The Annals of Statistics
2016, Vol. 44, No. 4, 1564–1592
DOI: 10.1214/15-AOS1427

THE TRACY–WIDOM LAW FOR THE LARGEST EIGENVALUE OF F TYPE MATRICES

BY XIAO HAN, GUANGMING PAN[1] AND BO ZHANG

Nanyang Technological University

Let $\mathbf{A}_p = \frac{\mathbf{YY}^*}{m}$ and $\mathbf{B}_p = \frac{\mathbf{XX}^*}{n}$ be two independent random matrices where $\mathbf{X} = (X_{ij})_{p\times n}$ and $\mathbf{Y} = (Y_{ij})_{p\times m}$ respectively consist of real (or complex) independent random variables with $\mathbb{E}X_{ij} = \mathbb{E}Y_{ij} = 0$, $\mathbb{E}|X_{ij}|^2 = \mathbb{E}|Y_{ij}|^2 = 1$. Denote by λ_1 the largest root of the determinantal equation $\det(\lambda\mathbf{A}_p - \mathbf{B}_p) = 0$. We establish the Tracy–Widom type universality for λ_1 under some moment conditions on X_{ij} and Y_{ij} when p/m and p/n approach positive constants as $p \to \infty$.

1. Introduction. High-dimensional data now commonly arise in many scientific fields such as genomics, image processing, microarray, proteomics and finance, to name but a few. It is well known that the classical theory of multivariate statistical analysis for the fixed dimension p and large sample size n may lose its validity when handling high-dimensional data. A popular tool in analyzing large covariance matrices, and hence high-dimensional data is random matrix theory. The spectral analysis of high-dimensional sample covariance matrices has attracted considerable interests among statisticians, probabilitists and mathematicians since the seminal work of Marčenko and Pastur [21] about the limiting spectral distribution for a class of sample covariance matrices. One can refer to the monograph of Bai and Silverstein [1] for a comprehensive summary and references therein.

The largest eigenvalue of covariance matrices plays an important role in multivariate statistical analysis such as principle component analysis (PCA), multivariate analysis of variance (MANOVA) and discriminant analysis. One may refer to [22] for more details. In this paper, we focus on the largest eigenvalue of the F type matrices. Suppose that

$$\mathbf{A}_p = \frac{\mathbf{YY}^*}{m}, \qquad \mathbf{B}_p = \frac{\mathbf{XX}^*}{n} \tag{1.1}$$

are two independent random matrices where $\mathbf{X} = (X_{ij})_{p\times n}$ and $\mathbf{Y} = (Y_{ij})_{p\times m}$ respectively consist of real (or complex) independent random variables with $\mathbb{E}X_{ij} =$

Received June 2015; revised December 2015.

[1]Supported in part by a MOE Tier 2 Grant 2014-T2-2-060 and by a MOE Tier 1 Grant RG25/14 at the Nanyang Technological University, Singapore.

MSC2010 subject classifications. Primary 60B20, 34K25; secondary 60F05, 62H10.

Key words and phrases. Tracy–Widom distribution, largest eigenvalue, sample covariance matrix, F matrix.

$\mathbb{E}Y_{ij} = 0$ and $\mathbb{E}|X_{ij}|^2 = \mathbb{E}|Y_{ij}|^2 = 1$. Consider the determinantal equation

$$\det(\lambda \mathbf{A}_p - \mathbf{B}_p) = 0. \tag{1.2}$$

When $\mathbf{A}_p$ is invertible, the roots to (1.2) are the eigenvalues of a F matrix

$$\mathbf{A}_p^{-1}\mathbf{B}_p, \tag{1.3}$$

referred to as a Fisher matrix in the literature. The determinantal equation (1.2) is closely connected with the generalized eigenproblem

$$\det\bigl[\lambda(\mathbf{A}_p + \mathbf{B}_p) - \mathbf{B}_p\bigr] = 0. \tag{1.4}$$

We illustrate this in the next section. Many classical multivariate statistical tests are based on the roots of (1.2) or (1.4). For instance, one may use them to test the equality of two covariance matrices and the general linear hypothesis. In the framework of multivariate analysis of variance (MANOVA), $\mathbf{A}_p$ represents the within group covariance matrix while $\mathbf{B}_p$ means the between groups covariance matrix. A one-way MANOVA can be used to examine the hypothesis of equality of the mean vectors of interest.

Tracy and Widom in [29, 30] first discovered the limiting distributions of the largest eigenvalue for the large Gaussian Wigner ensemble, thus named as Tracy–Widom's law. Since their pioneer work study toward the largest eigenvalues of large random matrices becomes flourishing. To name a few, we mention [6, 10, 13, 14] and [26]. Among them we would mention El Karoui [6] which handled the largest eigenvalue of Wishart matrices for the nonnull population covariance matrix and provided a kind of condition on the population covariance matrix to ensure the Tracy–Widow law [see (4.16) below].

A follow-up to the above results is to establish the so-called universality property for generally distributed large random matrices. Specifically speaking, the universality property states that the limiting behavior of an eigenvalue statistic usually is not dependent on the distribution of the matrix entries. Indeed, the Tracy–Widom law has been established for the general sample covariance matrices under very general assumptions on the distributions of the entries of $\mathbf{X}$. The readers can refer to [3, 7, 9, 17, 19, 25, 27, 28, 33] for some representative developments on this topic. When proving universality an important tool is the Lindeberg comparison strategy (see Tao and Vu in [27] and Erdos, Yau and Yin [9]) and an important input when applying Lindeberg's comparison strategy is the strong local law developed by Erdos, Schlein and Yau in [8] and Erdos, Yau and Yin in [9].

Johnstone in [15] proved that the largest root of (1.1) converges to Tracy and Widom's distribution of type one after appropriate centering and scaling when the dimension p of the matrices $\mathbf{A}_p$ and $\mathbf{B}_p$ is even, $\lim_{p\to\infty} p/m < 1$ and $\mathbf{B}_p$ and $\mathbf{A}_p$ are both Wishart matrices. It is believed that the limiting distribution should not be affected by the dimension p. Indeed, numerical investigations both in [15] and [16] suggest that the Tracy and Widom approximation in the odd dimension case works as well as in the even dimension case. Besides, as it can be guessed, the Tracy and Widom approximation should not rely on the Gaussian assumption. However, theoretical support for these remains open. Furthermore, when $\mathbf{A}_p$ is not invertible the limiting distribution of the largest root to (1.1) is unknown yet even under the Gaussian assumption.

In this paper, we prove the universality of the largest root of (1.2) by imposing some moment conditions on $\mathbf{A}_p$ and $\mathbf{B}_p$. Specifically speaking, we prove that the largest root of (1.2) converges in distribution to the Tracy and Widom law for the general distributions of the entries of $\mathbf{X}$ and $\mathbf{Y}$ no matter what the dimension p is, even or odd. Moreover, the result holds when $\lim_{p\to\infty} p/m < 1$ or $\lim_{p\to\infty} p/m > 1$, corresponding to invertible $\mathbf{A}_p$ and noninvertible $\mathbf{A}_p$. This result also implies the asymptotic distribution of the largest root of (1.4).

At this point, it is also appropriate to mention some related work about the roots of (1.2). The limiting spectral distribution of the roots was derived by [32] and [1]. One may also find the limits of the largest root and the smallest root in [1]. Central limit theorem about linear spectral statistics was established in [38]. Very recently, the so-called spiked F model has been investigated by [5] and [34]. We would like to point out that they prove the local asymptotic normality or asymptotic normality for the largest eigenvalue of the spiked F model, which is completely different from our setting.

We conclude this section by outlining some ideas in the proof and presenting the structure of the rest of the paper. When $\mathbf{A}_p$ is invertible, the roots to (1.2) become those of the F matrix $\mathbf{A}_p^{-1}\mathbf{B}_p$ so that we may work on $\mathbf{A}_p^{-1}\mathbf{B}_p$. Roughly speaking, $\mathbf{A}_p^{-1}\mathbf{B}_p$ can be viewed as a kind of general sample covariance matrix $\mathbf{T}_n^{1/2}\mathbf{X}\mathbf{X}^*\mathbf{T}_n^{1/2}$ with $\mathbf{T}_n$ being a population covariance matrix by conditioning on $\mathbf{A}_p$. Denote the largest root of (1.2) by λ_1. The key idea is to break λ_1 into a sum of two parts as follows:

$$\lambda_1 - \mu_p = (\lambda_1 - \hat{\mu}_p) + (\hat{\mu}_p - \mu_p), \tag{1.5}$$

where $\hat{\mu}_p$ is an appropriate value when $\mathbf{A}_p$ is given and μ_p is an appropriate value when $\mathbf{A}_p$ is not given (their definitions are given in the later sections). However, we cannot condition on $\mathbf{A}_p$ directly. Instead we first construct an appropriate event so that we can handle the first term on the right-hand side of (1.5) on the event to apply the earlier results about $\mathbf{T}_n^{1/2}\mathbf{X}\mathbf{X}^*\mathbf{T}_n^{1/2}$. Particularly, we need to verify the condition (4.16) below. Once this is done, the next step is to prove that the second term on the right-hand side of (1.5) after scaling converges to zero in probability. This approach is different from that used in the literature in proving universality for the local eigenvalue statistics.

Unfortunately, when $\mathbf{A}_p$ is not invertible we cannot work on F matrices $\mathbf{A}^{-1}\mathbf{B}_p$ anymore. To overcome the difficulty, we instead start from the determinantal equation (1.2). It turns out that the largest root λ_1 can then be linked to the largest root of some F matrix when $\mathbf{X}$ consists of Gaussian random variables. Therefore, the result about F matrices $\mathbf{A}^{-1}\mathbf{B}_p$ is applicable. For general distributions, we find that it is equivalent to working on such a "covariance-type" matrix

$$\mathbf{D}^{-1/2}\mathbf{U}_1\mathbf{X}(\mathbf{I}-\mathbf{X}^*\mathbf{U}_2^*(\mathbf{U}_2\mathbf{X}\mathbf{X}^*\mathbf{U}_2^*)^{-1}\mathbf{U}_2\mathbf{X})\mathbf{X}^*\mathbf{U}_1^*\mathbf{D}^{-1/2}. \tag{1.6}$$

The definitions of $\mathbf{D}$ and $\mathbf{U}_j$, $j=1,2$ are given in the later section. This matrix is much more complicated than general sample covariance matrices. To deal with (1.6), we construct a 3×3 block linearization matrix

$$\mathbf{H}=\mathbf{H}(\mathbf{X})=\begin{pmatrix} -z\mathbf{I} & 0 & \mathbf{D}^{-1/2}\mathbf{U}_1\mathbf{X} \\ 0 & 0 & \mathbf{U}_2\mathbf{X} \\ \mathbf{X}^T\mathbf{U}_1^T\mathbf{D}^{-1/2} & \mathbf{X}^T\mathbf{U}_2^T & -\mathbf{I} \end{pmatrix}, \tag{1.7}$$

where $z=E+i\eta$ is a complex number with a positive imaginary part. It turns out that the upper left block of the 3×3 block matrix $\mathbf{H}^{-1}$ is the Stieltjes transform of (1.6) by simple calculations. We next develop the strong local law around the right end support μ_p by using a type of Lindeberg's comparison strategy raised in [17] and then use it to prove edge universality by adapting the approach used in [9] and [3].

The paper is organized as follows. Section 2 is to give the main results. Statistical applications and Tracy–Widom approximation are discussed in Section 3. Section 4 is devoted to proving the main result when $\mathbf{A}_p$ is invertible. Sections 5 and 6 prove the main result when $\mathbf{A}_p$ is not invertible. Some lemmas (theorems) and their proof are provided in the supplementary material [12] (Sections 7–12).

2. The main results. Throughout the paper, we make the following conditions.

CONDITION 1. *Assume that $\{Z_{ij}\}$ are independent random variables with $\mathbb{E}Z_{ij}=0$, $\mathbb{E}|Z_{ij}|^2=1$. For all $k\in N$, there is a constant C_k such that $\mathbb{E}|Z_{ij}|^k\le C_k$. In addition, if $\{Z_{ij}\}$ are complex, then $\mathbb{E}Z_{ij}^2=0$.*

We say that a random matrix $\mathbf{Z}=(Z_{ij})$ satisfies Condition 1 if its entries $\{Z_{ij}\}$ satisfy Condition 1.

