Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations

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

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

Abstract

In this article, we propose a new estimation methodology to deal with PCA for high-dimension, low-sample-size (HDLSS) data. We first show that HDLSS datasets have different geometric representations depending on whether a ρ-mixing-type dependency appears in variables or not. When the ρ-mixing-type dependency appears in variables, the HDLSS data converge to an n-dimensional surface of unit sphere with increasing dimension. We pay special attention to this phenomenon. We propose a method called *the noise-reduction methodology* to estimate eigenvalues of a HDLSS dataset. We show that the eigenvalue estimator holds consistency properties along with its limiting distribution in HDLSS context. We consider consistency properties of PC directions. We apply the noise-reduction methodology to estimating PC scores. We also give an application in the discriminant analysis for HDLSS datasets by using the inverse covariance matrix estimator induced by the noise-reduction methodology.

Key words: Consistency; Discriminant analysis; Eigenvalue distribution; Geometric representation; HDLSS; Inverse matrix; Noise reduction; Principal component analysis.

1. Introduction

The high-dimension, low-sample-size (HDLSS) data situation occurs in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance,

Email address: aoshima@math.tsukuba.ac.jp (Makoto Aoshima)
[1]Phone & Fax: +81-298-53-6501

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 in the limit as $d \to \infty$ was studied by Johnstone [6], Baik et al. [2] and Paul [10] under Gaussian assumptions, and Baik and Silverstein [3] under non-Gaussian but i.i.d. assumptions when the dimension d and the sample size n increase at the same rate, i.e. $n/d \to c > 0$. In recent years, substantial work has been done on the HDLSS asymptotic theory, where only $d \to \infty$ while n is fixed, by Hall et al. [5], Ahn et al. [1], Jung and Marron [7], and Yata and Aoshima [14], [15] and [16]. Hall et al. [5] and Ahn et al. [1] explored conditions to give a geometric representation of HDLSS data. Jung and Marron [7] investigated consistency properties of both eigenvalues and eigenvectors of the sample covariance matrix in the HDLSS data situations. The HDLSS asymptotic theory had been created under the assumption that either the population distribution is normal or the random variables in the sphered data matrix have the ρ-mixing dependency (see Bradley [4]). However, Yata and Aoshima [14], [15] and [16] developed the HDLSS asymptotic theory without assuming either the normality or the ρ-mixing condition. Yata and Aoshima [14] gave consistency properties of both eigenvalues and eigenvectors of the sample covariance matrix together with PC scores. Yata and Aoshima [15] proposed a method for dimensionality estimation of HDLSS data, and Yata and Aoshima [16] generalized the method to create a new PCA called the cross-data-matrix methodology.

In this paper, suppose we have a $d \times n$ data matrix $\boldsymbol{X}_{(d)} = [\boldsymbol{x}_{1(d)}, ..., \boldsymbol{x}_{n(d)}]$ with $d > n$, 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 nonnegative 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 $\lambda_{1(d)} \geq \cdots \geq \lambda_{d(d)} (> 0)$ and $\boldsymbol{H}_d = [\boldsymbol{h}_{1(d)}, ..., \boldsymbol{h}_{d(d)}]$ is a matrix of corresponding eigenvectors. Then, $\boldsymbol{Z}_{(d)} = \boldsymbol{\Lambda}_d^{-1/2} \boldsymbol{H}_d^T \boldsymbol{X}_{(d)}$ is a $d \times n$ sphered data matrix from a distribution with the identity covariance matrix. Here, we write $\boldsymbol{Z}_{(d)} = [\boldsymbol{z}_{1(d)}, ..., \boldsymbol{z}_{d(d)}]^T$ and $\boldsymbol{z}_{j(d)} = (z_{j1(d)}, ..., z_{jn(d)})^T$, $j = 1, ..., d$. Hereafter, the subscript d will be omitted for the sake of simplicity when it does not cause any confusion. We assume that the fourth moments of each variable in $\boldsymbol{Z}$ are uniformly bounded. We assume that $||\boldsymbol{z}_j|| \neq 0$ for $j = 1, ..., d$, where $||\cdot||$ denotes the Euclidean norm. We consider a general setting as follows:

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

Here, $a_i(>0)$, $c_j(>0)$ and $\alpha_i(\alpha_1 \geq \cdots \geq \alpha_m > 0)$ are unknown constants preserving the order that $\lambda_1 \geq \cdots \geq \lambda_d$, and m is an unknown non-negative integer. We assume $n > m$.

In Section 2, we show that HDLSS datasets have different geometric representations depending on whether a ρ-mixing-type dependency appears in variables or not. When the ρ-mixing-type dependency appears in variables, the HDLSS data converge to an n-dimensional surface of unit sphere with increasing dimension. We pay special attention to this phenomenon. *After Section 3, we assume that* z_{jk}, $j = 1, ..., d$ $(k = 1, ..., n)$ *are independent.* Note that the assumption includes the case that $\boldsymbol{X}$ is Gaussian. In Section 3, we propose a method called *the noise-reduction methodology* to estimate eigenvalues of a HDLSS dataset. We show that the eigenvalue estimator holds consistency properties along with its limiting distribution in HDLSS context. In Section 4, we consider consistency properties of PC directions. In Section 5, we apply the noise-reduction methodology to estimating PC scores. In Section 6, we show performances of the noise-reduction methodology by conducting simulation experiments. In Section 7, we provide an inverse covariance matrix estimator induced by the noise-reduction methodology. Finally, in Section 8, we give an application in the discriminant analysis for HDLSS datasets by using the inverse covariance matrix estimator.

2. Geometric representations

In this section, we consider several geometric representations. 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 and $E\{(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D\} = \boldsymbol{I}_n$. Ahn et al. [1] and Jung and Marron [7] claimed that when the eigenvalues of $\boldsymbol{\Sigma}$ are sufficiently diffused in the sense that

$$\frac{\sum_{i=1}^{d}\lambda_i^2}{(\sum_{i=1}^{d}\lambda_i)^2} \to 0 \quad \text{as} \quad d \to \infty, \tag{2}$$

the sample eigenvalues behave as if they are from a scaled identity covariance matrix. When $\boldsymbol{X}$ is Gaussian or the components of $\boldsymbol{Z}$ are ρ-mixing, it follows that

$$\frac{n}{\sum_{i=1}^{d}\lambda_i}\boldsymbol{S}_D \to \boldsymbol{I}_n \tag{3}$$

in probability as $d \to \infty$ for a fixed n under (2).

Remark 1. The concept of ρ-mixing was first developed by Kolmogorov and Rozanov [8]. See Bradley [4] for a clear and insightful discussion. See also Jung and Marron [7]. For $-\infty \leq J \leq K \leq \infty$, let $\mathcal{F}_J^K$ denote that the σ-field of events generated by the random variables $(Y_i,\, J \leq i \leq K)$. For any σ-filed $\mathcal{A}$, let $L_2(\mathcal{A})$ denote the space of square-integrable, $\mathcal{A}$ measurable (real-valued) random variables. For each $r \geq 1$, define the maximal correlation coefficient

$$\rho(r) = \sup|\text{corr}(f, g)|, \quad f \in L_2(\mathcal{F}_{-\infty}^{j}), g \in L_2(\mathcal{F}_{j+r}^{\infty}),$$

where sup is over all f, g and j is a positive integer. The sequence $\{Y_i\}$ is said to be ρ-mixing if $\rho(r) \to 0$ as $r \to \infty$. Note that when $(z_{1k}, z_{2k}, ...)$ is ρ-mixing, it holds that for $j, j' = 1, 2, ...$ with $|j - j'| = r$,

$$|E\left((z_{jk}^2 - 1)(z_{j'k}^2 - 1)\right)| \leq \rho(r) \to 0 \quad \text{as } r \to \infty.$$

Remark 2. Let $\boldsymbol{R}_n = \{\boldsymbol{e}_n \in \boldsymbol{R}^n : ||\boldsymbol{e}_n|| = 1\}$. Let $\boldsymbol{w}_j = (n/\sum_{i=1}^{d}\lambda_i)\boldsymbol{S}_D\hat{\boldsymbol{u}}_j = (n/\sum_{i=1}^{d}\lambda_i)\hat{\lambda}_j\hat{\boldsymbol{u}}_j$. When $\boldsymbol{X}$ is Gaussian or the components of $\boldsymbol{Z}$ are ρ-mixing, it holds from (3) that

$$\boldsymbol{w}_j \in \boldsymbol{R}_n, \quad j = 1, ..., n \tag{4}$$

in probability as $d \to \infty$ for a fixed n under (2).

When $\boldsymbol{X}$ is non-Gaussian without ρ-mixing, Yata and Aoshima [15] claimed that

$$\frac{n}{\sum_{i=1}^{d}\lambda_i}\boldsymbol{S}_D \to \boldsymbol{D}_n \tag{5}$$

in probability as $d \to \infty$ for a fixed n under (2), where $\boldsymbol{D}_n$ is a diagonal matrix with any diagonal element having $O_p(1)$.

Now, let us further consider the geometric representations given by (3) and (5). Let $\boldsymbol{z}_{k*} = (z_{1k}^2 - 1, ..., z_{dk}^2 - 1)^T,\ k = 1, ..., n$. We denote the covariance matrix of $\boldsymbol{z}_{k*}$ by $\boldsymbol{\Phi}$. Note that when $\boldsymbol{X}$ is Gaussian (or $z_{jk},\ j = 1, ..., d$ $(k = 1, ..., n)$ are independent), $\boldsymbol{\Phi}$ is a diagonal matrix. Let $\boldsymbol{\Phi} = (\phi_{ij})$ and $r = |i - j|$. Note that when the components of $\boldsymbol{Z}$ are ρ-mixing, it holds that $\phi_{ij} \to 0$ as $r \to \infty$. When $\boldsymbol{X}$ is non-Gaussian without ρ-mixing, we may claim that $\phi_{ij} \neq 0$ for $i \neq j$. However, it should be noted that the geometric representation given by (3) is still claimed even in a case when $\boldsymbol{X}$ is non-Gaussian without ρ-mixing. Let us write $D_k = (\sum_{j=1}^d \lambda_j)^{-1} \sum_{j=1}^d \lambda_j z_{jk}^2$ as a diagonal element of $(n/\sum_{j=1}^d \lambda_j)\boldsymbol{S}_D$. Note that $D_k = ||\boldsymbol{x}_k||^2/\mathrm{tr}(\boldsymbol{\Sigma})$ and $E(D_k) = 1$. Let $V(x)$ denote the variance of a random variable x. We have for the variance of each D_k that

$$V(D_k) = \frac{E\left((\sum_{j=1}^d \lambda_j (z_{jk}^2 - 1))^2\right)}{(\sum_{j=1}^d \lambda_j)^2} = \frac{\sum_{i,j} \lambda_i \lambda_j \phi_{ij}}{(\sum_{j=1}^d \lambda_j)^2}.$$

Hence, we consider a regular condition of ρ-mixing-type dependency given by

$$\frac{\sum_{i,j} \lambda_i \lambda_j \phi_{ij}}{(\sum_{j=1}^d \lambda_j)^2} \to 0 \quad \text{as} \quad d \to \infty. \tag{6}$$

Note that it holds (6) under (2) when $\boldsymbol{X}$ is Gaussian or the components of $\boldsymbol{Z}$ are ρ-mixing. Then, we obtain the following theorem.

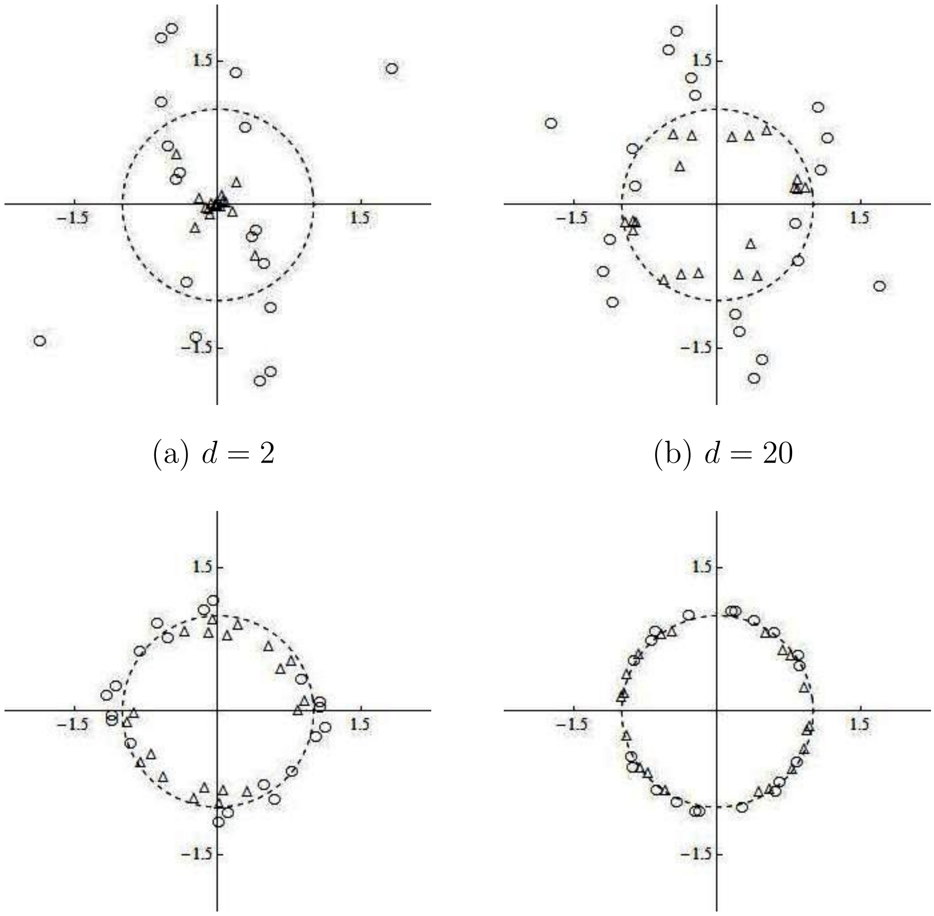
Theorem 1. *When the components of $\boldsymbol{Z}$ satisfy the condition given by (6), we have (3) as $d \to \infty$ for a fixed n. Otherwise, we have (5) as $d \to \infty$ for a fixed n under (2).*

Remark 3. We consider the case that $z_{jk},\ j = 1, ..., d$ $(k = 1, ..., n)$ are distributed as continuous distributions. Let $f(D_k)$ be the p.d.f. of D_k. Assume that $\boldsymbol{Z}$ does not satisfy (6). Assume further that $f(D_k) < \infty$ w.p.1 as $d \to \infty$. Let $\boldsymbol{R}_{n*} = \{\boldsymbol{e}_{(1)} = (1, 0, ..., 0)^T, \boldsymbol{e}_{(2)} = (0, 1, ..., 0)^T, ..., \boldsymbol{e}_{(n)} = (0, 0, ..., 1)^T\}$. Then, we have that

$$\hat{\boldsymbol{u}}_j \in \boldsymbol{R}_{n*}, \quad j = 1, ..., n. \tag{7}$$

in probability as $d \to \infty$ for a fixed n under (2).

