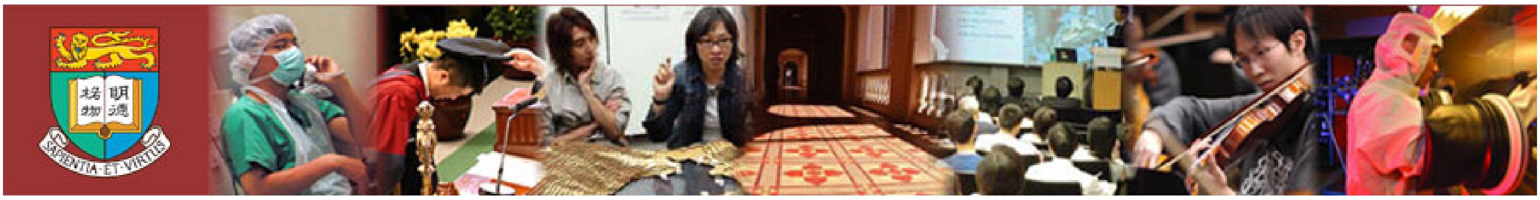

Title	Extreme eigenvalues of large-dimensional spiked Fisher matrices with application
Author(s)	Wang, Q; Yao, JJ
Citation	The Annals of Statistics, 2017, v. 45 n. 1, p. 415-460
Issued Date	2017
URL	http://hdl.handle.net/10722/231315
Rights	This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The Annals of Statistics
2017, Vol. 45, No. 1, 415–460
DOI: 10.1214/16-AOS1463

EXTREME EIGENVALUES OF LARGE-DIMENSIONAL SPIKED FISHER MATRICES WITH APPLICATION

BY QINWEN WANG AND JIANFENG YAO

University of Hong Kong

Consider two p-variate populations, not necessarily Gaussian, with covariance matrices Σ_1 and Σ_2, respectively. Let S_1 and S_2 be the corresponding sample covariance matrices with degrees of freedom m and n. When the difference Δ between Σ_1 and Σ_2 is of small rank compared to p, m and n, the Fisher matrix $S := S_2^{-1} S_1$ is called a *spiked Fisher matrix*. When p, m and n grow to infinity proportionally, we establish a phase transition for the extreme eigenvalues of the Fisher matrix: a displacement formula showing that when the eigenvalues of Δ (*spikes*) are above (or under) a critical value, the associated extreme eigenvalues of S will converge to some point outside the support of the global limit (LSD) of other eigenvalues (become outliers); otherwise, they will converge to the edge points of the LSD. Furthermore, we derive central limit theorems for those outlier eigenvalues of S. The limiting distributions are found to be Gaussian if and only if the corresponding population spike eigenvalues in Δ are *simple*. Two applications are introduced. The first application uses the largest eigenvalue of the Fisher matrix to test the equality between two high-dimensional covariance matrices, and explicit power function is found under the spiked alternative. The second application is in the field of signal detection, where an estimator for the number of signals is proposed while the covariance structure of the noise is arbitrary.

1. Introduction. Consider two p-variate populations with covariance matrices Σ_1 and Σ_2, and let S_1 and S_2 be the sample covariance matrices from samples of the two populations with degrees of freedom m and n, respectively. When the difference Δ between Σ_1 and Σ_2 is of finite rank, the Fisher matrix $S := S_2^{-1} S_1$ is called a *spiked Fisher matrix*. In this paper, we derive three results related to the extreme eigenvalues of the spiked Fisher matrix for general populations in the large-dimensional regime, that is, the dimension (p) grows to infinity together with the two sample sizes (m and n). Our first result is a phase transition phenomenon for the extreme eigenvalues of S: a displacement formula showing that when the eigenvalues of Δ (*spikes*) are above (or under) a critical value, the associated extreme eigenvalues of S will converge to some point outside the support of the global limit (LSD) of other eigenvalues (become outliers), and the location of this limit only depends on the corresponding population spike of Δ and

Received April 2015; revised March 2016.
MSC2010 subject classifications. Primary 62H12; secondary 60F05.
Key words and phrases. Large-dimensional Fisher matrices, spiked Fisher matrix, spiked population model, extreme eigenvalue, phase transition, central limit theorem, signal detection, high-dimensional data analysis.

two dimension-to-sample-size ratios; otherwise, they will converge to the edge points of the LSD. The second result is on the second-order behavior of those outlier eigenvalues of S. We show that after proper normalization, a packet of those outlier eigenvalues (corresponding to the same spike in Δ) converge to the eigenvalues' distribution of some structured Gaussian random matrix. In particular, the limiting distribution of the outlier eigenvalue of S (after normalization) is Gaussian if and only if the corresponding spike in Δ is simple. Finally, as an extension, we consider the joint distribution of all those outlier eigenvalues (correspond to different spikes in Δ) as a whole, and it is shown that those outlier eigenvalues (after normalization) converge to the eigenvalues' distribution of some block random matrix, whose structure can be fully identified. Also as a special case, if all the spikes in Δ are simple, then the joint distribution of the outlier eigenvalues of S is multivariate Gaussian.

There exists a vast literate on the spectral analysis of multivariate Fisher matrices under the assumption that both populations are Gaussian and share the same covariance matrix, that is, $\Sigma_1 = \Sigma_2$. The joint distribution of the eigenvalues of the corresponding Fisher matrix S was first simultaneously and independently published in 1939 by R. A. Fisher, S. N. Roy, P. L. Hsu and M. A. Girshick. Later in 1980, Wachter (1980) finds a deterministic limit, the celebrated Wacheter distribution, for the empirical measure of these eigenvalues when the dimension p grows to infinity proportionally with the degrees of freedom m and n (large-dimensional regime). Wachter's result has been later extended to non-Gaussian populations using the tools from the random matrix theory and two early examples of such extensions are Silverstein (1985) and Bai, Yin and Krishnaiah (1987).

In this paper, we are also interested in the large-dimensional regime, while allowing Σ_1 and Σ_2 to be separated by a (finite) rank-M matrix Δ. Besides, the two populations can have arbitrary distributions other than Gaussian. From the perturbation theory, when M is a fixed integer while p, m and n grow to infinity proportionally, the empirical measure of the p eigenvalues of S will be affected by a difference of order M/p ($\to 0$), so that its limit remains the Wachter distribution. Therefore, our main concern is the local asymptotic behaviors of the M extreme eigenvalues of S (other than the global limit). In a recent preprint Dharmawansa, Johnstone and Onatski (2014), by assuming both population are Gaussian and $M = 1$, these authors show that, when the norm of the rank-1 difference Δ (*spike*) exceeds a phase transition threshold, the asymptotic behavior of the log-ratio of the joint density of these characteristic roots under a local deviation from the spike depends only on the largest eigenvalue $l_{p,1}$ and the statistical experiment of observing all the eigenvalues is locally asymptotically normal (LAN). As a by-product of their analysis, the authors also establish joint asymptotic normality of a few of the largest eigenvalues when the corresponding spikes in Δ (with $M > 1$) exceed the phase transition threshold. The analysis given in this reference highly relies on the Gaussian assumption so that the joint density function of the eigenvalues has indeed an explicit form, and the main results are obtained via an accurate asymptotic approximation of the log-ratio of these density functions. Therefore, one of the main objectives of our work is to develop a general theory without such Gaussian assumption. It is thus apparent that the joint density of the eigenvalues of the Fisher matrix S has then no more an analytic formula and new techniques are needed to solve the questions.

Our approach relies on the tools borrowed from the theory of random matrices. A methodology particularly successful both in theory and applications within this approach relies on the *spiked population model* coined in Johnstone (2001). This model assumes the population covariance matrix has the structure $\Sigma_p = I_p + \Delta$ where the rank of Δ is M (M is a fixed integer). Again for small rank M, the empirical eigenvalue distribution of the corresponding sample covariance matrix remains the standard Marčenko–Pastur law. What makes a difference is the local asymptotic behaviors of the extreme sample eigenvalues. For example, the fluctuation of largest eigenvalues of a sample covariance matrix from a complex spiked Gaussian population is studied in Baik, Ben Arous and Péché (2005), where the authors uncover a phase transition phenomenon: the weak limit and the scaling of these extreme eigenvalues are different depending on whether the eigenvalues of Δ (*spikes*) are above, equal or below a critical value, situations refereed as *super-critical, critical* and *sub-critical*, respectively. In Baik and Silverstein (2006), the authors consider the spiked population model with general populations (not necessarily Gaussian). For the almost sure limits of the extreme sample eigenvalues, they find that if a population spike (in Δ) is large or small enough, the corresponding spiked sample eigenvalues will converge to a limit outside the support of the limiting spectrum (become *outliers*). In Paul (2007), a CLT is established for these outliers, that is, the super-critical case, under the Gaussian assumption and assuming that population spikes are simple (multiplicity 1). The CLT for super-critical outliers with general populations and arbitrary multiplicity numbers is developed in Bai and Yao (2008). Joint distributions for the outlier sample eigenvalues and eigenvectors can be found in Shi (2013) and Wang, Su and Yao (2014). A recent related application to high-dimensional regression can be found in Kargin (2015).

Within the theory of random matrices, the techniques we use in this paper for spiked models are closely connected to other random matrix ensembles through the concept of small-rank perturbations. Theories on perturbed Wigner matrices can be found in Péché (2006), Féral and Péché (2007), Capitaine, Donati-Martin and Féral (2009), Pizzo, Renfrew and Soshnikov (2013) and Renfrew and Soshnikov (2013). In a more general setting of finite-rank perturbation including both the additive and the multiplicative one, referees include Benaych-Georges and Nadakuditi (2011), Benaych-Georges, Guionnet and Maida (2011) and Capitaine (2013).

Apart from the theoretical results, we also propose two applications both in high-dimensional hypothesis testing and signal detection, respectively. The first application uses the largest eigenvalue of the Fisher matrix to test the following hypotheses:

$$H_0: \quad \Sigma_1 = \Sigma_2 \quad \text{vs.} \quad H_1: \quad \Sigma_1 = \Sigma_2 + \Delta, \tag{1.1}$$

where Δ is a nonnegative definite matrix of rank M. Under this spiked alternative H_1, explicit formula for the power function is derived. Our second application is to propose an estimator for the number of signals based on noisy observations. Other than the existing approaches [see, e.g., Kritchman and Nadler (2008), Nadler (2010), Passemier and Yao (2012, 2014)], our method allows the covariance structure of the noise to be arbitrary.

The rest of the paper is organized as follows. First, in Section 2, the exact setting of the spiked Fisher matrix $S = S_2^{-1}S_1$ is introduced. Then in Section 3, we establish the phase transition phenomenon for the extreme eigenvalues of S: a displacement formula is found as well as the transition boundary is explicitly obtained. Next, CLTs for those outlier eigenvalues fluctuating around their limit (i.e., in the super-critical case) are established first in Section 4 for one group of sample eigenvalues corresponding to a same population spike, and then in Section 6 for all the groups jointly. Section 5 contains numerical illustrations that demonstrate the finite sample performance of our results. In Section 7, we develop in detail two applications in high-dimensional statistics. Proofs of the main theorems (Theorems 3.1 and 4.1) are included in Section 8 while some technical lemmas are grouped in the Appendix.

2. Spiked Fisher matrix and preliminary results. In what follows, we will assume that $\Sigma_2 = I_p$. This assumption does not lose any generality since the eigenvalues of the Fisher matrix $S = S_2^{-1}S_1$ are invariant under the transformation

$$S_1 \mapsto \Sigma_2^{-1/2} S_1 \Sigma_2^{-1/2}, \qquad S_2 \mapsto \Sigma_2^{-1/2} S_2 \Sigma_2^{-1/2}. \tag{2.1}$$

Also we will write Σ_p for Σ_1 to signify the dependence on the dimension p. Let

$$Z = (z_1, \dots, z_n) = (z_{ij})_{1\le i\le p, 1\le j\le n} \tag{2.2}$$

and

$$W = (w_1, \dots, w_m) = (w_{kl})_{1\le k\le p, 1\le l\le m} \tag{2.3}$$

be two independent arrays, with respective size $p \times n$ and $p \times m$, of independent real-valued random variables with mean 0 and variance 1. Now suppose we have two samples $\{z_i\}_{1\le i\le n}$ and $\{x_i = \Sigma_p^{1/2} w_i\}_{1\le i\le m}$, where $\{z_i\}$ and $\{w_i\}$ are given by (2.2) and (2.3), and Σ_p is a rank M (M is a fixed integer) perturbation of I_p, that is,

$$\Sigma_p = \begin{pmatrix} \Omega_M & 0 \\ 0 & I_{p-M} \end{pmatrix}. \tag{2.4}$$

Here, Ω_M is a $M \times M$ covariance matrix, containing k nonzero and nonunit eigenvalues (a_i), with multiplicity numbers (n_i) $(n_1 + \cdots + n_k = M)$. That is, Ω_M has the eigen-decomposition $U \operatorname{diag}(\underbrace{a_1, \dots, a_1}_{n_1}, \dots, \underbrace{a_k, \dots, a_k}_{n_k})U^*$, where U is a $M \times M$ orthogonal matrix.

The sample covariance matrices of the two observations $\{x_i\}$ and $\{z_i\}$ are

$$S_1 = \frac{1}{m}\sum_{l=1}^{m} x_l x_l^* = \frac{1}{m}XX^* = \Sigma_p^{1/2}\left(\frac{1}{m}WW^*\right)\Sigma_p^{1/2} \tag{2.5}$$

and

$$S_2 = \frac{1}{n}\sum_{j=1}^{n} z_j z_j^* = \frac{1}{n}ZZ^*, \tag{2.6}$$

respectively.

Throughout the paper, we consider an asymptotic regime of Marčenko–Pastur-type, that is,

$$\begin{aligned} &p \wedge n \wedge m \to \infty, \qquad y_p := p/n \to y \in (0, 1) \quad \text{and} \\ &c_p := p/m \to c > 0. \end{aligned} \tag{2.7}$$

Recall that the *empirical spectral distribution* (ESD) of a $p \times p$ matrix A with eigenvalues $\{\lambda_j\}$ is the distribution $p^{-1}\sum_{j=1}^{p}\delta_{\lambda_j}$ where δ_a denotes the Dirac mass at a. Since the total rank M generated by the k spikes is fixed, the ESD of S will have the same limit (LSD) as there were no spikes in Σ_p. This limiting spectral distribution, which is the celebrated Wachter distribution, has been known for a long time.

PROPOSITION 2.1. *For the Fisher matrix $S = S_2^{-1}S_1$ with the sample covariance matrices S_i's given in* (2.5)–(2.6), *assume that the dimension p and the two sample sizes n, m grow to infinity proportionally as in* (2.7). *Then almost surely, the ESD of S weakly converges to a deterministic distribution $F_{c,y}$ with a bounded support* $[\alpha, \beta]$ *and a density function given by*

$$f_{c,y}(x) = \begin{cases} \dfrac{(1-y)\sqrt{(\beta - x)(x-\alpha)}}{2\pi x(c+xy)}, & \text{when } \alpha \le x \le \beta, \\ 0, & \text{otherwise}, \end{cases} \tag{2.8}$$

where

$$\alpha = \left(\frac{1-\sqrt{c+y-cy}}{1-y}\right)^2 \quad \text{and} \quad \beta = \left(\frac{1+\sqrt{c+y-cy}}{1-y}\right)^2. \tag{2.9}$$

Furthermore, if $c > 1$, then $F_{c,y}$ has a point mass $1 - 1/c$ at the origin. Also, the Stieltjes transform $s(z)$ of $F_{c,y}$ equals

$$\begin{aligned} s(z) &= \frac{1}{zc} - \frac{1}{z} \\ &\quad - \frac{c(z(1-y)+1-c)+2zy - c\sqrt{(1-c+z(1-y))^2 - 4z}}{2zc(c+zy)}, \\ &\qquad z \notin [\alpha, \beta]. \end{aligned} \tag{2.10}$$

REMARK 2.1. Assuming both populations are Gaussian, Wachter (1980), Theorem 3.1, derives the limiting distribution for roots of the determinental equation,

$$\left|mS_1 - x^2(mS_1 + nS_2)\right| = 0, \qquad x \in \mathbb{R}.$$

The continuous component of the distribution has a compact support $[A^2, B^2]$ with density function proportional to $\{(x - A^2)(B^2 - x)\}^{1/2}/\{x(1 - x^2)\}$. It can be readily checked that by the change of variable $z = cx^2/\{y(1 - x^2)\}$, the density of the continuous component of the LSD of S is exactly (2.8). The validity of this limit for general populations (nonnecessarily Gaussian) is due to Silverstein (1985) and Bai, Yin and Krishnaiah (1987).

3. Phase transition of the extreme eigenvalues of $S = S_2^{-1}S_1$. In this section, we establish a phase transition phenomenon for the extreme eigenvalues of $S = S_2^{-1}S_1$, that is, when a population spike a_i with multiplicity n_i is larger (or smaller) than a critical value, a packet of n_i corresponding sample eigenvalues of S will jump outside the support $[\alpha, \beta]$ of its LSD $F_{c,y}$ and converge all to a fixed limit $\phi(a_i)$, which is called the displacement of the population spike a_i. Otherwise, these associated sample eigenvalues will converge to one of the edges α and β.

