The Annals of Statistics
2015, Vol. 43, No. 1, 382–421
DOI: 10.1214/14-AOS1281
© Institute of Mathematical Statistics, 2015

UNIVERSALITY FOR THE LARGEST EIGENVALUE OF SAMPLE COVARIANCE MATRICES WITH GENERAL POPULATION

BY ZHIGANG BAO[1], GUANGMING PAN[2] AND WANG ZHOU[3]

Zhejiang University, Nanyang Technological University and National University of Singapore

This paper is aimed at deriving the universality of the largest eigenvalue of a class of high-dimensional real or complex sample covariance matrices of the form $\mathcal{W}_N = \Sigma^{1/2}XX^*\Sigma^{1/2}$. Here, $X = (x_{ij})_{M,N}$ is an $M \times N$ random matrix with independent entries $x_{ij}, 1 \le i \le M, 1 \le j \le N$ such that $\mathbb{E}x_{ij} = 0$, $\mathbb{E}|x_{ij}|^2 = 1/N$. On dimensionality, we assume that $M = M(N)$ and $N/M \to d \in (0, \infty)$ as $N \to \infty$. For a class of general deterministic positive-definite $M \times M$ matrices Σ, under some additional assumptions on the distribution of x_{ij}'s, we show that the limiting behavior of the largest eigenvalue of $\mathcal{W}_N$ is universal, via pursuing a Green function comparison strategy raised in [*Probab. Theory Related Fields* **154** (2012) 341–407, *Adv. Math.* **229** (2012) 1435–1515] by Erdős, Yau and Yin for Wigner matrices and extended by Pillai and Yin [*Ann. Appl. Probab.* **24** (2014) 935–1001] to sample covariance matrices in the null case ($\Sigma = I$). Consequently, in the standard complex case ($\mathbb{E}x_{ij}^2 = 0$), combing this universality property and the results known for Gaussian matrices obtained by El Karoui in [*Ann. Probab.* **35** (2007) 663–714] (nonsingular case) and Onatski in [*Ann. Appl. Probab.* **18** (2008) 470–490] (singular case), we show that after an appropriate normalization the largest eigenvalue of $\mathcal{W}_N$ converges weakly to the type 2 Tracy–Widom distribution TW_2. Moreover, in the real case, we show that when Σ is spiked with a fixed number of subcritical spikes, the type 1 Tracy–Widom limit TW_1 holds for the normalized largest eigenvalue of $\mathcal{W}_N$, which extends a result of Féral and Péché in [*J. Math. Phys.* **50** (2009) 073302] to the scenario of nondiagonal Σ and more generally distributed X. In summary, we establish the Tracy–Widom type universality for the largest eigenvalue of generally distributed sample covariance matrices under quite light assumptions on Σ. Applications of these limiting results to statistical signal detection and structure recognition of separable covariance matrices are also discussed.

1. Introduction. In recent decades, researchers working on multivariate analysis have a growing interest in data with large size arising from various fields such

Received December 2013; revised October 2014.

[1]Supported in part by NSFC Grant 11071213, ZJNSF Grant R6090034 and SRFDP Grant 20100101110001.

[2]Supported in part by the Ministry of Education, Singapore, under Grant ARC 14/11.

[3]Supported in part by the Ministry of Education, Singapore, under Grant ARC 14/11, and by a Grant R-155-000-131-112 at the National University of Singapore.

MSC2010 subject classifications. Primary 60B20; secondary 62H10, 15B52, 62H25.

Key words and phrases. Sample covariance matrices, edge universality, Tracy–Widom law.

as genomics, image processing, microarray, proteomics and finance, to name but a few. The classical setting of *fixed p and large n* may lose its validity in tackling some statistical problems for high-dimensional data, due to the so-called *curse of dimensionality*. As a feasible and useful way in dealing with high-dimensional data, the spectral analysis of high-dimensional sample covariance matrices has attracted considerable interests among statisticians, probabilitists and mathematicians. Study toward the eigenvalues of sample covariance matrices traces back to the works of Fisher [25], Hsu [26] and Roy [47], and becomes flourishing after the seminal work of Marčenko and Pastur [33], in which the authors established the limiting spectral distribution (MP type distribution) for a class of sample covariance matrices, under the setting that p and n are comparable. Since then, a lot of research has been devoted to understanding the asymptotic properties of various spectral statistics of high-dimensional sample covariance matrices. One can refer to the monograph of Bai and Silverstein [1] for a comprehensive summary and detailed references.

In this paper, we will focus on the limiting behavior of the largest eigenvalue of a class of high-dimensional sample covariance matrices, which is of great interest naturally from the principal component analysis point of view. The largest eigenvalue has been commonly used in hypothesis testing problems on the structure of the population covariance matrix. Not trying to be comprehensive, one can refer to [8, 12, 28, 40, 44] for instance. We also refer to the review paper of Johnstone [29] for further reading on the statistical motivations of the study on the largest eigenvalue of sample covariance matrices. Precisely, we will consider the sample covariance matrix of the form

$$\mathcal{W} = \mathcal{W}_N := \Sigma^{1/2} X X^* \Sigma^{1/2}, \qquad X = (x_{ij})_{M,N}, \tag{1.1}$$

where $\{x_{ij} := x_{ij}(N), 1 \le i \le M := M(N), 1 \le j \le N\}$ is a collection of independent real or complex variables such that

$$\mathbb{E}x_{ij} = 0, \qquad \mathbb{E}|x_{ij}|^2 = N^{-1}.$$

We call $\mathcal{W}_N$ a *standard* complex sample covariance matrix if there also exists

$$\mathbb{E}x_{ij}^2 = 0, \qquad 1 \le i \le M, 1 \le j \le N.$$

In addition, $\Sigma := \Sigma_N$ is assumed to be an $M \times M$ positive-definite matrix. In particular, if the columns of X are independently drawn from $\mathbf{h}/\sqrt{N}$ for some random vector $\mathbf{h}$ possessing covariance matrix I, $\mathcal{W}$ can then be viewed as the sample covariance matrix of N observations of the random vector $\Sigma^{1/2}\mathbf{h}$. Conventionally, we call $\mathcal{W}$ a Wishart matrix if x_{ij}'s are Gaussian. As is well known now, the limiting distributions of the largest eigenvalues for classical high-dimensional random matrices were originally discovered by Tracy and Widom in [52, 53] for Gaussian Wigner ensembles G(O/U/S)E, thus named as the Tracy–Widom law of type β ($\beta = 1, 2, 4$ for GOE, GUE, GSE, resp.), denoted by TW_β hereafter. The analogs in the context of sample covariance matrices with $\Sigma = I$ were carried out by Johansson [27] and Johnstone [28]. More specifically, the TW_2 and TW_1 limits were established for the largest eigenvalues of standard complex and real null Wishart matrices in [27] and [28], respectively.

For the nonnull population covariance matrix, that is, $\Sigma \neq I$, much work has been devoted to the so-called spiked model, introduced by Johnstone in [28]. We say $\mathcal{W}$ is spiked when a few eigenvalues of Σ are not equal to 1. On the spiked Wishart models, one can refer to [4] for the standard complex case and [9, 10, 34, 41, 55] for the real case. However, in most cases, Σ has more complicated structures. In this paper, a more general setting on Σ stated in (iii) of Condition 1.1 below will be employed, whereby El Karoui showed in [12] that the TW_2 limit holds for the standard complex nonnull Wishart matrices when $d > 1$ (nonsingular case), followed by Onatski's extension to the singular case ($0 < d \leq 1$) in [38].

With the above mentioned limiting results for the Wishart matrices at hand, a conventional sequel in the Random Matrix Theory is to establish the so-called universality property for generally distributed sample covariance matrices, which states that the limiting behavior of an eigenvalue statistic usually does not depend on the details of the distribution of the matrix entries. The universality property of the extreme eigenvalues is usually referred to as *edge universality*. Specifically, for sample covariance matrices in the null case, the Tracy–Widom law has been established for $\mathcal{W}$ under very general assumptions on the distribution of X. The readers can refer to [43, 46, 49, 56] for some representative developments on this topic. For generally distributed spiked models, the universality property was also partially obtained in [2] and [24]. Especially, in the latter, the authors proved that TW_1 also holds for real spiked sample covariance matrices with a finite number of *subcritical* spikes (see Corollary 1.7 for definition).

In this paper, armed with the condition on Σ, that is, Condition 1.1(iii), we will prove the universality of the largest eigenvalues of $\mathcal{W}$. It will be clear that such a class of Σ contains those spiked population covariance matrices with a finite number of subcritical spikes, and goes far beyond. This work can therefore be viewed as a substantial generalization of the Tracy–Widom type edge universality, verified for the null case in [46] and [56], to a class of nonnull sample covariance matrices under quite light assumptions on Σ. A direct consequence of such a universality property, together with the results in [12] and [38], is that the TW_2 also holds for generally distributed standard complex $\mathcal{W}$ under our setting on Σ; see Corollary 1.5. Moreover, by combining the aforementioned result in [24], we can also show that TW_1 holds for real sample covariance matrices with spiked Σ containing a fixed number of subcritical spikes; see Corollary 1.7. Note that Σ is required to be diagonal in [24] and all odd order moments of x_{ij}'s are assumed to vanish. We stress here, our result can remove these restrictions. Both Corollary 1.5 and Corollary 1.7 can be used in high-dimensional statistical inference then. In Section 2, we will introduce two applications, namely *Presence of signals in the* *correlated noise* and *one-sided identity of separable covariance matrix*. Related numerical simulations will also be conducted.

In the sequel, we will start by introducing some notation and then present our main results. Subsequently, we will give a brief introduction of the so-called *Green function comparison strategy*, and then sketch our new inputs for treating the general setting of Σ.

1.1. *Main results.* Henceforth, we will denote by $\lambda_n(A) \le \cdots \le \lambda_2(A) \le \lambda_1(A)$ the ordered eigenvalues of an $n \times n$ Hermitian matrix A. For simplicity, we set the dimensional ratio

$$d_N := N/M \to d \in (0, \infty) \qquad \text{as } N \to \infty.$$

The empirical spectral distribution (ESD) of Σ is

$$H_N(\lambda) := \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}_{\{\lambda_i(\Sigma) \le \lambda\}}$$

and that of $\mathcal{W}$ is

$$\underline{F}_N(\lambda) := \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}_{\{\lambda_i(\mathcal{W}) \le \lambda\}}.$$

Here and throughout the following, $\mathbf{1}_{\mathbb{S}}$ represents the indicator function of the event $\mathbb{S}$. In addition, we will need a crucial parameter $\mathbf{c} := \mathbf{c}(\Sigma, N, M) \in [0, 1/\lambda_1(\Sigma))$ satisfying the equation

$$\int \left(\frac{\lambda \mathbf{c}}{1 - \lambda \mathbf{c}} \right)^2 dH_N(\lambda) = d_N. \tag{1.2}$$

It is elementary to check that the solution to (1.2) in $[0, 1/\lambda_1(\Sigma))$ is unique. With the above notation at hand, we can state our main condition as follows.

CONDITION 1.1. Throughout the paper, we need the following conditions.

(i) (On dimensionality). We assume that there are some positive constants c_1 and C_1 such that $c_1 < d_N < C_1$.

(ii) (On X). We assume that $\{x_{ij} := x_{ij}(N), 1 \le i \le M, 1 \le j \le N\}$ is a collection of independent real or complex variables such that $\mathbb{E}x_{ij} = 0$ and $\mathbb{E}|x_{ij}|^2 = N^{-1}$. Moreover, we assume that $\sqrt{N}x_{ij}$'s have a uniform subexponential tail, that is, there exists some positive constant τ_0 independent of i, j, N such that for sufficiently large t, one has

$$\mathbb{P}\big(|\sqrt{N}x_{ij}| \ge t\big) \le \tau_0^{-1} \exp\big(-t^{\tau_0}\big). \tag{1.3}$$

(iii) (On Σ). We assume that $\liminf_N \lambda_M(\Sigma) > 0$, $\limsup_N \lambda_1(\Sigma) < \infty$ and

$$\limsup_N \lambda_1(\Sigma)\mathbf{c} < 1. \tag{1.4}$$

Besides, we also need the following ad hoc terminology.

DEFINITION 1.2 (Matching to order k). Let $X^{\mathbf{u}} = (x_{ij}^{\mathbf{u}})_{M,N}$ and $X^{\mathbf{v}} = (x_{ij}^{\mathbf{v}})_{M,N}$ be two matrices satisfying (ii) of Condition 1.1. We say $X^{\mathbf{u}}$ matches $X^{\mathbf{v}}$ to order k, if for all $1 \le i \le M$, $1 \le j \le N$ and nonnegative integers l, m with $l + m \le k$, there exists

$$\begin{aligned}\mathbb{E}\big(\Re(\sqrt{N}x_{ij}^{\mathbf{u}})^l \Im(\sqrt{N}x_{ij}^{\mathbf{u}})^m\big) \\ = \mathbb{E}\big(\Re(\sqrt{N}x_{ij}^{\mathbf{v}})^l \Im(\sqrt{N}x_{ij}^{\mathbf{v}})^m\big) + O\big(e^{-(\log N)^C}\big)\end{aligned} \tag{1.5}$$

with some positive constant $C > 1$. Alternatively, if (1.5) holds, we also say that $\mathcal{W}^{\mathbf{u}}$ matches $\mathcal{W}^{\mathbf{v}}$ to order k, where $\mathcal{W}^{\mathbf{u}} = \Sigma^{1/2} X^{\mathbf{u}} (X^{\mathbf{u}})^* \Sigma^{1/2}$ and $\mathcal{W}^{\mathbf{v}} = \Sigma^{1/2} X^{\mathbf{v}} (X^{\mathbf{v}})^* \Sigma^{1/2}$.

Our main theorem on edge universality of $\mathcal{W}$ can be formulated as follows.

THEOREM 1.3 (Universality for both real and complex cases). *Suppose that two sample covariance matrices $\mathcal{W}^{\mathbf{u}} = \Sigma^{1/2} X^{\mathbf{u}} (X^{\mathbf{u}})^* \Sigma^{1/2}$ and $\mathcal{W}^{\mathbf{v}} = \Sigma^{1/2} X^{\mathbf{v}} (X^{\mathbf{v}})^* \Sigma^{1/2}$ satisfy Condition 1.1, where $X^{\mathbf{u}} := (x_{ij}^{\mathbf{u}})_{M,N}$ and $X^{\mathbf{v}} := (x_{ij}^{\mathbf{v}})_{M,N}$. Let*

$$\lambda_r = \frac{1}{\mathbf{c}}\left(1 + d_N^{-1} \int \frac{\lambda \mathbf{c}}{1 - \lambda \mathbf{c}}\, dH_N(\lambda)\right). \tag{1.6}$$

Then for sufficiently large N and any real number s which may depend on N, there exist some positive constants $\varepsilon, \delta > 0$ such that

$$\begin{aligned}\mathbb{P}\big(N^{2/3}(\lambda_1(\mathcal{W}^{\mathbf{u}}) - \lambda_r) \le s - N^{-\varepsilon}\big) - N^{-\delta} \\ \le \mathbb{P}\big(N^{2/3}(\lambda_1(\mathcal{W}^{\mathbf{v}}) - \lambda_r) \le s\big) \\ \le \mathbb{P}\big(N^{2/3}(\lambda_1(\mathcal{W}^{\mathbf{u}}) - \lambda_r) \le s + N^{-\varepsilon}\big) + N^{-\delta}\end{aligned} \tag{1.7}$$

if one of the following two additional conditions holds:

A: Σ *is diagonal and $\mathcal{W}^{\mathbf{u}}$ matches $\mathcal{W}^{\mathbf{v}}$ to order* 2.
B: *$\mathcal{W}^{\mathbf{u}}$ matches $\mathcal{W}^{\mathbf{v}}$ to order* 4.

REMARK 1.4. Theorem 1.3 can be extended to the case of joint distribution of the largest k eigenvalues for any fixed positive integer k, namely, for any real numbers $s_1, \dots, s_k$ which may depend on N, there exist some positive constants $\varepsilon, \delta > 0$ such that

$$\begin{aligned}&\mathbb{P}\big(N^{2/3}(\lambda_1(\mathcal{W}^{\mathbf{u}}) - \lambda_r) \le s_1 - N^{-\varepsilon}, \dots, \\ &\quad N^{2/3}(\lambda_k(\mathcal{W}^{\mathbf{u}}) - \lambda_r) \le s_k - N^{-\varepsilon}\big) - N^{-\delta} \\ &\quad\le \mathbb{P}\big(N^{2/3}(\lambda_1(\mathcal{W}^{\mathbf{v}}) - \lambda_r) \le s_1, \dots, N^{2/3}(\lambda_k(\mathcal{W}^{\mathbf{v}}) - \lambda_r) \le s_k\big) \\ &\quad\le \mathbb{P}\big(N^{2/3}(\lambda_1(\mathcal{W}^{\mathbf{u}}) - \lambda_r) \le s_1 + N^{-\varepsilon}, \dots, \\ &\qquad\quad N^{2/3}(\lambda_k(\mathcal{W}^{\mathbf{u}}) - \lambda_r) \le s_k + N^{-\varepsilon}\big) + N^{-\delta}.\end{aligned}$$

Such an extension can be realized through a parallel discussion as that for the null case in [46]. One can refer to [46] for more details. Here, we do not reproduce it.

