The Annals of Applied Probability
2018, Vol. 28, No. 3, 1679–1738
https://doi.org/10.1214/17-AAP1341

A NECESSARY AND SUFFICIENT CONDITION FOR EDGE UNIVERSALITY AT THE LARGEST SINGULAR VALUES OF COVARIANCE MATRICES

BY XIUCAI DING[1] AND FAN YANG[2]

University of Toronto and University of Wisconsin–Madison

In this paper, we prove a necessary and sufficient condition for the edge universality of sample covariance matrices with general population. We consider sample covariance matrices of the form $\mathcal{Q} = TX(TX)^*$, where X is an $M_2 \times N$ random matrix with $X_{ij} = N^{-1/2}q_{ij}$ such that q_{ij} are *i.i.d.* random variables with zero mean and unit variance, and T is an $M_1 \times M_2$ deterministic matrix such that T^*T is diagonal. We study the asymptotic behavior of the largest eigenvalues of $\mathcal{Q}$ when $M := \min\{M_1, M_2\}$ and N tend to infinity with $\lim_{N\to\infty} N/M = d \in (0, \infty)$. We prove that the Tracy–Widom law holds for the largest eigenvalue of $\mathcal{Q}$ if and only if $\lim_{s\to\infty} s^4\mathbb{P}(|q_{ij}| \geq s) = 0$ under mild assumptions of T. The necessity and sufficiency of this condition for the edge universality was first proved for Wigner matrices by Lee and Yin [*Duke Math. J.* **163** (2014) 117–173].

CONTENTS

Received March 2017; revised August 2017.
[1]Supported in part by NSERC of Canada (Jeremy Quastel).
[2]Supported in part by NSF Career Grant (Jun Yin) DMS-1552192.
MSC2010 subject classifications. 15B52, 82B44.
Key words and phrases. Sample covariance matrices, edge universality, Tracy–Widom distribution, Marchenko–Pastur law.

1. Introduction. Sample covariance matrices are fundamental objects in modern multivariate statistics. In the classical setting [2], for an $M \times N$ sample matrix X, people focus on the asymptotic properties of XX^* when M is fixed and N goes to infinity. In this case, the central limit theorem and law of large numbers can be applied to the statistical inference procedure. However, the advance of technology has led to high dimensional data such that M is comparable to or even larger than N [26, 27]. This high dimensionality cannot be handled with the classical multivariate statistical theory.

An important topic in the statistical study of sample covariance matrices is the distribution of the largest eigenvalues, which have been playing essential roles in analyzing the data matrices. For example, they are of great interest to principal component analysis (PCA) [28], which is a standard technique for dimensionality reduction and provides a way to identify patterns from real data. Also, the largest eigenvalues are commonly used in hypothesis testing, such as the well-known Roy's largest root test [37]. For a detailed review, one can refer to [27, 41, 54].

In this paper, we study the largest eigenvalues of sample covariance matrices with comparable dimensions and general population (i.e., the expectation of the sample covariance matrices are nonscalar matrices). More specifically, we consider sample covariance matrices of the form $\mathcal{Q} = TX(TX)^*$, where the sample $X = (x_{ij})$ is an $M_2 \times N$ random matrix with *i.i.d.* entries such that $\mathbb{E}x_{11} = 0$ and $\mathbb{E}|x_{11}|^2 = N^{-1}$, and T is an $M_1 \times M_2$ deterministic matrix. On dimensionality, we assume that $N/M \to d$ as $N \to \infty$, where $M := \min\{M_1, M_2\}$. In the last decade, random matrix theory has been proved to be one of the most powerful tools in dealing with this kind of large dimensional random matrices. It is well known that the empirical spectral distribution (ESD) of $\mathcal{Q}$ converges to the (deformed) Marchenko–Pastur (MP) law [34], whose rightmost edge λ_r gives the asymptotic location of the largest eigenvalue. Furthermore, it was proved in a series of papers that under a proper $N^{2/3}$ scaling, the distribution of the largest eigenvalue λ_1 of $\mathcal{Q}$ around λ_r converges to the Tracy–Widom distribution [49, 50], which arises as the limiting distribution of the rescaled largest eigenvalues of the Gaussian orthogonal ensemble (GOE). This result is commonly referred to as the *edge universality*, in the sense that it is independent of the detailed distribution of the entries of X. The Tracy–Widom distribution of $(\lambda_1 - \lambda_r)$ was first proved for $\mathcal{Q}$ with X consisting of *i.i.d.* centered real or complex Gaussian random entries (i.e., X is a Wishart matrix) and with trivial population (i.e., $T = I$) [26]. The edge universality in the $T = I$ case was later proved for all random matrices X whose entries satisfy arbitrary sub-exponential distribution [42, 43]. When T is a (nonscalar) diagonal matrix, the Tracy–Widom distribution was first proved for Wishart matrix X in [14] (nonsingular T case) and [38] (singular T case). Later the edge universality in the case with diagonal T was proved in [6, 32] for random matrices X with sub-exponentially distributed entries. The most general case with rectangular and nondiagonal T is considered in [31], where the edge universality was proved for

X with sub-exponentially distributed entries. Similar results have been proved for Wigner matrices in [18, 33, 48].

In this paper, we prove a necessary and sufficient condition for the edge universality of sample covariance matrices with general population. Briefly speaking, we will prove the following result.

If T^*T is diagonal and satisfies some mild assumptions, then the rescaled eigenvalues $N^{\frac{2}{3}}(\lambda_1(\mathcal{Q}) - \lambda_r)$ converges weakly to the Tracy–Widom distribution if and only if the entries of X satisfy the following tail condition:

$$\lim_{s\to\infty} s^4\mathbb{P}\big(|q_{11}| \geq s\big) = 0. \tag{1.1}$$

For a precise statement of the result, one can refer to Theorem 2.7. Note that under the assumption that T^*T is diagonal, the matrix $\mathcal{Q}$ is equivalent (in terms of eigenvalues) to a sample covariance matrix with diagonal T. Hence our result is basically an improvement of the ones in [6, 32]. The condition (1.1) provides a simple criterion for the edge universality of sample covariance matrices without assuming any other properties of matrix entries.

Note that the condition (1.1) is slightly weaker than the finite fourth moment condition for $\sqrt{N}x_{11}$. In the null case with $T = I$, it was proved previously in [55] that $\lambda_1 \to \lambda_r$ almost surely if the fourth moment exists. Later the finite fourth moment condition is proved to be also necessary for the almost sure convergence of λ_1 in [5]. Our theorem, however, shows that the existence of finite fourth moment is not necessary for the Tracy–Widom fluctuation. In fact, one can easily construct random variables that satisfies condition (1.1) but has infinite fourth moment. For example, we can use the following probability density function with $x^{-5}(\log x)^{-1}$ tail:

$$\rho(x) = \frac{e^4(4\log x + 1)}{x^5(\log x)^2}\mathbf{1}_{\{x>e\}}.$$

Then in this case λ_1 does not converge to λ_r almost surely, but $N^{2/3}(\lambda_1 - \lambda_r)$ still converges weakly to the Tracy–Widom distribution. On the other hand, Silverstein proved that $\lambda_1 \to \lambda_r$ in probability under the condition (1.1) [44]. So our result can be also regarded as an improvement of the one in [44].

The necessity and sufficiency of the condition (1.1) for the edge universality of Wigner matrix ensembles has been proved by Lee and Yin in [33]. The main idea of our proof is similar to theirs. For the necessity part, the key observation is that if the condition (1.1) does not hold, then X has a large entry with nonzero probability. As a result, the largest eigenvalue of $\mathcal{Q}$ can be larger than C with nonzero probability for any fixed constant $C > \lambda_r$, that is, $\lambda_1 \not\to \lambda_r$ in probability. The sufficiency part is more delicate. A key observation of [33] is that if we introduce a "cutoff" on the matrix elements of X at the level $N^{-\varepsilon}$, then the matrix with cutoff can well approximate the original matrix in terms of the largest singular value if and only if the condition (1.1) holds. Thus our problem can be reduced to proving the edge universality of the sample covariance matrices whose entries have size (or support) $\le N^{-\varepsilon}$. In [6, 32], the edge universality for sample covariance matrices has been proved assuming a sub-exponential decay of the x_{ij} entries; see Lemma 3.13. Under their assumptions, the typical size of the x_{ij} entries are of order $O(N^{-1/2+\varepsilon})$ for any constant $\varepsilon > 0$. Then a major part of this paper is devoted to extending the "small" support case to the "large" support case where the x_{ij} entries have size $N^{-\varepsilon}$. This can be accomplished with a Green function comparison method, which has been applied successfully in proving the universality of covariance matrices [42, 43]. A technical difficulty is that the change of the sample covariance matrix $\mathcal{Q}$ is nonlinear in terms of the change of X. To handle this, we use the self-adjoint linearization trick; see Definition 3.4.

This paper is organized as follows. In Section 2, we define the deformed Marchenko–Pastur law and its rightmost edge (i.e., the soft edge) λ_r, and then state the main theorem—Theorem 2.7—of this paper. In Section 3, we introduce the notation and collect some tools that will be used to prove the main theorem. In Section 4, we prove Theorem 2.7. In Section 5 and Section 6, we prove some key lemmas and theorems that are used in the proof of main result. In particular, the Green function comparison is performed in Section 6. In Appendix A, we prove the local law of sample covariance matrices with support $N^{-\phi}$ for some constant $\phi > 0$.

REMARK 1.1. In this paper, we do not consider the edge universality at the leftmost edge for the smallest eigenvalues. It will be studied elsewhere. Let λ_l be the leftmost edge of the deformed Marchenko-Pastur law. It is worth mentioning that the condition (1.1) can be shown to be sufficient for the edge universality at λ_l if $\lambda_l \not\to 0$ as $N \to \infty$. However, it seems that (1.1) is not necessary. So far, there is no conjecture about the necessary and sufficient condition for the edge universality at the leftmost edge.

Conventions. All quantities that are not explicitly constant may depend on N, and we usually omit N from our notation. We use C to denote a generic large positive constant, whose value may change from one line to the next. Similarly, we use ε, τ and c to denote generic small positive constants. For two quantities a_N and b_N depending on N, the notation $a_N = O(b_N)$ means that $|a_N| \le C|b_N|$ for some constant $C > 0$, and $a_N = o(b_N)$ means that $|a_N| \le c_N|b_N|$ for some positive sequence $\{c_N\}$ with $c_N \to 0$ as $N \to \infty$. We also use the notation $a_N \sim b_N$ if $a_N = O(b_N)$ and $b_N = O(a_N)$. For a matrix A, we use $\|A\| := \|A\|_{l^2 \to l^2}$ to denote the operator norm and $\|A\|_{\mathrm{HS}}$ the Hilbert–Schmidt norm; for a vector $\mathbf{v} = (v_i)_{i=1}^n$, $\|\mathbf{v}\| \equiv \|\mathbf{v}\|_2$ stands for the Euclidean norm, while $|\mathbf{v}| \equiv \|\mathbf{v}\|_1$ stands for the l^1-norm. In this paper, we often write an $n \times n$ identity matrix $I_{n\times n}$ as 1 or I without causing any confusions. If two random variables X and Y have the same distribution, we write $X \stackrel{d}{=} Y$.

2. Definitions and main result.

2.1. *Sample covariance matrices with general populations.* We consider the $M_1 \times M_1$ sample covariance matrix $\mathcal{Q}_1 := TX(TX)^*$, where T is a deterministic $M_1 \times M_2$ matrix and X is a random $M_2 \times N$ matrix. We assume $X = (x_{ij})$ has entries $x_{ij} = N^{-1/2} q_{ij}$, $1 \le i \le M_2$ and $1 \le j \le N$, where q_{ij} are *i.i.d.* random variables satisfying

$$\mathbb{E} q_{11} = 0, \qquad \mathbb{E}|q_{11}|^2 = 1. \tag{2.1}$$

In this paper, we regard N as the fundamental (large) parameter and $M_{1,2} \equiv M_{1,2}(N)$ as depending on N. We define $M := \min\{M_1, M_2\}$ and the aspect ratio $d_N := N/M$. Moreover, we assume that

$$d_N \to d \in (0, \infty) \qquad \text{as } N \to \infty. \tag{2.2}$$

For simplicity of notation, we will almost always abbreviate d_N as d in this paper. We denote the eigenvalues of $\mathcal{Q}_1$ in decreasing order by $\lambda_1(\mathcal{Q}_1) \ge \cdots \ge \lambda_{M_1}(\mathcal{Q}_1)$. We will also need the $N \times N$ matrix $\mathcal{Q}_2 := (TX)^* TX$ and denote its eigenvalues by $\lambda_1(\mathcal{Q}_2) \ge \cdots \ge \lambda_N(\mathcal{Q}_2)$. Since $\mathcal{Q}_1$ and $\mathcal{Q}_2$ share the same nonzero eigenvalues, we will for simplicity write λ_j, $1 \le j \le \min\{N, M_1\}$, to denote the jth eigenvalue of both $\mathcal{Q}_1$ and $\mathcal{Q}_2$ without causing any confusion.

We assume that T^*T is diagonal. In other words, T has a singular decomposition $T = U\bar{D}$, where U is an $M_1 \times M_1$ unitary matrix and $\bar{D}$ is an $M_1 \times M_2$ rectangular diagonal matrix. Then it is equivalent to study the eigenvalues of $\bar{D}X(\bar{D}X)^*$. When $M_1 \le M_2$ (i.e., $M = M_1$), we can write $\bar{D} = (D, 0)$ where D is an $M \times M$ diagonal matrix such that $D_{11} \ge \cdots \ge D_{MM}$. Hence we have $\bar{D}X = D\tilde{X}$, where $\tilde{X}$ is the upper $M \times N$ block of X with *i.i.d.* entries x_{ij}, $1 \le i \le M$ and $1 \le j \le N$. On the other hand, when $M_1 \ge M_2$ (i.e., $M = M_2$), we can write $\bar{D} = \binom{D}{0}$ where D is an $M \times M$ diagonal matrix as above. Then $\bar{D}X = \binom{DX}{0}$, which shares the same nonzero singular values with DX. The above discussions show that we can make the following stronger assumption on T:

$$M_1 = M_2 = M \quad \text{and} \quad T \equiv D = \operatorname{diag}\big(\sigma_1^{1/2}, \sigma_2^{1/2}, \ldots, \sigma_M^{1/2}\big), \tag{2.3}$$

where

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_M \ge 0.$$

Under the above assumption, the population covariance matrix of $\mathcal{Q}_1$ is defined as

$$\Sigma := \mathbb{E}\mathcal{Q}_1 = D^2 = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_M). \tag{2.4}$$

We denote the empirical spectral density of Σ by

$$\pi_N := \frac{1}{M} \sum_{i=1}^{M} \delta_{\sigma_i}. \tag{2.5}$$

We assume that there exists a small constant $\tau > 0$ such that

$$\sigma_1 \le \tau^{-1} \quad \text{and} \quad \pi_N\big([0,\tau]\big) \le 1-\tau \qquad \text{for all } N. \tag{2.6}$$

Note the first condition means that the operator norm of Σ is bounded by τ^{-1}, and the second condition means that the spectrum of Σ cannot concentrate at zero.

For definiteness, in this paper we focus on the real case, that is, the random variable q_{11} is real. However, we remark that our proof can be applied to the complex case after minor modifications if we assume in addition that $\operatorname{Re} q_{11}$ and $\operatorname{Im} q_{11}$ are independent centered random variables with variance $1/2$.

We summarize our basic assumptions here for future reference.

ASSUMPTION 2.1. We assume that X is an $M \times N$ random matrix with real *i.i.d.* entries satisfying (2.1) and (2.2). We assume that T is an $M \times M$ deterministic diagonal matrix satisfying (2.3) and (2.6).

2.2. *Deformed Marchenko–Pastur law.* In this paper, we will study the eigenvalue statistics of $\mathcal{Q}_{1,2}$ through their *Green functions* or *resolvents*.

DEFINITION 2.2 (Green functions). For $z = E + i\eta \in \mathbb{C}_+$, where $\mathbb{C}_+$ is the upper half complex plane, we define the Green functions for $\mathcal{Q}_{1,2}$ as

$$\mathcal{G}_1(z) := \big(DXX^*D^* - z\big)^{-1}, \qquad \mathcal{G}_2(z) := \big(X^*D^*DX - z\big)^{-1}. \tag{2.7}$$

We denote the empirical spectral densities (ESD) of $\mathcal{Q}_{1,2}$ as

$$\rho_1^{(N)} := \frac{1}{M}\sum_{i=1}^{M} \delta_{\lambda_i(\mathcal{Q}_1)}, \qquad \rho_2^{(N)} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(\mathcal{Q}_2)}.$$

Then the Stieltjes' transforms of $\rho_{1,2}$ are given by

$$m_1^{(N)}(z) := \int \frac{1}{x-z}\rho_1^{(N)}(dx) = \frac{1}{M}\operatorname{Tr}\mathcal{G}_1(z),$$
$$m_2^{(N)}(z) := \int \frac{1}{x-z}\rho_2^{(N)}(dx) = \frac{1}{N}\operatorname{Tr}\mathcal{G}_2(z).$$

Throughout the rest of this paper, we omit the super-index N from our notation.

REMARK 2.3. Since the nonzero eigenvalues of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are identical, and $\mathcal{Q}_1$ has $M - N$ more (or $N - M$ less) zero eigenvalues, we have

$$\rho_1 = \rho_2 d + (1-d)\delta_0 \tag{2.8}$$

and

$$m_1(z) = -\frac{1-d}{z} + d m_2(z). \tag{2.9}$$

In the case $D = I_{M\times M}$, it is well known that the ESD of X^*X, ρ_2, converges weakly to the Marchenko–Pastur (MP) law [34]:

$$\rho_{\mathrm{MP}}(x)\,dx := \frac{1}{2\pi}\frac{\sqrt{[(\lambda_+ - x)(x-\lambda_-)]_+}}{x}\,dx, \tag{2.10}$$

where $\lambda_\pm = (1 \pm d^{-1/2})^2$. Moreover, $m_2(z)$ converges to the Stieltjes' transform $m_{\mathrm{MP}}(z)$ of $\rho_{\mathrm{MP}}(z)$, which can be computed explicitly as

$$m_{\mathrm{MP}}(z) = \frac{d^{-1} - 1 - z + i\sqrt{(\lambda_+ - z)(z-\lambda_-)}}{2z}, \qquad z \in \mathbb{C}_+. \tag{2.11}$$

Moreover, one can verify that $m_{\mathrm{MP}}(z)$ satisfies the self-consistent equation [6, 43, 45]

$$\frac{1}{m_{\mathrm{MP}}(z)} = -z + d^{-1}\frac{1}{1+m_{\mathrm{MP}}(z)}, \qquad \operatorname{Im} m_{\mathrm{MP}}(z) \ge 0 \text{ for } z \in \mathbb{C}_+. \tag{2.12}$$

Using (2.8) and (2.9), it is easy to get the expressions for ρ_{1c}, the asymptotic eigenvalue density of $\mathcal{Q}_1$, and m_{1c}, the Stieltjes' transform of ρ_{1c}.

If D is nonidentity but the ESD π_N in (2.5) converges weakly to some $\hat{\pi}$, then it was shown in [34] that the empirical eigenvalue distribution of $\mathcal{Q}_2$ still converges in probability to some deterministic distributions $\hat{\rho}_{2c}$, referred to as the *deformed Marchenko–Pastur law* below. It can be described through the Stieltjes' transform:

$$\hat{m}_{2c}(z) := \int_{\mathbb{R}} \frac{\hat{\rho}_{2c}(dx)}{x-z}, \qquad z = E + i\eta \in \mathbb{C}_+.$$

For any given probability measure $\hat{\pi}$ compactly supported on $\mathbb{R}_+$, we define $\hat{m}_{2c}$ as the unique solution to the self-consistent equation [6, 31, 32]

$$\frac{1}{\hat{m}_{2c}(z)} = -z + d^{-1}\int \frac{x}{1+\hat{m}_{2c}(z)x}\hat{\pi}(dx), \tag{2.13}$$

where the branch-cut is chosen such that $\operatorname{Im}\hat{m}_{2c}(z) \ge 0$ for $z \in \mathbb{C}_+$. It is well known that the functional equation (2.13) has a unique solution that is uniformly bounded on $\mathbb{C}_+$ under the assumptions (2.2) and (2.6) [34]. Letting $\eta \searrow 0$, we can recover the asymptotic eigenvalue density $\hat{\rho}_{2c}$ with the inverse formula

$$\hat{\rho}_{2c}(E) = \lim_{\eta\searrow 0}\frac{1}{\pi}\operatorname{Im}\hat{m}_{2c}(E+i\eta). \tag{2.14}$$

The measure $\hat{\rho}_{2c}$ is sometimes called the *multiplicative free convolution* of $\hat{\pi}$ and the MP law; see, for example, [1, 52]. Again with (2.8) and (2.9), we can easily obtain $\hat{m}_{1c}$ and $\hat{\rho}_{1c}(z)$.

Similar to (2.13), for any finite N we define $m_{2c}^{(N)}$ as the unique solution to the self-consistent equation

$$\frac{1}{m_{2c}^{(N)}(z)} = -z + d_N^{-1}\int \frac{x}{1+m_{2c}^{(N)}(z)x}\pi_N(dx), \tag{2.15}$$

and define $\rho_{2c}^{(N)}$ through the inverse formula as in (2.14). Then we define $m_{1c}^{(N)}$ and $\rho_{1c}^{(N)}(z)$ using (2.8) and (2.9). In the rest of this paper, we will always omit the super-index N from our notation. The properties of $m_{1,2c}$ and $\rho_{1,2c}$ have been studied extensively; see, for example, [3, 4, 7, 24, 31, 46, 47]. Here, we collect some basic results that will be used in our proof. In particular, we shall define the rightmost edge (i.e., the *soft edge*) of $\rho_{1,2c}$.

Corresponding to the equation in (2.15), we define the function

$$f(m) := -\frac{1}{m} + d_N^{-1} \int \frac{x}{1+mx} \pi_N(dx). \tag{2.16}$$

Then $m_{2c}(z)$ can be characterized as the unique solution to the equation $z = f(m)$ with $\operatorname{Im} m \geq 0$.

LEMMA 2.4 (Support of the deformed MP law). *The densities ρ_{1c} and ρ_{2c} have the same support on $\mathbb{R}_+$, which is a union of connected components*:

$$\operatorname{supp} \rho_{1,2c} \cap (0, \infty) = \bigcup_{k=1}^{p} [a_{2k}, a_{2k-1}] \cap (0, \infty), \tag{2.17}$$

where $p \in \mathbb{N}$ depends only on π_N. Here, a_k are characterized as following: there exists a real sequence $\{b_k\}_{k=1}^{2p}$ such that $(x, m) = (a_k, b_k)$ are the real solutions to the equations

$$x = f(m) \quad \text{and} \quad f'(m) = 0. \tag{2.18}$$

Moreover, we have $b_1 \in (-\sigma_1^{-1}, 0)$. Finally, under assumptions (2.2) and (2.6), we have $a_1 \leq C$ for some positive constant C.

For the proof of this lemma, one can refer to Lemma 2.6 and Appendix A.1 of [31]. It is easy to observe that $m_{2c}(a_k) = b_k$ according to the definition of f. We shall call a_k the edges of the deformed MP law ρ_{2c}. In particular, we will focus on the rightmost edge $\lambda_r := a_1$. To establish our result, we need the following extra assumption.

ASSUMPTION 2.5. For σ_1 defined in (2.3), we assume that there exists a small constant $\tau > 0$ such that

$$\left|1 + m_{2c}(\lambda_r)\sigma_1\right| \geq \tau \qquad \text{for all } N. \tag{2.19}$$

REMARK 2.6. The above assumption has previously appeared in [6, 14, 31]. It guarantees a regular square-root behavior of the spectral density ρ_{2c} near λ_r (see Lemma 3.6 below), which is used in proving the local deformed MP law at the soft edge. Note that $f(m)$ has singularities at $m = -\sigma_i^{-1}$ for nonzero σ_i, so the condition (2.19) simply rules out the singularity of f at $m_{2c}(\lambda_r)$.

2.3. *Main result*. The main result of this paper is the following theorem. It establishes the necessary and sufficient condition for the edge universality of the deformed covariance matrix $\mathcal{Q}_2$ at the soft edge λ_r. We define the following tail condition for the entries of X:

$$\lim_{s\to\infty} s^4\mathbb{P}\big(|q_{11}|\geq s\big)=0. \tag{2.20}$$

THEOREM 2.7. *Let $\mathcal{Q}_2 = X^*T^*TX$ be an $N\times N$ sample covariance matrix with X and T satisfying Assumptions 2.1 and 2.5. Let λ_1 be the largest eigenvalues of $\mathcal{Q}_2$.*

- Sufficient condition: *If the tail condition* (2.20) *holds*, *then we have*

$$\lim_{N\to\infty}\mathbb{P}\big(N^{2/3}(\lambda_1-\lambda_r)\leq s\big)=\lim_{N\to\infty}\mathbb{P}^G\big(N^{2/3}(\lambda_1-\lambda_r)\leq s\big) \tag{2.21}$$

for all $s\in\mathbb{R}$, where $\mathbb{P}^G$ denotes the law for X with i.i.d. *Gaussian entries*.

