Journal of Multivariate Analysis 101 (2010) 2060–2077

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix

Kazuyoshi Yata [a], Makoto Aoshima [b,*]

[a] Graduate School of Pure and Applied Sciences, University of Tsukuba, Ibaraki 305-8571, Japan
[b] Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan

ARTICLE INFO

*Article history:*
Received 4 June 2009
Available online 24 April 2010

*AMS subject classifications:*
primary 62H25
62H30
secondary 34L20

*Keywords:*
Consistency
Eigenvalue distribution
HDLSS
Microarray data analysis
Mixture model
Principal component analysis
Singular value

ABSTRACT

In this paper, we propose a new methodology to deal with PCA in high-dimension, low-sample-size (HDLSS) data situations. We give an idea of estimating eigenvalues via singular values of a cross data matrix. We provide consistency properties of the eigenvalue estimation as well as its limiting distribution when the dimension $d$ and the sample size $n$ both grow to infinity in such a way that $n$ is much lower than $d$. We apply the new methodology to estimating PC directions and PC scores in HDLSS data situations. We give an application of the findings in this paper to a mixture model to classify a dataset into two clusters. We demonstrate how the new methodology performs by using HDLSS data from a microarray study of prostate cancer.



## 1. Introduction

High-dimension, low-sample-size (HDLSS) data are emerging in various areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. The asymptotic studies of these types of data are becoming increasingly relevant. Principal Component Analysis (PCA) is an important tool of dimension reduction especially when the dimension is very high. PCA visualizes important underlying structures in the data by approximating the data with the first few principal components. Let us see Fig. 1. The data in Fig. 1, described in detail in Singh et al. [11] and Pochet et al. [10], are from a microarray study of prostate cancer. Different symbols correspond to cancer subtypes. The dataset contains 34 patients with 12 600 genes. There are 9 Normal Prostate (plotted as o) and 25 Prostate Tumors (plotted as $\times$). Fig. 1 shows the projections of the data onto the sub-spaces generated by the first three PC directions (PC1, PC2 and PC3). We carried out PCA using this data to reduce the high dimensionality to a few specified dimensions so that it could be visualized effectively.

As observed in Fig. 1, the first few PC directions seem to separate the normal and tumor samples. However, the separation between the two cases is not always clear. One of the causes of obscurity is in extreme high-dimensional setting in the sense of a small number of patients and a large number of gene expression levels for each patient. It is very crucial in studying PCA in HDLSS data situations.

In recent years, substantial work has been done on the asymptotic behavior of eigenvalues of the sample covariance matrix in the limit as $d \to \infty$; see Johnstone [6], Baik et al. [2] and Paul [9] for Gaussian assumptions, and Baik and

* Corresponding author.
*E-mail address:* aoshima@math.tsukuba.ac.jp (M. Aoshima).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.04.006

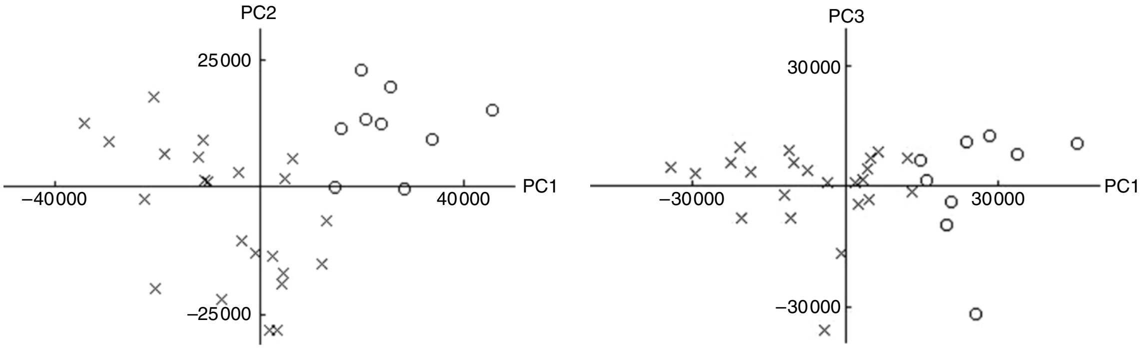


**Fig. 1.** Scatterplots of PC scores by PC1 and PC2 (left panel) or PC1 and PC3 (right panel). There are 9 Normal Prostate (plotted as o) and 25 Prostate Tumors (plotted as $\times$).

Silverstein [3] for non-Gaussian but i.i.d. assumptions when $d$ and $n$ increase at the same rate, i.e. $n/d \to c > 0$. On the other hand, Johnstone and Lu [7] have shown that the estimate of the leading principal component vector is consistent if and only if $d(n)/n \to 0$. Many ofthese focus on the spiked covariance model introduced by Johnstone [6]. In HDLSS settings, Hall et al. [4] and Ahn et al. [1] have studied the HDLSS asymptotics in which $d \to \infty$ while $n$ is fixed. They explored conditions to give a geometric representation of HDLSS data. The HDLSS asymptotics usually regulate either the population distribution by the normality or the dependency of the random variables in the sphered data matrix by the $\rho$-mixing condition as described, for example, on p. 440 in Hall et al. [4]. Those assumptions are somewhat too strict and have some obvious shortcomings. Yata and Aoshima [12] have developed the HDLSS asymptotics in more general settings without assuming either the normality or the $\rho$-mixing condition and applied to estimating the intrinsic dimension of a HDLSS dataset.

In this paper, suppose we have a $d \times n$ data matrix $\boldsymbol{X}_{(d)} = [\boldsymbol{x}_{1(d)}, \dots, \boldsymbol{x}_{n(d)}]$ with $d > n$, where $\boldsymbol{x}_{k(d)} = (x_{1k(d)}, \dots, x_{dk(d)})^T,\ k = 1, \dots, n$, are independent and identically distributed as a $d$-dimensional multivariate distribution with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}_d$. The eigen-decomposition of $\boldsymbol{\Sigma}_d$ is $\boldsymbol{\Sigma}_d = \boldsymbol{H}_d \boldsymbol{\Lambda}_d \boldsymbol{H}_d^T$, where $\boldsymbol{\Lambda}_d$ is a diagonal matrix of eigenvalues $\lambda_{1(d)} \ge \dots \ge \lambda_{d(d)} (>0)$ and $\boldsymbol{H}_d = [\boldsymbol{h}_{1(d)}, \dots, \boldsymbol{h}_{d(d)}]$ is a matrix of corresponding eigenvectors. Then, $\boldsymbol{Z}_{(d)} = \boldsymbol{\Lambda}_d^{-1/2} \boldsymbol{H}_d^T \boldsymbol{X}_{(d)}$ is a $d \times n$ sphered data matrix from a distribution with the identity covariance matrix. Here, we write $\boldsymbol{Z}_{(d)} = [\boldsymbol{z}_{1(d)}, \dots, \boldsymbol{z}_{d(d)}]^T$ and $\boldsymbol{z}_{j(d)} = (z_{j1(d)}, \dots, z_{jn(d)})^T,\ j = 1, \dots, d$. Hereafter, the subscript $d$ will be omitted for the sake of simplicity when it does not cause any confusion. We assume that the fourth moments of each variable in $\boldsymbol{Z}$ are uniformly bounded and $\|\boldsymbol{z}_j\| \ne 0$ for $j = 1, \dots, d$, where $\|\cdot\|$ denotes the Euclidean norm. The multivariate distribution assumed here does not have to be Gaussian and the random variables in $\boldsymbol{Z}$ do not have to be regulated by the $\rho$-mixing condition. Then, we consider a general setting as follows:

$$\lambda_j = a_j d^{\alpha_j} \quad (j = 1, \dots, m) \quad \text{and} \quad \lambda_j = c_j \quad (j = m+1, \dots, d). \tag{1}$$

Here, $a_j (>0)$, $c_j (>0)$ and $\alpha_j$ ($\alpha_1 \ge \dots \ge \alpha_m > 0$) are unknown constants preserving the ordering that $\lambda_1 \ge \dots \ge \lambda_d$, and $m$ is an unknown positive integer.

The sample covariance matrix is $\boldsymbol{S} = n^{-1}\boldsymbol{X}\boldsymbol{X}^T$ and its dual matrix is defined by $\boldsymbol{S}_D = n^{-1}\boldsymbol{X}^T\boldsymbol{X}$. Note that $\boldsymbol{S}_D$ and $\boldsymbol{S}$ share non-zero eigenvalues. Let $\hat{\lambda}_1 \ge \dots \ge \hat{\lambda}_n (\ge 0)$ be the eigenvalues of $\boldsymbol{S}_D$. Let us write the eigen-decomposition of $\boldsymbol{S}_D$ as $\boldsymbol{S}_D = \sum_{j=1}^n \hat{\lambda}_j \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j^T$. Jung and Marron [8] found it strongly inconsistent for estimating PC directions of HDLSS data satisfying (1) along with

$$\sum_{j=1}^d \acute{\lambda}_j^2 \to 0 \quad \text{as } d \to \infty, \text{ where } \acute{\lambda}_j = \lambda_j \Big/ \left( \sum_{j=1}^d \lambda_j \right). \tag{2}$$

We note that the formulation (1), provided that $\alpha_1 < 1$ and $c_d > 0$, includes the case satisfying (2). Recently, Yata and Aoshima [13] have given the convergence conditions with respect to $d$ and $n$ to claim the consistency properties for the sample eigenvalues as well as the PC directions and the PC scores: For $j = 1, \dots, m$, it holds that

$$\frac{\hat{\lambda}_j}{\lambda_j} = 1 + o_p(1) \tag{3}$$

under the conditions:

(YA-i) $d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1$;
(YA-ii) $d \to \infty$ and $d^{2-2\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (0, 1]$.

If $z_{jk},\ j = 1, \dots, d$ ($k = 1, \dots, n$) are independent, the above conditions are modified as

(YA-i′) $d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1$;
(YA-ii′) $d \to \infty$ and $d^{1-\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (0, 1]$.

In addition, they have given the limiting distribution of the sample eigenvalue. It should be noted that $n$ is free from $d$ in condition (YA-i) or (YA-i′). The condition for $\alpha_j > 1$ is more relaxed than that for $\alpha_j \in (0, 1]$ given by (YA-ii) or (YA-ii′). The facts described above draw our attention to the limitations of the capabilities of naive PCA in HDLSS data situations. Let us see a case, say, that $d = 1000$, $\lambda_1 = d^{2/3}$ and $\lambda_2 = \cdots = \lambda_d = 1$. Then, we observe from (YA-ii) that one requires the sample size to be $n \gg d^{2-2\alpha_1} = d^{2/3} = 100$. It is somewhat inconvenient for the experimenter to handle HDLSS data situations.

In this paper, we propose a new methodology to deal with PCA in HDLSS data situations. In Section 2, we give an idea of estimating eigenvalues via singular values of a cross data matrix. We provide consistency properties of the eigenvalue estimation as well as its limiting distribution. The new methodology is examined in its performance in Section 3. We apply the new methodology to estimating PC directions and PC scores in Sections 4 and 5. In Section 6, we give an application of the findings in this paper to a mixture model to classify a dataset into two clusters. In Section 7, we demonstrate how the new methodology performs in HDLSS data situations with the microarray data used in Fig. 1.

## 2. New estimation methodology

Suppose we have two $d \times n$ data matrices, $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in}]$, $i = 1, 2$, where $\boldsymbol{x}_{ik} = (x_{i1k}, \ldots, x_{idk})^T$, $i = 1, 2$; $k = 1, \ldots, n$, are independent and identically distributed as a $d$-dimensional multivariate distribution as stated before. Note that the size $n$ in $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ may be different. We define a cross data matrix by $\boldsymbol{S}_{D(1)} = n^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2$ or $\boldsymbol{S}_{D(2)} = n^{-1}\boldsymbol{X}_2^T\boldsymbol{X}_1$ $(=\boldsymbol{S}_{D(1)}^T)$. Let us write $\boldsymbol{Z}_i = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{H}^T\boldsymbol{X}_i$, $i = 1, 2$, as $d \times n$ sphered data matrices from a distribution with the identity covariance matrix. Note that $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are independent. Let $\boldsymbol{Z}_i = [\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{id}]^T$ and $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijn})^T$, $i = 1, 2$; $j = 1, \ldots, d$. Then, we have that $\boldsymbol{S}_{D(1)} = n^{-1}\sum_{j=1}^{d}\lambda_j\boldsymbol{z}_{1j}\boldsymbol{z}_{2j}^T$. When we consider the singular value decomposition of $\boldsymbol{S}_{D(1)}$, it follows that $\boldsymbol{S}_{D(1)} = \sum_{j=1}^{n}\tilde{\lambda}_j\tilde{\boldsymbol{u}}_{j(1)}\tilde{\boldsymbol{u}}_{j(2)}^T$, where $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ $(\geq 0)$ denote singular values of $\boldsymbol{S}_{D(1)}$, and $\tilde{\boldsymbol{u}}_{j(1)}$ (or $\tilde{\boldsymbol{u}}_{j(2)}$) denotes a unit left- (or right-) singular vector corresponding to $\tilde{\lambda}_j$ $(j = 1, \ldots, n)$.

Now, we consider an easy example such as $\lambda_1 = d^{\alpha_1}$, $\lambda_2 = \cdots = \lambda_d = 1$, where $\alpha_1 \in (1/2, 1)$. Note that it is satisfying (2). Let us write $\lambda_1^{-1}\boldsymbol{S}_{D(1)} = n^{-1}\boldsymbol{z}_{11}\boldsymbol{z}_{21}^T + (n\lambda_1)^{-1}\sum_{j=2}^{d}\boldsymbol{z}_{1j}\boldsymbol{z}_{2j}^T$. Here, by using Markov's inequality for any $\tau > 0$, one has for all elements of $(n\lambda_1)^{-1}\sum_{j=2}^{d}\boldsymbol{z}_{1j}\boldsymbol{z}_{2j}^T$ that

$$P\left(\sum_{i',j'}\left((n\lambda_1)^{-1}\sum_{j=2}^{d}z_{1ji'}z_{2jj'}\right)^2 > \tau\right) \leq \tau^{-1}d^{1-2\alpha_1} = o(1)$$

as $d \to \infty$ either when $n \to \infty$ or $n$ is fixed. Thus we have that $\sum_{i',j'}((n\lambda_1)^{-1}\sum_{j=2}^{d}z_{1ji'}z_{2jj'})^2 = o_p(1)$. Let $\boldsymbol{e}_{in} = (e_{i1}, \ldots, e_{in})^T$, $i = 1, 2$, be arbitrary unit $n$-vectors. Then, we have that

$$\left|\sum_{i',j'}e_{1i'}e_{2j'}\sum_{j=2}^{d}(n\lambda_1)^{-1}z_{1ji'}z_{2jj'}\right| \leq \left(\sum_{i',j'}\left((n\lambda_1)^{-1}\sum_{j=2}^{d}z_{1ji'}z_{2jj'}\right)^2\right)^{1/2} = o_p(1).$$

Thus it holds that $\lambda_1^{-1}\boldsymbol{e}_{1n}^T\boldsymbol{S}_{D(1)}\boldsymbol{e}_{2n} = \boldsymbol{e}_{1n}^Tn^{-1}\boldsymbol{z}_{11}\boldsymbol{z}_{21}^T\boldsymbol{e}_{2n} + o_p(1)$. Now, let us consider singular values of $\boldsymbol{S}_{D(1)}$. Noting that $\|n^{-1/2}\boldsymbol{z}_{ij}\| = 1 + o_p(1)$ as $n \to \infty$, we claim as $d \to \infty$ and $n \to \infty$ that

$$\frac{\tilde{\lambda}_1}{\lambda_1} = \max\left(\boldsymbol{e}_{1n}^Tn^{-1}\boldsymbol{z}_{11}\boldsymbol{z}_{21}^T\boldsymbol{e}_{2n} + o_p(1)\right) = 1 + o_p(1)$$

with respect to any unit $n$-vectors $\boldsymbol{e}_{1n}$ and $\boldsymbol{e}_{2n}$. When we compare that fact with (3), it is observed that the singular value $\tilde{\lambda}_1$ has consistency with $\lambda_1$ for $\alpha_1 \in (1/2, 1)$ under the condition that $d \to \infty$ and $n \to \infty$. The above convergence condition relaxes (YA-ii) for (3) in the sense that $n$ is chosen free from $d$. This is our motivation for the new estimation methodology to start with singular values of a cross data matrix $\boldsymbol{S}_{D(1)}$.