Combining Theorem 1.3 with Theorem 1 of [12] and Proposition 2 of [38] yields the following more concrete result in the standard complex case ($\mathbb{E}x_{ij}^2 = 0$).

COROLLARY 1.5 (Tracy–Widom limit for the standard complex case). *Let $\mathcal{W}_N^{\mathbf{g}\mathbb{C}}$ be a standard complex Wishart matrix and $\mathcal{W}_N$ be a general standard complex sample covariance matrix. Assume that both of them satisfy Condition 1.1. Denoting*

$$\sigma^3 = \frac{1}{\mathbf{c}^3}\left(1 + d_N^{-1}\int\left(\frac{\lambda\mathbf{c}}{1-\lambda\mathbf{c}}\right)^3 dH_N(\lambda)\right), \tag{1.8}$$

we have

$$N^{2/3}\left(\frac{\lambda_1(\mathcal{W}_N) - \lambda_r}{\sigma}\right) \Longrightarrow \mathrm{TW}_2$$

if either Σ is diagonal or $\mathcal{W}_N$ matches $\mathcal{W}_N^{\mathbf{g}\mathbb{C}}$ to order 4.

REMARK 1.6. According to Remark 1.4, we also have the fact that the joint distribution of

$$\left(\frac{\lambda_1(\mathcal{W}_N) - \lambda_r}{\sigma}, \dots, \frac{\lambda_k(\mathcal{W}_N) - \lambda_r}{\sigma}\right)$$

converges weakly to the k-dimensional joint TW_2.

For real sample covariance matrices, putting our Theorem 1.3 and Theorem 1.6 of [24] together, we can get the following corollary.

COROLLARY 1.7 (Tracy–Widom limit for the real spiked case). *Suppose that $\mathcal{W}_N$ is a real sample covariance matrix satisfying* (i) *and* (ii) *of Condition* 1.1. *Let r be some given positive integer. Assume that Σ is spiked in the sense that $\lambda_1(\Sigma) \geq \cdots \geq \lambda_r(\Sigma) \geq \lambda_{r+1}(\Sigma) = \cdots = \lambda_M(\Sigma) = 1$. Moreover, the r spikes $\lambda_i(\Sigma)$, $i = 1, \dots, r$ are fixed* (*independent of N*) *and subcritical, that is, $\lambda_1(\Sigma) < 1 + (\sqrt{d})^{-1}$. Let $\mathcal{W}_N^{\mathbf{g}\mathbb{R}}$ be a real Wishart matrix with population covariance matrix Σ. Then in the scenario of $d \in [1, \infty)$* (*i.e., nonsingular case*), *we have*

$$N^{2/3}\left(\frac{\lambda_1(\mathcal{W}_N) - \lambda_r}{\sigma}\right) \Longrightarrow \mathrm{TW}_1$$

if either Σ is diagonal or $\mathcal{W}_N$ matches $\mathcal{W}_N^{\mathbf{g}\mathbb{R}}$ to order 4, *where σ is defined in* (1.8). *In addition, we have*

$$\lambda_r = \left(1 + d_N^{-1/2}\right)^2 + O\left(N^{-1}\right), \qquad \sigma = d_N^{-1/2}\left(1 + d_N^{1/2}\right)^{4/3} + o(1). \tag{1.9}$$

REMARK 1.8. Analogously, under the assumption of Theorem 1.7 we can get that the joint distribution of the first k normalized eigenvalues converges weakly to the k-dimensional joint TW_1.

REMARK 1.9. Lemma 4.2 below will show that the special spiked Σ with a fixed number of subcritical spikes satisfies (iii) of Condition 1.1. It is known that if there is any spike on or above the critical value $1 + (\sqrt{d})^{-1}$, the limiting distribution of the largest eigenvalue will not be the classical Tracy–Widom law any more, assuming r is fixed. One can refer to [4] and [9] for such a phase transition phenomenon for the standard complex and real cases, respectively. Such a fact reflects that (iii) of Condition 1.1 is quite light for the Tracy–Widom type universality to hold.

REMARK 1.10. We conjecture that the TW_1 law holds for all Σ satisfying (iii) of Condition 1.1 and the restriction on the nonsingular case is also artificial. However, as far as we know, only [24] can provide us the reference matrix to use the universality property in the real case. This is why we just focus on the special real spiked sample covariance matrices here. Nevertheless, these restrictions do not conceal the generality of the universality result (Theorem 1.3) itself even in the real case.

1.2. *Basic notions.* We define the $N \times N$ matrix

$$W = W_N := X^* \Sigma X$$

which shares the same nonzero eigenvalues with $\mathcal{W}$. Denoting the ESD of W_N by F_N, we see

$$F_N = d_N^{-1} \underline{F}_N + (1 - d_N^{-1}) \mathbf{1}_{[0,\infty)}. \tag{1.10}$$

If there is some deterministic distribution H such that $H_N \Longrightarrow H$ as $N \to \infty$, it is well known that there are deterministic distributions $F_{d,H}$ and $\underline{F}_{d,H}$ such that $F_N \Longrightarrow F_{d,H}$ and $\underline{F}_N \Longrightarrow \underline{F}_{d,H}$ in probability. One can refer to [3] or [1] for detailed discussions. Analogous to (1.10), we have the relation

$$F_{d,H} = d^{-1} \underline{F}_{d,H} + (1 - d^{-1}) \mathbf{1}_{[0,\infty)}. \tag{1.11}$$

For any distribution function D, its Stieltjes transform $m_D(z)$ is defined by

$$m_D(z) = \int \frac{1}{\lambda - z} \, dD(\lambda)$$

for all $z \in \mathbb{C}^+ := \{\omega \in \mathbb{C}, \Im \omega > 0\}$. And for any square matrix A, its Green function is defined by $G_A(z) = (A - zI)^{-1}$, $z \in \mathbb{C}^+$. For convenience, we will denote the Green functions of W_N and $\mathcal{W}_N$, respectively, by

$$G(z) = G_N(z) := (W_N - z)^{-1} \quad \text{and} \quad \mathcal{G}(z) = \mathcal{G}_N(z) := (\mathcal{W}_N - z)^{-1}, \qquad z \in \mathbb{C}^+.$$

The Stieltjes transforms of F_N and $\underline{F}_N$ will be denoted by $m_N(z)$ and $\underline{m}_N(z)$, respectively. By definitions, obviously one has

$$m_N(z) = \frac{1}{N}\operatorname{Tr} G(z), \qquad \underline{m}_N(z) = \frac{1}{M}\operatorname{Tr} \mathcal{G}(z).$$

Here, we draw attention to the basic relation $\operatorname{Tr} G(z) - \operatorname{Tr} \mathcal{G}(z) = (M-N)/z$. Actually, what really pertains to our discussion in the sequel is the nonasymptotic version of $F_{d,H}$ which can be obtained via replacing d and H by d_N and H_N in $F_{d,H}$, and thus will be denoted by F_{d_N,H_N}. More precisely, F_{d_N,H_N} is the corresponding distribution function of the Stieltjes transform $m_{d_N,H_N}(z) := m_{F_{d_N,H_N}}(z) \in \mathbb{C}^+$ satisfying the following self-consistent equation:

$$m_{d_N,H_N}(z) = \frac{1}{-z + d_N^{-1}\int t/(tm_{d_N,H_N}(z)+1)\, dH_N(t)}, \qquad z \in \mathbb{C}^+. \tag{1.12}$$

Analogously, we can define the nonasymptotic versions of $\underline{F}_{d,H}$ and its Stieltjes transform, denoted by $\underline{F}_{d_N,H_N}$ and $\underline{m}_{d_N,H_N}(z)$, respectively. Then the N-dependent version of (1.11) is

$$F_{d_N,H_N} = d_N^{-1}\underline{F}_{d_N,H_N} + (1 - d_N^{-1})\mathbf{1}_{[0,\infty)}. \tag{1.13}$$

For simplicity, we will briefly use the notation

$$m_0(z) := m_{d_N,H_N}(z), \qquad \underline{m}_0(z) := \underline{m}_{d_N,H_N}(z),$$
$$F_0 := F_{d_N,H_N}, \qquad \underline{F}_0 := \underline{F}_{d_N,H_N}$$

in the sequel.

It has been discussed in [48] by Silverstein and Choi that F_0 has a continuous derivative ρ_0 on $\mathbb{R}\setminus\{0\}$ and the rightmost boundary of the support of ρ_0 is λ_r defined in (1.6), that is, $\lambda_r = \inf\{x \in \mathbb{R} : F_0(x) = 1\}$. Moreover, the parameter $\mathbf{c}$ defined by (1.2) satisfies $\mathbf{c} = -\lim_{z\in\mathbb{C}^+\to\lambda_r} m_0(z)$.

1.3. *Sketch of the proof route.* As mentioned above, Theorem 1.3 can be viewed as a substantial generalization of the edge universality for the null sample covariance matrices provided in [46]. However, the general machinery in [46], with the so-called Green function comparison approach at the core, still works well even for general nonnull case. The Green function comparison strategy was raised in the series of work [21–23] on the local eigenvalue statistics of Wigner matrices originally, and has shown its strong applicability on some other random matrix models or statistics; see [5, 15, 45] for its variants for sample covariance and correlation matrices and see [50] for an application on random determinant. We also refer to the survey [14] for an overview.

To be specific, the preliminary heuristic of the Green function comparison strategy for our objective can be roughly explained as follows. At first, the distribution function of $\lambda_1(\mathcal{W})$ can actually be approximated from above and from below by the expectations of two functionals of the Stieltjes transform $m_N(z)$, that is, the normalized trace of the Green function $G_N(z)$; see (4.4) below. Hence, the comparison between the distributions of the largest eigenvalues of $\mathcal{W}^{\mathbf{u}}$ and $\mathcal{W}^{\mathbf{v}}$ can then be reduced to the comparison between the expectations of the functionals of the Green functions. For the latter, a replacement method inherited from the classical Lindeberg swapping process (see [32]) can be employed. Together with the expansion formula of the Green function, such a replacement method can effectively lead to the universality property.

A main technical tool escorting the Green function comparison process is the so-called *strong law of local eigenvalue density*, which asserts that the limiting spectral law is even valid on short intervals which contain only N^{ε} eigenvalues for any constant $\varepsilon > 0$. Such a limiting law on microscopic scales was developed in a series of work [16–18, 23] for Wigner matrices originally and was shown to be crucial in recent work on universality problems of local eigenvalue statistics, one can refer to [19, 22, 50] for instance. For our purpose, we will need a *strong local MP type law around* λ_r, which was established in our recent paper [7] and is recorded as Theorem 3.2 below. The companion work [7] initiates the project of edge universality and provides essential technical inputs for the Green function comparison process. However, the strong law of local eigenvalue density is also of interest in its own right.

To lighten the notation, we make the convention $E = \Re z$ and $\eta = \Im z$ hereafter. And we also denote

$$\Delta(z) := \Sigma^{1/2}\mathcal{G}(z)\Sigma^{1/2}$$

for simplicity. It will be seen that, in our comparison process, we need to control the magnitude of the entries of $\Delta(z)$ in the regime $|E - \lambda_r| \le N^{-2/3+\varepsilon}$ and $\eta = N^{-2/3-\varepsilon}$ for some small positive constant ε. This issue turns out to be a new difficulty due to the complexity of Σ. We handle this main technical task for diagonal and nondiagonal Σ via substantially different approaches, which are sketched as follows.

Clearly, when Σ is diagonal, we can turn to bound the entries of $\mathcal{G}(z)$ instead. Invoking the spectral decomposition [see the first inequality of (5.3) below, e.g.], the desired bound can be obtained via providing (1): an accurate description of the locations of the eigenvalues; (2): an upper bound for the eigenvector coefficients. It will be clear that (1) can be transformed into the strong local MP type law which has already been established. Toward (2), we will prove the so-called *delocalization property*, which states the eigenvector coefficients are of order $O(N^{-1/2+\varepsilon})$ typically. The delocalization property was first derived in [16] and improved in the series of papers [17, 18, 23] for Wigner matrices, and extended to sample covariance matrices in the null case in [20, 46, 51, 56]. Here, we extend the delocalization property to $\mathcal{W}$ for those eigenvectors corresponding to the eigenvalues around λ_r.

However, for the nondiagonal Σ, we need to focus on the entries of $\Delta(z)$ themselves. Fortunately, it turns out that only the diagonal entries $\Delta_{kk}(z)$ should be bounded if we are additionally granted in the comparison process that two ensembles match to order 4. To this end, we can start from the spectral decomposition again [see (5.6) below, e.g.]. Analogous to the diagonal case, we could provide (1′): an accurate description of the locations of the eigenvalues; (2′): an upper bound for $(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{kk}$. Observe that (1′) is just the same as (1) for diagonal Σ, actually can also be ensured by the strong local MP type law. However, (2′) requires some totally novel ideas. More details in Section 5 will show that the spectral decomposition equality (5.6) can also be applied in a converse direction, to wit, with a bound on $\Delta_{kk}(z_0)$ for some appropriately chosen $z_0 := E_0 + \mathbf{i}\eta_0$, one can actually obtain a bound for $(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{kk}$ in turn. For $\eta_0 = N^{-2/3+\varepsilon} \gg \eta$, we will perform a novel bounding scheme for $\Delta_{kk}(z_0)$, based on the Schur complement and the concentration inequalities on quadratic forms (Lemma 3.4). Then, by the bound on $\Delta_{kk}(z_0)$ one can get a bound on $(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{kk}$, which together with (1′) implies the desired bound on $\Delta_{kk}(z)$. The choice of $\eta_0 \gg N^{-2/3}$ will be technically necessary for our bounding scheme on $\Delta_{kk}(z_0)$. Therefore, we adopt such a roundabout way to bound $\Delta_{kk}(z)$, owing to the fact that $\eta \ll N^{-2/3}$ is unaffordable for a direct application of our bounding scheme based on the Schur complement and the concentration inequalities.

1.4. *Notation and organization.* Throughout the paper, we use the notation $O(\cdot)$ and $o(\cdot)$ in the conventional sense. As usual, C, C_1, C_2 and C' stand for some generic positive constants whose values may differ from line to line. We say $x \sim y$ if there exist some positive constants C_1 and C_2 such that $C_1|y| \le |x| \le C_2|y|$. Generally, for two functions $f(z), g(z) : \mathbb{C} \to \mathbb{C}$, we say $f(z) \sim g(z)$ if there exist some positive constants C_1 and C_2 independent of z such that $C_1|g(z)| \le |f(z)| \le C_2|g(z)|$. Moreover, $\|A\|_{\mathrm{op}}$ and $\|A\|_{\mathrm{HS}}$ represent the operator norm and Hilbert–Schmidt norm of a matrix A, respectively, and $\|\mathbf{u}\|$ is the L_2 norm of a vector $\mathbf{u}$. We use $\mathbf{i}$ to denote the imaginary unit to release i which will be frequently used as index or subscript. In addition, we conventionally denote by $\mathbf{e}_i$ the vector with all 0's except for a 1 in the ith coordinate and by $\mathbf{1}$ the vector with 1 in each coordinate. The dimensions of these vectors are usually obvious according to the context thus just omitted from the notation. $\mathbf{0}_{\alpha\times\beta}$ will be used to represent the $\alpha \times \beta$ null matrix which will be abbreviated to $\mathbf{0}_\alpha$ if $\alpha = \beta$. In addition, we adopt the notation in [46] to set the frequently used parameter

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

For $\zeta > 0$, we say that an event $\mathbb{S}$ holds with ζ-high probability if there is some positive constant C such that for sufficiently large N,

$$\mathbb{P}(\mathbb{S}) \ge 1 - N^C \exp(-\varphi^\zeta).$$

We conclude this section by stating its organization. In Section 2, we will introduce some applications of our main results in high-dimensional statistical inference, and some related simulations will be conducted. Then we will turn to the theoretical part. In Section 3, we will recall the properties of $m_0(z)$ and the strong local MP type law around λ_r established in [7] as the preliminaries of our proofs for the main results. In Section 4, we will use the strong local MP type law and a Green function comparison approach to prove Theorem 1.3, Corollaries 1.5 and 1.7. Section 5 will be devoted to the aforementioned argument of bounding the entries of $\Delta(z)$.

2. Applications and simulations.

In this section, we introduce some applications of our universality results in high-dimensional statistical inference, and conduct related simulations to check the quality of the approximations of our limiting laws and discuss their utility in the concrete hypothesis testing problems. We remark here, though Corollary 1.7 and Remark 1.8 are only stated for the case of $d \geq 1$, we will also perform the simulations for the case of $d < 1$.

