The circular law for sparse random combinatorial matrices

Dongbin Li, Alexander E. Litvak, and Tingzhou Yu

Abstract

Let $\log^{2+\varepsilon} n \le d \le n/2$ for some fixed $\varepsilon \in (0,1)$, and let M_n 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. We show that the empirical spectral distribution of the appropriately rescaled matrix M_n converges in probability to the circular law provided that $d = o(n)$. As a crucial element of our proof, we obtain quantitative lower bounds on the smallest singular value of the shifted matrices $M_n - z\mathbf{I}_n$ whenever $|z| \le \sqrt{d} \log\log d$ and $C \log n \le d \le n/2$ for some absolute positive constant C.

AMS 2010 Classification: primary: 60B20, 15B52, 60B10; secondary: 60C05, 05C80, 46B06.
Keywords: circular law, invertibility, random graph, random matrix, regular graphs, singular probability, smallest singular value, sparse matrix.

Contents

1 Introduction

For an $n \times n$ matrix A, its empirical spectral distribution (ESD) is defined as the probability measure

$$\mu_A := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i},$$

where $(\lambda_i)_{1\le i\le n}$ denotes the eigenvalues of A. A fundamental phenomenon in the study of the ESD of a random matrix is universality, which asserts that, under very general conditions, the ESD behaves similarly to that of a Gaussian random matrix of the appropriate symmetry type. This phenomenon has been confirmed for various models, and interested readers are referred to [5, 6, 46] for further details.

A well-studied class of non-Hermitian random matrices is the one with i.i.d. entries. Denote by μ_{circ} the uniform probability measure on the unit disk of the complex plane, that is, the probability measure having the density $\mu_{circ} = \pi^{-1}\mathbf{1}_{|z|\le 1}$ with respect to the Lebesgue measure. After a long succession of important papers dating back to the 1960s [7, 19, 24, 25, 52, 45], Tao and Vu [53] finally established that the ESD of properly rescaled random matrices with i.i.d. entries distributed as a fixed random variable ξ converges almost surely to μ_{circ} as $n \to \infty$. While this result effectively captures the limiting spectral behavior of *dense random matrices*, it is less informative when it comes to *sparse random matrices*, such as Bernoulli random matrices with the parameter $p = p_n \to 0$ as $n \to \infty$. The sparse i.i.d. model for different ensembles (mostly for a product of a mean zero random variable and a Bernoulli random variable) was subsequently considered in papers [25, 52, 8, 47, 50, 56]. In particular, Theorem 1.4 in [50] yields that the circular law holds for sparse Bernoulli(p_n) matrices as long as $np_n \to \infty$ and $p_n \to 0$. We summarize these results in the following theorem.

Theorem 1.1. *Let $\{p(n)\}_{n\ge 1}$ be a sequence in $(0,1)$ satisfying two conditions: $p(n)n \to \infty$ and $p(n) \to 0$. For every $n \ge 1$, let $\mathcal{B}_{np(n)}$ be an $n \times n$ matrix with i.i.d. entries distributed as Bernoulli random variables with parameter p. Then the ESD of $\frac{\mathcal{B}_{np(n)}}{\sqrt{p(n)(1-p(n))n}}$ converges in probability to the circular law.*

Remark 1.2. Of course the term $(1-p(n))$ in the normalization of the Bernoulli random matrix can be removed as $p(n) \to 0$. However, we prefer to keep it, as Theorem 1.1 remains true for the constant sequence $\{p(n)\}_{n\ge 1}$ (that is, $p(n) = p(1)$ for every $n \ge 1$)) in which case this term is important. This follows from Corollary 1.12 in [53].

Non-Hermitian matrices with dependent entries have been studied extensively. In [12], Bordenave, Caputo and Chafaï showed that the random Markov matrices obey the circular law. Nguyen and Vu [44] established the circular law for random ± 1 matrices with prescribed row sums $|s| \le (1-\varepsilon)n$ for a fixed $\varepsilon \in (0,1]$. Later, Nguyen [43] proved the circular law for random doubly stochastic matrices. In [1] Adamczak and Chafaï showed that random real matrices having an unconditional log-concave distribution obey the circular law. Shortly after, Adamczak, Chafaï and Wolff [2] proved the circular law for random matrices with exchangeable entries having finite moments of order $20+\delta$. In [15, 38], the authors proved that the empirical spectral distribution of properly rescaled random d-regular matrices, that is, $0/1$ matrices in which every row and column sums to d, converges to the circular law as d grows to infinity with n.

In this paper, we consider the following model of non-Hermitian matrices: Let $\mathcal{M}_{n,d}$ consist of all $n \times n$ matrices with entries in $\{0, 1\}$ where each row has a fixed sum of d. Our random matrix M_n is uniformly drawn from the set $\mathcal{M}_{n,d}$. In some works such matrices are called random combinatorial matrices, so we follow this terminology. Note that the rows of M_n are independent random vectors. An element $M_n \in \mathcal{M}_{n,d}$ 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_n 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. Our first main result is the circular law for properly rescaled random combinatorial matrices.

Theorem 1.3. *Fix $\varepsilon \in (0, 1)$. For each $n \geq 1$ Let $d = d(n)$ satisfy $\min\{d, n-d\} \geq \log^{2+\varepsilon} n$ and $\min\{d, n-d\}/n \to 0$, and let M_n be drawn uniformly from $\mathcal{M}_{n,d}$. Denote*

$$\overline{M_n} = \frac{M_n}{\sqrt{d(1-d/n)}}.$$

Then the sequence of empirical spectral distributions $\left(\mu_{\overline{M_n}}\right)_{n\geq 1}$ converges in probability to μ_{circ} as $n \to \infty$.

Remark 1.4. Denote by E_n the matrix with all entries equal to 1. Similarly to [15, Remark 1.5]) (cf., [53, Corollary 1.12]), we observe that M_n satisfies circular law iff $E_n - M_n$ does. Thus, we may assume that $d \leq n/2$.

Remark 1.5. The result for the dense regime $\min(d, n-d) = cn$ for some $c \in (0, 1/2]$ was obtained by Nguyen and Vu [44]. In their work, random ± 1 matrices A_n with prescribed row sums s satisfying $|s| \leq (1-\varepsilon)n$ for some fixed $\varepsilon \in (0, 1]$ are considered. Since $(A_n + E_n)/2$ has the same distribution as our random combinatorial matrices M_n with $d = (n+s)/2 \in [\varepsilon n/2, (1-\varepsilon/2)n]$. Our proof also works for the case $\min(d, n-d) = cn$.

Remark 1.6. Denote by X the number of zero columns in the random matrix M_n. Clearly, $\mathbb{E}X = n(1-d/n)^n$, which implies that for any fixed d, the multiplicity of the zero eigenvalue is bounded from below by a constant proportion of n with a positive probability (see Section 8 for a proof). Therefore, one cannot expect that the circular law holds when d does not grow with n. This is also consistent with the numerical evidence shown in Figure 1, where for $d = 2, n = 5000$ the empirical spectral distribution is visibly inconsistent with the circular law. On the other hand, as d increases, the empirical spectrum becomes progressively closer to the unit disk. This suggests that Theorem 1.3 may remain valid as long as $d \to \infty$ with n at any speed as indicated by the graphs above, where relatively large d are chosen.

For a fixed d, determining the limiting behavior of the ESD of M_n, denoted by μ_{M_n}, remains a largely open problem. Interestingly, the limiting law can vary significantly across different random matrix models. For example, in a recent work, Sah, Sahasrabudhe, and Sawhney [49], showed that for fixed $d > 0$, the ESD of random matrices with i.i.d. Bernoulli(d/n) entries converges in probability to some distribution μ_d as n goes to infinity. Meanwhile, it is well known that the limiting ESD of the adjacency matrices of random undirected d-regular simple graphs converges, as $n \to \infty$, to the Kesten-McKay law on the real line with density

$$\frac{d}{2\pi} \frac{\sqrt{4(d-1) - x^2}}{d^2 - x^2} \mathbf{1}_{[-2\sqrt{d-1}, 2\sqrt{d-1}]}(x)$$

with respect to the Lebesgue measure [32, 41, 10]. In contrast, for random d-regular graphs with fixed $d \geq 3$, it was conjectured in [13, Section 7] that the ESD of their adjacency matrices

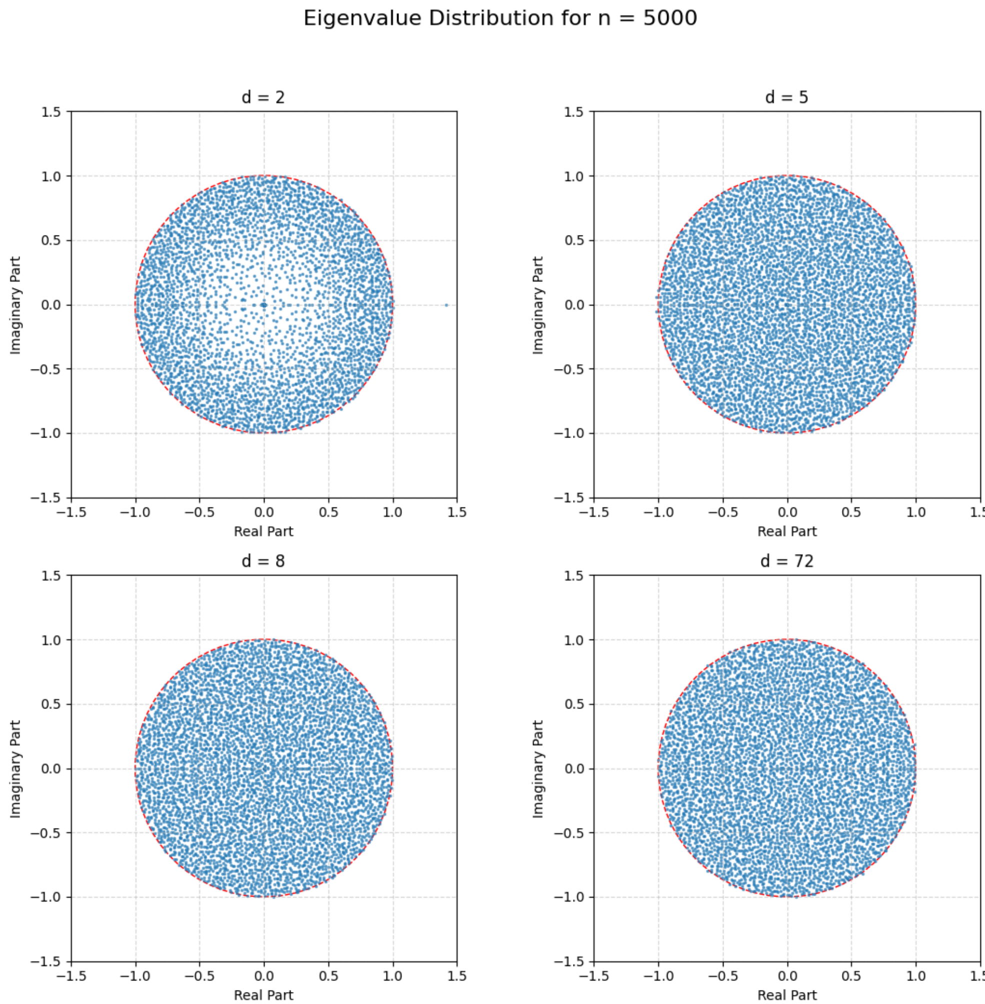

Figure 1: Empirical spectra of four independently generated 5000×5000 random combinatorial matrices where each row has exactly $d \in \{2, 5, 8, 72\}$ ones. In each panel, the plotted points are the eigenvalues of the normalized matrix $\overline{M_n} = M_n/\sqrt{d(1-d/n)}$. The red dashed circle is the unit circle $\{z \in \mathbb{C} : |z| = 1\}$. For $d = 2$, the empirical spectral distribution is visibly inconsistent with the circular law. However, as d increases, the spectrum becomes progressively closer to uniform on the unit disk.

converges in probability to the oriented Kesten-McKay law on $\mathbb{C}$ with density

$$\frac{1}{\pi}\frac{d^2(d-1)}{(d^2-|z|^2)^2}\mathbf{1}_{\{|z|\le\sqrt{d}\}}$$

with respect to the Lebesgue measure. This conjecture was recently confirmed for the case of random d-regular simple graphs (i.e., graphs without self-loops) in the work of Adhikari and Dembo [3] (see also related results by Beker [11]).

The proof of Theorem 1.3 crucially relies on establishing a quantitative lower bound for the smallest singular value of the shifted matrices $M_n - z\mathbf{I}_n$, which is another main contribution of our paper (see Theorem 1.7 below). We briefly outline the connection here. For many Hermitian random matrix models, techniques such as the method of moments and the Stieltjes transform yield precise asymptotic information about the empirical spectral distribution (ESD). However, these methods break down in the non-Hermitian setting. In such cases, the standard alternative is to convert the limiting ESD to the distribution of the singular values via the Hermitization technique going back to Girko [24]. The utility of this technique can also be captured by the *replacement principle*, which was first introduced by Tao and Vu in [53]. This gives a sufficient condition for the ESD of two random matrices to have the same limit.

To better illustrate the methods, we recall that the *logarithmic potential* $U_\mu : \mathbb{C} \to (-\infty, \infty]$ of a probability measure μ on $\mathbb{C}$ is defined by

$$U_\mu(z) = -\int_{\mathbb{C}} \log|z-\lambda| \, \mathrm{d}\mu(\lambda), \quad z \in \mathbb{C}.$$

For a given $n \times n$ matrix A, let $(s_i(A))_{1\le i\le n}$ denote the sequence of singular values (arranged in the non-increasing order) of A and ν_A denote the empirical singular value distribution of A, which is

$$\nu_A := \frac{1}{n}\sum_{i=1}^n \delta_{s_i(A)}.$$

Then one can check that

$$U_{\mu_{M_n}}(z) = -\frac{1}{n}\log|\det(M_n - z\mathbf{I}_n)| = -\frac{1}{n}\sum_{i=1}^n \log(s_i(M_n - z\mathbf{I}_n)) = -\int_0^\infty \log(t)\,\mathrm{d}\nu_{M_n - z\mathbf{I}_n}(t).$$

The logarithmic potential function uniquely determines the underlying measures, in particular, a sequence of probability measures $(\mu_n)_{n\ge1}$ converges to μ almost surely (resp., in probability) if $(U_{\mu_n}(z))_{n\ge1}$ converges almost surely (resp., in probability) to $U_\mu(z)$ for almost all complex numbers z [51, Theorem 2.8.3]. Therefore, to prove Theorem 1.3, it suffices to show that $U_{\mu_{\overline{M_n}}}(z)$ converges to $U_{\mu_{circ}}(z)$ in probability for almost all complex numbers z. To this end, one usually needs to establish

(a) $\nu_{\overline{M_n} - z\mathbf{I}_n}$ converges in probability to some probability measure ν_z such that

$$U_{\mu_{circ}}(z) = -\int_0^\infty \log(t)\,\mathrm{d}\nu_z(t)$$

and

(b) the uniform integrability of $\log(t)$ with respect to $\nu_{\overline{M_n} - z\mathbf{I}_n}$ (see, e.g., [13, Lemma 4.3]).

Part (a) is usually achieved via a comparison argument with Bernoulli and Gaussian random matrices (see, e.g., [15, Proposition 7.2]), while part (b) requires a good control on $s_n(\overline{M_n} - z\mathbf{I}_n)$ as well as on its intermediate singular values (see, e.g., [38, 8] for more details).

The replacement principle we use in the present paper has the same spirit as the Hermitization technique but enables us to make a more direct comparison with the Bernoulli random matrices. To be more precise, let $\mathcal{B}_{np}$ be an $n \times n$ random matrix with i.i.d. Bernoulli(p), $p = d/n$, entries, and normalize it as before,

$$\overline{\mathcal{B}_{np}} := \frac{\mathcal{B}_{np}}{\sqrt{d(1-d/n)}}.$$

One needs a good control of the Hilbert–Schmidt norms of $\overline{M_n}$ and $\overline{\mathcal{B}_{np}}$ and to verify that $U_{\mu_{\overline{M_n}}}(z) - U_{\mu_{\overline{\mathcal{B}_{np}}}}(z)$ converges to zero in probability for almost all $z \in \mathbb{C}$ (with respect to the Lebesgue measure). These two conditions guarantee that $\mu_{\overline{M_n}} - \mu_{\overline{\mathcal{B}_{np}}}$ converges weakly to zero in probability. Since Theorem 1.1 implies that $\mu_{\overline{\mathcal{B}_{np}}}$ converges to the circular law in probability, Theorem 1.3 now follows.

As mentioned above, one of the key steps in almost all known approaches to the circular law is obtaining quantitative lower tail bounds on the smallest singular value of shifted matrices $\overline{M_n} - z\mathbf{I}_n$ for a.e. $z \in \mathbb{C}$. Specifically, one needs to show that for a.e. $z \in \mathbb{C}$,

$$\mathbb{P}(s_n(\overline{M_n} - z\mathbf{I}_n) \le \varepsilon) = o(1), \tag{1.1}$$

for some $\varepsilon \ge \exp(-n^{o(1)})$. Obtaining (1.1) is a generalization of the following estimates,

$$\mathbb{P}(s_n(M_n) \le \varepsilon) = o(1) \qquad \text{or} \qquad \mathbb{P}(s_n(M_n) = 0) = o(1). \tag{1.2}$$

The problem of proving (1.2) with M_n being random matrices with i.i.d. entries has been studied extensively, and interested readers are referred to [48, 54, 9, 40, 28] and the references therein for a more comprehensive discussion. In contrast, solving problem (1.2) 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 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 [14], despite the analogous result for non-symmetric matrices being established nearly 17 years earlier by Rudelson and Vershynin [48]. 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 random d-regular matrices have been studied extensively (see, e.g., [17, 30, 36, 37]). For random combinatorial matrices, the study of (1.2) was initiated by Nguyen [42], who proved that if $d = n/2$, then $\mathbb{P}(s_n(M_n) = 0) = O(n^{-C})$ for any $C > 0$; moreover, in the same paper, it was conjectured that the sharp singularity probability is $(\frac{1}{2} + o(1))^n$. Nguyen's result has been strengthened by several authors [21, 27, 55], and finally, the sharp singularity probability was obtained by Jain, Sah and Sawhney [29]. In the dense regime $d = pn$ with $p \in (0, 1/2]$ fixed, a matching upper bound of order $s_n(M_n) = O(\sqrt{d}/n)$ with high probability was recently obtain in [33]. Combined with corresponding lower bounds (see, e.g., [55]), this identifies the correct scale $n^{-1/2}$ for the least singular value. In the sparse case, it was Aigner-Horev and Person [4] who first showed that $\mathbb{P}(s_n(M_n) = 0) = o(1)$ under the assumption that $\lim_{n\to\infty} d/(n^{1/2}\log^{3/2} n) = \infty$. Shortly after, it was shown by Ferber, Kwan and Sauermann [22] that M_n is asymptotically almost surely non-singular when $\min(d, n-d) \ge (1+\varepsilon)\log n$ for a fixed $\varepsilon > 0$. It is worth noting that the threshold $\log n$ is sharp, in the sense that M_n contains a zero column asymptotically almost surely when $\min(d, n-d) \le (1-\varepsilon)\log n$ for a fixed $\varepsilon > 0$ (see Section 8 below). However, obtaining a quantitative lower tail bound on the smallest singular value in the sparse regime remains largely unsolved. In this paper, we establish such a lower tail bound. Namely, we prove the following general theorem, which also plays a crucial role in the proof of the circular law.

Theorem 1.7. *There exist absolute positive constants C_1, C_2, C_3, $C_{1.7} \ge 1$ such that the following holds. Let n be a large enough integer and $C_{1.7}\log n \le d \le n/2$. Let M_n be an $n \times n$ random matrix uniformly drawn from the set $\mathcal{M}_{n,d}$. Let $z \in \mathbb{C}$ satisfy $|z| \le \sqrt{d}\log\log d$. Then*

$$\mathbb{P}(s_n(M_n - z\mathbf{I}_n) \le n^{-\beta}) \le C_1/\sqrt{d},$$

for some $\beta = \beta(n, d) \in (2, C_2 \log\log n]$. Moreover, if $d \ge (\log n)^2$ then $\beta < 9$, if $d \ge C_3(\log n)^{21}$ then $\beta < 5$.

Remark 1.8. Setting $z = 0$, we observe that

$$\mathbb{P}(s_n(M_n) \le n^{-\beta}) \le C_1/\sqrt{d}.$$

To the best of our knowledge, this is the first time a quantitative lower bound for $s_n(M_n)$ has been obtained in the sparse regime $d \ge C\log n$. Moreover, for any fixed $\varepsilon \in (0, 1)$, with probability close to 1, M_n is invertible whenever $\min(d, n-d) \geq (1+\varepsilon)\log n$ [22]; and with probability close to 1, M_n contains a zero column if $d \leq (1-\varepsilon)\log n$ (see Section 8 for a proof). This implies that at $z = 0$, the range of d is almost optimal in our theorem.

Remark 1.9. In fact our proof gives $\mathbb{P}(s_n(M_n - z\mathbf{I}_n) \leq \alpha(n,d)) \leq C_1/\sqrt{d}$, where

(1) if $d \geq c_0 n^{2/3}(\log n)^{1/3}$ then $\alpha(n,d) \geq \dfrac{1}{2n\, d^{3/2}} \geq \dfrac{1}{n^{5/2}}$;

(2) if $d \leq c_0 n^{2/3}(\log n)^{1/3}$ then $\alpha(n,d) \geq \dfrac{d^{3\alpha-5.5}}{n^{2\alpha}(\log n)^{\alpha-1/2}} \geq n^{-2\alpha}$;

for

$$\alpha = \frac{2\log(6d)}{\log(d/(C_2 \log n))} \in (2, C_4 \log\log n].$$

In particular, for large enough n,

(3) if $d \geq C_3(\log n)^{21}$ then $\alpha < 2.1$ and $\alpha(n,d) \geq \dfrac{d^{0.8}}{n^{4.2}(\log n)^{1.6}} \geq \dfrac{1}{n^{4.2}}$;

(4) if $d \geq (\log n)^2$ then $\alpha < 4.1$ and $\alpha(n,d) \geq \dfrac{d^{6.8}}{n^{8.2}(\log n)^{3.6}} \geq \dfrac{1}{n^{8.2}}$.

Here, c_0, C_1, C_3, C_4 are appropriate positive absolute constants.

The remainder of the introduction focuses on the main ideas behind the proof of Theorem 1.7, and for ease of exposition, we take $z = 0$. An often employed approach to estimating the least singular value, introduced in [39] (see also [35]), is to partition the complex unit sphere and to work separately with different types of vectors. In this context, one usually splits the sphere into vectors of small complexity (close to sparse vectors) and "spread" vectors with a relatively small ℓ_∞-norm. Such splitting was further formalized in [48] into a concept of "compressible" and "incompressible" vectors to obtain sharp lower tail bounds for the smallest singular value of random matrices with i.i.d. subgaussian entries. For non-Hermitian random matrices with dependent entries such as d-regular random matrices and random combinatorial matrices, the concept of compressible and incompressible vectors is not directly applicable. We use a different splitting by partitioning the complex unit sphere $\mathbb{S}^{2n-1} \subset \mathbb{C}^n$ into two parts. The first part consists of *almost constant vectors*, defined as follows:

$$\mathrm{Cons}(\delta,\rho) = \left\{ x \in \mathbb{S}^{2n-1} : \exists \lambda \in \mathbb{C} \text{ such that } \left| \left\{ i \leq n : |x_i - \lambda| \leq \frac{\rho}{\sqrt{n}} \right\} \right| > (1-\delta)n \right\},$$

where $\delta, \rho \in (0,1)$, and the second part is its complement. The prototype of this structural dichotomy dates back to Cook's work [17] and was also used and refined in various papers (see, e.g., [15, 36, 37, 34, 55]).

The set $\mathrm{Cons}(\delta,\rho)$, roughly speaking, consists of vectors obtained by shifting very sparse vectors by constant vectors. Therefore, it has low complexity, which enables us to use a standard ε-net argument to show that the images of such vectors are bounded away from 0. This result in turn helps to reveal the arithmetic structure of a unit normal vector to the subspace spanned by $n-1$ rows of M_n, which allows one to obtain a sharp small ball probability in the subsequent calculation (see, [55] for more details). Since our random matrices have neither i.i.d. entries nor d-regularity, M_n and M_n^T are not equally distributed. We therefore have to demonstrate that for all $x \in \mathrm{Cons}(\delta,\rho)$, $\|\bar{x}M_n\|_2$ and $\|M_n x\|_2$ are bounded away from zero separately, and different difficulties arise in establishing these bounds.

To bound $\|M_n x\|_2$, one may use the independence between rows; however, in the sparse regime, the individual probability for a fixed vector $x \in \mathbb{S}^{2n-1}$, as computed using the method for the dense regime (see, e.g., [55, 28]), is too large to beat the metric entropy of $\mathrm{Cons}(\delta,\rho)$ when employing the ε-net argument. If one tries to further reduce the complexity of almost constant vectors by shrinking δ and ρ, it consequently impairs the arithmetic structure of a unit normal vector, rendering the small ball estimation ineffective. Meanwhile, bounding $\|\bar{x}M_n\|_2$ is considerably more difficult because the columns of M_n (or the rows of M_n^T) are not independent. In [28, 55], the authors addressed the left-multiplication problem for the dense case $d = n/2$ via a conditioning argument: conditioning on the first $k-1$ columns of M with $1 \le k \le n/4$, the entries in the k-th column are independent Bernoulli variables with success probability $p \in [1/3, 2/3]$. Hence, the Paley-Zygmund inequality together with a tensorization argument yields a good individual probability. However, in the sparse regime, the sparsity leads to a lack of sufficient randomness, preventing the use of the conditioning argument to establish the desired individual probability necessary for the ε-net argument.

To overcome the difficulty due to the sparsity, we adopt the strategy from [37, 34, 40], which involves a refined version of the aforementioned splitting. Specifically, we define four (overlapping) classes on $\mathbb{C}^n$, which we call steep vectors (vectors possessing a significant jump) , gradual vectors (that are not steep), almost constant vectors, and non-almost constant vectors. To be more precise, given a vector $x = (x_i)_{1\le i\le n} \in \mathbb{C}^n$, denote by $(x_i^*)_{1\le i\le n}$ the non-increasing rearrangement of the sequence $(|x_i|)_{1\le i\le n}$. The vector x is said to be steep, if $x_k^* \gg x_m^*$ for some $k \ll m$.

To deal with steep vectors, we first explore the expansion properties of our random matrices using the negative association of suitably constructed random variables (a similar result was also proved in [36] for d-regular matrices but using a pure combinatorial approach). The expansion properties guarantee that there are relatively many rows that have exactly one entry equal to 1 within the set of columns corresponding to large coordinates of a steep vector x. Then the inner product of each such row with the steep vector x is dominated by the absolute value of the unique large coordinate due to the large jump and sparsity of the matrices. Therefore, to fully take advantage of the expansion properties, we need to choose the jump size and its location carefully. In addition, the selection of the size of the jump and its location also directly impacts the construction of the associated nets. Furthermore, to effectively implement the net argument, it is crucial to obtain a precise estimate of the operator norm $\|M_n - \mathbb{E}M_n\|$ (which is equal to the operator norm of M_n restricted to the subspace of vectors having zero sum of coordinates). This is accomplished via the Kahn-Szemerédi argument. The argument was first introduced by Kahn and Szemerédi in [23] to bound the second eigenvalue of a random graph from the permutation model, and later was adapted to establish bounds on the spectral gap for several other random graph models (see, e.g., [16, 57] and the references therein).

Bounding the magnitude of the matrix-vector product for almost constant gradual vectors is straightforward. Loosely speaking, without big jumps, any nonzero almost constant gradual vector x is essentially very close to a constant nonzero vector $\lambda_0\mathbf{1}$, so the ℓ_2-norms are also comparable. Moreover, by the concentration, each column contains approximately d ones, therefore the inner product of each row or column with an almost constant gradual vector is roughly $d\lambda_0$. Hence, we can deduce that both $\|\bar{x}M_n\|_2$ and $\|M_n x\|_2$ are bounded below.

After bounds for the above two classes are obtained, it remains to deal with non-almost constant vectors. The infimum over the non-almost constant vectors can be tackled by relating it to the average distance of a row of the matrix M_n to the subspace spanned by the rest of the rows (similar reductions were used in various papers, see, e.g., [48, 37, 53]). Note that the distance is further lower bounded by the absolute value of the inner product between a row and a unit normal vector to the subspace. Moreover, since the rows of our random matrices are independent, via conditioning, we may assume the row and the unit normal vector are independent. Now the bounds for the first two classes imply that, with high probability, the normal vector belongs to the class of non-almost constant vectors. This enables us to apply an anti-concentration inequality in [49, Lemma 8.5] to complete the proof of Theorem 1.7.

