On the sphericity test with large-dimensional observations

Qinwen Wang[*] and Jianfeng Yao[†]

Qinwen Wang
Department of Mathematics
Zhejiang University
e-mail: wqw8813@gmail.com

Jianfeng Yao
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam, Hong Kong
e-mail: jeffyao@hku.hk

Abstract: In this paper, we propose corrections to the likelihood ratio test and John's test for sphericity in large-dimensions. New formulas for the limiting parameters in the CLT for linear spectral statistics of sample covariance matrices with general fourth moments are first established. Using these formulas, we derive the asymptotic distribution of the two proposed test statistics under the null. These asymptotics are valid for general population, i.e. not necessarily Gaussian, provided a finite fourth-moment. Extensive Monte-Carlo experiments are conducted to assess the quality of these tests with a comparison to several existing methods from the literature. Moreover, we also obtain their asymptotic power functions under the alternative of a spiked population model as a specific alternative.

Contents

1. Introduction

Consider a sample $Y_1, \ldots, Y_n$ from a p-dimensional multivariate distribution with covariance matrix Σ_p. An important problem in multivariate analysis is to test the sphericity, namely the hypothesis $H_0 : \Sigma_p = \sigma^2 I_p$ where σ^2 is *unspecified*. If the observations represent a multivariate error with p components, the null hypothesis expresses the fact that the error is cross-sectionally uncorrelated (independent if in addition they are normal) and have a same variance (homoscedasticity).

Much of the existing theory about this test has been exposed first in details in [17] about Gaussian likelihood ratio test and later in [10, 11, 27] and also in textbooks like [18, Chapter 8] and [1, Chapter 10]. Assume for a moment that the sample has a normal distribution with mean zero and covariance matrix Σ_p. Let $S_n = n^{-1} \sum_i Y_i Y_i^*$ be the sample covariance matrix and denote its eigenvalues by $\{\ell_i\}_{1 \leq i \leq p}$. Two well established procedures for testing the sphericity are the likelihood ratio test (LRT) and a test devised in [10]. The likelihood ratio statistic is, see e.g. [1, §10.7.2],

$$L_n = \left(\frac{(\ell_1 \cdots \ell_p)^{1/p}}{\frac{1}{p}(\ell_1 + \cdots + \ell_p)} \right)^{\frac{1}{2} pn},$$

which is a power of the ratio of the geometric mean of the sample eigenvalues to the arithmetic mean. It is here noticed that in this formula it is necessary to assume that $p \leq n$ to avoid null eigenvalues in (the numerator of) L_n. If we let $n \to \infty$ while keeping p fixed, classical asymptotic theory indicates that under the null hypothesis, $-2 \log L_n \Longrightarrow \chi_f^2$, a chi-square distribution with degree of freedom $f = \frac{1}{2} p(p+1) - 1$. This asymptotic distribution is further refined by the following Box-Bartlett correction (referred as *BBLRT*):

$$P(-2\rho \log L_n \leq x) = P_f(x) + \omega_2 \{P_{f+4}(x) - P_f(x)\} + O(n^{-3}), \qquad (1.1)$$

where $P_k(x) = P(\chi_k^2 \leq x)$ and

$$\rho = 1 - \frac{2p^2 + p + 2}{6pn}, \qquad \omega_2 = \frac{(p+2)(p-1)(p-2)(2p^3 + 6p^2 + 3p + 2)}{288p^2 n^2 \rho^2}.$$

By observing that the asymptotic variance of $-2 \log L_n$ is proportional to $\text{tr}\{\Sigma(\text{tr}\,\Sigma)^{-1} - p^{-1} I_p\}^2$, [10] proposed to use the statistic

$$T_2 = \frac{p^2 n}{2} \text{tr} \left\{ S_n (\text{tr}\, S_n)^{-1} - p^{-1} I_p \right\}^2$$

(p, n)	(4,64)	(8,64)	(16,64)	(32,64)	(48,64)	(56,64)	(60,64)
BBLRT	0.0483	0.0523	0.0491	0.0554	0.1262	0.3989	0.7605
Nagao's test	0.0485	0.0495	0.0478	0.0518	0.0518	0.0513	0.0495

for testing sphericity. When p is fixed and $n \to \infty$, under the null hypothesis, it also holds that $T_2 \Longrightarrow \chi_f^2$, which we referred to as *John's test.* It is observed that T_2 is proportional to the square of the coefficient of variation of the sample eigenvalues, namely

$$T_2 = \frac{np}{2} \cdot \frac{p^{-1} \sum (\ell_i - \overline{\ell})^2}{\overline{\ell}^2}, \qquad \text{with } \overline{\ell} = \frac{1}{n} \sum_i \ell_i.$$

Following the idea of the Box-Bartlett correction, [19] established an expansion for the distribution function of the statistics T_2 (referred as *Nagao's test*),

$$P(T_2 \leq x) = P_f(x) + \frac{1}{n} \left\{ a_p P_{f+6}(x) + b_p P_{f+4}(x) + c_p P_{f+2}(x) + d_p P_f(x) \right\}$$
$$+ O(n^{-2}), \tag{1.2}$$

where

$$a_p = \frac{1}{12}(p^3 + 3p^2 - 12 - 200p^{-1}), \; b_p = \frac{1}{8}(-2p^3 - 5p^2 + 7p - 12 - 420p^{-1}),$$
$$c_p = \frac{1}{4}(p^3 + 2p^2 - p - 2 - 216p^{-1}), \; d_p = \frac{1}{24}(-2p^3 - 3p^2 + p + 436p^{-1}).$$

It has been well known that classical multivariate procedures are in general challenged by large-dimensional data. A small simulation experiment is conducted to explore the performance of the BBLRT and Nagao's test (two corrections) with growing dimension p. The sample size is set to $n = 64$ while dimension p increases from 4 to 60 (we have also run other experiments with larger sample sizes n but conclusions are very similar), and the nominal level is set to be $\alpha = 0.05$. The samples come from normal vectors with mean zero and identity covariance matrix, and each pair of (p, n) is assessed with 10000 independent replications.

Table 1 gives the empirical sizes of BBLRT and Nagao's test. It is found here that when the dimension to sample size ratio p/n is below $1/2$, both tests have an empirical size close to the nominal test level 0.05. Then when the ratio grows up, the BBLRT becomes quickly biased while Nagao's test still has a correct empirical size. It is striking that although Nagao's test is derived under classical "p fixed, $n \to \infty$" regime, it is remarkably robust against dimension inflation.

Therefore, the goal of this paper is to propose novel corrections to both LRT and John's test to cope with the large-dimensional context. Similar works have already been done in [15], which confirms the robustness of John's test in large-dimensions; however, these results assume a Gaussian population. In this

paper, we remove such a Gaussian restriction, and prove that the robustness of John's test is in fact general. Following the idea of [15], [9] proposed to use a family of well selected U-statistics to test the sphericity; however, as showed in our simulation study in Section 3, the powers of our corrected John's test are slightly higher than this test in most cases. More recently, [26] examined the performance of T_1 (a statistic first put forward in [24]) under non-normality, but with the moment condition $\gamma = 3 + O(p^{-\epsilon})$, which essentially matches the Gaussian case ($\gamma = 3$) asymptotically. We have also removed this moment restriction in our setting. In short, we have unveiled two corrections that have a better performance and removed the Gaussian or nearly Gaussian restriction found in the existing literature.

From the technical point of view, our approach differs from [15] and follows the one devised in [4] and [6]. The central tool is a CLT for linear spectral statistics of sample covariance matrices established in [2] and later refined in [21]. The paper also contains an original contribution on this CLT reported in the Appendix: new formulas for the limiting parameters in the CLT. Since such CLT's are increasingly important in large-dimensional statistics, we believe that these new formulas will be of independent interest for applications other than those considered in this paper.

The remaining of the paper is organized as follows. Large-dimensional corrections to LRT and John's test are introduced in Section 2. Section 3 reports a detailed Monte-Carlo study to analyze finite-sample sizes and powers of these two corrections under both normal and non-normal distributed data. Next, Section 4 gives the theoretical analysis of their asymptotic power under the alternative of a spiked population model. Section 5 generalizes our test procedures to populations with an unknown mean. Technical proofs and calculations are relegated to Section 6. The last Section contains some concluding remarks.

2. Large-dimensional corrections

From now on, we assume that the observations $Y_1, \ldots, Y_n$ have the representation $Y_j = \Sigma_p^{1/2} X_j$ where the $p \times n$ table $\{X_1, \ldots, X_n\} = \{x_{ij}\}_{1 \leq i \leq p, 1 \leq j \leq n}$ are made with an array of i.i.d. standardized random variables (mean 0 and variance 1). This setting is motivated by the random matrix theory and it is generic enough for a precise analysis of the sphericity test. Furthermore, under the null hypothesis $H_0 : \Sigma_p = \sigma^2 I_p$ (σ^2 is unspecified), we notice that both LRT and John's test are independent from the scale parameter σ^2 under the null. Therefore, we can assume w.l.o.g. $\sigma^2 = 1$ when dealing with the null distributions of these test statistics. This will be assumed in all the sections.

Throughout the paper we will use an indicator κ set to 2 when $\{x_{ij}\}$ are real and to 1 when they are complex as defined in [3]. Also, we define the kurtosis coefficient $\beta = E|x_{ij}|^4 - 1 - \kappa$ for both cases and note that for normal variables, $\beta = 0$ (recall that for a standard complex-valued normal random variable, its real and imaginary parts are two iid. $N(0, \frac{1}{2})$ real random variables).

2.1. The corrected likelihood ratio test (CLRT)

For the correction of LRT, let $\mathcal{L}_n = -2n^{-1}\log L_n$ be the test statistic for $n \geq 1$. Our first main result is the following.

Theorem 2.1. *Assume $\{x_{ij}\}$ are iid, satisfying $Ex_{ij} = 0, E|x_{ij}|^2 = 1, E|x_{ij}|^4 < \infty$. Then under H_0 and when $\frac{p}{n} = y_n \to y \in (0,1)$,*

$$\mathcal{L}_n + (p - n) \cdot \log\left(1 - \frac{p}{n}\right) - p$$
$$\Longrightarrow N\left\{-\frac{\kappa - 1}{2}\log(1 - y) + \frac{1}{2}\beta y, -\kappa \log(1 - y) - \kappa y\right\}. \qquad (2.1)$$

The test based on this asymptotic normal distribution will be hereafter referred as the *corrected likelihood-ratio test* (*CLRT*). One may observe that the limiting distribution of the test crucially depends on the limiting dimension-to-sample ratio y through the factor $-\log(1 - y)$. In particular, the asymptotic variance will blow up quickly when y approaches 1, so it is expected that the power will seriously break down. Monte-Carlo experiments in Section 3 will provide more details on this behavior.

