Rank deficiency of random matrices

Vishesh Jain[*]   Ashwin Sah[†]   Mehtaab Sawhney[‡]

Abstract

Let M_n be a random $n \times n$ matrix with i.i.d. $\mathrm{Bernoulli}(1/2)$ entries. We show that for fixed $k \geq 1$,

$$\lim_{n \to \infty} \frac{1}{n} \log_2 \mathbb{P}[\operatorname{corank} M_n \geq k] = -k.$$

1 Introduction

A fundamental, and intensely studied, problem in combinatorial random matrix theory is the determination of the probability of singularity of $n \times n$ random Bernoulli matrices (i.e. $n \times n$ matrices for which each entry is independently 0 or 1 with equal probability). The study of this problem was initiated in work of Komlós [10]. After intermediate works over a period of over 50 years [9, 14, 15, 2], the breakthrough work of Tikhomirov [16] showed that for any fixed $p \in (0, 1/2]$,

$$\mathbb{P}[M_n(\mathrm{Ber}(p)) \text{ is singular}] = (1 - p + o_n(1))^n,$$

where we use the notation $M_{m \times n}(\xi)$ to denote an $m \times n$ random matrix with i.i.d. entries distributed as ξ, and the lighter notation $M_n(\xi)$ for $M_{n \times n}(\xi)$. Also, $\mathrm{Ber}(p)$ is the random variable which takes on the value 1 with probability p and 0 with probability $1 - p$.

By considering the probability of a row of the matrix being 0, one sees that the result of Tikhomirov is optimal up to the $o_n(1)$ term. Recently, several works [6, 7, 8, 11, 1] have addressed the more refined question of determining the probability of singularity of $M_n(\xi)$ up to a $(1 + o_n(1))$ factor; in contrast, the aforementioned result of Tikhomirov determines this probability only up to a subexponential (in n) factor. While these works have succeeded in the case of sparse Bernoulli matrices (with sparsity allowed to depend on n) [6, 11, 1], as well as in the case of a fixed ξ which is not uniform on its support [7, 8], we note that for the case of $\mathrm{Ber}(1/2)$, the estimate of Tikhomirov remains essentially the best known.

[*]Department of Statistics, Stanford University, Stanford CA 94305, USA E-mail: visheshj@stanford.edu
[†]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
  E-mail: asah@mit.edu
[‡]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
  E-mail: msawhney@mit.edu

A condition equivalent to singularity of $M_n(\xi)$ is that the corank of $M_n(\xi)$ is at least 1. Given this formulation, the following question immediately suggests itself: given $k \geq 1$, what is the probability that the corank of $M_n(\xi)$ is at least k? In the case when $\xi = \mathrm{Ber}(p)$ for fixed $p \in (0, 1/2]$, by considering the event that first k rows of the matrix are identically 0, we see that this probability is at least $(1-p)^{nk}$. The previously best-known upper bound appears to be due to Kahn, Komlós, and Szemerédi [9], who showed that there exists a function $f \colon \mathbb{N} \to \mathbb{R}^+$ with $f(k) \to 0$ as $k \to \infty$ such that

$$\mathbb{P}[\mathrm{corank}\, M_n(\mathrm{Ber}(1/2)) \geq k] \leq f(k)^n.$$

The above simple lower bound shows that the decay of $f(k)$ can be at most 2^{-k}; it has been suggested (cf. [17, Section 4], 'It is tempting to conjecture...') that this rate of decay is essentially sharp, i.e. that

$$\mathbb{P}[\mathrm{corank}\, M_n(\mathrm{Ber}(1/2)) \geq k] = (1/2 + o_n(1))^{kn}.$$

The main result of this paper confirms this belief.

Theorem 1.1. *Fix $p \in (0, 1/2]$ and let $\xi = \mathrm{Ber}(p)$. Fix $k \geq 1$ and $\epsilon > 0$. Then, for $n \geq n_{1.1}(p, k, \epsilon)$, we have*

$$\mathbb{P}[\mathrm{corank}\, M_n(\xi) \geq k] = (1 - p + \epsilon)^{kn}.$$

Remark 1.2. A modification of our proof, with Theorem 2.8 replaced by the corresponding version in [8], shows that for any fixed ξ which is supported on finitely many points,

$$\mathbb{P}[\mathrm{corank}\, M_n(\xi) \geq k] \leq (\max_{z \in \mathbb{R}} \mathbb{P}[\xi = z] + o_n(1))^{kn}.$$

We conjecture that, in general, the following holds.

