the singularity probability of M is $(1/2 + o(1))^n$, as conjectured by Nguyen in [27].

Although the existing lower bounds provide insight into the anticipated scale of $s_n(M)$, a matching probabilistic upper bound has not been obtained for the $\mathcal{M}_{n,d}$ model in the dense regime. The main contribution of this paper is to establish this missing upper bound, thereby determining the sharp asymptotic order of $s_n(M)$ for this class of random matrices. Specifically, we prove the following main result in this note.

Theorem 1.1. *Let $p \in (0, 1/2]$ be a fixed constant and let d, n be integers such that $d = pn$. Let M be a random $n \times n$ matrix uniformly drawn from $\mathcal{M}_{n,d}$. Then there exists a constant $C_p > 0$ depending only on p such that for every $\varepsilon > 0$,*

$$\mathbb{P}\left(s_n(M) \le \frac{\sqrt{d}}{\varepsilon^2 n}\right) \ge 1 - C_p\left(\varepsilon + \frac{1}{\sqrt{d}}\right).$$

Remark 1.2. Theorem 1.1 shows that in the dense regime ($d = pn$, p is fixed) the upper bound $s_n(M) = O(\sqrt{d}/n)$ holds with high probability, in particular, it implies that $s_n(M) = O_p(n^{-1/2})$. The matching lower bound on $s_n(M)$ was proved in [43] (see the third part of the remark following Theorem 1.2 there, see also [14] for $d = n/2$). Thus, our upper bound confirms that in the dense regime the typical order of the least singular value $s_n(M)$ is of order $n^{-1/2}$.

Remark 1.3. By Theorem 4.1 in [19], for some absolute constant $C > 0$, $\|M - \mathbb{E}M\| \le C\sqrt{d}$ with probability at least $1 - 2/n$, whenever $d \ge \log n$ (for $d = n/2$ this bound with better probability was obtained in [43, Proposition 2.8] (see also [14]) and note that the proof in [43] can be extended to $d = pn$). Since all entries of $\mathbb{E}M$ are d/n, its norm is d. Therefore, by the triangle inequality, $\|M\| \le 2d$ with probability at least

$1-2/n$ for large enough d. Note that $M\mathbf{1}_n = d\mathbf{1}_n$, where $\mathbf{1}_n = (1,1,\dots,1)$, hence $\|M\| \geq d$. Thus, the largest singular value of M, $s_1(M) = \|M\|$ is of order d, therefore, in the dense regime $d = pn$, the condition number

$$\kappa(M) := s_1(M)/s_n(M) \approx n^{3/2}$$

(equivalence is up to constants depending only on p and with probability at least $1 - C'_p/\sqrt{n}$). This has important implications for numerical computation. Indeed, it is well known that condition number is closely linked to computational complexity, particularly in the context of numerical linear algebra and iterative methods. In general, a large condition number typically leads to slower convergence (see, e.g., [12, Chapters 2 and 11]). We would also like to mention that in general the operator norm of M could be as large as its Hilbert-Schmidt norm, $\sqrt{dn}$, as the example of matrix having first d columns equal to $\mathbf{1}_n$ and $n-d$ zero columns shows (in this case the norm is attained on the vector $e_1 + e_2 + ... + e_d$).

Remark 1.4. It is also natural to consider the behavior of $s_n(M)$ in the *sparse* regime, where $C_0 \log n \leq d = o(n)$ for an absolute constant $C_0 > 1$. For such sparse random matrices, it is speculated that the typical order of $s_n(M)$ is $\sqrt{d}/n$ (see also a similar conjecture in [21] for d-regular matrices). Notably, this conjectured order $\sqrt{d}/n$ appears in the upper bound stated in Theorem 1.1. However, despite the well-understood by now behavior of the least singular value in the dense case, its sparse counterpart has remained far less accessible. The only known quantitative lower bound for $s_n(M)$ has been recently established in [19], where the authors proved that for large enough n, some absolute positive constants C and c, and for $d \geq \log^2 n$,

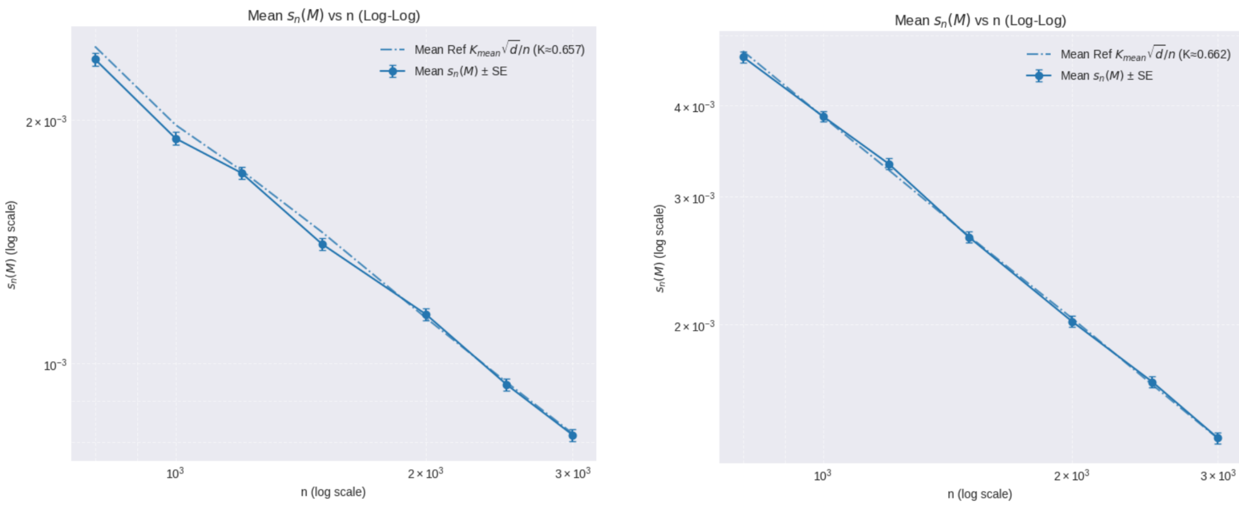
$$\mathbb{P}(s_n(M) \leq n^{-c}) \leq C/\sqrt{d}.$$

While the expected order $\sqrt{d}/n$ remains out of reach in the sparse regime, our numerical simulations offer supporting evidence for the conjecture. Figures 1a and 1b present log-log plots of the mean $s_n(M)$ versus n for the intermediate regimes $d = n^{1/3}$ and $d = 5\log n$, respectively. In both cases, the observed data aligns well with a reference line proportional to $\sqrt{d}/n$, indicating that this order likely holds in the sparse regime as well.

Conjecture 1.5. *For every $\varepsilon, \delta > 0$ there is a constant $C = C(\varepsilon, \delta)$ depending only on ε, δ such that the following holds. Let $d \leq n$ be integers that satisfy $\min(d, n-d) \geq (1+\varepsilon)\log n$. Let M be a random $n \times n$ matrix uniformly drawn from $\mathcal{M}_{n,d}$. Then*

$$\mathbb{P}\left(s_n(M) > C\,\frac{\sqrt{d}}{n}\right) \leq \delta.$$

Remark 1.6. For fixed $\varepsilon > 0$, and $d \leq (1-\varepsilon)\log n$, it is not difficult to check that with high probability, M contains a zero column. Therefore, $s_n(M) = 0$ with high probability whenever $\min(d, n-d) \leq (1-\varepsilon)\log n$ and Conjecture 1.5 holds in this regime.



(a) Simulation for $d = \lfloor n^{1/3} \rfloor$

(b) Simulation for $d = 5\lfloor \log n \rfloor$

Figure 1: Log-log plots of the mean smallest singular value $\mathbb{E}[s_n(M)]$ versus n. Points show simulation results (mean $\pm$ standard error). The dashed lines represent reference slopes proportional to $\sqrt{d}/n$. (a) Case $d = \lfloor n^{1/3} \rfloor$. (b) Case $d = 5\lfloor \log n \rfloor$.

