Regularized LRT for Large Scale Covariance Matrices:
One Sample Problem

Young-Geun Choi, Chi Tim Ng and Johan Lim

Abstract
The main theme of this paper is a modification of the likelihood ratio test (LRT) for testing high dimensional covariance matrix. In recent, a correction of the asymptotic of the LRT for a large-dimensional case (the case $p/n$ approaches to a constant $\gamma \in (0, 1]$) is proposed by Bai et al. (2009) (for $\gamma \in (0, 1)$) and Jang et al. (2012) (for $\gamma = 1$). The corrected procedure is named as corrected LRT. Despite its correction, however, the corrected LRT is a function of sample eigenvalues which are suffered from redundant variability from high dimensionality and, subsequently, still does not have full power in differentiating hypotheses on the covariance matrix. In this paper, we are motivated by the successes of the shrinkage covariance matrix estimator (simply shrinkage estimator) in various applications, we propose to use the regularized LRT which uses, in defining the LRT, the shrinkage estimator instead of the sample covariance matrix. We compute the asymptotic distribution of the regularized LRT when the true covariance matrix is the identity matrix and a spiked covariance matrix. The obtained asymptotic results have various applications in testing hypotheses on the covariance matrix. Here, we apply them to testing the identity of the true covariance matrix which is a long standing problem in the literature, and show that the regularized LRT outperforms the corrected LRT, which is its non-regularized counterpart. In addition, we compare the power of the regularized LRT to those of recent non-likelihood based procedures.

Keywords: Asymptotic normality; covariance matrix estimator; identity covariance matrix; high dimensional data; linear shrinkage estimator; linear spectral statistics; random matrix theory; regularized likelihood ratio test; spiked covariance matrix.

1 Introduction

In recent, high dimensional data are prevalent everywhere which include genomic data in biology, financial times series data in economics, and natural language processing data in machine learning and marketing. In analyzing them, the traditional procedures, which assume
that sample size \( n \) is large and dimension \( p \) is fixed, are not valid anymore. A significant amount of research are made to resolve the difficulty from the dimensionality of the data.

This paper is on the inference of large scale covariance matrix whose dimension \( p \) is large compared to the sample size \( n \). To be specific, we are interested in testing whether the covariance matrix equals to a given known matrix; \( \mathcal{H}_0 : \Sigma = \Sigma_0 \), where \( \Sigma_0 \) is a given matrix and is assumed to be \( I_p \) without loss of generality. The likelihood ratio test (LRT) statistic to test \( \mathcal{H}_0 : \Sigma = I_p \) is defined by

\[
LRT = \text{tr}(S) - \log |S| - p = \sum_{i=1}^{p} (l_i - \log l_i - 1),
\]

where \( S \) is the sample covariance matrix and \( l_i \) is the \( i \)-th largest eigenvalue of the sample covariance matrix. Its asymptotic null distribution for finite \( p \), which is the chi-square distribution with degrees of freedom \( p(p+1)/2 \), is not accurate any more, if \( p \) increases. Its correct asymptotic distribution is computed by Bai et al. (2009) for the case \( p/n \) approaches \( \gamma \in (0, 1) \) and both \( n \) and \( p \) increase. They further numerically show that their asymptotic normal distribution defines a valid procedure for testing \( \mathcal{H}_0 : \Sigma = I_p \). The results of Bai et al. (2009) are refined by Jiang et al. (2012) which include the asymptotic null distribution for the case \( \gamma = 1 \). Though the correction of the null distribution, the sample covariance is known to have redundant variability when \( p \) is large, and it still remains a general question that the LRT is asymptotically optimal for testing problem in the \( n, p \) large scheme.

In this paper, we pursue to improve the ‘corrected’ LRT using the regularized covariance matrix estimator. In detail, we consider a modification of the LRT, denoted by regularized LRT (rLRT), that is defined by

\[
rLRT = \text{tr}(\hat{\Sigma}) - \log |\hat{\Sigma}| - p = \sum_{i=1}^{p} (\psi_i - \log \psi_i - 1),
\]  

where \( \hat{\Sigma} \) is a regularized covariance matrix and \( \psi_i \) is the \( i \)-th largest eigenvalue of \( \hat{\Sigma} \). Here, we consider the regularization via linear shrinkage:

\[
\hat{\Sigma} \equiv \lambda S + (1 - \lambda)I_p.
\]

We also occasionally notate rLRT(\( \lambda \)) to emphasize the use of the value \( \lambda \). The linearly shrunken sample covariance matrix (simply shrinkage estimator) is known to reduce expected estimation loss of the sample covariance matrix (Ledoit and Wolf, 2004). It is also
successfully applied to many high-dimensional procedures to resolve the dimensionality problem. For example, Schäfer and Strimmer (2005) reconstruct a gene regulatory network from microarray gene expression data using the inverse of a regularized covariance matrix. Chen et al. (2011) propose a modified Hotelling’s $T^2$-statistic for testing high dimensional mean vectors and apply it to finding differentially expressed gene sets. We are motivated by the success of above examples and inspect whether the power can be improved by the reduced variability via linear shrinkage. To the best of our knowledge, our work is the first time to apply the linear shrinkage to the covariance matrix testing problem itself.

We derive the asymptotic distribution of the proposed rLRT($\lambda$) under two scenarios, (i) when $\Sigma = I_p$ for the null distribution, and additionally (ii) when $\Sigma = \Sigma_{\text{spike}}$ for power study. Here $\Sigma_{\text{spike}}$ means a covariance matrix from the spiked population model (Johnstone 2001), roughly it is defined by a covariance matrix whose eigenvalues are all 1’s but some finite nonunit ‘spike’. The spiked covariance matrix assumed here includes the well known compound symmetry matrix $\Sigma_{\text{cs}}(\rho) = I_p + \rho J_p$, where $J_p$ is $p \times p$ matrix of ones. The main results show that rLRT($\lambda$) has normal distribution in asymptotic under both (i) and (ii); their asymptotic means are different but the variances are same. The main results are useful in testing various one sample covariance matrix. To be specific, first, in testing $H_0 : \Sigma = I_p$, (i) provides the asymptotic null distribution of rLRT($\lambda$). Second, combining (i) and (ii) provides the asymptotic power for an arbitrary spiked alternative covariance matrix including $\Sigma_{\text{cs}}(\rho)$. Finally, the results with $\lambda = 1$ provide various asymptotic distributions of the LRT. Among these many applications, in this paper, we particularly focus on the LRT for testing $H_0 : \Sigma = I_p$ which has long been studies by many researchers (Anderson 2003; Bai et al. 2009; Chen et al., 2010; Jiang et al., 2012; Ledoit and Wolf, 2002).

The paper is organized as follows. In Section 2, we briefly review results of the random matrix theory which are the major basis of this paper. The results include the limit of empirical spectral distribution (ESD) of the sample covariance matrix and the central limit theorem (CLT) for linear spectral statistics (LSS). In Section 3, we formally define the rLRT, and prove the asymptotic normality of the rLRT when the true covariance matrix $\Sigma$ is $I_p$ or $\Sigma_{\text{spike}}$. In Section 4, we formally define the rLRT and hire the results in Section 2 to prove the asymptotic normality of the rLRT when the true covariance matrix $\Sigma$ is $I_p$ or $\Sigma_{\text{spike}}$. In Section 4, we apply our results in Section 3 to testing $H_0 : \Sigma = I_p$. In this section, we
numerically compare the powers of the LRT and other existing methods, particularly to the original LRT (say LRT) and the corrected LRT (cLRT) by Bai et al. (2009). In Section 5, we conclude the paper with discussions of several technical details of the rLRT which include close spiked eigenvalues and non-zero mean vector.

2 Preliminaries: Some results on random matrix theory

This section reviews some useful properties of linear spectral statistics of the sample covariance matrix when the true covariance matrix $\Sigma$ is identity or that from a spiked population model.

We start the section with introducing one notation to be used in the remainder of the paper. Let $F^M$ be the spectral distribution (SD) for a real symmetric or a Hermitian matrix $M$ of size $p \times p$, i.e.,

$$F^M(t) := \frac{1}{p} \sum_{j=1}^{p} I(\alpha_j(M) \leq t), \quad t \in \mathbb{R},$$

where $I(A)$ denotes the indicator function of a set $A$ and $\alpha_j(M)$ the $j$-th largest eigenvalue of the matrix $M$ with natural labeling $\alpha_p(M) \leq \cdots \leq \alpha_1(M)$.

