A Test of Sphericity for High-Dimensional Data and Its Application for Detection of Divergently Spiked Noise

Kazuyoshi Yata[a], Makoto Aoshima[a], and Yugo Nakayama[b]

[a] Institute of Mathematics, University of Tsukuba, Ibaraki, Japan
[b] Graduate School of Pure and Applied Sciences, University of Tsukuba, Ibaraki, Japan

Abstract: In this paper, we consider a test of the sphericity for high-dimensional covariance matrices. We produce a test statistic by using the extended cross-data-matrix (ECDM) methodology. We show that the ECDM test statistic is based on an unbiased estimator of a sphericity measure. In addition, the ECDM test statistic enjoys consistency properties and the asymptotic normality in high-dimensional settings. We propose a new test procedure based on the ECDM test statistic and evaluate its asymptotic size and power theoretically and numerically. We give a two-stage sampling scheme so that the test procedure can ensure a prespecified level both for the size and power. We apply the test procedure to detect divergently spiked noise in high-dimensional statistical analysis. We analyze gene expression data by the proposed test procedure.

Keywords: Cross-data-matrix method; Gene expression data; HDLSS; Noise detection; Noise-reduction method; Sphericity.

1. INTRODUCTION

High-dimension, low-sample-size (HDLSS) data situations occur in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. In the HDLSS situations, the large sample theory does not work and hence some new theories and methodologies are expected to develop for high-dimensional inferences. Aoshima and Yata (2011a,b) is pioneering work which established a variety of inference for HDLSS data such as given-bandwidth confidence regions, two-sample tests, tests of the equality of two covariance matrices, classification, variable selection, regression, tests of the correlation coefficients and so on, and discussed sample size determination to ensure prespecified accuracy for each inference. Afterward, those high-dimensional inferences have been further studied and developed by many researchers in the field of high-dimensional statistical analysis. In the current paper, we consider a test of the sphericity for high-dimensional covariance matrices.

Suppose we take samples, x_j, $j = 1, ..., n$, of size n (≥ 4) from a population, which are independent and identically distributed (i.i.d.) as a p-variate distribution. We assume that x_j has an unknown mean vector μ and unknown covariance matrix Σ. We denote the eigenvalue decomposition of Σ by $\Sigma = H\Lambda H^T$, where Λ is a diagonal matrix of eigenvalues, $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, and H is an orthogonal matrix of the corresponding eigenvectors. Let $x_j = H\Lambda^{1/2}z_j + \mu$, where $z_j = (z_{1j}, ..., z_{pj})^T$ is considered as a sphered

---

data vector having the zero mean vector and identity covariance matrix. Let $\sigma = \text{tr}(\boldsymbol{\Sigma})/p$. We assume that $\sigma \in (0, \infty)$ as $p \to \infty$. For a function, $f(\cdot)$, "$f(p) \in (0, \infty)$ as $p \to \infty$" implies that $\liminf_{p \to \infty} f(p) > 0$ and $\limsup_{p \to \infty} f(p) < \infty$. We consider the following model:

$$\boldsymbol{x}_j = \boldsymbol{\Gamma}\boldsymbol{w}_j + \boldsymbol{\mu}, \qquad (1.1)$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_q)$ is a $p \times q$ matrix for some $q > 0$ such that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \boldsymbol{\Sigma}$, and $\boldsymbol{w}_j = (w_{1j}, ..., w_{qj})^T$, $j = 1, \ldots, n$, are i.i.d. random vectors having $E(\boldsymbol{w}_j) = \boldsymbol{0}$ and $\text{Var}(\boldsymbol{w}_j) = \boldsymbol{I}_q$. Here, $\boldsymbol{I}_q$ denotes the identity matrix of dimension q. Let $\text{Var}(w_{rj}^2) = M_r$, $r = 1, \ldots, q$. We assume that $M_r \in (0, \infty)$ as $p \to \infty$ for all r. Similar to Bai and Saranadasa (1996) and Aoshima and Yata (2015), we assume that

(A-i) $E(w_{rj}^2 w_{sj}^2) = E(w_{rj}^2)E(w_{sj}^2) = 1$ and $E(w_{rj}w_{sj}w_{tj}w_{uj}) = 0$ for all $r \neq s, t, u$.

We assume the following assumption instead of (A-i) as necessary:

(A-ii) $E(w_{r_1j}^{\alpha_1} w_{r_2j}^{\alpha_2} \cdots w_{r_vj}^{\alpha_v}) = E(w_{r_1j}^{\alpha_1})E(w_{r_2j}^{\alpha_2}) \cdots E(w_{r_vj}^{\alpha_v})$ for all $r_1 \neq r_2 \neq \cdots \neq r_v \in [1, q]$ and $\alpha_i \in [1, 4]$, $i = 1, \ldots, v$, where $v \leq 8$ and $\sum_{i=1}^v \alpha_i \leq 8$.

See Chen and Qin (2010) about (A-ii). Note that (A-ii) implies (A-i). When $\boldsymbol{x}_j$ is Gaussian, it holds that $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{w}_j = \boldsymbol{z}_j$ in (1.1). Note that (A-ii) is naturally satisfied when $\boldsymbol{x}_j$ is Gaussian because the elements of $\boldsymbol{z}_j$ are independent and $M_r = 2$ for all r. We assume the following HDLSS divergence condition:

(A-iii) $p, n \to \infty$ and $n/p \to 0$.

In this paper, we are interested in testing the sphericity of $\boldsymbol{\Sigma}$:

$$H_0 : \boldsymbol{\Sigma} = \sigma \boldsymbol{I}_p \quad \text{vs.} \quad H_1 : \boldsymbol{\Sigma} \neq \sigma \boldsymbol{I}_p. \qquad (1.2)$$

We give a two-stage test procedure which can ensure a prespecified level both for the size and power. Most interestingly, we apply the test procedure to detect divergently spiked noise in high-dimensional statistical analysis.

When $n > p$ and p is fixed, Nagao (1973) and others gave test statistics for (1.2) by using the large sample theory. Ledoit and Wolf (2002) investigated asymptotic properties of the test statistics when $p/n \to c > 0$. Since the conventional test statistics do not work for HDLSS data, Srivastava et al. (2011) gave a test statistic under (A-iii). However, the test statistic is heavily biased for high-dimensional data unless $\boldsymbol{x}_j$ is Gaussian. On the other hand, Chen et al. (2010) gave a test statistic based on the U-statistic for high-dimensional data. In the current paper, we shall also pursue a non-parametric approach, but we produce a new test statistic by using the *extended cross-data-matrix (ECDM) methodology*. The ECDM method was developed by Yata and Aoshima (2013) and was motivated by the cross-data-matrix (CDM) method due to Yata and Aoshima (2010). One of the advantages of the ECDM method is that one can produce an unbiased estimator having a small variance at a low computational cost even for ultra high-dimensional data. In addition, the ECDM method possesses a high versatility in high-dimensional data analysis. See Yata and Aoshima (2016) for the details.