Organization of the paper. This paper is structured as follows. In Section 2, we introduce essential preliminaries, including key concepts such as biorthogonal systems, almost constant vectors, and concentration inequalities. In Section 3, we establish the invertibility of the matrix M on almost constant vectors. The proof of Theorem 1.1 is presented in Section 4.

2 Preliminaries

2.1 Notation.

We use the following standard notations. Let $\{e_1, e_2, \ldots, e_n\}$ be the canonical basis of $\mathbb{R}^n$ equipped with the canonical inner product $\langle \cdot, \cdot \rangle$ and the canonical Euclidean norm $\| \cdot \|_2$. We use $[n] := \{1, 2, \ldots, n\}$ to represent the set of the first n positive integers. The vector $\mathbf{1}_n \in \mathbb{R}^n$ denotes the vector with each component equals 1. The set

$$\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$$

represents the Euclidean unit sphere in $\mathbb{R}^n$. Matrices are denoted by uppercase letters, e.g., M, with M^T indicating the transpose of M. For a matrix M, $R_i = R_i(M)$ denotes its i-th row, and $X_i := R_i^T$ is the corresponding column vector. The distance from a vector v to a subspace H is denoted by

$$\text{dist}(v, H) := \inf_{w \in H} \|v - w\|_2.$$

The operator norm of an $m \times k$ matrix A, considered as an operator acting between corresponding Euclidean spaces (that is, $A : \ell_2^k \to \ell_2^m$) is denoted by $\|A\|$. Respectively, given a subspace $V \subset \mathbb{R}^k$, $\|A_{|V}\|$ denotes the operator norm of A restricted to V. By $C, c, C_0, C_1, \cdots$ we denote either positive absolute constants independent of all parameters or positive constants that may depend on the parameter $p = d/n$. They may vary from line to line.

2.2 Biorthogonal systems

Biorthogonal systems play an important role in the analysis of the invertibility of matrices and the structure of their inverses (see, e.g., [31, 40]). Let $\{E_k\}_{k=1}^m$ and $\{F_k\}_{k=1}^m$ be two sets of vectors in a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$. We say that the system $\{E_k, F_k\}_{k=1}^m$ is a *biorthogonal system* in $\mathcal{H}$ if it satisfies $\langle E_i, F_j \rangle = \delta_{ij}$ for all $i, j \in [m]$, where δ_{ij} is the Kronecker delta. Note that if $\{E_k, F_k\}_{k=1}^m$ is a biorthogonal system, then both $\{E_k\}_{k=1}^m$ and $\{F_k\}_{k=1}^m$ are linearly independent. Thus, if the dimension of $\mathcal{H}$ is m then $\{E_k\}_{k=1}^m$ spans $\mathcal{H}$ as well as $\{F_k\}_{k=1}^m$. We summarize key properties of biorthogonal systems (see, e.g., [31, Proposition 2.1]) relevant to the proof of our main result in the following proposition.

Proposition 2.1. *Let $\mathcal{H}$ be an n-dimensional Hilbert space.*

1. *For any invertible $n \times n$ matrix A, the system $\{Ae_k, (A^{-1})^T e_k\}_{k=1}^n$ forms a biorthogonal system in $\mathbb{R}^n$.*
2. *Given a linearly independent set of vectors $\{E_k\}_{k=1}^n$ in an n-dimensional Hilbert space $\mathcal{H}$, there exist unique vectors $\{F_k\}_{k=1}^n$ such that $\{E_k, F_k\}_{k=1}^n$ is a biorthogonal system in $\mathcal{H}$.*
3. *For a biorthogonal system $\{E_k, F_k\}_{k=1}^n$ in an n-dimensional Hilbert space $\mathcal{H}$, the Euclidean norm of each F_k satisfies*

$$\|F_k\|_2 = \frac{1}{dist(E_k, H_k)}, \qquad k = 1, \dots, n, \tag{2.1}$$

   where $H_k = span\{E_i\}_{i \neq k}$ denotes the subspace spanned by all E_i except E_k, and $\mathrm{dist}(E_k, H_k)$ is the Euclidean distance from E_k to H_k.

2.3 Almost constant vectors

To obtain an upper bound on the smallest singular value $s_n(M)$, we consider a decomposition of the unit sphere $\mathbb{S}^{n-1}$ into *almost constant* and *non-almost constant vectors*. The notion of almost constant vectors was introduced in a similar context in earlier works [6, 20] and then developed in [21, 23].

Definition 2.2. *For parameters $\delta, \rho \in (0,1)$, let $Cons_{\delta,\rho} \subset \mathbb{S}^{n-1}$ be the set of vectors $v \in \mathbb{S}^{n-1}$ for which there exists a real number λ such that $|v_i - \lambda| \leq \rho/\sqrt{n}$ holds for at least $(1-\delta)n$ coordinates $i \in [n]$. A vector $v \in \mathbb{S}^{n-1}$ is called non-almost constant if it does not belong to $Cons_{\delta,\rho}$.*

2.4 Concentration

Let M be an $m \times n$ random matrix with i.i.d. rows, such that each row is uniformly drawn from n-dimensional 0/1 vectors having exactly d ones. The main result of this section is an individual concentration bound for the norm $\|Mv\|_2$, where v is a fixed unit vector. We will combine this individual probability with net argument to obtain the invertibility on the set of almost constant vectors in Proposition 3.1 below.

For the case $d = n/2$, Lemma 2.9 in [43] established such a result using Bernstein's inequality. Here we provide an alternative proof using the Paley-Zygmund inequality and Chernoff bounds for the sum of Bernoulli random variables, extending the result to any fixed $p = d/n \in (0, 1/2]$.

Lemma 2.3. *Let $v \in \mathbb{S}^{n-1}$ be a fixed vector and $p \in (0, 1/2]$ be a fixed constant. Let $d \leq n$ be positive integers such that $d = pn$. There exist positive constants $C_{2.3}, c_{2.3}$ depending only on p such that the following holds. Let M be a random $m \times n$ matrix, $n/2 \leq m \leq n$, whose rows are independent random vectors uniformly distributed on the set of $\{0,1\}^n$ with exactly d ones. Then*

$$\mathbb{P}\left(\|Mv\|_2 \leq c_{2.3}\sqrt{pn}\right) \leq e^{-C_{2.3}n}.$$

The next lemma provides a concentration inequality for functions on the Boolean hypercube, which we will use in the proof of Lemma 2.3.

Lemma 2.4. *[18, Lemma 2.1] Let $w \in \mathbb{R}^n$ be a positive vector and $f : \{0,1\}^n \to \mathbb{R}$ be a function such that*

$$|f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)| \leq w_i, \tag{2.2}$$

for all $x \in \{0,1\}^n$ and all $i \in [n]$. Suppose that η is a random vector uniformly distributed over $\{0,1\}^n$ with exactly d ones, where $1 \leq d \leq n$. Then

$$\forall t \geq 0 \qquad \mathbb{P}(|f(\eta) - \mathbb{E}f(\eta)| \geq t) \leq 2\exp\left(-\frac{t^2}{8\sum_{i=1}^n w_i^2}\right).$$

Proof of Lemma 2.3. Let q be a random vector distributed uniformly on the set of $\{0,1\}^n$ vectors with exactly d ones. Define the function $f : \{0,1\}^n \to \mathbb{R}$ by $f(x) = \langle x, v\rangle$ for $x \in \{0,1\}^n$. Then (2.2) holds with $w_i = |v_i|$. Let $\mu := \mathbb{E}f(q)$ and $\sigma^2 := \mathrm{Var} f(q)$. Applying Lemma 2.4 and using $v \in \mathbb{S}^{n-1}$, we have

$$\forall t \geq 0 \qquad \mathbb{P}\left(|f(q) - \mu| \geq t\right) \leq 2\exp\left(-\frac{t^2}{8}\right). \tag{2.3}$$

Fix $\delta = \sqrt{8\log 4}$ and consider two cases based on the value of μ.