CONDITION 2. *Assume that random matrices $\mathbf{X}=(\mathbf{X}_{ij})_{p,n}$ and $\mathbf{Y}=(\mathbf{Y}_{ij})_{p,m}$ are independent.*

CONDITION 3. *Set $m=m(p)$ and $n=n(p)$. Suppose that*

$$\lim_{p\to\infty}\frac{p}{m}=d_1>0,\qquad \lim_{p\to\infty}\frac{p}{n}=d_2>0,\qquad 0<\lim_{p\to\infty}\frac{p}{m+n}<1.$$

To present the main results uniformly, we define $\breve{m} = \max\{m, p\}$, $\breve{n} = \min\{n, m+n-p\}$ and $\breve{p} = \min\{m, p\}$. Moreover, let

$$\sin^2(\gamma/2) = \frac{\min\{\breve{p}, \breve{n}\} - 1/2}{\breve{m} + \breve{n} - 1}, \qquad \sin^2(\psi/2) = \frac{\max\{\breve{p}, \breve{n}\} - 1/2}{\breve{m} + \breve{n} - 1}, \tag{2.1}$$

$$\begin{aligned} \mu_{J,p} &= \tan^2\Big(\frac{\gamma+\psi}{2}\Big), \\ \sigma_{J,p}^3 &= \mu_{J,p}^3 \frac{16}{(\breve{m} + \breve{n} - 1)^2} \frac{1}{\sin(\gamma)\sin(\psi)\sin^2(\gamma+\psi)}. \end{aligned} \tag{2.2}$$

Formulas (2.2) can be found in [15] when $d_1 < 1$.

We below present alternative expressions of $\mu_{J,p}$ and $\sigma_{J,p}$. To this end, define a modified density of the Marčenko–Pastur law [21] (M–P law) by

$$\varrho_p(x) = \frac{1}{2\pi x(\breve{p}/\breve{m})}\sqrt{(b_p - x)(x - a_p)}\mathbf{I}(a_p \le x \le b_p), \tag{2.3}$$

where $a_p = (1 - \sqrt{\frac{\breve{p}}{\breve{m}}})^2$ and $b_p = (1 + \sqrt{\frac{\breve{p}}{\breve{m}}})^2$. Let $\gamma_1 \ge \gamma_2 \ge \cdots \ge \gamma_{\breve{p}}$ satisfy

$$\int_{\gamma_j}^{+\infty} \varrho_p(x)\,dx = \frac{j}{\breve{p}}, \tag{2.4}$$

with $\gamma_0 = b_p$ and $\gamma_p = a_p$. Moreover, suppose that $c_p \in [0, a_p)$ satisfies the equation

$$\int_{-\infty}^{+\infty} \Big(\frac{c_p}{x - c_p}\Big)^2 \varrho_p(x)\,dx = \frac{\breve{n}}{\breve{p}}. \tag{2.5}$$

One may easily check the existence and uniqueness of c_p. Define

$$\mu_p = \frac{1}{c_p}\Big(1 + \frac{\breve{p}}{\breve{n}} \int_{-\infty}^{+\infty} \Big(\frac{c_p}{x - c_p}\Big)\varrho_p(x)\,dx\Big) \tag{2.6}$$

and

$$\frac{1}{\sigma_p^3} = \frac{1}{c_p^3}\Big(1 + \frac{\breve{p}}{\breve{n}} \int_{-\infty}^{+\infty} \Big(\frac{c_p}{x - c_p}\Big)^3 \varrho_p(x)\,dx\Big). \tag{2.7}$$

It turns out that (2.2) and (2.6)–(2.7) are equivalent subject to some scaling, which is verified in Section 7 in the supplementary material [12].

We also need the following moment match condition.

DEFINITION 1 (Moment matching). Let $\mathbf{X}^1 = (x_{ij}^1)_{M\times N}$ and $\mathbf{X}^0 = (x_{ij}^0)_{M\times N}$ be two matrices satisfying Condition 1. We say that $\mathbf{X}^1$ matches $\mathbf{X}^0$ to order q, if for the integers i, j, l and k satisfying $1 \le i \le M$, $1 \le j \le N$, $0 \le l, k$ and $l + k \le q$, they have the relationship

$$\mathbb{E}\big[(\Im x_{ij}^1)^l (\Re x_{ij}^1)^k\big] = \mathbb{E}\big[(\Im x_{ij}^0)^l (\Re x_{ij}^0)^k\big] + O\big(\exp(-(\log p)^C)\big), \tag{2.8}$$

where C is some positive constant bigger than one, $\Re x$ is the real part and $\Im x$ is the imaginary part of x.

Throughout the paper, we use $\mathbf{X}^0$ to stand for the random matrix consisting of independent Gaussian random variables with mean zero and variance one.

Denote the type-i Tracy–Widom distribution by F_i, $i = 1, 2$ (see [30]). Set $\mathbf{B}_p = \frac{\mathbf{X}\mathbf{X}^*}{\breve{n}}$ and $\mathbf{A}_p = \frac{\mathbf{Y}\mathbf{Y}^*}{\breve{m}}$. We are now in a position to state the main results about F type matrices.

THEOREM 2.1. *Suppose that the real random matrices* $\mathbf{X}$ *and* $\mathbf{Y}$ *satisfy Conditions* 1–3. *Moreover, suppose that* $0 < d_2 < \infty$. *Denote the largest root of* $\det(\lambda\mathbf{A}_p - \mathbf{B}_p) = 0$ *by* λ_1.

(i) *If* $0 < d_1 < 1$, *then*

$$\lim_{p\to\infty} P\left(\frac{(\breve{n}/\breve{m})\lambda_1 - \mu_{J,p}}{\sigma_{J,p}} \le s\right) = F_1(s). \tag{2.9}$$

(ii) *If* $d_1 > 1$ *and* $\mathbf{X}$ *matches the standard* $\mathbf{X}^0$ *to order* 3, *then* (2.9) *still holds*.

REMARK 1. When $\mathbf{X}$ and $\mathbf{Y}$ are complex random matrices, Theorem 2.1 still holds but the Tracy–Widom distribution $F_1(s)$ should be replaced by $F_2(s)$.

If $0 < d_1 < 1$, then $\mathbf{A}_p$ is invertible. In this case the largest eigenvalue λ_1 is that of F matrices $\mathbf{A}_p^{-1}\mathbf{B}_p$. If $d_1 > 1$, then $\mathbf{A}_p$ is not invertible.

REMARK 2. Theorem 2.1 immediately implies the distribution of the largest root of $\det(\lambda(\mathbf{B}_p + \mathbf{A}_p) - \mathbf{B}_p) = 0$. In fact, the largest root of $\det(\lambda(\mathbf{B}_p + \mathbf{A}_p) - \mathbf{B}_p) = 0$ is $\frac{\lambda_1}{1+\lambda_1}$ if λ_1 is the largest root of the F matrices $\mathbf{B}_p\mathbf{A}_p^{-1}$ in Theorem 2.1 when $0 < d_1 < 1$.

When $d_1 > 1$ the largest root of $\det(\lambda(\mathbf{B}_p + \mathbf{A}_p) - \mathbf{B}_p) = 0$ is one with multiplicity $(p - m)$. We instead consider the $(p - m + 1)$th largest root of $\det(\lambda(\mathbf{B}_p + \mathbf{A}_p) - \mathbf{B}_p) = 0$. It turns out that the $(p - m + 1)$th largest root of $\det(\lambda(\mathbf{B}_p + \mathbf{A}_p) - \mathbf{B}_p) = 0$ is $\frac{\lambda_1}{1+\lambda_1}$ if λ_1 is the largest root of $\det(\lambda\mathbf{A}_p - \mathbf{B}_p) = 0$.

Moreover, note the equality

$$(\mathbf{B}_p + \mathbf{A}_p)^{-1}\mathbf{B}_p + (\mathbf{B}_p + \mathbf{A}_p)^{-1}\mathbf{A}_p = I.$$

If $\mathbf{Y}$ matches $\mathbf{X}^0$ to order 3, then the smallest positive root of $\det(\lambda(\mathbf{B}_p + \mathbf{A}_p) - \mathbf{B}_p) = 0$ also tends to type-1 Tracy–Widom distribution after appropriate centralizing and rescaling by Theorem 2.1 when $d_1 > 1$ and $d_2 > 1$.

We would like to point out that Johnstone [15] proved part (i) of Theorem (2.1) when p is even, $\mathbf{A}_p$ and $\mathbf{B}_p$ are both Wishart matrices. Part (ii) of Theorem (2.1) is new even if $\mathbf{A}_p$ and $\mathbf{B}_p$ are both Wishart matrices. When proving Theorem 2.1, we have indeed obtained different asymptotic mean and variance. Precisely we have proved that

$$\lim_{p\to\infty} P\big(\sigma_p \breve{n}^{2/3}(\lambda_1 - \mu_p) \le s\big) = F_1(s) \tag{2.10}$$

and that

$$\left|\frac{\breve{m}}{\breve{n}}\mu_{J,p} - \mu_p\right| = O(p^{-1}), \qquad \lim_{p\to\infty} \sigma_p \frac{\breve{m}}{\breve{n}^{1/3}}\sigma_{J,p} = 1. \tag{2.11}$$

Equations (2.10) and (2.11) imply Theorem 2.1. The proof of (2.11) is provided in the supplementary material [12] and we prove (2.10) in the main paper.

3. Applications and simulations.

This section is to explore some applications of our universality results in high-dimensional statistical inference and conduct simulations to check the quality of the approximations of our limiting law.

3.1. *Long-side spherical test (LSST) for separable covariance matrices*.

Consider a data matrix $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)_{p\times N}$ as follows:

$$\mathbf{Y} = \mathbf{\Sigma}^{1/2}\mathbf{X}\mathbf{T}^{1/2}, \tag{3.1}$$

where $\mathbf{X}$ is a $p \times N$ matrix satisfying Condition 1, $\mathbf{X}$ matches $\mathbf{X}^0$ to order 3 (if $p > 2N$ or $2p < N$, this condition is not necessary), and $\mathbf{\Sigma}^{1/2}$ and $\mathbf{T}^{1/2}$ are positive definite matrices. We start from a special case $\mathbf{T} = \mathbf{I}$ and the model then becomes

$$\mathbf{Y} = \mathbf{\Sigma}^{1/2}\mathbf{X}.$$

For such a simplified model, the spherical test: $H_0 : \mathbf{\Sigma} = \sigma^2\mathbf{I}$ vs. $H_1 : \mathbf{\Sigma} \ne \sigma^2\mathbf{I}$ has been widely discussed in literature. When p is comparable to N, there are considerable work on it. To name a few, we mention [18] and Section 9.5 of [35]. We can extend this test to the more general model (3.1). In model (3.1), $\mathbf{Y}\mathbf{Y}^*$ is called the separable covariance matrix which is to depict the spatial temporal data. For high dimensional data, the spectral properties of $\mathbf{Y}\mathbf{Y}^*$ are studied in some recent papers such as [24] and [37]. In this section, we focus on testing whether $\mathbf{\Sigma}$ or $\mathbf{T}$ is proportional to $\mathbf{I}$. To be precise, we can conduct the following hypothesis testing problems:

$$\text{If } \lim_{p\to\infty}\frac{p}{N} < 1, \text{ we test } H_0 : \mathbf{T} = \sigma_1^2\mathbf{I} \text{ vs. } H_1 : \mathbf{T} \ne \sigma_1^2\mathbf{I}$$

or

$$\text{If } \lim_{p\to\infty}\frac{p}{N} > 1, \text{ we test } H_0 : \mathbf{\Sigma} = \sigma_2^2\mathbf{I} \text{ vs. } H_1 : \mathbf{\Sigma} \ne \sigma_2^2\mathbf{I}.$$

In the sequel, we focus on the first testing problem, that is, $\lim_{p\to\infty}\frac{p}{N} < 1$ and the second can be discussed similarly. We choose an index subset $\mathbf{S} \in \{1, \dots, N\}$ such that the cardinality of $\mathbf{S}$ is $\lfloor \frac{N}{2} \rfloor$ (we also suggest an approach for selection of $\mathbf{S}$ in stationary time series models in simulation). Moreover, we define $\mathbf{Z}_2\mathbf{Z}_2^* = \sum_{i\in\mathbf{S}} \mathbf{y}_i\mathbf{y}_i^*$, $\mathbf{Z}_1\mathbf{Z}_1^* = \sum_{i\notin\mathbf{S}} \mathbf{y}_i\mathbf{y}_i^*$, $n = \lfloor \frac{N}{2} \rfloor$ and $m = N - n$. We use the largest root λ_1 of $\det(\lambda\frac{\mathbf{Z}_1\mathbf{Z}_1^*}{\breve{m}} - \frac{\mathbf{Z}_2\mathbf{Z}_2^*}{\breve{n}}) = 0$ as a test statistic. Under the null hypothesis, λ_1 tends to Tracy–Widom's distribution after centralizing and rescalling for any selection $\mathbf{S}$. The key observation is that $\mathbf{\Sigma}$ can be eliminated in $\det(\lambda\frac{\mathbf{Z}_1\mathbf{Z}_1^*}{\breve{m}} - \frac{\mathbf{Z}_2\mathbf{Z}_2^*}{\breve{n}}) = 0$. But under the alternative hypothesis, for a suitably chosen $\mathbf{S}$, the correlation structure involved in $\mathbf{Z}_2$ can be much different from that of $\mathbf{Z}_1$, which implies the largest root of the above determinant deviates much from μ_p. This observation ensures that $n^{2/3}(\lambda_1 - \mu_p)$ will be very large when the null hypothesis does not hold. One can see that the restriction of $\lim_{p\to\infty} \frac{p}{N} < 1$ comes from the conditions of Theorem 2.1.