The paper is organized as follows: In Section 2, we produce a test statistic for (1.2) by using the ECDM method. We show that the ECDM test statistic is based on an unbiased estimator of a sphericity measure. In addition, the ECDM test statistic enjoys consistency properties and the asymptotic normality in high-dimensional settings. In Section 3, we propose a new test procedure based on the ECDM test statistic and evaluate its asymptotic size and power theoretically. In Section 4, we give a two-stage sampling scheme so that the test procedure can ensure a prespecified level both for the size and power. In Section 5, we apply the test procedure to detect divergently spiked noise in high-dimensional statistical analysis. In Section 6, we give simulation studies to investigate the performance of the proposed test procedure. Finally, in Section 7, we analyze gene expression data by the proposed test procedure.

2. UNBIASED ESTIMATOR OF SPHERICITY MEASURE

Let $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma} - \sigma \boldsymbol{I}_p$ and $\Delta = ||\boldsymbol{\Sigma}_*||_F^2 = \text{tr}(\boldsymbol{\Sigma}^2) - \sigma^2 p$, where $||\cdot||_F$ is the Frobenius norm. Note that $\Delta = 0$ under H_0 in (1.2) and Δ is regarded as a sphericity measure. See Ahn et al. (2007) for the sphericity measure. In this section, we give an unbiased estimator of Δ by using the extended cross-data-matrix (ECDM) methodology.

The ECDM methodology was developed by Yata and Aoshima (2013) as an extension of the CDM method due to Yata and Aoshima (2010). One of the advantages of the ECDM method is that one can produce an unbiased estimator having a small variance at a low computational cost even for ultra high-dimensional data. See Section 2.5 of Yata and Aoshima (2013) for the details. Let $n_{(1)} = \lceil n/2 \rceil$ and $n_{(2)} = n - n_{(1)}$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Let

$$\boldsymbol{V}_{n(1)(k)} = \begin{cases} \{\lfloor k/2 \rfloor - n_{(1)} + 1, \ldots, \lfloor k/2 \rfloor\} & \text{if } \lfloor k/2 \rfloor \geq n_{(1)}, \\ \{1, \ldots, \lfloor k/2 \rfloor\} \cup \{\lfloor k/2 \rfloor + n_{(2)} + 1, \ldots, n\} & \text{otherwise;} \end{cases}$$

$$\boldsymbol{V}_{n(2)(k)} = \begin{cases} \{\lfloor k/2 \rfloor + 1, \ldots, \lfloor k/2 \rfloor + n_{(2)}\} & \text{if } \lfloor k/2 \rfloor \leq n_{(1)}, \\ \{1, \ldots, \lfloor k/2 \rfloor - n_{(1)}\} \cup \{\lfloor k/2 \rfloor + 1, \ldots, n\} & \text{otherwise} \end{cases}$$

for $k = 3, \ldots, 2n-1$, where $\lfloor x \rfloor$ denotes the largest integer $\leq x$. Let $\#\boldsymbol{S}$ denote the number of elements in a set $\boldsymbol{S}$. Note that $\#\boldsymbol{V}_{n(l)(k)} = n_{(l)}, l = 1, 2, \boldsymbol{V}_{n(1)(k)} \cap \boldsymbol{V}_{n(2)(k)} = \emptyset$ and $\boldsymbol{V}_{n(1)(k)} \cup \boldsymbol{V}_{n(2)(k)} = \{1, \ldots, n\}$ for $k = 3, \ldots, 2n-1$. Also, note that

$$i \in \boldsymbol{V}_{n(1)(i+j)} \text{ and } j \in \boldsymbol{V}_{n(2)(i+j)} \text{ for } i < j \ (\leq n). \tag{2.1}$$

See Figure 1.

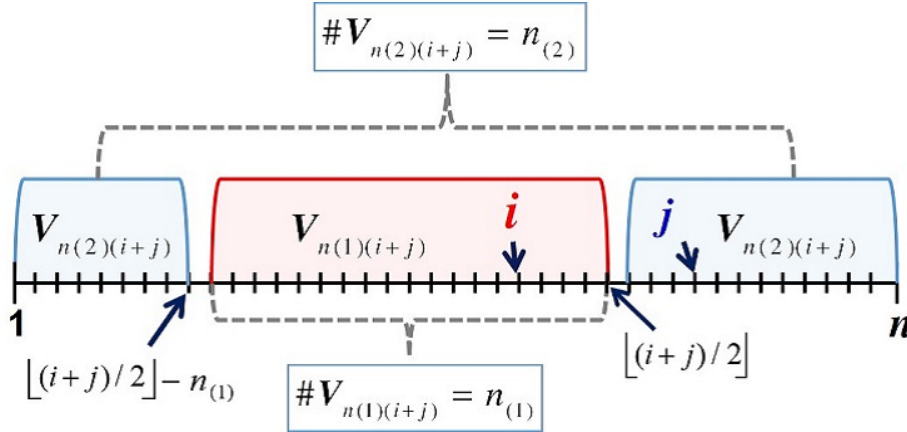**Figure 1.** Illustration of (2.1) when $\lfloor (i+j)/2 \rfloor > n_{(1)}$.

Let

$$\overline{\boldsymbol{x}}_{(1)(k)} = n_{(1)}^{-1} \sum_{j \in \boldsymbol{V}_{n(1)(k)}} \boldsymbol{x}_j \quad \text{and} \quad \overline{\boldsymbol{x}}_{(2)(k)} = n_{(2)}^{-1} \sum_{j \in \boldsymbol{V}_{n(2)(k)}} \boldsymbol{x}_j$$

for $k = 3, \ldots, 2n-1$. From (2.1), we note that $(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(1)(i+j)})$ and $(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{(2)(i+j)})$ are independent for all $i < j$. Then, Yata and Aoshima (2013) gave an estimator of $\text{tr}(\boldsymbol{\Sigma}^2)$ by the ECDM method as

$$W_n = \frac{2u_n}{n(n-1)} \sum_{i<j}^{n} \left\{ (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(1)(i+j)})^T (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{(2)(i+j)}) \right\}^2, \tag{2.2}$$

3

where $u_n = n_{(1)}n_{(2)}/\{(n_{(1)} - 1)(n_{(2)} - 1)\}$. Note that $E(W_n) = tr(\boldsymbol{\Sigma}^2)$. Aoshima and Yata (2015) and Yata and Aoshima (2016) gave the following result.

Lemma 2.1 (Aoshima and Yata (2015); Yata and Aoshima (2016)). *Assume (A-i). Then, it holds that as* $p, n \to \infty$

$$Var\Big(\frac{W_n}{tr(\boldsymbol{\Sigma}^2)}\Big) = \frac{4}{n^2}\{1 + o(1)\} + O\Big(\frac{tr(\boldsymbol{\Sigma}^4)}{tr(\boldsymbol{\Sigma}^2)^2 n}\Big) \to 0.$$

Also, we can give an estimator of $\sigma^2 p$ by the ECDM method as

$$U_n = \frac{2u_n}{pn(n-1)} \sum_{i<j}^{n} ||\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)}||^2 \, ||\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)}||^2, \tag{2.3}$$

where $|| \cdot ||$ denotes the Euclidean norm. Note that $E(U_n) = tr(\boldsymbol{\Sigma})^2/p = \sigma^2 p$. We have the following result.