2.1 Limiting spectral distribution of sample covariance matrix

Let $\{z_{ij}\}_{i,j \geq 1}$ be an infinite double array of independent and identically distributed (IID) real random variables with $Ez_{11} = 0$, $Ez_{11}^2 = 1$ and $Ez_{11}^4 = 3$. Let $Z_n = \{z_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, p\}$, i.e., $Z_n$ denotes the upper-left $n \times p$ block of the infinite double array. We assume that $p = p_n$ depends on $n$ and $\gamma_n := p/n$ converges to a positive constant $\gamma$. The data matrix $X_n$ and the sample covariance matrix are described by $X_n = Z_n \Sigma_n^{1/2}$ and $S_n = \frac{1}{n}X_n^\top X_n$, where $\{\Sigma_n, n = 1, 2, \ldots\}$ is a sequence of $p_n \times p_n$ nonrandom Hermitian matrices. We remark that the fourth moment condition $Ez_{11}^4 = 3$ is necessary for the forthcoming Proposition 1.

In the standard random matrix theory, the limiting distribution of empirical SD $F^{S_n}(\cdot)$ is known to be determined by both the limits of $p/n$ and $F^{\Sigma_n}(\cdot)$. More rigorously, if $H_n(\cdot) := F^{\Sigma_n}(\cdot)$ converges weakly to $H(\cdot)$, then, with probability one, $F^{S_n}(\cdot)$ converges weakly to a nonrandom distribution function, say $F^{\gamma \cdot H}(\cdot)$, whose Stieltjes transform $m^{\gamma \cdot H}(z)$ is given as
the unique solution of the following system of equations

\[ m(z) = \frac{1}{\gamma}m(z) + \frac{1 - \gamma}{\gamma z}; \]

\[ z = -\frac{1}{m^{\gamma,H}(z)} + \gamma \int \frac{t}{1 + tm^{\gamma,H}(z)}dH(t), \quad z : \text{Im}(z) > 0. \quad (3) \]

Generally, \( m^{\gamma,H}(z) \) is known as the Stieltjes transform of the limiting SD of \( S_n := \frac{1}{n}X_nX_n^T \), which is called as the companion matrix for \( S_n \). The limiting SD \( F^{\gamma,H} \) is identified from the Stieltjes transformation \( m^{\gamma,H} \) by the inversion formula:

\[ \frac{dF^{\gamma,H}}{dx}(x) = \lim_{z \to x} \frac{1}{\gamma \pi} \text{Im}[m^{\gamma,H}(z)], \quad x \in \mathbb{R}, z : \text{Im}(z) > 0. \quad (4) \]

In the special case \( \Sigma_n = I_p \), \( H_n(t) \equiv \delta_1(t) \) for \( \delta_1(t) = I(t \geq 1) \), and we can obtain the corresponding spectral distribution \( F^{\gamma,\delta_1} \) as the Marcenko-Pastur law. To see this, note that for \( H = \delta_1 \) and the implicit function (3) can be rewritten as

\[ z = -\frac{1}{m^{\gamma,\delta_1}(z)} + \frac{\gamma}{1 + m^{\gamma,\delta_1}(z)}, \quad z : \text{Im}(z) > 0. \quad (5) \]

In sequel, (4) shows the probability density function of the Marčenko-Pastur law indexed by \( \gamma \) when \( 0 < \gamma < 1 \),

\[ \frac{dF^{\gamma,\delta_1}}{dx}(x) = \frac{1}{2\pi \gamma x} \sqrt{(b(\gamma) - x)(x - a(\gamma))}, \quad a(\gamma) \leq x \leq b(\gamma), \quad (6) \]

where \( a(\gamma) := (1 - \sqrt{\gamma})^2 \) and \( b(\gamma) := (1 + \sqrt{\gamma})^2 \).

### 2.2 CLT for linear spectral statistics

Many multivariate statistical procedures is based on \( F^S \), the empirical SD of the sample covariance matrix. Specifically, of interest here is a family of functional of eigenvalues which is also called as linear spectral statistics (LSS) or linear eigenvalue statistics:

\[ \hat{\theta} = \frac{1}{p} \sum_{j=1}^{p} g(l_j) = \int g(x) dF^S(x) \]

for a function \( g \) with some complex-analytic conditions.

To state the CLT for the LSS, some terminologies in the previous section are still in use. The limiting SD \( H \) and the corresponding Stieltjes transformation \( m^{\gamma,H} \) are needed for the
computation of the mean and the variance of limiting normal distribution. The centering
term of the CLT possesses a finite-dimensional proxy $F^\gamma_{n,H_n}(\cdot)$ (Wang et al. 2014). The
proposition below is a modification of Theorem 2.1 in Bai et al. (2009), which is also a version
of Theorem 1.1 in Bai and Silverstein (2004) for the case of $H = \delta_1$ and $0 < \gamma < 1$.

**Proposition 1.** Let $T_n(g)$ be the functional

$$T_n(g) = p \int g(x) d\left\{ F_{S_n}(x) - F_{\gamma_n,H_n}(x) \right\}.$$

Suppose two functions $g_1$ and $g_2$, are complex analytic on an open domain containing an
closed interval $[a(\gamma), b(\gamma)]$ on the real axis. If $\gamma \in (0, 1)$ and $H_n$ converges weakly to $\delta_1$, then
the vector $(T_n(g_1), T_n(g_2))$ converges in distribution to a bivariate normal distribution with
mean

$$\mu(g_i) = g_i(a(\gamma)) + g_i(b(\gamma)) - \frac{1}{2\pi} \int_{a(\gamma)}^{b(\gamma)} \frac{g_i(x)}{\sqrt{4\gamma - (x - 1 - \gamma)^2}} \, dx$$

for $i = 1, 2$, and variance

$$v(g_1, g_2) = -\frac{1}{2\pi^2} \int \int \frac{g_1(z_1)g_2(z_2)}{(m(z_1) - m(z_2))^2} \, dm(z_1) \, dm(z_2),$$

where $m = m^{\gamma, \delta_1}$ is defined in (5). The contours in (8) are non-overlapping and both contain
$[a(\gamma), b(\gamma)]$, the support of $F_{\gamma, \delta_1}$.

The assumption $H = \delta_1$ roughly implies the spectrum of $\Sigma_n$ is eventually concentrated
at one or its neighborhood. One simple example is the covariance matrix of IID random
variables $\Sigma_n = I_p$, whose SD is $H_n = F_{\delta_1} = \delta_1$ and trivially converges to $\delta_1$ in distribution. In
addition, recalling tha $H = \delta_1$ is the limiting SD, one notes that the spiked population model
(Johnstone 2001) is applicable for Proposition 1. The spiked population model assumes the
$p_n \times p_n$ covariance matrix $\Sigma_n$ has the structure:

$$\Sigma_{\text{spike},n} = P_n D_n P_n^T,$$

$$D_n := \text{diag}(a_{n_1}, \ldots, a_{n_1}, a_{n_2}, \ldots, a_{n_2}, \ldots, a_{n_k}, \ldots, a_{n_k}, 1, \ldots, 1)$$

where $P_n$ denotes orthogonal matrices. Then the SD $H_n$ corresponded to $\Sigma_n$ is

$$H_n(t) = \frac{p_n - K}{p_n} \delta_1(t) + \frac{1}{p_n} \sum_{i=1}^{K} n_i I(t \geq a_i),$$

(10)
where $K = n_1 + \ldots + n_k$ is a fixed finite integer and independent of $n$ so that $p_n - K$ eigenvalues of $1$’s eventually dominate corresponding $H_n$. Thus, the limiting SD is still left unchanged as $H = \delta_1$.

Despite its simplicity of the limiting SD, the spiked population model has several difficulties along with Proposition 1. In the spiked model, $H_n$ in (10) has masses at $K + 1$ distinct points and $m_{\gamma_n^H}(z)$ is the solution to a polynomial equation of degree $K + 2$. A polynomial equation has an analytic solution only when its degree is less than or equal to 4. Therefore, if $K \geq 3$, we do not have analytic form of $m_{\gamma_n^H}(z)$. To resolve this difficulty, recently, [Wang et al. (2014)] provides an approximation formula to $\int g dF_{\gamma_n^H}$ for the spiked population model. They use the asymptotic expansion of $g(m_{\gamma_n^H})$ at $m_{\gamma_n^H, \delta_1}$ under the conception that $m_{\gamma_n^H, \delta_1}$ and $m_{\gamma_n^H}(z)$ would be close enough if $n$ is large.