Let us observe geometric representations induced by (3) with (4) and (5) with (7). Now, we consider an easy example such as $\lambda_1 = \cdots = \lambda_d = 1$ and $n = 2$. Note that it is satisfying (2). Figs. 1(a), 1(b), 1(c) and 1(d) give scatter plots of 20 independent pairs of $\pm \boldsymbol{w}_j$ $(j = 1, 2)$ generated from the normal distribution, $N_d(\mathbf{0}, \boldsymbol{I}_d)$, with mean zero and covariance matrix $\boldsymbol{I}_d$ in d $(= 2, 20, 200,$ and $2000)$-dimensional Euclidian space, respectively.



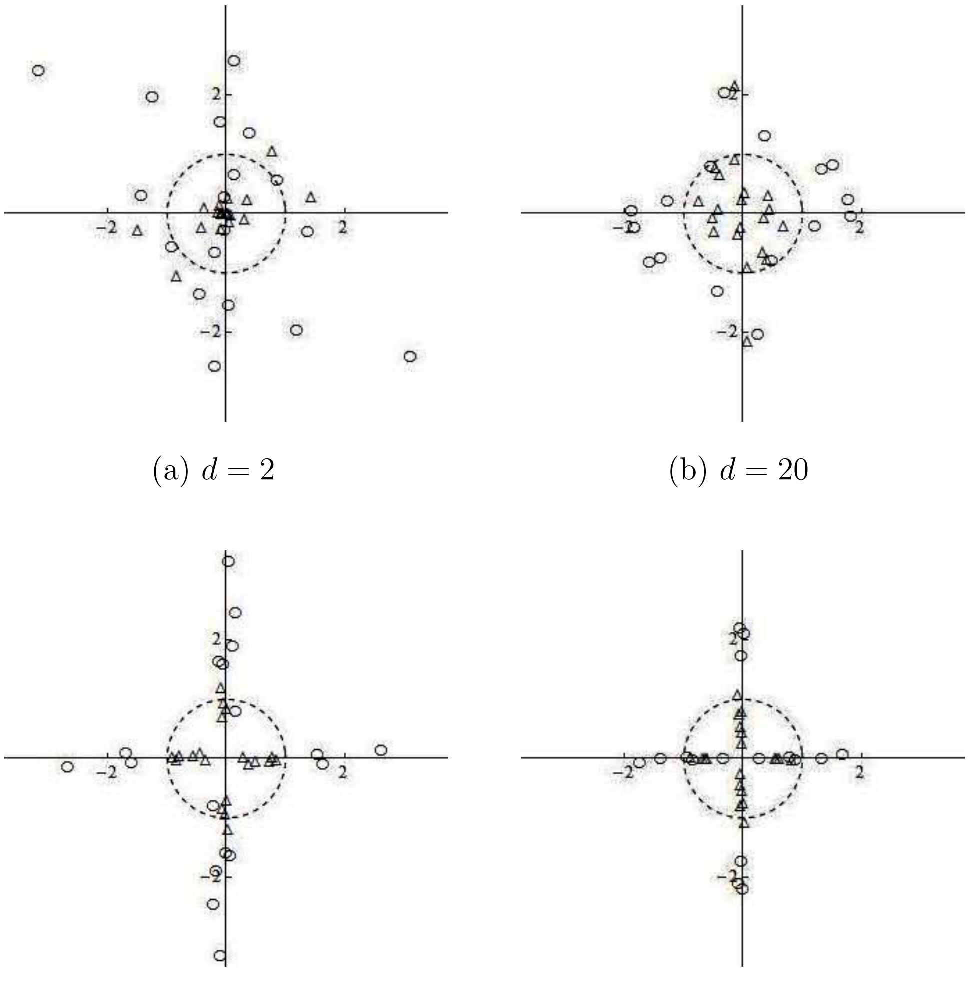
(a) $d = 2$ (b) $d = 20$

(c) $d = 200$ (d) $d = 2000$

Fig. 1. Gaussian toy example for $n = 2$, illustrating the geometric representation of $\boldsymbol{w}_1$ (plotted as $\bigcirc$) and $\boldsymbol{w}_2$ (plotted as $\triangle$), and the convergence to an n-dimensional surface of unit sphere with increasing dimension: (a) $d = 2$, (b) $d = 20$, (c) $d = 200$, and (d) $d = 2000$.

Fig. 1 shows the geometric representation induced by (3) with (4). When $d = 2$, the plots of $\boldsymbol{w}_1$ appeared quite random and the plots of $\boldsymbol{w}_2$ appeared around $\mathbf{0}$. However, when $d = 200$, the approximation in (3) with (4) became quite good. It reflected that the plots of $\boldsymbol{w}_i$ $(i = 1, 2)$ appeared around the surface of an n-dimensional unit sphere. As expected, when $d = 2000$, it showed an even more rigid geometric representation.

Figs. 2(a), 2(b), 2(c) and 2(d) give scatter plots of 20 independent pairs of $\pm\boldsymbol{w}_j$ $(j = 1, 2)$ generated from the d-variate t-distribution, $t_d(\mathbf{0}, \boldsymbol{I}_d, \nu)$ with mean zero, covariance matrix $\boldsymbol{I}_d$ and degree of freedom (d.f.) $\nu = 5$ in d $(= 2, 20, 200$, and $2000)$-dimensional Euclidian space.



(a) $d = 2$ (b) $d = 20$

(c) $d = 200$ (d) $d = 2000$

Fig. 2. Non-Gaussian toy example for $n = 2$, illustrating the geometric representation of $\boldsymbol{w}_1$

(plotted as $\bigcirc$) and $\boldsymbol{w}_2$ (plotted as $\triangle$), and the concentration on axes with increasing dimension: (a) $d = 2$, (b) $d = 20$, (c) $d = 200$, and (d) $d = 2000$.

Fig. 2 shows the geometric representation induced by (5) with (7). When $d = 2$, the plots of $\boldsymbol{w}_i$ ($i = 1, 2$) appeared quite random. When $d = 200$, the approximation in (5) with (7) became moderate. When $d = 2000$, the approximation became quite good. It reflected that the plots of $\boldsymbol{w}_i$ ($i = 1, 2$) appeared in close to axes.

Here, we consider the case that $d \to \infty$ and $n \to \infty$. Let $\boldsymbol{e}_n$ be an arbitrary element of $\boldsymbol{R}_n$ that is defined in Remark 2. Then, we have the following theorem.

Theorem 2. *We assume that*

$$n \frac{\sum_{i,j} \lambda_i \lambda_j \phi_{ij}}{(\sum_{j=1}^d \lambda_j)^2} \to 0 \quad \textit{and} \quad n^2 \frac{\sum_{i=1}^p \lambda_i^2}{(\sum_{j=1}^d \lambda_j)^2} \to 0 \tag{8}$$

when $d \to \infty$ and $n \to \infty$. Then, it holds that

$$\frac{n}{\sum_{i=1}^d \lambda_i} \boldsymbol{e}_n^T \boldsymbol{S}_D \boldsymbol{e}_n = 1 + o_p(1). \tag{9}$$

From Theorem 2, $\hat{\lambda}_j$'s are mutually equivalent under (8) in the sense that $n(\sum_{i=1}^d \lambda_i)^{-1} \hat{\lambda}_j = 1 + o_p(1)$ for all $j = 1, ..., n$. Note that $n/d \to 0$ under (8). Theorem 2 claims that a geometric representation appeared in Fig. 1 still remains even when $n \to \infty$ in the HDLSS context.

In this article, we pay special attention to the geometric representation given by (3) or (9), that is appeared in Fig. 1. *After Section 3, we assume that z_{jk}, $j = 1, ..., d$ ($k = 1, ..., n$) are independent.* This assumption is milder than that the population distribution is Gaussian. We propose a new estimation method called *the noise-reduction methodology* to deal with PCA in HDLSS data situations. When $\boldsymbol{X}$ may have the geometric representation given by (5), Yata and Aoshima [16] proposed a different method called the cross-data-matrix methodology. We compare those two methodologies by simulations in Section 6.

3. Noise-reduction methodology

Hereafter we assume that z_{jk}, $j = 1, ..., d$ ($k = 1, ..., n$) are independent. We denote n

by $n(d)$ only when $n = d^\gamma$, where γ is a positive constant. Yata and Aoshima [14] gave consistency properties of the sample eigenvalues. Their result is summarized as follows: It holds for j $(= 1, ..., m)$ that

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

under the conditions:

(YA-i) $d \to \infty$ and $n \to \infty$ for j such that $\alpha_j > 1$;

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

The condition described by both $d \to \infty$ and $n \to \infty$ is a mild condition for n in the sense that one can choose n free from d. The above result given by Yata and Aoshima [14] draws our attention to the limitations of the capabilities of naive PCA in HDLSS data situations. Let us see a case, say, that $d = 10000, \lambda_1 = d^{1/2}$ and $\lambda_2 = \cdots = \lambda_d = 1$. Then, we observe from (YA-ii) that one requires the sample of size $n >> d^{1-\alpha_1} = d^{1/2} = 100$. It is somewhat inconvenient for the experimenter to handle HDLSS data situations.

We have that $\boldsymbol{S}_D = n^{-1} \sum_{j=1}^{d} \lambda_j \boldsymbol{z}_j \boldsymbol{z}_j^T$. Let us write that $\boldsymbol{U}_1 = n^{-1} \sum_{j=1}^{m} \lambda_j \boldsymbol{z}_j \boldsymbol{z}_j^T$ and $\boldsymbol{U}_2 = n^{-1} \sum_{j=m+1}^{d} \lambda_j \boldsymbol{z}_j \boldsymbol{z}_j^T$ so that $\boldsymbol{S}_D = \boldsymbol{U}_1 + \boldsymbol{U}_2$. Here, we consider $\boldsymbol{U}_1$ as *intrinsic part* and $\boldsymbol{U}_2$ as *noise part*. Since it holds that

$$\frac{\sum_{j=m+1}^{d} \lambda_j^2}{(\sum_{j=m+1}^{d} \lambda_j)^2} \to 0 \quad \text{as} \quad d \to \infty, \tag{11}$$

the noise part holds the geometric representation similar to (3) or (9). Let $\boldsymbol{e}_n = (e_1, ..., e_n)^T$ be an arbitrary element of $\boldsymbol{R}_n$ that is defined in Remark 2. Then, from (3) and Theorem 2, we have that

$$\frac{n}{\sum_{j=m+1}^{d} \lambda_j} \boldsymbol{e}_n^T \boldsymbol{U}_2 \boldsymbol{e}_n = 1 + o_p(1) \tag{12}$$

as $d \to \infty$ either when n is fixed or $n = n(d)$ satisfying $n(d) \sum_{j=m+1}^{d} \lambda_j^2 / (\sum_{j=m+1}^{d} \lambda_j)^2 \to 0$. *This geometric representation for the noise part influences the estimation scheme proposed in this article.*

We consider an easy example such as $m=2$ and $\lambda_1 = d^{\alpha_1},\ \lambda_2 = d^{\alpha_2},\ \lambda_j = c_j,\ j=3,...,d$, where $\alpha_1 > \alpha_2 > 1/2$ and c_j's are positive constants. Note that it is satisfying (11). Then, we write that $\lambda_1^{-1}\boldsymbol{S}_D = n^{-1}\boldsymbol{z}_1\boldsymbol{z}_1^T + (n\lambda_1)^{-1}\lambda_2\boldsymbol{z}_2\boldsymbol{z}_2^T + (n\lambda_1)^{-1}\sum_{j=3}^d \lambda_j\boldsymbol{z}_j\boldsymbol{z}_j^T$. Let us consider the behavior of $\boldsymbol{e}_n^T(\boldsymbol{U}_2 - (\sum_{j=m+1}^d \lambda_j/n)\boldsymbol{I}_n)\boldsymbol{e}_n$ in (12). By using Chebyshev's inequality for any $\tau > 0$ and the uniform bound M (>0) for the fourth moments condition, one has for all diagonal elements of $\lambda_1^{-1}(\boldsymbol{U}_2 - (\sum_{j=m+1}^d \lambda_j/n)\boldsymbol{I}_n)$ that

$$\sum_{i=1}^n P\Big(\Big|(n\lambda_1)^{-1}\sum_{j=m+1}^d \lambda_j(z_{ji}^2 - 1)\Big| > \tau\Big) \le M\tau^{-2}n^{-1}\lambda_{m+1}^2 d^{1-2\alpha_1} = o(1) \tag{13}$$

as $d\to\infty$ either when $n\to\infty$ or n is fixed. Since we have that $(n\lambda_1)^{-1}\sum_{j=m+1}^d \lambda_j(z_{ji}^2 - 1) = o_p(1)$ for all $i=1,...,n$, it holds that all diagonal elements of $\lambda_1^{-1}(\boldsymbol{U}_2 - (\sum_{j=m+1}^d \lambda_j/n)\boldsymbol{I}_n)$ converge to 0 in probability. By using Markov's inequality for any $\tau > 0$, one has for all off-diagonal elements of $\lambda_1^{-1}(\boldsymbol{U}_2 - (\sum_{j=m+1}^d \lambda_j/n)\boldsymbol{I}_n)$ that

$$P\Big(\sum_{i\neq i'}\Big((n\lambda_1)^{-1}\sum_{j=m+1}^d \lambda_j z_{ji}z_{ji'}\Big)^2 > \tau\Big) \le \tau^{-1}\lambda_{m+1}^2 d^{1-2\alpha_1} = o(1).$$