Lemma 2.2. *Assume (A-i). Then, it holds that as* $p, n \to \infty$

$$Var\Big(\frac{U_n}{\sigma^2 p}\Big) = O\Big(\frac{tr(\boldsymbol{\Sigma}^2)}{tr(\boldsymbol{\Sigma})^2 n}\Big) \to 0.$$

Finally, we construct an estimator of Δ by the ECDM method as

$$T_n = W_n - U_n. \tag{2.4}$$

We note that $E(T_n) = \Delta$ without any assumptions. We have the following result.

Lemma 2.3. *Assume (A-i). Then, it holds that as* $p, n \to \infty$

$$Var(T_n) = 4\frac{tr(\boldsymbol{\Sigma}^2)^2}{n^2}\{1 + o(1)\} + O\Big(\frac{tr(\boldsymbol{\Sigma}^4)}{n^2} + \frac{tr\{(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_*)^2\}}{n}\Big).$$

3. NEW TEST PROCEDURE FOR (1.2)

For T_n given by (2.4), we have the following results.

Lemma 3.1. *Assume (A-i) and*

(A-iv) $\dfrac{tr(\boldsymbol{\Sigma}^2)}{n\Delta} \to 0$ *under (A-iii).*

Then, it holds that under (A-iii)

$$\frac{T_n}{\Delta} = 1 + o_P(1).$$

Lemma 3.2. *Assume (A-ii) and*

(A-v) $\limsup\Big\{\dfrac{n\Delta}{tr(\boldsymbol{\Sigma}^2)}\Big\} < \infty$ *under (A-iii).*

Then, it holds that under (A-iii)

$$\frac{T_n - \Delta}{2tr(\boldsymbol{\Sigma}^2)/n} \Rightarrow N(0,1),$$

where "$\Rightarrow$" denotes the convergence in distribution and $N(0,1)$ denotes a random variable distributed as the standard normal distribution.

4

Note that $\text{tr}(\boldsymbol{\Sigma}^2) = \sigma^2 p$ under H_0 in (1.2). From Lemma 3.2 we propose a test procedure for (1.2) by

$$\text{rejecting } H_0 \iff \frac{nT_n}{2U_n} > z_\alpha, \tag{3.1}$$

where z_α is a constant such that $P\{N(0,1) > z_\alpha\} = \alpha$ with $\alpha \in (0, 1/2)$. Then, we have the following result.

Theorem 3.1. *Assume (A-ii) and (A-v). For the test by (3.1), we have that under (A-iii)*

$$\textit{Size} = \alpha + o(1) \quad \textit{and} \quad \textit{Power} = \Phi\Big(\frac{n\Delta}{2tr(\boldsymbol{\Sigma}^2)} - z_\alpha\Big) + o(1), \tag{3.2}$$

where $\Phi(\cdot)$ denotes the c.d.f. of $N(0,1)$.

When (A-iv) is met, we have the following result.

Corollary 3.1. *Assume (A-i). Assume (A-iv) under H_1. For the test by (3.1), we have that under (A-iii)*

$$\textit{Power} = 1 + o(1).$$

Remark 3.1. Chen et al. (2010) gave a test procedure for (1.2) based on the following statistic:

$$T_{\text{CZZ}} = A_n - \text{tr}(\boldsymbol{S}_n)^2/p,$$

where $\boldsymbol{S}_n$ is the sample covariance matrix having $E(\boldsymbol{S}_n) = \boldsymbol{\Sigma}$, and

$$A_n = \frac{1}{n(n-1)} \sum_{j \neq j'}^{n} (\boldsymbol{x}_j^T \boldsymbol{x}_{j'})^2 - \frac{2}{n(n-1)(n-2)} \sum_{j \neq j' \neq j''}^{n} \boldsymbol{x}_{j'}^T \boldsymbol{x}_j \boldsymbol{x}_j^T \boldsymbol{x}_{j''}$$

$$+ \frac{1}{n(n-1)(n-2)(n-3)} \sum_{j \neq j' \neq l \neq l'}^{n_i} \boldsymbol{x}_j^T \boldsymbol{x}_{j'} \boldsymbol{x}_l^T \boldsymbol{x}_{l'}.$$

Note that $E(A_n) = \text{tr}(\boldsymbol{\Sigma}^2)$. However, T_{CZZ} is biased for high-dimensional data because $E\{\text{tr}(\boldsymbol{S}_n)^2\} > \text{tr}(\boldsymbol{\Sigma})^2$. Although the test by Chen et al. (2010) is asymptotically equivalent to the test by (3.1), the latter is much more applicable to the sequential analysis ensuring prespecified accuracy as seen in the next section.

4. TWO-STAGE SAMPLING SCHEME TO CONTROL BOTH SIZE AND POWER

We are interested in designing a test of (1.2) having size α and power no less than $1 - \beta$ when $\Delta \geq \Delta_L$, where $\alpha \in (0, 1/2)$, $\beta \in (0, 1/2)$ and $\Delta_L \, (> 0)$ are prespecified constants. We assume that $\Delta_L \to \infty$ and $\Delta_L = o(p)$ as $p \to \infty$.

From Theorem 3.1 we consider n satisfying

$$\frac{n\Delta}{2\text{tr}(\boldsymbol{\Sigma}^2)} - z_\alpha \geq z_\beta \quad \text{when } \Delta \geq \Delta_L.$$

Then, one finds the sample size as

$$n \geq \frac{2(z_\alpha + z_\beta)\text{tr}(\boldsymbol{\Sigma}^2)}{\Delta_L} \quad (= C, \text{ say}). \tag{4.1}$$

We note that $C \to \infty$ as $p \to \infty$ from the facts that $\text{tr}(\boldsymbol{\Sigma}^2) \geq \sigma^2 p$ and $\Delta_L = o(p)$ as $p \to \infty$. Also, note that $C/p \to 0$ as $p \to \infty$ under H_0 in (1.2) from the fact that $\Delta_L \to \infty$ as $p \to \infty$. Then, from Theorem 3.1, we have the following result.

Theorem 4.1. *Assume (A-ii) and (A-v). For the test by (3.1) with $n \geq C$ given by (4.1), we have under (A-iii)*

$$Size = \alpha + o(1), \quad and \quad Power \geq 1 - \beta + o(1) \quad when \ \Delta \geq \Delta_L. \tag{4.2}$$

Since C includes unknown parameter $\mathrm{tr}(\boldsymbol{\Sigma}^2)$, it is necessary to estimate C with some pilot samples. We proceed with the following two steps:

1. Choose $m(\geq 4)$ satisfying

$$\frac{m}{C} \leq 1, \quad \frac{C}{m^2} \to 0 \ \text{and} \ \frac{C}{m} \frac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2} \to 0 \ \text{as} \ p \to \infty \ \text{under} \ \frac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2} \to 0 \ \text{as} \ p \to \infty. \tag{4.3}$$

Take pilot samples, $\boldsymbol{x}_j, \ j = 1, ..., m$, of size m. Then, calculate W_m according to (2.2). Define the total sample size by

$$N = \max \left\{ m, \left\lceil \frac{2(z_\alpha + z_\beta) W_m}{\Delta_L} \right\rceil \right\}. \tag{4.4}$$

2. If $N = m$, do not take any additional samples and otherwise, that is if $N > m$, take additional samples, $\boldsymbol{x}_j, \ j = m+1, ..., N$, of size $N - m$. By combining the pilot samples and the additional samples, calculate U_N and T_N according to (2.3) and (2.4). Then, we propose a test procedure for (1.2) by

$$\text{rejecting } H_0 \iff \frac{NT_N}{2U_N} > z_\alpha. \tag{4.5}$$

We have the following result.