The proof of Theorem 2.1 is based on the following lemma. In all the following, F^y denotes the Marčenko-Pastur distribution of index y (> 0) which is introduced and discussed in the Appendix. And $F^y(f(x)) = \int f(x)F^y(dx)$ denotes the integral of function $f(x)$ with respect to F^y.

Lemma 2.1. *Let $\{\ell_i\}_{1\leq i\leq p}$ be the eigenvalues of the sample covariance matrix $S_n = n^{-1}\sum_i Y_i Y_i^*$. Then under H_0 and the conditions of Theorem 2.1, we have*

$$\left(\begin{array}{c} \sum_{i=1}^p \log \ell_i - pF^{y_n}(\log x) \\ \sum_{i=1}^p \ell_i - pF^{y_n}(x) \end{array}\right) \Longrightarrow N(\mu_1, V_1),$$

with

$$\mu_1 = \left(\begin{array}{c} \frac{\kappa-1}{2}\log(1 - y) - \frac{1}{2}\beta y \\ 0 \end{array}\right),$$

and

$$V_1 = \left(\begin{array}{cc} -\kappa \log(1 - y) + \beta y & (\beta + \kappa)y \\ (\beta + \kappa)y & (\beta + \kappa)y \end{array}\right).$$

The proof of this lemma is postponed to Section 6.

Proof of Theorem 2.1. Let $A_n \equiv \sum_{i=1}^p \log \ell_i - pF^{y_n}(\log x)$ and $B_n \equiv \sum_{i=1}^p \ell_i - pF^{y_n}(x)$. By Lemma 2.1,

$$\left(\begin{array}{c} A_n \\ B_n \end{array}\right) \Longrightarrow N(\mu_1, V_1).$$

Consequently, $-A_n + B_n$ is asymptotically normal with mean $-\frac{\kappa-1}{2}\log(1 - y) + \frac{1}{2}\beta y$ and variance

$$V_1(1,1) + V_1(2,2) - 2V_1(1,2) = -\kappa \log(1 - y) - \kappa y.$$

Besides,

$$\begin{aligned}
\mathcal{L}_n &= -\Sigma_{i=1}^{p} \log \ell_i + p \log\left(\frac{1}{p}\Sigma_{i=1}^{p}\ell_i\right) \\
&= -(A_n + pF^{y_n}(\log x)) + p\log\left(\frac{1}{p}(B_n + p)\right) \\
&= -A_n - pF^{y_n}(\log x) + p\log\left(1 + \frac{B_n}{p}\right).
\end{aligned}$$

Since $B_n \Rightarrow N(0, y(\beta + \kappa))$, $B_n = O_p(1)$ and $\log(1 + B_n/p) = B_n/p + O_p(1/p^2)$. Therefore,

$$\mathcal{L}_n = -A_n - pF^{y_n}(\log x) + B_n + O_p\left(\frac{1}{p}\right).$$

The conclusion follows with the following well-known integrals w.r.t. the Marčenko-Pastur distribution F^y $(y < 1)$ in [2],

$$F^y(\log x) = \frac{y-1}{y}\log(1-y) - 1, \quad F^y(x) = 1.$$

The proof of Theorem 2.1 is complete. □

2.2. The corrected John's test (CJ)

Earlier than the asymptotic expansion (1.2) given in [19], [10] proved that when the observations are normal, the sphericity test based on T_2 is a locally most powerful invariant test. It is also established in [11] that under these conditions, the limiting distribution of T_2 under H_0 is χ_f^2 with degree of freedom $f = \frac{1}{2}p(p+1) - 1$, or equivalently,

$$nU - p \Longrightarrow \frac{2}{p}\chi_f^2 - p,$$

where for convenience, we have let $U = 2(np)^{-1}T_2$. Clearly, this limit has been established for $n \to \infty$ and a fixed dimension p. However, if we now let $p \to \infty$ in the right-hand side of the above result, it is not hard to see that $\frac{2}{p}\chi_f^2 - p$ will tend to the normal distribution $N(1,4)$. It then seems "natural" to conjecture that when both p and n grow to infinity in some "proper" way, it may happen that

$$nU - p \Longrightarrow N(1,4). \tag{2.2}$$

This is indeed the main result of [15] where this asymptotic distribution was established assuming that data are normal-distributed and p and n grow to infinity in a proportional way (i.e. $p/n \to y > 0$).

In this section, we provide a more general result using our own approach. In particular, the distribution of the observation is arbitrary provided a finite fourth moment exists.

Theorem 2.2. *Assume $\{x_{ij}\}$ are iid, satisfying $Ex_{ij} = 0, E|x_{ij}|^2 = 1, E|x_{ij}|^4 < \infty$, and let $U = 2(np)^{-1}T_2$ be the test statistic. Then under H_0 and when*

$$p \to \infty, n \to \infty, \frac{p}{n} = y_n \to y \in (0, \infty),$$

$$nU - p \Longrightarrow N(\kappa + \beta - 1, 2\kappa). \tag{2.3}$$

The test based on the asymptotic normal distribution given in equation (2.3) will be hereafter referred as the *corrected John's test (CJ)*.

A striking fact in this theorem is that as in the normal case, the limiting distribution of CJ is *independent* of the dimension-to-sample ratio $y = \lim p/n$. In particular, the limiting distribution derived under classical scheme (p fixed, $n \to \infty$), e.g. the distribution $\frac{2}{p}\chi_f^2 - p$ in the normal case, when used for large p, stays very close to this limiting distribution derived for large-dimensional scheme ($p \to \infty, n \to \infty, p/n \to y \in (0, \infty)$). In this sense, Theorem 2.2 gives a theoretic explanation to the widely observed robustness of John's test against the dimension inflation. Moreover, CJ is also valid for the p larger (or much larger) than n case in contrast to the CLRT where this ratio should be kept smaller than 1 to avoid null eigenvalues.

It is also worth noticing that for real normal data, we have $\kappa = 2$ and $\beta = 0$ so that the theorem above reduces to $nU - p \Rightarrow N(1, 4)$. This is exactly the result discussed in [15]. Besides, if the data has a non-normal distribution but has the same first four moments as the normal distribution, we have again $nU - p \Rightarrow N(1, 4)$, which turns out to have a universality property.

The proof of Theorem 2.2 is based on the following lemma.

Lemma 2.2. *Let $\{\ell_i\}_{1 \le i \le p}$ be the eigenvalues of the sample covariance matrix $S_n = n^{-1} \sum_i Y_i Y_i^*$. Then under H_0 and the conditions of Theorem 2.2, we have*

$$\begin{pmatrix} \sum_{i=1}^p \ell_i^2 - p(1 + y_n) \\ \sum_{i=1}^p \ell_i - p \end{pmatrix} \Longrightarrow N(\mu_2, V_2),$$

with

$$\mu_2 = \begin{pmatrix} (\kappa - 1 + \beta)y \\ 0 \end{pmatrix},$$

and

$$V_2 = \begin{pmatrix} 2\kappa y^2 + 4(\kappa + \beta)(y + 2y^2 + y^3) & 2(\kappa + \beta)(y + y^2) \\ 2(\kappa + \beta)(y + y^2) & (\kappa + \beta)y \end{pmatrix}.$$

The proof of this lemma is postponed to Section 6.

Proof of Theorem 2.2. The result of Lemma 2.2 can be rewritten as:

$$n \begin{pmatrix} p^{-1} \sum_{i=1}^p \ell_i^2 - 1 - \frac{p}{n} - \frac{(\kappa + \beta - 1)y}{p} \\ p^{-1} \sum_{i=1}^p \ell_i - 1 \end{pmatrix} \Longrightarrow N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{y^2} \cdot V_2 \right).$$

Define the function $f(x, y) = \frac{x}{y^2} - 1$, then $U = f(p^{-1}\Sigma_{i=1}^p \ell_i^2, \ p^{-1}\Sigma_{i=1}^p \ell_i)$.
We have

$$\frac{\partial f}{\partial x}\left(1 + \frac{p}{n} + \frac{(\kappa + \beta - 1)y}{p}, 1\right) = 1,$$

$$\frac{\partial f}{\partial y}\left(1 + \frac{p}{n} + \frac{(\kappa + \beta - 1)y}{p}, 1\right) = -2\left(1 + \frac{p}{n} + \frac{(\kappa + \beta - 1)y}{p}\right),$$

$$f\left(1 + \frac{p}{n} + \frac{(\kappa + \beta - 1)y}{p}, 1\right) = \frac{p}{n} + \frac{(\kappa + \beta - 1)y}{p}.$$

By the delta method,

$$n\left(U - f\left(1 + \frac{p}{n} + \frac{(\kappa + \beta - 1)y}{p}, 1\right)\right) \Longrightarrow N(0, \lim C),$$

where

$$C = \left(\begin{array}{c} \frac{\partial f}{\partial x}(1 + \frac{p}{n} + \frac{(\kappa+\beta-1)y}{p}, 1) \\ \frac{\partial f}{\partial y}(1 + \frac{p}{n} + \frac{(\kappa+\beta-1)y}{p}, 1) \end{array}\right)^T \cdot \left(\frac{1}{y^2}V_2\right) \cdot \left(\begin{array}{c} \frac{\partial f}{\partial x}(1 + \frac{p}{n} + \frac{(\kappa+\beta-1)y}{p}, 1) \\ \frac{\partial f}{\partial y}(1 + \frac{p}{n} + \frac{(\kappa+\beta-1)y}{p}, 1) \end{array}\right)$$

$$\longrightarrow 2\kappa.$$

Therefore,

$$n\left(U - \frac{p}{n} - \frac{(\kappa + \beta - 1)y}{p}\right) \Longrightarrow N(0, 2\kappa),$$

that is,

$$nU - p \Longrightarrow N(\kappa + \beta - 1, 2\kappa).$$

The proof of Theorem 2.2 is complete. $\square$

Remark 2.1. Note that in Theorems 2.1 and 2.2 appears the parameter β, which is in practice unknown with real data. So we may estimate the parameter β using the fourth-order sample moment:

$$\widehat{\beta} = \frac{1}{np}\sum_{i=1}^{p}\sum_{j=1}^{n}|Y_j(i)|^4 - \kappa - 1.$$

According to the law of large numbers, $\hat{\beta} = \beta + o_p(1)$, so substituting $\hat{\beta}$ for β in Theorems 2.1 and 2.2 does not modify the limiting distribution.

3. Monte Carlo study

Monte Carlo simulations are conducted to find empirical sizes and powers of CLRT and CJ. In particular, here we want to examine the following questions: how robust are the tests against non-normal distributed data and what is the range of the dimension to sample ratio p/n where the tests are applicable.