In addition, we would point out that the extreme eigenvalue of sample covariance matrices is not a proper statistic for such a hypothesis test when there are no spiked eigenvalues, while our statistic is not dependent on the fact that whether $\mathbf{\Sigma}$ is spiked or not. We demonstrate the reasons below. First, one cannot directly use the largest eigenvalue of $\mathbf{YY}^*$ since $\mathbf{\Sigma}$ is unknown. Therefore, we have to apply the statistic [One-side identity test (OSI)] in Section 2.1 of [3], that is, $\frac{\lambda_1(\mathbf{YY}^*)-\lambda_2(\mathbf{YY}^*)}{\lambda_2(\mathbf{YY}^*)-\lambda_3(\mathbf{YY}^*)}$. When $\mathbf{T}$ is a spiked matrix, the statistic works well (see Table 4 in [3]). However, when $\mathbf{T}$ is not spiked and $\mathbf{\Sigma} = \mathbf{I}$ the statistic $\frac{\lambda_1(\mathbf{YY}^*)-\lambda_2(\mathbf{YY}^*)}{\lambda_2(\mathbf{YY}^*)-\lambda_3(\mathbf{YY}^*)}$ always tends to the same distribution for any $\mathbf{T}$ satisfying (1.4) of [3] (including $\mathbf{T} = \sigma^2\mathbf{I}$), which means the statistic does not work in this case. Table 4 below confirms this phenomenon.

3.2. *Equality of K covariance matrices* (*EOM*). Consider the model of the following form:

$$\mathbf{Z}_i = \mathbf{\Sigma}_i^{1/2}\mathbf{X}_i, \qquad i = 1, \dots, K,$$

where $\{\mathbf{X}_i\}$ are $p \times n_i$ random matrices and $\{\mathbf{\Sigma}_i\}$ are $p \times p$ invertible population covariance matrices and K is a positive integer. Moreover, we assume that there exists a k_0 such that $\lim_{p\to\infty} \frac{\sum_{i=1}^{k_0} n_i}{p} \in (0, \infty)$ and $\lim_{p\to\infty} \frac{\sum_{i=k_0+1}^{K} n_i}{p} \in (0, \infty)$. For simplicity and being consistent with the previous notation, we set $n = \sum_{i=1}^{k_0} n_i$, $m = \sum_{i=k_0+1}^{K} n_i$, $\mathbf{Y} = (\mathbf{X}_{k_0+1}, \dots, \mathbf{X}_K)$ and $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{k_0})$. We also assume that $\mathbf{X}$ and $\mathbf{Y}$ satisfy the conditions of Theorem 2.1.

We are interested in the following hypothesis test:

$$\mathbf{H}_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \Sigma_K \text{ vs. } \mathbf{H}_1 : \exists 1 \le i < j \le K \text{ such that } \mathbf{\Sigma}_i \ne \mathbf{\Sigma}_j.$$

Under the null hypothesis, we have

$$\det\left(\lambda\frac{\sum_{i=k_0+1}^{K} \mathbf{Z}_k\mathbf{Z}_k^*}{\breve{m}} - \frac{\sum_{i=1}^{k_0} \mathbf{Z}_k\mathbf{Z}_k^*}{\breve{n}}\right) = 0 \quad \Longleftrightarrow \quad \det\left(\lambda\frac{\mathbf{YY}^*}{\breve{m}} - \frac{\mathbf{XX}^*}{\breve{n}}\right) = 0.$$

In view of this, we may propose the largest root λ_1 of $\det(\lambda\frac{\sum_{i=k_0+1}^{K}\mathbf{Z}_k\mathbf{Z}_k^*}{\breve{m}} - \frac{\sum_{i=1}^{k_0}\mathbf{Z}_k\mathbf{Z}_k^*}{\breve{n}}) = 0$ as a test statistic. By Theorem 2.1, we see that λ_1 tends to Tracy–Widom's distribution after centralizing and rescaling.

We below investigate its power for a kind of sparse alternative hypothesis when $K = 2$. Specifically, we consider the alternative case

$$\mathbf{Z}_1 = \mathbf{\Sigma}^{1/2}\mathbf{X}, \qquad \mathbf{Z}_2 = \mathbf{Y}. \tag{3.2}$$

If $\mathbf{YY}^*$ is invertible, we choose $\mathbf{\Sigma} = \mathbf{I} + \tau\frac{p/m+r}{1-p/m}\mathbf{e}_1\mathbf{e}_1^T$ and $r = \sqrt{\frac{p}{m} + \frac{p}{n} - \frac{p^2}{mn}}$. The reason why we choose the factor $\frac{p/m+r}{1-p/m}$ is that it is a spiked F matrix when $\tau > 1$. The largest eigenvalue λ_1 converges to normal distribution weakly by Proposition 11 of [5] and Theorem 4.1 of [34]. In fact, by Proposition 5 of [5] and Theorem 3.1 of [34] we immediately have the following proposition.

PROPOSITION 1. *For the model* (3.2), *suppose* $\mathbf{YY}^*$ *is invertible and* $\mathbf{\Sigma} = \mathbf{I} + \tau\frac{p/m+r}{1-p/m}\mathbf{e}_1\mathbf{e}_1^T$. *Let* $\phi(x) = \frac{x(x-1+p/n)}{x(1-p/m)-1}$. *When* $\tau > 1$, *the largest eigenvalue of the spiked F matrix* $\frac{m}{n}(\mathbf{YY}^*)^{-1}\Sigma^{1/2}\mathbf{XX}^*\Sigma^{1/2}$ (*denoted by* λ_1) *almost surely converges to* $\frac{m}{n}\phi(1 + \tau\frac{p/m+r}{1-p/m})$, *and for any positive constant* C, *we have*

$$\lim_{p\to\infty} P\left(\frac{(n/m)\lambda_1 - \mu_{J,p}}{\sigma_{J,p}} > C\right) = 1 \tag{3.3}$$

(*the power of the test goes to* 1 *as* $p \to \infty$).

An interesting feature of this approach is that K can tend to infinity. Below we compare our statistic with Corrected Likelihood Ratio Test (CLRT) proposed in Chapter 9 of [35] when $K = 2$. First, we do not assume $\mathbb{E}|\mathbf{X}_{ij}|^4$ to be equal to some known constant β for all i and j unlike [35]. Moreover, the 4th moment assumption restricts the extension of their approach to the equality test of K matrices since it is not reasonable to make such an assumption when K is large. The advantage of CLRT is that it includes all information of the F matrix's spectrum such that their test is more powerful when the population eigenvalues are close to each other. But when $\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_2$ is a spiked matrix the largest eigenvalue of F type matrices works better. See Table 8 below.

3.3. *Correlated noise detection.* Let

$$\mathbf{y}_t = f(\mathbf{x}_t) + \Sigma^{1/2}\varepsilon_t, \qquad t = 1, 2, \ldots, T \tag{3.4}$$

be the signal received at time t where $\mathbf{y}_t$ is a p-dimensional real or complex vector and ε_t is a p-dimensional white noise vector (i.i.d.) satisfying Condition 1. Moreover, if $2p < T$, we assume that the third moments of the entries of ε_t are 0, Σ is an unknown $p \times p$ invertible matrix, $\lim_{p\to\infty}\frac{p}{T} \in (0, 1)$, $\mathbf{x}$ is a vector or matrix with arbitrary dimension (maybe correlated to $\mathbf{y}$) and that $f(\mathbf{x})$ is a given function such as regression model $f(\mathbf{x}) = \mathbf{x}^*\beta$. We are interested in whether there is "real signal" contained in $\mathbf{y}_t$. Our hypothesis testing problem is

$$H_0 : \mathbf{y}_t = \Sigma^{1/2}\varepsilon_t \quad \text{vs.} \quad H_1 : \mathbf{y} \neq \Sigma^{1/2}\varepsilon_t. \tag{3.5}$$

Let $\mathbf{X} = (\mathbf{y}_1, \dots, \mathbf{y}_{\lfloor T/2 \rfloor})$, $\mathbf{Y} = (\mathbf{y}_{\lfloor T/2 \rfloor+1}, \dots, \mathbf{y}_T)$, $n = \lfloor \frac{T}{2} \rfloor$ and $m = T - n$. As before use the largest root of $\det(\lambda \frac{\mathbf{YY}^*}{\breve{m}} - \frac{\mathbf{XX}^*}{\breve{n}}) = 0$ as a test statistic which converges to Tracy–Widom's distribution after centralizing and rescaling by Theorem 2.1.

In engineering, we do not need to split the sample into $\mathbf{X}$ and $\mathbf{Y}$. Specifically, in signal detection or cognitive radio, model (3.4) takes the form of

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \Sigma^{1/2}\varepsilon_t, \qquad t = 1, 2, \dots, T, \tag{3.6}$$

where $\mathbf{x}_t$ is a k-dimensional signal vector with covariance matrix $\mathbf{S}$, $\mathbf{A}$ is a $p \times k$ deterministic matrix and the other assumptions are the same as those in the previous model (3.4). We are also interested in the test (3.5). This is a widely discussed problem in cognitive radio. For the high dimensional setting, once may see [36] and [23]. We also refer to the recent paper [31] assuming the correlated noise. In engineering, there exists some method to get another signal-free sample, say $\mathbf{r}_t = \Sigma^{1/2}\mathbf{z}_t$, $t = 1, \dots, T_1$ where $\{\varepsilon_t\}_{t=1}^{T}$ and $\{\mathbf{z}_t\}_{t=1}^{T_1}$ are i.i.d. One can refer to [20, 34] and [23] for detailed discussions. Let $\mathbf{R}_1 = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ and $\mathbf{R}_2 = (\mathbf{r}_1, \dots, \mathbf{r}_{T_1})$. We use the largest root of $\det(\lambda \frac{\mathbf{R}_2\mathbf{R}_2^*}{T_1} - \frac{\mathbf{R}_1\mathbf{R}_1^*}{T}) = 0$ as a statistic. The power is sated in Table 9 below.

3.4. *Other applications under the Gaussian distribution.* There are many other applications which can be connected with the largest eigenvalue of the F matrices due to nice properties of the Gaussian distribution. We illustrate a multivariate ANOVA test below. One can refer to [15] for more applications. We consider the multivariate regression model

$$\mathbf{Y} = \mathbf{XB} + \mathbf{\Sigma Z},$$

where $\mathbf{Y}$ is an $N \times p$ response matrix, $\mathbf{X}$ is a known $N \times q$ design matrix, $\mathbf{B}$ is a $q \times p$ unknown regression matrix, $\mathbf{Z}$ is a $N \times p$ random matrix with i.i.d. Gaussian entries and Σ is an invertible deterministic matrix. We are interested in the following hypothesis test: given $g \times q$ matrix $\mathbf{C}$:

$$H_0 : \mathbf{CB} = 0 \quad \text{vs.} \quad H_1 : \mathbf{CB} \neq 0.$$

To explain the motivation behind the test, we consider a low dimensional example. If $q = 3$, $p = 1$, $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)^T$, $\mathbf{y}_i \sim N(b_j, \sigma^2)$, $N_{j-1} \leq i \leq N_j$, $1 = N_1 \leq \cdots \leq N_4 = N$ and $\mathbf{C} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix}$, the null hypothesis $\mathbf{CB} = 0$ is equivalent to $b_1 = b_2 = b_3$, that is, ANOVA test. Of course, we consider high dimensional setting in this paper. We assume $N > q$. The least square estimator is $\hat{\mathbf{B}} = (\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*\mathbf{Y}$. Under the null hypothesis, it is easy to see that the matrices

$$\mathbf{D} = \mathbf{Y}^*(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{Y} \sim W_p(\boldsymbol{\Sigma}, N - q),$$
$$\mathbf{E} = (\mathbf{C}\hat{\mathbf{B}})^*\big[\mathbf{C}\big(\mathbf{X}^*\mathbf{X}\big)^{-1}\mathbf{C}^*\big]^{-1}(\mathbf{C}\hat{\mathbf{B}}) \sim W_p(\boldsymbol{\Sigma}, g)$$

are independent where $\mathbf{P}_\mathbf{X}$ is the projection matrix generated by $\mathbf{X}$. Then the largest root of $\det(\lambda \frac{\mathbf{D}}{N-q} - \frac{\mathbf{E}}{g}) = 0$ can be used as a statistic for the test. One can further refer to pages 210–213 of [11] for constructing a confidence interval for the linear combination of the entries of $\mathbf{B}$.

