PCA consistency for the power spiked model in high-dimensional settings

Kazuyoshi Yata[a], Makoto Aoshima[a,1,]

[a] *Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan*

Abstract

In this paper, we propose a general spiked model called the power spiked model in high-dimensional settings. We derive relations among the data dimension, the sample size and the high-dimensional noise structure. We first consider asymptotic properties of the conventional estimator of eigenvalues. We show that the estimator is affected by the high-dimensional noise structure directly, so that it becomes inconsistent. In order to overcome such difficulties in a high-dimensional situation, we develop new principal component analysis (PCA) methods called the noise-reduction methodology and the cross-data-matrix methodology under the power spiked model. We show that the new PCA methods can enjoy consistency properties not only for eigenvalues but also for PC directions and PC scores in high-dimensional settings.

Keywords: Cross-data-matrix methodology; HDLSS; Large p small n; Microarray data; Noise-reduction methodology.

---

1. Introduction

The high-dimension, low-sample-size (HDLSS) data situations occur in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. The asymptotic studies of this type of data are becoming increasingly relevant. The asymptotic behavior of eigenvalues of the sample covariance matrix had been studied by several references when the data dimension, d, and the sample size, n,

Email address: aoshima@math.tsukuba.ac.jp (Makoto Aoshima)

[1] Phone & Fax: +81-298-53-6501

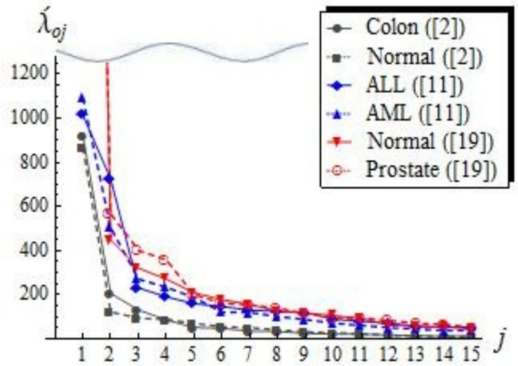


Figure 1: Estimates of the first fifteen eigenvalues for the microarray data sets. The estimates were given by the noise-reduction estimator.

increase at the same rate, i.e. $n/d \to c > 0$, under the assumption that all eigenvalues are just constants (see [7, 8, 13, 14, 17, 18]). Let us see Table 1 and Fig. 1. We observed eigenvalues for three well-known microarray data sets by using the three estimators: Conventional estimator, $\hat{\lambda}_{oj}$ (Corollary 3.3); noise-reduction estimator, $\acute{\lambda}_{oj}$ (Corollary 4.1); and cross-data-matrix estimator, $\tilde{\lambda}_{oj}$ (Corollary 5.1). The asymptotic properties of the estimators are given in Sections 3-5. The microarray data sets are as follows: Colon cancer data with 2000 genes consisting of colon tumor (40 samples) and normal colon (22 samples) given by Alon et al. [2]; leukemia data with 7129 genes consisting of ALL (47 samples) and AML (25 samples) given by Golub et al. [11]; and prostate cancer data with 12600 genes consisting of normal prostate (50 samples) and prostate tumor (52 samples) given by Singh et al. [19]. We obtained the estimates after normalizing the scale of each data set, that is, for instance, $\hat{\lambda}_{oj}$s are the sample eigenvalues of each correlation matrix. We summarized the results for the first three eigenvalues by using the three estimators in Table 1. We also visualized the first fifteen eigenvalues given by the noise-reduction estimator in Fig. 1. We observed that each n/d is quite small and the first several eigenvalues are much larger than the rest especially when d is very high. It is crucial to take the facts into account when constructing a model of eigenvalues.

In recent years, substantial work has been done on HDLSS asymptotic theory. Ahn et al. [1], Hall et al. [12], and Yata and Aoshima [22] explored several types of geometric representations on HDLSS data. Jung and Marron

Table 1: Estimates of the first three eigenvalues for the microarray data sets. The estimates were given by three methods: Conventional estimator, $\hat{\lambda}_{oj}$; noise-reduction estimator, $\acute{\lambda}_{oj}$; and cross-data-matrix estimator, $\tilde{\lambda}_{oj}$.

	n	n/d	$\hat{\lambda}_{o1},\ \hat{\lambda}_{o2},\ \hat{\lambda}_{o3}$	$\acute{\lambda}_{o1},\ \acute{\lambda}_{o2},\ \acute{\lambda}_{o3}$	$\tilde{\lambda}_{o1},\ \tilde{\lambda}_{o2},\ \tilde{\lambda}_{o3}$
Colon cancer data with 2000 (= d) genes given by Alon et al. [2]					
Colon	40	0.02	949, 228, 150	922, 205, 131	895, 194, 98
Normal	22	0.011	922, 170, 137	868, 122, 94	827, 137, 91
Two types of leukemia data with 7129 (= d) genes given by Golub et al. [11]					
ALL	47	0.0066	1148, 841, 342	1015, 724, 231	941, 685, 208
AML	25	0.0035	1344, 733, 488	1093, 504, 271	1004, 441, 260
Prostate cancer data with 12600 (= d) genes given by Singh et al. [19]					
Normal	50	0.004	6748, 561, 426	6626, 448, 320	6360, 331, 287
Prostate	52	0.0041	6095, 685, 511	5965, 566, 401	5987, 568, 370

[15] investigated consistency properties of the eigenvalues and eigenvectors of the sample covariance matrix. Yata and Aoshima [22] gave consistent estimators of both the eigenvalues and eigenvectors together with the principal component (PC) scores by a method called the *noise-reduction methodology*. The HDLSS asymptotic theory had been created under the assumption that either the population distribution is Gaussian or the random variables in a sphered data matrix have a ρ-mixing dependency. However, Yata and Aoshima [20] developed the asymptotic theory by assuming neither the Gaussian assumption nor the ρ-mixing condition. Moreover, Yata and Aoshima [21] created a new PCA called the *cross-data-matrix methodology* that provides consistent estimators of both the eigenvalues and eigenvectors together with PC scores and is applicable to constructing an unbiased estimator in nonparametric settings. Aoshima and Yata [3, 4] developed a variety of high-dimensional statistical inference based on the geometric representations by using the cross-data-matrix methodology.

In this paper, suppose we have a $d \times n$ data matrix $\boldsymbol{X}_{(d)} = [\boldsymbol{x}_{1(d)}, ..., \boldsymbol{x}_{n(d)}]$, where $\boldsymbol{x}_{j(d)} = (x_{1j(d)}, ..., x_{dj(d)})^T$, $j = 1, ..., n$, are independent and identically distributed (i.i.d.) as a d-dimensional 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, $\lambda_{1(d)} \ge \cdots \ge \lambda_{d(d)} (> 0)$, and $\boldsymbol{H}_d = [\boldsymbol{h}_{1(d)}, ..., \boldsymbol{h}_{d(d)}]$ is an orthogonal matrix of the corresponding eigenvectors. Let $\boldsymbol{X}_{(d)} = \boldsymbol{H}_d \boldsymbol{\Lambda}_d^{1/2} \boldsymbol{Z}_{(d)}$. Then, $\boldsymbol{Z}_{(d)}$ is a $d \times n$ sphered data matrix from a distribution with the identity covariance matrix. Here, we write $\boldsymbol{Z}_{(d)} = [\boldsymbol{z}_{1(d)}, ..., \boldsymbol{z}_{d(d)}]^T$ and $\boldsymbol{z}_{j(d)} = (z_{j1(d)}, ..., z_{jn(d)})^T,\ j = 1, ..., d$. Note that $E(z_{ji(d)} z_{j'i(d)}) = 0\ (j \neq j')$ and $\text{Var}(\boldsymbol{z}_{j(d)}) = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the n-dimensional identity matrix. 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. Note that if $\boldsymbol{X}$ is Gaussian, z_{ij}s are i.i.d. standard normal random variables.

We assume the following assumption as necessary:

(A-i) We consider a factor model as follows:

$$\boldsymbol{X} = \boldsymbol{\Gamma} \boldsymbol{W}, \tag{1}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_t)$ is a $d \times t$ matrix for some $t \ge d$ such that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \sum_{i=1}^t \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T = \boldsymbol{\Sigma}$, $E(w_{ij}) = 0$, $\text{Var}(w_{ij}) = 1$ and $E(w_{ij} w_{i'j}) = 0$ for $i \neq i'$. Here, w_{ij} $(i = 1, ..., t;\ j = 1, ..., n)$ is the (i, j) element of $\boldsymbol{W}$. Note that $\boldsymbol{x}_k = \sum_{i=1}^t \boldsymbol{\gamma}_i w_{ik}$. As for w_{ij}s, we assume that the fourth moments of w_{ij}s are uniformly bounded,

$$E(w_{pk}^2 w_{qk}^2) = 1 \quad \text{and} \quad E(w_{pk} w_{qk} w_{rk} w_{sk}) = 0$$

for all $p \neq q, r, s$ $(k = 1, ..., n)$.

See Bai and Saranadasa [5] and Chen and Qin [10] for the details about (1). On the other hand, Baik and Silverstein [8] and Yata and Aoshima [22] assumed that

(A-ii) $z_{jk},\ j = 1, ..., d$ $(k = 1, ..., n)$, are independent (or i.i.d. for [8]),

which includes the case that $\boldsymbol{X}$ is Gaussian. We emphasize that (A-i) is milder than (A-ii) from the fact that (1) holds when $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{W} = \boldsymbol{Z}$.

Remark 1. If there exists at least one coefficient vector $\boldsymbol{\gamma}_j$ such that $||\boldsymbol{\gamma}_j|| \to \infty$ as $d \to \infty$ in (1), it follows that $\lambda_i \to \infty$ as $d \to \infty$ for first several is, where $||\cdot||$ denotes the Euclidean norm. Namely, if there exists at least one factor having influences on many variates of $\boldsymbol{x}_k$, the first several eigenvalues become significantly large for high-dimensional data.

In Section 2, we propose a new model of eigenvalues, λ_js, called the power spiked model. In the following sections, we rigorously investigate the power spiked model. In Section 3, we show that the conventional estimator, $\hat{\lambda}_j$, is affected by the high-dimensional noise structure and the dependency of z_{ij}s directly. In Sections 4-5, we develop the noise-reduction methodology and the cross-data-matrix methodology under the power spiked model, and show that the methods can enjoy the consistency property and the asymptotic normality with respect to the eigenvalues. In Section 6, we verify performances of the methods by using simulation experiments. In Sections 7-8, we also provide consistent estimators of PC directions and PC scores and give their asymptotic properties.

2. The power spiked model

In this section, we introduce a new spiked model called the power spiked model. The sample covariance matrix is $\boldsymbol{S} = n^{-1}\boldsymbol{X}\boldsymbol{X}^T$. We consider the $n \times n$ dual sample covariance matrix defined by $\boldsymbol{S}_D = n^{-1}\boldsymbol{X}^T\boldsymbol{X}$. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n \geq 0$ be the eigenvalues of $\boldsymbol{S}_D$. Let us write the eigen-decomposition of $\boldsymbol{S}_D$ as $\boldsymbol{S}_D = \sum_{j=1}^{n} \hat{\lambda}_j \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j^T$. Note that $\boldsymbol{S}_D$ and $\boldsymbol{S}$ share non-zero eigenvalues.

Johnstone [13] considered a spiked model as follows:

$$\lambda_j \ (> 1),\ j = 1, ..., m, \text{ are fixed and } \lambda_j = 1 \ (j = m+1, ..., d). \qquad (2)$$

Here, m $(< d)$ is an unknown and fixed positive integer. Then, the asymptotic behavior of the eigenvalues of the sample covariance matrix was studied by several references when d and n increase at the same rate, i.e. $n/d \to c > 0$. See [7, 13, 14, 18] for Gaussian assumptions, and [8, 17] for non-Gaussian but i.i.d. assumptions as in (A-ii). Paul [18] also considered the asymptotic behavior of the eigenvectors and Lee et al. [17] considered that of the PC scores under (2). For the latter part in (2), the condition such as $\lambda_{m+1} = \cdots = \lambda_d = 1$ is quite strict. Under a mild condition without assuming $\lambda_{m+1} = \cdots = \lambda_d = 1$ in (2), Bai and Ding [6] considered the estimation of the forward eigenvalues. However, we note that the former part in (2) is also a strict condition since the eigenvalues probably depend on d and it is probable that $\lambda_j \to \infty$ as $d \to \infty$ for the first several js (see Table 1 and Fig. 1). We also note that the HDLSS context ($d >> n$) does not accept the convergence rate such as $n/d \to c > 0$.

Jung and Marron [15], Jung et al. [16] and Yata and Aoshima [20, 21, 22] considered different models such as $\lambda_j \to \infty$ as $d \to \infty$ for the first several

js when $d \to \infty$ while n is fixed in [15, 16, 22] or $n \to \infty$ in [20, 21, 22]. For example, Yata and Aoshima [20] considered a general spiked model:

$$\lambda_i = a_i d^{\alpha_i} \ (i = 1, ..., m) \quad \text{and} \quad \lambda_j = c_j \ (j = m+1, ..., d). \tag{3}$$

Here, $a_i(> 0)$, $c_j(> 0)$ and $\alpha_i(\alpha_1 \ge \cdots \ge \alpha_m > 0)$ are unknown constants preserving the order that $\lambda_1 \ge \cdots \ge \lambda_d$, and m $(< d)$ is an unknown and positive fixed integer. Then, Yata and Aoshima [20] showed that the sample eigenvalues are consistent under some conditions as follows: It holds for j $(\le m)$ that

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

under the conditions that

(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 $\alpha_j \in (0, 1]$.

The condition described by "$d \to \infty$ and $n \to \infty$" in (YA-i) is a mild condition in the sense that one can choose n free from d (e.g., n may be much smaller than d such as $n = \log d$). However, it should be noted that (YA-i)-(YA-ii) heavily depend on the latter (noise) part of (3) in which $\sum_{i=m+1}^{d} \lambda_i = O(d)$. Also, note that (3) does not always cover situations in which the first several eigenvalues are relatively large compared to the rest.

Now, we propose a new spiked model which develops (3) in order to study consistency properties of PCA in more extensive high-dimensional situations. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{(1)} + \boldsymbol{\Sigma}_{(2)}$, where $\boldsymbol{\Sigma}_{(1)} = \sum_{i=1}^{m} \lambda_i \boldsymbol{h}_i \boldsymbol{h}_i^T$ and $\boldsymbol{\Sigma}_{(2)} = \sum_{i=m+1}^{d} \lambda_i \boldsymbol{h}_i \boldsymbol{h}_i^T$ with some unknown and positive fixed integer m $(< d)$. Here, $\boldsymbol{\Sigma}_{(1)}$ is regarded as an intrinsic part and $\boldsymbol{\Sigma}_{(2)}$ is regarded as a noise part. Then, if there exists a positive fixed integer k_m such that

$$\lim_{d \to \infty} \frac{\text{tr}(\boldsymbol{\Sigma}_{(2)}^{k_m})}{\lambda_m^{k_m}} = 0, \tag{5}$$

we call such $\lambda_1 \ge \cdots \ge \lambda_d$ the *power spiked model*. Note that the spiked model given by (3) is one of the power spiked models when $k_m > 1/\alpha_m$. Under (5), it holds that

$$\lim_{d \to \infty} \frac{\text{tr}(\boldsymbol{\Sigma}_{(2)}^{k_m})^{1/k_m}}{\lambda_m} \ge \lim_{d \to \infty} \frac{\lambda_{m+1}}{\lambda_m} = 0.$$

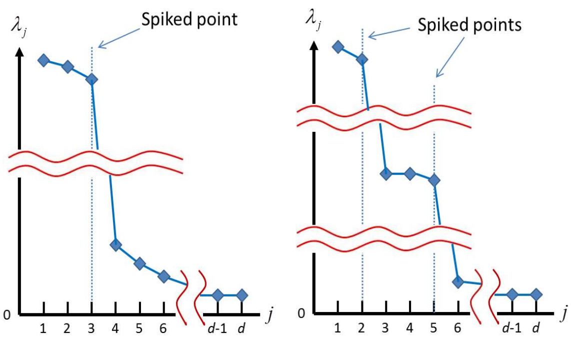


Figure 2: Illustrations of the power spiked model. The left panel has one spiked point in (5) with $m = 3$ and the right panel has two spiked points in (5) with $m = 2$ or $m = 5$.

Hence, there exists at least one spiked point such as $\lim_{d\to\infty} \lambda_{j+1}/\lambda_j = 0$ in (5). Remember that we observed a spiked point in Fig. 1 when $d = 12600$. Fig. 2 is illustrations of the spiked model.

Proposition 2.1. *If there exist some positive fixed integer j and constant α (> 0) such that*

$$\frac{\lambda_{j+1}}{\lambda_j} = O(d^{-\alpha}),$$

such eigenvalues satisfy the power spiked model given by (5) with $m = j$.

Remark 2. One of the advantages of the power spiked model is the flexibility to adapt to most high-dimensional data sets. One does not need to assume a specific function, such as $\lambda_i = a_i d^{\alpha_i}$ in (3), of d in the power spiked model. The power spiked model depends only on spiked points such as in Fig. 2. Also, one can check the validity of the parameter, k_m, in (5) by using the cross-data-matrix methodology. See Section 5.2 about the details.

Remark 3. Let us consider an interesting example such as $\lambda_1 = d$, $\lambda_2 = d^{1/2}$, $\lambda_3 = d^{1/3}, ..., \lambda_d = d^{1/d}$. From Proposition 2.1, the above example is included in the power spiked model given by (5) with any positive fixed integer m. Note that (3) cannot describe the above example.