[*New estimation methodology* (Cross-data-matrix methodology)]

(*Step* 1) Define a cross data matrix by $\boldsymbol{S}_{D(1)} = n^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2$.

(*Step* 2) Calculate the singular values $\tilde{\lambda}_j$'s of $\boldsymbol{S}_{D(1)}$ for the estimation of $\lambda_j$'s.

Yata and Aoshima [12] considered a dual square matrix defined by $\boldsymbol{S}_{D(1)}^2(=\boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T)$ for the estimation of the intrinsic dimension of a HDLSS dataset. We have the following theorem.

**Theorem 1.** *For $j = 1, \ldots, m$, we have that*

$$\frac{\tilde{\lambda}_j}{\lambda_j} = 1 + o_p(1) \tag{4}$$

*under the conditions:*

(i) *$d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1/2$;*

(ii) *$d \to \infty$ and $d^{2-2\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (0, 1/2]$.*

**Corollary 1.** *Assume further in Theorem 1 that $z_{ijk}$, $j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ are independent. Then, for $j = 1, \ldots, m$, we have (4) under the conditions:*
(i) *$d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1/2$;*
(ii) *$d \to \infty$ and there exists a positive constant $\varepsilon_j$ satisfying $d^{1-2\alpha_j}/n < d^{-\varepsilon_j}$ for $j$ such that $\alpha_j \in (0, 1/2]$.*

**Corollary 2.** *When the population mean may not be zero, let us write that $\boldsymbol{S}_{oD(1)} = n^{-1}(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2)$, where $\overline{\boldsymbol{X}}_i = [\bar{\boldsymbol{x}}_{i1}, \ldots, \bar{\boldsymbol{x}}_{id}]^T$ is having $n$-vector $\bar{\boldsymbol{x}}_{ij} = (\bar{x}_{ij}, \ldots, \bar{x}_{ij})^T$ with $\bar{x}_{ij} = \sum_{k=1}^n x_{ijk}/n$ $(j = 1, \ldots, d)$ for each $i$ $(=1, 2)$. Then, after replacing $\boldsymbol{S}_{D(1)}$ with $\boldsymbol{S}_{oD(1)}$, the assertion in Theorem 1 (or Corollary 1) is still justified under those conditions.*

**Theorem 2.** *Let $V(z_{ijk}^2) = M_j$ $(<\infty)$ for $j = 1, \ldots, m$ $(i = 1, 2;\ k = 1, \ldots, n)$. Assume that the first $m$ population eigenvalues are distinct. Then, under the conditions* (i)–(ii) *in Theorem 1, we have for $j = 1, \ldots, m$, that*

$$\sqrt{\frac{2n}{M_j}}\left(\frac{\tilde{\lambda}_j}{\lambda_j} - 1\right) \Rightarrow N(0, 1), \tag{5}$$

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

**Corollary 3.** *Assume further in Theorem 2 that $z_{ijk}$, $j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ are independent. Then, for $j = 1, \ldots, m$, we have (5) under the conditions:*
(i) *$d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1/2$;*
(ii) *$d \to \infty$ and $d^{2-4\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (0, 1/2]$.*

**Remark 1.** When the population eigenvalues are not distinct such as $\lambda_1 \geq \cdots \geq \lambda_m$, we can still claim both Theorem 2 and Corollary 3 for some $j$ such that $\lambda_j$ has multiplicity one. When the population mean may not be zero, we can still claim both Theorem 2 and Corollary 3 by using $\boldsymbol{S}_{oD(1)}$ defined in Corollary 2.

**Remark 2.** Suppose that we have a $d \times n$ data matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] = [\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1n_1}, \boldsymbol{x}_{21}, \ldots, \boldsymbol{x}_{2n_2}]$, where $n_1 + n_2 = n$ with $n_1 = O(n)$ and $n_2 = O(n)$ for a fixed $n$. One may define $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ by $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i}]$, $i = 1, 2$. Then, one may generally define $\boldsymbol{S}_{D(1)} = (n_1 n_2)^{-1/2}\boldsymbol{X}_1^T\boldsymbol{X}_2$. Then, we can claim both Theorem 1 and Corollary 1. For Theorem 2 and Corollary 3, the result (5) is modified by

$$2\sqrt{\frac{n_1 n_2}{nM_j}}\left(\frac{\tilde{\lambda}_j}{\lambda_j} - 1\right) \Rightarrow N(0, 1).$$

Hence, the variance of $\tilde{\lambda}_j/\lambda_j$ is approximately given by $M_j n/(4n_1 n_2)$ which has the minimum $M_j/n$ when $n_1 = n_2$. We suggest that one should divide $\boldsymbol{X}$ into $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with equally balanced $n_1 = n_2$ $(=n')$ when $n = 2n'$ or $n_1 = n' + 1$ and $n_2 = n'$ when $n = 2n' + 1$. Then, for Theorem 2 and Corollary 3, the result (5) is modified by

$$\sqrt{\frac{n}{M_j}}\left(\frac{\tilde{\lambda}_j}{\lambda_j} - 1\right) \Rightarrow N(0, 1).$$

**Remark 3.** The condition (ii) given by Theorem 1 (or Theorem 2) is a sufficient condition for the case of $\alpha_j \in (0, 1/2]$. If more information is available about the distribution of $\boldsymbol{X}_i$, the condition (ii) can be relaxed to give consistency under a broader set of $(d, n)$ for the case of $\alpha_j \in (0, 1/2]$. For example, when $\boldsymbol{X}_i$ is Gaussian, the asymptotic property is claimed under a broader set of $(d, n)$ given by the condition (ii) of Corollary 1 (or Corollary 3).

**Remark 4.** In view of Theorem 1 compared to (3), the cross-data-matrix methodology successfully relaxes the condition for the case that $\alpha_j > 1/2$. The conditions given by Theorem 1 are not continuous in $\alpha_j$ at $\alpha_j = 1/2$. When $\boldsymbol{X}_i$ is Gaussian, the conditions given by Corollaries 1 and 3 are continuous in $\alpha_j$.

**Remark 5.** One might recall that the Partial Least Squares Regression (PLSR) deals with the singular value decomposition of a cross covariance matrix defined by a response variables matrix and a predictor variables matrix. See, for example, Chapter 3 in Hastie et al. [5]. It should be noted that the cross data matrix used in the new estimation methodology is defined by two independent data matrices taken from a common dataset. The cross-data-matrix methodology given in this paper is conceptually different from PLSR.

## 3. Performances

We observe naive PCA that the sample size $n$ should be determined depending on $d$ for $\alpha_i \in (1/2, 1]$ in (3). On the other hand, the cross-data-matrix methodology allows the experimenter to choose $n$ free from $d$ for the case that $\alpha_i > 1/2$ as seen in Theorems 1 and 2. The cross-data-matrix methodology might make it possible to give feasible estimation of eigenvalues

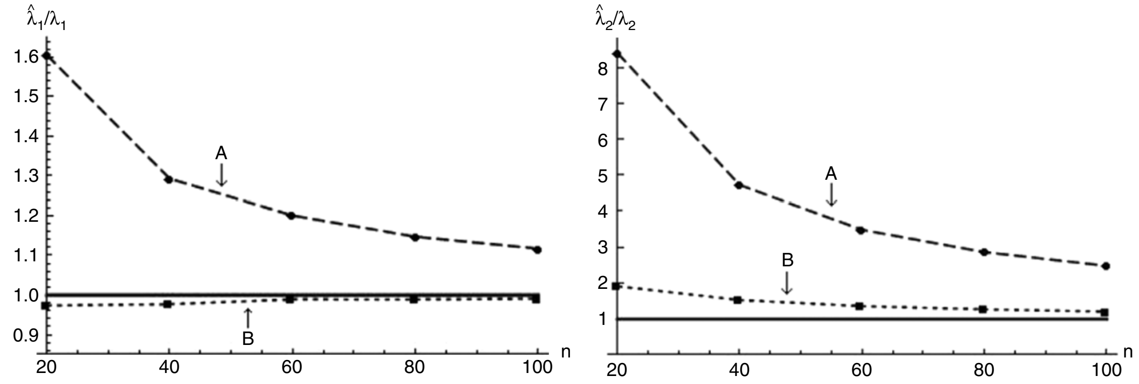


**Fig. 2.** The behaviors of $A$: $\hat{\lambda}_j/\lambda_j$ and $B$: $\tilde{\lambda}_j/\lambda_j$ for the first eigenvalue (left panel) and second eigenvalue (right panel) when the samples, of size $n = 20(20)100$, were taken from $N_d(\mathbf{0}, \boldsymbol{\Sigma})$ with $d = 1600$.

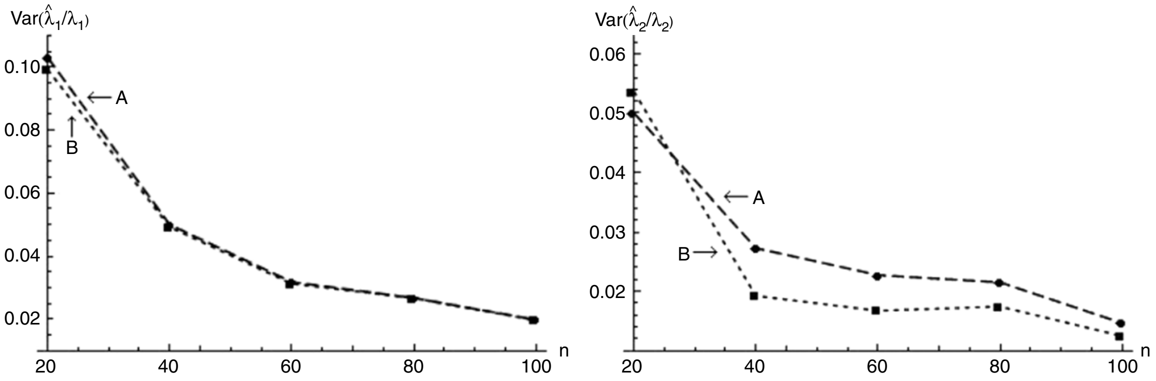


**Fig. 3.** The behaviors of $A$: $\mathrm{Var}(\hat{\lambda}_j/\lambda_j)$ and $B$: $\mathrm{Var}(\tilde{\lambda}_j/\lambda_j)$ for the first eigenvalue (left panel) and second eigenvalue (right panel) when the samples, of size $n = 20(20)100$, were taken from $N_d(\mathbf{0}, \boldsymbol{\Sigma})$ with $d = 1600$.

for HDLSS data with extremely small order of $n$ compared to $d$. In this section, we examine its performance with the help of Monte Carlo simulations.

We first consider the Gaussian case. Independent pseudorandom normal observations were generated from $N_d(\mathbf{0}, \boldsymbol{\Sigma})$ with $d = 1600$. We considered $\lambda_1 = d^{2/3}$, $\lambda_2 = d^{1/3}$ and $\lambda_3 = \cdots = \lambda_d = 1$ in (1). We used the sample of size $n = 20(20)100$ to define the data matrix $\boldsymbol{X}$:$d \times n$ for the calculation of $\boldsymbol{S}_D$, whereas we divided the sample into $\boldsymbol{X}_1$:$d \times (n/2)$ and $\boldsymbol{X}_2$:$d \times (n/2)$ for the calculation of $\boldsymbol{S}_{D(1)}$ in Theorem 1. The findings were obtained by averaging the outcomes from 1000 (=$R$, say) replications. Under a fixed scenario, suppose that the $r$-th replication ends with estimates of $\lambda_j$, $\hat{\lambda}_{jr}$ and $\tilde{\lambda}_{jr}$ ($r = 1, \ldots, R$), given by using (3) and Theorem 1, respectively. Let us simply write $\hat{\lambda}_j = R^{-1}\sum_{r=1}^{R}\hat{\lambda}_{jr}$ and $\tilde{\lambda}_j = R^{-1}\sum_{r=1}^{R}\tilde{\lambda}_{jr}$. We considered two quantities, $A$:$\hat{\lambda}_j/\lambda_j$ and $B$:$\tilde{\lambda}_j/\lambda_j$. Fig. 2 shows the behaviors of both $A$ and $B$ for the first two eigenvalues. By observing the behavior of $A$, (3) seems not to give a feasible estimation within the range of $n$. The sample size $n$ was not large enough to use the eigenvalues of $\boldsymbol{S}_D$ for such a high-dimensional space. On the other hand, in view of the behavior of $B$, Theorem 1 gives a reasonable estimation surprisingly well for such HDLSS datasets. The cross-data-matrix methodology seems to perform excellently as expected theoretically.

We also considered the Monte Carlo variability. Let $\mathrm{Var}(\hat{\lambda}_j/\lambda_j) = (R-1)^{-1}\sum_{r=1}^{R}(\hat{\lambda}_{jr} - \hat{\lambda}_j)^2/\lambda_j^2$ and $\mathrm{Var}(\tilde{\lambda}_j/\lambda_j) = (R-1)^{-1}\sum_{r=1}^{R}(\tilde{\lambda}_{jr} - \tilde{\lambda}_j)^2/\lambda_j^2$. We considered two quantities, $A$:$\mathrm{Var}(\hat{\lambda}_j/\lambda_j)$ and $B$:$\mathrm{Var}(\tilde{\lambda}_j/\lambda_j)$, in Fig. 3 to show the behaviors of sample variances of both $A$ and $B$ for the first two eigenvalues.

By observing the behaviors of the sample variances, both the behaviors seem not to make much difference between $A$ and $B$. From Theorem 2 of Yata and Aoshima [13], the limiting distribution of $(n/2)^{1/2}(\hat{\lambda}_j/\lambda_j - 1)$ is $N(0, 1)$, so that the variance of $A$ is approximately given by $\mathrm{Var}(\hat{\lambda}_j/\lambda_j) = 2/n$. On the other hand, in view of Theorem 2, noting that the sample is divided into two pieces of size $n/2$ for each in $B$, the limiting distribution of $(n/2)^{1/2}(\tilde{\lambda}_j/\lambda_j - 1)$ is $N(0, 1)$. Hence, the variance of $B$ is approximately given by $\mathrm{Var}(\tilde{\lambda}_j/\lambda_j) = 2/n$; that is approximately equal to the variance of $A$.