Thus we have that $\sum_{i\neq i'}((n\lambda_1)^{-1}\sum_{j=m+1}^d \lambda_j z_{ji}z_{ji'})^2 = o_p(1)$ so that

$$\Big|\sum_{i\neq i'} e_i e_{i'}\sum_{j=m+1}^d (n\lambda_1)^{-1}\lambda_j z_{ji}z_{ji'}\Big| \le \Big(\sum_{i\neq i'}\Big((n\lambda_1)^{-1}\sum_{j=m+1}^d \lambda_j z_{ji}z_{ji'}\Big)^2\Big)^{1/2} = o_p(1). \tag{14}$$

Then, we obtain that $\lambda_1^{-1}\boldsymbol{e}_n^T(\boldsymbol{U}_2 - (\sum_{j=m+1}^d \lambda_j/n)\boldsymbol{I}_n)\boldsymbol{e}_n = o_p(1)$ as $d\to\infty$ either when $n\to\infty$ or n is fixed. Note that $\lambda_1^{-1}\lambda_2 \to 0$ as $d\to\infty$ and $||n^{-1/2}\boldsymbol{z}_1|| = 1 + o_p(1)$ as $n\to\infty$. Hence, by noting that $\max_{\boldsymbol{e}_n}(\boldsymbol{e}_n^T\boldsymbol{S}_D\boldsymbol{e}_n) = \hat{\boldsymbol{u}}_1^T\boldsymbol{S}_D\hat{\boldsymbol{u}}_1$, it holds that

$$\lambda_1^{-1}\hat{\boldsymbol{u}}_1^T\left(\boldsymbol{S}_D - n^{-1}\sum_{j=m+1}^d \lambda_j\boldsymbol{I}_n\right)\hat{\boldsymbol{u}}_1 = \frac{\hat{\boldsymbol{u}}_1^T\boldsymbol{U}_1\hat{\boldsymbol{u}}_1}{\lambda_1} + o_p(1) = (\hat{\boldsymbol{u}}_1^T\boldsymbol{z}_1/n^{1/2})^2 + o_p(1) = 1 + o_p(1).$$

Hence, we claim as $d\to\infty$ and $n\to\infty$ that

$$\lambda_1^{-1}\left(\hat{\boldsymbol{u}}_1^T\boldsymbol{S}_D\hat{\boldsymbol{u}}_1 - n^{-1}\sum_{j=m+1}^d \lambda_j\right) = \frac{\hat{\lambda}_1 - n^{-1}\sum_{j=m+1}^d \lambda_j}{\lambda_1} = 1 + o_p(1).$$

From the proof of Corollary 4 and Theorem 6 in Appendix, we can obtain that $n^{-1/2}\boldsymbol{z}_1^T\hat{\boldsymbol{u}}_2 = o_p(d^{\alpha_2-\alpha_1})$ as $d \to \infty$ and $n \to \infty$. By noting that $||n^{-1/2}\boldsymbol{z}_2|| = 1 + o_p(1)$ as $n \to \infty$, we have that

$$\lambda_2^{-1}\left(\hat{\boldsymbol{u}}_2^T\boldsymbol{S}_D\hat{\boldsymbol{u}}_2 - n^{-1}\sum_{j=m+1}^{d}\lambda_j\right) = \hat{\boldsymbol{u}}_2^T\frac{\lambda_1\boldsymbol{z}_1\boldsymbol{z}_1^T}{\lambda_2 n}\hat{\boldsymbol{u}}_2 + \hat{\boldsymbol{u}}_2^T\frac{\boldsymbol{z}_2\boldsymbol{z}_2^T}{n}\hat{\boldsymbol{u}}_2 + o_p(1) = 1 + o_p(1).$$

Now, we consider estimating the noise part from the fact that as $d \to \infty$ and $n \to \infty$

$$\lambda_j^{-1}\left(\frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^{j}\hat{\lambda}_i}{n-j} - n^{-1}\sum_{i=m+1}^{d}\lambda_i\right) = o_p(1)$$

for $j = 1, 2$. (See Lemma 7 in Appendix for the details.) Then, we have as $d \to \infty$ and $n \to \infty$ that

$$\lambda_j^{-1}\left(\hat{\lambda}_j - \frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^{j}\hat{\lambda}_i}{n-j}\right) = 1 + o_p(1)$$

for $j = 1, 2$. Hence, we have a consistent estimator for $\lambda_j = d^{\alpha_j}$ with $\alpha_j > 1/2$ that is a milder condition than (YA-ii).

In general, we propose the new estimation methodology as follows:

[Noise-reduction methodology]

$$\tilde{\lambda}_j = \hat{\lambda}_j - \frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^{j}\hat{\lambda}_i}{n-j} \quad (j = 1, ..., n-1). \tag{15}$$

Note that $\tilde{\lambda}_j \geq 0$ $(j = 1, ..., n-1)$ w.p.1 for $n \leq d$. Then, we claim the following theorem.

Theorem 3. *For $j = 1, ..., m$, we have that*

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

under the conditions:

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

Theorem 4. *Let $V(z_{jk}^2) = M_j$ $(< \infty)$ for $j = 1, ..., m$ $(k = 1, ..., n)$. Assume that λ_j $(j \leq m)$ has multiplicity one. Under the conditions:*

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

(ii) *$d \to \infty$ and $d^{2-4\alpha_j}/n(d) \to 0$ for j such that $\alpha_j \in (0, 1/2]$,*

we have that

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

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

Remark 4. Yata and Aoshima [14] gave the asymptotic normality of $\hat{\lambda}_j$'s. Under the assumption that $\lambda_{m+1} = \cdots = \lambda_d = 1$, Lee et al. [9] considered an estimate of λ_j such as

$$\dot{\lambda}_j = \frac{\hat{\lambda}_j + 1 - \eta + \sqrt{(\hat{\lambda}_j + 1 - \eta)^2 - 4\hat{\lambda}_j}}{2},$$

where $d/n \to \eta \geq 0$ and $n \to \infty$. If $1/\hat{\lambda}_j = o_p(1)$, we claim that $\dot{\lambda}_j = (\hat{\lambda}_j - \eta)(1 + o_p(1))$. By noting that $n^{-1}\sum_{j=m+1}^{d} \lambda_j \to \eta$ when $\lambda_{m+1} = \cdots = \lambda_d = 1$, it holds that $\dot{\lambda}_j = (\hat{\lambda}_j - n^{-1}\sum_{j=m+1}^{d} \lambda_j)(1 + o_p(1))$. Hence, we may consider $\dot{\lambda}_j$ as a noise reduction. However, we emphasize that *the noise-reduction methodology allows users to have a consistent estimator, $\tilde{\lambda}_j$, when $\lambda_{m+1} \geq \cdots \geq \lambda_d$ (> 0) and $\lambda_j = c_j$ $(j = m + 1, ..., d)$ are unknown constants.*

Remark 5. When $\boldsymbol{X}$ is Gaussian, $\alpha_1 > 1/2$ and either when $\alpha_1 > \alpha_2$ or $m = 1$, we have as $d \to \infty$ that

$$\frac{\tilde{\lambda}_1}{\lambda_1} \Rightarrow \frac{\chi_n^2}{n}$$

for fixed n, where χ_n^2 denotes a random variable distributed as the χ^2 distribution with d.f. n. Jung and Marron [7] claimed a similar result for $\hat{\lambda}_j$ with $\alpha_j > 1$.

Remark 6. When z_{jk}, $j = 1, ..., d$ $(k = 1, ..., n)$ are not independent but the components of $\boldsymbol{Z}$ are ρ-mixing, we can claim the assertions similar to Theorems 3-4 under the conditions:

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

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

The conditions (i)-(ii) are milder than the ones given by Theorem 3.1 in Yata and Aoshima [14] for a non-ρ-mixing case.

Corollary 1. When the population mean may not be zero, let us write that $\boldsymbol{S}_{oD} = (n-1)^{-1}(\boldsymbol{X}-\overline{\boldsymbol{X}})^T(\boldsymbol{X}-\overline{\boldsymbol{X}})$, where $\overline{\boldsymbol{X}} = [\bar{\boldsymbol{x}}_n, ..., \bar{\boldsymbol{x}}_n]$ is having d-vector $\bar{\boldsymbol{x}}_n = \sum_{s=1}^n \boldsymbol{x}_s/n$. We redefine $\tilde{\lambda}_j$ $(j = 1, ..., m)$ in (15) by replacing $\boldsymbol{S}_D$ and n with $\boldsymbol{S}_{oD}$ and $n-1$. Then, the assertions in Theorems 3-4 are still justified under the convergence conditions.

4. Consistency properties of PC directions.

In this section, we consider PC direction vectors. Jung and Marron [7], and Yata and Aoshima [14] studied consistency properties of PC direction vectors in the context of naive PCA. Let $\hat{\boldsymbol{H}} = [\hat{\boldsymbol{h}}_1, \cdots, \hat{\boldsymbol{h}}_d]$ such that $\hat{\boldsymbol{H}}^T \boldsymbol{S}\hat{\boldsymbol{H}} = \hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Lambda}} = \mathrm{diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_d)$. 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 an eigenvector of $\boldsymbol{S}_D$. Then, Yata and Aoshima [14] gave consistency properties of the sample eigenvectors with their population counterparts: Assume that λ_j $(j \le m)$ has multiplicity one as $\lambda_j \ne \lambda_{j'}$ for all $j'(\ne j)$. Then, the first m sample eigenvectors are consistent in the sense that

$$\mathrm{Angle}(\hat{\boldsymbol{h}}_j, \boldsymbol{h}_j) \xrightarrow{p} 0 \quad (j = 1, ..., m)$$

under (YA-i)-(YA-ii). The following result can be obtained as a corollary of Theorem 4.1 in Yata and Aoshima [14].

Corollary 2. The first m sample eigenvectors are inconsistent in the sense that

$$\mathrm{Angle}(\hat{\boldsymbol{h}}_j, \boldsymbol{h}_j) \xrightarrow{p} \pi/2 \quad (j = 1, ..., m) \tag{16}$$

under the condition that $d \to \infty$ and $d/(n(d)\lambda_j) \to \infty$.

Remark 7. Under the condition described above, we have that $\hat{\boldsymbol{h}}_j^T \boldsymbol{h}_j = o_p(1)$ $(j = 1, ..., m)$. Jung and Marron [7] gave (16) as $d \to \infty$ for a fixed n.

Remark 8. When the population mean may not be zero, we still have Corollary 2 by using $\boldsymbol{S}_{oD}$ defined in Corollary 1.

5. PC scores with noise-reduction methodology

In this section, we apply the noise-reduction methodology to principal component scores (Pcss). The j-th Pcs of $\boldsymbol{x}_k$ is given by $\boldsymbol{h}_j^T\boldsymbol{x}_k = z_{jk}\sqrt{\lambda_j}$ $(= s_{jk}$, say). However, since $\boldsymbol{h}_j$ is unknown, one may use $\hat{\boldsymbol{h}}_j = (n\hat{\lambda}_j)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{u}}_j$ as a sample eigenvector. The j-th Pcs of $\boldsymbol{x}_k$ is estimated by $\hat{\boldsymbol{h}}_j^T\boldsymbol{x}_k = \hat{u}_{jk}\sqrt{n\hat{\lambda}_j}$ $(= \hat{s}_{jk}$, say), where $\hat{\boldsymbol{u}}_j^T = (\hat{u}_{j1}, ..., \hat{u}_{jn})$. Let us define a sample mean square error of the j-th Pcs by $\mathrm{MSE}(\hat{s}_j) = n^{-1}\sum_{k=1}^{n}(\hat{s}_{jk} - s_{jk})^2$. Then, Yata and Aoshima [14] evaluated the sample Pcs as follows: Assume that λ_j $(j \le m)$ has multiplicity one. Then, it holds that

$$\frac{\mathrm{MSE}(\hat{s}_j)}{\lambda_j} = o_p(1) \tag{17}$$

under (YA-i)-(YA-ii).

Now, we modify $\hat{s}_{jk}$ by using $\tilde{\lambda}_j$ defined by (15). Let us write that $\hat{u}_{jk}\sqrt{n\tilde{\lambda}_j}$ $(= \tilde{s}_{jk}$, say). Then, we obtain the following result.

Theorem 5. *Assume that λ_j $(j \le m)$ has multiplicity one. Then, we have that*

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

under the conditions (i)-(ii) of Theorem 3.

For $\hat{\boldsymbol{u}}_j$, we can claim the consistency for a Pcs vector $n^{-1/2}\boldsymbol{z}_j$.

Corollary 3. *Assume that λ_j $(j \le m)$ has multiplicity one. Then, the j-th sample eigenvector is consistent in the sense that*

$$Angle(\hat{\boldsymbol{u}}_j, n^{-1/2}\boldsymbol{z}_j) \xrightarrow{p} 0 \tag{19}$$

under the conditions (i)-(ii) of Theorem 3.

Remark 9. Lee et al. [9] gave a result similar to (19). Under the assumption that $\lambda_{m+1} = \cdots = \lambda_d = 1$ and the multiplicity one assumption, they claimed as $d/n \to \eta \ge 0$ and

$n \to \infty$ that

$$|\hat{\boldsymbol{u}}_j^T \boldsymbol{z}_j / n^{1/2}| = \begin{cases} \sqrt{1 - \frac{\eta}{(\lambda_j - 1)^2}} + o_p(1) & \text{when } \lambda_j > 1 + \eta, \\ o_p(1) & \text{when } 1 < \lambda_j \le 1 + \eta. \end{cases}$$

Here, it holds that $|\hat{\boldsymbol{u}}_j^T \boldsymbol{z}_j / n^{1/2}| = 1 + o_p(1)$ under $\eta/(\lambda_j - 1)^2 = O(d^{1-2\alpha_j}/n) \to 0$. Then, by noting that $||\boldsymbol{z}_j / n^{1/2}|| = 1 + o_p(1)$ as $n \to \infty$, we have (19) under the conditions (i)-(ii) of Theorem 3. Thus their result corresponds to Corollary 3 when $\lambda_{m+1} = \cdots = \lambda_d = 1$.