Proposition 2. Suppose that $\gamma_n < 1$ and that $H_n$ is in the form of (10), with $|a_i - 1| > \sqrt{\gamma}$ for all $i = 1, 2, \ldots, K$. If a complex function $g$ is analytic on an open domain containing the interval $[a(\gamma), b(\gamma)]$ and $k$ points $\varphi(a_i) := a_i + \frac{a_i}{a_i - 1}$, $i = 1, 2, \ldots, K$ on the real axis, Then

$$\int g(x) dF_{\gamma_n^H}(x) = -\frac{1}{2\pi i} \oint_C g\left(-\frac{1}{m} + \frac{\gamma_n}{1 + m}\right) \left(\frac{K}{\gamma_n m} - \sum_{i=1}^{K} \frac{n_i a_i^2 m}{(1 + a_i m)^2}\right) \frac{dm}{m}$$

(11)

$$+ \frac{1}{2\pi i} \oint_C f'\left(-\frac{1}{m} + \frac{\gamma_n}{1 + m}\right) \sum_{i=1}^{K} \frac{(1 - a_i) n_i}{(1 + a_i m)(1 + m)} \left(\frac{1}{m} - \frac{\gamma_n m}{(1 + m)^2}\right) \frac{dm}{m}$$

(12)

$$+ \left(1 - \frac{K}{p}\right) \int g(x) dF_{\gamma_n^H, \delta_1}(x) + \frac{1}{p} \sum_{i=1}^{K} n_i g(\varphi(a_i)) + O\left(\frac{1}{n^2}\right)$$

(13)

where $m = m_{\gamma_n, \delta_1}$ is defined in [3] by substituting $\gamma$ by $\gamma_n$, and $C$ is a counterclockwise contour enclosing the interval $\left[-\frac{1}{1 - \sqrt{\gamma}}, \frac{1}{1 - \sqrt{\gamma}}\right]$ on the real axis.

The above proposition is from Theorem 2 of [Wang et al. (2014)] and the condition $|a_i - 1| > \sqrt{\gamma}$ defines that $a_i$ is a distant spike. We remark that the original version of Theorem 2 allows the existence of some close spikes $a_i$ which is defined by $|a_i - 1| \leq \sqrt{\gamma}$. For this case, following [Wang et al. (2014)], we also have an extended results but are not able to make both the theoretical result and empirical simulation to coincide. The detailed discussion of this will be followed in Section 5.
3 Main results

In this section, the asymptotic results of the rLRT are presented. Here, the rLRT is defined via the linear shrinkage estimator instead of the sample covariance matrix:

$$rLRT(\lambda) := \text{tr}(\hat{\Sigma}) - \log |\hat{\Sigma}| - p,$$

where $\hat{\Sigma} := \lambda S_n + (1 - \lambda)I_p$.

The tuning parameter $\lambda$ is assumed to be a constant in $(0, 1)$. We define $\psi(x) = \lambda x + (1 - \lambda)$ and $g(x) = \psi(x) - \log\{\psi(x)\} - 1$. Then we can write $rLRT(\lambda) = p \int g(x) dF_{\Sigma_n}(x)$ and use Proposition 1 to have the following results.

**Theorem 1.** Let $g(x) = \psi(x) - \log\{\psi(x)\} - 1$ and $\psi(x) = \lambda x + (1 - \lambda)$ with fixed $\lambda \in (0, 1)$. Suppose that $\Sigma_n = I_p$. If $\gamma_n := p/n \to \gamma \in (0, 1)$, then

$$T_n(g) = rLRT(\lambda) - p \int g(x) dF_{\gamma_n, \delta_1}(x)$$

converges in distribution to the normal distribution with mean

$$\mu(g) = \frac{-\log \sqrt{(1 + \lambda \gamma)^2 - 4\lambda^2 \gamma}}{2} + \frac{1}{4\pi} \int_0^{2\pi} (\log(1 + \lambda \gamma - 2\lambda \sqrt{\gamma} \cos \theta)) d\theta$$

and variance

$$v(g) = 2 \left\{ -\frac{\lambda}{M} - \lambda(1 + \gamma - \lambda \gamma) + \frac{\lambda \gamma}{1 + N} - \log \frac{M - N}{M(1 + N)} \right\},$$

where

$$M, N = M(\lambda, \gamma), N(\lambda, \gamma) := -\frac{(1 - 2\lambda + \lambda \gamma) \pm \sqrt{(1 - 2\lambda + \lambda \gamma)^2 + 4\lambda(1 - \lambda)}}{2(1 - \lambda)}$$

The detailed proof of Theorem 1 is followed in Appendix A.

The result in Theorem 1 is consistent with the $\lambda = 1$ case presented in Bai et al. (2009), where $\mu(g) = -\log(1 - \gamma)/2$ and $v(g) = -2\gamma - 2\log(1 - \gamma)$. To see this, we note that the integral in the mean function $\int_0^{2\pi} \log(1 + \lambda \gamma - 2\lambda \sqrt{\gamma} \cos \theta) d\theta$ approaches zero by the dominate convergence theorem. In addition, we could show that $M$ goes to $-1/(1 - \gamma)$, and $N$ goes to $+\infty$ as $\lambda \to 1$ using the approximation formula $\sqrt{x + \Delta x} \approx \sqrt{x} + \frac{1}{2} x^{-1/2} \Delta x$.

We now move to the computation of the term

$$\int g(x) dF_{\gamma_n, \delta_1}(x).$$
From the density function of Marčenko-Pastur law [6] and its property $1 = \int x dF_{\gamma_n, \delta_1}(x) = \int 1 dF_{\gamma_n, \delta_1}(x)$, we get

\[
\int g(x) dF_{\gamma_n, \delta_1}(x) = \int_{a(\gamma_n)}^{b(\gamma_n)} \frac{\lambda x - \lambda - \log(\lambda x + 1 - \lambda)}{2\pi x\gamma_n} \sqrt{b(\gamma_n) - x} \{x - a(\gamma_n)\} dx
\]

\[
= - \int_{a(\gamma_n)}^{b(\gamma_n)} \frac{\log(\lambda x + 1 - \lambda)}{\frac{2\pi x\gamma_n}{\sqrt{b(\gamma_n) - x} \{x - a(\gamma_n)\}}} dx.
\]

By substituting $x = 1 + \gamma_n - 2\sqrt{\gamma_n} \cos \theta$, we have an alternative representation of the integral

\[
\int g(x) dF_{\gamma_n, \delta_1}(x) = - \frac{2}{\pi} \int_0^\pi \frac{\log(1 + \lambda \gamma_n - 2\lambda \sqrt{\gamma_n} \cos \theta)}{1 + \gamma_n - 2\sqrt{\gamma_n} \cos \theta} \cdot \sin^2 \theta d\theta.
\]

(17)

It is remarked that (17) is easily implemented via the standard numerical integration techniques.

Our next theorem shows that asymptotic normality of the rLRT when the true covariance matrix is from the spiked population model in Section 2. A direct application of Proposition 2 results in the following theorem.

**Theorem 2.** Let $g(x) = \psi(x) - \log\{\psi(x)\} - 1$ and $\psi(x) = \lambda x + (1 - \lambda)$ with fixed $\lambda \in (0, 1)$. Suppose that $\Sigma_n$ has $SD H_n(t) = \frac{\nu_n - K}{p_n} \delta_1(t) + \frac{1}{p_n} \sum_{i=1}^K n_i I(t \geq a_i)$ as in [10] with $|a_i - 1|$ for all $i = 1, 2, \ldots, K$. If $\gamma_n := \frac{p}{n} \to \gamma \in (0, 1)$, then

\[
\text{rLRT}(\lambda) - p \int g(x) dF_{\gamma_n, H_n}(x) \longrightarrow N(\mu(g), \psi(g)),
\]

where $\mu(g), \psi(g)$ are defined in Theorem 1 and

\[
p \int g(x) dF_{\gamma_n, H_n}(x) = \left(1 - \frac{K}{p}\right) \int g(x) dF_{\gamma_n, \delta_1}(x) + \frac{K}{p} C_n + O\left(\frac{1}{n^2}\right)
\]

with

\[
C_n = \lambda \cdot \frac{1}{K} \sum_{i=1}^K n_i a_i - \lambda - \frac{1}{K} \sum_{i=1}^K n_i \log\psi\{\varphi(a_i)\}
\]

\[
- \left[\frac{1}{\gamma_n} \log(-M) + \frac{1}{K} \sum_{i=1}^K n_i \log \left(\frac{1 - a_i}{1 + a_i M}\right) - \frac{1}{1 + a_i M}\right]
\]

\[
+ \frac{\lambda}{(1 - \lambda) K} \sum_{i=1}^K n_i \left[\frac{a_i(M + 1)}{M - N} \left(\frac{1}{1 + a_i M} - \frac{a_i \gamma_n M^2}{(1 + a_i M)(M + 1)}\right) - \frac{1}{(M + 1)(N + 1)} \left(\frac{a_i \gamma_n}{1 - a_i} + \frac{\gamma_n(2M N + M + N)}{(M + 1)(N + 1)}\right)\right],
\]
where \( \varphi(a) = a + \frac{\gamma_n}{a-1} \), \( M = M(\lambda, \gamma_n) \) and \( N = N(\lambda, \gamma_n) \) in (16).

The proof of the above theorem is also followed in Appendix A.