2.1. *Applications.*

- *Presence of signals in the correlated noise.*

Consider an M-dimensional signal-plus-noise vector $\mathbf{y} := A\mathbf{s} + \Sigma_a^{1/2}\mathbf{z}$ and its N i.i.d. samples, namely

$$\mathbf{y}_i = A\mathbf{s}_i + \Sigma_a^{1/2}\mathbf{z}_i, \qquad i = 1, \ldots, N,$$

where $\mathbf{s}$ is a k-dimensional real or complex mean zero signal vector with covariance matrix S; $\mathbf{z}$ is an M-dimensional real or complex random vector with independent mean zero and variance one coordinates; A is an $M \times k$ deterministic matrix which is of full column rank and Σ_a is an $M \times M$ deterministic positive-definite matrix. We call $\Sigma_a^{1/2}\mathbf{z}$ the noise vector. Moreover, the signal vector and the noise vector are assumed to be independent. Set the matrices $Z_N = [\mathbf{z}_1, \ldots, \mathbf{z}_N]$ and $Y_N = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$. Denoting the covariance matrix of $\mathbf{y}$ by R, we can get by assumption that

$$R = ASA^T + \Sigma_a.$$

Such a model stems from several statistical signal processing problems, and is used commonly in various fields such as wireless communications, bioinformatics and machine learning, to name a few. We refer to Kay [30] for a comprehensive overview. A fundamental target is to detect signals via data. Thus the very first step is to know whether there is any signal present, that is, $k = 0$ versus $k \geq 1$. Once signals are detected, one can take a step further to estimate the number k. Under the high-dimensional setting, Nadakuditi and Edelman in [35], and Bianchi et al. in [8] considered respectively to detect signals in the *white* Gaussian noise, that is, $\Sigma_a = I$ (or more generally, $\Sigma_a = cI$ with some positive number c) and $\mathbf{z}$ is Gaussian. Also under the Gaussian assumption on the noise, Nadakuditi and Silverstein in [36] considered this detection problem when the noise may be correlated, that is, Σ_a may not be a multiple of I. We also refer to the very recent work of Vinogradova, etc. [54] for the case of correlated noise. Our aim is to test, for generally distributed and correlated noise $\Sigma_a^{1/2}\mathbf{z}$, whether there is no signal present. Thus our hypothesis testing problem can be stated as

$$(\mathbf{Q}_a):\qquad \mathbf{H}_0:\qquad k=0\quad \text{vs.}\quad \mathbf{H}_1:\qquad k\geq 1.$$

- *One-sided identity of separable covariance matrices.*

Consider the data model of the form

$$\mathcal{Y}_N=\Sigma_b^{1/2}\mathcal{Z}_N T^{1/2},$$

where $\mathcal{Z}_N$ is an $M\times N$ random matrix sharing the same distribution as Z_N in the previous problem, T is an $N\times N$ deterministic positive-definite matrix and Σ_b is an $M\times M$ deterministic positive-definite matrix. $N^{-1}\mathcal{Y}_N\mathcal{Y}_N^*$ is then called the *separable covariance matrix* which is widely used for handling the spatiotemporal sampling data. Such a nomenclature is owing to the fact that the vectorization of the data matrix $\mathcal{Y}_N$ has a separable covariance $\Sigma_b\otimes T$. The spectral properties of $N^{-1}\mathcal{Y}_N\mathcal{Y}_N^*$ have been widely investigated in some recent work under the high-dimensional setting, for example, one can refer to [13, 42, 57, 58]. Without loss of generality, we regard T as the temporal covariance matrix and Σ_b as the spatial covariance matrix. In this paper, we are interested in whether the temporal identity (i.e., $T=I$) holds. Formally, we are concerned with the following hypothesis testing:

$$(\mathbf{Q}_b):\qquad \mathbf{H}_0:\qquad T=I\quad \text{vs.}\quad \mathbf{H}_1:\qquad T\neq I.$$

Actually, we can consider to test whether $T=T_0$ for any given positive-definite T_0, since considering the renormalized data matrix $\mathcal{Y}_N T_0^{-1/2}$ we can recover the testing problem $\mathbf{Q}_b$. A similar testing problem with T replaced by Σ_b can also be considered. We call this kind of hypothesis testing problem *one-sided identity test* for the separable covariance matrix.

- *Onatski's statistics.*

Note that under $\mathbf{H}_0$ of either $\mathbf{Q}_a$ or $\mathbf{Q}_b$, the involved sample covariance matrix $N^{-1}Y_NY_N^*$ or $N^{-1}\mathcal{Y}_N\mathcal{Y}_N^*$ is of the form $\mathcal{W}$ defined in (1.1). It is then natural to construct our test statistics for $\mathbf{Q}_a$ and $\mathbf{Q}_b$ from the largest eigenvalues of $N^{-1}Y_NY_N^*$ and $N^{-1}\mathcal{Y}_N\mathcal{Y}_N^*$, respectively, such that our universality results can be employed under $\mathbf{H}_0$. For simplicity, we will use $\mathcal{W}$ to represent either $N^{-1}Y_NY_N^*$ or $N^{-1}\mathcal{Y}_N\mathcal{Y}_N^*$ under $\mathbf{H}_0$, that is, we will regard (Σ, X) as $(\Sigma_a, Z_N/\sqrt{N})$ and $(\Sigma_b, \mathcal{Z}_N/\sqrt{N})$ when we refer to $\mathbf{Q}_a$ and $\mathbf{Q}_b$, respectively.

At first glance, it is natural to choose the normalized largest eigenvalue as our test statistic. Unfortunately, in the real system, Σ is usually unknown. Hence, a general result like Corollary 1.5, where the parameters λ_r and σ depend on Σ, cannot be used directly if no information of Σ is known priori. To eliminate the unknown parameters λ_r and σ, we adopt the strategy used by Onatski in [37, 39]. More specifically, we will use the statistics

$$\mathbf{T}_a = \frac{\lambda_1(Y_N Y_N^*) - \lambda_2(Y_N Y_N^*)}{\lambda_2(Y_N Y_N^*) - \lambda_3(Y_N Y_N^*)} \quad \text{and} \quad \mathbf{T}_b = \frac{\lambda_1(\mathcal{Y}_N \mathcal{Y}_N^*) - \lambda_2(\mathcal{Y}_N \mathcal{Y}_N^*)}{\lambda_2(\mathcal{Y}_N \mathcal{Y}_N^*) - \lambda_3(\mathcal{Y}_N \mathcal{Y}_N^*)}$$

for $\mathbf{Q}_a$ and $\mathbf{Q}_b$, respectively. In the sequel, we will call $\mathbf{T}_a$ and $\mathbf{T}_b$ *Onatski's statistics*. Note that under $\mathbf{H}_0$, $\mathbf{T}_a$ and $\mathbf{T}_b$ possess the same limiting distribution, determined by the joint TW_β laws, mentioned in Remarks 1.6 and 1.8. An obvious advantage of $\mathbf{T}_a$ or $\mathbf{T}_b$ is that its limiting distribution is independent of λ_r and σ under $\mathbf{H}_0$, which makes it asymptotically pivotal. Moreover, though the explicit formula for the limiting distribution function of Onatski's statistic under $\mathbf{H}_0$ is unavailable currently, one can approximate it via simulation, by generating the eigenvalues from high-dimensional GOE (resp., GUE) in the real (resp., complex) case. We will describe such an approximation in detail in the subsequent simulation study.

2.2. *Simulations*.

- *Accuracy of approximations for TW laws.*

We conduct some numerical simulations to check the accuracy of the distributional approximations in Corollaries 1.5 and 1.7, under various settings of (M, N), Σ and the distribution of X. Firstly, for each pair of (M, N), we generate an observation from $M \times M$ Haar distributed random orthogonal matrix and denote it by $U := U(M, N)$. To get such a U, we can generate in Matlab an $M \times M$ Gaussian matrix $\mathbf{G}$ with i.i.d. $N(0, 1)$ entries, and let $U = \mathbf{G}(\mathbf{G}^*\mathbf{G})^{-1/2}$ which is well defined with probability 1; refer to Section 7.1 of [11] for instance. Then we will fix this U for each pair of (M, N) as a deterministic orthogonal matrix. Next, we set some scenarios of Σ in Corollaries 1.5 and 1.7. To this end, we define

$$D_c := \mathrm{diag}(\underbrace{1, \dots, 1}_{\lfloor M/2 \rfloor}, \underbrace{2, \dots, 2}_{M - \lfloor M/2 \rfloor}), \qquad D_r := \mathrm{diag}\left(1 + \frac{(\sqrt{d_N})^{-1}}{2}, 1, \dots, 1\right)$$

and choose Σ to be some similar forms of D_c and D_r in Corollaries 1.5 and 1.7, respectively. More specifically, we will use the following four choices of population covariance matrix Σ, denoted by

$$\Sigma(c, 1) := D_c, \qquad \Sigma(c, 2) := U D_c U^*,$$
$$\Sigma(r, 1) := D_r, \qquad \Sigma(r, 2) := U D_r U^*.$$

Here, U is the orthogonal matrix generated priori.

Now, we state our choice for the distribution of X. For simplicity, we set $h_{ij} := \sqrt{N} x_{ij}$ for $i = 1, \dots, M$ and $j = 1, \dots, N$, and choose all these (h_{ij})'s to be i.i.d.

For standard complex Gaussian case, the numerical performance of the limiting law in Corollary 1.5 has been assessed; see Tables 1 and 2 of [12]. Here, we use a discrete distribution in our simulation study. Specifically, let $\mathfrak{s}_1$ and $\mathfrak{s}_2$ be two i.i.d. variables with the distribution

$$\mathfrak{u} = \frac{1}{12}\boldsymbol{\delta}_{-2} + \frac{4}{25}\boldsymbol{\delta}_{-1} + \frac{13}{24}\boldsymbol{\delta}_0 + \frac{16}{75}\boldsymbol{\delta}_{3/2} + \frac{1}{600}\boldsymbol{\delta}_4,$$

where $\boldsymbol{\delta}_a$ represents the Dirac measure at a. It is elementary to check that the first four moments of $\mathfrak{u}$ are the same as those of $N(0,1)$. Now we choose h_{11} for the standard complex and real cases respectively as

$$h_{11}(c) \stackrel{d}{=} \frac{1}{\sqrt{2}}(\mathfrak{s}_1 + \mathbf{i}\mathfrak{s}_2) \quad \text{and} \quad h_{11}(r) \stackrel{d}{=} \mathfrak{s}_1$$

and denote the corresponding X by $X(c)$ and $X(r)$, respectively. We conduct the simulations for the combinations $(\Sigma(c,1), X(c))$, $(\Sigma(c,2), X(c))$, $(\Sigma(r,1), X(r))$ and $(\Sigma(r,2), X(r))$ under various settings of (M, N). The results are provided in Table 1. It can be seen, in each case, the approximation is satisfactory even for relatively small M and N.

Next, a natural question is, to what extent can we weaken the assumptions imposed on X. Very recently, a necessary and sufficient condition for the Tracy–Widom limit of Wigner matrix with i.i.d. off-diagonal entries (up to symmetry) was established by Lee and Yin in [31], where the matrix entry is only required to have mean 0 and variance 1, and satisfies a tail condition slightly weaker than the existence of the 4th moment. It is reasonable to conjecture a similar moment condition is sufficient for the validity of Tracy–Widom laws for sample covariance matrices. To give a numerical evidence for such a conjecture, we also conduct some simulation for the largest eigenvalue of $\mathcal{W}$ whose entries possess a symmetric Pareto distribution. For simplicity, we only state the simulation result for the complex case with $\Sigma = \Sigma(c,2)$. We choose $\mathfrak{p}_1$ and $\mathfrak{p}_2$ to be i.i.d. variables with the symmetric Pareto distribution whose density is given by $f(x) = \frac{9}{10}\sqrt{\frac{3}{5}}|x|^{-6}$ when $|x| > \sqrt{\frac{3}{5}}$ and 0 otherwise. It is then elementary to see that

$$h_{11} \stackrel{d}{=} \frac{1}{\sqrt{2}}(\mathfrak{p}_1 + \mathbf{i}\mathfrak{p}_2) \tag{2.1}$$

has mean 0 and variance 1. Moreover, we see $\mathbb{E}|h_{11}|^4 < \infty$. We denote by $X(P)$ the corresponding X. The simulation results are stated in Table 2. It can be seen that the approximation is also very good even for small M and N.

- *Size and power study for* $\mathbf{T}_a$ *and* $\mathbf{T}_b$.

Now, we evaluate the sizes and powers of the statistics $\mathbf{T}_a$ and $\mathbf{T}_b$ for $\mathbf{Q}_a$ and $\mathbf{Q}_b$ respectively. For simplicity, we only report the results for the real case here. Note that, in the real case, we do not establish the TW_1 law for general Σ satisfying

TABLE 1

Simulated quantiles for four pairs of (Σ, X). The cases of $(\Sigma(r,1), X(r))$, $(\Sigma(r,2), X(r))$, $(\Sigma(c,1), X(c))$ and $(\Sigma(c,2), X(2))$ are titled by $R1$, $R2$, $C1$ and $C2$, respectively, for simplicity

$(\boldsymbol{\Sigma}, X)$	**Percentile**	**$\mathbf{TW_1}$**	**30 x 30**	**60 x 60**	**100 x 100**	**80 x 20**	**20 x 80**	**100 x 400**	**2 ∗ SE**
R1	−3.9000	0.0100	0.0053	0.0087	0.0114	0.0075	0.0076	0.0115	0.0020
	−3.1800	0.0500	0.0479	0.0523	0.0566	0.0493	0.0580	0.0601	0.0040
	−2.7800	0.1000	0.1070	0.1151	0.1151	0.1099	0.1197	0.1192	0.0060
	−1.9100	0.3000	0.3520	0.3524	0.3393	0.3535	0.3539	0.3352	0.0090
	−1.2700	0.5000	0.5762	0.5674	0.5457	0.5725	0.5714	0.5388	0.1000
	−0.5900	0.7000	0.7752	0.7547	0.7388	0.7685	0.7713	0.7372	0.0090
	0.4500	0.9000	0.9345	0.9260	0.9214	0.9347	0.9359	0.9171	0.0060
	0.9800	0.9500	0.9689	0.9650	0.9620	0.9706	0.9708	0.9620	0.0040
	2.0200	0.9900	0.9943	0.9929	0.9931	0.9938	0.9952	0.9905	0.0020
R2	−3.9000	0.0100	0.0054	0.0088	0.0098	0.0063	0.0086	0.0100	0.0020
	−3.1800	0.0500	0.0497	0.0533	0.0552	0.0504	0.0556	0.0572	0.0040
	−2.7800	0.1000	0.1077	0.1144	0.1143	0.1085	0.1226	0.1116	0.0060
	−1.9100	0.3000	0.3617	0.3460	0.3363	0.3470	0.3707	0.3392	0.0090
	−1.2700	0.5000	0.5784	0.5582	0.5506	0.5700	0.5834	0.5433	0.1000
	−0.5900	0.7000	0.7714	0.7568	0.7503	0.7731	0.7765	0.7404	0.0090
	0.4500	0.9000	0.9301	0.9258	0.9248	0.9334	0.9349	0.9166	0.0060
	0.9800	0.9500	0.9658	0.9649	0.9630	0.9671	0.9704	0.9605	0.0040
	2.0200	0.9900	0.9924	0.9929	0.9928	0.9934	0.9941	0.9937	0.0020
$(\boldsymbol{\Sigma}, X)$	**Percentile**	**$\mathbf{TW_2}$**	**30 x 30**	**60 x 60**	**100 x 100**	**80 x 20**	**20 x 80**	**100 x 400**	**2 ∗ SE**
C1	−3.7300	0.0100	0.0031	0.0053	0.0066	0.0037	0.0042	0.0082	0.0020
	−3.2000	0.0500	0.0266	0.0377	0.0363	0.0319	0.0326	0.0396	0.0040
	−2.9000	0.1000	0.0674	0.0812	0.0827	0.0749	0.0745	0.0870	0.0060
	−2.2700	0.3000	0.2573	0.2728	0.2819	0.2772	0.2648	0.2819	0.0090
	−1.8100	0.5000	0.4695	0.4804	0.4866	0.4838	0.4818	0.4861	0.1000
	−1.3300	0.7000	0.6913	0.6963	0.6950	0.6942	0.6928	0.6936	0.0090
	−0.6000	0.9000	0.9053	0.9004	0.9006	0.9012	0.9021	0.9025	0.0060
	−0.2300	0.9500	0.9549	0.9506	0.9489	0.9521	0.9525	0.9531	0.0040
	0.4800	0.9900	0.9913	0.9900	0.9886	0.9880	0.9924	0.9912	0.0020
C2	−3.7300	0.0100	0.0021	0.0056	0.0066	0.0032	0.0050	0.0071	0.0020
	−3.2000	0.0500	0.0234	0.0321	0.0399	0.0326	0.0321	0.0445	0.0040
	−2.9000	0.1000	0.0642	0.0754	0.0852	0.0781	0.0746	0.0926	0.0060
	−2.2700	0.3000	0.2639	0.2641	0.2805	0.2721	0.2734	0.2955	0.0090
	−1.8100	0.5000	0.4745	0.4756	0.4858	0.4874	0.4864	0.4933	0.1000
	−1.3300	0.7000	0.6875	0.6877	0.6930	0.7006	0.6923	0.6954	0.0090
	−0.6000	0.9000	0.9008	0.9012	0.8988	0.9055	0.8994	0.9028	0.0060
	−0.2300	0.9500	0.9490	0.9512	0.9493	0.9547	0.9517	0.9529	0.0040
	0.4800	0.9900	0.9899	0.9905	0.9894	0.9917	0.9891	0.9903	0.0020

The simulation was done in Matlab. In each of the above four cases, we generated 10,000 matrix X with the distribution defined above, and then calculated the largest eigenvalue of $\mathcal{W}$ and renormalized it with the parameters λ_r and σ according to Corollaries 1.5 and 1.7. In the column titled "Percentile," we listed the quantiles of TW_β law for $\beta = 1, 2$. Simulating 10,000 times gave us an empirical distribution of the renormalized largest eigenvalue. And we stated the values of this empirical distribution at the quantiles of the TW laws for various pairs of $(M, N) = (30, 30), (60, 60), (100, 100), (80, 20), (20, 80), (100, 400)$. The last column states the approximate standard errors based on binomial sampling.