Conjecture 1.3. *Fix a random variable ξ supported on finitely many points. Fix $k \geq 1$ and $\epsilon > 0$. Then for $n \geq n_{1.3}(\xi, k, \epsilon)$, we have*

$$\mathbb{P}[\mathrm{corank}\, M_n(\xi) \geq k] \leq (\mathbb{P}[\xi = 0] + \epsilon)^{kn} + (\mathbb{P}[\xi_1 = \cdots = \xi_{k+1}] + \epsilon)^n,$$

where $\xi_1, \ldots, \xi_{k+1}$ are independent samples of ξ.

A stronger conjecture is that the dominant contribution to the probability of the corank being at least k comes from the event of having groups of rows which coincide with each other, and a corresponding number of zero rows, or the same for columns (i.e., these 'local' events determine the probability up to an overall $(1 + o_n(1))$ factor). In the case when $\xi = \mathrm{Ber}(p)$, $p \in (0, 1/2]$, the event of having k rows which are identically zero is exponentially more likely than any of these other 'local' events. In concurrent and independent work, Huang [6] has proved this stronger conjecture (with accompanying singular value bounds) for sufficiently sparse Bernoulli matrices i.e. $\xi = \mathrm{Ber}(p_n)$ with

$$1 \leq \liminf_{n \to \infty} \frac{p_n \cdot n}{\log n} \leq \limsup_{n \to \infty} \frac{p_n \cdot n}{\log n} < \infty;$$

it is plausible that, combined with the techniques in [11], the upper bound on the $\limsup$ can be relaxed (perhaps even up to $p_n \leq c$ for some small constant c). In the complementary dense case considered here, we leave the resolution of this stronger conjecture as a subject for future research.

Finally, we mention that in recent years, there have been several other works on the (co)rank of random matrices (e.g. [4, 3]). However, the focus of these works is on the determination of the typical (co)rank of various models, which is completely different from our goal of (sharply) determining the rate of the probability of having corank at least k for matrices which are of full rank with high probability.

1.1 Notation

Given a positive integer $N \geq 1$, let $\mathbb{S}^{N-1}$ be the set of unit vectors in $\mathbb{R}^N$. Let $\|\cdot\|_2$ be the Euclidean norm. For a matrix $A = (A_{ij})$, let $\|A\|$ be its spectral norm (i.e., $\ell^2 \to \ell^2$ operator norm) and let $\|A\|_{\mathrm{HS}}$ be its Hilbert-Schmidt norm, defined as

$$\|A\|_{\mathrm{HS}}^2 = \sum A_{ij}^2.$$

We let $[N]$ denote the discrete interval $\{1, \ldots, N\}$. Given an $m \times n$ matrix A and a subset $S \subseteq [n]$ of columns, we let A_S be the $m \times |S|$ submatrix of A consisting only of the columns in S.

For an $\mathbb{R}^N$-valued random variable ξ and a real number $r \geq 0$, we define the Lévy concentration function by

$$\mathcal{L}(\xi, r) := \sup_{z \in \mathbb{R}^N} \mathbb{P}[\|\xi - z\|_2 \leq r].$$

Note that the case $N = 1$ coincides with the usual (scalar) Lévy concentration function. We let $\ell_1(\mathbb{Z})$ denote the set of functions $f \colon \mathbb{Z} \to \mathbb{R}$ satisfying $\sum_{z \in \mathbb{Z}} |f(z)| < \infty$.

We also make use of asymptotic notation. Given functions f, g, we write $f = O_\alpha(g)$ or $f \lesssim_\alpha g$ to mean $f \leq C_\alpha g$, where C_α is some constant depending on α. We write $f = \Omega_\alpha(g)$ or $f \gtrsim_\alpha g$ to mean $f \geq c_\alpha g$, where $c_\alpha > 0$ is some constant depending on α. We write $f = \Theta_\alpha(g)$ to mean that both $f = O_\alpha(g)$ and $f = \Omega_\alpha(g)$ hold.