Case 1: $|\mu| \geq 2\delta$. By the triangle inequality, we have

$$|f(q)| \geq |\mu| - |f(q) - \mu| \geq 2\delta - |f(q) - \mu|.$$

Thus if $|f(q) - \mu| \leq \delta$ then $|f(q)| \geq \delta$. Therefore, using (2.3)

$$\mathbb{P}\left(|f(q)| \geq \delta\right) \geq \mathbb{P}\left(|f(q) - \mu| \leq \delta\right) \geq 1 - 2\exp\left(-\frac{\delta^2}{8}\right) = \frac{1}{2}.$$

Case 2: $|\mu| < 2\delta$. We first compute the mean and the variance. Note that we have

$$\mu = \mathbb{E}\left(\sum_{i=1}^{n} q_i v_i\right) = \sum_{i=1}^{n} v_i \mathbb{E}(q_i) = \frac{d}{n}\sum_{i=1}^{n} v_i.$$

The variance is computed using $\mathrm{Var} q_i = p(1-p)$ and $\mathrm{Cov}(q_i, q_j) = -\frac{p(1-p)}{n-1}$ for $i \neq j$:

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^{n} v_i^2 \,\mathrm{Var}\, q_i + \sum_{i\neq j} v_i v_j \;\mathrm{Cov}(q_i, q_j) = p(1-p)\sum_{i=1}^{n} v_i^2 - \frac{p(1-p)}{n-1}\sum_{i\neq j} v_i v_j \\
&= p(1-p) - \frac{p(1-p)}{n-1}\left(\left(\sum_{i=1}^{n} v_i\right)^2 - \sum_{i=1}^{n} v_i^2\right) = p(1-p) - \frac{p(1-p)}{n-1}\left((\mu/p)^2 - 1\right) \\
&= p(1-p)\left(1 - \frac{\mu^2/p^2 - 1}{n-1}\right).
\end{aligned}$$

Since $|\mu| < 2\delta$ and $\delta = \sqrt{8\log 4}$ is a constant, $(\mu/p)^2$ is bounded by a constant depending only on p. Without loss of generality, we may assume that n is sufficiently large (bounded below by a constant depending only on p). Indeed, otherwise the conclusion of the lemma follows by making $c_{2.3}$ and $C_{2.3}$ small enough. Then the term $(\mu^2/p^2 - 1)/(n-1)$ is bounded above by $1/2$. Therefore,

$$\sigma^2 \geq \frac{1}{2}p(1-p).$$

Then we have

$$\mathbb{E}\, f(q)^2 = \mu^2 + \sigma^2 \geq \frac{1}{2}p(1-p).$$

Hence, by the Paley-Zygmund inequality,

$$\mathbb{P}\left(f(q)^2 > \frac{1}{4}p(1-p)\right) \geq \mathbb{P}\left(f(q)^2 > \frac{1}{2}\mathbb{E}\left(f(q)^2\right)\right) \geq \frac{1}{4}\frac{\left(\mathbb{E}\, f(q)^2\right)^2}{\mathbb{E}\, f(q)^4}.$$

Thus it remains to bound $\mathbb{E} f(q)^4$ from above. Applying (2.3) and the distribution formula, we observe

$$\mathbb{E}\,|f(q) - \mu|^4 = \int_0^\infty 4u^3\,\mathbb{P}\left(|f(q) - \mu| \geq u\right)\,du \leq 128.$$

Using $(a+b)^4 \leq 16(a^4 + b^4)$ for every $a, b \in \mathbb{R}$ and $|\mu| \leq 2\delta$,

$$\mathbb{E}\, f(q)^4 = \mathbb{E}\left(|f(q) - \mu + \mu|^4\right) \leq 16\,\mathbb{E}\left(f(q) - \mu\right)^4 + 16\mu^4 \leq 2^{11} + 256\delta^4.$$

Let $C_1' = 2^{11} + 256\delta^4$. Note that C_1' is an absolute constant. The Paley-Zygmund inequality yields

$$\mathbb{P}\left(f(q)^2 > \frac{1}{4}p(1-p)\right) \geq \frac{1}{4}\frac{\left(\mathbb{E}[f(q)^2]\right)^2}{\mathbb{E}[f(q)^4]} \geq \frac{1}{4}\cdot\frac{(p(1-p)/2)^2}{C_1'}.$$

Since $1-p \geq 1/2$,

$$\mathbb{P}\left(|f(q)| \geq \sqrt{p/8}\right) \geq \frac{(p(1-p))^2}{16C_1'} =: c \in (0, 1/2].$$

Thus, combining these two cases and using $\sqrt{p/8} \leq 1/4 \leq \delta$, we obtain

$$\mathbb{P}\left(|f(q)| \geq \sqrt{p/8}\right) \geq c. \tag{2.4}$$

Let $q_1, \dots, q_m$ be the independent rows of M, and let $f_i = f(q_i) = \langle q_i, v\rangle$. Define the indicator variables $\eta_i := \mathbb{1}_{\{|f_i|>\sqrt{p/8}\}}$ for $i \in [m]$. From (2.4), we have

$$\mathbb{P}(\eta_i = 1) = \mathbb{P}(|f(q_i)| > \sqrt{p/8}) \geq c.$$

Let $G := \sum_{i=1}^m \eta_i$. Then

$$\mathbb{E}G = \sum_{i=1}^m \mathbb{E}(\eta_i) \geq cm.$$

Since η_i are independent Bernoulli random variables, we can apply the Chernoff bound (see, e.g., [7, Theorem 1.1] with $\varepsilon = 1/2$), obtaining

$$\mathbb{P}\left(G \leq \frac{1}{2}\mathbb{E}G\right) \leq \exp\left(-\frac{1}{8}\mathbb{E}(G)\right).$$

Since $\mathbb{E}(G) \geq cm \geq cn/2$, we observe

$$\mathbb{P}\left(G \leq \frac{cn}{4}\right) \leq \mathbb{P}\left(G \leq \frac{1}{2}\,\mathbb{E}G\right) \leq \exp\left(-\frac{1}{8}\,\mathbb{E}G\right) \leq \exp\left(-\frac{cm}{8}\right) \leq \exp\left(-\frac{cn}{16}\right).$$

Finally, consider the norm $\|Mv\|_2^2 = \sum_{i=1}^m f_i^2$. If the event $\{G > cn/4\}$ occurs, then

$$\|Mv\|_2^2 = \sum_{i=1}^m f_i^2 \geq \sum_{i:\eta_i=1} f_i^2 \geq \sum_{i:\eta_i=1} \left(\sqrt{\frac{p}{8}}\right)^2 \geq \frac{cn}{4}\,\frac{p}{8}.$$

Therefore,

$$\mathbb{P}\left(\|Mv\|_2 > c_0\sqrt{pn}\right) \geq \mathbb{P}\left(G > \frac{cn}{4}\right) \geq 1 - \exp\left(-\frac{cn}{16}\right),$$

where $c_0 = \sqrt{c/32}$. This completes the proof. □

The next lemma provides a bound on the operator norm of the centered random matrix $\|M - \mathbb{E}M\|$. We first note that the expectation $\mathbb{E}M$ is the matrix where each entry is $d/n = p$, hence $\|\mathbb{E}M\| = d$, and that the direct calculations show that

$$\|M - \mathbb{E}M\| = \|M_{|\mathcal{H}}\|,$$

where

$$\mathcal{H} = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\} \tag{2.5}$$

(the subspace orthogonal to the vector $\mathbf{1}_n = (1, 1, \ldots, 1)$). For $d = n/2$, Tran [43, Proposition 2.8] used Bernstein's inequality and a net argument to bound the operator norm $\|M|_{\mathcal{H}}\|$, finding it to be typically $O(\sqrt{n})$ (see also [13, Lemma 5.1]). This type of result, bounding $\|M - \mathbb{E}M\|$, holds more generally for $p = d/n \in (0, 1/2]$ being a fixed constant as stated below. Since the proof follows the same lines, we refer the interested reader to [43, Proposition 2.8] for the core argument. Alternatively, one can use Theorem 4.1 in [19], where such a bound is obtained with a worse (but enough for our purposes) probability.