For comparison, we show both the performance of the LW test using the asymptotic $N(1, 4)$ distribution in (2.2) (Notice however this is the CJ test under normal distribution) and the Chen's test (denoted as C for short) using the asymptotic $N(0, 4)$ distribution derived in [9]. The nominal test level is set to be $\alpha = 0.05$, and for each pair of (p, n), we run 10000 independent replications.

We consider two scenarios with respect to the random vectors Y_i:

TABLE 2
Empirical sizes of LW, CJ, CLRT and C test at 5% significance level based on 10000
independent applications with real $N(0,1)$ random variables and with real Gamma(4,2)-2
random variables

(p,n)	$N(0,1)$				Gamma(4,2)-2			
	LW/CJ	CLRT	C		LW	CLRT	CJ	C
(4,64)	0.0498	0.0553	0.0523		0.1396	0.074	0.0698	0.0717
(8,64)	0.0545	0.061	0.0572		0.1757	0.0721	0.0804	0.078
(16,64)	0.0539	0.0547	0.0577		0.1854	0.0614	0.078	0.0756
(32,64)	0.0558	0.0531	0.0612		0.1943	0.0564	0.0703	0.0682
(48,64)	0.0551	0.0522	0.0602		0.1956	0.0568	0.0685	0.0652
(56,64)	0.0547	0.0505	0.0596		0.1942	0.0549	0.0615	0.0603
(60,64)	0.0523	0.0587	0.0585		0.194	0.0582	0.0615	0.0603
(8,128)	0.0539	0.0546	0.0569		0.1732	0.0701	0.075	0.0754
(16,128)	0.0523	0.0534	0.0548		0.1859	0.0673	0.0724	0.0694
(32,128)	0.051	0.0545	0.0523		0.1951	0.0615	0.0695	0.0693
(64,128)	0.0538	0.0528	0.0552		0.1867	0.0485	0.0603	0.0597
(96,128)	0.055	0.0568	0.0581		0.1892	0.0539	0.0577	0.0579
(112,128)	0.0543	0.0522	0.0591		0.1875	0.0534	0.0591	0.0593
(120,128)	0.0545	0.0541	0.0561		0.1849	0.051	0.0598	0.0596
(16,256)	0.0544	0.055	0.0574		0.1898	0.0694	0.0719	0.0716
(32,256)	0.0534	0.0515	0.0553		0.1865	0.0574	0.0634	0.0614
(64,256)	0.0519	0.0537	0.0522		0.1869	0.0534	0.0598	0.0608
(128,256)	0.0507	0.0505	0.0498		0.1858	0.051	0.0555	0.0552
(192,256)	0.0507	0.054	0.0518		0.1862	0.0464	0.052	0.0535
(224,256)	0.0503	0.0541	0.0516		0.1837	0.0469	0.0541	0.0538
(240,256)	0.0494	0.053	0.0521		0.1831	0.049	0.0533	0.0559
(32,512)	0.0542	0.0543	0.0554		0.1884	0.0571	0.0606	0.059
(64,512)	0.0512	0.0497	0.0513		0.1816	0.0567	0.0579	0.0557
(128,512)	0.0519	0.0567	0.0533		0.1832	0.0491	0.0507	0.0504
(256,512)	0.0491	0.0503	0.0501		0.1801	0.0504	0.0495	0.0492
(384,512)	0.0487	0.0505	0.0499		0.1826	0.051	0.0502	0.0507
(448,512)	0.0496	0.0495	0.0503		0.1881	0.0526	0.0482	0.0485
(480,512)	0.0488	0.0511	0.0505		0.1801	0.0523	0.053	0.0516

(a) Y_i is p-dimensional real random vector from the multivariate normal population $\mathcal{N}(0, I_p)$. In this case, $\kappa = 2$ and $\beta = 0$.

(b) Y_i consists of iid real random variables with distribution $Gamma(4,2) - 2$ so that y_{ij} satisfies $Ey_{ij} = 0, Ey_{ij}^4 = 4.5$. In this case, $\kappa = 2$ and $\beta = 1.5$.

Table 2 reports the sizes of the four tests in these two scenarios for different values of (p,n). We see that when $\{y_{ij}\}$ are normal, LW (=CJ), CLRT and C all have similar empirical sizes tending to the nominal level 0.05 as either p or n increases. But when $\{y_{ij}\}$ are Gamma-distributed, the sizes of LW are higher than 0.1 no matter how large the values of p and n are while the sizes of CLRT and CJ all converge to the nominal level 0.05 as either p or n gets larger. This empirically confirms that normal assumptions are needed for the result of [15] while our corrected statistics CLRT and CJ (also the C test) have no such distributional restriction.

As for empirical powers, we consider two alternatives (here, the limiting spectral distributions of Σ_p under these two alternatives differs from that under H_0):

TABLE 3
Empirical powers of LW, CJ, CLRT and C test at 5% significance level based on 10000 independent applications with real $N(0,1)$ random variables and with real Gamma(4,2)-2 random variables under two alternatives Power 1 and 2 (see the text for details)

$N(0,1)$

(p,n)	Power 1				Power 2		
	LW/CJ	CLRT	C		LW/CJ	CLRT	C
(4,64)	0.7754	0.7919	0.772		0.4694	0.6052	0.4716
(8,64)	0.8662	0.8729	0.8582		0.5313	0.6756	0.5308
(16,64)	0.912	0.9075	0.9029		0.5732	0.6889	0.5671
(32,64)	0.9384	0.8791	0.931		0.5868	0.6238	0.5775
(48,64)	0.9471	0.7767	0.9389		0.6035	0.5036	0.5982
(56,64)	0.949	0.6663	0.9411		0.6025	0.4055	0.5982
(60,64)	0.9501	0.5575	0.941		0.6048	0.3328	0.5989
(8,128)	0.9984	0.9989	0.9986		0.9424	0.9776	0.9391
(16,128)	0.9998	1	0.9998		0.9698	0.9926	0.9676
(32,128)	1	1	1		0.9781	0.9956	0.9747
(64,128)	1	1	1		0.9823	0.9897	0.9788
(96,128)	1	0.9996	1		0.9824	0.9532	0.9804
(112,128)	1	0.9943	1		0.9841	0.881	0.9808
(120,128)	1	0.9746	1		0.9844	0.7953	0.9817

$Gamma(4,2)-2$

(p,n)	Power 1				Power 2		
	CJ	CLRT	C		CJ	CLRT	C
(4,64)	0.6517	0.6826	0.6628		0.3998	0.5188	0.4204
(8,64)	0.7693	0.7916	0.781		0.4757	0.5927	0.4889
(16,64)	0.8464	0.8439	0.846		0.5327	0.633	0.5318
(32,64)	0.9041	0.848	0.9032		0.5805	0.5966	0.5667
(48,64)	0.9245	0.7606	0.924		0.5817	0.4914	0.5804
(56,64)	0.9267	0.6516	0.9247		0.5882	0.4078	0.583
(60,64)	0.9288	0.5547	0.9257		0.5919	0.3372	0.5848
(8,128)	0.9859	0.9875	0.9873		0.8704	0.9294	0.8748
(16,128)	0.999	0.999	0.9987		0.9276	0.9699	0.9311
(32,128)	0.9999	1	0.9999		0.9582	0.9873	0.9587
(64,128)	1	0.9998	1		0.9729	0.984	0.9727
(96,128)	1	0.999	1		0.9771	0.9482	0.9763
(112,128)	1	0.9924	1		0.9781	0.8747	0.9763
(120,128)	1	0.9728	1		0.9786	0.7864	0.977

(1) Σ_p is diagonal with half of its diagonal elements 0.5 and half 1. We denote its power by Power 1;

(2) Σ_p is diagonal with 1/4 of the elements equal 0.5 and 3/4 equal 1. We denote its power by Power 2.

Table 3 reports the powers of LW(=CJ), CLRT and C when $\{y_{ij}\}$ are distributed as $N(0,1)$, and of CJ, CLRT and C when $\{y_{ij}\}$ are distributed as Gamma(4,2)-2, for the situation when n equals 64 or 128, with varying values of p and under the above mentioned two alternatives. For $n = 256$ and p varying from 16 to 240, all the tests have powers around 1 under both alternatives so that these values are omitted. And in order to find the trend of these powers, we also present the results when $n = 128$ in Figure 1 and Figure 2.
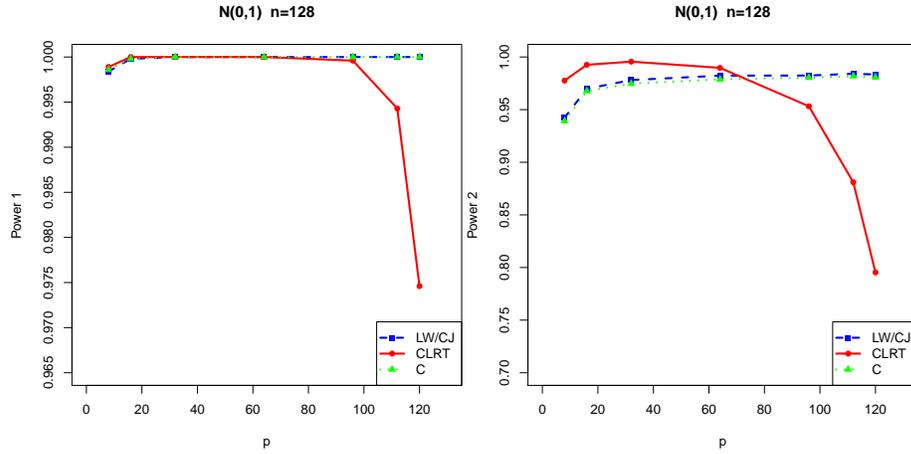
FIG 1. *Empirical powers of LW/CJ, CLRT and C test at* 5% *significance level based on 10000 independent applications with real N(0,1) random variables for fixed* $n = 128$ *under two alternatives Power 1 and 2 (see the text for details).*
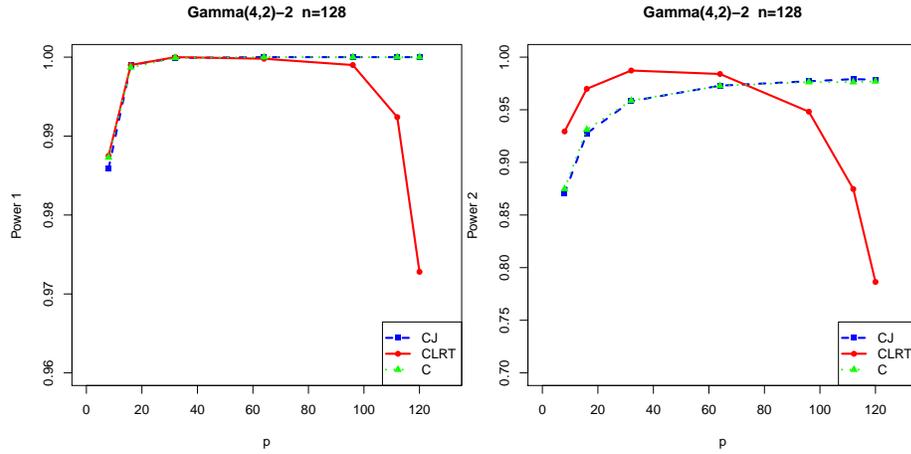