Finally, we will omit floors and ceilings when they make no essential difference.

2 Proof of Theorem 1.1

2.1 Preliminaries

We collect some (by now) standard notions in the non-asymptotic theory of random matrices. For parameters $\delta, \rho \in (0, 1)$ and an integer $n \geq 1$, $\mathrm{Comp}_n(\delta, \rho)$ denotes the set of unit vectors in $\mathbb{R}^n$ which have Euclidean distance at most ρ to the set of δn-sparse vectors. $\mathrm{Incomp}_n(\delta, \rho) := \mathbb{S}^{n-1} \setminus \mathrm{Comp}_n(\delta, \rho)$. When the ambient dimension is clear from context, we will drop the subscript n.

We will need to consider the anticoncentration behavior of a vector with respect to i.i.d. $\mathrm{Ber}(p)$ random variables. For this, we will use the threshold function, which was isolated in the work of Tikhomirov [16].

Definition 2.1. *For $p \in (0, 1/2]$, $L \geq 1$, and $x \in \mathbb{R}^n$, we define*

$$\mathcal{T}_p(x, L) := \sup\left\{ t \in (0, 1) : \mathcal{L}\left(\sum_{i=1}^n b_i x_i, t \right) > Lt \right\},$$

where $b_1, \ldots, b_n$ are independent $\mathrm{Ber}(p)$ random variables.

2.2 Overview of the proof

In this subsection, we present the (short) proof of Theorem 1.1, modulo the key Proposition 2.2 and Theorem 2.8, which we will formally state and prove in the subsequent subsections.

Proof of Theorem 1.1. For $\delta, \rho \in (0, 1)$, let $\mathcal{E}_C(\delta, \rho)$ be the event that for every $(n-k) \times n$ sub-matrix A of $M_n(\xi)$, and for all $x \in \mathbb{S}^{n-1}$ such that $Ax = 0$, we have that $x \in \mathrm{Comp}(\delta, \rho)$. In words, the right-kernel unit vectors of every $(n-k) \times n$ sub-matrix of $M_n(\xi)$ are in $\mathrm{Comp}(\delta, \rho)$.

Denote the rows of $M_n(\xi)$ by $R_1, \ldots, R_n$. If $\operatorname{rank} M_n(\xi) \leq n - k$, then there must be some k rows of $M_n(\xi)$ which are in the span of the remaining $n - k$ rows. Since the rows are i.i.d, it follows from the union bound that

$$\mathbb{P}[\operatorname{rank} M_n(\xi) \leq n - k] \leq \mathbb{P}[\mathcal{E}_C(\delta, \rho)] + \mathbb{P}[\operatorname{rank} M_n(\xi) \leq n - k \wedge \mathcal{E}_C(\delta, \rho)^c]$$

$$\leq \mathbb{P}[\mathcal{E}_C(\delta, \rho)] + \binom{n}{k} \mathbb{P}[\{R_1, \ldots, R_k \in \operatorname{span}(R_{k+1}, \ldots, R_n)\} \wedge \mathcal{E}_C(\delta, \rho)^c].$$

In Proposition 2.2, which is the key innovation of this work, we will show that there exist $\delta, \rho \in (0, 1)$ (depending on k, p, ϵ) such that for all sufficiently large n (depending on k, p, ϵ),

$$\mathbb{P}[\mathcal{E}_C(\delta, \rho)] \leq (1 - p + \epsilon)^{kn}.$$

Now, fix this choice of δ, ρ and denote the corresponding event $\mathcal{E}_C(\delta, \rho)$ simply by $\mathcal{E}_C$. Since $\binom{n}{k} \leq n^k \leq (1 + \epsilon)^n$ for n sufficiently large, it remains to show that for all sufficiently large n,

$$\mathbb{P}[\{R_1, \ldots, R_k \in \operatorname{span}(R_{k+1}, \ldots, R_n)\} \wedge \mathcal{E}_C^c] \leq (1 - p + \epsilon)^{nk}.$$

In Theorem 2.8, we show the following dichotomy: consider the $(n - k) \times n$ matrix $M_{(n-k) \times n}(\xi)$ formed by the rows $R_{k+1}, \ldots, R_n$. With probability at least $1 - 4^{-kn}$, either

- every unit vector in the right-kernel of $M_{(n-k) \times n}(\xi)$ is in $\operatorname{Comp}(\delta, \rho)$, or
- there is a unit vector $v = v(R_{k+1}, \ldots, R_n)$ in the right-kernel of $M_{(n-k) \times n}(\xi)$ with

$$\mathcal{T}_p(v, L_{2.8}) \leq (1 - p + \epsilon)^n,$$

where $L_{2.8}$ is a constant depending on k, p, ϵ.