We finally discuss the organization of the paper. In the next section we introduce notation and some preliminary results. In Section 3, we will prove Theorem 1.3 using Theorem 1.7 and Theorem 1.1. As mentioned in Section 1, one key condition that needs to be verified is that $U_{\mu_{\overline{M_n}}}(z) - U_{\mu_{\overline{\mathcal{B}_{np}}}}(z)$ converges to zero in probability for almost all z. This is equivalent to showing that

$$\frac{1}{n}\log|\det(\overline{M_n} - z\mathbf{I}_n)| - \frac{1}{n}\log|\det(\overline{\mathcal{B}_{np}} - z\mathbf{I}_n)| \longrightarrow 0 \tag{1.3}$$

in probability for almost all complex numbers z. To this end, we adapt the approach in [44] (see also [53]) by expressing the determinant as a product of distances from a fixed row vector to successive subspaces, in accordance with the base-times-height formula. We then control the distances using either the lower bound for the least singular value or Talagrand's concentration inequality (see Theorem 3.8 below). Then the rest of the paper will be mostly devoted to the proof of Theorem 1.7. In Section 4, we obtain a sharp upper bound on the operator norm $\|M_n - \mathbb{E}M_n\|$ using the Kahn-Szemerédi argument — a general approach for bounding the second largest eigenvalues of random regular graphs. In Section 5, we focus on the expansion properties of our random matrices, which also crucially determine the definition of the steep vectors in Section 6. To explore the expansion properties, we use the negative association of suitably constructed random variables. In Section 6, we show that for any almost constant non-zero vector x and any complex number z with $|z| \leq \sqrt{d}\log\log d$ with high probability both $\|\bar{x}(M_n - z\mathbf{I}_n)\|_2/\|\bar{x}\|_2$ and $\|(M_n - z\mathbf{I}_n)x\|_2/\|x\|_2$ are bounded from below by a constant depending on n and d. The argument is technically involved since we are required to distinguish several types of jumps as well as different jump locations and combine these with a very delicate construction of the ε-nets. More details will be discussed there. We prove Theorem 1.7 in Section 7. The key idea is to show that with large probability the distance of a row of M_n from the subspace spanned by the rest of the rows cannot be too small. This requires combining the results in Section 6 with an application of anti-concentration inequality in Lemma 6.4. Finally, in the last section, we briefly discuss the singularity threshold.

2 Notations and preliminaries

In the rest of the paper, we simply write our random matrix M_n as M omitting the index n.

2.1 Notations

Given a natural number n, we denote $[n] := \{1, 2, \dots, n\}$. For $x \in \mathbb{C}^n$, let $\sigma_x : [n] \to [n]$ be a permutation such that $|x_{\sigma_x(1)}| \geq |x_{\sigma_x(2)}| \geq \cdots \geq |x_{\sigma_x(n)}|$. Let $x^* \in \mathbb{C}^n$ be the non-increasing rearrangement of x defined in the following way: $x_i^* = |x_{\sigma_x(i)}|$ for $i \in [n]$.

For a vector $x \in \mathbb{C}^n$ (or $x \in \mathbb{R}^n$), its canonical Euclidean norm is denoted by $\|x\|_2$. With a slight abuse of notation we denote the unit sphere in $\mathbb{C}^n$ by

$$\mathbb{S}^{2n-1} := \{z \in \mathbb{C}^n : \|z\|_2 = 1\}$$

and the unit sphere in $\mathbb{R}^n$ by

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

We also denote

$$\mathbb{S}_0^{2n-1} := \Big\{v \in \mathbb{S}^{2n-1} : \sum_{i=1}^n v_i = 0\Big\} \qquad \text{and} \qquad \mathbb{S}_0^{n-1} := \Big\{v \in \mathbb{S}^{n-1} : \sum_{i=1}^n v_i = 0\Big\}.$$

Given a matrix A, we denote its i-th row by $R_i = R_i(A)$. and i-th column by $\mathbf{Col}_i(A)$. For $n \times n$ zero-one matrices we denote

$$\text{Supp } R_i = \text{Supp}(R_i(A)) := \{j \in [n] : A_{ij} = 1\}.$$

Denote $\mathbf{1} = (1, 1, ..., 1) \in \mathbb{C}^n$ and let $\mathbf{I}_n$ denote the $n \times n$ identity matrix. Given a matrix A, A^T is the transpose of A. Given a (column) vector $x \in \mathbb{C}^n$, the transpose of x is denoted by x^T.

Given an $m \times n$ matrix A, the operator norm of A acting from ℓ_2^n to ℓ_2^m is denoted by $\|A\|$, that is,

$$\|A\| = \sup_{\|x\|_2 = 1} \|Ax\|_2.$$

Given two real numbers a and b we use the standard notation $a \vee b = \max\{a, b\}$. For a complex number $z \in \mathbb{C}$, we denote by $\Re(z)$ its real part.

2.2 Concentration

Below we will use the following particular case of the Bennett inequality (see, e.g., [40, Lemma 3.4]).

Lemma 2.1. *Let $p \in (0, 1/2]$, $m \geq 1$. Let $\delta_1, ..., \delta_m$ be independent Bernoulli(p) random variables. Then for $\tau > e$,*

$$\mathbb{P}\left(\sum_{i=1}^m \delta_i > (1+\tau)pm\right) \leq \exp(-\tau \log(\tau/e)pm),$$

and for $0 < \tau < 1$,

$$\mathbb{P}\left(\sum_{i=1}^m \delta_i > (1+\tau)pm\right) \leq \exp\left(-\frac{mp\tau^2}{2(1-p)}(1-\tau/3)\right) \leq \exp\left(-\frac{mp\tau^2}{3(1-p)}\right).$$

The following immediate consequence of Bennett's lemma controls the number of ones in each column, which will be used later in several places, in particular in Theorem 1.3 and in Lemmas 4.9 and 6.9.

Lemma 2.2. *Let $n \geq 2$, $m \geq 1$, $\beta \geq 4$, and $1 \leq d \leq n/2$. Let M be an $m \times n$ 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. Given $\tau > 0$, consider the event*

$$\mathcal{E}_{2.2} = \mathcal{E}_{2.2}(\tau) := \left\{M = \{M_{ij}\}_{i \in [m], j \in [n]} : \sum_{i=1}^m M_{ij} \leq \frac{(1+\tau)md}{n} \quad \text{for every } j \leq n\right\}.$$

Then for $\tau = \max\left\{e^2, \frac{\beta n \log m}{md}\right\}$,

$$\mathbb{P}(\mathcal{E}_{2.2}) \geq 1 - n \cdot m^{-\beta},$$

and for $0 < \tau < 1$,

$$\mathbb{P}(\mathcal{E}_{2.2}) \geq 1 - n \exp\left(-\frac{md\tau^2}{3n}\right).$$

Proof. Fix $j \in [n]$. Note that $M_{1j}, \ldots, M_{mj}$ are independent Bernoulli(d/n) random variables. By Lemma 2.1, for $\tau = \max\left\{e^2, \frac{\beta n \log m}{md}\right\}$,

$$\mathbb{P}\left(\sum_{i=1}^m M_{ij} > \frac{(1+\tau)md}{n}\right) \leq \exp(-\tau \log(\tau/e)(dm/n)) \leq m^{-\beta},$$

and for $0 < \tau < 1$,

$$\mathbb{P}\left(\sum_{i=1}^{m} M_{ij} > \frac{(1+\tau)md}{n}\right) \le \exp\left(-\frac{md\tau^2}{3n(1-d/n)}\right) \le \exp\left(-\frac{md\tau^2}{3n}\right).$$

Then the union bound over $j \le n$ completes the proof . □

2.3 Negative association

In Sections 4 and 5 we will use the notion of negatively associated (NA) random variables, which allows us to work with lightly dependent random variables (such as entries of M) as if they were independent.

Definition 2.3. *(Negative association) Random variables $X_i, i \in [n]$, are said to be negatively associated (NA) if for all disjoint subsets $I, J \subset [n]$ and all coordinate-wise non-decreasing functions $f : \mathbb{R}^I \to \mathbb{R}$ and $g : \mathbb{R}^J \to \mathbb{R}$ one has*

$$\mathbb{E}f((X_i)_{i\in I})\, g((X_j)_{j\in J}) \le \mathbb{E}f((X_i)_{i\in I})\, \mathbb{E}g((X_j)_{j\in J}). \tag{2.1}$$

Remark 2.4. Not that equivalently (passing to functions $-f$ and $-g$) we can require that f and g are coordinate-wise non-increasing in the definition. (cf, [31]).

The following two properties of negative association are very useful in the applications (see, e.g., [31, Properties 4 and 6]).

1. **Closure under products.** If $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are two independent families of negatively associated random variables, then the family

$$X_1, \dots, X_n,\, Y_1, \dots, Y_m$$

   is also negatively associated.

2. **Disjoint monotone aggregation.** If $X_i, i \in [n]$, are negatively associated random variables, $\mathcal{A}$ is a family of disjoint subsets of $[n]$, and for each $A \in \mathcal{A}$, $f_A : \mathbb{R}^A \to \mathbb{R}$ is a coordinate-wise non-decreasing function, then the random variables

$$f_A((X_i)_{i\in A}), \quad A \in \mathcal{A},$$

   are also negatively associated.

Lemma 2.5. *[31, Theorem 2.6] Let $X_1, \dots, X_n$ be a sequence of independent random variables. Suppose that the conditional expectation $\mathbb{E}\left(f((X_i)_{i\in A}) \mid \sum_{i\in A} X_i\right)$ is increasing in $\sum_{i\in A} X_i$ for every proper subset A of $[n]$ and every coordinate-wise increasing function $f : \{0,1\}^{|A|} \to \mathbb{R}$. Then the conditional distributions of $X_1, \dots, X_n$ given $\sum_{i\in[n]} X_i$ is NA almost surely.*

Lemma 2.6. *Let $n, m \ge 1$ and let M be an $m \times n$ 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. Then the entries $(M_{ij})_{1\le i\le m, 1\le j\le n}$ are negatively associated.*

Proof. Let $\delta_1, \cdots, \delta_n$ be i.i.d. Bernoulli(p) random variables with $p = d/n$, note that for each $i \in [m]$, $(M_{i1}, \dots, M_{in})$ has the same distribution as the conditional distribution of $(\delta_1, \cdots, \delta_n)$ given $\sum_{i=1}^n \delta_i = d$. Therefore, by Lemma 2.5, it suffices to show that the conditional expectation

$$\mathbb{E}\left\{ f((\delta_i)_{i\in A}) \mid \sum_{i\in A} \delta_i \right\}$$

is increasing in $\sum_{i\in A}\delta_i$, for every $\ell \le n$, every proper subset A of $[n]$ with $|A| = \ell$ and every coordinate-wise increasing function $f : \{0,1\}^{|A|} \to \mathbb{R}$.

Fix ℓ, $A \subset [n]$ with $|A| = \ell$ and a coordinate-wise increasing function $f : \{0,1\}^{|A|} \to \mathbb{R}$. Consider the function g defined by

$$g(t) := \mathbb{E}\left\{ f((\delta_i)_{i\in A}) \mid \sum_{i\in A}\delta_i = t \right\}$$

for every $0 \le t \le \min(\ell, d)$. We first show that g is increasing. Indeed, given a set $B \subset A$, let $f(B)$ denote $f((w_i)_{i\in A})$, where $w_i = 1$ if $i \in B$ and $w_i = 0$ otherwise. Then

$$g(t) = \frac{1}{\binom{\ell}{t}} \sum_{S'\subset A, |S'|=t} f(S')$$

and

$$\begin{aligned} g(t+1) &= \frac{1}{\binom{\ell}{t+1}} \sum_{S\subset A, |S|=t+1} f(S) = \frac{1}{\binom{\ell}{t+1}} \frac{1}{t+1} \sum_{\substack{S\subset A\\ |S|=t+1}} \sum_{\substack{S'\subset S\\ |S'|=t}} f(S) \\ &= \frac{1}{\binom{\ell}{t+1}} \frac{1}{t+1} \sum_{\substack{S'\subset A\\ |S'|=t}} \sum_{\substack{S\supset S'\\ |S|=t+1}} f(S) \ge \frac{1}{\binom{\ell}{t+1}} \frac{1}{t+1} \sum_{\substack{S'\subset A\\ |S'|=t}} (\ell - t) f(S') = g(t). \end{aligned}$$

Thus, for every fixed row with index $i \in [m]$, the family $\mathcal{C}_i := \{M_{ij} : j \in [n]\}$ is NA. Meanwhile, since the rows of M are independent, the families $\{\mathcal{C}_i\}_{i\in[m]}$ are independent NA families. Therefore, by the first property ("closure under products") of NA,

$$\bigcup_{i\in[m]} \mathcal{C}_i = \{M_{ij} : i \in [m], j \in [n]\}$$

is an NA family. □

Let Q be an $m \times n$ deterministic matrix, and let M be as in Lemma 2.6. Denote

$$f_Q(M) := Q \circ M = \sum_{i\in[m], j\in[n]} Q_{ij} M_{ij}, \tag{2.2}$$

$$\mu := \mathbb{E} f_Q(M) = \frac{d}{n} \sum_{i\in[m], j\in[n]} Q_{ij}, \qquad \text{and} \qquad \sigma^2 := \frac{d}{n} \sum_{i\in[m], j\in[n]} Q_{ij}^2.$$

We show Bennett's inequality for $f_Q(M)$ using negative association.

Theorem 2.7. *Let $n, m \ge 1$. Let M be as in Lemma 2.6. Let Q be an $m \times n$ deterministic matrix with all entries in $[0, K]$. Then for all $t \ge 0$,*

$$\max\{\mathbb{P}(f_Q(M) - \mu \ge t), \mathbb{P}(f_Q(M) - \mu \le -t)\} \le \exp\left(-\frac{\sigma^2}{K^2} h\left(\frac{Kt}{\sigma^2}\right)\right) \le \exp\left(-\frac{3t^2}{6\sigma^2 + 2Kt}\right),$$

where $h(u) = (1+u)\log(1+u) - u$ for $u \ge 0$.

Proof. The second inequality follows from the elementary inequality $h(u) \ge \frac{u^2}{2(1+u/3)}$ for $u \ge 0$. We proceed to proving the first inequality. By homogeneity, we may assume that $K = 1$. Since Q_{ij} are nonnegative, by disjoint monotone aggregation, $(Q_{ij} M_{ij})_{i\in[m], j\in[n]}$ is again an NA family. We begin by proving a bound for the upper tail. For $\lambda \geq 0$, by Chebyshev's inequality, one has

$$\mathbb{P}(f_Q(M) - \mu \geq t) \leq e^{-\lambda(t+\mu)}\mathbb{E}e^{\lambda f_Q(M)} \leq e^{-\lambda(t+\mu)} \prod_{i\in[m],j\in[n]} \mathbb{E}e^{\lambda Q_{ij}M_{ij}},$$

where for the second inequality, we used that $(Q_{ij}M_{ij})_{i\in[m],j\in[n]}$ is an NA family.

Consider the function $\phi(u) := e^u - u - 1$ on $\mathbb{R}$. By taking a continuous extension at $u = 0$, one can readily verify that $\phi(u)/u^2$ is increasing on $[0, \infty)$. Applying this to $u = \lambda Q_{ij}M_{ij} \leq \lambda$, we observe

$$\frac{\phi(\lambda Q_{ij}M_{ij})}{(Q_{ij}M_{ij})^2} \leq \phi(\lambda).$$

This implies that

$$\begin{aligned}\mathbb{E}e^{\lambda Q_{ij}M_{ij}} &= \mathbb{E}\left(1 + \lambda Q_{ij}M_{ij} + \phi(\lambda Q_{ij}M_{ij})\right) \leq 1 + \lambda\mathbb{E}(Q_{ij}M_{ij}) + \phi(\lambda)\mathbb{E}(Q_{ij}M_{ij})^2 \\ &= 1 + \frac{\lambda d Q_{ij}}{n} + \frac{\phi(\lambda) d Q_{ij}^2}{n} \leq \exp\left(\frac{\lambda d Q_{ij}}{n} + \frac{\phi(\lambda) d Q_{ij}^2}{n}\right).\end{aligned}$$

Thus,

$$e^{-\lambda(t+\mu)} \prod_{i\in[m],j\in[n]} \mathbb{E}e^{\lambda Q_{ij}M_{ij}} \leq e^{-\lambda t + \sigma^2\phi(\lambda)}.$$

Optimizing over $\lambda \geq 0$, we obtain the desired bound on $\mathbb{P}(f_Q(M) - \mu \geq t)$.

Next we show the lower bound. For $\lambda \leq 0$, by Chebyshev's inequality, one has

$$\mathbb{P}(f_Q(M) - \mu \leq -t) \leq e^{\lambda(t-\mu)}\mathbb{E}e^{\lambda f_Q(M)}.$$

Since $\lambda \leq 0$, the map $u \mapsto e^{\lambda u}$ is coordinate-wise non-increasing on $[0, \infty)$. By Remark 2.4 we can define NA family using coordinate-wise non-increasing functions. Thus, using this, the fact that $(Q_{ij}M_{ij})_{i\in[m],j\in[n]}$ is an NA family, and the disjoint monotone aggregation property of an NA family, we obtain

$$\mathbb{P}(f_Q(M) - \mu \leq -t) \leq e^{\lambda(t-\mu)} \prod_{i\in[m],j\in[n]} \mathbb{E}e^{\lambda Q_{ij}M_{ij}}.$$

Note that for $u \leq 0$ one has $e^u \leq 1 + u + u^2/2$ and $e^{-u} \geq 1 - u + u^2/2$, therefore, for $\lambda \leq 0$,

$$e^{\lambda Q_{ij}M_{ij}} \leq 1 + \lambda Q_{ij}M_{ij} + \frac{\lambda^2(Q_{ij}M_{ij})^2}{2} \leq 1 + \lambda Q_{ij}M_{ij} + \phi(-\lambda)(Q_{ij}M_{ij})^2.$$

This implies

$$\mathbb{E}e^{\lambda Q_{ij}M_{ij}} \leq \exp\left(\frac{\lambda d Q_{ij}}{n} + \frac{\phi(-\lambda) d Q_{ij}^2}{n}\right).$$

Thus,

$$\mathbb{P}(f_Q(M) - \mu \leq -t) \leq e^{\lambda t + \sigma^2\phi(-\lambda)}.$$

Optimizing over $\lambda \leq 0$, we complete the proof. $\square$

3 Proof of Theorem 1.3

The main idea is to use the replacement principle, which is stated in Theorem 3.3 below. It enables us to compare our random combinatorial matrix with a random matrix having i.i.d. Bernoulli(d/n) entries. We then apply Theorem 1.1 for random matrices with i.i.d. Bernoulli(d/n) entries to complete the proof. Our exposition is similar to [53, 44]. The replacement principle requires two conditions to be checked. The first one is relatively straightforward in our model. To verify the second one, we need a bound on the smallest singular value similar to the one given in Theorem 1.7, but for random matrices with i.i.d. Bernoulli(d/n) entries. Such an estimate was first obtained by Basak and Rudelson [8, Theorem 2.2] in a much more general setting (when each entry of a matrix was the product of a Bernoulli random variable with a sub-Gaussian random variable). In fact, following similar lines to those of the proof of Theorem 1.7, we can obtain the following.

Theorem 3.1. *There exist absolute positive constants C, C_1, $C_2 \geq 1$ such that the following holds. Let n be a large enough integer and $C \log n \leq d \leq n/2$. Let $\mathcal{B}_n$ be an $n \times n$ random matrix having i.i.d. Bernoulli(d/n) entries. Let $z \in \mathbb{C}$ satisfy $|z| \leq \sqrt{d} \log\log d$. Then*

$$\mathbb{P}(s_n(\mathcal{B}_n - z\mathbf{I}_n) \leq n^{-\beta}) \leq \frac{C_1}{\sqrt{d}},$$

for some $\beta \in (2, C_2 \log\log n]$.

Remark 3.2. Similarly to Theorem 1.7, the bound $n^{-\beta}$ in Theorem 3.1 can be substituted with $\alpha(n,d)$ introduced in Remark 1.9.

Recall that a sequence of random variables $(X_n)_{n\geq 1}$ is said to be bounded in probability if we have

$$\lim_{K\to\infty} \liminf_{n\to\infty} \mathbb{P}(|X_n| \leq K) = 1$$

Theorem 3.3 (Replacement principle)**.** *[53, Theorem 2.1] For every $n \geq 1$ let A_n, B_n be two random $n \times n$ matrices. Assume that the following two conditions hold.*

1. *The sequence*
$$\frac{1}{n}\left(\|A_n\|_{HS}^2 + \|B_n\|_{HS}^2\right)$$
is bounded in probability, where $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm of a matrix.
2. *For almost all complex number z,*
$$\frac{1}{n}\log|\det(A_n - zI_n)| - \frac{1}{n}\log|\det(B_n - zI_n)|$$
converges in probability to 0.

Then $\mu_{A_n} - \mu_{B_n}$ converges in probability to 0.

As we mentioned above, the main difficulty is to verify the second condition. We need a few auxiliary results.

Theorem 3.4. *Let $1 \leq k < n$ and $p \in (0,1)$ satisfy $p(1-p)(n-k) \geq 1$. Let $V \subset \mathbb{C}^n$ be a (fixed) subspace of dimension k and $u \in \mathbb{C}^n$ be a (fixed) vector. Let $\delta = (\delta_1, \cdots, \delta_n)$ be a random vector consisting of i.i.d. Bernoulli(p) random variables, and denote the distance from $\delta + u$ to V by r. Furthermore, denote*

$$E(p) = \mathbb{E}r \qquad \textit{and} \qquad D(p) = \sqrt{p(1-p)(n-k) + d_{u'}^2},$$

where $u' = u + p\,\mathbf{1} = u + \mathbb{E}\delta$ and $d_{u'}$ is the distance from u' to V. Then

$$\tfrac{1}{2}D(p) \le E(p) \le D(p)$$

and for any $t > 0$,

$$\mathbb{P}_\delta\left(|r - E(p)| \ge t\right) \le Ce^{-ct^2},$$

where C, c are positive absolute constants.

Proof. Denote by $P = \{p_{ij}\}_{ij}$ the orthogonal projection matrix from $\mathbb{C}^n$ to the subspace $V^\perp$. Let $x = \delta - \mathbb{E}\delta$. Then

$$r^2 = \|P(\delta+u)\|_2^2 = \|P(x+\mathbb{E}\delta+u)\|_2^2 = \|P(x+u')\|_2^2 = \|Px\|_2^2 + 2\Re(\langle x, Pu'\rangle) + \|Pu'\|_2^2. \quad (3.1)$$

Since random vector x consists of i.i.d. mean-zero coordinates, we have

$$\mathbb{E}\,\Re(\langle x, Pu'\rangle) = 0 \quad (3.2)$$

and

$$\mathbb{E}\|Px\|_2^2 = \mathbb{E}\langle Px, x\rangle = \mathbb{E}\sum_{i,j} p_{ij}x_ix_j = p(1-p)\sum_{i=1}^n P_{ii} = p(1-p)\mathrm{Tr}(P) = p(1-p)(n-k). \quad (3.3)$$

Therefore,

$$\mathbb{E}\,r^2 = p(1-p)(n-k) + d_{u'}^2 = (D(p))^2.$$

Using that x is a $\mathbb{R}^n$-valued random vector consisting of i.i.d. mean zero coordinates again we obtain that

$$\mathbb{E}|\langle x, Pu'\rangle|^2 = \mathbb{E}\left|\sum_{i=1}^n x_i \sum_{j=1}^n \overline{p_{ij}u'_j}\right|^2 = \sum_{i=1}^n \mathbb{E}x_i^2 \left|\sum_{j=1}^n \overline{p_{ij}u'_j}\right|^2 \quad (3.4)$$

$$= p(1-p)\sum_{i=1}^n \left|\sum_{j=1}^n \overline{p_{ij}u'_j}\right|^2 = p(1-p)\|Pu'\|^2 = p(1-p)d_{u'}^2 \quad (3.5)$$

and

$$\mathbb{E}\left|\sum_{i\ne j} p_{ij}x_ix_j\right|^2 = \sum_{i\ne j}\sum_{k\ne \ell} p_{ij}\overline{p_{k\ell}}\,\mathbb{E}(x_ix_jx_kx_\ell) = p^2(1-p)^2\left(\sum_{i\ne j}|p_{ij}|^2 + \sum_{i\ne j} p_{ij}\overline{p_{ji}}\right).$$

Hence,

$$\mathbb{E}\left|\sum_{i\ne j} p_{ij}x_ix_j\right|^2 \le p^2(1-p)^2\left(\sum_{i\ne j}|p_{ij}|^2 + \sum_{i\ne j}|p_{ij}||p_{ji}|\right)$$

Using $|ab| \le (|a|^2+|b|^2)/2$ for $\forall a, b \in \mathbb{C}$, we observe $\sum_{i\ne j} p_{ij}\overline{p_{ji}} \le \sum_{i\ne j}|p_{ij}||p_{ji}| \le \sum_{i\ne j}|p_{ij}|^2$. Therefore,

$$\mathbb{E}\left|\sum_{i\ne j} p_{ij}x_ix_j\right|^2 \le 2p^2(1-p)^2\sum_{i\ne j}|p_{ij}|^2. \quad (3.6)$$

Since $x_i^2 - p(1-p)$ are also i.i.d. mean zero,

$$\mathbb{E}\left|\sum_{i=1}^n p_{ii}[x_i^2 - p(1-p)]\right|^2 = \sum_{i=1}^n |p_{ii}|^2\mathbb{E}[x_i^2 - p(1-p)]^2 = p(1-p)(1-2p)^2\sum_{i=1}^n |p_{ii}|^2. \quad (3.7)$$

Next, we evaluate the variance of r^2, which yields a lower bound on $\mathbb{E}r$. Using (3.1), (3.2), and (3.3), we obtain

$$\begin{aligned}
\mathrm{Var}(r^2) &= \mathbb{E}(r^2 - \mathbb{E}(r^2))^2 = \mathbb{E}\left(\|P(x+u')\|_2^2 - \mathbb{E}\|P(x+u')\|_2^2\right)^2 \\
&= \mathbb{E}\left(\|Px\|_2^2 + 2\Re(\langle x, Pu'\rangle) - \mathbb{E}\|Px\|_2^2\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{n} p_{ii}\left(|x_i|^2 - p(1-p)\right) + \sum_{i\neq j} p_{ij}x_ix_j + 2\Re(\langle x, Pu'\rangle)\right)^2 \\
&\leq 3\mathbb{E}\left|\sum_{i=1}^{n} p_{ii}[|x_i|^2 - p(1-p)]\right|^2 + 3\mathbb{E}\left|\sum_{i\neq j} p_{ij}x_ix_j\right|^2 + 12\mathbb{E}\left|\Re(\langle x, Pu'\rangle)\right|^2.
\end{aligned}$$

Since for every $z \in \mathbb{C}$, $|\Re(z)| \leq |z|$, the last term is bounded by $12\mathbb{E}\left|\langle x, Pu'\rangle\right|^2$. Therefore, using (3.4), (3.6), and (3.7), and $4p(1-p) \leq 1$,

$$\begin{aligned}
\mathrm{Var}(r^2) &\leq 3p(1-p)\Big((1-2p)^2\sum_{i=1}^{n}|p_{ii}|^2 + 2p(1-p)\sum_{i\neq j}|p_{ij}|^2 + 4d_{u'}^2\Big) \\
&\leq 3p(1-p)\sum_{1\leq i,j\leq n}|p_{ij}|^2 + 3d_{u'}^2 = 3\left(p(1-p)\mathrm{Tr}(PP^*) + d_{u'}^2\right) \\
&= 3\left(p(1-p)\mathrm{Tr}(P) + d_{u'}^2\right) = 3\left(p(1-p)(n-k) + d_{u'}^2\right) = 3(D(p))^2.
\end{aligned}$$

This implies

$$\mathbb{E}r^4 \leq 3(D(p))^2 + (\mathbb{E}r^2)^2 = (D(p))^2((D(p))^2 + 3).$$

Using the Hölder inequality, we observe

$$(D(p))^2 = \mathbb{E}\,r^2 = \mathbb{E}\,r^{2/3}r^{4/3} \leq (\mathbb{E}\,r)^{2/3}\left(\mathbb{E}r^4\right)^{1/3} \leq (\mathbb{E}\,r)^{2/3}\,(D(p))^{2/3}((D(p))^2+3)^{1/3}.$$

Using that under the assumptions of the theorem $(D(p)) \geq 1$, we obtain

$$\mathbb{E}\,r \geq \frac{(D(p))^2}{\sqrt{(D(p))^2+3}} \geq \frac{D(p)}{2}.$$

Since $\mathbb{E}\,r \leq \left(\mathbb{E}\,r^2\right)^{1/2} = D(p)$, this shows the first inequality of Theorem 3.4. The second one follows from Talagrand's concentration inequality (see, e.g., [51, Theorem 2.1.13]), as r is a 1-Lipschitz convex function of δ. □

Corollary 3.5. *Let $1 \leq d \leq n/2$ and $1 \leq k < n$. Denote $p = d/n$ and assume $p(1-p)(n-k) \geq 1$. Let $V \subset \mathbb{C}^n$ be a (fixed) subspace of dimension k and $u \in \mathbb{C}^n$ be a (fixed) vector. Let δ be a random vector drawn uniformly from the set of all vectors in $\{0,1\}^n$ containing exactly d ones and denote the distance from $\delta + u$ to V by r. Let $E(p) = E(d/n)$ be the parameter from Theorem 3.4 (note that to define it we need to deal with a Bernoulli random vector). Then for any $t > 0$ we have*

$$\mathbb{P}_\delta\left(|r - E(p)| \geq t\right) \leq C\sqrt{d}\,e^{-ct^2},$$

where C, c are positive absolute constants.