- Necessary condition: *If the condition* (2.20) *does not hold for X*, *then for any fixed $s>\lambda_r$*, *we have*

$$\limsup_{N\to\infty}\mathbb{P}(\lambda_1\geq s)>0. \tag{2.22}$$

REMARK 2.8. In [32], it was proved that there exists $\gamma_0\equiv\gamma_0(N)$ depending only on π_N and the aspect ratio d_N such that

$$\lim_{N\to\infty}\mathbb{P}^G\big(\gamma_0 N^{2/3}(\lambda_1-\lambda_r)\leq s\big)=F_1(s)$$

for all $s\in\mathbb{R}$, where F_1 is the type-1 Tracy–Widom distribution. The scaling factor γ_0 is given by [14]

$$\frac{1}{\gamma_0^3}=\frac{1}{d}\int\left(\frac{x}{1+m_{2c}(\lambda_r)x}\right)^3\pi_N(dx)-\frac{1}{m_{2c}(\lambda_r)^3},$$

and Assumption 2.5 assures that $\gamma_0\sim 1$ for all N. Hence (2.21) and (2.22) together show that the distribution of the rescaled largest eigenvalue of $\mathcal{Q}_2$ converges to the Tracy–Widom distribution if and only if the condition (2.20) holds.

REMARK 2.9. The universality result (2.21) can be extended to the joint distribution of the k largest eigenvalues for any fixed k:

$$\begin{aligned}&\lim_{N\to\infty}\mathbb{P}\big(\big(N^{2/3}(\lambda_i-\lambda_r)\leq s_i\big)_{1\leq i\leq k}\big)\\&\quad=\lim_{N\to\infty}\mathbb{P}^G\big(\big(N^{2/3}(\lambda_i-\lambda_r)\leq s_i\big)_{1\leq i\leq k}\big)\end{aligned} \tag{2.23}$$

for all $s_1,s_2,\dots,s_k\in\mathbb{R}$. Let H^{GOE} be an $N\times N$ random matrix belonging to the Gaussian orthogonal ensemble. The joint distribution of the k largest eigenvalues of H^{GOE}, $\mu_1^{\mathrm{GOE}} \geq \cdots \geq \mu_k^{\mathrm{GOE}}$, can be written in terms of the Airy kernel for any fixed k [23]. It was proved in [32] that

$$\lim_{N\to\infty} \mathbb{P}^G\big(\big(\gamma_0 N^{2/3}(\lambda_i - \lambda_r) \leq s_i\big)_{1\leq i\leq k}\big)$$
$$= \lim_{N\to\infty} \mathbb{P}\big(\big(N^{2/3}\big(\mu_i^{\mathrm{GOE}} - 2\big) \leq s_i\big)_{1\leq i\leq k}\big)$$

for all $s_1, s_2, \ldots, s_k \in \mathbb{R}$. Hence (2.23) gives a complete description of the finite-dimensional correlation functions of the largest eigenvalues of $\mathcal{Q}_2$.

2.4. *Statistical applications.* In this subsection, we briefly discuss possible applications of our results to high-dimensional statistics. Theorem 2.7 indicates that the Tracy–Widom distribution still holds true for the data with heavy tails as in (2.20). Heavy-tailed data is commonly collected in insurance, finance and telecommunications [11]. For example, the log-return of S&P500 index is a heavy-tailed time series and is usually calibrated using distributions with only few moments. For this type of data, many high dimensional statistical hypothesis tests that rely on some strong moment assumptions cannot be employed. For example, the sphericity test based on arithmetic mean of the eigenvalues and maximum likelihood ratio principle needs either Gaussian assumption or high moments assumption [22, 40, 54]. Hence, our result on the distribution of the largest singular value can serve as a valuable tool for many statistical applications.

We now give a few concrete examples of applications to multivariate statistics, empirical finance and signal processing. Consider the following model:

$$\mathbf{x} = \Gamma \mathbf{s} + T\mathbf{z}, \tag{2.24}$$

where $\mathbf{s}$ is a k-dimensional centered vector with population covariance matrix S, $\mathbf{z}$ is an M-dimensional random vector with *i.i.d.* mean zero and variance one entries, Γ is an $M \times k$ deterministic matrix of full rank and T is a $M \times M$ deterministic matrix. Moreover, we assume that the vectors $\mathbf{s}$ and $\mathbf{z}$ are independent. In practice, suppose we observe N such *i.i.d.* samples.

This model has many applications in statistics. One example is from multivariate statistics. It is important to determine if there exists any relation between two sets of variables. To test independence, we consider a multivariate multiple regression model (2.24) in the sense that $\mathbf{x}$, $\mathbf{s}$ are the two sets of variables for testing [25]. We wish to test the null hypothesis that these regression coefficients (entries of Γ) are all equal to zero:

$$\mathbf{H}_o : \Gamma = 0 \quad \text{vs.} \quad \mathbf{H}_a : \Gamma \neq 0. \tag{2.25}$$

Another example is from financial studies [19–21]. In the empirical research of finance, (2.24) is the factor model, where $\mathbf{s}$ is the common factor, Γ is the factor loading matrix and $\mathbf{z}$ is the idiosyncratic component. In order to analyze the stock return $\mathbf{x}$, we first need to know if the factor $\mathbf{s}$ is significant for the prediction. Here, the statistical test can also be constructed as (2.25). The third example is from classic signal processing [29], where (2.24) gives the standard signal-plus-noise model. A fundamental task is to detect the signals via observed samples, and the very first step is to know whether there exists any such signal. Hence our hypothesis testing problem can be formulated as

$$\mathbf{H}_o : k = 0 \quad \text{vs.} \quad \mathbf{H}_a : k \geq 1.$$

For the above hypothesis testing problems, under $\mathbf{H}_o$, the population covariance matrix of $\mathbf{x}$ is TT^*, and $\Gamma S \Gamma^* + TT^*$ under $\mathbf{H}_a$. The largest eigenvalue of the observed samples then serves as a natural choice for the tests. Under the high dimensional setting, this problem was studied in [8, 35] under the assumptions that $\mathbf{z}$ is Gaussian and $T = I$. Nadakuditi and Silverstein [36] also considered this problem with correlated Gaussian noise (i.e., T is not a multiple of I). Under the assumption that the entries have arbitrarily high moments, the problem beyond Gaussian assumption was considered in [6, 32]. Our result shows that, for the heavy-tailed data satisfying (2.20), we can still employ the previous statistical inference methods.

Unfortunately, in practice, T is usually unknown. In particular, the parameters λ_r and γ_0 in Theorem 2.7, Remark 2.8 and Remark 2.9 are unknown, and it would appear that our result cannot be applied directly. Following the strategy in [39], we can use the following statistics:

$$\mathbf{T}_1 := \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_3}. \tag{2.26}$$

The main advantage of $\mathbf{T}_1$ is that its limiting distribution is independent of λ_r and γ_0 under $\mathbf{H}_o$, which makes it asymptotically pivotal. As mentioned in Remark 2.9, we have a complete description of the limiting distribution of $\mathbf{T}_1$. Although the explicit formula is unavailable currently, one can approximate the limiting distribution of $\mathbf{T}_1$ using numerical simulations for the extreme eigenvalues of GOE or GUE.

Our result can be also used in model checking problems in time series analysis, especially the analysis of financial time series [13, 51]. In most of the model building processes, the last step is devoted to checking whether the residuals are white noise, which is an essential driving element of the time series. We assume the residuals have GARCH effect. Consider the GARCH$(1, 1)$ model, where the residuals r_t and the volatility σ_t satisfy

$$r_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where ε_t is a standard white noise. We want to check the null hypothesis that the residuals are white noise:

$$\mathbf{H}_o : \alpha = \beta = 0 \quad \text{vs.} \quad \mathbf{H}_a : \alpha\beta \neq 0.$$

Assuming that K points of r_t are available, we can construct an $M \times N$ matrix R with $MN = K$ [13], Section 3. Under $\mathbf{H}_o$, the population covariance matrix is ωI. Then the largest eigenvalue (or $\mathbf{T}_1$) of RR^* can be used as our test statistic.

3. Basic notation and tools.

3.1. *Notation.* Following the notation in [15, 17], we will use the following definition to characterize events of high probability.

DEFINITION 3.1 (High probability event). Define

$$\varphi := (\log N)^{\log \log N}. \tag{3.1}$$

We say that an N-dependent event Ω holds with ξ*-high probability* if there exist constant $c, C > 0$ independent of N, such that

$$\mathbb{P}(\Omega) \geq 1 - N^C \exp\bigl(-c\varphi^\xi\bigr) \tag{3.2}$$

for all sufficiently large N. For simplicity, for the case $\xi = 1$, we just say *high probability*. Note that if (3.2) holds, then $\mathbb{P}(\Omega) \geq 1 - \exp(-c'\varphi^\xi)$ for any constant $0 \leq c' < c$.

DEFINITION 3.2 (Bounded support condition). A family of $M \times N$ matrices $X = (x_{ij})$ are said to satisfy the bounded support condition with $q \equiv q(N)$ if

$$\mathbb{P}\Bigl(\max_{1 \leq i \leq M, 1 \leq j \leq N} |x_{ij}| \leq q\Bigr) \geq 1 - e^{-N^c} \tag{3.3}$$

for some $c > 0$. Here, $q \equiv q(N)$ depends on N and usually satisfies

$$N^{-1/2} \log N \leq q \leq N^{-\phi},$$

for some small constant $\phi > 0$. Whenever (3.3) holds, we say that X has support q.

REMARK 3.3. Note that the Gaussian distribution satisfies the condition (3.3) with $q < N^{-\phi}$ for any $\phi < 1/2$. We also remark that if (3.3) holds, then the event $\{|x_{ij}| \leq q, \forall 1 \leq i \leq M, 1 \leq j \leq N\}$ holds with ξ-high probability for any fixed $\xi > 0$ according to Definition 3.1. For this reason, the bad event $\{|x_{ij}| > q \text{ for some } i, j\}$ is negligible, and we will not consider the case where the band event happens throughout the proof.

Next, we introduce a convenient self-adjoint linearization trick, which has been proved to be useful in studying the local laws of random matrices of the A^*A type [12, 31, 53]. We define the following $(N + M) \times (N + M)$ block matrix, which is a linear function of X.

DEFINITION 3.4 (Linearizing block matrix). For $z \in \mathbb{C}_+$, we define the $(N + M) \times (N + M)$ block matrix

$$H \equiv H(X) := \begin{pmatrix} 0 & DX \\ (DX)^* & 0 \end{pmatrix}, \tag{3.4}$$

and its Green function

$$G \equiv G(X, z) := \begin{pmatrix} -I_{M\times M} & DX \\ (DX)^* & -zI_{N\times N} \end{pmatrix}^{-1}. \tag{3.5}$$

DEFINITION 3.5 (Index sets). We define the index sets:

$$\mathcal{I}_1 := \{1, \dots, M\}, \qquad \mathcal{I}_2 := \{M+1, \dots, M+N\}, \qquad \mathcal{I} := \mathcal{I}_1 \cup \mathcal{I}_2.$$

Then we label the indices of the matrices according to

$$X = (X_{i\mu} : i \in \mathcal{I}_1, \mu \in \mathcal{I}_2) \quad \text{and} \quad D = \operatorname{diag}(D_{ii} : i \in \mathcal{I}_1).$$

In the rest of this paper, whenever referring to the entries of H and G, we will consistently use the latin letters $i, j \in \mathcal{I}_1$, greek letters $\mu, \nu \in \mathcal{I}_2$ and $a, b \in \mathcal{I}$. For $1 \le i \le \min\{N, M\}$ and $M+1 \le \mu \le M + \min\{N, M\}$, we introduce the notation $\bar{i} := i + M \in \mathcal{I}_2$ and $\bar{\mu} := \mu - M \in \mathcal{I}_1$. For any $\mathcal{I} \times \mathcal{I}$ matrix A, we define the following 2×2 submatrices:

$$A_{[ij]} = \begin{pmatrix} A_{ij} & A_{i\bar{j}} \\ A_{\bar{i}j} & A_{\bar{i}\bar{j}} \end{pmatrix}, \qquad 1 \le i, j \le \min\{N, M\}. \tag{3.6}$$

We shall call $A_{[ij]}$ a *diagonal group* if $i = j$, and an *off-diagonal group* otherwise.

It is easy to verify that the eigenvalues $\lambda_1(H) \ge \dots \ge \lambda_{M+N}(H)$ of H are related to the ones of $\mathcal{Q}_2$ through

$$\lambda_i(H) = -\lambda_{N+M-i+1}(H) = \sqrt{\lambda_i(\mathcal{Q}_2)}, \qquad 1 \le i \le N \wedge M, \tag{3.7}$$

and

$$\lambda_i(H) = 0, \qquad N \wedge M + 1 \le i \le N \vee M,$$

where we used the notation $N \wedge M := \min\{N, M\}$ and $N \vee M := \max\{N, M\}$. Furthermore, by the Schur complement formula, we can verify that

$$\begin{aligned} G &= \begin{pmatrix} z(DXX^*D^* - z)^{-1} & (DXX^*D^* - z)^{-1}DX \\ X^*D^*(DXX^*D^* - z)^{-1} & (X^*D^*DX - z)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} z\mathcal{G}_1 & \mathcal{G}_1 DX \\ X^*D^*\mathcal{G}_1 & \mathcal{G}_2 \end{pmatrix} = \begin{pmatrix} z\mathcal{G}_1 & DX\mathcal{G}_2 \\ \mathcal{G}_2 X^*D^* & \mathcal{G}_2 \end{pmatrix}. \end{aligned} \tag{3.8}$$

Thus a control of G yields directly a control of the resolvents $\mathcal{G}_{1,2}$ defined in (2.7). By (3.8), we immediately get that

$$m_1 = \frac{1}{Mz} \sum_{i \in \mathcal{I}_1} G_{ii}, \qquad m_2 = \frac{1}{N} \sum_{\mu \in \mathcal{I}_2} G_{\mu\mu}.$$

Next, we introduce the spectral decomposition of G. Let

$$DX = \sum_{k=1}^{N\wedge M} \sqrt{\lambda_k}\xi_k\zeta_k^*,$$

be a singular value decomposition of DX, where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N\wedge M} \geq 0 = \lambda_{N\wedge M+1} = \cdots = \lambda_{N\vee M},$$

and $\{\xi_k\}_{k=1}^M$ and $\{\zeta_k\}_{k=1}^N$ are orthonormal bases of $\mathbb{R}^{\mathcal{I}_1}$ and $\mathbb{R}^{\mathcal{I}_2}$, respectively. Then using (3.8), we can get that for $i, j \in \mathcal{I}_1$ and $\mu, \nu \in \mathcal{I}_2$,

$$G_{ij} = \sum_{k=1}^{M} \frac{z\xi_k(i)\xi_k^*(j)}{\lambda_k - z}, \qquad G_{\mu\nu} = \sum_{k=1}^{N} \frac{\zeta_k(\mu)\zeta_k^*(\nu)}{\lambda_k - z}, \tag{3.9}$$

$$G_{i\mu} = \sum_{k=1}^{N\wedge M} \frac{\sqrt{\lambda_k}\xi_k(i)\zeta_k^*(\mu)}{\lambda_k - z}, \qquad G_{\mu i} = \sum_{k=1}^{N\wedge M} \frac{\sqrt{\lambda_k}\zeta_k(\mu)\xi_k^*(i)}{\lambda_k - z}. \tag{3.10}$$

3.2. *Main tools*. For small constant $c_0 > 0$ and large constants $C_0, C_1 > 0$, we define a domain of the spectral parameter $z = E + i\eta$ as

$$S(c_0, C_0, C_1) := \left\{z = E + i\eta : \lambda_r - c_0 \leq E \leq C_0\lambda_r, \frac{\varphi^{C_1}}{N} \leq \eta \leq 1\right\}. \tag{3.11}$$

We define the distance to the rightmost edge as

$$\kappa \equiv \kappa_E := |E - \lambda_r| \qquad \text{for } z = E + i\eta. \tag{3.12}$$

Then we have the following lemma, which summarizes some basic properties of m_{2c} and ρ_{2c}.

LEMMA 3.6 (Lemma 2.1 and Lemma 2.3 in [7]). *There exists sufficiently small constant $\tilde{c} > 0$ such that*

$$\rho_{2c}(x) \sim \sqrt{\lambda_r - x} \qquad \text{for all } x \in [\lambda_r - 2\tilde{c}, \lambda_r]. \tag{3.13}$$

The Stieltjes' transform m_{2c} satisfies that

$$|m_{2c}(z)| \sim 1, \tag{3.14}$$

and

$$\operatorname{Im} m_{2c}(z) \sim \begin{cases} \eta/\sqrt{\kappa + \eta}, & E \geq \lambda_r, \\ \sqrt{\kappa + \eta}, & E \leq \lambda_r \end{cases} \tag{3.15}$$

for $z = E + i\eta \in S(\tilde{c}, C_0, -\infty)$.

REMARK 3.7. Recall that a_k are the edges of the spectral density ρ_{2c}; see (2.17). Hence $\rho_{2c}(a_k) = 0$, and we must have $a_k < \lambda_r - 2\tilde{c}$ for $2 \leq k \leq 2p$. In particular, $S(c_0, C_0, C_1)$ is away from all the other edges if we choose $c_0 \leq \tilde{c}$.

DEFINITION 3.8 (Classical locations of eigenvalues). The classical location γ_j of the jth eigenvalue of $\mathcal{Q}_2$ is defined as

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} \rho_{2c}(x)\, dx > \frac{j-1}{N} \right\}. \tag{3.16}$$

In particular, we have $\gamma_1 = \lambda_r$.

REMARK 3.9. If γ_j lies in the bulk of ρ_{2c}, then by the positivity of ρ_{2c} we can define γ_j through the equation

$$\int_{\gamma_j}^{+\infty} \rho_{2c}(x)\, dx = \frac{j-1}{N}.$$

We can also define the classical location of the jth eigenvalue of $\mathcal{Q}_1$ by changing ρ_{2c} to ρ_{1c} and $(j-1)/N$ to $(j-1)/M$ in (3.16). By (2.8), this gives the same location as γ_j for $j \le N \wedge M$.

DEFINITION 3.10 (Deterministic limit of G). We define the deterministic limit Π of the Green function G in (3.8) as

$$\Pi(z) := \begin{pmatrix} -(1 + m_{2c}(z)\Sigma)^{-1} & 0 \\ 0 & m_{2c}(z) I_{N\times N} \end{pmatrix}, \tag{3.17}$$

where Σ is defined in (2.4).

In the rest of this section, we present some results that will be used in the proof of Theorem 2.7. Their proofs will be given in subsequent sections.

LEMMA 3.11 (Local deformed MP law). *Suppose the Assumptions 2.1 and 2.5 hold. Suppose X satisfies the bounded support condition (3.3) with $q \le N^{-\phi}$ for some constant $\phi > 0$. Fix $C_0 > 0$ and let $c_1 > 0$ be a sufficiently small constant. Then there exist constants $C_1 > 0$ and $\xi_1 \ge 3$ such that the following events hold with ξ_1-high probability*:

$$\bigcap_{z \in S(2c_1, C_0, C_1)} \left\{ |m_2(z) - m_{2c}(z)| \le \varphi^{C_1} \left(\min\left\{ q, \frac{q^2}{\sqrt{\kappa + \eta}} \right\} + \frac{1}{N\eta} \right) \right\}, \tag{3.18}$$

$$\bigcap_{z \in S(2c_1, C_0, C_1)} \left\{ \max_{a,b \in \mathcal{I}} |G_{ab}(z) - \Pi_{ab}(z)| \le \varphi^{C_1} \left(q + \sqrt{\frac{\operatorname{Im} m_{2c}(z)}{N\eta}} + \frac{1}{N\eta} \right) \right\}, \tag{3.19}$$

$$\{ \|H\|^2 \le \lambda_r + \varphi^{C_1}(q^2 + N^{-2/3}) \}. \tag{3.20}$$

The estimates in (3.18) and (3.19) are usually referred to as the *averaged local law* and *entrywise local law*, respectively. In fact, under different assumptions, they have been proved previously in different forms [6, 31]. For completeness, we will give a concise proof in Appendix A that fits into our setting.

The local laws (3.18) and (3.19) can be used to derive some important properties of the eigenvectors and eigenvalues of the random matrices. For instance, they lead to the following results about the delocalization of eigenvectors and the rigidity of eigenvalues. Note that (3.21) gives an almost optimal estimate on the flatness of the singular vectors of DX, while (3.22) gives some quite precise information on the locations of the singular values of DX. We will prove them in Section 5.

LEMMA 3.12. *Suppose the events* (3.18) *and* (3.19) *hold with* ξ_1*-high probability. Then there exists constant* $C_1' > 0$ *such that the following events hold with* ξ_1*-high probability*:

(1) *Delocalization*:

$$\bigcap_{k:\lambda_r - c_1 \le \gamma_k \le \lambda_r} \left\{ \max_i |\xi_k(i)|^2 + \max_\mu |\zeta_k(\mu)|^2 \le \frac{\varphi^{C_1'}}{N} \right\}. \tag{3.21}$$

(2) *Rigidity of eigenvalues*: *if* $q \le N^{-\phi}$ *for some constant* $\phi > 1/3$,

$$\bigcap_{j:\lambda_r - c_1 \le \gamma_j \le \lambda_r} \{ |\lambda_j - \gamma_j| \le \varphi^{C_1'} (j^{-1/3} N^{-2/3} + q^2) \}, \tag{3.22}$$

where λ_j *is the* j*th eigenvalue of* $(DX)^* DX$ *and* γ_j *is defined in* (3.16).

With Lemma 3.11, Lemma 3.12 and a standard Green function comparison method, one can prove the following edge universality result when the support q is small. For the details of the method, the reader can refer to for example, [6], Section 4, [32], Section 4, [15], Theorem 2.7, [18], Section 6 and [43], Section 4.

LEMMA 3.13 (Theorem 1.3 of [6]). *Let* X^W *and* X^V *be two sample covariance matrices satisfying the assumptions in Lemma* 3.11. *Moreover, suppose* $q \le \varphi^C N^{-1/2}$ *for some constant* $C > 0$. *Then there exist constants* $\varepsilon, \delta > 0$ *such that, for any* $s \in \mathbb{R}$, *we have*

$$\begin{aligned} &\mathbb{P}^V \left(N^{2/3}(\lambda_1 - \lambda_r) \le s - N^{-\varepsilon} \right) - N^{-\delta} \le \mathbb{P}^W \left(N^{2/3}(\lambda_1 - \lambda_r) \le s \right) \\ &\qquad \le \mathbb{P}^V \left(N^{2/3}(\lambda_1 - \lambda_r) \le s + N^{-\varepsilon} \right) + N^{-\delta}, \end{aligned} \tag{3.23}$$

where $\mathbb{P}^V$ *and* $\mathbb{P}^W$ *denote the laws of* X^V *and* X^W, *respectively.*

REMARK 3.14. As in [15, 18, 33], Lemma 3.13, as well as Theorem 3.16 below, can be can be generalized to finite correlation functions of the k largest eigenvalues for any fixed k:

$$\begin{aligned} &\mathbb{P}^V\big(\big(N^{2/3}(\lambda_i-\lambda_r)\le s_i-N^{-\varepsilon}\big)_{1\le i\le k}\big)-N^{-\delta} \\ &\quad\le \mathbb{P}^W\big(\big(N^{2/3}(\lambda_i-\lambda_r)\le s_i\big)_{1\le i\le k}\big) \\ &\quad\le \mathbb{P}^V\big(\big(N^{2/3}(\lambda_i-\lambda_r)\le s_i+N^{-\varepsilon}\big)_{1\le i\le k}\big)+N^{-\delta}. \end{aligned} \tag{3.24}$$

The proof of (3.24) is similar to that of (3.23) except that it uses a general form of the Green function comparison theorem; see, for example, [18], Theorem 6.4. As a corollary, we can then get the stronger universality result (2.23).