As an application of Theorem 2, we consider the case when the true covariance matrix is a compound symmetry, \( \Sigma_{cs,n}(\beta \frac{p_n}{\gamma_n}) \):

\[
\Sigma_{cs,n}(\beta \frac{p_n}{\gamma_n}) = \begin{pmatrix}
1 + \beta \frac{p_n}{\gamma_n} & \beta \frac{p_n}{\gamma_n} & \beta \frac{p_n}{\gamma_n} & \cdots & \beta \frac{p_n}{\gamma_n} \\
\beta \frac{p_n}{\gamma_n} & 1 + \beta \frac{p_n}{\gamma_n} & \ddots & \ddots & \\
\vdots & \ddots & \ddots & \ddots & \\
\beta \frac{p_n}{\gamma_n} & \cdots & \cdots & \cdots & 1 + \beta \frac{p_n}{\gamma_n}
\end{pmatrix} = \mathbf{I}_{p_n} + \frac{\beta}{p_n} \mathbf{1} \mathbf{1}^T, \tag{18}
\]

This has a spiked eigenvalue structure; \( 1 + \beta \) for one eigenvalue and \( 1 \) for the other \( p_n - 1 \) eigenvalues and corresponding SD is \( H_n(t) = \frac{p_n-1}{p_n} I(t \geq 1) + \frac{1}{p_n} I(t \geq 1 + \beta) \). Theorem 2 with \( K = 1, k = 1, n_1 = 1 \), and \( a_1 = 1 + \beta \), gives the following corollary.

**Corollary 1.** Let \( g(x) = \psi(x) - \log \{ \psi(x) \} - 1 \) and \( \psi(x) = \lambda x + (1 - \lambda) \) with fixed \( \lambda \in (0, 1) \). Suppose that \( \Sigma_n \) has SD \( H_n(t) = \frac{p_n-1}{p_n} I(t \geq 1) + \frac{1}{p_n} I(t \geq 1 + \beta) \) with \( \beta > \sqrt{\gamma} \). If \( \gamma_n := p/n \to \gamma \in (0, 1) \), then

\[
\text{rLRT}(\lambda) - p \int g(x) dF^{\gamma_n,H_n}(x) \longrightarrow N(\mu(g), v(g)),
\]

where \( \mu(g), v(g) \) are defined in Theorem 1 and

\[
\int g(x) dF^{\gamma_n,H_n}(x) = \left( 1 - \frac{1}{p} \right) \int g(x) dF^{\gamma_n,\delta_1}(x) + \frac{1}{p} C_n + O \left( \frac{1}{n^2} \right)
\]

with

\[
C_n = \lambda \beta - \log \psi \{ \varphi(1+\beta) \} + \frac{1}{\gamma_n} \left\{ \frac{1}{\lambda} + \log \left( -\frac{1}{M} \right) \right\} + \frac{1}{1 + (1 + \beta)M} - \log \left( -\frac{\beta}{1 + (1 + \beta)M} \right) + \frac{\lambda}{(1 - \lambda)(M - N)} \left\{ \frac{(1 + \beta)(M + 1)}{1 + (1 + \beta)M} - \frac{\gamma_n(1 + \beta)M^2}{(1 + (1 + \beta)M)(M + 1)} - 1 + \frac{\gamma_n M^2}{(M + 1)^2} \right\}.
\]

In the corollary, the condition \( \beta > \sqrt{\gamma} \) means that the spiked eigenvalue \( 1 + \beta \) is distant.
4 Testing the identity covariance matrix

As stated in Introduction, the results in Section 3 can be used for testing various hypotheses on one sample covariance matrix. In this section, we study the properties of the rLRT in testing $H_0 : \Sigma = I_p$ and compare its power to existing procedures including the cLRT by Bai et al. (2009). We briefly introduce existing procedures in below.

- **Ledoit and Wolf (2002)** assume $p/n \to \gamma \in (0, \infty)$ and propose a statistic that is
  \[
  T_{\text{LW}} = \frac{1}{p} \text{tr} \left\{ (S - I)^2 \right\} - \frac{p}{n} \left\{ \frac{1}{p} \text{tr}(S) \right\}^2 + \frac{p}{n}.
  \]
  The asymptotic distribution of $T_{\text{LW}}$ is, if both $n$ and $p$ increase with $p/n \to c \in (0, \infty)$,
  \[
  nT_{\text{LW}} - p
  \]
  converges in distribution to normal distribution with mean 1 and variance 4.

- **Bai et al. (2009)** proposes a corrected LRT for the case both $n$ and $p$ increases and $p/n$ converges to $\gamma \in (0, 1)$. The LRT statistic is
  \[
  T_{\text{lrt}} = \text{tr}(S) - \log |S| - p.
  \]
  As introduced in Section 3, they show that
  \[
  \tilde{T}_{\text{lrt}} = v(g)^{-1/2} \left\{ T_{\text{lrt}} - p \int g(x) dF^{\gamma_n,\delta_1}(x) dx - \mu(g) \right\}
  \]
  converges in distribution to the standard normal distribution, where $\mu(g) = -\frac{\log(1 - \gamma)}{2}$, $v(g) = -2 \log(1 - \gamma) - 2\gamma$, and $\int g(x) dF^{\gamma_n,\delta_1}(x) dx = 1 - \frac{\gamma}{\gamma} \log(1 - \gamma)$.

- **Finally, Chen et al. (2010)** proposes to use the statistic
  \[
  T_{\text{C}} = \frac{1}{p} V_{2,n} - \frac{2}{p} V_{1,n} + 1,
  \]
where

\[ V_{1,n} = \frac{1}{n} \sum_{i=1}^{n} X_i^\top X_i - \frac{1}{P^2_n} \sum_{i \neq j} X_i^\top X_j \]

\[ V_{2,n} = \frac{1}{P^2_n} \sum_{i \neq j} (X_i^\top X_j)^2 - 2 \frac{1}{P^3_n} \sum_{i,j,k}^* X_i^\top X_j X_j^\top X_k \]

\[ + \frac{1}{P^4_n} \sum_{i,j,k,l} X_i^\top X_j X_j^\top X_l \]

\[ P^r_n = n!/(n-r)! \]

and \( \sum^* \) is the summation over different indices. The asymptotic theory tells that, under the null, \( nT_C \) converges in distribution to the normal distribution with mean 0 and variance 4.

We remark that the sample covariance matrix is differently defined for different methods. The cLRT and the rLRT are based on \( S = \frac{1}{n} \sum_i X_i X_i^\top \), while \( S_{cen} = \frac{1}{n-1} \sum_i (X_i - \bar{X})(X_i - \bar{X})^\top \) is used for Ledoit and Wolf (2002) and Chen et al. (2010). Thus, in our simulation study in the following section, the rLRT and the cLRT are based on the sample covariance without centering term, \( S = \frac{1}{n} \sum_i X_i X_i^\top \), and those of Ledoit and Wolf (2002) and Chen et al. (2010) are based on the sample covariance with centering term, \( S_{cen} = \frac{1}{n-1} \sum_i (X_i - \bar{X})(X_i - \bar{X})^\top \). The use of \( S_{cen} = \frac{1}{n-1} \sum_i (X_i - \bar{X})(X_i - \bar{X})^\top \) in the cLRT and the rLRT will have different asymptotic results whose details are discussed in Section 5.

4.1 Power comparison with cLRT

By applying Corollary 1, the asymptotic power curves of the cLRT and the rLRT are available for the alternative hypothesis of the compound symmetry. Given \( \Sigma_n \equiv \Sigma_{cs,n}(\beta/p) \) and setting \( \gamma_n \to \gamma \), the rejection probability in testing \( H_0 : \Sigma_n = I_p \) at level \( \eta \) is

\[ 1 - \Phi \left[ \Phi^{-1}(\eta) + \frac{1}{v(g)} \left( \int g(x) dF_{\gamma,\delta_1}(x) - C(\gamma, \lambda) \right) \right], \quad \beta > \sqrt{\gamma} \quad (19) \]

where a constant \( C(\gamma, \lambda) \) is defined in Corollary 1 and \( \Phi(\cdot) \) denotes the cumulative distribution function of the standard normal distribution.

The powers of the cLRT and the rLRT with \( \lambda = 0.4 \) and \( \lambda = 0.7 \) are plotted in Figure 1. Each panel of Figure 1 compares the powers of the cLRT and the rLRT for different
sample size $n$. In each panel, we include the cases of $\beta < \sqrt[3]{\gamma}$ (close spike) which violates the assumption of distant spike in Corollary 1. However, the results in Figure 1 implicitly indicate that Theorem 2 would be applicable even for the existence of “close spike” eigenvalue. More detailed discussions are followed in Section 5 at the end of the paper.

We find that in all cases the rLRT has much higher power than the cLRT regardless of the choice of $\lambda$. We also find that the empirical power curve increases to 1 less rapidly if $\lambda$ or $\gamma$ is closer to 1. In addition, although we do not report the details, the the empirical curves converge fast and do have minor changes after $n = 80$ regardless of the values of $\lambda$ and $\gamma$.