3.5. *Simulations*. We conduct some numerical simulations to check the accuracy of the distributional approximations in Theorem 2.1 under various settings of (p, m, n) and the distribution of $\mathbf{X}$. We also study the power of the hypothesis tests in Sections 3.1–3.2.

As in [15], we below use $\ln(\lambda_1)$ to run simulations. To do so, we first give its distribution. By [15] and (2.10), we can find that

$$\lambda_1 = \mu_p + \frac{Z}{\sigma_p \breve{n}^{2/3}} + o_p\big(\breve{n}^{-2/3}\big), \tag{3.7}$$

where $Z = F_1^{-1}(U)$ and U is a $U(0, 1)$ random variable. By Taylor's expansion, we then have

$$\ln(\lambda_1) = \ln(\mu_p) + \frac{Z}{\mu_p\sigma_p \breve{n}^{2/3}} + o_p\big(\breve{n}^{-2/3}\big). \tag{3.8}$$

Recall $|\frac{\breve{m}}{\breve{n}}\mu_{J,p} - \mu_p| = O(p^{-1})$ and $\lim_{p\to\infty} \sigma_p \frac{\breve{m}}{\breve{n}^{1/3}}\sigma_{J,p} = 1$ in Section 2. Summarizing the above, we can find

$$\lim_{p\to\infty} P\big(\sigma_{pln}\big(\ln(\lambda_1) - \mu_{pln}\big) \le s\big) = F_1(s), \tag{3.9}$$

where

$$\mu_{pln} = \ln\bigg(\frac{\breve{m}}{\breve{n}}\mu_{J,p}\bigg), \qquad \sigma_{pln} = \frac{\mu_{J,p}}{\sigma_{J,p}}. \tag{3.10}$$

3.5.1. *Accuracy of approximations for TW laws and size*. We conduct some numerical simulations to check the accuracy of the distributional approximations in Theorem 2.1, which include the size of the tests as well.

Table 1 is done by the software R. We set two initial triples (p, m, n) of $M_0 = (5, 40, 10)$ and $M_1 = (30, 20, 25)$ and then consider $2M_i$, $3M_i$ and $4M_i$, $i = 0, 1$. The triples M_0 and M_1 correspond to invertible $\mathbf{YY}^*$ and noninvertible $\mathbf{YY}^*$ respectively. For each case, we generate 10,000 $(\mathbf{X}, \mathbf{Y})$ whose entries follow standard normal distribution. We calculate the largest root of $\det(\lambda\frac{\mathbf{YY}^*}{\breve{m}} - \frac{\mathbf{XX}^*}{\breve{n}}) = 0$ to get $\ln(\lambda_1)$ and renormalize it with μ_{pln} and σ_{pln}. In the "Percentile column",

TABLE 1
Standard quantiles for several triples (p, m, n): Gaussian case

		Initial triple $M_0 = (5, 40, 10)$				Initial triple $M_1 = (30, 20, 25)$				
Percentile	**TW**	M_0	$2M_0$	$3M_0$	$4M_0$	M_1	$2M_1$	$3M_1$	$4M_1$	**2*SE**
−3.9	0.01	0.0208	0.0133	0.0124	0.0115	0.0017	0.0035	0.0048	0.0060	0.002
−3.18	0.05	0.0680	0.0601	0.0562	0.0582	0.0210	0.0276	0.0327	0.0370	0.004
−2.78	0.1	0.1176	0.1120	0.1088	0.1095	0.0608	0.0712	0.0808	0.0842	0.006
−1.91	0.3	0.3154	0.3030	0.3080	0.3084	0.2641	0.2744	0.2864	0.2909	0.009
−1.27	0.5	0.5139	0.5070	0.5051	0.5082	0.4839	0.4904	0.4960	0.4964	0.01
−0.59	0.7	0.7073	0.7154	0.7012	0.7111	0.7055	0.7031	0.7019	0.7005	0.009
0.45	0.9	0.9083	0.9058	0.9047	0.9090	0.9040	0.9010	0.9016	0.9003	0.006
0.98	0.95	0.9561	0.9544	0.9517	0.9557	0.9489	0.9530	0.9504	0.9498	0.004
2.02	0.99	0.9919	0.9909	0.9913	0.9919	0.9878	0.9887	0.9897	0.9901	0.002

the quantiles of TW_1 law corresponding to the "TW" column are listed. We state the values of the empirical distributions of the renormalized λ_1 for various triples at the corresponding quantiles in columns 3–10 and the standard errors based on binomial sampling are listed in the last column. QQ-plots corresponding to the triples (20, 160, 40) and (120, 80, 100) are also stated in Figure 1.

Tables 2, 3 and Figures 2 and 3 are the same as Table 1 and the corresponding Figure 1 except that that we replace the Gaussian distribution by the some discrete distribution and uniform distribution.

When considering the tests in Sections 3.1–3.3, one may refer to Tables 1–3 as well for their sizes at the nominal significant levels.

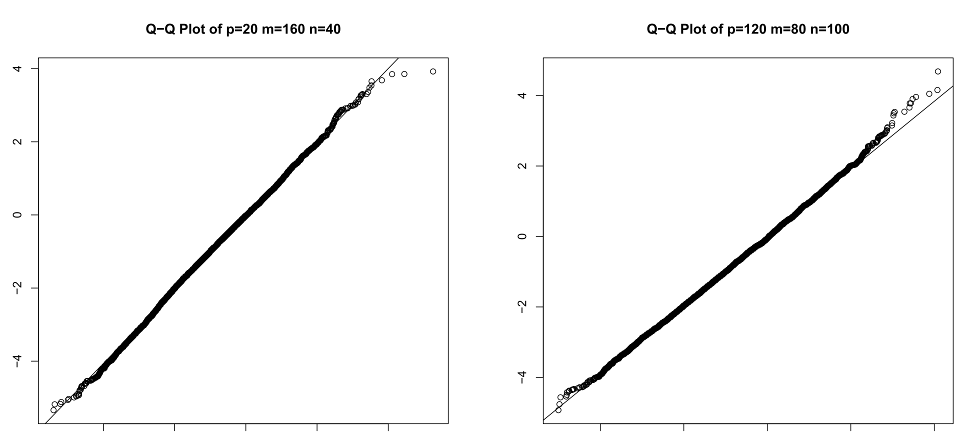

FIG. 1. *QQ plots of the triples* (20, 160, 40) *and* (120, 80, 100) *corresponding to Table* 1.

TABLE 2
Standard quantiles for several triples (p, m, n): *Discrete distribution with the probability mass function* $P(\mathbf{x} = \sqrt{3}) = P(\mathbf{x} = -\sqrt{3}) = 1/6$ *and* $P(\mathbf{x} = 0) = 2/3$

		Initial triple $M_0 = (5, 40, 10)$				Initial triple $M_1 = (30, 20, 25)$				
Percentile	**TW**	M_0	$2M_0$	$3M_0$	$4M_0$	M_1	$2M_1$	$3M_1$	$4M_1$	**2*SE**
−3.9	0.01	0.0192	0.0132	0.0136	0.0123	0.0006	0.0031	0.0046	0.0047	0.002
−3.18	0.05	0.0637	0.0581	0.0571	0.0573	0.0216	0.0302	0.0321	0.0356	0.004
−2.78	0.1	0.1147	0.1101	0.1099	0.1088	0.0626	0.0733	0.0757	0.0824	0.006
−1.91	0.3	0.3100	0.2966	0.3060	0.3029	0.2665	0.2721	0.2808	0.2827	0.009
−1.27	0.5	0.5000	0.4959	0.4969	0.4996	0.4841	0.4834	0.4985	0.4899	0.01
−0.59	0.7	0.7025	0.7013	0.7099	0.7018	0.6990	0.6992	0.7109	0.6975	0.009
0.45	0.9	0.9107	0.9061	0.9071	0.9036	0.9014	0.9040	0.9059	0.9001	0.006
0.98	0.95	0.9566	0.9546	0.9538	0.9546	0.9503	0.9527	0.9526	0.9512	0.004
2.02	0.99	0.9929	0.994	0.9903	0.9914	0.9890	0.9908	0.9901	0.9894	0.002

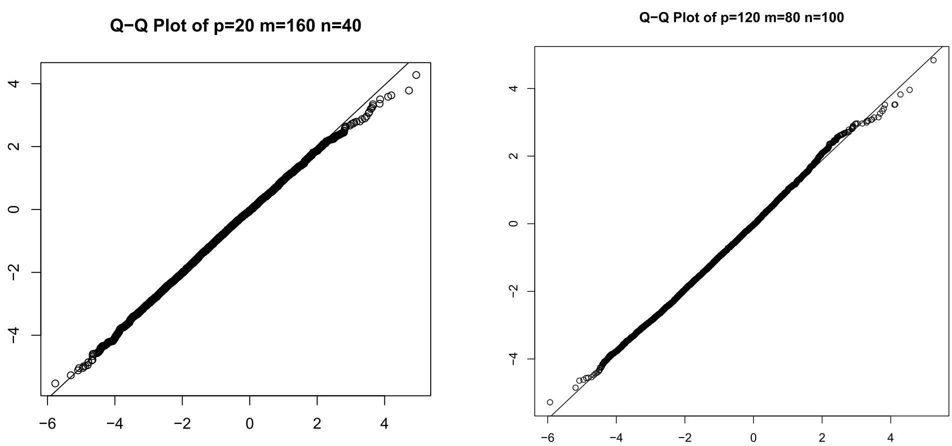

FIG. 2. *QQ plots of the triples* (20, 160, 40) *and* (120, 80, 100) *corresponding to Table* 2.

TABLE 3
Standard quantiles for several triples (p, m, n): *Continuous uniform distribution* $U(-\sqrt{3}, \sqrt{3})$

		Initial triple $M_0 = (5, 40, 10)$				Initial triple $M_1 = (30, 20, 25)$				
Percentile	**TW**	M_0	$2M_0$	$3M_0$	$4M_0$	M_1	$2M_1$	$3M_1$	$4M_1$	**2*SE**
−3.9	0.01	0.0098	0.0117	0.0122	0.0120	0.0101	0.0087	0.0092	0.0096	0.002
−3.18	0.05	0.0612	0.0632	0.0606	0.0592	0.0514	0.0462	0.0492	0.0482	0.004
−2.78	0.1	0.1205	0.1243	0.1208	0.1197	0.1023	0.0942	0.1033	0.0992	0.006
−1.91	0.3	0.3644	0.3542	0.351	0.3432	0.3132	0.2946	0.3101	0.3017	0.009
−1.27	0.5	0.5767	0.5575	0.5563	0.5496	0.516	0.5073	0.5151	0.5069	0.01
−0.59	0.7	0.7728	0.7540	0.7443	0.7440	0.7182	0.7123	0.714	0.7171	0.009
0.45	0.9	0.9397	0.9243	0.9181	0.9202	0.9141	0.9068	0.9071	0.9059	0.006
0.98	0.95	0.9722	0.9672	0.9599	0.9614	0.9584	0.9538	0.9556	0.9534	0.004
2.02	0.99	0.9959	0.9941	0.993	0.9922	0.9932	0.9912	0.9919	0.9916	0.002

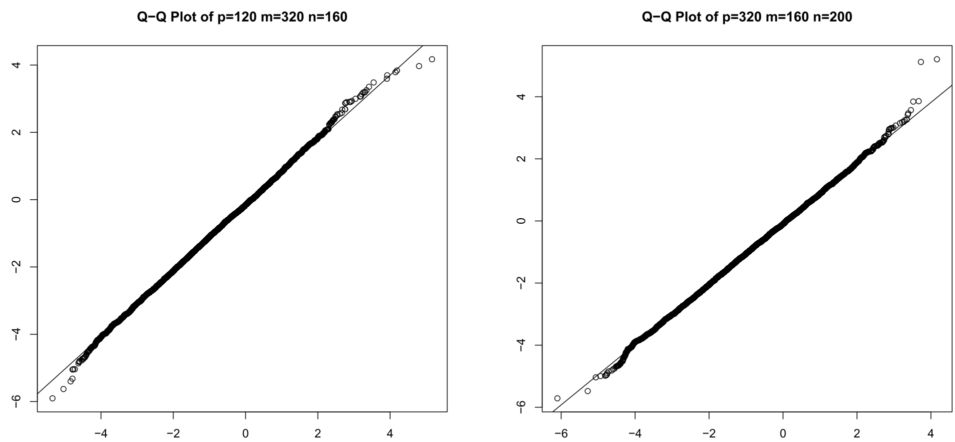

FIG. 3. *QQ plots of the triples* (120, 320, 160) *and* (320, 140, 200) *corresponding to Table* 3.