Next, we considered a non-Gaussian case. Independent pseudorandom observations were generated from a $d$-variate $t$-distribution, $t_d(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$, with mean zero, covariance matrix $\boldsymbol{\Sigma}$ and degree of freedom $\nu = 15$. We considered the case that $\lambda_1 = d^{2/3}$, $\lambda_2 = d^{1/3}$ and $\lambda_3 = \cdots = \lambda_d = 1$ in (1) as before. We fixed the sample size as $n = 60$. We set the dimension

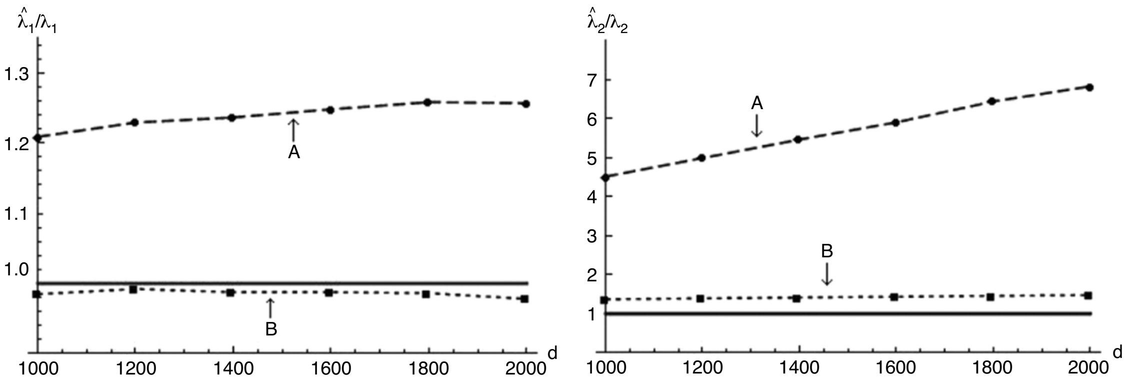


**Fig. 4.** The behaviors of $A$: $\hat{\lambda}_j/\lambda_j$ and $B$: $\tilde{\lambda}_j/\lambda_j$ for the first eigenvalue (left panel) and second eigenvalue (right panel) when the samples, of size $n = 60$, were taken from $t_d(\mathbf{0}, \mathbf{\Sigma}, \nu)$ with $\nu = 15$ and $d = 1000(200)2000$.

as $d = 1000(200)2000$. Similarly to Fig. 2, the findings were obtained by averaging the outcomes from 1000 replications. Fig. 4 shows the behaviors of two quantities, $A$:$\hat{\lambda}_j/\lambda_j$ and $B$:$\tilde{\lambda}_j/\lambda_j$, for the first two eigenvalues.

Again, the cross-data-matrix methodology seems to perform much better than naive PCA. One can observe the consistency of $\tilde{\lambda}_j$ for all $d = 1000(200)2000$. We conducted simulation studies for other settings as well and verified the superiority of the cross-data-matrix methodology to naive PCA in HDLSS data situations.

## 4. PC directions with the cross-data-matrix methodology

In this section, we apply the cross-data-matrix methodology to PC direction vectors. Jung and Marron [8], and Yata and Aoshima [13] studied consistency properties of PC direction vectors in the context of naive PCA. Let $\hat{\boldsymbol{H}} = [\hat{\boldsymbol{h}}_1, \ldots, \hat{\boldsymbol{h}}_d]$ such that $\hat{\boldsymbol{H}}^T\boldsymbol{S}\hat{\boldsymbol{H}} = \hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Lambda}} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_d)$. Then, Yata and Aoshima [13] gave consistency properties of the sample eigenvectors with their population counterparts: Assume that the first $m$ population eigenvalues are distinct such as $\lambda_1 > \cdots > \lambda_m$. Then, the first $m$ sample eigenvectors are consistent in the sense that

$$\mathrm{Angle}(\hat{\boldsymbol{h}}_j, \boldsymbol{h}_j) \xrightarrow{p} 0 \tag{6}$$

under the conditions (YA-i)–(YA-ii) appeared in Section 1. If $z_{jk},\ j = 1, \ldots, d\ (k = 1, \ldots, n)$ are independent, those conditions are modified as (YA-i′)–(YA-ii′).

Now, we consider applying the cross-data-matrix methodology to the PC direction vectors. Recall that $\boldsymbol{S}_{D(i)}^2 = \boldsymbol{S}_{D(i)}\boldsymbol{S}_{D(i)}^T$ $(i = 1, 2)$. We have the eigen-decomposition of $\boldsymbol{S}_{D(i)}^2$ as $\boldsymbol{S}_{D(i)}^2 = \sum_{j=1}^n \tilde{\lambda}_j^2 \tilde{\boldsymbol{u}}_{j(i)}\tilde{\boldsymbol{u}}_{j(i)}^T$. Let us define $\tilde{\boldsymbol{h}}_{j(i)} = (n\tilde{\lambda}_j)^{-1/2}$ $\boldsymbol{X}_i\tilde{\boldsymbol{u}}_{j(i)},\ i = 1, 2$. Since the sign of each eigenvector does not match the other, we adjust the sign of $\tilde{\boldsymbol{h}}_{j(2)}$ as $\tilde{\boldsymbol{h}}_{j(2)} = \mathrm{Sign}(\tilde{\boldsymbol{h}}_{j(1)}^T\tilde{\boldsymbol{h}}_{j(2)})\tilde{\boldsymbol{h}}_{j(2)}$. After the modification, we consider $\tilde{\boldsymbol{h}}_j = (\tilde{\boldsymbol{h}}_{j(1)} + \tilde{\boldsymbol{h}}_{j(2)})/2$ as an estimate of the PC direction vector, $\boldsymbol{h}_j$. Here, we also consider a unit vector, $\tilde{\boldsymbol{h}}_{j*} = \tilde{\boldsymbol{h}}_j/\|\tilde{\boldsymbol{h}}_j\|$.

**Theorem 3.** *Assume that the first $m$ population eigenvalues are distinct such as $\lambda_1 > \cdots > \lambda_m$. Then, the first $m$ sample eigenvectors are consistent in the sense that*

$$\mathrm{Angle}(\tilde{\boldsymbol{h}}_{j*}, \boldsymbol{h}_j) \xrightarrow{p} 0 \tag{7}$$

*under the conditions:*
(i) *$d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1$;*
(ii) *$d \to \infty$ and $d^{1-\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (1/2, 1]$;*
(iii) *$d \to \infty$ and $d^{2-2\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (0, 1/2]$.*

**Corollary 4.** *Assume further in Theorem 3 that $z_{ijk},\ j = 1, \ldots, d\ (i = 1, 2;\ k = 1, \ldots, n)$ are independent. Then, the first $m$ sample eigenvectors are consistent in the sense of (7) under the conditions:*
(i) *$d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1$;*
(ii) *$d \to \infty$ and $d^{1-\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (0, 1]$.*

**Remark 6.** Suppose the assumption in Theorem 3. Then, we claim that

$$\tilde{\boldsymbol{h}}_j^T\boldsymbol{h}_j = 1 + o_p(1)$$

under the conditions (i)–(ii) of Theorem 1. Suppose the assumption in Corollary 4. Then, the above assertion is justified under the conditions (i)–(ii) of Corollary 1.

**Remark 7.** When the population eigenvalues are not distinct such as $\lambda_1 \geq \cdots \geq \lambda_m$, we can still claim a set of the results described above for some $j$ such that $\lambda_j$ has multiplicity one. When the population mean may not be zero, we still have the above results by using $\boldsymbol{S}_{oD(1)}$ defined in Corollary 2.

## 5. PC scores with the cross-data-matrix methodology

The estimation of principal component scores (Pcs) is an important issue in PCA. The $j$-th Pcs of $\boldsymbol{x}_k$ is given by $\boldsymbol{h}_j^T\boldsymbol{x}_k = z_{jk}\sqrt{\lambda_j}$ $(=s_{jk}$, say). However, since $\boldsymbol{h}_j$ is unknown, one calculates $\boldsymbol{h}_j^T\boldsymbol{x}_k$ by using an estimate of $\boldsymbol{h}_j$. In HDLSS data situations, it is very crucial for the experimenter to choose some reasonable estimate of $\boldsymbol{h}_j$. Yata and Aoshima [13] gave a sample eigenvector by $\hat{\boldsymbol{h}}_j = (n\hat{\lambda}_j)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{u}}_j$, so that the $j$-th Pcs of $\boldsymbol{x}_k$ was estimated by $\hat{\boldsymbol{h}}_j^T\boldsymbol{x}_k = \hat{u}_{jk}\sqrt{n\hat{\lambda}_j}$ $(=\hat{s}_{jk}$, say), where $\hat{\boldsymbol{u}}_j^T = (\hat{u}_{j1}, \ldots, \hat{u}_{jn})$. Note that $\hat{\boldsymbol{h}}_j$ can be calculated by using a unit-norm eigenvector, $\hat{\boldsymbol{u}}_j$, of $\boldsymbol{S}_D$ whose size is much smaller than $\boldsymbol{S}$ especially for a HDLSS data matrix. They studied the Pcs of naive PCA in terms of the sample mean square error, $\mathrm{MSE}(\hat{s}_j) = n^{-1}\sum_{k=1}^n(\hat{s}_{jk} - s_{jk})^2$, of the $j$-th Pcs: Assume that the first $m$ population eigenvalues are distinct such that $\lambda_1 > \cdots > \lambda_m$. Then, for $j = 1, \ldots, m$, it holds that

$$\frac{\mathrm{MSE}(\hat{s}_j)}{\lambda_j} = o_p(1) \tag{8}$$

under the conditions (YA-i)–(YA-ii) appeared in Section 1. If $z_{jk}$, $j = 1, \ldots, d$ $(k = 1, \ldots, n)$ are independent, those conditions are modified as (YA-i′)–(YA-ii′). By noting that $V(s_{jk}) = \lambda_j$, one may observe from (8) that the average of the normalized square error, $\lambda_j^{-1}(\hat{s}_{jk} - s_{jk})^2$, tends to zero under the convergence conditions.

Now, we consider applying the cross-data-matrix methodology to principal component scores. Suppose we have $d \times n$ data matrices, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] = [\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1n_1}, \boldsymbol{x}_{21}, \ldots, \boldsymbol{x}_{2n_2}]$, where $n_1 + n_2 = n$ with $n_1 = O(n)$ and $n_2 = O(n)$. See Remark 2 about how to handle the general case that $n_1$ and $n_2$ may not be equal. Let $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i}]$, $i = 1, 2$. Let $\boldsymbol{z}_{1j}^T = (z_{j1}, \ldots, z_{jn_1})$ and $\boldsymbol{z}_{2j}^T = (z_{jn_1+1}, \ldots, z_{jn})$, $j = 1, \ldots, d$. Recall that $\tilde{\boldsymbol{u}}_{j(1)}$ (or $\tilde{\boldsymbol{u}}_{j(2)}$) is a unit left- (or right-) singular vector corresponding to the singular value $\tilde{\lambda}_j$ $(j = 1, \ldots, n)$ of $\boldsymbol{S}_{D(1)} = n^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2$. Note that $\tilde{\boldsymbol{u}}_{j(i)}$ is available as an eigenvector of $\boldsymbol{S}_{D(i)}^2 = \boldsymbol{S}_{D(i)}\boldsymbol{S}_{D(i)}^T$ for each $i$ $(=1, 2)$. Since the sign of each eigenvector does not match the other, we adjust the sign of $\tilde{\boldsymbol{u}}_{j(2)}$ as $\tilde{\boldsymbol{u}}_{j(2)} = \mathrm{Sign}(\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{X}_1^T\boldsymbol{X}_2\tilde{\boldsymbol{u}}_{j(2)})\tilde{\boldsymbol{u}}_{j(2)}$. After the modification, let us write $\tilde{\boldsymbol{u}}_{j(i)}^T = (\tilde{u}_{j1(i)}, \ldots, \tilde{u}_{jn_i(i)})$, $i = 1, 2$. Then, the $j$-th Pcs of $\boldsymbol{x}_{ik}$ is estimated by $\tilde{u}_{jk(i)}\sqrt{n_i\tilde{\lambda}_j}$ $(=\tilde{s}_{jk(i)}$, say). Here, we write that $\tilde{s}_{jk(1)} = \tilde{s}_{jk}$ and $\tilde{s}_{jk(2)} = \tilde{s}_{jk+n_1}$. Let $\mathrm{MSE}(\tilde{s}_j) = n^{-1}\sum_{k=1}^n(\tilde{s}_{jk} - s_{jk})^2$. Then, we obtain the following result on the Pcs given by the cross-data-matrix methodology.

**Theorem 4.** *Assume that the first $m$ population eigenvalues are distinct such that $\lambda_1 > \cdots > \lambda_m$. Then, for $j = 1, \ldots, m$, we have that*

$$\frac{\mathrm{MSE}(\tilde{s}_j)}{\lambda_j} = o_p(1) \tag{9}$$

*under the conditions* (i)–(ii) *in Theorem* 1.

**Corollary 5.** *Assume further that $z_{jk}$, $j = 1, \ldots, d$ $(k = 1, \ldots, n)$ are independent. Then, we have* (9) *under the conditions* (i)–(ii) *in Corollary* 1.

**Remark 8.** Assume that the first $m$ population eigenvalues are distinct such that $\lambda_1 > \cdots > \lambda_m$. Then, for any $k$ $(=1, \ldots, n)$, it holds that

$$\tilde{\lambda}_j^{-1/2}\tilde{s}_{jk} = \lambda_j^{-1/2}s_{jk} + o_p(1) = z_{jk} + o_p(1) \tag{10}$$

under the conditions (i)–(ii) of Theorem 1. If the assumption in Corollary 5 is supposed, we claim (10) under the conditions (i)–(ii) of Corollary 3.

For a singular vector $\tilde{\boldsymbol{u}}_{j(i)}$ $(i = 1, 2)$, we claim the following result.

**Corollary 6.** *Suppose the assumption in Theorem 4. Then, the first $m$ eigenvectors of $\boldsymbol{S}_D$ are consistent in the sense that*

$$\mathrm{Angle}(\tilde{\boldsymbol{u}}_{j(i)}, n^{-1/2}\boldsymbol{z}_{ij}) \xrightarrow{p} 0 \tag{11}$$

*for $j = 1, \ldots, m$ $(i = 1, 2)$, under the conditions* (i)–(ii) *in Theorem* 1. *If the assumption in Corollary* 5 *is supposed, we claim* (11) *under the conditions* (i)–(ii) *in Corollary* 1.

From Corollary 6, we have that a singular vector $\tilde{\boldsymbol{u}}_{j(i)}$ is consistent with a vector of Pcs, $n^{-1/2}\boldsymbol{z}_{ij}$.

**Remark 9.** When the population eigenvalues are not distinct such as $\lambda_1 \geq \cdots \geq \lambda_m$, we can still claim a set of the results described above for some $j$ such that $\lambda_j$ has multiplicity one. When the population mean may not be zero, we still have the above results by using $\boldsymbol{S}_{oD(1)}$ defined in Corollary 2.

It should be noted that the cross-data-matrix methodology successfully relaxes the convergence condition to hold the consistency properties for the case that $\alpha_j > 1/2$.

## 6. Application

In this section, we give an application of the findings in this paper to a mixture model to classify a dataset into two clusters. We assume that the observation is sampled with mixing proportions $w_j$'s from two populations, $\Pi_1$ and $\Pi_2$, and the label of the population is missing. We consider a mixture model whose p.d.f. (or p.f.) is given by

$$f(\boldsymbol{x}) = w_1\pi_1(\boldsymbol{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + w_1\pi_2(\boldsymbol{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \tag{12}$$

where $w_j$'s are positive constants such that $w_1 + w_2 = 1$ and $\pi_i(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$'s are $d$-variate p.d.f. (or p.f.) of $\Pi_i$ having mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. Let $\boldsymbol{\mu}$ be the mean vector and let $\boldsymbol{\Sigma}$ be the covariance matrix of the mixture model. Then, we have that $\boldsymbol{\mu} = w_1\boldsymbol{\mu}_1 + w_2\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma} = w_1w_2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T + w_1\boldsymbol{\Sigma}_1 + w_2\boldsymbol{\Sigma}_2$. We assume (1) about $\boldsymbol{\Sigma}$.

