Statistics and Its Interface Volume 15 (2022) 449–461

Testing high-dimensional covariance matrices with random projections and corrected likelihood ratio

Nan Sun* and Cheng Yong Tang

Testing the equality between two high-dimensional covariance matrices is challenging. As the efficient way to measure evidential discrepancy from observed data, the likelihood ratio test is expected to be powerful when the null hypothesis is violated. However, when the data dimensionality becomes large and may substantially exceed the sample size, likelihood ratio based approaches are encountering both practical and theoretical difficulties. To solve the problem, we propose in this study to first randomly project the original high-dimensional data to some lower-dimensional space, and then to apply the corrected likelihood ratio tests developed with the random matrix theory. We show that our test is consistent under the null hypothesis. Through evaluating the power function which is a challenging objective in this context, we show evidence that our test based on random projection matrix with reasonable column size is more powerful when the two covariance matrices are unequal but component-wise discrepancy could be small – a weak and dense signal setting. Numerical studies with simulations and a real data analysis confirm the merits of our test.

AMS 2000 subject classifications: 62H15.
Keywords and phrases: Corrected likelihood ratio test, Covariance matrix, Hypothesis testing, Random matrix theory, Random projections.

1. INTRODUCTION

Testing the equality between two covariance matrices is an important statistical problem. In classical multivariate analysis, such a problem is typically solved using the golden rule – the likelihood ratio approach – by assuming that the two-sample data independently follow some multivariate normal distributions; see the monograph [2]. It is well recognized that the likelihood ratio approach is powerful in the context of the conventional setting when the number of the observations goes to infinity, while the dimensionality of the random vectors is fixed. However, it is known that when the data dimensionality also diverges, both the practical applicability and statistical properties of the conventional likelihood ratio approach become challenging.

Recently, there are increasing interests and development for solving the testing problem with high-dimensional data; we refer to [16] and [7] as overviews on this topic. In attempts to address the challenges from large dimensional covariance matrices in testing problems, the random matrix theory [3] has been influential. In the context of likelihood ratio approaches, [4] proposed corrections to the likelihood ratio test statistic, and established the limiting distribution of the corrected test statistic. Along this line, [27] established the central limit theorem for linear spectral statistics with high-dimensional covariance matrices; see also [28].

Besides likelihood ratio based approaches, other discrepancy measures are also utilized in testing statistical hypotheses for covariance matrices. Among them, the maximum element-wise discrepancy is a distinguished candidate. [8] considered testing for sparse covariance matrices with the maximum element-wise discrepancy and established the limiting distribution of the test statistic. Accompanying this class of testing methods, the multiplier bootstrap approaches supported by the high-dimensional Gaussian approximation [13] have demonstrated useful for approximating the distribution of the maximum discrepancy measures. [10] investigated testing the equivalence of two high-dimensional covariance matrices with the Gaussian approximation method and theory; [11] constructed a class of simultaneous confidence regions for a subset of the entries of a large precision matrix based on multiplier bootstrap procedure. Intuitively, the maximum discrepancy is suitable for detecting sparse and relatively stronger signals when the null hypothesis is violated. Besides the maximum discrepancy measure, other discrepancy measures by designated aggregations have also been investigated in the literature for testing problems; see, among others, [12], [22], [18], [17], [8], [9], [15] and [24].

Practically, despite the merits of using the corrected likelihood ratio testing approaches supported by the random matrix theory, there are remaining limitations and challenges in high-dimensional testing problems. In particular, the likelihood ratio test is powerful; but it becomes not applicable, theoretically and practically, when the data dimensionality exceeds the sample size. This constitutes a major challenge for addressing high-dimensional testing problems. As for the approaches based on the maximum element-wise discrepancy, on the other hand, a natural concern is that they may be less powerful in case many but weaker element-wise violations occur when the null hypothesis is violated.

*Corresponding author.

In this study, we develop a new testing procedure with the corrected likelihood ratio approach, accommodating high-dimensional testing problems. The foundational building block of our methods is the random matrix theory supported corrected likelihood ratio tests of, for example, [4] and [27]. With the likelihood ratio based discrepancy measure, our procedure is capable of capturing the evidence from violating the null hypothesis in a most efficient way. To address the challenges from testing with high-dimensional random vectors, we propose to reduce the dimensionality of the problem with random projections – projecting the high-dimensional vectors to some lower-dimensional space. Practically, random projections form a class of computationally efficient dimension reduction approaches and they have appealing properties in maintaining the geometric properties of the high-dimensional data; see [6] and [25]. In the context of testing high-dimensional mean vectors, [21] proposed random projection based tests, and showed that the projected test has the potential of achieving higher power compared with alternative statistical tests for high-dimensional mean vectors; see also [23] for high-dimensional mean test with random projections. Nevertheless, the random projection based testing approaches remain less explored in the context of covariance matrices, in particular concerning the power.

Concretely, our approach first applies random projections to high-dimensional random vectors, and then applies the corrected likelihood ratio approach for testing the equality of the covariance matrices. We demonstrate that under the null hypothesis, the proposed test with projected high-dimensional random vectors follows the standard normal distribution with appropriate centering and normalizing. Since the asymptotic power function of the test takes a complicated form, we carefully develop a numerical evaluation. Our results show evidence that our approach can indeed be more powerful for testing high-dimensional covariance matrix. Compared with the maximum discrepancy based approaches, we show clear evidence favoring our test especially when the component-wise discrepancy is small. Our numerical studies with simulations and data analysis confirm the promising performance of the proposed testing approach.

The rest of this article is organized as follows. The methodology is presented in Section 2, followed by its properties studied in Section 3. Numerical investigations containing simulation studies and a real data analysis are given in Section 4. We conclude the paper with some discussions in Section 5.

2. METHODOLOGY

2.1 Overview

Let $X = (X_1, \cdots, X_p)^T$ and $Y = (Y_1, \cdots, Y_p)^T$ be two p-dimensional random vectors with means $\mu_1 = (\mu_{11}, \cdots, \mu_{1p})^T$ and $\mu_2 = (\mu_{21}, \cdots, \mu_{2p})^T$, and covariance matrices $\Sigma_1 = (\sigma_{1,kl})_{1\le k,l\le p}$ and $\Sigma_2 = (\sigma_{2,kl})_{1\le k,l\le p}$, respectively. We are interested in testing the equality of the two covariance matrices:

$$\mathrm{H}_0 : \Sigma_1 = \Sigma_2 \text{ vs } \mathrm{H}_1 : \Sigma_1 \neq \Sigma_2. \tag{1}$$

For testing (1), we observe two random samples, denoted by $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ that are mutually independent, and identically follow the respective distributions of X and Y.

Working with the multivariate normal distribution, the likelihood ratio test applies. In particular, with regular cases, the test statistic is:

$$\lambda = \frac{\left|\frac{N_1}{n_1}S_1\right|^{n_1/2}\left|\frac{N_2}{n_2}S_2\right|^{n_2/2}}{\left|\frac{N_1}{n_1+n_2}S_1 + \frac{N_2}{n_1+n_2}S_2\right|^{(n_1+n_2)/2}}, \tag{2}$$

where S_1, S_2 are the empirical covariance matrices of $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ respectively, and $|\cdot|$ is the determinant of a matrix, $N_i = n_i - 1$, $i = 1, 2$. Standard theory says that when p is fixed, as $\min(n_1, n_2) \to \infty$, $-2\log\lambda$ converges to $\chi^2_{p(p+1)/2}$ distribution asymptotically under H_0; see [2].

Correction and adjustment to the likelihood ratio test statistic (2) have been well studied. To correct the bias and improve the coverage accuracy, [5] proposed a modified likelihood ratio statistic $-2\log\lambda^*$ that simply replaces all n_i ($i = 1,\ 2$) appearing in (2) by the associative N_i.

Clearly, the chi-square limiting distribution results from asymptotic quadratic form. When p is also diverging with n_1, n_2, the chi-square limiting distribution may provide poor approximation to the test statistic; see [4]. Further, the likelihood ratio statistic (2) even fail to define when p is larger than n_i ($i = 1, 2$), because the determinants therein are seen as zeros.

Further adjustments are thus developed for large sample covariance matrices, supported with the random matrix theory [3]. A prominent result for the properties of the corrected likelihood ratio test statistic from [27] is presented as follows. Instead of the chi-square distribution, a standard normal distribution for the likelihood ratio statistic is established under the null hypothesis, with appropriate centering and normalization. [27] derived central limit theorems for linear spectral statistics from the model $\mu + \Sigma^{1/2}u$, where μ is unknown and u consists of independent and identically distributed random variables. For general u, the central limit theorems involve unknown parameters such as the fourth moment. For ease in our analysis and presentation without compromising the spirit of our approach, we consider normal distributions: $N(\mu, \Sigma)$.