3.5.2. *Power study of "Long-Side Spherical Test* (*LSST*) *for separable covariance matrices"* (*see Section* 3.1). We consider the alternative model as follows:

$$\mathbf{y}_1 = \mathbf{z}_1, \qquad \mathbf{y}_t = a\mathbf{y}_{t-1} + \sqrt{1-a^2}\mathbf{z}_t, \qquad t = 2, \ldots, N,$$

where $\{\mathbf{y}_t\}_{t=1}^N$ are p-dimensional vectors, $\{\mathbf{z}_t\}_{t=1}^N$ are independent noise vectors satisfying Condition 1 and $a \in (0, 1)$. It is easy to see that $\{\mathbf{y}_t\}_{t=1}^N$ are a stationary sequence, and hence the matrix $\mathbf{T}$ [see (3.1)] is a Toeplitz matrix. We then suggest to choose the set $\mathbf{S} = \{1, \ldots, \lfloor \frac{N}{4} \rfloor, \lfloor \frac{3N}{4} \rfloor, \ldots, N\}$ so that the correlation structures involved in $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are different. In this subsection, we also compare our approach (denoted by LSST) with that raised by BAO, etc. in [3] (denoted by OSI).

The power of the tests is listed in Table 4 below and the nominal significant level of the tests is 5%.

From the table, we can see that the OSI approach does not gain power under the alternative hypothesis, which is consistent with our analysis before. The power of LSST is larger when either dimension or a becomes larger. When a is small, say 0.1, the power is very poor. This phenomenon is easy to understand since $a^3 = 0.001 \approx 0$, that is, $\mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_{t+3}) = 0.001 \approx 0$, the data $\{\mathbf{y}_t\}_{t=1}^T$ look like independent.

3.5.3. *Power of "Equality of K covariance matrices* (*EOM*)*" for* $k = 2$. We study the power of the test and consider the alternative case (3.2).

When $\mathbf{YY}^*$ is invertible, we choose $\mathbf{\Sigma} = \mathbf{I} + \tau \frac{p/m+r}{1-p/m}\mathbf{e}_1\mathbf{e}_1^T$ mentioned below (3.2). When $\mathbf{YY}^*$ is not invertible, by Theorem 1.2 of [2] we can find out

TABLE 4
Power of several two-tuples (p, N): the entries of $\mathbf{z}_t$ follow continuous uniform distribution $U(-\sqrt{3}, \sqrt{3})$

		Initial two-tuples $M_0 = (5, 40)$ and $M_1 = (30, 40)$								
Triples	**Approach**	$a = 0.1$	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
M_0	LSST	0.0198	0.0226	0.0251	0.0363	0.0608	0.1223	0.2510	0.4812	0.7789
	OSI	0.0593	0.0618	0.0645	0.0669	0.0782	0.0841	0.0935	0.1204	0.1845
$2M_0$	LSST	0.0324	0.0339	0.0389	0.0709	0.1669	0.3910	0.7120	0.9584	0.9996
	OSI	0.0590	0.0607	0.0622	0.0601	0.0631	0.0694	0.0741	0.0815	0.1028
$3M_0$	LSST	0.0388	0.0386	0.0474	0.1275	0.3062	0.6668	0.9516	0.9996	1.0000
	OSI	0.0616	0.0593	0.0574	0.0607	0.0590	0.0655	0.0665	0.0766	0.0854
$4M_0$	LSST	0.0389	0.0380	0.0688	0.1805	0.4707	0.8653	0.9961	1.0000	1.0000
	OSI	0.0579	0.0606	0.0583	0.0623	0.0598	0.0609	0.0669	0.0672	0.0753
$5M_0$	LSST	0.0390	0.0409	0.0862	0.2463	0.6321	0.9567	1.0000	1.0000	1.0000
	OSI	0.0551	0.0589	0.0550	0.0570	0.0641	0.0618	0.0651	0.0671	0.0704
M_1	LSST	0.0293	0.0321	0.0364	0.0507	0.0730	0.1025	0.1438	0.1868	0.2267
	OSI	0.0613	0.0615	0.0668	0.0676	0.0695	0.0678	0.0667	0.0705	0.0904
$2M_1$	LSST	0.0360	0.0369	0.0499	0.0878	0.1503	0.2467	0.3531	0.4604	0.5364
	OSI	0.0550	0.0605	0.0565	0.0601	0.0595	0.0613	0.0632	0.0578	0.0577
$3M_1$	LSST	0.0388	0.0453	0.0689	0.1333	0.2550	0.4188	0.5903	0.7249	0.7883
	OSI	0.0562	0.0602	0.0566	0.0552	0.0611	0.0558	0.0601	0.0566	0.0518
$4M_1$	LSST	0.0439	0.0478	0.0920	0.1871	0.3556	0.5922	0.7885	0.8914	0.9316
	OSI	0.0587	0.0562	0.0583	0.0608	0.0611	0.0577	0.0556	0.0559	0.0540
$5M_1$	LSST	0.0396	0.0504	0.1107	0.2458	0.4794	0.7570	0.9068	0.9659	0.9831
	OSI	0.0566	0.0585	0.0579	0.0607	0.0580	0.0622	0.0603	0.0576	0.0538

that the smallest nonzero eigenvalue of $\frac{1}{m}\mathbf{\Sigma}^{-1/2}\mathbf{Y}\mathbf{Y}^*\mathbf{\Sigma}^{-1/2}$ is not spiked for the above $\mathbf{\Sigma}$. So it is hard to get a spiked F matrix. Therefore, we use another matrix

$$\mathbf{\Sigma}(\omega) = \begin{pmatrix} 1 & & & & & & \\ & \omega & & & & & \\ & & 1 & & & & \\ & & & \omega & & & \\ & & & & \ddots & & \\ & & & & & 1 & \\ & & & & & & \omega \end{pmatrix}.$$

In Tables 5–7, the data $\mathbf{X}$ and $\mathbf{Y}$ are generated as in Tables 1–3 and the nominal significant level of our test is 5%.

TABLE 5
Power of several triples (p, m, n): Gaussian distribution

	Initial triple $M_0 = (5, 40, 10)$					**Initial triple $M_1 = (30, 20, 25)$**			
τ	M_0	$2M_0$	$3M_0$	$4M_0$	ω	M_1	$2M_1$	$3M_1$	$4M_1$
0.5	0.0672	0.0585	0.0563	0.0593	0.3	0.2178	0.4934	0.7071	0.8419
2	0.2763	0.3801	0.4551	0.5067	0.6	0.0574	0.1332	0.2241	0.3106
4	0.6291	0.816	0.9072	0.9567	2	0.1037	0.2166	0.3463	0.5029
6	0.8162	0.9543	0.988	0.9967	3	0.2242	0.5521	0.8156	0.9537

TABLE 6
Power of several triples (p, m, n): Discrete distribution with the probability mass function $P(\mathbf{x} = \sqrt{3}) = P(\mathbf{x} = -\sqrt{3}) = 1/6$ and $P(\mathbf{x} = 0) = 2/3$

	Initial triple $M_0 = (5, 40, 10)$					**Initial triple $M_1 = (30, 20, 25)$**			
τ	M_0	$2M_0$	$3M_0$	$4M_0$	ω	M_1	$2M_1$	$3M_1$	$4M_1$
0.5	0.0674	0.0573	0.0576	0.0595	0.3	0.2101	0.4883	0.7024	0.8425
2	0.3045	0.397	0.4561	0.5171	0.6	0.057	0.1382	0.2176	0.3078
4	0.647	0.8137	0.8984	0.9478	2	0.1055	0.2232	0.3504	0.4974
6	0.8147	0.943	0.9813	0.9936	3	0.2254	0.5487	0.8211	0.9529

TABLE 7
Power of several triples (p, m, n): Continuous uniform distribution $U(-\sqrt{3}, \sqrt{3})$

	Initial triple $M_0 = (30, 80, 40)$					**Initial triple $M_1 = (80, 40, 50)$**			
τ	M_0	$2M_0$	$3M_0$	$4M_0$	ω	M_1	$2M_1$	$3M_1$	$4M_1$
0.5	0.0394	0.0452	0.0497	0.0440	0.3	0.8209	0.9865	0.9995	1.0000
2	0.4495	0.6622	0.7989	0.8708	0.6	0.2712	0.5675	0.7669	0.8765
4	0.9791	0.9998	1.0000	1.0000	2	0.3115	0.7163	0.9381	0.9937
6	1.0000	1.0000	1.0000	1.0000	3	0.7285	0.9933	1.0000	1.0000

In Tables 5–7, we can find that when $\tau = 0.5 < 1$, $(\mathbf{YY}^*)^{-1}\mathbf{\Sigma}^{1/2}\mathbf{XX}^*\mathbf{\Sigma}^{1/2}$ is not a spiked F matrix and the power is poor. When $\tau > 1$, it is a spiked F matrix and by Proposition 1 the power increases with the dimension and τ. This phenomenon is due to the fact that it may not cause significant change to the largest eigenvalue of F matrix when finite rank perturbation is weak enough. This phenomenon has been widely discussed for sample covariance matrices; see [10] and [3]. For the spiked F matrix, one can refer to [5] and [34]. For the non-invertible case when $\mathbf{\Sigma}$ is far away from $\mathbf{I}$ ($\omega = 0.3$ or 3), the power becomes better. This is because when the empirical spectral distribution (ESD) of $\mathbf{\Sigma}$ is

TABLE 8
Power comparison of several triples (p, m, n): *Gaussian distribution with* $M_0 = (5, 40, 10)$

	EOM				CLRT			
τ	M_0	$2M_0$	$3M_0$	$4M_0$	M_0	$2M_0$	$3M_0$	$4M_0$
0.5	0.0672	0.0585	0.0563	0.0593	0.1051	0.0909	0.0777	0.0787
2	0.2763	0.3801	0.4551	0.5067	0.2696	0.2614	0.2503	0.2500
4	0.6291	0.816	0.9072	0.9567	0.5385	0.5867	0.6126	0.6282
6	0.8162	0.9543	0.988	0.9967	0.7318	0.8157	0.8499	0.8676

very different from the M–P law λ_1 may tend to another point $\mu_{\boldsymbol{\Sigma}}$ instead of μ_p. Then we may gain good power because $\breve{n}^{2/3}(\mu_{\boldsymbol{\Sigma}} - \mu_p)$ may tend to infinity.

Table 8 lists the power comparison between our approach (denoted by EOM) and CLRT, where we only consider the alternative case (3.2) and $\boldsymbol{\Sigma} = \mathbf{I} + \tau \frac{p/m+r}{1-p/m}\mathbf{e}_1\mathbf{e}_1^T$. The nominal significant level is also 5%.

The comparison shows that when $\boldsymbol{\Sigma}$ is a spiked covariance matrix, our statistic performs better than CLRT, which is consistent with our discussion in Section 3.2.

3.5.4. *Power of "correlated noise detection"*. We consider the model (3.6), that is, $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \Sigma^{1/2}\varepsilon_t, t = 1, 2, \ldots, T$. Here, we choose $k = 1$, $\mathbf{A} = \sqrt{\tau \frac{p/m+r}{1-p/m}}\mathbf{e}_1$, $\Sigma = \mathbf{I}$, $\mathbf{x}_t \sim U(-\sqrt{3}, \sqrt{3})$ and the entries of ε_t also follow the uniform distribution $U(-\sqrt{3}, \sqrt{3})$. Then it is easy to see that $\mathrm{Cov}(\mathbf{y}_t) = \mathbf{I} + \tau \frac{p/m+r}{1-p/m}\mathbf{e}_1\mathbf{e}_1^T$, which is similar to Section 3.5.3. Since we can generate an i.i.d. copy of ε_t, say $\mathbf{z}_t$, $t = 1, \ldots, T_1$, in engineering as we discussed in Section 3.3, we always choose $T_1 > p$ for simplicity.

Table 9 lists the power study of the model and the nominal significant level of our test is 5%.

TABLE 9
Power of several triples (p, T_1, T): *Continuous uniform distribution* $U(-\sqrt{3}, \sqrt{3})$

	Initial triple $M_0 = (30, 80, 40)$			
τ	M_0	$2M_0$	$3M_0$	$4M_0$
0.5	0.0412	0.0436	0.0485	0.0469
2	0.4485	0.6399	0.7624	0.8401
4	0.9486	0.9983	0.9998	1.0000
6	0.9964	1.0000	1.0000	1.0000

The simulation result in Table 9 is similar to that for the spiked case in Table 7 above. But we would mention one difference. Write $\mathbf{y}_t = (\sqrt{\tau \frac{p/m+r}{1-p/m}}\mathbf{e}_1 \mathbf{I})\binom{\mathbf{s}_t}{\varepsilon_t}$ with $(\sqrt{\tau \frac{p/m+r}{1-p/m}}\mathbf{e}_1 \mathbf{I})$ being of size $p \times (p+1)$. However, Theorem 3.1 of [34] did not consider such a rectangular matrix. This means we cannot directly apply Proposition 1 to say the power tends to 1 when $\tau > 1$. But when the $(p+1)$-dimensional vector $\binom{\mathbf{s}_t}{\varepsilon_t}$ follows standard multivariate Gaussian distribution, we then have $\mathbf{y}_t \stackrel{d}{=} (\tau \frac{p/m+r}{1-p/m}\mathbf{e}_1\mathbf{e}_1^T + \mathbf{I})^{1/2}\mathbf{z}_t$, where $\mathbf{z}_t \stackrel{d}{=} \varepsilon_t$. This means that Proposition 1 holds under the Gaussian case. From the simulation result, we conjecture that Theorem 3.1 of [34] and Proposition 1 still hold for this model even if $(\sqrt{\tau \frac{p/m+r}{1-p/m}}\mathbf{e}_1 \mathbf{I})$ is not a square matrix.