By assumption, the k population spike eigenvalues $\{a_i\}$ are all positive and nonunit. We order them with their multiplicities in descending order together with the $p - M$ unit eigenvalues as

$$\begin{aligned} &a_1 = \cdots = a_1 > a_2 = \cdots = a_2 > \cdots > a_{k_0} = \cdots = a_{k_0} > 1 = \cdots = 1 \\ &\quad > a_{k_0+1} = \cdots = a_{k_0+1} > \cdots > a_k = \cdots = a_k. \end{aligned} \tag{3.1}$$

That is, k_0 of these population spike eigenvalues are larger than 1 while the other $k - k_0$ are smaller. Let

$$J_i = \begin{cases} [n_1 + \cdots + n_{i-1} + 1, n_1 + \cdots + n_i], & 1 \le i \le k_0, \\ [p - (n_i + \cdots + n_k) + 1, p - (n_{i+1} + \cdots + n_k)], & k_0 < i \le k. \end{cases}$$

Notice that the cardinality of each J_i is n_i. Next, the sample eigenvalues $\{l_{p,j}\}$ of the Fisher matrix $S_2^{-1}S_1$ are also sorted in the descending order as $l_{p,1} \ge l_{p,2} \ge \cdots \ge l_{p,p}$. Therefore, for each spike eigenvalue a_i, there are n_i associated sample eigenvalues $\{l_{p,j}, j \in J_i\}$. The phase transition for these extreme eigenvalues is given in the following Theorem 3.1.

THEOREM 3.1. *For the Fisher matrix $S = S_2^{-1}S_1$ with the sample covariance matrices S_i's given in* (2.5)–(2.6), *assume that the dimension p and the two sample sizes n, m grow to infinity proportionally as in* (2.7). *Then for any spike eigenvalue*

a_i $(i = 1, \dots, k)$, *it holds that for all* $j \in J_i$, $l_{p,j}$ *almost surely converges to a limit*

$$\lambda_i = \begin{cases} \phi(a_i), & |a_i - \gamma| > \gamma\sqrt{c + y - cy}, \\ \beta, & 1 < a_i \le \gamma\{1 + \sqrt{c + y - cy}\}, \\ \alpha, & \gamma\{1 - \sqrt{c + y - cy}\} \le a_i < 1, \end{cases} \tag{3.2}$$

where $\gamma := 1/(1 - y) \in (1, \infty)$ *and*

$$\phi(a_i) = \frac{a_i(a_i + c - 1)}{a_i - a_i y - 1} \tag{3.3}$$

is the displacement of the population spike a_i.

The proof of this Theorem is postponed to Section 8.1.

REMARK 3.1. Theorem 3.1 states that when the population spike a_i is large enough ($a_i > \gamma\{1 + \sqrt{c + y - cy}\}$) or small enough ($a_i < \gamma\{1 - \sqrt{c + y - cy}\}$), the corresponding extreme sample eigenvalues of the spiked Fisher matrix will converge to $\phi(a_i)$, which is located outside the support $[\alpha, \beta]$ of its LSD. Otherwise, they converge to one of its edges α and β. This phenomenon is depicted in Figure 1 for understanding.

REMARK 3.2. Using the notation $\gamma = 1/(1 - y)$, the function $\phi(x)$ in (3.3) could be expressed as

$$\phi(x) = \frac{\gamma x(x - 1 + c)}{x - \gamma}, \qquad x \neq \gamma, \tag{3.4}$$

which is a rational function with a single pole at $x = \gamma$. And the function asymptotically equals to $g(x) = \gamma(x + c - 1 + \gamma)$ when $|x| \to \infty$. On the other hand, since $\phi(\gamma\{1 - \sqrt{c + y - cy}\}) = \alpha$ and $\phi(\gamma\{1 + \sqrt{c + y - cy}\}) = \beta$, it can be checked that the points $A(\gamma\{1 - \sqrt{c + y - cy}, \alpha\})$ and $B(\gamma\{1 + \sqrt{c + y - cy}, \beta\})$ are exactly the two extreme points for the function ϕ. An example of $\phi(x)$ with parameters $(c, y) = (\frac{1}{5}, \frac{1}{2})$ is illustrated in Figure 2.

REMARK 3.3. It is worth observing that when $y \to 0$, the $\phi(x)$ function tends to the function well known in the literature for similar transition phenomenon of a spiked sample covariance matrix, that is,

$$\lim_{y \to 0} \phi(x) = x + \frac{cx}{x - 1}, \qquad x \neq 1, \tag{3.5}$$

see, for example, the ψ-function in Figure 4 of Bai and Yao (2012). These functions [(3.4) and (3.5)] share a same shape; however, the pole here equals 1, which is smaller than the pole $\gamma = 1/(1 - y)$ [in (3.4)] for the case of a spiked Fisher matrix.

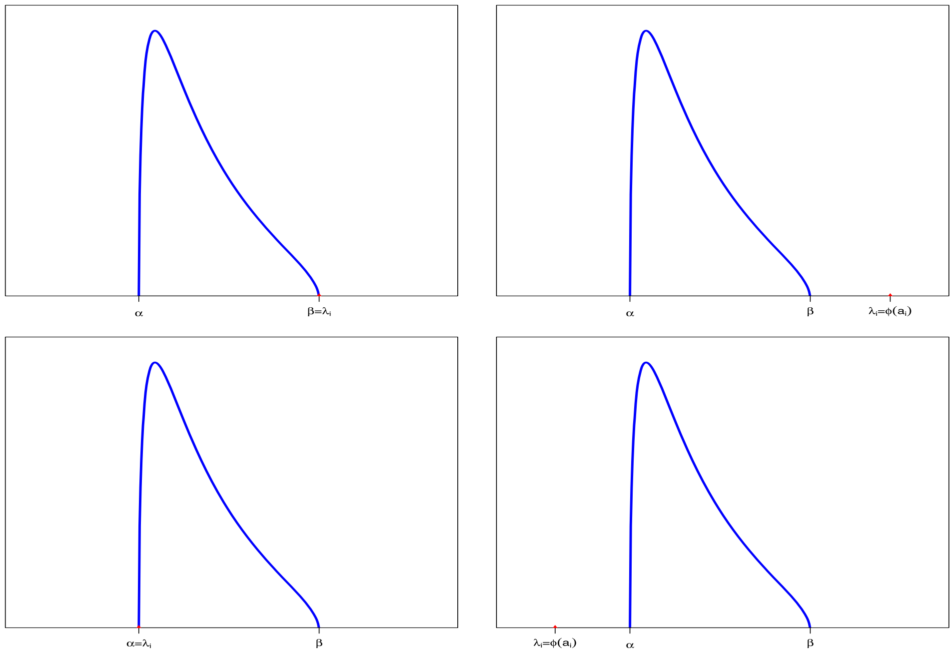

FIG. 1. *Phase transition of the extreme eigenvalues of the spiked Fisher matrix*: *upper-left panel*: *when* $1 < a_i \le \gamma\{1 + \sqrt{c + y - cy}\}$, *the limit of the corresponding extreme sample eigenvalue* $\{l_{p,j}, j \in J_i\}$ *is* β; *upper-right panel*: *when* $a_i > \gamma\{1 + \sqrt{c + y - cy}\}$, *the limit of* $\{l_{p,j}, j \in J_i\}$ *is larger than* β [*located at* $\lambda_i = \phi(\alpha_i)$]; *lower-left panel*: *when* $\gamma\{1 - \sqrt{c + y - cy}\} \le a_i < 1$, *the limit of* $\{l_{p,j}, j \in J_i\}$ *is* α; *lower-right panel*: *when* $0 < a_i < \gamma\{1 - \sqrt{c + y - cy}\}$, *the limit of* $\{l_{p,j}, j \in J_i\}$ *is smaller than* α [*located at* $\lambda_i = \phi(\alpha_i)$].

REMARK 3.4. As said in the Introduction, this phase transition phenomenon has already been established in a preprint Dharmawansa, Johnstone and Onatski (2014) (their Proposition 5) under Gaussian assumption and using a completely different approach. Theorem 3.1 proves that such a phase transition phenomenon is indeed universal.

4. Central limit theorem for the outlier eigenvalues of $S_2^{-1}S_1$. The aim of this section is to give a CLT for the n_i-packed outlier eigenvalues:

$$\sqrt{p}\{l_{p,j} - \phi(a_i), j \in J_i\}.$$

Denote $U = (U_1 \quad U_2 \quad \cdots \quad U_k)$, where each U_i is a matrix of size $M \times n_i$ that corresponds to the spike eigenvalue a_i.

THEOREM 4.1. *Assume the same assumptions as in Theorem* 3.1 *and in addition*, *the variables* (z_{ij}) [*in* (2.2)] *and* (w_{kl}) [*in* (2.3)] *have the same first four moments and denote* v_4 *as their common fourth moment*:

$$v_4 = \mathbb{E}|z_{ij}|^4 = \mathbb{E}|w_{kl}|^4, \qquad 1 \le i, k \le p, 1 \le j \le n, 1 \le l \le m.$$

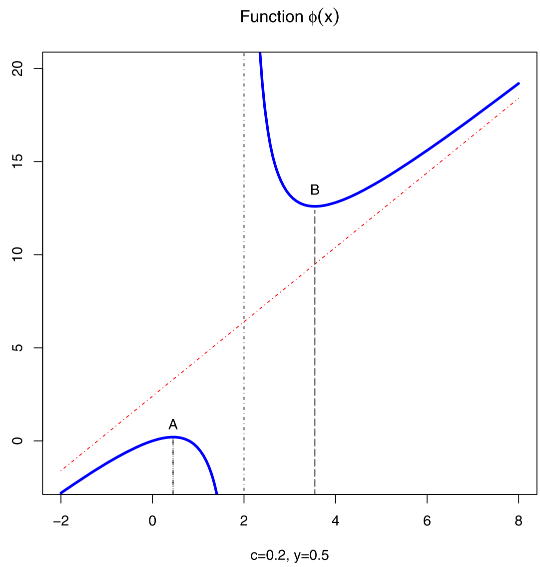

FIG. 2. *Example of the function* $\phi(x)$ *with* $(c, y) = (\frac{1}{5}, \frac{1}{2})$. *Its pole is at* $x = 2$. *When* $|x| \to \infty$, $\phi(x)$ *is getting close to the equation* $g(x) = 2x + \frac{12}{5}$ *(see the red line). The two extreme points are at* $A(0.450, 0.203)$ *and* $B(3.549, 12.597)$, *meaning that critical values for spikes are* 0.450 *and* 3.549 *while the support of the LSD is* [0.203,12.597].

Then for any population spike a_i *satisfying* $|a_i - \gamma| > \gamma\sqrt{c + y - cy}$, *the normalized* n_i*-packed outlier eigenvalues of* $S_2^{-1}S_1$: $\sqrt{p}\{l_{p,j} - \phi(a_i), j \in J_i\}$ *converge weakly to the distribution of the eigenvalues of the random matrix* $-U_i^* R(\lambda_i)U_i/\Delta(\lambda_i)$. *Here,*

$$\Delta(\lambda_i) = \frac{(1 - a_i - c)(1 + a_i(y - 1))^2}{(a_i - 1)(-1 + 2a_i + c + a_i^2(y - 1))}, \tag{4.1}$$

$R(\lambda_i) = (R_{mn})$ *is a* $M \times M$ *symmetric random matrix, made with independent Gaussian entries of mean zero and variance*

$$\mathbb{V}\mathrm{ar}(R_{mn}) = \begin{cases} 2\theta_i + (v_4 - 3)\omega_i, & m = n, \\ \theta_i, & m \neq n, \end{cases} \tag{4.2}$$

where

$$\omega_i = \frac{a_i^2(a_i + c - 1)^2(c + y)}{(a_i - 1)^2}, \tag{4.3}$$

$$\theta_i = \frac{a_i^2(a_i + c - 1)^2(cy - c - y)}{-1 + 2a_i + c + a_i^2(y - 1)}. \tag{4.4}$$

The proof of this theorem is postponed to Section 8.2.

REMARK 4.1. Notice that the result above involves the ith block U_i of the eigen-matrix U. When the spike a_i is simple, U_i is unique up to its sign, then $U_i^* R(\lambda_i) U_i$ is uniquely determined. But when a_i has multiplicities greater than 1, U_i is not unique; actually, any rotation of U_i can be an eigenvector corresponding to a_i. But, according to Lemma A.1 in the Appendix, such a rotation will not affect the eigenvalues of the matrix $U_i^* R(\lambda_i) U_i$.

Next, we consider a special case where Ω_M is diagonal ($U = I_M$), with distinct eigenvalues a_i, that is, $M = k$ and $n_i = 1$ for all $1 \le i \le M$. Using the previous result of Theorem 4.1, it can be shown that after normalization, the outlier eigenvalues $l_{p,i}$ of $S_2^{-1} S_1$ are asymptotically Gaussian when $|a_i - \gamma| > \gamma\sqrt{c + y - cy}$.

PROPOSITION 4.1. *Under the same assumptions as in Theorem 3.1, with additional conditions that Ω_M is diagonal and all its eigenvalues a_i $(1 \le i \le M)$ are simple, we have when $|a_i - \gamma| > \gamma\sqrt{c + y - cy}$, the outlier eigenvalue $l_{p,i}$ of $S_2^{-1} S_1$ is asymptotically Gaussian*:

$$\sqrt{p}\left(l_{p,i} - \frac{a_i(a_i - 1 + c)}{a_i - 1 - a_i y}\right) \Longrightarrow N(0, \sigma_i^2),$$

where

$$\sigma_i^2 = \frac{2a_i^2(cy - c - y)(a_i - 1)^2(-1 + 2a_i + c + a_i^2(y - 1))}{(1 + a_i(y - 1))^4} + (v_4 - 3) \cdot \frac{a_i^2(c + y)(-1 + 2a_i + c + a_i^2(y - 1))^2}{(1 + a_i(y - 1))^4}.$$

PROOF. Under the above assumptions, the random matrix $-U_i^* R(\lambda_i) U_i$ reduces to $-R(\lambda_i)(i, i)$, which is a Gaussian random variable of mean zero and variance

$$2\theta_i + (v_4 - 3)\omega_i = \frac{2a_i^2(a_i + c - 1)^2(cy - c - y)}{-1 + 2a_i + c + a_i^2(y - 1)} + (v_4 - 3) \cdot \frac{a_i^2(a_i + c - 1)^2(c + y)}{(a_i - 1)^2}.$$

Therefore, combining with the value of $\delta(\lambda_i)$ in (4.1) we have

$$\sqrt{p}\left(l_{p,i} - \frac{a_i(a_i - 1 + c)}{a_i - 1 - a_i y}\right) \Longrightarrow N(0, \sigma_i^2),$$

where

$$\sigma_i^2 = \frac{2a_i^2(cy - c - y)(a_i - 1)^2(-1 + 2a_i + c + a_i^2(y - 1))}{(1 + a_i(y - 1))^4} + (v_4 - 3) \cdot \frac{a_i^2(c + y)(-1 + 2a_i + c + a_i^2(y - 1))^2}{(1 + a_i(y - 1))^4}.$$

The proof of Proposition 4.1 is complete. □

REMARK 4.2. Notice that when the observations are standard Gaussian, we have $v_4 = 3$, then the above theorem reduces to

$$\sqrt{p}\left(l_{p,i} - \frac{a_i(a_i - 1 + c)}{a_i - 1 - a_i y}\right) \Longrightarrow N\left(0, \frac{2a_i^2(a_i - 1)^2(cy - c - y)(-1 + 2a_i + c + a_i^2(y - 1))}{(1 + a_i(y - 1))^4}\right),$$

which is exactly the result in Dharmawansa, Johnstone and Onatski (2014); see setting 1 in their Proposition 11.

5. Numerical illustrations. In this section, numerical results are provided to illustrate the results of our Theorem 4.1 and Proposition 4.1. We fix $p = 200$, $T = 1000$, $n = 400$ with 1000 replications, thus $y = 1/2$ and $c = 1/5$. The critical interval is then $[\gamma - \gamma\sqrt{c + y - cy}, \gamma + \gamma\sqrt{c + y - cy}] = [0.45, 3.55]$ and the limiting support $[\alpha, \beta] = [0.2, 12.6]$. Consider $k = 3$ spike eigenvalues $(a_1, a_2, a_3) = (20, 0.2, 0.1)$ with respective multiplicity $(n_1, n_2, n_3) = (1, 2, 1)$. Let $l_1 \geq \cdots \geq l_p$ be the ordered eigenvalues of the Fisher matrix $S_2^{-1}S_1$. We are particularly interested in the distributions of l_1, (l_{p-2}, l_{p-1}) and l_p, which corresponds to the spike eigenvalues a_1, a_2 and a_3, respectively.

5.1. *Case of $U = I_4$.* In this subsection, we consider a simple case that $U = I_4$. Therefore, following Theorem 4.1, we have:

- for $j = 1, p$, $\sqrt{p}\{l_j - \phi(a_i)\} \to N(0, \sigma_i^2)$. Here, for $j = 1, i = 1, \phi(a_1) = 42.67$ and $\sigma_1^2 = 4246.8 + 1103.5(v_4 - 3)$; and for $j = p$, $i = 3$, $\phi(a_3) = 0.07$ and $\sigma_3^2 = 7.2 \times 10^{-3} + 3.15 \times 10^{-3}(v_4 - 3)$;
- for $j = p - 2, p - 1$ and $i = 2$, the two-dimensional random vector $\sqrt{p}\{l_j - \phi(a_2)\}$ converges to the eigenvalues of the random matrix $-\frac{R_{mn}}{\Delta(\lambda_2)}$. Here, $\phi(a_2) = 0.13$, $\Delta(\lambda_2) = 1.45$ and R_{mn} is the 2×2 symmetric random matrix, made with independent Gaussian entries of mean zero and variance given by

$$\mathbb{V}\text{ar}(R_{mn}) = \begin{cases} 2\theta_2 + (v_4 - 3)\omega_2(= 0.04 + 0.016(v_4 - 3)), & m = n, \\ \theta_2(= 0.02), & m \neq n. \end{cases} \tag{5.1}$$

Simulations are conducted to compare the distributions of the empirical extreme eigenvalues with their limits.