FIG 2. *Empirical powers of CJ, CLRT and C test at* 5% *significance level based on 10000 independent applications with real Gamma(4,2)-2 random variables for fixed* $n = 128$ *under two alternatives Power 1 and 2 (see the text for details).*

The behavior of Power 1 and Power 2 in each figure related to the three statistics are similar, except that Power 1 is much higher compared with Power 2 for a given dimension design (p, n) and any given test for the reason that the first alternative differs more from the null than the second one. The powers of

TABLE 4
Empirical sizes and powers (Power 1 and 2) of CJ test and C test at 5% significance level based on 10000 independent applications with real Gamma(4,2)-2 random variables when $p \geq n$

(p,n)	CJ				C		
	Size	Power 1	Power 2		Size	Power 1	Power 2
(64,64)	0.0624	0.9282	0.5897		0.0624	0.9257	0.5821
(320,64)	0.0577	0.9526	0.612		0.0576	0.9472	0.6059
(640,64)	0.0558	0.959	0.6273		0.0562	0.9541	0.6105
(960,64)	0.0543	0.9631	0.6259		0.0551	0.955	0.6153
(1280,64)	0.0555	0.9607	0.6192		0.0577	0.9544	0.6067

LW (in the normal case), CJ (in the Gamma case) and C are all monotonically increasing in p for a fixed value of n. But for CLRT, when n is fixed, the powers first increase in p and then become decreasing when p is getting close to n. This can be explained by the fact that when p is close to n, some of the eigenvalues of S_n are getting close to zero, causing the CLRT nearly degenerate and losing power.

Besides, we find that in the normal case the trend of C's power is very much alike of those of LW while in the Gamma case it is similar with those of CJ under both alternatives. And in most of the cases (especially in large p case), the power of C test is slightly lower than LW (in the normal case) and CJ (in the Gamma case).

Lastly, we examine the performance of CJ and C when p is larger than n. Empirical sizes and powers are presented in Table 4. We choose the variables to be distributed as Gamma(4,2)-2 since CJ reduces to LW in the normal case, and [15] has already reported the performance of LW when p is larger than n. From the table, we see that when p is larger than n, the size of CJ is still correct and it is always around the nominal level 0.05 as the dimension p increases and the same phenomenon exists for C test.

When we evaluate the power, the same two alternatives Power 1 and Power 2 as above are considered. The sample size is fixed to $n = 64$ and the ratio p/n varies from 1 to 20. We see that Power 1 are much higher than Power 2 for the same reason that the first alternative is easier to be distinguished from H_0. Besides, the powers under both alternatives all increase monotonically for $1 \leq \frac{p}{n} \leq 15$. However, when p/n is getting larger, say $p/n = 20$, we can observe that its size is a little larger and powers a little drop (compared with $p/n = 15$) but overall, it still behaves well, which can be considered as free from the assumption constraint "$p/n \to y$". Besides, the powers of CJ are always slightly higher than those of C in this "large p small n" setting.

Since the asymptotic distribution for the CLRT and CJ are both derived under the "Marcenko-Pasture scheme" (i.e $p/n \to y \in (0, \infty)$), if p/n is getting too large ($p \gg n$), it seems that the limiting results provided in this paper will loose accuracy. It is worth noticing that [7] has extended the LW test to such a scheme ($p \gg n$) for multivariate normal distribution.

Summarizing all these findings from this Monte-Carlo study, the overall figure is the following: when the ratio p/n is much lower than 1 (say smaller than $1/2$),

it is preferable to employ CLRT (than CJ, LW or C); while this ratio is higher, CJ (or LW for normal data) becomes more powerful (slightly more powerful than C).

4. Asymptotic powers: under the spiked population alternative

In this section, we give an analysis of the powers of the two corrections: CLRT and CJ. To this end, we consider an alternative model that has attracted lots of attention since its introduction by [13], namely, the spiked population model. This model can be described as follows: the eigenvalues of Σ_p are all one except for a few fixed number of them. Thus, we restrict our sphericity testing problem to the following:

$$H_0 : \Sigma_p = I_p \quad vs \quad H_1^* : \Sigma_p = diag(\underbrace{a_1, \ldots, a_1}_{n_1}, \ldots, \underbrace{a_k, \ldots, a_k}_{n_k}, \underbrace{1, \ldots, 1}_{p-M}),$$

where the multiplicity numbers n_i's are fixed and satisfying $n_1 + \cdots + n_k = M$. We derive the explicit expressions of the power functions of CLRT and CJ in this section.

Under H_1^*, the empirical spectral distribution of Σ_p is

$$H_n = \frac{1}{p} \sum_{i=1}^{k} n_i \delta_{a_i} + \frac{p - M}{p} \delta_1, \tag{4.1}$$

and it will converge to δ_1, a Dirac mass at 1, which is the same as the limit under the null hypothesis $H_0 : \Sigma_p = I_p$. From this point of view, anything related to the limiting spectral distribution remains the same whenever under H_0 or H_1^*. Then recall the CLT for LSS of the sample covariance matrix, as provided in [2], is of the form:

$$p \int f(x) d(F^{S_n}(x) - F^{y_n}(x)) \Longrightarrow N(\mu, \sigma^2),$$

where the right side of this equation is determined only by the limiting spectral distribution. So we can conclude that the limiting parameters μ and σ^2 remain the same under H_0 and H_1^*, only the centering term $p \int f(x) dF^{y_n}(x)$ possibly makes a difference. Since there's a p in front of $\int f(x) dF^{y_n}(x)$, which tends to infinity as assumed, so knowing the convergence $H_n \to \delta_1$ is not enough and more details about the convergence are needed. In [28], we have established an asymptotic expansion for the centering parameter when the population has a spiked structure. We will use these formulas like equations (4.2), (4.3) and (4.6) in the following to derive the powers of the CLRT and CJ.

Lemma 2.1 remains the same under H_1^*, except that this time the centering terms become (see formulas (4.11) and (4.12) in [28]):

$$F^{y_n}(\log x) = \frac{1}{p} \sum_{i=1}^{k} n_i \log a_i - 1 + \left(1 - \frac{1}{y_n}\right) \log(1 - y_n) + O\left(\frac{1}{p^2}\right), \quad (4.2)$$

$$F^{y_n}(x) = 1 + \frac{1}{p} \sum_{i=1}^{k} n_i a_i - \frac{M}{p} + O\left(\frac{1}{p^2}\right). \tag{4.3}$$

Repeating the proof of Theorem 2.1, we can get:

$$\mathcal{L}_n - p + (p-n)\log\left(1 - \frac{p}{n}\right) + \sum_{i=1}^{k} n_i\left(\log a_i - a_i + 1\right)$$

$$\implies N\left(-\frac{\kappa - 1}{2}\log(1-y) + \frac{1}{2}\beta y, \ -\kappa\log(1-y) - \kappa y\right) \tag{4.4}$$

under H_1^*. As a result, the power of CLRT for testing H_0 against H_1^* can be expressed as:

$$\beta_1(\alpha) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sum_{i=1}^{k} n_i\left(a_i - \log a_i - 1\right)}{\sqrt{-\kappa\log(1-y) - \kappa y}}\right) \tag{4.5}$$

for a pre-given significance level α.

It is worth noticing here that if the alternative has only one simple spike, i.e. $k = 1$, $n_k = 1$, and assuming the real Gaussian variable case, i.e. $\kappa = 2$, (4.5) reduces to a result provided in [20]. However, our formula is valid for a general number of spikes with eventual multiplicities. Besides, these authors use some more sophisticated tools of asymptotic contiguity and Le Cam's first and third lemmas, which are totally different from ours.

In order to calculate the power function of CJ, we restated Lemma 2.2 as follows:

$$\begin{pmatrix} \sum_{i=1}^{p} l_i^2 - pF^{y_n}(x^2) \\ \sum_{i=1}^{p} l_i - pF^{y_n}(x) \end{pmatrix} \implies N(\mu_2, V_2),$$

where

$$F^{y_n}(x^2) = \frac{2}{n} \sum_{i=1}^{k} n_i a_i - \frac{2}{n} M + 1 + y_n - \frac{M}{p} + \frac{1}{p} \sum_{i=1}^{k} n_i a_i^2 + O\left(\frac{1}{p^2}\right), \tag{4.6}$$

$$F^{y_n}(x) = 1 + \frac{1}{p} \sum_{i=1}^{k} n_i a_i - \frac{M}{p} + O\left(\frac{1}{p^2}\right).$$

Using the delta method as in the proof of Theorem 2.2, this time, we have

$$nU - p - \frac{n}{p} \sum_{i=1}^{k} n_i(a_i - 1)^2 \implies N(\kappa + \beta - 1, 2\kappa) \tag{4.7}$$

under H_1^*.

Power function of CJ can be expressed as

$$\beta_2(\alpha) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{n}{p} \cdot \frac{\sum_{i=1}^{k} n_i(a_i - 1)^2}{\sqrt{2\kappa}}\right) \tag{4.8}$$

for a pre-given significance level α.
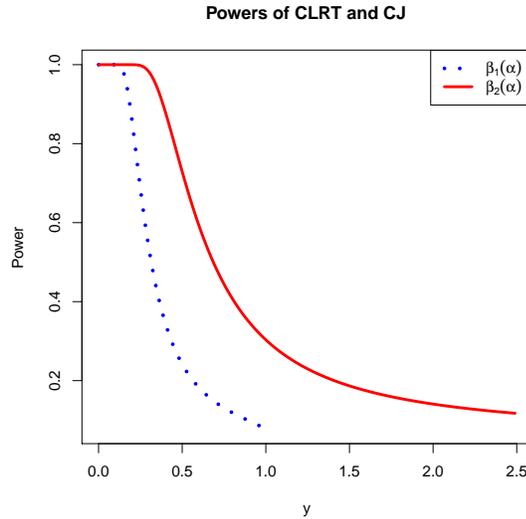
Powers of CLRT and CJ

FIG 3. *Theoretical powers of CLRT ($\beta_1(\alpha)$) and CJ ($\beta_2(\alpha)$) under the spiked population alternative.*

Now consider the functions $a_i - \log a_i - 1$ and $(a_i - 1)^2$ appearing in the expressions (4.5) and (4.8), they will achieve their minimum value 0 at $a_i = 1$, which is to say, once a_i's going away from 1, the powers $\beta_1(\alpha)$ and $\beta_2(\alpha)$ will both increase. This phenomenon agrees with our intuition, since the more a_i's deviate away from 1, the easier to distinguish H_0 from H_1^*. Therefore, the powers should naturally grow higher.