Lemma 2.5. *Let $p \in (0, 1/2]$ be a fixed constant. Let $1 \le d \le n$ be integers such that $d = pn$. Let M be a random $m \times n$ matrix, $1 \le m \le n$, whose rows are independent random 0/1 vectors having exactly d ones. Then there exist constants $C_{2.5} > 0$ and $c_{2.5} > 0$ depending only on p such that for all $t \ge C_{2.5}$ one has*

$$\mathbb{P}\left(\|M - \mathbb{E}M\| \ge t\sqrt{pn}\right) \le 2e^{-c_{2.5}t^2 n}.$$

2.5 Anticoncentration

Anticoncentration results provide upper bounds on the probability that a random variable falls within a small interval. Such results, applied to sums $\sum_{i=1}^n R_i v_i$, where R_i are rows of a random matrix and v is a fixed unit vector, are crucial for the study of the smallest singular value. Standard techniques often rely on the Least Common Denominator (LCD) for sums of independent random variables, as introduced in [32]. Several adjustments of LCD were used for different models of random matrices with independent entries (see, e.g., [25, 23, 11]). However, the components of the row vectors R_i in our model (uniformly sampled from 0/1-vectors with d ones) are not independent. To handle this dependency, Tran [43] introduced the combinatorial Least Common Denominator (CLCD). The CLCD measures, for a given vector $x \in \mathbb{R}^n$, the closeness of the scaled vector

$$(x_i - x_j)_{1 \le i < j \le n} \in \mathbb{R}^k \quad \text{to} \quad \mathbb{Z}^k, \quad \text{where} \quad k = \binom{n}{2}.$$

The notion of the combinatorial Least Common Denominator was also used in [16] to study the smallest singular value of adjacency matrices for random regular digraphs.

Definition 2.6. *For a vector $v \in \mathbb{R}^n$, and parameters $\gamma \in (0, 1)$, $\alpha > 0$, the combinatorial Least Common Denominator (CLCD) is defined as*

$$CLCD_{\alpha,\gamma}(v) = LCD_{\alpha,\gamma}(D(v)) = \inf_{\theta > 0} \left\{ \mathrm{dist}(\theta D(v), \mathbb{Z}^k) < \min\left\{\gamma\|\theta D(v)\|_2, \alpha\right\} \right\},$$

where $k = \binom{n}{2}$, $D(v)$ is the vector in $\mathbb{R}^k$ with coordinates $v_i - v_j$ for $i < j$.

The following lemma provides a lower bound on the CLCD for non-almost constant vectors (see [43, Lemma 2.15]).

Lemma 2.7. *Let $\delta, \rho \in (0, 1)$, and let $v \in \mathbb{S}^{n-1} \setminus Cons_{\delta,\rho}$. Then for every $\alpha > 0$ and every $\gamma \in (0, \delta\rho/12)$, one has*

$$CLCD_{\alpha,\gamma}(v) \ge \sqrt{\delta n}/7.$$

To quantify anticoncentration, the Lévy concentration function of a random variable X is used.

Definition 2.8. *For a random variable X and $\varepsilon \geq 0$, the Lévy concentration of X of width ε is*

$$\mathcal{L}(X, \varepsilon) = \sup_{x \in \mathbb{R}} \mathbb{P}(|X - x| < \varepsilon).$$

The connection between the CLCD and anticoncentration is established in the following key result which is analogous to standard properties of the LCD from [32]. Its proof is an adaptation of the proof of Theorems 1.5 and 3.2 in [43]. We skip the details.

Lemma 2.9. *Fix $b > 0$ and $\gamma \in (0,1)$. Let $d \leq n$ be positive integers. Let $v \in \mathbb{R}^n$ satisfy $\|D(v)\|_2 \geq b\sqrt{n}$. Suppose that η is a random vector uniformly distributed over n-dimensional $0/1$ vectors having exactly d ones. Then for every $\alpha > 0$ and $\varepsilon \geq 0$,*

$$\mathcal{L}\left(\sum_{i=1}^{n} v_i \eta_i, \varepsilon\sqrt{\frac{d}{n}}\right) \leq \frac{\varepsilon}{\gamma b} + \frac{1}{\gamma b}\sqrt{\frac{n}{d}} \cdot \frac{1}{CLCD_{\alpha,\gamma}(v)} + 2e^{-4\alpha^2 d/n^2}.$$

Remark 2.10. We apply Lemma 2.9 to non-almost constant vectors $v \in \mathbb{S}^{n-1}$. Note that for $v \in \mathbb{S}^{n-1} \setminus \mathrm{Cons}_{\delta,\rho}$, it follows from [21, Lemma 2.2] (cf., [43, Lemma 2.2]) that

$$\|D(v)\|_2 \geq \frac{\delta\rho\sqrt{n}}{4\sqrt{2}}.$$

Setting $\alpha = \mu n$ in CLCD, choosing $b = \delta\rho/4\sqrt{2}$ in Lemma 2.9, and using Lemma 2.7, we obtain for $0 < \gamma < \delta\rho/12$

$$\mathcal{L}\left(\sum_{i=1}^{n} v_i \eta_i, \varepsilon\sqrt{\frac{d}{n}}\right) \leq \frac{4\sqrt{2}\varepsilon}{\gamma\delta\rho} + \frac{28\sqrt{2}}{\gamma\delta^{3/2}\rho\sqrt{d}} + 2e^{-4\mu^2 d}.$$

3 Invertibility on almost constant vectors

In this section, we focus on the invertibility of a random matrix M on the set of almost constant vectors. The main result in this section establishes that with high probability $\|Mv\|_2$ is bounded away from zero on the set $\mathrm{Cons}_{\delta,\rho}$.

Proposition 3.1. *Let $p \in (0, 1/2]$ be a fixed constant. Let d, n be large enough positive integers such that $d = pn$. There exist constants $C_{3.1}, c_{3.1}, \delta, \rho \in (0,1)$ depending on p only such that the following holds. Let M be a random $m \times n$ matrix, $n/2 \leq m \leq n$, whose rows are independent random vectors drawn uniformly from the set of n-dimensional $0/1$ vectors having exactly d ones. Then*

$$\mathbb{P}\left(\inf_{v \in Cons_{\delta,\rho}} \|Mv\|_2 \leq c_{3.1}\sqrt{pn}\right) \leq e^{-C_{3.1} n}.$$

To establish Proposition 3.1, we use a net construction for almost constant vectors with respect to the pseudometric $d(x,y) := \|M(x-y)\|_2$, as provided by [43, Lemma 2.12]. Recall that $\mathcal{H}$ was introduced in (2.5).