Theorem 4.2. *Assume (A-ii). Assume also*

(A-vi) $\limsup\limits_{p \to \infty} \left\{ \dfrac{C\Delta}{tr(\boldsymbol{\Sigma}^2)} \right\} < \infty.$

For the test by (4.5), we have (4.2) as $p \to \infty$.

Remark 4.1. Under (A-vi), the condition "$\mathrm{tr}(\boldsymbol{\Sigma}^4)/\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 \to 0$ as $p \to \infty$" in (4.3) holds. See (A.5) in Appendix. From Lemma 2.1, under (A-i) and (4.3), we have that $W_m = \mathrm{tr}(\boldsymbol{\Sigma}^2)\{1 + o_P(C^{-1/2})\}$ as $p \to \infty$. Then, it holds that $N - C = o_P(C^{1/2})$ as $p \to \infty$.

5. DETECTION OF DIVERGENTLY SPIKED NOISE

In this section, we consider the detection of divergently spiked noise as an application of the sphericity test. Paul (2007) and Johnstone and Lu (2009) handled the following multicomponent covariance model:

$$\boldsymbol{x}_j = \boldsymbol{\mu} + \sum_{i=1}^{k} \boldsymbol{\rho}_i \xi_{ij} + \tau^{1/2} \boldsymbol{\varepsilon}_j \ \text{ for } j = 1, ..., n, \tag{5.1}$$

where $\tau \in (0, \infty)$ as $p \to \infty$, ξ_{ij}s are i.i.d. as $N(0, 1)$, $\boldsymbol{\varepsilon}_j$s are i.i.d. as $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, and ξ_{ij}s and $\boldsymbol{\varepsilon}_j$s are mutually independent. Here, k is a fixed positive integer (not depending on p) and $\boldsymbol{\rho}_i$s are mutually orthogonal with

$$||\boldsymbol{\rho}_1||^2 \geq \cdots \geq ||\boldsymbol{\rho}_k||^2 > 0.$$

Note that (A-ii) is met under (5.1). We have that $\boldsymbol{\Sigma} = \sum_{i=1}^{k} \boldsymbol{\rho}_i \boldsymbol{\rho}_i^T + \tau \boldsymbol{I}_p$ and

$$\lambda_j = ||\boldsymbol{\rho}_j||^2 + \tau \ \text{ for } j = 1, ..., k, \ \text{ and } \lambda_{k+1} = \cdots = \lambda_p = \tau. \tag{5.2}$$

In (5.2), the first k eigenvalues are spiked compared to the remaining. Johnstone (2001), Baik and Silverstein (2006), Paul (2007), and Johnstone and Lu (2009) considered the following spiked model:

$$\lambda_j \ (> \tau) \text{ is fixed (not depending on } p) \text{ for } j = 1, ..., k. \tag{5.3}$$

They studied asymptotic behaviors of the conventional principal component analysis (PCA) when $n/p \to c > 0$ under (5.3). However, high-dimensional eigenvalues naturally depend on p and it is probable that $\lambda_j \to \infty$ as $p \to \infty$ for the first several js. See Jung and Marron (2009), Yata and Aoshima (2009), Fan et al. (2013), Ishii et al. (2016), Shen et al. (2016), and Aoshima and Yata (2018) for the details. They considered the following spiked model in which the first k eigenvalues are divergently spiked:

$$\lambda_j = p^{\alpha_j} \quad \text{for } j = 1, ..., k. \tag{5.4}$$

Here, α_js are fixed positive constants (not depending on p) preserving the order that $\lambda_1 \geq \cdots \geq \lambda_k$. For such divergently spiked models, Yata and Aoshima (2010, 2012) developed new PCA methods. They showed that the new PCAs can enjoy consistency properties both for the eigenvalues and PC directions when $\lambda_j \to \infty$ as $p \to \infty$.

One would be interested in testing the following hypotheses:

$$H_0 : (5.3) \text{ holds} \quad \text{vs.} \quad H_1 : (5.4) \text{ holds}. \tag{5.5}$$

From (5.2) we have that

$$\text{tr}(\boldsymbol{\Sigma}^2) = (p - k)\tau^2 + \sum_{j=1}^{k} \lambda_j^2 \ \text{ and } \ \sigma = \frac{(p-k)\tau}{p} + \frac{\sum_{j=1}^{k} \lambda_j}{p}.$$

Under (5.3), we have that $\Delta = O(1)$ as $p \to \infty$, so that from Lemma 3.2 it holds that

$$\frac{T_n}{2\text{tr}(\boldsymbol{\Sigma}^2)/n} \Rightarrow N(0, 1)$$

under (A-iii) since $n\Delta/\text{tr}(\boldsymbol{\Sigma}^2) = O(n/p) \to 0$. Under (5.4), we have that

$$\Delta = \sum_{j=1}^{k} p^{2\alpha_j}\{1 + o(1)\} \to \infty \ \text{ as } p \to \infty.$$

Thus, for the test of (5.5), one can apply the test procedure (3.1) or (4.5).

Corollary 5.1. *The test procedure (3.1) for (5.5) has (3.2) under (A-iii) and (A-v).*

Corollary 5.2. *The test procedure (4.5) for (5.5) has (4.2) as $p \to \infty$ under (A-vi).*

We note that

$$\Delta \geq p^{2\alpha_1}\{1 + o(1)\} \tag{5.6}$$

as $p \to \infty$ under (5.4). Thus we can consider a lower bound of Δ as $\Delta > p^{2\alpha_1}$. Then, one may set $\Delta_L = p^{2\alpha_{1L}}$ with a prespecified constant $\alpha_{1L} \in (0, 1/2)$ in view of the assumptions that $\Delta_L \to \infty$ and $\Delta_L = o(p)$ as $p \to \infty$.

6. SIMULATION STUDIES

In order to investigate the performance of the test procedure (4.5) for (1.2) or (5.5), we used computer simulations.