5.1.1. *Gaussian case.* First, we assume all the z_{ij} and w_{ij} are i.i.d. standard Gaussian, thus $v_4 - 3 = 0$. And according to (5.1), $R_{mn}/\sqrt{0.04}$ is the standard 2×2 Gaussian Wigner matrix (GOE). Therefore, we have:

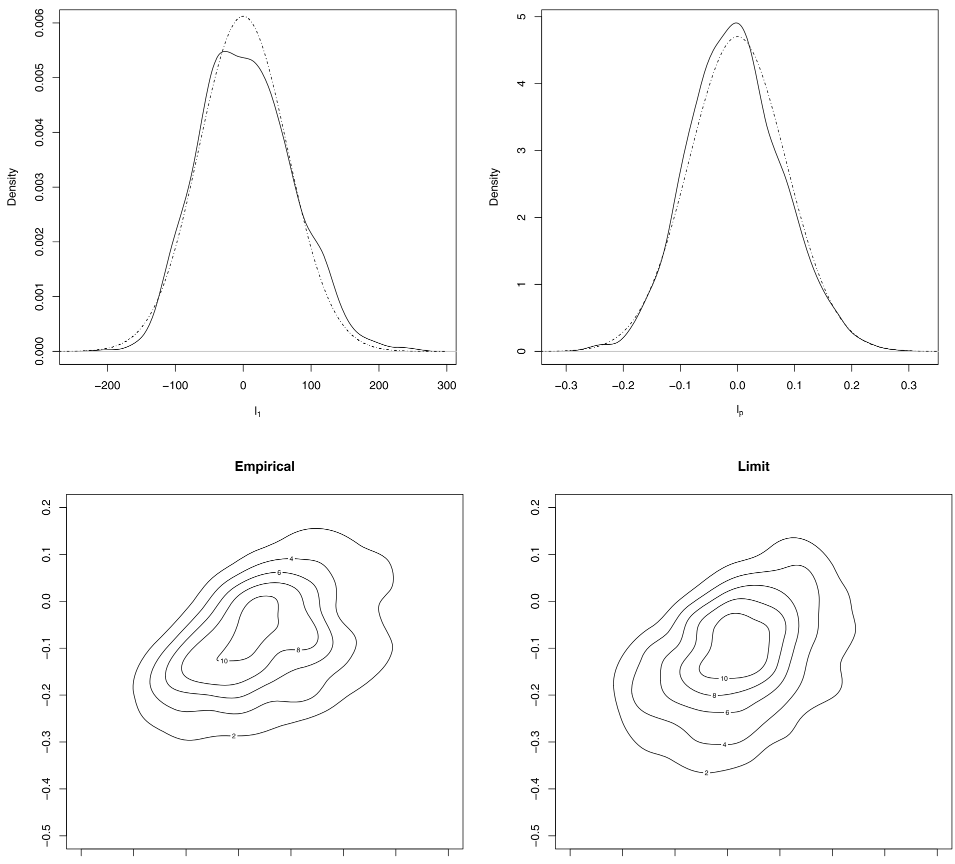

FIG. 3. *Upper panels show the empirical densities of l_1 and l_p (solid lines, after centralization and scaling) compared to their Gaussian limits (dashed lines). Lower panels show contour plots of empirical joint density function of (l_{p-2}, l_{p-1}) (left plot, after centralization and scaling) and contour plots of their limits (right plot). Both the empirical and limit joint density functions are displayed using the two-dimensional kernel density estimates. Samples are draw from i.i.d. standard Gaussian distribution with $U = I_4$. The replication number is* 1000.

- $\sqrt{p}\{l_1 - 42.67\} \to N(0, 4246.8)$,
- $\sqrt{p}\{l_p - 0.07\} \to N(0, 7.2 \times 10^{-3})$,
- the two-dimensional random vector $\sqrt{p}\{l_{p-2} - 0.13, l_{p-1} - 0.13\}$ converges to the eigenvalues of the random matrix $-0.138 \cdot W$; here, W is a 2×2 GOE.

We compare the empirical distributions with their limits in Figure 3. The upper panels show the empirical kernel density estimates (in solid lines) of $\sqrt{p}\{l_1 - 42.67\}$ and $\sqrt{p}\{l_p - 0.07\}$ from 1000 independent replications, compared to their Gaussian limits $N(0, 4246.8)$ and $N(0, 7.2 \times 10^{-3})$, respectively (dashed lines). When considering the empirical distribution of the two-dimensional random vector $\sqrt{p}\{l_{p-2} - 0.13, l_{p-1} - 0.13\}$, we run the two-dimensional kernel density estimation from 1000 independent replications and display their contour lines (see the lower-left panel of the figure), while the lower-right panel shows the contour lines of the kernel density estimation of the eigenvalues of the 2×2 random matrix $-0.138 \cdot GOE$ (their limits).

5.1.2. *Binary case.* Second, we assume all the z_{ij} and w_{ij} are i.i.d. binary variables taking values $\{1, -1\}$ with probability $1/2$, and in this case we have $v_4 = 1$. Similarly, we have:

- $\sqrt{p}\{l_1 - 42.67\} \to N(0, 2039.8)$,
- $\sqrt{p}\{l_p - 0.07\} \to N(0, 9 \times 10^{-4})$,
- the two-dimensional random vector $\sqrt{p}\{l_{p-2} - 0.13, l_{p-1} - 0.13\}$ converges to the eigenvalues of the random matrix $-R_{mn}/1.45$. Here, R_{mn} is the 2×2 symmetric random matrix, made with independent Gaussian entries of mean zero and variance

$$\mathbb{V}\mathrm{ar}(R_{mn}) = \begin{cases} 0.008, & m = n, \\ 0.02, & m \neq n. \end{cases}$$

Figure 4 compares the empirical distributions with their limits in this binary case. The upper panels show the empirical kernel density estimates of $\sqrt{p}\{l_1 - 42.67\}$ and $\sqrt{p}\{l_p - 0.07\}$ from 1000 independent replications (in solid lines), compared to their Gaussian limits (in dashed lines). Also, the lower panel shows the contour lines of the empirical joint density of the $\sqrt{p}\{l_{p-2} - 0.13, l_{p-1} - 0.13\}$ (the left plot), with the right plot displaying the contour lines of their limit.

5.2. *Case of general U.* In this subsection, we consider the following nonunit orthogonal matrix:

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}, \tag{5.2}$$

that is, we have

$$U_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad U_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad U_3 = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}.$$

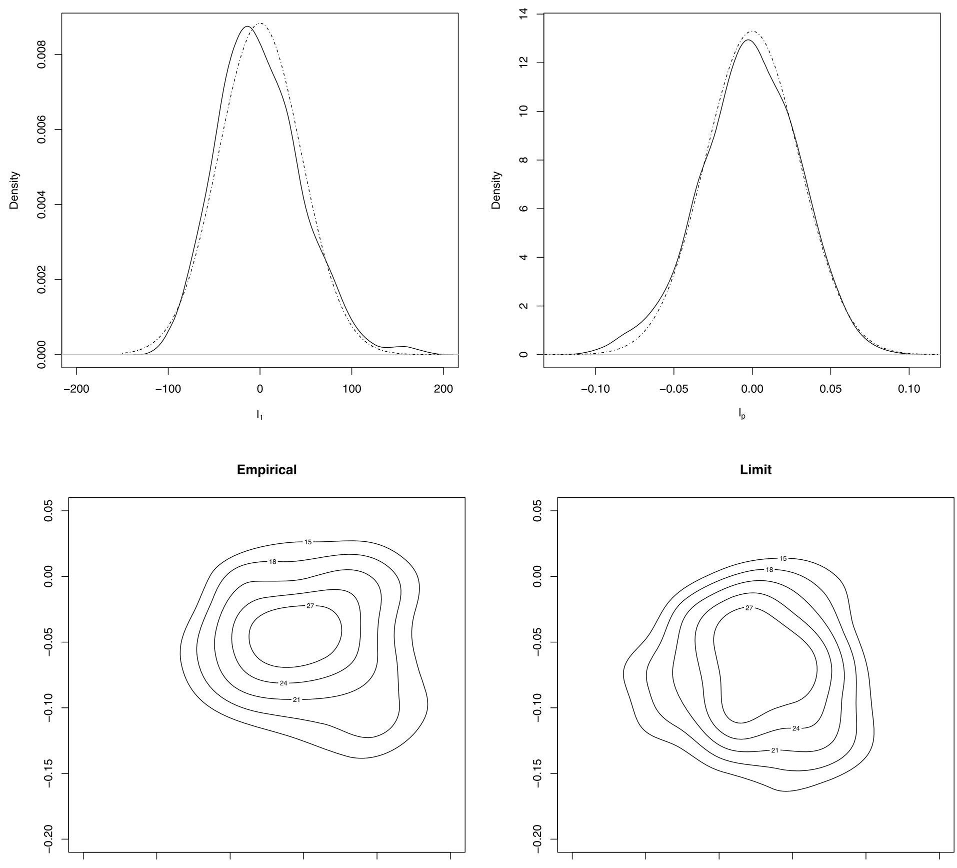

FIG. 4. *Upper panels show the empirical densities of l_1 and l_p (solid lines, after centralization and scaling) compared to their Gaussian limits (dashed lines). Lower panels show contour plots of empirical joint density function of (l_{p-2}, l_{p-1}) (left plot, after centralization and scaling) and contour plots of their limits (right plot). Both the empirical and limit joint density functions are displayed using the two-dimensional kernel density estimates. Samples are draw from i.i.d. binary distribution with $U = I_4$. The replication number is* 1000.

Since Gaussian distribution is invariant under orthogonal transformation, we only consider the case that all the z_{ij} and w_{ij} are i.i.d. binary variables taking values $\{1, -1\}$ with probability $1/2$, with all the other settings the same as in Section 5.1. Then according to Theorem 4.1, we have:

- $\sqrt{p}\{l_1 - 42.67\} \to N(0, 2039.8)$,
- $\sqrt{p}\{l_p - 0.07\} \to N(0, 0.004)$,
- the two-dimensional random vector $\sqrt{p}\{l_{p-2} - 0.13, l_{p-1} - 0.13\}$ converges to the eigenvalues of the random matrix $-U_2^* R(\lambda_2) U_2 / 1.45$. Here, $R(\lambda_2)$ is the 4×4 symmetric random matrix, made with independent Gaussian entries of

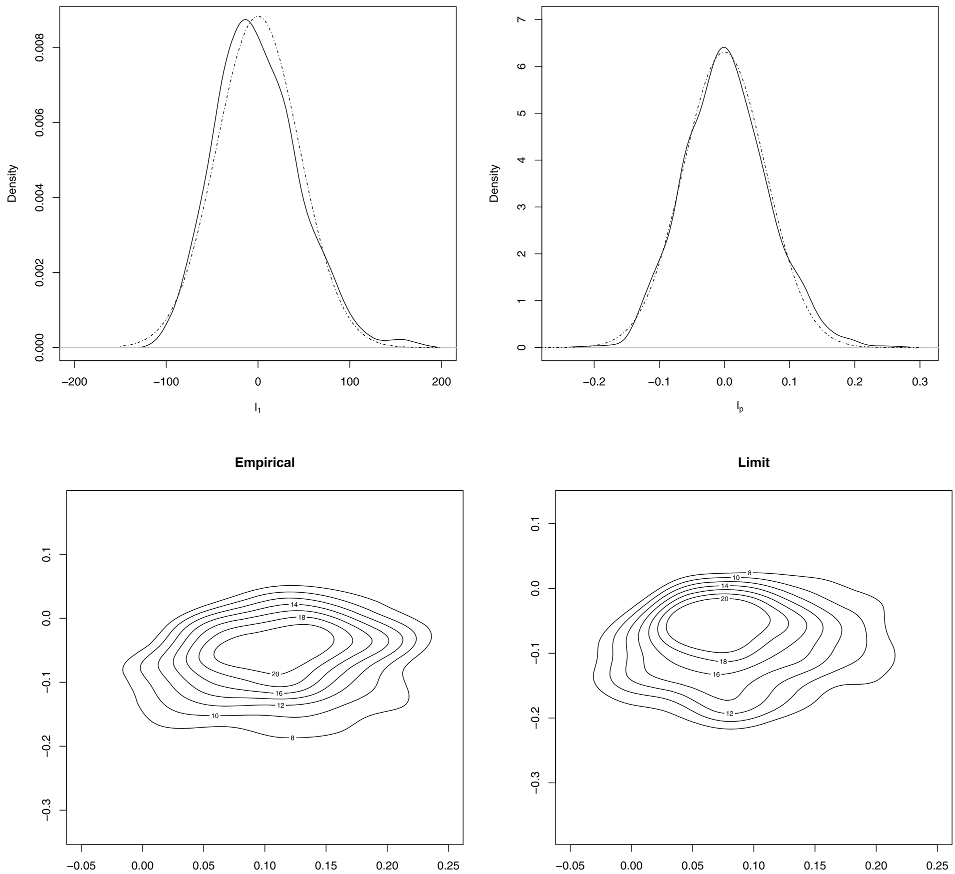

FIG. 5. *Upper panels show the empirical densities of l_1 and l_p (solid lines, after centralization and scaling) compared to their Gaussian limits (dashed lines). Lower panels show contour plots of empirical joint density function of (l_{p-2}, l_{p-1}) (left plot, after centralization and scaling) and contour plots of their limits (right plot). Both the empirical and limit joint density functions are displayed using the two-dimensional kernel density estimates. Samples are from i.i.d. binary distribution with U given by (5.2). The replication number is* 1000.

mean zero and variance

$$\mathbb{V}\mathrm{ar}(R_{mn}) = \begin{cases} 0.008, & m = n, \\ 0.02, & m \neq n. \end{cases}$$

Figure 5 compares the empirical distributions with their limits in this general U case. The upper panels show the empirical kernel density estimates of $\sqrt{p}\{l_1 - 42.67\}$ and $\sqrt{p}\{l_p - 0.07\}$ from 1000 independent replications (in solid lines), compared to their Gaussian limits (in dashed lines). Also, the lower panel of the figure shows the contour lines of the empirical joint density of $\sqrt{p}\{l_{p-2} -$

$0.13, l_{p-1} - 0.13\}$ (the lower-left plot), with the lower-right plot showing the contour lines of their limit.

6. Joint distribution of the outlier eigenvalues. In the previous section, we have obtained the following result for the outlier eigenvalues: the n_i-dimensional real random vector $\sqrt{p}\{l_{p,j} - \lambda_i, j \in J_i\}$ converges to the distribution of the eigenvalues of a random matrix $-U_i^* R(\lambda_i) U_i / \Delta(\lambda_i)$. It is in fact possible to derive their joint distribution, that is, the limit of the M-dimensional real random vector

$$\begin{pmatrix} \sqrt{p}\{l_{p,j_1} - \lambda_1, j_1 \in J_1\} \\ \vdots \\ \sqrt{p}\{l_{p,j_k} - \lambda_k, j_k \in J_k\} \end{pmatrix} \tag{6.1}$$

if all the spike eigenvalues a_i are above (or below) the phase transition threshold. Such joint convergence result is useful for inference procedures where consecutive sample eigenvalues are used such as their differences or ratios; see, for example, Onatski (2009) and Passemier and Yao (2014).

THEOREM 6.1. *Assume the same conditions as in Theorem 4.1 holds and all the population spikes a_i satisfy the condition $|a_i - \gamma| > \gamma\sqrt{c + y - cy}$. Then the M-dimensional random vector in (6.1) converges in distribution to the eigenvalues of the following $M \times M$ random matrix*:

$$\begin{pmatrix} \frac{-U_1^* R(\lambda_1) U_1}{\Delta(\lambda_1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{-U_k^* R(\lambda_k) U_k}{\Delta(\lambda_k)} \end{pmatrix}, \tag{6.2}$$

where the matrices $\{R(\lambda_i)\}$ are made with zero-mean independent Gaussian random variables, with the following covariance function between different blocks $(l \neq s)$: for $1 \le i \le j \le M$:

$$\mathrm{Cov}\big(R(\lambda_l)(i,j), R(\lambda_s)(i,j)\big) = \begin{cases} \theta(l,s), & i \neq j, \\ \omega(l,s)(v_4 - 3) + 2\theta(l,s), & i = j, \end{cases}$$

where

$$\theta(l,s) = \lim \frac{1}{n+T} \operatorname{tr} A_n(\lambda_l) A_n(\lambda_s),$$

$$\omega(l,s) = \lim \frac{1}{n+T} \sum_{i=1}^{n+T} A_n(\lambda_l)(i,i) A_n(\lambda_s)(i,i),$$

and $A_n(\lambda)$ is defined in (A.17).

The proof of this theorem is very close to that of Theorem 2.3 in Wang, Su and Yao (2014), thus omitted.

In principle, the limiting parameters $\theta(l,s)$ and $\omega(l,s)$ can be completely specified for a given spiked structure. However, this will lead to quite complex formula. Here, we prefer explaining a simple case where Ω_M is diagonal with simple eigenvalues (a_i), all satisfying the condition: $|a_i-\gamma|>\gamma\sqrt{c+y-cy}$ $(i=1,\dots,M)$. Therefore, $U_i^* R(\lambda_i)U_i$ in (6.2) reduces to the (i,i)th element of $R(\lambda_i)$, which is a Gaussian random variable. Besides, from Theorem 6.1, we see that the random variables $\{R(\lambda_i)(i,i)\}_{i=1,\dots,M}$ are jointly independent since the index sets (i,i) are disjoint. Finally, we have the following joint distribution of the M outlier eigenvalues of $S_2^{-1}S_1$.