TABLE 2
Simulated quantiles for the case of $(\Sigma, X) = (\Sigma(c, 2), X(P))$ (*CP for short*)

$(\boldsymbol{\Sigma}, X)$	Percentile	$\mathbf{TW_2}$	30 x 30	60 x 60	100 x 100	80 x 20	20 x 80	100 x 400	2 * SE
CP	−3.7300	0.0100	0.0016	0.0043	0.0062	0.0035	0.0044	0.0088	0.0020
	−3.2000	0.0500	0.0280	0.0409	0.0460	0.0345	0.0369	0.0512	0.0040
	−2.9000	0.1000	0.0776	0.0987	0.1037	0.0862	0.0894	0.1069	0.0060
	−2.2700	0.3000	0.3113	0.3311	0.3320	0.3201	0.3275	0.3235	0.0090
	−1.8100	0.5000	0.5517	0.5476	0.5507	0.5603	0.5628	0.5347	0.1000
	−1.3300	0.7000	0.7675	0.7472	0.7501	0.7736	0.7689	0.7335	0.0090
	−0.6000	0.9000	0.9392	0.9303	0.9228	0.9364	0.9380	0.9176	0.0060
	−0.2300	0.9500	0.9716	0.9658	0.9659	0.9714	0.9708	0.9580	0.0040
	0.4800	0.9900	0.9932	0.9909	0.9921	0.9928	0.9928	0.9912	0.0020

The simulation was taken analogously. We generated 10,000 matrix X with h_{11} following the distribution defined in (2.1). Each column has the same meaning as that in Table 1.

Condition 1.1(iii). However, in the sequel, we will also perform the simulation for Σ which is not spiked, such as $\Sigma = \Sigma(c, 1)$. More specifically, we will focus on two settings

$$\text{(I):} \qquad \Sigma_a = \Sigma_b = \Sigma(r, 1), \qquad Z_N \stackrel{d}{=} \mathcal{Z}_N \stackrel{d}{=} \sqrt{N} X(r),$$

and

$$\text{(II):} \qquad \Sigma_a = \Sigma_b = \Sigma(c, 1), \qquad Z_N \stackrel{d}{=} \mathcal{Z}_N \stackrel{d}{=} \sqrt{N} X(r).$$

For $\mathbf{Q}_a$, we choose the alternative with some positive number ρ_a as

$$\mathbf{H}_1(a, \rho_a): \qquad k = 1, \qquad A = \mathbf{e}_1' \quad \text{and} \quad \mathbf{s} \sim N(0, \rho_a),$$

where $\mathbf{e}_1$ is M-dimensional by the assumption on A. For $\mathbf{Q}_b$, we choose two alternatives parameterized by ρ_b as

$$\mathbf{H}_1(b, \rho_b, 1): \qquad T = I + \rho_b \mathbf{e}_1 \mathbf{e}_1' \quad \text{and} \quad \mathbf{H}_1(b, \rho_b, 2): \qquad T = I + \rho_b \frac{1}{N} \mathbf{1}\mathbf{1}',$$

where $\mathbf{e}_1$ and $\mathbf{1}$ are both N-dimensional by the assumption on T. Under the setting (I), for $\mathbf{H}_1(a, \rho_a)$, we set $\rho_a := \tau(\sqrt{d_N})^{-1}$, while for both $\mathbf{H}_1(b, \rho_b, 1)$ and $\mathbf{H}_1(b, \rho_b, 2)$, we set $\rho_b := \tau\sqrt{d_N}$ with some strength parameter $\tau > 0$. Under the setting (II), for $\mathbf{H}_1(a, \rho_a)$, we set $\rho_a := 2\tau(\sqrt{d_N})^{-1}$, while for both $\mathbf{H}_1(b, \rho_b, 1)$ and $\mathbf{H}_1(b, \rho_b, 2)$, we set $\rho_b := 2\tau\sqrt{d_N}$ with some strength parameter $\tau > 0$. We will choose $\tau = 0.5, 4, 6$ for each alternative above.

Now assuming that ξ_1, ξ_2 and ξ_3 have the joint TW_1 distribution, we approximate the percentiles of the distribution of $(\xi_1 - \xi_2)/(\xi_2 - \xi_3)$ as follows. We can simulate 30,000 independent matrices from GOE of dimension 1000 and numerically compute the ratio of the differences between the first and the second and the second and the third eigenvalues for each matrix, then we can get the percentiles of

TABLE 3
Simulated sizes for settings (I) *and* (II)

Setting	30 x 30	60 x 60	100 x 100	80 x 20	80 x 40	20 x 80	40 x 80	100 x 400	400 x 200
(I)	0.0522	0.0476	0.0490	0.0604	0.0526	0.0474	0.0521	0.0512	0.0486
(II)	0.0544	0.0511	0.0488	0.0543	0.0521	0.0493	0.0526	0.0478	0.0446

the empirical distribution of these 30,000 ratios. By doing the above in Matlab, we got that the approximate 95th percentile of the distribution of $(\xi_1 - \xi_2)/(\xi_2 - \xi_3)$ is 7.16. The nominal significant level of our tests is 5%. The results for the sizes are reported in Table 3, and the results for the powers are reported in Table 4 for setting (I) and Table 5 for (II), respectively. The small $\tau = 0.5$ is tailored for corroborating the following phenomenon, that is, *an additive or multiplicative finite rank perturbation may not cause significant change of the largest eigenvalue of a sample covariance matrix when the strength of the perturbation is weak enough.* This phenomenon has been explicitly verified for the spiked sample covariance matrices, see the aforementioned references on the spiked models [4] and [24]. Our Corollary 1.7 also confirms it again. However, for more complicated models such as $N^{-1}Y_N Y_N^*$ and $N^{-1}\mathcal{Y}_N\mathcal{Y}_N^*$ in our $\mathbf{Q}_a$ and $\mathbf{Q}_b$, given general Σ_a and Σ_b, the theoretical discussions on this phenomenon with respect to various As and T are still open. Under our choices of Σ, from the simulations we can see that when $\tau = 0.5$, the powers of both tests in various scenarios are very poor. However, when τ is relatively large, our tests are reliable. It can be seen from Tables 4 and 5, in the cases of $\tau = 4$ or 6, the powers are satisfactory, especially when N and M are relatively large.

3. Square root behavior and local MP type law. In this section, we will record several main results proved in our recent paper [7] which will serve as

TABLE 4
Simulated powers for $\mathbf{T}_a$ *and* $\mathbf{T}_b$ *under setting* (I), τ *is* 0.5, 4 *or* 6

τ	$\mathbf{H}_1$	30 x 30	60 x 60	100 x 100	80 x 20	80 x 40	20 x 80	40 x 80	100 x 400	400 x 200
0.5	$\mathbf{H}_1(a, \rho_a)$	0.0630	0.0577	0.0622	0.0604	0.0588	0.0589	0.0614	0.0541	0.0545
	$\mathbf{H}_1(b, \rho_b, 1)$	0.0541	0.0516	0.0497	0.0540	0.0521	0.0533	0.0522	0.0488	0.0482
	$\mathbf{H}_1(b, \rho_b, 2)$	0.0551	0.0463	0.0508	0.0540	0.0518	0.0530	0.0545	0.0506	0.0498
4	$\mathbf{H}_1(a, \rho_a)$	0.4825	0.6857	0.8421	0.5090	0.6454	0.5263	0.6680	0.9684	0.9929
	$\mathbf{H}_1(b, \rho_b, 1)$	0.3932	0.5775	0.7507	0.4243	0.5475	0.4262	0.5488	0.8998	0.9816
	$\mathbf{H}_1(b, \rho_b, 2)$	0.3983	0.5776	0.7511	0.4216	0.5529	0.4352	0.5427	0.8970	0.9812
6	$\mathbf{H}_1(a, \rho_a)$	0.7319	0.9089	0.9807	0.7434	0.8830	0.7861	0.8980	0.9999	1.0000
	$\mathbf{H}_1(b, \rho_b, 1)$	0.6556	0.8653	0.9628	0.7205	0.8539	0.6822	0.8247	0.9954	0.9998
	$\mathbf{H}_1(b, \rho_b, 2)$	0.6623	0.8647	0.9608	0.7235	0.8477	0.6888	0.8277	0.9955	1.0000

TABLE 5
Simulated powers for $\mathbf{T}_a$ *and* $\mathbf{T}_b$ *under setting* (II), τ *is* 0.5, 4 *or* 6

τ	$\mathbf{H}_1$	30 × 30	60 × 60	100 × 100	80 × 20	80 × 40	20 × 80	40 × 80	100 × 400	400 × 200
0.5	$\mathbf{H}_1(a, \rho_a)$	0.0583	0.0588	0.0534	0.0649	0.0557	0.0568	0.0554	0.0536	0.0512
	$\mathbf{H}_1(b, \rho_b, 1)$	0.0627	0.0583	0.0587	0.0614	0.0577	0.0607	0.0589	0.0572	0.0545
	$\mathbf{H}_1(b, \rho_b, 2)$	0.0620	0.0593	0.0532	0.0614	0.0618	0.0590	0.0565	0.0513	0.0502
4	$\mathbf{H}_1(a, \rho_a)$	0.4352	0.6207	0.7891	0.5269	0.6354	0.3476	0.5360	0.8278	0.9913
	$\mathbf{H}_1(b, \rho_b, 1)$	0.7650	0.9258	0.9870	0.8454	0.9277	0.7237	0.8776	0.9974	1.0000
	$\mathbf{H}_1(b, \rho_b, 2)$	0.7558	0.9249	0.9856	0.8480	0.9259	0.7263	0.8780	0.9978	1.0000
6	$\mathbf{H}_1(a, \rho_a)$	0.9255	0.9914	0.9996	0.9754	0.9954	0.9031	0.9817	1.0000	1.0000
	$\mathbf{H}_1(b, \rho_b, 1)$	0.9594	0.9978	1.0000	0.9854	0.9984	0.9595	0.9951	1.0000	1.0000
	$\mathbf{H}_1(b, \rho_b, 2)$	0.9285	0.9923	0.9997	0.9758	0.9956	0.9072	0.9831	1.0000	1.0000

fundamental inputs for the Green function comparison process. The main result established in [7] is the aforementioned strong local MP type law around λ_r; see Theorem 3.2 below. As a necessary input to the proof of the strong local MP type law around λ_r in [7], the so-called square root behavior of $m_0(z)$ has been derived therein, see Theorem 3.1 below. Then, as a direct consequence of the strong local MP type law around λ_r, a nearly optimal convergence rate of $F_N(x)$ around λ_r has also been obtained, see Theorem 3.3 below. All these results will play roles in our Green function comparison process.

In [7], it has been shown that $\mathbf{c}, \lambda_r \sim 1$. More precisely, there exist two positive constants $C_l \le C_r$ such that $\lambda_r \in [C_l/2, 2C_r]$. Here, C_l and C_r can be chosen appropriately such that $\lambda_1(\mathcal{W}) \in [C_l, C_r]$ with ζ-high probability for any given constant $\zeta > 0$. We will always write $z := E + \mathbf{i}\eta$, and use the notation

$$\kappa := \kappa(z) = |E - \lambda_r|.$$

We introduce for $\zeta \ge 0$ two sets of z,

$$S(\zeta) := \{z \in \mathbb{C} : C_l \le E \le C_r, \varphi^\zeta N^{-1} \le \eta \le 1\},$$
$$S_r(\tilde{c}, \zeta) := \{z \in \mathbb{C} : \lambda_r - \tilde{c} \le E \le C_r, \varphi^\zeta N^{-1} \le \eta \le 1\},$$

where $\tilde{c}$ is some positive constant.

The first main result we need is a collection of some crucial properties of $m_0(z)$, which are essentially guaranteed by (iii) of Condition 1.1, and can be inferred from the square root behavior of the limiting spectral density ρ_0 on its right edge λ_r. Informally, we can call it *square root behavior of* $m_0(z)$.

THEOREM 3.1 (Square root behavior of $m_0(z)$, Lemma 2.3 of [7]). *Under Condition* 1.1, *there exists some small but fixed positive constant* $\tilde{c}$ *such that the following three statements hold.*

(i) *For* $z \in S(0)$, *we have*

$$|m_0(z)| \sim 1;$$

(ii) *For* $z \in S_r(\tilde{c}, 0)$, *we have*

$$\Im m_0(z) \sim \begin{cases} \dfrac{\eta}{\sqrt{\kappa+\eta}}, & \text{if } E \geq \lambda_r + \eta, \\ \sqrt{\kappa+\eta}, & \text{if } E \in [\lambda_r - \tilde{c}, \lambda_r + \eta); \end{cases}$$

(iii) *For* $z \in S_r(\tilde{c}, 0)$, *we have*

$$|1 + t m_0(z)| \geq \hat{c}\big(1 + \lambda_1(\Sigma) m_0(\lambda_r)\big) \geq c_0, \qquad \forall t \in \big[\lambda_M(\Sigma), \lambda_1(\Sigma)\big]$$

for some small positive constants $\hat{c}, c_0$ *depending on* $\tilde{c}$.

The second necessary input is the strong local MP type law around λ_r. To state it, we also need to recall some additional notation from [7]. We denote by $\mathbf{x}_i$ the ith column of X and set $\mathbf{r}_i = \Sigma^{1/2}\mathbf{x}_i$. We introduce the notation $X^{(\mathbb{T})}$ to represent the $M \times (N - |\mathbb{T}|)$ minor of X obtained by deleting $\mathbf{x}_i$ from X if $i \in \mathbb{T}$. For convenience, $(\{i\})$ will be abbreviated to (i). Denoting

$$W^{(\mathbb{T})} = X^{(\mathbb{T})*}\Sigma X^{(\mathbb{T})}, \qquad \mathcal{W}^{(\mathbb{T})} = \Sigma^{1/2} X^{(\mathbb{T})} X^{(\mathbb{T})*} \Sigma^{1/2},$$

we can further set

$$G^{(\mathbb{T})}(z) = \big(W^{(\mathbb{T})} - z\big)^{-1}, \qquad \mathcal{G}^{(\mathbb{T})}(z) = \big(\mathcal{W}^{(\mathbb{T})} - z\big)^{-1},$$

$$m_N^{(\mathbb{T})}(z) = \frac{\operatorname{Tr} G^{(\mathbb{T})}(z)}{N}, \qquad \underline{m}_N^{(\mathbb{T})}(z) = \frac{\operatorname{Tr} \mathcal{G}^{(\mathbb{T})}(z)}{M}.$$

We emphasize here, in the sequel, the names of indices of X for $X^{(\mathbb{T})}$ will be kept, that is, $X_{ij}^{(\mathbb{T})} = \mathbf{1}_{\{j \notin \mathbb{T}\}} X_{ij}$. Correspondingly, we will denote the (i, j)th entry of $G^{(\mathbb{T})}(z)$ by $G_{ij}^{(\mathbb{T})}(z)$ for all $i, j \notin \mathbb{T}$. In addition, in light of the discussion in [7] [see the truncation issue above (3.3) therein], henceforth we can and do additionally assume that

$$\max_{i,j} |\sqrt{N} x_{ij}| \leq (\log N)^C \tag{3.1}$$

with some sufficiently large positive constant C. Then we have the following theorem.

THEOREM 3.2 (Strong local MP type law around λ_r, Theorem 3.2 of [7]). *Let* $\tilde{c}$ *be the constant in Theorem* 3.1. *Under Condition* 1.1 *and assumption* (3.1), *for any* $\zeta > 0$ *there exists some constant* C_ζ *such that*

(i)

$$\bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \left\{ |m_N(z) - m_0(z)| \leq \varphi^{C_\zeta} \frac{1}{N\eta} \right\} \tag{3.2}$$

holds with ζ*-high probability, and*

(ii)

$$\bigcap_{z\in S_r(\tilde{c},5C_\zeta)}\left\{\max_{i\neq j}|G_{ij}(z)|+\max_i|G_{ii}(z)-m_0(z)|\le\varphi^{C_\zeta}\left(\sqrt{\frac{\Im m_0(z)}{N\eta}}+\frac{1}{N\eta}\right)\right\} \tag{3.3}$$

holds with ζ-high probability.

For our purpose, the following result concerning the convergence rate of ESD around λ_r is also needed, which can be essentially derived from Theorem 3.2.