Suppose we have a $d\times n$ data matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, where $\boldsymbol{x}_k$, $k = 1, \ldots, n$, are independent and identically distributed as (12). Let $\Delta = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$. Let $\lambda_{(1)}$ and $\lambda_{(2)}$ be the largest eigenvalues of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. We assume that $\lambda_{(1)}/\Delta \to 0$ and $\lambda_{(2)}/\Delta \to 0$ as $d \to \infty$. Then, one claims that

$$\frac{\lambda_1}{w_1w_2\Delta} = 1 + o(1) \quad \text{and} \quad \text{Angle}(\boldsymbol{h}_1, (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)/\Delta^{1/2}) \to 0.$$

Hence, for $s_{1k}$ (the first Pcs of $\boldsymbol{x}_k - \boldsymbol{\mu}$), we have as $d \to \infty$ that

$$\frac{s_{1k}}{\sqrt{\lambda_1}} = \frac{\boldsymbol{h}_1^T(\boldsymbol{x}_k - \boldsymbol{\mu})}{\sqrt{\lambda_1}} = \frac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T(\boldsymbol{x}_k - \boldsymbol{\mu})}{\sqrt{w_1w_2\Delta}}(1 + o(1)).$$

When $\boldsymbol{x}_k \in \Pi_i (i = 1, 2)$, we have for any $\tau > 0$ as $d \to \infty$ that

$$P(|\Delta^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T(\boldsymbol{x}_k - \boldsymbol{\mu}_i)| > \tau) \le \tau^{-2}\Delta^{-2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \le \tau^{-2}\Delta^{-1}\lambda_{(i)} \to 0$$

by using Chebyshev's inequality. Then, by noting that $\boldsymbol{\mu}_1 - \boldsymbol{\mu} = w_2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $\boldsymbol{\mu}_2 - \boldsymbol{\mu} = -w_1(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, we have as $d \to \infty$ that

$$\frac{s_{1k}}{\sqrt{\lambda_1}} = \begin{cases} \sqrt{w_2/w_1} + o_p(1) & (\boldsymbol{x}_k \in \Pi_1), \\ -\sqrt{w_1/w_2} + o_p(1) & (\boldsymbol{x}_k \in \Pi_2). \end{cases}$$

Thus, from the first Pcs $s_{1k}$, one can classify the dataset $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ into two clusters. From Theorem 4 (or Remark 8) in Section 5, the first Pcs $s_{1k}$ can be estimated by $\tilde{s}_{1k}$ effectively.

## 7. Demonstration

In this section, we demonstrate how to apply the cross-data-matrix methodology to a real dataset. We make use of gene expression data, introduced in Section 1, that are from a microarray study of prostate cancer. Refer to Singh et al. [11] and Pochet et al. [10] for details of the dataset. The dataset consisted of 12 600 ($=d$) genes and 34 ($=n$) microarrays in which there were 9 Normal Prostate and 25 Prostate Tumors. We assume the mixture model (12) for the dataset. We started with data matrix $\boldsymbol{X}$:12 600 $\times$ 34 $= [\boldsymbol{X}_1, \boldsymbol{X}_2]$. Here, we set $(n_1, n_2) = (17, 17)$ to divide the whole sample into $\boldsymbol{X}_1$:12 600 $\times$ 17 and $\boldsymbol{X}_2$:12 600 $\times$ 17. We put 4 Normal Prostate and 13 Prostate Tumor samples in $\boldsymbol{X}_1$ and the others (that is, 5 Normal Prostate and 12 Prostate Tumor samples) in $\boldsymbol{X}_2$ so as to balance one thing with another. We focused on a three-dimensional (3D) sub-space. Refer to Yata and Aoshima [12] for the intrinsic dimensionality estimation. Let us define $\boldsymbol{S}_{oD(1)} = (n_1n_2)^{-1/2}(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2)$ and $\boldsymbol{S}_{oD(2)} = \boldsymbol{S}_{oD(1)}^T$ according to Corollary 2. We calculated eigenvalues of $\boldsymbol{S}_{oD(1)}^2 = \boldsymbol{S}_{oD(1)}\boldsymbol{S}_{oD(1)}^T$ as $(\tilde{\lambda}_1^2, \tilde{\lambda}_2^2, \tilde{\lambda}_3^2, \ldots) = (3.29^2 \times 10^{16}, 1.39^2 \times 10^{16}, 8.48^2 \times 10^{14}, \ldots)$. With the help of Theorem 1, we obtained the estimates of the first three eigenvalues as $(3.29 \times 10^8,\ 1.39 \times 10^8,\ 8.48 \times 10^7)$. Next, we considered the Pcs along the lines of Section 5. Let $\boldsymbol{S}_{oD(2)}^2 = \boldsymbol{S}_{oD(2)}\boldsymbol{S}_{oD(2)}^T$. Then, we calculated the first three eigenvectors of $\boldsymbol{S}_{oD(1)}^2$ and $\boldsymbol{S}_{oD(2)}^2$ as $(\tilde{\boldsymbol{u}}_{1(1)}, \tilde{\boldsymbol{u}}_{2(1)}, \tilde{\boldsymbol{u}}_{3(1)})$ and $(\tilde{\boldsymbol{u}}_{1(2)}, \tilde{\boldsymbol{u}}_{2(2)}, \tilde{\boldsymbol{u}}_{3(2)})$, respectively. For every $j$ (=1, 2, 3), we adjusted the sign of $\tilde{\boldsymbol{u}}_{j(2)}$ by multiplying $s_j = \text{Sign}(\tilde{\boldsymbol{u}}_{j(1)}^T(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2)\tilde{\boldsymbol{u}}_{j(2)})$ as $s_j\tilde{\boldsymbol{u}}_{j(2)}$. Let $\tilde{\boldsymbol{u}}_{j(1)}^T = (\tilde{u}_{j1}, \ldots, \tilde{u}_{j17})$ and $\tilde{\boldsymbol{u}}_{j(2)}^T = (\tilde{u}_{j18}, \ldots, \tilde{u}_{j34})$ after the modification described above. Then, the $j$-th Pcs of $k$-th sample was given by $\tilde{s}_{jk} = \tilde{u}_{jk}\sqrt{\tilde{\lambda}_j n_1}$ $(k = 1, \ldots, 17)$ and $\tilde{s}_{jk} = \tilde{u}_{jk}\sqrt{\tilde{\lambda}_j n_2}$ $(k = 18, \ldots, 34)$. Fig. 5 gives the scatterplots of the first three PC scores. As observed, Normal Prostate (plotted as o) and Prostate Tumor (plotted as $\times$) samples seem to be separated clearer than in Fig. 1 that was plotted by using naive PCA. It is obvious specially on the first Pcs (PC1) line. This observation is theoretically supported by the arguments in Section 6. We observed that the superiority of new PCA, given by using the cross-data-matrix methodology, to naive PCA was remarkable in many other HDLSS situations.

**Acknowledgments**

The authors would like to thank the anonymous referees for their valuable comments and careful reading.

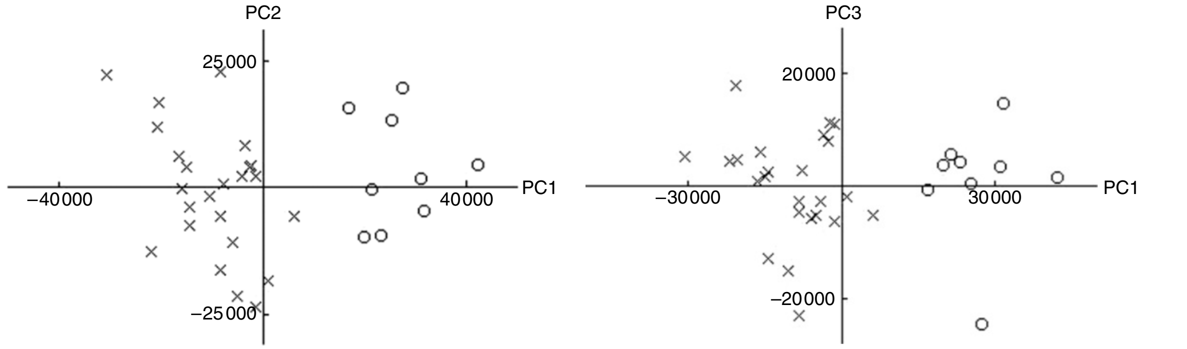


**Fig. 5.** Scatterplots of PC scores by PC1 and PC2 (left panel) or PC1 and PC3 (right panel) by using the cross-data-matrix methodology. There are 9 Normal Prostate (plotted as o) and 25 Prostate Tumors (plotted as ×).

The research of the second author was partially supported by Grant-in-Aid for Scientific Research (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 18300092.

## Appendix

Throughout this section, let $\boldsymbol{R}_n = \{\boldsymbol{e}_n \in \boldsymbol{R}^n : \|\boldsymbol{e}_n\| = 1\}$ and let $\boldsymbol{e}_{in}$, $i = 1, 2$, be arbitrary elements of $\boldsymbol{R}_n$. Let $\boldsymbol{V}_1 = n^{-1}\sum_{s=1}^{m}\lambda_s \boldsymbol{z}_{1s}\boldsymbol{z}_{2s}^T$, $\boldsymbol{V}_{2(1)} = n^{-1}\sum_{s=m+1}^{d}\lambda_s \boldsymbol{z}_{1s}\boldsymbol{z}_{2s}^T$ and $\boldsymbol{V}_{2(2)} = \boldsymbol{V}_{2(1)}^T$. Let us write $\boldsymbol{V}_{2(1)} = (v_{ij})$, where $v_{ij} = n^{-1}\sum_{s=m+1}^{d}\lambda_s z_{1si}z_{2sj}$. Let $\boldsymbol{U}_2 = n^{-1}\sum_{s=m+1}^{d}\lambda_s \boldsymbol{z}_s\boldsymbol{z}_s^T$. Let $\boldsymbol{U}_{21} = (u_{ij})$ be an $n \times n$ matrix such that

$$u_{ij} = \begin{cases} n^{-1}\displaystyle\sum_{s=m+1}^{d}\lambda_s z_{si}z_{sj} & (i \neq j), \\ 0 & (i = j). \end{cases}$$

Suppose that $\alpha_1 = \cdots = \alpha_{s_1} > \alpha_{s_1+1} = \cdots = \alpha_{s_2} > \cdots > \alpha_{s_{l-1}+1} = \cdots = \alpha_{s_l} (= \alpha_m)$, where $l \le m$. For every $i\ (=1, \ldots, l)$, let $\boldsymbol{V}_{1i} = n^{-1}\sum_{j=1}^{s_i}\lambda_j \boldsymbol{z}_{1j}\boldsymbol{z}_{2j}^T$. Let $\tilde{\lambda}_{i1} \ge \cdots \ge \tilde{\lambda}_{is_i}\ (\ge 0)$ be singular values of $\boldsymbol{V}_{1i}$. Let $\tilde{\boldsymbol{u}}_{ij(1)} \in \boldsymbol{R}_n$ be a left-singular vector and let $\tilde{\boldsymbol{u}}_{ij(2)} \in \boldsymbol{R}_n$ be a right-singular vector corresponding to $\tilde{\lambda}_{ij}$ $(j = 1, \ldots, s_i)$. Then, we have the singular value decomposition as $\boldsymbol{V}_{1i} = \sum_{j=1}^{s_i}\tilde{\lambda}_{ij}\tilde{\boldsymbol{u}}_{ij(1)}\tilde{\boldsymbol{u}}_{ij(2)}^T$. Let $\tilde{\boldsymbol{z}}_{ij} = (\|n^{-1/2}\boldsymbol{z}_{ij}\|)^{-1}n^{-1/2}\boldsymbol{z}_{ij}$ $(i = 1, 2;\ j = 1, \ldots, m)$.

The following three lemmas were obtained by Yata and Aoshima [13].

**Lemma 1.** *It holds for* $j = 1, \ldots, m$, *that* $\|d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{21}\|^2 = o_p(1)$ *under the conditions:*

(i) $d \to \infty$ *either when* $n \to \infty$ *or* $n$ *is fixed for* $j$ *such that* $\alpha_j > 1/2$;
(ii) $d \to \infty$ *and* $d^{2-2\alpha_j}/n \to 0$ *for* $j$ *such that* $\alpha_j \in (0, 1/2]$.

**Lemma 2.** *Assume that* $z_{jk}$, $j = 1, \ldots, d$ $(k = 1, \ldots, n)$ *are independent. It holds for* $\alpha_j \in (0, 1/2]$ *that* $\|d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{21}\|^2 = o_p(1)$ *under the conditions that* $d \to \infty$ *and there exists a positive constant* $\varepsilon_j$ *satisfying* $d^{1-2\alpha_j}/n < d^{-\varepsilon_j}$.

**Lemma 3.** *It holds for* $j = 1, \ldots, m$, *that* $d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{U}_2\boldsymbol{e}_{2n} = o_p(1)$ *and* $d^{-\alpha_j}n^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_2\boldsymbol{z}_{j'} = o_p(n^{-1/2})$ $(i' = 1, \ldots, m;\ j' = 1, \ldots, m)$ *under the conditions:*

(i) $d \to \infty$ *either when* $n \to \infty$ *or* $n$ *is fixed for* $j$ *such that* $\alpha_j > 1$;
(ii) $d \to \infty$ *and* $d^{2-2\alpha_j}/n \to 0$ *for* $j$ *such that* $\alpha_j \in (0, 1]$.

We will refer to the above three lemmas in the proofs of the followings.

**Lemma 4.** *It holds for* $j = 1, \ldots, m$, *that* $d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{V}_{2(1)}\boldsymbol{e}_{2n} = o_p(1)$ *under either* (i)–(ii) *of* Theorem 1 *or* (i)–(ii) *of* Corollary 1 *for the case that* $z_{ijk}$, $j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ *are independent.*

**Proof.** Let us write $\boldsymbol{V}_2 = \boldsymbol{V}_{2(1)} - \mathrm{diag}(v_{11}, \ldots, v_{nn})$. By using Chebyshev's inequality, for any $\tau > 0$, one has under either (i)–(ii) of Theorem 1 or (i)–(ii) of Corollary 1 that

$$\sum_{k=1}^{n}P\left(d^{-\alpha_j}v_{kk} > \tau\right) = \sum_{k=1}^{n}P\left((nd^{\alpha_j})^{-1}\sum_{s=m+1}^{d}\lambda_s z_{1sk}z_{2sk} > \tau\right)$$
$$\le (\tau n^{1/2}d^{\alpha_j})^{-2}\left(\sum_{s=m+1}^{d}\lambda_s^2\right) \le (\tau n^{1/2}d^{\alpha_j})^{-2}d\lambda_{m+1}^2 = O(d^{1-2\alpha_j}/n) = o(1).$$

Thus it holds that $d^{-\alpha_j} v_{kk} = o_p(1)$ for every $k$ $(=1, \ldots, n)$. From Lemmas 1 and 2, similarly to $\boldsymbol{U}_{21}$, we have that $\|d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{V}_2\|^2 = o_p(1)$, $j = 1, \ldots, m$, under either (i)–(ii) of Theorem 1 or (i)–(ii) of Corollary 1 for the case that $z_{ijk}$, $j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ are independent. Hence, we obtain that

$$d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{V}_{2(1)}\boldsymbol{e}_{2n} = d^{-\alpha_j}\left(\boldsymbol{e}_{1n}^T\boldsymbol{V}_2\boldsymbol{e}_{2n} + \boldsymbol{e}_{1n}^T\mathrm{diag}(v_{11}, \ldots, v_{nn})\boldsymbol{e}_{2n}\right) = o_p(1) \quad (j = 1, \ldots, m).$$

It concludes the result. $\square$

**Lemma 5.** *It holds for $j = 1, \ldots, m$, that*

$$d^{-\alpha_j}n^{-1}\boldsymbol{z}_{1i'}^T\boldsymbol{V}_{2(1)}\boldsymbol{z}_{2j'} = o_p(n^{-1/2}) \quad (i' = 1, \ldots, m;\ j' = 1, \ldots, m)$$

*under either* (i)–(ii) *of* Theorem 1 *or* (i)–(ii) *of* Corollary 1 *for the case that $z_{ijk}$, $j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ are independent.*

**Proof.** Let us write $d^{-\alpha_j}n^{-1}\boldsymbol{z}_{1i'}^T\boldsymbol{V}_{2(1)}\boldsymbol{z}_{2j'} = d^{-\alpha_j}\sum_{k_1,k_2} n^{-1}z_{1i'k_1}z_{2j'k_2}v_{k_1k_2}$. We first consider the case of $\alpha_j > 1/2$. Note that $E\left\{\left(\sum_{k_2=1}^n z_{2j'k_2}v_{k_1k_2}\right)^2\right\} \le Mn^{-1}\sum_{s=m+1}^d \lambda_s^2$ with the uniform bound $M$ for the fourth moments condition. Then, by using Markov's inequality and Schwarz's inequality, for any $\tau > 0$, one has under (i) of Theorem 1 (or Corollary 1) that

$$\begin{aligned}
P\left(\left|d^{-\alpha_j}\sum_{k_1,k_2} n^{-1}z_{1i'k_1}z_{2j'k_2}v_{k_1k_2}\right| > n^{-1/2}\tau\right) &\le P\left(d^{-\alpha_j}n^{-1}\sum_{k_1=1}^n |z_{1i'k_1}|\left|\sum_{k_2=1}^n z_{2j'k_2}v_{k_1k_2}\right| > n^{-1/2}\tau\right)\\
&\le \tau^{-1}d^{-\alpha_j}n^{-1/2}\sum_{k_1=1}^n E\left(|z_{1i'k_1}|\left|\sum_{k_2=1}^n z_{2j'k_2}v_{k_1k_2}\right|\right)\\
&\le \tau^{-1}d^{-\alpha_j}n^{-1/2}\sum_{k_1=1}^n\left(E(z_{1i'k_1}^2)E\left\{\left(\sum_{k_2=1}^n z_{2j'k_2}v_{k_1k_2}\right)^2\right\}\right)^{1/2} = O(d^{1/2-\alpha_j}) = o(1).
\end{aligned}$$

It concludes the result for the case of $\alpha_j > 1/2$.