For any matrix X satisfying Assumption 2.1 and the tail condition (2.20), we can construct a matrix X_1 that approximates X with probability $1-o(1)$, and satisfies Assumption 2.1, the bounded support condition (3.3) with $q\le N^{-\phi}$ for some small $\phi>0$, and

$$\mathbb{E}|x_{ij}|^3\le BN^{-3/2},\qquad \mathbb{E}|x_{ij}|^4\le B(\log N)N^{-2} \tag{3.25}$$

for some constant $B>0$ (see Section 4 for the details). We will need the following local law, eigenvalues rigidity and edge universality results for covariance matrices with large support and satisfying condition (3.25).

THEOREM 3.15 (Rigidity of eigenvalues: large support case). *Suppose the Assumptions 2.1 and 2.5 hold. Suppose X satisfies the bounded support condition (3.3) with $q\le N^{-\phi}$ for some constant $\phi>0$ and the condition (3.25). Fix the constants c_1, C_0, C_1 and ξ_1 as given in Lemma 3.11. Then there exists constant $C_2>0$, depending only on c_1, C_1, B and ϕ, such that with high probability we have*

$$\max_{z\in S(c_1,C_0,C_2)}\big|m_2(z)-m_{2c}(z)\big|\le\frac{\varphi^{C_2}}{N\eta} \tag{3.26}$$

for sufficiently large N. Moreover, (3.26) implies that for some constant $\tilde{C}>0$, the following events hold with high probability:

$$\bigcap_{j:\lambda_r-c_1\le\gamma_j\le\lambda_r}\big\{|\lambda_j-\gamma_j|\le\varphi^{\tilde{C}}j^{-1/3}N^{-2/3}\big\} \tag{3.27}$$

and

$$\Big\{\sup_{E\ge\lambda_r-c_1}\big|n(E)-n_c(E)\big|\le\frac{\varphi^{\tilde{C}}}{N}\Big\}, \tag{3.28}$$

where

$$n(E):=\frac{1}{N}\#\{\lambda_j\ge E\},\qquad n_c(E):=\int_E^{+\infty}\rho_{2c}(x)\,dx. \tag{3.29}$$

THEOREM 3.16. *Let X^W and X^V be two* i.i.d. *sample covariance matrices satisfying the assumptions in Theorem* 3.15. *Then there exist constants $\varepsilon, \delta > 0$ such that, for any $s \in \mathbb{R}$, we have*

$$\begin{aligned}\mathbb{P}^V\big(N^{2/3}(\lambda_1 - \lambda_r) \le s - N^{-\varepsilon}\big) - N^{-\delta} &\le \mathbb{P}^W\big(N^{2/3}(\lambda_1 - \lambda_r) \le s\big)\\ &\le \mathbb{P}^V\big(N^{2/3}(\lambda_1 - \lambda_r) \le s + N^{-\varepsilon}\big) + N^{-\delta},\end{aligned} \tag{3.30}$$

where $\mathbb{P}^V$ and $\mathbb{P}^W$ denote the laws of X^V and X^W, respectively.

LEMMA 3.17 (Bounds on G_{ij}: large support case). *Let X be a sample covariance matrix satisfying the assumptions in Theorem* 3.15. *Then for any $0 < c < 1$ and $z \in S(c_1, C_0, C_2) \cap \{z = E + i\eta : \eta \ge N^{-1+c}\}$, we have the following weak bound*:

$$\mathbb{E}\big|G_{ab}(z)\big|^2 \le \varphi^{C_3}\left(\frac{\operatorname{Im} m_{2c}(z)}{N\eta} + \frac{1}{(N\eta)^2}\right), \qquad a \ne b \tag{3.31}$$

for some constant $C_3 > 0$.

In proving Theorem 3.15, Theorem 3.16 and Lemma 3.17, we will make use of the results in Lemmas 3.11–3.13 for covariance matrices with small support. In fact, given any matrix X satisfying the assumptions in Theorem 3.15, we can construct a matrix $\tilde{X}$ having the same first four moments as X but with smaller support $q = O(N^{-1/2}\log N)$.

LEMMA 3.18 (Lemma 5.1 in [33]). *Suppose X satisfies the assumptions in Theorem* 3.15. *Then there exists another matrix $\tilde{X} = (\tilde{x}_{ij})$, such that $\tilde{X}$ satisfies the bounded support condition* (3.3) *with $q = O(N^{-1/2}\log N)$, and the first four moments of the entries of X and $\tilde{X}$ match*, *that is*,

$$\mathbb{E}x_{ij}^k = \mathbb{E}\tilde{x}_{ij}^k, \qquad k = 1, 2, 3, 4. \tag{3.32}$$

From Lemmas 3.11–3.13, we see that Theorems 3.15, 3.16 and Lemma 3.17 hold for $\tilde{X}$. Then due to (3.32), we expect that X has "similar properties" as $\tilde{X}$, so that these results also hold for X. This will be proved with a Green function comparison method, that is, we expand the Green functions with X in terms of Green functions with $\tilde{X}$ using resolvent expansions, and then estimate the relevant error terms; see Section 6 for more details.

4. Proof of of the main result. In this section, we prove Theorem 2.7 with the results in Section 3.2. We begin by proving the necessity part.

PROOF OF THE NECESSITY. Assume that $\lim_{s\to\infty} s^4 \mathbb{P}(|q_{11}| \ge s) \ne 0$. Then we can find a constant $0 < c_0 < 1/2$ and a sequence $\{r_n\}$ such that $r_n \to \infty$ as $n \to \infty$ and

$$\mathbb{P}\big(|q_{ij}| \ge r_n\big) \ge c_0 r_n^{-4}. \tag{4.1}$$

Fix any $s > \lambda_r$. We denote $L := \lfloor \tau M \rfloor$, $I := \sqrt{\tau^{-1} s}$ and define the event

$$\Gamma_N = \{\text{There exist } i \text{ and } j, 1 \le i \le L, 1 \le j \le N, \text{such that } |x_{ij}| \ge I\}.$$

We first show that $\lambda_1(\mathcal{Q}_2) \ge s$ when Γ_N holds. Suppose $|x_{ij}| \ge I$ for some $1 \le i \le L$ and $1 \le j \le N$. Let $\mathbf{u} \in \mathbb{R}^N$ such that $\mathbf{u}(k) = \delta_{kj}$. By assumption (2.6), we have $\sigma_i \ge \tau$ for $i \le L$. Hence

$$\lambda_1(\mathcal{Q}_2) \ge \langle \mathbf{u}, (DX)^*(DX)\mathbf{u} \rangle = \sum_{k=1}^{M} \sigma_k x_{kj}^2 \ge \sigma_i x_{ij}^2 \ge \tau \big(\sqrt{\tau^{-1} s}\big)^2 = s.$$

Now we choose $N \in \{\lfloor (r_n/I)^2 \rfloor : n \in \mathbb{N}\}$. With the choice $N = \lfloor (r_n/I)^2 \rfloor$, we have

$$\begin{aligned} 1 - \mathbb{P}(\Gamma_N) &= \big(1 - \mathbb{P}(|x_{11}| \ge I)\big)^{NL} \le \big(1 - \mathbb{P}(|q_{11}| \ge r_n)\big)^{NL} \\ &\le \big(1 - c_0 r_n^{-4}\big)^{NL} \le \big(1 - c_1 N^{-2}\big)^{c_2 N^2} \end{aligned} \tag{4.2}$$

for some constant $c_1 > 0$ depending on c_0 and I, and some constant $c_2 > 0$ depending on τ and d. Since $(1 - c_1 N^{-2})^{c_2 N^2} \le c_3$ for some constant $0 < c_3 < 1$ independent of N, the above inequality shows that $\mathbb{P}(\Gamma_N) \ge 1 - c_3 > 0$. This shows that $\limsup_{N \to \infty} \mathbb{P}(\Gamma_N) > 0$ and concludes the proof. $\square$

PROOF OF THE SUFFICIENCY. Given the matrix X satisfying Assumption 2.1 and the tail condition (2.20), we introduce a cutoff on its matrix entries at the level $N^{-\varepsilon}$. For any fixed $\varepsilon > 0$, define

$$\alpha_N := \mathbb{P}\big(|q_{11}| > N^{1/2-\varepsilon}\big), \qquad \beta_N := \mathbb{E}\big[\mathbf{1}\big(|q_{11}| > N^{1/2-\varepsilon}\big) q_{11}\big].$$

By (2.20) and integration by parts, we get that for any $\delta > 0$ and large enough N,

$$\alpha_N \le \delta N^{-2+4\varepsilon}, \qquad |\beta_N| \le \delta N^{-3/2+3\varepsilon}. \tag{4.3}$$

Let $\rho(x)$ be the distribution density of q_{11}. Then we define independent random variables q_{ij}^s, q_{ij}^l, c_{ij}, $1 \le i \le M$ and $1 \le j \le N$, in the following ways:

- q_{ij}^s has distribution density $\rho_s(x)$, where

$$\rho_s(x) = \mathbf{1}\left(\left| x - \frac{\beta_N}{1 - \alpha_N} \right| \le N^{1/2-\varepsilon} \right) \frac{\rho(x - \frac{\beta_N}{1-\alpha_N})}{1 - \alpha_N}; \tag{4.4}$$

- q_{ij}^l has distribution density $\rho_l(x)$, where

$$\rho_l(x) = \mathbf{1}\left(\left| x - \frac{\beta_N}{1 - \alpha_N} \right| > N^{1/2-\varepsilon} \right) \frac{\rho(x - \frac{\beta_N}{1-\alpha_N})}{\alpha_N}; \tag{4.5}$$

- c_{ij} is a Bernoulli 0–1 random variable with $\mathbb{P}(c_{ij} = 1) = \alpha_N$ and $\mathbb{P}(c_{ij} = 0) = 1 - \alpha_N$.

Let X^s, X^l and X^c be random matrices such that $X^s_{ij} = N^{-1/2}q^s_{ij}$, $X^l_{ij} = N^{-1/2}q^l_{ij}$ and $X^c_{ij} = c_{ij}$. By (4.4), (4.5) and the fact that X^c_{ij} is Bernoulli, it is easy to check that for independent X^s, X^l and X^c,

$$X_{ij} \overset{d}{=} X^s_{ij}(1 - X^c_{ij}) + X^l_{ij} X^c_{ij} - \frac{1}{\sqrt{N}} \frac{\beta_N}{1 - \alpha_N}, \tag{4.6}$$

where by (4.3), we have

$$\left| \frac{1}{\sqrt{N}} \frac{\beta_N}{1 - \alpha_N} \right| \le 2\delta N^{-2+3\varepsilon}.$$

Therefore, if we define the $M \times N$ matrix $Y = (Y_{ij})$ by

$$Y_{ij} = \frac{1}{\sqrt{N}} \frac{\beta_N}{1 - \alpha_N} \qquad \text{for all } i \text{ and } j,$$

we have $\|Y\| \le cN^{-1+3\varepsilon}$ for some constant $c > 0$. In the proof below, one will see that $\|D(X + Y)\| = \lambda_1^{1/2}((X + Y)^* D^* D(X + Y)) = O(1)$ with probability $1 - o(1)$, where $\lambda_1(\cdot)$ denotes the largest eigenvalue of the random matrix. Then it is easy to verify that with probability $1 - o(1)$,

$$\left|\lambda_1\big((X + Y)^* D^* D(X + Y)\big) - \lambda_1\big(X^* D^* DX\big)\right| = O\big(N^{-1+3\varepsilon}\big). \tag{4.7}$$

Thus the deterministic part in (4.6) is negligible under the scaling $N^{2/3}$.

By (2.20) and integration by parts, it is easy to check that

$$\begin{aligned} &\mathbb{E}q^s_{11} = 0, \qquad \mathbb{E}|q^s_{11}|^2 = 1 - O\big(N^{-1+2\varepsilon}\big), \\ &\mathbb{E}|q^s_{11}|^3 = O(1), \qquad \mathbb{E}|q^s_{11}|^4 = O(\log N). \end{aligned} \tag{4.8}$$

We note that $X_1 := (\mathbb{E}|q^s_{ij}|^2)^{-1/2} X^s$ is a matrix that satisfies the assumptions for X in Theorem 3.16. Together with the estimate for $\mathbb{E}|q^s_{ij}|^2$ in (4.8), we conclude that there exist constants $\varepsilon, \delta > 0$ such that for any $s \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}^G\big(N^{2/3}(\lambda_1 - \lambda_r) \le s - N^{-\varepsilon}\big) - N^{-\delta} &\le \mathbb{P}^s\big(N^{2/3}(\lambda_1 - \lambda_r) \le s\big) \\ &\le \mathbb{P}^G\big(N^{2/3}(\lambda_1 - \lambda_r) \le s + N^{-\varepsilon}\big) + N^{-\delta}, \end{aligned} \tag{4.9}$$

where $\mathbb{P}^s$ denotes the law for X^s and $\mathbb{P}^G$ denotes the law for a Gaussian covariance matrix. Now we write the first two terms on the right-hand side of (4.6) as

$$X^s_{ij}(1 - X^c_{ij}) + X^l_{ij} X^c_{ij} = X^s_{ij} + R_{ij} X^c_{ij},$$

where $R_{ij} := X^l_{ij} - X^s_{ij}$. It remains to show that the effect of the $R_{ij} X^c_{ij}$ terms on λ_1 is negligible. We call the corresponding matrix as $R^c := (R_{ij} X^c_{ij})$. Note that X^c_{ij} is independent of X^s_{ij} and R_{ij}.

We first introduce a cutoff on matrix X^c as $\tilde{X}^c := \mathbf{1}_A X^c$, where

$$A := \{\#\{(i,j) : X^c_{ij} = 1\} \le N^{5\varepsilon}\}$$
$$\cap \{X^c_{ij} = X^c_{kl} = 1 \Rightarrow \{i,j\} = \{k,l\} \text{ or } \{i,j\} \cap \{k,l\} = \varnothing\}.$$

If we regard the matrix X^c as a sequence $\mathbf{X}^c$ of NM *i.i.d.* Bernoulli random variables, it is easy to obtain from the large deviation formula that

$$\mathbb{P}\left(\sum_{i=1}^{MN} \mathbf{X}^c_i \le N^{5\varepsilon}\right) \ge 1 - \exp(-N^{\varepsilon}) \tag{4.10}$$

for sufficiently large N. Suppose the number m of the nonzero elements in X^c is given with $m \le N^{5\varepsilon}$. Then it is easy to check that

$$\begin{aligned} &\mathbb{P}\left(\exists i = k, j \ne l \text{ or } i \ne k, j = l \text{ such that } X^c_{ij} = X^c_{kl} = 1 \Bigm| \sum_{i=1}^{MN} \mathbf{X}^c_i = m\right) \\ &\quad = O(m^2 N^{-1}). \end{aligned} \tag{4.11}$$

Combining the estimates (4.10) and (4.11), we get that

$$\mathbb{P}(A) \ge 1 - O(N^{-1+10\varepsilon}). \tag{4.12}$$

On the other hand, by condition (2.20), we have

$$\mathbb{P}(|R_{ij}| \ge \omega) \le \mathbb{P}\left(|q_{ij}| \ge \frac{\omega}{2} N^{1/2}\right) = o(N^{-2}) \tag{4.13}$$

for any fixed constant $\omega > 0$. Hence if we introduce the matrix

$$E = \mathbf{1}\left(A \cap \left\{\max_{i,j} |R_{ij}| \le \omega\right\}\right) R^c,$$

then

$$\mathbb{P}(E = R^c) = 1 - o(1) \tag{4.14}$$

by (4.12) and (4.13). Thus we only need to study the largest eigenvalue of $(X^s + E)^* D^* D (X^s + E)$, where $\max_{i,j} |E_{ij}| \le \omega$ and the rank of E is less than $N^{5\varepsilon}$. In fact, it suffices to prove that

$$\mathbb{P}(|\lambda^s_1 - \lambda^E_1| \le N^{-3/4}) = 1 - o(1), \tag{4.15}$$

where $\lambda^s_1 := \lambda_1((X^s)^* D^* D X^s)$ and $\lambda^E_1 := \lambda_1((X^s + E)^* D^* D (X^s + E))$. The estimate (4.15), combined with (4.7), (4.9) and (4.14), concludes (2.21).

Now we prove (4.15). Note that $\tilde{X}^c$ is independent of X^s, so the positions of the nonzero elements of E are independent of X^s. Without loss of generality, we assume the m nonzero entries of DE are

$$e_{11}, e_{22}, \ldots, e_{mm}, \qquad m \le N^{5\varepsilon}. \tag{4.16}$$

For the other choices of the positions of nonzero entries, the proof is exactly the same. But we make this assumption to simplify the notation. By (2.6) and the definition of E, we have $|e_{ii}| \le \tau^{-1}\omega$ for $1 \le i \le m$.

We define the matrices:

$$H^s := \begin{pmatrix} 0 & DX^s \\ (DX^s)^* & 0 \end{pmatrix} \quad \text{and} \quad H^E := H^s + P, \qquad P := \begin{pmatrix} 0 & DE \\ (DE)^* & 0 \end{pmatrix}.$$

Then we have the eigendecomposition $P = VP_DV^*$, where P_D is a $2m \times 2m$ diagonal matrix

$$P_D = \operatorname{diag}(e_{11}, \dots, e_{mm}, -e_{11}, \dots, -e_{mm}),$$

and V is an $(M+N) \times 2m$ matrix such that

$$V_{ab} = \begin{cases} \delta_{a,i}/\sqrt{2} + \delta_{a,(M+i)}/\sqrt{2}, & b = i, i \le m, \\ \delta_{a,i}/\sqrt{2} - \delta_{a,(M+i)}/\sqrt{2}, & b = i+m, i \le m, \\ 0, & b \ge 2m+1. \end{cases}$$

With the identity

$$\det \begin{pmatrix} -I_{M\times M} & DX \\ (DX)^* & -zI_{N\times N} \end{pmatrix} = \det(-I_{M\times M}) \det\bigl(X^*D^*DX - zI_{N\times N}\bigr),$$

and Lemma 6.1 of [30], we find that if $\mu \notin \sigma((DX^s)^*DX^s)$, then μ is an eigenvalue of $\mathcal{Q}^\gamma := (X^s + \gamma E)^*D^*D(X^s + \gamma E)$ if and only if

$$\det\bigl(V^*G^s(\mu)V + (\gamma P_D)^{-1}\bigr) = 0, \tag{4.17}$$

where

$$G^s(\mu) := \left(H^s - \begin{pmatrix} I_{M\times M} & 0 \\ 0 & \mu I_{N\times N} \end{pmatrix}\right)^{-1}.$$

Define $R^\gamma := V^*G^sV + (\gamma P_D)^{-1}$ for $0 < \gamma < 1$. It has the following 2×2 blocks [recall the definition (3.6)]: for $1 \le i \le m$,

$$\begin{aligned} \begin{pmatrix} R^\gamma_{i,j} & R^\gamma_{i,j+m} \\ R^\gamma_{i+m,j} & R^\gamma_{i+m,j+m} \end{pmatrix} &= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} G_{[ij]} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &\quad + \delta_{ij} \begin{pmatrix} (\gamma e_{ii})^{-1} & 0 \\ 0 & -(\gamma e_{ii})^{-1} \end{pmatrix}. \end{aligned} \tag{4.18}$$

Now let $\mu := \lambda_1^s \pm N^{-3/4}$. We claim that

$$\mathbb{P}\bigl(\det R^\gamma(\mu) \ne 0 \text{ for all } 0 < \gamma \le 1\bigr) = 1 - o(1). \tag{4.19}$$

If (4.19) holds, then μ is not an eigenvalue of $\mathcal{Q}^\gamma$ with probability $1 - o(1)$. Denoting the largest eigenvalue of $\mathcal{Q}^\gamma$ by λ_1^γ, $0 < \gamma \le 1$, and defining $\lambda_1^0 := \lim_{\gamma\to 0}\lambda_1^\gamma$, we have $\lambda_1^0 = \lambda_1^s$ and $\lambda_1^1 = \lambda_1^E$ by definition. With the continuity of λ_1^γ with respect to γ and the fact that $\lambda_1^0 \in (\lambda_1^s - N^{-3/4}, \lambda_1^s + N^{-3/4})$, we find that

$$\lambda_1^E = \lambda_1^1 \in \big(\lambda_1^s - N^{-3/4}, \lambda_1^s + N^{-3/4}\big),$$

with probability $1 - o(1)$, that is, we have proved (4.15).

Finally, we prove the claim (4.19). Choose $z = \lambda_r + iN^{-2/3}$ and note that H^s has support $N^{-\varepsilon}$. Then by (3.19) and (3.15), we have with high probability,

$$\max_a \big|G_{aa}^s(z) - \Pi_{aa}(\lambda_r)\big| \le N^{-\varepsilon/2}, \tag{4.20}$$

where we also used the Assumption 2.5 and

$$\big|m_{2c}(z) - m_{2c}(\lambda_r)\big| \sim |z - \lambda_r|^{1/2},$$

which follows from (3.13). For the off-diagonal terms, we use (3.31), (3.15) and the Markov inequality to conclude that

$$\max_{a \ne b \in \{1,\dots,m\} \cup \{M+1,\dots,M+m\}} \big|G_{ab}^s(z)\big| \le N^{-1/6}, \tag{4.21}$$

holds with probability $1 - o(N^{-1/6})$. As pointed out in Remark 3.14, we can extend (4.9) to finite correlation functions of the largest eigenvalues. Since the largest eigenvalues in the Gaussian case are separated in the scale $\sim N^{-2/3}$, we conclude that

$$\mathbb{P}\Big(\min_i \big|\lambda_i\big((X^s)^* X^s\big) - \mu\big| \ge N^{-3/4}\Big) \ge 1 - o(1). \tag{4.22}$$

On the other hand, the rigidity result (3.27) gives that with high probability,

$$|\mu - \lambda_r| \le \varphi^{\tilde{C}} N^{-2/3} + N^{-3/4}. \tag{4.23}$$

Using (3.21), (4.22), (4.23) and the rigidity estimate (3.27), we can get that with probability $1 - o(1)$,

$$\max_{a,b} \big|G_{ab}^s(z) - G_{ab}^s(\mu)\big| < N^{-1/4+\varepsilon}. \tag{4.24}$$

For instance, for $\alpha, \beta \in \mathcal{I}_2$, small $c > 0$ and large enough $C > 0$, we have with probability $1 - o(1)$ that

$$\begin{aligned}
&\big|G_{\alpha\beta}(z) - G_{\alpha\beta}(\mu)\big| \\
&\quad \le \sum_k \big|\zeta_k(\alpha)\zeta_k^*(\beta)\big| \left|\frac{1}{\lambda_k - z} - \frac{1}{\lambda_k - \mu}\right| \\
&\quad \le \frac{C}{N^{2/3}} \sum_{\gamma_k \le \lambda_r - c} \big|\zeta_k(\alpha)\zeta_k^*(\beta)\big| + \frac{\varphi^C}{N^{5/3}} \sum_{\gamma_k > \lambda_r - c} \frac{1}{|\lambda_k - z||\lambda_k - \mu|} \\
&\quad \le \frac{C}{N^{2/3}} + \frac{\varphi^C}{N^{5/3}} \sum_{1 \le k \le \varphi^C} \frac{1}{|\lambda_k - z||\lambda_k - \mu|}
\end{aligned}$$

$$+\frac{\varphi^C}{N^{5/3}}\sum_{k>\varphi^C,\gamma_k>\lambda_r-c}\frac{1}{|\lambda_k-z||\lambda_k-\mu|}$$

$$\leq \frac{C}{N^{2/3}}+\frac{\varphi^{2C}}{N^{1/4}}+\frac{\varphi^C}{N^{2/3}}\left(\frac{1}{N}\sum_{k>\varphi^C,\gamma_k>\lambda_r-c}\frac{1}{|\lambda_k-z||\lambda_k-\mu|}\right)\leq N^{-1/4+\varepsilon},$$

where in the first step we used (3.9), in the second step (3.21) and $|\lambda_k - z| \times |\lambda_k-\mu| \gtrsim 1$ for $\gamma_k \leq \lambda_r - c$ due to (3.27), in the third step the Cauchy–Schwarz inequality, in the fourth step (4.22) and in the last step the rigidity estimate (3.27). For all the other choices of a and b, we can prove the estimate (4.24) in a similar way. Now by (4.24), we see that (4.20) and (4.21) still hold if we replace z by $\mu = \lambda_1^s \pm N^{-3/4}$ and double the right-hand sides. Then using $\max_i |e_{ii}| \leq \tau^{-1}\omega$ and (4.18), we get that for any $0<\gamma\leq 1$,

$$\min_{1\leq i\leq m,\gamma}\{|R^\gamma_{ii}|,|R^\gamma_{i+m,i+m}|\}\geq \tau\omega^{-1}-\frac{1}{2}\left|\Pi_{ii}(\lambda_r)+m_{2c}(\lambda_r)\right|-O\left(N^{-\varepsilon/2}\right),$$

$$\max_{1\leq i\leq m,\gamma}\{|R^\gamma_{i,i+m}|,|R^\gamma_{i+m,i}|\}\leq \frac{1}{2}\left|\Pi_{ii}(\lambda_r)-m_{2c}(\lambda_r)\right|+O\left(N^{-\varepsilon/2}\right)$$

and

$$\max_{1\leq i\neq j\leq m,\gamma}\left(|R^\gamma_{i,j}|+|R^\gamma_{i+m,j}|+|R^\gamma_{i,j+m}|+|R^\gamma_{i+m,j+m}|\right)=O\left(N^{-1/6}\right),$$

hold with probability $1-o(1)$. Thus R^γ is diagonally dominant with probability $1-o(1)$ (provided that ω is chosen to be sufficiently small). This proves the claim (4.19), which further gives (4.15) and completes the proof. □

5. Proof of Lemma 3.12, Theorem 3.15 and Theorem 3.16.

PROOF OF LEMMA 3.12. We first prove the delocalization result in (3.21) assuming that (3.22) holds with ξ_1-high probability. Choose $z_0 = E + i\eta_0 \in S(2c_1, C_0, C_1)$ with $\eta_0 = \varphi^{C_1}N^{-1}$. By (3.19), we have

$$\left|G_{\mu\mu}(z_0)\right| = O(1) \qquad \text{with } \xi_1\text{-high probability.}$$

Then using the spectral decomposition (3.9), we get

$$\sum_{k=1}^N \frac{\eta_0|\zeta_k(\mu)|^2}{(\lambda_k-E)^2+\eta_0^2} = \operatorname{Im} G_{\mu\mu}(z_0) = O(1). \tag{5.1}$$

By (3.22), it is easy to see that $\lambda_k + i\eta_0 \in S(2c_1, C_0, C_1)$ with ξ_1-high probability for every k such that $\lambda_r - c_1 \leq \gamma_k \leq \lambda_r$. Then choosing $E = \lambda_k$ in (5.1) yields that

$$\left|\zeta_k(\mu)\right|^2 \lesssim \eta_0 = \frac{\varphi^{C_1}}{N} \qquad \text{with } \xi_1\text{-high probability.}$$

The proof for $|\xi_k(i)|^2$ is similar.