Proof. Let $\delta' = (\delta'_1, \cdots, \delta'_n)$ be a random vector consisting of i.i.d. Bernoulli(d/n) random variables, and denote the distance from $\delta' + u$ to V by r'. Denote also $u' = u + p\,\mathbf{1} = u + \mathbb{E}\delta = u + \mathbb{E}\delta'$ and let $d_{u'}$ be the distance from u' to V. By Theorem 3.4, we have

$$\mathbb{P}_{\delta'}\left(|r' - E(p)| \geq t\right) \leq Ce^{-ct^2}$$

for some absolute positive constants C and c. Using Stirling's formula $m! = \sqrt{2\pi m}(m/e)^m e^{\theta_m/(12m)}$ for some $\theta_m \in (0,1)$, we observe

$$\mathbb{P}_{\delta'}\Big(\sum_{i=1}^n \delta'_i = d\Big) = \binom{n}{d}\left(\frac{d}{n}\right)^d\left(1-\frac{d}{n}\right)^{n-d} \ge \frac{1}{2}\sqrt{\frac{n}{2\pi d(n-d)}} \ge \frac{1}{4}\sqrt{\frac{1}{\pi d}}.$$

Therefore,

$$\mathbb{P}_\delta\left(|r-E(p)|\ge t\right) = \mathbb{P}_{\delta'}\left(|r'-E(p)|\ge t \;\middle|\; \sum_{i=1}^n \delta'_i = d\right) \le \frac{\mathbb{P}_{\delta'}\left(|r'-E(p)|\ge t\right)}{\mathbb{P}_{\delta'}(\sum_{i=1}^n \delta'_i = d)} \le 8C\sqrt{d}\,e^{-ct^2},$$

which completes the proof. □

We need two more well-known facts about singular values.

Lemma 3.6 (Cauchy interlacing law)**.** *[53, Lemma A.1] Let $1 \le m \le n$. Let A be an $n\times n$ matrix and A' be the sub-matrix formed by the first $n-m$ rows of A. Let $s_1(A)\ge\cdots\ge s_n(A)\ge 0$ and $s_1(A')\ge\cdots\ge s_n(A')\ge 0$ be the singular values of A and A', respectively. Then we have*

$$s_{i+m}(A) \le s_i(A') \le s_i(A),$$

for $1\le i\le n-m$.

Lemma 3.7 (Negative second moment)**.** *[53, Lemma A.4] Let $1\le k\le n$, and let A' be a $k\times n$ matrix of rank k with singular values $s_1(A')\ge\cdots\ge s_k(A')\ge 0$ and rows $R_1,\cdots,R_k$. For each $i\le k$, let W_i be the subspace generated by the $k-1$ rows $R_1,\cdots,R_{i-1},R_{i+1},\cdots,R_k$. Then we have*

$$\sum_{i=1}^k s_i^{-2}(A') = \sum_{i=1}^k \operatorname{dist}^{-2}(R_i,W_i).$$

We are now ready to verify the second condition in Theorem 3.3.

Theorem 3.8. *Let $\varepsilon\in(0,1)$ be fixed, $\log^{2+\varepsilon} n \le d \le n/2$, and $p = d/n$. Let M_n be uniformly drawn from $\mathcal{M}_{n,d}$ and $\mathcal{B}_{np}$ be a random $n\times n$ matrix with i.i.d. Bernoulli(p) entries that is independent of M_n. Then for every fixed $z\in\mathbb{C}$,*

$$\frac{1}{n}\log|\det(M_n - z\sqrt{d(1-p)}I_n)| - \frac{1}{n}\log|\det(\mathcal{B}_{np} - z\sqrt{d(1-p)}I_n)|$$

converges to zero in probability as $n\to\infty$.

Proof. Let $u_1,\cdots,u_n$ denote the rows of $-z\sqrt{d(1-p)}I_n$, and let $R_1,\cdots,R_n$ and $R'_1,\cdots,R'_n$ denote the rows of M_n and $\mathcal{B}_{np}$, respectively. Set $V_0=\{0\}$, and for $i\ge 1$, let V_i be the subspace spanned by $R_1+u_1,\cdots,R_i+u_i$. Denote the distance from R_i+u_i to V_{i-1} by d_i. Similarly, define V'_i and d'_i using the vectors R'_j+u_j, $j\le n$. Recall that the absolute value of the determinant of a matrix represents the volume of the n-dimensional parallelepiped generated by the row vectors. Therefore

$$\log|\det(M_n - z\sqrt{d(1-p)}I_n)| = \sum_{i=1}^n \log d_i \quad\text{and}\quad \log|\det(\mathcal{B}_{np} - z\sqrt{d(1-p)}I_n) = \sum_{i=1}^n \log d'_i.$$

Let $m = \lceil n - d^{-1/2+\varepsilon/6}n\rceil$. Then

$$\frac{1}{n}\log|\det(M_n - z\sqrt{d(1-p)}I_n)| - \frac{1}{n}\log|\det(\mathcal{B}_{np} - z\sqrt{d(1-p)}I_n)|$$

$$= \underbrace{\frac{1}{n}\Big(\log\prod_{i=1}^m d_i - \log\prod_{i=1}^m d_i'\Big)}_{T_1} + \underbrace{\frac{1}{n}\sum_{i=m+1}^n (\log d_i - \log d_i')}_{T_2}.$$

We first evaluate T_1. Fix $1 \le i \le m$. Denote by Vol_i the volume of the m-dimensional parallelepiped generated by $R_1+u_1,\cdots,R_i+u_i,R'_{i+1}+u_{i+1},\cdots,R'_m+u_m$ and by Vol_0 the one generated by $R'_1+u_1,\cdots,R'_m+u_m$. Let

$$\mathcal{F}_i = \{R_1+u_1,\cdots,R_{i-1}+u_{i-1},R'_{i+1}+u_{i+1},\cdots,R'_m+u_m\}$$

and U_i be the subspace generated by $\mathcal{F}_i$. Note that parallelepipeds corresponding to Vol_i and Vol_{i-1} are generated by $\mathcal{F}_i \cup \{R_i+u_i\}$ and $\mathcal{F}_i \cup \{R'_i+u_i\}$ respectively. Therefore,

$$\log \mathrm{Vol}_i - \log \mathrm{Vol}_{i-1} = \log r_i - \log r'_i,$$

where r_i and r'_i denote the distance from R_i+u_i and R'_i+u_i to U_i respectively.

For each $i \le m$ we apply Theorem 3.4 and Corollary 3.5 with $t = d^{1/4}$ and $V = U_i$ to vectors R'_i and R_i respectively (note that conditioning on $\mathcal{F}_i$, the subspace U_i becomes deterministic, and R_i, R'_i are independent of $\mathcal{F}_i$). Note that

$$p(1-p) = (d/n)(1-d/n) \ge d/(2n) \qquad \text{and} \qquad k := \dim U_i \le m-1.$$

Thus

$$(D(p))^2 \ge p(1-p)(n-k) \ge \frac{d}{2n}\, d^{-1/2+\varepsilon/6} n = \tfrac{1}{2} d^{1/2+\varepsilon/6}.$$

As we are looking for a limit result and $d \to \infty$ as $n \to \infty$, without loss of generality we assume that $d > 3^{12/\varepsilon}$. Theorem 3.4 and Corollary 3.5 imply that with probability at least $1 - Cm\sqrt{d}\exp(-c\sqrt{d})$ one has for every $i \le m$,

$$\frac{r_i}{r'_i} \le \frac{E(p)+d^{1/4}}{E(p)-d^{1/4}} = 1 + \frac{2d^{1/4}}{E(p)-d^{1/4}} \le 1 + \frac{2d^{1/4}}{D(p)/2 - d^{1/4}} = 1 + \frac{2}{(1/3)d^{\varepsilon/12}-1},$$

hence

$$|\log r_i - \log r'_i| \le \frac{2}{(1/3)d^{\varepsilon/12}-1}.$$

Note that $\mathrm{Vol}_m = \prod_i^m d_i$ and $\mathrm{Vol}_0 = \prod_i^m d'_i$. Therefore

$$|T_1| \le \frac{1}{n}\Big|\log \mathrm{Vol}_m - \log \mathrm{Vol}_0\Big| \le \frac{1}{n}\sum_{i=1}^m |\log \mathrm{Vol}_i - \log \mathrm{Vol}_{i-1}|$$
$$\le \frac{1}{n}\sum_{i=1}^m |\log r_i - \log r'_i| \le \frac{2}{(1/3)d^{\varepsilon/12}-1}.$$

with probability at least

$$1 - Cn\sqrt{d}\exp(-c\sqrt{d}) \longrightarrow 1 \quad \text{as} \quad n \to \infty$$

(we used that $d \ge \log^{2+\varepsilon} n$). This implies that T_1 converges to 0 as $n \to \infty$ in probability.

Now we deal with T_2. Fix $z \in \mathbb{C}$ and assume that $|z| \le \log\log d$ (we can do that as $d \to \infty$ as $n \to \infty$). By Theorem 1.7 (applied with $z' = z\sqrt{d(1-p)}$ instead of z) for some $2 < \beta \le 9$ one has $s_n(A) > n^{-\beta}$ with probability at least $1 - C_1/\sqrt{d}$, where $A = M_n - z\sqrt{d(1-p)}I_n$, C_1 and C_2 are absolute positive constants. Note that the rows of A are $R_i + u_i$, $i \le n$ and that A is not singular whenever $s_n(A) > n^{-\beta}$. Therefore, we may assume that V_i has full rank for all $i \le n$. Let $k \le n$. Applying Lemmas 3.6 and 3.7 to matrix A' generated by the first k rows $R_1 + u_1, \cdots, R_k + u_k$ we obtain

$$d_k^{-2} = \mathrm{dist}^{-2}(R_k + u_k, V_{k-1}) \le \sum_{i=1}^{k} s_i^{-2}(A') \le \sum_{i=1}^{k} s_{i+n-k}^{-2}(A) \le k\, s_n^{-2}(A) \le n^{1+2\beta}.$$

On the other hand, clearly $\|R_k\|_2 = \sqrt{d}$ and $\|u_k\|_2 \le |z|\sqrt{d} \le \sqrt{d}\log\log d$, therefore

$$d_k = \mathrm{dist}(R_k + u_k, V_{k-1}) \le \|R_k + u_k\|_2 \le 2\sqrt{d}\log\log d.$$

Thus, for any $k > n - d^{-1/3}n$,

$$n^{-1/2-\beta} \le d_k \le 2\sqrt{d}\log\log d.$$

For $M = \mathcal{B}_{np}$ we use a similar argument. For the lower bound on d'_k, we combine Theorem 3.1, Remark 3.2 (applied with $M = \mathcal{B}_{np}$ and with $z\sqrt{d(1-p)}$ instead of z) and Lemmas 3.6 and 3.7. For the upper bound, we note that by Lemmas 2.1 with very large probability we have $\|R'_k\|_2 \le \sqrt{d}\log\log d$ for all $k \le n$. Therefore, with probability at least $1 - C'_1/\sqrt{d}$,

$$n^{-1/2-\beta} \le d'_k \le 2\sqrt{d}\log\log d.$$

These bounds on d_k and d'_k together with $d \ge (\log n)^{2+\varepsilon}$, $m \ge n - d^{-1/2+\varepsilon/6}n$, and $\beta < 9$ imply

$$|T_2| \le \frac{n-m}{n}\max_{k\le n}\left|\log\frac{d_k}{d'_k}\right| \le \frac{\log(2n^{1+\beta}\log\log n)}{d^{1/2-\varepsilon/6}} \le \frac{2(1+\beta)\log n}{(\log n)^{1+(\varepsilon-\varepsilon^2)/6}} \le \frac{20}{(\log n)^{(\varepsilon-\varepsilon^2)/6}}$$

with probability at least $1-(C_1+C'_1)/\sqrt{d}$. Thus, T_2 converges to 0 in probability as $n \to \infty$. $\square$

Proof of Theorem 1.3. $p = d/n$, M_n is uniformly drawn from $\mathcal{M}_{n,d}$, and $\mathcal{B}_{np}$ is a random matrix with i.i.d. Bernoulli(p) entries that is independent of M_n. Denote

$$A_n := \frac{M_n}{\sqrt{np(1-p)}} \qquad \text{and} \qquad B_n := \frac{\mathcal{B}_{np}}{\sqrt{np(1-p)}}.$$

Then

$$\mathbb{E}\|A_n\|_{HS}^2 = \mathbb{E}\|B_n\|_{HS}^2 = \frac{n}{1-p} \le 2n.$$

Therefore, the first condition of Theorem 3.3 is satisfied by Markov's inequality. The second condition for matrices A_n, B_n follows by Theorem 3.8. Therefore, Theorem 3.3 implies that $\mu_{A_n} - \mu_{B_n}$ converges in probability to 0. Finally, Theorem 1.1 concludes the proof. $\square$

4 Restricted operator norm

The upper bound on the operator norm $\|M - \mathbb{E}M\|$ plays a crucial role in the construction of nets in Section 6. It is not difficult to see that $\|M - \mathbb{E}M\|$ equals the operator norm of M restricted to the subspace $\left\{v \in \mathbb{S}^{n-1} : \sum_{i=1}^n v_i = 0\right\}$ (see (4.1) below), so we call it *restricted operator norm.* For $d = n/2$, Tran [55, Proposition 2.8] employed Bernstein's inequality combined with a net argument to show that this norm is typically of order $O(\sqrt{n})$ (see also [27, Lemma 5.1]). The proof also works for $d = pn$, where $p \in (0,1)$ is a fixed constant. However, their approach does not extend to the sparse regime, particularly when d is of the order of $\log n$. In contrast, we apply the Kahn-Szemerédi argument to deduce that the restricted operator norm is $O_\lambda(\sqrt{\min\{d, n-d\} + \log n})$ for $1 \le d \le n$. The main result of this section is the following theorem.

Theorem 4.1. *There are absolute positive constants C_1 and C_2 such that the following holds. Let $1 \le m \le n$, $\gamma \ge 4$, $\beta \ge 1$, and let d be an integer satisfying $1 \le d \le n$. Let M be an $m \times n$ 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. Let*

$$N = \max\left\{9\min\{d, n-d\}, 2\gamma \log m\right\}.$$

Then, with probability at least $1 - n \cdot m^{-\gamma} - 2n^{-\beta}$, one has

$$\|M - \mathbb{E}M\| \le \frac{2\min\{d, n-d\}}{\sqrt{N}} + C_1 \beta \sqrt{N} \le C_2 \beta \left(\sqrt{\min\{d, n-d\} + \gamma \log m}\right).$$

In particular, taking $m = n$, $\gamma = 4$, $\beta = 2$, and $\log n \le d \le n - \log n$, we have

$$\|M - \mathbb{E}M\| = \|M^T - \mathbb{E}M^T\| \le C_{4.1}\sqrt{d}.$$

with probability at least $1 - 3n^{-1}$.

Remark 4.2. Denote by $E_{m\times n}$ the matrix with all entries equal to 1, by considering the matrix $A = E_{m\times n} - M$ and noting that $M - \mathbb{E}M = -(A - \mathbb{E}A)$, it is enough to prove Theorem 4.1 for $d \le n/2$.

Remark 4.3. Note that for $d \le n/2$ one has

$$\|M - EM\| = \|(M - EM)^T\| \ge \|(M - EM)^T e_1\|_2 = \sqrt{\frac{d(n-d)}{n}} \ge \sqrt{\frac{d}{2}},$$

where e_1 is the first element of the standard basis of $\mathbb{R}^n$. Thus, for $m = n$ and $\log n \le d \le n/2$, the bound in Theorem 4.1 is sharp up to a constant.

The proof of Theorem 4.1 is based on properties of negatively associated random variables discussed in Section 2.3 and on the Kahn-Szemerédi argument, which is based on splitting the sum under consideration into two sums depending on how large the corresponding summands are. In the rest of this section, we will assume that $d \le n/2$.

4.1 Kahn-Szemerédi argument

We follow the exposition in [16] (see also [57]) with the necessary modifications. In view of Remark 4.2 we assume that $d \le n/2$, so $\min\{d, n-d\} = d$. We also note that $(M - \mathbb{E}M)\mathbf{1} = 0$, hence

$$\|M - \mathbb{E}M\| = \sup_{y \in \mathbb{S}_0^{n-1}} \|(M - \mathbb{E}M)y\|_2 = \sup_{y \in \mathbb{S}_0^{n-1}} \|My\|_2 = \sup_{x \in \mathbb{S}^{m-1}, y \in \mathbb{S}_0^{n-1}} \langle x, My \rangle. \tag{4.1}$$

For a fixed $x \in \mathbb{S}^{m-1}, y \in \mathbb{S}_0^{n-1}$, consider the matrix $Q = Q(x,y) := xy^T$. Then

$$f_Q(M) = \langle x, My \rangle = \sum_{i \in [m], j \in [n]} M_{ij} x_i y_j,$$

where the function f_Q was defined in (2.2). The key point in Kahn-Szemerédi's argument is to split the sum $\sum_{1 \le i,j \le n} M_{ij} x_i y_j$ into two pieces according to absolute values of $x_i y_j$'s. Fix

$$N = \max\left\{d, \frac{(1+\tau)md}{n}\right\} \tag{4.2}$$

for some positive τ to be determined later (in fact, τ comes from Lemma 2.2). We define the light and heavy couples of coordinates by

$$\mathcal{L}(x,y)=\Big\{(i,j)\in[m]\times[n]:|x_iy_j|\le\frac{\sqrt{N}}{n}\Big\}\quad\text{and}\quad\mathcal{H}(x,y)=\Big\{(i,j)\in[m]\times[n]:|x_iy_j|>\frac{\sqrt{N}}{n}\Big\}.$$

We decompose $f_Q(M)$ as

$$f_Q(M)=\sum_{(i,j)\in\mathcal{L}(x,y)}M_{ij}x_iy_j+\sum_{(i,j)\in\mathcal{H}(x,y)}M_{ij}x_iy_j.$$

We first estimate the mean of the light part.

Lemma 4.4. *Let M be as in Theorem 4.1 and N as in (4.2). Then for every $x\in\mathbb{S}^{m-1},y\in\mathbb{S}_0^{n-1}$, we have*

$$\Big|\mathbb{E}\sum_{(i,j)\in\mathcal{L}(x,y)}M_{ij}x_iy_j\Big|\le\frac{d}{\sqrt{N}}.$$

Proof. Since $\mathbb{E}M_{ij}=d/n$, we have

$$\Big|\mathbb{E}\sum_{(i,j)\in\mathcal{L}(x,y)}M_{ij}x_iy_j\Big|=\frac{d}{n}\Big|\sum_{(i,j)\in\mathcal{L}(x,y)}x_iy_j\Big|\le\frac{d}{n}\sum_{i=1}^{m}\Big|\sum_{\{j:(i,j)\in\mathcal{L}(x,y)\}}x_iy_j\Big|.$$

For any $y\in\mathbb{S}_0^{n-1}$, we have $\sum_{j=1}^n y_j=0$. Hence for every (fixed) $i\in[m]$,

$$\sum_{\{j:(i,j)\in\mathcal{L}(x,y)\}}x_iy_j=-\sum_{\{j:(i,j)\in\mathcal{H}(x,y)\}}x_iy_j.$$

Therefore, using that $|x_iy_j|>\sqrt{N}/n$ on $\mathcal{H}(x,y)$ and that x,y are unit vectors,

$$\Big|\mathbb{E}\sum_{(i,j)\in\mathcal{L}(x,y)}M_{ij}x_iy_j\Big|\le\frac{d}{n}\sum_{i=1}^{m}\Big|\sum_{\{j:(i,j)\in\mathcal{H}(x,y)\}}x_iy_j\Big|\le\frac{d}{\sqrt{N}}\sum_{(i,j)\in\mathcal{H}(x,y)}|x_iy_j|^2\le\frac{d}{\sqrt{N}}.$$

□

Lemma 4.5. *Let M be as in Theorem 4.1 and N as in (4.2). Then for every $\beta>0$ and every $x\in\mathbb{S}^{m-1},y\in\mathbb{S}_0^{n-1}$,*

$$\mathbb{P}\left(\Big|\sum_{(i,j)\in\mathcal{L}(x,y)}M_{ij}x_iy_j\Big|\ge\frac{d}{\sqrt{N}}+\beta\sqrt{N}\right)\le4\exp\left(-\frac{3\beta^2n}{4(6+\beta)}\right).$$

Proof. We further split $\mathcal{L}(x,y)$ into "positive" and "negative" as $\mathcal{L}(x,y)=\mathcal{L}_+(x,y)\cup\mathcal{L}_-(x,y)$, where

$$\mathcal{L}_+(x,y)=\{(i,j):0\le x_iy_j\le\sqrt{N}/n\},\qquad\mathcal{L}_-(x,y)=\mathcal{L}(x,y)\setminus\mathcal{L}_+(x,y).$$

Applying Theorem 2.7 with $K=\sqrt{N}/n$ and

$$Q_{ij}=\begin{cases}x_iy_j, & \text{if }(i,j)\in\mathcal{L}_+(x,y),\\ 0, & \text{otherwise,}\end{cases}$$

and using $N \geq d$ together with

$$\sigma^2 = \frac{d}{n} \sum_{(i,j)\in\mathcal{L}_+(x,y)} (x_i y_j)^2 \leq d/n,$$

we obtain

$$\mathbb{P}\left(\left|\sum_{(i,j)\in\mathcal{L}_+(x,y)} M_{ij}x_iy_j - \mathbb{E}\sum_{(i,j)\in\mathcal{L}_+(x,y)} M_{ij}x_iy_j\right| \geq \beta\sqrt{N}/2\right) \leq 2\exp\left(-\frac{3\beta^2 n}{4(6+\beta)}\right),$$

Similarly applying Theorem 2.7 with

$$Q_{ij} = \begin{cases} -x_iy_j, & \text{if } (i,j)\in\mathcal{L}_-(x,y), \\ 0, & \text{otherwise,} \end{cases}$$

we obtain

$$\mathbb{P}\left(\left|\sum_{(i,j)\in\mathcal{L}_-(x,y)} M_{ij}x_iy_j - \mathbb{E}\sum_{(i,j)\in\mathcal{L}_-(x,y)} M_{ij}x_iy_j\right| \geq \beta\sqrt{N}/2\right) \leq 2\exp\left(-\frac{3\beta^2 n}{4(6+\beta)}\right).$$

Note that Lemma 4.4 implies

$$\begin{aligned} &\mathbb{P}\left(\left|\sum_{(i,j)\in\mathcal{L}(x,y)} M_{ij}x_iy_j\right| \geq \frac{d}{\sqrt{N}} + \beta\sqrt{N}\right) \\ &\leq \mathbb{P}\left(\left|\sum_{(i,j)\in\mathcal{L}(x,y)} M_{ij}x_iy_j - \mathbb{E}\sum_{(i,j)\in\mathcal{L}(x,y)} M_{ij}x_iy_j\right| \geq \beta\sqrt{N}\right). \end{aligned}$$

Therefore, the desired result follows by the triangle inequality. □

To bound the heavy part, we use the following version of *discrepancy property* for a matrix M, introduced in [16]. This concept originally dates back to the work of Kahn and Szemerédi (see [23, Lemma 2.5]), where they proved that for a fixed d and sufficiently large n, the second eigenvalue of a random matrix drawn uniformly from the permutation model is $O(\sqrt{d})$ with high probability. For an adjacency matrix of a random graph, it ensures that the number of edges between any two sets of vertices is comparable to the expected number.

Definition 4.6. *Let $n, m \geq 1$ and let M be an $m \times n$ matrix with non-negative entries. Given $S \subset [m], T \subset [n]$, define the edge count as*

$$e_M(S,T) := \sum_{i\in S}\sum_{j\in T} M_{ij}.$$

Furthermore, we say that the matrix M has the discrepancy property $DP(\delta, \kappa_1, \kappa_2)$ with parameters $\delta \in (0,1)$, $\kappa_1 > 0$, $\kappa_2 \geq 0$ if for all nonempty $S \subset [m]$ and $T \subset [n]$ either

(1) $\frac{e_M(S,T)}{\delta|S||T|} \leq \kappa_1$, *or*

(2) $e_M(S,T)\log\frac{e_M(S,T)}{\delta|S||T|} \leq \kappa_2(|S|\vee|T|)\log\frac{en}{|S|\vee|T|}$.

The following lemma asserts that our random matrices enjoy the discrepancy property with high probability.

Lemma 4.7. *Let $1 \le m \le n$. Let M be the random matrix as in Theorem 4.1. Then for any $\beta > 0$, with probability at least $1 - n^{-\beta}$, the discrepancy property DP$(\delta, \kappa_1, \kappa_2)$ holds for M with $\delta = d/n$, $\kappa_1 = e^2$, $\kappa_2 = 2(4+\beta)$.*

Proof. Fix $\beta > 0$. Let $Q = \mathbf{1}_S \mathbf{1}_T^\top$, where $\mathbf{1}_S \in \{0,1\}^m$ and $\mathbf{1}_T \in \{0,1\}^n$ are indicator vectors of the set S and T, respectively. We have

$$f_Q(M) = \sum_{i \in [m], j \in [n]} (\mathbf{1}_S)_i (\mathbf{1}_T)_j M_{ij} = e_M(S,T).$$

Denote

$$\mu(S,T) := \mathbb{E} f_Q(M) = \frac{d}{n}|S||T| = \delta |S||T|.$$

Let $\gamma_1 = e^2 - 1$. Recall that $h(u) = (1+u)\log(1+u) - u$ is strictly increasing and let $\gamma^*(S,T)$ be the (unique) solution of

$$h(u)\mu(S,T) = (\beta+4)(|S| \vee |T|) \log \frac{en}{|S| \vee |T|}.$$

Denote $\gamma(S,T) = \max\{\gamma^*(S,T), \gamma_1\}$. Note that

$$\sigma^2 := \frac{d}{n}\sum_{i,j} Q_{ij}^2 = \frac{d}{n}|S||T| = \mu(S,T).$$

Applying Theorem 2.7 with $K = 1$ and $t = \gamma(S,T)\mu(S,T)$, we have

$$\mathbb{P}(e_M(S,T) \ge (1+\gamma(S,T))\mu(S,T)) \le \exp(-\mu(S,T)h(\gamma(S,T))) \le \exp(-\mu(S,T)h(\gamma^*(S,T))).$$

Therefore,

$$\begin{aligned}
&\mathbb{P}(\exists S \subset [m],\, \exists T \subset [n],\, |S| = k,\, |T| = \ell : \quad e_M(S,T) \ge (1+\gamma(S,T))\mu(S,T)) \\
&\le \sum_{S \subset [m], |S|=k} \sum_{T \subset [n], |T| = \ell} \exp(-\mu(S,T)h(\gamma^*(S,T))) \\
&\le \binom{m}{k}\binom{n}{\ell} \exp\left(-(\beta+4)(k \vee \ell)\log\frac{en}{k \vee \ell}\right) \\
&\le \left(\frac{em}{k}\right)^k \left(\frac{en}{\ell}\right)^\ell \exp\left(-(\beta+4)(k \vee \ell)\log\frac{en}{k \vee \ell}\right) \\
&\le \exp\left(-(\beta+2)(k \vee \ell)\log\frac{en}{k \vee \ell}\right) \le \exp(-(\beta+2)\log(en)),
\end{aligned}$$

where we used that functions $h(u)$ and $u \mapsto u\log(e/u)$ are increasing. Now, by the union bound over the mn choices of $k \in [m]$, $\ell \in [n]$, we have that with probability at least $1 - n^{-\beta}$,

$$\forall S \subset [m]\ T \subset [n] \qquad e_M(S,T) \le (1+\gamma(S,T))\mu(S,T). \tag{4.3}$$

If S, T are such that $\gamma(S,T) = \gamma_1 = e^2 - 1$, then from (4.3), we have

$$e_M(S,T) \le e^2 \mu(S,T) = e^2 \delta |S||T|,$$

which implies that (1) in Definition 4.6 holds with $\kappa_1 = e^2$. Otherwise, we have

$$e_M(S,T) \le (1+\gamma^*(S,T))\mu(S,T). \tag{4.4}$$

Consequently,

$$\frac{h(\gamma^*(S,T))}{1+\gamma^*(S,T)} e_M(S,T) \le (\beta+4)(|S| \vee |T|)\log\frac{en}{|S| \vee |T|}$$

by the definition of $\gamma^*(S,T)$. Note that $\gamma^*(S,T) \geq \gamma_1 = e^2 - 1$, hence $\log(1+\gamma^*(S,T)) \geq 2$. Therefore, using (4.4), we obtain

$$\frac{h(\gamma^*(S,T))}{1+\gamma^*(S,T)} \geq \log(1+\gamma^*(S,T)) - 1 \geq \frac{\log(1+\gamma^*(S,T))}{2} \geq \frac{1}{2}\log\frac{e_M(S,T)}{\mu(S,T)}.$$

This implies that for $\gamma^* \geq \gamma_1$,

$$e_M(S,T)\log\frac{e_M(S,T)}{\mu(S,T)} \leq 2(\beta+4)(|S| \vee |T|)\log\frac{en}{|S| \vee |T|}.$$

As $\mu(S,T) = \delta|S||T|$, the second case in Definition 4.6 holds with $\kappa_2 = 2(\beta+4)$. This completes the proof. □

The following deterministic lemma shows that when the discrepancy property holds, the heavy part $\sum_{(i,j)\in\mathcal{H}(x,y)} M_{ij}x_iy_j$ is of order $O(\sqrt{N})$. We omit the proof, since it is very similar to the proof of Lemma 6.6 in [16], see also [57, Lemma 5.5].

Lemma 4.8. *Let $1 \leq m \leq n$, $\lambda > 0$, and N as in (4.2). Let M be a nonnegative $m \times n$ matrix with all row and column sums bounded by N. Suppose that M has DP$(\delta, \kappa_1, \kappa_2)$ with $\delta = \lambda N/n$, $\kappa_1 > 1$, $\kappa_2 \geq 0$. Then for any $x \in \mathbb{S}^{m-1}, y \in \mathbb{S}_0^{n-1}$, we have*

$$\left|\sum_{(i,j)\in\mathcal{H}(x,y)} M_{ij}x_iy_j\right| \leq \alpha\sqrt{N},$$

where $\alpha = 16 + 32\lambda(1+\kappa_1) + 64\kappa_2\left(1 + \frac{2}{\kappa_1\log\kappa_1}\right)$.