Assumption 1. *$X \sim N(\mu_1, \Sigma_1)$, and $Y \sim N(\mu_2, \Sigma_2)$.*

Assumption 2. *The ratio of dimension-to-sample size $\hat{y}_1 = p/N_1 \to y_1 > 0$ as $n_1, p \to \infty$, and $\hat{y}_2 = p/N_2 \to y_2 \in (0, 1)$ as $n_2, p \to \infty$.*

Assumption 3. *The sequences* $\left\{\Sigma_1 = \Gamma_1\Gamma_1^T\right\}_{p\geq 1}$ *and* $\left\{\Sigma_2 = \Gamma_2\Gamma_2^T\right\}_{p\geq 1}$ *are bounded in spectral norm, and empirical spectral distributions of both* $H_{1,p}(t) = \sum_{j=1}^{p} I\left(\gamma_{j,1} \leq t\right)/p$ *and* $H_{2,p}(t) = \sum_{j=1}^{p} I\left(\gamma_{j,2} \leq t\right)/p$ *converge weakly to limiting spectral distributions* $H_1(t)$ *and* $H_2(t)$ *respectively, where* $\{\gamma_{j,1}\}_{j=1}^{p}$ *and* $\{\gamma_{j,2}\}_{j=1}^{p}$ *are eigenvalues of* Σ_1 *and* Σ_2 *respectively, and* $I(\cdot)$ *is an indicator function.*

Proposition 1. *Assume that* X *and* Y *satisfy Assumptions 1–2–3. Then under the null hypothesis* $H_0 : \Sigma_1 = \Sigma_2$,

$$v^{-1/2}\left(-\frac{2\log\lambda^*}{N_1+N_2} - pF - m\right) \xrightarrow{\mathcal{D}} N(0,1),$$

where

$$\begin{aligned}
v =& v(\hat{y}_1, \hat{y}_2, y_1, y_2)\\
=& -\frac{2\hat{y}_2^2}{\left(\hat{y}_1+\hat{y}_2\right)^2}\log\left(1-\hat{y}_1\right)\\
&-\frac{2\hat{y}_1^2}{\left(\hat{y}_1+\hat{y}_2\right)^2}\log\left(1-\hat{y}_2\right)\\
&+2\log\left(\frac{y_1+y_2-y_1y_2}{\hat{y}_1+\hat{y}_2}\right),\\
m =& m(\hat{y}_1, \hat{y}_2, y_1, y_2)\\
=& \frac{1}{2}\log\left(\frac{y_1+y_2-y_1y_2}{\hat{y}_1+\hat{y}_2}\right)\\
&-\frac{1}{2}\frac{\hat{y}_1}{\hat{y}_1+\hat{y}_2}\log\left(1-\hat{y}_2\right)\\
&-\frac{1}{2}\frac{\hat{y}_2}{\hat{y}_1+\hat{y}_2}\log\left(1-\hat{y}_1\right),\\
F =& F(\hat{y}_1, \hat{y}_2, y_1, y_2)\\
=& \frac{y_1+y_2-y_1y_2}{\hat{y}_1\hat{y}_2}\log\left(\frac{\hat{y}_1+\hat{y}_2}{y_1+y_2-y_1y_2}\right)\\
&+\frac{\hat{y}_1\left(1-\hat{y}_2\right)}{\hat{y}_2\left(\hat{y}_1+\hat{y}_2\right)}\log\left(1-\hat{y}_2\right)\\
&+\frac{\hat{y}_2\left(1-\hat{y}_1\right)}{\hat{y}_1\left(\hat{y}_1+\hat{y}_2\right)}\log\left(1-\hat{y}_1\right).
\end{aligned}$$

2.2 Covariance test with random projections

As seen from (2) and the assumptions in Proposition 1, though the dimensionality of the random vector is allowed to diverge, it is required to be smaller than the sample size. Otherwise, the test statistic is not defined. Hence, the test is not applicable when p exceeds N_1 or N_2.

To address the problem, we propose to project the p-dimensional random vectors to a k-dimensional space. Let $R \in \mathbb{R}^{p\times k}$ $(k \in \{1, \cdots, \min\left(n_1, n_2, p\right)\})$ be a linear transformation matrix with orthogonal columns such that $R^T R = I_k$, where I_k is the identity matrix of size k. The orthogonality is not a strong requirement, because the Gram–Schmidt process can ensure it if otherwise. Clearly, $\Sigma_1 = \Sigma_2$ implies $R^T\Sigma_1 R = R^T\Sigma_2 R$. Hence, we consider testing

$$\begin{aligned}
&\mathrm{H}_{0,\mathrm{proj}} : R^T\Sigma_1 R = R^T\Sigma_2 R\\
&\mathrm{H}_{1,\mathrm{proj}} : R^T\Sigma_1 R \neq R^T\Sigma_2 R.
\end{aligned} \tag{3}$$

This test (3) serves for the same purpose as the original test (1) by validating that it is a necessary condition.

We propose to apply random projection in generating R. That is, to randomly generate the entries of R from some distributions, e.g., the standard normal distribution, and other sparse random matrices; see, for example [1]. For high-dimensional problems, though randomly generated R may not have exactly orthogonal columns, $R^T R$ is actually close to I_k; see [6]. In our implementation, we generate entries of R independently from standard normal distribution, and then apply the Gram–Schmidt process.

We propose to apply the corrected likelihood ratio test (CLRT) of [27] on the projected random vectors. In particular, we investigate the modified likelihood ratio statistic with projected data using R:

$$\lambda^{(R)} = \frac{\left|R^T S_1 R\right|^{N_1/2}\left|R^T S_2 R\right|^{N_2/2}}{\left|\frac{N_1}{N_1+N_2}R^T S_1 R + \frac{N_2}{N_1+N_2}R^T S_2 R\right|^{(N_1+N_2)/2}}. \tag{4}$$

Here we note that $R^T S_i R$, $(i = 1, 2)$, are invertible with probability 1, as implied from Lemma 1 in [23], provided that $R^T R = I_k$ and Σ_i's $(i = 1, 2)$ are positive definite. Based on (4) and following [27], we define a corrected likelihood ratio test statistic as

$$Z^{(R)} = v^{*-1/2}\left(-\frac{2\log\lambda^{(R)}}{N_1+N_2} - kF^* - m^*\right), \tag{5}$$

where $\hat{y}_i^* = k/N_i$,

$$\begin{aligned}
v^* &= v\left(\hat{y}_1^*, \hat{y}_2^*, y_1^*, y_2^*\right),\\
m^* &= m\left(\hat{y}_1^*, \hat{y}_2^*, y_1^*, y_2^*\right),\\
F^* &= F(\hat{y}_1^*, \hat{y}_2^*, y_1^*, y_2^*),
\end{aligned}$$

and y_i^* are limits of $\hat{y}_i^*(i = 1, 2)$.

Now, the dimensionality of the test statistic becomes k instead of p; k thus plays the same role as p in Proposition 1. Therefore, our approach is broadly applicable to solve high-dimensional testing problems for covariance matrices. Assumptions for Proposition 1 can be validated, so that $Z^{(R)}$ converges in distribution to standard normal under the null hypothesis; see Section 3.

Concretely, we summarize our procedure as follows.

1. Randomly generate matrix $R \in \mathbb{R}^{p\times k}$ with entries from the standard normal distribution; upon applying the Gram–Schmidt process, $R^T R = I_k$;
2. Compute modified likelihood ratio $\lambda^{(R)}$ by (4);

3. Compute the corrected likelihood ratio statistic $Z^{(R)}$ by (5) with y_i^* replaced by $\hat{y}_i^*$, $i = 1, 2$;
4. For Bartlett's modified likelihood ratio test, reject H_0 if $\lambda^{(R)}$ exceeds the $(1-\alpha)$-level quantile of $\chi^2_{k(k+1)/2}$ distribution; for corrected likelihood ratio test, reject H_0 if $\left|Z^{(R)}\right|$ exceeds the $(1-\alpha/2)$-level quantile of the standard normal distribution.

3. THEORY

3.1 Consistency of the test

Assumption 4. *The ratios of dimension-to-sample size $\hat{y}_1^* = k/N_1 \to y_1^* > 0$ as $n_1, k \to \infty$, and $\hat{y}_2^* = k/N_2 \to y_2^* \in (0,1)$ as $n_2, k \to \infty$.*

Assumption 5. *The sequences of both $\{R^T\Sigma_1 R = \Gamma_1^*\Gamma_1^{*T}\}_{k\geq 1}$ and $\left\{R^T\Sigma_2 R = \Gamma_2^*\Gamma_2^{*T}\right\}_{k\geq 1}$ are bounded in spectral norm, and empirical spectral distributions of both $H_{1,k}^*(t) = \sum_{j=1}^k I\left(\gamma_{j,1}^* \leq t\right)/k$ and $H_{2,k}^*(t) = \sum_{j=1}^k I\left(\gamma_{j,2}^* \leq t\right)/k$ converge weakly to limiting spectral distributions $H_1^*(t)$ and $H_2^*(t)$ respectively, where $\left\{\gamma_{j,1}^*\right\}_{j=1}^k$ and $\left\{\gamma_{j,2}^*\right\}_{j=1}^k$ are eigenvalues of $R^T\Sigma_1 R$ and $R^T\Sigma_2 R$ respectively.*

Theorem 1. *Assume that X and Y satisfy Assumptions 1–4–5. Conditioning on R and under the null hypothesis $H_0 : \Sigma_1 = \Sigma_2$,*

$$\mathbb{P}\left(Z^{(R)} \leq x \mid R\right) \to \Phi(x),$$

where $x \in \mathbb{R}$.

Theorem 1 shows that the random projection based test is consistent for any random projection matrix. Clearly, the random projection matrix R together with the specific form of the alternative hypothesis are jointly determining the power of the test; see our result in Section 3.2. We show that the random projection based test is advantageous when testing against alternatives with weak and dense signals, both approximately using theory and empirically using simulations; see our evaluations of the power functions in Section 4.

3.2 Power of the test

Clearly, the power of the test using $Z^{(R)}$ is determined by its distribution when the null hypothesis is violated. We have the following result for the distribution of $Z^{(R)}$ under a more general data generating process.

Assumption 6. *Define*

$$T^{(R)} = \left(R^T\Sigma_2 R\right)^{-1/2}\left(R^T\Sigma_1 R\right)\left(R^T\Sigma_2 R\right)^{-1/2}.$$

Sequence of $\left\{T^{(R)} = \Lambda\Lambda^T\right\}_{k\geq 1}$ is bounded in spectral norm, and the empirical spectral distribution of $H_k(t) = \sum_{j=1}^k I(\lambda_j \leq t)/k$ converges weakly to limiting spectral distribution $H(t)$, where $\{\lambda_j\}_{j=1}^k$ are eigenvalues of $T^{(R)}$.