PROPOSITION 6.1. *Under the same assumptions as in Theorem 4.1, then if Ω_M is diagonal with all its eigenvalues (a_i) being simple, satisfying: $|a_i-\gamma|>\gamma\sqrt{c+y-cy}$, then the M outlier eigenvalues $l_{p,j}$ $(j=1,\dots,M)$ of $S_2^{-1}S_1$ are asymptotically independent, having the joint distribution as follows*:

$$\begin{pmatrix} \sqrt{p}(l_{p,1}-\lambda_1) \\ \vdots \\ \sqrt{p}(l_{p,M}-\lambda_M) \end{pmatrix} \Longrightarrow \mathcal{N}\left(\mathbf{0}_M, \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_M^2 \end{pmatrix}\right),$$

where

$$\sigma_i^2 = \frac{2a_i^2(cy-c-y)(a_i-1)^2(-1+2a_i+c+a_i^2(y-1))}{(1+a_i(y-1))^4} + (v_4-3)\cdot\frac{a_i^2(c+y)(-1+2a_i+c+a_i^2(y-1))^2}{(1+a_i(y-1))^4}.$$

7. Applications. In this section, we present two applications of our previous results Theorem 4.1 and Proposition 4.1 in the areas of high-dimensional hypothesis testing and signal detection.

7.1. *Application* 1: *Power of testing the equality between two high-dimensional covariance matrices.* Let $(x_i)_{1\le i\le m}$ and $(z_j)_{1\le j\le n}$ be two p-dimensional observations from populations Σ_1 and Σ_2. This subsection considers the high-dimensional hypothesis testing for the equality between Σ_1 and Σ_2 against a specific alternative, that is, the difference between Σ_1 and Σ_2 is a finite rank covariance matrix. Put it in another way, we are concerned about the following testing problem:

$$H_0:\quad \Sigma_1=\Sigma_2 \quad \text{vs.} \quad H_1:\quad \Sigma_1=\Sigma_2+\Delta, \tag{7.1}$$

where $\operatorname{rank}(\Delta)=M$ (here M is a finite integer).

There exists a wide literature on testing the equality between two covariance matrices. In the classical large sample asymptotics, early works can be found in text books like Muirhead (1982) and Anderson (1984), where the authors find the limit distribution to be χ^2 [with degrees of freedom $p(p+1)/2$] for the likelihood ratio statistic under the Gaussian assumption. In recent years, this testing problem has been reconsidered but in a different asymptotic regime, that is, both the dimension and the two sample sizes are allowed to grow to infinity together. For example, in Bai et al. (2009), the authors prove that in the asymptotic regime of Marčenko–Pastur-type, the limiting distribution of the likelihood ratio statistic is Gaussian under H_0. Li and Chen (2012) propose a test based on some U-statistic, and its limiting distribution is derived under both the null and the alternative hypotheses in the high-dimensional framework. Cai, Liu and Xia (2013) proposes a test statistic based on the elements of the two sample covariance matrices and both its limiting distribution under the null hypothesis and its power are studied. And it is shown that their statistic enjoys certain optimality and especially powerful against sparse alternatives.

In the following, we consider a statistic based on the largest eigenvalue of the Fisher matrix and it will be shown that it is powerful against spiked alternatives. Now denote the sample covariance matrices of the two populations to be

$$S_1 = \frac{1}{m}\sum_{j=1}^{m} x_j x_j^* = \frac{1}{m} X X^* \tag{7.2}$$

and

$$S_2 = \frac{1}{n}\sum_{j=1}^{n} z_j z_j^* = \frac{1}{n} Z Z^* \tag{7.3}$$

respectively. When p, m and n are all growing to infinity proportionally while M is a fixed integer, the empirical measure of the p eigenvalues of $S_2^{-1}S_1$ (for simplicity, we assume $p < n$) will be affected by a difference of order M/p which vanishes, so that its limit remains the same as in the null hypothesis, that is, the Wachter distribution (see Proposition 2.1). In other words, such global limit from all the eigenvalues of $S_2^{-1}S_1$ will be of little help for distinguishing the two hypotheses (7.1). It happens that the useful information to detect a small rank alternative is actually encoded in a few largest eigenvalues of $S_2^{-1}S_1$.

Now denote l_1 as the largest eigenvalue of $S_2^{-1}S_1$. Notice that the eigenvalues of $S_2^{-1}S_1$ are invariant under the transformation (2.1), so without lose of generality, we can assume that under H_0, it holds $\Sigma_1 = \Sigma_2 = I_p$. Then according to Han, Pan and Zhang (2016), we have

$$\frac{l_1 - \beta}{s_p} \Longrightarrow F_1,$$

where $s_p = \frac{1}{m}(\sqrt{m}+\sqrt{p})(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{p}})^{1/3}$, which is the order of $p^{-2/3}$ and F_1 denotes the type-1 Tracy–Widom distribution. Consequently, we adopt the following decision rule:

$$\text{Reject } H_0: \quad \text{if } l_1 > q_\alpha s_p + \beta, \tag{7.4}$$

where q_α is the upper quantile at level α of the Tracy–Widom distribution F_1:

$$F_1(q_\alpha, \infty) = \alpha.$$

Once the largest eigenvalue a_1 of Δ is above the critical value for phase transition, this test will be able to detect the alternative hypothesis with a power tending to one as the dimension tends to infinity.

THEOREM 7.1. *Under the asymptotic scheme set in* (2.7), *assume the largest eigenvalue* a_1 *of* Δ *is above the critical value* $\frac{1+\sqrt{c+y-cy}}{1-y}$. *Then the power function of the test procedure* (7.4) *equals to*

$$\mathit{Power} = 1 - \Phi\left(\frac{\sqrt{p}}{\sigma_1} s_p q_\alpha + \frac{\sqrt{p}}{\sigma_1}\left(\beta - \frac{a_1(a_1 - 1 + c)}{a_1 - 1 - a_1 y}\right)\right) + o(1),$$

which will finally tend to one as the dimension tends to infinity.

PROOF. Under the alternative H_1 and according to our Proposition 4.1, the asymptotic distribution for l_1 is Gaussian:

$$\sqrt{p}\left(l_1 - \frac{a_1(a_1 - 1 + c)}{a_1 - 1 - a_1 y}\right) \Longrightarrow N(0, \sigma_1^2).$$

Therefore, the power can be calculated as

$$\text{Power} = 1 - \Phi\left(\frac{\sqrt{p}}{\sigma_1} s_p q_\alpha + \frac{\sqrt{p}}{\sigma_1}\left(\beta - \frac{a_1(a_1 - 1 + c)}{a_1 - 1 - a_1 y}\right)\right) + o(1), \tag{7.5}$$

where Φ is the standard normal cumulative distribution function. Since the order of s_p is $p^{-2/3}$ when $p \to \infty$, the first term $\frac{\sqrt{p}}{\sigma_1} s_p q_\alpha \to 0$ and the second term $\frac{\sqrt{p}}{\sigma_1}(\beta - \frac{a_1(a_1-1+c)}{a_1-1-a_1 y}) \to -\infty$ [when $a_1 > \frac{1+\sqrt{c+y-cy}}{1-y}$, $\frac{a_1(a_1-1+c)}{a_1-1-a_1 y}$ is always larger than the right edge point β]. Therefore, we have the right-hand side of (7.5) tend to one for any pre-given α when $p \to \infty$. The proof of Theorem 7.1 is complete. □

REMARK 7.1. In Li and Chen (2012), the authors use an U-statistic $T_{m,n}$ to test the hypothesis $H_0 : \Sigma_1 = \Sigma_2$. And its power is shown to be

$$\Phi\left(-\mathscr{L}_{m,n}(\Sigma_1, \Sigma_2) z_\alpha + \frac{\operatorname{tr}\{(\Sigma_1 - \Sigma_2)^2\}}{\sigma_{m,n}}\right), \tag{7.6}$$

where z_α is the upper-α quantile of $N(0, 1)$ and

$$\mathscr{L}_{m,n}(\Sigma_1, \Sigma_2) = \sigma_{m,n}^{-1}\left\{\frac{2}{m}\operatorname{tr}(\Sigma_2^2) + \frac{2}{n}\operatorname{tr}(\Sigma_1^2)\right\},$$

$$\sigma_{m,n}^2 = \frac{4}{n^2}\{\operatorname{tr}(\Sigma_2^2)\}^2 + \frac{8}{n}\operatorname{tr}(\Sigma_2^2 - \Sigma_1\Sigma_2)^2 + \frac{4}{m^2}\{\operatorname{tr}(\Sigma_1^2)\}^2 + \frac{8}{m}\operatorname{tr}(\Sigma_1^2 - \Sigma_1\Sigma_2)^2 + \frac{8}{mn}\{\operatorname{tr}(\Sigma_1\Sigma_2)\}^2.$$

If we restrict it to the specific alternative as in (7.1), then all the three parameters $\mathscr{L}_{m,n}(\Sigma_1, \Sigma_2)$, $\operatorname{tr}\{(\Sigma_1 - \Sigma_2)^2\}$ and $\sigma_{m,n}$ in (7.6) are of constant order. Therefore, against an alternative hypothesis of spiked type (7.1), our procedure is more powerful.

7.2. *Application* 2: *Determine the number of signals*. In this subsection, we consider an application of our results in the field of signal detection, where the spiked Fisher matrix arises naturally.

In a signal detection equipment, records are of form

$$x_i = As_i + e_i, \qquad i = 1, \dots, m, \tag{7.7}$$

where x_i is p-dimensional observations, s_i is a $k \times 1$ low-dimensional signal ($k \ll p$) with unit covariance matrix, A a $p \times k$ *mixing* matrix, and (e_i) is an i.i.d. noise with covariance matrix Σ_2. Therefore, the covariance matrix of x_i can be considered as a k-dimensional (low rank) perturbation of Σ_2, denoted as Σ_p in the following. Notice that none of the quantities at the right-hand side of (7.7) is observed. One of the fundamental problems here is to estimate k, the number of signals present in the system, which is challenging when the dimension p is large, say has a comparable magnitude with the sample size m. When the noise has the simplest covariance structure, that is, $\Sigma_2 = \sigma_e^2 I_p$, this problem has been much investigated recently and several solutions are proposed; see, for example, Kritchman and Nadler (2008), Nadler (2010), Passemier and Yao (2012, 2014). However, the problem with an *arbitrary* noise covariance matrix Σ_2, say diagonal to simplify, remains unsolved in the large-dimensional context (to the best of our knowledge).

Nevertheless, there exists an astute engineering device where the system can be tuned in a signal-free environment, for example, in laboratory: that is, we can directly record a sequence of pure-noise observations z_j, $j = 1, \dots, n$, which have the same distribution as the (e_i) above. These signal-free records can then be used to whiten the observations (x_i) thanks to the invariant property in (2.1), which states that the eigenvalues of $S_2^{-1}S_1$ [S_1 and S_2 are same defined as in (7.2) and (7.3)] are in fact *independent of* Σ_2. Therefore, these eigenvalues can be thought as if $\Sigma_2 = I_p$, that is, $S_2^{-1}S_1$ becomes a spiked Fisher matrix as introduced in Section 2. This is actually the reason why the two sample procedure developed here can deal with an arbitrary covariance matrix of the noise while the existing one-sample procedures cannot.

Based on Theorem 3.1, we propose our estimator of the number of signals as the number of eigenvalues of $S_2^{-1}S_1$ that is larger than the right edge point of the support of its LSD:

$$\hat{k} = \max\{i : l_i \geq \beta + d_n\}, \tag{7.8}$$

where (d_n) is a sequence of vanishing constants.

THEOREM 7.2. *Assume all the spike eigenvalues a_i $(i = 1, \dots, k)$ satisfy $a_i > \gamma + \gamma\sqrt{c + y - cy}$. Let d_n be a sequence of positive numbers such that $\sqrt{p} \cdot d_n \to 0$ and $p^{2/3} \cdot d_n \to +\infty$ as $p \to +\infty$, then the estimator $\hat{k}$ in (7.8) is consistent, that is, $\hat{k} \to k$ in probability as $p \to +\infty$.*

PROOF. Since

$$\begin{aligned}\{\hat{k} = k\} &= \big\{k = \max\{i : l_i \geq \beta + d_n\}\big\} \\ &= \big\{\forall j \in \{1, \dots, k\}, l_j \geq \beta + d_n\big\} \cap \{l_{k+1} < \beta + d_n\},\end{aligned}$$

we have

$$\begin{aligned}\mathbb{P}\{\hat{k} = k\} &= \mathbb{P}\Big(\bigcap_{1 \leq j \leq k} \{l_j \geq \beta + d_n\} \cap \{l_{k+1} < \beta + d_n\}\Big) \\ &= 1 - \mathbb{P}\Big(\bigcup_{1 \leq j \leq k} \{l_j < \beta + d_n\} \cup \{l_{k+1} \geq \beta + d_n\}\Big) \\ &\geq 1 - \sum_{j=1}^{k} \mathbb{P}(l_j < \beta + d_n) - \mathbb{P}(l_{k+1} \geq \beta + d_n).\end{aligned} \tag{7.9}$$

For $j = 1, \dots, k$,

$$\begin{aligned}\mathbb{P}(l_j < \beta + d_n) &= \mathbb{P}\big(\sqrt{p}\big(l_j - \phi(a_j)\big) < \sqrt{p}\big(\beta + d_n - \phi(a_j)\big)\big) \\ &\to \mathbb{P}\big(\sqrt{p}\big(l_j - \phi(a_j)\big) < \sqrt{p}\big(\beta - \phi(a_j)\big)\big),\end{aligned} \tag{7.10}$$

which is due to the assumption that $\sqrt{p} \cdot d_n \to 0$. Then the part $\sqrt{p}(\beta - \phi(a_j))$ in (7.10) will tend to $-\infty$ since we have always $\phi(a_j) > \beta$ when $a_i > \gamma + \gamma\sqrt{c + y - cy}$. On the other hand, by Theorem 4.1, $\sqrt{p}(l_j - \phi(a_j))$ in (7.10) has a limiting distribution; it is then bounded in probability. Therefore, we have

$$P(l_j < \beta + d_n) \to 0 \qquad \text{for } j = 1, \dots, k. \tag{7.11}$$

Also

$$\mathbb{P}(l_{k+1} \geq \beta + d_n) = \mathbb{P}\big(p^{2/3}(l_{k+1} - \beta) \geq p^{2/3} \cdot d_n\big),$$

and the part $p^{2/3}(l_{k+1}-\beta)$ is asymptotically Tracy–Widom distributed [see Han, Pan and Zhang (2016)]. As $p^{2/3}\cdot d_n$ tend to infinity as assumed, we have

$$\mathbb{P}(l_{k+1}\geq \beta+d_n)=0. \tag{7.12}$$

Combine (7.9), (7.11) and (7.12), we have $\mathbb{P}\{\hat{k}=k\}\to 1$ as $p\to+\infty$. The proof of Theorem 7.2 is complete. □

REMARK 7.2. Notice here that there is no need for those spikes a_i to be simple. The only requirement is that they should be properly strong enough ($a_i>\gamma+\gamma\sqrt{c+y-cy}$) for detection.

In the following, we will conduct a short simulation to illustrate the performance of our estimator. For comparison, we also show the performance of another estimator $\bar{k}$ that treats the noise covariance as known (using a plug-in estimator for this quantity). Detailed illustrations are as follows. Recall the model in (7.7), where $\mathrm{Cov}(e_i)=\Sigma_2$ is arbitrary. Now assume for a moment that Σ_2 is known, then we can multiply both sides of (7.7) by $\Sigma_2^{-1/2}$:

$$\Sigma_2^{-1/2}x_i=\Sigma_2^{-1/2}As_i+\Sigma_2^{-1/2}e_i,\qquad i=1,\dots,m,$$

where the left-hand side is still observable (simply multiply the original observations $\{x_i\}$ by $\Sigma_2^{-1/2}$). Denote $\tilde{x}_i=\Sigma_2^{-1/2}x_i$ and $\tilde{e}_i=\Sigma_2^{-1/2}e_i$, then $\mathrm{Cov}(\tilde{e}_i)=I_p$. On the other hand, due to the fact that the rank of $\Sigma_2^{-1/2}As_i$ is still k, the covariance matrix of the new observation $\tilde{x}_i$ is then a rank k perturbation of I_p. Therefore, the method in Kritchman and Nadler (2008) can be adopted. Their proposed estimator is

$$\bar{k}=\max\{k:l_k>(1+\sqrt{c})^2+d_n\}. \tag{7.13}$$

Besides, the $\{l_k\}$ in (7.13) are the eigenvalues of the sample covariance matrix of the observation $\tilde{x}_i$:

$$\Sigma_2^{-1/2}\cdot\left(\frac{1}{m}XX^T\right)\cdot\Sigma_2^{-1/2},$$

whose eigenvalues are the same as those of $\Sigma_2^{-1}S_1$. Since Σ_2 is actually unknown, here we simply use its plug-in estimator S_2. Therefore, the estimator in (7.13) for comparison is then

$$\bar{k}=\max\{k:l_k(S_2^{-1}S_1)>(1+\sqrt{c})^2+d_n\}. \tag{7.14}$$

The parameters for the simulation is set as follows. We fix $y=0.1$, $c=0.9$ and the value of p varies from 50 to 250, therefore, the critical value for a_i in the model (2.4) (after whitening) is $a_i>\gamma\{1+\sqrt{c+y-cy}\}=2.17$. For each given pair of (p,n,m) (we take floor if the values of n or m are nonintegers), we repeat 1000 times. The tuning parameter d_n is chosen to be $\log p/p^{2/3}$.