Remark 4. If all eigenvalues are bounded such as $\liminf_{d\to\infty} \lambda_j > 0$ and $\limsup_{d\to\infty} \lambda_j < \infty$, (5) does not hold. See Bickel and Levina [9] for such a situation. However, we emphasize that the first several eigenvalues should naturally depend on d such as $\lambda_j \to \infty$ as $d \to \infty$ for high-dimensional data as observed in Table 1 and Fig. 1.

3. Asymptotic properties of the sample eigenvalues

In this section, we consider the sample eigenvalues, $\hat{\lambda}_j$s, under the power spiked model in either case of the following for j ($\le m$):

$$\textbf{(B-i)}\ \lim_{d\to\infty} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{\lambda_j^2} = 0 \quad \text{or} \quad \textbf{(B-ii)}\ \limsup_{d\to\infty} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{\lambda_j^2} > 0.$$

3.1. In case of (B-i)

We assume the following conditions as necessary:

$$\textbf{(C-i)}\ \frac{\mathrm{Var}(\sum_{s=m+1}^d \lambda_s z_{sk}^2)}{n\lambda_j^2} = \frac{\sum_{r,s=m+1}^d \lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\}}{n\lambda_j^2} = o(1);$$

$$\textbf{(C-ii)}\ \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})}{n\lambda_j} = o(1).$$

Now, we consider an asymptotic property of $\boldsymbol{S}_D$. Let us write that $\boldsymbol{S}_D = n^{-1}\sum_{s=1}^m \lambda_s \boldsymbol{z}_s\boldsymbol{z}_s^T + n^{-1}\sum_{s=m+1}^d \lambda_s \boldsymbol{z}_s\boldsymbol{z}_s^T$. The second term is the noise part. Here, by using Markov's inequality, for any $\tau > 0$ and j ($\le m$) satisfying (B-i), one has as $d \to \infty$ that $P[\sum_{k\neq k'}^n \{\sum_{s=m+1}^d \lambda_s z_{sk}z_{sk'}/(n\lambda_j)\}^2 > \tau] \le \tau^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 \to 0$, and $P[\sum_{k=1}^n \{\sum_{s=m+1}^d \lambda_s(z_{sk}^2-1)/(n\lambda_j)\}^2 > \tau] \le \tau^{-1}\mathrm{Var}(\sum_{s=m+1}^d \lambda_s z_{sk}^2)/(n\lambda_j^2) \to 0$ under (C-i). Let $\boldsymbol{e}_n = (e_1, ..., e_n)^T$ be an arbitrary (random) n-vector such that $||\boldsymbol{e}_n|| = 1$. Then, we have that

$$\begin{aligned}
\Big|\sum_{k=1}^n e_k^2 \sum_{s=m+1}^d \frac{\lambda_s(z_{sk}^2-1)}{n\lambda_j}\Big| &\le \Big\{\sum_{k=1}^n e_k^4\Big\}^{1/2}\Big\{\sum_{k=1}^n\Big(\sum_{s=m+1}^d \frac{\lambda_s(z_{sk}^2-1)}{n\lambda_j}\Big)^2\Big\}^{1/2} \\
&= o_p(1), \\
\Big|\sum_{k\neq k'}^n e_ke_{k'} \sum_{s=m+1}^d \frac{\lambda_s z_{sk}z_{sk'}}{n\lambda_j}\Big| &\le \Big\{\sum_{k\neq k'}^n e_k^2e_{k'}^2\Big\}^{1/2}\Big\{\sum_{k\neq k'}^n\Big(\sum_{s=m+1}^d \frac{\lambda_s z_{sk}z_{sk'}}{n\lambda_j}\Big)^2\Big\}^{1/2} \\
&= o_p(1)
\end{aligned}$$

from the facts that $\sum_{k=1}^n e_k^4 \le 1$ w.p.1 and $\sum_{k\neq k'}^n e_k^2e_{k'}^2 \le 1$ w.p.1. Hence, the noise part is consistent in the sense that $\boldsymbol{e}_n^T(n^{-1}\sum_{s=m+1}^d \lambda_s\boldsymbol{z}_s\boldsymbol{z}_s^T)\boldsymbol{e}_n/\lambda_j =$

$\boldsymbol{e}_n^T\{n^{-1}\sum_{s=m+1}^d \lambda_s(\boldsymbol{z}_s\boldsymbol{z}_s^T - \boldsymbol{I}_n)\}\boldsymbol{e}_n/\lambda_j + \mathrm{tr}(\boldsymbol{\Sigma}_{(2)})/(n\lambda_j) = \mathrm{tr}(\boldsymbol{\Sigma}_{(2)})/(n\lambda_j) + o_p(1)$ under (C-i), so that

$$\frac{\hat{\lambda}_j}{\lambda_j} = \hat{\boldsymbol{u}}_j^T\frac{\boldsymbol{S}_D}{\lambda_j}\hat{\boldsymbol{u}}_j = \hat{\boldsymbol{u}}_j^T\frac{\sum_{s=1}^m \lambda_s\boldsymbol{z}_s\boldsymbol{z}_s^T}{n\lambda_j}\hat{\boldsymbol{u}}_j + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})}{n\lambda_j} + o_p(1). \tag{6}$$

Then, we have the following results.

Theorem 3.1. *For j $(\le m)$ satisfying (B-i), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\frac{\hat{\lambda}_j}{\lambda_j} = 1 + \frac{tr(\boldsymbol{\Sigma}_{(2)})}{n\lambda_j} + o_p(1) \tag{7}$$

under (C-i). In addition, if (C-ii) is satisfied for j $(\le m)$, $\hat{\lambda}_j$s are consistent in the sense of (4).

Remark 5. The asymptotic property in (7) is derived from the geometric representations given by Yata and Aoshima [22]. From (7), the sample eigenvalues are inconsistent in the sense that $\limsup \hat{\lambda}_j/\lambda_j > 1$ in probability under (C-i) when $\limsup \mathrm{tr}(\boldsymbol{\Sigma}_{(2)})/(n\lambda_j) > 0$ for j $(\le m)$ satisfying (B-i).

From Lemma 1 in Appendix, under (A-i), we note that

$$\sum_{r,s=m+1}^d \lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\} = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)\}.$$

Then, for j $(\le m)$ satisfying (B-i), (C-i) holds under (A-i). Hence, we have the following result.

Corollary 3.1. *Assume (A-i). Then, for j $(\le m)$ satisfying (B-i), (4) holds as $d \to \infty$ and $n \to \infty$ under (C-ii), and (7) holds as $d \to \infty$ and $n \to \infty$.*

Let $\mathrm{Var}(z_{jk}^2) = M_j$ $(< \infty)$ for $j = 1, ..., m$. We assume for j $(\le m)$ that λ_j has multiplicity one in the following sense:

(C-iii) $\liminf_{d\to\infty}\left|\frac{\lambda_{j'}}{\lambda_j} - 1\right| > 0$ for all $j'(\ne j) = 1, ..., m$.

We also assume for j $(\le m)$ that

(C-iv) $\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})^2}{n\lambda_j^2} = o(1).$

Then, we have the following result.

Theorem 3.2. *Assume* $\liminf_{d\to\infty} M_j > 0$. *Then, for* j $(\le m)$ *satisfying (B-i), it holds as* $d \to \infty$ *and* $n \to \infty$ *that*

$$\sqrt{\frac{n}{M_j}}\Big(\frac{\hat{\lambda}_j}{\lambda_j} - 1\Big) \Rightarrow N(0,1) \tag{8}$$

under (C-iii) and (C-iv). Here, "$\Rightarrow$*" denotes the convergence in distribution and* $N(0,1)$ *denotes a random variable distributed as the standard normal distribution.*

Remark 6. Note that $\sum_{r,s=m+1}^{d} \lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\} = O\{\mathrm{tr}(\mathbf{\Sigma}_{(2)})^2\}$. Thus for j $(\le m)$ satisfying (B-i), (4) holds as $d \to \infty$ and $n \to \infty$ under (C-iv).

3.2. In case of (B-ii)

We assume the following conditions as necessary:

(C-v) $$\frac{\mathrm{Var}(\sum_{p\neq q\ge m+1}^{d} \lambda_p\lambda_q z_{pk}z_{pk'}z_{qk}z_{qk'})}{n^2\lambda_j^4} = \frac{\sum_{p\neq q, r\neq s\ge m+1}^{d} \lambda_p\lambda_q\lambda_r\lambda_s\{E(z_{pk}z_{qk}z_{rk}z_{sk})\}^2}{n^2\lambda_j^4} = o(1) \quad (k \neq k');$$

(C-vi) $$\frac{\mathrm{tr}(\mathbf{\Sigma}_{(2)}^2)^2}{n\lambda_j^4} = o(1).$$

Here, $\sum_{p\neq q, r\neq s\ge m+1}^{d}$ denotes the summation of $p, q, r, s (= m+1, ..., d)$ such that $p \neq q$, $r \neq s$. (C-v) and (C-vi) are sufficient conditions to hold (6) in case of (B-ii). See Lemmas 2 and 3 in Appendix for the details. Then, we have the following results.

Theorem 3.3. *For* j $(\le m)$ *satisfying (B-ii), (4) holds as* $d \to \infty$ *and* $n \to \infty$ *under (C-i), (C-ii), (C-v) and (C-vi), and (7) holds as* $d \to \infty$ *and* $n \to \infty$ *under (C-i), (C-v) and (C-vi).*

Remark 7. Note that $\sum_{p\neq q, r\neq s\ge m+1}^{d} \lambda_p\lambda_q\lambda_r\lambda_s\{E(z_{pk}z_{qk}z_{rk}z_{sk})\}^2 = O\{\mathrm{tr}(\mathbf{\Sigma}_{(2)})^4\}$. Thus for j $(\le m)$ satisfying (B-ii), (4) holds as $d \to \infty$ and $n \to \infty$ under (C-iv).

Theorem 3.4. *Assume* $\liminf_{d\to\infty} M_j > 0$. *Then, for* j $(\le m)$ *satisfying (B-ii), (8) holds as* $d \to \infty$ *and* $n \to \infty$ *under (C-iii) and (C-iv).*

From Lemma 1, we note that

$$\sum_{p\neq q, r\neq s \ge m+1}^{d} \lambda_p\lambda_q\lambda_r\lambda_s\{E(z_{pk}z_{qk}z_{rk}z_{sk})\}^2 = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2\}$$

under (A-i). Then, (C-v) holds under (A-i) and (C-vi). Hence, we have the following result.

Corollary 3.2. *Assume (A-i). Then, for* j $(\le m)$ *satisfying (B-ii), (4) holds as* $d \to \infty$ *and* $n \to \infty$ *under (C-ii) and (C-vi), and (7) holds as* $d \to \infty$ *and* $n \to \infty$ *under (C-vi).*

Corollary 3.3. *When the population mean may not be zero, let* $\boldsymbol{S}_{oD} = (n-1)^{-1}(\boldsymbol{X} - \overline{\boldsymbol{X}})^T(\boldsymbol{X} - \overline{\boldsymbol{X}})$ *having* $d \times n$ *matrix,* $\overline{\boldsymbol{X}} = [\bar{\boldsymbol{x}}_n, ..., \bar{\boldsymbol{x}}_n]$, *with* $\bar{\boldsymbol{x}}_n = \sum_{k=1}^n \boldsymbol{x}_k/n$. *Let* $\hat{\lambda}_{o1} \ge \cdots \ge \hat{\lambda}_{on-1} \ge 0$ *be the eigenvalues of* $\boldsymbol{S}_{oD}$ *Then, after replacing* $\hat{\lambda}_j$*s with* $\hat{\lambda}_{oj}$*s, all the results in this section are still justified.*

4. Asymptotic properties of the noise-reduction methodology

Yata and Aoshima [22] proposed a method for eigenvalues estimation called the *noise-reduction (NR) methodology* that was brought by a certain geometric representation. The NR methodology gives an estimator of λ_j by

$$\acute{\lambda}_j = \hat{\lambda}_j - \frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^{j} \hat{\lambda}_i}{n-j} \quad (j = 1, ..., n-1). \tag{9}$$

Note that $\acute{\lambda}_j \ge 0$ $(j = 1, ..., n-1)$. Then, Yata and Aoshima [22] showed that $\acute{\lambda}_j$ has several consistency properties under (3) and (A-ii). In this section, we investigate $\acute{\lambda}_j$ under the power spiked model.

Now, we consider an easy example of (5) when $m = 1$ and $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_1^2 = 0$. It holds that $\mathrm{Var}(\sum_{k=1}^n \sum_{s=m+1}^d \lambda_s z_{sk}^2/n)/\lambda_j^2 = o(1)$ under (C-i). Thus from (7) in Theorem 3.1, we have as $d \to \infty$ and $n \to \infty$ that

$$\begin{aligned}\frac{\mathrm{tr}(\boldsymbol{S}_D) - \hat{\lambda}_1}{(n-1)\lambda_1} &= \frac{\lambda_1 + \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}) - \lambda_1\{1 + \mathrm{tr}(\boldsymbol{\Sigma}_{(2)})/(n\lambda_1)\}}{(n-1)\lambda_1} + o_p(1) \\ &= \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})}{n\lambda_1} + o_p(1)\end{aligned}$$

under (C-i) from the facts that $\text{tr}(\boldsymbol{S}_D) = \sum_{k=1}^{n}\sum_{s=1}^{d}\lambda_s z_{sk}^2/n$ and $\sum_{k=1}^{n} z_{1k}^2/n = 1 + o_p(1)$. Then, from (7), it holds that

$$\frac{\acute{\lambda}_1}{\lambda_1} = \frac{\hat{\lambda}_1}{\lambda_1} - \frac{\text{tr}(\boldsymbol{\Sigma}_{(2)})}{n\lambda_1} + o_p(1) = 1 + o_p(1).$$

Contrary to Theorem 3.1, $\acute{\lambda}_1$ is consistent with λ_1 as $d \to \infty$ and $n \to \infty$ under (B-i) and (C-i), but without (C-ii).

In general, we have the following results for the NR methodology.

Theorem 4.1. *For j $(\le m)$ satisfying (B-i), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\frac{\acute{\lambda}_j}{\lambda_j} = 1 + o_p(1) \tag{10}$$

under (C-i).

Theorem 4.2. *Assume* $\liminf_{d\to\infty} M_j > 0$. *Then, for j $(\le m)$ satisfying (B-i), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\sqrt{\frac{n}{M_j}}\Big(\frac{\acute{\lambda}_j}{\lambda_j} - 1\Big) \Rightarrow N(0,1) \tag{11}$$

under (C-i) and (C-iii).

Next, for j $(\le m)$ satisfying (B-ii), we can claim the following results.

Theorem 4.3. *For j $(\le m)$ satisfying (B-ii), (10) holds as $d \to \infty$ and $n \to \infty$ under (C-i), (C-v) and (C-vi), and (11) holds as $d \to \infty$ and $n \to \infty$ under (C-i), (C-iii), (C-v), (C-vi) and* $\liminf_{d\to\infty} M_j > 0$.

Remark 8. Assume (A-i). Then, for j $(\le m)$ satisfying (B-i), (10) holds as $d \to \infty$ and $n \to \infty$, and (11) holds as $d \to \infty$ and $n \to \infty$ under (C-iii) and $\liminf_{d\to\infty} M_j > 0$. On the other hand, for j $(\le m)$ satisfying (B-ii), (10) holds as $d \to \infty$ and $n \to \infty$ under (C-vi), and (11) holds as $d \to \infty$ and $n \to \infty$ under (C-iii), (C-vi) and $\liminf_{d\to\infty} M_j > 0$.

Corollary 4.1. *When the population mean may not be zero, we define $\acute{\lambda}_{oj} = \hat{\lambda}_{oj} - \{tr(\boldsymbol{S}_{oD}) - \sum_{i=1}^{j}\hat{\lambda}_{oi}\}/(n-1-j)$ $(j = 1, ..., n-2)$, where $\boldsymbol{S}_{oD}$ and $\hat{\lambda}_{oj}$s are given in Corollary 3.3. Then, after replacing $\acute{\lambda}_j$s with $\acute{\lambda}_{oj}$s, all the results of this section are still justified.*

5. Asymptotic properties of the cross-data-matrix methodology

5.1. Eigenvalues estimation

Yata and Aoshima [21] provided another method for eigenvalues estimation called the *cross-data-matrix (CDM) methodology*. Suppose that we divide a data matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]$, into $\boldsymbol{X}_1 = [\boldsymbol{x}_{11}, ..., \boldsymbol{x}_{1n_1}]$ and $\boldsymbol{X}_2 = [\boldsymbol{x}_{21}, ..., \boldsymbol{x}_{2n_2}]$ at random with $n_1 = \lceil n/2 \rceil$ and $n_2 = n - n_1$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Note that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent. Then, Yata and Aoshima [21] defined $\boldsymbol{S}_{D(1)} = (n_1 n_2)^{-1/2} \boldsymbol{X}_1^T \boldsymbol{X}_2$ or $\boldsymbol{S}_{D(2)} = (n_1 n_2)^{-1/2} \boldsymbol{X}_2^T \boldsymbol{X}_1$ $(= \boldsymbol{S}_{D(1)}^T)$ as a cross data matrix. Note that $\boldsymbol{S}_{D(1)}$ is an $n_1 \times n_2$ matrix and $\mathrm{rank}(\boldsymbol{S}_{D(1)}) \leq n_2$. The CDM methodology gives an estimator of λ_j by the singular value, $\tilde{\lambda}_j$, of $\boldsymbol{S}_{D(1)}$. Yata and Aoshima [21] showed that $\tilde{\lambda}_j$ has several consistency properties under (3). In this section, we investigate $\tilde{\lambda}_j$ under the power spiked model.