THEOREM 3.3 (Convergence rate around λ_r, Theorem 4.1 of [7]). *Under Condition 1.1 and the assumption (3.1), for any $\zeta>0$ there exists a constant C_ζ such that the following events hold with ζ-high probability.*

(i) *For the largest eigenvalue $\lambda_1(\mathcal{W})$, there exists*

$$|\lambda_1(\mathcal{W})-\lambda_r|\le N^{-2/3}\varphi^{C_\zeta}.$$

(ii) *For any $E_1,E_2\in[\lambda_r-\tilde{c},C_r]$, there exists*

$$\left|\big(F_N(E_1)-F_N(E_2)\big)-\big(F_0(E_1)-F_0(E_2)\big)\right|\le N^{-1}\varphi^{C_\zeta}. \tag{3.4}$$

In addition, we record the following concentration lemma on quadratic forms, whose proof can be found in Appendix B of [22] for instance.

LEMMA 3.4. *Let $\mathbf{x}_i,\mathbf{x}_j$, $i\neq j$ be two columns of the matrix X satisfying* (ii) *of Condition 1.1. Then for any M-dimensional vector $\mathbf{b}$ and $M\times M$ matrix $\mathbf{C}$ independent of $\mathbf{x}_i$ and $\mathbf{x}_j$, the following three inequalities hold with ζ-high probability*

(i)

$$\left|\mathbf{x}_i^*\mathbf{C}\mathbf{x}_i-\frac{1}{N}\operatorname{Tr}\mathbf{C}\right|\le\frac{\varphi^{\tau\zeta}}{N}\|\mathbf{C}\|_{\mathrm{HS}},$$

(ii)

$$|\mathbf{x}_i^*\mathbf{C}\mathbf{x}_j|\le\frac{\varphi^{\tau\zeta}}{N}\|\mathbf{C}\|_{\mathrm{HS}},$$

(iii)

$$|\mathbf{b}^*\mathbf{x}_i|\le\frac{\varphi^{\tau\zeta}}{\sqrt{N}}\|\mathbf{b}\|.$$

Here, $\tau:=\tau(\tau_0)>1$ is some positive constant [*see* (ii) *of Condition 1.1 for τ_0*]. *Let $\mathfrak{X}_i$ be the conjugate transpose of the ith row of the matrix X for $i=1,\dots,M$. If we replace $\mathbf{x}_i,\mathbf{x}_j$ by $\mathfrak{X}_i$ and $\mathfrak{X}_j$ respectively, the above three inequalities also hold if $\mathbf{b}$ is an N-dimensional vector and $\mathbf{C}$ is an $N\times N$ matrix which are both independent of $\mathfrak{X}_i$ and $\mathfrak{X}_j$.*

Finally, regarding the $\|\cdot\|_{\mathrm{HS}}$ norm of a Green function, we will frequently need the fact that for any Hermitian matrix A, there is

$$\begin{aligned}\|(A-z)^{-1}\|_{\mathrm{HS}}^2 &= \operatorname{Tr}|A-z|^{-2} = \operatorname{Tr}(A-z)^{-1}(A-\bar{z})^{-1} \\ &= \frac{1}{\eta}\Im\operatorname{Tr}(A-z)^{-1},\end{aligned} \tag{3.5}$$

which can be verified easily by the spectral decomposition.

4. Universality for the largest eigenvalue. With some bounds on the entries of $\Sigma^{1/2}\mathcal{G}\Sigma^{1/2}$ granted (see Lemma 4.7 below), we can successfully prove our main results in this section via pursuing a Green function comparison strategy tailored for edge universality, which is analogous to those in [45, 46]. The proof of the desired bounds of the entries of $\Sigma^{1/2}\mathcal{G}\Sigma^{1/2}$ will be postponed to the next section, which can be viewed as our main technical ingredient of this paper.

THEOREM 4.1 (Green function comparison theorem around λ_r). *Let $\mathcal{W}^{\mathbf{u}}$ and $\mathcal{W}^{\mathbf{v}}$ be two sample covariance matrices in Theorem 1.3. Let F be a real function satisfying*

$$\sup_x \bigl|F^{(k)}(x)\bigr|/\bigl(|x|+1\bigr)^C \le C, \qquad k=0,1,2,3,4 \tag{4.1}$$

for some positive constant C. There exist $\varepsilon_0>0$ and $N_0\in\mathbb{N}$, such that for any positive constant $\varepsilon<\varepsilon_0$, $N\ge N_0$ and for any real numbers E, E_1 and E_2 satisfying $|E-\lambda_r|, |E_1-\lambda_r|, |E_2-\lambda_r| \le N^{-2/3+\varepsilon}$ and $\eta=N^{-2/3-\varepsilon}$, we have for $z=E+\mathbf{i}\eta$,

$$\bigl|\mathbb{E}F\bigl(N\eta\Im m_N^{\mathbf{u}}(z)\bigr)-\mathbb{E}F\bigl(N\eta\Im m_N^{\mathbf{v}}(z)\bigr)\bigr| \le N^{-C'\varepsilon} \tag{4.2}$$

and

$$\left|\mathbb{E}F\left(N\int_{E_1}^{E_2}\Im m_N^{\mathbf{u}}(x+\mathbf{i}\eta)\,dx\right)-\mathbb{E}F\left(N\int_{E_1}^{E_2}\Im m_N^{\mathbf{v}}(x+\mathbf{i}\eta)\,dx\right)\right| \le N^{-C'\varepsilon} \tag{4.3}$$

with some positive constant C' if either **A** *or* **B** *in Theorem 1.3 holds.*

Now we are at the stage to prove our main results assuming Theorem 4.1.

PROOF OF THEOREM 1.3. Given Theorems 3.1–3.3 and 4.1, the proof of Theorem 1.3 is nearly the same as that for the null case in [46] (see the proof of Theorem 1.1 therein). Due to the similarity, here we only sketch the main route and leave the details to the reader. We start from Theorem 3.3(i), which states that for any $\zeta>0$, there exists some positive constant C_ζ such that $|\lambda_1(\mathcal{W})-\lambda_r|\le N^{-2/3}\varphi^{C_\zeta}$ with ζ-high probability. Hence, it suffices to verify (1.7) for $s\in[-\frac{3}{2}\varphi^{C_\zeta}, \frac{3}{2}\varphi^{C_\zeta}]$. To this end, we denote $E_\zeta=\lambda_r+2N^{-2/3}\varphi^{C_\zeta}$ and set $E=\lambda_r+sN^{-2/3}$. With the above restriction on s, one can always assume that $E \le E_\zeta - \frac{1}{2}N^{-2/3}\varphi^{C_\zeta}$. Denoting $\eta_1 = N^{-2/3-9\varepsilon_1}$ and $\ell = \frac{1}{2}N^{-2/3-\varepsilon_1}$ with any given small constant $\varepsilon_1 > 0$, we record the following inequality from Corollary 5.1 of [7]:

$$\begin{aligned} &\mathbb{E}h\left(\frac{N}{\pi}\int_{E-\ell}^{E_\zeta} \Im m_N(y+\mathbf{i}\eta_1)\,dy\right) \\ &\quad \le \mathbb{P}(\lambda_1(\mathcal{W}) \le E) \\ &\quad \le \mathbb{E}h\left(\frac{N}{\pi}\int_{E+\ell}^{E_\zeta} \Im m_N(y+\mathbf{i}\eta_1)\,dy\right) + O(\exp(-\varphi^{C_\zeta})), \end{aligned} \tag{4.4}$$

where h is a smooth cutoff function satisfying the condition of F in Theorem 4.1; see Corollary 5.1 of [7] for the definition of the function h. (4.4) states that $\mathbb{P}(\lambda_1(\mathcal{W}) \le E)$ can be squeezed by the expectations of two functionals of the Stieltjes transform. Now, setting $\varepsilon = 9\varepsilon_1$, $\eta = \eta_1$, $F(x) = h(x/\pi)$, $E_1 = E - \ell$ and $E_2 = E_\zeta$ in (4.3) we obtain

$$\begin{aligned} &\mathbb{E}h\left(\frac{N}{\pi}\int_{E-\ell}^{E_\zeta} \Im m_N^{\mathbf{u}}(x+\mathbf{i}\eta_1)\,dx\right) \\ &\quad \le \mathbb{E}h\left(\frac{N}{\pi}\int_{E-\ell}^{E_\zeta} \Im m_N^{\mathbf{v}}(x+\mathbf{i}\eta_1)\,dx\right) + \frac{1}{2}N^{-\delta}, \end{aligned} \tag{4.5}$$

for sufficiently large N, where we took $\delta = \frac{1}{2}C'\varepsilon$ (say). Employing the second inequality in (4.4) via replacing E by $E - 2\ell$, we also have

$$\mathbb{P}(\lambda_1(\mathcal{W}^{\mathbf{u}}) \le E - 2\ell) \le \mathbb{E}h\left(\frac{N}{\pi}\int_{E-\ell}^{E_\zeta} \Im m_N^{\mathbf{u}}(y+\mathbf{i}\eta_1)\,dy\right) + O(\exp(-\varphi^{C_\zeta})),$$

which together with (4.5) implies that for sufficiently large N,

$$\mathbb{P}(\lambda_1(\mathcal{W}^{\mathbf{u}}) \le E - 2\ell) \le \mathbb{E}F\left(\frac{N}{\pi}\int_{E-\ell}^{E_\zeta} \Im m_N^{\mathbf{v}}(x+\mathbf{i}\eta_1)\,dx\right) + N^{-\delta}.$$

Using the first inequality in (4.4) yields

$$\mathbb{P}(\lambda_1(\mathcal{W}^{\mathbf{u}}) \le E - 2\ell) \le \mathbb{P}(\lambda_1(\mathcal{W}^{\mathbf{v}}) \le E) + N^{-\delta}. \tag{4.6}$$

Switching the roles of $\mathbf{u}$ and $\mathbf{v}$, we can analogously derive that

$$\mathbb{P}(\lambda_1(\mathcal{W}^{\mathbf{v}}) \le E) \le \mathbb{P}(\lambda_1(\mathcal{W}^{\mathbf{u}}) \le E + 2\ell) + N^{-\delta}. \tag{4.7}$$

(4.6) and (4.7) then lead to (1.7). Hence, we conclude the sketch of the proof. □

PROOF OF COROLLARY 1.5. Corollary 1.5 follows from Theorem 1.3, Theorem 1 of [12] and Proposition 2 of [38] immediately. □

Now, before commencing the proof of Corollary 1.7, we record the following lemma whose proof will be provided in the supplementary material [6].

LEMMA 4.2. *Assume that* Σ *satisfies the assumption of Corollary* 1.7. *Then* Σ *also satisfies Condition* 1.1(iii). *In addition*, *we have* (1.9).

PROOF OF COROLLARY 1.7. With the aid of Lemma 4.2, we see that $\mathcal{W}$ satisfies Condition 1.1 thus Theorem 1.3 can be adopted. Now we invoke the fact that the real Wishart matrix with population covariance matrix $\mathrm{diag}(\lambda_1(\Sigma), \dots, \lambda_M(\Sigma))$ satisfy the conditions of Theorem 1.6 of [24]. Moreover, taking the property of orthogonal invariance for Gaussian matrices into account, we know the result of [24] holds for all Wishart matrices with population covariance matrix Σ (possibly not diagonal) whose eigenvalues satisfy the condition in Corollary 1.7. We remind here the parameters N and p in [24] are corresponding to our M and N, respectively. Hence, with (1.9) at hand, choosing the Wishart matrix with population covariance matrix Σ as the reference matrix and combining our Theorem 1.3 with Theorem 1.6 of [24], we can complete the proof. $\square$

It remains to prove Theorem 4.1 in this section.

PROOF OF THEOREM 4.1. To simplify the presentation, we will only verify (4.2) in detail below. The proof of (4.3) can be taken similarly, thus we just leave it to the reader. As a compensation, some necessary modifications for the proof of (4.3) will be highlighted in Remarks 4.4 and 4.8. Now, let $\gamma \in \{1, 2, \dots, N+1\}$ and set X_γ to be the matrix whose first $\gamma - 1$ columns are the same as those of $X^{\mathbf{v}}$ and the remaining $N - \gamma + 1$ columns are the same as those of $X^{\mathbf{u}}$. Especially, we have $X_1 = X^{\mathbf{u}}$ and $X_{N+1} = X^{\mathbf{v}}$. Correspondingly, we set

$$W_{N,\gamma} = X_\gamma^* \Sigma X_\gamma, \qquad \mathcal{W}_{N,\gamma} = \Sigma^{1/2} X_\gamma X_\gamma^* \Sigma^{1/2}.$$

Then (4.2) can be achieved via checking that for every γ the following estimate holds:

$$\mathbb{E}F\big(\eta \Im \operatorname{Tr}(W_{N,\gamma} - z)^{-1}\big) - \mathbb{E}F\big(\eta \Im \operatorname{Tr}(W_{N,\gamma+1} - z)^{-1}\big) = O\big(N^{-1-C'\varepsilon}\big).$$

Observing the fact that X_γ and $X_{\gamma+1}$ only differ in the γth column yields $X_\gamma^{(\gamma)} = X_{\gamma+1}^{(\gamma)}$, which directly implies $W_{N,\gamma}^{(\gamma)} = W_{N,\gamma+1}^{(\gamma)}$ and $\mathcal{W}_{N,\gamma}^{(\gamma)} = \mathcal{W}_{N,\gamma+1}^{(\gamma)}$. Therefore, we can also write

$$\begin{aligned}
&\mathbb{E}F\big(\eta \Im \operatorname{Tr}(W_{N,\gamma} - z)^{-1}\big) - \mathbb{E}F\big(\eta \Im \operatorname{Tr}(W_{N,\gamma+1} - z)^{-1}\big) \\
&\quad = \big(\mathbb{E}F\big(\eta \Im \operatorname{Tr}(W_{N,\gamma} - z)^{-1}\big) - \mathbb{E}F\big(\eta \Im \big[\operatorname{Tr}\big(W_{N,\gamma}^{(\gamma)} - z\big)^{-1} - z^{-1}\big]\big)\big) \\
&\qquad - \big(\mathbb{E}F\big(\eta \Im \operatorname{Tr}(W_{N,\gamma+1} - z)^{-1}\big) \\
&\qquad\quad - \mathbb{E}F\big(\eta \Im \big[\operatorname{Tr}\big(W_{N,\gamma+1}^{(\gamma)} - z\big)^{-1} - z^{-1}\big]\big)\big).
\end{aligned} \tag{4.8}$$

Since the comparison process will greatly rely on the moment matching condition, it will be more convenient to work with the following set:

$$\mathcal{M}_k(i) := \big\{\big(\mathbb{E}(\Re\sqrt{N} x_{ji})^l (\Im\sqrt{N} x_{ji})^m, j, l, m\big) : j = 1, \dots, M, m + l \le k\big\},$$

that is, the set of all moments up to order k of the entries of $\sqrt{N}\mathbf{x}_i$, where its elements are indexed by j, l, m. In the spirit of (4.8), it suffices to show, for any sample covariance matrix W_N satisfying Condition 1.1, the following Lemmas 4.3 and 4.5 hold.

LEMMA 4.3. *Let F be a real function satisfying* (4.1) *and $z = E + \mathbf{i}\eta$. For any random matrix W_N satisfying Condition* 1.1, *if $|E - \lambda_r| \le N^{-2/3+\varepsilon}$ and $N^{-2/3-\varepsilon} \le \eta \ll N^{-2/3}$ for some $\varepsilon > 0$, there exists some positive constant C independent of ε such that*

$$\begin{aligned}(4.9)\qquad &\mathbb{E}F(N\eta\Im m_N(z)) - \mathbb{E}F(N\eta\Im[m_N^{(i)}(z) - (Nz)^{-1}]) \\ &\quad = A(X^{(i)}, \mathcal{M}_2(i)) + N^{-1-C\varepsilon}\end{aligned}$$

when Σ is diagonal, *and*

$$\begin{aligned}(4.10)\qquad &\mathbb{E}F(N\eta\Im m_N(z)) - \mathbb{E}F(N\eta\Im[m_N^{(i)}(z) - (Nz)^{-1}]) \\ &\quad = B(X^{(i)}, \mathcal{M}_4(i)) + N^{-1-C\varepsilon}\end{aligned}$$

for general Σ, where $A(X^{(i)}, \mathcal{M}_2(i))$ is a functional depending on the distribution of $X^{(i)}$ and $\mathcal{M}_2(i)$ only and similarly $B(X^{(i)}, \mathcal{M}_4(i))$ is a functional depending on the distribution of $X^{(i)}$ and $\mathcal{M}_4(i)$ only.

REMARK 4.4. To verify (4.3), actually we need to show two equalities analogous to (4.9) and (4.10), obtained via replacing

$$F(N\eta\Im m_N(z) \text{ and } F(N\eta\Im[m_N^{(i)}(z) - (Nz)^{-1}]))$$

by

$$F\left(N\int_{E_1}^{E_2} \Im m_N(x+\mathbf{i}\eta)\,dx\right) \quad \text{and} \quad F\left(N\int_{E_1}^{E_2} \Im[m_N^{(i)}(x+\mathbf{i}\eta) - (Nz)^{-1}]\,dx\right),$$

respectively, in (4.9) and (4.10) and correspondingly replacing $A(\cdot,\cdot)$ and $B(\cdot,\cdot)$ by some other functionals $\tilde{A}(\cdot,\cdot)$ and $\tilde{B}(\cdot,\cdot)$.