Let $\boldsymbol{x}_{new}$ be a new sample from the distribution and independent of $\boldsymbol{X}$. The j-th Pcs of $\boldsymbol{x}_{new}$ is given by $\boldsymbol{h}_j^T \boldsymbol{x}_{new}$ $(= s_{j(new)},$ say). Note that $V(s_{j(new)}/\sqrt{\lambda_j}) = 1$. We consider a consistent estimator of $s_{j(new)}$. Then, we have the following result.

Corollary 4. *Assume that λ_j $(j \le m)$ has multiplicity one. For $\hat{\boldsymbol{h}}_j$, it holds that*

$$\frac{\hat{\boldsymbol{h}}_j^T \boldsymbol{x}_{new}}{\sqrt{\lambda_j}} = \frac{s_{j(new)}}{\sqrt{\lambda_j}} + o_p(1)$$

under the conditions that (i) $d \to \infty$ and $n \to \infty$ for j such that $\alpha_j > 1$, (ii) $d \to \infty$ and $d^{1-\alpha_j}/n(d) \to 0$ for j such that $\alpha_j \in (1/3, 1]$, and (iii) $d \to \infty$ and $d^{2-4\alpha_j}/n(d) \to 0$ for j such that $\alpha_j \in (0, 1/3]$,

Remark 10. Lee et al. [9] also considered a predict Pcs for $s_{j(new)}$ when $\lambda_{m+1} = \cdots = \lambda_d = 1$.

Now, we consider applying the noise-reduction methodology to the PC direction vectors. Let us define $\tilde{\boldsymbol{h}}_j = (n\tilde{\lambda}_j)^{-1/2} \boldsymbol{X} \hat{\boldsymbol{u}}_j$. Then, we consider $\tilde{\boldsymbol{h}}_j$ as an estimate of the PC direction vector, $\boldsymbol{h}_j$. By using $\tilde{\boldsymbol{h}}_j$, $j = 1, ..., m$, we have the following the theorem.

Theorem 6. *Assume that λ_j $(j \le m)$ has multiplicity one. For $\tilde{\boldsymbol{h}}_j$, it holds that*

$$\frac{\tilde{\boldsymbol{h}}_j^T \boldsymbol{x}_{new}}{\sqrt{\lambda_j}} = \frac{s_{j(new)}}{\sqrt{\lambda_j}} + o_p(1)$$

under the conditions (i)-(ii) of Theorem 4.

Remark 11. Assume that λ_j ($j \le m$) has multiplicity one. Then, the j-th sample eigenvector is consistent in the sense that

$$\tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j = 1 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j} n^{-1})$$

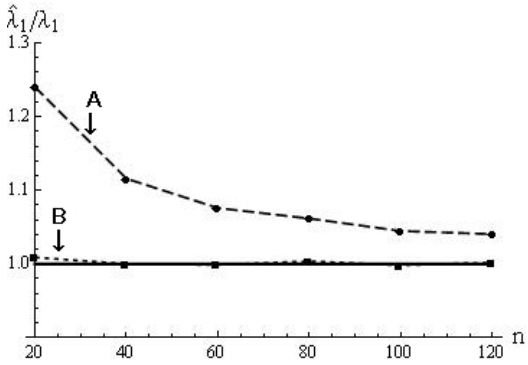
under the conditions (i)-(ii) of Theorem 3. For the norm, it holds that $||\tilde{\boldsymbol{h}}_j|| = 1 + o_p(1)$ under (YA-i)-(YA-ii).

Remark 12. When the population mean may not be zero, we still have the above results by using $\boldsymbol{S}_{oD}$ defined in Corollary 1.

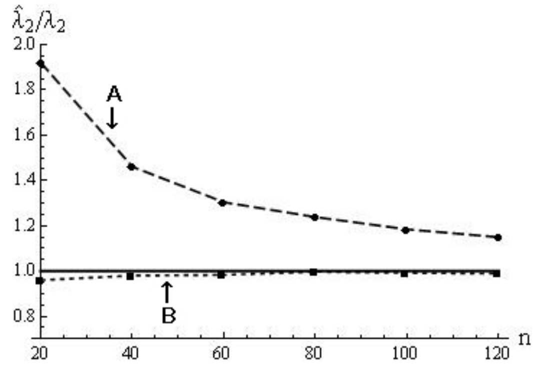
6. Performances of noise-reduction methodology

When we observe naive PCA, the sample size n should be determined depending on d for $\alpha_i \in (0, 1]$ in (YA-ii). On the other hand, the noise-reduction methodology allows the experimenter to choose n free from d for the case that $\alpha_i > 1/2$ as seen in the theorems given in Sections 3 and 5. The noise-reduction methodology is promising to give feasible estimation for HDLSS data with extremely small order of n compared to d. In this section, we examine its performance with the help of Monte Carlo simulations.

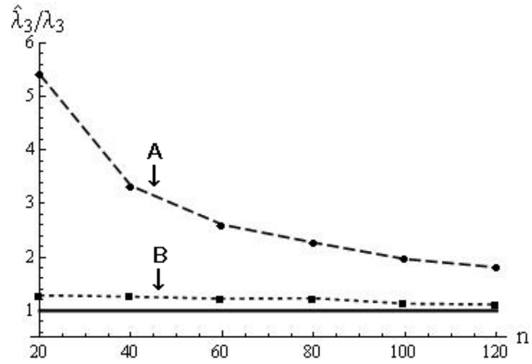
Independent pseudo-random normal observations were generated from $N_d(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $d = 1600$. We considered $\lambda_1 = d^{4/5}$, $\lambda_2 = d^{3/5}$, $\lambda_3 = d^{2/5}$ and $\lambda_4 = \cdots = \lambda_d = 1$ in (1). We used the sample of size $n \in [20, 120]$ to define the data matrix $\boldsymbol{X} : d \times n$ for the calculation of $\boldsymbol{S}_D$. The findings were obtained by averaging the outcomes from 2000 ($= R$, say) replications. Under a fixed scenario, suppose that the r-th replication ends with estimates of λ_j, $\hat{\lambda}_{jr}$ and $\tilde{\lambda}_{jr}$ ($r = 1, ..., R$), given by using (10) and (15). Let us simply write $\hat{\lambda}_j = R^{-1} \sum_{r=1}^R \hat{\lambda}_{jr}$ and $\tilde{\lambda}_j = R^{-1} \sum_{r=1}^R \tilde{\lambda}_{jr}$. We considered two quantities, A: $\hat{\lambda}_j/\lambda_j$ and B: $\tilde{\lambda}_j/\lambda_j$. Fig. 3 shows the behaviors of both A and B for the first three eigenvalues. By observing the behavior of A, (10) seems not to give a feasible estimation within the range of n. The sample size n was not large enough to use the eigenvalues of $\boldsymbol{S}_D$ for such a high-dimensional space. On the other hand, in view of the behavior of B, (15) gives a reasonable estimation surprisingly well for such HDLSS datasets. The noise-reduction methodology seems to perform excellently as expected theoretically.



(a) For the first eigenvalue



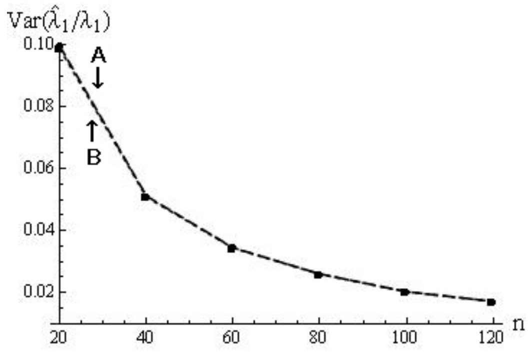
(b) For the second eigenvalue



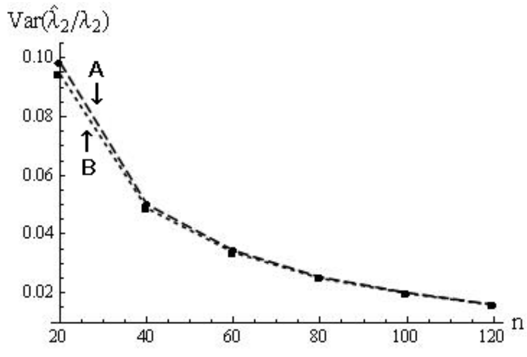
(c) For the third eigenvalue

Fig. 3. The behaviors of A: $\hat{\lambda}_j/\lambda_j$ and B: $\tilde{\lambda}_j/\lambda_j$ for (a) the first eigenvalue, (b) the second eigenvalue and (c) the third eigenvalue when the samples, of size $n = 20(20)120$, were taken from $N_d(\mathbf{0}, \mathbf{\Sigma})$ with $d = 1600$.

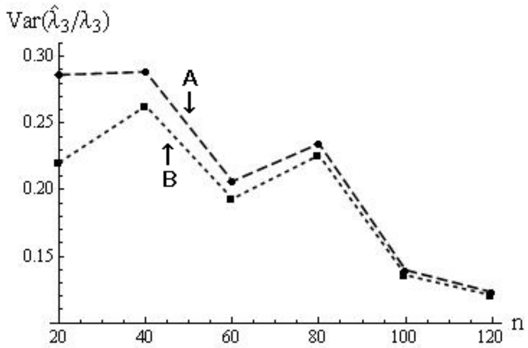
We also considered the Monte Carlo variability. Let $\text{Var}(\hat{\lambda}_j/\lambda_j) = (R-1)^{-1}\sum_{r=1}^{R}(\hat{\lambda}_{jr} - \hat{\lambda}_j)^2/\lambda_j^2$ and $\text{Var}(\tilde{\lambda}_j/\lambda_j) = (R-1)^{-1}\sum_{r=1}^{R}(\tilde{\lambda}_{jr} - \tilde{\lambda}_j)^2/\lambda_j^2$. We considered two quantities, A: $\text{Var}(\hat{\lambda}_j/\lambda_j)$ and B: $\text{Var}(\tilde{\lambda}_j/\lambda_j)$, in Fig. 4 to show the behaviors of sample variances of both A and B for the first three eigenvalues.



(a) For the first eigenvalue



(b) For the second eigenvalue



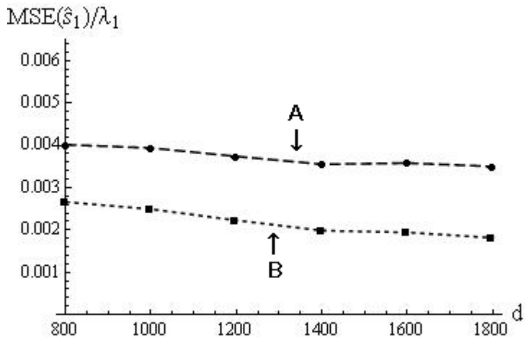
(c) For the third eigenvalue

Fig. 4. The behaviors of A: $\mathrm{Var}(\hat{\lambda}_j/\lambda_j)$ and B: $\mathrm{Var}(\tilde{\lambda}_j/\lambda_j)$ for (a) the first eigenvalue, (b) the second eigenvalue and (c) the third eigenvalue when the samples, of size $n = 20(20)120$, were taken from $N_d(\mathbf{0}, \mathbf{\Sigma})$ with $d = 1600$.

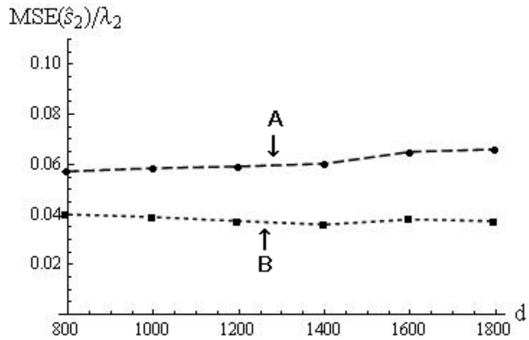
By observing the behaviors of the sample variances, both the behaviors seem not to make much difference between A and B. Note that it holds $M_j = 2$ $(j = 1, ..., m)$ for Gaussian $\boldsymbol{X}$. From Theorem 3.2 given in Yata and Aoshima [14], the limiting distribution of $(n/2)^{1/2}(\hat{\lambda}_j/\lambda_j - 1)$ is $N(0,1)$, so that the variance of A is approximately given by $\mathrm{Var}(\hat{\lambda}_j/\lambda_j) = 2/n$. On the other hand, in view of Theorem 4, the limiting distribution of $(n/2)^{1/2}(\tilde{\lambda}_j/\lambda_j - 1)$ is $N(0,1)$. Hence, the variance of B is approximately given by $\mathrm{Var}(\tilde{\lambda}_j/\lambda_j) = 2/n$; that is approximately equal to the variance of A.

Next, we considered the Pcs. Independent pseudo-random normal observations were generated from $N_d(\mathbf{0}, \mathbf{\Sigma})$. We considered the case that $\lambda_1 = d^{4/5}$, $\lambda_2 = d^{3/5}$ $\lambda_3 = d^{2/5}$ and $\lambda_4 = \cdots = \lambda_d = 1$ in (1) as before. We fixed the sample size as $n = 60$. We set the dimension as $d = 800(200)1800$. Under a fixed scenario, suppose that the r-th replication ends with

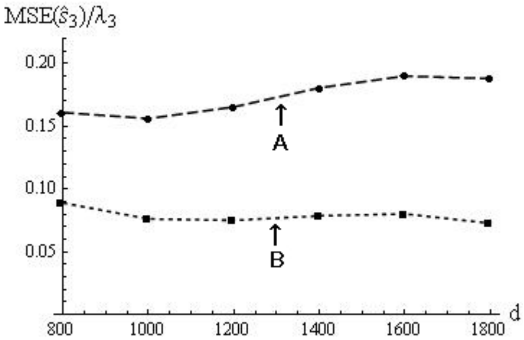
$\mathrm{MSE}(\hat{s}_j)_r$ and $\mathrm{MSE}(\tilde{s}_j)_r$ $(r = 1, ..., R)$, given by using (17) and (18). Let us simply write $\mathrm{MSE}(\hat{s}_j) = R^{-1}\sum_{r=1}^{R} \mathrm{MSE}(\hat{s}_j)_r$ and $\mathrm{MSE}(\tilde{s}_j) = R^{-1}\sum_{r=1}^{R} \mathrm{MSE}(\tilde{s}_j)_r$. We considered two quantities, A: $\mathrm{MSE}(\hat{s}_j)/\lambda_j$ and B: $\mathrm{MSE}(\tilde{s}_j)/\lambda_j$, in Fig. 5 to show the behaviors of both A and B for the first three Pcss.