Let $\boldsymbol{X}_i = \boldsymbol{H} \boldsymbol{\Lambda}^{1/2} \boldsymbol{Z}_i$, where $\boldsymbol{Z}_i = [\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{id}]^T$ and $\boldsymbol{z}_{ij} = (z_{ij1}, ..., z_{ijn_i})^T$, $i = 1, 2;\ j = 1, ..., d$. Then, we have $\boldsymbol{S}_{D(1)} = (n_1 n_2)^{-1/2} \sum_{j=1}^{d} \lambda_j \boldsymbol{z}_{1j} \boldsymbol{z}_{2j}^T$. When we consider the singular value decomposition of $\boldsymbol{S}_{D(1)}$, it follows that $\boldsymbol{S}_{D(1)} = \sum_{j=1}^{n_2} \tilde{\lambda}_j \tilde{\boldsymbol{u}}_{j(1)} \tilde{\boldsymbol{u}}_{j(2)}^T$, where $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_{n_2} (\geq 0)$ 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 $\tilde{\lambda}_j$ $(j = 1, ..., n_2)$. Note that $\tilde{\boldsymbol{u}}_{j(i)}$ is available as an eigenvector of $\boldsymbol{S}_{D(i)} \boldsymbol{S}_{D(i)}^T$ for each i. Then, we adjust the sign of $\tilde{\boldsymbol{u}}_{j(2)}$ by $\tilde{\boldsymbol{u}}_{j(2)} = \mathrm{Sign}(\tilde{\boldsymbol{u}}_{j(1)}^T \boldsymbol{S}_{D(1)} \tilde{\boldsymbol{u}}_{j(2)})\ \tilde{\boldsymbol{u}}_{j(2)}$.

Now, we consider an easy example when $m = 1$ and $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_1^2 = 0$. Let us write that $\lambda_1^{-1} \boldsymbol{S}_{D(1)} = (n_1 n_2)^{-1/2}\ \boldsymbol{z}_{11} \boldsymbol{z}_{21}^T + (n_1 n_2)^{-1/2} \lambda_1^{-1} \sum_{j=2}^{d} \lambda_j \boldsymbol{z}_{1j} \boldsymbol{z}_{2j}^T$. Here, by using Markov's inequality, for any $\tau > 0$, one has that

$$P\Big\{ \sum_{i'=1}^{n_1} \sum_{j'=1}^{n_2} \Big(\sum_{j=2}^{d} \frac{\lambda_j z_{1ji'} z_{2jj'}}{(n_1 n_2)^{1/2} \lambda_1} \Big)^2 > \tau \Big\} \leq \tau^{-1} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_1^2 = o(1)$$

as $d \to \infty$ when $n \to \infty$ or when n is fixed. Let $\boldsymbol{e}_{in_i} = (e_{i1}, ..., e_{in_i})^T$ be an arbitrary (random) n_i-vector such that $||\boldsymbol{e}_{in_i}|| = 1$ for $i = 1, 2$. Then, we have that

$$\begin{aligned} &\Big| \sum_{i'=1}^{n_1} \sum_{j'=1}^{n_2} e_{1i'} e_{2j'} \sum_{j=2}^{d} \frac{\lambda_j z_{1ji'} z_{2jj'}}{(n_1 n_2)^{1/2} \lambda_1} \Big| \\ &\leq \Big\{ \sum_{i'=1}^{n_1} \sum_{j'=1}^{n_2} e_{1i'}^2 e_{2j'}^2 \Big\}^{1/2} \Big\{ \sum_{i'=1}^{n_1} \sum_{j'=1}^{n_2} \Big(\frac{\lambda_j z_{1ji'} z_{2jj'}}{(n_1 n_2)^{1/2} \lambda_1} \Big)^2 \Big\}^{1/2} = o_p(1) \end{aligned}$$

from the fact that $\sum_{i'=1}^{n_1}\sum_{j'=1}^{n_2} e_{1i'}^2 e_{2j'}^2 = 1$. It follows that $\lambda_1^{-1}\boldsymbol{e}_{1n_1}^T\boldsymbol{S}_{D(1)}\boldsymbol{e}_{2n_2}$ $= (n_1n_2)^{-1/2}\boldsymbol{e}_{1n_1}^T\boldsymbol{z}_{11}\boldsymbol{z}_{21}^T\boldsymbol{e}_{2n_2} + o_p(1)$. Now, let us consider the largest singular value of $\boldsymbol{S}_{D(1)}$. Noting that $||n_i^{-1/2}\boldsymbol{z}_{ij}|| = 1 + o_p(1)$, $i = 1, 2$, as $n \to \infty$, we have as $d \to \infty$ and $n \to \infty$ that

$$\frac{\tilde{\lambda}_1}{\lambda_1} = \max_{\boldsymbol{e}_{1n_1}, \boldsymbol{e}_{2n_2}} \big\{(n_1n_2)^{-1/2}\boldsymbol{e}_{1n_1}^T\boldsymbol{z}_{11}\boldsymbol{z}_{21}^T\boldsymbol{e}_{2n_2} + o_p(1)\big\} = 1 + o_p(1).$$

Hence, the singular value, $\tilde{\lambda}_1$, is consistent with λ_1 as $d \to \infty$ and $n \to \infty$ under (B-i), but without (C-i)-(C-ii).

In general, we have the following results for the CDM methodology.

Theorem 5.1. *For j ($\le m$) satisfying (B-i), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\frac{\tilde{\lambda}_j}{\lambda_j} = 1 + o_p(1). \tag{12}$$

Theorem 5.2. *Assume* $\liminf_{d\to\infty} M_j > 0$. *Then, for j ($\le m$) satisfying (B-i), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\sqrt{\frac{n}{M_j}}\Big(\frac{\tilde{\lambda}_j}{\lambda_j} - 1\Big) \Rightarrow N(0, 1) \tag{13}$$

under (C-iii).

As for j ($\le m$) satisfying (B-ii), we can claim the following results.

Theorem 5.3. *For j ($\le m$) satisfying (B-ii), (12) holds as $d \to \infty$ and $n \to \infty$ under (C-v) and (C-vi), and (13) holds as $d \to \infty$ and $n \to \infty$ under (C-iii), (C-v), (C-vi) and* $\liminf_{d\to\infty} M_j > 0$.

Remark 9. Assume (A-i). Then, for j ($\le m$) satisfying (B-ii), (12) holds as $d \to \infty$ and $n \to \infty$ under (C-vi), and (13) holds as $d \to \infty$ and $n \to \infty$ under (C-iii), (C-vi) and $\liminf_{d\to\infty} M_j > 0$.

Corollary 5.1. *When the population mean may not be zero, let $\boldsymbol{S}_{oD(1)} = \{(n_1-1)(n_2-1)\}^{-1/2}(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2)$ having $d \times n_i$ matrices, $\overline{\boldsymbol{X}}_i = [\bar{\boldsymbol{x}}_{in}, ..., \bar{\boldsymbol{x}}_{in}]$, with $\bar{\boldsymbol{x}}_{in} = \sum_{k=1}^{n_i}\boldsymbol{x}_{ik}/n_i$, $i = 1, 2$. Let $\tilde{\lambda}_{o1} \ge \cdots \ge \tilde{\lambda}_{on_2-1} (\ge 0)$ be the singular values of $\boldsymbol{S}_{oD(1)}$. Then, after replacing $\tilde{\lambda}_j$s with $\tilde{\lambda}_{oj}$s, all the results in this section are still justified.*

5.2. Applications of the cross-data-matrix methodology

In this section, we provide some applications of the CDM methodology. It is crucial to estimate $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ in high-dimensional inference. For example, one may refer to Bai and Saranadasa [5], Chen and Qin [10], and Aoshima and Yata [3]. Aoshima and Yata [3] gave an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ such as $\mathrm{tr}(\boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T) = \sum_{i=1}^{n_2} \tilde{\lambda}_i^2$ by using the CDM methodology. Note that $E\{\mathrm{tr}(\boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T)\} = \mathrm{tr}(\boldsymbol{\Sigma}^2)$. Also, note that $\mathrm{Var}\{\mathrm{tr}(\boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T)/\mathrm{tr}(\boldsymbol{\Sigma}^2)\} \to 0$ as $d \to \infty$ and $n \to \infty$ under (A-i).

As another application, one can check whether $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 = 0$ holds or not by using the CDM methodology. We have the following proposition.

Proposition 5.1. *Assume (A-i). When it holds as $d \to \infty$ and $n \to \infty$ that*

$$\frac{tr(\boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T) - \sum_{i=1}^{j_*} \tilde{\lambda}_i^2}{\tilde{\lambda}_j^2} = o_p(1)$$

for some fixed integers j and j_* $(\geq j > 0)$, one can claim $\lim_{d\to\infty} tr(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 = 0$ with some fixed m $(\geq j)$, i.e., λ_is hold the power spiked model given by (5).

Remark 10. When the population mean may not be zero, replace $\boldsymbol{S}_{D(1)}$ and $\tilde{\lambda}_j$s with $\boldsymbol{S}_{oD(1)}$ and $\tilde{\lambda}_{oj}$s given in Corollary 5.1. Then, the result in Proposition 5.1 is still justified.

As for the three microarray data sets in Table 1, we checked whether $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 = 0$ holds or not by using Proposition 5.1 in view of Remark 10. We set $j_* = 5$. In Table 2, we calculated $\beta_j = \{\mathrm{tr}(\boldsymbol{S}_{oD(1)}\boldsymbol{S}_{oD(1)}^T) - \sum_{i=1}^{5} \tilde{\lambda}_{oi}^2\}/\tilde{\lambda}_{oj}^2,\ j = 1, 2, 3$. We observed that all β_1s and several β_2s are sufficiently small. Hence, from Proposition 5.1, it is probable that the three data sets have the power spiked model. Also, from Theorem 5.1, for j having sufficiently small β_j, one may claim that the CDM estimator, $\tilde{\lambda}_j$, is probably consistent in the sense of (12). As for the NR estimator, from Theorem 4.1, $\acute{\lambda}_j$ is probably consistent in the sense of (10) under (C-i).

6. Numerical comparisons of eigenvalue estimators

From Theorem 3.1, one needs to choose the sample size, n, depending on the noise, $\boldsymbol{\Sigma}_{(2)}$, so that the sample eigenvalue becomes a consistent estimate.

Table 2: Values of $\beta_j = \{\mathrm{tr}(\boldsymbol{S}_{oD(1)}\boldsymbol{S}_{oD(1)}^T) - \sum_{i=1}^5 \tilde{\lambda}_{oi}^2\}/\tilde{\lambda}_{oj}^2$, $j = 1, 2, 3$, for the microarray data sets in Table 1.

	n	β_1	β_2	β_3
Colon cancer data ([2]) with $d = 2000$				
Colon	40	0.00288	0.0614	0.239
Normal	22	0.00247	0.0906	0.204
Two types of leukemia data ([11]) with $d = 7129$				
ALL	47	0.0496	0.0935	1.018
AML	25	0.021	0.109	0.313
Prostate cancer data ([19]) with $d = 12600$				
Normal	50	0.00156	0.576	0.766
Prostate	52	0.00201	0.223	0.524

On the other hand, the NR methodology allows an experimenter to choose n free from the noise under (A-i) when $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 = 0$. See Theorem 4.1 or Remark 8. Moreover, the CDM methodology can claim the same argument without (A-i) (or (C-i)). See Theorem 5.1. It seems that the CDM methodology is promising to give robust estimation for HDLSS data. In this section, we examine their performances with the help of Monte Carlo simulations.

Independent pseudo-random normal observations were generated from $N_d(\mathbf{0}, \boldsymbol{\Sigma})$. Then, (A-i) holds. We considered that

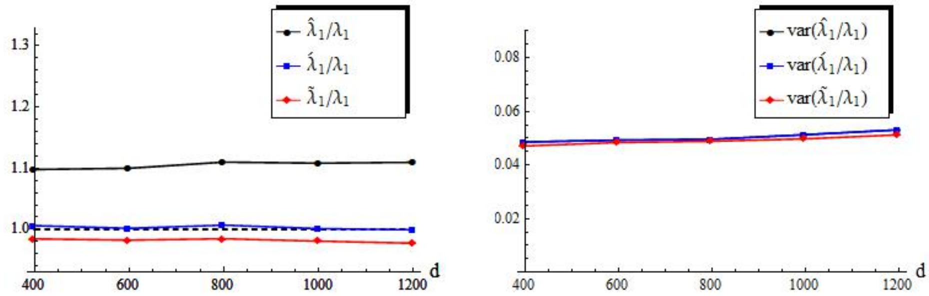
$$\boldsymbol{\Sigma}_{(1)} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, 0, ..., 0) \quad \text{and} \quad \boldsymbol{\Sigma}_{(2)} = \begin{pmatrix} \boldsymbol{O}_{3,3} & \boldsymbol{O}_{3,d-3} \\ \boldsymbol{O}_{d-3,3} & \boldsymbol{\Sigma}_* \end{pmatrix}, \tag{14}$$

where $\lambda_1 = d^{4/5}$, $\lambda_2 = d^{3/5}$, $\lambda_3 = d^{2/5}$, $\boldsymbol{\Sigma}_* = (\sigma_{ij})$ with $\sigma_{ij} = (|i-j|+1)^{-1}$ and $\boldsymbol{O}_{k,l}$ is the $k \times l$ zero matrix. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2) = O(d)$ and $\lambda_j = O(\log d)$ for some j (≥ 4). Then, this setting satisfies (5) with $m = 3$. We considered the cases of $d = 400(200)1200$. We set $n = 40$ and defined a data matrix as $\boldsymbol{X} : d \times n = [\boldsymbol{X}_1, \boldsymbol{X}_2]$ for the calculation of $\boldsymbol{S}_D$ and $\boldsymbol{S}_{D(1)}$. 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, $\hat{\lambda}_{jr}$, $\acute{\lambda}_{jr}$ and $\tilde{\lambda}_{jr}$ $(r = 1, ..., R)$. Let us simply write $\hat{\lambda}_j = R^{-1}\sum_{r=1}^{R}\hat{\lambda}_{jr}$, $\acute{\lambda}_j = R^{-1}\sum_{r=1}^{R}\acute{\lambda}_{jr}$ and $\tilde{\lambda}_j = R^{-1}\sum_{r=1}^{R}\tilde{\lambda}_{jr}$. We also considered the Monte Carlo variability. Let $\mathrm{var}(\hat{\lambda}_j/\lambda_j) = (R-1)^{-1}\sum_{r=1}^{R}(\hat{\lambda}_{jr} - \hat{\lambda}_j)^2/\lambda_j^2$, $\mathrm{var}(\acute{\lambda}_j/\lambda_j) = (R-1)^{-1}\sum_{r=1}^{R}(\acute{\lambda}_{jr} - \acute{\lambda}_j)^2/\lambda_j^2$ 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 six quantities, $(\hat{\lambda}_j/\lambda_j, \mathrm{var}(\hat{\lambda}_j/\lambda_j))$, $(\acute{\lambda}_j/\lambda_j, \mathrm{var}(\acute{\lambda}_j/\lambda_j))$ and $(\tilde{\lambda}_j/\lambda_j, \mathrm{var}(\tilde{\lambda}_j/\lambda_j))$. Fig. 3 shows the behavior of the six quantities for the first three eigenvalues. By observing the behavior of $\hat{\lambda}_j/\lambda_j$, the sample eigenvalue seems not to give a feasible estimation especially when d is very large. The sample size, $n = 40$, was not large enough to use the eigenvalues of $\boldsymbol{S}_D$ for such a high-dimensional data. On the other hand, in view of the behaviors of $\acute{\lambda}_j/\lambda_j$ and $\tilde{\lambda}_j/\lambda_j$, the NR and CDM methods give reasonable estimates surprisingly for such HDLSS data sets. The NR and CDM methods seem to perform excellently as expected for $\lambda_j,\ j = 1, 2$, that satisfy (B-i). It seems that the NR method performs a little better than the CDM method. However, for λ_3 satisfying (B-ii), those two methods do not always give such excellent performances because $n = 40$ is not large enough to claim the consistency for such a target. As for the sample variances, it seems not to make much difference among the three estimates.

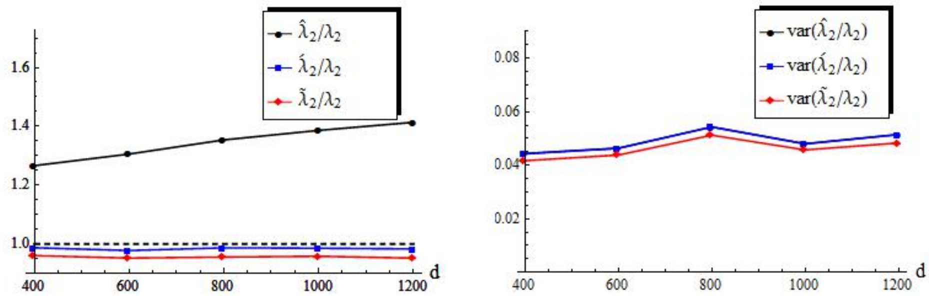
Next, we considered non-Gaussian cases that do not satisfy (A-i). Independent pseudo-random observations were generated from a d-variate t-distribution, $t_d(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$ with mean zero, covariance matrix $\boldsymbol{\Sigma}$ and degree of freedom ν. We considered $\boldsymbol{\Sigma}$ given by (14). We fixed $d = 1000$. We set the sample sizes as $n = 20(20)100$. We set $\nu = 10$ and 30. Similarly to Fig. 3, the findings were obtained by averaging the outcomes from 1000 replications. Fig. 4 shows the behaviors of three quantities, $\hat{\lambda}_j/\lambda_j$, $\acute{\lambda}_j/\lambda_j$ and $\tilde{\lambda}_j/\lambda_j$, for the first three eigenvalues. Note that $t_d(\mathbf{0}, \boldsymbol{\Sigma}, \nu) \Rightarrow N_d(\mathbf{0}, \boldsymbol{\Sigma})$ as $\nu \to \infty$. When $\nu = 10$, the NR method seems not to give a feasible estimation compared to the case of $\nu = 30$. This is probably due to the size of $\nu = 10$ is not large enough for $\boldsymbol{X}$ to satisfy (C-i). On the other hand, the CDM method does not require (A-i) (or (C-i)) in case of (B-i). As observed in Fig. 4, the CDM method seems to perform excellently even when $\nu = 10$.