Then, we consider the power functions $\beta_1(\alpha)$ and $\beta_2(\alpha)$ as functions of y, and see how they are going along with $y's$ changing. we see that in expression (4.5), $\sqrt{-\log(1-y) - y}$ is increasing when $y \in (0,1)$, so $\beta_1(\alpha)$ is decreasing as the function of y, which attains its maximum value 1 when $y \to 0_+$ and minimum value α when $y \to 1_-$. Also, expression (4.8) is obviously a decreasing function of y, attaining its maximum value 1 when $y \to 0_+$ and minimum value α when $y \to +\infty$. We present the trends of $\beta_1(\alpha)$ and $\beta_2(\alpha)$ (corresponding to the power of CLRT and CJ) in Figure 3 when only one spike $a = 2.5$ exists. It is however a little different from the non-spiked case as showed in the simulation in Section 3 (both Figure 1 and Figure 2), where the power of CLRT first increases then decreases while the power of CJ is always increasing along with the increase of the value of p. These power drops are due to the fact that when p increases, since only one spike eigenvalue is considered, it becomes harder to distinguish both hypotheses. Besides, an interesting finding here is that these power functions give a new confirmation of the fact that CLRT behaves quite badly when $y \to 1_-$, while CJ test has a reasonable power for a significant range of $y > 1$.

5. Generalization to the case when the population mean is unknown

So far, we have assumed the observation (Y_i) are centered. However, this is hardly true in practical situations when $\mu = EY_i$ is usually unknown. Therefore, the sample covariance matrix should be taken as

$$S_n^* = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})(Y_i - \overline{Y})^*,$$

where $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ is the sample mean. Because $\overline{YY}^*$ is a rank one matrix, substituting S_n for S_n^* when μ is unknown will not affect the limiting distribution in the CLT for LSS; while it is not the case for the centering parameter, for it has a p in front.

Recently, [22] shows that if we use S_n^* in the CLT for LSS when μ is unknown, the limiting variance remains the same as we use S_n; while the limiting mean has a shift which can be expressed as a complex contour integral. Later, [30] looks into this shift and finally derives a concise conclusion on the CLT corresponding to S_n^*: the random vector $\big(X_n^*(f_1), \ldots, X_n^*(f_k)\big)$ converges weakly to a Gaussian vector with the same mean and covariance function as given in Theorem A.1, where this time, $X_n^*(f) = p \int f(x)d(F^{S_n^*} - F^{y_{n-1}})$. It is here important to pay attention that the only difference is in the centering term, where we use the new ratio $y_{n-1} = \frac{p}{n-1}$ instead of the previous $y_n = \frac{p}{n}$, while leaving all the other terms unchanged.

Using this result, we can modify our Theorems 2.1 and 2.2 to get the CLT of CLRT and CJ under H_0 when μ is unknown only by considering the eigenvalues of S_n^* and substituting $n-1$ for n in the centering terms. More precisely, now equations (2.1) and (2.3) in Theorems 2.1 and 2.2 become

$$\mathcal{L}_n + (p - n + 1) \cdot \log\left(1 - \frac{p}{n-1}\right) - p$$
$$\Longrightarrow N\left\{-\frac{\kappa - 1}{2}\log(1-y) + \frac{1}{2}\beta y, -\kappa\log(1-y) - \kappa y\right\}$$

and

$$nU - \frac{np}{n-1} \Longrightarrow N(\kappa + \beta - 1, 2\kappa).$$

The same procedures can be applied to get the CLT of CLRT and CJ under H_1^* when μ is unknown. This time, equations (4.4) and (4.7) become

$$\mathcal{L}_n - p + (p - n + 1)\log\left(1 - \frac{p}{n-1}\right) + \sum_{i=1}^{k} n_i\big(\log a_i - a_i + 1\big)$$
$$\Longrightarrow N\left(-\frac{\kappa - 1}{2}\log(1-y) + \frac{1}{2}\beta y, \ -\kappa\log(1-y) - \kappa y\right)$$

and

$$nU - \frac{np}{n-1} - \frac{n}{p}\sum_{i=1}^{k} n_i(a_i - 1)^2 \Longrightarrow N(\kappa + \beta - 1, 2\kappa),$$

and therefore the powers of CLRT and CJ under the spiked alternative remain
unchanged as expressed in (4.5) and (4.8).

6. Additional proofs

We recall these two important formulas which appear in the Appendix as (A.2)
and (A.3) here for the convenience of reading:

$$
\begin{aligned}
\mathbb{E}[X_f] &= (\kappa - 1)I_1(f) + \beta I_2(f), \\
\mathrm{Cov}(X_f, X_g) &= \kappa J_1(f, g) + \beta J_2(f, g).
\end{aligned}
$$

6.1. Proof of Lemma 2.1

Let for $x > 0$, $f(x) = \log x$ and $g(x) = x$. Define A_n and B_n by the decompositions

$$
\sum_{i=1}^{p} \log \ell_i = p \int f(x) d(F_n(x) - F^{y_n}(x)) + pF^{y_n}(f) = A_n + pF^{y_n}(f),
$$

$$
\sum_{i=1}^{p} \ell_i = p \int g(x) d(F_n(x) - F^{y_n}(x)) + pF^{y_n}(g) = B_n + pF^{y_n}(g).
$$

Applying Theorem A.1 given in the Appendix to the pair (f, g), we have

$$
\begin{pmatrix} A_n \\ B_n \end{pmatrix} \Longrightarrow N\left(\begin{pmatrix} EX_f \\ EX_g \end{pmatrix}, \begin{pmatrix} \mathrm{Cov}(X_f, X_f) & \mathrm{Cov}(X_f, X_g) \\ \mathrm{Cov}(X_g, X_f) & \mathrm{Cov}(X_g, X_g) \end{pmatrix} \right).
$$

It remains to evaluate the limiting parameters and this results from the following
calculations where h is denoted as $\sqrt{y}$:

$$
\begin{aligned}
I_1(f, r) &= \frac{1}{2} \log \left(1 - h^2/r^2\right), & (6.1) \\
I_1(g, r) &= 0, & (6.2) \\
I_2(f) &= -\frac{1}{2} h^2, & (6.3) \\
I_2(g) &= 0, & (6.4) \\
J_1(f, g, r) &= \frac{h^2}{r^2}, & (6.5) \\
J_1(f, f, r) &= -\frac{1}{r} \log(1 - h^2/r), & (6.6) \\
J_1(g, g, r) &= \frac{h^2}{r^2}, & (6.7) \\
J_2(f, g) &= h^2, & (6.8) \\
J_2(f, f) &= h^2, & (6.9) \\
J_2(g, g) &= h^2. & (6.10)
\end{aligned}
$$

We now detail these calculations to complete the proof. They are all based on the formula given in Proposition A.1 in the Appendix and repeated use of the residue theorem.

Proof of (6.1). We have

$$
\begin{aligned}
I_1(f, r) &= \frac{1}{2\pi i} \oint_{|\xi|=1} f(|1 + h\xi|^2) \left[\frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right] d\xi \\
&= \frac{1}{2\pi i} \oint_{|\xi|=1} \log(|1 + h\xi|^2) \left[\frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right] d\xi \\
&= \frac{1}{2\pi i} \oint_{|\xi|=1} \left(\frac{1}{2} \log((1 + h\xi)^2) + \frac{1}{2} \log((1 + h\xi^{-1})^2) \right) \left[\frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right] d\xi \\
&= \frac{1}{2\pi i} \Bigg[\oint_{|\xi|=1} \log(1 + h\xi) \frac{\xi}{\xi^2 - r^{-2}} d\xi - \oint_{|\xi|=1} \log(1 + h\xi) \frac{1}{\xi} d\xi \\
&\quad + \oint_{|\xi|=1} \log(1 + h\xi^{-1}) \frac{\xi}{\xi^2 - r^{-2}} d\xi - \oint_{|\xi|=1} \log(1 + h\xi^{-1}) \frac{1}{\xi} d\xi \Bigg]
\end{aligned}
$$

For the first integral, note that as $r > 1$, the poles are $\pm \frac{1}{r}$ and we have by the residue theorem,

$$
\begin{aligned}
&\frac{1}{2\pi i} \oint_{|\xi|=1} \log(1 + h\xi) \frac{\xi}{\xi^2 - r^{-2}} d\xi \\
&= \left. \frac{\log(1 + h\xi) \cdot \xi}{\xi - r^{-1}} \right|_{\xi=-r^{-1}} + \left. \frac{\log(1 + h\xi) \cdot \xi}{\xi + r^{-1}} \right|_{\xi=r^{-1}} \\
&= \frac{1}{2} \log\left(1 - \frac{h^2}{r^2} \right).
\end{aligned}
$$

For the second integral,

$$
\frac{1}{2\pi i} \oint_{|\xi|=1} \log(1 + h\xi) \frac{1}{\xi} d\xi = \log(1 + h\xi) \Big|_{\xi=0} = 0.
$$

The third one is

$$
\begin{aligned}
&\frac{1}{2\pi i} \oint_{|\xi|=1} \log(1 + h\xi^{-1}) \frac{\xi}{\xi^2 - r^{-2}} d\xi \\
&= -\frac{1}{2\pi i} \oint_{|z|=1} \log(1 + hz) \frac{z^{-1}}{z^{-2} - r^{-2}} \cdot \frac{-1}{z^2} dz \\
&= \frac{1}{2\pi i} \oint_{|z|=1} \frac{\log(1 + hz) r^2}{z(z + r)(z - r)} dz = \left. \frac{\log(1 + hz) r^2}{(z + r)(z - r)} \right|_{z=0} = 0,
\end{aligned}
$$

where the first equality results from the change of variable $z = \frac{1}{\xi}$, and the third equality holds because $r > 1$, so the only pole is $z = 0$.

The fourth one equals

$$\frac{1}{2\pi i}\oint_{|\xi|=1}\log(1+h\xi^{-1})\frac{1}{\xi}d\xi = -\frac{1}{2\pi i}\oint_{|z|=1}\log(1+hz)\frac{-z}{z^2}dz$$
$$= \log(1+hz)\big|_{z=0} = 0.$$

Collecting the four integrals leads to the desired formula for $I_1(f,r)$.