Now we prove the rigidity results in (3.22) when $q \le N^{-1/3}$. Our argument basically follows from the ones in [17], Section 8, [18], Section 5 and [43], Section 8. Recall the notation in (3.29), we first claim the following lemma. It can be proved with a standard argument using the local law (3.18) and the Helffer–Sjöstrand calculus. For the reader's sake, we include its proof in Appendix B.

LEMMA 5.1. *Suppose the event* (3.18) *holds with* ξ_1*-high probability. Then there exists constant* $\tilde{C}_1 > 0$ *such that*

$$\bigcap_{E \ge \lambda_r - c_1} \{ |n(E) - n_c(E)| \le \varphi^{\tilde{C}_1} (N^{-1} + q^3 + q^2 \sqrt{\kappa_E}) \} \tag{5.2}$$

holds with ξ_1*-high probability, where* κ_E *is defined in* (3.12).

Now we derive the estimate (3.22) from Lemma 5.1. We define the event Ω as the intersection of the events on which (5.2) and (3.20) hold. We first assume that $\lambda_j, \gamma_j \ge \lambda_r - \varphi^K N^{-2/3}$ for some constant $K > \tilde{C}_1$. Then by (3.20) and $q \le N^{-1/3}$, we have that on Ω,

$$|\lambda_j - \gamma_j| \le \varphi^L N^{-2/3} \tag{5.3}$$

for some constant $L > \max\{K, C_1\}$. Note that by (3.13), $n_c(x) \sim (\lambda_r - x)^{3/2}$ for x near λ_r, that is,

$$n_c(\gamma_j) = \frac{j}{N} \sim (\lambda_r - \gamma_j)^{3/2}. \tag{5.4}$$

Then in this case, we have $j \le \varphi^{2K}$. Together with (5.3), we get (3.22) (for a sufficiently large constant $C_1' > 0$).

For the rest of j's, we use the dyadic decomposition

$$U_k := \{ j : \gamma_j \ge \lambda_r - c_1, 2^k \varphi^K N^{-2/3} < \lambda_r - \min\{\gamma_j, \lambda_j\} \le 2^{k+1} \varphi^K N^{-2/3} \}$$

for $k \ge 0$. By (5.2) and $q \le N^{-1/3}$, we find that on Ω,

$$\frac{j}{N} = n_c(\gamma_j) = n(\lambda_j) = n_c(\lambda_j) + \varphi^{\tilde{C}_1} O(N^{-1} + q^2 \sqrt{\kappa_{\lambda_j}}). \tag{5.5}$$

On Ω and for $j \in U_k$, the second term on the right-hand side of (5.5) can be estimated as

$$\varphi^{\tilde{C}_1} O(N^{-1} + q^2 \sqrt{\kappa_{\lambda_j}}) \le C \varphi^{\tilde{C}_1} N^{-1} + C 2^{k/2} \varphi^{\tilde{C}_1 + K/2} q^2 N^{-1/3}.$$

Moreover, on Ω and for $j \in U_k$ we have

$$n_c(\gamma_j) \ge c 2^{3k/2} \varphi^{3K/2} N^{-1} \gg \varphi^{\tilde{C}_1} O(N^{-1} + q^2 \sqrt{\kappa_{\lambda_j}}),$$

where we used (5.4) again. Then we deduce from (5.5) that

$$n_c(\lambda_j) = n_c(\gamma_j)[1 + O(\varphi^{\tilde{C}_1 - K})].$$

Thus on Ω and for $j \in U_k$, we have $\lambda_r - \lambda_j \sim \lambda_r - \gamma_j$, and hence $|n_c'(x)| \sim |n_c'(\gamma_j)|$ for any x between γ_j and λ_j. Thus mean-value theorem and (5.5) imply that on Ω and for $j \in U_k$,

$$\begin{aligned}
|\lambda_j - \gamma_j| &\le \frac{C|n_c(\lambda_j) - n_c(\gamma_j)|}{|n_c'(\gamma_j)|} \le \frac{C\varphi^{\tilde{C}_1}}{(j/N)^{1/3}}\big(N^{-1} + q^2\sqrt{\kappa_{\lambda_j}}\big) \\
&\le \frac{C\varphi^{\tilde{C}_1}}{(j/N)^{1/3}}\big(N^{-1} + q^2\sqrt{\kappa_{\gamma_j}} + q^2\sqrt{|\lambda_j - \gamma_j|}\big) \\
&\le C\varphi^{\tilde{C}_1}\big(j^{-1/3}N^{-2/3} + q^2\big) + C\varphi^{\tilde{C}_1}N^{1/3}q^2\sqrt{|\lambda_j - \gamma_j|} \\
&\le C\varphi^{\tilde{C}_1}\big(j^{-1/3}N^{-2/3} + q^2\big) + C\varphi^{2\tilde{C}_1}q^2 + \frac{1}{2}|\lambda_j - \gamma_j|,
\end{aligned}$$

where we used that $|n_c'(\gamma_j)| \sim (\lambda_r - \gamma_j)^{1/2} \sim (j/N)^{1/3}$, $\kappa_{\lambda_j} \le \kappa_{\gamma_j} + |\lambda_j - \gamma_j|$ and $\kappa_{\gamma_j} \sim (j/N)^{2/3}$. Thus the above inequality gives that on Ω and for $j \in U_k$,

$$|\lambda_j - \gamma_j| \le C\varphi^{2\tilde{C}_1}\big(j^{-1/3}N^{-2/3} + q^2\big).$$

This concludes (3.22). $\square$

With Lemma 3.18, given X satisfying the assumptions in Theorem 3.15, we can construct a matrix $\tilde{X}$ with support bounded by $q = O(N^{-1/2}\log N)$ and shares the same first four moments with X. We first prove Theorem 3.15 with the following lemma. Its proof will be given in Section 6.

LEMMA 5.2. *Let X, $\tilde{X}$ be two matrices as in Lemma 3.18, and $G \equiv G(X,z)$, $\tilde{G} \equiv G(\tilde{X},z)$ be the corresponding Green functions. For $z \in S(c_1, C_0, C_1)$ with large enough $C_1 > 0$, if there exist deterministic quantities $J \equiv J(N)$ and $K \equiv K(N)$ such that*

$$\max_{a \ne b}\big|\tilde{G}_{ab}(z)\big| \le J, \qquad \big|\tilde{m}_2(z) - m_{2c}(z)\big| \le K \tag{5.6}$$

hold with ξ_1-high probability for some $\xi_1 \ge 3$, then for any $p \in 2\mathbb{N}$ with $p \le \varphi$, we have

$$\mathbb{E}\big|m_2(z) - m_{2c}(z)\big|^p \le \mathbb{E}\big|\tilde{m}_2(z) - m_{2c}(z)\big|^p + (Cp)^{Cp}\big(J^2 + K + N^{-1}\big)^p. \tag{5.7}$$

PROOF OF THEOREM 3.15. By Lemma 3.18, $\tilde{X}$ has support bounded by $q = O(N^{-1/2}\log N)$. Together with (3.15), we get

$$q = O\bigg(\log N\sqrt{\frac{\operatorname{Im} m_{2c}}{N\eta}}\bigg), \qquad \frac{q^2}{\sqrt{\kappa + \eta}} = O\bigg(\frac{(\log N)^2}{N\eta}\bigg).$$

Then (3.18) and (3.19) show that we can choose

$$J = \varphi^{C'/2}\left(\frac{1}{N\eta} + \sqrt{\frac{\operatorname{Im} m_{2c}}{N\eta}}\right) \quad \text{and} \quad K = \frac{\varphi^{C'}}{N\eta}$$

for some large enough constant $C' > 0$ such that (5.6) holds with ξ_1-high probability. Then using Markov inequality and (5.7), we get that for sufficiently large constant $C_2 > 0$ and small constants $c, c' > 0$,

$$\begin{aligned} &\mathbb{P}\big(\big|m_2(z) - m_{2c}(z)\big| \geq \varphi^{C_2}(N\eta)^{-1}\big) \\ &\quad \leq \frac{(C\eta^{-1})^p \exp(-c\varphi^{\xi_1})}{\varphi^{C_2 p}(N\eta)^{-p}} + \frac{(Cp)^{Cp}\varphi^{C'p}}{\varphi^{C_2 p}} \leq \exp(-c'\varphi), \end{aligned}$$

where we used $p = \varphi$ and the trivial bound $|\tilde{m}_2(z) - m_{2c}(z)| \leq C\eta^{-1}$ [see (A.5)] on the bad event with probability $\leq \exp(-c\varphi^{\xi_1})$. This proves (3.26). Then using (3.26), one can derive (3.27) and (3.28) with the same arguments as in the previous proof of Lemma 3.12. In fact, comparing (3.26) with (3.18), it is easy to see that we can simply take $q = 0$ in (5.2) and (3.22) to get the desired bounds. □

For the matrix $\tilde{X}$ in Lemma 3.18, it satisfies the desired edge universality according to Lemma 3.13. Then Theorem 3.16 follows immediately from the following comparison lemma.

LEMMA 5.3. *Let X and $\tilde{X}$ be two matrices as in Lemma 3.18. Then there exist constants $\varepsilon, \delta > 0$ such that, for any $s \in \mathbb{R}$ we have*

$$\begin{aligned} &\mathbb{P}^{\tilde{X}}\big(N^{2/3}(\lambda_1 - \lambda_r) \leq s - N^{-\varepsilon}\big) - N^{-\delta} \leq \mathbb{P}^{X}\big(N^{2/3}(\lambda_1 - \lambda_r) \leq s\big) \\ &\quad \leq \mathbb{P}^{\tilde{X}}\big(N^{2/3}(\lambda_1 - \lambda_r) \leq s + N^{-\varepsilon}\big) + N^{-\delta}, \end{aligned} \tag{5.8}$$

where $\mathbb{P}^X$ and $\mathbb{P}^{\tilde{X}}$ are the laws for X and $\tilde{X}$, respectively.

By the rigidity result (3.27), we can assume that the parameter s satisfies

$$|s| \leq \varphi^{\tilde{C}}, \tag{5.9}$$

since otherwise (3.27) already gives the desired result. Our goal is to write the distribution of the largest eigenvalue in terms of a cutoff function depending only on the Green functions. Then it is natural to use the Green function comparison method to conclude the proof. Let

$$\mathcal{N}(E_1, E_2) := \#\{j : E_1 \leq \lambda_j \leq E_2\}$$

denote the number of eigenvalues of $\mathcal{Q}_2 = X^* D^* D X$ in $[E_1, E_2]$; similarly we define $\tilde{\mathcal{N}}$ for $\tilde{\mathcal{Q}}_2 = \tilde{X}^* D^* D \tilde{X}$. Then to quantify the distribution of λ_1, it is equivalent to use $\mathbb{P}(\mathcal{N}(E, \infty) = 0)$. Set

$$E_u := \lambda_r + 2N^{-2/3}\varphi^{\tilde{C}}, \tag{5.10}$$

and for any $E \leq E_u$, define $\mathcal{X}_E := \mathbf{1}_{[E,E_u]}$ to be the characteristic function of the interval $[E, E_u]$. For any $\eta > 0$, we define

$$\theta_\eta(x) := \frac{1}{\pi} \frac{\eta}{x^2 + \eta^2} = \frac{1}{\pi} \operatorname{Im} \frac{1}{x - i\eta},$$

to be an approximate delta function on scale η. Note that under the above definitions, we have $\mathcal{N}(E, E_u) = \operatorname{Tr} \mathcal{X}_E(\mathcal{Q}_2)$ and

$$\operatorname{Tr} \mathcal{X}_{E-l} * \theta_\eta(\mathcal{Q}_2) = N \frac{1}{\pi} \int_{E-l}^{E_u} \operatorname{Im} m_2(y + i\eta)\, dy \tag{5.11}$$

for any $l > 0$. Let $\mathbf{q}$ be a smooth symmetric cutoff function such that

$$\mathbf{q}(x) = \begin{cases} 1 & \text{if } |x| \leq 1/9, \\ 0 & \text{if } |x| \geq 2/9, \end{cases}$$

and we assume that $\mathbf{q}(x)$ is decreasing when $x \geq 0$. Then the following lemma provides a way to approximate $\mathbb{P}(\mathcal{N}(E, \infty) = 0)$ with a function depending only on Green functions.

LEMMA 5.4. *For $\varepsilon > 0$, let $\eta = N^{-2/3-9\varepsilon}$ and $l = N^{-2/3-\varepsilon}/2$. Suppose Theorem 3.15 holds. Then for all E such that*

$$|E - \lambda_r| \leq \frac{3}{2} \varphi^{\tilde{C}} N^{-2/3},$$

where the constant $\tilde{C}$ is as in (3.27), (3.28), (5.9) and (5.10), we have

$$\begin{aligned} \mathbb{E}\mathbf{q}\big(\operatorname{Tr} \mathcal{X}_{E-l} * \theta_\eta(\mathcal{Q}_2)\big) &\leq \mathbb{P}\big(\mathcal{N}(E, \infty) = 0\big) \\ &\leq \mathbb{E}\mathbf{q}\big(\operatorname{Tr} \mathcal{X}_{E+l} * \theta_\eta(\mathcal{Q}_2)\big) + \exp\big(-c\varphi^{\tilde{C}}\big) \end{aligned} \tag{5.12}$$

for some constant $c > 0$.

PROOF. See [43], Corollary 4.2, or [18], Corollary 6.2, for the proof. The key inputs are the rigidity estimates (3.27) and (3.28) in Theorem 3.15. □

To prove Lemma 5.3, we need the following Green function comparison result, which will be proved in Section 6.

LEMMA 5.5. *Let X and $\tilde{X}$ be two matrices as in Lemma 3.18. Suppose $F : \mathbb{R} \to \mathbb{R}$ is a function whose derivatives satisfy*

$$\sup_x \big|F^{(n)}(x)\big|\big(1 + |x|\big)^{-C_4} \leq C_4, \qquad n = 1, 2, 3 \tag{5.13}$$

for some constant $C_4 > 0$. Then for any sufficiently small constant $\varepsilon > 0$ and for any real numbers

$$E, E_1, E_2 \in I_\varepsilon := \big\{x : |x - \lambda_r| \leq N^{-2/3+\varepsilon}\big\} \quad \textit{and} \quad \eta := N^{-2/3-\varepsilon},$$

we have

$$\bigl|\mathbb{E}F\bigl(N\eta\operatorname{Im}m_2(z)\bigr)-\mathbb{E}F\bigl(N\eta\operatorname{Im}\tilde{m}_2(z)\bigr)\bigr|\le N^{-\phi+C_5\varepsilon},\qquad z=E+i\eta,\tag{5.14}$$

and

$$\begin{aligned}&\biggl|\mathbb{E}F\biggl(N\int_{E_1}^{E_2}\operatorname{Im}m_2(y+i\eta)\,dy\biggr)-\mathbb{E}F\biggl(N\int_{E_1}^{E_2}\operatorname{Im}\tilde{m}_2(y+i\eta)\,dy\biggr)\biggr|\\&\quad\le N^{-\phi+C_5\varepsilon},\end{aligned}\tag{5.15}$$

where ϕ is defined in Theorem 3.15 *and $C_5>0$ is some constant.*

PROOF OF LEMMA 5.3. Recall that we only need to consider s that satisfies (5.9). Thus it suffices to assume that $|E-\lambda_r|\le\varphi^{\tilde{C}}N^{-2/3}$. Then by Lemma 5.5 and (5.11), there exists constant $\delta>0$ such that

$$\mathbb{E}\mathbf{q}\bigl(\operatorname{Tr}\mathcal{X}_{E-l}*\theta_\eta(\tilde{\mathcal{Q}}_2)\bigr)\le\mathbb{E}\mathbf{q}\bigl(\operatorname{Tr}\mathcal{X}_{E-l}*\theta_\eta(\mathcal{Q}_2)\bigr)+N^{-\delta}.\tag{5.16}$$

For the choice $l=\frac{1}{2}N^{-2/3-\varepsilon}$, we also have $|E-l-\lambda_r|\le\frac{3}{2}\varphi^{\tilde{C}}N^{-2/3}$. Thus we can apply Lemma 5.4 to get

$$\mathbb{P}\bigl(\tilde{\mathcal{N}}(E-2l,\infty)=0\bigr)\le\mathbb{E}\mathbf{q}\bigl(\operatorname{Tr}\mathcal{X}_{E-l}*\theta_\eta(\tilde{\mathcal{Q}}_2)\bigr)+\exp\bigl(-c\varphi^{\tilde{C}}\bigr).\tag{5.17}$$

With (5.16), (5.17) and Lemma 5.4, we get that

$$\begin{aligned}\mathbb{P}\bigl(\tilde{\mathcal{N}}(E-2l,\infty)=0\bigr)-2N^{-\delta}&\le\mathbb{E}\mathbf{q}\bigl(\operatorname{Tr}\mathcal{X}_{E-l}*\theta_\eta(\mathcal{Q}_2)\bigr)\\&\le\mathbb{P}\bigl(\mathcal{N}(E,\infty)=0\bigr).\end{aligned}\tag{5.18}$$

If we choose $E=\lambda_r+sN^{-\frac{2}{3}}$, then (5.18) implies that

$$\mathbb{P}^{\tilde{X}}\bigl(N^{2/3}(\lambda_1-\lambda_r)\le s-N^{-\varepsilon}\bigr)-2N^{-\delta}\le\mathbb{P}^X\bigl(N^{2/3}(\lambda_1-\lambda_r)\le s\bigr).$$

This proves one of the inequalities in (5.3). The other inequality can be proved in a similar way using Lemma 5.4 and Lemma 5.5. This proves Lemma 5.3, which further completes the proof of Theorem 3.16. □

6. Proof of Lemma 5.2, Lemma 5.5 and Lemma 3.17. For the proof in this section, we will use the Green function comparison method developed in [33]. More specifically, we will apply the Lindeberg replacement strategy to G in (3.5). Let $X=(x_{i\mu})$ and $\tilde{X}=(\tilde{x}_{i\mu})$ be two matrices as in Lemma 3.18 (in this section we name the indices as in Definition 3.5). Define a bijective ordering map Φ on the index set of X as

$$\Phi:\bigl\{(i,\mu):1\le i\le M,\,M+1\le\mu\le M+N\bigr\}\to\{1,\dots,\gamma_{\max}=MN\}.$$

For any $1\le\gamma\le\gamma_{\max}$, we define the matrix $X^\gamma=(x_{i\mu}^\gamma)$ such that $x_{i\mu}^\gamma=x_{i\mu}$ if $\Phi(i,\mu)\le\gamma$, and $x_{i\mu}^\gamma=\tilde{x}_{i\mu}$ otherwise. Note we have that $X^0=\tilde{X}$, $X^{\gamma_{\max}}=X$, and

X^γ satisfies the bounded support condition with $q = N^{-\phi}$ for all $0 \le \gamma \le \gamma_{\max}$. Correspondingly, we define

$$H^\gamma := \begin{pmatrix} 0 & DX^\gamma \\ (DX^\gamma)^* & 0 \end{pmatrix}, \qquad G^\gamma := \begin{pmatrix} -I_{M\times M} & DX^\gamma \\ (DX^\gamma)^* & -zI_{N\times N} \end{pmatrix}^{-1}. \tag{6.1}$$

Note that H^γ and $H^{\gamma-1}$ differ only at the (i,μ) and (μ,i) elements, where $\Phi(i,\mu) = \gamma$. Then we define the $(N+M)\times(N+M)$ matrices V and W by

$$V_{ab} = (\mathbf{1}_{\{(a,b)=(i,\mu)\}} + \mathbf{1}_{\{(a,b)=(\mu,i)\}})\sqrt{\sigma_i}x_{i\mu}$$

and

$$W_{ab} = (\mathbf{1}_{\{(a,b)=(i,\mu)\}} + \mathbf{1}_{\{(a,b)=(\mu,i)\}})\sqrt{\sigma_i}\tilde{x}_{i\mu},$$

so that H^γ and $H^{\gamma-1}$ can be written as

$$H^\gamma = Q + V, \qquad H^{\gamma-1} = Q + W \tag{6.2}$$

for some $(N+M)\times(N+M)$ matrix Q satisfying $Q_{i\mu} = Q_{\mu i} = 0$. For simplicity of notation, we denote the Green functions by

$$S := G^\gamma, \qquad T := G^{\gamma-1}, \qquad R := \left(Q - \begin{pmatrix} I_{M\times M} & 0 \\ 0 & zI_{N\times N} \end{pmatrix}\right)^{-1}. \tag{6.3}$$

Under the above definitions, we can write

$$S = \left(Q - \begin{pmatrix} I_{M\times M} & 0 \\ 0 & zI_{N\times N} \end{pmatrix} + V\right)^{-1} = (I + RV)^{-1}R. \tag{6.4}$$

Thus we can expand S using the resolvent expansion till order m:

$$S = R - RVR + (RV)^2R + \cdots + (-1)^m(RV)^mR + (-1)^{m+1}(RV)^{m+1}S. \tag{6.5}$$

On the other hand, we can also expand R in terms of S,

$$R = (I - SV)^{-1}S = S + SVS + (SV)^2S + \cdots + (SV)^mS + (SV)^{m+1}R. \tag{6.6}$$

We have similar expansions for T and R by replacing V, S with W, T in (6.5) and (6.6).

By the bounded support condition, we have

$$\max_{a,b}|V_{ab}| = \sqrt{\sigma_i}|x_{i\mu}| = O(N^{-\phi}), \tag{6.7}$$

with ξ_1-high probability. Together with Lemma 3.11 and (6.6), it is easy to check that $\max_{a,b}|R_{ab}| = O(1)$ with ξ_1-high probability. Thus there exists a constant $C_6 > 0$ such that with ξ_1-high probability,

$$\sup_{z\in S(c_1,C_0,C_1)} \max_\gamma \max_{a,b} \max\{|S_{ab}|, |T_{ab}|, |R_{ab}|\} \le C_6, \tag{6.8}$$

where we used (3.19) and the fact that $m_{2c}(z)$ is uniformly bounded on $S(c_1, C_0, C_1)$. On the other hand, by (A.5) we have the following trivial deterministic bound for S, R and T:

$$\max_{\gamma} \max_{a,b} \max\{|S_{ab}|, |T_{ab}|, |R_{ab}|\} \le C\eta^{-1} \le N. \tag{6.9}$$

In the following discussions, we fix γ and i, μ such that $\Phi(i, \mu) = \gamma$. The expressions below will depend on γ, but we drop this dependence for convenience. For simplicity, we will use $|\mathbf{v}| \equiv \|\mathbf{v}\|_1$ to denote the l^1-norm for any vector $\mathbf{v}$.