4.2 Completing the proof.

We will finish the proof of Theorem 4.1 with the standard ε-net argument (as it is classical and used in many works, we omit the proof).

Lemma 4.9. *For $\varepsilon \in (0,1/2)$. Let $\mathcal{N}_\varepsilon$ and $\mathcal{N}_\varepsilon^0$ be nets for $\mathbb{S}^{m-1}$ and $\mathbb{S}_0^{n-1}$, respectively. Let A be an $m \times n$ matrix. Then*

$$\sup_{\substack{x\in\mathbb{S}^{m-1}\\ y\in\mathbb{S}_0^{n-1}}} \langle x, Ay\rangle \leq \frac{1}{1-2\varepsilon}\sup_{x\in\mathcal{N}_\varepsilon, y\in\mathcal{N}_\varepsilon^0} |\langle x, Ay\rangle|.$$

Proof of Theorem 4.1. In view of Remark 4.2 we may assume that $d \leq n/2$. Fix $\beta \geq 1$, $\beta_1 \geq 16$, and $\beta_2 \geq 4$. Let $\tau = \max\left\{e^2, \frac{\beta_2 n\log m}{md}\right\}$, and recall that $N = \max\left\{d, \frac{(1+\tau)md}{n}\right\}$. Note that

$$\text{if} \quad \frac{\beta_2 n\log m}{md} \geq e^2 \qquad \text{then} \qquad \frac{(1+\tau)md}{n} < 2\beta_2\log m$$

and otherwise, $(1+\tau)md/n \leq (1+\tau)d < 9d$. Therefore, $N \leq \max\{9d, 2\beta_2\log m\}$, and by Lemma 2.2, column sums of M do not exceed N with probability at least $1 - n\cdot m^{-\beta_2}$. Applying Lemma 4.7 we obtain that with probability at least $1 - n\cdot m^{-\beta_2} - n^{-\beta}$, the matrix M has DP$(\delta, \kappa_1, \kappa_2)$ with $\delta = d/n$, $\kappa_1 = e^2$, $\kappa_2 = 2(4+\beta)$. Then Lemma 4.8 implies that with probability at least $1 - n\cdot m^{-\beta_2} - n^{-\beta}$,

$$\sup_{\substack{x\in\mathbb{S}^{m-1}\\ y\in\mathbb{S}_0^{n-1}}} \left|\sum_{(i,j)\in\mathcal{H}(x,y)} M_{ij}x_iy_j\right| \leq \alpha\sqrt{N}, \tag{4.5}$$

where $\alpha = 16 + 32\frac{(1+e^2)d}{N} + 128(4+\beta)(1+e^{-2})$. Let $\mathcal{G}$ denote this event. As we mentioned above,

$$\|M - \mathbb{E}M\| = \sup_{\substack{x \in \mathbb{S}^{m-1} \\ y \in \mathbb{S}_0^{n-1}}} \langle x, My \rangle.$$

Fix $\varepsilon = 1/4$. By a standard volume argument, on $\mathbb{S}^{m-1}$ and $\mathbb{S}_0^{n-1}$ there exist ε-nets $\mathcal{N}_\varepsilon$ and $\mathcal{N}_\varepsilon^0$ of size at most 9^m and 9^n, respectively. Combining (4.5), Lemma 4.5, and the union bound, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{G} \cap \left\{\|M - \mathbb{E}M\| \geq 2\alpha\sqrt{N} + \frac{2d}{\sqrt{N}} + 2\beta_1\sqrt{N}\right\}\right) \\
&\quad \leq \mathbb{P}\left(\mathcal{G} \cap \left\{\sup_{\substack{x \in \mathcal{N}_\varepsilon \\ y \in \mathcal{N}_\varepsilon^0}} |\langle x, My \rangle| \geq \alpha\sqrt{N} + \frac{d}{\sqrt{N}} + \beta_1\sqrt{N}\right\}\right) \\
&\quad \leq \mathbb{P}\left(\mathcal{G} \cap \left\{\sup_{\substack{x \in \mathcal{N}_\varepsilon \\ y \in \mathcal{N}_\varepsilon^0}} \left|\sum_{(i,j) \in \mathcal{L}(x,y)} M_{ij}x_iy_j\right| \geq \alpha\sqrt{N} + \frac{d}{\sqrt{N}} + \beta_1\sqrt{N} - \sup_{\substack{x \in \mathcal{N}_\varepsilon \\ y \in \mathcal{N}_\varepsilon^0}} \left|\sum_{(i,j) \in \mathcal{H}(x,y)} M_{ij}x_iy_j\right|\right\}\right) \\
&\quad \leq \mathbb{P}\left(\mathcal{G} \cap \left\{\sup_{\substack{x \in \mathcal{N}_\varepsilon \\ y \in \mathcal{N}_\varepsilon^0}} \left|\sum_{(i,j) \in \mathcal{L}(x,y)} M_{ij}x_iy_j\right| \geq \frac{d}{\sqrt{N}} + \beta_1\sqrt{N}\right\}\right) \\
&\quad \leq 4 \cdot 9^{2n} \exp\left(-\frac{3\beta_1^2 n}{4(6+\beta_1)}\right) \leq \exp\left(-\beta_1 n/8\right).
\end{aligned}$$

Finally, to obtain the desired result, we choose $\gamma = \beta_2$ and $\beta_1 = 16\beta$. □

4.3 Tall rectangular matrices

Theorem 4.1 is stated (and used below) for $1 \leq m \leq n$. For the sake of completeness, we generalize it to the complementary regime $m > n$. This almost immediately follows by partitioning the rows of the matrix into blocks of size comparable to n, applying Theorem 4.1 to each block, and then combing the estimates.

Theorem 4.10. *Let $m, n \geq 1$, $\gamma \geq 4$, $\beta \geq 1$, and let d be an integer satisfying $1 \leq d \leq n$. Let M be an $m \times n$ 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. Denote $k := \lceil m/n \rceil$. Then, with probability at least $1 - k(2^\gamma n^{1-\gamma} + 2n^{-\beta})$, one has*

$$\|M - \mathbb{E}M\| \leq C_{4.1}\beta\sqrt{k}\sqrt{\min\{d, n-d\} + \gamma \log n}$$

where $C_{4.1}$ denotes the absolute constant from Theorem 4.1.

Remark 4.11. Taking $\gamma = 4$ in Theorem 4.10. If $\log n \leq d \leq n/2$ and $m \geq n$, then

$$\|M - \mathbb{E}M\| \leq C\beta\sqrt{\frac{md}{n}}$$

for some absolute constant $C > 0$.

On the other hand, assuming that $d \leq n/2$, note that each row of $M - \mathbb{E}M$ has the squared Euclidean norm $d(n-d)/n \geq d/2$. Thus, $\|M - \mathbb{E}M\|_{HS}^2 \geq md/2$. Moreover, $(M - \mathbb{E}M)\mathbf{1} = 0$, and hence, $r := \operatorname{rank}(M - \mathbb{E}M) \leq n - 1$. Denoting by $s_1 \geq \cdots \geq s_r > 0$ the non-zero singular values of $M - \mathbb{E}M$, we observe

$$\|M - \mathbb{E}M\|_{HS}^2 = \sum_{i=1}^{r} s_i^2 \leq r s_1^2 = r\|M - \mathbb{E}M\|^2.$$

This implies,

$$\|M-\mathbb{E}M\|^2 \geq \frac{\|M-\mathbb{E}M\|_{HS}^2}{r} \geq \frac{md}{2(n-1)}.$$

Thus, $\|M-\mathbb{E}M\| \geq \sqrt{md/(2(n-1))}$. Therefore, for $\log n \leq d \leq n/2$ and $m \geq n$, the bound in Theorem 4.10 is sharp up to an absolute constant.

Proof. First note that the case $m \leq n$ is covered by Theorem 4.1, so we assume that $m > n$ and hence $k \geq 2$. Denote $d_* := \min\{d, n-d\}$. The case $n = 1$ is trivial, so we also assume $n \geq 2$.

Note that for positive integers $m_1, \dots, m_k$ satisfying $n/2 \leq m_i \leq n$ their sum $\sum_{i=1}^k m_i$ can take any value between $k\lceil n/2\rceil$ and kn. Since $k(n+1)/2 \leq m \leq kn$ for $m > n > 1$, there exist positive integers $m_1, \dots, m_k$ such that

$$\sum_{i=1}^k m_i = m \quad \text{and} \quad \frac{n}{2} \leq m_i \leq n \text{ for all } i = 1, \dots, k.$$

Partition the set of row indices into pairwise disjoint consecutive blocks $I_1, \dots, I_k$ with $|I_j| = m_j$ for $j = 1, \dots, k$, and let $M^{(j)}$ be the $m_j \times n$ submatrix of M formed by rows indexed by I_j. Then

$$M - \mathbb{E}M = \begin{bmatrix} M^{(1)} - \mathbb{E}M^{(1)} \\ \vdots \\ M^{(k)} - \mathbb{E}M^{(k)} \end{bmatrix}.$$

For each $j = 1, \dots, k$, the rows of $M^{(j)}$ are still independent. Since $m_j \leq n$, we apply Theorem 4.1 to $M^{(j)}$. Therefore, for $j = 1, \dots, k$,

$$\mathbb{P}\left(\left\|M^{(j)} - \mathbb{E}M^{(j)}\right\| > C_{4.1}\beta\sqrt{d_* + \gamma \log m_j}\right) \leq n m_j^{-\gamma} + 2n^{-\beta}.$$

Because $m_j \leq n$, we have $\log m_j \leq \log n$, and because $m_j \geq n/2$, we have $n m_j^{-\gamma} \leq 2^\gamma n^{1-\gamma}$. Thus, for $j = 1, \dots, k$,

$$\mathbb{P}\left(\left\|M^{(j)} - \mathbb{E}M^{(j)}\right\| > C_{4.1}\beta\sqrt{d_* + \gamma \log n}\right) \leq 2^\gamma n^{1-\gamma} + 2n^{-\beta}.$$

By the union bound, with probability at least $1 - k(2^\gamma n^{1-\gamma} + 2n^{-\beta})$, one has

$$\left\|M^{(j)} - \mathbb{E}M^{(j)}\right\| \leq C_{4.1}\beta\sqrt{d_* + \gamma \log n} \quad \text{for all } j = 1, \dots, k.$$

Then, on this event we have for any $y \in \mathbb{S}^{n-1}$,

$$\begin{aligned} \|(M-\mathbb{E}M)y\|_2^2 &= \sum_{j=1}^k \left\|(M^{(j)} - \mathbb{E}M^{(j)})y\right\|_2^2 \\ &\leq \sum_{j=1}^k \left\|(M^{(j)} - \mathbb{E}M^{(j)})\right\|_2^2 \|y\|_2^2 \leq k C_{4.1}^2 \beta^2 (d_* + \gamma \log n). \end{aligned}$$

This implies the desired result. □

5 Concentration results and graph expansion properties

5.1 Auxiliary lemmas

In this section we present a pair of lemmas that control the connectivity between the rows or columns of M and an arbitrary subset of indices. The first lemma provides a probabilistic bound on the number of rows of M that have an atypically large number of non-zero entries within a given set of columns J. The second simple lemma offers a known deterministic statement for the columns of M. It is essentially [37, Claim 3.6] (where a different model was used). We provide its proof for the sake of completeness.

Lemma 5.1. *Let $1 \le d, k \le n$ and $r \ge 2$. Let $J \subset [n]$, $|J| = k$. Let M be uniformly drawn from $\mathcal{M}_{n,d}$. Then with probability at least $1 - \exp(-(8n/r))$,*

$$\left| \left\{ i \le n : |Supp(R_i(M)) \cap J| \ge \frac{rkd}{n} \right\} \right| \le \frac{9n}{r}.$$

Proof. Fix $J \subset [n]$ with $|J| = k$. Denote by $\mathcal{E}_i$ the event that $|\mathrm{Supp}(R_i(M)) \cap J| \ge \frac{rkd}{n}$. Note that the random variable $|\mathrm{Supp}\ R_i(M) \cap J|$ has a hypergeometric distribution with mean kd/n. Therefore, applying Markov's inequality, we have

$$q := \mathbb{P}(\mathcal{E}_i) \le \frac{\mathbb{E}\left(|\mathrm{Supp}(R_i(M)) \cap J|\right)}{rkd/n} \le \frac{1}{r}.$$

Let δ_i be the indicator function of the event $\mathcal{E}_i$, $i \le n$. Note that events $\mathcal{E}_i$'s (hence random variables δ_i's) are independent. Since $q \le 1/r$, $\tau := 9/(rq) - 1 \ge 8 \ge e^2$. Then, writing $9n/r = (1+\tau)qn$ and applying Lemma 2.1, we observe

$$\mathbb{P}\left(\sum_{i=1}^{n} \delta_i \ge \frac{9n}{r} \right) \le \exp(-(9/(rq) - 1)qn) \le \exp(-(8n/r)).$$

This completes the proof. □

The following lemma was proved in [37] (see Claim 3.6 there).

Lemma 5.2. *[37, Claim 3.6] Let $1 \le d, k, \le n$. Let $J \subset [n]$, $|J| = k$, and $A > 1$. Let $M \in \mathcal{M}_{n,d}$. Then*

$$\left| \left\{ i \le n : |Supp(R_i(M^T)) \cap J| \ge \frac{Akd}{n} \right\} \right| \le \frac{n}{A}.$$

Proof. By the definition, the total number of ones in the submatrix of M indexed by $J \times [n]$ is kd. Therefore,

$$\left| \left\{ i \le n : |\mathrm{Supp}(R_i(M^T)) \cap J| \ge \frac{Akd}{n} \right\} \right| \cdot \frac{Akd}{n} \le kd,$$

which yields the desired bound. □

5.2 The triple norm

Recall that $\mathbf{1} = (1, 1, ..., 1)$ and denote $\mathbf{e} = \mathbf{1}/\sqrt{n}$. Let $P_{\mathbf{e}}$ be the projection on $\mathbf{e}$, and let $P_{\mathbf{e}^\perp}$ be the projection on $\mathbf{e}^\perp$. Consider the triple norm on $\mathbb{C}^n$ defined by

$$|||x|||^2 := \|P_{\mathbf{e}^\perp} x\|_2^2 + d\|P_{\mathbf{e}} x\|_2^2.$$

The triple norm will be used later for net construction.

The following proposition is an adaptation of a result from [40, Proposition 3.14]. As the proof is analogous to that in [40], we omit it.

Proposition 5.3. *Let n be large enough and $\log n \le d \le n/2$. Denote*

$$\mathcal{E}_{5.3} := \left\{ M \in \mathcal{M}_{n,d} : \|M - \mathbb{E}M\| \le C_{4.1}\sqrt{d} \right\}.$$

Then for every $x \in \mathbb{C}^n$ and every $M \in \mathcal{E}_{5.3}$ one has $\|Mx\| \le (C_{4.1} + 1)\sqrt{d}\, |||x|||$.

A similar result holds for M^T.

Proposition 5.4. *Let n be large enough and $\log n \le d \le n/2$. Denote*

$$\mathcal{E}_{5.4} := \left\{ M \in \mathcal{M}_{n,d} : \|M^T - \mathbb{E}M^T\| \le C_{4.1}\sqrt{d} \quad \text{and} \quad \|M^T \mathbf{1}\|_2 \le (1 + C_{4.1}) d\sqrt{n} \right\}.$$

Then for every $x \in \mathbb{C}^n$ and every $M \in \mathcal{E}_{5.4}$ one has $\|M^T x\| \le (2C_{4.1} + 1)\sqrt{d}\,\|\!|x|\!\|$.

5.3 Vertex expansion property

In this section we establish two vertex expansion properties for in-neighbors and out-neighbors of the random matrix M uniformly drawn from $\mathcal{M}_{n,d}$. The first property, Lemma 5.5, is a consequence of the independence between the rows of M, which allows for a direct application of the classical Chernoff bound for independent random variables. In the contrast, the second property, Lemma 5.6, is more subtle, as the corresponding indicator variables are not independent. To overcome this difficulty, we leverage the property of negative association (NA) introduced in Section 2.3 and apply a Chernoff concentration for NA random variables.

Let $M \in \mathcal{M}_{n,d}$. For $J \subset [n]$ denote

$$S(J, M) := \{ i \in [n] : \mathrm{Supp}(R_i(M)) \cap J \neq \emptyset \}.$$

Given $1 \le k \le n$ and $\varepsilon \in (0, 1/2]$, let

$$\Omega_{k,\varepsilon}^{in} := \{ M \in \mathcal{M}_{n,d} : \forall J \subset [n], |J| = k \text{ one has } (1-\varepsilon)kd \le |S(J,M)| \le (1+\varepsilon)kd \}$$

and

$$\Omega_{k,\varepsilon}^{out} := \left\{ M \in \mathcal{M}_{n,d} : \forall J \subset [n], |J| = k \text{ one has } |S(J, M^T)| \ge (1-\varepsilon)kd \right\}.$$

We also denote

$$\Omega_{k,\varepsilon} := \Omega_{k,\varepsilon}^{in} \cap \Omega_{k,\varepsilon}^{out}. \tag{5.1}$$

Note that if we interpret a matrix $M \in \mathcal{M}_{n,d}$ as the adjacency matrix of a directed graph G with vertex set $[n]$, then $S(J, M)$ and $S(J, M^T)$ correspond to the sets of in-neighbors and out-neighbors, respectively, of the vertices in $J \subset [n]$. Moreover, $|S(J, M^T)| \le d|J|$, since each column of M^T contains exactly d ones.

Lemma 5.5. *Let $n \ge 3$, $192 \log n \le d \le n/2$, $\varepsilon \in \left[\sqrt{\frac{48 \log n}{d}}, 1\right)$, and $1 \le k \le \varepsilon n/d$. Then*

$$\mathbb{P}(\Omega_{k,\varepsilon}^{in}) \ge 1 - 2\binom{n}{k} \exp\left(-\frac{\varepsilon^2 kd}{24}\right).$$

Moreover,

$$\mathbb{P}\left(\bigcap_{k \le \varepsilon n/d} \Omega_{k,\varepsilon}^{in} \right) \ge 1 - 3\exp\left(-\frac{\varepsilon^2 d}{48}\right).$$

Proof. Fix a set $J \subset [n]$ with $|J| = k$. Let ξ_i, $i \le n$, be the indicator function on the event $\{i \in S(J, M)\}$. Then we have

$$|S(J, M)| = \sum_{i=1}^n \xi_i.$$

As rows of M are independent, ξ_i's are independent Bernoulli random variables with

$$\mathbb{P}(\xi_i = 0) = \mathbb{P}(\mathrm{Supp}(R_i(M)) \cap J = \emptyset) = \frac{\binom{n-k}{d}}{\binom{n}{d}}$$

and

$$q := \mathbb{P}(\xi_i = 1) = 1 - \frac{\binom{n-k}{d}}{\binom{n}{d}} = 1 - \prod_{i=1}^{k}\left(1 - \frac{d}{n-i+1}\right),$$

Let $\mu := \mathbb{E}\left(|S(J,M)|\right)$. Then

$$n\left(1-\left(1-\frac{d}{n}\right)^k\right) \le \mu = \sum_{i=1}^{n}\mathbb{E}\xi_i = nq \le n\left(1-\left(1-\frac{d}{n-k+1}\right)^k\right).$$

For $1 \le k \le \frac{\varepsilon n}{d}$ and $d \ge 4$, using the inequality $1 - kx \le (1-x)^k \le 1 - kx + \frac{k(k-1)x^2}{2}$, for $x \in (0,1)$, we obtain that

$$kd(1-\varepsilon/2) \le \mu \le kd(1+\varepsilon/3).$$

Chernoff bound (see, e.g., [18, Theorem 1.1]), for every $\delta > 0$,

$$\mathbb{P}(||S(J,M)| - \mu| \ge \delta\mu) \le 2e^{-\delta^2\mu/3}.$$

Set $\delta = \varepsilon/2$, and combining the estimation for μ, we have

$$\mathbb{P}\left(||S(J,M)| - kd| \ge \varepsilon kd\right) \le 2\exp\left(-\frac{\varepsilon^2(1-\varepsilon/2)kd}{12}\right) \le 2\exp\left(-\frac{\varepsilon^2 kd}{24}\right) := p_k.$$

Then the first result follows from the union bound over $J \subset [n]$, $|J| = k$. Using that $\varepsilon^2 d \ge 48\log n$,

$$\frac{\binom{n}{k+1}p_{k+1}}{\binom{n}{k}p_k} \le n\exp\left(-\varepsilon^2 d/24\right) \le \exp\left(-\varepsilon^2 d/48\right),$$

and the union bound over all k,

$$\mathbb{P}\left(\bigcup_{k\le\varepsilon n/d}(\Omega^{in}_{k,\varepsilon})^c\right) \le 2\sum_{k\le\varepsilon n/d}\binom{n}{k}\exp\left(-\frac{\varepsilon^2 kd}{24}\right) \le 3\exp\left(-\frac{\varepsilon^2 d}{48}\right)$$

for $n \ge 3$. This implies the moreover part. □

Lemma 5.6. *Let $n \ge 2$ and $128\log n \le d \le n/2$. Let $\varepsilon \in \left[\sqrt{\frac{32\log n}{d}}, 1\right)$. Let $1 \le k \le \varepsilon n/d$. Then*

$$\mathbb{P}(\Omega^{out}_{k,\varepsilon}) \ge 1 - \binom{n}{k}\exp\left(-\frac{\varepsilon^2 kd}{16}\right).$$

Moreover,

$$\mathbb{P}\left(\bigcap_{k\le\varepsilon n/d}\Omega^{out}_{k,\varepsilon}\right) \ge 1 - 2\exp\left(-\frac{\varepsilon^2 d}{32}\right).$$

Proof. Fix $J \subset [n]$ with $|J| = k$. Let ξ_i, $i \le n$, be the indicator function on the event $\{i \in S(J, M^T)\}$. Then we have

$$|S(J,M^T)| = \sum_{i=1}^{n}\xi_i.$$

Let $\mathcal{C}_j = \left\{M^T_{ij} : i \in [n]\right\}$, $j \le n$. Applying Lemma 2.6 to the matrix consisting of rows of M indexed by $j \in J$, we observe that

$$\bigcup_{j\in J}\mathcal{C}_j = \left\{M^T_{ij}\ :\ i\in[n], j\in J\right\}$$

is an NA family. Note that $\xi_i = \max_{j\in J} M_{ij}^T$ for $i \in [n]$. For each row i, define $A_i := \{(i,j) : j \in J\}$. Clearly, $A_i \cap A_{i'} = \emptyset$ whenever $i \neq i'$. Then the family $\mathcal{A} = \{A_1, \ldots, A_n\}$ consists of n disjoint subsets of $\{(i,j) : i \in [n], j \in J\}$. For each A_i, consider the function

$$h_{A_i}\left((M_{ij}^T)_{j\in J}\right) := \max_{j\in J} M_{ij}^T$$

which is coordinate-wise non-decreasing. By the disjoint monotone aggregation of NA in Section 2.3, the family

$$\left\{h_{A_i}\left(M_{ij}^T, j \in J\right) : i \in [n]\right\} = \{\xi_i : i \in [n]\}$$

is NA.

Then by Chernoff-Hoeffding bounds for NA random variables (see, e.g., [18, Theorem 3.1]), for every $\delta \in (0,1)$

$$\mathbb{P}\left(|S(J, M^T)| < (1-\delta)\mu\right) \le \exp\left(-\frac{1}{2}\delta^2\mu\right),$$

where $\mu := \mathbb{E}\left(|S(J, M^T)|\right)$.

Let $\mathbf{Col}_j = \mathbf{Col}_j(M^T)$ be the j-th column of M^T. Note that for every $i \in [n]$

$$\mathbb{E}\xi_i = \mathbb{P}(\xi_i = 1) = 1 - \mathbb{P}\left(i \notin \operatorname{Supp}\left(\mathbf{Col}_j\right), \forall j \in J\right) = 1 - \prod_{j\in J} \mathbb{P}\left(i \notin \operatorname{Supp}\left(\mathbf{Col}_j\right)\right) = 1 - \left(1 - \frac{d}{n}\right)^k.$$

Since $\mu = \mathbb{E}\sum_{i=1}^n \xi_i$, then for $1 \le k \le \varepsilon n/d$

$$\mu = n\left(1 - \left(1 - \frac{d}{n}\right)^k\right) \ge dk - \frac{k(k-1)d^2}{2n} \ge kd\left(1 - \frac{\varepsilon}{2}\right),$$

where we again used that for $x \in (0,1)$, $(1-x)^m \le 1 - mx + \binom{m}{2}x^2$ with $x = d/n$. Take $\delta = \varepsilon/2$. Since $(1-\delta)\mu \ge (1-\varepsilon)kd$, we have

$$\mathbb{P}\left(|S(J, M^T)| < (1-\varepsilon)kd\right) \le \exp\left(-\frac{\varepsilon^2(1-\varepsilon/2)}{8}kd\right) \le \exp\left(-\frac{\varepsilon^2 kd}{16}\right).$$

Therefore, by the union bound,

$$\mathbb{P}(\Omega_{k,\varepsilon}^{out}) \ge 1 - \binom{n}{k}\exp\left(-\frac{\varepsilon^2 kd}{16}\right).$$

The proof of the moreover part is similar to the end of the proof of Lemma 5.5. Using $\varepsilon^2 d \ge 32\ln n$, we observe

$$\mathbb{P}\left(\bigcup_{k\le\varepsilon n/d} (\Omega_{k,\varepsilon}^{out})^c\right) \le \sum_{k\le\varepsilon n/d} \binom{n}{k}\exp\left(-\frac{\varepsilon^2 kd}{16}\right) \le 2\exp\left(-\frac{\varepsilon^2 d}{32}\right).$$

for $n \ge 2$. This completes the proof. □

We need the following notation. Given two disjoint sets $J^\ell, J^r \subset [n]$ and a matrix A with $\{0,1\}$ entries, denote

$$I^\ell(A) = I^\ell(A, J^\ell, J^r) := \{i \le n : |\operatorname{Supp}(R_i(A)) \cap J^\ell| = 1 \quad \text{and} \quad \operatorname{Supp}(R_i(A)) \cap J^r = \emptyset\}$$

and

$$I^r(A) = I^r(A, J^\ell, J^r) := \{i \le n : \operatorname{Supp}(R_i(A)) \cap J^\ell = \emptyset \quad \text{and} \quad |\operatorname{Supp}(R_i(A)) \cap J^r| = 1\}.$$

We now combine the expansion properties established in Lemmas 5.5 and 5.6 to derive a key combinatorial property of matrices in the high-probability event $\mathcal{E}_{2.2} \cap \Omega^{in}_{s,\varepsilon} \cap \Omega^{out}_{s,\varepsilon}$, where s is some parameter and $\mathcal{E}_{2.2}$ is the event from Lemma 2.2 (in the case of square matrices, that is, when $m = n$ in Lemma 2.2). The following lemma is an adaptation of the argument in [37, Lemma 2.7] to our setting. While their result was for d-regular matrices, one can show that the same combinatorial counting argument holds for our model, by using the expansion properties for both in-neighbors and out-neighbors. The proof follows the same line of proof as in [37, Lemma 2.7], and for the sake of brevity we omit it.

Lemma 5.7. *Let $5000 \log n \le d \le n/2$ and $\varepsilon \in \left[\sqrt{\frac{48 \log n}{d}}, 0.1\right)$. Let $q \ge 2, m \ge 1$ be integers satisfying $qm \le \varepsilon n/d$. Let $J^\ell, J^r \subset [n]$ be such that*

$$J^\ell \cap J^r = \emptyset, \ |J^\ell| = m, \quad \text{and} \quad |J^r| = (q-1)m.$$

Let $M \in \mathcal{E}_{2.2} \cap \Omega_{qm,\varepsilon}$. Then

$$|I^\ell(M)| \ge (1 - 4\varepsilon q) d |J^\ell| \quad \text{and} \quad |I^\ell(M^T)| \ge (1 - 2\varepsilon q) d |J^\ell|.$$

In particular, if $|J^r| = |J^\ell| = m$ then

$$(1 - 8\varepsilon) dm \le \min\{|I^\ell(M)|, |I^r(M)|\} \le \max\{|I^\ell(M)|, |I^r(M)|\} \le (1 + \varepsilon) dm$$

and

$$(1 - 4\varepsilon) dm \le \min\{|I^\ell(M^T)|, |I^r(M^T)|\} \le \max\{|I^\ell(M^T)|, |I^r(M^T)|\} \le dm.$$

6 Invertibility on the almost constant vectors

In this section, we establish the invertibility of the shifted matrix $M - z\mathbf{I}_n$ and its transpose when acting on vectors from the class of almost-constant vectors, $\mathrm{Cons}(\delta, \rho)$. To this end, we provide separate lower bounds for both the right-multiplication $\|(M - z\mathbf{I}_n)x\|_2$ and the left-multiplication $\|(M^T - z\mathbf{I}_n)x\|_2$, which require different treatments due to asymmetry in the distribution of our model. We will split almost-constant vectors into two parts: almost constant steep vectors (those with "jump", that is, $x_i^* \gg x_j^*$ for some $i \ll j$) and almost constant gradual vectors (those without jumps). Then we further classify our steep vectors as $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ based on a careful selection of the magnitude and location of the first significant jump, see Section 6.1.1 below for precise definitions. We will proceed with our proof by using different techniques for almost constant gradual vectors and each subclass $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ of steep vectors.