(a) For the first Pcs



(b) For the second Pcs



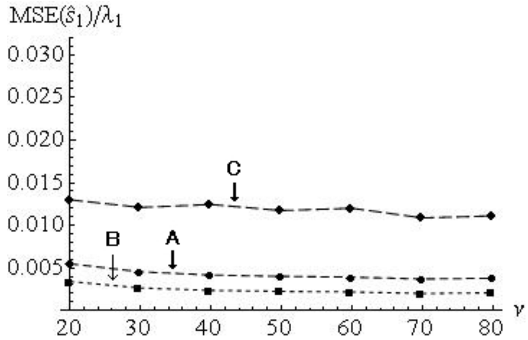
(c) For the third Pcs

Fig. 5. The behaviors of A: $\mathrm{MSE}(\hat{s}_j)/\lambda_j$ and B: $\mathrm{MSE}(\tilde{s}_j)/\lambda_j$ for (a) the first Pcs, (b) the second Pcs and (c) the third Pcs when the samples, of size $n = 60$, were taken from $N_d(\mathbf{0}, \mathbf{\Sigma})$ with $d = 800(200)1800$.

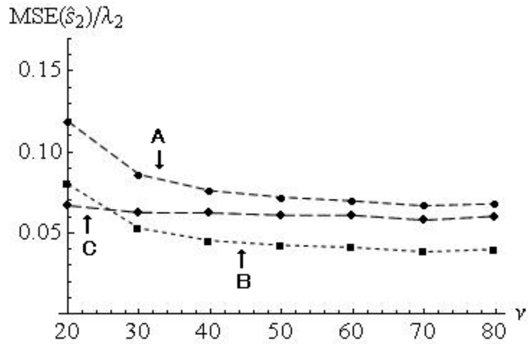
Again, the noise-reduction methodology seems to perform much better than naive PCA. We conducted simulation studies for other settings as well and verified the superiority of the noise-reduction methodology to naive PCA in HDLSS data situations.

Finally, we compare the noise-reduction methodology with the cross-data-matrix methodology. Yata and Aoshima [16] gave a Pcs estimator by using the cross-data-matrix methodology. Let $\acute{s}_{jk}$ be the j-th Pcs estimator of $\boldsymbol{x}_k$ given in Section 5 in Yata and Aoshima [16]. Independent pseudo-random observations were generated from the d-variate t-distribution,

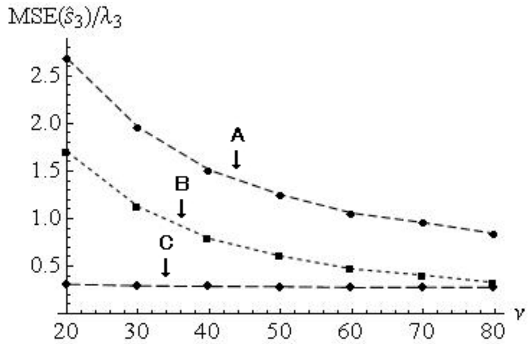
$t_d(\mathbf{0}, \mathbf{\Sigma}, \nu)$ with mean zero, covariance matrix $\mathbf{\Sigma}$ and d.f. $\nu \in [20, 80]$. We considered $\lambda_1 = d^{4/5}$, $\lambda_2 = d^{3/5}$, $\lambda_3 = d^{2/5}$ and $\lambda_4 = \cdots = \lambda_d = 1$ in (1). We set $d = 1600$ and $n = 60$. We considered three quantities, A: $\mathrm{MSE}(\hat{s}_j)/\lambda_j$, B: $\mathrm{MSE}(\tilde{s}_j)/\lambda_j$ and C: $\mathrm{MSE}(\acute{s}_j)/\lambda_j$, in Fig. 6 to show the behaviors of A, B and C for the first three Pcss.



(a) For the first Pcs



(b) For the second Pcs



(c) For the third Pcs

Fig. 6. The behaviors of A: $\mathrm{MSE}(\hat{s}_j)/\lambda_j$, B: $\mathrm{MSE}(\tilde{s}_j)/\lambda_j$ and C: $\mathrm{MSE}(\acute{s}_j)/\lambda_j$ for (a) the first Pcs, (b) the second Pcs and (c) the third Pcs when the samples, of size $n = 60$, were taken from $t_d(\mathbf{0}, \mathbf{\Sigma}, \nu)$ with $\nu = 20(10)80$.

Note that $t_d(\mathbf{0}, \mathbf{\Sigma}, \nu) \Rightarrow N_d(\mathbf{0}, \mathbf{\Sigma})$ as $\nu \to \infty$. When ν is small, $\boldsymbol{X}$ has the geometric representation given by (5). In the case, the cross-data-matrix methodology seems to perform better than the noise-reduction methodology. On the other hand, when ν is large, $\boldsymbol{X}$ has the geometric representation given by (3). In the case, the noise-reduction methodology seems to perform best among the three estimators.

7. Inverse covariance matrix estimator

In this section, we apply the noise-reduction methodology to estimating the inverse covariance matrix of $\boldsymbol{\Sigma}$. The inverse covariance matrix, $\boldsymbol{\Sigma}^{-1}$, is the key to constructing inference procedures in many statistical problems. However, it should be noted that $\boldsymbol{S}^{-1}$ does not exist in the HDLSS context. Srivastava [11] and [12] used the Moore-Penrose inverse of $\boldsymbol{S}$ for several inference problems. Srivastava and Kubokawa [13] proposed an empirical Bayes inverse matrix estimator of $\boldsymbol{\Sigma}$ for the discriminant analysis and compared the performance with that of the Moore-Penrose inverse or the inverse matrix defined by only diagonal elements of $\boldsymbol{S}$. Then, they concluded that the discriminant rule using the empirical Bayes inverse matrix estimator was better than the others. The empirical Bayes inverse matrix estimator was defined by $\boldsymbol{S}_\delta^{-1} = (\boldsymbol{S} + \delta \boldsymbol{I}_d)^{-1}$ with $\delta = \mathrm{tr}(\boldsymbol{S})/n$. Then, let us consider the eigen-decomposition of $\boldsymbol{S}_\delta^{-1}$ as

$$\boldsymbol{S}_\delta^{-1} = \sum_{j=1}^{n} (\hat{\lambda}_j + \delta)^{-1} \hat{\boldsymbol{h}}_j \hat{\boldsymbol{h}}_j^T + \delta^{-1} \left(\boldsymbol{I}_d - \sum_{j=1}^{n} \hat{\boldsymbol{h}}_j \hat{\boldsymbol{h}}_j^T \right).$$

Let $\boldsymbol{V}_\delta = (v_{ij(\delta)}) = \boldsymbol{\Lambda}^{1/2} \boldsymbol{H}^T \boldsymbol{S}_\delta^{-1} \boldsymbol{H} \boldsymbol{\Lambda}^{1/2}$, where $\boldsymbol{\Lambda}^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, ..., \lambda_d^{1/2})$. Note that $\boldsymbol{\Lambda}^{1/2} \boldsymbol{H}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{H} \boldsymbol{\Lambda}^{1/2} = \boldsymbol{I}_d$. Let us write $\kappa = n^{-1} \sum_{i=m+1}^{d} \lambda_i$. Then, we obtain the following result.

Theorem 7. *Assume that $n = n(d)$, $\alpha_1 < 1 - \gamma/2$, $\gamma < 1$ and the first m population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. Under the condition that (i) $d \to \infty$ and $\kappa/\lambda_j = O(d^{1-\gamma-\alpha_j}) < \infty$ for $j = 1, ..., m$, we have that*

$$v_{jj(\delta)} = \frac{2\lambda_j}{\lambda_j + 2\kappa} + o_p(1),$$
$$v_{jj'(\delta)} = o_p(1), \ j' = j+1, ..., d.$$

For j such that $\lambda_j/\kappa \to 0$ as $d \to \infty$, we have as $d \to \infty$ that

$$v_{jj(\delta)} = \lambda_j \kappa^{-1} + o_p(\lambda_j \kappa^{-1}),$$
$$v_{jj'(\delta)} = o_p(\lambda_j \kappa^{-1}), \ j' = j+1, ..., d.$$

Let j_δ be the maximum integer $j(\le m)$ such that $\alpha_j - 1 + \gamma > 0$. We assume that $\alpha_j - 1 + \gamma \neq 0$ $(j = 1, ..., m)$. We observe that $(v_{11(\delta)}, ..., v_{j_\delta j_\delta(\delta)}, v_{j_\delta+1\, j_\delta+1(\delta)}, ..., v_{dd(\delta)})$ is close to $(2, ..., 2, \lambda_{j_\delta+1}\kappa^{-1}, ..., \lambda_d\kappa^{-1})$. One should note that $\boldsymbol{V}_\delta$ is far from $\boldsymbol{I}_d$, so that $\boldsymbol{H}^T\boldsymbol{S}_\delta^{-1}\boldsymbol{H}$ is far from $\boldsymbol{\Lambda}^{-1}$. Let us consider a different inverse matrix estimator of $\boldsymbol{\Sigma}$ by using the noise-reduction methodology. Let $\omega = \min(\mathrm{tr}(\boldsymbol{S})/(d^{1/2}n^{1/4}),\ \delta)$ and $\acute{\lambda}_j = \max(\tilde{\lambda}_j,\ \omega)$. Then, we define a new inverse matrix estimator as

$$\boldsymbol{S}_\omega^{-1} = \sum_{j=1}^{n-1} \acute{\lambda}_j^{-1}\tilde{\boldsymbol{h}}_j\tilde{\boldsymbol{h}}_j^T + \omega^{-1}\left(\boldsymbol{I}_d - \sum_{j=1}^{n-1}\tilde{\boldsymbol{h}}_j\tilde{\boldsymbol{h}}_j^T\right), \tag{20}$$

where $\tilde{\boldsymbol{h}}_j$ is the same one as in Theorem 6. Let $\boldsymbol{V}_\omega = (v_{ij(\omega)}) = \boldsymbol{\Lambda}^{1/2}\boldsymbol{H}^T\boldsymbol{S}_\omega^{-1}\boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$. Then, we obtain the following theorem.

Theorem 8. *Assume that $n = n(d)$, $\gamma < 1$, $\alpha_1 < \min(1/2 + \gamma/4,\ 1 - \gamma/2)$ and the first m population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. Let $\psi = \min(n^{3/4}\kappa/d^{1/2},\ \kappa)$. Under the conditions that (i) $d \to \infty$ and $\psi/\lambda_j = O(d^{1/2-\gamma/4-\alpha_j}) < \infty$ for $\gamma < 2/3$, (ii) $d \to \infty$ and $\psi/\lambda_j = O(d^{1-\gamma-\alpha_j}) < \infty$ for $\gamma \in [2/3, 1)$, we have that*

$$v_{jj(\omega)} = \frac{1}{\max(1,\ \psi/\lambda_j)} + o_p(1),$$
$$v_{jj'(\omega)} = o_p(1),\ j' = j+1, ..., d.$$

For j such that $\lambda_j/\psi \to 0$ as $d \to \infty$, we have as $d \to \infty$ that

$$v_{jj(\omega)} = \frac{\lambda_j}{\psi} + o_p(\lambda_j/\psi),$$
$$v_{jj'(\omega)} = o_p(\lambda_j/\psi),\ j' = j+1, ..., d.$$

Let j_ω be the maximum integer $j(\le m)$ such that $\alpha_j - 1/2 + \gamma/4 > 0$. We assume that $\alpha_j - 1/2 + \gamma/4 \neq 0$ $(j = 1, ..., m)$. We observe that $(v_{11(\omega)}, ..., v_{j_\omega j_\omega(\omega)}, v_{j_\omega+1\, j_\omega+1(\omega)}, ..., v_{dd(\omega)})$ is close to $(1, ..., 1, \lambda_{j_\omega+1}\psi^{-1}, ..., \lambda_d\psi^{-1})$. Note that $\psi < \kappa$ w.p.1 when $\gamma < 2/3$. We can claim that $\boldsymbol{V}_\omega$ is surely closer to $\boldsymbol{I}_d$ than $\boldsymbol{V}_\delta$ under (i)-(ii) of Theorem 8.

Remark 13. It should be noted that $\hat{\boldsymbol{h}}_j^T \boldsymbol{S}_\omega^{-1} \hat{\boldsymbol{h}}_j \le 0$ w.p.1 as $\tilde{\lambda}_j/\hat{\lambda}_j = o_p(1)$. Let $\boldsymbol{e}_d$ be a d-dimensional unit vector. Assume further in Theorem 8 that $\boldsymbol{e}_d$ is a constant vector or $\boldsymbol{e}_d$ and $\boldsymbol{S}_\omega^{-1}$ are independent. Then, we claim as $d \to \infty$ that $\boldsymbol{e}_d^T \boldsymbol{S}_\omega^{-1} \boldsymbol{e}_d \ge 0$ w.p.1.

8. Application

In this section, we apply the inverse covariance matrix estimator given by (20) to the discriminant analysis. Suppose that we have two $d \times N_i$ data matrices, $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iN_i}]$, $i = 1, 2$. We assume that $\boldsymbol{x}_{11}, ..., \boldsymbol{x}_{1N_1}$ and $\boldsymbol{x}_{21}, ..., \boldsymbol{x}_{2N_2}$ are independent and identically distributed as $\pi_1 : N_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\pi_2 : N_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, respectively. Let us write the eigen-decomposition of $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \sum_{j=1}^d \lambda_j \boldsymbol{h}_j \boldsymbol{h}_j^T$. We assume (1) about $\boldsymbol{\Sigma}$. Let $\boldsymbol{x}_0$ be an observation vector on an individual belonging to π_1 or to π_2. We estimate $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ by

$$\bar{\boldsymbol{x}}_i = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{x}_{ij} \quad i = 1, 2, \quad \text{and} \quad \boldsymbol{S} = n^{-1} \sum_{i=1}^{2} \sum_{j=1}^{N_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)^T,$$

where $n = N_1 + N_2 - 2$. We assume $d > n$. We consider the discriminant rule based on the maximum likelihood ratio under which we classify $\boldsymbol{x}_0$ into π_1 if

$$(1 + N_1^{-1})^{-1} (\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_1)^T \boldsymbol{S}^{-1} (\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_1) < (1 + N_2^{-1})^{-1} (\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_2)^T \boldsymbol{S}^{-1} (\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_2), \tag{21}$$

and into π_2 otherwise.