Lemma 3.2. *[43, Lemma 2.12] Let $\delta, \rho \in (0, 1/12)$ and n be sufficiently large with respect to δ and ρ. Then there is a net $\mathcal{N} \subset \mathbb{S}^{n-1}$ of cardinality at most $\exp(2n\delta\log(5/\delta))$ such that for any $v \in \mathrm{Cons}_{\delta,\rho}$ there is $w \in \mathcal{N}$ so that for any deterministic $m \times n$ matrix A we have*

$$\|A(v-w)\|_2 \le (\delta + 2\rho)\left(2\|A_{|\mathcal{H}}\| + \frac{\|A\|}{n}\right).$$

Proof of Proposition 3.1. Let $C_{2.3}, c_{2.3}$ be the constants from Lemma 2.3, and $C_{2.5}, c_{2.5}$ be the constants from Lemma 2.5. We choose $\delta, \rho \in (0, 1/12)$ sufficiently small to satisfy conditions specified later. Our goal is to estimate the probability of the event

$$E := \left\{ \inf_{v \in \mathrm{Cons}_{\delta,\rho}} \|Mv\|_2 \le c_{2.3}\sqrt{pn} \right\}.$$

Consider the following event

$$\mathcal{E}_{2.5} := \left\{ M \in \mathcal{M}_{n,d} : \|M - \mathbb{E}M\| < C_{2.5}\sqrt{pn} \right\}.$$

By Lemma 2.5,

$$\mathbb{P}(\mathcal{E}_{2.5}^c) \le 2e^{-Cn}$$

for the constant $C = c_{2.5}C_{2.5}^2 > 0$, depending only on p. We split the event E based on whether $\mathcal{E}_{2.5}$ occurs, as follows,

$$E_1 := \left\{ \inf_{v \in \mathrm{Cons}_{\delta,\rho}} \|Mv\|_2 \le c_{2.3}\sqrt{pn} \quad \text{and} \quad \|M - \mathbb{E}M\| \ge C_{2.5}\sqrt{pn} \right\}$$

and

$$E_2 := \left\{ \inf_{v \in \mathrm{Cons}_{\delta,\rho}} \|Mv\|_2 \le c_{2.3}\sqrt{pn} \quad \text{and} \quad \|M - \mathbb{E}M\| < C_{2.5}\sqrt{pn} \right\}.$$

Then $E = E_1 \cup E_2$, and

$$\mathbb{P}(E) \le \mathbb{P}(E_1) + \mathbb{P}(E_2) \le \mathbb{P}(\mathcal{E}_{2.5}^c) + \mathbb{P}(E_2) \le \mathbb{P}(E_2) + 2e^{-Cn}.$$

Next we bound $\mathbb{P}(E_2)$. Note, $\|\mathbb{E}M\| = pn$. Suppose E_2 occurs, then

$$\|M - \mathbb{E}M\| < C_{2.5}\sqrt{pn}$$

(in particular, for large enough n, $\|M\| \le n$) and there exists a vector $v \in \mathrm{Cons}_{\delta,\rho}$ such that

$$\|Mv\|_2 \le c_{2.3}\sqrt{pn}.$$

Let $\mathcal{N}$ be the net from Lemma 3.2 of cardinality at most $\exp(2n\delta\log(5/\delta))$. Choose $w \in \mathcal{N}$ such that

$$\|M(v-w)\|_2 \le (\delta + 2\rho)\left(2\|M - \mathbb{E}M\| + \frac{\|M\|}{n}\right) \le (\delta + 2\rho)(2C_{2.5}\sqrt{pn} + 1).$$

By the triangle inequality we observe

$$\begin{aligned}\|Mw\|_2 &\le \|Mv\|_2 + \|M(v-w)\|_2 \le c_{2.3}\sqrt{pn} + (\delta+2\rho)(2C_{2.5}\sqrt{pn}+1)\\ &\le c_{2.3}\sqrt{pn} + (9\max\{\delta,\rho\})C_{2.5}\sqrt{pn} \le 2c_{2.3}\sqrt{pn},\end{aligned}$$

provided that

$$\max\{\delta,\rho\} \le c_{2.3}/(9C_{2.5}) \tag{3.1}$$

and d is large enough. Thus

$$E_2 \subset \left\{\inf_{w\in\mathcal{N}} \|Mw\|_2 \le 2c_{2.3}\sqrt{pn}\right\}.$$

By Lemma 2.3, for any fixed $w \in \mathcal{N}$ we have

$$\mathbb{P}\left(\|Mw\|_2 \le 2c_{2.3}\sqrt{pn}\right) \le e^{-C_{2.3}n}.$$

Applying the union bound over the net $\mathcal{N}$,

$$\mathbb{P}(E_2) \le \mathbb{P}\left(\inf_{w\in\mathcal{N}} \|Mw\|_2 \le 2c_{2.3}\sqrt{pn}\right) \le e^{2n\delta\log(5/\delta)}e^{-C_{2.3}n/4} \le e^{-C_1 n},$$

where we assume that δ satisfies

$$2\delta\log(5/\delta) \le C_{2.3}/8 \tag{3.2}$$

and $C_1 = C_{2.3}/8$. Thus, choosing δ and ρ to be small enough to satisfy (3.1) and (3.2), we obtain the desired result. □

4 Proof of Theorem 1.1

We follow the general scheme introduced in [31] (see also [40, 36]). It is based on the following simple implication: for an invertible matrix M and any $u, v > 0$,

$$\exists x \in \mathbb{R}^n : \|x\|_2 \le u,\ \|(M^T)^{-1}x\|_2 \ge v \quad \text{implies} \quad s_n(M) \le u/v.$$

Indeed, recall that $s_n(M) = 1/\|M^{-1}\| = 1/\|(M^T)^{-1}\|$, hence the existence of such x means

$$1/s_n(M) = \|(M^T)^{-1}\| \ge \|(M^T)^{-1}x\|_2/\|x\|_2 \ge v/u.$$

Note that the condition $d = pn$ for a fixed constant $p \in (0, 1/2]$ ensures that M is invertible with high probability. For simplicity, we will assume that M is invertible, with the understanding that this occurs on an event of probability at least $1 - 2e^{-c_p n}$ (see, e.g., [43])

Let R_i denote the i-th row of M and $X_i = R_i^T$, which we consider as the column vector in $\mathbb{R}^n$. Let $H_1 := \text{span}\{X_2, \ldots, X_n\}$. Let $P_1 : \mathbb{R}^n \to H_1$ be the orthogonal projection onto H_1. Let

$$x := X_1 - P_1 X_1.$$

Roughly speaking, this vector x quantifies the degree to which the first row X_1 is linearly independent from the remaining rows $X_2, \dots, X_n$.

Note that x is orthogonal to H_1. Since M is invertible, H_1 has dimension $n-1$, then its orthogonal complement $H_1^\perp$ has dimension 1. Let f_1 be a unit vector spanning $H_1^\perp$. Note $f_1 = x/\|x\|_2$ (up to the sign). Then we have

$$\|x\|_2 = |\langle X_1, f_1 \rangle| = \operatorname{dist}(X_1, H_1).$$

4.1 Bounding $\|x\|_2$

We first bound $\|x\|_2$ from above. By Chebyshev's inequality, for any $u > 0$

$$\mathbb{P}(\|x\|_2 \geq u) \leq \frac{\mathbb{E}\|x\|_2^2}{u^2}. \tag{4.1}$$

To apply this, we compute $\mathbb{E}\|x\|_2^2$ by conditioning on the sigma field generated by H_1. Denoting the k-th coordinates of X_1 and f_1 by $X_1(k)$ and $f_1(k)$ respectively, we observe