Proof of (6.2). We have

$$
\begin{aligned}
I_1(g,r) &= \frac{1}{2\pi i}\oint_{|\xi|=1}g(|1+h\xi|^2)\cdot\left[\frac{\xi}{\xi^2-r^{-2}}-\frac{1}{\xi}\right]d\xi \\
&= \frac{1}{2\pi i}\oint_{|\xi|=1}|1+h\xi|^2\cdot\left[\frac{\xi}{\xi^2-r^{-2}}-\frac{1}{\xi}\right]d\xi \\
&= \frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\xi+h+h\xi^2+h^2\xi}{\xi}\cdot\left[\frac{\xi}{\xi^2-r^{-2}}-\frac{1}{\xi}\right]d\xi \\
&= \frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\xi+h+h\xi^2+h^2\xi}{\xi^2-r^{-2}}d\xi \\
&\quad -\frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\xi+h+h\xi^2+h^2\xi}{\xi^2}d\xi.
\end{aligned}
$$

These two integrals are calculated as follows:

$$
\begin{aligned}
&\frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\xi+h+h\xi^2+h^2\xi}{\xi^2-r^{-2}}d\xi \\
&= \frac{\xi+h+h\xi^2+h^2\xi}{\xi-r^{-1}}\bigg|_{\xi=-r^{-1}}+\frac{\xi+h+h\xi^2+h^2\xi}{\xi+r^{-1}}\bigg|_{\xi=r^{-1}} \\
&= 1+h^2;
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\xi+h+h\xi^2+h^2\xi}{\xi^2}d\xi = \frac{\partial}{\partial\xi}(\xi+h+h\xi^2+h^2\xi)\bigg|_{\xi=0} \\
&= 1+h^2.
\end{aligned}
$$

Collecting the two terms leads to $I_1(g,r)=0$.

Proof of (6.3).

$$
\begin{aligned}
I_2(f) &= \frac{1}{2\pi i}\oint_{|\xi|=1}\log(|1+h\xi|^2)\frac{1}{\xi^3}d\xi \\
&= \frac{1}{2\pi i}\left[\oint_{|\xi|=1}\frac{\log(1+h\xi)}{\xi^3}d\xi+\oint_{|\xi|=1}\frac{\log(1+h\xi^{-1})}{\xi^3}d\xi\right].
\end{aligned}
$$

We have

$$\frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\log(1+h\xi)}{\xi^3}d\xi = \frac{1}{2}\frac{\partial^2}{\partial\xi^2}\log(1+h\xi)\bigg|_{\xi=0} = -\frac{1}{2}h^2;$$

$$\frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\log(1+h\xi^{-1})}{\xi^3}d\xi = -\frac{1}{2\pi i}\oint_{|z|=1}\frac{\log(1+hz)}{\frac{1}{z^3}}\cdot\frac{-1}{z^2}dz = 0.$$

Combining the two leads to $I_2(f) = -\frac{1}{2}h^2$.

Proof of (6.4).

$$I_2(g) = \frac{1}{2\pi i}\oint_{|\xi|=1}\frac{(1+h\overline{\xi})(1+h\xi)}{\xi^3}d\xi = \frac{1}{2\pi i}\oint_{|\xi|=1}\frac{\xi+h\xi^2+h+h^2\xi}{\xi^4}d\xi = 0.$$

Proof of (6.5).

$$J_1(f,g,r) = \frac{1}{2\pi i}\oint_{|\xi_2|=1}|1+h\xi_2|^2\cdot\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(|1+h\xi_1|^2)}{(\xi_1-r\xi_2)^2}d\xi_1 d\xi_2.$$

We have,

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(|1+h\xi_1|^2)}{(\xi_1-r\xi_2)^2}d\xi_1$$

$$= \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(1+h\xi_1)}{(\xi_1-r\xi_2)^2}d\xi_1 + \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(1+h\xi_1^{-1})}{(\xi_1-r\xi_2)^2}d\xi_1.$$

The first term

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(1+h\xi_1)}{(\xi_1-r\xi_2)^2}d\xi_1 = 0,$$

because for fixed $|\xi_2| = 1$, $|r\xi_2| = |r| > 1$, so $r\xi_2$ is not a pole.

The second term is

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(1+h\xi_1^{-1})}{(\xi_1-r\xi_2)^2}d\xi_1 = -\frac{1}{2\pi i}\oint_{|z|=1}\frac{\log(1+hz)}{(\frac{1}{z}-r\xi_2)^2}\cdot\frac{-1}{z^2}dz$$

$$= \frac{1}{2\pi i}\cdot\frac{1}{(r\xi_2)^2}\oint_{|z|=1}\frac{\log(1+hz)}{(z-\frac{1}{r\xi_2})^2}dz = \frac{1}{(r\xi_2)^2}\cdot\frac{\partial}{\partial z}\log(1+hz)\Big|_{z=\frac{1}{r\xi_2}}$$

$$= \frac{h}{r\xi_2(r\xi_2+h)},$$

where the first equality results from the change of variable $z = \frac{1}{\xi_1}$, and the third equality holds because for fixed $|\xi_2| = 1$, $|\frac{1}{r\xi_2}| = \frac{1}{|r|} < 1$, so $\frac{1}{r\xi_2}$ is a pole of second order.

Therefore,

$$J_1(f,g,r)$$

$$= \frac{h}{2\pi ir^2}\oint_{|\xi_2|=1}\frac{(1+h\xi_2)(1+h\overline{\xi_2})}{\xi_2(\xi_2+\frac{h}{r})}d\xi_2$$

$$= \frac{h}{2\pi ir^2}\oint_{|\xi_2|=1}\frac{\xi_2+h\xi_2^2+h+h^2\xi_2}{\xi_2^2(\xi_2+\frac{h}{r})}d\xi_2$$

$$= \frac{h}{2\pi i r^2} \left[\oint_{|\xi_2|=1} \frac{1+h^2}{\xi_2(\xi_2 + \frac{h}{r})} d\xi_2 + \oint_{|\xi_2|=1} \frac{h}{\xi_2 + \frac{h}{r}} d\xi_2 + \oint_{|\xi_2|=1} \frac{h}{\xi_2^2(\xi_2 + \frac{h}{r})} d\xi_2 \right].$$

Finally we find $J_1(f, g, r) = \frac{h^2}{r^2}$ since

$$\frac{h}{2\pi i r^2} \oint_{|\xi_2|=1} \frac{1+h^2}{\xi_2(\xi_2 + \frac{h}{r})} d\xi_2 = 0, \quad \frac{h}{2\pi i r^2} \oint_{|\xi_2|=1} \frac{h}{\xi_2 + \frac{h}{r}} d\xi_2 = \frac{h^2}{r^2},$$

$$\frac{h}{2\pi i r^2} \oint_{|\xi_2|=1} \frac{h}{\xi_2^2(\xi_2 + \frac{h}{r})} d\xi_2 = 0.$$

Proof of (6.6).

$$
\begin{aligned}
J_1(f, f, r) &= \frac{1}{2\pi i} \oint_{|\xi_2|=1} f(|1 + h\xi_2|^2) \cdot \frac{1}{2\pi i} \oint_{|\xi_1|=1} \frac{f(|1 + h\xi_1|^2)}{(\xi_1 - r\xi_2)^2} d\xi_1 d\xi_2 \\
&= \frac{1}{2\pi i} \oint_{|\xi_2|=1} f(|1 + h\xi_2|^2) \frac{h}{r\xi_2(r\xi_2 + h)} d\xi_2 \\
&= \frac{h}{2\pi i r^2} \oint_{|\xi_2|=1} \frac{\log(1 + h\xi_2)}{\xi_2(\frac{h}{r} + \xi_2)} d\xi_2 + \frac{h}{2\pi i r^2} \oint_{|\xi_2|=1} \frac{\log(1 + h\xi_2^{-1})}{\xi_2(\frac{h}{r} + \xi_2)} d\xi_2.
\end{aligned}
$$

We have

$$
\begin{aligned}
&\frac{h}{2\pi i r^2} \oint_{|\xi_2|=1} \frac{\log(1 + h\xi_2)}{\xi_2(\frac{h}{r} + \xi_2)} d\xi_2 \\
&= \frac{h}{r^2} \left[\frac{\log(1 + h\xi_2)}{\frac{h}{r} + \xi_2} \bigg|_{\xi_2=0} + \frac{\log(1 + h\xi_2)}{\xi_2} \bigg|_{\xi_2=-\frac{h}{r}} \right] \\
&= -\frac{1}{r} \log\left(1 - \frac{h^2}{r}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{h}{2\pi i r^2} \oint_{|\xi_2|=1} \frac{\log(1 + h\xi_2^{-1})}{\xi_2(\frac{h}{r} + \xi_2)} d\xi_2 = \frac{-h}{2\pi i r^2} \oint_{|z|=1} \frac{\log(1 + hz)}{\frac{1}{z}(\frac{h}{r} + \frac{1}{z})} \cdot \frac{-1}{z^2} dz \\
&= \frac{1}{2\pi i r} \oint_{|z|=1} \frac{\log(1 + hz)}{z + \frac{r}{h}} dz = 0,
\end{aligned}
$$

where the first equality results from the change of variable $z = \frac{1}{\xi_2}$, and the third equality holds because $|\frac{r}{h}| > 1$, so $\frac{r}{h}$ is not a pole.

Finally, we find $J_1(f, f, r) = -\frac{1}{r} \log(1 - \frac{h^2}{r})$.