7. Consistency of PC direction vectors

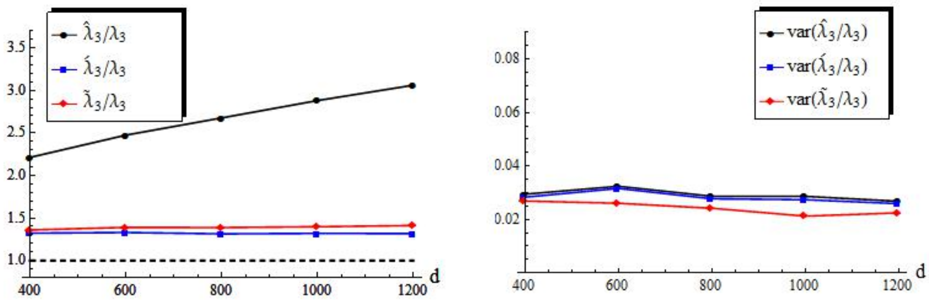
In this section, we consider PC direction vectors under the power spiked model. Jung and Marron [15] and Yata and Aoshima [20] investigated PC direction vectors in the context of conventional PCA. Let $\hat{\boldsymbol{H}} = [\hat{\boldsymbol{h}}_1, \cdots, \hat{\boldsymbol{h}}_d]$, where $\hat{\boldsymbol{H}}$ is a $d \times d$ orthogonal matrix of the sample eigenvectors such



(a) For the first eigenvalue.

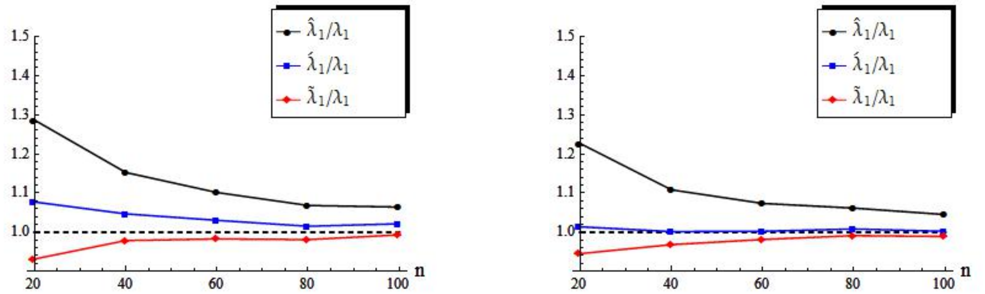


(b) For the second eigenvalue.

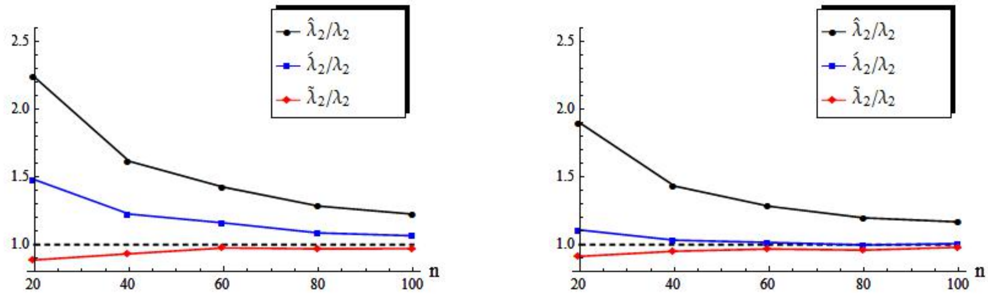


(c) For the third eigenvalue.

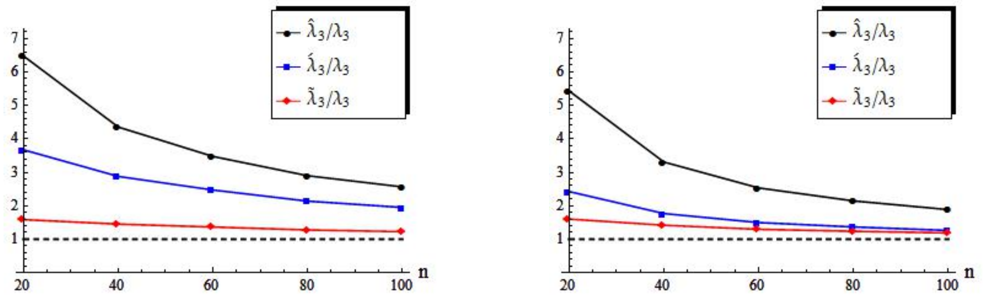
Figure 3: The behaviors of three estimates, $\hat{\lambda}_j/\lambda_j$, $\acute{\lambda}_j/\lambda_j$ and $\tilde{\lambda}_j/\lambda_j$, in the left panels and their sample variances, $\mathrm{var}(\hat{\lambda}_j/\lambda_j)$, $\mathrm{var}(\acute{\lambda}_j/\lambda_j)$ and $\mathrm{var}(\tilde{\lambda}_j/\lambda_j)$, in the right panels. The eigenvalue estimates were calculated based on samples of size $n = 40$ from $N_d(\mathbf{0}, \mathbf{\Sigma})$ having (14) with $\lambda_1 = d^{4/5}, \lambda_2 = d^{3/5}$ and $\lambda_3 = d^{2/5}$ for $d = 400(200)1200$.



(a) For the first eigenvalue when $\nu = 10$ (left) and $\nu = 30$ (right).



(b) For the second eigenvalue when $\nu = 10$ (left) and $\nu = 30$ (right).



(c) For the third eigenvalue when $\nu = 10$ (left) and $\nu = 30$ (right).

Figure 4: The behaviors of three estimates, $\hat{\lambda}_j/\lambda_j$, $\acute{\lambda}_j/\lambda_j$ and $\tilde{\lambda}_j/\lambda_j$, when $\nu = 10$ (left panels) and $\nu = 30$ (right panels). The eigenvalue estimates were calculated based on samples of sizes $n = 20(20)100$ from $t_d(\mathbf{0}, \mathbf{\Sigma}, \nu)$ having (14) with $\lambda_1 = d^{4/5}, \lambda_2 = d^{3/5}$ and $\lambda_3 = d^{2/5}$ for $d = 1000$.

that $\hat{\boldsymbol{H}}^T \boldsymbol{S}\hat{\boldsymbol{H}} = \hat{\boldsymbol{\Lambda}}$ having $\hat{\boldsymbol{\Lambda}} = \text{diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_d)$. We assume $\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_j \geq 0$ for all j without loss of generality. Note that $\hat{\boldsymbol{h}}_j$ can be calculated by $\hat{\boldsymbol{h}}_j = (n\hat{\lambda}_j)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{u}}_j$, where $\hat{\boldsymbol{u}}_j$ is a unit eigenvector of $\boldsymbol{S}_D$ corresponding to $\hat{\lambda}_j$. Yata and Aoshima [20] showed that the sample eigenvectors are consistent with their population counterparts under (3) as follows: Assume that λ_j $(j \leq m)$ has multiplicity one such as $\lambda_j \neq \lambda_{j'}$ for all $j'(\neq j)$. Then, $\hat{\boldsymbol{h}}_j$ is consistent with $\boldsymbol{h}_j$ in the sense that

$$\text{Angle}(\hat{\boldsymbol{h}}_j, \boldsymbol{h}_j) = o_p(1) \tag{15}$$

under (3) and (YA-i)-(YA-ii) in Section 1. Note that (15) is equivalent to the consistency in the sense that $||\hat{\boldsymbol{h}}_j - \boldsymbol{h}_j||^2 = o_p(1)$.

For the power spiked model, we have the following results.

Theorem 7.1. *For j $(\leq m)$ satisfying (C-i) to (C-iii), (15) holds as $d \to \infty$ and $n \to \infty$ without extra conditions in case of (B-i) or under (C-v) and (C-vi) in case of (B-ii).*

Remark 11. For j $(\leq m)$ satisfying (C-ii) and (C-iii), (15) holds as $d \to \infty$ and $n \to \infty$ under (A-i) in case of (B-i) or under (A-i) and (C-vi) in case of (B-ii).

If one cannot assume (C-ii), we have the following result.

Corollary 7.1. *For j $(\leq m)$ satisfying (C-i) and (C-iii), it holds as $d \to \infty$ and $n \to \infty$ that*

$$Angle(\hat{\boldsymbol{h}}_j, \boldsymbol{h}_j) = Arccos\big(1/\{1 + tr(\boldsymbol{\Sigma}_{(2)})/(n\lambda_j)\}^{1/2}\big) + o_p(1) \tag{16}$$

without extra conditions in case of (B-i) or under (C-v) and (C-vi) in case of (B-ii).

Remark 12. For j $(\leq m)$ satisfying (C-iii), (16) holds as $d \to \infty$ and $n \to \infty$ under (A-i) in case of (B-i) or under (A-i) and (C-vi) in case of (B-ii).

Remark 13. When the population mean may not be zero, let $\hat{\boldsymbol{h}}_{oj} = \{(n-1)\hat{\lambda}_{oj}\}^{-1/2}(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_{oj}$, where $\hat{\boldsymbol{u}}_{oj}$ is a unit eigenvector of $\boldsymbol{S}_{oD}$ corresponding to $\hat{\lambda}_{oj}$. Here, $\boldsymbol{S}_{oD}$ is defined in Corollary 3.3. Then, after replacing $\hat{\boldsymbol{h}}_j$s with $\hat{\boldsymbol{h}}_{oj}$s, the above results are still justified.

Next, we consider PC direction vectors by using the CDM methodology. Let $\tilde{\boldsymbol{h}}_{j(i)} = (n_i\tilde{\lambda}_j)^{-1/2}\boldsymbol{X}_i\tilde{\boldsymbol{u}}_{j(i)}$, $i = 1, 2$. We assume $\boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_{j(1)} \geq 0$ for all j without loss of generality. 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, $\boldsymbol{h}_j$. Let $\tilde{\boldsymbol{h}}_{j\star} = \tilde{\boldsymbol{h}}_j/||\tilde{\boldsymbol{h}}_j||$. Then, we have the following result.

Theorem 7.2. *For j ($\leq m$) satisfying (C-ii) and (C-iii), it holds as $d \to \infty$ and $n \to \infty$ that*

$$Angle(\tilde{\boldsymbol{h}}_{j\star}, \boldsymbol{h}_j) = o_p(1) \tag{17}$$

without extra conditions in case of (B-i) or under (C-v) and (C-vi) in case of (B-ii).

Note that (17) is equivalent to $||\tilde{\boldsymbol{h}}_{j\star} - \boldsymbol{h}_j||^2 = o_p(1)$.

Remark 14. Note that one can claim (17) without (C-i). For j ($\leq m$) satisfying (B-ii), (17) holds as $d \to \infty$ and $n \to \infty$ under (A-i), (C-ii), (C-iii) and (C-vi).

Remark 15. When the population mean may not be zero, let $\tilde{\boldsymbol{h}}_{oj(i)} = \{(n_i - 1)\tilde{\lambda}_{oj}\}^{-1/2}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)\tilde{\boldsymbol{u}}_{oj(i)}$, where $\tilde{\boldsymbol{u}}_{oj(1)}$ (or $\tilde{\boldsymbol{u}}_{oj(2)}$) is a unit left- (or right-) singular vector of $\boldsymbol{S}_{oD(1)}$ corresponding to $\tilde{\lambda}_{oj}$. Here, $\boldsymbol{S}_{oD(1)}$ is defined in Corollary 5.1. Then, after replacing $\tilde{\boldsymbol{h}}_{j(i)}$s with $\tilde{\boldsymbol{h}}_{oj(i)}$s, the results in Theorem 7.2 and Remark 14 are still justified.

8. Consistency of PC scores

The j-th PC score of $\boldsymbol{x}_k$ is given by $\boldsymbol{h}_j^T\boldsymbol{x}_k = z_{jk}\lambda_j^{1/2}$ ($= s_{jk}$, say). Note that $\text{Var}(s_{jk}) = \lambda_j$. Since $\boldsymbol{h}_j$ is unknown, one may use $\hat{\boldsymbol{h}}_j = (n\hat{\lambda}_j)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{u}}_j$ instead. The j-th PC score of $\boldsymbol{x}_k$ is estimated by $\hat{\boldsymbol{h}}_j^T\boldsymbol{x}_k = \hat{u}_{jk}(n\hat{\lambda}_j)^{1/2}$ ($= \hat{s}_{jk}$, say), where $\hat{\boldsymbol{u}}_j^T = (\hat{u}_{j1}, ..., \hat{u}_{jn})$. Then, Yata and Aoshima [20] evaluated conventional PC scores, $\hat{s}_{jk}$s, under (3) as follows: Assume that λ_j ($j \leq m$) has multiplicity one. Define a sample mean square error of the j-th PC scores by $\text{MSE}(\hat{s}_j) = n^{-1}\sum_{k=1}^{n}(\hat{s}_{jk} - s_{jk})^2$. Then, it holds that

$$\frac{\text{MSE}(\hat{s}_j)}{\lambda_j} = o_p(1) \tag{18}$$

under (YA-i)-(YA-ii) in Section 1. In this section, we consider PC scores under the power spiked model.

8.1. Conventional PC scores

As for $\hat{s}_{jk}$s, we have the following results.

Theorem 8.1. *For j ($\le m$) satisfying (C-i) to (C-iii), (18) holds as $d \to \infty$ and $n \to \infty$ without extra conditions in case of (B-i) or under (C-v) and (C-vi) in case of (B-ii).*

Corollary 8.1. *For j ($\le m$) satisfying (C-iii) and (C-iv), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\frac{MSE(\hat{s}_j)}{\lambda_j} = o_p(n^{-1/2}).$$

Remark 16. For j ($\le m$) satisfying (C-ii) and (C-iii), (18) holds as $d \to \infty$ and $n \to \infty$ under (A-i) in case of (B-i) or under (A-i) and (C-vi) in case of (B-ii).

Remark 17. When the population mean may not be zero, let $\hat{s}_{ojk} = \hat{u}_{ojk}\{(n-1)\hat{\lambda}_{oj}\}^{1/2}$, where $\hat{\lambda}_{oj}$ and $\hat{\boldsymbol{u}}_{oj}^T = (\hat{u}_{oj1}, ..., \hat{u}_{ojn})$ are given in Corollary 3.3 and Remark 13, respectively. Then, after replacing $\hat{s}_{jk}$s with $\hat{s}_{ojk}$s, the above results are still justified.

8.2. PC scores by the noise-reduction methodology

When we use the NR methodology, $\hat{s}_{jk}$ can be modified by $\hat{u}_{jk}(n\acute{\lambda}_j)^{1/2}$ ($= \acute{s}_{jk}$, say). A sample mean square error of the j-th PC scores is given by $\text{MSE}(\acute{s}_j) = n^{-1}\sum_{k=1}^{n}(\acute{s}_{jk} - s_{jk})^2$. Then, we have the following results.

Theorem 8.2. *For j ($\le m$) satisfying (C-i) and (C-iii), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\frac{MSE(\acute{s}_j)}{\lambda_j} = o_p(n^{-1/2}) \tag{19}$$

without extra conditions in case of (B-i) or under (C-v) and (C-vi) in case of (B-ii).

Remark 18. Note that one can claim (19) without (C-ii). For j ($\le m$) satisfying (C-iii), (19) holds as $d \to \infty$ and $n \to \infty$ under (A-i) in case of (B-i) or under (A-i) and (C-vi) in case of (B-ii).

Remark 19. When the population mean may not be zero, let $\acute{s}_{ojk} = \hat{u}_{ojk}\{(n-1)\acute{\lambda}_{oj}\}^{1/2}$, where $\acute{\lambda}_{oj}$ and $\hat{u}_{ojk}$ are given in Corollary 4.1 and Remark 17, respectively. Then, after replacing $\acute{s}_{jk}$s with $\acute{s}_{ojk}$s, the above results are still justified.

8.3. PC scores by the cross-data-matrix methodology

When we use the CDM methodology, 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, $\tilde{\lambda}_j$ ($j = 1, ..., n_2$), of $\boldsymbol{S}_{D(1)} = (n_1 n_2)^{-1/2}\boldsymbol{X}_1^T\boldsymbol{X}_2$. Let $\tilde{\boldsymbol{u}}_{j(i)}^T = (\tilde{u}_{j1(i)}, ..., \tilde{u}_{jn_i(i)})$, $i = 1, 2$. Then, the j-th PC score of $\boldsymbol{x}_{ik}$ can be estimated by $\tilde{u}_{jk(i)}(n_i\tilde{\lambda}_j)^{1/2}$ ($= \tilde{s}_{jk(i)}$, say). We denote $\tilde{s}_{jk} = \tilde{s}_{jk'(i')}$ with some k', i' for $k = 1, ..., n$, according to the relation that $\boldsymbol{x}_k = \boldsymbol{x}_{i'k'}$. A sample mean square error of the j-th PC scores is given by $\text{MSE}(\tilde{s}_j) = n^{-1}\sum_{k=1}^n(\tilde{s}_{jk} - s_{jk})^2$. Then, we have the following results.

Theorem 8.3. *For j ($\le m$) satisfying (C-iii), it holds as $d \to \infty$ and $n \to \infty$ that*

$$\frac{MSE(\tilde{s}_j)}{\lambda_j} = o_p(n^{-1/2}) \qquad (20)$$

without extra conditions in case of (B-i) or under (C-v) and (C-vi) in case of (B-ii).