The following lemma gives a simple estimate for the remainder terms in (6.5) and (6.6).

LEMMA 6.1. *There exists constant $C > 0$ such that for any $m \in \mathbb{N}$,*

$$\max_{a,b} \max\{|((RV)^m S)_{ab}|, |((SV)^m R)_{ab}|\} = O(C^m N^{-m\phi}), \tag{6.10}$$

with ξ_1-high probability.

PROOF. By the definition of V, we have, for example,

$$((RV)^m S)_{ab} = \sum_{(a_l, b_l) \in \{(i,\mu),(\mu,i)\}: 1 \le l \le m} (\sqrt{\sigma_i} x_{i\mu})^m R_{aa_1} R_{b_1 a_2} \cdots S_{b_m b}. \tag{6.11}$$

Since there are 2^m terms in the above sum, the conclusion follows immediately from (6.7) and (6.8). $\square$

From the expression (6.11), one can see that it is helpful to introduce the following notation.

DEFINITION 6.2 (Matrix operators $*_\gamma$). For any two $(N + M) \times (N + M)$ matrices A and B, we define $A *_\gamma B$ as

$$A *_\gamma B := A I_\gamma B, \qquad (I_\gamma)_{ab} = \mathbf{1}_{\{(a,b)=(i,\mu)\}} + \mathbf{1}_{\{(a,b)=(\mu,i)\}}, \tag{6.12}$$

where (i, μ) is such that $\Phi(i, \mu) = \gamma$. In other words, we have

$$(A *_\gamma B)_{ab} = A_{ai} B_{\mu b} + A_{a\mu} B_{ib}.$$

When γ is fixed, we often drop the subscript γ and write $A * B$ for simplicity. Also we denote the mth power of A under the $*_\gamma$-product by A^{*m}, that is,

$$A^{*m} \equiv A^{*_\gamma m} := \underbrace{A * A * A * \cdots * A}_{m}. \tag{6.13}$$

DEFINITION 6.3 ($\mathcal{P}_{\gamma,\mathbf{k}}$ and $\mathcal{P}_{\gamma,k}$ notation). For $k \in \mathbb{N}$ and $\mathbf{k} = (k_1, \ldots, k_s) \in \mathbb{N}^s$, $\gamma = \Phi(i, \mu)$, we define

$$\mathcal{P}_{\gamma,k} G_{ab} := G_{ab}^{*_\gamma (k+1)} \tag{6.14}$$

and

$$\mathcal{P}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^{s} G_{a_t b_t}\right) := \prod_{t=1}^{s}(\mathcal{P}_{\gamma,k_t} G_{a_t b_t}) = \prod_{t=1}^{s} G_{a_t b_t}^{*\gamma(k_t+1)}. \tag{6.15}$$

If G_1 and G_2 are products of matrix entries as above, then we define

$$\mathcal{P}_{\gamma,\mathbf{k}}(G_1 + G_2) := \mathcal{P}_{\gamma,\mathbf{k}} G_1 + \mathcal{P}_{\gamma,\mathbf{k}} G_2. \tag{6.16}$$

Similarly, for the product of the entries of $G - \Pi$, we define

$$\tilde{\mathcal{P}}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^{s}(G-\Pi)_{a_t b_t}\right) := \prod_{t=1}^{s}(\tilde{\mathcal{P}}_{\gamma,k_t}(G-\Pi)_{a_t b_t}), \tag{6.17}$$

where

$$\tilde{\mathcal{P}}_{\gamma,k}(G-\Pi)_{ab} := \begin{cases} (G-\Pi)_{ab} & \text{if } a=b \text{ and } k=0, \\ G_{ab}^{*(k+1)} & \text{otherwise.} \end{cases}$$

Again, we will often drop the subscript γ whenever there is no confusion about it.

REMARK 6.4. Note that $\mathcal{P}_{\gamma,\mathbf{k}}$ and $\tilde{P}_{\gamma,\mathbf{k}}$ are not linear operators acting on matrices, but just notation we use for simplification. Moreover, for $k, l \in \mathbb{N}$ and $\mathbf{k} \in \mathbb{N}^{k+1}$, it is easy to verify that

$$G^{*l} I_\gamma G^{*k} = G^{*(l+k)}, \qquad \mathcal{P}_{\gamma,\mathbf{k}}(\mathcal{P}_{\gamma,k} G_{ab}) = \mathcal{P}_{\gamma,k+|\mathbf{k}|} G_{ab}. \tag{6.18}$$

For the second equality, note that $\mathcal{P}_{\gamma,k} G_{ab}$ is a sum of products of the entries of G, where each product contains $k+1$ matrix entries.

With the above definitions and bound (6.8), it is easy to prove the following lemma.

LEMMA 6.5. *There exists constant $\tilde{C}_6 > 0$ such that for any $\mathbf{k} \in \mathbb{N}^s$, γ, and $a_1, b_1, \ldots, a_s, b_s$, we have*

$$\max\left\{\left|\mathcal{P}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^{s} A_{a_t b_t}\right)\right|, \left|\tilde{P}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^{s}(A-\Pi)_{a_t b_t}\right)\right|\right\} \le \tilde{C}_6^{s+|\mathbf{k}|+1}, \tag{6.19}$$

with ξ_1-high probability, where A can be R, S or T.

Now we begin to perform the Green function comparison strategy. The basic idea is to expand S and T in terms of R using the resolvent expansions as in (6.5) and (6.6), and then compare the two expressions. We expect that the main terms will cancel since $x_{i\mu}$ and $\tilde{x}_{i\mu}$ have the same first four moments, while the remaining error terms will be sufficiently small since $x_{i\mu}$ and $\tilde{x}_{i\mu}$ have support bounded by $N^{-\phi}$. The key is the following Lemma 6.6, whose proof is the same as the one for Lemma 6.5 in [33].

LEMMA 6.6 (Green function representation theorem). *Let $z \in S(c_1, C_0, C_1)$ and $\Phi(i,\mu) = \gamma$. Fix $s = O(\varphi)$ and $\zeta = O(\varphi)$. Then we have*

$$\begin{aligned}\mathbb{E}\prod_{t=1}^{s} S_{a_t b_t} = {} & \sum_{0\le k\le 4} A_k \mathbb{E}\big[(-\sqrt{\sigma_i} x_{i\mu})^k\big] \\ & + \sum_{5\le|\mathbf{k}|\le 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2} \mathcal{A}_{\mathbf{k}} \mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{s} S_{a_t b_t} + O(N^{-\zeta}),\end{aligned} \tag{6.20}$$

where A_k, $0 \le k \le 4$, depend only on R, $\mathcal{A}_{\mathbf{k}}$'s are independent of (a_t, b_t), $1 \le t \le s$, and

$$|\mathcal{A}_{\mathbf{k}}| \le N^{-|\mathbf{k}|\phi/10-2}. \tag{6.21}$$

Similarly, we have

$$\begin{aligned}\mathbb{E}\prod_{t=1}^{s} (S-\Pi)_{a_t b_t} = {} & \sum_{0\le k\le 4} \tilde{A}_k \mathbb{E}\big[(-\sqrt{\sigma_i} x_{i\mu})^k\big] \\ & + \sum_{5\le|\mathbf{k}|\le 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2} \mathcal{A}_{\mathbf{k}} \mathbb{E}\tilde{\mathcal{P}}_{\gamma,\mathbf{k}} \prod_{t=1}^{s} S_{a_t b_t} + O(N^{-\zeta}),\end{aligned} \tag{6.22}$$

where $\tilde{A}_k$, $0 \le k \le 4$, depend only on R, and $\mathcal{A}_{\mathbf{k}}$'s are the same as above.

Finally, as (6.20), *we have*

$$\begin{aligned}\mathbb{E}\prod_{t=1}^{s} S_{a_t b_t} = {} & \mathbb{E}\prod_{t=1}^{s} R_{a_t b_t} \\ & + \sum_{1\le|\mathbf{k}|\le 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2} \tilde{\mathcal{A}}_{\mathbf{k}} \mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{s} S_{a_t b_t} + O(N^{-\zeta}),\end{aligned} \tag{6.23}$$

where $\tilde{\mathcal{A}}$ are independent of (a_t, b_t), $1 \le t \le s$, and

$$|\tilde{\mathcal{A}}_{\mathbf{k}}| \le N^{-|\mathbf{k}|\phi/10}. \tag{6.24}$$

Note that the terms $\mathcal{A}$ and $\tilde{\mathcal{A}}$ do depend on γ and we have omitted this dependence in the above expressions.

REMARK 6.7. We emphasize that most of the comparison arguments in [33], Section 6, can be carried over here due to the introduction of the linearized block matrix in Definition 3.4 and the local laws given in Lemma 3.11. To make this point clear, let $H = (h_{ij})$ and $\tilde{H} = (\tilde{h}_{ij})$ be two $N \times N$ real Wigner matrices. Suppose one would like to compare their Green functions $G = (H-z)^{-1}$ and $\tilde{G} = (\tilde{H}-z)^{-1}$ through the Lindeberg replacement strategy. Then consider a bijective ordering map $\Phi : \{(i, j) : 1 \le i \le j \le N\} \to \{1, 2, \dots, N(N+1)/2\}$, we have for $i < j$ and $\Phi(i, j) = \gamma$,

$$H^{\gamma} = Q + V, \qquad G^{\gamma} = (H^{\gamma} - z)^{-1},$$
$$H^{\gamma-1} = Q + W, \qquad G^{\gamma-1} = (H^{\gamma-1} - z)^{-1},$$

where

$$V_{kl} = (\mathbf{1}_{\{(k,l)=(i,j)\}} + \mathbf{1}_{\{(k,l)=(j,i)\}})h_{ij}, \qquad W_{kl} = (\mathbf{1}_{\{(k,l)=(i,j)\}} + \mathbf{1}_{\{(k,l)=(j,i)\}})\tilde{h}_{ij},$$

and Q is an $N \times N$ matrix satisfying $Q_{ij} = Q_{ji} = 0$. Compared with (6.2) and (6.3), we observe the obvious similarity between these two settings. In particular, the key resolvent expansions (6.5) and (6.6) take the same form as in the Wigner case. Thus most of the comparison arguments in [33] can be used in our paper, as long as we have some appropriate estimates on the R, S, T entries, which have been provided by Lemma 3.11. Due to this reason, we omit the details for the proof. (In fact, the comparison argument in our paper is a little simpler than the one in [33], since we do not need to include a separate argument for the diagonal entries as in the Wigner case.)

It is clear that a result similar to Lemma 6.6 also holds for the product of T entries. Thus as in (6.20), we define the notation $\mathcal{A}^{\gamma,a}$, $a = 0, 1$ as follows:

$$\begin{aligned}\mathbb{E}\prod_{t=1}^{s} S_{a_t b_t} &= \sum_{0\le k\le 4} A_k \mathbb{E}\big[(-\sqrt{\sigma_i}x_{i\mu})^k\big] \\ &\quad + \sum_{5\le|\mathbf{k}|\le 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2}\mathcal{A}_{\mathbf{k}}^{\gamma,0}\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} S_{a_t b_t} + O(N^{-\zeta}),\end{aligned} \tag{6.25}$$

$$\begin{aligned}\mathbb{E}\prod_{t=1}^{s} T_{a_t b_t} &= \sum_{0\le k\le 4} A_k \mathbb{E}\big[(-\sqrt{\sigma_i}\tilde{x}_{i\mu})^k\big] \\ &\quad + \sum_{5\le|\mathbf{k}|\le 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2}\mathcal{A}_{\mathbf{k}}^{\gamma,1}\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} T_{a_t b_t} + O(N^{-\zeta}).\end{aligned} \tag{6.26}$$

Since A_k, $0 \le k \le 4$, depend only on R and $x_{i\mu}$, $\tilde{x}_{i\mu}$ have the same first four moments, we get from (6.25) and (6.26) that for $s = O(\varphi)$ and $\zeta = O(\varphi)$,

$$\begin{aligned}&\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t} - \mathbb{E}\prod_{t=1}^{s} \tilde{G}_{a_t b_t} \\ &\quad = \sum_{\gamma=1}^{\gamma_{\max}}\left(\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma} - \mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma-1}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\gamma=1}^{\gamma_{\max}} \sum_{5\le|\mathbf{k}|\le 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2} \\
&\quad\times \left(\mathcal{A}_{\mathbf{k}}^{\gamma,0}\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma} - \mathcal{A}_{\mathbf{k}}^{\gamma,1}\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma-1}\right) \\
&\quad+ O(N^{-\zeta+2}),
\end{aligned}
\tag{6.27}
$$

where we abbreviate $G := G(X, z)$ and $\tilde{G} := G(\tilde{X}, z)$. Then we obtain that

$$
\begin{aligned}
\left|\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma_{\max}}\right| &\le \left|\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{0}\right| \\
&\quad+ \sum_{\gamma=1}^{\gamma_{\max}} \sum_{a=0,1} \sum_{5\le|\mathbf{k}|\le 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2} \left|\mathcal{A}_{\mathbf{k}}^{\gamma,a}\right| \left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma-a}\right| \\
&\quad+ O(N^{-\zeta+2}).
\end{aligned}
\tag{6.28}
$$

By (6.19) and (6.21), the second term in (6.28) is bounded by

$$
\begin{aligned}
&\sum_{5\le k\le 2\zeta/\phi} \sum_{\gamma=1}^{\gamma_{\max}} \sum_{a=0,1} \sum_{|\mathbf{k}|=k, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{k/2} \left|\mathcal{A}_{\mathbf{k}}^{\gamma,a}\right| \left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma-a}\right| \\
&\quad\le \sum_{5\le k\le 2\zeta/\phi} N^{-k\phi/10} s^k (C_6')^{s+k+1} \\
&\quad\le 2N^{-5\phi/10} s^5 (C_6')^{s+6} \le N^{-5\phi/20}(C_6')^s
\end{aligned}
\tag{6.29}
$$

for some constant $C_6' > 0$, where we used the rough bound $\#\{\mathbf{k}\in\mathbb{N}^s : |\mathbf{k}| = k\} \le s^k$ and $s = O(\varphi)$.

However, the bound in (6.29) is not good enough. To improve it, we iterate the above arguments as following. Recall that $\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t,b_t}^{\gamma-a}$ is also a sum of products of G entries. Applying (6.27) again to the term $\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t,b_t}^{\gamma-a}$ and replacing $\gamma_{\max}$ in (6.28) with $\gamma - a$, we obtain that

$$
\begin{aligned}
\left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma-a}\right| &\le \left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t b_t}^{0}\right| \\
&\quad+ \sum_{\gamma'=1}^{\gamma-a} \sum_{a'=0,1} \sum_{5\le|\mathbf{k}'|\le 2\zeta/\phi, \mathbf{k}'\in\mathbb{N}^{s+|\mathbf{k}|}} \sigma_{i'}^{|\mathbf{k}'|/2} \\
&\quad\times \left|\mathcal{A}_{\mathbf{k}'}^{\gamma',a'}\right| \left|\mathbb{E}\mathcal{P}_{\gamma',\mathbf{k}'}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma'-a'}\right| \\
&\quad+ O(N^{-\zeta}).
\end{aligned}
$$

Together with (6.28), we have

$$\begin{aligned}\left|\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma_{\max}}\right| \leq \left|\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{0}\right| &+ \sum_{\gamma=1}^{\gamma_{\max}} \sum_{a=0,1} \sum_{5\leq|\mathbf{k}|\leq 2\zeta/\phi, \mathbf{k}\in\mathbb{N}^s} \sigma_i^{|\mathbf{k}|/2} |\mathcal{A}_{\mathbf{k}}^{\gamma,a}| \left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{s} G_{a_t,b_t}^{0}\right| \\ &+ \sum_{\gamma,\gamma'}\sum_{a,a'}\sum_{\mathbf{k},\mathbf{k}'} \sigma_i^{|\mathbf{k}|/2}\sigma_{i'}^{|\mathbf{k}'|/2} |\mathcal{A}_{\mathbf{k}}^{\gamma,a}\mathcal{A}_{\mathbf{k}'}^{\gamma',a'}| \left|\mathbb{E}\mathcal{P}_{\gamma',\mathbf{k}'}\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{s} G_{a_t b_t}^{\gamma'-a'}\right| \\ &+ O(N^{-\zeta+2}).\end{aligned}$$

Again using (6.19) and (6.21), it is easy to see that

$$\sum_{\gamma,\gamma'}\sum_{a,a'}\sum_{\mathbf{k},\mathbf{k}'} \sigma_i^{|\mathbf{k}|/2}\sigma_{i'}^{|\mathbf{k}'|/2} |\mathcal{A}_{\mathbf{k}}^{\gamma,a}\mathcal{A}_{\mathbf{k}'}^{\gamma',a'}| \left|\mathbb{E}\mathcal{P}_{\gamma',\mathbf{k}'}\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{s} G_{a_t b_t}^{\gamma'-a'}\right| \leq N^{-\frac{10\phi}{20}} (C_6')^s,$$

where we used that $k'+k \geq 10$. Repeating the above process for $n \leq 6\zeta/\phi$ times, we obtain that

$$\begin{aligned}\left|\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma_{\max}}\right| \leq \sum_{n=0}^{6\zeta/\phi} \sum_{\gamma_1,\dots,\gamma_n} \sum_{a_1,\dots,a_n} \sum_{\mathbf{k}_1,\dots,\mathbf{k}_n} &\left|\prod_j \sigma_{m_j}^{|\mathbf{k}_j|} \mathcal{A}_{\mathbf{k}_j}^{\gamma_j,a_j}\right| \\ \times &\left|\mathbb{E}\mathcal{P}_{\gamma_n,\mathbf{k}_n}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1} \prod_{t=1}^{s} G_{a_t b_t}^{0}\right| \\ + &O(\zeta N^{-\zeta+2}),\end{aligned}$$

where

$$\mathbf{k}_1 \in \mathbb{N}^s, \qquad \mathbf{k}_2 \in \mathbb{N}^{s+|\mathbf{k}_1|}, \qquad \mathbf{k}_3 \in \mathbb{N}^{s+|\mathbf{k}_1|+|\mathbf{k}_2|}, \qquad \dots, \qquad \text{and} \tag{6.30}$$
$$5 \leq |\mathbf{k}_i| \leq \frac{2\zeta}{\phi}.$$

Again using (6.19), (6.21) and $s, \zeta = O(\varphi)$, we obtain that

$$\begin{aligned}\left|\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{\gamma_{\max}}\right| \leq \left|\mathbb{E}\prod_{t=1}^{s} G_{a_t b_t}^{0}\right| &+ C_7^s \max_{\mathbf{k},n} (N^{-2})^n (N^{-\phi/20})^{\sum_i |\mathbf{k}_i|} \\ &\times \sum_{\gamma_1,\dots,\gamma_n} \left|\mathbb{E}\mathcal{P}_{\gamma_n,\mathbf{k}_n}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1} \prod_{t=1}^{s} G_{a_t b_t}^{0}\right| \\ &+ O(\zeta N^{-\zeta+2})\end{aligned} \tag{6.31}$$

for some constant $C_7 > 0$. We remark that the above estimate still holds if we replace some of the G entries with $\overline{G}$ entries, since we have only used the absolute bounds for the relevant entries.

Now we use Lemma 6.6 and (6.31) to complete the proof of Lemma 3.17, Lemma 5.2 and Lemma 5.5.

PROOF OF LEMMA 3.17. We apply (6.31) to $G_{ab}\overline{G}_{ab}$ with $s = 2$ and $\zeta = 3$. Recall that $\tilde{X}$ is a bounded support matrix with $q = O(N^{-1/2}\log N)$. Then by (3.19), we have with ξ_1-high probability,

$$|\tilde{G}_{ab}| \le \varphi^{C_1+1}\left(\sqrt{\frac{\operatorname{Im} m_{2c}(z)}{N\eta}} + \frac{1}{N\eta}\right), \qquad a \ne b \tag{6.32}$$

for $z \in S(c_1, C_0, C_1)$, where we used that

$$N^{-1/2} \le C\sqrt{\frac{\operatorname{Im} m_{2c}(z)}{N\eta}},$$

which follows from (3.15). On the other hand, we have the trivial bound $\max_{a,b}|\tilde{G}_{ab}| \le N$ on the bad event [see (A.5)]. Hence we can get the bound

$$\mathbb{E}|\tilde{G}_{ab}|^2 \le C\varphi^{2C_1+2}\left(\frac{\operatorname{Im} m_{2c}(z)}{N\eta} + \frac{1}{(N\eta)^2}\right), \qquad a \ne b.$$

Again with (3.15), it is easy to check that the right-hand side is larger than N^{-1}. Thus the remainder term $O(N^{-\zeta+2})$ in (6.31) is negligible.

It remains to handle the second term on the right-hand side of (6.31). Let $\Phi(i_t, \mu_t) = \gamma_t$. Then we have

$$\begin{aligned} &\max_{\gamma_1,\dots,\gamma_n : a,b \notin \bigcup_{1\le t\le n}\{i_t,\mu_t\}} \left|\mathbb{E}\mathcal{P}_{\gamma_n,\mathbf{k}_n}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}(\tilde{G}_{ab}\overline{\tilde{G}}_{ab})\right| \\ &\qquad \le C\varphi^{2C_1+2}\left(\frac{\operatorname{Im} m_{2c}(z)}{N\eta} + \frac{1}{(N\eta)^2}\right), \end{aligned} \tag{6.33}$$

since $\mathcal{P}_{\gamma_n,\mathbf{k}_n}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\tilde{G}_{ab}\overline{\tilde{G}}_{ab}$ is a finite sum of the products of the matrix entries of $\tilde{G}$ and $\overline{\tilde{G}}$, and there are at least two off diagonal terms in each product. This bound immediately gives that

$$(N^{-2})^n \sum_{\gamma_1,\gamma_2,\dots,\gamma_n} \left|\mathbb{E}\mathcal{P}_{\gamma_n,\mathbf{k}_n}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}(\tilde{G}_{ab}\overline{\tilde{G}}_{ab})\right| \le \varphi^C\left(\frac{\operatorname{Im} m_{1c}(z)}{N\eta} + \frac{1}{(N\eta)^2}\right)$$

for some large enough constant $C > 0$. Here, if $\{a,b\} \cap \{i_t,\mu_t\} \ne \varnothing$ for some $1 \le t \le n$, we bound the above sum by N^{-1}, which is due to the lost of a free index. Plugging this estimate into (6.31), we conclude Lemma 3.17. $\square$

PROOF OF LEMMA 5.2. For simplicity, instead of (5.7), we shall prove that

$$
\begin{aligned}
&\left|\mathbb{E}\big(m_2(z)-m_{2c}(z)\big)^p\right| \\
&\quad\leq \left|\mathbb{E}\big(\tilde{m}_2(z)-m_{2c}(z)\big)^p\right| + (Cp)^{Cp}\big(J^2+K+N^{-1}\big)^p.
\end{aligned} \tag{6.34}
$$

The proof for (5.7) is exactly the same but with slightly heavier notation.