Next, we consider the case of $\alpha_j \in (0, 1/2]$. Note that $E(z_{1i'k_1}^2 z_{1s_1k_1}z_{1s_2k_1}) \le M$ for $s_1, s_2 = m+1, \ldots, d$. By using Chebyshev's inequality, for any $\tau > 0$, one has under (ii) of Theorem 1 that

$$\begin{aligned}
P\left(\left|d^{-\alpha_j}\sum_{k_1,k_2} n^{-1}z_{1i'k_1}z_{2j'k_2}v_{k_1k_2}\right| > n^{-1/2}\tau\right) &\le \tau^{-2}d^{-2\alpha_j}n^{-1}E\left(\left|\sum_{k_1,k_2} z_{1i'k_1}z_{2j'k_2}v_{k_1k_2}\right|^2\right)\\
&\le \tau^{-2}d^{-2\alpha_j}n^{-1}M^2\sum_{s_1,s_2(\ge m+1)}^d \lambda_{s_1}\lambda_{s_2}\\
&= O(d^{2-2\alpha_j}/n) = o(1).
\end{aligned}$$

It concludes the result for the case of $\alpha_j \in (0, 1/2]$ under (ii) of Theorem 1.

Finally, we consider the case when $z_{ijk}$, $j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ are independent. Note that $E(z_{1i'k_1}^2 z_{1s_1k_1}z_{1s_2k_1}) = 0$ for $s_1 \ne s_2$. By using Chebyshev's inequality, for any $\tau > 0$, one has under (ii) of Corollary 1 that

$$\begin{aligned}
P\left(\left|d^{-\alpha_j}\sum_{k_1,k_2} n^{-1}z_{1i'k_1}z_{2j'k_2}v_{k_1k_2}\right| > n^{-1/2}\tau\right) &\le \tau^{-2}d^{-2\alpha_j}n^{-1}\sum_{s_1=m+1}^d \lambda_{s_1}^2\\
&= O(d^{1-2\alpha_j}/n) = o(1).
\end{aligned}$$

It concludes the result for the case of $\alpha_j \in (0, 1/2]$ under (ii) of Corollary 1. In conclusion, we obtain the results. $\square$

**Lemma 6.** *It holds for $j = 1, \ldots, m$, that*

$$\|d^{-\alpha_j}n^{-1/2}\boldsymbol{z}_{ii'}^T\boldsymbol{V}_{2(i)}\| = o_p(n^{-1/4}) \quad (i = 1, 2;\ i' = 1, \ldots, m)$$

*under either* (i)–(ii) *of* Theorem 1 *or* (i)–(ii) *of* Corollary 3 *for the case that $z_{ijk}$, $j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ are independent.*

**Proof.** When $i = 1$, we have that

$$\|d^{-\alpha_j}n^{-1/2}\boldsymbol{z}_{1i'}^T\boldsymbol{V}_{2(1)}\|^2 = d^{-2\alpha_j}\left(\sum_{k_1=1}^n n^{-1}z_{1i'k_1}^2\sum_{k_2=1}^n v_{k_1k_2}^2 + \sum_{k_1\ne k_2} n^{-1}z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^n v_{k_1k_3}v_{k_2k_3}\right). \tag{A.1}$$

We first consider the first term in (A.1). By using Markov's inequality, for any $\tau > 0$ and the uniform bound $M$ for the fourth moments condition, one has under either (i)–(ii) of Theorem 1 or (i)–(ii) of Corollary 3 that

$$P\left(d^{-2\alpha_j}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n} n^{-1}z_{1i'k_1}^2 v_{k_1k_2}^2 > n^{-1/2}\tau\right) \le \tau^{-1}d^{-2\alpha_j}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n} n^{-1/2}E(z_{1i'k_1}^2 v_{k_1k_2}^2)$$
$$= O(d^{1-2\alpha_j}/n^{1/2}) = o(1). \tag{A.2}$$

Next, we consider the second term in (A.1). Let us write $\psi_{ijk} = n^{-2}\sum_{s=m+1}^{d}\lambda_s^2 z_{1si}z_{1sj}z_{2sk}^2$ and $\omega_{ijk} = n^{-2}\sum_{s_1\neq s_2(\ge m+1)}^{d}\lambda_{s_1}\lambda_{s_2}z_{1s_1i}z_{1s_2j}z_{2s_1k}z_{2s_2k}$. Then, by using Chebyshev's inequality, for any $\tau > 0$, one has under either (i)–(ii) of Theorem 1 or (i)–(ii) of Corollary 3 that

$$P\left(d^{-2\alpha_j}\left|\sum_{k_1\neq k_2} n^{-1}z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^{n}\psi_{k_1k_2k_3}\right| > n^{-1/2}\tau\right) \le 2\tau^{-2}n^{-1}d^{-4\alpha_j}M^3\sum_{s_1,s_2(\ge m+1)}^{d}\lambda_{s_1}^2\lambda_{s_2}^2 = O(d^{2-4\alpha_j}/n) = o(1).$$

Next, we consider $\omega_{ijk}$ for the case of $\alpha_j > 1/2$. By using Markov's inequality and Schwarz's inequality, for any $\tau > 0$, one has under (i) of Theorem 1 (or Corollary 3) that

$$P\left(d^{-2\alpha_j}\left|\sum_{k_1\neq k_2} n^{-1}z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^{n}\omega_{k_1k_2k_3}\right| > n^{-1/2}\tau\right)$$
$$\le (\tau n^{1/2}d^{2\alpha_j})^{-1}\sum_{k_1\neq k_2}\left(E(z_{1i'k_1}^2 z_{1j'k_2}^2)E\left\{\left(\sum_{k_3=1}^{n}\omega_{k_1k_2k_3}\right)^2\right\}\right)^{1/2} = O(d^{1-2\alpha_j}) = o(1).$$

Finally, we consider $\omega_{ijk}$ for the case of $\alpha_j \in (0, 1/2]$. We have under (ii) of Theorem 1 that

$$P\left(d^{-2\alpha_j}\left|\sum_{k_1\neq k_2} n^{-1}z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^{n}\omega_{k_1k_2k_3}\right| > n^{-1/2}\tau\right) = O(d^{4-4\alpha_j}/n^2) = o(1).$$

On the other hand, we have that

$$P\left(d^{-2\alpha_j}\left|\sum_{k_1\neq k_2} n^{-1}z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^{n}\omega_{k_1k_2k_3}\right| > n^{-1/2}\tau\right) = O(d^{2-4\alpha_j}/n^2) = o(1)$$

under (ii) of Corollary 3 for the case that $z_{ijk},\ j = 1,\ldots,d\ (i = 1,2;\ k = 1,\ldots,n)$ are independent. Thus we claim that

$$d^{-2\alpha_j}\sum_{k_1\neq k_2} n^{-1}z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^{n}v_{k_1k_3}v_{k_2k_3} = d^{-2\alpha_j}\sum_{k_1\neq k_2} n^{-1}z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^{n}(\psi_{k_1k_2k_3}+\omega_{k_1k_2k_3})$$
$$= o_p(n^{-1/2}). \tag{A.3}$$

By combining (A.2)–(A.3) with (A.1), we conclude the result. □

**Lemma 7.** *Assume that the first $s_1$ population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_{s_1}$. Then, it holds under* (i)–(ii) *of Theorem* 1 *that*

$$\frac{\tilde{\lambda}_j}{\lambda_j} = \left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right) + o_p(n^{-1/2}) = 1 + o_p(1),$$
$$\tilde{\boldsymbol{u}}_{j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2}) \quad (i = 1,2;\ j = 1,\ldots,s_1).$$

**Proof.** By using Chebyshev's inequality, for any $\tau$ (>0) and the uniform bound $M$ for the fourth moments condition, one has as $n \to \infty$ that

$$P(|n^{-1}\boldsymbol{z}_{ij}^T\boldsymbol{z}_{ij'}| > n^{-1/4}\tau) = P\left(\left|n^{-1}\sum_{k=1}^{n}z_{ijk}z_{ij'k}\right| > n^{-1/4}\tau\right) \le \tau^{-2}Mn^{-1/2} = o(1) \quad (i = 1,2;\ j \neq j').$$

Thus we claim as $n \to \infty$ that $n^{-1}\boldsymbol{z}_{ij}^T\boldsymbol{z}_{ij'} = o_p(n^{-1/4})$ $(i = 1,2;\ j \neq j')$. Note that $\|n^{-1/2}\boldsymbol{z}_{ij}\|^2 = 1 + o_p(1)$ $(i = 1,2)$ as $n \to \infty$. Here, we have that

$$\max(\boldsymbol{e}_{1n}^T\boldsymbol{S}_{D(1)}\boldsymbol{e}_{2n}) = \tilde{\boldsymbol{u}}_{1(1)}^T\boldsymbol{S}_{D(1)}\tilde{\boldsymbol{u}}_{1(2)} = \tilde{\lambda}_1$$

with respect to any $\boldsymbol{e}_{1n}$ and $\boldsymbol{e}_{2n}$. Next, we have that

$$\max(\boldsymbol{e}_{1n}^T\boldsymbol{S}_{D(1)}\boldsymbol{e}_{2n}) = \tilde{\boldsymbol{u}}_{2(1)}^T\boldsymbol{S}_{D(1)}\tilde{\boldsymbol{u}}_{2(2)} = \tilde{\lambda}_2$$

with respect to any $\boldsymbol{e}_{1n}$ and $\boldsymbol{e}_{2n}$, provided that $\tilde{\boldsymbol{u}}_{1(1)}^T\boldsymbol{e}_{1n} = 0$ and $\tilde{\boldsymbol{u}}_{1(2)}^T\boldsymbol{e}_{2n} = 0$. Similarly, we have $\tilde{\lambda}_j, j = 1, \dots, m$.

For $\lambda_j$ ($j = 1, \dots, s_1$) that holds power $\alpha_{s_1}$, we have from Lemma 4 that $\lambda_j^{-1}\boldsymbol{e}_{1n}^T\boldsymbol{V}_{2(1)}\boldsymbol{e}_{2n} = o_p(1)$ under (i)–(ii) of Theorem 1. Then, it holds that $\lambda_1(\|n^{-1/2}\boldsymbol{z}_{11}\|)(\|n^{-1/2}\boldsymbol{z}_{21}\|) > \cdots > \lambda_m(\|n^{-1/2}\boldsymbol{z}_{1m}\|)(\|n^{-1/2}\boldsymbol{z}_{2m}\|)$ and $\lambda_{s_1}(\|n^{-1/2}\boldsymbol{z}_{1s_1}\|)(\|n^{-1/2}\boldsymbol{z}_{2s_1}\|) > \boldsymbol{e}_{1n}^T\boldsymbol{V}_{2(1)}\boldsymbol{e}_{2n}$ w.p.1. Then, it holds that

$$\begin{aligned}
\frac{\tilde{\lambda}_1}{\lambda_1} &= \tilde{\boldsymbol{u}}_{1(1)}^T\frac{\boldsymbol{S}_{D(1)}}{\lambda_1}\tilde{\boldsymbol{u}}_{1(2)} = \tilde{\boldsymbol{u}}_{1(1)}^T\left(\sum_{j=1}^m\frac{\lambda_j}{\lambda_1 n}\boldsymbol{z}_{1j}\boldsymbol{z}_{2j}^T\right)\tilde{\boldsymbol{u}}_{1(2)} + \lambda_1^{-1}\tilde{\boldsymbol{u}}_{1(1)}^T\boldsymbol{V}_{2(1)}\tilde{\boldsymbol{u}}_{1(2)} \\
&= \tilde{\boldsymbol{u}}_{1(1)}^T\left(\sum_{j=1}^m\frac{\lambda_j}{\lambda_1}\left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right)\tilde{\boldsymbol{z}}_{1j}\tilde{\boldsymbol{z}}_{2j}^T\right)\tilde{\boldsymbol{u}}_{1(2)} + o_p(1) \\
&= \left(\|n^{-1/2}\boldsymbol{z}_{11}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{21}\|\right) + o_p(1) = 1 + o_p(1).
\end{aligned} \tag{A.4}$$

Then, it holds that $\tilde{\boldsymbol{u}}_{1(i)}^T\tilde{\boldsymbol{z}}_{i1} = 1 + o_p(1)$ ($i = 1, 2$). For $i$ (=1, 2) there exists a random variable $\varepsilon_i \in [0, 1]$ and $\boldsymbol{y}_{i1} \in \boldsymbol{R}_n$ such that $\tilde{\boldsymbol{u}}_{1(i)} = \tilde{\boldsymbol{z}}_{i1}\sqrt{1-\varepsilon_i^2} + \varepsilon_i\boldsymbol{y}_{i1}$ and $\tilde{\boldsymbol{z}}_{i1}^T\boldsymbol{y}_{i1} = 0$. Here, from Lemmas 5 and 6, we have under (i)–(ii) of Theorem 1 that

$$\lambda_j^{-1}\tilde{\boldsymbol{z}}_{11}^T\boldsymbol{V}_{2(1)}\tilde{\boldsymbol{z}}_{21} = o_p(n^{-1/2}), \qquad \lambda_j^{-1}\tilde{\boldsymbol{z}}_{i1}^T\boldsymbol{V}_{2(i)}\boldsymbol{y}_{i1} = o_p(n^{-1/4}) \quad (i = 1, 2).$$