Theorem 2. *Assume that X and Y satisfy Assumptions 1–4–6. Conditioning on R and under the alternative hypothesis $H_1 : \Sigma_1 \neq \Sigma_2$, $v^{*1/2}Z^{(R)} + kF^* + m^*$ converges in distribution to a normal distribution with mean $kF_1^* + m_1^*$ and variance v_1^*. That is,*

$$\mathbb{P}\left(v^{*1/2}Z^{(R)} + kF^* + m^* \leq x \mid R\right) \to \Phi\left(\frac{x - kF_1^* - m_1^*}{v_1^*}\right),$$

where $x \in \mathbb{R}$.

Define

$$f(x) = \log\left(\hat{y}_1^* + \hat{y}_2^* x\right) - \frac{\hat{y}_2^*}{\hat{y}_1^* + \hat{y}_2^*}\log x - \log\left(\hat{y}_1^* + \hat{y}_2^*\right),$$

then

$$\begin{aligned}(6)\quad m_1^* = &-\frac{1}{4\pi i}\oint_{\mathcal{C}} f'(z)\log\left\{\frac{y_1^* + y_2^* - y_1^* y_2^*}{y_2^*}\right.\\ &\left.-\frac{y_1^*}{y_2^*}\frac{\left(1 - y_2^*\int\frac{m_0(z)}{t+m_0(z)}dH(t)\right)^2}{1 - y_2^*\int\frac{m_0^2(z)}{(t+m_0(z))^2}dH(t)}\right\}dz\\ &-\frac{1}{4\pi i}\oint_{\mathcal{C}} f'(z)\log\left\{1 - y_2^*\int\frac{m_0^2(z)}{\left(t+m_0(z)\right)^2}dH(t)\right\}dz,\end{aligned}$$

where

$$(7)\quad m_0(z) = z\left\{\frac{y_1^* + y_2^* - y_1^* y_2^*}{y_2^*\left(-1 + y_2^*\int\frac{m_0(z)}{t+m_0(z)}dH(t)\right)} + \frac{y_1^*}{y_2^*}\right\}^{-1},$$

and $H(t)$ is the limiting spectral distribution of $T^{(R)}$.

Besides, by Theorem B.10 of [3], define

$$(8)\qquad \underline{m}(z) = \frac{1}{m_0(z)} - y_2^*\int\frac{1}{t+m_0(z)}dH(t),$$

$$(9)\qquad u^*(x) = \frac{1}{\pi y_1^*}\lim_{\epsilon\to 0+}\Im\left\{\underline{m}(x+\epsilon i)\right\},$$

$$(10)\qquad \underline{m}(x) = \lim_{\epsilon\to 0+}\underline{m}(x+\epsilon i),$$

then

$$(11)\qquad F_1^* = \int_{c_1}^{c_2} f(x)u^*(x)dx,$$

where (c_1, c_2) is the support of the limiting spectral density u^* of the general Fisher matrix $\left(R^T S_1 R\right)\left(R^T S_2 R\right)^{-1}$ conditional on R, see the definition of general Fisher matrix in [28]. $\mathcal{C}$ in the formula of m_1^* is the contour enclosing the support set (c_1, c_2). $\Im(\cdot)$ denotes the imaginary part of a complex number.

Finally,

$$v_1^* = \frac{1}{\pi^2} \int\int f'(x) f'(y) \log \left| \frac{\underline{m}(x) - \overline{\underline{m}(y)}}{\underline{m}(x) - \underline{m}(y)} \right| dxdy, \tag{12}$$

where $\overline{z}$ denotes conjugation of a complex number z.

From Theorem 2, we have the following result.

Corollary 1. *Under the alternative hypothesis $H_1 : \Sigma_1 \neq \Sigma_2$,*

$$\mathbb{P}\left\{\Phi^{-1}\left(\frac{\alpha}{2}\right) < Z^{(R)} < \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \mid R\right\} =$$
$$\Phi\left[v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\}\right]$$
$$-\Phi\left[v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\}\right].$$

The corollary follows by observing that

$$\begin{aligned}
&v^{*1/2}\Phi^{-1}\left(\frac{\alpha}{2}\right) + kF^* + m^* \\
&< v^{*1/2} Z^{(R)} + kF^* + m^* \\
&< v^{*1/2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) + kF^* + m^*,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\} \\
&< v_1^{*-1/2}\left\{v^{*1/2} Z^{(R)} + kF^* - kF_1^* + m^* - m_1^*\right\} \\
&< v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\}.
\end{aligned}$$

Then applying the asymptotic normality result from Theorem 2, the corollary follows.

Here from Theorem 2, we note that the power of the test is decided by the spectrum properties of $T^{(R)} = \left(R^T\Sigma_2 R\right)^{-1/2}\left(R^T\Sigma_1 R\right)\left(R^T\Sigma_2 R\right)^{-1/2}$. However, it is known that it is generally difficult to evaluate the quantities in Theorem 2, so that evaluating the power function is a challenging problem; see [3]. In Section 3.3, we develop a framework to numerically approximate the power function.

3.3 Evaluating the power function

To evaluate the power function given in Corollary 1, we need to approximate the limiting mean and covariance functions in the central limit theorem for general Fisher matrix; see those items in Theorem 2. General Fisher matrix is defined as the product of one sample covariance matrix and inverse of another sample covariance matrix, where the population covariance matrices Σ_1 and Σ_2 can be arbitrary. When the population covariance matrices are equal, the general Fisher matrix has a simplified name as the Fisher matrix. Evaluating the limiting mean and covariance functions for the Fisher matrix has been studied in [14] and [20], which is a known challenging problem. [26] derived asymptotic power for corrected likelihood ratio test based on random matrix theory concerning one-sample covariance test, however, an explicit power expression is much harder for two-sample covariance test.

Evaluating the limiting mean and covariance functions for general Fisher matrix is nevertheless much harder, because those functions are complex contour integrals though their values are real. Furthermore, since those functions have a very complex structure depending on non-trivial contour integrals, it is more challenging to accurately approximate them, see [28].

To approximate the power function in Corollary 1, we develop a numerical approximation procedure inspired by the approach of [28]. The starting point is to estimate the limiting spectral distribution H of $T^{(R)}$. We use the empirical spectral distribution $H_k(t) = \sum_{j=1}^k \mathrm{I}\left(\lambda_j \le t\right)/k$ to approximate the limiting spectral distribution $H(t)$. The next and an important step is to detect support (c_1, c_2) where $H(t)$ is defined. For such a purpose, we compute the Stieltjes transform $\underline{m}(z)$ by combining fixed point algorithm on $m_0(z)$, given by (7) and (8), on a grid of points $z = x + i\epsilon$, for a range of x, and a small ϵ, e.g., 0.001. Then the support is decided by examining the imaginary part of the $\underline{m}(z)$. That is, when the imaginary part is small, e.g., less than ϵ, we declare that the corresponding value of x is outside of the support.

On the detected support set (c_1, c_2), we specify $\{z_q = x_q + i\epsilon,\ x_q = c_1 + (c_2 - c_1)\, q/Q\}_{q=0}^{Q}$, where ϵ is a small step size, e.g., 0.001, and Q is a large number, e.g., 1000. Then, we compute the Stieltjes transform $\underline{m}\left(z_q\right)$ by combining fixed point algorithm on $m_0(z_q)$ with (7) and (8). $u^*\left(x_q\right)$ is approximated by $\Im\left(\underline{m}\left(z_q\right)\right)/\left(\pi y_1^*\right)$ according to (9). F_1^* is approximated by $\left(c_2 - c_1\right)\sum_{q=0}^{Q} f\left(x_q\right) u^*\left(x_q\right)/(Q+1)$ according to equation (11). To approximate m_1^* and v_1^*, we denote two grid sets as follows:

$$\begin{aligned}
\mathcal{A}_1 = \Bigg\{ & z_k = c_1 - \epsilon + \left(\zeta - \frac{2\zeta k}{K_1}\right)i, \\
& k = 0, \cdots, K_1, \\
& z_{K_1+j} = c_1 - \epsilon + \frac{\left(c_2 - c_1 + 2\epsilon\right)j}{K_2} - \zeta i, \\
& j = 1, \cdots, K_2 - 1, \\
& z_{K_1+K_2+k} = c_2 + \epsilon + \left(-\zeta + \frac{2\zeta k}{K_1}\right)i, \\
& k = 0, \cdots, K_1, \\
& z_{2K_1+K_2+j} = c_2 + \epsilon - \frac{\left(c_2 - c_1 + 2\epsilon\right)j}{K_2} + \zeta i, \\
& j = 1, \cdots, K_2 \Bigg\},
\end{aligned}$$

$$\mathcal{A}_2 = \Big\{ z_k = c_1 - \frac{\epsilon}{2} + \left(\frac{\zeta}{2} - \frac{\zeta k}{K_1}\right) i,$$
$$k = 0, \cdots, K_1,$$
$$z_{K_1+j} = c_1 - \frac{\epsilon}{2} + \frac{(c_2 - c_1 + \epsilon)\, j}{K_2} - \frac{\zeta}{2} i,$$
$$j = 1, \cdots, K_2 - 1,$$
$$z_{K_1+K_2+k} = c_2 + \frac{\epsilon}{2} + \left(-\frac{\zeta}{2} + \frac{\zeta k}{K_1}\right) i,$$
$$k = 0, \cdots, K_1$$
$$z_{2K_1+K_2+j} = c_2 + \frac{\epsilon}{2} - \frac{(c_2 - c_1 + \epsilon)\, j}{K_2} + \frac{\zeta}{2} i,$$
$$j = 1, \cdots, K_2 \Big\},$$

where K_1 and K_2 are large integers, e.g., 1,000, and ϵ and ζ are small numbers, e.g., 0.001.