Define a function $f(I,J)$ such that

$$
\sum_{I,J} f(I,J)=1, \qquad f(I,J)\geq 0 \tag{6.35}
$$

for $I=(a_1,a_2,\dots,a_s)$ and $J=(b_1,b_2,\dots,b_s)$. Since $\mathcal{A}$ are independent of a_t and b_t $(1\leq t\leq s)$, we may consider a linear combination of (6.31) with coefficients given by $f(I,J)$. Moreover, with (6.22), we can extend (6.31) to the product of $(G-\Pi)$ entries, that is,

$$
\begin{aligned}
&\left|\mathbb{E}\sum_{I,J} f(I,J)\prod_{t=1}^{s}(G-\Pi)_{a_t b_t}\right| \\
&\quad\leq \left|\mathbb{E}\sum_{I,J} f(I,J)\prod_{t=1}^{s}(\tilde{G}-\Pi)_{a_t b_t}\right| \\
&\qquad + C_8^s \max_{\mathbf{k},n,\gamma}\big(N^{-\phi/20}\big)^{\sum_i |\mathbf{k}_i|} \\
&\qquad \times \left|\mathbb{E}\sum_{I,J} f(I,J)\tilde{\mathcal{P}}_{\gamma_n,\mathbf{k}_n}\cdots\tilde{\mathcal{P}}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{s}(\tilde{G}-\Pi)_{a_t b_t}\right| \\
&\qquad + O\big(\zeta N^{-\zeta+2}\big)
\end{aligned} \tag{6.36}
$$

for some constant $C_8>0$. If we take $a_t=b_t\in\mathcal{I}_2$, $s=p$, $\zeta=p+2$ and $f(I,J)=N^{-p}\prod\delta_{a_t b_t}$, it is easy to check that

$$
\mathbb{E}\sum_{I,J} f(I,J)\prod_{t=1}^{s}\big(G^{\alpha}-\Pi\big)_{a_t b_t}=\mathbb{E}\big(m_2^{\alpha}-m_{2c}\big)^p, \qquad \alpha=0,\gamma_{\max}. \tag{6.37}
$$

Now to conclude (6.34), it suffices to control the second term on the right-hand side of (6.36). We consider the terms

$$
\tilde{\mathcal{P}}_{\gamma_n,\mathbf{k}_n}\cdots\tilde{\mathcal{P}}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}(\tilde{G}_{\mu_t\mu_t}-m_{2c}) \tag{6.38}
$$

for $\mathbf{k}_1,\dots,\mathbf{k}_n$ satisfying (6.30). By definition of $\tilde{\mathcal{P}}$, (6.38) is a sum of at most $C^{\sum|\mathbf{k}_i|}$ products of $\tilde{G}_{\mu\nu}$ and $(\tilde{G}_{\mu\mu}-m_{2c})$ terms, where the total number of $\tilde{G}_{\mu\nu}$ and $(\tilde{G}_{\mu\mu}-m_{2c})$ terms in each product is $\sum|\mathbf{k}_i|+p=O(\varphi^2)$. Due to the rough bound (6.9), (6.38) is always bounded by $N^{O(\varphi^2)}$. Then by the assumption that

(5.6) and (6.8) hold with ξ_1-high probability with $\xi_1 \geq 3$, we see that the expectation over the event that (5.6) or (6.8) does not hold is negligible. Furthermore, for each product in (6.38) and any $1 \leq t \leq p$, there are two μ_t's in the indices of G. These two μ_t's can only appear as (1) $(\tilde{G}_{\mu_t\mu_t} - m_{2c})$ in the product, or (2) $G_{\mu_t a}G_{b\mu_t}$, where a, b come from some γ_k and γ_l via $\tilde{\mathcal{P}}$ (see Definition 6.3). Then after averaging over $N^{-p}\sum_{\mu_1,\dots,\mu_p}$, this term becomes (1) $\tilde{m}_2 - m_{2c}$, which is bounded by K by (5.6), or (2) $N^{-1}\sum_{\mu_t} G_{\mu_t a}G_{b,\mu_t}$, which is bounded by $J^2 + CN^{-1}$ by (5.6). For any other G's in the product with no μ_t, we simply bound them by C using (6.8). Then, for any fixed $\gamma_1, \dots, \gamma_n$, $\mathbf{k}_1, \dots, \mathbf{k}_n$, we have proved that

$$\begin{aligned}\text{(6.39)} \qquad & \left|\frac{1}{N^p}\sum_{\mu_1,\dots,\mu_p}\mathbb{E}\tilde{\mathcal{P}}_{\gamma_n,\mathbf{k}_n}\cdots\tilde{\mathcal{P}}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}(\tilde{G}_{\mu_t\mu_t} - m_{2c})\right| \\ & \qquad \leq C^{\sum|\mathbf{k}_i|+p}\big(J^2 + K + N^{-1}\big)^p.\end{aligned}$$

Together with (6.36), this concludes (6.34). $\square$

Recall that with Lemma 5.2, we can prove Theorem 3.15 (see Section 5). Now we prove Lemma 5.5 with the help of Theorem 3.15.

PROOF OF LEMMA 5.5. For simplicity, we only prove (5.14). The proof for (5.15) is similar. By (A.6), we have

$$\text{(6.40)} \qquad \|\mathcal{G}_2(z)\|_{\mathrm{HS}}^2 = \sum_{\mu,\nu}|G_{\mu\nu}|^2 = \frac{N\operatorname{Im} m_2(z)}{\eta}.$$

Hence it is equivalent to prove that

$$\text{(6.41)} \qquad \left|\mathbb{E}F\Big(\eta^2\sum_{\mu,\nu}G_{\mu\nu}\overline{G}_{\mu\nu}\Big) - \mathbb{E}F\Big(\eta^2\sum_{\mu,\nu}\tilde{G}_{\mu\nu}\overline{\tilde{G}}_{\mu\nu}\Big)\right| \leq N^{-\phi+C_5\varepsilon}$$

for $z = E + i\eta$ with $E \in I_\varepsilon$ and $\eta = N^{-2/3-\varepsilon}$. Corresponding to the notation in (6.3), we denote

$$\begin{aligned}\text{(6.42)} \qquad & x^S := \eta^2\sum_{\mu,\nu}S_{\mu\nu}\overline{S}_{\mu\nu}, \qquad x^R := \eta^2\sum_{\mu,\nu}R_{\mu\nu}\overline{R}_{\mu\nu}, \\ & x^T := \eta^2\sum_{\mu,\nu}T_{\mu\nu}\overline{T}_{\mu\nu}.\end{aligned}$$

Applying (6.40) to S, T and using (3.26) and (3.15), we get that with high probability

$$\text{(6.43)} \qquad \max_{\gamma}\{|x^S| + |x^T|\} \leq N^{C\varepsilon}$$

for some constant $C > 0$. Since the rank of $H^\gamma - Q$ is at most 2, by the Cauchy interlacing theorem we have

$$|\operatorname{Tr} S - \operatorname{Tr} R| \le C\eta^{-1}. \tag{6.44}$$

Together with (6.43), we also get that

$$\max_\gamma |x^R| \le N^{C\varepsilon} \qquad \text{with high probability.} \tag{6.45}$$

By (3.19), (6.7) and the expansion (6.6), we get that with high probability,

$$\max_\gamma \{|S_{\mu\nu}| + |R_{\mu\nu}|\} \le N^{-\phi+C\varepsilon} + C\delta_{\mu\nu}. \tag{6.46}$$

Moreover, by (6.9) we have the trivial bounds

$$|x^S|+|x^R| = O(\eta^2 N^2 \eta^{-2}) = O(N^2), \qquad \max_{\mu,\nu}\{|S_{\mu\nu}|+|R_{\mu\nu}|\} = O(N), \tag{6.47}$$

on the bad event. Since the bad event holds with exponentially small probability, we can ignore it in the proof.

Applying the Lindeberg replacement strategy, we get that

$$\begin{aligned} &\mathbb{E}F\biggl(\eta^2 \sum_{\mu,\nu} G_{\mu\nu}\overline{G}_{\mu\nu}\biggr) - \mathbb{E}F\biggl(\eta^2 \sum_{\mu,\nu} \tilde{G}_{\mu\nu}\overline{\tilde{G}}_{\mu\nu}\biggr) \\ &\quad = \sum_{\gamma=1}^{\gamma_{\max}} [\mathbb{E}F(x^S) - \mathbb{E}F(x^T)]. \end{aligned} \tag{6.48}$$

From the Taylor expansion, we have

$$\begin{aligned} &F(x^S) - F(x^R) \\ &\quad = \sum_{l=1}^{2} \frac{1}{l!} F^{(l)}(x^R)(x^S - X^R)^s + \frac{1}{3!} F^{(3)}(\zeta_S)(x^S - x^R)^3, \end{aligned} \tag{6.49}$$

where ζ_S lies between x^S and x^R. We have a similar expansion for $F(x^T) - F(x^R)$ with ζ_S replaced by ζ_T.

Let $\Phi(i,\mu) = \gamma$ and fix $m \in \mathbb{N}$. We perform the expansion (6.5) and use Lemma 6.1 to get that with ξ_1-high probability,

$$S_{a_t b_t} = \sum_{0\le k\le m} (-\sqrt{\sigma_i} x_{i\mu})^k \mathcal{P}_k R_{a_t b_t} + O(C^m N^{-m\phi}). \tag{6.50}$$

Using this expansion and bound (6.8), we have that with ξ_1-high probability,

$$\prod_{t=1}^{s} S_{a_t b_t} = \sum_{0\le k\le ms} \sum_{\mathbf{k}\in I^s_{m,k}} \biggl(\mathcal{P}_{\mathbf{k}} \prod_{t=1}^{s} R_{a_t b_t}\biggr)(-\sqrt{\sigma_i} x_{i\mu})^k + O(C^{m+s} N^{-m\phi}), \tag{6.51}$$

where

$$\mathbf{k} := (k_1, \dots, k_s), \qquad I^s_{m,k} = \Bigl\{\mathbf{k} \in \mathbb{N}^s : 0 \le k_i \le m, \sum k_i = k\Bigr\}. \tag{6.52}$$

From the above definition, we have the rough bound

$$|I_{m,k}^s| \le s^k. \tag{6.53}$$

By Lemma 6.5 and (6.53), the $k > m$ terms in (6.51) can be bounded by

$$\begin{aligned}\left|\sum_{k>m}\sum_{\mathbf{k}\in I_{m,k}^s}\left(\mathcal{P}_{\mathbf{k}}\prod_{t=1}^s R_{a_t b_t}\right)(-\sqrt{\sigma_i}x_{i\mu})^k\right| &\le \sum_{k>m} s^k \tilde{C}_6^{k+s+1}(CN^{-\phi})^k \\ &= O\big(s^m C^{m+s} N^{-m\phi}\big),\end{aligned}$$

with ξ_1-high probability. Hence with ξ_1-high probability,

$$\begin{aligned}\prod_{t=1}^s S_{a_t b_t} = \prod_{t=1}^s R_{a_t b_t} &+ \sum_{1\le k\le m}(-\sqrt{\sigma_i}x_{i\mu})^k\left(\sum_{\mathbf{k}\in I_{m,k}^s}\mathcal{P}_{\mathbf{k}}\prod_{t=1}^s R_{a_t b_t}\right) \\ &+ O\big(s^m C^{m+s}N^{-m\phi}\big).\end{aligned} \tag{6.54}$$

Similarly, we also have with ξ_1-high probability,

$$\begin{aligned}\prod_{t=1}^s T_{a_t b_t} = \prod_{t=1}^s R_{a_t b_t} &+ \sum_{1\le k\le m}(-\sqrt{\sigma_i}\tilde{x}_{i\mu})^k\left(\sum_{\mathbf{k}\in I_{m,k}^s}\mathcal{P}_{\mathbf{k}}\prod_{t=1}^s R_{a_t b_t}\right) \\ &+ O\big(s^m C^{m+s}N^{-m\phi}\big).\end{aligned} \tag{6.55}$$

Again we can replace some of the resolvent entries with their complex conjugates by modifying the notation slightly.

Now we apply (6.54) and (6.55) with $s = 2$ and $m := 3/\phi$ to get that

$$\begin{aligned}x^S = x^R &+ \sum_{1\le k\le 3/\phi}\left(\sum_{\mathbf{k}\in I_{3/\phi,k}^2}\eta^2\sum_{\mu,\nu}\mathcal{P}_{\gamma,\mathbf{k}}(R_{\mu\nu}\overline{R}_{\mu\nu})\right)(-\sqrt{\sigma_i}x_{i\mu})^k \\ &+ O(CN^{-3}),\end{aligned} \tag{6.56}$$

$$\begin{aligned}x^T = x^R &+ \sum_{1\le k\le 3/\phi}\left(\sum_{\mathbf{k}\in I_{3/\phi,k}^2}\eta^2\sum_{\mu,\nu}\mathcal{P}_{\gamma,\mathbf{k}}(R_{\mu\nu}\overline{R}_{\mu\nu})\right)(-\sqrt{\sigma_i}\tilde{x}_{i\mu})^k \\ &+ O(CN^{-3}),\end{aligned} \tag{6.57}$$

with high probability. To control the second term in (6.56), we need the following lemma.

LEMMA 6.8. *For any fixed $\mathbf{k} \ne 0$, $\mathbf{k} \in I_{3/\phi,k}^2$, and $p = O(1)$ with $p \in 2\mathbb{N}$, we have*

$$\mathbb{E}\left|\sum_{\mu,\nu}\mathcal{P}_{\gamma,\mathbf{k}}(R_{\mu\nu}\overline{R}_{\mu\nu})\right|^p \le \big(N^{1+C\varepsilon}\big)^p. \tag{6.58}$$

PROOF. This is (6.89) of [33], we can repeat the proof there with minor modifications. In fact, its proof is similar to the one for Lemma 5.2 given above. Thus we omit the details. □

Given (6.58), with Markov inequality we find that for any fixed $\mathbf{k} \neq 0$ with $\mathbf{k} \in I^2_{3/\phi,k}$, there exists constant $C > 0$ such that

$$|\mathcal{P}_{\gamma,\mathbf{k}} x^R| := \left|\eta^2 \sum_{\mu,\nu} \mathcal{P}_{\gamma,\mathbf{k}}(R_{\mu\nu}\overline{R}_{\mu\nu})\right| \leq N^{-1/3+C\varepsilon}, \tag{6.59}$$

with probability with $1 - N^{-A}$ for any fixed $A > 0$, where we used that $\eta = N^{-2/3-\varepsilon}$. Combining (6.56), (6.59), (6.46) and (3.25), we see that there exists a constant $C > 0$ such that

$$\mathbb{E}|x^S - x^R|^3 \leq N^{-5/2+C\varepsilon}, \tag{6.60}$$

for sufficiently large N independent of γ, where we used the bound (6.47) on the bad event with probability $O(N^{-A})$. Since ζ_S is between x^S and x^R, we have $|\zeta_S| \leq N^{C\varepsilon}$ with high probability by (6.43). Together with (6.60) and the assumption (5.13), we get

$$\left|\sum_{\gamma=1}^{\gamma_{\max}} \mathbb{E}\big[F^{(3)}(\zeta_S)\big(x^S - x^R\big)^3\big]\right| \leq N^{-1/2+C\varepsilon}. \tag{6.61}$$

We have a similar estimate for $\mathbb{E}[F^{(3)}(\zeta_T)(x^T - x^R)^3]$.

Now it only remains to deal with the first term on the right-hand side of (6.49). Using (6.56), (6.57) and the fact that the first four moments of $x_{i\mu}$ and $\tilde{x}_{i\mu}$ match, we obtain that for $l = 1, 2$,

$$\begin{aligned}
&\big|\mathbb{E}\big[F^{(l)}\big(x^R\big)\big(x^S - x^R\big)^l\big] - \mathbb{E}\big[F^{(l)}\big(x^R\big)\big(x^T - x^R\big)^l\big]\big| \\
&\quad \leq \left|\sum_{k=5}^{6/\phi} \sum_{\sum_{t=1}^s |\mathbf{k}_t| = k} \sum_{\mathbf{k}_t \in I^{2l}_{3/\phi,k}} \mathbb{E}\prod_{t=1}^{s}\big(\mathcal{P}_{\gamma,\mathbf{k}_t} x^R\big)\right| \big(|\mathbb{E}(-\sqrt{\sigma_i}x_{i\mu})^k| + |\mathbb{E}(-\sqrt{\sigma_i}\tilde{x}_{i\mu})^k|\big) \\
&\qquad + O\big(CN^{-3+C\varepsilon}\big).
\end{aligned}$$

Recall that (3.25) holds for $x_{i\mu}$ and $\tilde{x}_{i\mu}$, $x_{i\mu}$ has support bounded by $O(N^{-\phi})$, and $\tilde{x}_{i\mu}$ has support bounded by $O(N^{-1/2}\log N)$. Then it is easy to check that $|\mathbb{E}(-\tilde{x}_{i\mu})^k| \leq (\log N)^C N^{-5/2}$ and $|\mathbb{E}(-x_{i\mu})^k| \leq (\log N)^C N^{-2-\phi}$ for $k \geq 5$. Using (6.59), we obtain that for $1 \leq l \leq 2$

$$\big|\mathbb{E}\big[F^{(l)}\big(x^R\big)\big(x^S - x^R\big)^l\big] - \mathbb{E}\big[F^{(l)}\big(x^R\big)\big(x^T - x^R\big)^l\big]\big| \leq N^{-2-\phi+C_5\varepsilon}.$$

Together with (6.48), (6.49) and (6.61), this concludes the proof. □

APPENDIX A: PROOF OF LEMMA 3.11

Throughout the proof, we denote the spectral parameter by $z = E + i\eta$.

A.1. Basic tools. In this subsection, we collect some tools that will be used in the proof. For simplicity, we denote $Y := DX$.

DEFINITION A.1 (Minors). For $\mathbb{T} \subseteq \mathcal{I}$, we define the minor $H^{(\mathbb{T})} := (H_{ab} : a, b \in \mathcal{I} \setminus \mathbb{T})$ obtained by removing all rows and columns of H indexed by $a \in \mathbb{T}$. Note that we keep the names of indices of H when defining $H^{(\mathbb{T})}$, that is, $(H^{(\mathbb{T})})_{ab} = \mathbf{1}_{\{a,b \notin \mathbb{T}\}} H_{ab}$. Correspondingly, we define the Green function

$$G^{(\mathbb{T})} := (H^{(\mathbb{T})})^{-1} = \begin{pmatrix} z\mathcal{G}_1^{(\mathbb{T})} & \mathcal{G}_1^{(\mathbb{T})} Y^{(\mathbb{T})} \\ (Y^{(\mathbb{T})})^* \mathcal{G}_1^{(\mathbb{T})} & \mathcal{G}_2^{(\mathbb{T})} \end{pmatrix} = \begin{pmatrix} z\mathcal{G}_1^{(\mathbb{T})} & Y^{(\mathbb{T})} \mathcal{G}_2^{(\mathbb{T})} \\ \mathcal{G}_2^{(\mathbb{T})} (Y^{(\mathbb{T})})^* & \mathcal{G}_2^{(\mathbb{T})} \end{pmatrix},$$

and the partial traces

$$m_1^{(\mathbb{T})} := \frac{1}{M} \operatorname{Tr} \mathcal{G}_1^{(\mathbb{T})} = \frac{1}{Mz} \sum_{i \notin \mathbb{T}} G_{ii}^{(\mathbb{T})}, \qquad m_2^{(\mathbb{T})} := \frac{1}{N} \operatorname{Tr} \mathcal{G}_2^{(\mathbb{T})} = \frac{1}{N} \sum_{\mu \notin \mathbb{T}} G_{\mu\mu}^{(\mathbb{T})},$$

where we adopt the convention that $G_{ab}^{(T)} = 0$ if $a \in \mathbb{T}$ or $b \in \mathbb{T}$. We will abbreviate $(\{a\}) \equiv (a)$, $(\{a, b\}) \equiv (ab)$, and

$$\sum_{a \notin \mathbb{T}} \equiv \sum_{a}^{(\mathbb{T})}, \qquad \sum_{a,b \notin \mathbb{T}} \equiv \sum_{a,b}^{(\mathbb{T})}.$$

LEMMA A.2 (Resolvent identities). (i) *For $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_2$, we have*

$$\frac{1}{G_{ii}} = -1 - (Y G^{(i)} Y^*)_{ii}, \qquad \frac{1}{G_{\mu\mu}} = -z - (Y^* G^{(\mu)} Y)_{\mu\mu}. \tag{A.1}$$

(ii) *For $i \neq j \in \mathcal{I}_1$ and $\mu \neq \nu \in \mathcal{I}_2$, we have*

$$G_{ij} = G_{ii} G_{jj}^{(i)} (Y G^{(ij)} Y^*)_{ij}, \qquad G_{\mu\nu} = G_{\mu\mu} G_{\nu\nu}^{(\mu)} (Y^* G^{(\mu\nu)} Y)_{\mu\nu}. \tag{A.2}$$

For $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_2$, we have

$$\begin{aligned} G_{i\mu} &= G_{ii} G_{\mu\mu}^{(i)} (-Y_{i\mu} + (Y G^{(i\mu)} Y)_{i\mu}), \\ G_{\mu i} &= G_{\mu\mu} G_{ii}^{(\mu)} (-Y_{\mu i}^* + (Y^* G^{(\mu i)} Y^*)_{\mu i}). \end{aligned} \tag{A.3}$$

(iii) *For $a \in \mathcal{I}$ and $b, c \in \mathcal{I} \setminus \{a\}$,*

$$G_{bc}^{(a)} = G_{bc} - \frac{G_{ba} G_{ac}}{G_{aa}}, \qquad \frac{1}{G_{bb}} = \frac{1}{G_{bb}^{(a)}} - \frac{G_{ba} G_{ab}}{G_{bb} G_{bb}^{(a)} G_{aa}}. \tag{A.4}$$

(iv) *All of the above identities hold for $G^{(\mathbb{T})}$ instead of G for $\mathbb{T} \subset \mathcal{I}$.*

PROOF. All these identities can be proved using Schur's complement formula. The reader can refer to, for example, [31], Lemma 4.4. □

LEMMA A.3. *Fix constants $c_0, C_0, C_1 > 0$. The following estimates hold uniformly for all $z \in S(c_0, C_0, C_1)$:*

$$\|G\| \le C\eta^{-1}, \qquad \|\partial_z G\| \le C\eta^{-2}. \tag{A.5}$$

Furthermore, we have the following identities:

$$\sum_{\mu\in\mathcal{I}_2} |G_{\nu\mu}|^2 = \sum_{\mu\in\mathcal{I}_2} |G_{\mu\nu}|^2 = \frac{\operatorname{Im} G_{\nu\nu}}{\eta}, \tag{A.6}$$

$$\sum_{i\in\mathcal{I}_1} |G_{ji}|^2 = \sum_{i\in\mathcal{I}_1} |G_{ij}|^2 = \frac{|z|^2}{\eta} \operatorname{Im}\left(\frac{G_{jj}}{z}\right), \tag{A.7}$$

$$\sum_{i\in\mathcal{I}_1} |G_{\mu i}|^2 = \sum_{i\in\mathcal{I}_1} |G_{i\mu}|^2 = G_{\mu\mu} + \frac{\bar{z}}{\eta} \operatorname{Im} G_{\mu\mu}, \tag{A.8}$$

$$\sum_{\mu\in\mathcal{I}_2} |G_{i\mu}|^2 = \sum_{\mu\in\mathcal{I}_2} |G_{\mu i}|^2 = \frac{G_{ii}}{z} + \frac{\bar{z}}{\eta} \operatorname{Im}\left(\frac{G_{ii}}{z}\right). \tag{A.9}$$

All of the above estimates remain true for $G^{(\mathbb{T})}$ instead of G for any $\mathbb{T} \subseteq \mathcal{I}$.

PROOF. These estimates and identities can be proved through simple calculations with (3.8), (3.9) and (3.10). We refer the reader to [31], Lemma 4.6, and [53], Lemma 3.5. □

LEMMA A.4. *Fix constants $c_0, C_0, C_1 > 0$. For any $\mathbb{T} \subseteq \mathcal{I}$, the following bounds hold uniformly in $z \in S(c_0, C_0, C_1)$:*

$$\left|m_2 - m_2^{(\mathbb{T})}\right| \le \frac{2|\mathbb{T}|}{N\eta} \tag{A.10}$$

and

$$\left|\frac{1}{N}\sum_{i=1}^{M} \sigma_i \left(G_{ii}^{(\mathbb{T})} - G_{ii}\right)\right| \le \frac{C|\mathbb{T}|}{N\eta}, \tag{A.11}$$

where $C > 0$ is a constant depending only on τ.

PROOF. For $\mu \in \mathcal{I}_2$, we have

$$\left|m_2 - m_2^{(\mu)}\right| = \frac{1}{N}\left|\sum_{\nu\in\mathcal{I}_2} \frac{G_{\nu\mu}G_{\mu\nu}}{G_{\mu\mu}}\right| \le \frac{1}{N|G_{\mu\mu}|} \sum_{\nu\in\mathcal{I}_2} |G_{\nu\mu}|^2 = \frac{\operatorname{Im} G_{\mu\mu}}{N\eta|G_{\mu\mu}|} \le \frac{1}{N\eta},$$

where in the first step we used (A.4), and in the second and third steps we used the equality (A.6). Similarly, using (A.4) and (A.8) we get

$$|m_2 - m_2^{(i)}| = \frac{1}{N}\left|\sum_{\nu\in\mathcal{I}_2}\frac{G_{\nu i}G_{i\nu}}{G_{ii}}\right| \le \frac{1}{N|G_{ii}|}\left(\frac{G_{ii}}{z} + \frac{\bar{z}}{\eta}\operatorname{Im}\left(\frac{G_{ii}}{z}\right)\right) \le \frac{2}{N\eta}.$$

Then we can prove (A.10) by induction on the indices in $\mathbb{T}$. The proof for (A.11) is similar except that one needs to use the assumption (2.6). □

The following large deviation bounds for bounded supported random variables are proved in [17], Lemma 3.8.