7

6.1. Test Procedure (4.5) for (1.2)

We set $\Delta_L = p^{2/3}$. We considered constructing a test having size $\alpha = 0.01$ and power no less than $1 - \beta = 0.8$ when $\Delta \geq \Delta_L$. We set $p = 250,\ 500,\ 1000,\ 2000$ and 4000. We put $p_1 = \lceil (6/5)p^{2/3} \rceil$. The following two cases were considered:

$$(a)\ \boldsymbol{\Sigma} = \boldsymbol{I}_p \quad \text{and} \quad (b)\ \boldsymbol{\Sigma} = \boldsymbol{I}_p + \boldsymbol{G},$$

where $\boldsymbol{G} = \mathrm{diag}(1, ..., 1, 0, ..., 0)$ whose first p_1 elements are 1. Note that $\Delta = 3p_1 + p - (p_1 + p)^2/p = p_1 - p_1^2/p \geq \Delta_L$ when p is large for (b). Also, note that (A-vi) is met both for (a) and (b). We considered a non-Gaussian case by setting $q = p$, $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $w_{rj} = (v_{rj} - 5)/10^{1/2}$ in (1.1), where v_{rj}s are i.i.d. as the chi-squared distribution with 5 degrees of freedom. Note that (A-ii) is met. We set $m = \lceil C/2 \rceil$. Note that (4.3) is met both for (a) and (b).

In Tables 1 and 2, we summarized the findings obtained by averaging the outcomes from 2000 $(= 2R$, say) replications, where the first 1000 replications were generated for (a) and the last 1000 replications were generated for (b). Under a fixed scenario, suppose that the rth replication ends with $N = N_r$ observations given by (4.4) and the test result given by (4.5). We defined $P_r = 1$ (or 0) accordingly as H_0 was falsely rejected (or not) and H_1 was falsely rejected (or not). We defined $\overline{\alpha} = R^{-1}\sum_{r=1}^{R} P_r$ to estimate the size and $1 - \overline{\beta} = 1 - R^{-1}\sum_{r=R+1}^{2R} P_r$ to estimate the power when $\Delta \geq \Delta_L$, while their estimated standard errors, $s(\overline{\alpha})$ and $s(\overline{\beta})$, were given by $s^2(\overline{\alpha}) = R^{-1}\overline{\alpha}(1 - \overline{\alpha})$ and $s^2(\overline{\beta}) = R^{-1}\overline{\beta}(1 - \overline{\beta})$. For (a), we calculated $\overline{N} = R^{-1}\sum_{r=1}^{R} N_r$ and $\mathrm{MSE}(N/C) = R^{-1}\sum_{r=1}^{R}(N_r/C - 1)^2$. Similarly, we calculated $\overline{N}$ and $\mathrm{MSE}(N/C)$ for (b).

Table 1. Required sample size and average size by the test procedure (4.5) for (1.2) in case of (a).

p	m	C	$\overline{N}$	$\overline{N} - C$	$\mathrm{MSE}(N/C)$	$\overline{\alpha}$	$s(\overline{\alpha})$
250	20	39.91	39.91	-0.01	0.0173	0.017	0.0041
500	26	50.29	50.63	0.34	0.0094	0.014	0.0037
1000	32	63.36	63.65	0.29	0.0052	0.016	0.004
2000	40	79.83	79.95	0.12	0.0033	0.011	0.0033
4000	51	100.58	101.23	0.66	0.0018	0.014	0.0037

We observed that the test procedure (4.5) for (1.2) provides good performances especially when p is large.

6.2. Test Procedure (4.5) for (5.5)

We set $\Delta_L = (5/6)p^{3/4}$. We considered constructing a test having size $\alpha = 0.05$ and power no less than $1 - \beta = 0.9$ when $\Delta \geq \Delta_L$. We set $p = 250,\ 500,\ 1000,\ 2000$ and 4000. We handled (5.1) with (5.2). The following two cases were considered:

$$(c)\ \boldsymbol{\Sigma} = \mathrm{diag}(2, 1, 0, ..., 0) + \boldsymbol{I}_p \quad \text{and} \quad (d)\ \boldsymbol{\Sigma} = \mathrm{diag}(p^{3/8}, p^{1/4}, 0, ..., 0) + \boldsymbol{I}_p$$

for (5.3) and (5.4), respectively. Note that (A-vi) is met both for (c) and (d), and $\Delta = p^{3/4} + p^{1/2} + o(1) \geq \Delta_L$ when p is large for (d). We set $m = \lceil C/2 \rceil$.

8

Table 2. Required sample size and average power by the test procedure (4.5) for (1.2) in case of (b).

p	m	C	$\overline{N}$	$\overline{N} - C$	MSE(N/C)	$1 - \overline{\beta}$	$s(\overline{\beta})$
250	32	62.9	63.29	0.39	0.0079	0.771	0.0133
500	37	73.22	73.5	0.28	0.0048	0.809	0.0124
1000	44	86.17	86.43	0.26	0.0028	0.835	0.0117
2000	52	102.7	103.21	0.51	0.0021	0.871	0.0106
4000	62	123.43	124	0.56	0.0013	0.897	0.0096

Similar to Tables 1 and 2, we calculated $\overline{\alpha}$, $1 - \overline{\beta}$, $s(\overline{\alpha})$, $s(\overline{\beta})$, $\overline{N}$ and MSE(N/C), by 2000 replications. In Tables 3 and 4, we summarized the results. We observed that the test procedure (4.5) for (5.5) provides good performances especially when p is large.

Table 3. Required sample size and average size by the test procedure (4.5) for (5.5) in case of (c).

p	m	C	$\overline{N}$	$\overline{N} - C$	MSE(N/C)	$\overline{\alpha}$	$s(\overline{\alpha})$
250	15	29.16	29.82	0.66	0.031	0.115	0.0101
500	17	33.94	34.38	0.43	0.02	0.072	0.0082
1000	20	39.93	40.52	0.59	0.013	0.056	0.0073
2000	24	47.23	47.9	0.68	0.01	0.06	0.0075
4000	29	56.01	56.67	0.66	0.0062	0.058	0.0074

Throughout the simulations, we observed that the test procedure (4.5) meets the required accuracy successfully.

7. ACTUAL DATA ANALYSIS

We analyzed gene expression data for the test of (5.5). We handled microarray data of Naderi et al. (2007) with $47293 \, (= p)$ genetic probes. We used the data set of luminal group (84 samples). We set $\alpha = 0.05$ and $\beta = 0.1$. From (5.6) we set $\Delta_L = p^{4/5}$, that is, we designed the test of (5.5) to have size 0.05 and power no less than 0.9 when $\lambda_1 \geq p^{2/5}$. We set $m = 30$. We took the first 30 samples as a pilot sample. We calculated $W_m = 35079$ according to (2.2). From (4.4) the total sample size was calculated as

$$N = \max\left\{30, \; \left\lceil \frac{2(z_\alpha + z_\beta)W_m}{\Delta_L} \right\rceil\right\} = 38.$$

Thus we took the next $8 \, (= 38 - 30)$ samples. We calculated U_N and T_N according to (2.3) and (2.4). Then, it follows that

$$\frac{NT_N}{2U_N} > z_\alpha \, (= 1.64),$$

9

Table 4. Required sample size and average power by the test procedure (4.5) for (5.5) in case of (d).

p	m	C	$\overline{N}$	$\overline{N} - C$	MSE(N/C)	$1 - \overline{\beta}$	$s(\overline{\beta})$
250	20	39.38	39.65	0.28	0.0385	0.888	0.001
500	22	43.71	44.35	0.63	0.026	0.924	0.0084
1000	25	49.27	49.77	0.51	0.0193	0.914	0.0089
2000	29	56.17	56.57	0.4	0.0119	0.917	0.0087
4000	33	64.61	65.4	0.79	0.0088	0.94	0.0075

so that H_0 in (5.5) was rejected in terms of (4.2). We concluded that $\lambda_j \to \infty$ for the first several js and $\lambda_1^{-1} = O(p^{-2/5})$. Hence, we recommend to use new PCA methods given by Yata and Aoshima (2010, 2012) because $\lambda_j \to \infty$ for the first several js.