![Graphs showing analytical and empirical power curves for rLRT and cLRT.]

Figure 1: Analytic and empirical power curves for rLRT and cLRT.
To understand better the power gain by the rLRT, we plot the empirical probability distributions of the rLRT and cLRT for the null and the alternative hypotheses. They are plotted in Figure 2. The regularized estimators or the procedures based on them are typically believed to have smaller variability and performance gain from it. This is also true in some testing procedures including Chen et al. (2011) and the proposed rLRT as shown in Figure 2. The figure shows that (i) the variances of the rLRT are smaller than the cLRT under both null and alternative hypotheses and (ii) the distances between the null and the alternative distributions are more larger in the rLRT than in the cLRT.

![Figure 2: n = 40, p = 32. Empirical distribution functions of the rLRT and the cLRT under the null and three alternative hypotheses.](image)

### 4.2 Power comparison with other existing procedures

In this section, we numerically compare the empirical sizes and powers of rLRT statistic for the identity null $H_0 : \Sigma = I_p$ to other existing tests, the corrected LRT (cLRT) by Bai et al. (2009), the invariant test by Ledoit and Wolf (2002), and the non-parametric test by Chen et al. (2010). Let $p$-dimensional IID random vectors $X_1, X_2, \ldots, X_n$ be generated from multivariate normal distribution $\text{MVN}_p(0, \Sigma)$.

In the study, we first generate the data under $\Sigma = I_p$ to report the empirical sizes. The sample size $n$ was set to be 20, 40, 80, and 160, and, for each $n$, $\gamma = p/n$ was chosen as
0.2, 0.5, and 0.8. For example, the case of $n = 160$ gives the simulation results of $p = 32$, 80, and 128. The size of test was designed to be 0.05 based on the asymptotic result. We also set the tuning parameter $\lambda$ of rLRT($\lambda$) was put as 0.2, 0.5, and 0.8 to investigate the effect of the magnitude of the linear shrinkage. Thus, we compare the empirical sizes of cLRT, rLRT(0.2), rLRT(0.5), and rLRT(0.8) under varying $n$ and $\gamma$.

To compare the powers, we consider the following two alternatives, (1) Independent but heteroscedastic variance as $\Sigma = \text{diag}(2, 2, \cdots, 2, 1, 1, \cdots, 1)$ where the number of 2’s is approximately one-fifth of $p$ and (2) Compound symmetry $\Sigma = \Sigma_{cs}(0.1)$ as defined in (18). The sample size $n$ and $\gamma$ are chosen as the same with the above.

The empirical sizes and powers of the listed methods are reported in Table 1. Some interesting results we find from Table 1 are as follows. First, the table confirms our finding in Section 4.1 that the powers of the rLRTs are higher than those of the cLRT in all cases we considered. In addition, it is interesting to note that the power improvement is specially higher in the case $\gamma = 0.8$ (when $p$ is relatively large). Second, the empirical sizes of the invariant test by Ledoit and Wolf (2002) tend to approach to the aimed level 0.05 as $n$ increases. However, the convergence is slow and they are biased in upward in all cases we considered here. For this reason, in Table 1, we also report the powers after correcting the size (empirically) for a fair comparison, where the cut off value is decided based on 5000 simulated test statistics under the null for each simulation setting. In comparison to the invariant test by Ledoit and Wolf (2002), the proposed rLRT has higher power for most of the cases. Thirdly, in comparison to the non-parametric test by Chen et al. (2010), the proposed rLRT shows higher power when the samples are from the independent heterogeneous alternatives but has lower power for other cases where variables are globally dependent to each other. We conjecture this may be due to the difference of how the tests reflect the eigen-structures of the alternative covariance matrix. Roughly speaking, the rLRT depends on the eigenvalues via the proxy $g_1(x) = \lambda x + (1 - \lambda) - \log (\lambda x + (1 - \lambda)) - 1$, whereas Chen et al. (2010) depends on $g_2(x) = x^2 - 2x + 1$. Finally, we remark that the computational cost of Chen et al. (2010) is at least $O(pn^4)$ due to the 4–th moment calculation so it is not suitable for data of large $n$. In fact, to test the data of $n = 500$, Chen et al. (2010) needs nearly tens of hours to finish the job (using C coding and standard Core-i7 CPU), whereas the other tests only need just times of seconds.
| $\gamma$ | $n$ | cLRT | $rLRT$ (0.8) | $rLRT$ (0.5) | $rLRT$ (0.2) | Chen | LW |
|------|----|------|-------------|-------------|-------------|------|----|
|      |    |      |             |             |             |      |    |
| Null | 0.2| 20   | 0.072       | 0.069       | 0.071       | 0.079| 0.088|
|      |    | 40   | 0.076       | 0.074       | 0.072       | 0.079| 0.079|
|      |    | 80   | 0.057       | 0.059       | 0.063       | 0.059| 0.075|
|      |    | 160  | 0.070       | 0.059       | 0.059       | 0.049| 0.073|
|      | 0.5| 20   | 0.073       | 0.068       | 0.068       | 0.074| 0.084|
|      |    | 40   | 0.072       | 0.059       | 0.065       | 0.071| 0.090|
|      |    | 80   | 0.061       | 0.071       | 0.071       | 0.074| 0.102|
|      |    | 160  | 0.048       | 0.040       | 0.047       | 0.057| 0.088|
|      | 0.8| 20   | 0.067       | 0.059       | 0.072       | 0.086| 0.149|
|      |    | 40   | 0.055       | 0.056       | 0.065       | 0.050| 0.116|
|      |    | 80   | 0.052       | 0.053       | 0.052       | 0.056| 0.112|
|      |    | 160  | 0.053       | 0.050       | 0.057       | 0.059| 0.101|
| Indep.| 0.2| 20   | 0.388       | 0.542       | 0.574       | 0.474| 0.516 (0.433)|
| with |    | 40   | 0.410       | 0.589       | 0.645       | 0.543| 0.596 (0.524)|
| hetero. |  | 80   | 0.949       | 0.995       | 1.000       | 0.990| 0.990 (0.985)|
| variance |    | 160  | 1.000       | 1.000       | 1.000       | 1.000| 1.000 (1.000)|
|      | 0.5| 20   | 0.285       | 0.640       | 0.694       | 0.494| 0.589 (0.484)|
|      |    | 40   | 0.576       | 0.966       | 0.970       | 0.836| 0.910 (0.881)|
|      |    | 80   | 0.965       | 1.000       | 1.000       | 0.999| 0.998 (0.997)|
|      |    | 160  | 1.000       | 1.000       | 1.000       | 1.000| 1.000 (1.000)|
|      | 0.8| 20   | 0.195       | 0.716       | 0.759       | 0.505| 0.641 (0.487)|
|      |    | 40   | 0.407       | 0.989       | 0.992       | 0.855| 0.924 (0.857)|
|      |    | 80   | 0.872       | 1.000       | 1.000       | 0.998| 1.000 (1.000)|
|      |    | 160  | 1.000       | 1.000       | 1.000       | 1.000| 1.000 (1.000)|
| Compound | 0.2| 20   | 0.126       | 0.124       | 0.130       | 0.134| 0.155 (0.086)|
| symmetry |    | 40   | 0.292       | 0.319       | 0.323       | 0.370| 0.357 (0.268)|
| with $\rho = 0.1$ |    | 80   | 0.826       | 0.878       | 0.893       | 0.961| 0.961 (0.946)|
|      |    | 160  | 1.000       | 1.000       | 1.000       | 1.000| 1.000 (1.000)|
|      | 0.5| 20   | 0.159       | 0.204       | 0.219       | 0.275| 0.284 (0.189)|
|      |    | 40   | 0.419       | 0.554       | 0.586       | 0.800| 0.789 (0.728)|
|      |    | 80   | 0.962       | 0.985       | 0.991       | 1.000| 1.000 (1.000)|
|      |    | 160  | 1.000       | 1.000       | 1.000       | 1.000| 1.000 (1.000)|
|      | 0.8| 20   | 0.137       | 0.252       | 0.271       | 0.342| 0.348 (0.326)|
|      |    | 40   | 0.394       | 0.643       | 0.683       | 0.941| 0.941 (0.906)|
|      |    | 80   | 0.956       | 0.997       | 0.999       | 1.000| 1.000 (1.000)|
|      |    | 160  | 1.000       | 1.000       | 1.000       | 1.000| 1.000 (1.000)|

Table 1: Summary of sizes and powers. We use centred sample covariance matrix $S_{cen}$ for Ledoit and Wolf (2002) and Chen et al. (2010), and uncentered $S$ for the cLRT and the rLRT. The empirically corrected powers of Ledoit and Wolf (2002) are reported in the parenthesis.
5 Discussion

We conclude the paper with some detailed discussions on the proposed rLRT not covered in the main body of the paper.