Proof of (6.7).

$$J_1(g, g, r) = \frac{1}{2\pi i} \oint_{|\xi_2|=1} |1 + h\xi_2|^2 \cdot \frac{1}{2\pi i} \oint_{|\xi_1|=1} \frac{|1 + h\xi_1|^2}{(\xi_1 - r\xi_2)^2} d\xi_1 d\xi_2.$$

We have

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{|1+h\xi_1|^2}{(\xi_1-r\xi_2)^2}d\xi_1 = \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\xi_1+h\xi_1^2+h+h^2\xi_1}{\xi_1(\xi_1-r\xi_2)^2}d\xi_1$$

$$= \frac{1}{2\pi i}\left[\oint_{|\xi_1|=1}\frac{1+h^2}{(\xi_1-r\xi_2)^2}d\xi_1 + \oint_{|\xi_1|=1}\frac{h\xi_1}{(\xi_1-r\xi_2)^2}d\xi_1\right.$$

$$\left. + \oint_{|\xi_1|=1}\frac{h}{\xi_1(\xi_1-r\xi_2)^2}d\xi_1\right]$$

$$= \frac{h}{r^2\xi_2^2},$$

since

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{1+h^2}{(\xi_1-r\xi_2)^2}d\xi_1 = 0, \quad \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{h\xi_1}{(\xi_1-r\xi_2)^2}d\xi_1 = 0,$$

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{h}{\xi_1(\xi_1-r\xi_2)^2}d\xi_1 = \frac{h}{(\xi_1-r\xi_2)^2}\bigg|_{\xi_1=0} = \frac{h}{r^2\xi_2^2},$$

where the equality above holds because for fixed $|\xi_2|=1$, $|r\xi_2|=|r|>1$, so $r\xi_2$ is not a pole. Therefore,

$$
\begin{aligned}
J_1(g,g,r) &= \frac{h}{2\pi i r^2}\oint_{|\xi_2|=1}\frac{\xi_2+h\xi_2^2+h+h^2\xi_2}{\xi_2^3}d\xi_2 \\
&= \frac{h}{2\pi i r^2}\left[\oint_{|\xi_2|=1}\frac{1+h^2}{\xi_2^2}d\xi_2 + \oint_{|\xi_2|=1}\frac{h}{\xi_2}d\xi_2 + \oint_{|\xi_2|=1}\frac{h}{\xi_2^3}d\xi_2\right], \\
&= \frac{h^2}{r^2}.
\end{aligned}
$$

Proof of (6.8), (6.9), (6.10). We have

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{f(|1+h\xi_1|^2)}{\xi_1^2}d\xi_1 = \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(|1+h\xi_1|^2)}{\xi_1^2}d\xi_1$$

$$= \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(1+h\xi_1)+\log(1+h\xi_1^{-1})}{\xi_1^2}d\xi_1 = h,$$

since

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(1+h\xi_1)}{\xi_1^2}d\xi_1 = \frac{\partial}{\partial\xi_1}\Big(\log(1+h\xi_1)\Big)\bigg|_{\xi_1=0} = h,$$

$$\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\log(1+h\xi_1^{-1})}{\xi_1^2}d\xi_1 = -\frac{1}{2\pi i}\oint_{|z|=1}\frac{\log(1+hz)}{\frac{1}{z^2}}\cdot\Big(-\frac{1}{z^2}dz\Big)$$

$$= \frac{1}{2\pi i}\oint_{|z|=1}\log(1+hz)dz = 0.$$

Similarly,

$$\frac{1}{2\pi i}\oint_{|\xi_2|=1}\frac{g(|1+h\xi_2|^2)}{\xi_2^2}d\xi_2 = \frac{1}{2\pi i}\oint_{|\xi_2|=1}\frac{\xi_2+h\xi_2^2+h+h^2\xi_2}{\xi_2^3}d\xi_2 = h.$$

Therefore,

$$J_2(f,g) = \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{f(|1+h\xi_1|^2)}{\xi_1^2}d\xi_1 \cdot \frac{1}{2\pi i}\oint_{|\xi_2|=1}\frac{g(|1+h\xi_2|^2)}{\xi_2^2}d\xi_2 = h^2,$$

$$J_2(f,f) = \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{f(|1+h\xi_1|^2)}{\xi_1^2}d\xi_1 \cdot \frac{1}{2\pi i}\oint_{|\xi_2|=1}\frac{f(|1+h\xi_2|^2)}{\xi_2^2}d\xi_2 = h^2,$$

$$J_2(g,g) = \frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{g(|1+h\xi_1|^2)}{\xi_1^2}d\xi_1 \cdot \frac{1}{2\pi i}\oint_{|\xi_2|=1}\frac{g(|1+h\xi_2|^2)}{\xi_2^2}d\xi_2 = h^2.$$

The proof of Lemma 2.1 is complete.

6.2. *Proof of Lemma 2.2*

Let $f(x) = x^2$ and $g(x) = x$. Define C_n and B_n by the decompositions

$$\sum_{i=1}^{p}\ell_i^2 = p\int f(x)d(F_n(x) - F^{y_n}(x)) + pF^{y_n}(f) = C_n + pF^{y_n}(f),$$

$$\sum_{i=1}^{p}\ell_i = p\int g(x)d(F_n(x) - F^{y_n}(x)) + pF^{y_n}(g) = B_n + pF^{y_n}(g).$$

Applying Theorem A.1 given in the Appendix to the pair (f,g), we have

$$\begin{pmatrix} C_n \\ B_n \end{pmatrix} \Longrightarrow N\left(\begin{pmatrix} EX_f \\ EX_g \end{pmatrix}, \begin{pmatrix} \mathrm{Cov}(X_f, X_f) & \mathrm{Cov}(X_f, X_g) \\ \mathrm{Cov}(X_g, X_f) & \mathrm{Cov}(X_g, X_g) \end{pmatrix}\right).$$

It remains to evaluate the limiting parameters and this results from the following calculations:

$$I_1(f,r) = \frac{h^2}{r^2}, \tag{6.11}$$

$$I_1(g,r) = 0, \tag{6.12}$$

$$I_2(f) = h^2, \tag{6.13}$$

$$I_2(g) = 0, \tag{6.14}$$

$$J_1(f,g,r) = \frac{2h^2 + 2h^4}{r^2}, \tag{6.15}$$

$$J_1(f,f,r) = \frac{2h^4 + (2h+2h^3)^2 r}{r^3}, \tag{6.16}$$

$$J_1(g,g,r) = \frac{h^2}{r^2}, \tag{6.17}$$

$$J_2(f,g) = 2h^2 + 2h^4, \tag{6.18}$$
$$J_2(f,f) = (2h + 2h^3)^2, \tag{6.19}$$
$$J_2(g,g) = h^2. \tag{6.20}$$

The results (6.12), (6.14), (6.17) and (6.20) are exactly the same as those found in (6.2), (6.4), (6.7) and (6.10) in the proof of Lemma 2.1. The remaining results are found by similar calculations using Proposition A.1 in the Appendix and their details are omitted.

7. Concluding remarks

Using recent central limit theorems for eigenvalues of large sample covariance matrices, we are able to find new asymptotic distributions for two major procedures to test the sphericity of a large-dimensional distribution. Although the theory is developed under the scheme $p \to \infty$, $n \to \infty$ and $p/n \to y > 0$, our Monte-Carlo study has proved that: on the one hand, both CLRT and CJ are already very efficient for middle dimension such as $(p,n) = (96, 128)$ both in size and power, see Table 2 and Table 3; and on the other hand, CJ also behaves very well in most of "large p, small n" situation, see Table 4.

Three characteristic features emerge from our findings:

(a) These asymptotic distributions are universal in the sense that they depend on the distribution of the observations only through its first four moments;
(b) The new test procedures improve quickly when either the dimension p or the sample size n gets large. In particular, for a given sample size n, within a wide range of values of p/n, higher dimensions p lead to better performance of these corrected test statistics.
(c) CJ is particularly robust against the dimension inflation. Our Monte-Carlo study shows that for a small sample size $n = 64$, the test is effective for $0 < p/n \le 20$.

In a sense, these new procedures have benefited from the "blessings of dimensionality".

Appendix A: Formula for limiting parameters in the CLT for eigenvalues of a sample covariance matrix with general fourth moments

Given a sample covariance matrix S_n of dimension p with eigenvalues $\lambda_1, \ldots, \lambda_p$, linear spectral statistics of the form $F_n(g) = p^{-1} \sum_{i=1}^{p} g(\lambda_j)$ for suitable functions g are of central importance in multivariate analysis. Such CLT's have been successively developed since the pioneering work of [12], see [2] and [16] for a recent account on the subject.

The CLT in [2] (see also an improved version in [5]) has been widely used in applications as this CLT also provides, for the first time, explicit formula for the mean and covariance parameters of the normal limiting distribution. In the

special case with an array $\{x_{ij}\}$ of independent variables, this CLT assumes the following moment conditions:

(a) For each n, $x_{ij} = x_{ij}^n, i \leqslant p, j \leqslant n$ are independent.
(b) $Ex_{ij} = 0, E|x_{ij}|^2 = 1, \max_{i,j,n} E|x_{ij}|^4 < \infty$.
(c) If $\{x_{ij}\}$'s are real, then $Ex_{ij}^4 = 3$; otherwise (complex variables), $Ex_{ij}^2 = 0$ and $E|x_{ij}|^4 = 2$.

In Condition (c), the fourth moments of the entries are set to the values 3 or 2 matching the normal case. This is indeed a quite demanding and restrictive condition since in the real case for example, it is incredibly hard to find a non-normal distribution with mean 0, variance 1 and fourth moment equaling 3. As a consequence, most of if not all applications published in the literature using this CLT assumes a normal distribution for the observations. Recently, effort have been made in [21, 16] and [29] to overcome these moment restrictions. We present below such a CLT with general forth moments that will be used for the sphericity test.

In all the following, we use an indicator κ set to 2 when $\{x_{ij}\}$ are real and to 1 when they are complex. Define $\beta = E|x_{ij}|^4 - 1 - \kappa$ for both cases and $h = \sqrt{y}$.

For the presentation of the results, let be the sample covariance matrix $S_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^*$ where $X_i = (x_{ki})_{1 \leq k \leq p}$ is the i-th observed vector. It is then well-known that when $p \to \infty$, $n \to \infty$ and $p/n \to y > 0$, the distribution of its eigenvalues converges to a nonrandom distribution, namely the Marčenko-Pastur distribution F^y with support $[a, b] = [(1 \pm \sqrt{y})^2]$ (an additional mass at the origin when $y > 1$). Moreover, the Stieltjes transform $\underline{m}$ of a companion distribution defined by $\underline{F}^y = (1 - y)\delta_0 + yF_c$ satisfies an inverse equation for $z \in \mathbb{C}^+$,

$$z = -\frac{1}{\underline{m}} + \frac{y}{1 + \underline{m}}. \tag{A.1}$$

The following CLT is a particular instance of Theorem 1.4 in [21].