In the HDLSS context ($d > n$), there does not exist the inverse matrix of $\boldsymbol{S}$. We observe from Theorems 7-8 that the inverse matrix estimator $\boldsymbol{S}_\omega^{-1}$ given by (20) is better than the empirical Bayes inverse matrix estimator $\boldsymbol{S}_\delta^{-1}$. Let us compare the performances of $\boldsymbol{S}_\omega^{-1}$ and $\boldsymbol{S}_\delta^{-1}$ by conducting simulation studies.

Let $\boldsymbol{S} = (s_{ij})$. Define $\boldsymbol{S}_{diag}^{-1}$ by $\boldsymbol{S}_{diag}^{-1} = \text{diag}(s_{11}^{-1}, ..., s_{dd}^{-1})$. We considered the discriminant rule given by applying $\boldsymbol{S}_\omega^{-1}$, $\boldsymbol{S}_\delta^{-1}$ and $\boldsymbol{S}_{diag}^{-1}$ to $\boldsymbol{S}^{-1}$ in (21). We examined its performance with the help of Monte Carlo simulations. We set $d = 1600$. We set $\boldsymbol{\mu}_1 = (1, ..., 1, 0, ..., 0)^T$ whose first 80 elements are 1, and $\boldsymbol{\mu}_2 = (0, ..., 0)^T$. We generated the datasets $(\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iN_i})$, $i = 1, 2$, by setting a common covariance matrix as $\boldsymbol{\Sigma} = (\rho^{|i-j|^{1/7}})$, where $\rho \in (0, 1)$. Note that $\text{tr}(\boldsymbol{\Sigma}) = d$. We considered three levels of correlation as $\rho = 0.2, 0.4, 0.6$. Then, the eigenvalues, $(\lambda_1, \lambda_2, \lambda_3, ...)$, of $\boldsymbol{\Sigma}$ were calculated as (44.88, 19.07, 13.55,...) when $\rho = 0.2$, (198.05, 47.32, 29.77,...) when $\rho = 0.4$, and (491.78, 64.75, 37.80,...) when $\rho = 0.6$. We used a testing sample, $\boldsymbol{x}_0$ in (21), by generating 50 times randomly from π_1 or π_2. The experiment was iterated 100 times. The correct classification was estimated by the average rate of correct classification over the 5000 iterations. Note that the standard deviation of this simulation study is less than 0.0071. We denoted the error of misclassifying an individual from π_1 (into π_2) and from π_2 (into π_1) by e_1 and e_2, respectively. We also considered the correct discriminant rule (CDR) given by replacing (21) with

$$(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu}_1) < (\boldsymbol{x}_0 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu}_2).$$

In Table 1, we reported the correct classification rate, $1 - e_1$, when $(N_1, N_2) = (10, 10)$ and (20, 20). In Table 2, we reported the correct classification rates, $(1 - e_1,\ 1 - e_2)$, when $(N_1, N_2) = (10, 20)$ and (20, 40). When the correlation was low such as $\rho = 0.2$, we observed that the rule given by $\boldsymbol{S}_{diag}^{-1}$ is as good as the others except CDR. This result is quite natural because $\boldsymbol{\Sigma}$ becomes close to a diagonal matrix as $\rho \to 0$. As the variables were highly correlated, the rule given by $\boldsymbol{S}_{diag}^{-1}$ became worse. It should be noted that the variables in actual HDLSS situations are highly correlated each other. When the correlation was high such as $\rho = 0.4,\ 0.6$, we observed that the rule given by $\boldsymbol{S}_{\omega}^{-1}$ was best among them. It should be noted that as the correlation between variables gets high, the first few eigenvalues of $\boldsymbol{\Sigma}$ tend to become extremely large. We may observe that the noise-reduction methodology effectively works for estimating eigenvalues in $\boldsymbol{S}_{\omega}^{-1}$.

Table 1. The correct classification rate, $1 - e_1$, when $(N_1, N_2) = (10, 10)$ and (20, 20).

$(N_1, N_2) = (10, 10)$

ρ	$\boldsymbol{S}_{\omega}^{-1}$	$\boldsymbol{S}_{\delta}^{-1}$	$\boldsymbol{S}_{diag}^{-1}$	CDR
0.2	0.864	0.849	0.841	0.977
0.4	0.806	0.770	0.716	0.953
0.6	0.787	0.717	0.623	0.952

$(N_1, N_2) = (20, 20)$

ρ	$\boldsymbol{S}_{\omega}^{-1}$	$\boldsymbol{S}_{\delta}^{-1}$	$\boldsymbol{S}_{diag}^{-1}$	CDR
0.2	0.905	0.897	0.882	0.975
0.4	0.849	0.817	0.740	0.949
0.6	0.839	0.811	0.666	0.949

Table 2. The correct classification rates, $(1 - e_1,\ 1 - e_2)$,

when $(N_1, N_2) = (10, 20)$ and (20, 40).

$(N_1, N_2) = (10, 20)$

ρ	$\boldsymbol{S}_{\omega}^{-1}$	$\boldsymbol{S}_{\delta}^{-1}$	$\boldsymbol{S}_{diag}^{-1}$	CDR
0.2	(0.893, 0.872)	(0.881, 0.860)	(0.865, 0.848)	(0.977, 0.973)
0.4	(0.815, 0.830)	(0.783, 0.781)	(0.722, 0.707)	(0.950, 0.948)
0.6	(0.817, 0.814)	(0.769, 0.759)	(0.642, 0.636)	(0.945, 0.957)

$(N_1, N_2) = (20, 40)$

ρ	$\boldsymbol{S}_{\omega}^{-1}$	$\boldsymbol{S}_{\delta}^{-1}$	$\boldsymbol{S}_{diag}^{-1}$	CDR
0.2	(0.924, 0.920)	(0.922, 0.913)	(0.904, 0.901)	(0.973, 0.974)
0.4	(0.861, 0.865)	(0.837, 0.841)	(0.761, 0.752)	(0.949, 0.950)
0.6	(0.855, 0.856)	(0.831, 0.835)	(0.655, 0.663)	(0.954, 0.951)

A. Appendix

Throughout this section, let $\boldsymbol{e}_{1n}$ and $\boldsymbol{e}_{2n}$ be arbitrary elements of $\boldsymbol{R}_n$. Let $u_{ij} = n^{-1}\sum_{s=m+1}^{d}\lambda_s z_{si} z_{sj}$, $\boldsymbol{U}_{21} = \boldsymbol{U}_2 - \mathrm{diag}(u_{11}, ..., u_{nn})$ and $\boldsymbol{U}_{22} = \boldsymbol{U}_2 - \kappa \boldsymbol{I}_n$, where $\kappa = n^{-1}\sum_{i=m+1}^{d}\lambda_i$. Suppose that $\alpha_1 = \cdots = \alpha_{s_1} > \alpha_{s_1+1} = \cdots = \alpha_{s_2} > \cdots > \alpha_{s_{l-1}+1} = \cdots = \alpha_{s_l} (= \alpha_m)$, where $l \le m$. For every i $(= 1, ..., l)$, let $\boldsymbol{U}_{1i} = n^{-1}\sum_{j=1}^{s_i}\lambda_j \boldsymbol{z}_j \boldsymbol{z}_j^T$. Let $\tilde{\lambda}_{i1} \ge \cdots \ge \tilde{\lambda}_{is_i}$ be eigenvalues of $\boldsymbol{U}_{1i}$. Let $\tilde{\boldsymbol{u}}_{ij} (\in \boldsymbol{R}_n)$ be an eigenvector corresponding to $\tilde{\lambda}_{ij}$ $(j = 1, ..., s_i)$. Then, we have the eigen-decomposition as $\boldsymbol{U}_{1i} = \sum_{j=1}^{s_i}\tilde{\lambda}_{ij}\tilde{\boldsymbol{u}}_{ij}\tilde{\boldsymbol{u}}_{ij}^T$. Let $\tilde{\boldsymbol{z}}_j = (||n^{-1/2}\boldsymbol{z}_j||)^{-1} n^{-1/2}\boldsymbol{z}_j$ $(j = 1, ..., d)$.

Proof of Theorem 1. By Chebyshev's inequality, for any $\tau > 0$, one has for each off-diagonal element $(i' \ne j')$ of $(n/\sum_{i=1}^{d}\lambda_i)\boldsymbol{S}_D$ that $P((\sum_{i=1}^{d}\lambda_i)^{-1}|\sum_{i=1}^{d}\lambda_i z_{ii'} z_{ij'}| > \tau) \le \tau^{-2}(\sum_{i=1}^{d}\lambda_i)^{-2}\sum_{i=1}^{d}\lambda_i^2 \to 0$ as $d \to \infty$ under (2). Thus each off-diagonal element of $(n/\sum_{i=1}^{d}\lambda_i)\boldsymbol{S}_D$ converges to 0 in probability as $d \to \infty$ under (2). Thus we have that

$$\frac{n}{\sum_{i=1}^{d}\lambda_i}\boldsymbol{S}_D \to \mathrm{diag}\,(D_1, ..., D_n)$$

in probability. Here, we have that $P(|D_k - 1| \le \tau) = 1 - P(|D_k - 1| > \tau) \ge 1 - \tau^{-2}V(D_k)$. When the components of $\boldsymbol{Z}$ satisfy (6), it holds that $V(D_k) \to 0$. Thus we have that

$D_k,\ k=1,...,n$, converge to 1 in probability. When the components of $\boldsymbol{Z}$ do not satisfy (6), it holds that D_k has $O_p(1)$ for $k=1,...,n$. It concludes the result. $\square$

Proof of Theorem 2. By Chebyshev's inequality and Markov's inequality, for any $\tau>0$, we have that $P(\sum_{i',j'}(\sum_{i=1}^d \lambda_i)^{-2}(\sum_{i=1}^d \lambda_i z_{ii'}z_{ij'})^2>\tau)\le n^2\tau^{-1}(\sum_{i=1}^d\lambda_i)^{-2}\sum_{i=1}^d\lambda_i^2\to 0$ and $\sum_{k=1}^n P(|D_k-1|>\tau)\le n\tau^{-2}V(D_k)\to 0$ under (8). Thus, in a way to similar to (13)-(14), it concludes the result. $\square$

The following lemma was obtained by Yata and Aoshima [14].

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

(i) *$d\to\infty$ either when $n\to\infty$ or n is fixed for j such that $\alpha_j>1/2$;*

(ii) *$d\to\infty$ and there exists a positive constant ε_j satisfying $d^{1-2\alpha_j}/n<d^{-\varepsilon_j}$.*

Lemma 2. *It holds that $d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{2n}=o_p(1)$ $(j=1,...,m)$ under (i)-(ii) of Theorem 3.*

Proof. By using Chebyshev's inequality, for any $\tau>0$ and the uniform bound $M(>0)$ for the fourth moments condition, one has under (i)-(ii) of Theorem 3 that

$$\begin{aligned}\sum_{k=1}^n P\Big(d^{-\alpha_j}|u_{kk}-\kappa|>\tau\Big) &= \sum_{k=1}^n P\Big((nd^{\alpha_j})^{-1}\Big|\sum_{s=m+1}^d\lambda_s(z_{sk}^2-1)\Big|>\tau\Big)\\ &\le (\tau n^{1/2}d^{\alpha_j})^{-2}M\Big(\sum_{s=m+1}^d\lambda_s^2\Big)\\ &\le(\tau n^{1/2}d^{\alpha_j})^{-2}Md\lambda_{m+1}^2=O(d^{1-2\alpha_j}/n)=o(1).\end{aligned}$$

Thus it holds that $d^{-\alpha_j}(u_{kk}-\kappa)=o_p(1)$ for every k $(=1,...,n)$. Note that $d^{1-2\alpha_i}/n(d)=d^{1-2\alpha-\gamma}$. From (ii) of Theorem 3, there exists a positive constant ε_j satisfying $1-2\alpha-\gamma<-\varepsilon_j$. Thus we have $d^{1-2\alpha_j}/n(d)<d^{-\varepsilon_j}$. We claim Lemma 1 under (i)-(ii) of Theorem 3. Then, we obtain for $j=1,...,m$, that

$$d^{-\alpha_j}\left(\boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{2n}\right)=d^{-\alpha_j}\left(\boldsymbol{e}_{1n}^T\boldsymbol{U}_{21}\boldsymbol{e}_{2n}+\boldsymbol{e}_{1n}^T\mathrm{diag}(u_{11}-\kappa,...,u_{nn}-\kappa)\boldsymbol{e}_{2n}\right)=o_p(1).$$

It concludes the result. $\square$

Lemma 3. *It holds as $d \to \infty$ and $n \to \infty$ that*

$$\boldsymbol{z}_i^T \boldsymbol{U}_{22} \boldsymbol{z}_{i'} = O_p(d^{1/2}) \quad (i = 1, ..., m;\ i' = 1, ..., m).$$

Proof. One can write that

$$\boldsymbol{z}_i^T \boldsymbol{U}_{22} \boldsymbol{z}_{i'} = \sum_{k_1 \neq k_2}^{n} z_{ik_1} z_{i'k_2} u_{k_1 k_2} + \sum_{k=1}^{n} z_{ik} z_{i'k} (u_{kk} - \kappa).$$