Next, suppose $k = 3$ and A is a $p \times 3$ matrix of form $A = (\sqrt{c_1}v_1, \sqrt{c_2}v_2)$, where $c_1 = 10$, $c_2 = 5$,

$$v_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}^* \quad \text{and} \quad v_2 = \begin{pmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & \cdots & 0 \end{pmatrix}^* .$$

So we have two spike eigenvalues $c_1 = 10$, $c_2 = 5$ (before whitening) with multiplicity $n_1 = 1$, $n_2 = 2$, respectively.

Besides, assume $\mathrm{Cov}(s_i) = I_3$ and we run both the Gaussian (s_i is multivariate Gaussian) and non-Gaussian (each component of s_i is i.i.d., taking value 1 or -1 with equal probability) cases. Finally, we set e_i to be multivariate Gaussian distributed with covariance matrix $\mathrm{Cov}(e_i)$ either diagonal or nondiagonal as in the following two cases:

- Case 1: $\mathrm{Cov}(e_i) = \mathrm{diag}(\underbrace{1, \ldots, 1}_{p/2}, \underbrace{2, \ldots, 2}_{p/2})$. In this case, we have the three nonzero eigenvalues of $(c_1 v_1 v_1^* + c_2 v_2 v_2^*) \cdot [\mathrm{Cov}(e_i)]^{-1}$ equal 10, 5, 5, respectively, which are all larger than the critical value $2.17 - 1$, therefore, the number of detectable signals is three;
- Case 2: $\mathrm{Cov}(e_i)$ is compound symmetric with all the diagonal elements equal 1 and all the off-diagonal elements equal 0.1. In this case, we have for each given p, the three nonzero eigenvalues of $(c_1 v_1 v_1^* + c_2 v_2 v_2^*) \cdot [\mathrm{Cov}(e_i)]^{-1}$ are all larger than 5.36 ($> 2.17 - 1$). The number of detectable signals is again three.

Tables 1 and 2 report the empirical frequency of our estimator $\hat{k}$ in Case 1 and Case 2. For comparison, we also report the frequency of the plug-in estimator

TABLE 1

Frequency of our estimator and the plug-in estimator defined in (7.14) *for Case* 1

	Gaussian					**Non-Gaussian**				
p	**50**	**100**	**150**	**200**	**250**	**50**	**100**	**150**	**200**	**250**
$\hat{k} = 2$	0.029	0.001	0	0	0	0.011	0	0	0	0
$\hat{k} = 3$	**0.971**	**0.997**	**0.997**	**0.995**	**0.998**	**0.985**	**0.997**	**0.993**	**0.998**	**0.998**
$\hat{k} = 4$	0	0.002	0.003	0.005	0.002	0.004	0.003	0.007	0.002	0.002
$\bar{k} = 3$	**0.603**	0.037	0	0	0	**0.654**	0.051	0	0	0
$\bar{k} = 4$	0.387	**0.485**	0.03	0	0	0.334	**0.514**	0.026	0	0
$\bar{k} = 5$	0.01	0.439	0.375	0.016	0	0.012	0.394	0.392	0.009	0
$\bar{k} = 6$	0	0.039	**0.508**	0.194	0.008	0	0.041	**0.481**	0.253	0.002
$\bar{k} = 7$	0	0	0.084	**0.566**	0.125	0	0	0.096	**0.56**	0.108
$\bar{k} = 8$	0	0	0.003	0.204	**0.463**	0	0	0.005	0.163	**0.518**
$\bar{k} = 9$	0	0	0	0.02	0.369	0	0	0	0.015	0.334
$\bar{k} = 10$	0	0	0	0	0.035	0	0	0	0	0.038

TABLE 2
Frequency of our estimator and the plug-in estimator defined in (7.14) for Case 2

	Gaussian					Non-Gaussian				
p	**50**	**100**	**150**	**200**	**250**	**50**	**100**	**150**	**200**	**250**
$\hat{k}=2$	0.018	0	0	0	0	0.003	0	0	0	0
$\hat{k}=3$	**0.982**	**0.995**	**0.996**	**0.995**	**0.998**	**0.993**	**0.997**	**0.993**	**0.998**	**0.998**
$\hat{k}=4$	0	0.005	0.004	0.005	0.002	0.004	0.003	0.007	0.002	0.002
$\bar{k}=3$	**0.6**	0.034	0	0	0	**0.644**	0.048	0.026	0	0
$\bar{k}=4$	0.39	**0.477**	0.03	0	0	0.345	**0.511**	0.382	0.008	0
$\bar{k}=5$	0.01	0.449	0.36	0.016	0	0.011	0.399	**0.491**	0.243	0
$\bar{k}=6$	0	0.04	**0.518**	0.193	0.007	0	0.042	0.096	**0.564**	0.002
$\bar{k}=7$	0	0	0.088	**0.559**	0.116	0	0	0.005	0.169	0.103
$\bar{k}=8$	0	0	0.004	0.207	**0.465**	0	0	0	0.016	**0.516**
$\bar{k}=9$	0	0	0	0.025	0.377	0	0	0	0	0.341
$\bar{k}=10$	0	0	0	0	0.035	0	0	0	0	0.038

defined in (7.14). According to our set up, the true number of signals is $k=3$. From these two tables, we see that the frequency of correct estimation of our estimator $\hat{k}$ ($\hat{k}=3$) is always around some value close to 1 in the two cases (both for Gaussian signal and non-Gaussian signal), which confirms the consistency of our estimator. While the plug-in estimator will always overestimate the number of signals in both cases. This overestimation phenomenon gets more and more striking when the value of p gets larger.

8. Proofs of the main results.

8.1. *Proof of Theorem* 3.1. For notation convenience, first we define some integrals with respect to $F_{c,y}(x)$ as follows: for a complex number $z \notin [\alpha, \beta]$,

$$\begin{aligned} s(z) &:= \int \frac{1}{x-z}\, dF_{c,y}(x), & m_1(z) &:= \int \frac{1}{(z-x)^2}\, dF_{c,y}(x), \\ m_2(z) &:= \int \frac{x}{z-x}\, dF_{c,y}(x), & m_3(z) &:= \int \frac{x}{(z-x)^2}\, dF_{c,y}(x), \\ m_4(z) &:= \int \frac{x^2}{(z-x)^2}\, dF_{c,y}(x). \end{aligned} \tag{8.1}$$

PROOF. The proof is divided into the following three steps:

- Step 1: we derive the almost sure limit of an outlier eigenvalue of $S_2^{-1}S_1$;

- Step 2: we show that in order for the extreme eigenvalue of $S_2^{-1}S_1$ to be an outlier, the population spike a_i should be larger (or smaller) than a critical value;
- Step 3: if not so, the extreme eigenvalue of $S_2^{-1}S_1$ will converge to one of the edge points α and β.

Step 1: Let $l_{p,j}(j \in J_i)$ be the outlier eigenvalue of $S_2^{-1}S_1$ corresponding to the population spike a_i. Then $l_{p,j}$ must satisfy the following equation:

$$|l_{p,j}I_p - S_2^{-1}S_1| = 0,$$

and it is equivalent to

$$|l_{p,j}S_2 - S_1| = 0. \tag{8.2}$$

Now we make some shorthand. Denote $Z = \binom{Z_1}{Z_2}$, where Z_1 is the n observations of its first M coordinates and Z_2 the remaining. We partition X accordingly as $X = \binom{X_1}{X_2}$, where X_1 is the m observations of its first M coordinates and X_2 the remaining. Using such a representation, we have

$$\begin{aligned} S_1 &= \frac{1}{m}XX^* = \frac{1}{m}\begin{pmatrix} X_1X_1^* & X_1X_2^* \\ X_2X_1^* & X_2X_2^* \end{pmatrix}, \\ S_2 &= \frac{1}{n}ZZ^* = \frac{1}{n}\begin{pmatrix} Z_1Z_1^* & Z_1Z_2^* \\ Z_2Z_1^* & Z_2Z_2^* \end{pmatrix}. \end{aligned} \tag{8.3}$$

Then (8.2) could be written in the block form:

$$\left|\begin{pmatrix} \frac{l_{p,j}}{n}Z_1Z_1^* - \frac{1}{m}X_1X_1^* & \frac{l_{p,j}}{n}Z_1Z_2^* - \frac{1}{m}X_1X_2^* \\ \frac{l_{p,j}}{n}Z_2Z_1^* - \frac{1}{m}X_2X_1^* & \frac{l_{p,j}}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^* \end{pmatrix}\right| = 0. \tag{8.4}$$

Since $l_{p,j}$ is an outlier, it holds $|l_{p,j} \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*| \neq 0$, and for block matrix, we have $\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det D \cdot \det(A - BD^{-1}C)$ when D is invertible. Therefore, (8.4) reduces to

$$\begin{aligned} &\bigg|\frac{l_{p,j}}{n}Z_1Z_1^* - \frac{1}{m}X_1X_1^* \\ &\quad - \left(\frac{l_{p,j}}{n}Z_1Z_2^* - \frac{1}{m}X_1X_2^*\right)\left(\frac{l_{p,j}}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1} \\ &\quad \times \left(\frac{l_{p,j}}{n}Z_2Z_1^* - \frac{1}{m}X_2X_1^*\right)\bigg| = 0. \end{aligned}$$

More specifically, we have

$$\det\Bigg(\underbrace{\frac{l_{p,j}}{n}Z_1\Big[I_n - Z_2^*\Big(l_{p,j}I_p - \Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2X_2^*\Big)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{l_{p,j}}{n}Z_2\Big]Z_1^*}_{(I)}$$

$$-\underbrace{\frac{1}{m}X_1\Big[I_m + X_2^*\Big(l_{p,j}I_p - \Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2X_2^*\Big)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2\Big]X_1^*}_{(II)}$$

$$+\underbrace{\frac{l_{p,j}}{n}Z_1Z_2^*\Big(l_{p,j}I_p - \Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2X_2^*\Big)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2X_1^*}_{(III)} \tag{8.5}$$

$$+\underbrace{\frac{1}{m}X_1X_2^*\Big(l_{p,j}I_p - \Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2X_2^*\Big)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{l_{p,j}}{n}Z_2Z_1^*}_{(IV)}\Bigg)$$

$$= 0.$$

In all the following, we denote by S the Fisher matrix $(\frac{1}{n}Z_2Z_2^*)^{-1}\frac{1}{m}X_2X_2^*$, which has a LSD $F_{c,y}(x)$. And in order to find the limit of $l_{p,j}$, we simply find the limit on the left-hand side of (8.5), then it will generate an equation. Solving this equation will give the value of its limit.

First, consider the terms (III) and (IV). Since (Z_1, X_1) is independent of (Z_2, X_2), using Lemma A.2, we see these two terms will converge to some constant multiplied by the covariance matrix between X_1 and Z_1. On the other hand, X_1 is also independent of Z_1, we have

$$\operatorname{Cov}(X_1, Z_1) = \mathbb{E}X_1Z_1 - \mathbb{E}X_1\mathbb{E}Z_1 = \mathbb{E}X_1\mathbb{E}Z_1 - \mathbb{E}X_1\mathbb{E}Z_1 = \mathbf{0}_{M\times M}.$$

Therefore, these two terms will both tend to a zero matrix $\mathbf{0}_{M\times M}$ almost surely.

So the remaining task is to find the limit of (I) and (II). We recall the expression of X_1 and Z_1 that

$$\operatorname{Cov}(X_1) = U\operatorname{diag}(\underbrace{a_1,\dots,a_1}_{n_1},\dots,\underbrace{a_k,\dots,a_k}_{n_k})U^*, \qquad \operatorname{Cov}(Z_1) = I_M.$$

According to Lemma A.2, we have

$$\begin{aligned}(I) &= \frac{l_{p,j}}{n}Z_1\Big[I_n - Z_2^*(l_{p,j}I_p - S)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{l_{p,j}}{n}Z_2\Big]Z_1^* \\ &\to \frac{\lambda_i}{n}\Big\{\mathbb{E}\operatorname{tr}\Big[I_n - Z_2^*(\lambda_iI_p - S)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{\lambda_i}{n}Z_2\Big]\Big\}\cdot I_M \\ &= \lambda_i\big(1 + y\lambda_i s(\lambda_i)\big)\cdot I_M,\end{aligned} \tag{8.6}$$

here, we denote λ_i as the limit of the outlier $\{l_{p,j}, j \in J_i\}$. For the same reason,

$$\begin{aligned}(II) &= -\frac{1}{m}X_1\Big[I_m + X_2^*(l_{p,j}I_p - S)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2\Big]X_1^* \\ &\to -\frac{1}{m}\Big\{\mathbb{E}\,\mathrm{tr}\Big[I_m + X_2^*(\lambda_i I_p - S)^{-1}\Big(\frac{1}{n}Z_2Z_2^*\Big)^{-1}\frac{1}{m}X_2\Big]\Big\} \\ &\quad \times U\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k\end{pmatrix}U^* \\ &= U\big(-1 + c + c\lambda_i s(\lambda_i)\big)\cdot\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k\end{pmatrix}U^*.\end{aligned} \tag{8.7}$$

Therefore, combining (8.5), (8.6) and (8.7), we have the determinant of the following $M \times M$ matrix:

$$U\begin{pmatrix}\lambda_i\big(1 + y\lambda_i s(\lambda_i)\big) + \big(-1 + c + c\lambda_i s(\lambda_i)\big)a_1 & & 0\\ \vdots & \ddots & \vdots \\ 0 & \lambda_i\big(1 + y\lambda_i s(\lambda_i)\big) + \big(-1 + c + c\lambda_i s(\lambda_i)\big)a_k\end{pmatrix}U^*$$

equal to zero, which is also to say that λ_i satisfies the equation:

$$\lambda_i\big(1 + y\lambda_i s(\lambda_i)\big) + \big(-1 + c + c\lambda_i s(\lambda_i)\big)a_i = 0. \tag{8.8}$$

Finally, together with the expression of the Stieltjes transform of a Fisher matrix in (2.10), we have

$$\lambda_i = \frac{a_i(a_i + c - 1)}{a_i - a_i y - 1} = \phi(a_i). \tag{8.9}$$

Step 2: Define $\underline{s}(z)$ as the Stieltjes transform of the LSD of $\frac{1}{m}X_2^*(\frac{1}{n}Z_2Z_2^*)^{-1}X_2$, who shares the same nonzero eigenvalues as $S_2^{-1}S_1$. Then we have the relationship:

$$\underline{s}(z) + \frac{1}{z}(1 - c) = cs(z). \tag{8.10}$$

Recall the expression of $s(z)$ in (2.10), we have

$$\underline{s}(z) = -\frac{c(z(1-y) + 1 - c) + 2zy - c\sqrt{(1 - c + z(1-y))^2 - 4z}}{2z(c + zy)}. \tag{8.11}$$

On the other hand, due to (8.8) and (8.10), we have the value for $\underline{s}(\lambda_i)$:

$$\underline{s}(\lambda_i) = \frac{yc - y - c}{y\lambda_i + a_i c}. \tag{8.12}$$

Since λ_i is outside the support of the LSD, we have

$$\underline{s}^{-1}\left(\frac{yc-y-c}{y\lambda_i+a_ic}\right)=\lambda_i>\beta \quad \text{or} \quad \underline{s}^{-1}\left(\frac{yc-y-c}{y\lambda_i+a_ic}\right)=\lambda_i<\alpha,$$

which is also to say that

$$\underline{s}(\beta)<\frac{yc-y-c}{y\lambda_i+a_ic} \tag{8.13}$$

or

$$\underline{s}(\alpha)>\frac{yc-y-c}{y\lambda_i+a_ic}. \tag{8.14}$$

Then (8.13) says that $\underline{s}(\beta)$ must be smaller than the minimum value on its right-hand side, whose minimum value is attained when $\lambda_i=\beta$ [the right-hand side of (8.13) is a decreasing function of λ_i]. Similarly, (8.14) says that $\underline{s}(\alpha)$ must be larger than the maximum value on its right-hand side, which is attained when $\lambda_i=\alpha$. Therefore, the condition for λ_i be an outlier is

$$\underline{s}(\beta)<\frac{yc-y-c}{y\beta+a_ic} \quad \text{or} \quad \underline{s}(\alpha)>\frac{yc-y-c}{y\alpha+a_ic}. \tag{8.15}$$

Finally, using (8.11) together with the value of α and β, we have

$$a_i>\frac{1+\sqrt{c+y-cy}}{1-y} \quad \text{or} \quad a_i<\frac{1-\sqrt{c+y-cy}}{1-y},$$

which is equivalent to say that [recall the expression of γ that $\gamma=1/(1-y)$] the condition to allow for the outlier is

$$|a_i-\gamma|>\gamma\sqrt{c+y-cy}.$$

Step 3: In this step, we show that if the condition in Step 2 is not fulfilled, then the extreme eigenvalues of $S_2^{-1}S_1$ will tend to one of the edge points α and β. For simplicity, we only show the convergence to the right edge β: the proof for the convergence to the left edge α is similar. Thus suppose all the $a_i>1$ for $i=1,\dots,k$. Let

$$S_1=\frac{1}{m}XX^*=\frac{1}{m}\begin{pmatrix}X_1X_1^* & X_1X_2^*\\ X_2X_1^* & X_2X_2^*\end{pmatrix}:=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix}$$

and

$$S_2=\frac{1}{n}ZZ^*=\frac{1}{n}\begin{pmatrix}Z_1Z_1^* & Z_1Z_2^*\\ Z_2Z_1^* & Z_2Z_2^*\end{pmatrix}:=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},$$

where B_{11} and A_{11} are the blocks of size $M\times M$. Using the inverse formula for block matrix, the $(p-M)\times(p-M)$ major sub-matrix of $S_2^{-1}S_1$ is

$$\begin{aligned}&-\left(A_{22}-A_{21}A_{11}^{-1}A_{12}\right)^{-1}A_{21}A_{11}^{-1}B_{12}+\left(A_{22}-A_{21}A_{11}^{-1}A_{12}\right)^{-1}B_{22}\\ &\quad:=C.\end{aligned} \tag{8.16}$$