We first prove a bound for almost constant gradual vectors in Theorem 6.7, which is less involved as discussed in the introduction.

For vectors in $\mathcal{T}_1$, which have a large jump (of order d), our approach is to partition the initial block of coordinates (from x_1^* to $x_{n_1}^*$) into segments whose lengths are powers of ℓ_0, and then iteratively apply Lemma 5.7. This approach leverages the expansion properties to find a large set of "good" rows where the inner product with a vector $x \in \mathcal{T}_1$ is dominated by its unique large coordinates. A similar idea was also used in [37, Lemma 3.7] for random d-regular matrices. However, unlike random d-regular matrices, we must handle the left- and right-multiplication separately due to asymmetry in distribution (see Section 6.3). To implement this idea, the parameters n_1, ℓ_0 must be carefully chosen. In subclass $\mathcal{T}_1$, the parameter $\varepsilon_0 \lesssim \sqrt{\log n/d}$ is selected as the minimum value of ε to effectively apply the expansion properties in Lemmas 5.5 and 5.6. This selection of ε_0 in turn determines the choice of $n_1 \le \epsilon_0 n/d \lesssim (\log n/d)^{1/2}(n/d)$. The choice of $\ell_0 \lesssim 1/\varepsilon_0$ comes from $\ell_0 \varepsilon_0 < 1/4$ needed in Lemma 5.7.

For vectors in $\mathcal{T}_2$ and $\mathcal{T}_3$, where the jumps are moderate, the direct combinatorial argument is insufficient. We therefore use the ε-net argument, adapting the strategy from [40, Lemma

6.7]. To construct the nets, the triple norm is utilized, and the fact that $\|M - \mathbb{E}M\|$ is bounded by $O(\sqrt{d})$ and our choice of n_1 leads to the choice of $\varepsilon = 1/d^{3/4}$, which satisfies the condition of Lemma 6.11. This, in turn, constrains the scale of $n_2 \lesssim n/d^{3/4}$ and the magnitude of the shift $|z| \lesssim \sqrt{d}\log\log d$ in the proof. The parameter $n_3 \approx \delta n$ is chosen to align with the definition of almost-constant vectors. To optimize the final probability bound, δ is chosen to be a positive absolute constant. To effectively apply the ε-net argument, we need to balance the size of the nets and the individual probability bound. Here again, due to the asymmetry in the distribution of our model, to obtain individual probability bounds, we need to deal with left- and right-multiplication separately. For the right multiplication, the individual probability bound is attained by anti-concentration type techniques in companion with row independence. In contrast, for the left multiplication, due to the lack of independence, the individual probability bound is obtained by anti-concentration type techniques together with a switching argument. The delicate interplay between the expansion properties and the jump structures is essential for the argument. All technical details of this approach are presented in Section 6.4.

6.1 Almost-constant vectors

As in [36, 37], for a lower bound on the smallest singular value $s_n(M - zI_n)$, we consider a decomposition of the complex unit sphere $\mathbb{S}^{2n-1}$ into *almost-constant* and *non-almost-constant vectors.*

Definition 6.1. *Let $\delta, \rho \in (0,1)$. We define the set of almost-constant $Cons(\delta,\rho) \subset \mathbb{C}^n$ as the set of vectors $v \in \mathbb{C}^n$ for which there exists a complex number λ such that the number of coordinates $i \in [n]$ satisfying $|v_i - \lambda| \le \rho\|v\|_2/\sqrt{n}$ is strictly larger than $(1-\delta)n$. The vectors in the remaining set $(Cons(\delta,\rho))^c = \mathbb{C}^n \setminus Cons(\delta,\rho)$ are called non-almost-constant.*

We recall the definition of the Lévy concentration function, which roughly measures the spread of a random variable.

Definition 6.2. *For a random vector $X \in \mathbb{R}^n$ and $\varepsilon \ge 0$, the Lévy concentration of X of width ε is*

$$\mathcal{L}(X,\varepsilon) := \sup_{x\in\mathbb{R}^n} \mathbb{P}(\|X - x\|_2 < \varepsilon).$$

We will use the following bound on the Lévy concentration function of sums of independent random variables.

Lemma 6.3. *[20, Corollary 1 of Theorem 6.1] There exists an absolute positive constant $C_{6.3}$ such that the following holds. Let $n \ge 1$ and let $\xi_1, \dots, \xi_n$ be independent random vectors in $\mathbb{R}^2$. Then for any $t \ge t_0 > 0$, one has*

$$\mathcal{L}\left(\sum_{i=1}^n \xi_i, t\right) \le \frac{C_{6.3}(t/t_0)^2}{\sqrt{n - \sum_{i=1}^n \mathcal{L}(\xi_i, t_0)}}.$$

In particular, if $\alpha \ge \max_{1\le i\le n} \mathcal{L}(\xi_i, t_0)$, then

$$\mathcal{L}\left(\sum_{i=1}^n \xi_i, t\right) \le \frac{C_{6.3}}{\sqrt{n(1-\alpha)}}.$$

For non-almost-constant vectors, we have the following small ball estimate.

Lemma 6.4. *[49, Lemma 8.5] There exists an absolute constant $C_{6.4} > 0$ such that the following holds. Let $1 \le d \le n/2$ and $\delta, \rho > 0$. Sample $\xi \in \{0,1\}^n$ uniformly at random with exactly d ones. Let $v \in (Cons(\delta,\rho))^c$, that is*

$$\sup_{\lambda\in\mathbb{C}} \left|\left\{i \in [n] : |v_i - \lambda| \le \rho/\sqrt{n}\right\}\right| \le (1-\delta)n.$$

Then

$$\mathcal{L}\left(\langle \xi, v\rangle, \rho/\sqrt{n}\right) \le \frac{C_{6.4}}{\sqrt{\delta d}} + C_{6.4} \exp(-\delta^2 d/C_{6.4}).$$

Next we introduce parameters which will be used in the remaining part of the paper. We always assume that $C_{1.7}$ is a large enough absolute constant, that n is large enough, and that $C_{1.7} \log n \le d \le n/2$. Fix a sufficiently small absolute positive constant δ. In order to use Lemma 5.6, we fix

$$\varepsilon_0 = \sqrt{\frac{48 \log n}{d}} \qquad \text{and} \qquad \ell_0 = \lfloor 1/(100\varepsilon_0) \rfloor = \left\lfloor \frac{\sqrt{d}}{100\sqrt{48 \log n}} \right\rfloor \ge 2.$$

Let a_1, a_2, a_3 be three positive, sufficiently small, absolute constants, satisfying

$$a_3 \le a_2/30 \le a_1/900.$$

Set

$$n_1 := \left\lceil \frac{a_1 \varepsilon_0 n}{d} \right\rceil, \qquad n_2 := \left\lfloor \frac{a_2 n}{d^{3/4}} \right\rfloor, \quad \text{and} \quad n_3 := \lfloor a_3 n \rfloor.$$

Finally, if $n_1 > 1$, fix the integer $r \ge 0$ satisfying

$$\ell_0^r < n_1 \le \ell_0^{r+1}.$$

6.1.1 $\mathcal{T}$-vectors

We describe the class of steep vectors, $\mathcal{T}$, which is the union of three parts: $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}_3$. If $n_1 = 1$ set $\mathcal{T}_1 = \emptyset$. Otherwise, set

$$\mathcal{T}_{10} = \left\{ x \in \mathbb{C}^n : x_1^* > 6d x^*_{\min\{\ell_0, n_1\}} \right\},$$

and, for $1 \le i \le r-1$,

$$\mathcal{T}_{1i} := \left\{ x \in \mathbb{C}^n : x \notin \bigcup_{j=0}^{i-1} \mathcal{T}_{1j} \text{ and } x^*_{\ell_0^i} > 6d x^*_{\ell_0^{i+1}} \right\},$$

for $i = r$

$$\mathcal{T}_{1r} := \left\{ x \in \mathbb{C}^n : x \notin \bigcup_{j=0}^{r-1} \mathcal{T}_{1j} \text{ and } x^*_{\ell_0^r} > 6d x^*_{n_1} \right\}.$$

Then

$$\mathcal{T}_1 := \bigcup_{i=0}^{r} \mathcal{T}_{1i}.$$

Finally set

$$\mathcal{T}_2 := \left\{ x \in \mathbb{C}^n : x \notin \mathcal{T}_1 \text{ and } x^*_{n_1} > d^{3/4} x^*_{n_2} \right\},$$

$$\mathcal{T}_3 := \left\{ x \in \mathbb{C}^n : x \notin \mathcal{T}_1 \cup \mathcal{T}_2 \text{ and } x^*_{n_2} > 4x^*_{n_3} \right\}, \quad \text{and} \quad \mathcal{T} := \bigcup_{i=1}^{3} \mathcal{T}_i.$$

Figure 2 offers a visual guide to this classification scheme.

In the case $n_1 > 1$ we denote

$$B(\mathcal{T}_2) = \frac{(6d)^{r+1} d^{1/4}}{26 \log^{1/4} n}, \quad B(\mathcal{T}_3) = \frac{(6d)^{r+1} d}{26 \log^{1/4} n} \quad \text{and} \quad B_{\mathcal{T}} = B(\mathcal{T}^c) = \frac{(6d)^{r+2}}{36(\log n)^{1/4}} \tag{6.1}$$

and for $n_1 = 1$,

$$B(\mathcal{T}_2) = \sqrt{n}, \qquad B(\mathcal{T}_3) = B_{\mathcal{T}} = \sqrt{2n}. \tag{6.2}$$

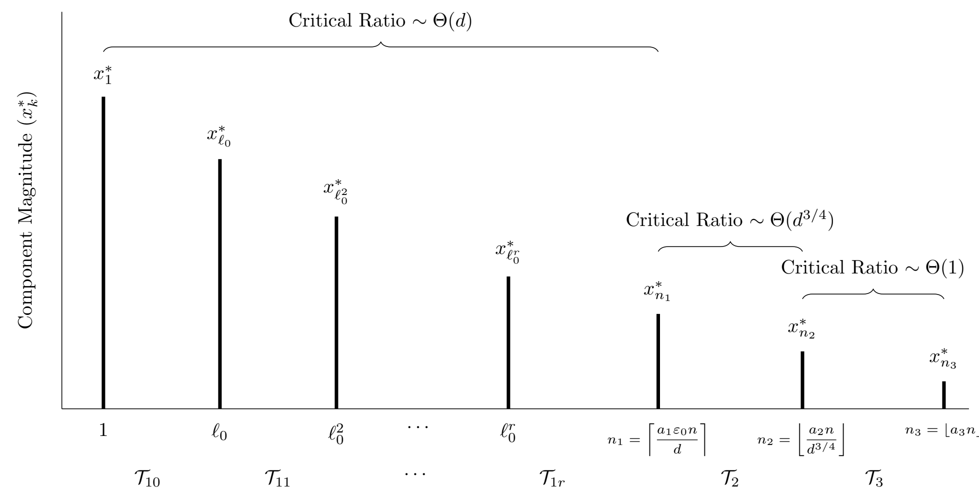

Figure 2: An illustration of the classification of steep vectors into the sets $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$. The vertical axis represents the magnitude of the sorted components (x_k^*) and the horizontal axis represents the component index k. The picture is drawn for the regime $\ell_0 < n_1$, so that $\mathcal{T}_{10}$ is represented by the jump condition $x_1^* > 6d\, x_{\ell_0}^*$. A vector is classified based on the location and the size of the first jump.

6.1.2 Bounds for ℓ_2-norm

The following lemma provides a bound on Euclidean norms of vectors in the class $\mathcal{T}$ and its complement in terms of their order statistics. This result is similar to [40, Lemma 6.4] (see also [37, Lemma 3.5], [34, Lemma 4.3]). Since our choice of parameters as well as the definition of steep vectors is slightly different, we provide the proof for the sake of completeness.

Lemma 6.5. *There exists a positive absolute constant C such that the following hold. Let n be large enough and let $C\log n \le d \le n/2$. Let $y \in \mathcal{T}_2$, $z \in \mathcal{T}_3$, and $w \in \mathcal{T}^c$. Then*

$$\|y\|_2 \le B(\mathcal{T}_2)\, y_{n_1}^*, \qquad \|z\|_2 \le B(\mathcal{T}_3)\, z_{n_2}^*, \qquad \text{and} \qquad \|w\|_2 \le B_{\mathcal{T}}\, w_{n_3}^*.$$

Moreover, if $n_1 > 1$ (so $\mathcal{T}_1 \ne \emptyset$) and $x \in \mathcal{T}_{1j}$ for some $0 \le j \le r$, then

$$\|x\|_2 \le \max\left\{\frac{d^{1/4}\,(6d)^j}{15\log^{1/4} n}, \sqrt{2n}\right\} x_{\ell_0^j}^*.$$

Proof. We first consider the case $n_1 > 1$. Note that in this case $(6d)^{2r+2} \ge 2n$ for large enough n. Indeed, if $d^2 \ge n$, then

$$(6d)^{2r+2} \ge (6d)^2 \ge 36n;$$

if $d^2 \le n$, then, using the definitions of r and n_1,

$$(6d)^{2r+2} \ge \exp\left(2(\log(6d)(\log n_1)/\log \ell_0\right) \ge n_1^4 \ge 2n.$$

Let $z \in \mathcal{T}_3$. Then

$$z_1^* \le (6d) z_{\ell_0}^* \le (6d)^2 z_{\ell_0^2}^* \le \cdots \le (6d)^r z_{\ell_0^r}^* \le (6d)^{r+1} z_{n_1}^* \le (6d)^{r+1}\, d^{3/4}\, z_{n_2}^*.$$

Recall that $\ell_0^r < n_1 \le \ell_0^{r+1}$ and $\ell_0 \le \sqrt{d}/(690\sqrt{\log n})$. Therefore, for large enough d,

$$\begin{aligned}\|z\|_2^2 &= \sum_{s=1}^{\ell_0}(z_s^*)^2 + \sum_{s=\ell_0+1}^{\ell_0^2}(z_s^*)^2 + \sum_{s=\ell_0^2+1}^{\ell_0^3}(z_s^*)^2 + \cdots + \sum_{s=\ell_0^r+1}^{n_1}(z_s^*)^2 + \sum_{s=n_1+1}^{n}(z_s^*)^2\\ &\le \left(\ell_0(6d)^{2r+2} + \ell_0^2(6d)^{2r} + \ell_0^3(6d)^{2r-2} + \cdots + \ell_0^{r+1}(6d)^2 + n\right) d^{3/2}\,(z_{n_2}^*)^2\\ &\le \Big(\ell_0(6d)^{2r+2}\sum_{s\ge 0}(6d)^{-2s}\ell_0^s + n\Big) d^{3/2}\,(z_{n_2}^*)^2 \le \Big(1.5\ell_0(6d)^{2r+2} + n\Big) d^{3/2}\,(z_{n_2}^*)^2\\ &\le 2\ell_0(6d)^{2r+2}d^{3/2}\,(z_{n_2}^*)^2 \le (6d)^{2r+2}d^2\,(z_{n_2}^*)^2/(690\sqrt{\log n}).\end{aligned}$$

Thus,

$$\|z\|_2 \le (6d)^{r+1}\,d\,z_{n_2}^*/(26\log^{1/4} n).$$

This completes the proof of the bound of z. The bounds for y and w are obtained similarly.

Let $x \in \mathcal{T}_{1j}$. If $j = 0$ we clearly have $\|x\|_2 \le \sqrt{n}x_1^*$. If $1 \le j \le r$, then

$$x_1^* \le (6d)x_{\ell_0}^* \le (6d)^2 x_{\ell_0^2}^* \le \cdots \le (6d)^j x_{\ell_0^j}^*.$$

Similarly to the bound for $\|z\|_2$, we obtain

$$\begin{aligned}\|x\|_2^2 &\le \Big(\ell_0(6d)^{2j} + \ell_0^2(6d)^{2j-2} + \cdots + \ell_0^j(6d)^2 + n\Big)\,(x_{\ell_0^j}^*)^2\\ &\le \Big(\ell_0(6d)^{2j}\sum_{s\ge 0}(6d)^{-2s}\ell_0^s + n\Big)\,(x_{\ell_0^j}^*)^2 \le \Big(1.5\ell_0(6d)^{2j} + n\Big)\,(x_{\ell_0^j}^*)^2\\ &\le \Big(\sqrt{d}\,(6d)^{2j}/(460\sqrt{\log n}) + n\Big)\,(x_{\ell_0^j}^*)^2.\end{aligned}$$

This implies the desired bound.

The case $n_1 = 1$ is similar but simpler. Indeed, the bound for y is trivial, for $w \in \mathcal{T}^c$ we have $w_j^* \le d^{3/4}w_{n_2}^* \le 4d^{3/4}w_{n_3}^*$ for $j < n_2$ and $w_j^* \le 4w_{n_3}^*$ for $n_2 < j \le n_3$, hence

$$\|w\|_2^2 \le (4d^{3/4}n_2 + 4n_3 + n)(w_{n_3}^*)^2 \le (4a_2 n + 4a_3 n + n)(w_{n_3}^*)^2 \le 2n(w_{n_3}^*)^2.$$

The bound for z is similar. □

For the class of steep vectors $\mathcal{T}$, we prove the following lower bound, in which we denote

$$\mathcal{T}' := \mathcal{T}_1 \cup \mathcal{T}_2 \cup (\mathcal{T}_3 \cap \mathrm{Cons}(a_3, \rho))$$

and

$$\theta(n,d) = \min\left\{\frac{\sqrt{d}}{4\sqrt{n}},\ \frac{5\,\ell_0^{r/2}\,(d\log n)^{1/4}}{(6d)^r},\ \frac{54(\log n)^{1/4}\sqrt{n}}{(6d)^{r+2}}\right\},$$

whenever $n_1 > 1$,

$$\theta(n,d) = c_1\left((\log n)/d\right)^{1/4}$$

for sufficiently small absolute constant $c_1 > 0$ in the case $n_1 = 1$.

Theorem 6.6. *Let n be a sufficiently large integer and $C\log n \le d \le n/2$ for large enough absolute constant C. Let $p = d/n$, $M \in \mathcal{M}_{n,d}$, $0 < \rho < \sqrt{n}/(B_{\mathcal{T}}d^{3/4})$, and let $z \in \mathbb{C}$ be such that $|z| \le \sqrt{d}\log\log d$. Denote*

$$\mathcal{E}_{steep}(M) := \{\exists x \in \mathcal{T}' \text{ such that } \|(M - z\mathbf{I}_n)x\|_2 < \theta(n,d)\|x\|_2\},$$

and

$$\mathcal{E}_{steep}(M^T) := \left\{\exists x \in \mathcal{T}' \text{ such that } \|(M^T - z\mathbf{I}_n)x\|_2 < \theta(n,d)\|x\|_2\right\},$$

Then

$$\mathbb{P}\left(\mathcal{E}_{steep}(M)\bigcup\mathcal{E}_{steep}(M^T)\right) \le 22/n.$$

6.1.3 Gradual vectors

We now establish the lower bound for the almost-constant gradual vectors $\mathrm{Cons}(a_3,\rho)\setminus\mathcal{T}$, which by definition lack the significant jumps that characterize steep vectors. The proof of this theorem adapts the strategy from [37, Theorem 3.1]. However, key modifications are necessary for our model. To deal with the right multiplication $\|(M-z\mathbf{I}_n)x\|_2$, we use Lemma 5.1 to handle the non-fixed column sums of our matrix. Moreover, the argument is necessarily adapted to align with our specific selection of the jump size and its location for steep vectors. Recall that $B_{\mathcal{T}}$ was defined in (6.1).

Theorem 6.7. *There exists a positive absolute constant C such that the following hold. Let n be large enough, $C\log n \le d \le n/2$ and $M \in \mathcal{M}_{n,d}$. Let $0<\rho\le\sqrt{n}/(5B_{\mathcal{T}})$ and let $z\in\mathbb{C}$ be such that $|z|\le 4d/25$. Then for all vectors $x\in(\mathrm{Cons}(a_3,\rho)\setminus\mathcal{T})$ with probability at least $1-4\exp(-(8n/81))-n\exp(-d/108)$,*

$$\|(M-z\mathbf{I}_n)x\|_2 \ge \frac{3d\sqrt{n}}{25B_{\mathcal{T}}}\|x\|_2 \ge \theta(n,d)\|x\|_2,$$

and deterministically,

$$\|(M^T-z\mathbf{I}_n)x\|_2 \ge \frac{3d\sqrt{n}}{25B_{\mathcal{T}}}\|x\|_2 \ge \theta(n,d)\|x\|_2.$$

To combine Theorems 6.6 and 6.7, we introduce the following two events. Let $\rho>0$ and $z\in\mathbb{C}$ satisfy assumptions of both theorems. Denote

$$\mathcal{E}_{6.8}(M,\rho) := \{M\in\mathcal{M}_{n,d} : \exists x\in \mathrm{Cons}(a_3,\rho) \text{ such that } \|(M-z\mathbf{I}_n)x\|_2\le\theta(n,d)\|x\|_2\},$$

and

$$\mathcal{E}_{6.8}(M^T,\rho) := \left\{M\in\mathcal{M}_{n,d} : \exists x\in \mathrm{Cons}(a_3,\rho) \text{ such that } \|(M^T-z\mathbf{I}_n)x\|_2\le\theta(n,d)\|x\|_2\right\}.$$

The following is an immediate consequence of Theorems 6.6 and 6.7.

Corollary 6.8. *There exist positive absolute constants $C_{6.8}, c_{6.8}>0$ such that the following holds. Let n be large enough and $C_{6.8}\log n\le d\le n/2$. Let $0<\rho\le\sqrt{n}/(B_{\mathcal{T}}d^{3/4})$. Let $z\in\mathbb{C}$ be such that $|z|\le\sqrt{d}\log\log d$. Then*

$$\mathbb{P}\left(\mathcal{E}_{6.8}(M,\rho)\bigcup\mathcal{E}_{6.8}(M^T,\rho)\right)\le 23/n.$$

6.2 Proof of Theorem 6.7

Without loss of generality, we assume that $x\neq 0$. Fix a permutation $\sigma=\sigma_x$ of $[n]$ such that $|x_{\sigma(i)}|=x_i^*$. Since $x\notin\mathcal{T}\cup\{0\}$, we have $x^*_{n_3}\neq 0$. Since $x\in\mathrm{Cons}(a_3,\rho)$ there exists $\lambda_0\in\mathbb{C}$ such that the cardinality of

$$J_0 := \left\{i\in[n] : |x_i-\lambda_0|\le\frac{\rho\|x\|_2}{\sqrt{n}}\right\}$$

is at least $n-n_3+1$. Therefore, there exists k,ℓ such that $k\le n_3<\ell$ and $\sigma(k),\sigma(\ell)\in J_0$. By Lemma 6.5, $\|x\|_2\le B_{\mathcal{T}}x^*_{n_3}$, hence, using $\rho\le\sqrt{n}/(5B_{\mathcal{T}})$,

$$\frac{\rho\|x\|_2}{\sqrt{n}}\le\frac{x^*_{n_3}}{5}.$$

Thus,

$$x^*_{n_3} \le x^*_k \le |\lambda_0| + \frac{\rho\|x\|_2}{\sqrt{n}} \le |\lambda_0| + \frac{x^*_{n_3}}{5}.$$

Similarly, we have

$$x^*_{n_3} \ge x^*_\ell \ge |\lambda_0| - \frac{\rho\|x\|_2}{\sqrt{n}} \ge |\lambda_0| - \frac{x^*_{n_3}}{5}.$$

Therefore

$$\frac{5}{6}|\lambda_0| \le x^*_{n_3} \le \frac{5}{4}|\lambda_0| \qquad \text{and} \qquad \frac{\rho\|x\|_2}{\sqrt{n}} \le \frac{x^*_{n_3}}{5} \le \frac{1}{4}|\lambda_0|. \tag{6.3}$$

Next, we split $[n] \setminus J_0$ as follows. Set

$$J_1 = \sigma([n_1]) \setminus J_0, J_2 = \sigma([n_2]) \setminus (J_0 \cup J_1),\ J_3 = \sigma([n_3]) \setminus (J_0 \cup J_1 \cup J_2),\ J_4 = \sigma([n]) \setminus \left(\cup_{i=0}^3 J_i\right).$$

Then $|J_1| \le n_1, |J_2| \le n_2, |J_3| \le n_3,\ |J_4| \le n_3,\ [n] = \bigcup_{i=0}^4 J_i$. One has

$$\forall j \in J_2 : \quad |x_j| \le x^*_{n_1} \le d^{3/4} x^*_{n_2} \le 4 d^{3/4} x^*_{n_3} \le 5 d^{3/4} |\lambda_0|,$$

$$\forall j \in J_3 : \quad |x_j| \le x^*_{n_2} \le 4 x^*_{n_3} \le 5|\lambda_0|,$$

and

$$\forall j \in J_4 : \quad |x_j| \le x^*_{n_3} \le \frac{5}{4}|\lambda_0|.$$

Now, given a matrix $M \in \mathcal{M}_{n,d}$, we define

$$I_1 = \left\{i \in [n] : \mathrm{Supp}(R_i(M^T)) \cap J_1 \ne \emptyset\right\}, \quad I_1' = \{i \in [n] : \mathrm{Supp}(R_i(M)) \cap J_1 \ne \emptyset\}$$

and, denoting $n_4 = n_3$, for $\ell = 2, 3, 4$ define

$$I_\ell = \left\{i \in [n] : \left|\mathrm{Supp}\ R_i(M^T) \cap J_\ell\right| \ge \frac{9 n_\ell d}{n}\right\}, \quad I_\ell' = \left\{i \in [n] : |\mathrm{Supp}(R_i(M)) \cap J_\ell| \ge \frac{81 n_\ell d}{n}\right\}.$$

Since columns of M^T contain exactly d ones, we first note that for M^T one has

$$|I_1| \le d n_1 \le \frac{n}{9},$$

provided that a_1 is small enough. Then applying Lemma 5.2 to $J = J_2, J_3, J_4$,

$$\left|\left\{i \le n : |\mathrm{Supp}(R_i(M^T)) \cap J_\ell| \ge \frac{9|J_\ell| d}{n}\right\}\right| \le \frac{n}{9}.$$

Since $|J_\ell| \le n_\ell$ for $\ell = 2, 3, 4$, we obtain $|I_\ell| \le n/9$, for $1 \le \ell \le 4$.

For M, applying Lemma 5.1, we observe that with probability at least $1 - 4\exp(-(8n/81))$,

$$\forall \ell \in \{1, 2, 3, 4\} : \quad \left|\left\{i \le n : |\mathrm{Supp}(R_i(M)) \cap J_\ell| \ge \frac{81|J_\ell| d}{n}\right\}\right| \le \frac{n}{9}.$$

Note that

$$\frac{81|J_1| d}{n} \le \frac{81 n_1 d}{n} < 1,$$

provided that a_1 is small enough. Since $|J_\ell| \le n_\ell$ for $\ell = 2, 3, 4$, defining the event

$$\mathcal{E} := \{\forall \ell \in \{1, 2, 3, 4\} : \quad |I_\ell'| \le n/9\},$$

we obtain

$$\mathbb{P}(\mathcal{E}) \geq 1 - 4\exp(-(8n/81)).$$

Set $I := [n] \backslash \big(\cup_{i=1}^4 I_i \cup \sigma([n_3])\big)$ and $I' := [n] \backslash \big(\cup_{i=1}^4 I_i' \cup \sigma([n_3])\big)$. Then, assuming $a_3 \leq 0.001$,

$$|I| \geq n - \frac{4n}{9} - n_3 \geq \frac{n}{2},$$

and, assuming that $\mathcal{E}$ occurs,

$$|I'| \geq n - \frac{4n}{9} - n_3 \geq \frac{n}{2}.$$

Note that

$$\forall i \in I \cup I' : \qquad |x_i| \leq x_{n_3}^* \leq \frac{5}{4}|\lambda_0|.$$

In the case of M^T, for $i \in I$, denote

$$J_\ell' = J_\ell'(i) := \mathrm{Supp}(R_i(M^T)) \cap J_\ell, \quad \text{ for } \ell \in \{0,1,2,3,4\}$$

and note that $J_1' = \emptyset$ since $i \notin I_1$. Then using the triangle inequality, we observe that for every $i \in I$,

$$|\langle R_i((M^T - z\mathbf{I}_n)), x\rangle| \geq |\sum_{j \in J_0'} x_j| - \sum_{j \in J_2'} |x_j| - \sum_{j \in J_3'} |x_j| - \sum_{j \in J_4'} |x_j| - |zx_i|.$$

We estimate each term separately. First note that similarly to Lemma 2.2 one observes that with probability at least $1 - n\exp(-d/108)$ for every $i \leq n$ one has $|\mathrm{Supp}(R_i(M^T))| \geq 5d/6$. Thus, for $i \notin I_1 \cup I_2 \cup I_3 \cup I_4$ (in particular, $J_1' = \emptyset$) one has

$$|J_0'| \geq 5d/6 - |J_2'| - |J_3'| - |J_4'| \geq 5d/6 - \frac{9n_2 d}{n} - \frac{18n_3 d}{n} \geq 5d/6 - 27a_3 d \geq 4d/5$$

(recall, $a_3 \leq 1/1000$). Therefore, using the definition of J_0, we obtain

$$|\sum_{j \in J_0'} x_j| \geq |\lambda_0||J_0'| - \sum_{j \in J_0'} |x_j - \lambda_0| \geq \frac{4d}{5}\left(|\lambda_0| - \frac{\rho}{\sqrt{n}}\|x\|_2\right).$$

Meanwhile, one has

$$\sum_{j \in J_2'} |x_j| + \sum_{j \in J_3'} |x_j| + \sum_{j \in J_4'} |x_j| \leq |J_2'|x_{n_1}^* + |J_3'|x_{n_2}^* + |J_4'|x_{n_3}^*$$
$$\leq \left(\frac{45n_2 d^{7/4}}{n} + \frac{57n_3 d}{n}\right)|\lambda_0| \leq \frac{d\,|\lambda_0|}{5},$$

provided $a_2 \leq 1/250, a_3 \leq a_2/30$. Putting together the above estimates, and using that for $i \in I$, $i \notin \sigma([n_3])$, hence $|x_i| \leq x_{n_3}^* \leq (5/4)|\lambda_0|$, we obtain for $d \geq 7^4$

$$|\langle R_i((M^T - z\mathbf{I}_n)), x\rangle| \geq \frac{4d}{5}\left(|\lambda_0| - \frac{\rho}{\sqrt{n}}\|x\|_2\right) - \frac{d}{5}|\lambda_0| - \frac{5}{4}|\lambda_0||z| \geq \frac{d}{5}|\lambda_0|,$$

where we used (6.3) and $|z| \leq 4d/25$. Since $|I| \geq n/2$, by (6.3), this implies that

$$\|(M^T - z\mathbf{I}_n)x\|_2 \geq \frac{|\lambda_0|d}{5}\sqrt{\frac{n}{2}} \geq \frac{4d\sqrt{n}}{25\sqrt{2}}x_{n_3}^* \geq \frac{3d\sqrt{n}}{25B_{\mathcal{T}}}\|x\|_2,$$

where we again used $\|x\|_2 \leq B_{\mathcal{T}} x_{n_3}^*$.