4. Proof of part (i) of Theorem 2.1. In this section, we focus on (i) of Theorem 2.1, that is, $(\check{m}, \check{n}, \check{p}) = (m, n, p)$. Actually, Lemmas 1 and 2 below hold if we replace (m, n, p) by $(\check{m}, \check{n}, \check{p})$.

4.1. *Two key lemmas.* This subsection is to first prove two key lemmas for proving part (i) of Theorem 2.1. We begin with some notation and definitions. Throughout the paper, we use $M, M_0, M_0', M_0'', M_1, M_1''$ to denote some generic positive constants whose values may differ from line to line. We also use D to denote sufficiently large positive constants whose values may differ from line to line. We say that an event Λ holds with high probability if for any big positive constant D

$$P(\Lambda^c) \le n^{-D},$$

for sufficiently large n. Recall the definition of γ_j in (2.4). Let $c_{p,0} \in [0, a_p)$ satisfy

$$\frac{1}{p}\sum_{j=1}^{p}\left(\frac{c_{p,0}}{\gamma_j - c_{p,0}}\right)^2 = \frac{n}{p}. \tag{4.1}$$

Existence of $c_{p,0}$ will be verified in Lemma 1 below. Moreover, define

$$\begin{aligned} \mu_{p,0} &= \frac{1}{c_{p,0}}\left(1 + \frac{1}{n}\sum_{j=1}^{p}\left(\frac{c_{p,0}}{\gamma_j - c_{p,0}}\right)\right), \\ \frac{1}{\sigma_{p,0}^3} &= \frac{1}{c_{p,0}^3}\left(1 + \frac{1}{n}\sum_{j=1}^{p}\left(\frac{c_{p,0}}{\gamma_j - c_{p,0}}\right)^3\right). \end{aligned} \tag{4.2}$$

Set $\mathbf{A}_p = \frac{1}{m}\mathbf{Y}\mathbf{Y}^*$ and $\mathbf{B}_p = \frac{1}{n}\mathbf{X}\mathbf{X}^*$. Rank the eigenvalues of the matrix $\mathbf{A}_p$ as $\hat{\gamma}_1 \ge \hat{\gamma}_2 \ge \cdots \ge \hat{\gamma}_p$. Let $\hat{c}_p \in [0, \hat{\gamma}_p)$ satisfy

$$\frac{1}{p}\sum_{j=1}^{p}\left(\frac{\hat{c}_p}{\hat{\gamma}_j - \hat{c}_p}\right)^2 = \frac{n}{p}. \tag{4.3}$$

The existence of $\hat{c}_p$ with high probability will be given in Lemma 2 below. Moreover, set

$$\begin{aligned}\hat{\mu}_p &= \frac{1}{\hat{c}_p}\left(1 + \frac{1}{n}\sum_{j=1}^{p}\left(\frac{\hat{c}_p}{\hat{\gamma}_j - \hat{c}_p}\right)\right),\\ \frac{1}{\hat{\sigma}_p^3} &= \frac{1}{\hat{c}_p^3}\left(1 + \frac{1}{n}\sum_{j=1}^{p}\left(\frac{\hat{c}_p}{\hat{\gamma}_j - \hat{c}_p}\right)^3\right).\end{aligned} \tag{4.4}$$

We now discuss the properties of $c_p, c_{p,0}, \hat{c}_p, \mu_p, \mu_{p,0}, \hat{\mu}_p, \sigma_p, \sigma_{p,0}$ defined (2.5)–(2.7), (4.1)–(4.4) in the next two lemmas. These lemmas are crucial to the proof strategy which transforms F matrices into an appropriate sample covariance matrix.

LEMMA 1. *Under the conditions in Theorem 2.1, there exists a constant M_0 such that*

$$\sup_p\left\{\frac{c_p}{a_p - c_p}\right\} \le M_0, \qquad \sup_p\left\{\frac{c_{p,0}}{a_p - c_{p,0}}\right\} \le M_0, \tag{4.5}$$

$$\lim_{p\to\infty} n^{2/3}|\mu_p - \mu_{p,0}| = 0 \tag{4.6}$$

and

$$\lim_{p\to\infty}\frac{\sigma_p}{\sigma_{p,0}} = 1, \qquad \limsup_p \frac{c_{p,0}}{a_p} < 1. \tag{4.7}$$

LEMMA 2. *Under the conditions in Theorem 2.1, for any $\zeta > 0$ there exists a constant $M_\zeta \ge M_0$ such that*

$$\sup_p\left\{\frac{\hat{c}_p}{\hat{\gamma}_p - \hat{c}_p}\right\} \le M_\zeta, \qquad \limsup_p \frac{\hat{c}_p}{\hat{\gamma}_p} < 1 \tag{4.8}$$

and

$$\lim_{p\to\infty} n^{2/3}|\hat{\mu}_p - \mu_{p,0}| = 0, \qquad \lim_{p\to\infty}\frac{\hat{\sigma}_p}{\sigma_{p,0}} = 1 \tag{4.9}$$

hold with high probability. Indeed (4.8) and (4.9) hold on the event S_ζ defined by

$$S_\zeta = \{\forall j, 1 \le j \le p, |\hat{\gamma}_j - \gamma_j| \le p^\zeta p^{-2/3}\tilde{j}^{-1/3}\}, \tag{4.10}$$

where ζ is a sufficiently small positive constant and $\tilde{j} = \min\{\min\{m, p\} + 1 - j, j\}$.

The proofs of Lemmas 1 and 2 are given in the supplementary material [12].

4.2. *Proof of part* (i) *of Theorem* 2.1. Recall the definition of the matrices $\mathbf{A}_p$ and $\mathbf{B}_p$ above (4.3). Define a F matrix $\mathbf{F} = \mathbf{A}_p^{-1}\mathbf{B}_p$ whose largest eigenvalue is λ_1 according to the definition of λ_1 in Theorem 2.1. It then suffices to find the asymptotic distribution of λ_1 to prove Theorem 2.1.

Recalling the definition of the event S_ζ in (4.10), we may write

$$\begin{aligned} &P\big(\sigma_p n^{2/3}(\lambda_1 - \mu_p) \le s\big) \\ &\quad = P\big((\sigma_p n^{2/3}(\lambda_1 - \mu_p) \le s) \cap S_\zeta\big) + P\big((\sigma_p n^{2/3}(\lambda_1 - \mu_p) \le s) \cap S_\zeta^c\big). \end{aligned}$$

This implies that (2.10) is equivalent to

$$\lim_{p\to\infty} P\big((\sigma_p n^{2/3}(\lambda_1 - \mu_p) \le s) \cap S_\zeta\big) = F_1(s), \tag{4.11}$$

where we use the fact that $P(S_\zeta^c) \le p^{-D}$ for any positive D by Theorem 3.3 of [25].

Write

$$\sigma_p n^{2/3}(\lambda_1 - \mu_p) = \frac{\sigma_p}{\hat{\sigma}_p}\hat{\sigma}_p n^{2/3}(\lambda_1 - \hat{\mu}_p) + \sigma_p n^{2/3}(\hat{\mu}_p - \mu_p) \tag{4.12}$$

[see (4.3) and (4.4) for $\hat{\sigma}_p$ and $\hat{\mu}_p$]. Note that the eigenvalues of $\mathbf{A}_p^{-1}$ are $\frac{1}{\hat{\gamma}_1} \le \frac{1}{\hat{\gamma}_2} \le \cdots \le \frac{1}{\hat{\gamma}_p}$. Rewrite (4.3) as

$$\frac{1}{p}\sum_{j=1}^{p}\left(\frac{(1/\hat{\gamma}_j)\hat{c}_p}{1-(1/\hat{\gamma}_p)\hat{c}_p}\right)^2 = \frac{n}{p}. \tag{4.13}$$

Also recast (4.4) as

$$\begin{aligned} \hat{\mu}_p &= \frac{1}{\hat{c}_p}\left(1 + \frac{p}{n}\frac{1}{p}\sum_{j=1}^{p}\frac{(1/\hat{\gamma}_j)\hat{c}_p}{1-(1/\hat{\gamma}_p)\hat{c}_p}\right), \\ \frac{1}{\hat{\sigma}_p^3} &= \frac{1}{\hat{c}_p^3}\left(1 + \frac{p}{n}\frac{1}{p}\sum_{j=1}^{p}\frac{(1/\hat{\gamma}_j)\hat{c}_p}{1-(1/\hat{\gamma}_p)\hat{c}_p}\right)^3. \end{aligned} \tag{4.14}$$

Up to this stage, the result about the largest eigenvalue of the sample covariance matrices $\mathbf{Z}\mathbf{Z}^*\mathbf{\Sigma}$ with $\mathbf{\Sigma}$ being the population covariance matrix comes into play where $\mathbf{Z}$ is of size $p \times n$ satisfying Condition 1 and $\mathbf{\Sigma}$ is of size $p \times p$. A key condition to ensure Tracy–Widom's law for the largest eigenvalue is that if $\rho \in (0, 1/\sigma_1)$ is the solution to the equation

$$\int\left(\frac{t\rho}{1-t\rho}\right)^2 dF^{\mathbf{\Sigma}}(t) = \frac{n}{p} \tag{4.15}$$

then

$$\limsup_p \rho\sigma_1 < 1 \tag{4.16}$$

(one may see [6], Conditions 1.2 and 1.4 and Theorem 1.3 [3], Definition 2.7(i) and Corollary 3.19 of [17]). Here, $F^{\mathbf{\Sigma}}(t)$ denotes the empirical spectral distribution of $\mathbf{\Sigma}$ and σ_1 means the largest eigenvalue of $\mathbf{\Sigma}$. Now given $\mathbf{A}_p$, if we treat $\mathbf{A}_p^{-1}$ as $\mathbf{\Sigma}$, then (4.16) is satisfied on the event S_ζ due to (4.3) and (4.8) in Lemma 2. It follows from Theorem 1.3 of [3] and Corollary 3.19 of [17] that

$$\lim_{p\to\infty} P\big(\big(\hat{\sigma}_p n^{2/3}(\lambda_1 - \hat{\mu}_p) \le s\big) \cap S_\zeta | \mathbf{A}_p\big) = F_1(s), \tag{4.17}$$

which implies that

$$\lim_{p\to\infty} P\big(\big(\hat{\sigma}_p n^{2/3}(\lambda_1 - \hat{\mu}_p) \le s\big) \cap S_\zeta\big) = F_1(s). \tag{4.18}$$

Moreover, by Lemmas 1 and 2 we obtain on the event S_ζ

$$\lim_{p\to\infty} \frac{\sigma_p}{\hat{\sigma}_p} = 1 \tag{4.19}$$

and

$$\lim_{p\to\infty} \sigma_p n^{2/3}(\hat{\mu}_p - \mu_p) = 0. \tag{4.20}$$

Equation (4.11) then follows from (4.12), (4.17)–(4.20) and Slutsky's theorem. The proof is complete.

5. Proof of part (ii) of Theorem 2.1: Standard Gaussian distribution. This section is to consider the case when $\{X_{ij}\}$ follow normal distribution with mean zero and variance one. We below first introduce more notation. Let $\mathbf{A} = (A_{ij})$ be a matrix. We define the following norms:

$$\|\mathbf{A}\| = \max_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|, \qquad \|\mathbf{A}\|_\infty = \max_{i,j} |A_{ij}|, \qquad \|\mathbf{A}\|_F = \sqrt{\sum_{ij} |A_{ij}|^2},$$

where $|\mathbf{x}|$ represents the Euclidean norm of a vector $\mathbf{x}$. Notice that $\|\cdot\|$ is a "pseudo norm" and we have a simple relationship among these norms

$$\|\mathbf{A}\|_\infty \le \|\mathbf{A}\| \le \|\mathbf{A}\|_F.$$

We also need the following commonly used definition about stochastic domination to simplify the statements.

DEFINITION 2 (Stochastic domination). Let

$$\xi = \{\xi^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\}, \qquad \zeta = \{\zeta^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\}$$

be two families of random variables, where $U^{(n)}$ is a n-dependent parameter set (or independent of n). If for sufficiently small positive ε and sufficiently large σ,

$$\sup_{u\in U^{(n)}} \mathbb{P}\big[\big|\xi^{(n)}(u)\big| > n^\varepsilon \big|\zeta^{(n)}(u)\big|\big] \le n^{-\sigma}$$

for large enough $n \geq n(\varepsilon, \sigma)$, then we say that ζ stochastically dominates ξ uniformly in u. We denote this relationship by $|\xi| \prec \zeta$ and also write it as $\xi = O_{\prec}(\zeta)$. Furthermore, we also write it as $|x| \prec y$ if x and y are both nonrandom and $|x| \leq n^{\varepsilon}|y|$ for sufficiently small positive ε.