Then approximation to m_1^* is as follows according to equation (6):

$$\begin{aligned} m_1^* \approx & \\ & -\frac{1}{4\pi} \sum_{j=0}^{2K_1+2K_2-1} \Im\Bigg[f'(z_j)(z_{j+1} - z_j) \\ & \times \log\left\{ \frac{h^{*2}}{y_2^*} - \frac{y_1^*}{y_2^*} \frac{\left(1 - y_2^* \int \frac{m_0(z_j)}{t + m_0(z_j)} dH(t)\right)^2}{1 - y_2^* \int \frac{m_0^2(z_j)}{(t + m_0(z_j))^2} dH(t)} \right\}\Bigg] \\ & -\frac{1}{4\pi} \sum_{j=0}^{2K_1+2K_2-1} \Im\Bigg[f'(z_j)(z_{j+1} - z_j) \\ & \times \log\left\{ 1 - y_2^* \int \frac{m_0^2(z_j)}{(t + m_0(z_j))^2} dH(t) \right\}\Bigg], \end{aligned}$$

where $h^{*2} = y_1^* + y_2^* - y_1^* y_2^*$. And approximation to m_1^* can be computed based on either $\mathcal{A}_1$ or $\mathcal{A}_2$.

Approximation to v_1^* is as follows by equation (10) and (12):

$$\begin{aligned} v_1^* \approx & \frac{(c_2 - c_1)^2}{\pi^2 K_1^2} \sum_{\substack{j,k=0 \\ j \neq k}}^{K_2} \Re\Bigg(f'(z_{K_1+j}) f'(z_{K_1+k}) \\ & \times \log\left| \frac{\underline{m}(z_j) - \overline{\underline{m}(z_k)}}{\underline{m}(z_j) - \underline{m}(z_k)} \right| \Bigg), \end{aligned}$$

where $z_{K_1+j},\ z_j \in \mathcal{A}_1$ and $z_{K_1+k},\ z_k \in \mathcal{A}_2$, and $\Re(\cdot)$ is the real part of a complex number.

With the above procedure, we numerically evaluate the theoretical power function which is one minus the result in Corollary 1 in Section 3.2.

4. NUMERICAL EXAMPLES

4.1 Consistency of the test

We conduct simulation studies to validate the performance of our two-sample covariance matrix test. In our numerical examples, we generate independent samples from normal distributions with zero mean, and set $n_1 = n_2 = n$. For all settings, the simulations are repeated for 1,000 times.

To validate the consistency of the test under the null hypothesis, we generate data from with equal covariance matrices $\Sigma_1 = \Sigma_2 = \Sigma$. Let $D = \text{diag}(d_{11}, \ldots, d_{pp})$ be a diagonal matrix with d_{ii} $(i = 1, \ldots, p)$ randomly generated from Unif$(0.5, 2.5)$ distribution. Denote by $\lambda_p(A)$ the smallest eigenvalue of a symmetric matrix A. Inspired by the study of [8], our simulations evaluate four settings with $\Sigma = \Sigma^{(i)}$ $(i = 1, \ldots, 4)$:

1. $\Sigma^{*(1)} = (\sigma_{ij}^{*(1)})$, where $\sigma_{ii}^{*(1)} = 1$, $\sigma_{ij}^{*(1)} = 0.5$ for $5(k-1)+1 \leq i \neq j \leq 5k$, where $k = 1, \cdots, p/5$ and $\sigma_{ij}^{*(1)} = 0$ otherwise. $\Sigma^{(1)} = D^{1/2}\Sigma^{*(1)}D^{1/2}$.
2. $\Sigma^{*(2)} = (\sigma_{ij}^{*(2)})$, where $\sigma_{ij}^{*(2)} = 0.5^{|i-j|}$ for $1 \leq i, j \leq p$. $\Sigma^{(2)} = D^{1/2}\Sigma^{*(2)}D^{1/2}$.
3. $\Sigma^{*(3)} = (\sigma_{ij}^{*(3)})$, where $\sigma_{ii}^{*(3)} = 1$, $\sigma_{ij}^{*(3)} = 0.5 *$ Bernoulli$(1, 0.05)$ for $i < j$ and $\sigma_{ji}^{*(3)} = \sigma_{ij}^{*(3)}$. $\Sigma^{(3)} = D^{1/2}(\Sigma^{*(3)} + \delta I)/(1+\delta)D^{1/2}$ with $\delta = |\lambda_p(\Sigma^{*(3)})| + 0.05$.
4. $\Sigma^{(4)} = O\Delta O$, where $O = \text{diag}(\omega_1, \cdots, \omega_p)$ and $\omega_1, \cdots, \omega_p \sim$ Unif$(1, 5)$ independently and $\Delta = (a_{ij})$ and $a_{ij} = (-1)^{i+j} 0.4^{|i-j|^{1/10}}$.

The tests are then evaluated from different combinations of n and p. Our method is implemented as described in Section 2.2. We implement the random projections with different k's: $k = [qn^{1/3}]$ with $q = 5, 6, 7$, where $[\cdot]$ denotes the operation of taking integer.

In each simulation, we generate a matrix, of which entries are independently from standard normal distribution, upon applying Gram–Schmidt process on this matrix to get the random projection matrix R. Then, we implement the Bartlett's modified likelihood ratio statistic (4) and compare with the chi-square distribution, and denote it by "RP-BLRT". We implement the corrected likelihood ratio test (5) with the random matrix theory, and denote it by "RP-CLRT".

For comparisons, our test is compared with two tests based on $\max_{1 \leq k, l \leq p} |\sigma_{1,kl} - \sigma_{2,kl}|$, one is of [8], denoted by "CLX", and one is the test of [10], a perturbed variation of "CLX", denoted by "CZZW" in this subsection.

The results are reported in Table 1, Table 2, and Table 3. From these tables, we observe that the corrected test adjusted with the random matrix theory works satisfactorily with the randomly projected vectors. From these tables, we also see that the adjustment with the random matrix theory is necessary, because we observe that the sizes with the Bartlett's correction are way off from the nominal level. This is due to the fact that the chi-square distribution poorly approximates that of the likelihood ratio even when the data dimensionality is moderate.

For all models, we also see that the empirical sizes of the tests of [8], and [10] are satisfactory.

Table 1. Empirical percentage of rejecting H_0 when H_0 is true; $\alpha = 0.05$; $k = [7n^{1/3}]$

n	p	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2				
60	RP-BLRT	97.3	97.4	97.7	97.6	97.4	96.9	97.3	97.6	97.2	97.6
	RP-CLRT	5.8	6.0	5.5	6.0	5.8	6.5	6.2	5.6	6.7	5.1
	CLX	3.8	4.6	5.1	4.6	3.5	3.9	5.9	5.4	3.2	5.6
	CZZW	4.9	5.7	6.3	6.3	5.8	4.9	7.0	7.3	5.3	7.8
100	RP-BLRT	85.4	86.7	86.3	87.8	85.1	86.8	86.7	85.8	85.8	84.1
	RP-CLRT	5.5	6.5	5.4	5.1	5.2	5.2	5.4	4.2	6.2	5.3
	CLX	3.6	4.4	4.2	4.9	3.1	4.6	4.5	4.4	5.1	5.3
	CZZW	4.2	6.2	5.1	5.5	4.4	5.0	5.2	5.1	5.2	5.4
300	RP-BLRT	52.6	55.2	53.1	51.2	54.2	52.7	54.2	50.4	53.0	51.9
	RP-CLRT	5.3	4.8	5.9	3.9	3.8	6.1	4.9	4.5	4.8	5.3
	CLX	3.5	4.5	4.2	3.4	4.6	4.0	4.2	5.2	2.4	7.6
	CZZW	3.8	5.1	4.5	3.4	4.8	4.5	4.8	5.6	2.4	9.0
		Model 3					Model 4				
60	RP-BLRT	96.4	96.8	96.5	96.8	97.5	97.1	96.7	98.1	97.1	97.4
	RP-CLRT	6.2	6.6	4.8	4.7	5.2	4.9	5.1	6.3	6.1	5.4
	CLX	6.0	5.5	5.5	5.6	3.4	4.9	4.0	4.5	4.1	4.0
	CZZW	6.3	6.3	6.6	7.6	4.8	6.6	6.5	6.9	9.1	9.1
100	RP-BLRT	85.2	86.3	85.7	86.8	86.1	84.4	85.9	84.9	85.7	85.5
	RP-CLRT	5.5	6.0	5.8	4.5	4.5	4.9	5.1	4.3	5.5	5.5
	CLX	4.1	3.8	4.7	6.3	4.9	3.9	2.6	5.1	4.7	4.2
	CZZW	4.6	4.8	5.2	6.6	6.7	5.4	4.4	7.1	5.8	5.3
300	RP-BLRT	51.8	55.0	51.0	51.9	53.8	50.4	53.7	50.8	54.1	52.4
	RP-CLRT	6.1	5.9	4.9	4.5	4.9	4.6	5.6	5.2	5.7	5.0
	CLX	3.9	3.4	5.3	4.3	3.8	4.2	3.3	3.3	4.2	6.6
	CZZW	4.2	4.0	5.7	4.6	3.7	5.6	5.2	4.5	5.0	7.6