The part

$$-(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}B_{12}$$
$$=-(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}\cdot\frac{1}{m}X_1X_2^*$$

is of rank M; besides, we have

$$\operatorname{tr}\left\{(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}\frac{1}{m}X_1X_2^*\right\}\to 0,$$

since X_1 is independent of X_2. Therefore, the M nonzero eigenvalues of the matrix $-(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}B_{12}$ will all tend to zero (so is its largest one). Then consider the second part of (8.16) as follows:

$$A_{22}-A_{21}A_{11}^{-1}A_{12}=\frac{1}{n}Z_2\left[I_n-Z_1^*\left(\frac{1}{n}Z_1Z_1^*\right)^{-1}\frac{1}{n}Z_1\right]Z_2^*:=\frac{1}{n}Z_2PZ_2.$$

Since $P=I_n-Z_1^*(\frac{1}{n}Z_1Z_1^*)^{-1}\frac{1}{n}Z_1$ is a projection matrix of rank $p-M$, it has the spectral decomposition:

$$P=V\begin{pmatrix}0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & I_{n-M}\end{pmatrix}V^*,$$

where V is an $n\times n$ orthogonal matrix. Since M is fixed, the ESD of P tends to δ_1, which leads to the fact that the LSD of the matrix $\frac{1}{n}Z_2PZ_2^*$ is the standard Marčenko–Pastur law. Then the matrix $(\frac{1}{n}Z_2PZ_2^*)^{-1}B_{22}$ is a standard Fisher matrix, and its $M+1$ largest eigenvalues $\alpha_1(C)\ge\cdots\ge\alpha_{M+1}(C)$ all converge to the right edge β of the limiting Wachter distribution. Meanwhile, since C is the $(p-M)\times(p-M)$ major sub-matrix of $S_2^{-1}S_1$, we have by Cauchy interlacing theorem

$$\alpha_{M+1}(C)\le l_{p,M+1}\le\alpha_1(C)\le l_{p,1}.$$

Thus $l_{p,M+1}\to\beta$ either. On the other hand, we have

$$l_{p,1}=\|S_2^{-1}S_1\|_{op}\le\|S_2^{-1}\|_{op}\cdot\|S_1\|_{op},$$

so that for some positive constant θ, $\limsup l_{p,1}\le\theta$. Consequently, almost surely,

$$\beta\le\liminf l_{p,M}\le\cdots\le\limsup l_{p,1}\le\theta<\infty;$$

in particular the whole family $\{l_{p,j},1\le j\le M\}$ is bounded. Now let $1\le j\le M$ be fixed and assume that a subsequence $(l_{p_k,j})_k$ converges to a limit $\tilde\beta\in[\beta,\theta]$. Either $\tilde\beta=\phi(a_i)>\beta$ or $\tilde\beta=\beta$. However, according to Step 2, $\tilde\beta>\beta$ implies that $a_i>\gamma\{1+\sqrt{c+y-cy}\}$, and otherwise, we have $a_i\le\gamma\{1+\sqrt{c+y-cy}\}$. Therefore, accordingly to one of these two conditions, all subsequences converge to a *same* limit $\phi(a_i)$ or β, which is thus also the unique limit of the whole sequence $(l_{p,j})_p$.

The proof of Theorem 3.1 is complete. □

8.2. *Proof of Theorem 4.1*. *Step* 1: *Convergence to the eigenvalues of the random matrix* $-U_i^* R(\lambda_i)U_i/\Delta(\lambda_i)$. We start from (8.5). Define

$$
\begin{aligned}
A(\lambda) &= I_n - Z_2^*\left[\lambda I_p - \left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2X_2^*\right]^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{\lambda}{n}Z_2,\\
B(\lambda) &= I_m + X_2^*\left[\lambda I_p - \left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2X_2^*\right]^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2,\\
C(\lambda) &= Z_2^*\left[\lambda I_p - \left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2X_2^*\right]^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2,\\
D(\lambda) &= X_2^*\left[\lambda I_p - \left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2X_2^*\right]^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{n}Z_2,
\end{aligned}
\tag{8.17}
$$

then (8.5) could be written as

$$
\begin{aligned}
&\det\Bigg(\underbrace{\frac{l_{p,j}}{n}Z_1A(l_{p,j})Z_1^*}_{(i)} - \underbrace{\frac{1}{m}X_1B(l_{p,j})X_1^*}_{(ii)}\\
&\qquad + \underbrace{\frac{l_{p,j}}{n}Z_1C(l_{p,j})X_1^*}_{(iii)} + \underbrace{\frac{l_{p,j}}{m}X_1D(l_{p,j})Z_1^*}_{(iv)}\Bigg) = 0.
\end{aligned}
\tag{8.18}
$$

The remaining is to find second-order approximation of the four terms on the left-hand side of (8.18).

Using Lemma A.5 in the Appendix, we have

$$
\begin{aligned}
(i) &= \mathbb{E}\frac{\lambda_i}{n}Z_1A(\lambda_i)Z_1^* + \frac{l_{p,j}}{n}Z_1A(l_{p,j})Z_1^* - \mathbb{E}\frac{\lambda_i}{n}Z_1A(\lambda_i)Z_1^*\\
&= \big(\lambda_i + y\lambda_i^2 s(\lambda_i)\big)\cdot I_M + \frac{l_{p,j}}{n}Z_1A(l_{p,j})Z_1^* - \frac{\lambda_i}{n}Z_1A(\lambda_i)Z_1^*\\
&\quad + \frac{\lambda_i}{n}Z_1A(\lambda_i)Z_1^* - \mathbb{E}\frac{\lambda_i}{n}Z_1A(\lambda_i)Z_1^*\\
&= \big(\lambda_i + y\lambda_i^2 s(\lambda_i)\big)\cdot I_M + \frac{l_{p,j}-\lambda_i}{n}Z_1A(l_{p,j})Z_1^*\\
&\quad + \frac{\lambda_i}{n}Z_1\big(A(l_{p,j}) - A(\lambda_i)\big)Z_1^*\\
&\quad + \frac{\lambda_i}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}Z_1A(\lambda_i)Z_1^* - \mathbb{E}\frac{1}{\sqrt{n}}Z_1A(\lambda_i)Z_1^*\right]\\
&\to \big(\lambda_i + y\lambda_i^2 s(\lambda_i)\big)\cdot I_M + (l_{p,j}-\lambda_i)\cdot\big(1 + 2y\lambda_i s(\lambda_i) + \lambda_i^2 y m_1(\lambda_i)\big)\cdot I_M\\
&\quad + \frac{\lambda_i}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}Z_1A(\lambda_i)Z_1^* - \mathbb{E}\frac{1}{\sqrt{n}}Z_1A(\lambda_i)Z_1^*\right],
\end{aligned}
\tag{8.19}
$$

$$
\begin{aligned}
(ii) &= \mathbb{E}\frac{1}{m}X_1B(\lambda_i)X_1^* + \frac{1}{m}X_1B(l_{p,j})X_1^* - \mathbb{E}\frac{1}{m}X_1B(\lambda_i)X_1^* \\
&= U\big(1-c-c\lambda_i s(\lambda_i)\big)\cdot\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k\end{pmatrix}U^* + \frac{1}{m}X_1\big(B(l_{p,j})-B(\lambda_i)\big)X_1^* \\
&\quad + \frac{1}{\sqrt{m}}\bigg[\frac{1}{\sqrt{m}}X_1B(\lambda_i)X_1^* - \mathbb{E}\frac{1}{\sqrt{m}}X_1B(\lambda_i)X_1^*\bigg] \\
&\to U\big(1-c-c\lambda_i s(\lambda_i)\big)\cdot\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k\end{pmatrix}U^* \\
&\quad - U(l_{p,j}-\lambda_i)\cdot cm_3(\lambda_i)\cdot\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k\end{pmatrix}U^* \\
&\quad + \frac{1}{\sqrt{m}}\bigg[\frac{1}{\sqrt{m}}X_1B(\lambda_i)X_1^* - \mathbb{E}\frac{1}{\sqrt{m}}X_1B(\lambda_i)X_1^*\bigg],
\end{aligned} \tag{8.20}
$$

$$
\begin{aligned}
(iii) &= \frac{l_{p,j}}{n}Z_1C(l_{p,j})X_1^* - \mathbb{E}\frac{\lambda_i}{n}Z_1C(\lambda_i)X_1^* \\
&= \frac{l_{p,j}}{n}Z_1C(l_{p,j})X_1^* - \frac{\lambda_i}{n}Z_1C(\lambda_i)X_1^* + \frac{\lambda_i}{n}Z_1C(\lambda_i)X_1^* \\
&\quad - \mathbb{E}\frac{\lambda_i}{n}Z_1C(\lambda_i)X_1^* \\
&= \frac{l_{p,j}}{n}Z_1\big(C(l_{p,j})-C(\lambda_i)\big)X_1^* + \frac{l_{p,j}-\lambda_i}{n}Z_1C(\lambda_i)X_1^* \\
&\quad + \frac{\lambda_i}{n}\cdot\big[Z_1C(\lambda_i)X_1^* - \mathbb{E}Z_1C(\lambda_i)X_1^*\big] \\
&\to \frac{\lambda_i}{n}\cdot\big[Z_1C(\lambda_i)X_1^* - \mathbb{E}Z_1C(\lambda_i)X_1^*\big],
\end{aligned} \tag{8.21}
$$

$$
\begin{aligned}
(iv) &= \frac{l_{p,j}}{m}X_1D(l_{p,j})Z_1^* - \mathbb{E}\frac{\lambda_i}{m}X_1D(\lambda_i)Z_1^* \\
&= \frac{l_{p,j}}{m}X_1D(l_{p,j})Z_1^* - \frac{\lambda_i}{m}X_1D(\lambda_i)Z_1^* + \frac{\lambda_i}{m}X_1D(\lambda_i)Z_1^* \\
&\quad - \mathbb{E}\frac{\lambda_i}{m}X_1D(\lambda_i)Z_1^* \\
&= \frac{l_{p,j}}{m}X_1\big(D(l_{p,j})-D(\lambda_i)\big)Z_1^* + \frac{l_{p,j}-\lambda_i}{m}X_1D(\lambda_i)Z_1^*
\end{aligned} \tag{8.22}
$$

$$
\begin{aligned}
&+\frac{\lambda_i}{m}\cdot\big[X_1D(\lambda_i)Z_1^*-\mathbb{E}X_1D(\lambda_i)Z_1^*\big]\\
\to&\ \frac{\lambda_i}{m}\cdot\big[X_1D(\lambda_i)Z_1^*-\mathbb{E}X_1D(\lambda_i)Z_1^*\big].
\end{aligned}
$$

Denote

$$
\begin{aligned}
R_n(\lambda_i)=\lambda_i\sqrt{\frac{p}{n}}\bigg[\frac{1}{\sqrt{n}}Z_1A(\lambda_i)Z_1^*\bigg]-\sqrt{\frac{p}{m}}\bigg[\frac{1}{\sqrt{m}}X_1B(\lambda_i)X_1^*\bigg]\\
+\lambda_i\sqrt{\frac{p}{n}}\bigg[\frac{1}{\sqrt{n}}Z_1C(\lambda_i)X_1^*\bigg]\\
+\lambda_i\sqrt{\frac{p}{m}}\bigg[\frac{1}{\sqrt{m}}X_1D(\lambda_i)Z_1^*\bigg]-\mathbb{E}[\cdot],
\end{aligned}
\tag{8.23}
$$

where $\mathbb{E}[\cdot]$ denotes the total expectation of all the preceding terms in the equation, and

$$
\Delta(\lambda_i)=1+2y\lambda_i s(\lambda_i)+\lambda_i^2 ym_1(\lambda_i)+a_i cm_3(\lambda_i).
$$

Combining (8.18), (8.19), (8.20), (8.21), (8.22) and considering the diagonal block that corresponds to the row and column index in $J_i\times J_i$ leads to

$$
\big|\sqrt{p}(l_{p,j}-\lambda_i)\cdot\Delta(\lambda_i)\cdot I_{n_i}+U_i^*R_n(\lambda_i)U_i\big|\to 0. \tag{8.24}
$$

Furthermore, it will be established in Step 2 below that

$$
U_i^*R_n(\lambda_i)U_i\longrightarrow U_i^*R(\lambda_i)U_i\qquad\text{in distribution,} \tag{8.25}
$$

for some random matrix $R(\lambda_i)$. Using the device of Skorokhod strong representation [Hu and Bai (2014), Skorokhod (1956)], we may assume that this convergence hold almost surely by considering an enlarged probability space. Under this device, (8.24) is equivalent to say that $\sqrt{p}(l_{p,j}-\lambda_i)$ tends to an eigenvalue of the matrix $-U_i^*R(\lambda_i)U_i/\Delta(\lambda_i)$. Finally, as the index j is arbitrary over the set J_i, all the n_i random variables

$$
\{\sqrt{p}(l_{p,j}-\lambda_i),\, j\in J_i\}
$$

converge almost surely to the set of eigenvalues of the random matrix $-\frac{U_i^*R(\lambda_i)U_i}{\Delta(\lambda_i)}$. Besides, due to Lemma A.3, we have

$$
\begin{aligned}
\Delta(\lambda_i)&=1+2y\lambda_i s(\lambda_i)+\lambda_i^2 ym_1(\lambda_i)+acm_3(\lambda_i)\\
&=\frac{(1-a_i-c)(1+a_i(y-1))^2}{(a_i-1)(-1+2a_i+c+a_i^2(y-1))}.
\end{aligned}
$$

Step 2: *Proof of the convergence* (8.25) *and structure of the random matrix* $R(\lambda_i)$. In the second step, we aim to find the matrix limit of the random matrix $U_i^*R_n(\lambda_i)U_i$. First, we show $U_i^*R_n(\lambda_i)U_i$ equals to another random matrix

$U_i^* \tilde{R}_n(\lambda_i) U_i$, here $\tilde{R}_n(\lambda_i)$ is the type of random sesquilinear form. Then using the results in Bai and Yao (2008) (Proposition 3.1 and Remark 1), we are able to find the matrix limit of $\tilde{R}_n(\lambda_i)$.

By assumption (b) that $x_i = \Sigma_p^{1/2} w_i$, we have its first M components:

$$X_1 = \Omega_M^{1/2} W_1 = U \begin{pmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_k} \end{pmatrix} U^* W_1.$$

Recall the definition of $R_n(\lambda_i)$ in (8.23), we have

$$\begin{aligned}
&U_i^* R_n(\lambda_i) U_i \\
&= U_i^* \frac{\sqrt{p}\lambda_i}{n} Z_1 A(\lambda_i) Z_1^* U_i - \frac{\sqrt{p}}{m} \begin{pmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_k} \end{pmatrix} U_i^* W_1 B(\lambda_i) \\
&\quad \times W_1^* U_i \begin{pmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_k} \end{pmatrix} \\
&\quad + U_i^* \frac{\sqrt{p}\lambda_i}{n} Z_1 C(\lambda_i) W_1^* U_i \begin{pmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_k} \end{pmatrix} \\
&\quad + \frac{\lambda_i \sqrt{p}}{m} \begin{pmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_k} \end{pmatrix} U_i^* W_1 D(\lambda_i) Z_1^* U_i - \mathbb{E}[\cdot] \qquad (8.26) \\
&= U_i^* \biggl\{ \lambda_i \frac{\sqrt{p}}{n} Z_1 A(\lambda_i) Z_1^* - a_i \frac{\sqrt{p}}{m} W_1 B(\lambda_i) W_1^* + \sqrt{a_i}\lambda_i \frac{\sqrt{p}}{n} Z_1 C(\lambda_i) W_1^* \\
&\quad + \sqrt{a_i}\lambda_i \frac{\sqrt{p}}{m} W_1 D(\lambda_i) Z_1^* \biggr\} U_i - \mathbb{E}[\cdot] \\
&= U_i^* \begin{pmatrix} Z_1 & W_1 \end{pmatrix} \begin{pmatrix} \dfrac{\lambda_i \sqrt{p} A(\lambda_i)}{n} & \dfrac{\lambda_i \sqrt{a_i p} C(\lambda_i)}{n} \\ \dfrac{\lambda_i \sqrt{a_i p} D(\lambda_i)}{m} & \dfrac{-a_i \sqrt{p} B(\lambda_i)}{m} \end{pmatrix} \begin{pmatrix} Z_1^* \\ W_1^* \end{pmatrix} U_i - \mathbb{E}[\cdot] \\
&:= U_i^* \tilde{R}_n(\lambda_i) U_i,
\end{aligned}$$

where

$$\tilde{R}_n(\lambda_i) := \begin{pmatrix} Z_1 & W_1 \end{pmatrix} \begin{pmatrix} \dfrac{\lambda_i \sqrt{p} A(\lambda_i)}{n} & \dfrac{\lambda_i \sqrt{a_i p} C(\lambda_i)}{n} \\ \dfrac{\lambda_i \sqrt{a_i p} D(\lambda_i)}{m} & \dfrac{-a_i \sqrt{p} B(\lambda_i)}{m} \end{pmatrix} \begin{pmatrix} Z_1^* \\ W_1^* \end{pmatrix} - \mathbb{E}[\cdot].$$

Finally, using Lemma A.6 in the Appendix leads to the result. The proof of Theorem 4.1 is complete.