For instance, Yata and Aoshima (2012) developed a new PCA called the noise-reduction (NR) methodology. In the NR method, λ_js are estimated by

$$\tilde{\lambda}_j = \hat{\lambda}_j - \frac{\text{tr}(\boldsymbol{S}_n) - \sum_{i=1}^{j} \hat{\lambda}_i}{n - 1 - j} \quad (j = 1, ..., n - 2),$$

where $\hat{\lambda}_j$ is the j-th eigenvalue of $\boldsymbol{S}_n$. We note that $\tilde{\lambda}_j$ has a consistency property in the sense that

$$\tilde{\lambda}_j / \lambda_j = 1 + o_P(1) \quad \text{when } \lambda_j \to \infty \text{ as } p \to \infty$$

under some regularity conditions. On the other hand, the conventional estimator, $\hat{\lambda}_j$, includes a large bias in the sense that

$$\hat{\lambda}_j / \lambda_j = 1 + \kappa_j + o_P(1) \quad \text{when } \lambda_j \to \infty \text{ as } p \to \infty, \tag{7.1}$$

where $\kappa_j = \lambda_j^{-1} \sum_{i=k+1}^{p} \lambda_i / (n - 1)$. See Yata and Aoshima (2012) and Aoshima and Yata (2018) for the details. In Table 5, we estimated the first five eigenvalues for the data set ($38 (= n)$) samples) both by the NR method and the conventional PCA. We observed that $\hat{\lambda}_j$ is quite large compared with $\tilde{\lambda}_j$ for all j. This

Table 5. Estimates of the first five eigenvalues by the NR method and conventional PCA together with their ratios for the data set in Naderi et al. (2007).

j	1	2	3	4	5
$\tilde{\lambda}_j$	225.2	120.1	89.3	66	54.3
$\hat{\lambda}_j$	278.8	170.4	137	111.8	98.5
$\hat{\lambda}_j / \tilde{\lambda}_j$	1.238	1.419	1.535	1.694	1.813

is probably because the bias in (7.1) is quite large for each j. On the other hand, $\tilde{\lambda}_j$ does not depend on the bias under (5.4). Thus, we recommend to use the NR method (or the CDM method by Yata and Aoshima (2010)) when H_0 in (5.5) is rejected.

A. APPENDIX

We put $K = 4\mathrm{tr}(\mathbf{\Sigma}^2)^2/n^2$ throughout this section. Let $\widehat{\mathbf{\Sigma}}_{ij(1)} = n_{(1)}(\mathbf{x}_i - \overline{\mathbf{x}}_{(1)(i+j)})(\mathbf{x}_i - \overline{\mathbf{x}}_{(1)(i+j)})^T/(n_{(1)} - 1)$, $\widehat{\mathbf{\Sigma}}_{ij(2)} = n_{(2)}(\mathbf{x}_j - \overline{\mathbf{x}}_{(2)(i+j)})(\mathbf{x}_j - \overline{\mathbf{x}}_{(2)(i+j)})^T/(n_{(2)} - 1)$, $A_{ij} = \mathrm{tr}\{(\widehat{\mathbf{\Sigma}}_{ij(1)} - \sigma\mathbf{I}_p)(\widehat{\mathbf{\Sigma}}_{ij(2)} - \sigma\mathbf{I}_p)\}$ and $B_{ij} = \mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(1)} - \sigma\mathbf{I}_p)\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(2)} - \sigma\mathbf{I}_p)/p$ for all $i < j$.

Proofs of Lemmas 2.2 and 2.3. We write that

$$\varepsilon_{ij} = A_{ij} + \sigma\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(1)}) + \sigma\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(2)}) - \sigma^2 p \quad \text{and}$$
$$\zeta_{ij} = B_{ij} + \sigma\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(1)}) + \sigma\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(2)}) - \sigma^2 p$$

for all $i < j$. Note that $W_n = 2\sum_{i<j}^n \varepsilon_{ij}/\{n(n-1)\}$ and $U_n = 2\sum_{i<j}^n \zeta_{ij}/\{n(n-1)\}$. Thus, it holds that

$$T_n = 2\sum_{i<j}^n \frac{A_{ij}}{n(n-1)} - 2\sum_{i<j}^n \frac{B_{ij}}{n(n-1)}. \tag{A.1}$$

Here, we can evaluate that

$$\mathrm{Var}\left(2\sum_{i<j}^n \frac{B_{ij}}{n(n-1)}\right) = O\left(\frac{\mathrm{tr}(\mathbf{\Sigma}^2)^2}{p^2 n^2}\right) = o(K) \quad \text{and} \tag{A.2}$$
$$\mathrm{Var}\left(2\sum_{i<j}^n \frac{\sigma\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(1)}) + \sigma\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{ij(2)})}{n(n-1)}\right) = O\left(\frac{\sigma^2\mathrm{tr}(\mathbf{\Sigma}^2)}{n}\right)$$

under (A-i) and (A-iii). Thus, we conclude the result of Lemma 2.2. On the other hand, from Lemma 5.1 in Yata and Aoshima (2016), we have that

$$\mathrm{Var}\left(2\sum_{i<j}^n \frac{A_{ij}}{n(n-1)}\right)$$
$$= \left\{\frac{8\mathrm{tr}\{(\mathbf{\Sigma}\mathbf{\Sigma}_*)^2\} + 4\sum_{j=1}^q (M_j - 2)(\boldsymbol{\gamma}_j^T \mathbf{\Sigma}_* \boldsymbol{\gamma}_j)^2}{n} + K\right\}\{1 + o(1)\} + O\left(\frac{\mathrm{tr}(\mathbf{\Sigma}^4)}{n^2}\right) \tag{A.3}$$

under (A-i) and (A-iii). Then, by noting that $\sum_{j=1}^q (\boldsymbol{\gamma}_j^T \mathbf{\Sigma}_* \boldsymbol{\gamma}_j)^2 \le \sum_{j,j'=1}^q (\boldsymbol{\gamma}_j^T \mathbf{\Sigma}_* \boldsymbol{\gamma}_{j'})^2 = \mathrm{tr}\{(\mathbf{\Sigma}\mathbf{\Sigma}_*)^2\}$ and

$$E\left\{\left(2\sum_{i<j}^n \frac{A_{ij}}{n(n-1)} - \Delta\right)\left(2\sum_{i<j}^n \frac{B_{ij}}{n(n-1)}\right)\right\} = o\left\{\mathrm{Var}\left(2\sum_{i<j}^n \frac{A_{ij}}{n(n-1)}\right)^{1/2} K^{1/2}\right\}$$

under (A-i) and (A-iii), from (A.2) and (A.3), we can conclude the result of Lemma 2.3. $\square$