Theorem A.1 ([21]). *Assume that for each n, the variables $x_{ij} = x_{ij}^n, i \leqslant p, j \leqslant n$ are independent and identically distributed satisfying $Ex_{ij} = 0$, $E|x_{ij}|^2 = 1$, $E|x_{ij}|^4 = \beta + 1 + \kappa < \infty$ and in case of they are complex, $Ex_{ij}^2 = 0$. Assume moreover,*

$$p \to \infty, \ n \to \infty, \ p/n \to y > 0.$$

Let $f_1, \dots f_k$ be functions analytic on an open region containing the support of F^y. The random vector $\{X_n(f_1), \dots X_n(f_k)\}$ where

$$X_n(f) = p\{F_n(f) - F^{y_n}(f)\}$$

converges weakly to a normal vector $(X_{f_1}, \dots X_{f_k})$ with mean function and co-variance function:

$$\mathbb{E}[X_f] = (\kappa - 1)I_1(f) + \beta I_2(f), \tag{A.2}$$

$$Cov(X_f, X_g) = \kappa J_1(f, g) + \beta J_2(f, g), \tag{A.3}$$

where

$$
\begin{aligned}
I_1(f) &= -\frac{1}{2\pi i}\oint \frac{y\left\{\underline{m}/(1+\underline{m})\right\}^3(z)f(z)}{\left[1-y\left\{\underline{m}/(1+\underline{m})\right\}^2\right]^2}dz,\\[2mm]
I_2(f) &= -\frac{1}{2\pi i}\oint \frac{y\left\{\underline{m}/(1+\underline{m})\right\}^3(z)f(z)}{1-y\left\{\underline{m}/(1+\underline{m})\right\}^2}dz,\\[2mm]
J_1(f,g) &= -\frac{1}{4\pi^2}\oint\oint \frac{f(z_1)g(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\underline{m}'(z_1)\underline{m}'(z_2)dz_1 dz_2,\\[2mm]
J_2(f,g) &= \frac{-y}{4\pi^2}\oint f(z_1)\frac{\partial}{\partial z_1}\left\{\frac{\underline{m}}{1+\underline{m}}(z_1)\right\}dz_1\cdot\oint g(z_2)\frac{\partial}{\partial z_2}\left\{\frac{\underline{m}}{1+\underline{m}}(z_2)\right\}dz_2,
\end{aligned}
$$

where the integrals are along contours (non overlapping in J_1) enclosing the support of F^y.

However, concrete applications of this CLT are not easy since the limiting parameters are given through those integrals on contours that are only vaguely defined. The purpose of this appendix is to go a step further by providing alternative formula for these limiting parameters. These new formulas, presented in the following Proposition convert all these integral along the unit circle; they are much easier to use for concrete applications, see for instance the proofs of Lemma 2.1 and 2.2 in the paper. Furthermore, these formulas will be of independent interest for applications other than those developed in this paper.

Proposition A.1. *The limiting parameters in Theorem A.1 can be expressed as following:*

$$
\begin{aligned}
I_1(f) &= \lim_{r\downarrow 1}I_1(f,r), && \text{(A.4)}\\[2mm]
I_2(f) &= \frac{1}{2\pi i}\oint_{|\xi|=1}f(|1+h\xi|^2)\frac{1}{\xi^3}d\xi, && \text{(A.5)}\\[2mm]
J_1(f,g) &= \lim_{r\downarrow 1}J_1(f,g,r), && \text{(A.6)}\\[2mm]
J_2(f,g) &= -\frac{1}{4\pi^2}\oint_{|\xi_1|=1}\frac{f(|1+h\xi_1|^2)}{\xi_1^2}d\xi_1\oint_{|\xi_2|=1}\frac{g(|1+h\xi_2|^2)}{\xi_2^2}d\xi_2, && \text{(A.7)}
\end{aligned}
$$

with

$$
\begin{aligned}
I_1(f,r) &= \frac{1}{2\pi i}\oint_{|\xi|=1}f(|1+h\xi|^2)\left[\frac{\xi}{\xi^2-r^{-2}}-\frac{1}{\xi}\right]d\xi,\\[2mm]
J_1(f,g,r) &= -\frac{1}{4\pi^2}\oint_{|\xi_1|=1}\oint_{|\xi_2|=1}\frac{f(|1+h\xi_1|^2)g(|1+h\xi_2|^2)}{(\xi_1-r\xi_2)^2}d\xi_1 d\xi_2.
\end{aligned}
$$

Proof. We start with the simplest formula $I_2(f)$ to explain the main argument and indeed, the other formula are obtained similarly. The idea is to introduce the change of the variable $z=1+hr\xi+hr^{-1}\overline{\xi}+h^2$ with $r>1$ but close to 1 and $|\xi|=1$ (recall $h=\sqrt{y}$). Note that this idea has been employed in [29].

It can be readily checked that when ξ runs counterclockwisely the unit circle, z will run a contour $\mathcal{C}$ that encloses closely the support interval $[a, b] = [(1 \pm h)^2]$ (recall $h = \sqrt{y}$). Moreover, by the Eq. (A.1), we have on $\mathcal{C}$

$$\underline{m} = -\frac{1}{1 + hr\xi}, \quad \text{and } dz = h(r - r^{-1}\xi^{-2})d\xi.$$

Applying this variable change to the formula of $I_2(f)$ given in Theorem A.1, we have

$$
\begin{aligned}
I_2(f) &= \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_{|\xi|=1} f(z) \frac{1}{\xi^3} \frac{r\xi^2 - r^{-1}}{r(r^2\xi^2 - 1)} d\xi \\
&= \frac{1}{2\pi i} \oint_{|\xi|=1} f(|1 + h\xi|^2) \frac{1}{\xi^3} d\xi.
\end{aligned}
$$

This proves the formula (A.5). For (A.4), we have similarly

$$
\begin{aligned}
I_1(f) &= \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_{|\xi|=1} f(z) \frac{1}{\xi^3} \frac{r\xi^2 - r^{-1}}{r(r^2\xi^2 - 1)} \frac{1}{1 - r^{-2}\xi^{-2}} d\xi \\
&= \lim_{r \downarrow 1} \frac{1}{2\pi i} \oint_{|\xi|=1} f(|1 + h\xi|^2) \frac{1}{\xi(\xi^2 - r^{-2})} \\
&= \lim_{r \downarrow 1} I_1(f, r).
\end{aligned}
$$

Formula (A.7) for $J_2(f, g)$ is calculated in a same fashion by observing that we have

$$\frac{\partial}{\partial z}\left\{\frac{\underline{m}}{1 + \underline{m}}(z)\right\} dz = \frac{\partial}{\partial \xi}\left\{\frac{\underline{m}}{1 + \underline{m}}(\xi)\right\} d\xi = \frac{\partial}{\partial \xi}\left\{\frac{1}{-hr\xi}\right\} d\xi = \frac{1}{hr\xi^2} d\xi.$$

Finally for (A.6), we use two non-overlapping contours defined by $z_j = 1 + hr_j\xi_j + hr_j^{-1}\bar{\xi}_j + h^2$, $j = 1, 2$ where $r_2 > r_1 > 1$. By observing that

$$\underline{m}'(z_j)dz_j = \left(\frac{\partial}{\partial \xi_j}\underline{m}\right) d\xi_j = \frac{hr_j}{(1 + hr_j\xi_j)^2} d\xi_j,$$

we find

$$
\begin{aligned}
J_1(f, g) &= \lim_{\substack{r_2 > r_1 > 1 \\ r_2 \downarrow 1}} -\frac{1}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{f(z_1)g(z_2)}{\{\underline{m}(z_1) - \underline{m}(z_2)\}^2} \\
&\qquad \cdot \frac{hr_1}{(1 + hr_1\xi_1)^2} \cdot \frac{hr_2}{(1 + hr_2\xi_2)^2} d\xi_1 d\xi_2 \\
&= \lim_{\substack{r_2 > r_1 > 1 \\ r_2 \downarrow 1}} -\frac{1}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{f(z_1)g(z_2)}{\{r_1\xi_1 - r_2\xi_2\}^2} d\xi_1 d\xi_2 \\
&= \lim_{r \downarrow 1} -\frac{1}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{f(|1 + h\xi_1|^2)g(|1 + h\xi_2|^2)}{\{\xi_1 - r\xi_2\}^2} d\xi_1 d\xi_2.
\end{aligned}
$$

The proof is complete.

Acknowledgements

References

[1] ANDERSON, T.W. (1984). *An introduction to Multivariate Statistical Analysis* (2nd edition). Wiley, New York. MR0771294

[2] BAI, Z.D. AND SILVERSTEIN, J.W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, **32**, 553–605. MR2040792

[3] BAI, Z.D. AND YAO, J.F. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, **11**(6), 1059–1092. MR2189081

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

[5] BAI, Z.D. AND SILVERSTEIN, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices* (2nd edition). Springer, 20. MR2567175

[6] BAI, Z.D., JIANG, D.D., YAO, J.F., AND ZHENG, S.R. (2012). Testing linear hypotheses in high-dimensional regressions. *Statistics*, doi:10.1080/02331888.2012.708031.

[7] BIRKE, M. AND DETTE, H. (2005). A note on testing the covariance matrix for large dimension. *Statistics and Probability Letters*, **74**, 281–289. MR2189467

[8] CHEN, S.X. AND QIN, Y.L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.*, **38**, 808–835. MR2604697

[9] CHEN, S.X., ZHANG, L.X., AND ZHONG, P.S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.*, **105**, 810–819. MR2724863

[10] JOHN, S. (1971). Some optimal multivariate tests. *Biometrika*, **58**, 123–127. MR0275568

[11] JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, **59**, 169–173. MR0312619

[12] JONSSON, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, **12**(1), 1–38. MR0650926

[13] JOHNSTONE, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, **29**(2), 295–327. MR1863961

[14] JOHNSTONE, I.M. AND TITTERINGTON, D.M. (2009). Statistical challenges of high-dimensional data. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, **367**, 4237–4253. MR2546386

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

[16] Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of the Wigner and the sample covariance random matrices. *Metrika*, **69**, 153–172. MR2481919

[17] Mauchly, J.W. (1940). Test for sphericity of a normal n-variate distribution. *Ann. Mathe. Statist.*, **11**, 204–209. MR0002084

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

[19] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, **1**, 700–709. MR0339405

[20] Onatski, A., Moreira, M.J., and Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.*, **41**(3), 1204–1231.

[21] Pan, G.M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.*, **18**, 1232-1270. MR2418244

[22] Pan, G.M. (2012). Comparision between two types of large sample covariance matrices. To appear in *Ann. Institut Henri Poincaré*.

[23] Schott, J.R. (2007). Some high-dimensional tests for a one-way MANOVA. *J. Multivariate Anal.*, **98**, 1825–1839. MR2392435

[24] Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, **35**, 251–272. MR2328427

[25] Srivastava, M.S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *J. Multiv. Anal.*, **97**, 1927–1940. MR2301265

[26] Srivastava, M.S., Kollo, T., and Rosen, D. (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *J. Multiv. Anal.*, **102**, 1090–1103. MR2793878

[27] Sugiura, N. (1972). Locally best invariant test for sphericity and the limiting distributions. *Ann. Mathe. Statist.*, **43**, 1312-1316. MR0311032

[28] Wang, Q.W., Silverstein, J.W., and Yao, J.F. (2013). A note on the CLT of the LSS for sample covariance matrix from a spiked population model. *Preprint*, available at arXiv:1304.6164.

[29] Zheng, S.R. (2012). Central limit theorem for linear spectral statistics of large dimensional F matrix. *Ann. Institut Henri Poincaré Probab. Statist.*, **48**, 444–476. MR2954263

[30] Zheng, S.R. and Bai, Z.D. (2013). A note on central limit theorems for linear spectral statistics of large dimensional F Matrix. *Preprint*, available at arXiv:1305.0339.