APPENDIX A: SOME LEMMAS

LEMMA A.1. *Let R be a $M \times M$ real-valued matrix, $U = (U_1 \quad \cdots \quad U_k)$ and $V = (V_1 \quad \cdots \quad V_k)$ be two orthogonal bases of some subspace $E \subseteq \mathbb{R}^M$ of dimension M, where both U_i and V_i are of size $M \times n_i$, satisfying $n_1 + \cdots + n_k = M$. Then the eigenvalues of the two $n_i \times n_i$ matrices $U_i^* R U_i$ and $V_i^* R V_i$ are the same.*

PROOF. It is sufficient to prove that there exists a $n_i \times n_i$ orthogonal matrix A, such that

$$V_i = U_i \cdot A. \tag{A.1}$$

If it is true, then $V_i^* R V_i = A^*(U_i^* R U_i)A$. Since A is orthogonal, we have the eigenvalues of $V_i^* R V_i$ and $U_i^* R U_i$ are the same. Therefore, it only remains to show (A.1). Let $U_i = (u_1 \quad \cdots \quad u_{n_i})$ and $V_i = (v_1 \quad \cdots \quad v_{n_i})$. Define $A = (a_{ls})_{1 \le l, s \le n_i}$, such that

$$\begin{cases} v_1 = a_{11} u_1 + \cdots + a_{n_i 1} u_{n_i}, \\ \quad \vdots \\ v_{n_i} = a_{1 n_i} u_1 + \cdots + a_{n_i n_i} u_{n_i}. \end{cases}$$

Put in matrix form:

$$(v_1 \quad \cdots \quad v_{n_i}) = (u_1 \quad \cdots \quad u_{n_i}) \begin{pmatrix} a_{11} & \cdots & a_{1n_i} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{n_i n_i} \end{pmatrix},$$

that is, $V_i = U_i \cdot A$. Since $\langle v_i, v_j \rangle = \langle a_{\cdot i}, a_{\cdot j} \rangle$ (by orthogonality of $\{u_j\}$), where $a_{\cdot k} = (a_{lk})_{1 \le k \le n_i}$, the matrix A is then orthogonal. □

LEMMA A.2. *Suppose $X = (x_1, \ldots, x_n)$ is a $p \times n$ matrix, with each columns $\{x_i\}$ being independent random vectors. $Y = (y_1, \ldots, y_n)$ is defined similarly. Let Σ_p be the covariance matrix between x_i and y_i, A is a deterministic matrix, then we have*

$$XAY^* \longrightarrow \operatorname{tr} A \cdot \Sigma_p.$$

Moreover, if A is random but independent of X and Y, then we have

$$XAY^* \longrightarrow \mathbb{E} \operatorname{tr} A \cdot \Sigma_p. \tag{A.2}$$

PROOF. We consider the (i, j)th entry of XAY^*:

$$XAY^*(i, j) = \sum_{k,l=1}^{n} X(i, k)A(k, l)Y^*(l, j) = \sum_{k,l=1}^{n} X_{ik}Y_{jl}A_{kl}. \tag{A.3}$$

Since $X_{ik}Y_{jl} \to \Sigma_p(i, j)$ when $k = l$, the right-hand side of (A.3) tends to $\Sigma_p(i, j) \cdot \sum_{k=1}^{n} A_{kk}$, which is equivalent to say that

$$XAY^* \to \operatorname{tr} A \cdot \Sigma_p.$$

Equation (A.2) is simply due to the conditional expectation. The proof of Lemma A.2 is complete. □

In all the following, write λ as the outlier limit $\phi(a)$ in (3.4), that is,

$$\lambda := \frac{a(a-1+c)}{a-1-ay}.$$

LEMMA A.3. *With $s(z)$, $m_1(z) - m_4(z)$ defined in (8.1), we have*

$$s(\lambda) = \frac{a(y-1)+1}{(a-1)(a+c-1)},$$

$$m_1(\lambda) = \frac{(a(y-1)+1)^2(-1+2a+a^2(y-1)+y(c-1))}{(a-1)^2(a+c-1)^2(-1+2a+c+a^2(y-1))},$$

$$m_2(\lambda) = \frac{1}{a-1},$$

$$m_3(\lambda) = \frac{-(a(y-1)+1)^2}{(a-1)^2(-1+2a+c+a^2(y-1))},$$

$$m_4(\lambda) = \frac{-1+2a+c+a^2(-1+c(y-1))}{(a-1)^2(-1+2a+c+a^2(y-1))}.$$

SKETCH OF THE PROOF OF LEMMA A.3. In this short proof, we skip all the detailed calculations. Recall the definition of $\underline{s}(z)$ in (8.11), its value at λ is

$$\underline{s}(\lambda) = \frac{a(y-1)+1}{(a-1)(a+c-1)}. \tag{A.4}$$

Also, (8.11) says that $\underline{s}(z)$ is the solution of the following equation:

$$z(c+zy)\underline{s}^2(z) + \big(c\big(z(1-y)+1-c\big)+2zy\big)\underline{s}(z) + c + y - cy = 0. \tag{A.5}$$

Taking derivatives on both sides of (A.5) and combing with (A.4) will give the value of $\underline{s}'(\lambda)$. On the other hand, according to (8.10), it holds

$$\underline{s}(z) + \frac{1}{z}(1-c) = cs(z), \tag{A.6}$$

taking derivatives on both sides again will give the value of $s'(\lambda)$. Finally, the above five values are all some linear combinations of $s(\lambda)$ and $s'(\lambda)$. The proof of Lemma A.3 is complete. □

LEMMA A.4. *Under assumptions* (a)–(d),

$$\frac{1}{p}\operatorname{tr}\left\{\left(\lambda\cdot\frac{1}{n}Z_2Z_2^*-\frac{1}{m}X_2X_2^*\right)^{-1}\right\}\xrightarrow{a.s.}\frac{1}{a+c-1}.$$

PROOF. We first condition on Z_2, then we can use the result in Zheng, Bai and Yao (2013) (Lemma 4.3), which says that

$$\frac{1}{p}\operatorname{tr}\left(\frac{1}{z}\cdot\frac{1}{n}Z_2Z_2^*-\frac{1}{m}X_2X_2^*\right)^{-1}\to\tilde{m}(z)\qquad\text{a.s.},$$

where $\tilde{m}(z)$ is the unique solution to the equation

$$\tilde{m}(z)=\int\frac{1}{\frac{x}{z}-\frac{1}{1-c\tilde{m}(z)}}\,dF_y(x)\tag{A.7}$$

satisfying

$$\Im(z)\cdot\Im\bigl(\tilde{m}(z)\bigr)\geq 0,$$

here, $F_y(x)$ is the LSD of $\frac{1}{n}Z_2Z_2^*$ (deterministic), which is the standard M–P law with parameter y. Besides, if we denote its Stieltjes transform as $s(z):=\int\frac{1}{x-z}\,dF_y(x)$, then (A.7) could be written as

$$\tilde{m}(z)=\int\frac{z}{x-\frac{z}{1-c\tilde{m}(z)}}\,dF_y(x)=z\cdot s\left(\frac{z}{1-c\tilde{m}(z)}\right).\tag{A.8}$$

Since we know that the Stieltjes transform of the LSD of a standard sample covariance matrix satisfies:

$$s(z)=\frac{1}{1-y-yzs(z)-z},\tag{A.9}$$

then bringing (A.8) into (A.9) leads to

$$\frac{\tilde{m}(z)}{z}=\frac{1}{1-y-y\cdot\frac{z}{1-c\tilde{m}(z)}\cdot\frac{\tilde{m}(z)}{z}-\frac{z}{1-c\tilde{m}(z)}},$$

whose nonnegative solution is unique, which is

$$\tilde{m}(z)=\frac{-1+y+z-zc+\sqrt{(1-y-z+zc)^2+4z(yc-y-c)}}{2(yc-y-c)}.\tag{A.10}$$

Therefore, we have for fixed $\frac{1}{n}Z_2Z_2^*$,

$$\frac{1}{p}\operatorname{tr}\left(\lambda\cdot\frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1} \to \tilde{m}\left(\frac{1}{\lambda}\right) = \frac{1}{a+c-1}$$

almost surely. Finally, due to the fact that for each ω, the ESD of $\frac{1}{n}Z_2Z_2^*(\omega)$ will tend to the same limit (standard M–P distribution), which is independent of the choice of ω. Therefore, we have for all $\frac{1}{n}Z_2Z_2^*$ (not necessarily deterministic but only independent of $\frac{1}{m}X_2X_2^*$),

$$\frac{1}{p}\operatorname{tr}\left(\lambda\cdot\frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1} \to \frac{1}{a+c-1}$$

almost surely. The proof of Lemma A.4 is complete. □

LEMMA A.5. *$A(\lambda)$, $B(\lambda)$, $C(\lambda)$ and $D(\lambda)$ are defined in* (8.17), *then*

$$(l-\lambda)\cdot\frac{1}{n}Z_1A(l)Z_1^* \to (l-\lambda)\cdot\big(1+y\lambda s(\lambda)\big)\cdot I_M, \tag{A.11}$$

$$\frac{\lambda}{n}Z_1\big[A(l)-A(\lambda)\big]Z_1^* \to (l-\lambda)\cdot\big(\lambda y s(\lambda)+\lambda^2 y m_1(\lambda)\big)\cdot I_M, \tag{A.12}$$

$$\frac{1}{m}X_1\big(B(l)-B(\lambda)\big)X_1^* \to -(l-\lambda)\cdot c m_3(\lambda)\cdot U\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k\end{pmatrix}U^*, \tag{A.13}$$

$$\frac{l}{n}Z_1\big(C(l)-C(\lambda)\big)X_1^* + \frac{l-\lambda}{n}Z_1C(\lambda)X_1^* \to (l-\lambda)\cdot\mathbf{0}_{M\times M}, \tag{A.14}$$

$$\frac{l}{m}X_1\big(D(l)-D(\lambda)\big)Z_1^* + \frac{l-\lambda}{m}X_1D(\lambda)Z_1^* \to (l-\lambda)\cdot\mathbf{0}_{M\times M}. \tag{A.15}$$

PROOF. *Proof of* (A.11): Since Z_1 is independent of A and $\operatorname{Cov}(Z_1) = I_M$, we combine this fact with Lemma A.2:

$$(l-\lambda)\cdot\frac{1}{n}Z_1A(l)Z_1^* \to (l-\lambda)\cdot\frac{1}{n}\mathbb{E}\operatorname{tr}A(l)\cdot I_M. \tag{A.16}$$

Considering the expression of $A(l)$, we have

$$\begin{aligned}\frac{1}{n}\mathbb{E}\operatorname{tr}A(\lambda) &= \frac{1}{n}\mathbb{E}\operatorname{tr}\left[I_n - Z_2^*(\lambda I_p - S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{\lambda}{n}Z_2\right]\\ &= 1-\frac{\lambda}{n}\mathbb{E}\operatorname{tr}(\lambda I_p - S)^{-1}\\ &= 1-y\lambda\int\frac{1}{\lambda-x}\,dF_{c,y}(x)\\ &= 1+y\lambda s(\lambda).\end{aligned}$$

Therefore, combine with (A.16), we have

$$(l-\lambda)\cdot\frac{1}{n}Z_1A(l)Z_1^* \to (l-\lambda)\big(1+y\lambda s(\lambda)\big)\cdot I_M.$$

Proof of (A.12): Bringing the expression of $A(l)$ into consideration, we first have

$$\begin{aligned}
&A(l)-A(\lambda)\\
&\quad= Z_2^*(\lambda I_p-S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{\lambda}{n}Z_2\\
&\qquad - Z_2^*(lI_p-S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{l}{n}Z_2\\
&\quad= Z_2^*(\lambda I_p-S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{\lambda-l}{n}Z_2\\
&\qquad + Z_2^*\big[(\lambda I_p-S)^{-1}-(lI_p-S)^{-1}\big]\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{l}{n}Z_2\\
&\quad= (l-\lambda)\cdot\left[-Z_2^*(\lambda I_p-S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{n}Z_2\right.\\
&\qquad \left.+ Z_2^*(\lambda I_p-S)^{-1}(lI_p-S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{l}{n}Z_2\right].
\end{aligned}$$

Then using Lemma A.2 for the same reason, we have

$$\frac{\lambda}{n}Z_1\big[A(l)-A(\lambda)\big]Z_1^* \to \frac{\lambda}{n}\big\{\mathbb{E}\,\mathrm{tr}\big(A(l)-A(\lambda)\big)\big\}\cdot I_M$$

and

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\,\mathrm{tr}\big(A(l)-A(\lambda)\big)\\
&\quad= (l-\lambda)\cdot\left[-\frac{1}{n}\mathbb{E}\,\mathrm{tr}\left\{Z_2^*(\lambda I_p-S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{n}Z_2\right\}\right.\\
&\qquad \left.+\frac{1}{n}\mathbb{E}\,\mathrm{tr}\left\{Z_2^*(\lambda I_p-S)^{-1}(lI_p-S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{l}{n}Z_2\right\}\right]\\
&\quad= (l-\lambda)\cdot\left[-\frac{1}{n}\mathbb{E}\,\mathrm{tr}(\lambda I_p-S)^{-1}+\frac{\lambda}{n}\mathbb{E}\,\mathrm{tr}(\lambda I_p-S)^{-2}+o(1)\right]\\
&\quad= (l-\lambda)\cdot\left[y\int\frac{1}{x-\lambda}\,dF_{c,y}(x)+\lambda y\int\frac{1}{(\lambda-x)^2}\,dF_{c,y}(x)+o(1)\right]\\
&\quad= (l-\lambda)\cdot\big[ys(\lambda)+\lambda y m_1(\lambda)+o(1)\big].
\end{aligned}$$

Therefore, we have

$$\frac{\lambda}{n}Z_1\big[A(l)-A(\lambda)\big]Z_1^* \to (l-\lambda)\cdot\big(y\lambda s(\lambda)+\lambda^2 y m_1(\lambda)\big)\cdot I_M.$$

Proof of (A.13): First, recall the fact that

$$\mathrm{Cov}(X_1) = U \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k \end{pmatrix} U^*$$

and X_1 is independent of B. Using Lemma A.2, we have

$$\frac{1}{m} X_1 \big(B(l) - B(\lambda)\big) X_1^* \to \frac{1}{m} \mathbb{E}\,\mathrm{tr}\big(B(l) - B(\lambda)\big) \cdot U \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k \end{pmatrix} U^*.$$

The part

$$\begin{aligned} &\frac{1}{m}\mathbb{E}\,\mathrm{tr}\big(B(l) - B(\lambda)\big) \\ &\quad = \frac{1}{m}\mathbb{E}\,\mathrm{tr}\bigg\{X_2^*[(lI_p - S)^{-1} - (\lambda I_p - S)^{-1}]\bigg(\frac{1}{n}Z_2Z_2^*\bigg)^{-1}\frac{1}{m}X_2\bigg\} \\ &\quad = (l-\lambda)\cdot\bigg[-\frac{1}{m}\mathbb{E}\,\mathrm{tr}\{(\lambda I_p - S)^{-2}S\} + o(1)\bigg] \\ &\quad = (l-\lambda)\cdot\bigg[-c\int\frac{x}{(\lambda - x)^2}\,dF_{c,y}(x) + o(1)\bigg] \\ &\quad = (l-\lambda)\cdot\big(-cm_3(\lambda) + o(1)\big). \end{aligned}$$

Therefore, we have

$$\frac{1}{m} X_1 \big(B(l) - B(\lambda)\big) X_1^* \to -c(l-\lambda)m_3(\lambda) \cdot U \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_k \end{pmatrix} U^*.$$

Proof of (A.14) *and* (A.15): (A.14) and (A.15) are derived simply due to the fact that $\mathrm{Cov}(X_1, Z_1) = \mathbf{0}_{M\times M}$. The proof of Lemma A.5 is complete. □

LEMMA A.6. *Define*

$$\tilde{R}_n(\lambda_i) := \begin{pmatrix} Z_1 & W_1 \end{pmatrix} \begin{pmatrix} \dfrac{\lambda_i\sqrt{p}A(\lambda_i)}{n} & \dfrac{\lambda_i\sqrt{a_i p}C(\lambda_i)}{n} \\ \dfrac{\lambda_i\sqrt{a_i p}D(\lambda_i)}{m} & \dfrac{-a_i\sqrt{p}B(\lambda_i)}{m} \end{pmatrix} \begin{pmatrix} Z_1^* \\ W_1^* \end{pmatrix} - \mathbb{E}[\cdot]$$

then $\tilde{R}_n(\lambda_i)$ *weakly converges to a* $M \times M$ *symmetric random matrix* $R(\lambda_i) = (R_{mn})$, *which is made with independent Gaussian entries of mean zero and variance*

$$\mathbb{V}\mathrm{ar}(R_{mn}) = \begin{cases} 2\theta_i + (v_4 - 3)\omega_i, & m = n, \\ \theta_i, & m \neq n, \end{cases}$$

where

$$\omega_i = \frac{a_i^2(a_i+c-1)^2(c+y)}{(a_i-1)^2},$$
$$\theta_i = \frac{a_i^2(a_i+c-1)^2(cy-c-y)}{-1+2a_i+c+a_i^2(y-1)}.$$

PROOF. Since Z_1 and W_1 are independent, both are made with i.i.d. components, having the same first four moments, we can now view $\begin{pmatrix} Z_1 & W_1 \end{pmatrix}$ as a $M \times (n+m)$ table ξ, made with i.i.d elements of mean 0 and variance 1. Besides, we can rewrite the expression of $A(\lambda)$, $B(\lambda)$, $C(\lambda)$ and $D(\lambda)$ as follows:

$$A(\lambda) = I_n - Z_2^*\left(\lambda \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1}\frac{\lambda}{n}Z_2,$$
$$B(\lambda) = I_m + X_2^*\left(\lambda \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1}\frac{1}{m}X_2,$$
$$C(\lambda) = Z_2^*\left(\lambda \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1}\frac{1}{m}X_2,$$
$$D(\lambda) = X_2^*\left(\lambda \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1}\frac{1}{n}Z_2.$$

It holds

$$A(\lambda)^* = A(\lambda), \qquad B(\lambda)^* = B(\lambda), \qquad m \cdot C(\lambda)^* = n \cdot D(\lambda),$$

therefore, the matrix

$$\begin{pmatrix} \frac{\lambda_i\sqrt{p}A(\lambda_i)}{n} & \frac{\lambda_i\sqrt{a_i p}C(\lambda_i)}{n} \\ \frac{\lambda_i\sqrt{a_i p}D(\lambda_i)}{m} & \frac{-a_i\sqrt{p}B(\lambda_i)}{m} \end{pmatrix}$$

is symmetric. Define

$$A_n(\lambda_i) = \sqrt{n+m} \cdot \begin{pmatrix} \frac{\lambda_i\sqrt{p}A(\lambda_i)}{n} & \frac{\lambda_i\sqrt{a_i p}C(\lambda_i)}{n} \\ \frac{\lambda_i\sqrt{a_i p}D(\lambda_i)}{m} & \frac{-a_i\sqrt{p}B(\lambda_i)}{m} \end{pmatrix}. \tag{A.17}$$

Now we can apply the results in Bai and Yao (2008) (Proposition 3.1 and Remark 1), which says that $\tilde{R}_n(\lambda_i)$ weakly converges to a $M \times M$ symmetric random matrix $R(\lambda_i) = (R_{mn})$, which is made with i.i.d. Gaussian entries of mean zero and variance

$$\mathbb{V}\mathrm{ar}(R_{mn}) = \begin{cases} 2\theta_i + (v_4 - 3)\omega_i, & m = n, \\ \theta_i, & m \neq n. \end{cases}$$

The following is devoted to the calculation of the values of θ_i and ω_i.