Table 2. Empirical percentage of rejecting H_0 when H_0 is true; $\alpha = 0.05$; $k = [6n^{1/3}]$

n	p	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2				
60	RP-BLRT	79.8	78.5	80.1	80.1	80.2	79.4	78.3	82.8	78.8	75.8
	RP-CLRT	6.9	7.2	5.2	5.4	4.3	5.9	4.3	5.9	5.6	6.3
	CLX	3.8	4.6	5.1	4.6	3.5	3.9	5.9	5.4	3.2	5.6
	CZZW	4.9	5.7	6.3	6.3	5.8	4.9	7.0	7.3	5.3	7.8
100	RP-BLRT	64.4	64.6	63.2	65.0	66.5	64.7	64.2	64.9	63.4	63.7
	RP-CLRT	5.2	4.5	6.5	6.4	5.6	5.6	5.6	5.8	4.8	5.1
	CLX	3.6	4.4	4.2	4.9	3.1	4.6	4.5	4.4	5.1	5.3
	CZZW	4.2	6.2	5.1	5.5	4.4	5.0	5.2	5.1	5.2	5.4
300	RP-BLRT	33.7	32.0	29.7	30.9	28.5	31.9	34.6	33.5	30.5	30.0
	RP-CLRT	6.1	5.0	6.0	5.4	5.5	6.4	5.1	5.1	4.8	6.0
	CLX	3.5	4.5	4.2	3.4	4.6	4.0	4.2	5.2	2.4	7.6
	CZZW	3.8	5.1	4.5	3.4	4.8	4.5	4.8	5.6	2.4	9.0
		Model 3					Model 4				
60	RP-BLRT	80.0	79.9	78.9	78.5	81.2	77.2	79.0	80.0	80.9	79.6
	RP-CLRT	4.9	6.7	5.1	5.0	5.5	4.7	6.5	6.3	5.7	3.4
	CLX	6.0	5.5	5.5	5.6	3.4	4.9	4.0	4.5	4.1	4.0
	CZZW	6.3	6.3	6.6	7.6	4.8	6.6	6.5	6.9	9.1	9.1
100	RP-BLRT	62.7	65.5	63.3	64.5	65.1	65.3	65.0	61.2	64.5	64.1
	RP-CLRT	5.0	4.9	5.0	5.9	5.9	6.6	5.0	5.9	5.5	4.7
	CLX	4.1	3.8	4.7	6.3	4.9	3.9	2.6	5.1	4.7	4.2
	CZZW	4.6	4.8	5.2	6.6	6.7	5.4	4.4	7.1	5.8	5.3
300	RP-BLRT	34.2	30.3	29.5	28.5	31.1	34.2	30.8	32.5	29.8	32.0
	RP-CLRT	6.1	6.4	4.5	6.7	5.3	6.5	3.7	6.2	4.3	6.8
	CLX	3.9	3.4	5.3	4.3	3.8	4.2	3.3	3.3	4.2	6.6
	CZZW	4.2	4.0	5.7	4.6	3.7	5.6	5.2	4.5	5.0	7.6

4.2 Power comparisons with additive, dense and weak signals

To evaluate the empirical powers of the tests, we generate data from scenarios with H_0 violated. We consider a setting where many components in Σ_2 differ from those in Σ_1. In particular, we generate Σ_1 respectively from the four settings in Section 4.1, and let $\Sigma_2 = \Sigma_1 + U$. Here U is a symmetric matrix with upper triangular as well as diagonal nonzero entries independently generated from $0.1 \times V \times \max_{1 \le j \le p} \sigma_{jj}$, where $\{\sigma_{jj}\}_{j=1}^{p}$ are diagonal entries of Σ_1 and V follows the uniform distribution over the interval $(0, 4)$.

Results of the simulations are reported in Table 4, Table 5, and Table 6. From these tables, we can see that the corrected test adjusted with the random matrix theory is either more powerful than or comparable to tests of [8] and [10] under all different combinations of n and p, and all four different settings, when the signals are additive, dense and weak. In particular, when $n = 300$, $p = 400$, the power superiority of the corrected test adjusted with the random matrix theory over "CLX" and "CZZW" are over 50% for $k = [7n^{1/3}]$, over 45% for $k = [6n^{1/3}]$, over 35% for $k = [5n^{1/3}]$, under all four different settings with tests designed for detecting sparse and strong signals having power around 35%. This reflects the merit of the likelihood ratio approach in aggregating the evidence from violating the null hypothesis.

4.3 Power comparisons with rotational transformations

In this case, we generate Σ_2 by sequentially rotating Σ_1 in multiple subspaces with a small angle θ accompanied with a small extension factor $e = (1 + d)^{1/2}$. Then we evaluate the powers of the tests in this case. We employ the Givens rotation matrix for such a purpose. Let

$$
G(i, j, \theta, d) = \begin{array}{c} \begin{array}{ccccccc} & & j & & i & & \end{array} \\ \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & e\cos(\theta) & \dots & -e\sin(\theta) & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & e\sin(\theta) & \dots & e\cos(\theta) & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ \\ \\ j \\ \\ i \\ \\ \\ \end{array},
$$

Table 3. *Empirical percentage of rejecting H_0 when H_0 is true;* $\alpha = 0.05$; $k = [5n^{1/3}]$

n	p	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2				
60	RP-BLRT	55.3	55.1	56.0	54.0	53.2	53.5	57.4	51.7	56.2	55.9
	RP-CLRT	4.4	5.6	5.2	6.0	6.1	4.7	5.2	7.1	6.6	3.8
	CLX	3.8	4.6	5.1	4.6	3.5	3.9	5.9	5.4	3.2	5.6
	CZZW	4.9	5.7	6.3	6.3	5.8	4.9	7.0	7.3	5.3	7.8
100	RP-BLRT	34.0	33.3	35.8	34.1	33.2	34.5	34.8	33.2	35.1	34.1
	RP-CLRT	5.9	5.1	6.5	4.9	5.1	4.6	4.6	5.1	4.9	5.3
	CLX	3.6	4.4	4.2	4.9	3.1	4.6	4.5	4.4	5.1	5.3
	CZZW	4.2	6.2	5.1	5.5	4.4	5.0	5.2	5.1	5.2	5.4
300	RP-BLRT	15.5	15.4	14.8	17.6	18.0	16.0	17.2	17.4	15.8	15.1
	RP-CLRT	4.8	4.9	4.3	5.7	5.8	6.2	4.2	5.1	5.5	4.9
	CLX	3.5	4.5	4.2	3.4	4.6	4.0	4.2	5.2	2.4	7.6
	CZZW	3.8	5.1	4.5	3.4	4.8	4.5	4.8	5.6	2.4	9.0
		Model 3					Model 4				
60	RP-BLRT	54.5	55.3	56.1	58.6	55.9	54.3	56.1	57.1	53.5	51.7
	RP-CLRT	6.2	4.9	5.9	6.0	7.2	6.1	5.9	5.0	5.5	5.9
	CLX	6.0	5.5	5.5	5.6	3.4	4.9	4.0	4.5	4.1	4.0
	CZZW	6.3	6.3	6.6	7.6	4.8	6.6	6.5	6.9	9.1	9.1
100	RP-BLRT	37.2	34.5	35.3	33.4	31.9	35.4	35.4	36.1	35.1	34.7
	RP-CLRT	5.7	6.2	4.6	5.8	6.3	7.0	5.3	5.4	6.7	6.1
	CLX	4.1	3.8	4.7	6.3	4.9	3.9	2.6	5.1	4.7	4.2
	CZZW	4.6	4.8	5.2	6.6	6.7	5.4	4.4	7.1	5.8	5.3
300	RP-BLRT	15.9	16.8	18.2	16.8	16.3	16.7	17.0	15.8	17.0	17.7
	RP-CLRT	5.4	5.3	6.0	6.0	6.2	4.9	5.6	5.4	4.3	4.2
	CLX	3.9	3.4	5.3	4.3	3.8	4.2	3.3	3.3	4.2	6.6
	CZZW	4.2	4.0	5.7	4.6	3.7	5.6	5.2	4.5	5.0	7.6

Table 4. *Empirical percentage of rejecting H_0 when H_0 is not true.* $\alpha = 0.05$; $k = [7n^{1/3}]$

n	p	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2				
60	RP-BLRT	100	99.6	99.4	98.9	98.4	100	99.9	99.7	99.2	99.2
	RP-CLRT	65.3	23.6	13.9	8.2	6.3	71.6	29.7	10.5	6.8	6.1
	CLX	29.6	19.6	13.5	7.6	6.9	27.9	20.9	11.8	9.3	11.1
	CZZW	33.7	22.0	16.5	9.8	9.2	31.6	23.2	13.4	11.5	14.3
100	RP-BLRT	100	100	99.6	96.6	92.8	100	100	99.3	97.8	95.7
	RP-CLRT	99.7	68.0	39.6	15.4	9.0	99.3	77.0	32.2	18.8	8.6
	CLX	59.9	35.5	19.1	10.4	5.4	63.3	36.5	19.1	16.9	3.9
	CZZW	62.4	39.4	21.1	11.6	6.9	66.0	39.6	21.5	18.6	5.4
300	RP-BLRT	100	100	100	99.8	97.4	100	100	100	99.9	98.1
	RP-CLRT	100	100	100	95.1	61.4	100	100	100	84.3	50.7
	CLX	100	98.4	74.1	33.3	23.9	100	99.0	78.9	34.7	22.1
	CZZW	100	98.5	76.5	34.1	24.6	100	98.9	80.7	36.1	23.1
		Model 3					Model 4				
60	RP-BLRT	100	100	99.4	99.0	99.5	100	99.8	99.8	98.6	98.2
	RP-CLRT	69.8	30.3	14.9	8.3	5.8	77.2	26.2	11.1	7.6	7.3
	CLX	35.4	20.8	12.2	7.7	4.6	35.3	19.2	12.4	7.7	7.8
	CZZW	38.2	24.0	14.4	9.6	6.2	37.6	23.2	15.4	11.1	9.5
100	RP-BLRT	100	100	99.2	96.7	93.2	100	100	99.3	97.5	94.9
	RP-CLRT	99.6	76.5	34.5	18.8	9.3	100	76.9	43.8	15.4	9.6
	CLX	69.0	41.3	19.8	11.4	4.6	68.1	35.5	17.9	10.1	4.4
	CZZW	70.9	43.8	21.9	12.9	5.0	70.2	38.7	19.6	10.7	5.7
300	RP-BLRT	100	100	100	100	97.9	100	100	100	100	97.5
	RP-CLRT	100	100	100	95.5	45.0	100	100	100	93.8	42.8
	CLX	100	98.7	78.2	44.0	14.5	100	99.4	78.1	37.3	21.3
	CZZW	100	98.9	80.2	44.8	17.0	100	99.7	80.2	37.6	22.9

then we set $\Sigma_2 = \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\}^T$. We consider two cases for Σ_1 respectively being the identity matrix of appropriate size, and $\Sigma_1 = 0.8I + 0.2 \times 11^T$.