Note that, on the event $\mathcal{E}_C^c$, the first case cannot occur. Let $\mathcal{A}$ denote the set of possible realizations of $R_{k+1}, \ldots, R_n$ for which the second case occurs; for every such realization $a \in \mathcal{A}$, we have a unit vector $v = v(a)$ satisfying the conclusion of the second case. Then,

$$\mathbb{P}[R_1, \ldots, R_k \in \operatorname{span}(R_{k+1}, \ldots, R_n) \wedge \mathcal{E}_C^c]$$

$$\leq \sup_{a \in \mathcal{A}} \mathbb{P}[\langle R_1, v(a) \rangle = \cdots = \langle R_k, v(a) \rangle = 0] + 4^{-kn}$$

$$\leq L_{2.8}^k (1 - p + \epsilon)^{kn} + 4^{-kn}.$$

The last inequality uses the independence of the rows, the definition of the threshold function, and the property $\mathcal{T}_p(v(a), L_{2.8}) \leq (1 - p + \epsilon)^n$. The result now follows upon rescaling ϵ. $\qquad \square$

2.3 Compressible vectors

In this subsection, we prove Proposition 2.2. Recall the event $\mathcal{E}_C(\delta, \rho)$ defined at the start of the proof of Theorem 1.1.

Proposition 2.2. *Fix* $p \in (0, 1/2]$, $k \geq 1$, *and* $\epsilon > 0$. *There exist* $\delta, \rho \in (0, 1)$ *and* n_0 *(depending on* p, k, ϵ*) such that for all* $n \geq n_0$,

$$\mathbb{P}[\mathcal{E}_C(\delta, \rho)] \leq (1 - p + \epsilon)^{kn}.$$

The proof of Proposition 2.2 requires two ingredients, the first of which is the classical Kolmogorov-Lévy-Rogozin anticoncentration inequality.

Lemma 2.3 ([12])**.** *Let* $\xi_1, \ldots, \xi_n$ *be independent random variables. Then, for any real numbers* $r_1, \ldots, r_n > 0$ *and any real number* $r \geq \max_{i \in [n]} r_i$,

$$\mathcal{L}\left(\sum_{i=1}^n \xi_i, r\right) \leq \frac{C_{2.3} r}{\sqrt{\sum_{i=1}^n (1 - \mathcal{L}(\xi_i, r_i)) r_i^2}},$$

where $C_{2.3} > 0$ *is an absolute constant.*

The second ingredient is a version of restricted invertibility which allows one to select a subset of columns of full rank.

Lemma 2.4 ([5, Theorem 1]). *Let U be an $n \times m$ matrix of rank n. Then, there exists a subset $\mathcal{S} \subseteq [m]$ of columns of size $|\mathcal{S}| = n$ such that*

$$\|U_{\mathcal{S}}^{-1}\|_{\mathrm{HS}}^2 \le (m - n + 1) \cdot \mathrm{Tr}[(UU^T)^{-1}].$$

The following is the key lemma in the proof of Proposition 2.2.

Lemma 2.5. *Fix $p \in (0, 1/2]$ and $k \ge 1$. There exists $\theta = \theta_{2.5}(p, k) > 0$ for which the following holds. If M is a $k \times n$ matrix whose rows are orthonormal vectors and x is an n-dimensional random vector with independent $\mathrm{Ber}(p)$ components, then*

$$\mathcal{L}(Mx, \theta) \le (1 - p)^k.$$

Remark 2.6. Our proof shows that $\theta(p, k)$ can be taken to be of size $\mathrm{poly}(p)\exp(-\Omega(k))$. We suspect this is far from the truth and conjecture that one can take $\theta(p, k)$ to be of size $\mathrm{poly}(p, 1/k)$.