Remark 20. Note that one can claim (20) without (C-i) and (C-ii). For j ($\le m$) satisfying (B-ii), (20) holds as $d \to \infty$ and $n \to \infty$ under (A-i), (C-iii) and (C-vi).

Remark 21. When the population mean may not be zero, let $\tilde{s}_{ojk(i)} = \tilde{u}_{ojk(i)}\{(n_i - 1)\tilde{\lambda}_{oj}\}^{1/2}$, where $\tilde{\lambda}_{oj}$ and $\tilde{\boldsymbol{u}}_{oj(i)}^T = (\tilde{u}_{oj1(i)}, ..., \tilde{u}_{ojn_i(i)})$ are given in Corollary 5.1 and Remark 15, respectively. Then, after replacing $\tilde{s}_{jk(i)}$s with $\tilde{s}_{ojk(i)}$s, the above results are still justified.

Appendix A.

Throughout, let $\boldsymbol{e}_{jn} = (e_{j1}, ..., e_{jn})^T$, $j = 1, 2$, be arbitrary unit random vectors, where $\sum_{k=1}^n e_{jk}^2 = 1$. Let M be a uniform bound for the fourth moment of z_{ij}s such that $E(z_{ij}^4) < M$ for all i, j. Let us write that

$$\boldsymbol{U}_1 = n^{-1}\sum_{s=1}^m \lambda_s\boldsymbol{z}_s\boldsymbol{z}_s^T, \quad \boldsymbol{U}_2 = n^{-1}\sum_{s=m+1}^d \lambda_s\boldsymbol{z}_s\boldsymbol{z}_s^T,$$

$$\boldsymbol{V}_1 = (n_1n_2)^{-1/2}\sum_{s=1}^m \lambda_s\boldsymbol{z}_{1s}\boldsymbol{z}_{2s}^T \quad \text{and} \quad \boldsymbol{V}_2 = (n_1n_2)^{-1/2}\sum_{s=m+1}^d \lambda_s\boldsymbol{z}_{1s}\boldsymbol{z}_{2s}^T.$$

Note that $\boldsymbol{S}_D = \boldsymbol{U}_1 + \boldsymbol{U}_2$ and $\boldsymbol{S}_{D(1)} = \boldsymbol{V}_1 + \boldsymbol{V}_2$. Let us write $u_i = n^{-1}\sum_{s=m+1}^{d}\lambda_s z_{si}^2$ as a diagonal element of $\boldsymbol{U}_2$. Let $\kappa = n^{-1}\sum_{s=m+1}^{d}\lambda_s$ and $\boldsymbol{U}_{22} = \boldsymbol{U}_2 - \kappa\boldsymbol{I}_n$. Let $\boldsymbol{U}_{2(t)} = (u_{ij(t)})$, $t = 1, 2, ...$, be $n \times n$ matrices such that

$$u_{ij(t)} = \begin{cases} n^{-1}\sum_{s=m+1}^{d}\lambda_s^{2^{t-1}} z_{si}z_{sj} & (i \neq j), \\ 0 & (i = j). \end{cases}$$

Let us write that $K_1 = \sum_{r,s=m+1}^{d}\lambda_r\lambda_s E\{(z_{rk}^2 - 1)(z_{sk}^2 - 1)\}$ and $K_2 = \sum_{p\neq q, r\neq s \geq m+1}^{d}\lambda_p\lambda_q\lambda_r\lambda_s\{E(z_{pk}z_{qk}z_{rk}z_{sk})\}^2$. Let $\tilde{\boldsymbol{z}}_j = (||n^{-1/2}\boldsymbol{z}_j||)^{-1}n^{-1/2}\boldsymbol{z}_j$, $j = 1, ..., m$. In Appendix, we consider the following conditions:

- **(D-i)** (C-i) in case of (B-i) or (C-i), (C-v) and (C-vi) in case of (B-ii);
- **(D-ii)** No extra conditions in case of (B-i) or (C-v) and (C-vi) in case of (B-ii).

Lemma 1. *It holds that $K_1 = O\{tr(\boldsymbol{\Sigma}_{(2)}^2)\}$ and $K_2 = O\{tr(\boldsymbol{\Sigma}_{(2)}^2)^2\}$ under (A-i).*

Proof of Lemma 1. Let $\boldsymbol{P} = \boldsymbol{I}_d - \sum_{i=1}^{m}\boldsymbol{h}_i\boldsymbol{h}_i^T$. Note that $\boldsymbol{P}^2 = \boldsymbol{P}$, $\boldsymbol{P}\boldsymbol{x}_k = \sum_{s=m+1}^{d}\lambda_s^{1/2}\boldsymbol{h}_s z_{sk}$ and $\boldsymbol{x}_k^T\boldsymbol{P}\boldsymbol{x}_k = \sum_{s=m+1}^{d}\lambda_s z_{sk}^2$. Under (A-i), we write that $\boldsymbol{x}_k^T\boldsymbol{P}\boldsymbol{x}_k = \sum_{r,s=1}^{t}\boldsymbol{\gamma}_r^T\boldsymbol{P}\boldsymbol{\gamma}_s w_{rk}w_{sk}$. Note that $E(\boldsymbol{x}_k^T\boldsymbol{P}\boldsymbol{x}_k) = \sum_{s=1}^{t}\boldsymbol{\gamma}_s^T\boldsymbol{P}\boldsymbol{\gamma}_s = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{P}) = \mathrm{tr}(\boldsymbol{\Sigma}_{(2)})$. Then, it holds under (A-i) that

$$K_1 = \mathrm{Var}(\boldsymbol{x}_k^T\boldsymbol{P}\boldsymbol{x}_k) = O\Big(\sum_{r,s=1}^{t}(\boldsymbol{\gamma}_r^T\boldsymbol{P}\boldsymbol{\gamma}_s)^2\Big) = O[\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{P})^2\}] = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)\}.$$

On the other hand, we write $(\boldsymbol{x}_k^T\boldsymbol{P}\boldsymbol{x}_{k'})^2 = \sum_{r,s=m+1}^{d}\lambda_r\lambda_s z_{rk}z_{rk'}z_{sk}z_{sk'} = \sum_{p,q,r,s=1}^{t}\boldsymbol{\gamma}_p^T\boldsymbol{P}\boldsymbol{\gamma}_q\boldsymbol{\gamma}_r^T\boldsymbol{P}\boldsymbol{\gamma}_s w_{pk}w_{qk'}w_{rk}w_{sk'}$ under (A-i). Then, for $k \neq k'$, it holds under (A-i) that

$$\begin{aligned} E\{(\boldsymbol{x}_k^T\boldsymbol{P}\boldsymbol{x}_{k'})^4\} &= \sum_{p,q,r,s=m+1}^{d}\lambda_p\lambda_q\lambda_r\lambda_s\{E(z_{pk}z_{qk}z_{rk}z_{sk})\}^2 \ (\leq K_2) \\ &= O[\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{P})^2\}^2] + O[\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{P})^4\}] = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2\} + O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^4)\}. \end{aligned}$$

Thus from the fact that $\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^4) \leq \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2$, it concludes the results. $\square$

Lemma 2. *It holds as $d \to \infty$ and $n \to \infty$ that*

$$||\lambda_j^{-2^{t-1}} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t)}||^2 = \lambda_j^{-2^t} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t+1)} \boldsymbol{e}_{1n} + o_p(1), \quad t = 1, 2, ...$$

for j $(\le m)$ under (D-ii).

Proof of Lemma 2. For every t $(= 1, 2, ...)$, we write that

$$||\boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t)}||^2 = \sum_{i'=1}^n e_{1i'}^2 \sum_{k=1(\backslash i')}^n u_{i'k(t)}^2 + \sum_{i' \ne j'}^n e_{1i'} e_{1j'} \sum_{k=1(\backslash i',j')}^n u_{i'k(t)} u_{j'k(t)}, \quad \text{(A.1)}$$

where $(\backslash i)$ excludes number i and $(\backslash i, j)$ excludes numbers i, j. For the second term in (A.1), in a way similar to the proof of Lemma A.1 given in Yata and Aoshima [20], we can obtain under (D-ii) that

$$\lambda_j^{-2^t} \sum_{i' \ne j'}^n e_{1i'} e_{1j'} \sum_{k=1(\backslash i',j')}^n u_{i'k(t)} u_{j'k(t)} = \lambda_j^{-2^t} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t+1)} \boldsymbol{e}_{1n} + o_p(1). \quad \text{(A.2)}$$

Next, we consider the first term in (A.1). In case of (B-i), by using Markov's inequality for any τ (> 0), one has as $d \to \infty$ that

$$\sum_{i'=1}^n P\Big(\lambda_j^{-2^t} \sum_{k=1(\backslash i')}^n u_{i'k(t)}^2 > \tau\Big) \le n \sum_{k=1(\backslash i')}^n \frac{E(u_{i'k(t)}^2)}{\tau \lambda_j^{2^t}} = O\Big(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{\lambda_j^2}\Big) \to 0.$$

In case of (B-ii), by using Chebyshev's inequality for any τ (> 0), one has that

$$\begin{aligned}
\sum_{i'=1}^n P\Big(\lambda_j^{-2^t} \sum_{k=1(\backslash i')}^n u_{i'k(t)}^2 > \tau\Big) &\le \sum_{i'=1}^n P\Big(\sum_{k=1(\backslash i')}^n \sum_{s=m+1}^d \frac{\lambda_s^{2^t} z_{si'}^2 z_{sk}^2}{n^2 \lambda_j^{2^t}} > \tau/2\Big) \\
&\quad + \sum_{i'=1}^n P\Big(\Big|\sum_{k=1(\backslash i')}^n \sum_{p \ne q \ge m+1}^d \frac{\lambda_p^{2^{t-1}} \lambda_q^{2^{t-1}} z_{pi'} z_{qi'} z_{pk} z_{qk}}{n^2 \lambda_j^{2^t}}\Big| > \tau/2\Big) \\
= O\Big\{\sum_{p \ne q, r \ne s \ge m+1}^d & \frac{\lambda_p^{2^{t-1}} \lambda_q^{2^{t-1}} \lambda_r^{2^{t-1}} \lambda_s^{2^{t-1}} \{E(z_{pk} z_{qk} z_{rk} z_{sk})\}^2}{n^2 \lambda_j^{2^{t+1}}}\Big\} \\
+ O\Big(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^{2^t})^2}{n \lambda_j^{2^{t+1}}}\Big) &= O\Big(\frac{K_2}{n^2 \lambda_j^4} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2}{n \lambda_j^4}\Big) \to 0
\end{aligned}$$

under (C-v) and (C-vi). Thus we obtain that

$$\lambda_j^{-2t} \sum_{i'=1}^{n} e_{1i'}^2 \sum_{k=1(\backslash i')}^{n} u_{i'k(t)}^2 = o_p(1). \tag{A.3}$$

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

Lemma 3. *It holds as $d \to \infty$ and $n \to \infty$ that $||\lambda_j^{-1} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(1)}||^2 = o_p(1)$ for $j\ (\le m)$ under (D-ii).*

Proof of Lemma 3. From (5), there is at least one positive integer t_j (≥ 2) satisfying $\lim_{d\to\infty} \lambda_j^{-2^{t_j}} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^{2^{t_j}}) = 0$. Then, by using Markov's inequality, we have for $i' \neq j'$ under (D-ii) that $P(\lambda_j^{-2^{t_j}} \sum_{i'\neq j'}^{n} |u_{i'j'(t_j)}|^2 > \tau) \le \tau^{-1} \lambda_j^{-2^{t_j}} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^{2^{t_j}}) \to 0$ for any $\tau > 0$. Then, from the fact that $\sum_{i'\neq j'}^{n} e_{1i'}^2 e_{1j'}^2 \le 1$ w.p.1, we obtain that

$$\frac{\boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t_j)} \boldsymbol{e}_{1n}}{\lambda_j^{2^{t_j-1}}} = \frac{\sum_{i'\neq j'}^{n} e_{1i'} e_{1j'} u_{i'j'(t_j)}}{\lambda_j^{2^{t_j-1}}} \le \Big(\frac{\sum_{i'\neq j'}^{n} |u_{i'j'(t_j)}|^2}{\lambda_j^{2^{t_j}}} \Big)^{1/2} = o_p(1).$$

Thus from Lemma 2, we have under (D-ii) that

$$||\lambda_j^{-2^{t_j-2}} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t_j-1)}||^2 = \lambda_j^{-2^{t_j-1}} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t_j)} \boldsymbol{e}_{1n} + o_p(1) = o_p(1),$$

so that $\lambda_j^{-2^{t_j-2}} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(t_j-1)} \boldsymbol{e}_{1n} = o_p(1)$. Similarly, we have under (D-ii) that $||\lambda_j^{-1} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(1)}\ ||^2 = \lambda_j^{-2} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{2(2)} \boldsymbol{e}_{1n} + o_p(1) = o_p(1)$, which concludes the result. □

Lemma 4. *It holds as $d \to \infty$ and $n \to \infty$ that $||\lambda_j^{-1} \boldsymbol{e}_{1n_1}^T \boldsymbol{V}_2||^2 = o_p(1)$ for $j\ (\le m)$ under (D-ii).*

Proof of Lemma 4. Let $\boldsymbol{V}_{2(t)} = (v_{ij(t)})$, $t = 1, 2, ...$, where $v_{ij(t)} = (n_1 n_2)^{-1/2} \sum_{s=m+1}^{d} \lambda_s^{2^{t-1}} z_{1si} z_{2sj}$. For every t, we write that $||\boldsymbol{e}_{1n_1}^T \boldsymbol{V}_{2(t)}||^2 = \sum_{i'=1}^{n_1} e_{1i'}^2 \sum_{k=1}^{n_2} v_{i'k(t)}^2 + \sum_{i'\neq j'}^{n_1} e_{1i'} e_{1j'} \sum_{k=1}^{n_2} v_{i'k(t)} v_{j'k(t)}$. Thus in a way similar to the proofs of Lemmas 2 and 3, we can claim the result. □

Lemma 5. *It holds as $d \to \infty$ and $n \to \infty$ that $\lambda_j^{-1} \boldsymbol{e}_{1n}^T \boldsymbol{U}_{22} \boldsymbol{e}_{2n} = o_p(1)$ for $j\ (\le m)$ under (D-i).*

Proof of Lemma 5. By using Markov's inequality, for any $\tau > 0$, one has under (C-i) that $P\{\sum_{k=1}^n \lambda_j^{-2}(u_k - \kappa)^2 > \tau\} \le \tau^{-1}K_1/(n\lambda_j^2) \to 0$. Thus it holds that $|\sum_{k=1}^n e_{1k}e_{2k}\lambda_j^{-1}(u_k - \kappa)| \le \{\sum_{k=1}^n \lambda_j^{-2}(u_k - \kappa)^2\}^{1/2} = o_p(1)$ from the fact that $\sum_{k=1}^n e_{1k}^2 e_{2k}^2 \le 1$ w.p.1. From Lemma 3, we have under (D-i) that $\lambda_j^{-1}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{2n} = \lambda_j^{-1}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{2(1)}\boldsymbol{e}_{2n} + \sum_{k=1}^n e_{1k}e_{2k}\lambda_j^{-1}(u_k - \kappa) = o_p(1)$ for j $(\le m)$. It concludes the result. □

Lemma 6. *It holds as $d \to \infty$ and $n \to \infty$ that*

$$(n\lambda_j)^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_{2(1)}\boldsymbol{z}_{j'} = o_p(n^{-1/2}) \quad (i' = 1, ..., m;\ j' = 1, ..., m);$$
$$(n_1n_2)^{-1/2}\lambda_j^{-1}\boldsymbol{z}_{1i'}^T\boldsymbol{V}_2\boldsymbol{z}_{2j'} = o_p(n^{-1/2}) \quad (i' = 1, ..., m;\ j' = 1, ..., m)$$

for j $(\le m)$ under (D-ii).

Proof of Lemma 6. We consider the first result. One can write that $\boldsymbol{z}_{i'}^T\boldsymbol{U}_{2(1)}\boldsymbol{z}_{j'} = \sum_{k_1 \ne k_2}^n z_{i'k_1}z_{j'k_2}u_{k_1k_2(1)}$. Note $E\{(\sum_{k_2=1(\backslash k_1)}^n z_{j'k_2}u_{k_1k_2(1)})^2\} = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/n\}$ in case of (B-i). Then, for any $\tau > 0$, by using Schwarz's inequality, one has as $d \to \infty$ and $n \to \infty$ that

$$\begin{aligned}
&P\Big(\big|(n\lambda_j)^{-1}\sum_{k_1 \ne k_2}^n z_{i'k_1}z_{j'k_2}u_{k_1k_2(1)}\big| > n^{-1/2}\tau\Big)\\
&\le P\Big((n\lambda_j)^{-1}\sum_{k_1=1}^n |z_{i'k_1}|\big|\sum_{k_2=1(\backslash k_1)}^n z_{j'k_2}u_{k_1k_2(1)}\big| > n^{-1/2}\tau\Big)\\
&\le \frac{n^{1/2}E(z_{i'k_1}^2)^{1/2}E\{(\sum_{k_2=1(\backslash k_1)}^n z_{j'k_2}u_{k_1k_2(1)})^2\}^{1/2}}{\tau\lambda_j} = O\Big(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^{1/2}}{\lambda_j}\Big) \to 0.
\end{aligned}$$

In case of (B-ii), note that $n^2E(z_{i'k_1}^2z_{j'k_2}^2u_{k_1k_2(1)}^2) \le M^2\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2) + E(z_{i'k_1}^2z_{j'k_2}^2 \sum_{p \ne q \ge m+1}^d \lambda_p\lambda_q z_{pk_1}z_{pk_2}z_{qk_1}z_{qk_2}) \le M^2\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2) + E(z_{i'k_1}^4z_{j'k_2}^4)^{1/2}K_2^{1/2} = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2) + K_2^{1/2}\}$ and $n^2E(z_{i'k_1}z_{j'k_1}z_{i'k_2}z_{j'k_2}u_{k_1k_2(1)}^2) = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2) + K_2^{1/2}\}$ $(k_1 \ne k_2)$ from Schwarz's inequality. Then, for any $\tau > 0$, one has under (C-v) and (C-vi) that

$$\begin{aligned}
&P\Big(\big|(n\lambda_j)^{-1}\sum_{k_1 \ne k_2}^n z_{i'k_1}z_{j'k_2}u_{k_1k_2(1)}\big| > n^{-1/2}\tau\Big)\\
&\le \frac{nE\{(z_{i'k_1}^2z_{i'k_2}^2 + z_{i'k_1}z_{j'k_1}z_{i'k_2}z_{j'k_2})u_{k_1k_2(1)}^2\}}{\tau^2\lambda_j^2} = O\Big(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2) + K_2^{1/2}}{n\lambda_j^2}\Big) \to 0.
\end{aligned}$$

Therefore, it concludes the first result.