Calculating of θ_i: From the definition of θ [see Bai and Yao (2008) for details], we have

$$\begin{aligned}
\theta_i &= \lim \frac{1}{n+m} \operatorname{tr} A_n^2(\lambda_i) \\
&= \lim \operatorname{tr} \begin{pmatrix} \frac{\lambda_i \sqrt{p} A(\lambda_i)}{n} & \frac{\lambda_i \sqrt{a_i p} C(\lambda_i)}{n} \\ \frac{\lambda_i \sqrt{a_i p} D(\lambda_i)}{m} & \frac{-a_i \sqrt{p} B(\lambda_i)}{m} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \frac{\lambda_i \sqrt{p} A(\lambda_i)}{n} & \frac{\lambda_i \sqrt{a_i p} C(\lambda_i)}{n} \\ \frac{\lambda_i \sqrt{a_i p} D(\lambda_i)}{m} & \frac{-a_i \sqrt{p} B(\lambda_i)}{m} \end{pmatrix} \qquad \text{(A.18)} \\
&= \lim \operatorname{tr} \begin{pmatrix} \frac{p\lambda_i^2}{n^2} A^2(\lambda_i) + \frac{\lambda_i^2 a_i p}{nm} C(\lambda_i) D(\lambda_i) & \star \\ \star & \frac{\lambda_i^2 a_i p}{nm} D(\lambda_i) C(\lambda_i) + \frac{a_i^2 p}{m^2} B^2(\lambda_i) \end{pmatrix} \\
&= \lim \left[\frac{p\lambda_i^2}{n^2} \operatorname{tr} A^2(\lambda_i) + \frac{2\lambda_i^2 a_i p}{nm} \operatorname{tr} C(\lambda_i) D(\lambda_i) + \frac{a_i^2 p}{m^2} \operatorname{tr} B^2(\lambda_i) \right],
\end{aligned}$$

$$\begin{aligned}
\operatorname{tr} A^2(\lambda_i) &= \operatorname{tr} \biggl[I_n + Z_2^* (\lambda_i I_p - S)^{-1} \left(\frac{1}{n} Z_2 Z_2^* \right)^{-1} \\
&\quad \times \frac{\lambda_i}{n} Z_2 Z_2^* (\lambda_i I_p - S)^{-1} \left(\frac{1}{n} Z_2 Z_2^* \right)^{-1} \frac{\lambda_i}{n} Z_2 \\
&\quad - 2 Z_2^* (\lambda_i I_p - S)^{-1} \left(\frac{1}{n} Z_2 Z_2^* \right)^{-1} \frac{\lambda_i}{n} Z_2 \biggr] \qquad \text{(A.19)} \\
&= n + \lambda_i^2 \operatorname{tr}(\lambda_i I_p - S)^{-2} - 2\lambda_i \operatorname{tr}(\lambda_i I_p - S)^{-1} \\
&= n + p\lambda_i^2 m_1(\lambda_i) + 2p\lambda_i s(\lambda_i),
\end{aligned}$$

$$\begin{aligned}
\operatorname{tr} C(\lambda_i) D(\lambda_i) &= \operatorname{tr} \biggl\{ Z_2^* (\lambda_i I_p - S)^{-1} \left(\frac{1}{n} Z_2 Z_2^* \right)^{-1} \\
&\qquad \times \frac{1}{m} X_2 X_2^* (\lambda_i I_p - S)^{-1} \left(\frac{1}{n} Z_2 Z_2^* \right)^{-1} \frac{1}{n} Z_2 \biggr\} \qquad \text{(A.20)} \\
&= \operatorname{tr}(\lambda_i I_p - S)^{-1} S (\lambda_i I_p - S)^{-1} = p m_3(\lambda_i),
\end{aligned}$$

$$\begin{aligned}
\operatorname{tr} B^2(\lambda_i) &= \operatorname{tr} \biggl[I_m + X_2^* (\lambda_i I_p - S)^{-1} \left(\frac{1}{n} Z_2 Z_2^* \right)^{-1} \\
&\qquad \times \frac{1}{m} X_2 X_2^* (\lambda_i I_p - S)^{-1} \left(\frac{1}{n} Z_2 Z_2^* \right)^{-1} \frac{1}{m} X_2
\end{aligned}$$

$$+ 2X_2^*(\lambda_i I_p - S)^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2\Big] \tag{A.21}$$

$$= m + \operatorname{tr}(\lambda_i I_p - S)^{-1}F(\lambda_i I_p - S)^{-1}S + 2\operatorname{tr}(\lambda_i I_p - S)^{-1}S$$

$$= m + pm_4(\lambda_i) + 2pm_2(\lambda_i).$$

Combining (A.18), (A.19), (A.20) and (A.21), we have

$$\begin{aligned}\theta_i &= \lambda_i^2 y\big(1 + y\lambda_i^2 m_1(\lambda_i) + 2y\lambda_i s(\lambda_i)\big) \\ &\quad + 2\lambda_i^2 a_i cym_3(\lambda_i) + a_i^2 c\big(1 + cm_4(\lambda_i) + 2cm_2(\lambda_i)\big) \\ &= \frac{a_i^2(a_i + c - 1)^2(cy - c - y)}{-1 + 2a_i + c + a_i^2(y-1)}.\end{aligned}$$

Calculating of ω_i:

$$\begin{aligned}\omega_i &= \lim \frac{1}{n+m}\sum_{i=1}^{n+m}\big(A_n(\lambda_i)(i,i)\big)^2 \\ &= \lim\left[\sum_{i=1}^{n}\frac{\lambda_i^2 p}{n^2}A^2(i,i) + \sum_{i=1}^{m}\frac{a_i^2 p}{m^2}B^2(i,i)\right].\end{aligned} \tag{A.22}$$

In the following, we will show that $A(i,i)$ and $B(i,i)$ both tend to some limits that is independent of i:

$$\begin{aligned}&A(i,i) \\ &= 1 - \left[Z_2^*\left[\lambda_i I_p - \left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{1}{m}X_2X_2^*\right]^{-1}\left(\frac{1}{n}Z_2Z_2^*\right)^{-1}\frac{\lambda_i}{n}Z_2\right](i,i) \\ &= 1 - \frac{\lambda_i}{n}\left[Z_2^*\left[\lambda_i \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right]^{-1}Z_2\right](i,i).\end{aligned} \tag{A.23}$$

If we denote η_i as the ith column of Z_2, we have

$$\frac{1}{n}Z_2Z_2^* = \frac{1}{n}\begin{pmatrix}\eta_1 & \cdots & \eta_n\end{pmatrix}\cdot\begin{pmatrix}\eta_1^* \\ \vdots \\ \eta_n^*\end{pmatrix} = \frac{1}{n}\eta_i\eta_i^* + \frac{1}{n}Z_{2i}Z_{2i}^*,$$

where Z_{2i} is independent of η_i. Since

$$\begin{aligned}&\left(\lambda_i \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1} - \left(\lambda_i \cdot \frac{1}{n}Z_{2i}Z_{2i}^* - \frac{1}{m}X_2X_2^*\right)^{-1} \\ &\quad = -\left(\lambda_i \cdot \frac{1}{n}Z_2Z_2^* - \frac{1}{m}X_2X_2^*\right)^{-1}\frac{\lambda_i}{n}\eta_i\eta_i^*\left(\lambda_i \cdot \frac{1}{n}Z_{2i}Z_{2i}^* - \frac{1}{m}X_2X_2^*\right)^{-1},\end{aligned}$$

we have

$$\begin{aligned}\left(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_2 X_2^*\right)^{-1} \\ = \frac{(\lambda_i \cdot \frac{1}{n} Z_{2i} Z_{2i}^* - \frac{1}{m} X_2 X_2^*)^{-1}}{1 + \frac{\lambda_i}{n} \eta_i^* (\lambda_i \cdot \frac{1}{n} Z_{2i} Z_{2i}^* - \frac{1}{m} X_2 X_2^*)^{-1} \eta_i}.\end{aligned} \tag{A.24}$$

Bringing (A.24) into (A.23),

$$\begin{aligned} A(i,i) &= 1 - \frac{\lambda_i}{n} \eta_i^* \left[\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_2 X_2^*\right]^{-1} \eta_i \\ &= 1 - \frac{\frac{\lambda_i}{n} \eta_i^* (\lambda_i \cdot \frac{1}{n} Z_{2i} Z_{2i}^* - \frac{1}{m} X_2 X_2^*)^{-1} \eta_i}{1 + \frac{\lambda_i}{n} \eta_i^* (\lambda_i \cdot \frac{1}{n} Z_{2i} Z_{2i}^* - \frac{1}{m} X_2 X_2^*)^{-1} \eta_i} \\ &= \frac{1}{1 + \frac{\lambda_i}{n} \eta_i^* (\lambda_i \cdot \frac{1}{n} Z_{2i} Z_{2i}^* - \frac{1}{m} X_2 X_2^*)^{-1} \eta_i}, \end{aligned}$$

whose denominator of (A.25) equals

$$1 + \frac{\lambda_i}{n} \operatorname{tr}\left(\lambda_i \cdot \frac{1}{n} Z_{2i} Z_{2i}^* - \frac{1}{m} X_2 X_2^*\right)^{-1} \eta_i \eta_i^*. \tag{A.25}$$

Since η_i is independent of $(\lambda_i \cdot \frac{1}{n} Z_{2i} Z_{2i}^* - \frac{1}{m} X_2 X_2^*)^{-1}$, (A.25) converges to the value $1 + \lambda_i y \cdot \frac{1}{a_i + c - 1}$ according to Lemma A.4. Therefore, we have

$$A(i,i) \to \frac{1}{1 + \lambda_i y \cdot \frac{1}{a_i + c - 1}}, \tag{A.26}$$

which is independent of the choice of i.

For the same reason, we have

$$\begin{aligned} &B(i,i) \\ &= 1 + \left[X_2^* \left[\lambda_i I_p - \left(\frac{1}{n} Z_2 Z_2^*\right)^{-1} \frac{1}{m} X_2 X_2^*\right]^{-1} \left(\frac{1}{n} Z_2 Z_2^*\right)^{-1} \frac{1}{m} X_2\right](i,i) \\ &= 1 + \left[X_2^* \left[\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_2 X_2^*\right]^{-1} \frac{1}{m} X_2\right](i,i). \end{aligned} \tag{A.27}$$

If we denote δ_i as the ith column of X_2, then we have

$$\frac{1}{m} X_2 X_2^* = \frac{1}{m} (\delta_1 \quad \cdots \quad \delta_m) \cdot \begin{pmatrix} \delta_1^* \\ \vdots \\ \delta_m^* \end{pmatrix} = \frac{1}{m} \delta_i \delta_i^* + \frac{1}{m} X_{2i} X_{2i}^*$$

and

$$\left(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_2 X_2^*\right)^{-1} - \left(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*\right)^{-1}$$
$$= \left(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_2 X_2^*\right)^{-1} \frac{1}{m} \delta_i \delta_i^* \left(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*\right)^{-1}.$$

So we have

$$\left(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_2 X_2^*\right)^{-1} = \frac{(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*)^{-1}}{1 - \frac{1}{m} \delta_i^* (\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*)^{-1} \delta_i}. \tag{A.28}$$

Combine (A.27) and (A.28), we have

$$\begin{aligned} B(i,i) &= 1 + \delta_i^* \left[\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_2 X_2^*\right]^{-1} \frac{1}{m} \delta_i \\ &= 1 + \frac{\frac{1}{m} \delta_i^* (\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*)^{-1} \delta_i}{1 - \frac{1}{m} \delta_i^* (\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*)^{-1} \delta_i} \\ &= \frac{1}{1 - \frac{1}{m} \delta_i^* (\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*)^{-1} \delta_i}. \end{aligned} \tag{A.29}$$

Using the independence between δ_i and $(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{T} X_{2i} X_{2i}^*)^{-1}$ and Lemma A.4 again, we have

$$\frac{1}{m} \delta_i^* \left(\lambda_i \cdot \frac{1}{n} Z_2 Z_2^* - \frac{1}{m} X_{2i} X_{2i}^*\right)^{-1} \delta_i \to c \cdot \frac{1}{a_i + c - 1}.$$

Therefore, we have

$$B(i,i) \to \frac{1}{1 - \frac{c}{a_i + c - 1}},$$

which is also independent of the choice of i.

Finally, taking the definition of ω_i in (A.22) into consideration, we have

$$\begin{aligned} \omega_i &= \frac{\lambda_i^2 y}{(1 + y\lambda_i \cdot \frac{1}{a_i + c - 1})^2} + \frac{a_i^2 c}{(1 - \frac{c}{a_i + c - 1})^2} \\ &= \frac{a_i^2 (a_i + c - 1)^2 (c + y)}{(a_i - 1)^2}. \end{aligned} \tag{A.30}$$

The proof of Lemma A.6 is complete. □

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York. MR0771294

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

BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686

BAI, Z. D., YIN, Y. Q. and KRISHNAIAH, P. R. (1987). On limiting empirical distribution function of the eigenvalues of a multivariate F matrix. *Theory Probab. Appl.* **32** 490–500.

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

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

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

BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. MR2835249

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618

CAPITAINE, M. (2013). Additive/multiplicative free subordination property and limiting eigenvectors of spiked additive deformations of Wigner matrices and spiked sample covariance matrices. *J. Theory Probab.* **26** 595–648. MR3090543

CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158

DHARMAWANSA, P., JOHNSTONE, I. M. and ONATSKI, A. (2014). Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios. Preprint. Available at arXiv:1411.3875.

FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807

HAN, X., PAN, G. and ZHANG, B. (2016). The Tracy–Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44** 1564–1592. MR3519933

HU, J. and BAI, Z. (2014). Strong representation of weak convergence. *Sci. China Math.* **57** 2399–2406. MR3266500

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

KARGIN, V. (2015). On estimation in the reduced-rank regression with a large number of responses and predictors. *J. Multivariate Anal.* **140** 377–394. MR3372575

KRITCHMAN, S. and NADLER, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemom. Intell. Lab. Syst.* **94** 19–32.

LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938

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

NADLER, B. (2010). Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator. *IEEE Trans. Signal Process.* **58** 2746–2756. MR2789420

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

PASSEMIER, D. and YAO, J.-F. (2012). On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices Theory Appl*. **1** 1150002, 19. MR2930380

PASSEMIER, D. and YAO, J. (2014). Estimation of the number of spikes, possibly equal, in the high-dimensional case. *J. Multivariate Anal*. **127** 173–183. MR3188885

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist*. *Sinica* **17** 1617–1642. MR2399865

PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787

PIZZO, A., RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat*. **49** 64–94. MR3060148

RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices II: Delocalized perturbations. *Random Matrices Theory Appl*. **2** 1250015, 36. MR3039820

SHI, D. (2013). Asymptotic joint distribution of extreme sample eigenvalues and eigenvectors in the spiked population model. Preprint. Available at arXiv:1304.6113.

SILVERSTEIN, J. W. (1985). The limiting eigenvalue distribution of a multivariate F matrix. *SIAM J. Math. Anal*. **16** 641–646. MR0783987

SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl*. **1** 261–290.

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

WANG, Q., SU, Z. and YAO, J. (2014). Joint CLT for several random sesquilinear forms with applications to large-dimensional spiked population models. *Electron. J. Probab.* **19** 1–28. MR3275855

ZHENG, S. R., BAI, Z. D. and YAO, J. F. (2013). CLT for linear spectral statistics of random matrix $S^{-1}T$. Preprint. Available at arXiv:1305.1376.

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
UNIVERSITY OF HONG KONG
POKFULAM
HONG KONG
E-MAIL: wqw8813@gmail.com
jeffyao@hku.hk