We first consider the case of $i = i'$. Note that $E(z_{ik_1}^2 z_{ik_2} z_{ik_3} u_{k_1 k_2} u_{k_1 k_3}) = 0$ $(k_1 \neq k_2 \neq k_3)$, $E(z_{ik_1}^2 z_{ik_2}^2 u_{k_1 k_2}^2) = n^{-2} \sum_{s=m+1}^{d} \lambda_s^2$ $(k_1 \neq k_2)$, $E(z_{ik_1}^2 z_{ik_2}^2 (u_{k_1 k_1} - \kappa)(u_{k_2 k_2} - \kappa)) = 0$ $(k_1 \neq k_2)$ and $E(z_{ik}^4 (u_{kk} - \kappa)^2) \leq M^2 n^{-2} \sum_{s=m+1}^{d} \lambda_s^2$ for the uniform bound M for the fourth moments condition. Then, for any $\tau > 0$, one has as $d \to \infty$ and $n \to \infty$ that

$$P\Big(|\sum_{k_1 \neq k_2} z_{ik_1} z_{ik_2} u_{k_1 k_2}| > \tau d^{1/2}\Big) \leq \tau^{-2} d^{-1} \sum_{k_1 \neq k_2} E(z_{ik_1}^2 z_{ik_2}^2 u_{k_1 k_2}^2) = O(\tau^{-2}),$$

$$P\Big(|\sum_{k=1}^{n} z_{ik}^2 (u_{kk} - \kappa)| > \tau d^{1/2}\Big) \leq \tau^{-2} n^{-1} d^{-1} M^2 \sum_{s=m+1}^{d} \lambda_s^2 = O(n^{-1}) = o(1).$$

Thus it holds that

$$\boldsymbol{z}_i^T \boldsymbol{U}_{22} \boldsymbol{z}_i = O_p(d^{1/2}) \quad (i = 1, ..., m).$$

As for the case of $i \neq i'$, note that

$$P\Big(|\sum_{k_1 \neq k_2} z_{ik_1} z_{i'k_2} u_{k_1 k_2}| > \tau d^{1/2}\Big) = O(\tau^{-2}), \quad P\Big(|\sum_{k=1}^{n} z_{ik} z_{i'k} (u_{kk} - \kappa)| > \tau d^{1/2}\Big) = o(1).$$

Therefore, we conclude the result. □

Lemma 4. *It holds as $d \to \infty$ and $n \to \infty$ that*

$$n^{-1/2} \boldsymbol{z}_i^T \boldsymbol{U}_{22} \boldsymbol{e}_{1n} = O_p((d/n)^{1/2}) \quad (i = 1, ..., m).$$

Proof. We have that

$$||n^{-1/2} \boldsymbol{z}_i^T \mathrm{diag}(u_{11} - \kappa, ..., u_{nn} - \kappa)||^2 = \sum_{k=1}^{n} n^{-1} z_{ik}^2 (u_{kk} - \kappa)^2.$$

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

$$P\Big(\sum_{k=1}^{n} n^{-1} z_{ik}^2 (u_{kk} - \kappa)^2 > \tau d/n\Big) \le \tau^{-1} d^{-1} \sum_{k=1}^{n} E(u_{kk} - \kappa)^2 = O(1/n) = o(1).$$

Thus it holds that $||n^{-1/2} \boldsymbol{z}_i^T \mathrm{diag}(u_{11} - \kappa, ..., u_{nn} - \kappa)|| = o_p((d/n)^{1/2})$. Next, we have that

$$||n^{-1/2} \boldsymbol{z}_i^T \boldsymbol{U}_{21}||^2 = \sum_{k_1 \ne k_2} n^{-1} z_{ik_1}^2 u_{k_1 k_2}^2 + \sum_{k_1 \ne k_2} n^{-1} z_{ik_1} z_{ik_2} \sum_{k_3 (\backslash k_1, k_2)} u_{k_1 k_3} u_{k_2 k_3}, \tag{22}$$

where $(\backslash i, j)$ excludes numbers i, j. We consider the first term in (22). We have as $d \to \infty$ and $n \to \infty$ that

$$P(\sum_{k_1 \ne k_2} n^{-1} z_{ik_1}^2 u_{k_1 k_2}^2 > \tau d/n) \le \tau^{-1} d^{-1} \sum_{k_1 \ne k_2} E(u_{k_1 k_2}^2) = O(\tau^{-1}). \tag{23}$$

Now, we consider the second term in (22). Note that $E(u_{k_1 k_3}^2 u_{k_2 k_3}^2) = O(d^2/n^4)$ and $E(u_{k_1 k_3} u_{k_2 k_3} u_{k_1 k_4} u_{k_2 k_4}) = O(d/n^4)$ for $k_1 \ne k_2 \ne k_3 \ne k_4$. By using Chebyshev's inequality, we have that

$$\begin{aligned} &P\Big(|\sum_{k_1 \ne k_2} n^{-1} z_{ik_1} z_{ik_2} \sum_{k_3 (\backslash k_1, k_2)}^{n} u_{k_1 k_3} u_{k_2 k_3}| > \tau d/n\Big) \\ &\le \tau^{-2} d^{-2} (n^3 E(u_{k_1 k_3}^2 u_{k_2 k_3}^2) + n^4 E(u_{k_1 k_3} u_{k_2 k_3} u_{k_1 k_4} u_{k_2 k_4})) = O(n^{-1}) + O(d^{-1}) = o(1). \end{aligned} \tag{24}$$

By combining (23)-(24) with (22), it holds that $||n^{-1/2} \boldsymbol{z}_i^T \boldsymbol{U}_{21}|| = O_p((d/n)^{1/2})$. Thus we have that

$$n^{-1/2} \boldsymbol{z}_i^T \boldsymbol{U}_{22} \boldsymbol{e}_{1n} = n^{-1/2} \boldsymbol{z}_i^T (\mathrm{diag}(u_{11} - \kappa, ..., u_{nn} - \kappa) + \boldsymbol{U}_{21}) \boldsymbol{e}_{1n} = O_p((d/n)^{1/2}).$$

It concludes the result. □

Lemma 5. *Assume that the first m population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. Then, it holds under (i)-(ii) of Theorem 3 that*

$$\frac{\hat{\lambda}_j - \kappa}{\lambda_j} = ||n^{-1/2} \boldsymbol{z}_j||^2 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j} n^{-1}), \quad \hat{\boldsymbol{u}}_j^T \tilde{\boldsymbol{z}}_j = 1 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j} n^{-1})$$
$$(j = 1, ..., m). \tag{25}$$

Proof. By using Chebyshev's inequality, for any τ (>0), one has as $n\to\infty$ that

$$P(|n^{-1}\boldsymbol{z}_j^T\boldsymbol{z}_{j'}|>n^{-1/2}\tau)=P\Big(\Big|n^{-1}\sum_{k=1}^{n}z_{jk}z_{j'k}\Big|>n^{-1/2}\tau\Big)=O(\tau^{-2})\quad(j\neq j').$$

Thus we claim as $n\to\infty$ that $n^{-1}\boldsymbol{z}_j^T\boldsymbol{z}_{j'}=O_p(n^{-1/2})$ $(j\neq j')$. Note that $||n^{-1/2}\boldsymbol{z}_j||^2=1+o_p(1)$ as $n\to\infty$. Let us consider that $\boldsymbol{S}_D-\kappa\boldsymbol{I}_n=\boldsymbol{U}_1+\boldsymbol{U}_{22}$. For λ_j $(j=1,...,s_1)$ that holds power α_{s_1}, we have from Lemma 2 that $d^{-\alpha_j}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{2n}=o_p(1)$ under (i)-(ii) of Theorem 3. Then, it holds that $\lambda_1||n^{-1/2}\boldsymbol{z}_1||^2>\cdots>\lambda_m||n^{-1/2}\boldsymbol{z}_m||^2$ and $\lambda_1||n^{-1/2}\boldsymbol{z}_1||^2>\boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{1n}$ w.p.1. Thus we have under (i)-(ii) of Theorem 3 that

$$\begin{aligned}\frac{\hat{\lambda}_1-\kappa}{\lambda_1}&=\lambda_1^{-1}\hat{\boldsymbol{u}}_1^T\left(\boldsymbol{U}_1+\boldsymbol{U}_{22}\right)\hat{\boldsymbol{u}}_1=\hat{\boldsymbol{u}}_1^T\sum_{s=1}^{m}\left(\frac{\lambda_s||n^{-1/2}\boldsymbol{z}_s||^2\tilde{\boldsymbol{z}}_s\tilde{\boldsymbol{z}}_s^T}{\lambda_1}\right)\hat{\boldsymbol{u}}_1+o_p(1)\\&=||n^{-1/2}\boldsymbol{z}_1||^2+o_p(1)=1+o_p(1).\end{aligned}$$

Then, it holds that $\hat{\boldsymbol{u}}_1^T\tilde{\boldsymbol{z}}_1=1+o_p(1)$. There exists a random variable $\varepsilon_1\in[0,1]$ and $\boldsymbol{y}_1\in\boldsymbol{R}_n$ such that $\hat{\boldsymbol{u}}_1=\tilde{\boldsymbol{z}}_1\sqrt{1-\varepsilon_1^2}+\varepsilon_1\boldsymbol{y}_1$ and $\tilde{\boldsymbol{z}}_1^T\boldsymbol{y}_1=0$. Here, we first consider the case when $\alpha_{s_1}\geq 1/2$. Then, from Lemmas 3-4, we have under (i)-(ii) of Theorem 3 that

$$\lambda_1^{-1}\tilde{\boldsymbol{z}}_1^T\boldsymbol{U}_{22}\tilde{\boldsymbol{z}}_1=O_p(n^{-1}),\quad \lambda_1^{-1}\tilde{\boldsymbol{z}}_1^T\boldsymbol{U}_{22}\boldsymbol{y}_1=O_p(n^{-1/2}).$$

By noting that $\varepsilon_1=o_p(1)$, it holds that $\sqrt{1-\varepsilon_1^2}=1+o_p(1)$. Then, we have that

$$\begin{aligned}\frac{\hat{\lambda}_1-\kappa}{\lambda_1}=&\hat{\boldsymbol{u}}_1^T\left(\sum_{j=1}^{m}\frac{\lambda_j}{\lambda_1}||n^{-1/2}\boldsymbol{z}_1||^2\tilde{\boldsymbol{z}}_j\tilde{\boldsymbol{z}}_j^T+\lambda_1^{-1}\boldsymbol{U}_{22}\right)\hat{\boldsymbol{u}}_1\\=&||n^{-1/2}\boldsymbol{z}_1||^2+\max_{\varepsilon_1}\Bigg\{-\varepsilon_1^2||n^{-1/2}\boldsymbol{z}_1||^2+O_p(\varepsilon_1 n^{-1/2})\\&+\varepsilon_1^2\boldsymbol{y}_1^T\left(\sum_{j=2}^{m}\frac{\lambda_j}{\lambda_1}||n^{-1/2}\boldsymbol{z}_j||^2\tilde{\boldsymbol{z}}_j\tilde{\boldsymbol{z}}_j^T\right)\boldsymbol{y}_1\Bigg\}+O_p(n^{-1}).\end{aligned}$$

From the fact that $||n^{-1/2}\boldsymbol{z}_1||^2>\lambda_1^{-1}\lambda_2||n^{-1/2}\boldsymbol{z}_2||^2$ w.p.1, we have under (i)-(ii) of Theorem 3 that

$$\begin{aligned}&\max_{\varepsilon_1}\Bigg\{-\varepsilon_1^2||n^{-1/2}\boldsymbol{z}_1||^2+O_p(\varepsilon_1 n^{-1/2})+\varepsilon_1^2\boldsymbol{y}_1^T\left(\sum_{j=2}^{m}\frac{\lambda_j}{\lambda_1}||n^{-1/2}\boldsymbol{z}_j||^2\tilde{\boldsymbol{z}}_j\tilde{\boldsymbol{z}}_j^T\right)\boldsymbol{y}_1\Bigg\}\\&\leq\max_{\varepsilon_1}\Bigg\{-\varepsilon_1^2||n^{-1/2}\boldsymbol{z}_1||^2+O_p(\varepsilon_1 n^{-1/2})+\varepsilon_1^2\frac{\lambda_2}{\lambda_1}||n^{-1/2}\boldsymbol{z}_2||^2\Bigg\}=O_p(n^{-1}),\end{aligned}$$

so that $\varepsilon_1 = O_p(n^{-1/2})$. Thus it holds under (i)-(ii) of Theorem 3 that

$$\frac{\hat{\lambda}_1 - \kappa}{\lambda_1} = ||n^{-1/2}\boldsymbol{z}_1||^2 + O_p(n^{-1})$$

together with that $\hat{\boldsymbol{u}}_1^T\tilde{\boldsymbol{z}}_1 = 1+O_p(n^{-1})$, $\hat{\boldsymbol{u}}_2^T\tilde{\boldsymbol{z}}_1 = O_p(n^{-1/2})$ and $\hat{\boldsymbol{u}}_1^T\tilde{\boldsymbol{z}}_2 = O_p(n^{-1/2})$. Similarly, we claim under (i)-(ii) of Theorem 3 that

$$\frac{\hat{\lambda}_j - \kappa}{\lambda_j} = ||n^{-1/2}\boldsymbol{z}_j||^2 + O_p(n^{-1}), \quad \hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_j = 1 + O_p(n^{-1}) \quad (j = 1, ..., s_1). \tag{26}$$

Next, we consider the case when $\alpha_1 \in (0, 1/2)$. Then, from Lemmas 3-4, we have under (i)-(ii) of Theorem 3 that

$$\lambda_1^{-1}\tilde{\boldsymbol{z}}_1^T\boldsymbol{U}_{22}\tilde{\boldsymbol{z}}_1 = O_p(d^{1/2-\alpha_1}n^{-1}), \quad \lambda_1^{-1}\tilde{\boldsymbol{z}}_1^T\boldsymbol{U}_{22}\boldsymbol{y}_1 = O_p(d^{1/2-\alpha_1}n^{-1/2}).$$

In a way similar to the case when $\alpha_1 \geq 1/2$, we have that $\varepsilon_1 = O_p(d^{1/2-\alpha_1}n^{-1/2})$. Thus it holds under (i)-(ii) of Theorem 3 that

$$\frac{\hat{\lambda}_j - \kappa}{\lambda_j} = ||n^{-1/2}\boldsymbol{z}_j||^2 + O_p(d^{1-2\alpha_j}n^{-1}), \quad \hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_j = 1 + O_p(d^{1-2\alpha_j}n^{-1}) \quad (j = 1, ..., s_1). \tag{27}$$

By combining (26)-(27), we can write that

$$\frac{\hat{\lambda}_j - \kappa}{\lambda_j} = ||n^{-1/2}\boldsymbol{z}_j||^2 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j}n^{-1}), \quad \hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_j = 1 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j}n^{-1})$$
$$(j = 1, ..., s_1) \tag{28}$$

under (i)-(ii) of Theorem 3.