First, in the cLRT and the rLRT, we may use sample covariance matrix of the version \( S_{cen} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^\top \) instead of \( S = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top \) (or more generally \( S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)(X_i - \mu)^\top \) with known population mean vector \( \mu \)). The numerical results not reported in the paper show the empirical sizes of the cLRT and the rLRT with \( S_{cen} \) are biased upward; the bias of the cLRT is much larger than that of the rLRT and it increases as \( \gamma \) increases. The biases of both tests are not very persuasive at a first glance since the matrix \( S_{cen} \) is an one-rank perturbation of \( S \) in the sense that \( S_{cen} = S - XX^\top \). Thus, \( \sup_t |F_S(t) - F_{S_{cen}}(t)| \leq 1/n \) and \( S_{cen} \) has the same limiting SD with \( S \). However, the LSS is the sum of perturbed eigenvalues where the perturbations are amplified and no longer negligible. Recently, Pan (2014) provides a modified CLT of the LSS of the matrix \( S_{cen} \) and Wang et al. (2013) applies it to finding the asymptotic distribution of the cLRT with \( S_{cen} \).

The second discussion is on the case \( \gamma \), the limit of \( p/n \), is larger than 1. The cLRT can be defined only in \( \gamma \in (0, 1) \) because of the log-term, \( \log |S| = \sum_i \log (l_i) \); if some of the eigenvalues are 0, the statistic is not well defined anymore. However, in the rLRT, the corresponding term \( \sum_i \log \psi(l_i) \) is positive after a suitable shrinkage. Thus, it is natural to question whether the results of this paper can be extended to \( \gamma \in [1, \infty) \). The answer is positive, since the CLT of the LSS, including the rLRT statistic, is based on the results of Bai and Silverstein (2004) which assume \( \gamma \in [0, \infty) \). In truth, if \( 1 \leq \gamma < \infty \), the LSD of \( S_n \) has a point mass at zero and extra terms are added to the results of Bai et al. (2009) based on Bai and Silverstein (2004); so do to our Theorems 1 and 2.

The third discussion is on the closely spiked eigenvalues. In the compound symmetry model of (18), the close spike eigenvalues are those smaller than \( 1 + \sqrt{\gamma} \). Thus, in Figure 1, the power curves for the interval \( \beta \in (0, \sqrt{\gamma}) \) are not supported by our main results. The powers in the given regions could be obtained simply by extending the formula of Corollary 1 to \( (0, \sqrt{\gamma}) \). We remark that, if we follow Wang et al. (2014), the term \( \log \psi(1 + \beta) \) in Corollary 1 should be omitted on \( (0, \sqrt{\gamma}) \).

Our final discussion is given to the selection of the tuning parameter \( \lambda \). The selection of \( \lambda \)
for the purpose of a good estimator of $\Sigma$ has been well studied in the literature. Particularly, that for the linear shrinkage covariance matrix is studied by Ledoit and Wolf (2004), Schäfer and Strimmer (2005) and Warton (2008). However, our additional numerical investigation shows that this choice for the estimation purpose does not deliver us a power gain in testing and the optimal selection of $\lambda$ for the hypothesis testing clearly needs further research. For this reason, at this time, based on the numerical study in Section 4.2 we recommend readers use simply $\lambda = 0.5$ which provides reasonable powers in all cases.

A Proof of the main theorems

A.1 Proof of Theorem 1

Theorem 1 is a direct consequence of Proposition 1. Here, we calculate the integrals (7) and (8) for
g$(z) = g_1(z) = g_2(z) = \psi(z) - \log(\psi(z)) - 1$,
where $\psi(z) = \lambda z + (1 - \lambda)$. Mean: Using Proposition 1 and the substitution $x = 1 + \gamma - 2\sqrt{\gamma}\cos\theta$,

$$
\mu(g_i) = \frac{g_i(a(\gamma)) + g_i(b(\gamma))}{4} - \frac{1}{2\pi} \int_{a(\gamma)}^{b(\gamma)} \frac{g_i(x)}{\sqrt{4\gamma - (x - 1 - \gamma)^2}} dx
$$

$$
= \frac{\lambda\gamma - \log \sqrt{(1 + \lambda\gamma)^2 - 4\lambda^2\gamma}}{2}
$$

$$
- \frac{1}{2\pi} \int_{0}^{\pi} (\lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta - \log(1 + \lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta)) d\theta
$$

$$
= -\frac{\log \sqrt{(1 + \lambda\gamma)^2 - 4\lambda^2\gamma}}{2} + \frac{1}{2\pi} \int_{0}^{\pi} (\log(1 + \lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta)) d\theta
$$

$$
= -\frac{\log \sqrt{(1 + \lambda\gamma)^2 - 4\lambda^2\gamma}}{2} + \frac{1}{4\pi} \int_{0}^{2\pi} (\log(1 + \lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta)) d\theta.
$$

Variance: We write $m_1 := m(z_1)$ and $m_2 = m(z_2)$ for notational simplicity. We have

$$
v(g) = -\frac{1}{2\pi^2} \int g(z_2(m_2)) \int \frac{g(z_1(m_1))}{(m_1 - m_2)^2} dm_1 dm_2,
$$

(20)

To evaluate this integral with Cauchy’s formula, we need to identify the points of singularity in $g(z(m))$. It can be seen that there is singularity when $\psi(z(m)) = 0$. Rewrite $\psi(z(m))$
as
\[ \psi(z(m_1)) = \lambda \left( -\frac{1}{m_1} + \frac{\gamma}{m_1 + 1} \right) + 1 - \lambda = \frac{(1 - \lambda)(m_1 - M)(m_1 - N)}{m_1(m_1 + 1)} \]
where
\[ M, N = \frac{-(1 - 2\lambda + \lambda \gamma) \pm \sqrt{(1 - 2\lambda + \lambda \gamma)^2 + 4\lambda(1 - \lambda)}}{2(1 - \lambda)} \]

Then, \( M, N \) are points of singularity. Next, choose contours \( C_1 \) and \( C_2 \) enclosing \(-1\) and \( M\), but not \( 0 \) and \( N \), such that on the contours, the logarithm in \( g(z) \) is single-valued. In addition, \( C_1 \) and \( C_2 \) are chosen so that they do not overlap. Applying integration by parts and Cauchy’s formula, we have
\[
\oint \frac{g(z(m_1))}{(m_1 - m_2)^2} \, dm_1 = \int \left\{ \frac{\lambda}{m_1^2} - \frac{\lambda \gamma}{(m_1 + 1)^2} + \frac{1}{m_1} + \frac{1}{m_1 + 1} - \frac{1}{m_1 - M} - \frac{1}{m_1 - N} \right\} \frac{1}{m_1 - m_2} \, dm_1
\]
\[ = 2\pi i \left\{ \frac{\lambda \gamma}{(m_2 + 1)^2} - \frac{1}{m_2 + 1} + \frac{1}{m_2 - M} \right\} \quad \text{(21)} \]

Then,
\[
v(g) = -\frac{1}{2\pi^2} \oint g(z_2(m_2)) \oint \frac{g(z_1(m_1))}{(m_1 - m_2)^2} \, dm_1 \, dm_2
\]
\[ = -\frac{i}{\pi} \oint g(z_2(m_2)) \left\{ \frac{\lambda \gamma}{(m_2 + 1)^2} - \frac{1}{m_2 + 1} + \frac{1}{m_2 - M} \right\} \, dm_2.
\]

Here,
\[
\oint \frac{g(z(m_2))}{(m_2 + 1)^2} \, dm_2
\]
\[ = \oint \left\{ \frac{\lambda}{m_2^2} - \frac{\lambda \gamma}{(m_2 + 1)^2} + \frac{1}{m_2} + \frac{1}{m_2 + 1} - \frac{1}{m_2 - M} - \frac{1}{m_2 - N} \right\} \frac{1}{m_2 + 1} \, dm_2
\]
\[ = \frac{2\pi i}{1 + N} \quad \text{(22)} \]

Applying integration by parts, Cauchy’s formula, and Lemma \( \Pi \) we have
\[
\oint \frac{g(z(m_1))}{(m_1 - m_2)^2} \, dm_1
\]
\[ = \oint \left\{ \frac{\lambda}{m_1^2} - \frac{\lambda \gamma}{(m_1 + 1)^2} + \frac{1}{m_1} + \frac{1}{m_1 + 1} - \frac{1}{m_1 - M} - \frac{1}{m_1 - N} \right\} \frac{1}{m_1 - m_2} \, dm_1
\]
\[ = 2\pi i \left\{ \frac{\lambda \gamma}{(m_2 + 1)^2} - \frac{1}{m_2 + 1} + \frac{1}{m_2 - M} \right\} \quad \text{(23)} \]
Then,
\[ v(g) = -\frac{1}{2\pi^2} \oint g(z_2(m_2)) \int \frac{g(z_1(m_1))}{(m_1 - m_2)^2} dm_1 dm_2 \]
\[ = -\frac{i}{\pi} \oint g(z_2(m_2)) \left\{ \frac{\lambda \gamma}{(m_2 + 1)^2} - \frac{1}{m_2 + 1} + \frac{1}{m_2 - M} \right\} dm_2. \]