PROOF OF PART (ii) OF THEOREM 2.1. We start the proof by reminding readers that $m < p$ and $m + n > p$. Since $m < p$, the limit of the empirical distribution function of $\frac{1}{p}\mathbf{Y}^*\mathbf{Y}$ is the M–P law and we denote its density by $\rho_{pm}(x)$. We define $\gamma_{m,1} \geq \gamma_{m,2} \geq \cdots \geq \gamma_{m,m}$ to satisfy

$$\int_{\gamma_{m,j}}^{+\infty} \rho_{pm}\, dx = \frac{j}{m}, \tag{5.1}$$

with $\gamma_{m,0} = (1 + \sqrt{\frac{m}{p}})^2$, $\gamma_{m,m} = (1 - \sqrt{\frac{m}{p}})^2$. Correspondingly, denote the eigenvalues of $\frac{1}{p}\mathbf{Y}^*\mathbf{Y}$ by $\hat{\gamma}_{m,1} \geq \hat{\gamma}_{m,2} \geq \cdots \geq \hat{\gamma}_{m,m}$. Here, we would remind the readers that $\rho_{pm}(x), \gamma_{m,j}, \hat{\gamma}_{m,1}$ are similar to those in (2.3), below (2.3) and above (4.3) except that we are interchanging the role of p and m because we are considering $\frac{1}{p}\mathbf{Y}^*\mathbf{Y}$ rather than $\frac{1}{m}\mathbf{Y}\mathbf{Y}^*$. Moreover, by Theorem 3.3 of [25] and (4.10) for any sufficiently small $\zeta > 0$ and big $D > 0$ there exists an event S_ζ (here with a bit abuse of notion S_ζ) such that

$$S_\zeta = \{\forall j, 1 \leq j \leq m, |\hat{\gamma}_{m,j} - \gamma_{m,j}| \leq p^{\zeta - 2/3}\tilde{j}^{-1/3}\} \tag{5.2}$$

and

$$P(S_\zeta^c) \leq p^{-D}. \tag{5.3}$$

Note that $\frac{1}{p}\mathbf{Y}\mathbf{Y}^*$ and $\frac{1}{p}\mathbf{Y}^*\mathbf{Y}$ have the same nonzero eigenvalues. To simplify notation, let $m_p = m + n - p$. Write

$$\frac{1}{p}\mathbf{Y}\mathbf{Y}^* = \mathbf{U}^*\begin{pmatrix} \mathbf{D} & 0 \\ 0 & 0 \end{pmatrix}\mathbf{U}, \tag{5.4}$$

with $\mathbf{D} = \operatorname{diag}\{\hat{\gamma}_{m,1}, \hat{\gamma}_{m,2}, \ldots, \hat{\gamma}_{m,m}\}$ and $\mathbf{U}$ is an orthogonal matrix. Then $\det(\lambda\frac{\mathbf{Y}\mathbf{Y}^*}{p} - \frac{\mathbf{X}\mathbf{X}^*}{m_p}) = 0$ is equivalent to

$$\det\left(\lambda\begin{pmatrix} \mathbf{D} & 0 \\ 0 & 0 \end{pmatrix} - \frac{1}{m_p}U\mathbf{X}\mathbf{X}^*U^*\right) = 0.$$

Moreover, since $\{\mathbf{X}_{ij}\}$ are independent standard normal random variables and $\mathbf{U}$ is an orthogonal matrix we have $U\mathbf{X} \stackrel{d}{=} \mathbf{X}$ so that it suffices to consider the following determinant:

$$\det\left(\lambda\begin{pmatrix} \mathbf{D} & 0 \\ 0 & 0 \end{pmatrix} - \frac{1}{m_p}\mathbf{X}\mathbf{X}^*\right) = 0. \tag{5.5}$$

Here, $\stackrel{d}{=}$ means having the identical distribution.

Now rewrite $\mathbf{X}$ as $\mathbf{X} = \binom{\mathbf{X}_1}{\mathbf{X}_2}$, where $\mathbf{X}_1$ is a $m \times n$ matrix and $\mathbf{X}_2$ is a $(p-m) \times n$ matrix. It follows that

$$\mathbf{X}\mathbf{X}^* = \begin{pmatrix} \mathbf{X}_1\mathbf{X}_1^* & \mathbf{X}_1\mathbf{X}_2^* \\ \mathbf{X}_2\mathbf{X}_1^* & \mathbf{X}_2\mathbf{X}_2^* \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{pmatrix}. \tag{5.6}$$

Equation (5.5) can be rewritten as

$$\det \begin{pmatrix} \frac{1}{m_p}\mathbf{X}_{11} - \lambda \mathbf{D} & \frac{1}{m_p}\mathbf{X}_{12} \\ \frac{1}{m_p}\mathbf{X}_{21} & \frac{1}{m_p}\mathbf{X}_{22} \end{pmatrix} = 0.$$

Since $m + n > p$, $\mathbf{X}_{22}$ is invertible. Equation (5.5) is further equivalent to

$$\det\left(\frac{1}{m_p}\mathbf{X}_{11} - \lambda \mathbf{D} - \frac{1}{m_p}\mathbf{X}_{12}\mathbf{X}_{22}^{-1}\mathbf{X}_{21}\right) = 0. \tag{5.7}$$

Moreover,

$$\begin{aligned} \mathbf{X}_{11} - \mathbf{X}_{12}\mathbf{X}_{22}^{-1}\mathbf{X}_{21} &= \mathbf{X}_1\mathbf{X}_1^* - \mathbf{X}_1\mathbf{X}_2^*(\mathbf{X}_2\mathbf{X}_2^*)^{-1}\mathbf{X}_2\mathbf{X}_1^* \\ &= \mathbf{X}_1(I_n - \mathbf{X}_2^*(\mathbf{X}_2\mathbf{X}_2^*)^{-1}\mathbf{X}_2)\mathbf{X}_1^*. \end{aligned}$$

Since $\operatorname{rank}(I_n - \mathbf{X}_2^*(\mathbf{X}_2\mathbf{X}_2^*)^{-1}\mathbf{X}_2) = m + n - p = m_p$, we can write

$$\mathbf{I}_n - \mathbf{X}_2^*(\mathbf{X}_2\mathbf{X}_2^*)^{-1}\mathbf{X}_2 = \mathbf{V}\begin{pmatrix} \mathbf{I}_{m_p} & 0 \\ 0 & 0 \end{pmatrix}\mathbf{V}^*,$$

where $\mathbf{V}$ is an orthogonal matrix. In view of the above, we can construct a $m \times m_p$ matrix $\mathbf{Z} = (Z_{ij})_{m,m_p}$ consisting of independent standard normal random variables so that

$$\mathbf{X}_{11} - \mathbf{X}_{12}\mathbf{X}_{22}^{-1}\mathbf{X}_{21} \stackrel{d}{=} \mathbf{Z}\mathbf{Z}^*. \tag{5.8}$$

It follows that (5.7), and hence (5.5) are equivalent to

$$\det\left(\frac{1}{m_p}\mathbf{Z}\mathbf{Z}^* - \lambda \mathbf{D}\right) = 0. \tag{5.9}$$

It then suffices to consider the largest eigenvalue of $\frac{1}{m_p}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^*$. Denote by λ_1 the largest eigenvalue of $\frac{1}{m_p}D^{-1}ZZ^*$. As in (4.3) and (4.4) define $\hat{c}_m \in [0, \hat{\gamma}_{m,m})$ to satisfy

$$\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\hat{c}_m}{\hat{\gamma}_{m,j} - \hat{c}_m}\right)^2 = \frac{m_p}{m} \tag{5.10}$$

and $\hat{\mu}_p$ and $\hat{\sigma}_p$ by

$$\hat{\mu}_m = \frac{1}{\hat{c}_m}\left(1 + \frac{1}{m_p}\sum_{j=1}^{m}\left(\frac{\hat{c}_m}{\hat{\gamma}_{m,j} - \hat{c}_m}\right)\right),$$

$$\frac{1}{\hat{\sigma}_m^3} = \frac{1}{\hat{c}_m^3}\left(1 + \frac{1}{m_p}\sum_{j=1}^{m}\left(\frac{\hat{c}_m}{\hat{\gamma}_{m,j} - \hat{c}_m}\right)^3\right).$$

From Lemma 2, we have on the event S_ζ

$$\limsup_p \frac{\hat{c}_m}{\hat{\gamma}_{m,m}} < 1, \tag{5.11}$$

which implies condition (4.16). It follows from Theorem 1.3 of [3] and Corollary 3.19 of [17] that

$$\lim_{p\to\infty} P\big(\hat{\sigma}_m(m+n-p)^{2/3}(\lambda_1 - \hat{\mu}_m) \le s\big) = F_1(s). \tag{5.12}$$

As in the proof of Theorem 2.1, by Lemmas 1 and 2 one may further conclude that

$$\lim_{p\to\infty} P\big(\sigma_p(m+n-p)^{2/3}(\lambda_1 - \mu_p) \le s\big) = F_1(s). \tag{5.13}$$

□

6. Proof of part (ii) of Theorem 2.1: General distributions. The aim of this section is to relax the Gaussian assumption on $\mathbf{X}$. We below assume that $\mathbf{X}$ and $\mathbf{Y}$ are real matrices. The complex case can be handled similarly, and hence we omit it here. In the sequel, we absorb $\frac{1}{\sqrt{m+n-p}}$ and $\frac{1}{\sqrt{p}}$ into $\mathbf{X}$ and $\mathbf{Y}$, respectively [i.e., $\operatorname{Var}(X_{ij}) = \frac{1}{m+n-p}$, $\operatorname{Var}(Y_{st}) = \frac{1}{p}$] for convenience.

In terms of the notation in this section [$\operatorname{Var}(\mathbf{Y}_{st}) = \frac{1}{p}$], (5.4) can be rewritten as

$$\mathbf{YY}^* = \mathbf{U}^* \begin{pmatrix} \mathbf{D} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{U}.$$

Break $\mathbf{U}$ as $\binom{\mathbf{U}_1}{\mathbf{U}_2}$ where $\mathbf{U}_1$ and $\mathbf{U}_2$ are $m \times p$ and $(p-m) \times p$, respectively. By (5.4)–(5.7) (note that here we cannot omit $\mathbf{U}$ by $\mathbf{UX} \overset{d}{=} \mathbf{X}$), the maximum eigenvalue of $\det(\lambda \mathbf{YY}^* - \mathbf{XX}^*)$ is equivalent to that of the following matrix:

$$\begin{aligned} \mathbf{A} &= \mathbf{D}^{-1/2}\mathbf{U}_1\mathbf{X}\big(I - \mathbf{X}^T\mathbf{U}_2^T\big(\mathbf{U}_2\mathbf{XX}^T U_2^T\big)^{-1} U_2\mathbf{X}\big)\mathbf{X}^T\mathbf{U}_1^T\mathbf{D}^{-1/2} \\ &\triangleq \mathbf{D}^{-1/2}\mathbf{U}_1\mathbf{X}(I - P_{\mathbf{X}^T U_2^T})\mathbf{X}^T\mathbf{U}_1^T\mathbf{D}^{-1/2}, \end{aligned} \tag{6.1}$$

where $\mathbf{P}_{\mathbf{X}^T\mathbf{U}_2^T}$ is the projection matrix. It is not necessary to assume that $\mathbf{U}_2\mathbf{XX}^T\mathbf{U}_2^T$ is invertible since $P_{\mathbf{X}^T U_2^T}$ is unique even if $(\mathbf{U}_2\mathbf{XX}^T\mathbf{U}_2^T)^-$ is the generalized inverse matrix of $\mathbf{U}_2\mathbf{XX}^T\mathbf{U}_2^T$. Moreover, we indeed have the following lemma to control the smallest eigenvalue of $\mathbf{U}_2\mathbf{XX}^T\mathbf{U}_2^T$.

LEMMA 3. *Suppose that $(m+n-p)^{1/2}\mathbf{X}$ satisfies Condition 1. Then $\mathbf{U}_2\mathbf{XX}^T\mathbf{U}_2^T$ is invertible and*

$$\big\|\big(\mathbf{U}_2\mathbf{XX}^T\mathbf{U}_2^T\big)^{-1}\big\| \le M \tag{6.2}$$

for a large constant M with high probability. Moreover,

$$\|\mathbf{XX}^*\| \le M \tag{6.3}$$

with high probability under conditions in Theorem 2.1.