$$\begin{aligned}
\mathbb{E}(\|x\|_2^2 \,|\, H_1) = \mathbb{E}(\langle X_1, f_1 \rangle^2 \,|\, H_1) &= \mathbb{E}\left(\left(\sum_{k=1}^n X_1(k) f_1(k) \right)^2 \,|\, H_1 \right) \\
&= \sum_{k=1}^n \mathbb{E}(X_1(k)^2 \,|\, H_1) f_1(k)^2 + \sum_{i \neq j} \mathbb{E}(X_1(i) X_1(j) \,|\, H_1) f_1(i) f_1(j) \\
&= \sum_{k=1}^n \mathbb{E}(X_1(k)^2) f_1(k)^2 + \sum_{i \neq j} \mathbb{E}(X_1(i) X_1(j)) f_1(i) f_1(j),
\end{aligned}$$

where the last line follows from the fact that X_1 is independent of H_1 (spanned by $X_2, \dots, X_n$). Note that

$$\mathbb{E}X_1(k)^2 = \mathbb{E}X_1(k) = \mathbb{P}(X_1(k) = 1) = d/n$$

and for $i \neq j$,

$$\mathbb{E}X_1(i)X_1(j) = \mathbb{P}(X_1(i) = 1 \quad \text{and} \quad X_1(j) = 1) = \binom{n-2}{d-2} \Big/ \binom{n}{d}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}(\|x\|_2^2 \,|\, H_1) &= \frac{d}{n} \sum_{k=1}^n f_1(k)^2 + \frac{d(d-1)}{n(n-1)} \sum_{i \neq j} f_1(i) f_1(j) \\
&= \frac{d}{n} \sum_{k=1}^n f_1(k)^2 + \frac{d(d-1)}{n(n-1)} \left(\left(\sum_{k=1}^n f_1(k) \right)^2 - \sum_{k=1}^n f_1(k)^2 \right) \\
&= \frac{d}{n} \cdot \frac{n-d}{n-1} + \frac{d(d-1)}{n(n-1)} \left(\sum_{k=1}^n f_1(k) \right)^2,
\end{aligned}$$

where the last equality relies on the fact that $\|f_1\|_2^2 = \sum_{k=1}^n f_1(k)^2 = 1$. Thus we get

$$\mathbb{E}\|x\|_2^2 = \mathbb{E}(\mathbb{E}(\|x\|_2^2 \,|\, H_1)) = \frac{d(n-d)}{n(n-1)} + \frac{d(d-1)}{n(n-1)}\,\mathbb{E}\left(\sum_{k=1}^n f_1(k)\right)^2.$$

To bound $\mathbb{E}\left(\sum_{k=1}^n f_1(k)\right)^2$, note that f_1 is orthogonal to H_1, so $\langle f_1, X_i\rangle = 0$ for every $i = 2, \dots, n$. Thus,

$$\left\langle f_1, \sum_{i=2}^n X_i \right\rangle = 0.$$

Therefore,

$$\mathbb{E}\left(\sum_{k=1}^n f_1(k)\right)^2 = \mathbb{E}\,\langle f_1, \mathbf{1}_n\rangle^2 = \mathbb{E}\left(\left\langle f_1, \mathbf{1}_n - \frac{n}{d}\cdot\frac{1}{n-1}\sum_{i=2}^n X_i\right\rangle\right)^2.$$

Since $\|f\|_2 = 1$,

$$\begin{aligned}\mathbb{E}\left(\left\langle f_1, \mathbf{1}_n - \frac{n}{d}\cdot\frac{1}{n-1}\sum_{i=2}^n X_i\right\rangle\right)^2 &\le \mathbb{E}\left\|\mathbf{1}_n - \frac{n}{d}\cdot\frac{1}{n-1}\sum_{i=2}^n X_i\right\|_2^2,\\ &= \sum_{j=1}^n \mathbb{E}\left(1 - \frac{n}{d}\cdot\frac{1}{n-1}\sum_{i=2}^n X_i(j)\right)^2\\ &= \sum_{j=1}^n \frac{n^2}{(d(n-1))^2}\mathbb{E}\left(\frac{d(n-1)}{n} - \sum_{i=2}^n X_i(j)\right)^2.\end{aligned}$$

Since each row R_i is chosen independently, the entries $M_{2j}, \dots, M_{nj}$ in j-th column are independent Bernoulli random variables with the mean d/n. Thus, $\sum_{i=2}^n X_i(j) = \sum_{i=2}^n M_{ij}$ is Binomial$(n-1, d/n)$. Therefore,

$$\mathbb{E}\left(\frac{d(n-1)}{n} - \sum_{i=2}^n X_i(j)\right)^2 = \mathrm{Var}\left(\sum_{i=2}^n X_i(j)\right) = (n-1)\cdot\frac{d}{n}\left(1 - \frac{d}{n}\right).$$

Hence,

$$\mathbb{E}\left\|\mathbf{1}_n - \frac{n}{d}\cdot\frac{1}{n-1}\sum_{i=2}^n X_i\right\|_2^2 = \frac{n(n-d)}{d(n-1)}.$$

This implies

$$\mathbb{E}(\|x\|_2^2) = \frac{d(n-d)}{n(n-1)} + \frac{(d-1)(n-d)}{(n-1)^2} = \frac{n-d}{n-1}\left(\frac{d}{n} + \frac{d-1}{n-1}\right) \le \frac{d}{n} + \frac{d-1}{n-1} \le \frac{3d}{n}.$$

Therefore, by (4.1), for every $u > 0$,

$$\mathbb{P}(\|x\|_2 \ge u) \le \frac{3\,d}{u^2\,n}. \tag{4.2}$$

4.2 Bounding $\|(M^T)^{-1}x\|_2$

Note that

$$\|(M^T)^{-1}x\|_2 = \|(M^T)^{-1}X_1 - (M^T)^{-1}P_1X_1\|_2 = \|e_1 - (M^T)^{-1}P_1X_1\|_2.$$

Since $P_1X_1 \in \text{span}\{X_2,\dots,X_n\}$, the vector $(M^T)^{-1}P_1X_1 \in \text{span}\{e_2,\dots,e_n\}$ is orthogonal to e_1. Denote by $X_k^* = M^{-1}e_k$ the k-th column of the inverse matrix M^{-1} for $k = 1,\dots,n$. We have

$$\|P_1X_1^*\|_2^2 = \langle P_1M^{-1}e_1, P_1M^{-1}e_1\rangle = \langle M^{-1}e_1, P_1M^{-1}e_1\rangle = \langle e_1, (M^{-1})^T P_1M^{-1}e_1\rangle = 0,$$

where we used the fact that $P_1M^{-1}e_1$ belongs to $\text{span}\{X_2,\dots,X_n\}$, hence

$$(M^{-1})^T P_1M^{-1}e_1 \in \text{span}\{e_2,\dots,e_n\}.$$

Therefore, denoting $Y_k := P_1X_k^*$,

$$\begin{aligned}\|(M^T)^{-1}x\|_2^2 &= \|e_1\|_2^2 + \|(M^T)^{-1}P_1X_1\|_2^2 \geq \|(M^T)^{-1}P_1X_1\|_2^2\\ &= \sum_{k=1}^n \langle (M^T)^{-1}P_1X_1, e_k\rangle^2 = \sum_{k=1}^n \langle X_1, P_1X_k^*\rangle^2\\ &= \sum_{k=2}^n \langle X_1, \underbrace{P_1X_k^*}_{=:Y_k}\rangle^2 = \sum_{k=2}^n \langle X_1, Y_k\rangle^2.\end{aligned}$$

The next lemma follows from [31, Lemma 2.1] which works for any invertible matrix. By the result in [43], M is invertible with probability at least $1 - 2e^{-c_p n}$, therefore, we may assume without loss of generality that M^T is invertible in our lemma. In particular, $H_1 := \text{span}\{X_2,\dots,X_n\}$ is $(n-1)$-dimensional.