We set $n = 200$, $p = 400$, $k = [7n^{1/3}]$, and compare the powers with different settings of θ. The results are reported in Figure 1 and Figure 2. Since it is evidential that the BLRT test using (4) and chi-square distribution as the reference distribution does not work well, we do not include corresponding results in our figures.

Both Figures 1 and 2 show promising performance of our test. From Figure 1, we can see that the corrected likelihood ratio test adjusted with the random matrix theory is more powerful than the tests designed for detecting sparse and strong signals of [8] and [10] under all combinations of angle θ and extension $e = (1 + d)^{1/2}$, under the setting $\Sigma_1 = I$. Under the setting $\Sigma_1 = I$, sequentially rotating Σ_1 in multiple subspaces with a small angle θ accompanied by small extension e will result in the signals differing Σ_1 and Σ_2 are dense and weak. From Figure 2, we see same phenomenon when $\theta = 0.01\pi$, 0.03π, accompanied with all pre-specified values of small extension e under the setting $\Sigma_1 = 0.8I + 0.2 \times 11^T$. Since sequential rotation by not so small angle θ accompanied with quite small extension e on matrix with all non-zero off-diagonal entries tend to generate sparse signals, for the case $\theta = 0.05\pi$, 0.07π, the tests of [8] and [10] display its sensitivity to sparse signals, but still show unsatisfactory power. On the other hand, when $\theta = 0.05\pi$, 0.07π, accompanied with not so small extension e, the corrected test adjusted with the random matrix theory has better capability to capture those dense and weak signals generated by getting Σ_2 from sequential rotation by not so small angle θ accompanied with not so small extension e on $\Sigma_1 = 0.8I + 0.2 \times 11^T$.

This setting is also informative in demonstrating the merits of our approach compared with others, including those targeting at aggregating the component-wise discrepancies, for example, [24] and [19]. We compare with [24], whose test is based on $\operatorname{tr}\left(\Sigma_1^2\right)/\operatorname{tr}^2\left(\Sigma_1\right) - \operatorname{tr}\left(\Sigma_2^2\right)/\operatorname{tr}^2\left(\Sigma_2\right)$, and we denote it by "SY". We consider the case for $\Sigma_1 = 0.8I + 0.2 \times 11^T$, and then $\Sigma_2 = \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\}^T$ as well as $\Sigma_2 = \left\{\prod_{g=1}^{p} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p} G\left(i_g, j_g, \theta, d\right)\right\}^T$, of which two settings of $\theta = 0.01\pi, 0.07\pi$ are considered.

We set (n, p) as $(100, 200)$ and $k = n/2$. We compare the powers in two settings: 1) with d ranging in

Table 5. Empirical percentage of rejecting H_0 when H_0 is not true. $\alpha = 0.05$; $k = [6n^{1/3}]$

n	p	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2				
60	RP-BLRT	99.7	97.0	92.4	88.9	85.1	99.7	97.4	92.3	89.6	84.7
	RP-CLRT	61.6	25.5	8.4	7.8	6.6	59.7	18.6	12.1	8.2	6.4
	CLX	29.6	19.6	13.5	7.6	6.9	27.9	20.9	11.8	9.3	11.1
	CZZW	33.7	22.0	16.5	9.8	9.2	31.6	23.2	13.4	11.5	14.3
100	RP-BLRT	100	99.9	94.4	86.5	79.2	100	99.1	94.1	85.5	79.0
	RP-CLRT	98.8	63.4	22.3	17.2	8.7	98.0	66.7	41.8	12.8	12.1
	CLX	59.9	35.5	19.1	10.4	5.4	63.3	36.5	19.1	16.9	3.9
	CZZW	62.4	39.4	21.1	11.6	6.9	66.0	39.6	21.5	18.6	5.4
300	RP-BLRT	100	100	100	99.4	88.7	100	100	100	99.5	88.9
	RP-CLRT	100	100	99.7	85.4	36.6	100	100	99.8	96.3	37.6
	CLX	100	98.4	74.1	33.3	23.9	100	99.0	78.9	34.7	22.1
	CZZW	100	98.5	76.5	34.1	24.6	100	98.9	80.7	36.1	23.1
		Model 3					Model 4				
60	RP-BLRT	99.7	98.2	92.8	87.1	85.8	99.8	98.0	92.2	89.2	85.3
	RP-CLRT	75.7	31.7	17.2	6.3	6.5	67.8	28.0	17.9	6.7	6.7
	CLX	35.4	20.8	12.2	7.7	4.6	35.3	19.2	12.4	7.7	7.8
	CZZW	38.2	24.0	14.4	9.6	6.2	37.6	23.2	15.4	11.1	9.5
100	RP-BLRT	100	99.7	95.5	89.1	78.8	100	99.8	94.5	86.7	77.2
	RP-CLRT	99.7	72.4	26.2	12.5	7.8	99.6	67.8	53.5	11.6	6.1
	CLX	69.0	41.3	19.8	11.4	4.6	68.1	35.5	17.9	10.1	4.4
	CZZW	70.9	43.8	21.9	12.9	5.0	70.2	38.7	19.6	10.7	5.7
300	RP-BLRT	100	100	100	98.9	88.9	100	100	100	99.6	90.3
	RP-CLRT	100	100	99.9	91.2	48.1	100	100	99.9	93.7	65.7
	CLX	100	98.7	78.2	44.0	14.5	100	99.4	78.1	37.3	21.3
	CZZW	100	98.9	80.2	44.8	17.0	100	99.7	80.2	37.6	22.9

Table 6. Empirical percentage of rejecting H_0 when H_0 is not true. $\alpha = 0.05$; $k = [5n^{1/3}]$

n	p	50	100	200	400	800	50	100	200	400	800
		Model 1					Model 2				
60	RP-BLRT	96.7	89.5	78.1	70.4	63.2	98.0	88.7	75.5	71.0	64.2
	RP-CLRT	48.5	18.9	14.3	9.6	5.4	49.3	28.0	11.0	5.9	8.7
	CLX	29.6	19.6	13.5	7.6	6.9	27.9	20.9	11.8	9.3	11.1
	CZZW	33.7	22.0	16.5	9.8	9.2	31.6	23.2	13.4	11.5	14.3
100	RP-BLRT	99.7	94.4	77.8	61.2	50.2	100	93.4	78.2	62.7	51.3
	RP-CLRT	97.1	65.6	17.7	16.2	9.0	95.9	49.3	20.4	10.6	5.1
	CLX	59.9	35.5	19.1	10.4	5.4	63.3	36.5	19.1	16.9	3.9
	CZZW	62.4	39.4	21.1	11.6	6.9	66.0	39.6	21.5	18.6	5.4
300	RP-BLRT	100	100	100	94.5	72.6	100	100	100	94.1	72.5
	RP-CLRT	100	100	99.8	90.5	32.2	100	100	99.5	65.6	63.5
	CLX	100	98.4	74.1	33.3	23.9	100	99.0	78.9	34.7	22.1
	CZZW	100	98.5	76.5	34.1	24.6	100	98.9	80.7	36.1	23.1
		Model 3					Model 4				
60	RP-BLRT	97.9	88.1	78.1	69.7	63.8	98.8	87.4	77.9	67.6	65.2
	RP-CLRT	53.8	29.9	15.6	7.7	8.0	72.4	31.0	10.3	8.5	7.1
	CLX	35.4	20.8	12.2	7.7	4.6	35.3	19.2	12.4	7.7	7.8
	CZZW	38.2	24.0	14.4	9.6	6.2	37.6	23.2	15.4	11.1	9.5
100	RP-BLRT	100	95.0	78.4	61.4	51.1	100	96.2	78.6	61.7	47.6
	RP-CLRT	98.8	62.4	39.7	9.7	6.5	97.6	68.2	43.1	11.5	7.1
	CLX	69.0	41.3	19.8	11.4	4.6	68.1	35.5	17.9	10.1	4.4
	CZZW	70.9	43.8	21.9	12.9	5.0	70.2	38.7	19.6	10.7	5.7
300	RP-BLRT	100	100	99.9	93.9	75.9	100	100	100	96.8	75.9
	RP-CLRT	100	100	99.9	92.8	44.3	100	100	99.6	88.1	46.6
	CLX	100	98.7	78.2	44.0	14.5	100	99.4	78.1	37.3	21.3
	CZZW	100	98.9	80.2	44.8	17.0	100	99.7	80.2	37.6	22.9

$0.03, 0.06, \cdots, 0.57, 0.6$ and $\Sigma_2 = \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\}^T$; 2) with d ranging in $0.07, 0.14, \cdots, 1.26, 1.40$ and $\Sigma_2 = \left\{\prod_{g=1}^{p} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p} G\left(i_g, j_g, \theta, d\right)\right\}^T$. Here a larger d means relatively more discrepancy between the two covariance matrices; and Σ_2 in the 2nd setting differs more substantially from Σ_1 than that in the 1st setting. The results are reported in Figure 3 and Figure 4, respectively. From Figure 3, we can see that our test is becoming more and more powerful when d is increasing, whereas "SY" remains not powerful in this setting. In Figure 4, though "SY" can pick up its power with relatively larger d in this setting with stronger signal, our test reaches 100% very quickly. This demonstrates the advantage of our approach for detecting weak and dense signal, thanks to the powerful nature of the corrected likelihood ratio statistics.

4.4 Impact from different k's

We examine the impact from different number of k's on the power of the tests. For each k, we repeat the simulation 1,000 times, and report the averaged powers and their standard deviations.

We conduct simulations with different k's being $0.05n$, $0.1n, \cdots, 0.45n$ and $0.5n$ for different combinations of (n, p) being $(100, 200)$, $(200, 400)$, and $(300, 600)$. We use the rotational alternative setting of $\Sigma_1 = 0.8I + 0.2 \times 11^T$ and $\Sigma_2 = \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\}^T$ with $\theta = 0.07\pi$ as well as $d = 0.25$, and 0.3.