Here, using integration by parts,
\[ \oint \frac{g(z(m_2))}{(m_2 + 1)^2} dm_2 \]
\[ = \oint \left\{ \frac{\lambda}{m_2^2} - \frac{\lambda \gamma}{(m_2 + 1)^2} + \frac{1}{m_2 + 1} - \frac{1}{m_2 - M} - \frac{1}{m_2 - N} \right\} \frac{1}{m_2 + 1} dm_2 \]
\[ = 2\pi i \left( \lambda - 1 + \frac{1}{1 + N} \right). \quad (24) \]

Applying Lemma 1 and Cauchy’s formula, obtain
\[ \oint \frac{g(z(m_2))}{m_2 + 1} dm_2 = -2\pi i \{ \log(1 - \lambda)(1 + N) \} \]
\[ = -2\pi i \{ \log(1 - \lambda)(1 + N) \} \quad (25) \]
and
\[ \oint \frac{g(z(m_2))}{m_2 - M} dm_2 = 2\pi i \left\{ -\frac{\lambda}{M} - \lambda - \log \left( \frac{1 - \lambda}{M} \right) \right\} \]
\[ = 2\pi i \left\{ -\frac{\lambda}{M} - \lambda - \log \left( \frac{1 - \lambda}{M} \right) \right\} \quad (26) \]

Combining the results (23) - (26), we have the desired result of variance. \[ \square \]

**A.2 Proof of Theorem 2**

We compute
\[ \int g(x) dF_{\gamma_n, H_n}(x) = \int \{ \lambda x - \lambda - \log(\lambda x + 1 - \lambda) \} dF_{\gamma_n, H_n}(x), \]
where \( H_n \) is the SD of spiked population model:
\[ H_n(t) = \frac{p - K}{p} \delta_1(t) + \frac{1}{p} \sum_{i=1}^{K} n_i I(t \geq a_i). \]

We utilize the calculation techniques of the Section 3 in [Wang et al. (2014)](Wang et al. (2014)). First, the calculation for polynomial integrands are direct from Section 3.1 of [Wang et al. (2014)](Wang et al. (2014)),
\[ \int (\lambda x - \lambda) dF_{\gamma_n, H_n}(x) = \frac{\lambda}{p} \sum_{i=1}^{K} n_i a_i - \frac{\lambda K}{p} + O \left( \frac{1}{n^2} \right). \]
The complicated part is to evaluate the integration of log-term. We rewrite it following the labeling convention of Proposition 2 and discuss them separately:

$$
\int \log(\lambda x + 1 - \lambda) dF_{\gamma_n,H_n}(x) = (11) + (12) + (13).
$$

From now on, we write \( m \) instead of \( m_0 \) for convenience. Evaluating the terms (11) and (12) involves the contour integral. Recall that the contour \( C \) on (11) and (12) encloses the closed interval \([-\frac{1}{1-\sqrt{\gamma}}, \frac{1}{1+\sqrt{\gamma}}]\) on the real axis of the complex plane and has poles of \( \{m = -1\} \), \( \{m = M\} \). Indeed, it is easy to show that \(-\frac{1}{1-\sqrt{\gamma}} < M < \frac{1}{1+\sqrt{\gamma}} \) and \( N > 0 \) provided \( \gamma \in (0,1) \) and \( n \) is large.

Now we take a strategy similar to Section 3.3 of Wang et al. (2014). Recall that \( \lambda(-\frac{1}{m} + \frac{\gamma_n}{1+m}) + (1 - \lambda) = \frac{(1-\lambda)(m-M)(m-N)}{m(m+1)} \). We have

$$
\begin{align*}
(11) &= \frac{-1}{2\pi ip} \int_C \log \left( \lambda(-\frac{1}{m} + \frac{\gamma_n}{1+m}) + (1 - \lambda) \right) \left( K - \sum_{i=1}^{K} \frac{n_i a_i^2 m}{(1 + a_i m)^2} \right) dm \\
&= \frac{-1}{2\pi ip\gamma_n} \int_C \log \left( \frac{(1-\lambda)(m-N)}{m} \right) \frac{(1 - \lambda)}{m} \left( K - \sum_{i=1}^{K} \frac{n_i a_i^2 m^2 \gamma_n}{(1 + a_i m)^2} \right) dm \\
&= \frac{-K}{2\pi ip\gamma_n} \int_C \log \left( \frac{m-M}{m+1} \right) dm + \frac{1}{2\pi ip\gamma_n} \int_C \log \left( \frac{m-M}{m+1} \right) \sum_{i=1}^{K} \frac{n_i a_i^2 m \gamma_n}{(1 + a_i m)^2} dm \\
&= A_1 + A_2.
\end{align*}
$$

Here,

\[
A_1 = -\frac{K}{2\pi ip\gamma_n} \int_C \log \left( \frac{m-M}{m+1} \right) d\log m,
\]

\[
A_2 = \frac{K}{2\pi ip\gamma_n} \int_C \log m \cdot d\log \left( \frac{m-M}{m+1} \right)
\]

\[
= \frac{K}{2\pi ip\gamma_n} \cdot (M+1) \cdot \int_C \frac{\log m}{(m+1)(m-M)} dm
\]

\[
= \frac{K}{p\gamma_n} \log(-M),
\]
\[
A_2 = \frac{1}{2\pi ip} \oint_C \log \left( \frac{m - M}{m + 1} \right) \sum_{i=1}^{K} \frac{n_i a_i m}{(1 + a_i m)^2} dm \\
= \frac{1}{2\pi ip} \sum_{i=1}^{K} \oint_C \log \left( \frac{m - M}{m + 1} \right) \cdot n_i a_i \left( \frac{1}{1 + a_i m} - \frac{1}{(1 + a_i m)^2} \right) dm \\
\triangleq A_3 - A_4,
\]

where
\[
A_3 = \frac{1}{2\pi ip} \sum_{i=1}^{K} \oint_C \log \left( \frac{m - M}{m + 1} \right) \frac{n_i a_i}{1 + a_i m} dm \\
= \frac{1}{2\pi ip} \sum_{i=1}^{K} \oint_C n_i \log \left( \frac{m - M}{m + 1} \right) d\log(1 + a_i m) \\
= -\frac{1}{2\pi ip} \sum_{i=1}^{K} \oint_C n_i \log(1 + a_i m) \cdot d\log \left( \frac{m - M}{m + 1} \right) \\
= -\frac{1}{2\pi ip} \cdot (M + 1) \sum_{i=1}^{K} \oint_C n_i \log(1 + a_i m) \cdot \frac{1}{(m + 1)(m - M)} dm \\
= \frac{1}{p} \sum_{i=1}^{K} n_i \log(1 - a_i) - \frac{1}{p} \sum_{i=1}^{K} n_i \log(1 + a_i M),
\]

and
\[
A_4 = \frac{1}{2\pi ip} \sum_{i=1}^{K} \oint_C \log \left( \frac{m - M}{m + 1} \right) \frac{n_i a_i}{(1 + a_i m)^2} dm \\
= \frac{1}{2\pi ip} \sum_{i=1}^{K} \oint_C \frac{n_i a_i}{1 + a_i m} \cdot d\log \left( \frac{m - M}{m + 1} \right) \\
= \frac{M + 1}{2\pi ip} \sum_{i=1}^{K} \oint_C \frac{n_i}{(1 + a_i m)(m - M)(m + 1)} dm \\
= \frac{1}{p} \sum_{i=1}^{K} n_i \left( \frac{1}{1 + a_i M} - \frac{1}{1 - a_i} \right).
\]

Combine the results of \( A_1 + A_2 = A_1 + A_3 - A_4 \) to get:
\[
(11) = \frac{K}{p\gamma n} \log(-M) + \frac{1}{p} \sum_{i=1}^{K} n_i \log \left( \frac{1 - a_i}{1 + a_i M} \right) - \frac{1}{p} \sum_{i=1}^{K} n_i \left( \frac{1}{1 + a_i M} - \frac{1}{1 - a_i} \right).
\]

22
Next, we consider the part (12). First, note that we substitute $f = \log \circ \psi$ now, and

\[
(\log \circ \psi)'(\frac{-1}{m} + \frac{\gamma}{1+m}) = \frac{\lambda}{\lambda x + (1-\lambda)} \bigg|_{x=-\frac{1}{m} + \frac{\gamma}{1+m}} = \frac{\lambda m(m+1)}{(1-\lambda)(m-M)(m-N)}.
\]