In the case of M, on event $\mathcal{E}$, for $k \in I'$, define $J_\ell'' = J_\ell''(k) := \mathrm{Supp} R_k(M) \cap J_\ell$ for $\ell \in \{0,1,2,3,4\}$. Repeating the above argument (note that $|\mathrm{Supp} R_k(M)| = d$) yields,

$$|\langle R_k((M - z\mathbf{I}_n)), x\rangle| \geq \frac{d}{5}\,|\lambda_0|.$$

Hence, conditioned on $\mathcal{E}$,

$$\|(M - z\mathbf{I}_n)x\|_2 \geq \frac{3d\sqrt{n}}{25B_{\mathcal{T}}}\,\|x\|_2.$$

This completes the proof.

6.3 Lower bounds for vectors from $\mathcal{T}_1$

In this section, we provide lower bounds on $\|(M - z\mathbf{I}_n)x\|_2$ and $\|(M^T - z\mathbf{I}_n)x\|_2$ for vectors from $\mathcal{T}_1$. As $\mathcal{T}_1 = \emptyset$ in the case $n_1 = 1$, the results of this section are relevant only in the case $n_1 > 1$. We essentially follow the proof in [37, Lemma 3.7] with necessary modifications. The primary difference lies in the analysis of the left–multiplication $\|\bar{x}(M - z\mathbf{I}_n)\|_2$, where fixed column sums of M are not fixed as in the d-regular case. We apply a probabilistic argument involving Lemma 2.2 to resolve this issue.

Lemma 6.9. *There exists an absolute constant $C > 0$ such that the following holds. Let $n \geq 3$. Let $C \log n \leq d \leq n/2$ and $z \in \mathbb{C}$ satisfy $|z| \leq d$. Consider the events*

$$\mathcal{E}_{6.9}(M^T) := \left\{\exists x \in \mathcal{T}_1 \quad \text{such that} \quad \|(M^T - z\mathbf{I}_n)x\|_2 \leq \omega(n,d)\|x\|_2\right\},$$

and

$$\mathcal{E}_{6.9}(M) := \{\exists x \in \mathcal{T}_1 \quad \text{such that} \quad \|(M - z\mathbf{I}_n)x\|_2 \leq \omega(n,d)\|x\|_2\},$$

where

$$\omega(n,d) = \min\left\{\frac{\sqrt{d}}{4\sqrt{n}},\ \frac{5\,\ell_0^{r/2}\,(d\log n)^{1/4}}{(6d)^r}\right\}.$$

Then

$$\mathbb{P}\left(\mathcal{E}_{6.9}(M)\bigcup\mathcal{E}_{6.9}(M^T)\right) \leq \frac{7e}{n}.$$

Proof. We first estimate the probability of the event $\mathcal{E}_{6.9}(M^T)$. Recall that the parameters $\ell_0, \varepsilon_0, n_1$ are defined in the beginning of Section 6 and the events $\Omega^{in}_{k,\varepsilon_1}$ and $\Omega^{out}_{k,\varepsilon_2}$ for $1 \leq k \leq n$, $\varepsilon_1 \in [\varepsilon_0, 1/2]$, $\varepsilon_2 \in [\varepsilon_0, 1)$ — in Section 5.3. Without loss of generality assume $n_1 > 1$ (otherwise, $\mathcal{T}_1 = \emptyset$). For each $0 \leq i \leq r-1$, let

$$\mathcal{E}_i := \Omega^{in}_{\ell_0^i,\varepsilon_0} \cap \Omega^{out}_{\ell_0^i,\varepsilon_0}, \quad \mathcal{E}_r := \Omega^{in}_{n_1,\varepsilon_0} \cap \Omega^{out}_{n_1,\varepsilon_0}, \quad \text{and} \quad \mathcal{E} := \bigcap_{i=0}^{r}\mathcal{E}_i.$$

We choose $\tau = 1/2$ in Lemma 2.2, and set

$$\mathcal{E}' := \mathcal{E} \cap \mathcal{E}_{2.2}.$$

By Lemma 5.5, Lemma 5.6, and since $\varepsilon_0 < 1/2$, for every $1 \leq k \leq \varepsilon_0 n/d$ one has

$$\begin{aligned}\mathbb{P}\big((\Omega^{in}_{k,\varepsilon_0} \cap \Omega^{out}_{k,\varepsilon_0})^c\big) &\leq 2\binom{n}{k}\exp\Big(-\frac{\varepsilon_0^2}{24}kd\Big) + \binom{n}{k}\exp\Big(-\frac{\varepsilon_0^2}{16}kd\Big)\\ &\leq 3\binom{n}{k}\exp\Big(-\frac{\varepsilon_0^2}{24}kd\Big) \leq 3\Big(\frac{en}{k}\Big)^k\exp\Big(-\frac{\varepsilon_0^2}{24}kd\Big).\end{aligned}$$

Using that $\varepsilon_0^2 d = 48 \log n$, we obtain

$$\mathbb{P}\big((\Omega_{k,\varepsilon_0}^{in} \cap \Omega_{k,\varepsilon_0}^{out})^c\big) \leq 3\Big(\frac{en}{k}\Big)^k n^{-2k} = 3\Big(\frac{e}{kn}\Big)^k$$

Using that $\ell_0 \geq 2$ and $n_1 > \ell_0^r \geq 1$, by the union bound,

$$\mathbb{P}(\mathcal{E}^c) \leq \sum_{j=0}^{r} \mathbb{P}(\mathcal{E}_j^c) \leq 3\sum_{j=0}^{r-1}\Big(\frac{e}{n\ell_0^j}\Big)^{\ell_0^j} + 3\Big(\frac{e}{nn_1}\Big)^{n_1} \leq \frac{3e}{n}\Big(\sum_{j=0}^{r-1}\frac{1}{\ell_0^j} + \frac{1}{n_1}\Big) \leq \frac{3e}{n}\sum_{j=0}^{r}\frac{1}{\ell_0^j} \leq \frac{6e}{n}.$$

Then, by Lemma 2.2 applied with $m = n$,

$$\mathbb{P}((\mathcal{E}')^c) \leq \mathbb{P}(\mathcal{E}^c) + \mathbb{P}(\mathcal{E}_{2.2}^c) \leq \frac{7e}{n}.$$

Fix a realization $M = (M_{ij}) \in \mathcal{E}'$. For $0 \leq i \leq r$, set

$$m_1 = \ell_0^i, \qquad m_2 = \begin{cases} \ell_0^{i+1}, & \text{if } 0 \leq i \leq r-1 \\ n_1 & \text{if } i = r \end{cases}$$

Recall that, given x, the permutation $\sigma = \sigma_x$ denotes a permutation $[n]$ such that $x_i^* = |x_{\sigma(i)}|$ for $i \in [n]$. Define the index sets

$$J^\ell = J^\ell(x) = \sigma([m_1]), \quad J^r = J^r(x) = \sigma([m_2] \setminus [m_1]), \quad J = J(x) = (J^\ell \cup J^r)^c.$$

Fix $0 \leq i \leq r$ and $x \in \mathcal{T}_{1i}$. Then for all $j \in J^\ell$ we have

$$|x_j| \geq x_{m_1}^* \geq 6dx_{m_2}^*, \quad \text{and} \quad \max_{k \in J} |x_k| \leq x_{m_2}^* \leq x_{m_1}^*/(6d).$$

Let A be either M or M^T. Recall that $I^\ell(A) = I^\ell(A, J^\ell, J^r)$ is the (random) set of rows of A having exactly one 1 in $J^\ell(x)$ and no 1's in $J^r(x)$. Next we apply Lemma 5.7 with $q = \ell_0$, $m = m_1$ (with corresponding adjustment for $i = r$, where we have to take $qm = m_2$ and $q = m_2/m_1 \leq \ell_0$ — we may assume that it is integer) to obtain

$$|I^\ell(M^T)| \geq (1 - 2\varepsilon_0\ell_0)d|J^\ell| \geq 3dm_1/5,$$

and

$$|I^\ell(M)| \geq (1 - 4\varepsilon_0\ell_0)d|J^\ell| \geq 3dm_1/5.$$

Let $I_0^\ell(A) := I^\ell(A) \setminus (J^\ell \cup J^r)$. In both cases ($A = M$ or $A = M^T$), one has

$$|I_0^\ell(A)| \geq 3dm_1/5 - m_2 \geq dm_1/2$$

provided that the absolute constant in $d \geq C \log n$ is large enough.

Now let B denote $(M - z\mathbf{I}_n)$ if $A = M$ or $(M^T - z\mathbf{I}_n)$ if $A = M^T$. For any $s \in I_0^\ell(A)$ there exists $j(s) \in J^\ell$ such that the s-th row of A satisfies

$$\text{Supp } R_s(A) \cap J^\ell = \{j(s)\} \qquad \text{and} \qquad \text{Supp } R_s(A) \cap J^r = \emptyset.$$

Then one has

$$|\langle R_s(B), x\rangle| = \Big|x_{j(s)} + \sum_{k \in J} A_{sk}x_k - zx_s\Big| \geq \big|x_{j(s)}\big| - \sum_{k \in J} A_{sk}|x_k| - |z||x_s|.$$

If $A = M^T$, $R_s(A)$ is the s-th row of M^T (i.e., s-th column of M). Since $M \in \mathcal{E}_{2.2}$,

$$\sum_{k\in J} M^T_{sk} = \sum_{k\in J} M_{ks} \le \sum_{k=1}^{n} M_{ks} \le 3d/2.$$

If $A = M$, $R_s(A)$ is the s-th row of M, hence

$$\sum_{k\in J} M_{sk} \le \sum_{k=1}^{n} M_{sk} = d.$$

Using $|x_k| \le x^*_{m_2}$ for $k \in J$, $x^*_s \le x^*_{m_2}$ (since $s \notin J^\ell \cup J^r$), $x^*_{m_1} \ge 6dx^*_{m_2}$, and $|z| \le d$, we obtain

$$|\langle R_s(B), x\rangle| \ge x^*_{m_1} - x^*_{m_2} \sum_{k\in J} A_{sk} - |z|x^*_{m_2} \ge x^*_{m_1} - \frac{x^*_{m_1}}{6d} \cdot \frac{3d}{2} - \frac{x^*_{m_1}}{6} \ge \frac{1}{2} x^*_{m_1}.$$

Since this holds for at least $|I^\ell_0(A)| \ge dm_1/2$ indices s, we have

$$\|Bx\|_2 \ge \frac{1}{2} x^*_{m_1} \sqrt{\frac{dm_1}{2}} = \frac{\sqrt{d}\, \ell_0^{i/2}}{2\sqrt{2}}\, x^*_{m_1}.$$

Applying Lemma 6.5 for $x \in \mathcal{T}_{1i}$, we obtain that in both cases $B = (M - z\mathbf{I}_n)$ and $B = (M^T - z\mathbf{I}_n)$,

$$\|Bx\|_2 \ge \min\left\{ \frac{\sqrt{d}\, \ell_0^{i/2}}{4\sqrt{n}}, \frac{5\sqrt{d}\, \ell_0^{i/2} \log^{1/4} n}{d^{1/4}\,(6d)^i} \right\} \|x\|_2 \ge \min\left\{ \frac{\sqrt{d}}{4\sqrt{n}}, \frac{5\, \ell_0^{r/2}\, (d\log n)^{1/4}}{(6d)^r} \right\} \|x\|_2.$$

This completes the proof. $\square$

6.4 Lower bounds for vectors from $\mathcal{T}_2 \cup (\mathcal{T}_3 \cap \mathbf{Cons}(a_3, \rho))$

In this section, we provide lower bounds on $\|(M - z\mathbf{I}_n)x\|_2$ and $\|(M^T - z\mathbf{I}_n)x\|_2$ for vectors from $\mathcal{T}_2 \cup (\mathcal{T}_3 \cap \mathrm{Cons}(a_3, \rho))$. We first estimate the cardinality of certain nets — discretizations of the sets of steep vectors $\mathcal{T}_2$ and $\mathcal{T}_3$ introduced earlier. Then we obtain individual probability bounds for a fixed vector in a net. Finally we conclude with the standard ε-net argument.

6.4.1 Nets

Recall that $\mathbf{1} = (1, 1, \ldots, 1)$, $\mathbf{e}$ denotes the vector $\mathbf{1}/\sqrt{n}$, $P_{\mathbf{e}}$ denotes the projection on span of $\mathbf{e}$, and $P_{\mathbf{e}^\perp}$ denotes the projection on $\mathbf{e}^\perp$. The triple norm on $\mathbb{C}^n$ is defined by

$$|||x||| = \sqrt{\|P_{\mathbf{e}^\perp} x\|_2^2 + d\|P_{\mathbf{e}} x\|_2^2}.$$

For convenience, we will consider the following normalization:

$$\mathcal{T}_2' := \left\{x \in \mathcal{T}_2 : x^*_{n_1} = 1\right\} = \left\{\frac{x}{x^*_{n_1}} : x \in \mathcal{T}_2\right\} \quad \text{and} \quad \mathcal{T}_3' := \left\{x \in \mathcal{T}_3 : x^*_{n_2} = 1\right\} \cap \mathrm{Cons}(a_3, \rho),$$

where $0 < \rho \le \sqrt{n}/(B_{\mathcal{T}} d^{3/4})$.

Lemma 6.10. *There exists an absolute constant $C > 0$ such that the following holds. Let $n \ge 1$ and let $C\log n \le d \le n/2$. Let $0 < \rho \le \sqrt{n}/(B_{\mathcal{T}} d^{3/4})$. Let $i \in \{2, 3\}$. There exists a set $\mathcal{N}_i = \mathcal{N}_i' + \mathcal{N}_i''$, $\mathcal{N}_i'' \subset$ span$\{\mathbf{1}\}$, with the following properties:*

1. $|\mathcal{N}_i| \le \exp(6n_i \log d)$.

2. For every $u \in \mathcal{N}_i'$ *one has* $u_j^* \le 1/4 + d^{-3/4}$ *for all* $j \ge n_i$.

3. For every $x \in \mathcal{T}_i'$ *there is* $u \in \mathcal{N}_i'$ *and* $w \in \mathcal{N}_i''$ *such that*

$$\|x-u\|_\infty \le \frac{1}{d^{3/4}}, \qquad \|w\|_\infty \le \frac{1}{d^{3/4}}, \qquad \text{and} \qquad \|\!|x-u-w|\!\| \le \frac{\sqrt{2n}}{d^{3/4}}.$$

As the proof of this lemma repeats the proofs from [37, Lemma 3.8] and [40, Lemma 6.7], we omit it.

6.4.2 Individual probability

We turn now to the individual probability bounds. Following the proof in [34, 40], we consider any $n \times n$ complex matrix W instead of the shift $z\mathbf{I}_n$. Recall that $\mathcal{M}_{n,d}$ denotes the class of all $n \times n$ matrices with entries in $\{0,1\}$ for which each row sums to d. Recall that the probability measure $\mathbb{P}$ on $\mathcal{M}_{n,d}$ is the uniform probability measure, in particular rows of a random matrix from $\mathcal{M}_{n,d}$ are independent. For a fixed vector x from the nets constructed earlier, we need to bound from below both $\|(M+W)x\|_2$ and $\|(M^T+W)x\|_2$ for $M \in \mathcal{M}_{n,d}$. As M has independent rows, whereas M^T does not, we will treat $A = M$ and $A = M^T$ separately. To obtain a lower bound, it suffices to obtain an anti-concentration bound for the inner products $\langle R_i(A+W), x^\dagger \rangle$ for indices $i \in [n]$, where A stands either for M or for M^T, given a vector $z \in \mathbb{C}^n$,

$$x^\dagger = \bar{x} = (\bar{x}_i)_{i=1}^n$$

for $x \in \mathbb{C}^n$ with $\bar{x}_i$ is the complex conjugate of $x_i \in \mathbb{C}$. To estimate this inner product, we will apply Esseen inequalities (see Lemma 6.3) for sums of independent random variables. To create independent random variables, we will use Lemma 5.7 for $2m$ columns of A corresponding to m largest and m smallest (in modulus) coordinates of x, where in our application $m = n_1$ or $m = n_2$. The combinatorial input Lemma 5.7 applies only to the sets of columns of size at most $\varepsilon_0 n/d$. This restriction is satisfied when $m = n_1$, but fails for $m = n_2$. To overcome this problem, following the idea from [34, Section 4.3], we split the selected $2m$ columns into disjoint blocks of size at most $\varepsilon_0 n/d$ each. Conditioning on this splitting, we obtain independent random variables such that their sum is essentially $\langle R_i(A+W), x^\dagger \rangle$. This allows us to apply Lemma 6.3 for sums of independent random variables.

Recall that given $J \subset [n]$ and an $n \times n$ matrix A, by $I(J,A)$ we denote the set of row indices where the row has exactly one 1 in the columns indexed by J, that is,

$$I(J,A) = \{i \in [n] : |\mathrm{Supp}(R_i(A)) \cap J| = 1\}.$$

Given a subset $J \subset [n]$, we define two families of $n \times |J|$ matrices,

$$\mathcal{M}_J^{\mathrm{col}} := \left\{ V \in \mathbb{R}^{n \times |J|} : \exists M \in \mathcal{M}_{n,d} \text{ such that } V = M_{[:,J]} \right\}$$

and

$$\mathcal{M}_J^{\mathrm{row}} := \left\{ V \in \mathbb{R}^{n \times |J|} : \exists M \in \mathcal{M}_{n,d} \text{ such that } V = (M^T)_{[:,J]} \right\},$$

where for a matrix A, $A_{[:,J]}$ denotes the submatrix consisting of the columns of A with indices in J.

Fix a positive integer $1 \le q_0 \le n$ and a disjoint partition $J_0, J_1, \ldots, J_{q_0}$ of $[n]$. Given subsets $I_1, I_2, \ldots, I_{q_0}$ of $[n]$, set $\mathcal{I} = (I_1, I_2, \ldots, I_{q_0})$ and consider the following two classes. For $V \in \mathcal{M}_{J_0}^{\mathrm{col}}$, denote

$$\mathcal{F}(\mathcal{I}, V) := \left\{ M \in \mathcal{M}_{n,d} : M_{[:,J_0]} = V \text{ and } I(J_k, M) = I_k, \forall k \in [q_0] \right\}$$

and for $V \in \mathcal{M}^{\text{row}}_{J_0}$, denote

$$\mathcal{G}(\mathcal{I}, V) := \left\{ M \in \mathcal{M}_{n,d} : (M^T)_{[:,J_0]} = V \text{ and } I(J_k, M^T) = I_k, \forall k \in [q_0] \right\}.$$

Note that these two classes depend on the choice of $\mathcal{I}$ and they can be empty. In words, we start with a matrix A such that either $A = M$ or $A = M^T$, for $M \in \mathcal{M}_{n,d}$ and proceed as follows. First, we fix the columns of A indexed by J_0 to be V. Then for each $k \in [q_0]$, we also fix the row indices of A that have exactly one 1 in columns of A indexed by J_k.

In both cases, for any fixed partition $J_0, J_1, \dots, J_{q_0}$ of $[n]$, the families $\{\mathcal{F}(\mathcal{I}, V)\}_{\mathcal{I},V}$ and $\{\mathcal{G}(\mathcal{I}, V)\}_{\mathcal{I},V}$ form disjoint decompositions of $\mathcal{M}_{n,d}$:

$$\mathcal{M}_{n,d} = \bigcup_{V \in \mathcal{M}^{\text{col}}_{J_0}} \bigcup_{\mathcal{I} \in (\mathcal{P}([n]))^{q_0}} \mathcal{F}(\mathcal{I}, V) \qquad \text{and} \qquad \mathcal{M}_{n,d} = \bigcup_{V \in \mathcal{M}^{\text{row}}_{J_0}} \bigcup_{\mathcal{I} \in (\mathcal{P}([n]))^{q_0}} \mathcal{G}(\mathcal{I}, V),$$

where $\mathcal{P}([n])$ is the power set of $[n]$.

Moreover, given $\mathcal{I}$ and V, we further split each class $\mathcal{F}(\mathcal{I}, V)$ (and similarly $\mathcal{G}(\mathcal{I}, V)$) into smaller equivalence classes using an equivalence relation defined below.

Equivalence for $\mathcal{F}(\mathcal{I}, V)$: Fix a row index $i_0 \in [n]$ and a subset $K \subset [q_0]$. Let

$$K_0 := \{0\} \cup ([q_0] \setminus K), \qquad J^* = J^*(K) := \bigcup_{q \in K_0} J_q = J_0 \cup \bigcup_{q \in [q_0] \setminus K} J_q.$$

We say that two matrices $M = (M_{ij}), M' = (M'_{ij}) \in \mathcal{F}(\mathcal{I}, V)$ are equivalent if

(i) For all $1 \le s \le i_0 - 1$ and for all $j \in [n]$, $M_{sj} = M'_{sj}$.

(ii) For all $s \in [n]$ and all $j \in J^*$, $M_{sj} = M'_{sj}$.

(iii) For every row $s \in [n]$ and for every $q \in K$, the number of 1's in columns indexed by J_q is the same:
$$\sum_{j \in J_q} M_{sj} = \sum_{j \in J_q} M'_{sj}.$$

Denote by $\mathcal{H} := \mathcal{H}(\mathcal{F}(\mathcal{I}, V), i_0, K)$ the set of all equivalence classes corresponding to this relation. Note that
$$\mathcal{F}(\mathcal{I}, V) = \bigcup_{H \in \mathcal{H}^M} H.$$

Equivalence for $\mathcal{G}(\mathcal{I}, V)$: Fix a row index $i_0 \in [n]$ and a subset $K \subset [q_0]$, and define K_0 and $J^*(K)$ as above. For $M, M' \in \mathcal{G}(\mathcal{I}, V)$, we say that two matrices M, M' are equivalent if their transposes M^T and $(M')^T$ satisfy the same three conditions above with $M^T, (M')^T$ in place of M, M'. Denote by $\mathcal{H}^{M^T} := \mathcal{H}(\mathcal{G}(\mathcal{I}, V), i_0, K)$ the set of all equivalence classes corresponding to this relation, then
$$\mathcal{G}(\mathcal{I}, V) = \bigcup_{H \in \mathcal{H}^{M^T}} H.$$

Note that for matrices in a given class $H \in \mathcal{H}^M$ (or $H \in \mathcal{H}^{M^T}$), all columns in $J^*(K)$ and the first $i_0 - 1$ rows are fixed, thus the randomness comes only from the positions on the 1's inside the blocks $[n] \times J_q, q \in K$, subject to a prescribed sum in each row. Thus, in a sense, these blocks are independent of each other on H.

Finally, we fix a vector $x \in \mathbb{C}^n$, a fixed constant vector $y = y_1 \mathbf{1} \in \text{span}\{\mathbf{1}\}, y_1 \in \mathbb{C}$, and a deterministic matrix $W = (w_{ij})$. Let $A = (A_{ij})$ denote either $M \in \mathcal{M}_{n,d}$ (when conditioning on $\mathcal{F}(\mathcal{I}, V)$) or M^T for $M \in \mathcal{M}_{n,d}$ (when conditioning on $\mathcal{G}(\mathcal{I}, V)$). Given a class $H \in \mathcal{H}^M$ or $H \in \mathcal{H}^{M^T}$ (in particular, $V, \mathcal{I}, i_0, K$ are fixed), and for each $k \in K$, define the random variable ξ_k on H by
$$\xi_k = \xi_k(A) := \sum_{j \in J_k} A_{i_0 j} x_j.$$

In words, ξ_k represents the inner product of x with the restriction of the i_0-th row of A to J_k. Later we will use this construction in the case $i_0 \in I_k$ for all $k \in K$, so that $|\mathrm{Supp}(R_{i_0}(A)) \cap J_k| = 1$ and each ξ_k selects exactly one coordinate x_j from the block $[n] \times J_k$. As we have already mentioned, by construction, for matrices $A \in H$, the rows $s < i_0$ are fixed, the entries in columns J^* are fixed, and each row's sum over J_k for $k \in K$ is prescribed. Thus, the blocks $[n] \times J_k$ for $k \in K$ are independent on H, and the random variables ξ_k, $k \in K$, are independent.

Define also

$$\xi' = \xi'(A) := y_1 \sum_{k\in K} \sum_{j \in J_k} A_{i_0 j} + \sum_{k \in K_0} \sum_{j\in J_k} A_{i_0 j}(x_j + y_j) + \sum_{j=1}^{n} w_{i_0 j}(x_j + y_j).$$

By definition of the equivalence classes, ξ' is constant on H. Therefore,

$$\langle R_{i_0}(A+W), (x+y)^{\dagger} \rangle = \sum_{k\in K} \xi_k + \xi'.$$

Suppose there exists $\alpha > 0$ such that $\mathcal{L}(\xi_k, 1/3) \le \alpha$ for every $k \in K$. Then, applying Lemma 6.3 with $t_0 = 1/3$, we obtain

$$\mathbb{P}\left(\left|\langle R_{i_0}(A+W), (x+y)^{\dagger}\rangle\right| \le 1/3\right) \le \frac{C_{6.3}}{\sqrt{(1-\alpha)|K|}}. \tag{6.4}$$

Recall that $\Omega_{k,\varepsilon} := \Omega^{in}_{k,\varepsilon} \cap \Omega^{out}_{k,\varepsilon}$ was defined in (5.1) and the event $\mathcal{E}_{2.2}(\tau)$ in Lemma 2.2.

Lemma 6.11. *There exist absolute constants $C_1, C_2, C_3, c_{6.11} > 0$ such that the following holds. Let $C_3 \log n \le d \le n/2$, and let $\varepsilon \in [\varepsilon_0, 0.01]$. Set $m_0 = \max\{1, \lfloor \varepsilon n/2d \rfloor\}$, and let m_1, m_2 be such that*

$$1 \le m_1 \le m_2 \le n - m_1.$$

Assume that $x \in \mathbb{C}^n$ satisfies

$$\forall i > m_2 : \; x^*_{m_1} > 2/3 + x^*_i.$$

Let $y \in$ span$\{\mathbf{1}\}$ and $z \in \mathbb{C}$. Let $m = \min\{m_0, m_1\}$ and $\mathcal{E}_{2.2} = \mathcal{E}_{2.2}(\varepsilon)$. Consider the event

$$\mathcal{E}'_{6.11}(x,y,A,k) = \left(\mathcal{E}_{6.11}(x,y,M) \cap \Omega_{2k,\varepsilon} \cap \mathcal{E}_{2.2}\right) \cup \left(\mathcal{E}_{6.11}(x,y,M^T) \cap \Omega_{2k,\varepsilon} \cap \mathcal{E}_{2.2}\right),$$

where $k = m_0$ or $k = m_1$ and for $A = M$ or $A = M^T$,

$$\mathcal{E}_{6.11}(x,y,A) := \left\{ M \in \mathcal{M}_{n,d} : \; \|(A - z\mathbf{I}_n)(x+y)\|_2 \le \sqrt{c_{6.11} m d} \right\}.$$

Then

1. *in the case $m_1 \le m_0$, one has* $\mathbb{P}\left(\mathcal{E}'_{6.11}(x,y,A,m_1)\right) \le (5/6)^{m_1 d/2}$;
2. *in the case $m_1 > C_1 m_0$ and $\varepsilon = 0.01$, one has* $\mathbb{P}\left(\mathcal{E}'_{6.11}(x,y,A,m_0)\right) \le \left(\frac{C_2 n}{m_1 d}\right)^{m_0 d/4}$.

Remark 6.12. We apply this lemma for $\mathcal{T}_i$ with the following choice of parameters. For $i = 2$, we set $m_1 = n_1 < m_0$, $m_2 = n_2$, $\varepsilon = 0.01$, obtaining

$$\mathbb{P}\left((\mathcal{E}_{6.11}(x,y,M) \cap \Omega_{2n_1,0.01} \cap \mathcal{E}_{2.2}) \bigcup (\mathcal{E}_{6.11}(x,y,M^T) \cap \Omega_{2n_1,0.01} \cap \mathcal{E}_{2.2}) \right) \le (5/6)^{n_1 d/2}.$$

For $i = 3$, we set $m_1 = n_2 > m_0$, $m_2 = n_3$, $\varepsilon = 0.01$, obtaining

$$\mathbb{P}\Big((\mathcal{E}_{6.11}(x,y,M) \cap \Omega_{2m_0,0.01} \cap \mathcal{E}_{2.2} \bigcup (\mathcal{E}_{6.11}(x,y,M^T) \cap \Omega_{2m_0,0.01} \cap \mathcal{E}_{2.2}) \Big)$$
$$\le \left(\frac{C_2 n}{n_2 d}\right)^{0.01 n/8} \le \left(\frac{a_2 d^{1/5}}{C_2}\right)^{-n/800}.$$

Recall that given two disjoint subsets $J^\ell, J^r \subset [n]$ and a matrix $M \in \mathcal{M}_{n,d}$,

$$I^\ell = I^\ell(J^\ell, J^r, M) = \left\{i \in [n] : \left|\mathrm{Supp}(R_i) \cap J^\ell\right| = 1 \text{ and } \mathrm{Supp}(R_i) \cap J^r = \emptyset\right\}$$

and

$$I^r = I^r(J^r, J^\ell, M) = \left\{i \in [n] : |\mathrm{Supp}(R_i) \cap J^r| = 1 \text{ and } \mathrm{Supp}(R_i) \cap J^\ell = \emptyset\right\}.$$

Proof. Fix $x \in \mathbb{C}^n$ and $y \in \mathrm{span}\{\mathbf{1}\}$ satisfying the condition of the lemma. Let $\sigma = \sigma_x$ be defined as above. Denote $q_0 = m_1/m$ and without loss of generality, assume that either $q_0 = 1$ or that q_0 is a large enough integer. Let $J_1^\ell, \ldots, J_{q_0}^\ell$ be a partition of $\sigma[m_1]$ into sets of cardinality m each, and let $J_1^r, \ldots, J_{q_0}^r$ be a partition of $\sigma([n - m_1 + 1, n])$ into sets of cardinality m each. Define for $q \in [q_0]$

$$J_q := J_q^\ell \cup J_q^r \quad \text{ and } \quad J_0 := [n] \setminus \bigcup_{q=1}^{q_0} J_q.$$

Then $J_0, J_1, \ldots, J_{q_0}$ forms a disjoint partition of $[n]$.

We deal simultaneously with the cases of M and M^T, so let A denote either M or M^T. All definitions and notations involving A apply to both cases. Let

$$I_q^\ell(A) = I_q^\ell(J_q^\ell, J_q^r, A) \quad \text{ and } \quad I_q^r(A) = I_q^r(J_q^r, J_q^\ell, A)$$

be defined as above, and for every $1 \le q \le q_0$ denote

$$I_q(A) = I_q(J_q, A) := I_q^\ell(A) \cup I_q^r(A).$$

Note that by definition, $I_q^\ell(A) \cap I_q^r(A) = \emptyset$ for each $1 \le q \le q_0$. Since the cardinality of J_q is $2m \le 2m_0 \le \varepsilon n/d$, for $M \in \mathcal{E}_{2.2} \cap \Omega_{2m,\varepsilon}$, by Lemma 5.7, we observe

$$|I_q^\ell(M)|,\ |I_q^r(M)| \in [(1 - 8\varepsilon)dm, (1 + \varepsilon)dm],$$

and

$$|I_q^\ell(M^T)|,\ |I_q^r(M^T)| \in [(1 - 4\varepsilon)dm, dm].$$

Therefore, as we deal with disjoint unions,

$$|I_q(M)| \in [2(1 - 8\varepsilon)dm, 2(1 + \varepsilon)dm], \tag{6.5}$$

and

$$|I_q(M^T)| \in [2(1 - 4\varepsilon)dm, 2dm]. \tag{6.6}$$

Split $\mathcal{M}_{n,d}$ into disjoint union of classes $\mathcal{F}(\mathcal{I}, V)$ defined at the beginning of this section with the fixed disjoint partition $J_0, J_1, \ldots, J_{q_0}$ chosen earlier. Let

$$\mathcal{F} := \{\mathcal{F}(\mathcal{I}, V) : \Omega_{2m,\varepsilon} \cap \mathcal{E}_{2.2} \cap \mathcal{F}(\mathcal{I}, V) \neq \emptyset\}.$$

Since $\mathcal{E}_{2.2} \cap \Omega_{2m,\varepsilon} \cap \mathcal{F}(\mathcal{I}, V) \neq \emptyset$ implies that $|I_q|$ satisfies (6.5) for each $1 \le q \le q_0$, we have

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}_{6.11}(x, y, M) \cap \Omega_{2m,\varepsilon} \cap \mathcal{E}_{2.2}) \\
&\qquad = \sum_{\mathcal{F}(\mathcal{I},V) \in \mathcal{F}} \mathbb{P}\left(\mathcal{E}_{6.11}(x, y, M) \cap \Omega_{2m,\varepsilon} \cap \mathcal{E}_{2.2} \mid \mathcal{F}(\mathcal{I}, V)\right) \mathbb{P}\left(\mathcal{F}(\mathcal{I}, V)\right) \\
&\qquad \le \max_{\mathcal{F}(\mathcal{I},V)} \mathbb{P}\left(\mathcal{E}_{6.11}(x, y, M) \cap \Omega_{2m,\varepsilon} \cap \mathcal{E}_{2.2} \mid \mathcal{F}(\mathcal{I}, V)\right) \\
&\qquad \le \max_{\mathcal{F}(\mathcal{I},V)} \mathbb{P}\left(\mathcal{E}_{6.11}(x, y, M) \mid \mathcal{F}(\mathcal{I}, V)\right),
\end{aligned}$$

where the maximum is taking over $\mathcal{F}(\mathcal{I}, V) \in \mathcal{F}$ with I_q's satisfying (6.5) for $1 \le q \le q_0$.

Similarly, we may also split $\mathcal{M}_{n,d}$ into the disjoint union of classes $\mathcal{G}(\mathcal{I}, V)$ defined at the beginning of this section with the fixed disjoint partition $J_0, J_1, \ldots, J_{q_0}$, and deduce that

$$\mathbb{P}\left(\mathcal{E}_{6.11}(x,y,M^T) \cap \Omega_{2m,\varepsilon} \cap \mathcal{E}_{2.2}\right) \leq \max_{\mathcal{G}(\mathcal{I},V)} \mathbb{P}\left(\mathcal{E}_{6.11}(x,y,M^T) \mid \mathcal{G}(\mathcal{I},V)\right), \tag{6.7}$$

where the maximum is taking over $\mathcal{G}(\mathcal{I}, V)$ with I_q's satisfying (6.6) for $1 \leq q \leq q_0$.

Thus, to prove our lemma, it suffices to prove a uniform upper bound for the conditional probabilities. Since we partition $\mathcal{M}_{n,d}$ using two different types of equivalent classes, to avoid potential confusion, we will treat those conditional probabilities separately, although the proofs are very similar. We start by fixing an admissible class $\mathcal{F}(\mathcal{I}_1, V_1)$ such that for $1 \leq q \leq q_0$, $I_q \in \mathcal{I}_1$ satisfies (6.5). Denote the corresponding induced probability measure:

$$\mathbb{P}_{\mathcal{F}}(\cdot) := \mathbb{P}(\cdot \mid \mathcal{F}(\mathcal{I}_1, V_1))$$

Let

$$I := \bigcup_{q=1}^{q_0} I_q.$$

We first show that the set of i's belonging to many I_q's is large. More precisely, given $i \in [n]$, denote

$$B_i := \{q \in [q_0] : i \in I_q\} \qquad \text{and} \qquad I_0 := \left\{i \in [n] : |B_i| \geq \frac{3mdq_0}{5n}\right\}.$$

By (6.5) we have

$$2(1-8\varepsilon)dmq_0 \leq \sum_{q=1}^{q_0} |I_q| = \sum_{i=1}^{n} |B_i| = \sum_{i \in I_0} |B_i| + \sum_{i \notin I_0} |B_i| \leq |I_0| q_0 + n \cdot \frac{3mdq_0}{5n}.$$

Thus, one has $|I_0| \geq 2md/3$. Without loss of generality, assume that $I_0 = [I_0] = \{1, \ldots, |I_0|\}$ and we only consider the first $k := \lceil 2md/3 \rceil$ indices from I_0. For every $M \in \mathcal{E}_{6.11}(x, y, M)$ we have

$$\|(M - z\mathbf{I}_n)(x+y)\|_2^2 = \sum_{i=1}^{n} \left|\left\langle R_i(M - z\mathbf{I}_n), (x+y)^\dagger \right\rangle\right|^2 \leq c_{6.11} md.$$

Thus, there are at most $9c_{6.11}md$ rows with $\left|\left\langle R_i(M - z\mathbf{I}_n), (x+y)^\dagger\right\rangle\right| \geq 1/3$. Hence,

$$\left|\left\{ i \in [k] : \left|\left\langle R_i(M - z\mathbf{I}_n), (x+y)^\dagger\right\rangle\right| < \frac{1}{3}\right\}\right| \geq \left(\frac{2}{3} - 9c_{6.11}\right) md.$$

Let $k_0 = \lceil \left(\frac{2}{3} - 9c_{6.11}\right) md \rceil$ ($k_0 \geq md/2$ for small enough $c_{6.11}$). For every $i \leq k$, we denote

$$\Omega_i := \left\{M \in \mathcal{F}(\mathcal{I}_1, V_1) : \left|\left\langle R_i(M - z\mathbf{I}_n), (x+y)^\dagger\right\rangle\right| < 1/3\right\} \quad \text{and} \quad \Omega_0 := \mathcal{F}(\mathcal{I}_1, V_1). \tag{6.8}$$

Then

$$\mathbb{P}_{\mathcal{F}}(\mathcal{E}_{6.11}(x,y,M)) \leq \sum_{\substack{B \subset [k] \\ |B| = k_0}} \mathbb{P}_{\mathcal{F}}\left(\bigcap_{i \in B} \Omega_i\right) \leq \binom{k}{k_0} \max_{\substack{B \subset [k] \\ |B| = k_0}} \mathbb{P}_{\mathcal{F}}\left(\bigcap_{i \in B} \Omega_i\right).$$

Without loss of generality, we assume that the maximum above is attained at $B = [k_0]$. Then

$$\begin{aligned} \mathbb{P}_{\mathcal{F}}(\mathcal{E}_{6.11}(x,y,M)) &\leq \left(\frac{ek}{k-k_0}\right)^{k-k_0} \mathbb{P}_{\mathcal{F}}\left(\bigcap_{i \in [k_0]} \Omega_i\right) \leq \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \mathbb{P}_{\mathcal{F}}\left(\bigcap_{i \in [k_0]} \Omega_i\right) \\ &= \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \prod_{i=1}^{k_0} \mathbb{P}_{\mathcal{F}}\left(\Omega_i | \Omega_1 \cap \cdots \cap \Omega_{i-1}\right). \end{aligned}$$

Now we estimate the factors in the product. Fix $i \in [k_0]$ and B_i. As $i \in I_0$, $|B_i| \geq \frac{3mdq_0}{5n}$. Consider the partition of $\mathcal{F}(\mathcal{I}_1, V_1)$ into equivalent classes $H \in \mathcal{H}(\mathcal{F}(\mathcal{I}_1, V_1), i, B_i)$ (such classes were defined at the beginning of this section). Let $\mathbb{P}_H$ denote the conditional probability measure $\mathbb{P}(\cdot|H)$ on a class H. Since all matrices in H have their first $i-1$ rows fixed, for every H, the intersection $H_i := H \cap \Omega_1 \cap \cdots \cap \Omega_{i-1}$ is either H or $\emptyset$. Thus,

$$\mathbb{P}_{\mathcal{F}}\left(\Omega_i | \Omega_1 \cap \cdots \cap \Omega_{i-1}\right) \leq \max_{H: H_i \neq \emptyset} \mathbb{P}_H(\Omega_i).$$

To bound the probability from above, we define a random variable $\xi_q(M, i), q \in B_i$ by

$$\xi_q(M, i) := \sum_{j \in J_q} M_{ij} x_j.$$

Fix H such that $H_i \neq \emptyset$. For any matrix in the equivalent class H, by the definition, every block $[n] \times J_q$ has a prescribed sum in each row. Recall that

$$J^*(B_i) := J_0 \cup \bigcup_{k \in [q_0] \setminus B_i} J_k.$$

Thus, conditional on H, random variables $\xi_q(M, i)$'s are independent for $q \in B_i$. Defining

$$\alpha := \max_{q \in B_i} \mathcal{L}_H(\xi_q(M, i), 1/3),$$

by (6.4) we obtain

$$\mathbb{P}_H(\Omega_i) = \mathbb{P}_H\left(\left|\left\langle R_i(M - z\mathbf{I}_n), (x+y)^\dagger \right\rangle\right| \leq \frac{1}{3}\right) \leq \frac{C_{6.3}}{\sqrt{(1-\alpha)|B_i|}} \leq \frac{C_{6.3}}{\sqrt{(1-\alpha)3mdq_0/(5n)}}.$$

Moreover, in the case $q_0 = 1$ we simply have

$$\mathbb{P}_H(\Omega_i) = \alpha = \mathcal{L}_H(\xi_1(M, i), 1/3).$$

Now it remains to bound α. Fix $q \in B_i$ such that $i \in I_q$. Since the intersection of the support of $R_i(M)$ with J_q is a singleton, say indexed by

$$j_r(q) = j_r(q, M, i) \in J_q^r \qquad \text{or} \qquad j_\ell(q) = j_\ell(q, M, i) \in J_q^\ell,$$

we have

$$|\xi_q(M, i)| = \begin{cases} |x_{j_r(q)}|, \\ |x_{j_\ell(q)}|. \end{cases}$$

Since each row of M is independently and uniformly sampled from the vectors in $\{0,1\}^n$ containing d ones, we observe that conditioned on H, the single 1 appears in either J_q^r or J_q^ℓ with equal probability, i.e., $\mathbb{P}_H\left(j_r(q) \in J_q^r\right) = \mathbb{P}_H\left(j_\ell(q) \in J_q^\ell\right) = 1/2$. Moreover, since $\left|x_{j_\ell(q)}\right| - \left|x_{j_r(q)}\right| > 2/3$, we have $\alpha \leq 1/2$. Therefore, we have in the case $q_0 = 1$,

$$\mathbb{P}_{\mathcal{F}}(\mathcal{E}_{6.11}(x, y, M)) \leq \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \left(\frac{1}{2}\right)^{(2/3 - 9c_{6.11})md}. \tag{6.9}$$

In the case $q_0 = m_1/m > C_1$ we obtain,

$$\mathbb{P}_{\mathcal{F}}(\mathcal{E}_{6.11}(x, y, M)) \leq \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \left(\frac{C_{6.3}\sqrt{10n}}{\sqrt{3m_1 d}}\right)^{(2/3 - 9c_{6.11})md}. \tag{6.10}$$

Now we turn to bounding the conditional probability in (6.7). Fix an admissible class $\mathcal{G}(\mathcal{I}_2, V_2)$ such that for $1 \le q \le q_0$, $I_q \in \mathcal{I}_2$ satisfies (6.6). Denote the corresponding induced probability measure by

$$\mathbb{P}_{\mathcal{G}}(\cdot) := \mathbb{P}(\cdot \mid \mathcal{G}(\mathcal{I}_2, V_2)).$$

Define the same sets I, B_i, and I_0 as before. By (6.6), we can again deduce that $|I_0| \ge 2md/3$. Now let

$$\Omega_i' := \left\{M \in \mathcal{G}(\mathcal{I}_2, V_2) : \left|\left\langle R_i(M^T - z\mathbf{I}_n), (x+y)^\dagger \right\rangle\right| < 1/3\right\} \quad \text{and} \quad \Omega_0' := \mathcal{G}(\mathcal{I}_2, V_2). \quad (6.11)$$

Repeating the above proof, we obtain

$$\begin{aligned}\mathbb{P}_{\mathcal{G}}(\mathcal{E}_{6.11}(x, y, M^T)) &\le \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \prod_{i=1}^{k_0} \mathbb{P}_{\mathcal{G}}\left(\Omega_i' | \Omega_1' \cap \cdots \cap \Omega_{i-1}'\right) \\ &\le \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \left(\max_{H: H_i \ne \emptyset} \mathbb{P}_H(\Omega_i')\right)^{k_0},\end{aligned}$$

where $H \in \mathcal{H}(\mathcal{G}(\mathcal{I}_2, V_2), i, B_i)$ and $H_i := H \cap \Omega_1' \cap \cdots \cap \Omega_{i-1}'$. To estimate the maximum probability, we again fix an H such that $H_i \ne \emptyset$ and define

$$\xi_q(M^T, i) := \sum_{j \in J_q} M_{ij}^T x_j.$$

Repeating the above proof, we have

$$\mathbb{P}_H(\Omega_i') = \mathbb{P}_H\left(\left|\left\langle R_i(M^T - z\mathbf{I}_n), (x+y)^\dagger\right\rangle\right| \le \frac{1}{3}\right) \le \frac{C_{6.3}}{\sqrt{(1-\alpha')3mdq_0/(5n)}},$$

where $\alpha' := \max_{q \in B_i} \mathcal{L}_H(\xi_q(M^T, i), 1/3)$. And in the case $q_0 = 1$,

$$\mathbb{P}_H(\Omega_i') = \alpha' = \mathcal{L}_H(\xi_1(M^T, i), 1/3).$$

To bound from above α', we follow our previous proof by indexing the single 1 by

$$j_r(q) = j_r(q, M^T, i) \in J_q^r \qquad \text{or} \qquad j_\ell(q) = j_\ell(q, M^T, i) \in J_q^\ell,$$

and defining

$$|\xi_q(M^T, i)| = \begin{cases} |x_{j_r(q)}|, \\ |x_{j_\ell(q)}|. \end{cases}$$

Since $\left|x_{j_\ell(q)}\right| - \left|x_{j_r(q)}\right| > 2/3$, we have

$$\mathcal{L}_H(\xi_q(M^T, i), 1/3) \le \max(\mathbb{P}_H(H_l), \mathbb{P}_H(H_r))$$

where $H_\ell := \{M \in H : j_\ell(q) \in J_q^\ell\}$ and $H_r := \{M \in H : j_r(q) \in J_q^r\}$. Since the rows of M^T are not independent, to evaluate the probabilities, one needs to use a "simple switching" argument, the details of which can be found in [34, Claim 4.8]. Applying the claim, we get for $i \in [k_0]$,

$$\mathcal{L}_H(\xi_q(M^T, i), 1/3) \le \max(\mathbb{P}_H(H_l), \mathbb{P}_H(H_r)) \le 4/5,$$

which implies that $\alpha' \le 4/5$. Hence, we have in the case $q_0 = 1$,

$$\mathbb{P}_{\mathcal{G}}(\mathcal{E}_{6.11}(x, y, M^T)) \le \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \left(\frac{4}{5}\right)^{(2/3 - 9c_{6.11})md}. \quad (6.12)$$

In the case $q_0 = m_1/m > C_1$ we obtain,

$$\mathbb{P}_{\mathcal{G}}(\mathcal{E}_{6.11}(x, y, M^T)) \le \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \left(\frac{5C_{6.3}\sqrt{n}}{\sqrt{3m_1 d}}\right)^{(2/3-9c_{6.11})md}. \tag{6.13}$$

Combining (6.9) and (6.12), in the case $q_0 = 1$ we obtain that

$$\mathbb{P}_{\mathcal{F}}(\mathcal{E}_{6.11}(x, y, M)) + \mathbb{P}_{\mathcal{G}}(\mathcal{E}_{6.11}(x, y, M^T)) \le \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \left(\left(\frac{4}{5}\right)^{(2/3-9c_{6.11})md} + \left(\frac{1}{2}\right)^{(2/3-9c_{6.11})md}\right) \le \left(\frac{5}{6}\right)^{md/2},$$

provided that $c_{6.11}$ is small enough. In the case $q_0 = m_1/m > C_1$, by combining (6.10) and (6.13), we obtain

$$\max\{\mathbb{P}_{\mathcal{F}}(\mathcal{E}_{6.11}(x, y, M)), \mathbb{P}_{\mathcal{G}}(\mathcal{E}_{6.11}(x, y, M^T))\} \le \left(\frac{1}{c_{6.11}}\right)^{9c_{6.11}md} \left(\frac{5C_{6.3}\sqrt{n}}{\sqrt{3m_1 d}}\right)^{(2/3-9c_{6.11})md} \le \frac{1}{2}\left(\frac{C_2 n}{m_1 d}\right)^{md/4},$$

provided $c_{6.11}$ is small enough and C_1, C_2 are large enough. This completes the proof. □

6.5 Proof of Theorem 6.6

Set $\varepsilon := 0.01$ and $m_0 := \max\{1, \lfloor \varepsilon n/(2d) \rfloor\} = \max\{1, \lfloor n/(200d) \rfloor\}$. By Lemma 6.9, we have

$$\mathbb{P}\left(\mathcal{E}_{6.9}(M) \cup \mathcal{E}_{6.9}(M^T)\right) \le \frac{7e}{n}. \tag{6.14}$$

Thus it is enough to consider vectors from $\mathcal{T}_2 \cup (\mathcal{T}_3 \cap \mathrm{Cons}(a_3, \rho))$. Recall that $B(\mathcal{T}_2)$ and $B(\mathcal{T}_3)$ were defined in (6.1) and (6.2). For $j \in \{2, 3\}$ denote

$$\mathcal{E}_j(M) =: \left\{M \in \mathcal{M}_{n,d} : \exists x \in \mathcal{T}_j \text{ such that } \|(M - z\mathbf{I}_n)x\|_2 \le \frac{\sqrt{c_{6.11}md}}{2B(\mathcal{T}_j)}\|x\|_2\right\}$$

and

$$\mathcal{E}_j(M^T) =: \left\{M \in \mathcal{M}_{n,d} : \exists x \in \mathcal{T}_j \text{ such that } \|(M^T - z\mathbf{I}_n)x\|_2 \le \frac{\sqrt{c_{6.11}md}}{2B(\mathcal{T}_j)}\|x\|_2\right\},$$

where

$$m = \begin{cases} n_1, & \text{when } j = 2, \\ m_0, & \text{when } j = 3. \end{cases}$$

We prove Theorem 6.6 by bounding from above the probability of the event

$$\mathcal{E} := \mathcal{E}_{6.9}(M) \cup \mathcal{E}_{6.9}(M^T) \cup \mathcal{E}_2(M) \cup \mathcal{E}_2(M^T) \cup \mathcal{E}_3(M) \cup \mathcal{E}_3(M^T).$$

Note that in the case $n_1 > 1$ bounds on the ratios of norms

$$\|(M - z\mathbf{I}_n)x\|_2/\|x\|_2 \qquad \text{and} \qquad \|(M^T - z\mathbf{I}_n)x\|_2/\|x\|_2$$

coming from $\mathcal{E}$ is the minimum between $\omega(n, d)$,

$$\frac{13(\log n)^{1/4}\sqrt{c_{6.11}n_1 d}}{(6d)^{r+1}\, d^{1/4}} \ge \frac{c_0\sqrt{n\log n}}{(6d)^{r+1}\sqrt{d}}, \quad \text{and} \quad \frac{13(\log n)^{1/4}\sqrt{c_{6.11}m_0\, d}}{(6d)^{r+1}\, d} \ge \frac{54(\log n)^{1/4}\sqrt{n}}{(6d)^{r+2}},$$

where $c_0 > 0$ is an absolute constant. In the case $n_1 = 1$ we have $\mathcal{T}_1 = \emptyset$, so we may assume $\mathcal{E}_{6.9}(M) = \mathcal{E}_{6.9}(M^T) = \emptyset$, therefore bounds on the ratios of norms above coming from $\mathcal{E}$ is the minimum between

$$\frac{\sqrt{c_{6.11} n_1 d}}{2\sqrt{2n}} \geq \frac{c_1 (\log n)^{1/4}}{d^{1/4}} \qquad \text{and} \qquad \frac{\sqrt{c_{6.11} m_0\, d}}{2\sqrt{2n}} \geq c_1,$$

where $c_0 > 1$ is an absolute constant. This leads to the bound announced in the Theorem 6.6. Thus, it remains to estimate probabilities.

Fix $j \in \{2, 3\}$. Let A be either M or M^T. Define the corresponding "norm control" event

$$\mathcal{E}_{\mathrm{nrm}}(A) := \begin{cases} \mathcal{E}_{5.3}, & \text{when } \ A = M, \\ \mathcal{E}_{5.4}, & \text{when } \ A = M^T. \end{cases}$$

Assume $M \in \mathcal{E}_j(A) \cap \mathcal{E}_{\mathrm{nrm}}(A)$. Then there exists $x = x(A)$ (which we fix now) such that

$$x \in \begin{cases} \mathcal{T}_2, & \text{when } \ j = 2, \\ \mathcal{T}_3 \cap \mathrm{Cons}(a_3, \rho), & \text{when } \ j = 3 \end{cases} \qquad \text{and} \qquad \|(A - z\mathbf{I}_n)x\|_2 \leq \tfrac{\sqrt{c_{6.11} md}}{2B(\mathcal{T}_j)} \|x\|_2.$$

Without loss of generality we assume that $x^*_{n_{j-1}} = 1$, which means that $x \in \mathcal{T}'_j$ (recall that $\mathcal{T}'_j$ is defined in Section 6.4.1 for $j = 2, 3$). By Lemma 6.5, we have $\|x\|_2 \leq B(\mathcal{T}_j)$ for $j = 2, 3$. Thus,

$$\|(A - z\mathbf{I}_n)x\|_2 \leq \frac{\sqrt{c_{6.11} md}}{2}.$$

Let $\mathcal{N}_j := \mathcal{N}'_j + \mathcal{N}''_j$ be the net constructed in Lemma 6.10. (Here and below, for fixed j, we write $u \in \mathcal{N}'_j$ and $w \in \mathcal{N}''_j$. When j changes, we reuse the notation.) For the $x = x(A) \in \mathcal{T}'_j$ fixed above, there exist $u \in \mathcal{N}'_j$ and $w \in \mathcal{N}''_j \subset \mathrm{span}\{\mathbf{1}\}$ such that

$$u^*_{n_{j-1}} \geq 1 - \frac{1}{d^{3/4}}, \quad u^*_\ell \leq \frac{1}{4} + \frac{1}{d^{3/4}} \quad \forall \ell \geq n_j, \quad |||x - (u + w)||| \leq \frac{\sqrt{2n}}{d^{3/4}}. \tag{6.15}$$

Moreover, for any $\ell \geq n_j$ and d large enough,

$$u^*_{n_{j-1}} - u^*_\ell \geq \Big(1 - \frac{1}{d^{3/4}}\Big) - \Big(\frac{1}{4} + \frac{1}{d^{3/4}}\Big) = \frac{3}{4} - \frac{2}{d^{3/4}} > \frac{2}{3}.$$

Hence, $u^*_{n_{j-1}} > 2/3 + u^*_\ell$ for all $\ell \geq n_j$.

Thus, for either $A = M$ or $A = M^T$, there exist $u = u(A) \in \mathcal{N}'_j$ and $w = w(A) \in \mathcal{N}''_j$ such that

$$\begin{aligned} \|(A - z\mathbf{I}_n)(u + w)\|_2 &\leq \|(A - z\mathbf{I})x\|_2 + \|(A - z\mathbf{I}_n)(x - u - w)\|_2 \\ &\leq \frac{\sqrt{c_{6.11} md}}{2} + \|A(x - u - w)\|_2 + |z|\|x - u - w\|_2. \end{aligned}$$

Applying Proposition 5.3 for $A = M$ and Proposition 5.4 for $A = M^T$, together with (6.15),

$$\|A(x - u - w)\|_2 \leq (2C_{4.1} + 1)\sqrt{d} \cdot |||x - u - w||| \leq (2C_{4.1} + 1)\,\frac{\sqrt{2n}}{d^{1/4}}.$$

Since $\|x - u - w\|_2 \leq |||x - u - w||| \leq \sqrt{2n}/d^{3/4}$ and $|z| \leq \sqrt{d}\log\log d$,

$$|z|\|x - u - w\|_2 \leq \frac{\sqrt{2n}\log\log d}{d^{1/4}}.$$

Note that $c_{6.11}md \geq cn\sqrt{(\log n)/d}$ for small enough absolute constant $c > 0$. Therefore, for large enough n,

$$\|(A - z\mathbf{I}_n)(u+w)\|_2 \leq \frac{\sqrt{c_{6.11}md}}{2} + (2C_{4.1}+1)\,\frac{\sqrt{2n}}{d^{1/4}} + \frac{\sqrt{2n}\log\log d}{d^{1/4}} \leq \sqrt{c_{6.11}md}.$$

Next we apply Lemma 6.10 and Lemma 6.11 twice with the choice of parameters as in Remark 6.12.

In the case $j = 2$, by Lemma 6.10, $|\mathcal{N}_2| \leq \exp(6n_2 \log d)$, and by Lemma 6.11 (with $m_1 = n_1 < m_0, m_2 = n_2$, $\varepsilon = 0.01$ — see Remark 6.12), for each fixed $u \in \mathcal{N}_2'$ and $w \in \mathcal{N}_2''$,

$$\mathbb{P}\Big(\big(\mathcal{E}_{6.11}(u,w,M) \cap \Omega_{2n_1,0.01} \cap \mathcal{E}_{2.2}\big) \cup \big(\mathcal{E}_{6.11}(u,w,M^T) \cap \Omega_{2n_1,0.01} \cap \mathcal{E}_{2.2}\big)\Big) \leq \Big(\frac{5}{6}\Big)^{n_1 d/2}.$$

Hence, by the union bound over the net $\mathcal{N}_2$, the probabilities $\mathbb{P}(\mathcal{E}_2(M) \cap \Omega_{2n_1,0.01} \cap \mathcal{E}_{5.3} \cap \mathcal{E}_{2.2})$ and $\mathbb{P}(\mathcal{E}_2(M^T) \cap \Omega_{2n_1,0.01} \cap \mathcal{E}_{5.4} \cap \mathcal{E}_{2.2})$ are both bounded by

$$e^{6n_2 \log d} \cdot (5/6)^{n_1 d/2} \leq e^{-n_1 d/20}$$

provided that d is large enough and a_2 is small enough. Since $n_1 = \lceil a_1 \varepsilon_0 n/d \rceil$ and $d \leq n/2$, one has

$$e^{-n_1 d/20} \leq \exp\Big(-\frac{a_1}{20}\varepsilon_0 n\Big) = \exp\Big(-\frac{a_1 n}{20}\sqrt{\frac{48\log n}{d}}\Big) \leq \exp\Big(-c\sqrt{n\log n}\Big)$$

for some absolute constant $c > 0$.

For $j = 3$, we argue similarly. By Lemma 6.10, $|\mathcal{N}_3| \leq \exp(6n_3 \log d)$, and by Lemma 6.11 (with $m_1 = n_2 > m_0$, $m_2 = n_3$, $\varepsilon = 0.01$ — see Remark 6.12), for each fixed $u \in \mathcal{N}_3'$ and $w \in \mathcal{N}_3''$, for some absolute constant $C_2 > 0$,

$$\mathbb{P}\Big(\big(\mathcal{E}_{6.11}(u,w,M) \cap \Omega_{2m_0,0.01} \cap \mathcal{E}_{2.2}\big) \bigcup \big(\mathcal{E}_{6.11}(u,w,M^T) \cap \Omega_{2m_0,0.01} \cap \mathcal{E}_{2.2}\big)\Big) \leq \left(\frac{a_2 d^{1/5}}{C_2}\right)^{-n/800}.$$

Thus, by the union bound over the net $\mathcal{N}_3$, the probabilities $\mathbb{P}(\mathcal{E}_3(M) \cap \Omega_{2m_0,0.01} \cap \mathcal{E}_{5.3} \cap \mathcal{E}_{2.2})$ and $\mathbb{P}(\mathcal{E}_3(M^T) \cap \Omega_{2m_0,0.01} \cap \mathcal{E}_{5.4} \cap \mathcal{E}_{2.2})$ are both bounded by

$$e^{6n_3 \log d} \cdot \left(\frac{a_2 d^{1/5}}{C_2}\right)^{-n/800} \leq e^{-c' n \log d},$$

for d is large enough and a_3 is sufficiently small, with an absolute constant $c' > 0$.

Finally, we bound the remaining probabilities. By Lemma 2.2 (with $\tau = 0.01, m = n$),

$$\mathbb{P}(\mathcal{E}_{2.2}^c) \leq n e^{-d/30000}.$$

Note that $\mathcal{E}_{5.3} \subset \mathcal{E}_{5.4}$. Indeed, if $\mathcal{E}_{5.3}$ occurs, then by the triangle inequality

$$\begin{aligned}\|M^T\mathbf{1}\|_2 &\leq \|(M^T - \mathbb{E}M^T)\mathbf{1}\|_2 + \|\mathbb{E}M^T\mathbf{1}\|_2 \leq \|(M^T - \mathbb{E}M^T)\| \cdot \|\mathbf{1}\|_2 + d\|\mathbf{1}\|_2 \\ &\leq (C_{4.1}\sqrt{d} + d)\sqrt{n} \leq (C_{4.1}+1)d\sqrt{n}.\end{aligned}$$

Thus, by Theorem 4.1, one has $\mathbb{P}(\mathcal{E}_{5.3}^c) = \mathbb{P}((\mathcal{E}_{5.3} \cup \mathcal{E}_{5.3})^c) \leq 2/n$.

By Lemma 5.5 and Lemma 5.6 (with $\varepsilon = 0.01$),

$$\mathbb{P}(\Omega_{2n_1,0.01}^c) + \mathbb{P}(\Omega_{2m_0,0.01}^c) \leq 10\, e^{-d/480000}.$$

Collecting the bounds for $j = 2, 3$ and adding the probability in (6.14), we obtain that the total probability is at most $22/n$ for $d \geq C\log n$ with large enough absolute constant C and large enough n. This completes the proof. □

7 Proof of Theorem 1.7

The infimum over the non-almost-constant vectors is tackled by associating it with the average distance of a row of the matrix $M - z\mathbf{I}_n$ from the subspace spanned by the rest of the rows. We adapt the result in [48, Lemma 3.5], where their original argument applies to incompressible vectors.

Lemma 7.1 (Invertibility via distance)**.** *Let A be a $n \times n$ random matrix. Let $R_1, \ldots, R_n$ be the row vectors of A, and let $H_k = span\{R_i : i \neq k\}$ be the span of all row vectors except the k-th row. Then for every $\delta, \rho \in (0,1)$ and every $\varepsilon \geq 0$, one has*

$$\mathbb{P}\left(\inf_{x \in \mathbb{S}^{2n-1} \cap (\mathrm{Cons}(\delta,\rho))^c} \|\bar{x}A\|_2 \leq \frac{\varepsilon\,\rho}{\sqrt{n}}\right) \leq \frac{1}{\delta n}\sum_{k=1}^{n} \mathbb{P}\left(\mathrm{dist}\left(R_k, H_k\right) \leq \varepsilon\right).$$

Proof. Let $x \in \mathbb{S}^{2n-1} \cap (\mathrm{Cons}(\delta,\rho))^c$. Let $X_k := R_k^T$ be the column vector corresponding R_k for $k \in [n]$. Note that for any vector u, and any linear subspace H, $\mathrm{dist}(u, H) \leq \|u\|_2$. Therefore,

$$\|\bar{x}A\|_2 \geq \max_{k=1,\ldots,n} \mathrm{dist}(\bar{x}A, H_k) = \max_{k=1,\ldots,n} \mathrm{dist}\Big(\sum_{i=1}^{n} \bar{x}_i X_i, H_k\Big),$$

where x_i is the i-th coordinate of the vector x. Note that by the definition $X_j \in H_k$ unless $j = k$. Therefore $\mathrm{dist}(X_j, H_k) = 0$ for every $j \neq k$. Thus,

$$\|\bar{x}A\|_2 \geq \max_{k=1,\ldots,n} \mathrm{dist}(\bar{x}_k X_k, H_k) = \max_{k=1,\ldots,n} |x_k|\,\mathrm{dist}(X_k, H_k).$$

Let $p_k := \mathbb{P}\left(\mathrm{dist}(X_k, H_k) \leq \varepsilon\right)$. Let $\sigma_1 := \{k \in [n] : \mathrm{dist}(X_k, H_k) > \varepsilon\}$, and denote by $\mathcal{E}$ the event that σ_1 contains more than $(1-\delta)n$ elements. By Markov's inequality,

$$\mathbb{P}(\mathcal{E}^c) = \mathbb{P}(|\sigma_1^c| \geq \delta n) \leq \frac{1}{\delta n}\mathbb{E}\left(\sum_{k=1}^{n} \mathbf{1}_{\{\mathrm{dist}(X_k,H_k) \leq \varepsilon\}}\right) \leq \frac{1}{\delta n}\sum_{k=1}^{n} p_k.$$

Let $\sigma_2(x) := \{k \in [n] : |x_k| \geq \rho/\sqrt{n}\}$. Note that $|\sigma_2^c(x)| \leq (1-\delta)n$ (otherwise,

$$|\{k \in [n] : |x_k - 0| \leq \rho/\sqrt{n}\}| \geq (1-\delta)n,$$

contradicts that x is a non-almost-constant vector). Thus, $|\sigma_2(x)| \geq \delta n$.

Now, suppose that $\mathcal{E}$ occurs. Then

$$|\sigma_1| + |\sigma_2(x)| > (1-\delta)n + \delta n = n,$$

which implies that $\sigma_1 \cap \sigma_2(x) \neq \emptyset$. Let $\ell \in \sigma_1 \cap \sigma_2(x)$. Then by the definitions of the sets σ_1 and $\sigma_2(x)$ one has

$$\|\bar{x}A\|_2 \geq |x_\ell|\,\mathrm{dist}(X_\ell, H_\ell) > \frac{\rho\,\varepsilon}{\sqrt{n}}.$$

Thus,

$$\mathbb{P}\left(\inf_{x \in \mathbb{S}^{2n-1} \cap (\mathrm{Cons}(\delta,\rho))^c} \|\bar{x}A\|_2 \leq \frac{\varepsilon\,\rho}{\sqrt{n}}\right) \leq \mathbb{P}(\mathcal{E}^c) \leq \frac{1}{\delta n}\sum_{k=1}^{n} p_k.$$

This completes the proof. ☐

To complete the picture, it remains to bound the probabilities appearing on the right-hand side in Lemma 7.1. Recall that $B_{\mathcal{T}}$ is defined in (6.1) and a_3 is a small enough positive constant introduced in Section 6 to define n_3.

Lemma 7.2. *Let $C_{6.8}\log n \le d \le n/2$. Let $M \in \mathcal{M}_{n,d}$. Let $z \in \mathbb{C}$. Let $R_1, \dots, R_n$ be the row vectors of $M - z\mathbf{I}_n$ and let $H_1 = span\{R_2, \dots, R_n\}$. Let R^* be a random unit vector orthogonal to H_1. Let $0 < \rho \le \sqrt{n}/(B_{\mathcal{T}} d^{3/4})$. Then $R^* \in \mathbb{S}^{2n-1} \cap (Cons(a_3, \rho))^c$ with probability at least $1 - 23/n$ and*

$$\mathbb{P}(\mathrm{dist}(R_1, H_1) \le \rho/\sqrt{n}) \le \frac{23}{n} + \frac{C_{6.4}}{\sqrt{a_3 d}} + C_{6.4} \exp\left(-a_3^2 d/C_{6.4}\right).$$

Proof. It is clear that R^* is a random vector that depends only on $R_2, \dots, R_n$ (hence independent of R_1) and that

$$\mathrm{dist}(R_1, H_1) \ge |\langle R^*, R_1\rangle|.$$

Let

$$\mathcal{E} := \{R^* \in \mathbb{S}^{2n-1} \cap (\mathrm{Cons}(a_3, \rho))^c\}.$$

Then

$$\mathbb{P}\Big(\mathrm{dist}(R_1, H_1) \le \frac{\rho}{\sqrt{n}}\Big) \le \mathbb{P}\left(|\langle R^*, R_1\rangle| \le \frac{\rho}{\sqrt{n}}\right) \le \mathbb{P}\left(\left\{|\langle R^*, R_1\rangle| \le \frac{\rho}{\sqrt{n}}\right\} \cap \mathcal{E}\right) + \mathbb{P}(\mathcal{E}^c)$$
$$= \mathbb{P}\Big(\, |\langle R^*, R_1\rangle| \le \frac{\rho}{\sqrt{n}} \mid \mathcal{E}\Big)\mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}^c).$$

Since R^* is independent of R_1, conditioning on $\mathcal{E}$, we have

$$\mathbb{P}\Big(\, |\langle R^*, R_1\rangle| \le \frac{\rho}{\sqrt{n}} \mid \mathcal{E}\Big) \le \mathcal{L}(\langle R^*, \xi\rangle, \frac{\rho}{\sqrt{n}}),$$

where $\xi := R_1 + ze_1$ is uniformly distributed on the set of all vectors in $\{0,1\}^n$ containing exactly d ones. Therefore, by Lemma 6.4,

$$\mathbb{P}\Big(\, |\langle R^*, R_1\rangle| \le \frac{\rho}{\sqrt{n}} \mid \mathcal{E}\Big) \le \frac{C_{6.4}}{\sqrt{a_3 d}} + C_{6.4} \exp\left(-a_3^2 d/C_{6.4}\right).$$

To bound $\mathbb{P}(\mathcal{E}^c)$, let M' be the $(n-1)\times n$ random matrix with row vectors $R_2, \dots, R_n$, i.e., the submatrix of $M - z\mathbf{I}_n$ obtained by removing the first row. By the definition of R^*,

$$M'R^* = 0.$$

Thus, on the event $\mathcal{E}^c$, there exists $x \in \mathbb{S}^{2n-1} \cap \mathrm{Cons}(a_3, \rho)$ such that $M'x = 0$. By replacing n with $n-1$, one can check that the proof of Corollary 6.8 remains valid for M' instead of $M - z\mathbf{I}_n$. Thus,

$$\mathbb{P}(\mathcal{E}^c) \le 23/n.$$

This completes the proof. □

We are now ready to prove the lower bound on the smallest singular value.

Proof of Theorem 1.7. We choose parameters

$$\rho = \sqrt{n}\,\min\left\{\frac{1}{B_{\mathcal{T}}\, d^{3/4}}, \sqrt{\theta(n,d)}\right\} \qquad \text{and} \qquad \varepsilon = \frac{\rho}{\sqrt{n}}$$

and set

$$A = M^T - \bar{z}\mathbf{I}_n, \;\; S_1 = \mathbb{S}^{2n-1} \cap \mathrm{Cons}(a_3, \rho), \quad \text{and} \quad S_2 = \mathbb{S}^{2n-1} \cap (\mathrm{Cons}(a_3, \rho))^c.$$

Consider the event

$$\mathcal{E} := \left\{\inf_{v \in \mathbb{S}^{2n-1}} \|Av\|_2 \le \frac{\varepsilon\,\rho}{\sqrt{n}}\right\} = \mathcal{E}_1 \cup \mathcal{E}_2,$$

where

$$\mathcal{E}_1 := \left\{ \inf_{x\in S_1} \|Ax\|_2 \le \frac{\varepsilon\,\rho}{\sqrt{n}} \right\} \qquad \text{and} \qquad \mathcal{E}_2 := \left\{ \inf_{x\in S_2} \|Ax\|_2 \le \frac{\varepsilon\,\rho}{\sqrt{n}} \right\}.$$

Since $\rho \le \sqrt{n}/(B_{\mathcal{T}} d^{3/4})$, $\varepsilon\rho/\sqrt{n} = \rho^2/n \le \theta(n,d)$, and $|\bar{z}| = |z| \le \sqrt{d}\log\log d$, Corollary 6.8 implies that

$$\mathbb{P}(\mathcal{E}_1) \le 23/n.$$

Let $H_k = \operatorname{span}\{R_i : i \ne k\}$, $k \le n$. Using that $\operatorname{dist}(R_k(A), H_k)$, $k \le n$, have the same distributions and applying Lemma 7.1 with $\delta = a_3$, we observe

$$\mathbb{P}(\mathcal{E}_2) \le \frac{1}{a_3}\mathbb{P}\left(\operatorname{dist}(R_1(A), H_1) \le \varepsilon\right).$$

Applying Lemma 7.2 (note again that $\rho \le \sqrt{n}/(B_{\mathcal{T}} d^{3/4})$)),

$$\mathbb{P}(\mathcal{E}) \le \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) \le \frac{23}{n} + \frac{1}{a_3}\left(\frac{23}{n} + \frac{C_{6.4}}{\sqrt{a_3 d}} + C_{6.4}\exp\left(-a_3^2 d/C_{6.4}\right)\right) \le \frac{C}{\sqrt{d}},$$

where C is a large enough absolute constant. This proves the bound on the singular value

$$\frac{\varepsilon\,\rho}{\sqrt{n}} = \frac{\rho^2}{n} = \min\left\{\frac{1}{B_{\mathcal{T}}^2\, d^{3/2}},\, \theta(n,d)\right\}$$

Finally we analyze this minimum.

First assume that $n_1 = 1$ (that is, $d \ge c\, n^{2/3}(\log n)^{1/3}$). In this case, for large n,

$$\min\left\{\frac{1}{B_{\mathcal{T}}^2\, d^{3/2}},\, \theta(n,d)\right\} = \min\left\{\frac{1}{2n\, d^{3/2}},\, \frac{c_1(\log n)^{1/4}}{d^{1/4}}\right\} = \frac{1}{2n\, d^{3/2}} \ge \frac{1}{n^{5/2}}.$$

Next assume that $n_1 > 1$ (that is, $d \le c\, n^{2/3}(\log n)^{1/3}$). It is not difficult to see that

$$\min\left\{\frac{1}{B_{\mathcal{T}}^2\, d^{3/2}},\, \theta(n,d)\right\} = \min\left\{\frac{6^4\sqrt{\log n}}{(6d)^{2r+4}\, d^{3/2}},\, \frac{\sqrt{d}}{4\sqrt{n}}\right\}. \tag{7.1}$$

Recall that $\ell_0^r < n_1 \le \ell_0^{r+1}$. Denote

$$\alpha := \frac{\log(6d)}{\log \ell_0} \in (2, C\log\log n]$$

for some absolute constant $C \ge 1$. Then

$$(6d)^{2r+4}\, d^{3/2} = \ell_0^{2\alpha(r+1)} d^{7/2} \ge n_1^{2\alpha} d^{7/2} \ge n_1^4 d^{7/2} \ge c_0 n^4 (\ln n)^2/d^{5/2} \ge c_0 n^{3/2},$$

for some absolute constant $c_0 > 0$. This shows that for large enough n, the minimum in (7.1) attains at the first term. On the other hand,

$$(6d)^{2r+4}\, d^{3/2} = \ell_0^{2\alpha r} d^{11/2} \le n_1^{2\alpha} d^{11/2} \le \left(n^2(\log n)/d^3\right)^{\alpha} d^{11/2},$$

Therefore the minimum in (7.1) is bounded below by

$$\frac{d^{3\alpha - 5.5}}{n^{2\alpha}(\log n)^{\alpha - 1/2}} \ge n^{-2\alpha}.$$

Finally note that
(1) if $d \ge C_1(\log n)^{21}$ then $\alpha \le 2.1$, so the bound is larger than

$$\frac{d^{0.8}}{n^{4.2}(\log n)^{1.6}};$$

(2) if $d \geq (\log n)^2$ then $\alpha \leq 4.1$ (for large enough n), so the bound is larger than

$$\frac{d^{6.8}}{n^{8.2}(\log n)^{3.6}};$$

This completes the proof of Theorem 1.7 and the bound in Remark 1.9. ☐

8 Singularity threshold

In [22], Ferber, Kwan, and Sauermann established that a random matrix uniformly drawn from $\mathcal{M}_{n,d}$ is asymptotically almost surely non-singular whenever $\min(d, n-d) \geq (1+\varepsilon)\log n$. In [4], Aigner–Horev and Person mentioned that this threshold is sharp, moreover, when $d \leq (1-\varepsilon)\log n$, a random matrix asymptotically almost surely contains a zero column. However, it seems that a proof of this fact does not appear in the existing literature, so we provide a direct proof of this threshold.

Proposition 8.1. *Let d, n be positive integers and $d \leq n/2$. Let M be an $n \times n$ random matrix where each row is independently and uniformly chosen from all n-dimensional binary vectors $\{0,1\}^n$ with exactly d ones. Let X denote the number of zero columns in M. Then the following holds.*

1. *Assume that $d = \log n - g(n)$ for some non-negative function g satisfying $g(n) \longrightarrow \infty$ as $n \to \infty$. Then,*
$$\lim_{n\to\infty} \mathbb{P}(X \geq 1) = 1.$$
That is, asymptotically (in n) the matrix M almost surely has at least one zero column. In particular, it holds for $d \leq (1-\varepsilon)\log n$ (for any fixed $\varepsilon > 0$).
2. *For any fixed $d \geq 1$ and fixed $\theta \in (0,1)$,*
$$\liminf_{n\to\infty} \mathbb{P}(X > \theta \mathbb{E}X) \geq (1-\theta)^2.$$
In particular,
$$\liminf_{n\to\infty} \mathbb{P}(X > \theta e^{-d} n/2) \geq (1-\theta)^2,$$
that is, for large enough n, the nullity of the matrix M is bounded from below by a constant proportion of n with a positive probability.

Proof. In both cases we will use the Paley-Zygmund inequality, which states that for any $\lambda \in [0,1]$,

$$\mathbb{P}(X > \lambda\, \mathbb{E}X) \geq (1-\lambda)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}.$$

We estimate $\mathbb{E}X$ and $\mathbb{E}X^2$ separately. Note that $X = \sum_{i=1}^n I_i$ where I_i is the indicator function of the event that i-th column is zero. Thus,

$$q := \mathbb{E}X = \sum_{i=1}^n E(I_i) = n(1-d/n)^n$$

and

$$\mathbb{E}X^2 = \sum_i \mathbb{E}I_i^2 + \sum_{i\neq j} \mathbb{E}I_i I_j = q + \sum_{i\neq j} \mathbb{P}(I_i = 1, I_j = 1).$$

Since $I_i = 1, I_j = 1$ means that i-th and j-th columns of M are zero columns and since $\mathbb{P}$ is the uniform probability on $\mathcal{M}_{n,d}$,

$$\mathbb{P}(I_i = 1, I_j = 1) = \left(\frac{\binom{n-2}{d}}{\binom{n}{d}}\right)^n = \left(\frac{(n-d)(n-d-1)}{n(n-1)}\right)^n \leq (1-d/n)^{2n} = (q/n)^2.$$

Therefore,

$$\mathbb{E}X^2 \le q + n(n-1)(q/n)^2 = q + \frac{n-1}{n}q^2,$$

and, consequently,

$$\frac{(\mathbb{E}X)^2}{\mathbb{E}X^2} \ge \frac{q^2}{q + \frac{n-1}{n}q^2} = \frac{1}{1/q + \frac{n-1}{n}}.$$

Case 1: *$d = \log n - g(n)$, $g(n) \ge 0$, $g(n) \longrightarrow \infty$.* In this case, using $\log(1-x) \ge -x - x^2$ on $[0, 1/2]$ and $d \le n/2$, we have

$$n\log(1-d/n) \ge -d - d^2/n \ge -\log n + g(n) - (\log^2 n)/n \ge -\log n + g(n) - \log 2$$

for large enough n. Thus,

$$q = \mathbb{E}X \ge ne^{-\log n + g(n) - \log 2} = e^{g(n)}/2 \longrightarrow \infty \quad \text{ as } n \to \infty.$$

Applying the Paley-Zygmund inequality with $\lambda = 0$,

$$\mathbb{P}(X \ge 1) = \mathbb{P}(X > 0) \ge \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2} \ge \frac{1}{1/q + \frac{n-1}{n}} \longrightarrow 1 \quad \text{ as } n \to \infty.$$

Case 2: *$d \ge 1$ and $\theta \in (0,1)$ are fixed.* Note that in this case $q/n = (1-d/n)^n \to e^{-d}$ as $n \to \infty$, in particular, $q \to \infty$. Therefore, applying the Paley-Zygmund inequality with $\lambda = \theta$, we have

$$\mathbb{P}(X > \theta\mathbb{E}X) \ge (1-\theta)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2} \ge \frac{(1-\theta)^2}{1/q + \frac{n-1}{n}} \longrightarrow (1-\theta)^2 \quad \text{ as } n \to \infty.$$

The "in particular" part follows as $q/n \to e^{-d}$. □

Finally, given the resemblance between random combinatorial matrices and i.i.d. Bernoulli random matrices, we would like to formulate the following conjecture. For corresponding results on Bernoulli(d/n) random matrices we refer to [9, 40, 26, 28, 29] and references therein.

Conjecture 8.2. *Let Ω_0, Ω_r, and Ω_c denote the events that M has a zero column, that M contains two identical rows, and M contains two identical columns respectively. Then*

$$\mathbb{P}(M \text{ is singular}) = (1+o(1))\left(\mathbb{P}(\Omega_0) + \mathbb{P}(\Omega_r) + \mathbb{P}(\Omega_c)\right). \tag{8.1}$$

Analyzing probabilities in Conjecture 8.2, note that M is singular iff $E_n - M$ is singular, where E_n denotes the matrix with all entries equal to 1. Therefore, we assume that $d \le n/2$. For $d = n/2$, Nguyen [42] conjectured that a random combinatorial matrix M is singular with probability $(\frac{1}{2} + o(1))^n$, which has recently been confirmed by Jain, Sah, and Sawhney [29, Theorem 1.8]. Note that for $d = n/2$, this bound is weaker than (8.1). Furthermore, for $d < c_0 n$ (for a suitable $c_0 > 0$) we have $\mathbb{P}(\Omega_c) > \mathbb{P}(\Omega_r)$ and if $d \ge 2\log n$ then $\mathbb{P}(\Omega_0)$ dominates $\mathbb{P}(\Omega_c)$. Moreover, for any fixed positive constant $c < 1/2$, if $d \le cn$ then $\mathbb{P}(\Omega_0)$ dominates $\mathbb{P}(\Omega_r)$. Therefore, we may formulate a weaker conjecture.

Conjecture 8.3. *Let $c \in (0, 1/2)$ be a fixed constant, assume that $2\log n \le d \le cn$. Then*

$$\mathbb{P}(M \text{ is singular}) = (1+o(1))\mathbb{P}(\Omega_0).$$

References

[1] R. Adamczak and D. Chafaï. Circular law for random matrices with unconditional log-concave distribution. *Commun. Contemp. Math.*, 17(4):1550020, 22, 2015.

[2] R. Adamczak, D. Chafaï, and P. Wolff. Circular law for random matrices with exchangeable entries. *Random Structures & Algorithms*, 48(3):454–479, 2016.

[3] A. Adhikari and A. Dembo. Spectral measure for uniform d-regular digraphs. *arXiv:2310.14132*, 2023.

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

[5] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[6] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[7] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[8] A. Basak and M. Rudelson. The circular law for sparse non-Hermitian matrices. *Ann. Probab.*, 47(4):2359–2416, 2019.

[9] 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.

[10] R. Bauerschmidt, J. Huang, and H.-T. Yau. Local Kesten-McKay law for random regular graphs. *Comm. Math. Phys.*, 369(2):523–636, 2019.

[11] A. Beker. Limiting spectral laws for sparse random circulant matrices. *arXiv:2504.13833*, 2025.

[12] C. Bordenave, P. Caputo, and D. Chafaï. Spectrum of Markov generators on sparse random graphs. *Comm. Pure Appl. Math.*, 67(4):621–669, 2014.

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

[14] 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.

[15] N. Cook. The circular law for random regular digraphs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(4):2111–2167, 2019.

[16] N. Cook, L. Goldstein, and T. Johnson. Size biased couplings and the spectral gap for random regular graphs. *Ann. Probab.*, 46(1):72–125, 2018.

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

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

[19] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[20] C.-G. Esseen. On the concentration function of a sum of independent random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 9(4):290–308, 1968.

[21] 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.

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

[23] J. Friedman, J. Kahn, and E. Szemeredi. On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 587–598, 1989.

[24] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984.

[25] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.

[26] H. Huang. Rank of sparse Bernoulli matrices. *Pure Appl. Funct. Anal.*, 10(6):1365–1441, 2025.

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

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

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

[30] 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.

[31] K. Joag-Dev and F. Proschan. Negative association of random variables with applications. *The Annals of Statistics*, pages 286–295, 1983.

[32] H. Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.

[33] D. Li, A. E. Litvak, and T. Yu. An upper bound on the smallest singular value of dense random combinatorial matrices. *Journal of Complexity*, 95:102031, 2026.

[34] A. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. Structure of eigenvectors of random regular digraphs. *Transactions of the American Mathematical Society*, 371(11):8097–8172, 2019.

[35] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, and R. Vershynin. Random Euclidean embeddings in spaces of bounded volume ratio. *C. R. Math. Acad. Sci. Paris*, 339(1):33–38, 2004.

[36] 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.

[37] 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.

[38] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. Circular law for sparse random regular digraphs. *J. Eur. Math. Soc. (JEMS)*, 23(2):467–501, 2021.

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

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

[41] B. D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40:203–216, 1981.

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

[43] H. H. Nguyen. Random doubly stochastic matrices: the circular law. *Ann. Probab.*, 42(3):1161–1196, 2014.

[44] 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.

[45] G. Pan and W. Zhou. Circular law, extreme singular values and potential theory. *J. Multivariate Anal.*, 101(3):645–656, 2010.

[46] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

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

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

[49] A. Sah, J. Sahasrabudhe, and M. Sawhney. The limiting spectral law for sparse iid matrices. *arXiv preprint arXiv:2310.17635, Forum of Math Pi, to appear*, 2023.

[50] A. Sah, J. Sahasrabudhe, and M. Sawhney. The sparse circular law, revisited. *Bull. Lond. Math. Soc.*, 57(2):330–358, 2025.

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

[52] T. Tao and V. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.

[53] 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.

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

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

[56] P. M. Wood. Universality and the circular law for sparse random matrices. *Ann. Appl. Probab.*, 22(3):1266–1300, 2012.

[57] Y. Zhu. On the second eigenvalue of random bipartite biregular graphs. *Journal of Theoretical Probability*, 36(2):1269–1303, 2023.

Dongbin Li
Email address: dongbinli@snnu.edu.cn
School of Mathematics and Statistics, Shaanxi Normal University, Xi'an, 710119, China.

Alexander E. Litvak
Email address: alitvak@ualberta.ca
Dept. of Math. and Stat. Sciences, University of Alberta, Edmonton, AB, T6G 2R3, Canada.

Tingzhou Yu
Email address: tingzho1@ualberta.ca
Dept. of Math. and Stat. Sciences, University of Alberta, Edmonton, AB, T6G 2R3, Canada.