Now, to differentiate, we denote the set $\mathcal{M}_k(i)$ for $X^{\mathbf{u}}$ and $X^{\mathbf{v}}$ by $\mathcal{M}_k^{\mathbf{u}}(i)$ and $\mathcal{M}_k^{\mathbf{v}}(i)$, respectively. Then, we also have the following lemma.

LEMMA 4.5. *Under Condition* 1.1 *and the assumptions in Lemma* 4.3, *there exist some positive constants c_0 and $C > 1$, such that the following statements hold. If $\mathcal{W}^{\mathbf{u}}$ matches $\mathcal{W}^{\mathbf{v}}$ to order* 2, *we have*

$$(4.11)\qquad \max_\gamma |A(X_\gamma^{(\gamma)}, \mathcal{M}_2^{\mathbf{u}}(\gamma)) - A(X_\gamma^{(\gamma)}, \mathcal{M}_2^{\mathbf{v}}(\gamma))| = O(e^{-c_0(\log N)^C}).$$

If $\mathcal{W}^{\mathbf{u}}$ matches $\mathcal{W}^{\mathbf{v}}$ to order 4, *we have*

$$(4.12)\qquad \max_\gamma |B(X_\gamma^{(\gamma)}, \mathcal{M}_4^{\mathbf{u}}(\gamma)) - B(X_\gamma^{(\gamma)}, \mathcal{M}_4^{\mathbf{v}}(\gamma))| = O(e^{-c_0(\log N)^C}).$$

Here, *$A(\cdot,\cdot)$ and $B(\cdot,\cdot)$ are the functionals in Lemma* 4.3.

It is obvious that (4.2) follows from Lemma 4.3 and Lemma 4.5. The proof of (4.3) can be taken analogously. Thus, we conclude the proof of Theorem 4.1 assuming the validity of Lemmas 4.3 and 4.5. □

We leave the proof of Lemma 4.5 to the supplementary material [6] and prove Lemma 4.3 in the sequel. Without loss of generality, we will just check the statements in Lemma 4.3 for $i = 1$. The others are just analogous. To commence the proof, we will need the following lemma as an input, whose proof will also appear in the supplementary material [6].

LEMMA 4.6. *Under the assumptions on z and F in Lemma 4.3, for any given $\zeta > 0$, there exists some positive constant C, such that*

$$\begin{aligned} &F\big(N\eta\Im m_N(z)\big) - F\big(N\eta\Im\big[m_N^{(1)}(z) - (Nz)^{-1}\big]\big) \\ &\quad = \sum_{k=1}^{3} \frac{1}{k!} F^{(k)}\big(N\eta\Im\big[m_N^{(1)}(z) - (Nz)^{-1}\big]\big)(\Im y)^k + O\big(N^{-4/3+C\varepsilon}\big) \end{aligned} \tag{4.13}$$

holds with ζ-high probability, where

$$y := \eta z G_{11} \mathbf{r}_1^* \big(\mathcal{G}^{(1)}\big)^2 \mathbf{r}_1. \tag{4.14}$$

Moreover, we have

$$\big|\mathbf{r}_1^* \big(\mathcal{G}^{(1)}\big)^2 \mathbf{r}_1\big| \le N^{1/3+C\varepsilon}, \qquad |y| \le N^{-1/3+C\varepsilon} \tag{4.15}$$

with ζ-high probability.

With Lemma 4.6, we now start to prove Lemma 4.3 for $i = 1$.

PROOF OF LEMMA 4.3 (FOR $i = 1$). Now, starting from (4.14), we further decompose y and then pick out the leading terms in the decomposition. Specifically, we set

$$D := \frac{m_0 - G_{11}}{G_{11}} = -m_0 \cdot \big(z + z\mathbf{r}_1^* \mathcal{G}^{(1)} \mathbf{r}_1\big) - 1, \tag{4.16}$$

which is implied by the Schur complement $G_{11} = -1/(z + z\mathbf{r}_1^* \mathcal{G}^{(1)} \mathbf{r}_1)$. At first, by (i) of Theorem 3.1 and (ii) of Theorem 3.2 we can see that $G_{ii}(z) \sim 1$ with ζ-high probability. Moreover, with ζ-high probability we can write

$$G_{11} = \frac{m_0}{D+1} = m_0 \sum_{k=0}^{\infty} (-D)^k \tag{4.17}$$

since $|D| \le N^{-1/3+C\varepsilon}$ for some positive constant C, which is implied by Theorem 3.1(ii) and Theorem 3.2(ii). Inserting (4.17) into (4.14), we can write

$$y = \sum_{k=1}^{\infty} y_k, \qquad y_k := \eta z m_0 (-D)^{k-1} \mathbf{r}_1^* \big(\mathcal{G}^{(1)}\big)^2 \mathbf{r}_1. \tag{4.18}$$

By (4.15) and the aforementioned bound for D, we can easily get

$$|y_k| = O(N^{-k/3+C\varepsilon}) \tag{4.19}$$

with ζ-high probability, which directly implies that

$$\begin{aligned} \Im y &= \Im y_1 + \Im y_2 + \Im y_3 + O(N^{-4/3+C\varepsilon}), \\ (\Im y)^2 &= (\Im y_1)^2 + 2\Im y_1 \Im y_2 + O(N^{-4/3+C\varepsilon}), \\ (\Im y)^3 &= (\Im y_1)^3 + O(N^{-4/3+C\varepsilon}) \end{aligned} \tag{4.20}$$

hold with ζ-high probability. By the discussions in the proof of Lemma 4.6 in the supplementary material [6], one can see that $N\eta\Im m_N(z) = O(N^{C\varepsilon})$ and $N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}] = O(N^{C\varepsilon})$ with ζ-high probability for any given $\zeta > 0$. Consequently, in light of the assumption (4.1), we see that for any real number t_N between $N\eta\Im m_N(z)$ and $N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}]$, there exist

$$|F^{(k)}(t_N)| = O(N^{C\varepsilon}), \qquad k = 0, 1, 2, 3, 4 \tag{4.21}$$

with ζ-high probability. Moreover, we have the deterministic bound $|m_N(z)|$, $|m_N^{(1)}(z)| = O(\eta^{-1})$, which implies $|N\eta\Im m_N(z)|, |N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}]| = O(N)$. Thus, using (4.1) again we have the deterministic bound $|F^{(k)}(t_N)| = O(N^C), k = 0, 1, 2, 3, 4$, for any real number t_N between $N\eta\Im m_N(z)$ and $N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}]$. Analogously, by using the fact that $\|\mathcal{G}^{-1}\|_{\text{op}} = O(\eta^{-1})$ and condition (3.1), we can get the deterministic bound $|y| = O(N^C)$, $|y_k| = O(N^{C(k)})$ with some positive constants C and $C(k)$ (depending on k), for $k = 0, 1, 2, 3, 4$. Then by (4.13), (4.18)–(4.21) and the deterministic bounds above, it is not difficult to find that

$$\begin{aligned} &\mathbb{E}F(N\eta\Im m_N(z)) - \mathbb{E}F(N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}]) \\ &\quad = \mathbb{E}F^{(1)}(N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}])(\Im y_1 + \Im y_2 + \Im y_3) \\ &\qquad + \mathbb{E}F^{(2)}(N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}])\left(\tfrac{1}{2}(\Im y_1)^2 + \Im y_1 \Im y_2\right) \\ &\qquad + \mathbb{E}F^{(3)}(N\eta\Im[m_N^{(1)}(z) - (Nz)^{-1}])\left(\tfrac{1}{6}\Im y_1\right)^3 + O(N^{-4/3+C\varepsilon}). \end{aligned} \tag{4.22}$$

Toward the right-hand side of (4.22), our task is to extract the terms depending on $X^{(1)}$ and $\mathcal{M}_k(1)$ ($k = 2$ or 4) only and bound the remaining terms. For the latter, we will need the following crucial lemma on bounding the entries of $\Sigma^{1/2}\mathcal{G}^{(1)}\Sigma^{1/2}$.

LEMMA 4.7. *Let $z = E + \mathbf{i}\eta$ with $|E - \lambda_r| \le N^{-2/3+\varepsilon}$ and $N^{-2/3-\varepsilon} \le \eta \ll N^{-2/3}$ for some $\varepsilon > 0$. When Σ is diagonal, for any given $\zeta > 0$, we have*

$$\begin{aligned} &|(\mathcal{G}^{(1)}(z))_{ij}| \le N^{C\varepsilon} \quad \text{and} \\ &|([\mathcal{G}^{(1)}(z)]^2)_{ij}| \le N^{1/3+C\varepsilon}, \qquad i, j \in \{1, \dots, M\} \end{aligned} \tag{4.23}$$

hold with ζ-high probability for some positive constant C independent of ε.

For general Σ *satisfying Condition* 1.1(iii), *we have for any given* $\zeta > 0$,

$$\big|\big(\Sigma^{1/2}\mathcal{G}^{(1)}(z)\Sigma^{1/2}\big)_{kk}\big| \le N^{1/3+C\varepsilon}, \qquad k \in \{1, \dots, M\} \tag{4.24}$$

hold with ζ*-high probability for some positive constant* C *independent of* ε.

REMARK 4.8. When we prove (4.3), as mentioned above, we actually need to verify the statement in Remark 4.4. To this end, we need to strengthen the bounds in (4.23) and (4.24) to hold with ζ-high probability uniformly on the set $\{z = E + \mathbf{i}\eta : |E - \lambda_r| \le N^{-2/3+\varepsilon}$ and $N^{-2/3-\varepsilon} \le \eta \ll N^{-2/3}\}$. These uniform bounds are necessary for the proof of the statement in Remark 4.4, since some integrations taken w.r.t. the real part of z are involved in the discussion. These stronger bounds can be obtained from the bounds for single point in (4.23) and (4.24) through some routine ε-net and Lipschitz continuity argument. One can refer to the extension from (5.1) to (5.2) below for a similar argument.

Lemma 4.7 is our main technical task whose proof will be postponed to the next section separately. Now, with Lemma 4.7 granted, we prove (4.9) and (4.10) in the sequel. At first, we will verify (4.9) for diagonal Σ. We start with the third term on the r.h.s. of (4.22). Denoting $\varpi := \Im \eta z m_0$ and $\varrho := \Re \eta z m_0$, we have

$$\Im y_1 = \varpi\big(\Re \mathbf{r}_1^*\big(\mathcal{G}^{(1)}\big)^2\mathbf{r}_1\big) + \varrho\big(\Im \mathbf{r}_1^*\big(\mathcal{G}^{(1)}\big)^2\mathbf{r}_1\big).$$

To further simplify the exposition, we denote the real part and imaginary part of a complex number A by $A[0]$ and $A[1]$ respectively. Introducing the notation $\mathbb{E}_i$ to denote the expectation with respect to $\mathbf{x}_i$, we see that $\mathbb{E}_1(\Im y_1)^3$ is a summation of finite terms of the form

$$\mathbf{1}_{\{a+b=3\}}\varpi^a\varrho^b \sum_{k_1,\dots,k_6}\prod_{i=1}^{3}\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2\Sigma^{1/2}\big)_{k_{2i-1}k_{2i}}[\alpha_i]\mathbb{E}\prod_{l=1}^{6}x_{k_l,1}[\beta_l], \tag{4.25}$$

where α_i, β_l are 0 or 1 and a, b are nonnegative integers. Hence, it suffices to analyze the quantities of the form (4.25) below.

We classify the terms in the summation (4.25) by various coincidence conditions of the indices $k_1, \dots, k_6$. If there is a k_j appearing only once in $\{k_1, \dots, k_6\}$, then this term is zero obviously, due to the independence and centering of the entries of X. Now we proceed to those terms in which each k_j appears exactly twice. Apparently, these terms only depend on $X^{(1)}$ and $\mathcal{M}_2(1)$. Finally, it remains to consider the terms that there is at least one k_j appearing at least three times and no k_j appearing only once. It is obviously that the total number of such terms is $O(N^2)$. Putting this observation and (4.23) in Lemma 4.7 together yields the fact that the total contribution of these terms is less than

$$\begin{aligned} CN^{-1}|\eta z m_0|^3 \max_{ij}\big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2\Sigma^{1/2}\big)_{ij}\big|^3 &\le CN^{-1}|\eta z m_0|^3 \max_{ij}\big|\big(\big(\mathcal{G}^{(1)}\big)^2\big)_{ij}\big|^3 \\ &= O\big(N^{-2+C\varepsilon}\big) \end{aligned}$$

with ζ-high probability. Since $|(\Sigma^{1/2}\mathcal{G}^{(1)}\Sigma^{1/2})_{ij}|$ are trivially bounded by $O(\eta^{-1})$, one can see that the above bound also holds in expectation by the definition of ζ-high probability. Therefore, we deduce that

$$\mathbb{E}(\Im y_1)^3 = A_1\big(X^{(1)}, \mathcal{M}_2(1)\big) + O\big(N^{-2+C\varepsilon}\big) \tag{4.26}$$

for some functional A_1 depending on the distribution of $X^{(1)}$ and $\mathcal{M}_2(1)$ only.

Now, for the first and second term on the right-hand side of (4.22) we can deal with them analogously. Note that by (4.16) and the definitions of y_2, y_3, one can see that

$$y_2 = \eta z^2 m_0^2 \mathbf{r}_1^* \mathcal{G}^{(1)} \mathbf{r}_1 \cdot \mathbf{r}_1^* \big(\mathcal{G}^{(1)}\big)^2 \mathbf{r}_1 + \eta z m_0 \cdot (1 + z m_0) \cdot \mathbf{r}_1^* \big(\mathcal{G}^{(1)}\big)^2 \mathbf{r}_1,$$

$$y_3 = \eta z^3 m_0^3 \big(\mathbf{r}_1^* \mathcal{G}^{(1)} \mathbf{r}_1\big)^2 \cdot \mathbf{r}_1^* \big(\mathcal{G}^{(1)}\big)^2 \mathbf{r}_1 + 2\eta z^2 m_0 \cdot (1 + z m_0) \cdot \mathbf{r}_1^* \mathcal{G}^{(1)} \mathbf{r}_1 \cdot \mathbf{r}_1^* \big(\mathcal{G}^{(1)}\big)^2 \mathbf{r}_1 + \eta z m_0 \cdot (1 + z m_0) \mathbf{r}_1^* \mathcal{G}^{(1)} \mathbf{r}_1.$$

Expanding each term above, then by a routine but detailed discussion on the coincidence condition of the indices as what we have done to the third term on the right-hand side of (4.22) above, we can actually get

$$\mathbb{E}\big((\Im y_1)^2 + 2(\Im y_1)(\Im y_2)\big) = A_2\big(X^{(1)}, \mathcal{M}_2(1)\big) + O\big(N^{-5/3+C\varepsilon}\big) \tag{4.27}$$

and

$$\mathbb{E}(\Im y_1 + \Im y_2 + \Im y_3) = A_3\big(X^{(1)}, \mathcal{M}_2(1)\big) + O\big(N^{-4/3+C\varepsilon}\big) \tag{4.28}$$

for some functionals A_2 and A_3 depending on $X^{(1)}$ and $\mathcal{M}_2(1)$ only. Inserting (4.26)–(4.28) into (4.22), we obtain (4.9).