Then

\[
(12) = -\frac{1}{2\pi ip} \oint_c (\log \circ \psi)'(\frac{-1}{m} + \frac{\gamma}{1+m}) \sum_{i=1}^{K} \left( \frac{n_i a_i}{1 + a_i m} - \frac{n_i}{1 + m} \right) \left( \frac{1}{m} - \frac{\gamma m}{(1+m)^2} \right) dm
\]

\[
= -\frac{1}{2\pi ip} \cdot \frac{\lambda}{1-\lambda} \sum_{i=1}^{K} n_i \oint_c \frac{m(m+1)}{(m-M)(m-N)} \left( \frac{a_i}{1 + a_i m} - \frac{1}{1 + m} \right) \left( \frac{1}{m} - \frac{\gamma m}{(1+m)^2} \right) dm
\]

\[
= -\frac{1}{2\pi ip} \cdot \frac{\lambda}{1-\lambda} \sum_{i=1}^{K} n_i (B_1 - B_2 - B_3 + B_4),
\]

where

\[
B_1 = \oint_c \frac{a_i (m+1)}{(m-M)(m-N)(1 + a_i m)} dm
\]

\[
= 2\pi i \cdot \left( \frac{a_i (M+1)}{(1 + a_i M)(M-N)} \right),
\]

\[
B_2 = \oint_c \frac{a_i \gamma n m^2}{(m-M)(m-N)(1 + a_i m)(m+1)} dm
\]

\[
= 2\pi i \cdot \left( \frac{a_i \gamma_n M^2}{(M-N)(1 + a_i M)(M+1)} + \frac{a_i \gamma_n}{(M+1)(N+1)(1-a_i)} \right),
\]

\[
B_3 = \oint_c \frac{1}{m-N} dm = 2\pi i \cdot \frac{1}{M-N},
\]

and

\[
B_4 = \oint_c \frac{\gamma_n m^2}{(m-M)(m-N)(m+1)^2} dm
\]

\[
= 2\pi i \cdot \left( \frac{\gamma_n M^2}{(M-N)(M+1)^2} - \frac{\gamma_n (2MN + M + N)}{(M+1)^2(N+1)^2} \right).
\]

Collecting the four terms, we have :

\[
(12) = -\frac{\lambda}{p(1-\lambda)} \sum_{i=1}^{K} n_i \left[ \frac{1}{M-N} \left\{ \frac{a_i (M+1)}{1 + a_i M} - \frac{a_i \gamma_n M^2}{(1 + a_i M)(M+1)} - 1 + \frac{\gamma_n M^2}{(M+1)^2} \right\} 
- \frac{1}{(M+1)(N+1)} \left\{ \frac{a_i \gamma_n}{1-a_i} + \frac{\gamma_n (2MN + M + N)}{(M+1)(N+1)} \right\} \right].
\]

23
For the last part (13), note that the integration term \( \int \log(\psi(x))dF_{\gamma_n,\delta_1}(x) \) is exactly equal to \(- \int \{ \psi(x) + \log(\psi(x)) - 1 \}dF_{\gamma_n,\delta_1}(x) \) since M-P law has the property \( \int xdF_{\gamma_n,\delta_1}(x) = \int 1dF_{\gamma_n,\delta_1}(x) \). So we obtain :

\[
(13) = -(1 - \frac{K}{p}) \int g(x)dF_{\gamma_n,\delta_1}(x) + \frac{1}{p} \sum_{i=1}^{K} n_i \log \psi(\varphi(a_i)) + O(\frac{1}{n^2})
\]

where \( \varphi(a_i) = a_i + \frac{\gamma a_i}{a_i - 1} \).

Finally, combining the four results, we finally obtain the centering term :

\[
\int \{ \psi(x) - \log(\psi(x)) - 1 \}dF_{\gamma_n,H_n}(x)
= \int (\lambda x - \lambda)dF_{\gamma_n,\delta_1}(x) - (11) - (12) - (13).
= \left(1 - \frac{K}{p}\right) \int g(x)dF_{\gamma_n,\delta_1}(x) + \frac{1}{p} C_n + O(\frac{1}{n^2}),
\]

where

\[
C_n = \lambda \cdot \frac{1}{K} \sum_{i=1}^{K} n_i a_i - \lambda - \frac{1}{K} \sum_{i=1}^{K} n_i \log \psi(\varphi(a_i))
\]

\[
- \left[ \frac{1}{\gamma_n} \log(-M) + \frac{1}{K} \sum_{i=1}^{K} n_i \log \left( \frac{1 - a_i}{1 + a_i M} \right) - \frac{1}{K} \sum_{i=1}^{K} n_i \left( \frac{1}{1 + a_i M} - \frac{1}{1 - a_i} \right) \right]
\]

\[
+ \frac{\lambda}{(1 - \lambda) K} \sum_{i=1}^{K} n_i \left[ \frac{1}{M - N} \left\{ \frac{a_i(M + 1)}{1 + a_i M} - \frac{a_i \gamma_n M^2}{(1 + a_i M)(M + 1)} \right\} - \frac{1}{(M + 1)(N + 1)} \left\{ \frac{a_i \gamma_n}{1 - a_i} + \frac{\gamma_n (2MN + M + N)}{(M + 1)(N + 1)} \right\} \right].
\]

A.3 Technical lemma

**Lemma 1.** Let \( z_A \) and \( z_B \) be any two different fixed complex numbers. Then, (a) for any contour \( C \) enclosing \( z_A \) and \( z_B \) such that

\[
\log \frac{z - z_A}{z - z_B}
\]

is single-valued on the contour, we have

\[
\int \frac{1}{z - z_A} \log \frac{z - z_A}{z - z_B} dz = 0,
\]

24
(b) for any contour $C$ enclosing $z_A$ and $z_B$, we have

$$\oint \frac{dz}{(z-z_A)(z-z_B)} = 0$$

Figure 3: Illustration of the path of the contour integral

**Proof.** (a) Let $C^*$ be a contour enclosing $C$ such that both $C$ and $C^*$ are clockwise (anticlockwise), and on $C^*$, we have $|z - z_A| < |z_B - z_A|$ (see Figure 3). Then, $D$, the singly connected region between $C$ and $C^*$ as indicated in Figure 3 contains no singularity. Therefore, the integral on $C$ and $C^*$ are the same. Next, consider the power series expansion

$$\log \frac{z - z_A}{z - z_B} = -\log \left(1 + \frac{z_A - z_B}{z - z_A}\right) = - \left\{ \frac{z_A - z_B}{z - z_A} - \frac{1}{2} \left( \frac{z_A - z_B}{z - z_A} \right)^2 + \frac{1}{3} \left( \frac{z_A - z_B}{z - z_A} \right)^3 - \ldots \right\}.$$ 

Such power series converges on $C^*$. The desired result is a consequence of

$$\oint \frac{dz}{(z - z_A)^i} = 0$$

for all $i = 2, 3, \ldots$. 

(b) Applying Cauchy’s formula, we have

$$\oint \frac{dz}{(z-z_A)(z-z_B)} = \frac{1}{z_A - z_B} \oint \left\{ \frac{1}{z-z_A} - \frac{1}{z-z_B} \right\} dz = \frac{1 - 1}{z_A - z_B} = 0.$$ 

\[\square\]
References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, 3rd edition.

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics*, 37(6B):3822–3840.

Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177.

Bai, Z. D. and Silverstein, J. W. J. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605.

Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388.

Chen, L. S., Paul, D., Prentice, R. L., and Wang, P. (2011). A Regularized Hotellings T² Test for Pathway Analysis in Proteomic Studies. *Journal of the American Statistical Association*, 106(496):1345–1360.

Chen, S., Zhang, L., and Zhong, P. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Jiang, D., Jiang, T., and Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference*, 142(8):2241–2256.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

26
Pan, G. (2014). Comparison between two types of large sample covariance matrices. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 50(2):655–677.

Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *The Annals of Applied Probability*, 18(3):1232–1270.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).

Wang, C., Yang, J., Miao, B., and Cao, L. (2013). Identity tests for high dimensional data using RMT. *Journal of Multivariate Analysis*, 118:128–137.

Wang, Q., Silverstein, J. W., and Yao, J.-f. (2014). A note on the CLT of the LSS for sample covariance matrix from a spiked population model. *Journal of Multivariate Analysis*, 130:194–207.

Warton, D. I. (2008). Penalized Normal Likelihood and Ridge Regularization of Correlation and Covariance Matrices. *Journal of the American Statistical Association*, 103(481):340–349.

Won, J.-H., Lim, J., Kim, S.-J., & Rajaratnam, B. (2013). Condition Number Regularized Covariance Estimation. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 75(3):427-450.