Next, we consider the second result. One can write that $\boldsymbol{z}_{1i'}^T\boldsymbol{V}_2\boldsymbol{z}_{2j'} = \sum_{k=1}^{n_2} z_{1i'k}z_{2j'k}v_{kk(1)} + \sum_{k_1=1}^{n_1}\sum_{k_2=1(\backslash k_1)}^{n_2} z_{1i'k_1}z_{2j'k_2}v_{k_1k_2(1)}$. For the first term, by using Schwarz's inequality, we have that

$$P\Big(\big|(n_1n_2)^{-1/2}\lambda_j^{-1}\sum_{k=1}^{n_2} z_{1i'k}z_{2j'k}v_{kk(1)}\big| > n^{-1/2}\tau\Big)$$
$$\le (nn_2/n_1)^{1/2}E(|z_{1i'k}z_{2j'k}||v_{kk(1)}|)/(\tau\lambda_j) = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^{1/2}/(n^{1/2}\lambda_j)\} \to 0$$

under (D-ii) for any $\tau > 0$. Similarly to the first result, we can claim for the second term that $(n_1n_2)^{-1/2}\lambda_j^{-1}\sum_{k_1=1}^{n_1}\sum_{k_2=1(\backslash k_1)}^{n_2} z_{1i'k_1}z_{2j'k_2}v_{k_1k_2(1)} = o_p(n^{-1/2})$. It concludes the second result. □

Lemma 7. *For i' $(\le m)$, it holds as $d\to\infty$ and $n\to\infty$ that*

$$||n^{-1/2}\lambda_j^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_{2(1)}||^2 = o_p(n^{-1/2}) \quad \textit{and} \quad ||n_1^{-1/2}\lambda_j^{-1}\boldsymbol{z}_{1i'}^T\boldsymbol{V}_2||^2 = o_p(n^{-1/2})$$

for j $(\le m)$ under (D-ii).

Proof of Lemma 7. We consider the first result. One can write that

$$||n^{-1/2}\lambda_j^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_{2(1)}||^2 = n^{-1}\lambda_j^{-2}\sum_{k_1,k_2}^{n} z_{i'k_1}z_{i'k_2}\sum_{k_3=1(\backslash k_1,k_2)}^{n} u_{k_1k_3(1)}u_{k_2k_3(1)}.$$

First, we consider the case of $k_1 = k_2$. Note that $E(z_{i'k_1}^2\sum_{k_3=1(\backslash k_1)}^{n} u_{k_1k_3(1)}^2) = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/n\}$. For any $\tau > 0$, one has under (D-ii) that

$$P\Big(\sum_{k_1=1}^{n} z_{i'k_1}^2\sum_{k_3=1(\backslash k_1)}^{n}\frac{u_{k_1k_3(1)}^2}{n\lambda_j^2} > \tau n^{-1/2}\Big) = O\Big(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{n^{1/2}\lambda_j^2}\Big) \to 0. \tag{A.4}$$

Next, we consider the case of $k_1 \ne k_2$. Let $u_{(k_1k_2k_3)} = \sum_{p\ne q\ge m+1}^{d}\lambda_p\lambda_q z_{pk_1}z_{qk_2}z_{pk_3}z_{qk_3}/n^2$. Note that $E\{(\sum_{k_3=1(\backslash k_1,k_2)}^{n} u_{(k_1k_2k_3)})^2\} = O\{n^{-3}\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2\}$ and $E\{z_{i'k_1}^2z_{i'k_2}^2(\sum_{k_3=1(\backslash k_1,k_2)}^{n}\sum_{s=m+1}^{d}\lambda_s^2z_{sk_1}z_{sk_2}z_{sk_3}^2/n^2)^2\} = O\{n^{-2}\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2\}$. In case of (B-i), for any $\tau > 0$, one has as $d \to \infty$ and $n \to \infty$ that

$$
\begin{aligned}
&P\Big((n^{1/2}\lambda_j)^{-2}\sum_{k_1\neq k_2}^{n} z_{i'k_1}z_{i'k_2}\sum_{k_3=1(\backslash k_1,k_2)}^{n} u_{(k_1k_2k_3)} > \tau n^{-1/2}\Big)\\
&\le P\Big(\sum_{k_1\neq k_2}^{n} |\frac{z_{i'k_1}z_{i'k_2}}{n^2}||\frac{n^{3/2}}{\lambda_j^2}\sum_{k_3=1(\backslash k_1,k_2)}^{n} u_{(k_1k_2k_3)}| > \tau\Big) = O\Big(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{\lambda_j^2}\Big) \to 0;\\
&P\Big((n^{1/2}\lambda_j)^{-2}\sum_{k_1\neq k_2}^{n} z_{i'k_1}z_{i'k_2}\sum_{k_3=1(\backslash k_1,k_2)}^{n}\sum_{s=m+1}^{d}\lambda_s^2 z_{sk_1}z_{sk_2}z_{sk_3}^2/n^2 > \tau n^{-1/2}\Big)\\
&= O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/(n\lambda_j^2)\} \to 0.
\end{aligned}
$$

In case of (B-ii), one has under (C-v) and (C-vi) that

$$
\begin{aligned}
&P\Big((n^{1/2}\lambda_j)^{-2}\sum_{k_1\neq k_2}^{n} z_{i'k_1}z_{i'k_2}\sum_{k_3=1(\backslash k_1,k_2)}^{n} u_{k_1k_3(1)}u_{k_2k_3(1)} > \tau n^{-1/2}\Big)\\
&\le \tau^{-2}n^2\lambda_j^{-4}\{E(z_{i'k_1}^2z_{i'k_2}^2u_{k_1k(1)}^2u_{k_2k(1)}^2) + M^2\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2/n^3\}\\
&= \tau^{-2}n^2\lambda_j^{-4}\{E(z_{i'k_1}^4u_{k_2k(1)}^4)E(z_{i'k_2}^4u_{k_1k(1)}^4)\}^{1/2} + o(1)\\
&= O\{n^2\lambda_j^{-4}E(u_{k_1k(1)}^4)\} + o(1) \to 0 \qquad\qquad (A.5)
\end{aligned}
$$

from the fact that $n^2E(u_{k_1k(1)}^4) = O\{K_2/n^2 + \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2/n^2\}$ for $k \neq k_1, k_2$. From (A.4) and (A.5), it concludes the first result.

Next, we consider the second result. One can write that $||n_1^{-1/2}\lambda_j^{-1}\boldsymbol{z}_{1i'}^T\boldsymbol{V}_2||^2 = n_1^{-1}\lambda_j^{-2}\sum_{k_1,k_2}^{n_1} z_{1i'k_1}z_{1i'k_2}\sum_{k_3=1}^{n_2} v_{k_1k_3}v_{k_2k_3(1)}$. Then, similarly to the first result, it concludes the second result. □

Lemma 8. *It holds as $d \to \infty$ and $n \to \infty$ that*

$$
\begin{aligned}
&||n^{-1/2}\lambda_j^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_{22}||^2 = o_p(n^{-1/2}) \quad (i' = 1, ..., m);\\
&(n\lambda_j)^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_{22}\boldsymbol{z}_{j'} = o_p(n^{-1/2}) \quad (i' = 1, ..., m;\ j' = 1, ..., m)
\end{aligned}
$$

for j ($\le m$) under (D-i).

Proof of Lemma 8. We consider the first result. One can write that

$$||n^{-1/2}\lambda_j^{-1}\boldsymbol{z}_{i'}^T\mathrm{diag}(u_1 - \kappa, ..., u_n - \kappa)||^2 = (n\lambda_j^2)^{-1}\sum_{k=1}^{n} z_{i'k}^2(u_k - \kappa)^2.$$

Let $\eta = K_1^{1/2}/(n\lambda_j^2)^{1/2}$. When $\eta = 0$, the result is obvious. We assume $\eta > 0$. Note that $\eta \to 0$ under (C-i). Here, it holds that $\sum_{k=1}^n P(z_{i'k}^2/n^{1/2} > 1/\eta) \le M\eta^2 \to 0$. Thus it holds that $z_{i'k}^2/n^{1/2} \le 1/\eta$ for all $k = 1, ..., n$ with probability going to 1 as $\eta \to 0$. Then, by noting that $E\{(u_k - \kappa)^2\} = K_1/n^2$, it holds for any $\tau > 0$ that $P\{(n\lambda_j^2)^{-1}\sum_{k=1}^n z_{i'k}^2(u_k-\kappa)^2 > \tau n^{-1/2}\} \le P\{\lambda_j^{-2}\sum_{k=1}^n \eta^{-1}(u_k-\kappa)^2 > \tau\} + o(1) = O(\eta) + o(1) \to 0$ under (C-i). Hence, we obtain from Lemma 7 that $||(n^{1/2}\lambda_j)^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_{22}||^2 = ||(n^{1/2}\lambda_j)^{-1}\boldsymbol{z}_{i'}^T\{\text{diag}(u_1-\kappa, ..., u_n-\kappa) + \boldsymbol{U}_{2(1)}\}||^2 = o_p(n^{-1/2})$ under (D-i), which concludes the first result.

Next, we consider the second result. One can write that $\boldsymbol{z}_{i'}^T\text{diag}(u_1 - \kappa, ..., u_n - \kappa)\boldsymbol{z}_{j'} = \sum_{k=1}^n z_{i'k}z_{j'k}(u_k-\kappa)$. Then, it holds under (C-i) that

$$
\begin{aligned}
&P\Big((n\lambda_j)^{-1}\Big|\sum_{k=1}^n z_{i'k}z_{j'k}(u_k-\kappa)\Big| > n^{-1/2}\tau\Big)\\
&\le P(n^{-1/2}\lambda_j^{-1}\sum_{k=1}^n |z_{i'k}z_{j'k}||u_k-\kappa| > \tau) \le n^{1/2}E(|z_{i'k}z_{j'k}||u_k-\kappa|)/(\tau\lambda_j)\\
&\le n^{1/2}\{E(z_{i'k}^2z_{j'k}^2)E(|u_k-\kappa|^2)\}^{1/2}/(\tau\lambda_j) = O\{K_1^{1/2}/(n^{1/2}\lambda_j)\} = o(1).
\end{aligned}
$$

Hence, we obtain from Lemma 6 that $(n\lambda_j)^{-1}\boldsymbol{z}_{i'}^T\boldsymbol{U}_{22}\boldsymbol{z}_{j'} = (n\lambda_j)^{-1}\sum_{k=1}^n z_{i'k}z_{j'k}(u_k-\kappa) + o_p(n^{-1/2}) = o_p(n^{-1/2})$. It concludes the second result. □

Lemma 9. *Assume that the first m population eigenvalues are distinct in the sense that* $\liminf_{d\to\infty}|\lambda_{j'}/\lambda_j - 1| > 0$ *for all* $j \ne j' = 1, ..., m$. *Then, it holds as* $d \to \infty$ *and* $n \to \infty$ *that*

$$(\hat\lambda_j - \kappa)/\lambda_j = ||n^{-1/2}\boldsymbol{z}_j||^2 + o_p(n^{-1/2}) \quad \text{and} \quad \hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_j = 1 + o_p(n^{-1/2})$$

for j $(\le m)$ *under (D-i).*

Proof of Lemma 9. Let us note that $\boldsymbol{S}_D - \kappa\boldsymbol{I}_n = \boldsymbol{U}_1 + \boldsymbol{U}_{22}$. Similarly to the proof of Lemma 5 given in Yata and Aoshima [22], by using Lemmas 5 and 8, we can claim under (D-i) with $j = 1$ that

$$(\hat\lambda_1 - \kappa)/\lambda_1 = \hat{\boldsymbol{u}}_1^T(\boldsymbol{U}_1 + \boldsymbol{U}_{22})\hat{\boldsymbol{u}}_1/\lambda_1 = ||n^{-1/2}\boldsymbol{z}_1||^2 + o_p(n^{-1/2})$$

and $\hat{\boldsymbol{u}}_1^T\tilde{\boldsymbol{z}}_1 = 1 + o_p(n^{-1/2})$. For λ_2, under (D-i) with $j = 2$, it holds from Lemma 8 that $\hat{\boldsymbol{u}}_1^T\boldsymbol{U}_{22}\hat{\boldsymbol{u}}_2/\lambda_2 = \tilde{\boldsymbol{z}}_1^T\boldsymbol{U}_{22}\hat{\boldsymbol{u}}_2/\lambda_2 + o_p(n^{-1/4}) = o_p(n^{-1/4})$. Then, from the fact that $n^{-1}\boldsymbol{z}_j^T\boldsymbol{z}_{j'} = o_p(n^{-1/4})$ $(j \ne j')$, it holds that

$$0 = \hat{\boldsymbol{u}}_1^T(\boldsymbol{U}_1 + \boldsymbol{U}_{22})\hat{\boldsymbol{u}}_2/\lambda_1 = \{1 + o_p(1)\}\tilde{\boldsymbol{z}}_1^T\hat{\boldsymbol{u}}_2 + o_p(n^{-1/4}\lambda_2/\lambda_1), \tag{A.6}$$

so that $\tilde{\boldsymbol{z}}_1^T\hat{\boldsymbol{u}}_2 = o_p(n^{-1/4}\lambda_2/\lambda_1)$. Thus we have under (D-i) with $j=2$ that

$$\hat{\boldsymbol{u}}_2^T(\boldsymbol{U}_1+\boldsymbol{U}_{22})\hat{\boldsymbol{u}}_2/\lambda_2 = \hat{\boldsymbol{u}}_2^T\Big(\sum_{s=2}^m \lambda_s||n^{-1/2}\boldsymbol{z}_s||^2\tilde{\boldsymbol{z}}_s\tilde{\boldsymbol{z}}_s^T + \boldsymbol{U}_{22}\Big)\hat{\boldsymbol{u}}_2/\lambda_2 + o_p(n^{1/2}).$$

Thus similar to the case of λ_1, we obtain that $(\hat{\lambda}_2-\kappa)/\lambda_2 = ||n^{-1/2}\boldsymbol{z}_2||^2 + o_p(n^{-1/2})$ and $\hat{\boldsymbol{u}}_2^T\tilde{\boldsymbol{z}}_2 = 1+o_p(n^{-1/2})$. Similarly to (A.6), we can claim under (D-i) with $j=3$ that

$$\begin{aligned}0 = \hat{\boldsymbol{u}}_1^T(\boldsymbol{U}_1+\boldsymbol{U}_{22})\hat{\boldsymbol{u}}_3/\lambda_1 &= \{1+o_p(1)\}\tilde{\boldsymbol{z}}_1^T\hat{\boldsymbol{u}}_3 + \tilde{\boldsymbol{z}}_2^T\hat{\boldsymbol{u}}_3 o_p(n^{-1/4}\lambda_2/\lambda_1)\\ &\quad + o_p(n^{-1/4}\lambda_3/\lambda_1);\\ 0 = \hat{\boldsymbol{u}}_2^T(\boldsymbol{U}_1+\boldsymbol{U}_{22})\hat{\boldsymbol{u}}_3/\lambda_2 &= \{1+o_p(1)\}\tilde{\boldsymbol{z}}_2^T\hat{\boldsymbol{u}}_3 + \tilde{\boldsymbol{z}}_1^T\hat{\boldsymbol{u}}_3 o_p(1) + o_p(n^{-1/4}\lambda_3/\lambda_2),\end{aligned}$$

so that $\tilde{\boldsymbol{z}}_s^T\hat{\boldsymbol{u}}_3 = o_p(n^{-1/4}\lambda_3/\lambda_s)$, $s=1,2$. Thus it holds that $\hat{\boldsymbol{u}}_3^T(\boldsymbol{U}_1+\boldsymbol{U}_{22})\hat{\boldsymbol{u}}_3/\lambda_3 = \hat{\boldsymbol{u}}_3^T(\sum_{s=3}^m \lambda_s||n^{-1/2}\boldsymbol{z}_s||^2\tilde{\boldsymbol{z}}_s\tilde{\boldsymbol{z}}_s^T + \boldsymbol{U}_{22})\hat{\boldsymbol{u}}_3/\lambda_3 + o_p(n^{-1/2})$. For $j\geq 3$, in a way similar to the case of λ_2, we have that $(\hat{\lambda}_j-\kappa)/\lambda_j = ||n^{-1/2}\boldsymbol{z}_j||^2 + o_p(n^{-1/2})$ and $\hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_j = 1+o_p(n^{-1/2})$ under (D-i). It concludes the results. □

Lemma 10. *Under (D-i), it holds as $d\to\infty$ and $n\to\infty$ that*

$$\begin{aligned}&(\hat{\lambda}_j-\kappa)/\lambda_j = 1+o_p(1) \quad \text{for } j=1,...,m;\\ &(\hat{\lambda}_j-\kappa)/\lambda_j = ||n^{-1/2}\boldsymbol{z}_j||^2 + o_p(n^{-1/2}) \quad \text{and} \quad \hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_j = 1+o_p(n^{-1/2})\\ &\qquad\qquad \text{for } j\ (\leq m) \text{ satisfying (C-iii)}.\end{aligned}$$

Proof of Lemma 10. We consider an example that $\liminf_{d\to\infty}|\lambda_2/\lambda_1-1| = 0$ and $\liminf_{d\to\infty}|\lambda_j/\lambda_3-1|>0$ for $j(\neq 3)=1,...,m\ (\geq 3)$.