Now, we go ahead to investigate (4.10) for more general Σ. At first, we revisit the canonical form of the terms in the expansion of $\mathbb{E}_1(\Im y_1)^3$, that is, (4.25). Note that for (4.25), it suffices to bound the terms in which all $k_i, i = 1, \dots, 6$ are the same, since all the other terms only depend on the distribution of $X^{(1)}$ and $\mathcal{M}_4(1)$. In other words, we need to bound the terms in which $\mathbb{E}|x_{k1}|^6$ appears. Analogously, writing $\mathbb{E}_1((\Im y_1)^2 + 2(\Im y_1)(\Im y_2))$ and $\mathbb{E}_1(\Im y_1 + \Im y_2 + \Im y_3)$ as some summations of terms in the forms similar to (4.25), again, we only need to address the terms containing $\mathbb{E}|x_{k1}|^6$ as a factor. It is not difficult to see after simple calculations that the total contribution of such terms in $\mathbb{E}_1(\Im y_1)^3$, $\mathbb{E}_1((\Im y_1)^2 + 2(\Im y_1)(\Im y_2))$ and $\mathbb{E}_1(\Im y_1 + \Im y_2 + \Im y_3)$ can be bounded by

$$\begin{aligned} & CN^{-3} \sum_k \eta^3 \big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2 \Sigma^{1/2}\big)_{kk}\big|^3 \\ & \quad + CN^{-3} \sum_k \eta^2 \big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2 \Sigma^{1/2}\big)_{kk}\big| \big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2 \Sigma^{1/2}\big)_{kk}\big|^2 \\ & \quad + CN^{-3} \sum_k \eta \big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big) \Sigma^{1/2}\big)_{kk}\big|^2 \big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2 \Sigma^{1/2}\big)_{kk}\big| \end{aligned} \tag{4.29}$$

for some positive constant C. Noticing the elementary relation

$$\begin{aligned}\big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2\Sigma^{1/2}\big)_{kk}\big| &\le \eta^{-1}\big|\Im\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)\Sigma^{1/2}\big)_{kk}\big| \\ &\le \eta^{-1}\big|\big(\Sigma^{1/2}\mathcal{G}^{(1)}\Sigma^{1/2}\big)_{kk}\big|,\end{aligned} \tag{4.30}$$

it thus suffices to bound the last term of (4.29). In addition, combining (4.30) and (4.24) we see

$$\big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2\Sigma^{1/2}\big)_{kk}\big| \le N^{1+C\varepsilon}$$

holds with ζ-high probability. Finally, the estimate of the last term of (4.29) can be addressed as follows:

$$\begin{aligned}&N^{-3}\sum_k \eta\big|\big(\Sigma^{1/2}\mathcal{G}^{(1)}\Sigma^{1/2}\big)_{kk}\big|^2\big|\big(\Sigma^{1/2}\big(\mathcal{G}^{(1)}\big)^2\Sigma^{1/2}\big)_{kk}\big| \\ &\quad\le N^{-7/3+C\varepsilon}\sum_k\big|\big(\Sigma^{1/2}\mathcal{G}^{(1)}\Sigma^{1/2}\big)_{kk}\big| \le N^{-7/3+C\varepsilon}\,\mathrm{Tr}\big(\Sigma^{1/2}\big|\mathcal{G}^{(1)}\big|\Sigma^{1/2}\big) \\ &\quad\le C'N^{-7/3+C\varepsilon}\,\mathrm{Tr}\big|\mathcal{G}^{(1)}\big| = O\big(N^{-4/3+C\varepsilon}\big)\end{aligned}$$

holds with ζ-high probability, where in the last step we used the fact that $\mathrm{Tr}\,|\mathcal{G}^{(1)}| \le N^{1+\varepsilon}$ with ζ-high probability for any fixed $\zeta > 0$, which has been proved in [7] (see Lemma 3.10 therein). Again, since $|(\Sigma^{1/2}\mathcal{G}^{(1)}\Sigma^{1/2})_{ij}|$ are trivially bounded by $O(\eta^{-1})$, the above bound also holds in expectation. Thus, we complete the proof of (4.10). □

5. Bounds on the entries of $\boldsymbol{\Sigma^{1/2}\mathcal{G}\Sigma^{1/2}}$. In this section, we prove Lemma 4.7. Substantially different strategies will be adopted for the proofs of (4.23) and (4.24). Thus we will perform them separately. Moreover, since $\mathcal{G}^{(i)}$ and $\mathcal{G}$ are only different in dimension (observing that they share the same population covariance matrix Σ), we will harmlessly work on $\mathcal{G}$ for simplicity.

PROOF OF (4.23) (WITH $\mathcal{G}^{(1)}$ REPLACED BY $\mathcal{G}$). Note when Σ is diagonal, we can denote it as $\Sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_M^2)$. Let $\Sigma^{[j]}$ be the $(M-1)\times(M-1)$ minor of Σ, obtained by deleting the jth column and row of Σ. Moreover, we denote the jth row of X by $\mathfrak{X}_j^*$ thus its conjugate transpose by $\mathfrak{X}_j$, and denote by $X^{[j]}$ the $(M-1)\times N$ submatrix obtained via deleting $\mathfrak{X}_j^*$ from X. Correspondingly, we will use the notation

$$\mathcal{W}^{[j]} = \big(\Sigma^{[j]}\big)^{1/2}X^{[j]}\big(X^{[j]}\big)^*\big(\Sigma^{[j]}\big)^{1/2}, \qquad W^{[j]} = \big(X^{[j]}\big)^*\Sigma^{[j]}X^{[j]},$$

$$\mathcal{G}^{[j]} = \big(\mathcal{W}^{[j]} - z\big)^{-1}, \qquad G^{[j]} = \big(W^{[j]} - z\big)^{-1}.$$

Employing the Schur complement yields

$$\mathcal{G}_{ii} = \frac{1}{-z - z\sigma_i^2\mathfrak{X}_i^*((X^{[i]})^*\Sigma^{[i]}X^{[i]} - z)^{-1}\mathfrak{X}_i} = \frac{1}{-z - z\sigma_1^2\mathfrak{X}_i^*G^{[i]}\mathfrak{X}_i}.$$

Then by using Lemma 3.4 and Theorem 3.2 again, we can actually get the following lemma, whose proof will be provided in the supplementary material [6] in detail.

LEMMA 5.1. *For any $\zeta > 0$ given, there exists some positive constant C_ζ, such that for any $z \in S_r(\tilde{c}, 5C_\zeta)$,*

$$\mathcal{G}_{ii} = 1/\big(-z - z\sigma_i^2 m_0(z) + o(1)\big)$$

holds with ζ-high probability.

Now we proceed to the proof of (4.23). Ensured by (iii) of Theorem 3.1, we deduce from Lemma 5.1 that for $z \in S_r(\tilde{c}, 5C_\zeta)$,

$$\big|\mathcal{G}_{ii}(z)\big| \le C \tag{5.1}$$

with ζ-high probability for some positive constant C independent of z. Therefore, we get the bound for $\mathcal{G}_{ii}$ when Σ is diagonal. Actually we can strengthen (5.1) to the uniform bound as

$$\sup_{z \in S_r(\tilde{c}, 5C_\zeta)} \big|\mathcal{G}_{ii}(z)\big| = O(1) \tag{5.2}$$

with ζ-high probability. To see this, we can assign an ε-net on the region $S_r(\tilde{c}, 5C_\zeta)$ with $\varepsilon = N^{-100}$ (say). Then by the definition of ζ-high probability, we see that (5.1) holds for all z in this ε-net uniformly with ζ-high probability. Moreover, note $|\mathcal{G}'_{ii}(z)| \le N^2$ for $z \in S_r(\tilde{c}, 5C_\zeta)$, thus by the Lipschitz continuity, we can extend the bound to the whole region $S_r(\tilde{c}, 5C_\zeta)$ easily.

Now, we are ready to use (5.2) to derive the aforementioned delocalization property for the eigenvectors of $\mathcal{W}$ in the edge case. Then we use the delocalization result to bound $\mathcal{G}_{ij}$ and $(\mathcal{G}^2)_{ij}$ in return. Denoting the unit eigenvector of $\mathcal{W}$ corresponding to $\lambda_k(\mathcal{W})$ by

$$\mathbf{u}_k = (u_{k1}, \dots, u_{kM})^T,$$

we can formulate the following lemma.

LEMMA 5.2. *When Σ is diagonal, for $\lambda_k(\mathcal{W}) \in [\lambda_r - \tilde{c}/2, C_r]$, we have*

$$\max_i |u_{ki}|^2 \le \varphi^{C_\zeta} N^{-1}$$

with ζ-high probability.

PROOF. By (5.2) and the spectral decomposition, we have

$$\Im \mathcal{G}_{ii}(z) = \sum_{k=1}^{M} \frac{\eta}{(\lambda_k(\mathcal{W}) - E)^2 + \eta^2} |u_{ki}|^2 = O(1),$$

with ζ-high probability. Now we set $\eta = \varphi^{C_\zeta} N^{-1}$. In light of (5.2), we can set $E = \lambda_k(\mathcal{W}$ if $\lambda_k(\mathcal{W}) \in [\lambda_r - \tilde{c}/2, C_r]$. Then with ζ-high probability,

$$\frac{\eta}{(\lambda_k(\mathcal{W}) - E)^2 + \eta^2} |u_{ki}|^2 = \varphi^{-C_\zeta} N |u_{ki}|^2 = O(1),$$

which implies Lemma 5.2 immediately. Thus, we complete the proof. □

Now relying on the above delocalization property, we proceed to prove (4.23). Note that by the spectral decomposition, for z satisfying the assumption in Lemma 4.7 and $\alpha = 1, 2$ we have

$$\begin{aligned} \left|\left(\mathcal{G}^\alpha(z)\right)_{ij}\right| &\le \sum_{k=1}^M \frac{1}{|\lambda_k(\mathcal{W}) - z|^\alpha} |u_{ki}||u_{kj}| \\ &\le \sum_{k:\lambda_k \in [\lambda_r - \tilde{c}/2, C_r]} \frac{1}{|\lambda_k - z|^\alpha} |u_{ki}||u_{kj}| + O(1) \\ &\le \varphi^{C_\zeta} \frac{1}{N} \sum_{k=1}^M \frac{1}{|\lambda_k(\mathcal{W}) - z|^\alpha} + O(1) \end{aligned} \tag{5.3}$$

with ζ-high probability. When $\alpha = 2$, we see that with ζ-high probability,

$$\frac{1}{N} \sum_{k=1}^M \frac{1}{|\lambda_k(\mathcal{W}) - z|^2} = \eta^{-1} \Im m_N(z) = \eta^{-1} \left(\Im m_0(z) + O\left(\frac{\varphi^{C_\zeta}}{N\eta} \right) \right)$$

according to (3.5) and (3.2). From (ii) of Theorem 3.1 we have $\Im m_0(z) \sim \eta/\sqrt{\kappa + \eta}$. Noticing the assumptions on E and η in Lemma 4.7, we immediately get that the second inequality in (4.23) holds. Now, when $\alpha = 1$, we claim that for some sufficiently large constant $C_\zeta > 0$,

$$\frac{1}{N} \sum_{k=1}^M \frac{1}{|\lambda_k(\mathcal{W}) - z|} = \frac{1}{N} \operatorname{Tr}\left|\mathcal{G}(z)\right| = O\left((\log N)^{C_\zeta}\right) \tag{5.4}$$

holds with ζ-high probability. Such a bound has been established in Lemma 3.10 of [7] for $\frac{1}{N} \operatorname{Tr} |\mathcal{G}^{(i)}(z)|$ by using the strong local MP type law. It is just the same to check its validity for $\frac{1}{N} \operatorname{Tr} |\mathcal{G}(z)|$ [bearing in mind that for (4.23) what we really need is the bound for $\frac{1}{N} \operatorname{Tr} |\mathcal{G}^{(1)}(z)|$]. So we will not reproduce the details here. Therefore, we complete the proof of (4.23). □

Now we start to tackle the much more complicated case, that is, (4.24) for general Σ.

PROOF OF (4.24) (WITH $\mathcal{G}^{(1)}$ REPLACED BY $\mathcal{G}$). For simplicity, we will also work with $\mathcal{G}$ instead of $\mathcal{G}^{(1)}$. Note that

$$\Sigma^{1/2} \mathcal{G} \Sigma^{1/2} = \Sigma^{1/2} \left(\Sigma^{1/2} X X^* \Sigma^{1/2} - zI\right)^{-1} \Sigma^{1/2} = \left(X X^* - z\Sigma^{-1}\right)^{-1}. \tag{5.5}$$

For convenience, we use the notation $\Phi := \Sigma^{-1}$ and recall $\Delta = \Delta(z) := \Sigma^{1/2}\mathcal{G}(z)\Sigma^{1/2}$ defined in Introduction. Thus, we have $\Delta_{kk} := \Delta_{kk}(z) = [(XX^* - z\Phi)^{-1}]_{kk}$. An elementary observation from the spectral decomposition is

$$\Delta_{kk} = \sum_{i=1}^{M} \frac{1}{\lambda_i(\mathcal{W}) - z} (\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{kk}. \tag{5.6}$$

We will only provide the estimate for Δ_{11} in the sequel, since the others can be handled analogously. The following lemma lies at the core of our subsequent discussion.

LEMMA 5.3. *Let $z_0 := E_0 + \mathbf{i}\eta_0$ satisfy $E_0 \in [\lambda_r - \tilde{c}, \lambda_r + N^{-2/3+\varepsilon}]$ and $\eta_0 := N^{-2/3+A_0\varepsilon}$ for some positive constant $A_0 > 1$ independent of ε. Under Condition 1.1, for any given constant $\zeta > 0$ we have*

$$\sup_{E_0 \in [\lambda_r - \tilde{c}, \lambda_r + N^{-2/3+\varepsilon}]} |\Delta_{11}(z_0)| \le C\eta_0^{-1/2} \tag{5.7}$$

with ζ-high probability for some positive constant C.

We postpone the proof of Lemma 5.3 to the end of this section and proceed to prove (4.24) by assuming Lemma 5.3. By (5.7) and the spectral decomposition we have

$$C\eta_0^{-1/2} \ge \Im\Delta_{11}(z_0) = \sum_i \frac{\eta_0}{(\lambda_i(\mathcal{W}) - E_0)^2 + \eta_0^2} (\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11} \tag{5.8}$$

with ζ-high probability. We set in (5.8) that $E_0 = \lambda_i(\mathcal{W})$ for some $\lambda_i(\mathcal{W}) \in [\lambda_r - \tilde{c}, \lambda_r + N^{-2/3+\varepsilon}]$. Immediately, (5.8) implies that

$$(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11} \le C\eta_0^{1/2} \le N^{-1/3+A_0\varepsilon} \tag{5.9}$$

holds with ζ-high probability. (5.9) together with (5.7) can then be employed to bound $\Delta_{11}(z)$, for all z satisfying the assumption of Lemma 4.7. We perform it as follows. At first, according to (i) of Theorem 3.3, we can assume that $\lambda_1(\mathcal{W}) \le \lambda_r + N^{-2/3+\varepsilon}$. Now, for $z = E + \mathbf{i}\eta$, we choose $E_0 = E$ thus $z_0 = E + \mathbf{i}\eta_0$. Again, by the spectral decomposition, we see that

$$\begin{aligned} |\Delta_{11}(z) - \Delta_{11}(z_0)| &= \left| \sum_{i=1}^{M} \left(\frac{1}{\lambda_i(\mathcal{W}) - z} - \frac{1}{\lambda_i(\mathcal{W}) - z_0} \right) (\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11} \right| \\ &\le (\eta_0 - \eta) \sum_{i=1}^{M} \frac{(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11}}{|(\lambda_i(\mathcal{W}) - z)(\lambda_i(\mathcal{W}) - z_0)|} \\ &\le (\eta_0 - \eta) \sum_{i=1}^{M} \frac{(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11}}{|\lambda_i(\mathcal{W}) - z|^2} \end{aligned} \tag{5.10}$$

$$= (\eta_0 - \eta)\eta^{-1} \sum_{i=1}^{M} \Im \frac{(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11}}{\lambda_i(\mathcal{W}) - z}$$
$$\le N^{2A_0\varepsilon} \sum_{i=1}^{M} \Im \frac{(\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11}}{\lambda_i(\mathcal{W}) - z}$$

with ζ-high probability. Now we split the index collection $\{1, \dots, M\}$ into two parts as

$$I_1 := \{i : \lambda_i(\mathcal{W}) \in [\lambda_r - \tilde{c}, \lambda_r + N^{-2/3+\varepsilon}]\}, \qquad I_2 := \{i : \lambda_i(\mathcal{W}) < \lambda_r - \tilde{c}\}.$$

Combining (5.9), (5.10) and the assumption on z yields

$$\begin{aligned}
&|\Delta_{11}(z) - \Delta_{11}(z_0)| \\
&\quad \le N^{-1/3+3A_0\varepsilon} \sum_{i \in I_1} \Im \frac{1}{\lambda_i(\mathcal{W}) - z} + CN^{2A_0\varepsilon}\eta \sum_{i \in I_2} (\Sigma^{1/2}\mathbf{u}_i\mathbf{u}_i^*\Sigma^{1/2})_{11} \\
&\quad \le N^{2/3+3A_0\varepsilon} \Im m_N(z) + CN^{2A_0\varepsilon}\eta\Sigma_{11} \\
&\quad \le N^{2/3+4A_0\varepsilon}\left(\Im m_0(z) + \frac{1}{N\eta}\right) + CN^{-2/3+2A_0\varepsilon}\Sigma_{11} \\
&\quad \le N^{1/3+5A_0\varepsilon}
\end{aligned}$$

with ζ-high probability. Here in the last two inequalities we used (3.2) and (ii) of Theorem 3.1, along with the fact that Σ_{11} is bounded. Hence, we have

$$|\Delta_{11}(z)| \le |\Delta_{11}(z_0)| + N^{1/3+5A_0\varepsilon} \le N^{1/3+6A_0\varepsilon} \tag{5.11}$$

with ζ-high probability. Thus (4.24) follows if we replace $\mathcal{G}$ by $\mathcal{G}^{(1)}$. □

The remaining part of this section will be devoted to the proof of Lemma 5.3.