Lemma 4.1. *Recall that $X_k = R_k^T$ is the k-th column vector of M^T for $k \in [n]$, and P_1 is the orthogonal projection onto H_1. If $Y_k = P_1(M^{-1}e_k)$ for $k = 2,\dots,n$ is defined as above, then $\{X_k, Y_k\}_{k=2}^n$ is a biorthogonal system in H_1.*

The following is a consequence of the uniqueness in Proposition 2.1.

Corollary 4.2. *The system of vectors $\{Y_k\}_{k=2}^n$ defined as in Lemma 4.1 is uniquely determined by the system $\{X_k\}_{k=2}^n$ within the subspace H_1. In particular, the system $\{Y_k\}_{k=2}^n$ and the vector X_1 are independent.*

Denote $H_{i,j} = \text{span}\{X_k : k \notin \{i,j\}\}$. By (2.1), we have $\|Y_k\|_2 = 1/\text{dist}(X_k, H_{1,k})$. Therefore,

$$\|(M^T)^{-1}x\|_2^2 \geq \sum_{k=2}^n \langle X_1, Y_k\rangle^2 = \sum_{k=2}^n \left\langle X_1, \frac{Y_k}{\|Y_k\|_2}\right\rangle^2 \|Y_k\|_2^2 = \sum_{k=2}^n \frac{a_k^2}{b_k^2},$$

where we denoted

$$a_k = |\langle X_1, Y_k\rangle| / \|Y_k\|_2 \quad \text{and} \quad b_k = 1/\|Y_k\|_2 = \text{dist}(X_k, H_{1,k}) \quad \text{for} \quad k \in \{2,\dots,n\}.$$

4.3 Probabilistic bounds for a_k and b_k

Next, we show that for each k, with high probability a_k is bounded from below and b_k is bounded from above. Without loss of generality, we will do this for $k = 2$. The argument for any $k \in \{2, \dots, n\}$ is the same as $k = 2$.

We split the unit sphere into sets of almost constant vectors and non-almost constant vectors which were introduced in Definition 2.2. First, we will show that $H_{1,2}^{\perp}$ consists of non-almost constant vectors with high probability. Consider an $(n-2) \times n$ matrix B with rows $X_3, \dots, X_n$. Since the subspace $H_{1,2}$ is the span of random vectors $X_3, \dots, X_n$, we observe $H_{1,2}^{\perp} \subset \ker(B)$.

By Proposition 3.1 applied to $(n-2) \times n$ matrix B, there exist positive constants $C_{3.1}, c_{3.1}, \delta, \rho \in (0, 1)$ depending only on p, such that

$$\mathbb{P}\left(H_{1,2}^{\perp} \cap \mathrm{Cons}_{\delta,\rho} = \emptyset\right) \geq \mathbb{P}\left(\inf_{v \in \mathrm{Cons}_{\delta,\rho}} \|Bv\|_2 \geq c_{3.1}\sqrt{pn}\right) \geq 1 - e^{-C_{3.1}n}, \tag{4.3}$$

which exactly means that the subspace $H_{1,2}^{\perp}$ consists of not almost constant vectors with high probability.

Fix an absolute constant $\mu > 0$ and let $\gamma \in (0, \delta\rho/12)$. Denote

$$\mathcal{E} := \left\{v \in H_{1,2}^{\perp} \cap \mathbb{S}^{n-1} : \ \mathrm{CLCD}_{\mu n, \gamma}(v) \geq \sqrt{\delta n}/7\right\}$$

By Lemma 2.7 and (4.3),

$$\mathbb{P}(\mathcal{E}) \geq 1 - e^{-C_{3.1}n}. \tag{4.4}$$

Conditioning on the subspace H_1, we fix a realization of vectors $\{X_i\}_{i=2}^n$. Then by the uniqueness in Corollary 4.2 the vector Y_2 is also fixed. Denote $Y := Y_2/\|Y_2\|_2$. By Lemma 4.1, $\{X_i\}_{i=2}^n$ and $\{Y_i\}_{i=2}^n$ form a biorthogonal system. In particular, Y_2 is orthogonal to $\{X_i\}_{i=3}^n$. Thus, $Y \in H_{1,2}^{\perp} \cap \mathbb{S}^{n-1}$. Conditioning on the event $\mathcal{E}$, we know that

$$\mathrm{CLCD}_{\mu n, \gamma}(Y) \geq \sqrt{\delta n}/7.$$

By Lemma 2.9 and parameters chosen as in Remark 2.10, for every $\varepsilon \geq 0$, there exist constants $C_1, C_2 > 0$ depending on ρ, δ, γ such that

$$\mathbb{P}\left(a_2 \leq \varepsilon\sqrt{p} \,|\, X_2, \dots, X_n\right) \leq C_1\varepsilon + \frac{C_2}{\sqrt{d}} + 2e^{-4\mu^2 pn}.$$

Now we unfix all random vectors $X_2, \dots, X_n$ and then by (4.4) we get for every $\varepsilon \geq 0$

$$\begin{aligned}
\mathbb{P}(a_2 \leq \varepsilon\sqrt{p}) &= \mathbb{P}(\{a_2 \leq \varepsilon\sqrt{p}\} \cap \mathcal{E}) + \mathbb{P}(\mathcal{E}^c) \\
&\leq \mathbb{E}_{X_2,\dots,X_n}\mathbb{P}(\{a_2 \leq \varepsilon\sqrt{p} \,|\, X_2, \dots, X_n\} \cap \mathcal{E}) + \mathbb{P}(\mathcal{E}^c) \\
&\leq C_1\varepsilon + \frac{C_2}{\sqrt{d}} + 2e^{-4\mu^2 pn} + e^{-C_{3.1}n}.
\end{aligned}$$

To show a lower bound on b_2, we use a similar argument as in the estimate (4.2) for $\|x\|_2$. We condition on $H_{1,2}$, use Chebyshev's inequality and then take the expectation to obtain for every $t > 0$,

$$\mathbb{P}\left(b_2 \geq t\right) = \mathbb{P}(\operatorname{dist}(X_2, H_{1,2}) \geq t) \leq \frac{3\,d}{t^2\,n}.$$

Hence,

$$\forall t > 0 \qquad \mathbb{P}\left(b_2 \geq t\sqrt{3p}\right) \leq \frac{1}{t^2}.$$

Combining the probability estimates for a_2 and b_2, we obtain that for every $t > 0$ and every $\varepsilon \geq 0$ one has

$$\mathbb{P}\left(a_2 \leq \varepsilon\sqrt{p} \quad \text{or} \quad b_2 \geq t\sqrt{3p}\right) \leq C_1\varepsilon + \frac{C_2}{\sqrt{d}} + 2e^{-4\mu^2 pn} + e^{-C_{3.1}n} + \frac{1}{t^2},$$

where $C_1, C_2 > 0$ are constants depending only on p.

Repeating the above argument for all $k \in \{2, \ldots, n\}$ instead of $k = 2$ we conclude that for any $t > 0$, $\varepsilon \geq 0$, and for large enough d and a large enough constant $C_3 > 0$ (which may depend on p),

$$\mathbb{P}\left(\frac{a_k}{b_k} \leq \frac{\varepsilon}{\sqrt{3}\,t}\right) \leq C_1\varepsilon + \frac{C_3}{\sqrt{d}} + \frac{1}{t^2}. \tag{4.5}$$

To complete the proof we need the following general statement.