We first consider the case that $\lim_{d\to\infty}\lambda_2/\lambda_1 = 1$. Note that $||n^{-1/2}\boldsymbol{z}_j||^2 = 1+o_p(1)$ as $n\to\infty$. We have from Lemma 5 that $\lambda_j^{-1}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{2n} = o_p(1)$ under (D-i). Then, it holds for $j=1,2$, and $j'=3,...,m$, that $\lambda_j||n^{-1/2}\boldsymbol{z}_j||^2 > \lambda_{j'}||n^{-1/2}\boldsymbol{z}_{j'}||^2$ and $\lambda_j||n^{-1/2}\boldsymbol{z}_j||^2 > \boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{1n}$ with probability going to 1. Then, we have under (D-i) that

$$\frac{\hat{\lambda}_j-\kappa}{\lambda_j} = \frac{\hat{\boldsymbol{u}}_j^T(\boldsymbol{U}_1+\boldsymbol{U}_{22})\hat{\boldsymbol{u}}_j}{\lambda_j} = \sum_{s=1}^2\frac{\lambda_s||n^{-1/2}\boldsymbol{z}_s||^2(\hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_s)^2}{\lambda_j} + o_p(1) = 1+o_p(1)$$

for $j=1,2$. Then, there exist random variables $\varepsilon_{1j},\varepsilon_{2j},\varepsilon_{3j}\in[-1,1]$ and a random unit vector $\boldsymbol{y}_j$ such that $\hat{\boldsymbol{u}}_j = \varepsilon_{1j}\tilde{\boldsymbol{z}}_1 + \varepsilon_{2j}\tilde{\boldsymbol{z}}_{2(1)} + \varepsilon_{3j}\boldsymbol{y}_j$ and $\tilde{\boldsymbol{z}}_1^T\boldsymbol{y}_j =$

$\tilde{\boldsymbol{z}}_{2(1)}^T\boldsymbol{y}_j = 0$ for $j = 1, 2$, where $\tilde{\boldsymbol{z}}_{2(1)} = \{\tilde{\boldsymbol{z}}_2 - (\tilde{\boldsymbol{z}}_1^T\tilde{\boldsymbol{z}}_2)\tilde{\boldsymbol{z}}_1\}/||\tilde{\boldsymbol{z}}_2 - (\tilde{\boldsymbol{z}}_1^T\tilde{\boldsymbol{z}}_2)\tilde{\boldsymbol{z}}_1||$. Note that $\varepsilon_{1j}^2 + \varepsilon_{2j}^2 = 1 + o_p(1)$, $\varepsilon_{3j} = o_p(1)$ and $\varepsilon_{3j}^2 = 1 - \varepsilon_{1j}^2 - \varepsilon_{2j}^2$. Hence, from Lemma 8, it holds for $j = 1, 2$, that

$$\frac{\hat{\lambda}_j - \kappa}{\lambda_j} = \sum_{s=1}^{2}\{1 + o_p(1)\}\varepsilon_{sj}^2 + \varepsilon_{3j}^2\sum_{s=3}^{m}\frac{\lambda_s}{\lambda_j}||n^{-1/2}\boldsymbol{z}_j||^2(\boldsymbol{y}_j^T\tilde{\boldsymbol{z}}_s)^2 + \varepsilon_{3j}^2 o_p(1) + \varepsilon_{3j}o_p(n^{-1/4}) + o_p(n^{-1/2}). \tag{A.7}$$

Note that $\lambda_j^{-1}\sum_{s=3}^m \lambda_s||n^{-1/2}\boldsymbol{z}_j||^2(\boldsymbol{y}_j^T\tilde{\boldsymbol{z}}_s)^2 < 1$ with probability going to 1 for $j = 1, 2$. Then, we obtain $\varepsilon_{3j} = o_p(n^{-1/4})$, $j = 1, 2$, from the fact that $\hat{\lambda}_j = \max(\boldsymbol{e}_{1n}^T\boldsymbol{S}_D\boldsymbol{e}_{1n})$ with respect to any $\boldsymbol{e}_{1n}$, provided that $\boldsymbol{e}_{1n}^T\hat{\boldsymbol{u}}_i = 0,\ i = 1, ..., j-1$. Thus similarly to (A.6), it holds for $j = 1, 2$, that

$$0 = \hat{\boldsymbol{u}}_j^T(\boldsymbol{U}_1 + \boldsymbol{U}_{22})\hat{\boldsymbol{u}}_3/\lambda_j = \sum_{i=1}^{2}\{\varepsilon_{ij} + o_p(1)\}\tilde{\boldsymbol{z}}_i^T\hat{\boldsymbol{u}}_3 + o_p(n^{-1/4}\lambda_3/\lambda_j).$$

Similarly to the proof of Lemma 9, we can claim that $(\hat{\lambda}_3 - \kappa)/\lambda_3 = ||n^{-1/2}\boldsymbol{z}_3||^2 + o_p(n^{-1/2})$ and $\hat{\boldsymbol{u}}_3^T\tilde{\boldsymbol{z}}_3 = 1 + o_p(n^{-1/2})$.

Next, we consider the case that $\liminf_{d\to\infty}|\lambda_2/\lambda_1 - 1| > 0$. From Lemma 9, it holds that $(\hat{\lambda}_j - \kappa)/\lambda_j = ||n^{-1/2}\boldsymbol{z}_j||^2 + o_p(n^{-1/2}) = 1 + o_p(1),\ j = 1, 2, 3$, and $\hat{\boldsymbol{u}}_3^T\tilde{\boldsymbol{z}}_3 = 1 + o_p(n^{-1/2})$. Hence, we obtain the results by considering the convergent subsequence of $|\lambda_2/\lambda_1 - 1|$. In general cases, in a way similar to the above and the proof of Lemma 9, we can claim the results. □

Lemma 11. *Let* $\delta_j = \{(n-j)\lambda_j\}^{-1}\{tr(\boldsymbol{S}_D) - \sum_{i=1}^j \hat{\lambda}_i\} - \lambda_j^{-1}\kappa$, $j = 1, ..., m$. *Then, it holds as* $d \to \infty$ *and* $n \to \infty$ *that* $\delta_j = O_p(n^{-1})$ *for* $j\ (\le m)$ *under (D-i).*

Proof of Lemma 11. Note that $\mathrm{tr}(\boldsymbol{S}_D) = \sum_{s=1}^d \lambda_s \sum_{k=1}^n z_{sk}^2/n$ and $\mathrm{tr}(\boldsymbol{U}_2) = \sum_{s=m+1}^d \lambda_s \sum_{k=1}^n z_{sk}^2/n$. By using Chebyshev's inequality, for any $\tau > 0$, one has under (C-i) that $P(\lambda_j^{-1}|n^{-1}\mathrm{tr}(\boldsymbol{U}_2) - \kappa| > \tau n^{-1}) = O\{K_1/(n\lambda_j^2)\} \to 0$. Note that $\lambda_j^{-1}n^{-1}\sum_{s=j+1}^m \lambda_s \sum_{k=1}^n z_{sk}^2/n = O_p(n^{-1})$ for $j = 1, ..., m-1$. Let $\omega_j = \sum_{s=1}^j \lambda_s \sum_{k=1}^n z_{sk}^2/n$, $j = 1, ..., m$. Then, it holds that

$$\lambda_j^{-1}\{n^{-1}\mathrm{tr}(\boldsymbol{S}_D) - n^{-1}\omega_j - \kappa\} = O_p(n^{-1}). \tag{A.8}$$

Here, from Lemma 5, it holds that $\lambda_j^{-1}\boldsymbol{e}_{1n}^T(\boldsymbol{S}_D - \kappa\boldsymbol{I}_n)\boldsymbol{e}_{1n} = \lambda_j^{-1}\boldsymbol{e}_{1n}^T\boldsymbol{U}_1\boldsymbol{e}_{1n} + o_p(1)$ under (D-i). Note that $\mathrm{tr}(\sum_{s=1}^j \lambda_s\boldsymbol{z}_s\boldsymbol{z}_s^T/n) = \omega_j$. Then, we can claim that $\omega_j/\lambda_j + o_p(1) \le \sum_{i=1}^{j}(\hat{\lambda}_i - \kappa)/\lambda_j \le \mathrm{tr}(\boldsymbol{U}_1)/\lambda_j + o_p(1)$. Thus it holds that $\lambda_j^{-1}\{\omega_j - \sum_{i=1}^{j}(\hat{\lambda}_i - \kappa)\} = O_p(1)$. Then, from (A.8), we have that

$$\begin{aligned}\delta_j &= \frac{\mathrm{tr}(\boldsymbol{S}_D) - \omega_j + \omega_j - \sum_{i=1}^{j}(\hat{\lambda}_i - \kappa)}{(n-j)\lambda_j} - \frac{n\kappa}{(n-j)\lambda_j}\\ &= \Big(\frac{n}{n-j}\Big)\frac{n^{-1}\mathrm{tr}(\boldsymbol{S}_D) - n^{-1}\omega_j - \kappa}{\lambda_j} + O_p(n^{-1}) = O_p(n^{-1}) \quad \text{for } j\ (\le m).\end{aligned}$$

It concludes the result. □

Lemma 12. *Under (D-ii), it holds as $d \to \infty$ and $n \to \infty$ that*

$\tilde{\lambda}_j/\lambda_j = 1 + o_p(1)$ *for* $j = 1, ..., m$;
$\tilde{\boldsymbol{u}}_{j(i)}^T \tilde{\boldsymbol{z}}_{ij'} = o_p(n^{-1/4}\lambda_j/\lambda_{j'})$ $(i = 1, 2)$ *for* $j' < j\ (\le m)$ *satisfying (C-iii)*;
$\tilde{\lambda}_j/\lambda_j = ||n_1^{-1/2}\boldsymbol{z}_{1j}|| \cdot ||n_2^{-1/2}\boldsymbol{z}_{2j}|| + o_p(n^{-1/2})$ *and* $\tilde{\boldsymbol{u}}_{j(i)}^T \tilde{\boldsymbol{z}}_{ij} = 1 + o_p(n^{-1/2})$
$(i = 1, 2)$ *for* $j\ (\le m)$ *satisfying (C-iii),*

where $\tilde{\boldsymbol{z}}_{ij} = ||n_i^{-1/2}\boldsymbol{z}_{ij}||^{-1} n_i^{-1/2}\boldsymbol{z}_{ij}$, $i = 1, 2$; $j = 1, ..., m$.

Proof of Lemma 12. By using Lemma 4, under (D-ii) with $j = 1$, we have that

$$\begin{aligned}\tilde{\lambda}_1/\lambda_1 = \tilde{\boldsymbol{u}}_{1(1)}^T \boldsymbol{S}_{D(1)} \tilde{\boldsymbol{u}}_{1(2)}/\lambda_1 &= \tilde{\boldsymbol{u}}_{1(1)}^T \boldsymbol{V}_1 \tilde{\boldsymbol{u}}_{1(2)}/\lambda_1 + o_p(1)\\ &= ||n_1^{-1/2}\boldsymbol{z}_{1j}|| \cdot ||n_2^{-1/2}\boldsymbol{z}_{2j}|| + o_p(1) = 1 + o_p(1)\end{aligned}$$

in case of $\liminf_{d\to\infty} |\lambda_j/\lambda_1 - 1| > 0$ for $j = 2, ..., m$. Then, there exist a random variable $\varepsilon_i \in [0, 1]$ and a random unit vector $\boldsymbol{y}_i$ such that $\tilde{\boldsymbol{u}}_{1(i)} = \tilde{\boldsymbol{z}}_{i1}(1-\varepsilon_i^2)^{1/2} + \varepsilon_i \boldsymbol{y}_i$ and $\tilde{\boldsymbol{z}}_{i1}^T \boldsymbol{y}_i = 0$ for $i = 1, 2$. Note that $(1-\varepsilon_i^2)^{1/2} = 1 - \varepsilon_i^2/2 + o_p(\varepsilon_i^2)$, $i = 1, 2$. Then, by using Lemmas 4, 6 and 7, we have that

$$\begin{aligned}\frac{\tilde{\lambda}_1}{\lambda_1} =& ||n_1^{-1/2}\boldsymbol{z}_{11}|| \cdot ||n_2^{-1/2}\boldsymbol{z}_{21}|| + \max_{\varepsilon_1,\varepsilon_2}\Big\{ - \frac{\varepsilon_1^2 + \varepsilon_2^2}{2}||n_1^{-1/2}\boldsymbol{z}_{11}|| \cdot ||n_2^{-1/2}\boldsymbol{z}_{21}||\\ &+ (\varepsilon_1 + \varepsilon_2)o_p(n^{-1/4}) + (\varepsilon_1^2 + \varepsilon_2^2)o_p(1)\\ &+ \varepsilon_1\varepsilon_2 \sum_{s=2}^{m} \frac{\lambda_s}{\lambda_1}||n_1^{-1/2}\boldsymbol{z}_{1s}|| \cdot ||n_2^{-1/2}\boldsymbol{z}_{2s}||(\boldsymbol{y}_1^T\tilde{\boldsymbol{z}}_{1s})(\boldsymbol{y}_2^T\tilde{\boldsymbol{z}}_{2s})\Big\} + o_p(n^{-1/2})\end{aligned}$$

in case of $\liminf_{d\to\infty} |\lambda_j/\lambda_1 - 1| > 0$ for $j = 2, ..., m$. Then, by noting that $(\varepsilon_1^2 + \varepsilon_2^2)/2 \ge \varepsilon_1\varepsilon_2$, similarly to (A.7), we have that $\tilde{\lambda}_1/\lambda_1 = ||n_1^{-1/2}\boldsymbol{z}_{11}|| \cdot$

$||n_2^{-1/2}\boldsymbol{z}_{21}|| + o_p(n^{-1/2})$ together with that $\varepsilon_i = o_p(n^{-1/4})$ and $\tilde{\boldsymbol{u}}_{1(i)}^T\tilde{\boldsymbol{z}}_{i1} = 1 + o_p(n^{-1/2})$, $i = 1, 2$. Thus, under (D-ii) with $j = 2$, it holds for $i' \neq i$ that

$$0 = \lambda_1^{-1}\tilde{\boldsymbol{u}}_{1(i)}^T\boldsymbol{S}_{D(i)}\tilde{\boldsymbol{u}}_{2(i')} = \{1 + o_p(1)\}\tilde{\boldsymbol{z}}_{i'1}^T\tilde{\boldsymbol{u}}_{2(i')} + o_p(n^{-1/4}\lambda_2/\lambda_1),$$

so that $\tilde{\boldsymbol{z}}_{i1}^T\tilde{\boldsymbol{u}}_{2(i)} = o_p(n^{-1/4}\lambda_2/\lambda_1)$, $i = 1, 2$, and $\tilde{\boldsymbol{u}}_{2(1)}^T\boldsymbol{S}_{D(1)}\tilde{\boldsymbol{u}}_{2(2)}/\lambda_2 = \sum_{s=2}^m \lambda_s||n_1^{-1/2}\boldsymbol{z}_{1s}||\cdot||n_2^{-1/2}\boldsymbol{z}_{2s}||(\tilde{\boldsymbol{u}}_{2(1)}^T\tilde{\boldsymbol{z}}_{1s})(\tilde{\boldsymbol{u}}_{2(2)}^T\tilde{\boldsymbol{z}}_{2s})/\lambda_2 + o_p(1) = 1 + o_p(1)$. In general cases, in a way similar to the proofs of Lemmas 9 and 10, we can claim the results. □

Proof of Proposition 2.1. For $k > 1/\alpha$, we have as $d \to \infty$ that $\lambda_j^{-k}\sum_{i=j+1}^d \lambda_i^k \leq d\lambda_{j+1}^k/\lambda_j^k = O(d^{1-\alpha k}) \to 0$. Thus there exist m and k_m satisfying (5). It concludes the result. □

Proofs of Theorems 3.1 to 3.4. Note that $\kappa/\lambda_j \to 0$ under (C-ii). Thus from Lemma 10, we can claim (4) under (C-ii) and (D-i). On the other hand, we can claim (7) from Lemma 10. Thus it concludes the results of Theorems 3.1 and 3.3.