The results are reported in Table 7. Since all standard deviations in Table 7 are quite small, it can be concluded that the powers are stable for a given k.

We also examine the powers with different k's with Table 7. Collectively, we observe that the powers of our approach is stable for k in a reasonable range. Nonetheless, a remark is that when the sample size is not large enough, and k becomes excessively large, there is some power loss, compared with using smaller k. As demonstrated in our previous examples, our test takes advantage of the likelihood ratio approach in detecting signals violating the null hypothesis. When k becomes large with given n, it may exceed the asymptotic powerful zone of the likelihood ratio test statistic, especially when n is not large enough. It is our experience that as long as k not exceeding $n/2$, the tests work reasonably well; and we recommend checking multiple choices in the range between $n/4$ and $n/2$ for its stability.

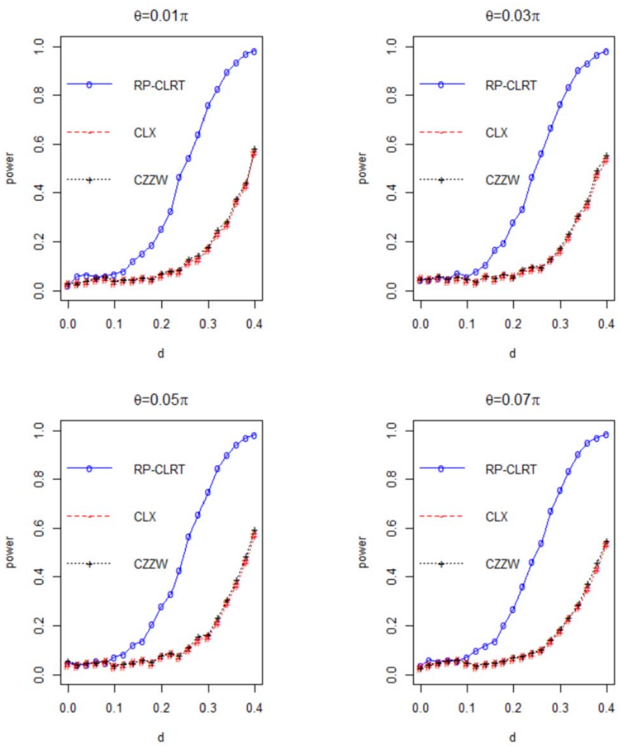

Figure 1. Graphs of divergence among RP-CLRT, CLX, and CZZW when $\Sigma_1 = I$.

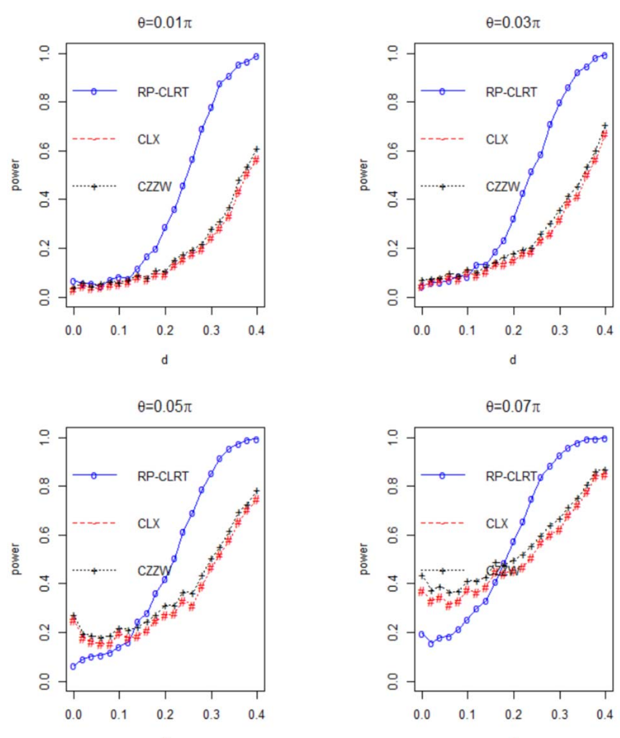

Figure 2. Graphs of divergence among RP-CLRT, CLX, and CZZW when $\Sigma_1 = 0.8I + 0.2 \times 11^T$.

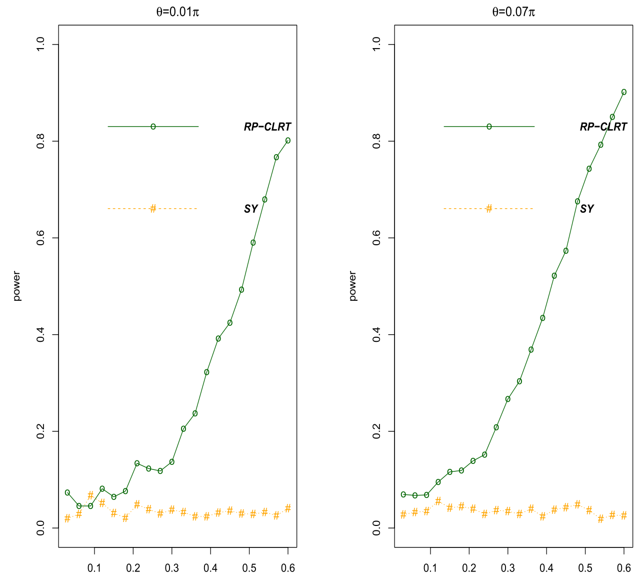

Figure 3. Graphs of divergence between RP-CLRT, and SY when $\Sigma_2 = \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\}^T$.

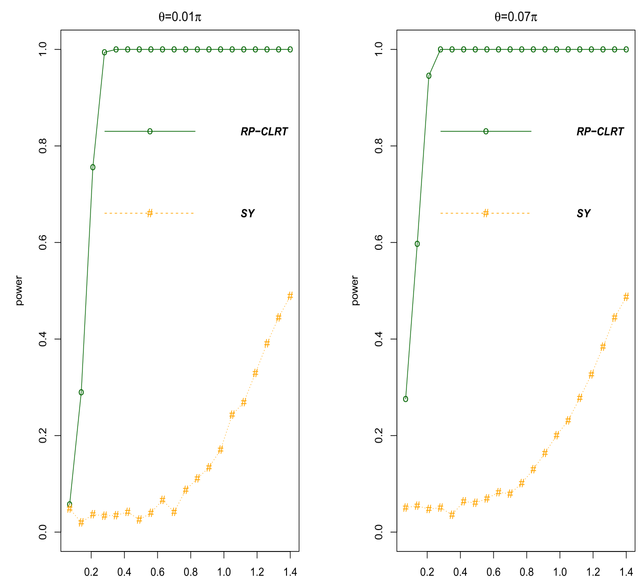

Figure 4. Graphs of divergence between RP-CLRT, and SY when $\Sigma_2 = \left\{\prod_{g=1}^{p} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p} G\left(i_g, j_g, \theta, d\right)\right\}^T$.

4.5 Numerical approximation of the power function

In Theorem 2, we have shown that the power of the new test depends on the matrix $T^{(R)}$. Given the order of k, we could utilize the numerical evaluation, which is introduced

Table 7. Averages ($avg.$) and standard deviations ($std.dev.$) of empirical percentage of rejecting H_0 under rotational alternative of $\Sigma_1 = 0.8I + 0.2 \times 11^T$, $\theta = 0.07\pi$; $\alpha = 0.05$; $k = yn$

(n, p)		(100, 200)									
d	y	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
0.25	avg.	34.5	32.1	31.6	28.8	27.5	27.1	25.3	22.8	20.4	18.9
	std. dev.	2.3	2.4	2.2	2.0	2.1	2.2	2.3	2.0	2.2	2.1
0.3	avg.	45.1	44.4	43.9	42.5	39.4	38.0	35.7	33.1	31.8	29.4
	std. dev.	2.1	2.4	2.3	2.3	2.2	2.4	2.0	2.3	2.4	2.2
(n, p)		(200, 400)									
d	y	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
0.25	avg.	75.3	81.2	79.5	79.0	78.3	74.6	72.8	69.2	64.1	59.6
	std. dev.	2.1	2.0	1.9	2.3	2.1	2.0	2.3	2.1	2.0	1.9
0.3	avg.	91.0	92.9	95.1	93.7	92.1	89.3	88.6	83.5	80.0	76.4
	std. dev.	1.2	1.2	1.1	1.3	1.2	1.2	1.4	1.4	1.9	2.0
(n, p)		(300, 600)									
d	y	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
0.25	avg.	97.9	98.3	98.1	98.4	98.3	97.0	95.6	92.8	91.1	90.0
	std. dev.	0.2	0.3	0.1	0.2	0.3	0.2	0.4	0.4	0.3	0.2
0.3	avg.	99.5	99.9	99.9	99.9	99.7	99.8	99.5	99.3	98.3	98.1
	std. dev.	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

in Section 3.3, of the power of the new test with any random projection matrix $R \in \mathbb{R}^{p\times k}$. Fixing the order of k being $[7n^{1/3}]$, we now demonstrate that the proposed test is able to asymptotically distinguish H_1 from H_0 with random projection matrix R of column size $k = [7n^{1/3}]$.

Denote $U = U^T = \Sigma_2 - \Sigma_1 = (u_{kl})_{1\le k,l\le p}$, and $\{\sigma_{jj}\}_{j=1}^p$ being diagonal entries of Σ_1, we consider testing the following class

$$\begin{aligned}&\mathcal{W}(s, \lambda)\\ &= \left\{u_{kl} \sim s \times W \times \max_{1\le j\le p} \sigma_{jj}, \lambda = \frac{p}{n}, W \sim \mathrm{Unif}(0,1)\right\},\end{aligned}$$

where s denotes the order of the signal and λ denotes the dimension-to-sample size ratio when both n and p tend to be large.