Noting that $\varepsilon_i = o_p(1),\ i = 1, 2$, it holds that $\sqrt{1-\varepsilon_1^2}\sqrt{1-\varepsilon_2^2} = 1 - \varepsilon_1^2/2 - \varepsilon_2^2/2 + o_p(\varepsilon_1^2) + o_p(\varepsilon_2^2)$. Then, we have that

$$\begin{aligned}
\frac{\tilde{\lambda}_1}{\lambda_1} &= \tilde{\boldsymbol{u}}_{1(1)}^T\left(\sum_{j=1}^m\frac{\lambda_j}{\lambda_1}\left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right)\tilde{\boldsymbol{z}}_{1j}\tilde{\boldsymbol{z}}_{2j}^T + \lambda_1^{-1}\boldsymbol{V}_{2(1)}\right)\tilde{\boldsymbol{u}}_{1(2)} \\
&= \left(\|n^{-1/2}\boldsymbol{z}_{11}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{21}\|\right) + \max_{\varepsilon_1,\varepsilon_2}\Bigg\{(-\varepsilon_1^2/2 - \varepsilon_2^2/2)\left(\|n^{-1/2}\boldsymbol{z}_{11}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{21}\|\right) \\
&\quad + o_p(\varepsilon_1 n^{-1/4}) + o_p(\varepsilon_2 n^{-1/4}) + \varepsilon_1\varepsilon_2\boldsymbol{y}_{11}^T\left(\sum_{j=2}^m\frac{\lambda_j}{\lambda_1}\left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right)\tilde{\boldsymbol{z}}_{1j}\tilde{\boldsymbol{z}}_{2j}^T\right)\boldsymbol{y}_{21} \\
&\quad + o_p(\varepsilon_1\varepsilon_2) + o_p(\varepsilon_1^2) + o_p(\varepsilon_2^2)\Bigg\} + o_p(n^{-1/2}).
\end{aligned}$$

Noting that $\varepsilon_i \in [0, 1],\ i = 1, 2$, it holds that $\varepsilon_1\varepsilon_2 \le \varepsilon_1^2/2 + \varepsilon_2^2/2$. From the fact that $(\|n^{-1/2}\boldsymbol{z}_{11}\|)(\|n^{-1/2}\boldsymbol{z}_{21}\|) > \lambda_1^{-1}\lambda_2(\|n^{-1/2}\boldsymbol{z}_{12}\|)(\|n^{-1/2}\boldsymbol{z}_{22}\|)$ w.p.1., we have under (i)–(ii) of Theorem 1 that

$$\begin{aligned}
&\max_{\varepsilon_1,\varepsilon_2}\Bigg\{(-\varepsilon_1^2/2 - \varepsilon_2^2/2)\left(\|n^{-1/2}\boldsymbol{z}_{11}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{21}\|\right) + o_p(\varepsilon_1 n^{-1/4}) + o_p(\varepsilon_2 n^{-1/4}) \\
&\quad + \varepsilon_1\varepsilon_2\boldsymbol{y}_{11}^T\left(\sum_{j=2}^m\frac{\lambda_j}{\lambda_1}\left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right)\tilde{\boldsymbol{z}}_{1j}\tilde{\boldsymbol{z}}_{2j}^T\right)\boldsymbol{y}_{21} + o_p(\varepsilon_1\varepsilon_2) + o_p(\varepsilon_1^2) + o_p(\varepsilon_2^2)\Bigg\} \\
&\le \max_{\varepsilon_1,\varepsilon_2}\Bigg\{(-\varepsilon_1^2/2 - \varepsilon_2^2/2)\left(\|n^{-1/2}\boldsymbol{z}_{11}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{21}\|\right) + o_p(\varepsilon_1 n^{-1/4}) + o_p(\varepsilon_2 n^{-1/4}) \\
&\quad + (\varepsilon_1^2/2 + \varepsilon_2^2/2)\frac{\lambda_2}{\lambda_1}\left(\|n^{-1/2}\boldsymbol{z}_{12}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{22}\|\right) + o_p(\varepsilon_1^2) + o_p(\varepsilon_2^2)\Bigg\} = o_p(n^{-1/2}),
\end{aligned}$$

so that $\varepsilon_1 = o_p(n^{-1/4})$ and $\varepsilon_2 = o_p(n^{-1/4})$. Thus we have under (i)–(ii) of Theorem 1 that

$$\frac{\tilde{\lambda}_1}{\lambda_1} = \left(\|n^{-1/2}\boldsymbol{z}_{11}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{21}\|\right) + o_p(n^{-1/2}) \tag{A.5}$$

together with that $\tilde{\boldsymbol{u}}_{1(i)}^T\tilde{\boldsymbol{z}}_{i1} = 1 + o_p(n^{-1/2})$, $\tilde{\boldsymbol{u}}_{2(i)}^T\tilde{\boldsymbol{z}}_{i1} = o_p(n^{-1/4})$ and $\tilde{\boldsymbol{u}}_{1(i)}^T\tilde{\boldsymbol{z}}_{i2} = o_p(n^{-1/4})$ for $i = 1, 2$. Now, similarly to (A.4)–(A.5), we have under (i)–(ii) of Theorem 1 that

$$\begin{aligned}\frac{\tilde{\lambda}_2}{\lambda_2} &= \tilde{\boldsymbol{u}}_{2(1)}^T\left(\sum_{j=2}^{m}\frac{\lambda_j}{\lambda_2}\left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right)\tilde{\boldsymbol{z}}_{1j}\tilde{\boldsymbol{z}}_{2j}^T\right)\tilde{\boldsymbol{u}}_{2(2)} + \lambda_2^{-1}\tilde{\boldsymbol{u}}_{2(1)}^T\boldsymbol{V}_{2(1)}\tilde{\boldsymbol{u}}_{2(2)} + o_p(n^{-1/2})\\ &= \left(\|n^{-1/2}\boldsymbol{z}_{12}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{22}\|\right) + o_p(n^{-1/2}) = 1 + o_p(1)\end{aligned}$$

together with that $\tilde{\boldsymbol{u}}_{2(i)}^T\tilde{\boldsymbol{z}}_{i2} = 1 + o_p(n^{-1/2})$, $i = 1, 2$. Similarly, we claim until $s_1$ to obtain under (i)–(ii) of Theorem 1 that

$$\begin{aligned}&\frac{\tilde{\lambda}_j}{\lambda_j} = \left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right) + o_p(n^{-1/2}) = 1 + o_p(1),\\ &\tilde{\boldsymbol{u}}_{j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2}) \quad (i = 1, 2;\ j = 1, \ldots, s_1).\end{aligned} \tag{A.6}$$

It concludes the results. □

**Lemma 8.** *Assume that the first m population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. Then, it holds under* (i)–(ii) *of Theorem* 1 *that*

$$\begin{aligned}&\frac{\tilde{\lambda}_j}{\lambda_j} = \left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right) + o_p(n^{-1/2}) = 1 + o_p(1),\\ &\tilde{\boldsymbol{u}}_{j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2}) \quad (i = 1, 2;\ j = 1, \ldots, m).\end{aligned}$$

**Proof.** First, we consider $\boldsymbol{V}_{11} = \sum_{j=1}^{s_1}\tilde{\lambda}_{1j}\tilde{\boldsymbol{u}}_{1j(1)}\tilde{\boldsymbol{u}}_{1j(2)}^T$. Similarly to Lemma 7, we obtain as $n \to \infty$ that

$$\frac{\tilde{\lambda}_{1j}}{\lambda_j} = 1 + o_p(1), \qquad \tilde{\boldsymbol{u}}_{1j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2}) \quad (i = 1, 2;\ j = 1, \ldots, s_1). \tag{A.7}$$

Next, we consider the case that $\lambda_j$ $(j = s_1 + 1, \ldots, s_2)$ holds power $\alpha_{s_2}$. Let us denote $\eta_{ij} = \lambda_j^{-1}\tilde{\boldsymbol{u}}_{1i(1)}^T\boldsymbol{V}_{2(1)}\tilde{\boldsymbol{u}}_{j(2)}$, $i = 1, \ldots, s_1$. Then, from Lemmas 4, 6 and (A.7), it holds under (i)–(ii) of Theorem 1 that

$$\eta_{ij} = \lambda_j^{-1}\tilde{\boldsymbol{z}}_{1i}^T\boldsymbol{V}_{2(1)}\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/4}) = o_p(n^{-1/4}) \quad (i = 1, \ldots, s_1;\ j = s_1 + 1, \ldots, s_2).$$

Note that $\tilde{\boldsymbol{u}}_{1i(1)}^T\tilde{\boldsymbol{z}}_{1j} = o_p(n^{-1/4})$ $(i = 1, \ldots, s_1;\ j = s_1 + 1, \ldots, s_2)$ in view of (A.7). Thus we have under (i)–(ii) of Theorem 1 that

$$\begin{aligned}\tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{S}_{D(1)}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} &= \tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{V}_{11}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} + \tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{V}_1 - \boldsymbol{V}_{11}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} + \eta_{ij}O(d^{\alpha_{s_2}-\alpha_{s_1}})\\ &= \tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{V}_{11}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} + \tilde{\boldsymbol{u}}_{1i(1)}^T\left(\sum_{s=s_1+1}^{m}\frac{\lambda_s}{nd^{\alpha_{s_1}}}\boldsymbol{z}_{1s}\boldsymbol{z}_{2s}^T\right)\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}})\\ &= \tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{V}_{11}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}}) \quad (i = 1, \ldots, s_1;\ j = s_1 + 1, \ldots, s_2).\end{aligned} \tag{A.8}$$

Hence, from (A.7) and (A.8), we obtain under (i)–(ii) of Theorem 1 that

$$\begin{aligned}\tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{S}_{D(1)}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} &= \frac{\tilde{\lambda}_j}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(1)}^T\tilde{\boldsymbol{u}}_{j(1)},\\ \tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{S}_{D(1)}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} &= \tilde{\boldsymbol{u}}_{1i(1)}^T\frac{\boldsymbol{V}_{11}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}}) = \frac{\tilde{\lambda}_{1i}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(2)}^T\tilde{\boldsymbol{u}}_{j(2)}\\ &\quad + o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}}) \quad (i = 1, \ldots, s_1;\ j = s_1 + 1, \ldots, s_2).\end{aligned}$$

Similarly, we obtain under (i)–(ii) of Theorem 1 that

$$\begin{aligned}\tilde{\boldsymbol{u}}_{1i(2)}^T\frac{\boldsymbol{S}_{D(2)}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(1)} &= \frac{\tilde{\lambda}_j}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(2)}^T\tilde{\boldsymbol{u}}_{j(2)},\\ \tilde{\boldsymbol{u}}_{1i(2)}^T\frac{\boldsymbol{S}_{D(2)}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{j(1)} &= \frac{\tilde{\lambda}_{1i}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(1)}^T\tilde{\boldsymbol{u}}_{j(1)} + o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}}) \quad (i = 1, \ldots, s_1;\ j = s_1 + 1, \ldots, s_2).\end{aligned}$$

Thus we have for every $i$ (=1, ..., $s_1$) and $j$ (=$s_1+1, \ldots, s_2$) that

$$\frac{\tilde{\lambda}_j}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(1)}^T\tilde{\boldsymbol{u}}_{j(1)} = \frac{\tilde{\lambda}_{1i}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(2)}^T\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}}), \tag{A.9}$$

$$\frac{\tilde{\lambda}_j}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(2)}^T\tilde{\boldsymbol{u}}_{j(2)} = \frac{\tilde{\lambda}_{1i}}{d^{\alpha_{s_1}}}\tilde{\boldsymbol{u}}_{1i(1)}^T\tilde{\boldsymbol{u}}_{j(1)} + o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}}). \tag{A.10}$$

From (A.6), we claim under (i)–(ii) of Theorem 1 that $\tilde{\boldsymbol{u}}_{j(i')}^T\tilde{\boldsymbol{z}}_{i'i} = o_p(1)$ $(i=1,\ldots,s_1;\ j=s_1+1,\ldots,s_2;\ i'=1,2)$. Thus it holds that $d^{-\alpha_{s_1}}\tilde{\lambda}_j = d^{-\alpha_{s_1}}\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{S}_{D(1)}\tilde{\boldsymbol{u}}_{j(2)} = o_p(1)$ for $j=s_1+1,\ldots,s_2$. Then, one has from (A.9)–(A.10) that

$$\left(\frac{\tilde{\lambda}_j}{d^{\alpha_{s_1}}} + o_p(1)\right)\tilde{\boldsymbol{u}}_{1i(i')}^T\tilde{\boldsymbol{u}}_{j(i')} = o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}}), \quad \text{i.e. } \tilde{\boldsymbol{u}}_{1i(i')}^T\tilde{\boldsymbol{u}}_{j(i')} = o_p(n^{-1/4}d^{\alpha_{s_2}-\alpha_{s_1}})$$
$$(i=1,\ldots,s_1;\ j=s_1+1,\ldots,s_2;\ i'=1,2).$$

So, we have under (i)–(ii) of Theorem 1 that

$$\tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{V}_{11}}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{j(2)} = \sum_{s=1}^{s_1}\frac{\tilde{\lambda}_{1s}}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{j(1)}^T\tilde{\boldsymbol{u}}_{1s(1)}\tilde{\boldsymbol{u}}_{1s(2)}^T\tilde{\boldsymbol{u}}_{j(2)} = o_p(n^{-1/2}d^{\alpha_{s_2}-\alpha_{s_1}}) \quad (j=s_1+1,\ldots,s_2). \tag{A.11}$$

Then, we obtain for $j=s_1+1,\ldots,s_2$, that

$$\begin{aligned}\tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{S}_{D(1)}}{\lambda_j}\tilde{\boldsymbol{u}}_{j(2)} &= \tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{V}_1-\boldsymbol{V}_{11}}{\lambda_j}\tilde{\boldsymbol{u}}_{j(2)} + \tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{V}_{2(1)}}{\lambda_j}\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/2})\\ &= \tilde{\boldsymbol{u}}_{j(1)}^T\left(\sum_{s=s_1+1}^{m}\frac{\lambda_s}{\lambda_j}\left(\|n^{-1/2}\boldsymbol{z}_{1s}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2s}\|\right)\tilde{\boldsymbol{z}}_{1s}\tilde{\boldsymbol{z}}_{2s}^T\right)\tilde{\boldsymbol{u}}_{j(2)} + \tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{V}_{2(1)}}{\lambda_j}\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/2}).\end{aligned} \tag{A.12}$$

Similarly to (A.4)–(A.5), for $j=s_1+1,\ldots,s_2$, it holds (A.6) under (i)–(ii) of Theorem 1. Then, for $\boldsymbol{V}_{12} = \sum_{j=1}^{s_2}\tilde{\lambda}_{2j}\tilde{\boldsymbol{u}}_{2j(1)}\tilde{\boldsymbol{u}}_{2j(2)}^T$, we obtain as $d\to\infty$ and $n\to\infty$ that

$$\frac{\tilde{\lambda}_{2j}}{\lambda_j} = 1 + o_p(1), \qquad \tilde{\boldsymbol{u}}_{2j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2}) \quad (i=1,2;\ j=1,\ldots,s_2). \tag{A.13}$$

As for $\lambda_j$ $(j=s_2+1,\ldots,s_3)$ that holds power $\alpha_{s_3}$, note that $\tilde{\boldsymbol{u}}_{2i(i')}^T\tilde{\boldsymbol{z}}_{i'j'} = o_p(n^{-1/4})$ $(i=1,\ldots,s_2;\ j'=s_2+1,\ldots,m;\ i'=1,2)$ in view of (A.13). Thus we have that

$$\tilde{\boldsymbol{u}}_{2i(1)}^T\frac{\boldsymbol{S}_{D(1)}}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{j(2)} = \tilde{\boldsymbol{u}}_{2i(1)}^T\frac{\boldsymbol{V}_{12}}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/4}d^{\alpha_{s_3}-\alpha_{s_2}}),$$
$$\tilde{\boldsymbol{u}}_{2i(2)}^T\frac{\boldsymbol{S}_{D(2)}}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{j(1)} = \tilde{\boldsymbol{u}}_{2i(2)}^T\frac{\boldsymbol{V}_{12}^T}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{j(1)} + o_p(n^{-1/4}d^{\alpha_{s_3}-\alpha_{s_2}}) \quad (i=1,\ldots,s_2).$$

Similarly to (A.9)–(A.10), we have for every $i$ (=1, ..., $s_2$) and $j$ (=$s_2+1,\ldots,s_3$) under (i)–(ii) of Theorem 1 that

$$\frac{\tilde{\lambda}_j}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{2i(1)}^T\tilde{\boldsymbol{u}}_{j(1)} = \frac{\tilde{\lambda}_{2i}}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{2i(2)}^T\tilde{\boldsymbol{u}}_{j(2)} + o_p(n^{-1/4}d^{\alpha_{s_3}-\alpha_{s_2}}),$$
$$\frac{\tilde{\lambda}_j}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{2i(2)}^T\tilde{\boldsymbol{u}}_{j(2)} = \frac{\tilde{\lambda}_{2i}}{d^{\alpha_{s_2}}}\tilde{\boldsymbol{u}}_{2i(1)}^T\tilde{\boldsymbol{u}}_{j(1)} + o_p(n^{-1/4}d^{\alpha_{s_3}-\alpha_{s_2}}).$$

Since it holds for $j=s_2+1,\ldots,s_3$ $(i'=1,2)$ that

$$\tilde{\boldsymbol{u}}_{2i(i')}^T\tilde{\boldsymbol{u}}_{j(i')} = \begin{cases} o_p(n^{-1/4}d^{\alpha_{s_3}-\alpha_{s_1}}) & (i=1,\ldots,s_1),\\ o_p(n^{-1/4}d^{\alpha_{s_3}-\alpha_{s_2}}) & (i=s_1+1,\ldots,s_2),\end{cases}$$

we obtain (A.6) for $j=s_2+1,\ldots,s_3$, in a way similar to (A.11)–(A.12).