PROOF OF LEMMA 5.3. At first, analogous to the derivation of (5.2) via (5.1), the verification of (5.7) can be reduced to providing the desired bound on $|\Delta_{11}(z_0)|$ for any single z_0 with $E_0 \in [\lambda_r - \tilde{c}, \lambda_r + N^{-2/3+\varepsilon}]$. Hence, in the sequel, we will just fix E_0. The extension to the uniform bound via Lipschitz continuity and ε-net argument is just routine. We recall the notation $\mathfrak{X}_j$ and $X^{[j]}$ in the proof of (4.23). For simplicity, we further write

$$\Sigma^{-1} = \Phi := \begin{pmatrix} \phi_{11} & \Phi_1^* \\ \Phi_1 & \Phi^{[1]} \end{pmatrix},$$

where ϕ_{11} is the $(1, 1)$th entry of Φ and Φ_1 is its first column with ϕ_{11} removed. As the inverse of Σ, we know that Φ is also positive-definite and its eigenvalues are bounded both from below and above, in light of Condition 1.1. Consequently, its entries are also bounded, so is $\|\Phi_1\|$. Now by using Schur complement to (5.5) we can deduce that

$$\begin{aligned}\Delta_{11}(z_0)& \\ &= 1/\big(\mathfrak{X}_1^*\mathfrak{X}_1 - z_0\phi_{11} \\ &\qquad - \big(\mathfrak{X}_1^*(X^{[1]})^* - z_0\Phi_1^*\big)\big(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]}\big)^{-1}\big(X^{[1]}\mathfrak{X}_1 - z_0\Phi_1\big)\big) \\ &:= \frac{1}{D_1 + D_2 + D_3},\end{aligned}$$

where $D_i := D_i(z_0), i = 1, 2, 3$, whose explicit formulas are as follows,

$$\begin{aligned}D_1 &:= \mathfrak{X}_1^*\mathfrak{X}_1 - z_0\phi_{11} - \mathfrak{X}_1^*(X^{[1]})^*\big(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]}\big)^{-1}X^{[1]}\mathfrak{X}_1, \\ D_2 &:= -z_0^2\Phi_1^*\big(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]}\big)^{-1}\Phi_1, \\ D_3 &:= z_0\mathfrak{X}_1^*(X^{[1]})^*\big(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]}\big)^{-1}\Phi_1 \\ &\qquad + z_0\Phi_1^*\big(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]}\big)^{-1}X^{[1]}\mathfrak{X}_1.\end{aligned}$$

Our starting point is the following elementary inequality:

$$|\Delta_{11}(z_0)| \le \min\big\{\big(|\Im(D_1 + D_2 + D_3)|\big)^{-1}, |\Re(D_1 + D_2 + D_3)|^{-1}\big\}. \tag{5.12}$$

Observe that if $|\Re(D_1 + D_2 + D_3)| > N^{1/6}$, the bound for $|\Delta_{11}(z_0)|$ in (5.7) automatically holds. Hence, it suffices to show that with ζ-high probability,

$$|\Im(D_1 + D_2 + D_3)| \ge C\eta_0^{1/2} \tag{5.13}$$

when

$$|\Re(D_1 + D_2 + D_3)| \le N^{1/6} \tag{5.14}$$

for some positive constant C. In order to verify (5.13) under assumption (5.14), a careful analysis on the real and imaginary parts of D_1, D_2, D_3 is required. We perform it as follows. We start from the following reduction on D_1,

$$\begin{aligned}D_1 &= \mathfrak{X}_1^*\mathfrak{X}_1 - z_0\phi_{11} \\ &\qquad - \mathfrak{X}_1^*(X^{[1]})^*(\Phi^{[1]})^{-1/2}\big((\Phi^{[1]})^{-1/2}X^{[1]}(X^{[1]})^*(\Phi^{[1]})^{-1/2} - z_0\big)^{-1} \\ &\qquad \times (\Phi^{[1]})^{-1/2}X^{[1]}\mathfrak{X}_1 \\ &= \mathfrak{X}_1^*\mathfrak{X}_1 - z_0\phi_{11} - \mathfrak{X}_1^*(X^{[1]})^*(\Phi^{[1]})^{-1}X^{[1]}\big((X^{[1]})^*(\Phi^{[1]})^{-1}X^{[1]} - z_0\big)^{-1}\mathfrak{X}_1 \\ &= -z_0\phi_{11} - z_0\mathfrak{X}_1^*\big((X^{[1]})^*(\Phi^{[1]})^{-1}X^{[1]} - z_0\big)^{-1}\mathfrak{X}_1.\end{aligned}$$

In the second equality above, we have used the elementary fact that for any $m \times n$ matrix A

$$A\big(A^*A - z_0I_n\big)^{-1}A^* = AA^*\big(AA^* - z_0I\big)^{-1},$$

which can be checked by the singular decomposition easily. To abbreviate, we use the notation

$$\widetilde{G}^{[1]}(z_0) := ((X^{[1]})^*(\Phi^{[1]})^{-1}X^{[1]} - z_0)^{-1}.$$

Adopting Lemma 3.4 again, we obtain

$$D_1 = -z_0\phi_{11} - z_0\frac{1}{N}\operatorname{Tr}\widetilde{G}^{[1]}(z_0) + O\left(\frac{\varphi^{C_\zeta}}{N}\|\widetilde{G}^{[1]}(z_0)\|_{\mathrm{HS}}\right) \tag{5.15}$$

with ζ-high probability. Now, we need the following lemma whose proof will be also stated in the supplementary material [6].

LEMMA 5.4. *Under the above notation, we can show that*

$$\frac{1}{N}\operatorname{Tr}\widetilde{G}^{[1]}(z_0) = m_N(z_0) + O\left(\frac{1}{N\eta_0}\right). \tag{5.16}$$

Denoting $\kappa_0 := |\lambda_r - E_0|$, we deduce from Lemma 5.4 that

$$\left|\frac{1}{N}\operatorname{Tr}\widetilde{G}^{[1]}(z_0)\right| = O(1), \qquad \frac{1}{N}\Im\operatorname{Tr}\widetilde{G}^{[1]}(z_0) \sim \sqrt{\kappa_0+\eta_0} \tag{5.17}$$

hold with ζ-high probability, by combining (3.2) and (i)–(ii) of Theorem 3.1. By (3.5), we have $\|\widetilde{G}^{[1]}(z_0)\|_{\mathrm{HS}} = \sqrt{\Im\operatorname{Tr}\widetilde{G}^{[1]}(z_0)/\eta_0}$, which together with (5.15) and (5.17) implies that

$$|\Re D_1(z_0)| \le |D_1(z_0)| = O(1) \tag{5.18}$$

and

$$\Im D_1(z_0) = -E_0\frac{1}{N}\Im\operatorname{Tr}\widetilde{G}^{[1]}(z_0) + O\left(\varphi^{C_\zeta}\sqrt{\frac{\Im\operatorname{Tr}\widetilde{G}^{[1]}(z_0)}{N^2\eta_0}}\right) + O(\eta_0) \tag{5.19}$$

with ζ-high probability. Here, we also used the fact that $|\phi_{11}|$ is bounded. Then by (5.16), (5.19) and (3.2) we have

$$\Im D_1 = -E_0\Im m_0(z_0) + O(N^{-1/3-C\varepsilon}) \tag{5.20}$$

with ζ-high probability.

We proceed to the analysis toward D_2 and D_3. For D_2, by definition we have

$$\begin{aligned}\Re D_2(z_0) = {} & -(E_0^2-\eta_0^2)\Re\Phi_1^*(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]})^{-1}\Phi_1 \\ & + 2E_0\eta_0\Im\Phi_1^*(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]})^{-1}\Phi_1,\end{aligned} \tag{5.21}$$

$$\begin{aligned}\Im D_2(z_0) = {} & -(E_0^2-\eta_0^2)\Im\Phi_1^*(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]})^{-1}\Phi_1 \\ & - 2E_0\eta_0\Re\Phi_1^*(X^{[1]}(X^{[1]})^* - z_0\Phi^{[1]})^{-1}\Phi_1.\end{aligned} \tag{5.22}$$

Now, for D_3, we have the following lemma whose proof will be presented in the supplementary material [6].

LEMMA 5.5. *Assume that z_0 satisfies the assumption in Lemma 5.3. For any $\zeta > 0$, there exists some constant C_ζ such that*

$$|D_3| \le \frac{\varphi^{C_\zeta}}{\sqrt{N}} \sqrt{\eta_0^{-1} \Im \Phi_1^* \big(X^{[1]} \big(X^{[1]}\big)^* - z_0 \Phi^{[1]}\big)^{-1} \Phi_1} \tag{5.23}$$

holds with ζ-high probability.

Now we invoke the crude bound

$$\Im \Phi_1^* \big(X^{[1]} \big(X^{[1]}\big)^* - z_0 \Phi^{[1]}\big)^{-1} \Phi_1 \le C \eta_0^{-1} \big\| \big(\Phi^{[1]}\big)^{-1/2} \Phi_1 \big\| \le C_1 \eta_0^{-1} \tag{5.24}$$

with some positive constants C and C_1, which trivially implies that

$$|D_3| \le \frac{C \varphi^{C_\zeta}}{\sqrt{N}} \eta_0^{-1} = O\big(N^{1/6 - C_2 \varepsilon}\big) \tag{5.25}$$

with ζ-high probability for some positive constant C_2. In addition, plugging (5.24) into (5.21) yields that

$$\Re D_2(z_0) = -\big(E_0^2 - \eta_0^2\big) \Re \Phi_1^* \big(X^{[1]} \big(X^{[1]}\big)^* - z_0 \Phi^{[1]}\big)^{-1} \Phi_1 + O(1) \tag{5.26}$$

with ζ-high probability. Now, we are ready to provide a bound for $\Re \Phi_1^* (X^{[1]} (X^{[1]})^* - z_0 \Phi^{[1]})^{-1} \Phi_1$ which is needed to estimate $\Im D_2$ according to (5.22). Combining (5.18), (5.25) and (5.26), we can see that

$$\big|\Re(D_1 + D_2 + D_3)\big| = \big|\big(E_0^2 - \eta_0^2\big) \Re \Phi_1^* \big(X^{[1]} \big(X^{[1]}\big)^* - z_0 \Phi^{[1]}\big)^{-1} \Phi_1\big| + O\big(N^{1/6 - C_2 \varepsilon}\big)$$

with ζ-high probability. Now, invoking assumption (5.14), we obtain

$$\big|\Re \Phi_1^* \big(X^{[1]} \big(X^{[1]}\big)^* - z_0 \Phi^{[1]}\big)^{-1} \Phi_1\big| = O\big(N^{1/6}\big) \tag{5.27}$$

with ζ-high probability. Inserting (5.27) into (5.22) we have

$$\Im D_2 = -\big(E_0^2 - \eta_0^2\big) \Im \Phi_1^* \big(X^{[1]} \big(X^{[1]}\big)^* - z_0 \Phi^{[1]}\big)^{-1} \Phi_1 + O\big(N^{-1/2 + C\varepsilon}\big). \tag{5.28}$$

For convenience, we set $t_0 = \Im \Phi_1^* (X^{[1]} (X^{[1]})^* - z_0 \Phi^{[1]})^{-1} \Phi_1$. Putting (5.23), (5.28) and (5.20) together, we get

$$\begin{aligned} &\Im(D_1 + D_2 + D_3) \\ &\quad = -E_0 \Im m_0(z_0) - \big(E_0^2 - \eta_0^2\big) t_0 + O\bigg(\frac{\varphi^{C_\zeta}}{\sqrt{N \eta_0}} t_0^{1/2}\bigg) + O\big(N^{-1/3 - C\varepsilon}\big) \end{aligned}$$

with ζ-high probability. Now observe that $E_0 \Im m_0(z_0)$ and $(E_0^2 - \eta_0^2) t_0$ are both positive. Moreover, by (ii) of Theorem 3.1 we see that

$$\Im m_0(z_0) \sim \sqrt{\kappa_0 + \eta_0}. \tag{5.29}$$

Now we split the discussion into two cases according to whether

$$t_0 \gg \frac{\varphi^{C_\zeta}}{\sqrt{N \eta_0}} t_0^{1/2}, \tag{5.30}$$

holds. If (5.30) is valid, then we deduce from (5.29) that (5.13) holds. If (5.30) fails, we claim that one must have

$$\frac{\varphi^{C_\zeta}}{\sqrt{N\eta_0}} t_0^{1/2} \ll \sqrt{\kappa_0+\eta_0}. \tag{5.31}$$

Since if (5.30) does not hold, there exists some positive constant C such that $t_0 \le C\varphi^{2C_\zeta}/N\eta_0$, which implies (5.31) immediately by our choice of η_0. Now (5.29) and (5.31) imply (5.13) again. Then by (5.12), we complete the proof. □

Acknowledgements. The authors are very grateful to Laszlo Erdös, Horng-Tzer Yau and Jun Yin for useful comments on a former version of this paper which helped to improve the organization significantly. We also thank the anonymous referees for careful reading and valuable comments, especially the suggestions on statistical signal processing.

SUPPLEMENTARY MATERIAL

Supplement: Proofs of some lemmas (DOI: 10.1214/14-AOS1281SUPP; .pdf). In the supplementary material [6], we will provide the proofs of Lemmas 4.2, 4.5, 4.6, 5.1, 5.4 and 5.5.

REFERENCES

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

[2] BAI, Z. and YAO, J.-F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053

[3] BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. MR1711663

[4] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575

[5] BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 1–32. MR2988403

[6] BAO, Z., PAN, G. and ZHOU, W. (2014). Supplement to "Universality for the largest eigenvalue of sample covariance matrices with general population." DOI:10.1214/14-AOS1281SUPP.

[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. and VIRÁG, B. (2011). Limits of spiked random matrices II. Preprint. Available at arXiv:1109.3704.

[10] BLOEMENDAL, A. and VIRÁG, B. (2013). Limits of spiked random matrices I. *Probab. Theory Related Fields* **156** 795–825. MR3078286

[11] EATON, M. L. (1989). *Group Invariance Applications in Statistics*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **1**. IMS, Hayward, CA. MR1089423

[12] 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

[13] EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248

[14] ERDŐS, L. (2011). Universality of Wigner random matrices: A survey of recent results. *Russ. Math. Surv.* **66** 507.

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

[16] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2009). Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37** 815–852. MR2537522

[17] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2009). Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.* **287** 641–655. MR2481753

[18] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2010). Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not. IMRN* **2010** 436–479. MR2587574

[19] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2011). Universality of random matrices and local relaxation flow. *Invent. Math.* **185** 75–119. MR2810797

[20] ERDŐS, L., SCHLEIN, B., YAU, H.-T. and YIN, J. (2012). The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1–46. MR2919197

[21] ERDŐS, L., YAU, H.-T. and YIN, J. (2011). Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.* **2** 15–82. MR2847916

[22] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. MR2981427

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

[24] FÉRAL, D. and PÉCHÉ, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. *J. Math. Phys.* **50** 073302. MR2548630

[25] FISHER, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugenics* **9** 238–249. MR0001499

[26] HSU, P. L. (1939). On the distribution of roots of certain determinantal equations. *Ann. Eugenics* **9** 250–258. MR0001500

[27] JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991

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

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

[30] KAY, S. M. (1998). *Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory*. Prentice Hall, Upper Saddle River, NJ.

[31] 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

[32] LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. MR1544569

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

[34] MO, M. Y. (2012). Rank 1 real Wishart spiked model. *Comm. Pure Appl. Math.* **65** 1528–1638. MR2969495

[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] ONATSKI, A. (2007). A formal statistical test for the number of factors in the approximate factor models. Unpublished manuscript.
[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. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865
[42] PAUL, D. and SILVERSTEIN, J. W. (2009). No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *J. Multivariate Anal.* **100** 37–57. MR2460475
[43] PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670
[44] PENNA, F. and GARELLO, R. (2009). Theoretical performance analysis of eigenvalue-based detection. Available at arXiv:0907.1523.
[45] PILLAI, N. S. and YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042
[46] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978
[47] ROY, S. N. (1939). p-Statistics and some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā* **4** 381–396.
[48] 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
[49] SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444
[50] TAO, T. and VU, V. (2011). Random matrices: Universality of local eigenvalue statistics. *Acta Math.* **206** 127–204. MR2784665
[51] TAO, T. and VU, V. (2012). Random covariance matrices: Universality of local statistics of eigenvalues. *Ann. Probab.* **40** 1285–1315. MR2962092
[52] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246
[53] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083
[54] VINOGRADOVA, J., COUILLET, R. and HACHEM, W. (2013). Statistical inference in large antenna arrays under unknown noise pattern. *IEEE Trans. Signal Process.* **61** 5633–5645. MR3130031
[55] WANG, D. (2012). The largest eigenvalue of real symmetric, Hermitian and Hermitian self-dual random matrix models with rank one external source, Part I. *J. Stat. Phys.* **146** 719–761. MR2916094
[56] WANG, K. (2012). Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices Theory Appl.* **1** 1150005, 24. MR2930383
[57] WANG, L. and PAUL, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.* **126** 25–52. MR3173080

[58] ZHANG, L. X. (2006). Spectral analysis of large dimensional random matrices. Ph.D. thesis, National University of Singapore.

Z. BAO
DEPARTMENT OF MATHEMATICS
ZHEJIANG UNIVERSITY
HANGZHOU, 310027
P. R. CHINA
E-MAIL: zhigangbao@zju.edu.cn

G. PAN
DIVISION OF MATHEMATICAL SCIENCES
SCHOOL OF PHYSICAL
AND MATHEMATICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE 637371
SINGAPORE
E-MAIL: gmpan@ntu.edu.sg

W. ZHOU
DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
SINGAPORE
E-MAIL: stazw@nus.edu.sg
URL: http://www.sta.nus.edu.sg/~stazw/