LEMMA A.5. *Let (x_i), (y_i) be independent families of centered and independent random variables, and (A_i), (B_{ij}) be families of deterministic complex numbers. Suppose the entries x_i and y_j have variance at most N^{-1} and satisfies the bounded support condition (3.3) with $q \le N^{-\varepsilon}$ for some constant $\varepsilon > 0$. Then for any fixed $\xi > 0$, the following bounds hold with ξ-high probability*:

$$\left|\sum_i A_i x_i\right| \le \varphi^{\xi}\left[q\max_i |A_i| + \frac{1}{\sqrt{N}}\left(\sum_i |A_i|^2\right)^{1/2}\right], \tag{A.12}$$

$$\left|\sum_{i,j} x_i B_{ij} y_j\right| \le \varphi^{2\xi}\left[q^2 B_d + q B_o + \frac{1}{N}\left(\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right], \tag{A.13}$$

$$\left|\sum_i \bar{x}_i B_{ii} x_i - \sum_i (\mathbb{E}|x_i|^2) B_{ii}\right| \le \varphi^{\xi} q B_d, \tag{A.14}$$

$$\left|\sum_{i\neq j} \bar{x}_i B_{ij} x_j\right| \le \varphi^{2\xi}\left[q B_o + \frac{1}{N}\left(\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right], \tag{A.15}$$

where

$$B_d := \max_i |B_{ii}|, \qquad B_o := \max_{i\neq j}|B_{ij}|.$$

Finally, we have the following lemma, which is a consequence of the Assumption 2.5.

LEMMA A.6. *There exists constants $c_0, \tau' > 0$ such that*

$$\left|1 + m_{2c}(z)\sigma_k\right| \ge \tau' \tag{A.16}$$

for all $z \in S(c_0, C_0, C_1)$ and $1 \le k \le M$.

PROOF. By Assumption 2.5 and the fact $m_{2c}(\lambda_r) \in (-\sigma_1^{-1}, 0)$, we have

$$\left|1 + m_{2c}(\lambda_r)\sigma_k\right| \ge \tau, \qquad 1 \le k \le M.$$

Applying (3.13) to the Stieltjes' transform

$$m_{2c}(z) := \int_{\mathbb{R}} \frac{\rho_{2c}(dx)}{x-z}, \tag{A.17}$$

one can verify that $m_{2c}(z) \sim \sqrt{z-\lambda_r}$ for z close to λ_r. Hence if $\kappa + \eta \leq 2c_0$ for some sufficiently small constant $c_0 > 0$, we have

$$\left|1 + m_{2c}(z)\sigma_k\right| \geq \tau/2.$$

Then we consider the case with $E - \lambda_r \geq c_0$ and $\eta \leq c_1$ for some constant $c_1 > 0$. In fact, for $\eta = 0$ and $E \geq \lambda_r + c_0$, $m_{2c}(E)$ is real and it is easy to verify that $m'_{2c}(E) \geq 0$ using the formula (A.17). Hence we have

$$\left|1 + \sigma_k m_{2c}(E)\right| \geq \left|1 + \sigma_k m_{2c}(\lambda_r + c_0)\right| \geq \tau/2 \qquad \text{for } E \geq \lambda_r + c_0.$$

Using (A.17) again, we can get that

$$\left|\frac{dm_{2c}(z)}{dz}\right| \leq c_0^{-2} \qquad \text{for } E \geq \lambda_r + c_0.$$

So if c_1 is sufficiently small, we have

$$\left|1 + \sigma_k m_{2c}(E + i\eta)\right| \geq \frac{1}{2}\left|1 + \sigma_k m_{2c}(E)\right| \geq \tau/4$$

for $E \geq \lambda_r + c_0$ and $\eta \leq c_1$. Finally, it remains to consider the case with $\eta \geq c_1$. If $\sigma_k \leq |2m_{2c}(z)|^{-1}$, then we have $|1 + \sigma_k m_{2c}(z)| \geq 1/2$. Otherwise, we have $\operatorname{Im} m_{2c}(z) \sim 1$ by (3.15). Together with (3.14), we get that

$$\left|1 + \sigma_k m_{2c}(z)\right| \geq \sigma_k \operatorname{Im} m_{2c}(z) \geq \frac{\operatorname{Im} m_{2c}(z)}{2m_{2c}(z)} \geq \tau'$$

for some constant $\tau' > 0$. $\square$

A.2. Proof of the local laws. Our goal is to prove that G is close to Π in the sense of entrywise and averaged local laws. Hence it is convenient to introduce the following random control parameters.

DEFINITION A.7 (Control parameters). We define the entrywise and averaged errors

$$\Lambda := \max_{a,b\in\mathcal{I}}\left|(G-\Pi)_{ab}\right|, \qquad \Lambda_o := \max_{a\neq b\in\mathcal{I}}|G_{ab}|, \qquad \theta := |m_2 - m_{2c}|. \tag{A.18}$$

Moreover, we define the random control parameter

$$\Psi_\theta := \sqrt{\frac{\operatorname{Im} m_{2c} + \theta}{N\eta}} + \frac{1}{N\eta}, \tag{A.19}$$

and the deterministic control parameter

$$\Psi := \sqrt{\frac{\operatorname{Im} m_{2c}}{N\eta}} + \frac{1}{N\eta}. \tag{A.20}$$

In analogy to [17], Section 3 and [31], Section 5, we introduce the Z variables

$$Z_a^{(\mathbb{T})} := (1-\mathbb{E}_a)\big(G_{aa}^{(\mathbb{T})}\big)^{-1}, \qquad a \notin \mathbb{T},$$

where $\mathbb{E}_a[\cdot] := \mathbb{E}[\cdot \mid H^{(a)}]$, that is, it is the partial expectation over the randomness of the ath row and column of H. By (A.1), we have

$$Z_i = (\mathbb{E}_i - 1)\big(YG^{(i)}Y^*\big)_{ii} = \sigma_i \sum_{\mu,\nu\in\mathcal{I}_2} G_{\mu\nu}^{(i)}\left(\frac{1}{N}\delta_{\mu\nu} - X_{i\mu}X_{i\nu}\right), \tag{A.21}$$

and

$$\begin{aligned} Z_\mu &= (\mathbb{E}_\mu - 1)\big(Y^*G^{(\mu)}Y\big)_{\mu\mu} \\ &= \sum_{i,j\in\mathcal{I}_1} \sqrt{\sigma_i\sigma_j}\, G_{ij}^{(\mu)}\left(\frac{1}{N}\delta_{ij} - X_{i\mu}X_{j\mu}\right). \end{aligned} \tag{A.22}$$

The following lemma plays a key role in the proof of local laws.

LEMMA A.8. *Let $c_0 > 0$ be a sufficiently small constant and fix $C_0, C_1, \xi > 0$. Define the z-dependent event $\Xi(z) := \{\Lambda(z) \le (\log N)^{-1}\}$. Then there exists constant $C > 0$ such that the following estimates hold for all $a \in \mathcal{I}$ and $z \in S(c_0, C_0, C_1)$ with ξ-high probability*:

$$\mathbf{1}(\Xi)\big(\Lambda_o + |Z_a|\big) \le C\varphi^{2\xi}(q + \Psi_\theta), \tag{A.23}$$

$$\mathbf{1}(\eta \ge 1)\big(\Lambda_o + |Z_a|\big) \le C\varphi^{2\xi}(q + \Psi_\theta). \tag{A.24}$$

PROOF. Applying the large deviation Lemma A.5 to Z_i in (A.21), we get that on Ξ,

$$\begin{aligned} |Z_i| &\le C\varphi^{2\xi}\left[q + \frac{1}{N}\left(\sum_{\mu,\nu}|G_{\mu\nu}^{(i)}|^2\right)^{1/2}\right] \\ &= C\varphi^{2\xi}\left[q + \frac{1}{N}\left(\sum_{\mu}\frac{\operatorname{Im} G_{\mu\mu}^{(i)}}{\eta}\right)^{1/2}\right] = C\varphi^{2\xi}\left[q + \sqrt{\frac{\operatorname{Im} m_2^{(i)}}{N\eta}}\right], \end{aligned} \tag{A.25}$$

holds with ξ-high probability, where we used (2.6), (A.6) and the fact that $\max_{a,b}|G_{ab}| = O(1)$ on event Ξ. Now by (A.18), (A.19) and the bound (A.10), we have that

$$\sqrt{\frac{\operatorname{Im} m_2^{(i)}}{N\eta}} = \sqrt{\frac{\operatorname{Im} m_{2c} + \operatorname{Im}(m_2^{(i)} - m_2) + \operatorname{Im}(m_2 - m_{2c})}{N\eta}} \le C\Psi_\theta. \tag{A.26}$$

Together with (A.25), we conclude that

$$\mathbf{1}(\Xi)|Z_i| \le C\varphi^{2\xi}(q + \Psi_\theta),$$

with ξ-high probability. Similarly, we can prove the same estimate for $\mathbf{1}(\Xi)|Z_\mu|$. In the proof, we also need to use (2.9) and

$$\operatorname{Im}\left(-\frac{d-1}{z}\right) = O(\eta) = O\big(\operatorname{Im} m_{2c}(z)\big).$$

If $\eta \ge 1$, we always have $\max_{a,b}|G_{ab}| = O(1)$ by (A.5). Then repeating the above proof, we obtain that

$$\mathbf{1}(\eta \ge 1)|Z_a| \le C\varphi^{2\xi}(q + \Psi_\theta),$$

with ξ-high probability.

Similarly, using (A.2) and Lemmas A.3–A.5, we can prove that with ξ-high probability,

$$\mathbf{1}(\Xi)\big(|G_{ij}| + |G_{\mu\nu}|\big) \le C\varphi^{2\xi}(q + \Psi_\theta), \tag{A.27}$$

holds uniformly for $i \ne j$ and $\mu \ne \nu$. It remains to prove the bound for $G_{i\mu}$ and $G_{\mu i}$. Using (A.3), the bounded support condition (3.3) for $X_{i\mu}$, the bound $\max_{a,b}|G_{ab}| = O(1)$ on Ξ, Lemma A.3 and Lemma A.5, we get that with ξ-high probability,

$$\begin{aligned}
|G_{i\mu}| &\le C\left(q + \left|\sum_{j,\nu}^{(i\mu)} X_{i\nu} G_{\nu j}^{(i\mu)} X_{j\mu}\right|\right) \\
&\le C\varphi^{2\xi}\left[q + \frac{1}{N}\left(\sum_{j,\nu}^{(i\mu)} |G_{\nu j}^{(i\mu)}|^2\right)^{1/2}\right] \\
&\le C\varphi^{2\xi}\left[q + \frac{1}{N}\left(\sum_{\nu}^{(\mu)}\left(G_{\nu\nu}^{(i\mu)} + \frac{\bar{z}}{\eta}\operatorname{Im} G_{\nu\nu}^{(i\mu)}\right)\right)^{1/2}\right] \\
&\le C\varphi^{2\xi}\left[q + \sqrt{\frac{|m_2^{(i\mu)}|}{N}} + \sqrt{\frac{\operatorname{Im} m_2^{(i\mu)}}{N\eta}}\right].
\end{aligned} \tag{A.28}$$

As in (A.26), we can show that

$$\sqrt{\frac{\operatorname{Im} m_2^{(i\mu)}}{N\eta}} = O(\Psi_\theta). \tag{A.29}$$

For the other term, we have

$$\begin{aligned}
\sqrt{\frac{|m_2^{(i\mu)}|}{N}} &\le \sqrt{\frac{|m_{2c}| + |m_2^{(i\mu)} - m_2| + |m_2 - m_{2c}|}{N}} \\
&\le C\left(\frac{1}{N\sqrt{\eta}} + \sqrt{\frac{\theta}{N}} + \sqrt{\frac{|m_{2c}|}{N}}\right) \le C\Psi_\theta,
\end{aligned} \tag{A.30}$$

where we used (A.10), and that

$$\frac{|m_{2c}|}{N} = O\left(\frac{\operatorname{Im} m_{2c}}{N\eta}\right),$$

since $|m_{2c}| = O(1)$ and $\operatorname{Im} m_{2c} \gtrsim \eta$ by Lemma 3.6. From (A.28), (A.29) and (A.30), we obtain that

$$\mathbf{1}(\Xi)|G_{i\mu}| \le C\varphi^{2\xi}(q + \Psi_\theta),$$

with ξ-high probability. Together with (A.27), we get the estimate in (A.23) for Λ_o. Finally, the estimate (A.24) for Λ_o can be proved in a similar way with the bound $\mathbf{1}(\eta \ge 1)\max_{a,b}|G_{ab}| = O(1)$. □

Our proof of the local law starts with an analysis of the self-consistent equation. Recall that $m_{2c}(z)$ is the solution to the equation $z = f(m)$ for f defined in (2.16).

LEMMA A.9. *Let $c_0 > 0$ be sufficiently small. Fix $C_0 > 0$, $\xi \ge 3$ and $C_1 \ge 8\xi$. Then there exists $C > 0$ such that the following estimates hold uniformly in $z \in S(c_0, C_0, C_1)$ with ξ-high probability*:

$$\mathbf{1}(\eta \ge 1)\big|z - f(m_2)\big| \le C\varphi^{2\xi}\big(q + N^{-1/2}\big), \tag{A.31}$$

$$\mathbf{1}(\Xi)\big|z - f(m_2)\big| \le C\varphi^{2\xi}(q + \Psi_\theta), \tag{A.32}$$

where Ξ is as given in Lemma A.8. Moreover, we have the finer estimates

$$\mathbf{1}(\Xi)\big(z - f(m_2)\big) = \mathbf{1}(\Xi)\big([Z]_1 + [Z]_2\big) + O\big(\varphi^{4\xi}\big(q^2 + \Psi_\theta^2\big)\big), \tag{A.33}$$

with ξ-high probability, where

$$[Z]_1 := \frac{1}{N}\sum_{i\in\mathcal{I}_1}\frac{\sigma_i}{(1 + m_2\sigma_i)^2}Z_i, \qquad [Z]_2 := \frac{1}{N}\sum_{\mu\in\mathcal{I}_2}Z_\mu. \tag{A.34}$$

PROOF. We first prove (A.33), from which (A.32) follows due to (A.23) and (A.16). By (A.1), (A.21) and (A.22), we have

$$\frac{1}{G_{ii}} = -1 - \frac{\sigma_i}{N}\sum_{\mu\in\mathcal{I}_2}G_{\mu\mu}^{(i)} + Z_i = -1 - \sigma_i m_2 + \varepsilon_i \tag{A.35}$$

and

$$\frac{1}{G_{\mu\mu}} = -z - \frac{1}{N}\sum_{i\in\mathcal{I}_1}\sigma_i G_{ii}^{(\mu)} + Z_\mu = -z - \frac{1}{N}\sum_{i\in\mathcal{I}_1}\sigma_i G_{ii} + \varepsilon_\mu, \tag{A.36}$$

where

$$\varepsilon_i := Z_i + \sigma_i\big(m_2 - m_2^{(i)}\big) \quad \text{and} \quad \varepsilon_\mu := Z_\mu + \frac{1}{N}\sum_{i\in\mathcal{I}_1}\sigma_i\big(G_{ii} - G_{ii}^{(\mu)}\big).$$

By (A.10), (A.11) and (A.23), we have for all i and μ,

$$\mathbf{1}(\Xi)\big(|\varepsilon_i| + |\varepsilon_\mu|\big) \le C\varphi^{2\xi}(q + \Psi_\theta), \tag{A.37}$$

with ξ-high probability. Then using (A.36), we get that for any μ and ν,

$$\mathbf{1}(\Xi)(G_{\mu\mu} - G_{\nu\nu}) = \mathbf{1}(\Xi)G_{\mu\mu}G_{\nu\nu}(\varepsilon_\nu - \varepsilon_\mu) = O\big(\varphi^{2\xi}(q + \Psi_\theta)\big), \tag{A.38}$$

with ξ-high probability. This implies that

$$\mathbf{1}(\Xi)|G_{\mu\mu} - m_2| \le C\varphi^{2\xi}(q + \Psi_\theta), \qquad \mu \in \mathcal{I}_2, \tag{A.39}$$

with ξ-high probability.

Now we plug (A.35) into (A.36) and take the average $N^{-1}\sum_\mu$. Note that we can write

$$\frac{1}{G_{\mu\mu}} = \frac{1}{m_2} - \frac{1}{m_2^2}(G_{\mu\mu} - m_2) + \frac{1}{m_2^2}(G_{\mu\mu} - m_2)^2 \frac{1}{G_{\mu\mu}}.$$

After taking the average, the second term on the right-hand side vanishes and the third term provides a $O(\varphi^{4\xi}(q + \Psi_\theta)^2)$ factor by (A.39). On the other hand, using (A.4) and (A.23) we get that

$$1(\Xi)\left|\frac{1}{N}\sum_{i\in\mathcal{I}_1}\sigma_i\big(G_{ii}^{(\mu)} - G_{ii}\big)\right| \le 1(\Xi)\frac{1}{N}\sum_{i\in\mathcal{I}_1}\sigma_i\left|\frac{G_{i\mu}G_{\mu i}}{G_{\mu\mu}}\right| \le C\varphi^{4\xi}(q + \Psi_\theta)^2$$

and

$$\mathbf{1}(\Xi)\big|m_2 - m_2^{(i)}\big| \le \mathbf{1}(\Xi)\frac{1}{N}\sum_{\mu\in\mathcal{I}_2}\left|\frac{G_{\mu i}G_{i\mu}}{G_{ii}}\right| \le C\varphi^{4\xi}(q + \Psi_\theta)^2,$$

with ξ-high probability. Hence the average of (A.36) gives

$$\mathbf{1}(\Xi)\frac{1}{m_2} = \mathbf{1}(\Xi)\left\{-z + \frac{1}{N}\sum_{i\in\mathcal{I}_1}\frac{\sigma_i}{1 + \sigma_i m_2 - Z_i + O(\varphi^{4\xi}(q + \Psi_\theta)^2)} + [Z]_2\right\}$$
$$+ O\big(\varphi^{4\xi}(q + \Psi_\theta)^2\big),$$

with ξ-high probability. Finally, using (A.16) and the definition of Ξ we can expand the fractions in the sum to get that

$$\mathbf{1}(\Xi)\left\{z + \frac{1}{m_2} - \frac{1}{N}\sum_{i\in\mathcal{I}_1}\frac{\sigma_i}{1 + \sigma_i m_2}\right\} = \mathbf{1}(\Xi)\big([Z]_1 + [Z]_2\big) + O\big(\varphi^{4\xi}(q + \Psi_\theta)^2\big).$$

This concludes (A.33).

Then we prove (A.31). Using the bound $\mathbf{1}(\eta \ge 1)\max_{a,b}|G_{ab}| = O(1)$, it is easy to get that $|m_2| = O(1)$ and $\theta = O(1)$. Thus we have $\mathbf{1}(\eta \ge 1)\Psi_\theta = O(N^{-1/2})$ and (A.37) gives

$$\mathbf{1}(\eta \ge 1)\big(|\varepsilon_i| + |\varepsilon_\mu|\big) \le C\varphi^{2\xi}\big(q + N^{-1/2}\big), \tag{A.40}$$

with ξ-high probability. First, we claim that for $\eta \geq 1$,

$$|m_2| \geq \operatorname{Im} m_2 \geq c \qquad \text{with } \xi\text{-high probability} \tag{A.41}$$

for some constant $c > 0$. By the spectral decomposition (3.9), we have

$$\operatorname{Im} G_{ii} = \operatorname{Im} \sum_{k=1}^{M} \frac{z|\xi_k(i)|^2}{\lambda_k - z} = \sum_{k=1}^{M} |\xi_k(i)|^2 \operatorname{Im}\left(-1 + \frac{\lambda_k}{\lambda_k - z}\right) \geq 0.$$

Then by (A.36), $G_{\mu\mu}^{-1}$ is of order $O(1)$ and has an imaginary part $\leq -\eta + O(\varphi^{2\xi}(q + N^{-1/2}))$ with ξ-high probability. This implies that $\operatorname{Im} G_{\mu\mu} \gtrsim \eta$ with ξ-high probability, which concludes (A.41). Next, we claim that

$$|1 + \sigma_i m_2| \geq c' \qquad \text{with } \xi\text{-high probability} \tag{A.42}$$

for some constant $c' > 0$. In fact, if $\sigma_i \leq |2m_2|^{-1}$, we trivially have $|1 + \sigma_i m_2| \geq 1/2$. Otherwise, we have $\sigma_i \gtrsim 1$ [since $|m_2| = O(1)$], which gives that

$$|1 + \sigma_i m_2| \geq \sigma_i \operatorname{Im} m_2 \geq c'.$$

Finally, with (A.40), (A.41) and (A.42), we can repeat the previous arguments to get (A.31). $\square$

The following lemma gives the stability of the equation $z = f(m)$. Roughly speaking, it states that if $z - f(m_2(z))$ is small and $m_2(\tilde{z}) - m_{2c}(\tilde{z})$ is small for $\operatorname{Im} \tilde{z} \geq \operatorname{Im} z$, then $m_2(z) - m_{2c}(z)$ is small. For an arbitrary $z \in S(c_0, C_0, C_1)$, we define the discrete set

$$L(w) := \{z\} \cup \{z' \in S(c_0, C_0, C_1) : \operatorname{Re} z' = \operatorname{Re} z, \operatorname{Im} z' \in [\operatorname{Im} z, 1] \cap (N^{-10}\mathbb{N})\}.$$

Thus, if $\operatorname{Im} z \geq 1$, then $L(z) = \{z\}$; if $\operatorname{Im} z < 1$, then $L(z)$ is a 1-dimensional lattice with spacing N^{-10} plus the point z. Obviously, we have $|L(z)| \leq N^{10}$.

LEMMA A.10. *The self-consistent equation $z - f(m) = 0$ is stable on $S(c_0, C_0, C_1)$ in the following sense. Suppose the z-dependent function δ satisfies $N^{-2} \leq \delta(z) \leq (\log N)^{-1}$ for $z \in S(c_0, C_0, C_1)$ and that δ is Lipschitz continuous with Lipschitz constant $\leq N^2$. Suppose moreover that for each fixed E, the function $\eta \mapsto \delta(E + i\eta)$ is nonincreasing for $\eta > 0$. Suppose that $u_2 : S(c_0, C_0, C_1) \to \mathbb{C}$ is the Stieltjes' transform of a probability measure. Let $z \in S(c_0, C_0, C_1)$ and suppose that for all $z' \in L(z)$ we have*

$$|z - f(u_2)| \leq \delta(z). \tag{A.43}$$

Then we have

$$|u_2(z) - m_{2c}(z)| \leq \frac{C\delta}{\sqrt{\kappa + \eta + \delta}} \tag{A.44}$$

for some constant $C > 0$ independent of z and N, where κ is defined in (3.12).

PROOF. This result is proved in [31], Appendix A.2. □

Note that by Lemma A.10 and (A.31), we immediately get that

$$\mathbf{1}(\eta \geq 1)\theta(z) \leq C\varphi^{2\xi}\big(q + N^{-1/2}\big), \tag{A.45}$$

with ξ-high probability. From (A.24), we obtain the off-diagonal estimate

$$\mathbf{1}(\eta \geq 1)\Lambda_o(z) \leq C\varphi^{2\xi}\big(q + N^{-1/2}\big), \tag{A.46}$$

with ξ-high probability. Using (A.39), (A.35) and (A.45), we get that

$$\mathbf{1}(\eta \geq 1)\big(|G_{ii} - \Pi_{ii}| + |G_{\mu\mu} - m_{2c}|\big) \leq C\varphi^{2\xi}\big(q + N^{-1/2}\big), \tag{A.47}$$

with ξ-high probability, which gives the diagonal estimate. These bounds can be easily generalized to the case $\eta \geq c$ for some fixed $c > 0$. Comparing with (3.19), one can see that the bounds (A.46) and (A.47) are optimal for the $\eta \geq c$ case. Now it remains to deal with the small η case (in particular, the local case with $\eta \ll 1$). We first prove the following weak bound.

LEMMA A.11. *Let $c_0 > 0$ be sufficiently small. Fix $C_0 > 0$, $\xi \geq 3$ and $C_1 \geq 8\xi$. Then there exists $C > 0$ such that with ξ-high probability,*

$$\Lambda(z) \leq C\varphi^{2\xi}\big(\sqrt{q} + (N\eta)^{-1/3}\big), \tag{A.48}$$

holds uniformly in $z \in S(c_0, C_0, C_1)$.

PROOF. One can prove this lemma using a continuity argument as in [9], Section 4.1 or [17], Section 3.6. The key inputs are Lemmas A.8–A.10, the diagonal estimate (A.39) and the estimates (A.45)–(A.47) for the $\eta \geq 1$ case. All the other parts of the proof are essentially the same. □

To get stronger local laws in Lemma 3.11, we need stronger bounds on $[Z]_1$ and $[Z]_2$ in (A.33). They follow from the following *fluctuation averaging lemma*.