Proposition 4.3. *[31, Proposition 2.2] Let Z_k be non-negative random variables for $k \in [n]$. Then for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} Z_k \leq \varepsilon\right) \leq \frac{2}{n}\sum_{k=1}^{n}\mathbb{P}(Z_k \leq 2\varepsilon).$$

We use Proposition 4.3 for $Z_k = (a_k/b_k)^2$ to get

$$\mathbb{P}\left(\|(M^T)^{-1}x\|_2 \leq \frac{\varepsilon\sqrt{n}}{\sqrt{3}\,t}\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{k=2}^{n} Z_k \leq \frac{\varepsilon^2}{3t^2}\right) \leq \frac{2}{n}\sum_{k=2}^{n}\mathbb{P}\left(Z_k \leq \frac{2\varepsilon^2}{3t^2}\right).$$

By (4.5) this implies

$$\mathbb{P}\left(\|(M^T)^{-1}x\|_2 \leq \frac{\varepsilon\sqrt{n}}{\sqrt{3}\,t}\right) \leq 2\sqrt{2}C_1\varepsilon + \frac{2C_3}{\sqrt{d}} + \frac{2}{t^2}.$$

Combining the above estimate with (4.2), we obtain that for every $\varepsilon > 0, u > 0, t > 0$,

$$\begin{aligned}\mathbb{P}\left(s_n(M) \leq \frac{ut}{\varepsilon}\sqrt{\frac{p}{n}}\right) &\geq \mathbb{P}\left(\|x\|_2 \leq u\sqrt{p}/\sqrt{3} \quad \text{and} \quad \|(M^T)^{-1}x\|_2 \geq \frac{\varepsilon\sqrt{n}}{\sqrt{3}\,t}\right)\\ &\geq 1 - C_4\left(\varepsilon + \frac{1}{\sqrt{d}} + \frac{1}{t^2} + \frac{1}{u^2}\right),\end{aligned}$$

where $C_4 > 0$ is a constant depending only on p. To complete the proof, we fix $\varepsilon > 0$, choose $u = t = 1/\sqrt{\varepsilon}$, and adjust the constants.

References

[1] L. Addario-Berry and L. Eslava. Hitting time theorems for random matrices. *Combinatorics, Probability and Computing*, 23(5):635–669, 2014.

[2] E. Aigner-Horev and Y. Person. On sparse random combinatorial matrices. *Discrete Math.*, 345(11):Paper No. 113017, 11, 2022.

[3] A. Basak and M. Rudelson. Sharp transition of the invertibility of the adjacency matrices of sparse random graphs. *Probability Theory and Related Fields*, 180:233–308, 2021.

[4] C. Bordenave and D. Chafai. Around the circular law. *Probability Surveys*, 9:1–89, 2012.

[5] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe. The least singular value of a random symmetric matrix. In *Forum of Mathematics, Pi*, volume 12, page e3. Cambridge University Press, 2024.

[6] N. A. Cook. On the singularity of adjacency matrices for random regular digraphs. *Probability Theory and Related Fields*, 167(1-2):143–200, 2017.

[7] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms.* Cambridge University Press, 2009.

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

[9] A. Ferber, V. Jain, K. Luh, and W. Samotij. On the counting problem in inverse Littlewood–Offord theory. *Journal of the London Mathematical Society*, 103(4):1333–1362, 2021.

[10] A. Ferber, M. Kwan, and L. Sauermann. Singularity of sparse random matrices: simple proofs. *Combinatorics, Probability and Computing*, 31(1):21–28, 2022.

[11] M. Fernandez V. A distance theorem for inhomogenous random rectangular matrices. *arXiv:2408.06309*, 2024.

[12] G. H. Golub and C. F. Van Loan. *Matrix computations.* Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.

[13] V. Jain. Approximate Spielman-Teng theorems for the least singular value of random combinatorial matrices. *Israel Journal of Mathematics*, 242(1):461–500, 2021.

[14] V. Jain, A. Sah, and M. Sawhney. Sharp invertibility of random Bernoulli matrices. *arXiv:2010.06553*, 2020.

[15] V. Jain, A. Sah, and M. Sawhney. Singularity of discrete random matrices. *Geom. Funct. Anal.*, 31(5):1160–1218, 2021.

[16] V. Jain, A. Sah, and M. Sawhney. The smallest singular value of dense random regular digraphs. *International Mathematics Research Notices*, 2022(24):19300–19334, 2022.

[17] J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random ±1 matrix is singular. *Journal of the American Mathematical Society*, 8(1):223–240, 1995.

[18] M. Kwan, B. Sudakov, and T. Tran. Anticoncentration for subgraph statistics. *J. Lond. Math. Soc. (2)*, 99(3):757–777, 2019.

[19] D. Li, A. E. Litvak, and T. Yu. The circular law for sparse random combinatorial matrices. *arXiv:2604.10446*, 2026.

[20] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. Adjacency matrices of random digraphs: singularity and anti-concentration. *Journal of Mathematical Analysis and Applications*, 445(2):1447–1491, 2017.

[21] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. The smallest singular value of a shifted d-regular random square matrix. *Probability Theory and Related Fields*, 173:1301–1347, 2019.

[22] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Advances in Mathematics*, 195(2):491–523, 2005.

[23] A. E. Litvak and K. E. Tikhomirov. Singularity of sparse Bernoulli matrices. *Duke Mathematical Journal*, 171(5):1135–1233, 2022.

[24] G. V. Livshyts. The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding. *Journal d'Analyse Mathématique*, 145(1):257–306, 2021.

[25] G. V. Livshyts, K. Tikhomirov, and R. Vershynin. The smallest singular value of inhomogeneous square random matrices. *The Annals of Probability*, 49(3):1286–1309, 2021.

[26] S. Mendelson and G. Paouris. On the singular values of random matrices. *Journal of the European Mathematical Society (EMS Publishing)*, 16(4), 2014.

[27] H. H. Nguyen. On the singularity of random combinatorial matrices. *SIAM Journal on Discrete Mathematics*, 27(1):447–458, 2013.

[28] H. H. Nguyen and V. H. Vu. Circular law for random discrete matrices of given row sum. *Journal of Combinatorics*, 4(1):1–30, 2013.

[29] E. Rebrova and K. Tikhomirov. Coverings of random ellipsoids, and invertibility of matrices with iid heavy-tailed entries. *Israel Journal of Mathematics*, 227:507–544, 2018.

[30] M. Rudelson and K. Tikhomirov. The sparse circular law under minimal assumptions. *Geom. Funct. Anal.*, 29(2):561–637, 2019.

[31] M. Rudelson and R. Vershynin. The least singular value of a random square matrix is $O(n^{-1/2})$. *Comptes Rendus Mathematique*, 346(15-16):893–896, 2008.

[32] M. Rudelson and R. Vershynin. The Littlewood–Offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(2):600–633, 2008.

[33] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 62(12):1707–1739, 2009.

[34] D. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[35] S. J. Szarek. Condition numbers of random matrices. *Journal of Complexity*, 7(2):131–149, 1991.

[36] T. Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[37] T. Tao and V. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, pages 595–632, 2009.

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

[39] T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *The Annals of Probability*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[40] K. Tatarko. An upper bound on the smallest singular value of a square random matrix. *Journal of Complexity*, 48:119–128, 2018.

[41] K. Tikhomirov. Singularity of random Bernoulli matrices. *Annals of Mathematics*, 191(2):593–634, 2020.

[42] K. Tikhomirov. Quantitative invertibility of non-Hermitian random matrices. *to appear in the ICM 2022 proceedings, arXiv:2206.00601*, 2022.

[43] T. Tran. The smallest singular value of random combinatorial matrices. *arXiv:2007.06318*, 2020.

[44] J. von Neumann. *Collected works. Vol. V: Design of computers, theory of automata and numerical analysis.* The Macmillan Company, New York, 1963. General editor: A. H. Taub.

[45] J. von Neumann and H. H. Goldstine. Numerical inverting of matrices of high order. *Bulletin of the American Mathematical Society*, 53(11):1021 – 1099, 1947.

Alexander E. Litvak
Email address: `alitvak@ualberta.ca`

Dongbin Li
Email address: `dongbin@ualberta.ca`

Tingzhou Yu
Email address: `tingzho1@ualberta.ca`

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, T6G 2R3, Canada.