Proof. For each $1 \le i \le k$, let $\mathcal{T}_i \subseteq [n]$ be the indices $j \in [n]$ corresponding to the $\lfloor C_{2.3}^2 25^k/p \rfloor$ (which is $\ge k$) largest values $|M_{ij}|$, where (for concreteness) we break ties according to the natural ordering of the integers. Let $\mathcal{T} = \cup_{i=1}^k \mathcal{T}_i$. Since $\sum_{j=1}^n |M_{ij}|^2 = 1$ for every $i \in [k]$, it follows that for every $j \in \mathcal{T}^c$,

$$|M_{ij}| \le 5^{-k}\sqrt{p}/C_{2.3} \quad \forall i \in [k].$$

We have two cases.

Case I: There exists some $i \in [k]$ such that $\sum_{j \in \mathcal{T}^c} |M_{ij}|^2 \ge 4^{-k}$. In this case, by applying Lemma 2.3 with $r = 5^{-k}\sqrt{p}/(3C_{2.3})$ and $r_j = |M_{ij}|/3$, we find that

$$\mathcal{L}\left(\sum_{j \in \mathcal{T}^c} M_{ij}x_j, r\right) \le C_{2.3}\frac{C_{2.3}^{-1}5^{-k}\sqrt{p}/3}{2^{-k}\sqrt{p/9}} < 2^{-k} \le (1 - p)^k.$$

Since

$$\mathcal{L}(Mx, r) \le \mathcal{L}((Mx)_i, r) \le \mathcal{L}\left(\sum_{j \in \mathcal{T}^c} M_{ij}x_j, r\right),$$

we have the required conclusion with $\theta = r = 5^{-k}\sqrt{p}/(3C_{2.3})$.

Case II: For every $i \in [k]$, $\sum_{j \in \mathcal{T}^c} |M_{ij}|^2 \le 4^{-k}$. Let $U = M_{\mathcal{T}}$ denote the $k \times |\mathcal{T}|$ matrix formed by the columns of M corresponding to $\mathcal{T}$. Note that $k \le |\mathcal{T}| \le C_{2.3}^2 k 25^k/p$. Let $C_1, \dots, C_n$ denote the columns of M. Since $MM^T = I$, we have

$$\|I - UU^T\| = \|MM^T - UU^T\| \le \sum_{j \in \mathcal{T}^c} \|C_j C_j^T\|$$

$$= \sum_{j \in \mathcal{T}^c} \|C_j\|_2^2 = \sum_{i \in [k]} \sum_{j \in \mathcal{T}^c} |M_{ij}|^2$$

$$\le k4^{-k} < 1/2.$$

In particular, the eigenvalues of UU^T are in $(1/2, 1]$ so that

$$\mathrm{Tr}[(UU^T)^{-1}] < 2k.$$

Therefore, by Lemma 2.4, there exists a set of columns $\mathcal{S} \subseteq \mathcal{T}$ of size $|\mathcal{S}| = k$ such that

$$\|M_{\mathcal{S}}^{-1}\|_{\mathrm{HS}}^2 = \|U_{\mathcal{S}}^{-1}\|_{\mathrm{HS}}^2 \le 2C_{2.3}^2 k^2 25^k/p.$$

Hence, the smallest singular value of $M_{\mathcal{S}}$, denoted by $\sigma_k(M_{\mathcal{S}})$, must satisfy

$$\sigma_k(M_{\mathcal{S}}) \geq \frac{\sqrt{p}}{2C_{2.3}k5^k}.$$

We claim that for

$$r = \frac{\sqrt{p}}{5C_{2.3}k5^k},$$

we have that

$$\mathcal{L}(Mx, r) \leq (1-p)^k.$$

Since

$$\mathcal{L}(Mx, r) \leq \mathcal{L}(M_{\mathcal{S}}x_{\mathcal{S}}, r),$$

it suffices to show that

$$\mathcal{L}(M_{\mathcal{S}}x_{\mathcal{S}}, r) \leq (1-p)^k.$$

Since $x_{\mathcal{S}}$ is a k-dimensional vector with i.i.d. $\mathrm{Ber}(p)$ entries, it follows that $x_{\mathcal{S}}$ is supported on $\{0,1\}^k$ with maximum atom probability $(1-p)^k$. Moreover, by definition of the smallest singular value, we see that for any $x \neq y \in \{0,1\}^k$,