Proof of Lemma 3.1. Note that $\mathrm{tr}(\mathbf{\Sigma}^4) \le \mathrm{tr}(\mathbf{\Sigma}^2)^2$ and $\mathrm{tr}\{(\mathbf{\Sigma}\mathbf{\Sigma}_*)^2\} \le \lambda_1\mathrm{tr}(\mathbf{\Sigma}_*\mathbf{\Sigma}\mathbf{\Sigma}_*) \le \lambda_1^2\Delta \le \mathrm{tr}(\mathbf{\Sigma}^2)\Delta$. Then, from Lemma 2.3, it holds that

$$\mathrm{Var}(T_n/\Delta) = O\left\{\mathrm{tr}(\mathbf{\Sigma}^2)^2/(n^2\Delta^2) + \mathrm{tr}\{(\mathbf{\Sigma}\mathbf{\Sigma}_*)^2\}/(n\Delta^2)\right\} \to 0$$

under (A-i), (A-iii) and (A-iv), so that $T_n/\Delta = 1 + o_P(1)$. It concludes the result. $\square$

Proof of Lemma 3.2. If $\liminf_{p \to \infty} \text{tr}(\boldsymbol{\Sigma}^2)/(\sigma^2 p) > 1$, it holds that $\liminf_{p \to \infty} \Delta/\text{tr}(\boldsymbol{\Sigma}^2) > 0$, so that (A-iv) holds. Thus under (A-v), it holds that as $p \to \infty$

$$\text{tr}(\boldsymbol{\Sigma}^2)/(\sigma^2 p) \to 1. \tag{A.4}$$

If $\liminf_{p \to \infty} \lambda_1^2/(\sigma^2 p) > 0$, it holds that $\liminf_{p \to \infty} \Delta/(\sigma^2 p) > 0$ from the fact that $\sigma p = \sum_{j=1}^{p} \lambda_j$. Thus, under (A-v) it follows that $\lambda_1^2/\text{tr}(\boldsymbol{\Sigma}^2) \to 0$ as $p \to \infty$, so that

$$\frac{\text{tr}(\boldsymbol{\Sigma}^4)}{\text{tr}(\boldsymbol{\Sigma}^2)^2} \le \frac{\lambda_1^2 \text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}(\boldsymbol{\Sigma}^2)^2} \to 0 \tag{A.5}$$

as $p \to \infty$ under (A-v). Then, from Corollary 5.2 in Yata and Aoshima (2016), we have that

$$2 \sum_{i<j}^{n} \frac{A_{ij}}{K^{1/2} n(n-1)} \Rightarrow N(0,1)$$

under (A-ii), (A-iii) and (A-v). Thus, from (A.1) and (A.2) we conclude the result. $\qquad\square$

Proofs of Theorem 3.1 and Corollary 3.1. First, we consider Corollary 3.1. From Lemmas 2.2 and 3.1 we have that

$$P\left(\frac{nT_n}{2U_n} > z_\alpha\right) = P\left(\frac{T_n}{\Delta} > z_\alpha \frac{2\sigma^2 p\{1 + o_P(1)\}}{n\Delta}\right)$$
$$= P\{1 + o_P(1) > o_P(1)\} \to 1$$

under (A-i), (A-iii) and (A-iv) from the fact that $\sigma^2 p \le \text{tr}(\boldsymbol{\Sigma}^2)$. It concludes the result of Corollary 3.1.

Next, we consider Theorem 3.1. From Lemmas 2.2, 3.2 and (A.4) we have that

$$P\left(\frac{nT_n}{2U_n} > z_\alpha\right) = P\left(\frac{T_n - \Delta}{2\text{tr}(\boldsymbol{\Sigma}^2)/n} > z_\alpha \frac{U_n}{\text{tr}(\boldsymbol{\Sigma}^2)} - \frac{n\Delta}{2\text{tr}(\boldsymbol{\Sigma}^2)}\right) \tag{A.6}$$
$$= \Phi\left(\frac{n\Delta}{2\text{tr}(\boldsymbol{\Sigma}^2)} - z_\alpha\right) + o(1)$$

under (A-ii), (A-iii) and (A-v). Hence, we conclude the result of Theorem 3.1. The proofs are completed. $\qquad\square$

Proof of Theorem 4.1. From Theorem 3.1, the result of Theorem 4.1 is obtained straightforwardly. $\qquad\square$

Proof of Theorem 4.2. We assume that $\boldsymbol{\mu} = \boldsymbol{0}$ without loss of generality. Let $C_L = \lfloor C - (\omega C)^{1/2} \rfloor$ and $C_U = \lceil C + (\omega C)^{1/2} \rceil$, where $\omega\ (> 0)$ is a variable such that $\omega \to 0$ as $p \to \infty$. Under (A-vi), (A.5) holds as $p \to \infty$. Then, from the proof of Theorem 5 in Aoshima and Yata (2014), it holds that under (A-i) and (A-vi)

$$\max\{m, C_L\} \le N < C_U \tag{A.7}$$

as $p \to \infty$ with probability tending to 1. Let $A_{oij} = \text{tr}\{(\boldsymbol{x}_i \boldsymbol{x}_i^T - \sigma \boldsymbol{I}_p)(\boldsymbol{x}_j \boldsymbol{x}_j^T - \sigma \boldsymbol{I}_p)\}$ and $B_{oij} = \text{tr}(\boldsymbol{x}_i \boldsymbol{x}_i^T - \sigma \boldsymbol{I}_p)\text{tr}(\boldsymbol{x}_j \boldsymbol{x}_j^T - \sigma \boldsymbol{I}_p)/p$ for all $i < j$. Now, we write that

$$T_{oN} = \sum_{i<j}^{C_L} \frac{2(A_{oij} - B_{oij})}{N(N-1)} + \sum_{j=C_L+1}^{N} \sum_{i=1}^{C_L} \frac{2(A_{oij} - B_{oij})}{N(N-1)} + \sum_{i \ne j(>C_L)}^{N} \frac{(A_{oij} - B_{oij})}{N(N-1)}. \tag{A.8}$$