As for $\lambda_j$ $(j=s_{l-1}+1,\ldots,s_l)$ that holds power $\alpha_{s_l}$ $(l\ge 4)$ as well, we can obtain (A.6). Therefore, for every $j$ (=1, ..., $m$) and $i$ (=1, 2), we claim under (i)–(ii) of Theorem 1 that

$$\frac{\tilde{\lambda}_j}{\lambda_j} = \left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right) + o_p(n^{-1/2}) = 1 + o_p(1), \qquad \tilde{\boldsymbol{u}}_{j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2}). \tag{A.14}$$

It concludes the results. □

**Lemma 9.** *Assume that the first m population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. Then, it holds that*

$$\frac{\tilde{\lambda}_j}{\lambda_j} = \left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right) + o_p(1) = 1 + o_p(1),$$
$$\tilde{\boldsymbol{u}}_{j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(1) \quad (i = 1, 2;\ j = 1, \ldots, m)$$

*under* (ii) *of* Corollary 1 *for the case that $z_{ijk},\ j = 1, \ldots, d\ (i = 1, 2;\ k = 1, \ldots, n)$ are independent.*

**Proof.** It should be noted that Lemma 6 cannot be claimed under (ii) of Corollary 1. Hence, similarly to the proof of Lemma 8, it concludes the results. □

**Lemma 10.** *Assume that the first m population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. Then, it holds that*

$$\frac{\tilde{\lambda}_j}{\lambda_j} = \left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right) + o_p(n^{-1/2}) = 1 + o_p(1),$$
$$\tilde{\boldsymbol{u}}_{j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2}) \quad (i = 1, 2;\ j = 1, \ldots, m)$$

*under* (i)–(ii) *of* Corollary 3 *for the case that $z_{ijk},\ j = 1, \ldots, d\ (i = 1, 2;\ k = 1, \ldots, n)$ are independent.*

**Proof.** Similarly to the proof of Lemma 8, it concludes the result. □

**Remark 10.** Assume that the first $m$ population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. For $\tilde{\lambda}_{i'j}$ $(i' = 1, \ldots, l;\ j = 1, \ldots, s_{i'})$ it holds as $d \to \infty$ and $n \to \infty$ that $\lambda_j^{-1}\tilde{\lambda}_{i'j} = 1 + o_p(1)$. For $\tilde{\boldsymbol{u}}_{i'j'(i)}$ and $\tilde{\boldsymbol{u}}_{j(i)}$ $(i = 1, 2;\ i' = 1, \ldots, l-1;\ j \in [s_{i'}+1, s_{i'+1}];\ j' = 1, \ldots, s_{i'})$ it holds that $\tilde{\boldsymbol{u}}_{i'j'(i)}^T\tilde{\boldsymbol{u}}_{j(i)} = o_p(d^{\alpha_j - \alpha_{j'}})$ under either (i)–(ii) of Theorem 1 or (i)–(ii) of Corollary 1 for the case that $z_{ijk},\ j = 1, \ldots, d\ (i = 1, 2;\ k = 1, \ldots, n)$ are independent.

**Remark 11.** When the population eigenvalues are not distinct such as $\lambda_1 \geq \cdots \geq \lambda_m$, we consider the case as follows: Suppose that $\lambda_1 = \cdots = \lambda_{t_1} > \lambda_{t_1+1} = \cdots = \lambda_{t_2} > \cdots > \lambda_{t_{r-1}+1} = \cdots = \lambda_{t_r}$ $(=\lambda_m)$, where $r \leq m$. We can claim that

$$\frac{\tilde{\lambda}_{t_{i'-1}+j}}{\lambda_{t_{i'-1}+j}} = 1 + o_p(1), \qquad \tilde{\boldsymbol{u}}_{t_{i'-1}+j(i)} \in \left\{\sum_{s=1}^{t_{i'}-t_{i'-1}} b_s\tilde{\boldsymbol{z}}_{it_{i'-1}+s} : \sum_{s=1}^{t_{i'}-t_{i'-1}} b_s^2 = 1\right\}$$
$$(i = 1, 2;\ i' = 1, \ldots, r;\ j = 1, \ldots, t_{i'} - t_{i'-1}),$$

where $t_0 = 0$, under either (i)–(ii) of Theorem 1 or (i)–(ii) of Corollary 1 for the case that $z_{ijk},\ j = 1, \ldots, d\ (i = 1, 2;\ k = 1, \ldots, n)$ are independent. Then, it holds that

$$\mathrm{Angle}(\tilde{\boldsymbol{u}}_{t_{i'-1}+j(i)},\ \mathrm{span}\{\tilde{\boldsymbol{z}}_{it_{i'-1}+1}, \ldots, \tilde{\boldsymbol{z}}_{it_{i'}}\}) \xrightarrow{p} 0$$
$$(i = 1, 2;\ i' = 1, \ldots, r;\ j = 1, \ldots, t_{i'} - t_{i'-1}).$$

**Proof of Theorem 1.** The result is obtained straightforwardly by combining Lemma 8 with Remark 11. □

**Proof of Theorem 2.** We use the Taylor expansion to claim that

$$\|n^{-1/2}\boldsymbol{z}_{ij}\| = 1 + \frac{1}{2}\left(\|n^{-1/2}\boldsymbol{z}_{ij}\|^2 - 1\right) - \frac{1}{8}\varepsilon_{ij}^{-3/2}\left(\|n^{-1/2}\boldsymbol{z}_{ij}\|^2 - 1\right)^2 \tag{A.15}$$

with suitable random variable $\varepsilon_{ij}$ between 1 and $\|n^{-1/2}\boldsymbol{z}_{ij}\|^2$. Noting that $\|n^{-1/2}\boldsymbol{z}_{ij}\|^2 = 1 + o_p(1)$ as $n \to \infty$, one has $\varepsilon_{ij} = 1 + o_p(1)$. By using Markov's inequality, for any $\tau$ $(>0)$ and the uniform bound $M$ for the fourth moments condition, one has as $n \to \infty$ that

$$P\left((\|n^{-1/2}\boldsymbol{z}_{ij}\|^2 - 1)^2 > n^{-1/2}\tau\right) = O(n^{-1/2}) = o(1) \quad (i = 1, 2). \tag{A.16}$$

By combining (A.15) with (A.16), we have as $n \to \infty$ that

$$\|n^{-1/2}\boldsymbol{z}_{ij}\| = 1 + \frac{1}{2}\left(\|n^{-1/2}\boldsymbol{z}_{ij}\|^2 - 1\right) + o_p(n^{-1/2}).$$

Noting that $(\|n^{-1/2}\boldsymbol{z}_{1j}\|^2 - 1)(\|n^{-1/2}\boldsymbol{z}_{2j}\|^2 - 1) = o_p(n^{-1/2})$ as $n \to \infty$, we claim that

$$\left(\|n^{-1/2}\boldsymbol{z}_{1j}\|\right)\left(\|n^{-1/2}\boldsymbol{z}_{2j}\|\right) - 1 = \frac{1}{2}\left(\|n^{-1/2}\boldsymbol{z}_{1j}\|^2 - 1 + \|n^{-1/2}\boldsymbol{z}_{2j}\|^2 - 1\right) + o_p(n^{-1/2}). \tag{A.17}$$

Recall that $V(z_{ijk}^2) = M_j$ $(j = 1, \ldots, m)$. For each $i$, by using the central limiting theorem, one has as $n \to \infty$ that $(nM_j)^{-1/2}(\sum_{k=1}^n z_{ijk}^2 - n) \Rightarrow N(0, 1)$. Note that $\|n^{-1/2}\boldsymbol{z}_{1j}\|$ and $\|n^{-1/2}\boldsymbol{z}_{2j}\|$ are independent. Thus by combining (A.14) with (A.17), we have under (i)–(ii) of Theorem 1 that

$$\sqrt{\frac{2n}{M_j}}\left(\frac{\tilde{\lambda}_j}{\lambda_j} - 1\right) \Rightarrow N(0, 1) \quad (j = 1, \ldots, m). \tag{A.18}$$

It concludes the result. $\square$

**Proof of Corollary 1.** The result is obtained from Lemmas 9 and 10 and Remark 11 straightforwardly. $\square$

**Proof of Corollary 2.** Let us write $\boldsymbol{\Lambda}^{-1/2}\boldsymbol{H}^T(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i) = [\acute{\boldsymbol{z}}_{i1}, \ldots, \acute{\boldsymbol{z}}_{id}]^T$ and $\acute{\boldsymbol{z}}_{ij} = (\acute{z}_{ij1}, \ldots, \acute{z}_{ijn})^T$ for $i = 1, 2$ and $j = 1, \ldots, d$. Then, we have that $\acute{z}_{ijk} = z_{ijk} - \bar{z}_{ij}$ for $k = 1, \ldots, n$, where $\bar{z}_{ij} = \sum_{k=1}^n z_{ijk}/n$. Let $E(z_{ijk}) = \mu_j$ for $j = 1, \ldots, d$. We write that $\acute{z}_{ijk} = \ddot{z}_{ijk} + z_{oij}$, where $\ddot{z}_{ijk} = z_{ijk} - \mu_j$ and $z_{oij} = \mu_j - \bar{z}_{ij}$ $(i = 1, 2;\ j = 1, \ldots, d;\ k = 1, \ldots, n)$. Now, let us write $n$-vectors $\ddot{\boldsymbol{z}}_{ij} = (\ddot{z}_{ij1}, \ldots, \ddot{z}_{ijn})^T$ and $\boldsymbol{z}_{oij} = (z_{oij}, \ldots, z_{oij})^T$ for $i = 1, 2$ and $j = 1, \ldots, d$. Then, we can write that $(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2) = \sum_{j=1}^d \lambda_j(\ddot{\boldsymbol{z}}_{1j} + \boldsymbol{z}_{o1j})(\ddot{\boldsymbol{z}}_{2j} + \boldsymbol{z}_{o2j})^T$. Let $\boldsymbol{V}_o = n^{-1}\sum_{s=m+1}^d \lambda_s(\ddot{\boldsymbol{z}}_{1s} + \boldsymbol{z}_{o1s})(\ddot{\boldsymbol{z}}_{2s} + \boldsymbol{z}_{o2s})^T$. Let $\boldsymbol{V}_{o1} = n^{-1}\sum_{s=m+1}^d \lambda_s \ddot{\boldsymbol{z}}_{1s}\boldsymbol{z}_{o2s}^T$, $\boldsymbol{V}_{o2} = n^{-1}\sum_{s=m+1}^d \lambda_s \boldsymbol{z}_{o1s}\boldsymbol{z}_{o2s}^T$ and $\boldsymbol{V}_{o3} = n^{-1}\sum_{s=m+1}^d \lambda_s \ddot{\boldsymbol{z}}_{1s}\ddot{\boldsymbol{z}}_{2s}^T$. We first consider $\boldsymbol{V}_{o1}$. Let us write $v_{ij(1)} = n^{-1}\sum_{s=m+1}^d \lambda_s \ddot{z}_{1si} z_{o2s}$ as $(i, j)$ element of $\boldsymbol{V}_{o1}$. Then, we have as $d^{1-2\alpha_j}/n \to 0$ that $E\{n^2(d^{-\alpha_j}v_{i'j'(1)})^2\} = O(d^{1-2\alpha_j}/n) = o(1)$. Hence, for any $\tau$ $(>0)$ and the uniform bound $M$ for the fourth moments condition, it holds that

$$P\left(\sum_{i',j'}|d^{-\alpha_j}v_{i'j'(1)}|^2 > \tau\right) = o(1)$$

by using Markov's inequality. Thus we have that $d^{-2\alpha_j}\sum_{i',j'}v_{i'j'(1)}^2 = o_p(1)$. Let $\boldsymbol{e}_{in} = (e_{i1}, \ldots, e_{in})^T$ $(i = 1, 2)$, where $\sum_{k=1}^n e_{ik}^2 = 1$. Then, we obtain as $d^{1-2\alpha_j}/n \to 0$ that

$$d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{V}_{o1}\boldsymbol{e}_{2n} = d^{-\alpha_j}\sum_{i',j'}e_{1i'}e_{2j'}v_{i'j'(1)} = o_p(1)$$

for any $\boldsymbol{e}_{1n}, \boldsymbol{e}_{2n} \in \boldsymbol{R}_n$. Similarly, we claim as $d^{1-2\alpha_j}/n \to 0$ that $d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{V}_{o2}\boldsymbol{e}_{2n} = o_p(1)$. Thus we have as $d^{1-2\alpha_j}/n \to 0$ that $d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{V}_{o}\boldsymbol{e}_{2n} = d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{V}_{o3}\boldsymbol{e}_{2n} + o_p(1)$. Then, note that $\boldsymbol{V}_{o3}$ is essentially equal to $\boldsymbol{V}_{2(1)}$. Hence, we can claim the assertion in Lemma 4 by replacing $\boldsymbol{V}_{2(1)}$ with $\boldsymbol{V}_o$. Note that $n^{-1}\acute{\boldsymbol{z}}_{ij}^T\acute{\boldsymbol{z}}_{ij'} = o_p(n^{-1/4})$ $(j \neq j')$ and $\|n^{-1/2}\acute{\boldsymbol{z}}_{ij}\| = \|n^{-1/2}\ddot{\boldsymbol{z}}_{ij}\| + o_p(n^{-1/2}) = 1 + o_p(1)$ for $i = 1, 2$. Then, by replacing $\boldsymbol{S}_{D(1)}$ with $\boldsymbol{S}_{oD(1)}$, we can claim the assertions in Theorem 1 and Corollary 1. $\square$

**Proof of Corollary 3.** With the help of Lemma 10, similarly to the proof of Theorem 2, it concludes the result. $\square$

**Proof of Theorem 3.** We first consider $\boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_{j(1)}$. We claim for $j$ $(=1, \ldots, n)$ that

$$\boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_{j(1)} = (n\tilde{\lambda}_j)^{-1/2}\boldsymbol{h}_j^T\boldsymbol{X}_1\tilde{\boldsymbol{u}}_{j(1)} = \sqrt{\frac{\lambda_j}{\tilde{\lambda}_j}}\frac{\boldsymbol{z}_{1j}^T}{\sqrt{n}}\tilde{\boldsymbol{u}}_{j(1)}.$$

Then, we have (A.14) under (i)–(ii) of Theorem 1. Thus we have under (i)–(iii) of Theorem 3 that

$$\boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_{j(1)} = 1 + o_p(1). \tag{A.19}$$

Next, we consider $\|\tilde{\boldsymbol{h}}_{j(1)}\|$. Now, we can write that

$$\|\tilde{\boldsymbol{h}}_{j(1)}\|^2 = (n\tilde{\lambda}_j)^{-1}\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{X}_1^T\boldsymbol{X}_1\tilde{\boldsymbol{u}}_{j(1)} = (n\tilde{\lambda}_j)^{-1}\tilde{\boldsymbol{u}}_{j(1)}^T\sum_{s=1}^m \lambda_s\boldsymbol{z}_{1s}\boldsymbol{z}_{1s}^T\tilde{\boldsymbol{u}}_{j(1)} + \tilde{\lambda}_j^{-1}\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{U}_{2(1)}\tilde{\boldsymbol{u}}_{j(1)}, \tag{A.20}$$

where $\boldsymbol{U}_{2(1)} = n^{-1}\sum_{s=m+1}^d \lambda_s\boldsymbol{z}_{1s}\boldsymbol{z}_{1s}^T$. First, we consider the second term in (A.20). From Lemma 3, we have for $i = 1, \ldots, m$, that

$$n^{-1/2}d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{2(1)}\boldsymbol{e}_{2n} = o_p(1), \tag{A.21}$$

$$d^{-\alpha_j}n^{-1}\boldsymbol{z}_{1j'}^T\boldsymbol{U}_{2(1)}\boldsymbol{z}_{1j'} = o_p(1) \quad (j' = 1, \ldots, m) \tag{A.22}$$

under the conditions:

(i′) $d \to \infty$ and $n \to \infty$ for $j$ such that $\alpha_j > 1$;
(ii′) $d \to \infty$ and $d^{1-\alpha_j}/n \to 0$ for $j$ such that $\alpha_j \in (0, 1]$.