$$\|M_{\mathcal{S}}x - M_{\mathcal{S}}y\|_2 \geq \sigma_k(M_{\mathcal{S}})\|x-y\|_2 \geq \sigma_k(M_{\mathcal{S}}),$$

which shows that

$$\mathcal{L}(M_{\mathcal{S}}x_{\mathcal{S}}, r) \leq \max_{x \in \{0,1\}^k} \mathbb{P}[x_{\mathcal{S}} = x] = (1-p)^k.$$

Thus, in either case, we can take

$$\theta = \frac{\sqrt{p}}{5C_{2.3}k5^k}. \qquad \square$$

The previous lemma allows us to quickly deduce the following which, in the special case $k = 1$, is the usual 'invertibility with respect to a single vector' (cf. [16, Lemma 3.5]).

Lemma 2.7. *Fix $p \in (0, 1/2]$, $k \geq 1$, and $\epsilon > 0$. Let $\xi = \mathrm{Ber}(p)$. There exists $c_{2.7} = c_{2.7}(p, k, \epsilon) > 0$ for which the following holds. Let $n \geq n_{2.7}(p, k, \epsilon)$ and let $V = [v_1, \ldots, v_k]$ be an $n \times k$ matrix with orthonormal columns. Then,*

$$\mathbb{P}[\|M_{(n-k)\times n}(\xi)V\|_{\mathrm{HS}} \leq c_{2.7}\sqrt{n}] \leq (1-p+\epsilon)^{kn}.$$

Proof. Denote the rows of $M_{(n-k)\times n}(\xi)$ by $R_1, \ldots, R_{n-k}$. Let $\theta = \theta_{2.5}(p, k)$. Note that if

$$\|M_{(n-k)\times n}(\xi)V\|_{\mathrm{HS}} \leq \theta\sqrt{\epsilon'(n-k)},$$

then at most $\epsilon'(n-k)$ rows R_i can satisfy $\|R_iV\|_2 > \theta$. Denote the set of these rows by $\mathcal{I}$. By Lemma 2.5 and the independence of the rows, we have for all $I \subseteq [n]$, $|I| \leq \epsilon'(n-k)$ that

$$\mathbb{P}[\mathcal{I} = I] \leq \mathbb{P}[\|R_iV\|_2 \leq \theta \ \forall i \in I^c] \leq \prod_{i \in I^c} \mathbb{P}[\|V^T(R_i)^T\|_2 \leq \theta] \leq (1-p)^{k\cdot(1-\epsilon')(n-k)}.$$

Therefore, by the union bound over the choice of I, we have

$$\mathbb{P}[\|M_{(n-k)\times n}(\xi)V\|_{\mathrm{HS}} \leq \theta\sqrt{\epsilon'(n-k)}] \leq n \cdot \binom{n-k}{\epsilon'(n-k)}(1-p)^{k\cdot(1-\epsilon')(n-k)}.$$

Therefore taking $c_{2.7} = \theta\sqrt{\epsilon'}/2$ for sufficiently small $\epsilon' = \epsilon'(p, k, \epsilon)$ and taking n sufficiently large gives the desired conclusion. $\qquad \square$

Combining this with a standard epsilon-net argument allows us to prove Proposition 2.2.

Proof of Proposition 2.2. Let $\mathcal{E}(\delta, \rho)$ denote the event that for every $x \in \mathbb{S}^{n-1}$ such that

$$(M_{(n-k)\times n}(\xi))x = 0,$$

we have $x \in \operatorname{Comp}(\delta, \rho)$. By the union bound,

$$\mathbb{P}[\mathcal{E}_C(\delta, \rho)] \leq \binom{n}{k} \mathbb{P}[\mathcal{E}(\delta, \rho)],$$