In particular, we numerically approximate the power functions of five random projection matrices of column size k being determined as $[7n^{1/3}]$ for the testing class $\mathcal{W}(0.04, 2)$, where the order of the signal for establishing the consistency of the new test is set as 0.04, and $p = 2n$, a high-dimensional testing scenario, where a set of (n, p) combination is $(100, 200)$, $(150, 300)$, $(200, 400)$, $(300, 600)$, $(400, 800)$, $(600, 1200)$, $(800, 1600)$, $(1200, 2400)$ and $(1600, 3200)$.

The results for Setting 1 are reported in Figure 5. Since the results for all four settings are quite similar, we do not include corresponding results in our figures. From Figure 5, it can be seen that as long as the column size of the random projection matrix R is set as $k = [7n^{1/3}]$, the approximated power functions for given combination of (n, p) and

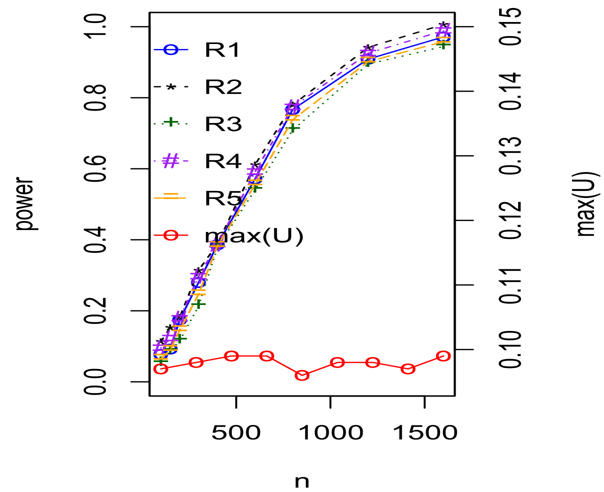

Figure 5. Numerical evaluation of the theoretical power functions when $k = [7n^{1/3}]$ under Setting 1 in $\mathcal{W}(0.04, 2)$.

given random projection matrix R_i of column size $[7n^{1/3}]$, $i = 1, 2, 3, 4, 5$, of the testing class $\mathcal{W}(0.04, 2)$ are quite close to each other. The similarity among approximated theoretical power functions along with Table 7 together deliver the message that for reasonable number of k's, the powers are stable among different random projects.

We observe that as both n and p increase in the testing class $\mathcal{W}(0.04, 2)$, the new test has its power approaching 1. We also report the maximum element-wise difference in Figure 5, which is seen relatively stable even as both n and p getting larger. This means that our test can still achieve promising power with high-dimensional data in a setting with weak maximum discrepancy between Σ_1 and Σ_2, thanks to the aggregation of evidence from high-dimensional data.

4.6 Real data analysis

We apply our test on the "chr1qseg" data in the R package "highD2pop". The data list contain $n_1 = 92$ by $p = 400$ matrix of copy number measurements for 92 long term survivors, and $n_2 = 138$ by $p = 400$ matrix of copy number measurements for 138 short term survivors. Around 3% of the values are missing. We replace missing copy number measurements at a certain location with average copy number measurements at a certain location for long term and short term survivors respectively. We apply the corrected test adjusted with the random matrix theory with $k = [7n^{1/3}]$ where $n = \min(n_1, n_2)$ to test the null hypothesis. The value of test statistic is 8.1194 with p-value being 4.4409e-16, then the hypothesis is rejected. The value of test statistic in [8] is 15.124 with p-value being 0.9989, then it does not reject the hypothesis. The value of test statistic in [10] is 3.889 with p-value being 0.1773, then it also does not reject the hypothesis. It is reasonable to reject the hypothesis since the two covariance matrices are constructed from copy number measurements for short-term and long-term survivors respectively. The hypothesis testing oriented to sparse and relatively strong signals, e.g., "CLX" and "CZZW", cannot detect the difference between two covariance matrices in this specific example. But our proposed corrected test adjusted with the random matrix theory confirms the difference between short-term and long-term survivors, which shows the potential merits of corrected test adjusted with the random matrix theory in aggregating many not so strong signals in contrast to test oriented to sparse signals.

5. DISCUSSION

We propose to apply random projections on high-dimensional random vectors and then testing the equality of their covariance matrices with the random matrix theory based testing procedure. We establish the consistency of our test, and demonstrate that it is more powerful when the alternatives contain weak and dense signals.

As a first step of this methodology, our test is based on a single random projection, which is computationally very efficient. Our aim is attempting to ensure the applicability of the random matrix assisted corrected likelihood ratio tests. Our theoretical and numerical evaluations justify our method. A further possibility is to explore a test procedure with multiple random projections so as to more efficiently utilize the data information; we plan to pursue such a study in a future project.

APPENDIX A. PROOFS

A.1 Proof of Theorem 1

Proof. Assuming that Assumptions 1–4–5 hold, we express X_i as $\mu_1 + \Sigma_1^{1/2} u_i$, and Y_i as $\mu_2 + \Sigma_2^{1/2} v_i$. Conditioning on R, $R^T u_i,\ i = 1, \cdots, n_1$, and $R^T v_i,\ i = 1, \cdots, n_2$, both independently follows $N(0, I)$ subject to $R^T R = I_k$, which validates assumption (a)–(a') in [27] with the case $\beta_x = \beta_y = 0$; Assumptions (b)–(c) and assumptions (b')–(c') have also been assumed explicitly here. Then Theorem 4.2 in [27] guarantees Theorem 1. □

A.2 Proof of Theorem 2

Proof. Assuming that Assumptions 1–4–6 hold, we express X_i as $\mu_1 + \Sigma_1^{1/2} u_i$, and Y_i as $\mu_2 + \Sigma_2^{1/2} v_i$. Conditioning on R, $R^T u_i,\ i = 1, \cdots, n_1$ and $R^T v_i,\ i = 1, \cdots, n_2$ both independently follow $N(0, I)$ subject to $R^T R = I_k$, which validates assumption (a)–$(b1)$–$(b2)$ in [28] with the case $\beta_x = \beta_y = 0$; Assumption (c) in [28] has also been assumed explicitly here; Conditioning on R, assumption (d) in [28] has also been assumed explicitly here. Then Theorem 3.1 in [28] guarantees Theorem 2. □

ACKNOWLEDGEMENTS

We thank the two reviewers for their insightful comments on our paper, which provide valuable inputs into the improvement of this paper.

Received 1 May 2021

REFERENCES

[1] Achlioptas, D. (2003). Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences* **66** 671–687. MR2005771

[2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley. MR1990662

[3] Bai, Z. and Silverstein, J. W. (2009). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer New York. MR2567175

[4] 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** 3822–3840. MR2572444

[5] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A – Mathematical and Physical Sciences* **160** 268–282.

[6] Bingham, E. and Mannila, H. (2001). Random projection in dimensionality reduction. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Press.

[7] Cai, T. T. (2017). Global Testing and Large-Scale Multiple Testing for High-Dimensional Covariance Structures. *Annual Review of Statistics and Its Application* **4** 423–446.

[8] Cai, T., Liu, W. and Xia, Y. (2013). Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings. *Journal of the American Statistical Association* **108** 265–277. MR3174618

[9] Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557

[10] Chang, J., Zhou, W., Zhou, W.-X. and Wang, L. (2016). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73** 31–41. MR3632349

[11] Chang, J., Qiu, Y., Yao, Q. and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics* **206** 57–82. MR3840783

[12] Chen, S. X., Zhang, L.-X. and Zhong, P.-S. (2010). Tests for High-Dimensional Covariance Matrices. *Journal of the American Statistical Association* **105** 810–819. MR2724863

[13] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41** 2786–2819. MR3161448

[14] Dobriban, E. (2015). Efficient computation of limit spectra of sample covariance matrices. *Random Matrices: Theory and Applications* **4** 1550019. MR3418848

[15] He, J. and Chen, S. X. (2016). Testing super-diagonal structure in high dimensional covariance matrices. *Journal of Econometrics* **194** 283–297. MR3536981

[16] Hu, J. and Bai, Z. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Science China Mathematics* **59** 2281–2300. MR3578957

[17] 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** 2241–2256. MR2911842

[18] Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics* **40** 908–940. MR2985938

[19] Li, W. and Qin, Y. (2014). Hypothesis testing for high-dimensional covariance matrices. *J. Multivariate Anal.* **128** 108–119. MR3199831

[20] Lin, R., Liu, Z., Zheng, S. and Yin, G. (2016). Power computation for hypothesis testing with high-dimensional covariance matrices. *Computational Statistics & Data Analysis* **104** 10–23. MR3540983

[21] Lopes, M., Jacob, L. and Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems* 1206–1214.

[22] Qiu, Y. and Chen, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics* **40** 1285–1314. MR3015026

[23] Srivastava, R., Li, P. and Ruppert, D. (2016). RAPTT: An Exact Two-Sample Test in High Dimensions Using Random Projections. *Journal of Computational and Graphical Statistics* **25** 954–970. MR3533647

[24] Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *J. Multivariate Anal.* **101** 1319–1329. MR2609494

[25] Vempala, S. S. (2005). *The Random Projection Method (Dimacs Series in Discrete Math)*. American Mathematical Society. MR2073630

[26] Wang, C. (2014). Asymptotic power of likelihood ratio tests for high dimensional data. *Statist. Probab. Lett.* **88** 184–189. MR3178349

[27] Zheng, S., Bai, Z. and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics* **43** 546–591. MR3316190

[28] Zheng, S., Bai, Z. and Yao, J. (2017). CLT for eigenvalue statistics of large-dimensional general Fisher matrices with applications. *Bernoulli* **23** 1130–1178. MR3606762

Nan Sun
1626, Newport Place, Philadelphia, PA, 19122
USA
E-mail address: snofnckustat@gmail.com

Cheng Yong Tang
Room 376, Building 1810 Liacouras Walk, Philadelphia, PA, 19122-6083
USA
E-mail address: yongtang@temple.edu