PROOF. One may check that the conditions in Theorem 3.12 in [17] are satisfied when considering $\mathbf{U}_2\mathbf{X}\mathbf{X}^T\mathbf{U}_2^T$. Applying Theorem 3.12 in [17], then yields

$$\left|\lambda_{\min}(\mathbf{U}_2\mathbf{X}\mathbf{X}^T\mathbf{U}_2^T) - \left(1 - \sqrt{\frac{n}{p-m}}\right)^2\right| \prec n^{-2/3},$$

where $(1 - \sqrt{\frac{n}{p-m}})^2$ can be obtained when considering the special case when the entries of $\mathbf{X}$ are Gaussian. As for (6.3), see Lemma 4.8 in [17]. □

Since the matrix in (6.1) is quite complicated, we construct a linearization matrix for it

$$\mathbf{H} = \mathbf{H}(\mathbf{X}) = \begin{pmatrix} -z\mathbf{I} & 0 & \mathbf{D}^{-1/2}\mathbf{U}_1\mathbf{X} \\ 0 & 0 & \mathbf{U}_2\mathbf{X} \\ \mathbf{X}^T\mathbf{U}_1^T\mathbf{D}^{-1/2} & \mathbf{X}^T\mathbf{U}_2^T & -\mathbf{I} \end{pmatrix}. \tag{6.4}$$

The connection between $\mathbf{H}$ and the matrix in (6.1) is that the upper left block of the 3×3 block matrix $\mathbf{H}^{-1}$ is the Stieltjes transform of (6.1) by simple calculations. We next give the limit of the Stieltjes transform of (6.1) and need the following well-known result (see [1]). There exists a unique solution $m(z) : \mathcal{C}^+ \to \mathcal{C}$ such that

$$\frac{1}{m(z)} = -z + \frac{m}{m+n-p} \int \frac{t}{1+tm(z)}\, dH_n(t), \tag{6.5}$$

where H_n is the empirical distribution function of $\mathbf{D}^{-1}$. Moreover, we set

$$\underline{m}(z) = -\frac{1}{m}\operatorname{Tr}\big(z\big(1 + m(z)\mathbf{D}^{-1}\big)\big)^{-1}, \qquad \rho(x) = \lim_{z \in \mathcal{C}^+ \to x} \Im m(z).$$

From the end of the last section, we see that under the Gaussian case (6.1) $\stackrel{d}{=}$ $\mathbf{D}^{-1/2}\mathbf{Z}\mathbf{Z}^*\mathbf{D}^{-1/2}$. Hence, it is easy to see that $\hat{\mu}_m$ defined above (5.11) is the right most end point of the support of $\rho(x)$.

For any small positive constant τ, we define the domains

$$E(\tau, n) = \{z = E + i\eta \in \mathbb{C}^+ : |z| \ge \tau, |E| \le \tau^{-1}, n^{-1+\tau} \le \eta \le \tau^{-1}\}, \tag{6.6}$$

$$E_+ = E_+(\tau, \tau', n) = \{z \in E(\tau, n) : E \ge \hat{\mu}_m - \tau'\}, \tag{6.7}$$

where τ' is a sufficiently small positive constant.

Set

$$\begin{aligned} &\Psi = \Psi(z) = \sqrt{\frac{\Im m(z)}{n\eta}} + \frac{1}{n\eta}, \\ &\mathbf{G}(z) = \mathbf{H}^{-1}, \qquad \boldsymbol{\Sigma} = \Sigma(z) = z^{-1}\big(1 + m(z)\mathbf{D}^{-1}\big)^{-1}, \end{aligned} \tag{6.8}$$

and

$$\mathbf{F}(z) = \begin{pmatrix} -\boldsymbol{\Sigma} & \boldsymbol{\Sigma}\mathbf{D}^{-1/2}\mathbf{U}_1\mathbf{X}\mathbf{X}^T\mathbf{U}_2^T\boldsymbol{\Gamma} & 0 \\ \boldsymbol{\Gamma}\mathbf{U}_2\mathbf{X}\mathbf{X}^T\mathbf{U}_1^T\mathbf{D}^{-1/2}\Sigma & \boldsymbol{\Gamma} + \boldsymbol{\Gamma}\mathbf{U}_2\mathbf{X}\mathbf{X}^T\mathbf{U}_1^T\mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}\mathbf{U}_1\mathbf{X}\mathbf{X}^T\mathbf{U}_2^T\boldsymbol{\Gamma} & \boldsymbol{\Gamma}\mathbf{U}_2\mathbf{X} \\ 0 & \mathbf{X}^T\mathbf{U}_2^T\Gamma & (zm(z)+1)(\mathbf{I} - \mathbf{P}_{\mathbf{X}^T\mathbf{U}_2^T}) \end{pmatrix},$$

where $\boldsymbol{\Gamma} = (\mathbf{U}_2\mathbf{X}\mathbf{X}^T\mathbf{U}_2^T)^{-1}$. In fact, $\mathbf{F}(z)$ is close to $\mathbf{G}(z)$ with high probability.

We are now in a position to state our main result about the local law near $\hat{\mu}_m$, the right end point of the support of the limit of the ESD of $\mathbf{A}$ in (6.1).

THEOREM 6.1 (Strong local law). *Suppose that* $(m+n-p)^{1/2}\mathbf{X}$ *and* $p^{1/2}\mathbf{Y}$ *satisfy the conditions of Theorem* 2.1. *Then*:

(i) *For any deterministic unit vectors* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{p+n}$.

$$\langle \mathbf{v}, (\mathbf{G}(z) - \mathbf{F}(z))\mathbf{w}\rangle \prec \Psi \tag{6.9}$$

uniformly $z \in E_+$ *and*

(ii)

$$|\underline{m}_n(z) - \underline{m}(z)| \prec \frac{1}{n\eta} \tag{6.10}$$

uniformly in $z \in E_+$, *where* $\underline{m}_n(z) = \frac{1}{m}\sum_{i=1}^m G_{ii}$.

The proof of Theorem 6.1 is provided in the supplementary material [12].

6.1. *Convergence rate on the right edge and universality.*

6.1.1. *Convergence rate on the right edge.* We state the following lemma and the proof is given in the supplementary material [12].

LEMMA 4. *Denote by* λ_1 *the largest eigenvalue of* $\mathbf{A}$ *in* (6.1). *Under conditions of Theorem* 2.1,

$$\lambda_1 - \hat{\mu}_m = O_\prec(n^{-2/3}).$$

6.1.2. *Universality.* The aim of this subsection is to prove (ii) of Theorem 2.1. By (5.12) and (5.13), it suffices to prove edge universality at the rightmost edge of the support $\hat{\mu}_m$. In other words, the asymptotic distribution of λ_1 is not affected by the distribution of the entries of $\mathbf{X}$ under the 3rd moment matching condition. Similar to Theorem 6.4 of [9], we first show the following Green function comparison theorem.

THEOREM 6.2. *There exists $\varepsilon_0 > 0$. For any $\varepsilon < \varepsilon_0$, set $\eta = n^{-2/3-\varepsilon}$, E_1, $E_2 \in \mathbb{R}$ with $E_1 < E_2$ and*

$$|E_1 - \hat{\mu}_m|, |E_2 - \hat{\mu}_m| \leq n^{-2/3+\varepsilon}.$$

Suppose that $K : \mathbb{R} \to \mathbb{R}$ is a smooth function with bounded derivatives up to fifth order. Then there exists a constant $\phi > 0$ such that for large enough n

$$\begin{aligned}&\left|\mathbb{E}K\left(n\int_{E_1}^{E_2} \Im m_{\mathbf{X}^1}(x+i\eta)\,dx\right) - \mathbb{E}K\left(n\int_{E_1}^{E_2} \Im m_{\mathbf{X}^0}(x+i\eta)\,dx\right)\right| \\ &\quad\leq n^{-\phi}\end{aligned} \tag{6.11}$$

[*see Definition* 1 *or* (2.8) *for* $\mathbf{X}^1$ *and* $\mathbf{X}^0$].

The proof of Theorem 6.2 is given in the supplementary material [12].

In order to prove the Tracy–Widom law, we need to connect the probability $\mathbb{P}(\lambda_1 \leq E)$ with Theorem 6.2.

By Lemma 4, we can fix $E^* \prec n^{-2/3}$ such that it suffices to consider $\lambda_1 \leq \hat{\mu}_m + E^*$. Choosing $|E - \hat{\mu}_m| \prec n^{-2/3}$, $\eta = n^{-2/3-9\varepsilon}$ and $l = \frac{1}{2}n^{-2/3-\varepsilon}$, then for some sufficiently small constant $\varepsilon > 0$ and sufficiently large constant D, there exists a constant $n_0(\varepsilon, D)$ such that

$$\begin{aligned}&\mathbb{E}K\left(\frac{n}{\pi}\int_{E-l}^{\hat{\mu}_m+E^*} \Im m_{\mathbf{X}^1}(x+i\eta)\,dx\right) \\ &\quad\leq \mathbb{P}(\lambda_1 \leq E) \leq \mathbb{E}K\left(\frac{n}{\pi}\int_{E+l}^{\hat{\mu}_m+E^*} \Im m_{\mathbf{X}^1}(x+i\eta)\,dx\right) + n^{-D},\end{aligned} \tag{6.12}$$

where $n \geq n_0(\varepsilon, D)$ and K is a smooth cutoff function satisfying the condition of K in Theorem 6.2. We omit the proof of (6.12) because it is a standard procedure and one can refer to [9] or Corollary 5.1 of [4], for instance. Combining (6.12) with Theorem 6.2 one can prove Tracy–Widom's law directly (see the proof of Theorem 1.3 of [3]).

SUPPLEMENTARY MATERIAL

Supplement: Proof of some lemmas and theorems (DOI: 10.1214/15-AOS1427SUPP; .pdf). In the supplementary file [12], we provide the proofs of (2.11), Lemmas 1, 2 and 4, Theorems 6.1 and 6.2.

REFERENCES

[1] BAI, Z. and SILVERSTEIN, J. W. (2006). *Spectral Analysis of Large Dimensional Random Matrices*, 1st ed. Springer, New York.

[2] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

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

[4] 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. Available at http://www.ntu.edu.sg/home/gmpan/publications.html.

[5] DHARMAWANSA, P., JOHNSTONE, I. M. and ONATSKI, A. (2014). Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios. Available at http://arxiv.org/pdf/1411.3875.pdf.

[6] 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

[7] ERDŐS, L., KNOWLES, A. and YAU, H. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922

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

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

[10] FÉRAL, D. and PÉCHÉ, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. *J. Math. Phys.* **50** 073302, 33. MR2548630

[11] FUJIKOSHI, Y., ULYANOV, V. V. and SHIMIZU, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, NJ. MR2640807

[12] HAN, X., PAN, G. and ZHANG, B. (2016). Supplement to "The Tracy–Widom law for the largest eigenvalue of F type matrices." DOI:10.1214/15-AOS1427SUPP.

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

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

[15] JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010

[16] JOHNSTONE, I. M. (2009). Approximate null distribution of the largest root in multivariate analysis. *Ann. Appl. Stat.* **3** 1616–1633. MR2752150

[17] KNOWLES, A. and YIN, J. (2015). Anisotropic local laws for random matrices. Available at arXiv:1410.3516v3.

[18] LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

[19] LEE, J. O. and SCHNELLI, K. (2014). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. Available at arXiv:1409.4979v1.

[20] LEVANON, N. (1988). *Radar Principles*. Wiley, New York.

[21] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb. Math.* **4** 457–483.

[22] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

[23] NADAKUDITI, R. R. and SILVERSTEIN, J. W. 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.

[24] 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

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

[26] 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

[27] TAO, T. and VU, V. (2011). Random matrices: Universality of local eigenvalue statistics. *Acta Math.* **206** 127–204. MR2784665

[28] TAO, T. and VU, V. (2012). Random covariance matrices: Universality of local statistics of eigenvalues. *Ann*. *Probab*. **40** 1285–1315. MR2962092

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

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

[31] 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

[32] WACHTER, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann*. *Statist*. **8** 937–957. MR0585695

[33] WANG, K. (2012). Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices Theory Appl*. **1** 1150005, 24. MR2930383

[34] WANG, Q. and YAO, J. (2015). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. Available at http://arxiv.org/pdf/1504.05087.pdf.

[35] YAO, J. F., ZHENG, S. R. and BAI, Z. D. (2015). *Large Sample Covariance Matrices and High-Dimemnsional Data Analysis*. Cambridge Univ. Press, Cambridge.

[36] ZENG, Y. H. and LIANG, Y. C. (2009). Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Transactions Communications* **57** 1784–1793.

[37] ZHANG, L. X. (2006). Spectral Analysis of Large Dimensional Random Matrices. Ph.D. Thesis, National University of Singapore.

[38] ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **48** 444–476. MR2954263

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE
E-MAIL: xhan011@e.ntu.edu.sg
gmpan@ntu.edu.sg
bzhang007@e.ntu.edu.sg