so that (after rescaling ϵ) it suffices to show that

$$\mathbb{P}[\mathcal{E}(\delta, \rho)] \leq (1 - p + \epsilon)^{kn}.$$

Since the right-kernel of $M_{(n-k)\times n}(\xi)$ has dimension at least k, it follows that on the event $\mathcal{E}(\delta, \rho)$, we can find k orthonormal vectors, $v_1, \ldots, v_k$, such that $v_i \in \operatorname{Comp}(\delta, \rho)$ and $(M_{(n-k)\times n}(\xi))v_i = 0$. Let V denote the $n \times k$ matrix with columns $v_1, \ldots, v_k$. Then,

$$M_{(n-k)\times n}(\xi)V = 0. \tag{2.1}$$

Let $H \subseteq \mathbb{R}^n$ denote the subspace of vectors $x = (x_1, \ldots, x_n)$ such that $x_1 + \cdots + x_n = 0$. Let $\mathcal{E}_K$ be the event that the operator norm of $(M_{(n-k)\times n}(\xi))|_H$ (i.e., the linear operator from H to $\mathbb{R}^{n-k}$ which coincides with $M_{(n-k)\times n}(\xi)$ on H) is at most $K\sqrt{n}$. Since ξ is sub-Gaussian, it follows from standard estimates (cf. [16, Lemma 3.4]) that

$$\mathbb{P}[\mathcal{E}_K] \geq 1 - \exp(-c_\xi K^2 n).$$

In particular, by choosing K to be of order $\sqrt{k}$, we can ensure that this probability is at least $1 - 4^{-kn}$. Then, by the union bound, it suffices to show that

$$\mathbb{P}[\mathcal{E}(\delta, \rho) \cap \mathcal{E}_K] \leq (1 - p + \epsilon)^{kn}.$$

We will show this by combining Lemma 2.7 with a standard epsilon-net argument. Let $\epsilon' > 0$ be a sufficiently small parameter to be chosen later. A standard volumetric net, 'densified' in the all-ones direction shows (cf. [16, Proposition 3.6]) that there exists a (deterministic) net $\mathcal{N}$ of $\operatorname{Comp}(\delta, \rho)$, of size at most $(C/\epsilon')^{\delta n}$ (where C is allowed to depend on K), such that for any $x \in \operatorname{Comp}(\delta, \rho)$, there exists $y \in \mathcal{N}$ such that on the event $\mathcal{E}_K$,

$$\|M_{(n-k)\times n}(\xi)(x - y)\|_2 \leq (\epsilon' + \rho)K\sqrt{n}.$$

Let $\mathcal{V}$ denote the set of all $n \times k$ matrices whose columns are orthonormal vectors in $\operatorname{Comp}(\delta, \rho)$. Then, by considering the k-fold product of $\mathcal{N}$, we obtain a net of $\mathcal{V}$. Using the standard trick of replacing points in this net by the closest point in $\mathcal{V}$ (see e.g. [13, Lemma 4.2]), we can obtain a (deterministic) net $\mathcal{M} \subseteq \mathcal{V}$ of size $|\mathcal{M}| \leq (C/\epsilon')^{\delta kn}$ such that for every $V \in \mathcal{V}$, there exists $V' \in \mathcal{M}$ such that, on the event $\mathcal{E}_K$,

$$\|M_{(n-k)\times n}(\xi)(V - V')\|_{\mathrm{HS}} \leq 2(\epsilon' + \rho)K\sqrt{kn}.$$

Therefore, by (2.1) and the union bound, we have

$$\mathbb{P}[\mathcal{E}(\delta, \rho) \cap \mathcal{E}_K] \leq \mathbb{P}[\mathcal{E}_K \cap \{\exists V \in \mathcal{V} : M_{(n-k)\times n}(\xi)V = 0\}]$$

$$\leq \sum_{V' \in \mathcal{M}} \mathbb{P}[\|M_{(n-k)\times n}(\xi)V'\|_{\mathrm{HS}} \leq 2(\epsilon' + \rho)K\sqrt{kn}]$$

$$\leq \left(\frac{C}{\epsilon'}\right)^{\delta kn} \cdot (1 - p + \epsilon/2)^{kn}$$
$$\leq (1 - p + \epsilon)^{kn},$$

where the penultimate line follows from Lemma 2.7 by taking $\rho = \epsilon'$, n sufficiently large, and ϵ' sufficiently small depending on k, p, ϵ, and the last line follows by taking n sufficiently large and δ sufficiently small depending on ϵ', ϵ. $\qquad\square$

2.4 Incompressible vectors

Recall the notion of the threshold of a vector (Definition 2.1). The following is the structure theorem/dichotomy used in the proof of Theorem 1.1. The case $k = 1$ is implicit in [16], although the statement given here is closer to the one in work of the authors [8, Proposition 3.7].