Next, we consider Theorems 3.2 and 3.4. Note that $\kappa/\lambda_j = o(n^{-1/2})$ under (C-iv). Also, note that (C-iv) implies (D-i) from the facts that $K_1 = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})^2\}$ and $K_2 = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})^4\}$. Here, we recall that $\mathrm{Var}(z_{jk}^2) = M_j$. By using the central limiting theorem, one has that $(n/M_j)^{1/2}(||n^{-1/2}\boldsymbol{z}_j||^2 - 1) = (nM_j)^{-1/2}\sum_{k=1}^n(z_{jk}^2 - 1) \Rightarrow N(0, 1)$ under $\liminf M_j > 0$. Hence, under (C-iii) and (C-iv), we have from Lemma 10 that

$$(n/M_j)^{1/2}(\hat{\lambda}_j/\lambda_j - 1) = (n/M_j)^{1/2}(||n^{-1/2}\boldsymbol{z}_j||^2 - 1) + o_p(1) \Rightarrow N(0, 1).$$

It concludes the results. □

Proofs of Corollaries 3.1 and 3.2. From Lemma 1, (C-vi) implies (C-i) and (C-v) under (A-i). Thus the results are obtained straightforwardly. □

Proofs of Corollaries 3.3 and 4.1. In a way similar to the proof of Corollary 1 given in Yata and Aoshima [22], we can obtain the results. □

Proofs of Theorems 4.1 to 4.3. We write that $\acute{\lambda}_j/\lambda_j = (\hat{\lambda}_j - \kappa)/\lambda_j - \delta_j$, where δ_j is given in Lemma 11. Then, by combining Lemma 11 with Lemma 10, the results are obtained straightforwardly. □

Proofs of Theorems 5.1 to 5.3. From the facts that $n/n_i = 2+o(1),\ i = 1, 2$, we have that

$$||n_1^{-1/2}\boldsymbol{z}_{1j}|| \cdot ||n_2^{-1/2}\boldsymbol{z}_{2j}|| - 1 = \sum_{i=1}^{2}(||n_i^{-1/2}\boldsymbol{z}_{ij}||^2 - 1)/2 + o_p(n^{-1/2})$$
$$= \sum_{i=1}^{2}\sum_{k=1}^{n_i}(z_{ijk}^2 - 1)/(2n_i) + o_p(n^{-1/2}) = \sum_{k=1}^{n}(z_{jk}^2 - 1)/n + o_p(n^{-1/2}). \quad \text{(A.9)}$$

Then, it holds that $(n/M_j)^{1/2}(||n_1^{-1/2}\boldsymbol{z}_{1j}|| \cdot ||n_2^{-1/2}\boldsymbol{z}_{2j}|| - 1) \Rightarrow N(0, 1)$. Then, by using Lemma 12, the results are obtained straightforwardly. □

Proof of Corollary 5.1. In a way similar to the proof of Corollary 2 given in Yata and Aoshima [21], we can obtain the results. □

Proof of Proposition 5.1. We first consider the case that there exists a fixed integer j' such that $\limsup_{d\to\infty}\sum_{s=j'+1}^{d}\lambda_s^2/\lambda_{j'}^2 < \infty$ and $\lim_{d\to\infty}\lambda_{j'+1}/\lambda_{j'} = 0$. Here, we set $m = j'$. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^4)/\lambda_m^4 \le \lambda_{m+1}^2\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_m^4 = o(1)$. Then, from Lemma 4, we can claim that $\boldsymbol{e}_{1n_1}^T\boldsymbol{V}_2\boldsymbol{e}_{2n_2}/\lambda_m = o_p(1)$ as $d \to \infty$ and $n \to \infty$ under (A-i), so that $\boldsymbol{e}_{1n_1}^T\boldsymbol{S}_{D(1)}\boldsymbol{e}_{2n_2}/\lambda_m = \boldsymbol{e}_{1n_1}^T\boldsymbol{V}_1\boldsymbol{e}_{2n_2}/\lambda_m + o_p(1)$. Let $\widetilde{\boldsymbol{V}}_1 = \boldsymbol{V}_1 - \sum_{s=1}^{m}\tilde{\lambda}_s\tilde{\boldsymbol{u}}_{s(1)}\tilde{\boldsymbol{u}}_{s(2)}^T$. Then, it holds that $\boldsymbol{e}_{1n_1}^T\widetilde{\boldsymbol{V}}_1\boldsymbol{e}_{2n_2}/\lambda_m = o_p(1)$, so that all the singular values of $\widetilde{\boldsymbol{V}}_1/\lambda_m$ are of the order $o_p(1)$. Then, from the fact that $\mathrm{rank}(\widetilde{\boldsymbol{V}}_1) \le 2m$, it holds that $\mathrm{tr}(\widetilde{\boldsymbol{V}}_1\widetilde{\boldsymbol{V}}_1^T)/\lambda_m^2 = o_p(1)$. Note that $E\{\mathrm{tr}(\boldsymbol{V}_2\boldsymbol{V}_2^T)\} = \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)$. Here, we can claim that $\mathrm{Var}\{\mathrm{tr}(\boldsymbol{V}_2\boldsymbol{V}_2^T)/\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)\} \to 0$ under (A-i), so that $\mathrm{tr}(\boldsymbol{V}_2\boldsymbol{V}_2^T) = \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)\{1+o_p(1)\}$. Then, it holds that $|\mathrm{tr}(\widetilde{\boldsymbol{V}}_1\boldsymbol{V}_2^T)|/\lambda_m^2 \le \mathrm{tr}(\widetilde{\boldsymbol{V}}_1\widetilde{\boldsymbol{V}}_1^T)^{1/2}\mathrm{tr}(\boldsymbol{V}_2\boldsymbol{V}_2^T)^{1/2}/\lambda_m^2 = o_p(1)$. Hence, we obtain for $j\ (\le m)$ that

$$\sum_{s=m+1}^{n_2}\frac{\tilde{\lambda}_s^2}{\lambda_j^2} = \frac{\mathrm{tr}\{(\widetilde{\boldsymbol{V}}_1 + \boldsymbol{V}_2)(\widetilde{\boldsymbol{V}}_1 + \boldsymbol{V}_2)^T\}}{\lambda_j^2} = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{\lambda_j^2} + o_p(1). \quad \text{(A.10)}$$

Here, from Lemma 12, it holds that $\tilde{\lambda}_j^2/\lambda_j^2 = 1 + o_p(1)$ for $j \le m$. On the other hand, we have for $j > m$ that $\tilde{\lambda}_j/\lambda_m = \tilde{\boldsymbol{u}}_{j(1)}^T\boldsymbol{V}_1\tilde{\boldsymbol{u}}_{j(2)}/\lambda_m + o_p(1) = o_p(1)$ from the fact that $\mathrm{rank}(\boldsymbol{V}_1) \le m$. Note that $\mathrm{tr}(\boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T) - \sum_{s=1}^{m}\tilde{\lambda}_s^2 = \sum_{s=m+1}^{n_2}\tilde{\lambda}_s^2$. Thus we have for $j \le m$ and any fixed $j_*\ (\ge m)$ that

$$\frac{\mathrm{tr}(\boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T) - \sum_{s=1}^{j_*}\tilde{\lambda}_s^2}{\tilde{\lambda}_j^2} = \sum_{s=m+1}^{n_2}\frac{\tilde{\lambda}_s^2}{\tilde{\lambda}_j^2} + o_p(1) = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{\lambda_j^2} + o_p(1). \quad \text{(A.11)}$$

Next, we consider the case that there exists a fixed integer j' such that $\lim_{d\to\infty} \lambda_{j'}^2/\sum_{s=j'+1}^d \lambda_s^2 = 0$. We set $m = j'$. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^4)/\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2 \le \lambda_{m+1}^2/\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2) = o(1)$. Then, from Lemma 4, we can obtain under (A-i) that

$$\boldsymbol{u}_{j(1)}^T \frac{\boldsymbol{S}_{D(1)}}{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^{1/2}} \boldsymbol{u}_{j(2)} = \boldsymbol{u}_{j(1)}^T \frac{\sum_{s=1}^{m-1} \lambda_s \boldsymbol{z}_{1s} \boldsymbol{z}_{2s}^T}{(n_1 n_2)^{1/2} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^{1/2}} \boldsymbol{u}_{j(2)} + o_p(1),$$

so that $\tilde{\lambda}_j/\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^{1/2} = o_p(1)$ for $j \ge m$. Hence, in a way similar to (A.10), we have for any fixed $j \ge m$ and j_* $(\ge j)$ that

$$\frac{\mathrm{tr}(\boldsymbol{S}_{D(1)} \boldsymbol{S}_{D(1)}^T) - \sum_{s=1}^{j_*} \tilde{\lambda}_s^2}{\tilde{\lambda}_j^2} \ge \frac{\mathrm{tr}\{(\widetilde{\boldsymbol{V}}_1 + \boldsymbol{V}_2)(\widetilde{\boldsymbol{V}}_1 + \boldsymbol{V}_2)^T\}}{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)} + o_p(1) = 1 + o_p(1). \tag{A.12}$$

When it holds that $\{\mathrm{tr}(\boldsymbol{S}_{D(1)} \boldsymbol{S}_{D(1)}^T) - \sum_{s=1}^{j_*} \tilde{\lambda}_s^2\}/\tilde{\lambda}_j^2 = o_p(1)$ for fixed j and $j_* \ge j$, we can claim that $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 = 0$ with some m $(\ge j)$ by combining (A.11) with (A.12). Thus it concludes the result. □

Proofs of Theorem 7.1 and Corollary 7.1. From Lemma 10, we have that $\boldsymbol{h}_j^T \hat{\boldsymbol{h}}_j = (n\hat{\lambda}_j)^{-1/2} \lambda_j^{1/2} \boldsymbol{z}_j^T \hat{\boldsymbol{u}}_j = (\lambda_j/\hat{\lambda}_j)^{1/2} n^{-1/2} \boldsymbol{z}_j^T \hat{\boldsymbol{u}}_j = (1 + \kappa/\lambda_j)^{-1/2} + o_p(1)$ under (C-iii) and (D-i). It concludes the result of Corollary 7.1. On the other hand, from $\kappa/\lambda_j \to 0$ under (C-ii), it concludes the result of Theorem 7.1. □

Proof of Theorem 7.2. By using Lemma 12, we have that

$$\boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_{j(i)} = (\lambda_j/\tilde{\lambda}_j)^{1/2} n_i^{-1/2} \boldsymbol{z}_{ij}^T \tilde{\boldsymbol{u}}_{j(i)} = 1 + o_p(1) \quad (i = 1, 2).$$

Let $\boldsymbol{U}_{22(i)} = n_i^{-1} \sum_{s=m+1}^d \lambda_s \boldsymbol{z}_{is} \boldsymbol{z}_{is}^T - \kappa \boldsymbol{I}_{n_i}$, $i = 1, 2$. Note that $||\tilde{\boldsymbol{h}}_{j(i)}||^2 = (n_i \tilde{\lambda}_j)^{-1} \tilde{\boldsymbol{u}}_{j(i)}^T \boldsymbol{X}_i^T \boldsymbol{X}_i \tilde{\boldsymbol{u}}_{j(i)}$. From Lemma 12, we have that

$$\begin{aligned} ||\tilde{\boldsymbol{h}}_{j(i)}||^2 &= \sum_{s=j}^m \lambda_s (\tilde{\boldsymbol{u}}_{j(i)}^T \boldsymbol{z}_{is}/n_i^{1/2})^2/\tilde{\lambda}_j + \tilde{\boldsymbol{u}}_{j(i)}^T \boldsymbol{U}_{22(i)} \tilde{\boldsymbol{u}}_{j(i)}/\tilde{\lambda}_j + \kappa/\tilde{\lambda}_j + o_p(1) \\ &= 1 + o_p(1) + \tilde{\boldsymbol{u}}_{j(i)}^T \boldsymbol{U}_{22(i)} \tilde{\boldsymbol{u}}_{j(i)}/\tilde{\lambda}_j \quad (i = 1, 2) \end{aligned}$$

under (C-ii), (C-iii) and (D-ii). Note that $K_1/(n\lambda_j)^2 = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})^2/(n\lambda_j)^2\} \to 0$ under (C-ii). In a way similar to the proof of Lemma 8, we can claim that

$$||(n_i\lambda_j)^{-1} \boldsymbol{z}_{ij}^T \boldsymbol{U}_{22(i)}||^2 = o_p(1) \quad \text{and} \quad (n_i\lambda_j)^{-1} \boldsymbol{z}_{ij}^T \boldsymbol{U}_{22(i)} \boldsymbol{z}_{ij} = o_p(1) \tag{A.13}$$

under (C-ii) and (D-ii). From the fact that $\tilde{\boldsymbol{u}}_{j(i)}^T\tilde{\boldsymbol{z}}_{ij} = 1+o_p(n^{-1/2})$ in Lemma 12, there exists a random unit vector $\boldsymbol{y}_i$ such that $\tilde{\boldsymbol{u}}_{j(i)} = \tilde{\boldsymbol{z}}_{ij}\{1+o_p(n^{1/2})\}+\boldsymbol{y}_i o_p(n^{1/2})$ and $\tilde{\boldsymbol{z}}_{ij}^T\boldsymbol{y}_i = 0$. From Lemma 5, it holds that $\lambda_j^{-1}n^{-1}\boldsymbol{e}_{in_i}^T\boldsymbol{U}_{22(i)}\boldsymbol{e}_{in_i} = o_p(1)$ under (C-ii) and (D-ii). Thus from (A.13), it holds that $\tilde{\boldsymbol{u}}_{j(i)}^T\boldsymbol{U}_{22(i)}\tilde{\boldsymbol{u}}_{j(i)}/\tilde{\lambda}_j = o_p(1)$, so that $||\tilde{\boldsymbol{h}}_{j(i)}||^2 = 1+o_p(1)$ $(i=1,2)$. Then, we have that $\boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_j = 1+o_p(1)$ and $||\tilde{\boldsymbol{h}}_j||^2 = 1+o_p(1)$. Thus it concludes the result. □

Proofs of Theorems 8.1 to 8.3 and Corollary 8.1. For j $(\le m)$, we write that

$$\begin{aligned}\mathrm{MSE}(\hat{s}_j)/\lambda_j &= n^{-1}\sum_{k=1}^{n}\{z_{jk}-(n\hat{\lambda}_j/\lambda_j)^{1/2}\hat{u}_{jk}\}^2\\ &= ||n^{-1/2}\boldsymbol{z}_j||^2+\hat{\lambda}_j/\lambda_j-2(\hat{\lambda}_j/\lambda_j)^{1/2}||n^{-1/2}\boldsymbol{z}_j||\tilde{\boldsymbol{z}}_j^T\hat{\boldsymbol{u}}_j. \qquad \text{(A.14)}\end{aligned}$$

With the help of Lemma 10, we have that $\mathrm{MSE}(\hat{s}_j)/\lambda_j = o_p(1)$ under (C-ii), (C-iii) and (D-i). It concludes the result of Theorem 8.1. On the other hand, under (C-iii) and (C-iv), it holds that $\hat{\lambda}_j/\lambda_j = ||n^{-1/2}\boldsymbol{z}_j||^2+o_p(n^{-1/2})$ and $\tilde{\boldsymbol{z}}_j^T\hat{\boldsymbol{u}}_j = 1+o_p(n^{-1/2})$. Then, from (A.14), it holds that $\mathrm{MSE}(\hat{s}_j)/\lambda_j = o_p(n^{-1/2})$. It concludes the result of Corollary 8.1. Similarly, we can obtain the results of Theorems 8.2 and 8.3 by using (A.9). □

Acknowledgment

The authors would like to thank three anonymous referees for their constructive comments. Research of the first author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 23740066. Research of the second author was partially supported by Grants-in-Aid for Scientific Research (B) and Challenging Exploratory Research, JSPS, under Contract Numbers 22300094 and 23650142.

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] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, A.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proc. Natl. Acad. Sci. USA 96 (1999) 6745-6750.

[3] M. Aoshima, K. Yata, Two-stage procedures for high-dimensional data, Sequential Anal. (*Editor's special invited paper*) 30 (2011) 356-399.

[4] M. Aoshima, K. Yata, Authors' response, Sequential Anal. 30 (2011) 432-440.

[5] Z. Bai, H. Saranadasa, Effect of high dimension: By an example of a two sample problem, Statist. Sinica 6 (1996) 311-329.

[6] Z. Bai, X. Ding, Estimation of spiked eigenvalues in spiked models, Random Matrices: Theory Appl. 1 (2012) [21 pages].

[7] J. Baik, G. Ben Arous, S. Péché, Phase transition of the largest eigenvalue for non-null complex sample covariance matrices, Ann. Probab. 33 (2005) 1643-1697.

[8] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivariate Anal. 97 (2006) 1382-1408.

[9] P.J. Bickel and E. Levina, Regularized estimation of large covariance matrices, Ann. Statist. 36 (2008), 199-227.

[10] S.X. Chen, Y.-L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, Ann. Statist. 38 (2010) 808-835.

[11] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfield, E. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531-537.

[12] P. Hall, J.S. Marron, A. Neeman, Geometric representation of high dimension, low sample size data, J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2005) 427-444.

[13] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2001) 295-327.

[14] I.M. Johnstone, Y. Lu, On consistency and sparsity for principal components analysis in high dimensions, J. Amer. Statist. Assoc. 104 (2009) 682-693.

[15] S. Jung, J.S. Marron, PCA consistency in high dimension, low sample size context, Ann. Statist. 37 (2009) 4104-4130.

[16] S. Jung, A. Sen, J.S. Marron, Boundary behavior in high dimension, low sample size asymptotics of PCA, J. Multivariate Anal. 109 (2012) 190-203.

[17] S. Lee, F. Zou, F.A. Wright, Convergence and prediction of principal component scores in high-dimensional settings, Ann. Statist. 38 (2010) 3605-3629.

[18] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statist. Sinica 17 (2007) 1617-1642.

[19] 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 (2002) 203-209.

[20] K. Yata, M. Aoshima, PCA consistency for non-Gaussian data in high dimension, low sample size context, Commun. Statist. Theory Methods, Special Issue Honoring S. Zacks (Ed. N. Mukhopadhyay) 38 (2009) 2634-2652.

[21] K. Yata, M. Aoshima, Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix, J. Multivariate Anal. 101 (2010) 2060-2077.

[22] K. Yata, M. Aoshima, Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations, J. Multivariate Anal. 105 (2012) 193-215.