LEMMA A.12. *Fix a constant $\xi > 0$. Suppose $q \leq \varphi^{-5\xi}$ and that there exists $\tilde{S} \subseteq S(c_0, C_0, L)$ with $L \geq 18\xi$ such that with ξ-high probability,*

$$\Lambda(z) \leq \gamma(z) \qquad \text{for } z \in \tilde{S}, \tag{A.49}$$

where γ is a deterministic function satisfying $\gamma(z) \leq \varphi^{-\xi}$. Then we have that with $(\xi - \tau_N)$-high probability,

$$\big|[Z]_1(z)\big| + \big|[Z]_2(z)\big| \leq \varphi^{18\xi}\bigg(q^2 + \frac{1}{(N\eta)^2} + \frac{\operatorname{Im} m_{2c}(z) + \gamma(z)}{N\eta}\bigg) \tag{A.50}$$

for $z \in \tilde{S}$, where $\tau_N := 2/\log\log N$.

PROOF. We suppose that the event Ξ holds. The bound for $[Z]_2$ is proved in Lemma 4.1 of [17]. The bound for $[Z]_1$ can be proved in a similar way, except that the coefficients $\sigma_i/(1+m_2\sigma_i)^2$ are random and depend on i. This can be dealt with by writing, for any $i \in \mathcal{I}_1$,

$$m_2 = m_2^{(i)} + \frac{1}{N}\sum_{\mu\in\mathcal{I}_2}\frac{G_{\mu i}G_{i\mu}}{G_{ii}} = m_2^{(i)} + O(\Lambda_o^2),$$

where by Lemma A.8, we have

$$\Lambda_o^2 \le C\varphi^{4\xi}(q^2+\Psi_\theta^2) \le C\varphi^{4\xi}\left(q^2+\frac{1}{(N\eta)^2}+\frac{\operatorname{Im} m_{2c}(z)+\gamma(z)}{N\eta}\right),$$

with ξ-high probability. Then we write

$$\begin{aligned}[Z]_1 &= \frac{1}{N}\sum_{i\in\mathcal{I}_1}\frac{\sigma_i}{(1+m_2^{(i)}\sigma_i)^2}Z_i + O(\Lambda_o^2)\\ &= \frac{1}{N}\sum_{i\in\mathcal{I}_1}(1-\mathbb{E}_i)\left[\frac{\sigma_i}{(1+m_2^{(i)}\sigma_i)^2}G_{ii}^{-1}\right] + O(\Lambda_o^2)\\ &= \frac{1}{N}\sum_{i\in\mathcal{I}_1}(1-\mathbb{E}_i)\left[\frac{\sigma_i}{(1+m_2\sigma_i)^2}G_{ii}^{-1}\right] + O(\Lambda_o^2).\end{aligned} \tag{A.51}$$

The method to bound the first term in the line (A.51) is a slight modification of the one in [17] or the simplified proof given in [16], Appendix B. For a demonstration of this process, one can also refer to the proof of Lemma 4.9 of [53]. Finally, one can use that the event Ξ holds with ξ-high probability by Lemma A.11 to conclude the proof. □

PROOF OF (3.18) AND (3.19). Fix $c_0, C_0 > 0$, $\xi > 3$ and set

$$L := 120\xi, \qquad \tilde{\xi} := 2/\log 2 + \xi.$$

Hence we have $\tilde{\xi} \le 2\xi$ and $L \ge 60\tilde{\xi}$. Then to prove (3.19), it suffices to prove

$$\bigcap_{z\in S(c_0,C_0,L)}\left\{\Lambda(z) \le C\varphi^{20\tilde{\xi}}\left(q+\sqrt{\frac{\operatorname{Im} m_{2c}(z)}{N\eta}}+\frac{1}{N\eta}\right)\right\}, \tag{A.52}$$

with ξ-high probability.

By Lemma A.11, the event Ξ holds with $\tilde{\xi}$-high probability. Then together with Lemma A.12 and (A.33), we get that with $(\tilde{\xi}-\tau_N)$-high probability,

$$\begin{aligned}|z-f(m_2)| &\le C\varphi^{18\tilde{\xi}}\left[q^2+\frac{1}{(N\eta)^2}+\frac{\operatorname{Im} m_{2c}+C\varphi^{2\tilde{\xi}}(\sqrt{q}+(N\eta)^{-1/3})}{N\eta}\right]\\ &\le C\left[\varphi^{20\tilde{\xi}}\left(q^2+\frac{1}{(N\eta)^{4/3}}\right)+\varphi^{18\tilde{\xi}}\frac{\operatorname{Im} m_{2c}}{N\eta}\right],\end{aligned}$$

where we used Young's inequality for the $\sqrt{q}/(N\eta)$ term. Now applying Lemma A.10, we get that with $(\tilde{\xi} - \tau_N)$-high probability,

$$\theta \le C\varphi^{10\tilde{\xi}}\left(q + \frac{1}{(N\eta)^{2/3}}\right) + C\varphi^{18\tilde{\xi}}\frac{\operatorname{Im} m_c}{N\eta\sqrt{\kappa+\eta}}$$
$$\le C\varphi^{18\tilde{\xi}}\left(q + \frac{1}{(N\eta)^{2/3}}\right),$$

where we used (3.15) in the second step. Then using Lemma A.8, (A.35) and (A.39), it is easy to obtain that

$$\Lambda \le C\varphi^{2\tilde{\xi}}(q + \Psi_\theta) + \theta \le C\varphi^{18\tilde{\xi}}\left(q + \frac{1}{(N\eta)^{2/3}}\right) + C\varphi^{2\tilde{\xi}}\sqrt{\frac{\operatorname{Im} m_{2c}}{N\eta}}$$
$$\le \varphi^{20\tilde{\xi}}\left(q + \frac{1}{(N\eta)^{2/3}}\right) + \varphi^{3\tilde{\xi}}\sqrt{\frac{\operatorname{Im} m_{2c}}{N\eta}}$$

uniformly in $z \in S(c_0, C_0, L)$ with $(\tilde{\xi} - \tau_N)$-high probability, which is a better bound than the one in (A.48). We can repeat this process M times, where each iteration yields a stronger bound on Λ which holds with a smaller probability. More specifically, suppose that after k iterations we get the bound

$$\Lambda \le \varphi^{20\tilde{\xi}}\left(q + \frac{1}{(N\eta)^{1-\tau}}\right) + \varphi^{3\tilde{\xi}}\sqrt{\frac{\operatorname{Im} m_{2c}}{N\eta}} \tag{A.53}$$

uniformly in $z \in S(c_0, C_0, L)$ with $\tilde{\xi}'$-high probability. Then by Lemma A.12 and (A.33), we have with $(\tilde{\xi}' - \tau_N)$-high probability,

$$|z - f(m_2)| \le C\varphi^{18\tilde{\xi}}\bigg[q^2 + \frac{1}{(N\eta)^2} + \frac{\operatorname{Im} m_{2c}}{N\eta}$$
$$+ \frac{\varphi^{20\tilde{\xi}}}{N\eta}\left(q + \frac{1}{(N\eta)^{1-\tau}}\right) + \frac{\varphi^{3\tilde{\xi}}}{N\eta}\sqrt{\frac{\operatorname{Im} m_{2c}}{N\eta}}\bigg]$$
$$\le C\left[\varphi^{38\tilde{\xi}}\left(q^2 + \frac{1}{(N\eta)^{2-\tau}}\right) + \varphi^{18\tilde{\xi}}\frac{\operatorname{Im} m_{2c}}{N\eta}\right].$$

Then using Lemma A.10, we get that with $(\tilde{\xi}' - \tau_N)$-high probability,

$$\theta \le C\varphi^{19\tilde{\xi}}\left(q + \frac{1}{(N\eta)^{1-\tau/2}}\right) + C\varphi^{18\tilde{\xi}}\frac{\operatorname{Im} m_c}{N\eta\sqrt{\kappa+\eta}}$$
$$\le C\varphi^{19\tilde{\xi}}\left(q + \frac{1}{(N\eta)^{1-\tau/2}}\right).$$

Again with Lemma A.8, (A.35) and (A.39), we obtain that

$$\begin{aligned}\Lambda &\le C\varphi^{2\tilde{\xi}}(q+\Psi_\theta)+\theta \le C\varphi^{19\tilde{\xi}}\left(q+\frac{1}{(N\eta)^{1-\tau/2}}\right)+C\varphi^{2\tilde{\xi}}\sqrt{\frac{\operatorname{Im} m_c}{N\eta}}\\ &\le \varphi^{20\tilde{\xi}}\left(q+\frac{1}{(N\eta)^{1-\tau/2}}\right)+\varphi^{3\tilde{\xi}}\sqrt{\frac{\operatorname{Im} m_c}{N\eta}}\end{aligned} \tag{A.54}$$

uniformly in $z \in S(c_0, C_0, L)$ with $(\tilde{\xi}' - \tau_N)$-high probability. Comparing with (A.53), we see that the power of $(N\eta)^{-1}$ is increased from $1-\tau$ to $1-\tau/2$, and moreover, there is no extra constant C appearing on the right-hand side of (A.54). Thus after M iterations, we get

$$\Lambda \le \varphi^{20\tilde{\xi}}\left(q+\frac{1}{(N\eta)^{1-(1/2)^{M-1}/3}}\right)+\varphi^{3\tilde{\xi}}\sqrt{\frac{\operatorname{Im} m_c}{N\eta}} \tag{A.55}$$

uniformly in $z \in S(c_0, C_0, L)$ with $(\tilde{\xi} - M\tau_N)$-high probability. Taking $M = \lfloor \log\log N/\log 2\rfloor$ such that

$$\tilde{\xi} - M\tau_N \ge \xi, \qquad \frac{1}{(N\eta)^{-(1/2)^{M-1}/3}} \le (N\eta)^{4/(3\log N)} \le C,$$

we can then conclude (A.52) and hence (3.19). Finally, to prove (3.18), we only need to plug (A.52) into Lemma A.12 and then apply Lemma A.10. $\square$

PROOF OF (3.20). The bound in (3.20) follows from a standard application of the local laws (3.18) and (3.19). The proof is exactly the same as the one for Lemma 4.4 in [17]. We omit the details here. $\square$

APPENDIX B: PROOF OF LEMMA 5.1

By (3.20), we have $\lambda_1 = \|H\|^2 \le \lambda_r + \varphi^{C_1}(q^2 + N^{-2/3})$ with ξ_1-high probability. Hence it suffices to prove (5.2) for $E \le \lambda_r + \varphi^{C_1}(q^2 + N^{-2/3})$. We define $\rho(x) := N^{-1}\sum_j \delta(x-\lambda_j)$. It is easy to see that $n(E) := \int_E^\infty \rho(x)\,dE$ and $m_2(z)$ is the Stieltjes' transform of $\rho(x)$. Then we introduce the differences:

$$\Delta\rho(x) := \rho(x) - \rho_{2c}(x), \qquad \Delta m(z) := m_2(z) - m_{2c}(z).$$

Fix $\eta_0 = \varphi^{C_1}N^{-1}$, $E_2 = \lambda_r + \varphi^{C_1+1}(q^2+N^{-2/3})$ and $\lambda_r - c_1 \le E_1 < E_2 - 2\eta_0$. We denote $\mathcal{E} := \max\{E_2 - E_1, \eta_0\}$ and $\kappa := \min\{\kappa_{E_1}, \kappa_{E_2}\}$. Let $\chi(y)$ be a smooth cutoff function with $\chi(y) = 0$ for $|y| \ge 2\mathcal{E}$, $\chi(y) = 1$ for $|y| \le \mathcal{E}$, and $|\chi'(y)| \le C\mathcal{E}^{-1}$. Let $f \equiv f_{E_1,E_2,\eta_0}$ be a smooth function supported in $[E_1 - \eta_0, E_2 + \eta_0]$ such that $f(x) = 1$ if $x \in [E_1 + \eta_0, E_2 - \eta_0]$, and $|f'| \le C\eta_0^{-1}$, $|f''| \le C\eta_0^{-2}$ if $|x - E_i| \le \eta_0$. Using the Helffer–Sjöstrand calculus (see, e.g., [10]), we have

$$f(E) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{iyf''(x)\chi(y) + i(f(x) + iyf'(x))\chi'(y)}{E - x - iy}\,dx\,dy.$$

Then we obtain that

$$
\begin{aligned}
&\left|\int f(E)\Delta\rho(E)\,dE\right| \\
&\quad\le C\int_{\mathbb{R}^2}\big(|f(x)|+|y||f'(x)|\big)|\chi'(y)||\Delta m(x+iy)|\,dx\,dy && \text{(B.1)}\\
&\qquad + C\sum_i\left|\int_{|y|\le\eta_0}\int_{|x-E_i|\le\eta_0} yf''(x)\chi(y)\,\mathrm{Im}\,\Delta m(x+iy)\,dx\,dy\right| && \text{(B.2)}\\
&\qquad + C\sum_i\left|\int_{|y|\ge\eta_0}\int_{|x-E_i|\le\eta_0} yf''(x)\chi(y)\,\mathrm{Im}\,\Delta m(x+iy)\,dx\,dy\right|. && \text{(B.3)}
\end{aligned}
$$

Using (3.18) with $\eta=\eta_0$ and (3.15), we get that

$$\eta_0\,\mathrm{Im}\,m_2(E+i\eta_0)=O\big(\varphi^{C_1}N^{-1}\big) \tag{B.4}$$

with ξ_1-high probability. Since $\eta\,\mathrm{Im}\,m_{2c}(E+i\eta)$ is increasing with η, we obtain that with ξ_1-high probability,

$$\eta\big|\mathrm{Im}\,\Delta m(E+i\eta)\big|=O\big(\varphi^{C_1}N^{-1}\big)\qquad\text{for all } 0\le\eta\le\eta_0. \tag{B.5}$$

Moreover, since $G(X,z)^*=G(X,\bar z)$, the estimates (3.18) and (B.5) also hold for $z\in\mathbb{C}_-$.

Now we bound the terms (B.1), (B.2) and (B.3). Using (3.18) and that the support of χ' is in $\mathcal{E}\le|y|\le 2\mathcal{E}$, the term (B.1) is bounded by

$$
\begin{aligned}
&\int_{\mathbb{R}^2}\big(|f(x)|+|y||f'(x)|\big)|\chi'(y)||\Delta m(x+iy)|\,dx\,dy \\
&\quad\le C\varphi^{C_1}\left(\frac{1}{N}+\frac{q^2\mathcal{E}}{\sqrt{\kappa+\mathcal{E}}}\right),
\end{aligned} \tag{B.6}
$$

with ξ_1-high probability. Using $|f''|\le C\eta_0^{-2}$ and (B.5), we can bound the terms in (B.2) by

$$\left|\int_{|y|\le\eta_0}\int_{|x-E_i|\le\eta_0} yf''(x)\chi(y)\,\mathrm{Im}\,\Delta m(x+iy)\,dx\,dy\right|\le C\frac{\varphi^{C_1}}{N},$$

with ξ_1-high probability. Finally, we integrate the term (B.3) by parts first in x, and then in y [and use the Cauchy–Riemann equation $\partial\,\mathrm{Im}(\Delta m)/\partial x=-\partial\,\mathrm{Re}(\Delta m)/\partial y$] to get that

$$
\begin{aligned}
&\int_{y\ge\eta_0}\int_{|x-E_i|\le\eta_0} yf''(x)\chi(y)\,\mathrm{Im}\,\Delta m(x+iy)\,dx\,dy \\
&\quad=\int_{y\ge\eta_0}\int_{|x-E_i|\le\eta_0} yf'(x)\chi(y)\frac{\partial\,\mathrm{Re}\,\Delta m(x+iy)}{\partial y}\,dx\,dy
\end{aligned}
$$

$$= -\int_{|x-E_i|\le\eta_0} \eta_0\chi(\eta_0) f'(x)\,\mathrm{Re}\,\Delta m(x+i\eta_0)\,dx \tag{B.7}$$

$$- \int_{y\ge\eta_0}\int_{|x-E_i|\le\eta_0} \big(y\chi'(y)+\chi(y)\big) f'(x)\,\mathrm{Re}\,\Delta m(x+iy)\,dx\,dy. \tag{B.8}$$

We bound the term in (B.7) by $O(\varphi^{C_1}N^{-1})$ using (3.18) and $|f'|\le C\eta_0^{-1}$. The first term in (B.8) can be estimated by $O(\varphi^{C_1}(\frac{1}{N}+\frac{q^2\mathcal{E}}{\sqrt{\kappa+\mathcal{E}}}))$ as in (B.6). For the second term in (B.8), we use (3.18) and $|f'|\le C\eta_0^{-1}$ to get that with ξ_1-high probability,

$$\begin{aligned}
&\left|\int_{y\ge\eta_0}\int_{|x-E_i|\le\eta_0} \chi(y) f'(x)\,\mathrm{Re}\,\Delta m(x+iy)\,dx\,dy\right| \\
&\quad\le C\int_{\eta_0}^{2\mathcal{E}} \varphi^{C_1}\left(\frac{1}{Ny}+\frac{q^2}{\sqrt{\kappa+y}}\right)dy \\
&\quad\le C\varphi^{C_1}\left(\frac{\log N}{N}+\frac{q^2\mathcal{E}}{\sqrt{\kappa+\mathcal{E}}}\right).
\end{aligned}$$

Obviously, the same bounds hold for the $y\le-\eta_0$ part. Combining the above estimates, we get that with ξ_1-high probability,

$$\begin{aligned}
\left|\int f(E)\Delta\rho(E)\,dE\right| &\le \varphi^{C_1+1}\left(\frac{1}{N}+\frac{q^2\mathcal{E}}{\sqrt{\kappa+\mathcal{E}}}\right) \\
&\le \varphi^{C_1+1}\left(\frac{1}{N}+q^2\sqrt{E_2-E_1+\eta_0}\right) \\
&\le \varphi^{2C_1+2}\left(\frac{1}{N}+q^3+q^2\sqrt{\kappa_{E_1}}\right),
\end{aligned} \tag{B.9}$$

where we used $E_2=\lambda_r+\varphi^{C_1+1}(q^2+N^{-2/3})$ in the last step.

For any interval $I:=[E-\eta_0,E+\eta_0]$ with $E\in[\lambda_r-c_1,E_2]$, we have

$$\begin{aligned}
n(E+\eta_0)-n(E-\eta_0) &\le \sum_k \frac{2\eta_0^2}{(\lambda_k-E)^2+\eta_0^2} \\
&= 2\eta_0\,\mathrm{Im}\, m_2(E+i\eta_0) = O\big(\varphi^{C_1}N^{-1}\big),
\end{aligned} \tag{B.10}$$

with ξ_1-high probability, where we used (B.4) in the last step. On the other hand, we trivially have

$$n_c(E+\eta_0)-n_c(E-\eta_0) = O(\eta_0) = O\big(\varphi^{C_1}N^{-1}\big), \tag{B.11}$$

with ξ_1-high probability, since the density $\rho_{2c}(x)$ is bounded by (3.13). Now with (B.9), (B.10) and (B.11), we get that with ξ_1-high probability,

$$\big|\big(n(E_2)-n(E)\big)-\big(n_c(E_2)-n_c(E)\big)\big| \le C\varphi^{2C_1+2}\left(\frac{1}{N}+q^3+q^2\sqrt{\kappa_E}\right)$$

for all $\lambda_r-c_1\le E<E_2-2\eta_0$. This concludes (5.2) since E_2 is chosen such that $n(E_2)=n_c(E_2)=0$ with ξ_1-high probability.

Acknowledgments. The authors would like to thank Jeremy Quastel, Bálint Virág and Jun Yin for fruitful discussions and valuable suggestions, which have significantly improved this paper. The first author wants to thank Jun Yin for the hospitality when he visited Madison. The second author would like to thank Jun Yin for the financial support during the writing of this paper. Finally, we are grateful to the referee for carefully reading our manuscript and suggesting several improvements.

REFERENCES

[1] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[2] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, New York. MR1990662

[3] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

[4] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051

[5] BAI, Z. D., SILVERSTEIN, J. W. and YIN, Y. Q. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166–168. MR0963829

[6] BAO, Z., PAN, G. and ZHOU, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864

[7] BAO, Z. G., PAN, G. M. and ZHOU, W. (2013). Local density of the spectrum on the edge for sample covariance matrices with general population. Preprint.

[8] BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inform. Theory* **57** 2400–2419. MR2809098

[9] BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** Art. ID 33. MR3183577

[10] DAVIES, E. B. (1995). The functional calculus. *J. Lond. Math. Soc.* (2) **52** 166–176. MR1345723

[11] DAVIS, R., HEINY, J., MIKOSCH, T. and XIE, X. (2016). Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes* **19** 517–547. MR3535965

[12] DING, X. C. (2016). Singular vector distribution of sample covariance matrices. Available at arXiv:1611.01837.

[13] EL KAROUI, N. (2005). Recent results about the largest eigenvalue of random covariance matrices and statistical application. *Acta Phys. Polon. B* **36** 2681–2697. MR2188088

[14] EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592

[15] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2012). Spectral statistics of Erdős–Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640. MR2964770

[16] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18**. Art. ID 59. MR3068390

[17] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41** 2279–2375. MR3098073

[18] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147

[19] FAMA, E. and FRENCH, K. (1992). The cross-section of expected stock returns. *J. Finance* **47** 427–465.

[20] FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991

[21] FAN, J., LIAO, Y. and LIU, H. (2016). An overview on the estimation of large covariance and precision matrices. *Econom. J.* **19** 1–32. MR3501529

[22] FISHER, J. T., SUN, X. Q. and GALLAGHER (2010). A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivariate Anal.* **101** 2554–2570. MR2719881

[23] FORRESTER, P. J. (1993). The spectrum edge of random matrix ensembles. *Nuclear Phys. B* **402** 709–728. MR1236195

[24] HACHEM, W., HARDY, A. and NAJIM, J. (2016). Large complex correlated Wishart matrices: Fluctuations and asymptotic independence at the edges. *Ann. Probab.* **44** 2264–2348. MR3502605

[25] JOHNSON, R. A. and WICHERN, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th ed. Pearson Prentice Hall, Upper Saddle River, NJ. MR2372475

[26] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[27] JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians*, *Vol. I* 307–333. Eur. Math. Soc., Zürich. MR2334195

[28] JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed. Springer, New York. MR2036084

[29] KAY, S. M. (1998). *Fundamentals of Statistical Signal Processing*, *Volume* 2: *Detection Theory*. Prentice-Hall, Upper Saddle River, NJ.

[30] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. MR3103909

[31] KNOWLES, A. and YIN, J. (2016). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* 1–96.

[32] LEE, J. O. and SCHNELLI, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** 3786–3839. MR3582818

[33] LEE, J. O. and YIN, J. (2014). A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.* **163** 117–173. MR3161313

[34] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457–483. MR208649

[35] NADAKUDITI, R. R. and EDELMAN, A. (2008). Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Trans. Signal Process.* **56** 2625–2638. MR1500236

[36] NADAKUDITI, R. R. and SILVERSTEIN, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE J. Sel. Top. Signal Process.* **4** 468–480.

[37] NADLER, B. and JOHNSTONE, I. M. (2011). On the distribution of Roy's largest root test in MANOVA and in signal detection in noise. Technical Report 2011-04.

[38] ONATSKI, A. (2008). The Tracy–Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.* **18** 470–490. MR2398763

[39] ONATSKI, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070

[40] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808

[41] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718

[42] PILLAI, N. S. and YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042

[43] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978

[44] SILVERSTEIN, J. W. (1989). On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivariate Anal.* **30** 307–311. MR1015375

[45] SILVERSTEIN, J. W. (2009). The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices. In *Random Matrix Theory and Its Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **18** 1–25. World Sci. Publ., Hackensack, NJ. MR2603192

[46] SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534

[47] SILVERSTEIN, J. W. and CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. MR1345541

[48] TAO, T. and VU, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449

[49] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246

[50] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

[51] TSAY, R. S. (2002). *Analysis of Financial Time Series*, 3rd ed. Wiley, New York. MR2778591

[52] VOICULESCU, D. V., DYKEMA, K. J. and NICA, A. (1992). *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras, and Harmonic Analysis on Free Groups*. Amer. Math. Soc., Providence, RI. MR1217253

[53] XI, H., YANG, F. and YIN, J. (2017). Local circular law for the product of a deterministic matrix with a random matrix. *Electron. J. Probab.* **22** Art. ID 60.

[54] YAO, J. F., BAI, Z. D. and ZHENG, S. R. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge Univ. Press, Cambridge. MR3468554

[55] YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
100 ST. GEORGE ST.
TORONTO, ONTARIO M5S 3G3
CANADA
E-MAIL: xiucai.ding@mail.utoronto.ca

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN–MADISON
480 LINCOLN ST.
MADISON, WISCONSIN 53706
USA
E-MAIL: fyang75@math.wisc.edu