Let $K_C = 4\mathrm{tr}(\boldsymbol{\Sigma}^2)^2/C_L^2$. By using Chebyshev's inequality and Schwarz's inequality, for any $\eta > 0$, from (A.7), we have that as $p \to \infty$

$$P\Big(\Big|\sum_{j=C_L+1}^{N}\sum_{i=1}^{C_L}(A_{oij}-B_{oij}-\Delta)/C^2\Big| > \eta K_C^{1/2}\Big)$$

$$\leq P\Big(\sum_{j=C_L+1}^{C_U}\Big|\sum_{i=1}^{C_L}(A_{oij}-B_{oij}-\Delta)/C^2\Big| > \eta K_C^{1/2}\Big)$$

$$= O\{\omega(\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 + \mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_*)^2\})/(C^2 K_C)\} \to 0 \text{ and}$$

$$P\Big(\Big|\sum_{i\neq j(>C_L)}^{N}(A_{oij}-B_{oij}-\Delta)/C^2\Big| > \eta K_C^{1/2}\Big)$$

$$\leq P\Big(\sum_{i\neq j(>C_L)}^{C_U}\Big|(A_{oij}-B_{oij}-\Delta)/C^2\Big| > \eta K_C^{1/2}\Big) = O\{\omega^2(\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 + \mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_*)^2\})/(C^2 K_C)\} \to 0$$

under (A-ii) and (A-vi) from the fact that $\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_*)^2\} = O\{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2\}$. Thus, from (A.8) and Lemma 3.2, we have that

$$\frac{T_{oN}-\Delta}{K_C^{1/2}} = \sum_{i<j}^{C_L}\frac{2(A_{oij}-B_{oij}-\Delta)}{K_C^{1/2}N(N-1)} + o_P(1) = \sum_{i<j}^{C_L}\frac{2(A_{oij}-B_{oij}-\Delta)}{K_C^{1/2}C_L(C_L-1)} + o_P(1) \Rightarrow N(0,1) \quad \text{(A.9)}$$

under (A-ii) and (A-vi) from the fact that

$$T_{C_L}-\Delta = \sum_{i<j}^{C_L}\frac{2(A_{oij}-B_{oij}-\Delta)}{C_L(C_L-1)} + o_P(K_C^{1/2}).$$

Here, in a way similar to the proof of Lemma A.5 in Yata and Aoshima (2013), we have that

$$T_N = T_{oN} + o_P(K_C^{1/2}) \tag{A.10}$$

under (A-ii) and (A-vi). By combining (A.9) with (A.10), we conclude the result. $\square$

Proofs of Corollaries 5.1 and 5.2. Under (5.1) and (5.3), it holds that $n\Delta/\mathrm{tr}(\boldsymbol{\Sigma}^2) \to 0$ under (A-iii) and $C\Delta/\mathrm{tr}(\boldsymbol{\Sigma}^2) = O(1/\Delta_L) \to 0$ as $p \to \infty$. Then, from Theorems 3.1 and 4.2, we conclude the results. $\square$

ACKNOWLEDGEMENTS

REFERENCES

Ahn, J., Marron, J. S., Muller, K. M., and Chi, Y.-Y. (2007). The High-Dimension, Low-Sample-Size Geometric Representation Holds Under Mild Conditions, *Biometrika* 94: 760–766.

Aoshima, M. and Yata, K. (2011a). Two-Stage Procedures for High-Dimensional Data, *Editor's special invited paper, Sequential Analysis* 30: 356–399.

Aoshima, M. and Yata, K. (2011b). Authors' Response, *Sequential Analysis* 30: 432–440.

Aoshima, M. and Yata, K. (2014). A Distance-Based, Misclassification Rate Adjusted Classifier for Multi-class, High-Dimensional Data, *Annals of Institute of Statistical Mathematics* 66: 983–1010.

Aoshima, M. and Yata, K. (2015). Asymptotic Normality for Inference on Multisample, High-Dimensional Mean Vectors under Mild Conditions, *Methodology and Computing in Applied Probability* 17: 419–439.

Aoshima, M. and Yata, K. (2018). Two-Sample Tests for High-Dimension, Strongly Spiked Eigenvalue Models. *Statistica Sinica* 28, 43–62.

Bai, Z. and Saranadasa, H. (1996). Effect of High Dimension: By an Example of a Two Sample Problem, *Statistica Sinica* 6: 311–329.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models, *Journal of Multivariate Analysis* 97: 1382–1408.

Chen, S. X. and Qin, Y.-L. (2010). A Two-Sample Test for High-Dimensional Data with Applications to Gene-Set Testing, *Annals of Statistics* 38: 808–835.

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.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large Covariance Estimation by Thresholding Principal Orthogonal Complements, *Journal of the Royal Statistal Society, Series B* 75: 603-680.

Ishii, A., Yata, K., and Aoshima, M. (2016). Asymptotic Properties of the First Principal Component and Equality Tests of Covariance Matrices in High-Dimension, Low-Sample-Size Context, *Journal of Statistical Planning and Inference* 170: 186–199.

Johnstone, I. M. (2001). On the Distribution of the Largest Eigenvalue in Principal Components Analysis, *Annals of Statistics* 29: 295–327.

Johnstone, I. M. and Lu, Y. (2009). On Consistency and Sparsity for Principal Components Analysis in High Dimensions, *Journal of the American Statistical Association* 104: 682–693.

Jung, S. and Marron, J. S. (2009). PCA Consistency in High Dimension, Low Sample Size Context, *Annals of Statistics* 37: 4104–4130.

Ledoit, O. and Wolf, M. (2002). Some Hypothesis Tests for the Covariance Matrix When the Dimension Is Large Compared to the Sample Size, *Annals of Statistics* 30: 1081–1102.

Naderi, A., Teschendorff, A. E., Barbosa-Morais, N. L., Pinder, S. E., Green, A. R., Powe, D. G., Robertson, J. F., Aparicio, S., Ellis, I. O., Brenton, J. D., and Caldas, C. (2007). A Gene-Expression Signature to Predict Survival in Breast Cancer Across Independent Data Sets, *Oncogene* 26: 1507–1516.

Nagao, H. (1973). On Some Test Criteria for Covariance Matrix, *Annals of Statistics* 1: 700–709.

Paul, D. (2007). Asymptotics of Sample Eigenstructure for a Large Dimensional Spiked Covariance Model, *Statistica Sinica* 17: 1617–1642.

Shen, D., Shen, H., Zhu, H., and Marron, J. S. (2016). The Statistics and Mathematics of High Dimension Low Sample Size Asymptotics, *Statistica Sinica* 26: 1747–1770.

Srivastava, M. S., Kollo, T., and von Rosen, D. (2011). Some Tests for the Covariance Matrix with Fewer Observations Than the Dimension under Non-Normality, *Journal of Multivariate Analysis* 102: 1090–1103.

Yata, K. and Aoshima, M. (2009). PCA Consistency for Non-Gaussian Data in High Dimension, Low Sample Size Context, *Communications in Statistics - Theory and Methods, Special Issue Honoring Zacks, S. (ed. Mukhopadhyay, N.)* 38: 2634–2652.

Yata, K. and Aoshima, M. (2010). Effective PCA for High-Dimension, Low-Sample-Size Data with Singular Value Decomposition of Cross Data Matrix, *Journal of Multivariate Analysis* 101: 2060–2077.

Yata, K. and Aoshima, M. (2012). Effective PCA for High-Dimension, Low-Sample-Size Data with Noise Reduction via Geometric Representations, *Journal of Multivariate Analysis* 105: 193–215.

Yata, K. and Aoshima, M. (2013). Correlation Tests for High-Dimensional Data Using Extended Cross-Data-Matrix Methodology, *Journal of Multivariate Analysis* 117: 313–331.

Yata, K. and Aoshima, M. (2016). High-Dimensional Inference on Covariance Structures via the Extended Cross-Data-Matrix Methodology, *Journal of Multivariate Analysis* 151: 151–166.