Theorem 2.8 (Modification of [8, Proposition 3.7])**.** *Let $\delta, \rho, \epsilon \in (0, 1)$ and $k \geq 1$. There exist $L_{2.8} = L_{2.8}(k, \delta, \rho, p, \epsilon)$ and $n_{2.8} = n_{2.8}(k, \delta, \rho, p, \epsilon)$ such that for all $n \geq n_{2.8}$, with probability at least $1 - 4^{-kn}$, exactly one of the following holds.*

- *Every unit vector v in the right-kernel of $M_{(n-k) \times n}(\xi)$ is in $\mathrm{Comp}(\delta, \rho)$, or*
- *there is a unit vector v in the right-kernel of $M_{(n-k) \times n}(\xi)$ with $\mathcal{T}_p(v, L_{2.8}) \leq (1 - p + \epsilon)^n$.*

Remark 2.9. The proof of the above theorem, for the case $k = 1$, is the main contribution of [16] and requires several deep and powerful ideas. This is, by far, the most technically difficult part of the argument. However, given [16], the modification presented above follows in a straightforward manner. For concreteness, we compare the statement with that of [8, Proposition 3.7]. The first difference is that in Theorem 2.8, we are missing k rows, as opposed to 1 row in [8], but since k is sufficiently small compared to n, this difference in row counts has essentially no effect on the union bound computation presented in [8]. The second difference is that in Theorem 2.8, we are considering the independent threshold model rather than the "multislice" models considered in [8], which actually simplifies the proof.

References

[1] Anirban Basak and Mark Rudelson, *Sharp transition of the invertibility of the adjacency matrices of sparse random graphs*, arXiv:1809.08454. MR4265022

[2] Jean Bourgain, Van H. Vu, and Philip Matchett Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), no. 2, 559–603. MR2557947

[3] Amin Coja-Oghlan, Alperen A Ergür, Pu Gao, Samuel Hetterich, and Maurice Rolvien, *The rank of sparse random matrices*, Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2020, pp. 579–591. MR4141217

[4] Kevin P Costello and Van H Vu, *The rank of random graphs*, Random Structures & Algorithms **33** (2008), no. 3, 269–285. MR2446482

[5] E. Gluskin and A. Olevskii, *Invertibility of sub-matrices and the octahedron width theorem*, Israel J. Math. **186** (2011), 61–68. MR2852316

[6] Han Huang, *Rank of sparse Bernoulli matrices*, arXiv:2009.13726.

[7] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Sharp invertibility of random Bernoulli matrices*, arXiv:2010.06553.

[8] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Singularity of discrete random matrices*, arXiv:2010.06554.

[9] Jeff Kahn, János Komlós, and Endre Szemerédi, *On the probability that a random ± 1-matrix is singular*, J. Amer. Math. Soc. **8** (1995), no. 1, 223–240. MR1260107

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

[11] Alexander E Litvak and Konstantin E Tikhomirov, *Singularity of sparse Bernoulli matrices*, arXiv:2004.03131. MR4076632

[12] B. A. Rogozin, *On the increase of dispersion of sums of independent random variables*, Teor. Verojatnost. i Primenen **6** (1961), 106–108. MR0131894

[13] Mark Rudelson, *Recent developments in non-asymptotic theory of random matrices*, Modern aspects of random matrix theory, Proc. Sympos. Appl. Math., vol. 72, Amer. Math. Soc., Providence, RI, 2014, pp. 83–120. MR3288229

[14] Terence Tao and Van Vu, *On random* ± 1 *matrices: singularity and determinant*, Random Structures Algorithms **28** (2006), no. 1, 1–23. MR2187480

[15] Terence Tao and Van Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628. MR2291914

[16] Konstantin Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), no. 2, 593–634. MR4076632

[17] Van Vu, *Recent progress in combinatorial random matrix theory*, arXiv:2005.02797. MR4260513

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm