Bernoulli **30**(4), 2024, 2904–2931
https://doi.org/10.3150/23-BEJ1699

Extreme singular values of inhomogeneous sparse random rectangular matrices

IOANA DUMITRIU[1,a] and YIZHE ZHU[2,b]

[1] *Department of Mathematics, University of California San Diego, La Jolla, CA 92093, USA,* [a] *idumitriu@ucsd.edu*
[2] *Department of Mathematics, University of California Irvine, Irvine, CA 92697, USA,* [b] *yizhe.zhu@uci.edu*

We develop a unified approach to bounding the largest and smallest singular values of an inhomogeneous random rectangular matrix, based on the non-backtracking operator and the Ihara-Bass formula for general random Hermitian matrices with a bipartite block structure. We obtain probabilistic upper (respectively, lower) bounds for the largest (respectively, smallest) singular values of a large rectangular random matrix X. These bounds are given in terms of the maximal and minimal ℓ_2-norms of the rows and columns of the variance profile of X. The proofs involve finding probabilistic upper bounds on the spectral radius of an associated non-backtracking matrix B. The two-sided bounds can be applied to the centered adjacency matrix of sparse inhomogeneous Erdős-Rényi bipartite graphs for a wide range of sparsity, down to criticality. In particular, for Erdős-Rényi bipartite graphs $\mathcal{G}(n,m,p)$ with $p=\omega(\log n)/n$, and $m/n \to y \in (0,1)$, our sharp bounds imply that there are no outliers outside the support of the Marčenko-Pastur law almost surely. This result extends the Bai-Yin theorem to sparse rectangular random matrices.

Keywords: Extreme singular value; inhomogeneous random matrix; non-backtracking operator; random bipartite graph

1. Introduction

1.1. Extreme singular values of random matrices

The asymptotic and non-asymptotic behavior of extreme singular values of random matrices is a fundamental topic in random matrix theory (Rudelson and Vershynin, 2010, Vershynin, 2012). They are crucial quantities used to provide theoretical guarantees for randomized linear algebra algorithms on large data sets, with applications in machine learning, signal processing, and data science.

Consider an $n \times m$ random matrix X with $m/n \to y \in (0,1)$ with i.i.d. entries. Let $\sigma_{\max}(X)$ and $\sigma_{\min}(X)$ be the largest and smallest singular value of X, respectively. The classical Bai-Yin theorem (Bai and Yin, 1993) says that, under the finite fourth-moment assumption of the distribution of entries, almost surely,

$$\frac{1}{\sqrt{n}}\sigma_{\max}(X) \to 1+\sqrt{y}, \quad \frac{1}{\sqrt{n}}\sigma_{\min}(X) \to 1-\sqrt{y}. \tag{1}$$

This implies that there are no outliers outside the support of the Marčenko-Pastur law for $\frac{1}{n}X^*X$. A non-asymptotic version of the Bai-Yin theorem with a sharp constant for Gaussian matrices can be obtained from Gordon's inequality (Gordon, 1985, Han, 2022, Vershynin, 2012); beyond the Gaussian case, similar results were given in (Feldheim and Sodin, 2010) using the moment method for symmetric sub-Gaussian distributions (in addition, Tracy-Widom fluctuations were also proved). Under the more relaxed, finite second moment assumption, the convergence of the smallest singular value to the left

1350-7265

edge of the Marčenko-Pastur law was proved in (Tikhomirov, 2015). Later, the convergence of the smallest and largest singular values to the edge of the spectrum was proved in (Chafaï and Tikhomirov, 2018, Heiny and Mikosch, 2018), for more general models. Finally, for sparse, heavy-tailed random matrices, the convergence of the largest singular value was considered in (Auffinger and Tang, 2016).

Besides the sharp asymptotic behavior for the extreme singular values, non-asymptotic bounds (which do not capture the sharp constants, but the correct order) for $\sigma_{\max}(X)$ and $\sigma_{\min}(X)$ were constructed by using other arguments, including ε-nets (Vershynin, 2012), matrix deviation inequalities (Vershynin, 2018), and the variational principle (Zhivotovskiy, 2021). Largest singular values can also be bounded using the moment method (Latała, van Handel and Youssef, 2018) or by using the spectral norm bound for Hermitian matrices of size $(n+m)\times(n+m)$ (Bandeira and van Handel, 2016, Benaych-Georges, Bordenave and Knowles, 2020).

Similar results for the smallest singular values of rectangular random matrices are harder to obtain, especially when the matrices are sparse, partly because there are fewer methods of approach. A lower bound without forth-moment assumptions was given by (Koltchinskii and Mendelson, 2015, Tikhomirov, 2016), and for heavy-tailed distributions in (Guédon, Litvak and Tatarko, 2020, Guédon et al., 2017, Tikhomirov, 2018); None of the results above capture the sharp constant in (1). Litvak and Rivasplata (2012) considered the smallest singular values for random matrices with a prescribed pattern of zeros (which does not cover sparse Bernoulli random matrices); and Götze and Tikhomirov (2023) considered sparse Bernoulli random matrices with $p=\omega(\log^4 n/n)$. Very recently, Brailovskaya and van Handel (2022) provided a very general, non-asymptotic universality principle on the spectrum of inhomogeneous random matrices that captures the sharp constant for extreme singular values in a general setting, including inhomogeneous and sparse random rectangular matrices when $p=\omega(\log^4 n/n)$. We compare their results with ours in Remarks 2.2.

We should also note that the smallest singular value of a square random matrix behaves differently from the rectangular one in (1), and our lower bounds on the smallest singular value (Theorems 2.3 and 2.5) do not cover the square case. Specifically, in the square case, with high probability, $\sigma_{\min}(X)$ is of order $1/\sqrt{n}$. A unified bound in both square and rectangular cases can be found in (Rudelson and Vershynin, 2009), which gives a lower bound $\Omega(\sqrt{n}-\sqrt{m-1})$. For the square random matrices, the smallest singular value bounds were proved in (Cook, 2018, Livshyts, 2021, Livshyts, Tikhomirov and Vershynin, 2021) for inhomogeneous random matrices, and in (Basak and Rudelson, 2017, 2021, Che and Lopatto, 2019) for sparse random matrices.

Another related topic is the study of the concentration of spectral norm for inhomogeneous random matrices, including sparse random matrices (Alt, Ducatez and Knowles, 2021, Benaych-Georges, Bordenave and Knowles, 2019, 2020, Le, Levina and Vershynin, 2017, Tikhomirov and Youssef, 2021), Gaussian matrices with independent entries (Bandeira and van Handel, 2016, Latała, van Handel and Youssef, 2018, van Handel, 2017), Wishart-type matrices (Cai, Han and Zhang, 2022), general random matrices (Bandeira, Boedihardjo and van Handel, 2023, Brailovskaya and van Handel, 2022, Tropp, 2015), and non-backtracking matrices (Benaych-Georges, Bordenave and Knowles, 2020, Bordenave, Coste and Nadakuditi, 2023, Stephan and Massoulié, 2022).

1.2. Sparse random bipartite graphs

Extreme singular values of sparse random bipartite graphs are important quantities in the study of community detection (Florescu and Perkins, 2016, Wan and Meila, 2015, Zhou and Amini, 2019), coding theory (Janwa and Lal, 2003), matrix completion (Bhojanapalli and Jain, 2014, Brito, Dumitriu and Harris, 2022), numerical linear algebra (Avron, Druinsky and Toledo, 2019), and theoretical computer science (Deshpande et al., 2019, Guruswami, Manohar and Mosheiff, 2022). However, classical estimates for sub-Gaussian random matrices (Vershynin, 2012) cannot be directly applied to sparse random matrices due to the lack of concentration.

Recently, the non-backtracking operator has proved to be a powerful tool in the study of spectra of sparse random graphs, specifically, when the average degree of the random graph is bounded (Bordenave, 2020, Bordenave, Coste and Nadakuditi, 2023, Bordenave, Lelarge and Massoulié, 2018, Brito, Dumitriu and Harris, 2022, Dumitriu and Zhu, 2021, Stephan and Zhu, 2022) or slowly growing (Alt, Ducatez and Knowles, 2021, Benaych-Georges, Bordenave and Knowles, 2020, Coste and Zhu, 2021, Stephan and Massoulié, 2022, Wang and Wood, 2023). Most results obtained with the help of non-backtracking operators are concerned with the largest eigenvalues and spectral gaps, with the exception of (Brito, Dumitriu and Harris, 2022), which gives a lower bound on the smallest singular value of a random biregular bipartite graph, and (Coste and Zhu, 2021), which gives the location of isolated real eigenvalues inside the bulk of the spectrum for the non-backtracking operator. Lower bounds on smallest singular values for sparse random rectangular matrices were also considered in (Götze and Tikhomirov, 2023, Guruswami, Manohar and Mosheiff, 2022, Zhu, 2023) by other methods for various models.

1.3. Contributions of this paper

In this paper, we provide new non-asymptotic bounds on the extreme singular values of inhomogeneous sparse rectangular matrices. Our main tool is the non-backtracking operator for a general $n \times n$ matrix defined as follows.

Definition 1.1 (Non-backtracking operator). Let $H \in M_n(\mathbb{C})$. For $e = (i, j), f = (k, l)$, define the non-backtracking operator of H as an $n^2 \times n^2$ matrix B such that

$$B_{ef} := H_{kl}\mathbf{1}_{j=k}\mathbf{1}_{i\neq l}. \tag{2}$$

To associate a non-backtracking operator with a rectangular $n \times m$ random matrix X, we work with the non-backtracking operator B of an $(n + m) \times (n + m)$ matrix

$$H = \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}. \tag{3}$$

We can summarize the major steps in our proofs as follows:

1. We improve the deterministic bound given in (Benaych-Georges, Bordenave and Knowles, 2020) on the largest singular value of rectangular matrices, in terms of the spectral radius $\rho(B)$ of B;
2. We provide a new deterministic lower bound on the smallest singular value in terms of $\rho(B)$;
3. We give an improved probabilistic bound on $\rho(B)$ for inhomogeneous random rectangular matrices;
4. Combining the deterministic and probabilistic results, we prove, in a unified way, two-sided probabilistic bounds for a general inhomogeneous rectangular random matrix model; we also specialize them for inhomogeneous sparse random matrices for a wide range of sparsity. Our main results are stated in Section 2.

Although some partial efforts were previously made in the literature ((Auffinger and Tang, 2016, Cai, Han and Zhang, 2022, Götze and Tikhomirov, 2023)) and some of our tools have been developed in (Benaych-Georges, Bordenave and Knowles, 2020), the crucial contribution of this paper is a deep and unified understanding of the relationship between the singular values of a rectangular matrix and the eigenvalues of its associated non-backtracking operator, which allows us to give a unified treatment of extreme singular values. For the first time and with a minimal set of conditions, this, in turn, allows us to extend the Bai-Yin theorem to sparse random bipartite graphs (Corollary 2.1) with average degree $\omega(\log n)$. For inhomogeneous sparse random bipartite graphs, the smallest singular value bound explicitly depends on the maximal and minimal expected degrees without extra constant factors. The upper bound on the largest singular value is valid when the maximal average degree $d = \Omega(1)$, and the lower bound on the smallest singular value is valid down to the critical regime $d = \Omega(\log n)$.

Our proof relies on the connection between extreme singular values of X and the spectral radius of the corresponding non-backtracking operator B. In the biregular bipartite graph case (Brito, Dumitriu and Harris, 2022), this relation is described by the Ihara-Bass formula (Bass, 1992), which results in algebraic equations involving the spectrum of X and B. For inhomogeneous Erdős-Rényi bipartite graphs, exact algebraic equations no longer work. Instead, we find deterministic inequalities between the extreme singular values of X and the spectral radius of B, using a block version of the generalized Ihara-Bass formula given in Lemma 3.2.

The proof of the spectral norm bound for Hermitian random matrices in (Benaych-Georges, Bordenave and Knowles, 2020) relies on the relation between the largest eigenvalue of a Hermitian matrix and the largest real eigenvalue of its associated non-backtracking matrix. In our case, to get the lower bound on $\sigma_{\min}(X)$, we connected the small singular values of X to the largest purely imaginary eigenvalue (in modulus) of B (see Lemma 3.5). This idea helped us establish here a similar phenomenon in a general inhomogeneous setting beyond the random bipartite biregular graph case studied in (Brito, Dumitriu and Harris, 2022).

Based on (Benaych-Georges, Bordenave and Knowles, 2020), a more refined phase transition behavior of extreme eigenvalues for homogeneous Erdős-Rényi graphs at $d = \log n/(\log 4 - 1)$ was shown in (Alt, Ducatez and Knowles, 2021), and the same threshold was also obtained in (Tikhomirov and Youssef, 2021) with a different method. It is possible to combine the techniques in (Alt, Ducatez and Knowles, 2021) with our Theorems 2.2 and 2.3 to study the phase transition behavior for $d = c \log n$ in homogeneous Erdős-Rényi bipartite graphs, and we intend to consider this problem in subsequent work.

Linear algebra notation

We say $X \preceq Y$ for two Hermitian matrices X and Y if $Y - X$ is a positive semidefinite matrix. For $c \in \mathbb{R}$ $X \preceq c$ means $X \preceq cI$. $\|X\|$ is the spectral norm of X, and for a square matrix B, $\rho(B)$ is the spectral radius of B, and $\sigma(B)$ is the set of all eigenvalues of B. We denote $\sigma_{\max}(X), \sigma_{\min}(X)$ the largest and smallest singular values of a matrix X, respectively. All C, c, C_i, c_i for $i \in \mathbb{N}$ are universal constants. $x \vee y, x \wedge y$ are the maximum and minimum of x and y, respectively. Denote $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ for $x < 0$.

Organization of the paper

The rest of the paper is structured as follows. In Section 2, we state our main results for sparse inhomogeneous Erdős-Rényi bipartite graphs and general rectangular random matrices. In Section 3, we connect the spectra of H defined in (3) and that of its non-backtracking operator B and prove some deterministic bounds on the extreme singular values of X. In Sections 4 and 5, we give a probabilistic upper bound on $\rho(B)$ for a general random matrix X with the moment method. In Sections 6 and 7, we give proofs for the probabilistic bounds on extreme singular values of X and specialize them for inhomogeneous sparse bipartite random graphs.

2. Main results

2.1. Inhomogeneous Erdős-Rényi bipartite graphs

Definition 2.1 (Inhomogeneous Erdős-Rényi bipartite graph). An *inhomogeneous Erdős-Rényi bipartite graph* $G \sim \mathcal{G}(n,m,p_{ij})$ is a random bipartite graph defined on a vertex set $V = V_1 \cup V_2$, where $|V_1| = n, |V_2| = m$ such that an edge ij, $i \in [n], j \in [m]$ is included independently with probability p_{ij}. Let $A \in \{0,1\}^{n\times m}$ be the *biadjacency matrix* of G such that $A_{ij} = 1$ if ij is an edge in G and $A_{ij} = 0$ otherwise. The *adjacency matrix* of G is given by $\begin{bmatrix} 0 & A \\ A^\top & 0 \end{bmatrix}$.

Notation and assumptions

In Section 2.1, we will use the following notation and assumptions and specify occasional, limited-use additional assumptions as necessary.

- The maximal expected degree among all vertices from $V_1 \cup V_2$ is denoted by

$$d := \max_{i\in[n], j\in[m]} \left(\sum_{k=1}^{m} p_{ik}, \sum_{k=1}^{n} p_{kj} \right)$$

- The normalized maximal expected degrees from V_2 (respectively, V_1) as

$$\rho_{\max} := \frac{1}{d} \max_{i\in[n]} \sum_{j\in[m]} p_{ij}, \quad \tilde{\rho}_{\max} := \frac{1}{d} \max_{j\in[m]} \sum_{i\in[n]} p_{ij}, \tag{4}$$

  and the normalized minimal expected degree in V_2 is defined as

$$\tilde{\rho}_{\min} = \frac{1}{d} \min_{j\in[m]} \sum_{i\in[n]} p_{ij}(1-p_{ij}).$$

- Denote $N := n \vee m$ and $\eta := \sqrt{\log N/d}$.

Assumption 1. We let $\gamma := \rho_{\max} \wedge \tilde{\rho}_{\max}$, where $\gamma \in (0,1]$ and $\gamma = \Omega(1)$. Note that when all $p_{ij} = p$, $\gamma = \frac{n\wedge m}{n\vee m}$ is the aspect ratio. Assume $d \geq \gamma^{-1/2}$ and $d^{\frac{7}{2}} \max_{ij} p_{ij} \leq N^{-\frac{1}{10}}$.

We first state the following upper bound on the largest singular values of $A - \mathbb{E}A$ for a wide range of sparsity down to $d = \Omega(1)$.

Theorem 2.2 (Largest singular value for inhomogeneous sparse random matrices). *Let A be the biadjacency matrix of an inhomogeneous random graph. Under Assumption 1,*

$$\frac{1}{\sqrt{d}} \mathbb{E}[\sigma_{\max}(A - \mathbb{E}A)] \leq 1 + \sqrt{\gamma} + O\left(\frac{\eta}{1 \vee \sqrt{\log \eta}} \right). \tag{5}$$

Moreover, with probability at least $1 - O(N^{-3})$, we have

$$\frac{1}{\sqrt{d}} \sigma_{\max}(A - \mathbb{E}A) \leq 1 + \sqrt{\gamma} + O\left(\frac{\eta}{1 \vee \sqrt{\log \eta}} \right). \tag{6}$$

Remark 2.1. Below, we qualitatively explain Theorem 2.2 in the various d regimes. When $d = \omega(\log N)$ and $d \le N^a$ for some constant $a > 0$, (5) is dominated by the first term, which gives $1 + \sqrt{\gamma} + o(1)$. This is sharp, and it recovers the results of (Latała, van Handel and Youssef, 2018, Theorem 4.9 and Example 4.10), and (Cai, Han and Zhang, 2022, Theorem 3.5). Moreover, our model can be seen as a specific case considered in (Benaych-Georges, Bordenave and Knowles, 2019), but their bound yields the weaker $2 + o(1)$ on the right-hand side of (5). The fact that we cannot cover denser regimes is an artifact of our proof method (see Theorem 4.1).

When $d = O(\log N)$, Theorem 2.2 is optimal up to a constant factor. When $d = o(\log N)$, the second term in (5) is dominating. Our results yield the sharp bound $O(\eta/\sqrt{\log \eta})$. This is tight up to a constant factor, down to $d \ge \gamma^{-\frac{1}{2}}$, and it agrees with the results in (Benaych-Georges, Bordenave and Knowles, 2019, Krivelevich and Sudakov, 2003) for non-bipartite graphs. Note that the results in (Latała, van Handel and Youssef, 2018) and (Cai, Han and Zhang, 2022) both imply an $O(\eta)$ upper bound, which is strictly weaker.

The real novelty arises in the following theorem, which provides the smallest singular value bound on $A - \mathbb{E}A$, down to $d = \Omega(\log n)$, and it is sharp in certain cases. Our sparsity assumption on d is optimal up to a constant factor: when $d < (1-\varepsilon)\log n$, with high probability there are isolated vertices, and that implies $\sigma_{\min}(A - \mathbb{E}A) = 0$.

For the next result, we need the following additional assumption.

Assumption 2. Let $n \ge m$ and $\delta \in (0,1)$. Assume $\tilde{\rho}_{\max} \ge \rho_{\max}$, $\tilde{\rho}_{\min} > \sqrt{\gamma}$, and $\min\{1 - \sqrt{\gamma}, \tilde{\rho}_{\min} - \sqrt{\gamma}\} = \Omega(1)$, $d^{\frac{7}{2}} \max_{ij} p_{ij} \le n^{-\frac{1}{10}}$, and there exists an absolute constant $C > 0$ such that

$$d \ge C \max\left\{\delta^{-1/2}, \delta^{-2}\gamma^{-1}\log n\right\}.$$

Theorem 2.3 (Smallest singular value for inhomogeneous sparse random matrices). *Under Assumptions 1 and 2,*

$$\frac{1}{\sqrt{d}}\mathbb{E}\sigma_{\min}(A - \mathbb{E}A) \ge \sqrt{(1-\sqrt{\gamma})(\tilde{\rho}_{\min} - \sqrt{\gamma})} - O(\delta^{1/4}). \tag{7}$$

Moreover, with probability at least $1 - O(n^{-3})$,

$$\frac{1}{\sqrt{d}}\sigma_{\min}(A - \mathbb{E}A) \ge \sqrt{(1-\sqrt{\gamma})(\tilde{\rho}_{\min} - \sqrt{\gamma})} - O(\delta^{1/4}). \tag{8}$$

Remark 2.2. The very general universality principle proved in (Brailovskaya and van Handel, 2022, Theorem 2.13) implies

$$d^{-1/2}\sigma_{\min}(A - \mathbb{E}A) \ge \sqrt{\tilde{\rho}_{\min}} - \sqrt{\gamma} - O(d^{-1/6}\log^{2/3}(n)) \tag{9}$$

with high probability. The leading constant $\sqrt{\tilde{\rho}_{\min}} - \sqrt{\gamma}$ in (9) is strictly better than our leading constant $\sqrt{(1-\sqrt{\gamma})(\tilde{\rho}_{\min} - \sqrt{\gamma})}$ in (8) for inhomogeneous random graphs and it matches our leading constant in the homogeneous case when $\tilde{\rho}_{\min} = 1$. However, their results are only valid for the regime when $d = \omega(\log^4 n)$—the benefit of our method is that our results extend down to $d = \Omega(\log n)$.

Cai, Han and Zhang (2022) studied concentration inequalities for inhomogeneous Wishart-type matrices. When $d = \omega(\log n)$, by triangle inequality, (Cai, Han and Zhang, 2022, Theorem 3.5) implies $d^{-1}\mathbb{E}[\sigma^2_{\min}(A - \mathbb{E}A)] \ge \tilde{\rho}_{\min} - 2\sqrt{\gamma} - \gamma - O(\sqrt{\eta})$. Their leading constant is strictly weaker than ours in all regimes. In (Götze and Tikhomirov, 2023), the authors considered the smallest singular value of a sparse random matrix $X = A \circ Y$, where A is a sparse Bernoulli matrix, and Y is a Wigner matrix whose entries have bounded support, and $\circ$ is the Hadamard product. This covers the homogeneous Erdős-Rényi case, but not the inhomogeneous one. The authors showed that the smallest singular value is $\Omega(\sqrt{np})$ when $p = \omega(\log^4 n)$ (see (Götze and Tikhomirov, 2023, Theorem 1.2)). Note that our Theorem 2.5 covers a more general inhomogeneous model, which includes (Götze and Tikhomirov, 2023).

Remark 2.3. The lower bound in (8) is $\sqrt{(1-\sqrt{\gamma})(\tilde{\rho}_{\min} - \sqrt{\gamma})} - o(1)$ when $d = \omega(\log n)$. When $d = \Omega(\log n)$, we can obtain an $\Omega(1)$ lower bound when δ is sufficiently small. We have assumed $\gamma = \Omega(1)$ for simplicity throughout. However, Theorems 2.2 and 2.3 also work for $\gamma = o(1)$, see Sections 6 and 7 with weaker bounds. It remains an open question to find the optimal dependence on γ when $\gamma = o(1)$ in (6) and (8).

For the next result, we assume $d = \omega(\log n)$, and the expected degree of each vertex concentrates. The Marčenko-Pastur law for the matrix $\frac{1}{d}(A - \mathbb{E}A)^\top (A - \mathbb{E}A)$ can be proved in the same way the semicircle law was proved in (Zhu, 2020, Corollary 4.3) via graphon theory. This implies the upper bound and lower bound given by (6) and (8) are tight. We state the generalization of the Bai-Yin theorem to sparse random bipartite graphs in the following corollary.

Corollary 2.1 (Bai-Yin theorem for supercritical random bipartite graphs). *Let A be the biadjacency matrix of an inhomogeneous Erdős-Rényi bipartite graph sampled from $\mathcal{G}(n,m,p_{ij})$. Assume $\frac{m}{n} \to y \in (0,1)$ as $n \to \infty$ and the parameters $\{p_{ij}\}$ satisfy*

$$\max_{i\in[n]} \left| \frac{1}{d} \sum_{j\in[m]} p_{ij} - 1 \right| = o(1), \quad \max_{j\in[m]} \left| \frac{1}{d} \sum_{i\in[n]} p_{ij} - y \right| = o(1). \tag{10}$$

Assume $d = \omega(\log n), d \le n^{1/5}$ and $\max_{ij} p_{ij} = O(d/n)$. Then, almost surely,

$$\lim_{n\to\infty} \frac{1}{\sqrt{d}} \sigma_{\min}(A - \mathbb{E}A) = 1 - \sqrt{y}, \qquad \lim_{n\to\infty} \frac{1}{\sqrt{d}} \sigma_{\max}(A - \mathbb{E}A) = 1 + \sqrt{y}. \tag{11}$$

Corollary 2.1 covers the homogeneous Erdős-Rényi bipartite graph $\mathcal{G}(n,m,p)$ with $p = \omega(\log n/n)$ and $p \le n^{-4/5}$. For denser cases when $p \ge n^{-4/5}$, the right edge limit (11) has been obtained in (Latała, van Handel and Youssef, 2018), and the left edge limit in (11) may be obtainable by moment methods without using the non-backtracking operator (Bai and Yin, 1993, Feldheim and Sodin, 2010).

2.2. Inhomogeneous random rectangular matrices

To prove our results in Section 2.1, we work with a more general matrix model whose entries have bounded support. Theorem 2.2 and Theorem 2.3 are obtained from Theorem 2.4 and Theorem 2.5 by specifying model parameters and taking $q = \sqrt{d}$.

Notation and assumptions

In Section 2.2, we use the following notation and assumptions:

- The minimal column sum of variances is denoted by

$$\tilde{\rho}_{\min} := \min_{j\in[m]} \sum_{i\in[n]} \mathbb{E}|X_{ij}|^2.$$

- Denote $N := n \vee m$ and $\eta := \sqrt{\log N}/q$.

Assumption 3. We assume there exist $q > 0, \kappa \geq 1$, $\rho_{\max}, \tilde{\rho}_{\max}$ such that

$$\max_{ij} |X_{ij}| \leq \frac{1}{q}, \quad \max_{ij} \mathbb{E}|X_{ij}|^2 \leq \frac{\kappa}{N}, \tag{12}$$

$$\max_{j\in[m]} \sum_{i\in[n]} \mathbb{E}|X_{ij}|^2 \leq \rho_{\max}, \quad \max_{i\in[n]} \sum_{j\in[m]} \mathbb{E}|X_{ij}|^2 \leq \tilde{\rho}_{\max}, \tag{13}$$

where $\rho_{\max} \vee \tilde{\rho}_{\max} = 1$ and $\rho_{\max} \wedge \tilde{\rho}_{\max} = \gamma \in (0,1]$, and $\gamma = \Omega(1)$. Assume

$$\gamma^{-\frac{1}{4}} \leq q \leq N^{\frac{1}{10}} \kappa^{-\frac{1}{9}} \gamma^{-\frac{1}{18}}.$$

Theorem 2.4 (Largest singular value). *Let X be an $n \times m$ random matrix with independent entries, and $\mathbb{E}X = 0$. Then under Assumption 3,*

$$\mathbb{E}[\sigma_{\max}(X)] \leq \sqrt{\gamma} + 1 + O\left(\frac{\eta}{1 \vee \sqrt{\log \eta}}\right). \tag{14}$$

Moreover, with probability at least $1 - O(N^{-3})$,

$$\sigma_{\max}(X) \leq \sqrt{\gamma} + 1 + O\left(\frac{\eta}{1 \vee \sqrt{\log \eta}}\right). \tag{15}$$

Assumption 4. Let $n \geq m$ and $\delta \in (0,1)$. Assume $\tilde{\rho}_{\max} \geq \rho_{\max}$, $\tilde{\rho}_{\min} > \sqrt{\gamma}$, and $\min\{1 - \sqrt{\gamma}, \tilde{\rho}_{\min} - \sqrt{\gamma}\} = \Omega(1)$. Assume for $\delta \in [0,1)$, q satisfies

$$C \max\left\{\delta^{-1/2}, \delta^{-1}\gamma^{-1/2}\sqrt{\log n}\right\} \leq q \leq n^{\frac{1}{10}} \kappa^{-\frac{1}{9}} \gamma^{-\frac{1}{18}}$$

for an absolute constant $C > 0$.

Theorem 2.5 (Smallest singular value). *Let X be an $n \times m$ random matrix with independent entries and $\mathbb{E}X = 0$. Under Assumptions 3 and 4, we have*

$$\mathbb{E}[\sigma_{\min}(X)] \geq \sqrt{(1 - \sqrt{\gamma})(\tilde{\rho}_{\min} - \sqrt{\gamma})} - O(\delta^{1/4}). \tag{16}$$

Moreover, with probability at least $1 - O(n^{-3})$, we have

$$\sigma_{\min}(X) \geq \sqrt{(1 - \sqrt{\gamma})(\tilde{\rho}_{\min} - \sqrt{\gamma})} - O(\delta^{1/4}). \tag{17}$$

Remark 2.4. Through standard truncation arguments, described in detail, e.g., in (Borodin, Corwin and Guionnet, 2019, Pages 41-73) and (Bai and Silverstein, 2010, Chapter 5), our results in Section 2.2 can be applied to random variables with unbounded support. This includes, for example, the dense Gaussian case and any other cases satisfying certain Lindeberg's conditions (Bai and Silverstein, 2010).

3. Spectral relation between X and B

3.1. Generalized Ihara-Bass formula

We will make use of the following generalized Ihara-Bass formula proved in (Benaych-Georges, Bordenave and Knowles, 2020, Watanabe and Fukumizu, 2009). When H is the adjacency matrix of a graph, Lemma 3.1 reduces to the classical Ihara-Bass formula in (Bass, 1992, Kotani and Sunada, 2000).

Lemma 3.1 (Lemma 4.1 in (Benaych-Georges, Bordenave and Knowles, 2020)). *Let $H \in M_n(\mathbb{C})$ with associated non-backtracking matrix B. Let $\lambda \in \mathbb{C}$ satisfying $\lambda^2 \neq H_{ij}H_{ji}$ for all $i, j \in [n]$. Define $H(\lambda)$ and $M(\lambda) = \operatorname{diag}(m_i(\lambda))_{i\in[n]}$ as*

$$H_{ij}(\lambda) := \frac{\lambda H_{ij}}{\lambda^2 - H_{ij}H_{ji}}, \quad m_i(\lambda) := 1 + \sum_{k\in[n]} \frac{H_{ik}H_{ki}}{\lambda^2 - H_{ik}H_{ki}}. \tag{18}$$

Then $\lambda \in \sigma(B)$ if and only if $\det(M(\lambda) - H(\lambda)) = 0$.

By itself, Lemma 3.1 is not sharp enough to yield a tight upper bound for $\sigma_{\max}(X)$, and it cannot yield any results for the smallest singular values. Therefore, we have developed a customized approach for the block matrix model, including a sharp analysis of the non-backtracking operator, which will lead to tight results in both cases. The first step in this approach is the following customized version of Lemma 3.1.

Lemma 3.2. *Let X be an $n \times m$ complex matrix and $H = \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}$. Let B be the non-backtracking operator associated with H. Define an $n \times m$ matrix $X(\lambda)$, and two diagonal matrices $M_1(\lambda) = \operatorname{diag}(m_i^{(1)}(\lambda))_{i\in[n]}$, $M_2(\lambda) = \operatorname{diag}(m_i^{(2)}(\lambda))_{i\in[m]}$ as follows:*

$$X_{ij}(\lambda) = \frac{\lambda X_{ij}}{\lambda^2 - |X_{ij}|^2}, \quad m_i^{(1)}(\lambda) = 1 + \sum_{k\in[m]} \frac{|X_{ik}|^2}{\lambda^2 - |X_{ik}|^2}, \quad m_j^{(2)}(\lambda) = 1 + \sum_{k\in[n]} \frac{|X_{kj}|^2}{\lambda^2 - |X_{kj}|^2}.$$

Assume $M_1(\lambda)$ is non-singular. Then $\lambda \in \sigma(B)$ if and only if

$$\det(M_2(\lambda) - X(\lambda)^* M_1(\lambda)^{-1} X(\lambda)) \det(M_1(\lambda)) = 0. \tag{19}$$

Proof. From Lemma 3.1, $\lambda \in \sigma(B)$ if and only if $\det \begin{bmatrix} M_1(\lambda) & -X(\lambda) \\ -X^*(\lambda) & M_2(\lambda) \end{bmatrix} = 0$. Since $M_1(\lambda)$ is non-singular, by the determinant formula for block matrices, (19) holds. □

3.2. Deterministic upper bound on the largest singular value

Using Lemma 3.1, we bound $\sigma_{\max}(X)$ in terms of the maximal Euclidean norm of rows and columns of X and $\rho(B)$ as follows.

Lemma 3.3. *Let H, X, and B be defined as in Lemma 3.2. Suppose*

$$\max_{ij} |X_{ij}| \le \delta, \quad \max_{j\in[m]} \sum_{i\in[n]} |X_{ij}|^2 \le \tilde{\rho}_{\max}(1+\delta), \tag{20}$$

$$\max_{i\in[n]} \sum_{j\in[m]} |X_{ij}|^2 \le \rho_{\max}(1+\delta), \tag{21}$$

with $\rho_{\max} \vee \tilde{\rho}_{\max} = 1, \rho_{\max} \wedge \tilde{\rho}_{\max} = \gamma$ and $\delta \in [0, \gamma^{\frac{1}{2}}]$. Let $\lambda \ge \max\{\gamma^{\frac{1}{4}}(1+\sqrt{\delta}), \rho(B)\}$. Then

$$\sigma_{\max}(X)^2 \le \left(\lambda + \frac{1}{\lambda}\right)\left(\lambda + \frac{\gamma}{\lambda}\right) + 6\gamma^{-1}\delta\left(2\lambda + \frac{1+\gamma}{\lambda}\right) + 36\gamma^{-2}\delta^2.$$

Proof. From Lemma 3.1, by continuity, $M(\lambda) - H(\lambda) \ge 0$ for $\lambda \ge \lambda_0 := \max\{\gamma^{\frac{1}{4}}(1+\sqrt{\delta}), \rho(B)\}$. For any $\lambda \ge \lambda_0$, we have

$$|\lambda H_{ij}(\lambda) - H_{ij}| = \frac{|H_{ij}|^3}{\lambda^2 - |H_{ij}|^2} \le \frac{\delta |H_{ij}|^2}{(\lambda^2 - \delta^2)} \le \gamma^{-\frac{1}{2}}\delta |H_{ij}|^2. \tag{22}$$

By Gershgorin circle theorem, it implies that

$$\|\lambda H(\lambda) - H\|_2 \le \gamma^{-\frac{1}{2}}\delta \max_i \sum_j |H_{ij}|^2 \le \gamma^{-\frac{1}{2}}\delta(1+\delta) \le 2\gamma^{-\frac{1}{2}}\delta. \tag{23}$$

For any $i \in V_1$, from (18), for any $\lambda \ge \lambda_0$,

$$\lambda m_i(\lambda) - \left(\lambda + \frac{\rho_{\max}}{\lambda}\right) = \sum_{k\in V_2} \frac{\lambda |H_{ik}|^2}{\lambda^2 - |H_{ik}|^2} - \frac{\rho_{\max}}{\lambda} \le \frac{\rho_{\max}}{\lambda}\left(\frac{\lambda^2(1+\delta)}{\lambda^2 - \delta^2} - 1\right) = \rho_{\max}\delta \frac{\lambda(1+\lambda^{-2}\delta)}{\lambda^2 - \delta^2}$$

$$\le \rho_{\max}\delta \frac{\lambda(1+\gamma^{-\frac{1}{2}})}{\lambda^2 - \delta^2} \le 2\rho_{\max}\delta(\gamma^{-\frac{1}{2}} + \gamma^{-1}) \le 4\gamma^{-1}\delta, \tag{24}$$

where in the last step we consider the cases $\lambda \ge 2$ and $\lambda < 2$ and use the inequality $\lambda^2 - \delta^2 \ge \gamma^{\frac{1}{2}}$ in the second case. For any $i \in V_2$, similarly,

$$\lambda m_i(\lambda) - \left(\lambda + \frac{\tilde{\rho}_{\max}}{\lambda}\right) = \sum_{k\in V_1} \frac{\lambda |H_{ik}|^2}{\lambda^2 - |H_{ik}|^2} - \frac{\tilde{\rho}_{\max}}{\lambda} \le \frac{\tilde{\rho}_{\max}}{\lambda}\left(\frac{1+\delta}{1-\lambda^{-2}\delta^2} - 1\right) \le 4\gamma^{-1}\delta. \tag{25}$$

Then for $\lambda \ge \lambda_0$, with (23), (24), and (25),

$$0 \le \lambda(M(\lambda) - H(\lambda)) \le \begin{bmatrix} \lambda + \frac{\rho_{\max}}{\lambda} + 6\gamma^{-1}\delta & 0 \\ 0 & \lambda + \frac{\tilde{\rho}_{\max}}{\lambda} + 6\gamma^{-1}\delta \end{bmatrix} - H. \tag{26}$$

Let $d_1 := \lambda + \frac{\rho_{\max}}{\lambda} + 6\gamma^{-1}\delta, d_2 := \lambda + \frac{\tilde{\rho}_{\max}}{\lambda} + 6\gamma^{-1}\delta$. Then from (26), $\Delta = \begin{bmatrix} d_1 I & -X \\ -X^* & d_2 I \end{bmatrix} \ge 0$. Note that the following matrix factorization holds:

$$\begin{bmatrix} d_1 I & -X \\ -X^* & d_2 I \end{bmatrix} = \begin{bmatrix} \sqrt{d_1} I & 0 \\ 0 & \sqrt{d_2} I \end{bmatrix} \cdot \begin{bmatrix} I & -(d_1 d_2)^{-\frac{1}{2}} X \\ -(d_1 d_2)^{-\frac{1}{2}} X^* & I \end{bmatrix} \cdot \begin{bmatrix} \sqrt{d_1} I & 0 \\ 0 & \sqrt{d_2} I \end{bmatrix}.$$

Therefore, on the right-hand side of the above equation, the second matrix is positive semidefinite. We obtain

$$X^*X \le d_1 d_2 = \left(\lambda + \frac{1}{\lambda} + 6\gamma^{-1}\delta\right)\left(\lambda + \frac{\gamma}{\lambda} + 6\gamma^{-1}\delta\right)$$
$$= \left(\lambda + \frac{1}{\lambda}\right)\left(\lambda + \frac{\gamma}{\lambda}\right) + 6\gamma^{-1}\delta\left(2\lambda + \frac{1+\gamma}{\lambda}\right) + 36\gamma^{-2}\delta^2.$$

This gives the desired upper bound on $\sigma_{\max}(X)$. □

Lemma 3.3 gives an upper bound on $\sigma_{\max}(X)$ when the parameter $\delta \in [0, \gamma^{\frac{1}{2}}]$. By rescaling the entries in H, we can obtain a general bound depending on the following quantities:

$$\|H\|_{1,\infty} = \max_{ij} |H_{ij}|, \quad \|H\|_{2,\infty} := \max_i \left(\sum_j |H_{ij}|^2\right)^{\frac{1}{2}},$$

without the restriction on the range of δ. This is more convenient for us to handle sparse random bipartite graphs in the critical and subcritical regimes. Define

$$f(x) = \begin{cases} (x + x^{-1})\left(x + \frac{\gamma}{x}\right), & x \ge \gamma^{\frac{1}{4}} \\ (\gamma^{\frac{1}{2}} + 1)^2, & 0 \le x \le \gamma^{\frac{1}{4}} \end{cases}, \quad g(x) = \begin{cases} 2\left(x + \frac{1}{x}\right), & x \ge \gamma^{\frac{1}{4}} \\ 4 & 0 \le x \le \gamma^{\frac{1}{4}} \end{cases},$$

where $\gamma \in (0, 1]$ is a constant such that

$$\|X\|_{2,\infty} \wedge \|X^*\|_{2,\infty} \le \gamma^{\frac{1}{2}} \|H\|_{2,\infty}. \tag{27}$$

Lemma 3.4 (Deterministic upper bound on $\sigma_{\max}(X)$). *Let X be an $n \times m$ matrix, $H = \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}$, and γ be the constant in (27). The following inequality holds:*

$$\sigma_{\max}^2(X) \le \|H\|_{2,\infty}^2 f\left(\frac{\rho(B)}{\|H\|_{2,\infty}}\right) + 12\gamma^{-\frac{5}{4}} g\left(\frac{\rho(B)}{\|H\|_{2,\infty}}\right) \|H\|_{2,\infty}\|H\|_{1,\infty} + 36\gamma^{-2}\|H\|_{1,\infty}^2. \tag{28}$$

Proof. First assume $\|H\|_{2,\infty} = 1$. Set $\delta = \|H\|_{1,\infty}$. From (27), $\delta \le \gamma^{\frac{1}{2}}$. By Lemma 3.3, for $\lambda_0 = \max\{\gamma^{\frac{1}{4}}(1 + \sqrt{\delta}), \rho(B)\}$, we have

$$\sigma_{\max}^2(X) \le f(\lambda_0) + 6\gamma^{-1}\delta\left(2\lambda_0 + \frac{1+\gamma}{\lambda_0}\right) + 36\gamma^{-2}\delta^2. \tag{29}$$

When $\lambda_0 = \rho(B)$, it implies

$$\sigma_{\max}^2(X) \le f(\rho(B)) + 6\gamma^{-1}\delta g(\rho(B)) + 36\gamma^{-2}\delta^2. \tag{30}$$

When $\lambda_0 = \gamma^{\frac{1}{4}}(1 + \sqrt{\delta})$, from (29),

$$\sigma_{\max}^2(X) \le \lambda_0^2 + 1 + \gamma + \frac{\gamma}{\lambda_0^2} + 6\gamma^{-1}\delta\left(2\lambda_0 + \frac{1+\gamma}{\lambda_0}\right) + 36\gamma^{-2}\delta^2$$

$$\leq (\sqrt{\gamma}+1)^2 + 45\gamma^{-\frac{5}{4}}\delta + 36\gamma^{-2}\delta^2 \leq f(\rho(B)) + 12\gamma^{-\frac{5}{4}}\delta g(\rho(B)) + 36\gamma^{-2}\delta^2. \tag{31}$$

Combining (30) and (31), we have

$$\sigma_{\max}(X)^2 \leq f(\rho(B)) + 12\gamma^{-\frac{5}{4}} g(\rho(B))\|H\|_{1,\infty} + 36\gamma^{-2}\|H\|_{1,\infty}^2.$$

Then, for general H, by considering $\frac{H}{\|H\|_{2,\infty}}$ and repeating the proof above, we get the desired bound. □

3.3. Deterministic lower bound on the smallest singular value

The following lemma gives us a connection between the spectral radius of B and the smallest singular value of X. The proof relies on finding a relation between purely imaginary eigenvalues of B and singular values of X.

Lemma 3.5 (Deterministic lower bound on $\sigma_{\min}(X)$). *Let H, X and B be defined as in Lemma 3.2 and $n \geq m$. Let $0 < \gamma < 1$, $\delta \in [0,1)$, $C_1 > 0$ such that*

$$\max_{ij} |X_{ij}| \leq \delta, \quad \|X\|_2 \leq C_1,$$

$$\max_{i\in[n]} \sum_{j\in[m]} |X_{ij}|^2 \leq \gamma(1+\delta), \quad \tilde{\rho}_{\min}(1-\delta) \leq \sum_{i\in[n]} |X_{ij}|^2 \leq 1+\delta, \quad \forall j \in [m].$$

Define $\beta_0 = \max\{\gamma^{\frac{1}{4}}(1+\sqrt{\delta}), \rho(B)\}$. Then for $\beta \geq \beta_0$,

$$\sigma_{\min}^2(X) \geq \frac{\sqrt{\gamma}-\gamma}{\sqrt{\gamma}+\delta}\left(\frac{\beta^2}{\beta^2+\delta^2}\tilde{\rho}_{\min} - \beta^2 - C_\gamma\delta^2 - \delta\right), \tag{32}$$

where $C_\gamma = 4\gamma^{-\frac{1}{2}}(C_1+\gamma^{-1}\delta)\frac{\sqrt{\gamma}+\delta}{\sqrt{\gamma}-\gamma}$.

Remark 3.1. This bound is only informative when the right-hand side is positive, which necessitates $\gamma < 1$ and $\tilde{\rho}_{\min} > \sqrt{\gamma}$.

Proof. Take $\lambda = i\beta$, with $\beta \geq \beta_0 = \max\{\gamma^{\frac{1}{4}} + \sqrt{\delta}, \rho(B)\}$. Then

$$m_i^{(1)}(\lambda) = 1 - \sum_{k\in[m]} \frac{|X_{ik}|^2}{\beta^2+|X_{ik}|^2} \geq 1 - \frac{c}{\beta^2}(1+\delta)$$

$$\geq 1 - \frac{\gamma(1+\delta)}{\sqrt{\gamma}+\delta} = \frac{\delta(1-\gamma)+\sqrt{\gamma}-\gamma}{\sqrt{\gamma}+\delta} \geq \frac{\sqrt{\gamma}-\gamma}{\sqrt{\gamma}+\delta} =: C_2, \tag{33}$$

where C_2 is lower bounded by $\frac{\sqrt{\gamma}-\gamma}{1+\sqrt{\gamma}}$. This implies $M_1(\lambda)$ is invertible.

Define $H_2(\lambda) = X(\lambda)^* M_1(\lambda)^{-1} X(\lambda)$. From (19), $\lambda \in \sigma(B)$ if and only if $\det(M_2(\lambda) - H_2(\lambda)) = 0$. Recall that when $\lambda = i\beta$, $M_2(\lambda)$ is a real diagonal matrix, then $H_2(\lambda)$ is Hermitian. As $\beta \to \infty$, $M_2(\lambda) -$

$H_2(\lambda) = I + O(\beta^{-2})$. By continuity, $\det(M_2(\lambda) - H_2(\lambda)) > 0$ for $\beta > \beta_0$ and $M_2(\lambda) - H_2(\lambda)$ is positive semidefinite for any $\beta \geq \beta_0$. Since

$$\begin{aligned}&\beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) + X^* M_1(\lambda)^{-1}X\\ &= \beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) + \lambda X^*(\lambda)M_1(\lambda)^{-1}X - \lambda X^*(\lambda)M_1(\lambda)^{-1}X + X^* M_1(\lambda)^{-1}X,\end{aligned}$$

by triangle inequality,

$$\|\beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) + X^* M_1(\lambda)^{-1}X\| \leq \|\beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) + \lambda X^*(\lambda)M_1(\lambda)^{-1}X\| \tag{34}$$

$$+ \|\lambda X^*(\lambda)M_1(\lambda)^{-1}X - X^* M_1(\lambda)^{-1}X\|. \tag{35}$$

For the term in (34),

$$\begin{aligned}&\|\beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) + \lambda X^*(\lambda)M_1(\lambda)^{-1}X\| = \|\lambda^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) - \lambda X^*(\lambda)M_1(\lambda)^{-1}X\|\\ &\leq \|\lambda X(\lambda) - X\|\|\lambda X(\lambda)\|\|M_1(\lambda)^{-1}\| \leq \|\lambda X(\lambda) - X\|\|M_1(\lambda)^{-1}\|(\|X\| + \|\lambda X(\lambda) - X\|).\end{aligned} \tag{36}$$

Rewriting (22) with $\lambda = i\beta$, we obtain

$$|\lambda H_{ij}(\lambda) - H_{ij}| = \frac{|H_{ij}|^3}{\beta^2 + |H_{ij}|^2} \leq \frac{\delta |H_{ij}|^2}{\beta^2}. \tag{37}$$

Then, applying the Gershgorin circle theorem to the row of H yields,

$$\|\lambda X(\lambda) - X\| = \|\lambda H(\lambda) - H\| \leq \frac{\delta(1+\delta)}{\beta^2} \leq \frac{1}{\sqrt{\gamma}}\delta(1+\delta) \leq 2\gamma^{-\frac{1}{2}}\delta. \tag{38}$$

Then with (36) and (38), the term in (34) satisfies

$$\|\beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) + \lambda X^*(\lambda)M_1(\lambda)^{-1}X\| \leq (2\gamma^{-\frac{1}{2}}\delta)(C_1 + 2\gamma^{-1}\delta)C_2^{-1}. \tag{39}$$

Similarly, the second term in (35) satisfies

$$\|\lambda X^*(\lambda)M_1(\lambda)^{-1}X - X^* M_1(\lambda)^{-1}X\| \leq \|\lambda X^*(\lambda) - X^*\|\|X\|\|M_1(\lambda)^{-1}\| \leq (2\gamma^{-\frac{1}{2}}\delta)C_1 C_2^{-1}. \tag{40}$$

Therefore from (36), (39), and (40),

$$\|\beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) + X^* M_1(\lambda)^{-1}X\| \leq 4\delta\gamma^{-\frac{1}{2}}(C_1 + \gamma^{-1}\delta)C_2^{-1} = C_\gamma \delta.$$

which implies

$$X^* M_1(\lambda)^{-1}X \geq -\beta^2 X^*(\lambda)M_1(\lambda)^{-1}X(\lambda) - C_\gamma\delta = -\beta^2 H_2(\lambda) - C_\gamma\delta \geq -\beta^2 M_2(\lambda) - C_\gamma\delta, \tag{41}$$

where we used the condition $M_2(\lambda) - H_2(\lambda) \geq 0$ for $\beta \geq \beta_0$. On the other hand, for any $j \in [m]$,

$$\beta^2 m_j^{(2)}(\lambda) = \beta^2 + \sum_{k\in[n]} \frac{\beta^2 X_{kj}^2}{-\beta^2 - X_{kj}^2} \leq \beta^2 - \frac{\beta^2}{\beta^2+\delta^2}\sum_{k\in[n]} X_{kj}^2 \leq \beta^2 - \frac{\beta^2}{\beta^2+\delta^2}\tilde{\rho}_{\min}(1-\delta).$$

Then

$$-\beta^2 M_2(\lambda) \geq -\left(\beta^2 - \frac{\beta^2}{\beta^2+\delta^2}(1-\delta)\tilde{\rho}_{\min}\right). \tag{42}$$

Hence, with (42) and (41), $X^* M_1(\lambda)^{-1} X \geq \frac{\beta^2}{\beta^2+\delta^2}\tilde{\rho}_{\min} - \beta^2 - C_\gamma\delta - \delta$. From (33), we obtain $X^*X \geq C_2\left(\frac{\beta^2}{\beta^2+\delta^2}\tilde{\rho}_{\min} - \beta^2 - C_\gamma\delta - \tilde{\rho}_{\min}\delta\right)$. Thus, the lower bound on $\sigma_{\min}(X)$ in (32) holds. □

4. Probabilistic upper bound on $\rho(B)$

In this section, we provide a probabilistic bound on the spectral radius $\rho(B)$ for a random matrix H. This involves sharp estimates of the trace of high powers of B. We give our main result below.

Theorem 4.1. *Let* $H = \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}$ *be an* $(n+m)\times(n+m)$ *random Hermitian matrix with associated non-backtracking matrix* B*, where* X *is an* $n\times m$ *random matrix with centered independent entries. Under Assumption 3, for* $\varepsilon > 0$*, there exist universal constants* $C, c_1 > 0$ *such that*

$$\mathbb{P}(\rho(B) \geq \gamma^{\frac{1}{4}}(1+\varepsilon)) \leq C\gamma^{-\frac{5}{6}} N^{3-c_1 q\log(1+\varepsilon)}.$$

Remark 4.1. The bound on $\rho(B)$ in Theorem 4.1 is an inhomogenous analog to (Brito, Dumitriu and Harris, 2022, Theorem 3) for the non-backtracking operator of random biregular bipartite graphs, and for a broader range of the (sparsity) parameter q. Compared to (Benaych-Georges, Bordenave and Knowles, 2020, Theorem 2.5), we improve a factor $\gamma^{1/4}$ in the bound of $\rho(B)$ when H has a bipartite block structure.

The key estimate to prove Theorem 4.1 is the following trace bound on a high power of B.

Lemma 4.2. *Suppose* X *satisfies Assumption 3. There exist universal constants* c_0, C *such that for any* $\delta \in (0, \frac{1}{3})$*, and any odd positive integer* l *satisfying*

$$l \leq c_0 \min\left\{\delta q \log N, \frac{N^{\frac{1}{3}-\delta}}{\kappa^{\frac{1}{3}}\gamma^{\frac{1}{6}} q^2}\right\}, \tag{43}$$

we have

$$\mathbb{E}\mathrm{Tr}[B^l (B^l)^*] \leq C l^4 q^2 mn \gamma^{(l-1)/2}. \tag{44}$$

Assuming Lemma 4.2, we can now prove Theorem 4.1.

Proof of Theorem 4.1. Choose $0 < \delta < \frac{1}{30}$ in (43), we can take $l = \lceil \frac{1}{2} c_1 q \log N\rceil$ for some sufficiently small constant $c_1 > 0$ then (43) is satisfied. By Markov's inequality,

$$\begin{aligned}\mathbb{P}(\rho(B) \geq \gamma^{\frac{1}{4}}(1+\varepsilon)) &\leq \frac{\mathbb{E}\mathrm{Tr}[B^l (B^l)^*]}{\gamma^{l/2}(1+\varepsilon)^{2l}} \leq C l^4 q^2 mn\gamma^{-\frac{1}{2}}\left(\frac{1}{1+\varepsilon}\right)^{2l} \\ &\leq CN^2 q^6 (\log N)^4 \gamma^{-\frac{1}{2}}(1+\varepsilon)^{-c_1 q\log(1+\varepsilon)} \leq C\gamma^{-\frac{5}{6}} N^{3-c_1 q \log(1+\varepsilon)}\end{aligned}$$

for some absolute constant C. In the last inequality, we used the upper bound on q in the assumption of Theorem 4.1. This finishes the proof. □

5. Proof of Lemma 4.2

We now provide the proof Lemma 4.2. The proof is adapted from (Benaych-Georges, Bordenave and Knowles, 2020, Proposition 5.1) to the non-backtracking operator B of a Hermitian random matrix H with a bipartite block structure. The improvement compared to (Benaych-Georges, Bordenave and Knowles, 2020) is the factor $\gamma^{(l-1)/2}$ in (44), by doing path counting on a complete bipartite graph instead of a complete graph.

Let V_1, V_2 be the left and right vertex sets on the complete graph $K_{n,m}$, and

$$\vec{E} = \{(u,v) : u \in V_1, v \in V_2\} \cup \{(u,v) : u \in V_2, v \in V_1\}$$

be the set of all oriented edges in $K_{n,m}$. For any $e \in \vec{E}, f \in \vec{E}$, from the definition of B,

$$(B^l)_{ef} = \sum_{a_1,\dots,a_{l-1}\in\vec{E}} B_{ea_1}B_{a_1a_2}\cdots B_{a_lf} = \sum_{\xi} H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{l-1}\xi_l},$$

where the sum on the right hand side runs over $\xi = (\xi_{-1}, \xi_0, \dots, \xi_l)$ as a path of length $l+1$ in $K_{n,m}$ with $(\xi_{-1}, \xi_0) = e, (\xi_{l-1}, \xi_l) = f$, and $\xi_{i-1} \neq \xi_{i+1}$ for $0 \le i \le l-1$. Therefore, we have

$$\operatorname{Tr}[B^l(B^l)^*] = \sum_{e,f\in\vec{E}} |(B^l)_{ef}|^2 \tag{45}$$

$$= \sum_{\xi^1,\xi^2} H_{\xi_0^1\xi_1^1}H_{\xi_1^1\xi_2^1}\cdots H_{\xi_{l-1}^1\xi_l^1}H_{\xi_l^2\xi_{l-1}^2}\cdots H_{\xi_1^2\xi_0^2}, \tag{46}$$

where the sum runs over paths $\xi^1 = (\xi_{-1}^1, \dots, \xi_l^1), \xi^2 = (\xi_{-1}^2, \dots, \xi_l^2)$ of length $l+1$ such that $(\xi_{-1}^1, \xi_0^1) = (\xi_{-1}^2, \xi_0^2)$, $(\xi_{l-1}^1, \xi_l^1) = (\xi_{l-1}^2, \xi_l^2)$ and $\xi_{i-1}^j \neq \xi_{i+1}^j$ for $j = 1,2$ and $0 \le i \le l-1$. From (45), for fixed ξ_{-1}^1, ξ_{-1}^2, the sum in (46) is nonnegative and does not depend on ξ_{-1}^1, ξ_{-1}^2. Therefore, we can bound it by

$$\begin{aligned}\operatorname{Tr}[B^l(B^l)^*] \le{}& m \sum_{\substack{\xi^1,\xi^2:\\ \xi_0^1,\xi_0^2\in V_1}} H_{\xi_0^1\xi_1^1}H_{\xi_1^1\xi_2^1}\cdots H_{\xi_{l-1}^1\xi_l^1}H_{\xi_l^2\xi_{l-1}^2}\cdots H_{\xi_1^2\xi_0^2} \\ &+ n \sum_{\substack{\xi^1,\xi^2:\\ \xi_0^1,\xi_0^2\in V_2}} H_{\xi_0^1\xi_1^1}H_{\xi_1^1\xi_2^1}\cdots H_{\xi_{l-1}^1\xi_l^1}H_{\xi_l^2\xi_{l-1}^2}\cdots H_{\xi_1^2\xi_0^2},\end{aligned} \tag{47}$$

where the sum is over $\xi^1 = (\xi_0^1, \dots, \xi_l^1), \xi^2 = (\xi_0^2, \dots, \xi_l^2)$ such that $(\xi_0^1, \xi_{l-1}^1, \xi_l^1) = (\xi_0^2, \xi_{l-1}^2, \xi_l^2)$ and $\xi_{i-1}^j \neq \xi_{i+1}^j$ for $j = 1,2$ and $1 \le i \le l-1$. For $j = 1,2$, define

$$\tilde{C}_j = \{\xi = (\xi_0, \dots, \xi_{2l}) : \xi_0 = \xi_{2l} \in V_j, \xi_{l-1} = \xi_{l+1}, \xi_{i-1} \neq \xi_{i+1}, \forall 1 \le i \le 2l-1, i \neq l\}.$$

We can combine each pair of ξ^1, ξ^2 in (47) into a path of length $2l$ and simplify the bound as

$$\operatorname{Tr}[B^l(B^l)^*] \le m \sum_{\xi\in\tilde{C}_1} H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}} + n \sum_{\xi\in\tilde{C}_2} H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}}.$$

Since the entries of H are independent up to symmetry with mean zero, taking the expectation yields

$$\mathbb{E}\mathrm{Tr}[B^l(B^l)^*] \leq m \sum_{\xi\in C_1} \mathbb{E}H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}} + n \sum_{\xi\in C_2} \mathbb{E}H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}}, \tag{48}$$

where C_i is a subset of $\tilde{C}_i$ for $i = 1,2$ such that each edge in the graph spanned by ξ is visited at least twice by the path defined by ξ.

We will combine paths according to their equivalence classes defined as follows.

Definition 5.1 (Equivalence class). For each $\xi \in C_1 \cup C_2$, we define a graph

$$G_\xi = (V_1(\xi), V_2(\xi), E(\xi))$$

spanned by ξ. Let $g(\xi) = |E(\xi)| - |V(\xi)| + 1$ be the genus of G_ξ and $e(\xi) = |E(\xi)|$. We say two paths ξ_1,ξ_2 are equivalent if they are the same up to a relabelling of vertices (with the restriction that after relabelling, vertices in V_i stay in V_i for $i = 1,2$). Define $[\xi]$ the equivalence class of a path ξ.

Lemma 5.2. *For any $\overline{\xi} \in C_1$ with $|V_1(\overline{\xi})| = s_1(\overline{\xi}), |V_2(\overline{\xi})| = s_2(\overline{\xi})$, we have*

$$\mathbb{E} \sum_{\xi\in[\overline{\xi}]} H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}} \leq n(\kappa/N)^{g(\overline{\xi})} q^{2e(\overline{\xi})-2l} \rho_{\max}^{s_1(\overline{\xi})} \tilde{\rho}_{\max}^{s_2(\overline{\xi})-1}. \tag{49}$$

And for any $\overline{\xi} \in C_2$,

$$\mathbb{E} \sum_{\xi\in[\overline{\xi}]} H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}} \leq m(\kappa/N)^{g(\overline{\xi})} q^{2e(\overline{\xi})-2l} \rho_{\max}^{s_1(\overline{\xi})-1} \tilde{\rho}_{\max}^{s_2(\overline{\xi})}. \tag{50}$$

Proof. We consider a breadth-first search ordering of vertices in $\overline{\xi}$. Let $a = |E(\overline{\xi})|$ and $s = s_1 + s_2$, where the drop the dependence on $\overline{\xi}$ in $s_1(\overline{\xi})$ and $s_2(\overline{\xi})$ for convenience. We order the edges in $G(\overline{\xi})$ such that the edges $e_1,\dots,e_{s-1}$ form a spanning tree of $G(\overline{\xi})$, and let m_t be the number of times that e_t appear in $\overline{\xi}$ for $1 \leq t \leq s-1$. Let $\mathcal{I}_{s_1,s_2}$ be the set of injection maps from $[s_1]\times[s_2]$ to $[n]\times[m]$. Then for any $\overline{\xi} \in C_1$,

$$\mathbb{E} \sum_{\xi\in[\overline{\xi}]} H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}} = \sum_{\tau\in\mathcal{I}_{s_1,s_2}} \mathbb{E}H_{\tau(\overline{\xi}_0)\tau(\overline{\xi}_1)}\cdots H_{\tau(\overline{\xi}_{2l-1})\tau(\overline{\xi}_{2l})} \leq \sum_{\tau\in\mathcal{I}_{s_1,s_2}} \prod_{t=1}^{a} \mathbb{E}|H_{\tau(e_t)}^{m_t}|.$$

From Assumption 3, we can use the estimates

$$\max_{j\in V_2} \sum_{i\in V_1} \mathbb{E}|H_{ij}|^k \leq \frac{\tilde{\rho}_{\max}}{q^{k-2}}, \quad \max_{i\in V_1} \sum_{j\in V_2} \mathbb{E}|H_{ij}|^k \leq \frac{\rho_{\max}}{q^{k-2}}$$

for contribution from edges $e_1,\dots,e_{s-1}$, and the estimate $\max_{ij} \mathbb{E}|H_{ij}|^k \leq \frac{\kappa}{Nq^{k-2}}$ for $e_s,\dots,e_a$ to obtain

$$\begin{aligned}\mathbb{E} \sum_{\xi\in[\overline{\xi}]} H_{\xi_0\xi_1}H_{\xi_1\xi_2}\cdots H_{\xi_{2l-1}\xi_{2l}} &\leq \prod_{t=s}^{a} \frac{\kappa}{Nq^{m_t-2}} \sum_{\tau\in\mathcal{I}_{s_1,s_2}} \prod_{t=1}^{s-1} \mathbb{E}|H_{\tau(e_t)}^{m_t}| \\ &\leq \left(\prod_{t=s}^{a} \frac{\kappa}{Nq^{m_t-2}}\right) n\rho_{\max}^{s_1}\tilde{\rho}_{\max}^{s_2-1} \frac{1}{q^{m_1-2}}\cdots\frac{1}{q^{m_{s-1}-2}} \end{aligned} \tag{51}$$

$$= nq^{2a-\sum_{t=1}^{a} m_t} \left(\frac{\kappa}{N}\right)^{a-s+1} \rho_{\max}^{s_1} \tilde{\rho}_{\max}^{s_2-1} = n(\kappa/N)^g q^{2a-2l} \rho_{\max}^{s_1} \tilde{\rho}_{\max}^{s_2-1},$$

where $g = a - v + 1$ is the genus of $G(\overline{\xi})$. The factor n in (51) comes from the processing of pruning trees, and the root of the spanning tree has at most n choices of labeling since $\overline{\xi} \in C_1$. Then (49) holds. (50) follows similarly. □

We can further simplify the upper bound on $\mathbb{E}\mathrm{Tr}[B^l(B^l)^*]$ by counting the contributions from normal graphs defined below.

Definition 5.3 (Normal graph). Denote a path as $w = w_0 w_{01} w_1 w_{12} \cdots w_{l-1,l} w_l$ in a multigraph G, where $w_0, \dots, w_l$ are the vertices and $w_{i,i+1}, 0 \le i \le l-1$ are the edges visited by the path w. We say w in G is *normal* if

- $V(G) = [s]$ where $s = |V(G)|$;
- the vertices in $V(G)$ are visited in increasing order by w.

Each equivalent class of C_i has a unique representative ξ such that ξ is normal in G_ξ. For $i = 1, 2$, denote $C_{0,i} := \{\xi \in C_i : \xi \text{ is normal in } G_\xi\}$. From (48) and Lemma 5.2, we obtain

$$\begin{aligned} \mathbb{E}\mathrm{Tr}[B^l(B^l)^*] \le mn &\sum_{\xi \in C_{0,1}} (\kappa/N)^{g(\xi)} q^{2e(\xi)-2l} \rho_{\max}^{s_1(\xi)} \tilde{\rho}_{\max}^{s_2(\xi)-1} \\ &+ mn \sum_{\xi \in C_{0,2}} (\kappa/N)^{g(\xi)} q^{2e(\xi)-2l} \rho_{\max}^{s_1(\xi)-1} \tilde{\rho}_{\max}^{s_2(\xi)} = S_1 + S_2. \end{aligned} \tag{52}$$

From now on, we only treat S_1, and S_2 can be estimated similarly. We now introduce a parametrization of $C_{0,1}$ following (Benaych-Georges, Bordenave and Knowles, 2020).

Definition 5.4 (Equivalence class). Let G be a graph and $\mathcal{V} \subset V(G)$. Define

$$\mathcal{I}_{\mathcal{V}}(G) = \{v \in V(G) \setminus \mathcal{V} : \deg(v) = 2\}.$$

Let $\Sigma_{\mathcal{V}}(G)$ be the set of paths $w = w_0 \cdots w_l$ in G such that $w_1, \dots, w_{l-1}$ are pairwise distinct and belong to $\mathcal{I}_{\mathcal{V}}(G)$ and $w_0, w_l \notin \mathcal{I}_{\mathcal{V}}(G)$. We define an equivalence relation on $\Sigma_{\mathcal{V}}(G)$ such that the path $w_0 \dots w_l$ and the reverse ordered path $w_l \cdots w_1$ are equivalent. Denote $\Sigma'_{\mathcal{V}}(G) = \{[w] : w \in \Sigma_{\mathcal{V}}(G)\}$ the set of equivalence classes.

We can construct a multigraph $\hat{G}_\xi$ from G_ξ by replacing every $[w] \in \Sigma'_{\xi_0,\xi_l}(G)$ with an edge in $E(\hat{G}_\xi)$ as follows.

Definition 5.5 (Multigraph $\hat{G}_\xi$ from G_ξ). Let $\xi \in C_{0,1}$. Define $V(\hat{G}_\xi) = V(G_\xi) \setminus \mathcal{I}_{\xi_0,\xi_\ell}(G_\xi)$ and $E(\hat{G}_\xi) = \Sigma'_{\xi_0,\xi_l}(G)$. The endpoints of $[w]$ in $E(\hat{G}_\xi)$ are labeled w_0, w_l. Assign each edge $[w] \in E(\hat{G}_\xi)$ the weight $\hat{k}_w$, which is the length of the path w.

From (Benaych-Georges, Bordenave and Knowles, 2020), any $\xi \in C_{0,1}$ as a closed path $\xi_0 \xi_1 \cdots \xi_{2l}$ in G_ξ gives rise to a closed path $\hat{\xi} = \hat{\xi}_0 \hat{\xi}_{01} \hat{\xi}_1 \hat{\xi}_{12} \cdots \hat{\xi}_{r-1} \hat{\xi}_r$ on the multigraph $\hat{G}_\xi$. Now for any $\xi \in C_{0,1}$, we have constructed a triple $(\hat{G}_\xi, \hat{\xi}, \hat{k})$, where $\hat{G}_\xi$ is a multigraph, $\hat{\xi}$ is a closed path in $\hat{G}_\xi$. Set τ to be the unique increasing bijection from $V(\hat{G}_\xi)$ to $\{1, \dots, |V(\hat{G}_\xi)|\}$. Denote by $(U, \zeta, k) := (U(\xi), \zeta(\xi), k(\xi))$ the triple obtained from the triple $(\hat{G}_\xi, \hat{\xi}, \hat{k})$ by relabelling the vertices using τ. By Definition, $\zeta_0 = \tau(\xi_0) = 1$. We set $v = v(\xi) = \tau(\xi_l)$. Altogether, the construction above gives a map $\xi \mapsto (U, \zeta, k)$. The following lemma collects some properties of this map.

Lemma 5.6 (Lemma 5.3 in (Benaych-Georges, Bordenave and Knowles, 2020)). *The map $\xi \mapsto (U, \zeta, k)$ satisfies the following properties:*

1. *The map $\xi \mapsto (U, \zeta, k)$ is an injection on $C_{0,1}$.*
2. *$g(U) = g(G_\xi)$.*
3. *ζ is a closed path in the multigraph U and it is normal in U.*
4. *Every vertex of $V(U) \setminus \{1, v\}$ has degree at least three. The vertices 1 and v have degrees at least one.*
5. *$|E(G_\xi)| = \sum_{e \in E(U)} k_e$.*
6. *$m_e(\zeta) \geq 2$ for all $e \in E(U)$ and $2l = \sum_{e \in E(U)} m_e(\zeta) k_e$.*

For a given $\xi \in C_{0,1}$ with given $s_1(\xi), s_2(\xi)$, we know

$$s_1 + s_2 = |V(G_\xi)| = e(\xi) - g(U) + 1 \quad \text{and } |s_1 - s_2| \leq g(U), \tag{53}$$

where we use $g(\xi) = g(U)$ from Lemma 5.6. The second claim is due to the fact that since ℓ is odd, any imbalance between s_1, s_2 is from creating a new cycle in ξ. Moreover,

$$e(\xi) = \sum_{e \in E(U)} k_e, \quad 2l = \sum_{e \in E(U)} m_e(\zeta) k_e. \tag{54}$$

Since $k_e \geq 1, m_e(\zeta) \geq 2$, we obtain

$$2e(\xi) - 2l \leq 2|E(U)| - |\zeta|, \tag{55}$$

where $|\zeta|$ is the length of the closed path ζ. Equipped with the construction of the triple (U, ζ, k), we continue to estimate S_1 in (52). With (53),

$$\begin{aligned}
S_1 =& mn \sum_{\xi \in C_{0,1}} (\kappa/N)^{g(\xi)} q^{2e(\xi)-2l} \rho_{\max}^{s_1} \tilde{\rho}_{\max}^{s_2-1} \\
\leq& mn \sum_{\xi \in C_{0,1}} (\kappa/N)^{g(\xi)} q^{2e(\xi)-2l} (\rho_{\max}\tilde{\rho}_{\max})^{(s_1+s_2-1)/2} \max\{\rho_{\max}/\tilde{\rho}_{\max}, \tilde{\rho}_{\max}/\rho_{\max}\}^{|s_1-s_2+1|/2} \\
\leq& mn \sum_{\xi \in C_{0,1}} (\kappa/N)^{g(\xi)} q^{2e(\xi)-2l} \gamma^{(e(\xi)-g(\xi))/2} \left(\frac{1}{c}\right)^{(g(\xi)+1)/2} \\
=& mn\gamma^{(l-1)/2} \sum_{\xi \in C_{0,1}} (\kappa/N)^{g(\xi)} q^{2e(\xi)-2l} \gamma^{(e(\xi)-l)/2-g(\xi)}.
\end{aligned} \tag{56}$$

From (1) in Lemma 5.6, we can upper bound (56) by summing over (U, ζ, k) instead of ξ. Then with (55), and the assumption $q \geq \gamma^{-\frac{1}{4}}$, we find

$$S_1 \leq mn\gamma^{(l-1)/2} \sum_{(U,\zeta,k)} \left(\frac{\kappa}{Nc}\right)^{g(U)} \left(q\gamma^{\frac{1}{4}}\right)^{2|E(U)|-|\zeta|},$$

where (U,ζ,k) is obtained from all $\zeta \in C_{0,1}$. Since $\sum_{e\in E(U)} k_e m_e(\zeta) = 2l$ and $m_e(\zeta) \geq 2$, for a given (U,ζ), the number of choices for $k = (k_e)_{e\in E(U)}$ can be bounded by the number of k such that $\sum_{e\in E(U)} k_e = 2l$. For fixed (U,ζ), the number of choices for such k is bounded by

$$\binom{2l-1}{|E(U)|-1} \leq \left(\frac{6l}{|E(U)|}\right)^{|E(U)|}.$$

Therefore

$$S_1 \leq mn\gamma^{(l-1)/2} \sum_{(U,\zeta)} \left(\frac{6l}{|E(U)|}\right)^{|E(U)|} \left(\frac{\kappa}{Nc}\right)^{g(U)} \left(q\gamma^{\frac{1}{4}}\right)^{2|E(U)|-|\zeta|}. \tag{57}$$

From (Benaych-Georges, Bordenave and Knowles, 2020, Lemma 5.8), we have

$$|E(U)| \leq 3g(U)+1, \quad |V(U)| \leq 2g(U)+2. \tag{58}$$

Also, $|E(U)| \geq g(U) \vee 1$ and $g(U) \leq l$ by definition. Therefore, (57) can be further bounded by

$$S_1 \leq 6lq^2 mn\gamma^{l/2} \sum_{(U,\zeta)} \left(\frac{12l}{g(U)+1}\right)^{3g(U)} \left(\frac{\kappa}{Nc}\right)^{g(U)} \left(q\gamma^{\frac{1}{4}}\right)^{6g(U)-|\zeta|}. \tag{59}$$

(58) implies the number of pairs (U,ζ) such that U has genus g and ζ has length t is bounded by

$$(3g+1)^t (2g+2)^{3g+1}. \tag{60}$$

With (60) and (59), we find for some absolute constant $C > 0$,

$$\begin{aligned} S_1 &\leq 6lq^2 mn\gamma^{(l-1)/2} \sum_{g=0}^{l} \sum_{t=1}^{2l} (3g+1)^t (2g+2)^{3g+1} \left(\frac{24l}{2g+2}\right)^{3g} \left(\frac{\kappa}{Nc}\right)^{g} \left(q\gamma^{\frac{1}{4}}\right)^{6g-t} \\ &\leq Clq^2 mn\gamma^{(l-1)/2} \sum_{t=1}^{2l} (q\gamma^{\frac{1}{4}})^{-t} + Cl^2 q^2 mn\gamma^{(l-1)/2} \sum_{g=1}^{l} \left(\frac{Cl^3\kappa\gamma^{\frac{1}{2}}q^6}{N}\right)^{g} \sum_{t=1}^{2l} \left(\frac{4g}{q\gamma^{\frac{1}{4}}}\right)^{t}. \end{aligned} \tag{61}$$

Since $q \geq \gamma^{-\frac{1}{4}}$, the first term is bounded by $Cl^2 q^2 mn\gamma^{(l-1)/2}$. For the second term, using

$$\sum_{m=1}^{2l} x^m \leq 2l(1+x^{2l}),$$

it is bounded by

$$Cl^3 q^2 mn\gamma^{(l-1)/2} \sum_{g=1}^{l} \left(\frac{Cl^3\kappa\gamma^{\frac{1}{2}}q^6}{N}\right)^{g} + Cl^3 q^2 mn\gamma^{(l-1)/2} \sum_{g=1}^{l} \left(\frac{Cl^3\kappa\gamma^{\frac{1}{2}}q^6}{N}\right)^{g} \left(\frac{4g}{q\gamma^{\frac{1}{4}}}\right)^{2l}. \tag{62}$$

From the assumption (43), we obtain

$$N \geq \left(\frac{\kappa\gamma^{\frac{1}{2}}q^6 l^3}{c_0^3}\right)^{1/(1-3\delta)}, \quad N^{3\delta} \leq \frac{c_0^3 N}{\kappa\gamma^{\frac{1}{2}}q^6 l^3},$$

$$l \le c_0 \delta q \log N = \frac{c_0 q}{3} \log(N^{3\delta}) \le \frac{c_0 q}{3} \log\left(\frac{c_0^3 N}{\kappa \gamma^{\frac{1}{2}} q^6 l^3}\right).$$

By choosing c_0 small enough, we have the following inequalities:

$$N \ge 2Cl^3 \kappa \gamma^{\frac{1}{2}} q^6, \quad l \le \frac{1}{8} q \gamma^{\frac{1}{4}} \log\left(\frac{N}{Cl^3 \kappa \gamma^{\frac{1}{2}} q^6}\right). \tag{63}$$

The first term in (62) is bounded by $Cl^3 q^2 mn\gamma^{(l-1)/2}$. The second term can be written as

$$Cl^3 q^2 mn\gamma^{(l-1)/2} \sum_{g=1}^{l} \exp\left(-g \log\left(\frac{N}{Cl^3 \kappa \gamma^{\frac{1}{2}} q^6}\right) + 2l \log \frac{4g}{q\gamma^{\frac{1}{4}}}\right).$$

The argument in the exponential function is maximized at

$$g = \frac{2l}{\log\left(\frac{N}{Cl^3 \kappa \gamma^{\frac{1}{2}} q^6}\right)}.$$

From (63), this maximizer is reached for $g \le \frac{1}{4} q \gamma^{\frac{1}{4}}$, and the second term in (62) can be bounded by $Cl^4 q^2 mn\gamma^{(l-1)/2}$.

Therefore, with the bound on (61) and (62), we obtain $S_1 \le Cl^4 q^2 mn\gamma^{(l-1)/2}$. Repeating the same argument for S_2 yields the same upper bound. This completes the proof of Lemma 4.2.

6. Probabilistic bounds on the largest singular value

In this section, we first prove Theorem 2.4 for general rectangular random matrices, then specify the model parameters to prove Theorem 2.2 for sparse rectangular random matrices. The proof is based on the deterministic spectral relations between B and H in Section 3, and the bound on $\rho(B)$ in Section 4.

Proof of Theorem 2.4. From the assumption of Theorem 2.4, we have

$$|X_{ij}|^2 \le \frac{1}{q^2}, \quad \max_i \sum_j \mathbb{E}|X_{ij}|^2 \le \rho_{\max}, \quad \text{and} \quad \max_i \sum_j \mathbb{E}|X_{ij}|^4 \le \frac{\rho_{\max}}{q^2}.$$

Since $\sum_j |X_{ij}|^2$ is a sum of independent bounded random variables, applying Bennett's inequality (Boucheron, Lugosi and Massart, 2013, Theorem 2.9), we obtain, for $\delta > 0$,

$$\mathbb{P}\left(\sum_j |X_{ij}|^2 \ge \rho_{\max}(1+\delta)\right) \le \exp\left(-q^2 \rho_{\max} h(\delta)\right), \tag{64}$$

where $h(\delta) = (1+\delta)\log(1+\delta) - \delta$. Similarly,

$$\mathbb{P}\left(\sum_i |X_{ij}|^2 \ge \tilde{\rho}_{\max}(1+\delta)\right) \le \exp\left(-q^2 \tilde{\rho}_{\max} h(\delta)\right). \tag{65}$$

Then, by taking a union bound,

$$\mathbb{P}\left(\|X\|_{2,\infty} \geq \sqrt{\rho_{\max}}(1+\delta)\right) \leq \mathbb{P}\left(\|X\|_{2,\infty} \geq \sqrt{\rho_{\max}} \cdot \sqrt{1+\delta \vee \delta^2}\right)$$
$$= \mathbb{P}\left(\|X\|_{2,\infty}^2 \geq \rho_{\max}(1+\delta \vee \delta^2)\right) \leq n \exp(-q^2 \rho_{\max} h(\delta \vee \delta^2)),$$

and $\mathbb{P}\left(\|X^*\|_{2,\infty} \geq \sqrt{\tilde{\rho}_{\max}}(1+\delta)\right) \leq m \exp(-q^2 \tilde{\rho}_{\max} h(\delta \vee \delta^2))$. From Theorem 4.1,

$$\mathbb{P}(\rho(B) \geq \gamma^{\frac{1}{4}}(1+\delta)) \leq C\gamma^{-\frac{5}{6}} N^{3-c_1 q \log(1+\delta)}.$$

Therefore, conditioned on a high probability event, we have

$$\|H\|_{1,\infty} \leq q^{-1}, \quad \|H\|_{2,\infty} \leq 1+\delta, \quad \text{and} \quad \rho(B) \leq \gamma^{\frac{1}{4}}(1+\delta). \tag{66}$$

Now we apply the deterministic upper bound on $\sigma_{\max}(X)$ given in (28) conditioned on (66). If $\rho(B) \leq \gamma^{\frac{1}{4}} \|H\|_{2,\infty}$, then

$$\sigma_{\max}^2(X) \leq (1+\delta)^2(\sqrt{\gamma}+1)^2 + 48\gamma^{-\frac{5}{4}}(1+\delta)q^{-1} + 36\gamma^{-2}q^{-2} \leq \left((\sqrt{\gamma}+1) + C_1(\delta + \gamma^{-\frac{5}{4}}q^{-1})\right)^2$$

for some universal constant $C_1 > 0$. If instead $\gamma^{\frac{1}{4}}\|H\|_{2,\infty} < \rho(B) \leq \gamma^{\frac{1}{4}}(1+\delta)$, we find

$$\sigma_{\max}^2(X) \leq \|H\|_{2,\infty}^2 f\left(\frac{\gamma^{\frac{1}{4}}(1+\delta)}{\|H\|_{2,\infty}}\right) + 12\gamma^{-\frac{5}{4}} g\left(\frac{\gamma^{\frac{1}{4}}(1+\delta)}{\|H\|_{2,\infty}}\right)\|H\|_{2,\infty} q^{-1} + 36\gamma^{-2}q^{-2}$$
$$\leq \left((\sqrt{\gamma}+1) + C_1(\delta + \gamma^{-\frac{3}{2}}q^{-1})\right)^2.$$

Combining both cases, for any $\delta > 0$, with probability at least

$$1 - C\gamma^{-\frac{5}{6}} N^{3-c_1 q \log(1+\delta)} - 2Ne^{-\gamma q^2 h(\delta \vee \delta^2)}, \tag{67}$$

$$\sigma_{\max}(X) \leq \sqrt{\gamma} + 1 + C_1(\delta + \gamma^{-\frac{3}{2}}q^{-1}). \tag{68}$$

Next, we simplify the probability tail bound (67) by picking a specific δ. Let $K \geq 1$ and

$$\delta = \frac{K\eta}{\sqrt{1 \vee \log \eta}} \quad \text{where} \quad \eta = \frac{\sqrt{\log N}}{q}. \tag{69}$$

From the condition that $q \geq \gamma^{-1/4} \geq 1$, by considering $\eta \geq e, \eta < e$ separately, we have for $N \geq 2$, there exists a constant $c_2 = \frac{2}{e}$ such that

$$q\delta = \frac{K\sqrt{\log N}}{\sqrt{1 \vee \log \eta}} = \frac{K\eta q}{\sqrt{1 \vee \log \eta}} \geq c_2 K. \tag{70}$$

Therefore

$$q^{-1} \leq \frac{\delta}{c_2 K}. \tag{71}$$

Moreover, from (70), using the fact that $\frac{\log(1+x)}{x}$ is decreasing on $(0,\infty)$ and $q \geq \gamma^{-1/4} \geq 1$, we obtain

$$q\log(1+\delta) \geq c_2 K \cdot \frac{\log(1+c_2K/q)}{c_2K/q} \geq \log(1+c_2K). \tag{72}$$

Now we give a lower bound on $a_1 := \gamma q^2 h(\delta \vee \delta^2)$. If $\eta \leq e$, using the fact that $h(x) \geq c(x^2 \wedge x)$ for all $x \geq 0$ and some universal constant $c > 0$, we find

$$a_1 \geq c\gamma q^2\delta^2 = c\gamma K^2 \log N. \tag{73}$$

When $\eta \geq e$, using the inequality $h(x) \geq c(x^2 \wedge x)(1 \vee \log x)$ for all $x \geq 0$, we obtain for $\delta \geq e$,

$$h(\delta \vee \delta^2) \geq 2c\delta^2 \log(\delta).$$

Since from (69), $\log(\delta) \geq c' \log(\eta)$ for some absolute constant $c' > 0$, we obtain

$$a_1 \geq 2c\gamma q^2\delta^2 \log(\delta) \geq 2cc'\gamma q^2\delta^2 \log(\eta) \geq 2cc'\gamma K^2 \log N. \tag{74}$$

From (73) and (74), we conclude

$$a_1 \geq c\gamma K^2 \log N \tag{75}$$

for an absolute constant $c > 0$. With (71), (72), and (75), we can simplify (67) and (68) to conclude that with probability at least

$$1 - C\left(\gamma^{-5/6} N^{3-c_1\log(1+c_2K)} + N^{1-c'\gamma K^2}\right), \tag{76}$$

$\sigma_{\max}(X)$ satisfies

$$\begin{aligned}\sigma_{\max}(X) &\leq \sqrt{\gamma} + 1 + C_1(\delta + \gamma^{-\frac{3}{2}} q^{-1}) \leq \sqrt{\gamma} + 1 + C_1'(1 + K^{-1}\gamma^{-3/2})\delta \\ &= \sqrt{\gamma} + 1 + (K + \gamma^{-3/2}) \frac{C_1'\eta}{\sqrt{1 \vee \log \eta}}.\end{aligned}$$

This finishes the proof of (15). Now we turn to the expectation bound (14). Since entries in X are bounded by q^{-1}, from the concentration of operator norm in (Boucheron, Lugosi and Massart, 2013, Example 8.7),

$$\mathbb{P}(|\sigma_{\max}(X) - \mathbb{E}[\sigma_{\max}(X)]| \geq \delta) \leq 2\exp(-q^2\delta^2/4) \leq 2\exp(-c_2^2K^2/4), \tag{77}$$

where the last inequality is due to (70). From (76) and (77), we can take $K = \gamma^{-1/2}K_0$ for an large enough absolute constant $K_0 > 0$ such that

$$\mathbb{P}(|\sigma_{\max}(X) - \mathbb{E}[\sigma_{\max}(X)]| \leq \delta) + \mathbb{P}\left(\sigma_{\max}(X) \leq \sqrt{\gamma} + 1 + C_1'(1 + K_0^{-1}\gamma^{-1})\delta\right) > 1. \tag{78}$$

This implies the intersection of the two events

$$\{|\sigma_{\max}(X) - \mathbb{E}[\sigma_{\max}(X)]| \leq \delta\} \quad \text{and} \quad \left\{\sigma_{\max}(X) \leq \sqrt{\gamma} + 1 + C_1'(1 + K_0^{-1}\gamma^{-1})\delta\right\}$$

are non-empty. Hence for some absolute constants $C_2', C_2 > 0$,

$$\mathbb{E}[\sigma_{\max}(X)] \le \sqrt{\gamma} + 1 + C_2'\gamma^{-1}\delta = \sqrt{\gamma} + 1 + \frac{C_2\gamma^{-3/2}\eta}{\sqrt{1 \vee \log \eta}}. \tag{79}$$

This finishes the proof of Theorem 2.4 by using the assumption $\gamma = \Omega(1)$. □

Based on Theorem 2.4, we prove Theorem 2.2.

Proof of Theorem 2.2. Take $q = \sqrt{d}$ and $X = \frac{1}{\sqrt{d}}(A - \mathbb{E}A)$ in Theorem 2.4. We have

$$|X_{ij}| \le \frac{1}{\sqrt{d}}, \quad \mathbb{E}|X_{ij}|^2 \le \frac{\kappa}{N}, \quad \text{with} \quad \kappa = \frac{\max_{ij} p_{ij}}{d/N}.$$

Also

$$\max_{i\in[n]} \sum_{j\in[m]} \mathbb{E}|X_{ij}|^2 \le \frac{1}{d} \max_{i\in[m]} \sum_{j\in[n]} p_{ij} = \rho_{\max},$$

$$\max_{j\in[m]} \sum_{i\in[n]} \mathbb{E}|X_{ij}|^2 \le \frac{1}{d} \max_{i\in[n]} \sum_{j\in[m]} p_{ij} = \tilde{\rho}_{\max}.$$

Equation (6) follows from (15) by taking $K = C_3\gamma^{-3/2}$ for a sufficiently large constant C_3, and the probability estimate can be lower bounded by $1 - C\gamma^{-5/6}N^{-3}$ for some absolute constants $C > 0$. The expectation bound in (5) follows directly from (14). □

Remark 6.1. The probability bound in the statement of Theorem 2.2 can be improved to $1 - O(\gamma^{-5/6}N^{-a})$ for any constant $a > 0$ by taking a larger constant C_3 in the proof.

7. Probabilistic bounds on the smallest singular value

We now turn to the probabilistic lower bound on the smallest singular values for a general random matrix model.

Proof of Theorem 2.5. Using Bennett's inequality (Vershynin, 2018, Theorem 2.9.2), we obtain for any $j \in [m]$ and $t \ge 0$,

$$\mathbb{P}\left(\sum_i \left(|X_{ij}|^2 - \mathbb{E}|X_{ij}|^2\right) \le -t\right) = \mathbb{P}\left(\sum_i \left(-|X_{ij}|^2 + \mathbb{E}|X_{ij}|^2\right) \ge t\right) \le \exp\left(-\frac{1}{2}q^2 h(2t)\right).$$

Taking $t = \delta\tilde{\rho}_{\min}$ implies for any $j \in [m]$,

$$\mathbb{P}\left(\sum_i |X_{ij}|^2 \le \tilde{\rho}_{\min}(1-\delta)\right) \le \exp\left(-\frac{1}{4}q^2 h(2\delta\tilde{\rho}_{\min})\right). \tag{80}$$

Combined with (64) and (65), after a union bound, with probability at least

$$1 - m\exp\left(-\frac{1}{4}q^2 h(2\delta\tilde{\rho}_{\min})\right) - 2n\exp\left(-\gamma q^2 h(\delta)\right), \tag{81}$$

we have for all $j \in [m]$,

$$\tilde{\rho}_{\min}(1-\delta) \leq \sum_{i\in[n]} |X_{ij}|^2 \leq 1+\delta, \quad \text{and} \quad \max_{i\in[n]} \sum_{j\in[m]} |X_{ij}|^2 \leq \gamma(1+\delta).$$

Moreover, from the concentration of spectral norm of H given in (Benaych-Georges, Bordenave and Knowles, 2020, Equations (2.4) and (2.6)), for $q \geq \sqrt{\log n}$, with probability at least $1-2\exp(-q^2)$, $\|X\| \leq C_4$ for some absolute constant $C_4 > 0$. From Theorem 4.1, with probability $1 - C\gamma^{-\frac{5}{6}} n^{3-c_1 q \log(1+\sqrt{\delta})}$, $\rho(B) \leq \gamma^{\frac{1}{4}}(1+\sqrt{\delta})$. Note that for $x \in (0,2]$,

$$h(x) = (1+x)\log(1+x) - x \geq \frac{x^2}{2(1+x/3)} \geq \frac{3}{10}x^2$$

and $\log(1+\sqrt{\delta}) \geq \frac{\sqrt{\delta}}{1+\sqrt{\delta}} \geq \frac{1}{2}\sqrt{\delta}$.

With the assumption that $\tilde{\rho}_{\min} \geq \sqrt{\gamma}$, conditioned on all events above, from Lemma 3.5, we have with probability at least

$$1 - 3n\exp\left(-\frac{3}{10}\gamma q^2\delta^2\right) - 2\exp(-q^2) - C\gamma^{-\frac{5}{6}} n^{3-\frac{1}{2}q\sqrt{\delta}}, \tag{82}$$

$$\sigma^2_{\min}(X) \geq \frac{\sqrt{\gamma}-\gamma}{\sqrt{\gamma}+\delta}\left(\frac{\beta^2}{\beta^2+\delta^2}\tilde{\rho}_{\min} - \beta^2 - C_3\delta^2 - \delta\right)_+, \tag{83}$$

where $\beta = \gamma^{1/4}(1+\sqrt{\delta})$ and $C_3 = 4\gamma^{-\frac{1}{2}}(C_4 + \gamma^{-1}\delta)\frac{\sqrt{\gamma}+\delta}{\sqrt{\gamma}-\gamma}$.

Since $\delta \in (0,1]$ and $\tilde{\rho}_{\min} \leq 1$, (83) implies

$$\sigma^2_{\min}(X) \geq \frac{\sqrt{\gamma}-\gamma}{\sqrt{\gamma}+\delta}\left(\tilde{\rho}_{\min} - \sqrt{\gamma}\left(1+3\sqrt{\delta}\right) - \delta - \frac{C_5}{\gamma^2(1-\sqrt{\gamma})}\delta^2\right)_+, \tag{84}$$

where C_5 is an absolute constant. Using the Assumption 4, (84) implies (17).

Next, we consider the expectation bound. Repeating the proof of (Boucheron, Lugosi and Massart, 2013, Example 8.7), we have the following concentration inequality for $\sigma_{\min}(X)$:

$$\mathbb{P}(|\sigma_{\min}(X) - \mathbb{E}[\sigma_{\min}(X)]| \geq \delta^{1/4}) \leq 2\exp(-q^2\delta^{1/2}/4). \tag{85}$$

We can take $q \geq C_0 \max\left\{\delta^{-1/2}, \delta^{-1}\gamma^{-1/2}\sqrt{\log n}\right\}$ for a sufficiently large C_0 such that

$$\begin{aligned}&\mathbb{P}\left(|\sigma_{\min}(X) - \mathbb{E}[\sigma_{\min}(X)]| \leq \delta^{1/4}\right)\\ &+ \mathbb{P}\left(\sigma^2_{\min}(X) \geq \frac{\sqrt{\gamma}-\gamma}{\sqrt{\gamma}+\delta}\left(\tilde{\rho}_{\min} - \sqrt{\gamma}\left(1+3\sqrt{\delta}\right) - \delta - \frac{C_5}{\gamma^2(1-\sqrt{\gamma})}\delta^2\right)\right) > 1.\end{aligned}$$

This implies the intersection of the event $\left\{|\sigma_{\min}(X) - \mathbb{E}[\sigma_{\min}(X)]| \leq \delta^{1/4}\right\}$ and (84) is nonempty. Therefore, under Assumption 4, we obtain

$$\mathbb{E}[\sigma_{\min}(X)] \geq \sqrt{(1-\sqrt{\gamma})(\tilde{\rho}_{\min} - \sqrt{\gamma}) - O(\delta^{1/4})}.$$

This finished the proof of (16). □

Proof of Theorem 2.3. We take $q = \sqrt{d}$ in Theorem 2.5. Then, under Assumption 2, with probability at least $1 - O(n^{-3})$, (8) holds. (7) follows directly from Theorem 2.5. □

With Theorems 2.2 and 2.3, we prove Corollary 2.1.

Proof of Corollary 2.1. From the assumption (10), Theorem 2.2 implies with probability $1 - O(n^{-3})$,

$$\frac{1}{\sqrt{d}}\sigma_{\max}(A - \mathbb{E}A) \le 1 + \sqrt{y} + o(1). \tag{86}$$

From (10), $\tilde{\rho}_{\min} = 1 + o(1)$. Taking $\delta^2 = \frac{\log n}{d}$ in Theorem 2.3, we obtain with probability $1 - O(n^{-3})$,

$$\frac{1}{\sqrt{d}}\sigma_{\min}(A - \mathbb{E}A) \ge 1 - \sqrt{y} - o(1). \tag{87}$$

We can apply the proof of (Zhu, 2020, Corollary 4.3 and Theorem 8.2) to inhomogeneous Erdős-Rényi bipartite graphs. One can show in the same way that, almost surely, the empirical spectral distribution of $\frac{1}{d}(A - \mathbb{E}A)^\top(A - \mathbb{E}A)$ converges to the Marčenko-Pastur law supported on the interval $[(1 - \sqrt{y})^2, (1 + \sqrt{y})^2]$. Therefore almost surely,

$$\frac{1}{\sqrt{d}}\sigma_{\max}(A - \mathbb{E}A) \ge 1 + \sqrt{y} - o(1), \quad \frac{1}{\sqrt{d}}\sigma_{\min}(A - \mathbb{E}A) \le 1 - \sqrt{y} + o(1). \tag{88}$$

From (86), (87), and (88), the convergence results in (11) hold. □

Acknowledgements

We are grateful to Antti Knowles and Ramon van Handel for helpful discussions and to the latter for a detailed explanation of the work (Brailovskaya and van Handel, 2022). Y.Z. thanks Ludovic Stephan for a careful reading of this paper and many useful suggestions. We thank the editor and anonymous referees for helping us improve the presentation.

Funding

I.D. and Y.Z. acknowledge support from NSF DMS-1928930 during their participation in the program Universality and Integrability in Random Matrix Theory and Interacting Particle Systems hosted by the Mathematical Sciences Research Institute in Berkeley, California, during the Fall semester of 2021. I.D. is partially supported by NSF DMS-2154099. Y.Z. is partially supported by NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning. This work was done in part while Y.Z. was visiting the Simons Institute for the Theory of Computing in the Fall of 2022.

References

Alt, J., Ducatez, R. and Knowles, A. (2021). Extremal eigenvalues of critical Erdős-Rényi graphs. *Ann. Probab.* **49** 1347–1401. MR4255147 https://doi.org/10.1214/20-aop1483

Auffinger, A. and Tang, S. (2016). Extreme eigenvalues of sparse, heavy tailed random matrices. *Stochastic Process. Appl.* **126** 3310–3330. MR3549709 https://doi.org/10.1016/j.spa.2016.04.029

Avron, H., Druinsky, A. and Toledo, S. (2019). Spectral condition-number estimation of large sparse matrices. *Numer. Linear Algebra Appl.* **26** e2235, 19. MR3946070 https://doi.org/10.1002/nla.2235

Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

Bandeira, A.S., Boedihardjo, M.T. and van Handel, R. (2023). Matrix concentration inequalities and free probability. *Invent. Math.* **234** 419–487. MR4635836 https://doi.org/10.1007/s00222-023-01204-6

Bandeira, A.S. and van Handel, R. (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.* **44** 2479–2506. MR3531673 https://doi.org/10.1214/15-AOP1025

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

Basak, A. and Rudelson, M. (2021). Sharp transition of the invertibility of the adjacency matrices of sparse random graphs. *Probab. Theory Related Fields* **180** 233–308. MR4265022 https://doi.org/10.1007/s00440-021-01038-4

Bass, H. (1992). The Ihara-Selberg zeta function of a tree lattice. *Internat. J. Math.* **3** 717–797. MR1194071 https://doi.org/10.1142/S0129167X92000357

Benaych-Georges, F., Bordenave, C. and Knowles, A. (2019). Largest eigenvalues of sparse inhomogeneous Erdős-Rényi graphs. *Ann. Probab.* **47** 1653–1676. MR3945756 https://doi.org/10.1214/18-AOP1293

Benaych-Georges, F., Bordenave, C. and Knowles, A. (2020). Spectral radii of sparse random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2141–2161. MR4116720 https://doi.org/10.1214/19-AIHP1033

Bhojanapalli, S. and Jain, P. (2014). Universal matrix completion. In *International Conference on Machine Learning* 1881–1889. PMLR.

Bordenave, C. (2020). A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér. (4)* **53** 1393–1439. MR4203039 https://doi.org/10.24033/asens.2450

Bordenave, C., Coste, S. and Nadakuditi, R.R. (2023). Detection thresholds in very sparse matrix completion. *Found. Comput. Math.* **23** 1619–1743. MR4649432 https://doi.org/10.1007/s10208-022-09568-6

Bordenave, C., Lelarge, M. and Massoulié, L. (2018). Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *Ann. Probab.* **46** 1–71. MR3758726 https://doi.org/10.1214/16-AOP1142

Borodin, A., Corwin, I. and Guionnet, A., eds. (2019). *Random Matrices*. *IAS/Park City Mathematics Series* **26**. Providence, RI: Amer. Math. Soc.

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

Brailovskaya, T. and van Handel, R. (2022). Universality and sharp matrix concentration inequalities. Available at arXiv:2201.05142.

Brito, G., Dumitriu, I. and Harris, K.D. (2022). Spectral gap in random bipartite biregular graphs and applications. *Combin. Probab. Comput.* **31** 229–267. MR4390294 https://doi.org/10.1017/s0963548321000249

Cai, T.T., Han, R. and Zhang, A.R. (2022). On the non-asymptotic concentration of heteroskedastic Wishart-type matrix. *Electron. J. Probab.* **27** Paper No. 29, 40. MR4385832 https://doi.org/10.1214/22-ejp758

Chafaï, D. and Tikhomirov, K. (2018). On the convergence of the extremal eigenvalues of empirical covariance matrices with dependence. *Probab. Theory Related Fields* **170** 847–889. MR3773802 https://doi.org/10.1007/s00440-017-0778-9

Che, Z. and Lopatto, P. (2019). Universality of the least singular value for sparse random matrices. *Electron. J. Probab.* **24** Paper No. 9, 53. MR3916329 https://doi.org/10.1214/19-EJP269

Cook, N. (2018). Lower bounds for the smallest singular value of structured random matrices. *Ann. Probab.* **46** 3442–3500. MR3857860 https://doi.org/10.1214/17-AOP1251

Coste, S. and Zhu, Y. (2021). Eigenvalues of the non-backtracking operator detached from the bulk. *Random Matrices Theory Appl.* **10** Paper No. 2150028, 21. MR4302283 https://doi.org/10.1142/S2010326321500283

Deshpande, Y., Montanari, A., O'Donnell, R., Schramm, T. and Sen, S. (2019). The threshold for SDP-refutation of random regular NAE-3SAT. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2305–2321. Philadelphia, PA: SIAM. MR3909609 https://doi.org/10.1137/1.9781611975482.140

Dumitriu, I. and Zhu, Y. (2021). Spectra of random regular hypergraphs. *Electron. J. Combin.* **28** Paper No. 3.36, 25. MR4295659 https://doi.org/10.37236/8741

Feldheim, O.N. and Sodin, S. (2010). A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.* **20** 88–123. MR2647136 https://doi.org/10.1007/s00039-010-0055-x

Florescu, L. and Perkins, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory* 943–959. PMLR.

Gordon, Y. (1985). Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** 265–289. MR0800188 https://doi.org/10.1007/BF02759761

Götze, F. and Tikhomirov, A. (2023). On the largest and the smallest singular value of sparse rectangular random matrices. *Electron. J. Probab.* **28** Paper No. 27, 18. MR4548419 https://doi.org/10.1214/23-ejp919

Guédon, O., Litvak, A.E. and Tatarko, K. (2020). Random polytopes obtained by matrices with heavy-tailed entries. *Commun. Contemp. Math.* **22** 1950027, 28. MR4106820 https://doi.org/10.1142/S0219199719500275

Guédon, O., Litvak, A.E., Pajor, A. and Tomczak-Jaegermann, N. (2017). On the interval of fluctuation of the singular values of random matrices. *J. Eur. Math. Soc. (JEMS)* **19** 1469–1505. MR3635358 https://doi.org/10.4171/JEMS/697

Guruswami, V., Manohar, P. and Mosheiff, J. (2022). ℓ_p-spread and restricted isometry properties of sparse random matrices. In *37th Computational Complexity Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **234** Art. No. 7, 17. Wadern: Schloss Dagstuhl. Leibniz-Zent. Inform. MR4481615 https://doi.org/10.4230/lipics.ccc.2022.7

Han, Q. (2022). Exact spectral norm error of sample covariance. Available at arXiv:2207.13594.

Heiny, J. and Mikosch, T. (2018). Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices. *Stochastic Process. Appl.* **128** 2779–2815. MR3811704 https://doi.org/10.1016/j.spa.2017.10.002

Janwa, H. and Lal, A.K. (2003). On Tanner codes: Minimum distance and decoding. *Appl. Algebra Engrg. Comm. Comput.* **13** 335–347. MR1959169 https://doi.org/10.1007/s00200-003-0098-4

Koltchinskii, V. and Mendelson, S. (2015). Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN* **23** 12991–13008. MR3431642 https://doi.org/10.1093/imrn/rnv096

Kotani, M. and Sunada, T. (2000). Zeta functions of finite graphs. *J. Math. Sci. Univ. Tokyo* **7** 7–25. MR1749978

Krivelevich, M. and Sudakov, B. (2003). The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** 61–72. MR1967486 https://doi.org/10.1017/S0963548302005424

Latała, R., van Handel, R. and Youssef, P. (2018). The dimension-free structure of nonhomogeneous random matrices. *Invent. Math.* **214** 1031–1080. MR3878726 https://doi.org/10.1007/s00222-018-0817-x

Le, C.M., Levina, E. and Vershynin, R. (2017). Concentration and regularization of random graphs. *Random Structures Algorithms* **51** 538–561. MR3689343 https://doi.org/10.1002/rsa.20713

Litvak, A.E. and Rivasplata, O. (2012). Smallest singular value of sparse random matrices. *Studia Math.* **212** 195–218. MR3009072 https://doi.org/10.4064/sm212-3-1

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

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

Rudelson, M. and Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707–1739. MR2569075 https://doi.org/10.1002/cpa.20294

Rudelson, M. and Vershynin, R. (2010). Non-asymptotic theory of random matrices: Extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III* 1576–1602. New Delhi: Hindustan Book Agency. MR2827856

Stephan, L. and Massoulié, L. (2022). Non-backtracking spectra of weighted inhomogeneous random graphs. *Math. Stat. Learn.* **5** 201–271. MR4526300 https://doi.org/10.4171/msl/34

Stephan, L. and Zhu, Y. (2022). Sparse random hypergraphs: Non-backtracking spectra and community detection. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS 2022* 567–575. Los Alamitos, CA: IEEE Computer Soc. MR4537236

Tikhomirov, K. (2015). The limit of the smallest singular value of random matrices with i.i.d. entries. *Adv. Math.* **284** 1–20. MR3391069 https://doi.org/10.1016/j.aim.2015.07.020

Tikhomirov, K.E. (2016). The smallest singular value of random rectangular matrices with no moment assumptions on entries. *Israel J. Math.* **212** 289–314. MR3504328 https://doi.org/10.1007/s11856-016-1287-8

Tikhomirov, K. (2018). Sample covariance matrices of heavy-tailed distributions. *Int. Math. Res. Not. IMRN* **20** 6254–6289. MR3872323 https://doi.org/10.1093/imrn/rnx067

Tikhomirov, K. and Youssef, P. (2021). Outliers in spectrum of sparse Wigner matrices. *Random Structures Algorithms* **58** 517–605. MR4234995 https://doi.org/10.1002/rsa.20982

Tropp, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

van Handel, R. (2017). On the spectral norm of Gaussian random matrices. *Trans. Amer. Math. Soc.* **369** 8161–8178. MR3695857 https://doi.org/10.1090/tran/6922

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge: Cambridge Univ. Press. MR2963170

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge: Cambridge Univ. Press. MR3837109 https://doi.org/10.1017/9781108231596

Wan, Y. and Meila, M. (2015). A class of network models recoverable by spectral clustering. *Adv. Neural Inf. Process* **28**.

Wang, K. and Wood, P.M. (2023). Limiting empirical spectral distribution for the non-backtracking matrix of an Erdős-Rényi random graph. *Combin. Probab. Comput.* **32** 956–973. MR4653733 https://doi.org/10.1017/s096354832300024x

Watanabe, Y. and Fukumizu, K. (2009). Graph Zeta function in the Bethe free energy and loopy belief propagation. *Adv. Neural Inf. Process* **22**.

Zhivotovskiy, N. (2021). Dimension-free bounds for sums of independent matrices and simple tensors via the variational principle. Available at arXiv:2108.08198.

Zhou, Z. and Amini, A.A. (2019). Analysis of spectral clustering algorithms for community detection: The general bipartite setting. *J. Mach. Learn. Res.* **20** Paper No. 47, 47. MR3948087

Zhu, Y. (2020). A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures Algorithms* **56** 251–279. MR4052853 https://doi.org/10.1002/rsa.20894

Zhu, Y. (2023). On the second eigenvalue of random bipartite biregular graphs. *J. Theoret. Probab.* **36** 1269–1303. MR4591874 https://doi.org/10.1007/s10959-022-01190-0

Received November 2022 and revised November 2023