Let us write $u_{ij(1)} = n^{-1}\sum_{s=m+1}^{d}\lambda_s z_{1si}z_{1sj}$ as $(i,j)$ element of $\boldsymbol{U}_{2(1)}$. From Lemma 1, we can claim under (i)–(ii) of Theorem 1 that

$$\|d^{-\alpha_j}n^{-1/2}\boldsymbol{z}_{1j'}^T(\boldsymbol{U}_{2(1)} - \mathrm{diag}(u_{11(1)},\ldots,u_{nn(1)}))\| = o_p(1) \quad (j' = 1,\ldots,m). \tag{A.23}$$

We have that $\|d^{-\alpha_j}\eta n^{-1/2}\boldsymbol{z}_{1j'}^T\mathrm{diag}(u_{11(1)},\ldots,u_{nn(1)})\|^2 = d^{-2\alpha_j}n^{-1}\eta^2\sum_{k=1}^{n}z_{1j'k}^2u_{kk(1)}^2$, where $\eta = o(n^{-1/4})$. Here, by using Chebyshev's inequality, for any $\tau$ ($>0$) and the uniform bound $M$ for the fourth moments condition, one has as $n\to\infty$ that

$$\sum_{k=1}^{n}P(z_{1j'k}^2\eta^2 > \tau) \le n\tau^{-2}\eta^4M = o(1). \tag{A.24}$$

Thus it holds that $z_{1j'k}^2\eta^2 = o_p(1)$ for every $k$ (=1, ..., $n$). Here, by using Markov's inequality, one has under (i′)–(ii′) that

$$\begin{aligned} P\left(\sum_{k=1}^{n}d^{-2\alpha_j}n^{-1}u_{kk(1)}^2 > \tau\right) &= P\left(\sum_{k=1}^{n}d^{-2\alpha_j}n^{-3}\left(\sum_{s=m+1}^{d}\lambda_s z_{1sk}^2\right)^2 > \tau\right) \\ &= O(d^{2-2\alpha_j}/n^2) = o(1). \end{aligned} \tag{A.25}$$

Thus it holds that $\sum_{k=1}^{n}d^{-2\alpha_j}n^{-1}u_{kk(1)}^2 = o_p(1)$. By combining (A.24) with (A.25), we claim under (i′)–(ii′) that $d^{-2\alpha_j}n^{-1}\sum_{k=1}^{n}z_{1j'k}^2\eta^2u_{kk(1)}^2 = o_p(1)$. Thus we claim that

$$\|d^{-\alpha_j}\eta n^{-1/2}\boldsymbol{z}_{1j'}^T\mathrm{diag}(u_{11(1)},\ldots,u_{nn(1)})\| = o_p(1).$$

Then, from (A.23), we claim under (i′)–(ii′) that

$$d^{-\alpha_j}\eta n^{-1/2}\boldsymbol{z}_{1j'}^T\boldsymbol{U}_{2(1)}\boldsymbol{e}_{1n} = o_p(1). \tag{A.26}$$

From (A.14), under (i)–(ii) of Theorem 1, there exists a random variable $\varepsilon_j \in [0,1]$ and $\boldsymbol{y}_{1j} \in \boldsymbol{R}_n$ such that $\tilde{\boldsymbol{u}}_{j(1)} = \tilde{\boldsymbol{z}}_{1j}\sqrt{1-\varepsilon_j^2} + \varepsilon_j\boldsymbol{y}_{1j}$ and $\tilde{\boldsymbol{z}}_{1j}^T\boldsymbol{y}_{1j} = 0$, where $\varepsilon_j = o_p(n^{-1/4})$. Thus it holds from (A.21), (A.22) and (A.26) that

$$d^{-\alpha_j}\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{U}_{2(1)}\tilde{\boldsymbol{u}}_{j(1)} = d^{-\alpha_j}\tilde{\boldsymbol{z}}_{1j}^T\boldsymbol{U}_{2(1)}\tilde{\boldsymbol{z}}_{1j} + o_p(1) = o_p(1) \tag{A.27}$$

under the conditions given by combining (i′)–(ii′) with (i)–(ii) of Theorem 1 (that is, (i), (ii) and (iii) of the present theorem).

Next, we consider the first term in (A.20). With the help of Remark 10, we have under (i)–(ii) of Theorem 1 that

$$\tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{V}_{1i}\boldsymbol{V}_{1i}^T}{\lambda_j^2}\tilde{\boldsymbol{u}}_{j(1)} = \tilde{\boldsymbol{u}}_{j(1)}^T\frac{\sum_{s=1}^{s_i}\tilde{\lambda}_{is}^2\tilde{\boldsymbol{u}}_{is(1)}\tilde{\boldsymbol{u}}_{is(1)}^T}{\lambda_j^2}\tilde{\boldsymbol{u}}_{j(1)} = o_p(1)$$

for $i$ (=1, ..., $l-1$) and $j$ ($=s_i+1,\ldots,s_{i+1}$). Thus we have that

$$\begin{aligned} \tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{V}_{1i}\boldsymbol{V}_{1i}^T}{\lambda_j^2}\tilde{\boldsymbol{u}}_{j(1)} &= \tilde{\boldsymbol{u}}_{j(1)}^T\left(n^{-1}\sum_{s=1}^{s_i}\lambda_j^{-1}\lambda_s\boldsymbol{z}_{1s}\boldsymbol{z}_{2s}^T\right)\left(n^{-1}\sum_{s=1}^{s_i}\lambda_j^{-1}\lambda_s\boldsymbol{z}_{2s}\boldsymbol{z}_{1s}^T\right)\tilde{\boldsymbol{u}}_{j(1)} \\ &= \sum_{s=1}^{s_i}\left(\lambda_j^{-1}\lambda_s\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{z}_{1s}/n^{1/2}\right)^2 + o_p(1)\sum_{s,s'}\left(\lambda_j^{-1}\lambda_s\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{z}_{1s}/n^{1/2}\right)\left(\lambda_i^{-1}\lambda_{s'}\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{z}_{1s'}/n^{1/2}\right) = o_p(1). \end{aligned}$$

Note that if it holds that $\lambda_j^{-1}\lambda_s\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{z}_{1s}/n^{1/2} \neq o_p(1)$ for some $s$ (=1, ..., $s_i$), we can claim that

$$\tilde{\boldsymbol{u}}_{j(1)}^T\frac{\boldsymbol{V}_{1i}\boldsymbol{V}_{1i}^T}{\lambda_j^2}\tilde{\boldsymbol{u}}_{j(1)} \neq o_p(1).$$

Hence, we have under (i)–(ii) of Theorem 1 that

$$\lambda_j^{-1}\lambda_s\tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{z}_{1s}/n^{1/2} = o_p(1) \quad (j = s_i+1,\ldots,s_{i+1};\ s = 1,\ldots,s_i;\ i = 1,\ldots,l-1).$$

Thus from (A.14), it holds under (i)–(iii) of Theorem 3 that

$$\lambda_j^{-1}n^{-1}\tilde{\boldsymbol{u}}_{j(1)}^T\sum_{s=1}^{m}\lambda_s\boldsymbol{z}_{1s}\boldsymbol{z}_{1s}^T\tilde{\boldsymbol{u}}_{j(1)} = 1 + o_p(1). \tag{A.28}$$

Note that $\lambda_j^{-1}\tilde{\lambda}_j = 1 + o_p(1)$. By combining (A.27) and (A.28) with (A.20), we have under (i)–(iii) of Theorem 3 that

$$\|\tilde{\boldsymbol{h}}_{j(1)}\|^2 = 1 + o_p(1) \quad (j = 1,\ldots,m). \tag{A.29}$$

Thus from (A.19) and (A.29), we claim under (i)–(iii) of Theorem 3 that $\text{Angle}(\boldsymbol{h}_j, \tilde{\boldsymbol{h}}_{j(1)}) = o_p(1)$ for $j = 1, \ldots, m$. Similarly, we claim that $\text{Angle}(\boldsymbol{h}_j, \tilde{\boldsymbol{h}}_{j(2)}) = o_p(1)$ for $j = 1, \ldots, m$. Note that $\tilde{\boldsymbol{h}}_{j(1)}^T \tilde{\boldsymbol{h}}_{j(2)} = 1 + o_p(1)$ for $j = 1, \ldots, m$. Hence, it holds that $\text{Angle}(\boldsymbol{h}_j, (\tilde{\boldsymbol{h}}_{j(1)} + \tilde{\boldsymbol{h}}_{j(2)})/(\|\tilde{\boldsymbol{h}}_{j(1)} + \tilde{\boldsymbol{h}}_{j(2)}\|)) = \text{Angle}(\boldsymbol{h}_j, \tilde{\boldsymbol{h}}_{j*}) = o_p(1)$ for $i = 1, \ldots, m$. It concludes the result. □

**Proof of Corollary 4.** In view of the proof of Corollary 1 given by Yata and Aoshima [13], one can claim that $d^{-\alpha_j} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(1)} \boldsymbol{e}_{2n} = o_p(1)$ $(j = 1, \ldots, m)$ under (i)–(ii) of Corollary 4 for the case that $z_{ijk},\ j = 1, \ldots, d$ $(i = 1, 2;\ k = 1, \ldots, n)$ are independent. Then, we claim (A.27) under (i)–(ii) of Corollary 4. Hence, similarly to the proof of Theorem 3, it concludes the result. □

**Proof of Theorem 4.** For each $j$ $(=1, \ldots, n)$, let us write

$$\begin{aligned}
\text{MSE}(\tilde{s}_j) &= \lambda_j n^{-1} \sum_{k=1}^{n_1} \left( z_{jk} - \sqrt{n_1 \frac{\tilde{\lambda}_j}{\lambda_j}} \tilde{u}_{jk(1)} \right)^2 + \lambda_j n^{-1} \sum_{k=n_1+1}^{n} \left( z_{jk} - \sqrt{n_2 \frac{\tilde{\lambda}_j}{\lambda_j}} \tilde{u}_{jk-n_1(2)} \right)^2 \\
&= \lambda_j \frac{n_1}{n} \left( n_1^{-1} \sum_{k=1}^{n_1} z_{jk}^2 + \frac{\tilde{\lambda}_j}{\lambda_j} \sum_{k=1}^{n_1} \tilde{u}_{jk(1)}^2 - 2\sqrt{\frac{\tilde{\lambda}_j}{\lambda_j}} \left( n_1^{-1/2} \boldsymbol{z}_{1j}^T \boldsymbol{u}_{j(1)} \right) \right) \\
&\quad + \lambda_j \frac{n_2}{n} \left( n_2^{-1} \sum_{k=n_1+1}^{n} z_{jk}^2 + \frac{\tilde{\lambda}_j}{\lambda_j} \sum_{k=n_1+1}^{n} \tilde{u}_{jk-n_1(2)}^2 - 2\sqrt{\frac{\tilde{\lambda}_j}{\lambda_j}} \left( n_2^{-1/2} \boldsymbol{z}_{2j}^T \boldsymbol{u}_{j(2)} \right) \right).
\end{aligned}$$

We have (A.14) under (i)–(ii) of Theorem 1. The result is obtained by noting that $n_1^{-1} \sum_{k=1}^{n_1} z_{jk}^2 = 1 + o_p(1)$ and $n_2^{-1} \sum_{k=n_1+1}^{n} z_{jk}^2 = 1 + o_p(1)$ as $n \to \infty$ for each $j$ $(=1, \ldots, m)$. □

**Proof of Corollary 5.** With the help of Lemma 9, similarly to the proof of Theorem 4, it concludes the result. □

**Proof of Corollary 6.** From (A.14) and Lemma 9, it concludes the result. □

## References


[1] J. Ahn, J.S. Marron, K.M. Muller, Y.-Y. Chi, The high-dimension, low-sample-size geometric representation holds under mild conditions, Biometrika 94 (2007) 760–766.
[2] J. Baik, G. Ben Arous, S. Péché, Phase transition of the largest eigenvalue for non-null complex covariance matrices, Ann. Probab. 33 (2005) 1643–1697.
[3] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivariate Anal. 97 (2006) 1382–1408.
[4] P. Hall, J.S. Marron, A. Neeman, Geometric representation of high dimension, low sample size data, J. R. Stat. Soc. Ser. B 67 (2005) 427–444.
[5] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed., Springer, New York, 2009.
[6] I. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2001) 295–327.
[7] I. Johnstone, A. Lu, On consistency and sparsity for principal components analysis in high dimensions, J. Amer. Statist. Assoc. 104 (2009) 682–693.
[8] S. Jung, J.S. Marron, PCA consistency in high dimension, low sample size context, Ann. Statist. 37 (2009) 4104–4130.
[9] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked cavariance model, Statist. Sinica 17 (2007) 1617–1642.
[10] N. Pochet, F. De Smet, J.A. Suykens, B.L. De Moor, Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction, Bioinformatics 20 (2004) 3185–3195.
[11] D. Singh, P.G. Febbo, K. Ross, D.G. Jackson, J. Manola, C. Ladd, P. Tamayo, A.A. Renshaw, A.V. D'Amico, J.P. Richie, E.S. Lander, M. Loda, P.W. Kantoff, T.R. Golub, W.R. Sellers, Gene expression correlates of clinical prostate cancer behavior, Cancer Cell 1 (2) (2002) 203–209.
[12] K. Yata, M. Aoshima, Intrinsic dimensionality estimation of high dimension, low sample size data with $d$-asymptotics, Comm. Statist. Theory Methods 39 (2010) 1511–1521.
[13] K. Yata, M. Aoshima, PCA consistency for non-Gaussian data in high dimension, low sample size context, Comm. Statist. Theory Methods 38 (2009) 2634–2652.