Finally, we consider the case that λ_j $(j = s_2, ..., m)$ that holds power $\leq \alpha_{s_2}$. Then, in a way similar to the proof of Theorems 3.1-3.2 in Yata and Aoshima [14], in view of Remark 14, it holds (28) under (i)-(ii) of Theorem 3. It concludes the results. □

Remark 14. Assume that the first m population eigenvalues are distinct as $\lambda_1 > \cdots > \lambda_m$. For $\tilde{\lambda}_{ij}$ $(i = 1, ..., l;\ j = 1, ..., s_i)$ it holds as $d \to \infty$ and $n \to \infty$ that $\lambda_j^{-1}\tilde{\lambda}_{ij} = 1 + o_p(1)$ and $\tilde{\boldsymbol{u}}_{ij}^T\tilde{\boldsymbol{z}}_j = 1+O_p(n^{-1})$. For $\tilde{\boldsymbol{u}}_{ij'}$ and $\hat{\boldsymbol{u}}_j$ $(i = 1, ..., l-1;\ j \in [s_i+1, s_{i+1}];\ j' = 1, ..., s_i)$ it holds that $\tilde{\boldsymbol{u}}_{ij'}^T\hat{\boldsymbol{u}}_j = O_p(d^{\alpha_j-\alpha_{j'}}/n^{1/2}) + O_p(d^{1/2-\alpha_{j'}}/n^{1/2})$ under (i)-(ii) of Theorem 3.

Remark 15. When the population eigenvalues are not distinct, we consider the case as follows: Suppose that $\lambda_1 = \cdots = \lambda_{t_1} > \lambda_{t_1+1} = \cdots = \lambda_{t_2} > \cdots > \lambda_{t_{r-1}+1} = \cdots = \lambda_{t_r}$ $(= \lambda_m)$, where $r \le m$. We can claim under (i)-(ii) of Theorem 3 that

$$\frac{\hat{\lambda}_j - \kappa}{\lambda_j} = \sum_{i'=t_{i-1}+1}^{t_i} ||n^{-1/2}\boldsymbol{z}_{i'}||^2(\hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_{i'})^2 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j}n^{-1}) = 1 + o_p(1)$$

$$(i = 1, ..., r;\ j = t_{i-1}+1, ..., t_i),$$

where $t_0 = 0$.

Lemma 6. *Assume that λ_j $(j \le m)$ has multiplicity one. Then, for the subscript j, we have (25) under (i)-(ii) of Theorem 3.*

Proof. From Remark 15, we have $\hat{\boldsymbol{u}}_j^T\tilde{\boldsymbol{z}}_{j'} = O_p(n^{-1/2}) + O_p(d^{1/2-\alpha_j}n^{-1/2})$ for $j' < j$ and $j', j \in [s_{i-1}+1, ..., s_i]$ $(i = 1, ..., r)$, where $s_0 = 0$. Then, in a way similar to the proof of Lemma 5, we obtain the result. □

Lemma 7. *Let*

$$\delta_j = \frac{tr(\boldsymbol{S}_D) - \sum_{i=1}^{j}\hat{\lambda}_i}{(n-j)\lambda_j} - \frac{\kappa}{\lambda_j} \quad (j = 1, ..., m).$$

Then, we have under (i)-(ii) of Theorem 3 that $\delta_j = O_p(n^{-1})$ $(j = 1, ..., m)$.

Proof. Note that $\mathrm{tr}(\boldsymbol{S}_D) = \sum_{j=1}^{d}\lambda_j||n^{-1/2}\boldsymbol{z}_j||^2$. By using Chebyshev's inequality, for any $\tau > 0$ and the uniform bound M for the fourth moments condition, one has under (i)-(ii) of Theorem 3 that

$$\begin{aligned}
&P\left(d^{-\alpha_j}\left|n^{-1}\sum_{s=m+1}^{d}\lambda_s||n^{-1/2}\boldsymbol{z}_s||^2 - \kappa\right| > \tau n^{-1}\right)\\
&= P\left(d^{-\alpha_j}\left|n^{-1}\sum_{s=m+1}^{d}\lambda_s\sum_{k=1}^{n}(z_{sk}^2 - 1)\right| > \tau\right) = O(d^{1-2\alpha_j}/n) = o(1).
\end{aligned}$$

Note that $\lambda_j^{-1}n^{-1}\sum_{i=j+1}^{m}\lambda_i||n^{-1/2}\boldsymbol{z}_i||^2 = O_p(n^{-1})$ for $j = 1, ..., m-1$. Thus it holds that

$$\lambda_j^{-1}\left(n^{-1}\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^{j}n^{-1}\lambda_i||n^{-1/2}\boldsymbol{z}_i||^2 - \kappa\right) = O_p(n^{-1}) \quad (j = 1, ..., m). \tag{29}$$

Let $s_0=0$. Here, for $i=1,...,l$ and $j=s_{i-1}+1,...,s_i$, from Lemma 2, we have under (i)-(ii) of Theorem 3 that $\lambda_j^{-1}\boldsymbol{e}_{1n}^T(\boldsymbol{S}_D-\kappa\boldsymbol{I}_n)\boldsymbol{e}_{1n}=\lambda_j^{-1}\boldsymbol{e}_{1n}^T\boldsymbol{U}_{1i}\boldsymbol{e}_{1n}+o_p(1)$. Note that $\mathrm{rank}(\boldsymbol{U}_{1i})=\mathrm{rank}(\sum_{j=1}^{s_i}\lambda_j||n^{-1/2}\boldsymbol{z}_j||^2\tilde{\boldsymbol{z}}_j\tilde{\boldsymbol{z}}_j^T)=s_i$ w.p.1. Thus it holds that

$$d^{-\alpha_{s_i}}\left(\mathrm{tr}(\boldsymbol{U}_{1i})-\sum_{i=1}^{s_i}(\hat{\lambda}_i-\kappa)\right)=o_p(1).$$

Then, from Lemma 5 and Remark 15, for $i=1,...,l$ and $j=s_{i-1}+1,...,s_i-1$, we have under (i)-(ii) of Theorem 3 that

$$\begin{aligned}
&d^{-\alpha_j}\left(\sum_{j'=1}^{j}(\lambda_{j'}||n^{-1/2}\boldsymbol{z}_{j'}||^2-(\hat{\lambda}_{j'}-\kappa))\right)\\
&=d^{-\alpha_j}\left(\mathrm{tr}(\boldsymbol{U}_{1i})-\sum_{j'=1}^{s_i}(\hat{\lambda}_{j'}-\kappa)\right)-d^{-\alpha_j}\left(\sum_{j'=j+1}^{s_i}(\lambda_{j'}||n^{-1/2}\boldsymbol{z}_{j'}||^2-(\hat{\lambda}_{j'}-\kappa))\right)\\
&=o_p(1). \qquad (30)
\end{aligned}$$

When $j=s_i$, we can claim (30). By combining (29) with (30), it holds under (i)-(ii) of Theorem 3 that

$$\begin{aligned}
&\frac{\mathrm{tr}(\boldsymbol{S}_D)-\sum_{i=1}^{j}\hat{\lambda}_i}{(n-j)\lambda_j}-\frac{\kappa}{\lambda_j}\\
&=\frac{\mathrm{tr}(\boldsymbol{S}_D)-\sum_{i=1}^{j}\lambda_i||n^{-1/2}\boldsymbol{z}_i||^2+\sum_{i=1}^{j}(\lambda_i||n^{-1/2}\boldsymbol{z}_i||^2-(\hat{\lambda}_i-\kappa))}{(n-j)\lambda_j}-\frac{n\kappa}{(n-j)\lambda_j}\\
&=\Big(\frac{n}{n-j}\Big)\frac{n^{-1}\mathrm{tr}(\boldsymbol{S}_D)-\sum_{i=1}^{j}n^{-1}\lambda_i||n^{-1/2}\boldsymbol{z}_i||^2-\kappa}{\lambda_j}+o_p(n^{-1})\\
&=O_p(n^{-1})\quad(j=1,...,m).
\end{aligned}$$

It concludes the result. $\square$

Proof of Theorems 3 and 4. We first consider the case when λ_j $(j\le m)$ has multiplicity one. We write that

$$\frac{\tilde{\lambda}_j}{\lambda_j}=\frac{\hat{\lambda}_j-\kappa}{\lambda_j}-\delta_j.$$

By combining Lemma 6 with Lemma 7, we have under (i)-(ii) of Theorem 3 that

$$\frac{\tilde{\lambda}_j}{\lambda_j}=||n^{-1/2}\boldsymbol{z}_j||^2+O_p(n^{-1})+O_p(d^{1-2\alpha_j}n^{-1}). \qquad (31)$$

Here, as for Theorem 4, recall that $V(z_{jk}^2) = M_j$. By using the central limiting theorem, one has as $n \to \infty$ that $(nM_j)^{-1/2}(||\boldsymbol{z}_j||^2 - n) = (nM_j)^{-1/2}(\sum_{k=1}^n z_{jk}^2 - n) \Rightarrow N(0,1)$. Note that $d^{1-2\alpha_j} n^{-1} = o_p(n^{-1/2})$ under (i)-(ii) of Theorem 4. Hence, under (i)-(ii) of Theorem 4, we have from (31) that

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

It concludes the result of Theorem 4. On the other hand, we can claim under (i)-(ii) of Theorem 3 that

$$\frac{\tilde{\lambda}_j}{\lambda_j} = 1 + o_p(1). \tag{32}$$

Next, we consider the case when $\lambda_j = \lambda_{j'}$ ($j \le m$) for some j'. One may refer to Remark 15. Since we can claim that $\tilde{\lambda}_j/\lambda_j = 1 + o_p(1)$, under (i)-(ii) of Theorem 3, in a way similar to (32), it concludes the result of Theorem 3. □

Proof of Corollary 1. Let us write that $\boldsymbol{\Lambda}^{-1/2}\boldsymbol{H}^T(\boldsymbol{X} - \overline{\boldsymbol{X}}) = [\acute{\boldsymbol{z}}_1, ..., \acute{\boldsymbol{z}}_d]^T$ and $\acute{\boldsymbol{z}}_j = (\acute{z}_{j1}, ..., \acute{z}_{jn})^T$ for $j = 1, ..., d$. Then, we have that $\acute{z}_{jk} = z_{jk} - \bar{z}_j$ for $k = 1, ..., n$, where $\bar{z}_j = \sum_{k=1}^n z_{jk}/n$. Let $E(z_{jk}) = \mu_j$ for $j = 1, ..., d$. We write that $\acute{z}_{jk} = \ddot{z}_{jk} + z_{oj}$, where $\ddot{z}_{jk} = z_{jk} - \mu_j$ and $z_{oj} = \mu_j - \bar{z}_j$ ($j = 1, ..., d;\ k = 1, ..., n$). Now, let us write that n-vectors $\ddot{\boldsymbol{z}}_j = (\ddot{z}_{j1}, ..., \ddot{z}_{jn})^T$ and $\boldsymbol{z}_{oj} = (z_{oj}, ..., z_{oj})^T$ for $j = 1, ..., d$. Then, we have that

$$\boldsymbol{S}_{oD} = (n-1)^{-1}\Big(\sum_{s=1}^m \lambda_s \acute{\boldsymbol{z}}_s \acute{\boldsymbol{z}}_s^T + \sum_{s=m+1}^d \lambda_s(\ddot{\boldsymbol{z}}_s + \boldsymbol{z}_{os})(\ddot{\boldsymbol{z}}_s + \boldsymbol{z}_{os})^T\Big).$$

Let $\boldsymbol{1}_n = n^{-1/2}(1, ..., 1)^T$. Then, it holds that $\boldsymbol{1}_n^T \boldsymbol{S}_{oD} \boldsymbol{1}_n = 0$. Thus we may write that $\hat{\boldsymbol{u}}_n = \boldsymbol{1}_n$. By noting that $\hat{\boldsymbol{u}}_n^T \hat{\boldsymbol{u}}_j = 0$ for $j = 1, ..., n-1$, it holds that $\hat{\boldsymbol{u}}_j^T \boldsymbol{z}_{os} = 0$ for $j = 1, ..., n-1$ ($s = 1, ..., d$). We have that $\hat{\boldsymbol{u}}_j^T \sum_{s=m+1}^d \lambda_s(\ddot{\boldsymbol{z}}_s + \boldsymbol{z}_{os})(\ddot{\boldsymbol{z}}_s + \boldsymbol{z}_{os})^T \hat{\boldsymbol{u}}_j = \hat{\boldsymbol{u}}_j^T (\sum_{s=m+1}^d \lambda_s \ddot{\boldsymbol{z}}_s \ddot{\boldsymbol{z}}_s^T)\hat{\boldsymbol{u}}_j$, $j = 1, ..., n-1$. Let us write that $\ddot{\boldsymbol{U}}_{22} = (n-1)^{-1}\sum_{s=m+1}^d \lambda_s \ddot{\boldsymbol{z}}_s \ddot{\boldsymbol{z}}_s^T - (n-1)^{-1} n\kappa \boldsymbol{I}_n$. Similarly to Lemma 2, we have that $\boldsymbol{e}_{1n}^T \ddot{\boldsymbol{U}}_{22} \boldsymbol{e}_{1n} = o_p(1)$. By noting that $n^{-1}\acute{\boldsymbol{z}}_j^T \acute{\boldsymbol{z}}_{j'} = O_p(n^{-1/2})$ ($j \ne j'$) and $||n^{-1/2}\acute{\boldsymbol{z}}_j||^2 = ||n^{-1/2}\ddot{\boldsymbol{z}}_j||^2 + O_p(n^{-1}) = 1 + o_p(1)$, we can claim Lemmas 3-7 as well. Then, by replacing $\boldsymbol{S}_D$ with $\boldsymbol{S}_{oD}$, we claim the assertions of Theorems 3-4. □

Proof of Corollary 2. With the help of Lemma 5 and Remark 15, we have that $\hat{\lambda}_j/\kappa =$

$1+o_p(1)$ under the condition that $d\to\infty$ and $d/(n\lambda_j)\to\infty$. Then, we have that

$$\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_j=(n\hat{\lambda}_j)^{-1/2}\lambda_j^{1/2}\boldsymbol{z}_j^T\hat{\boldsymbol{u}}_j=\left(\frac{\lambda_j}{\kappa}\right)^{1/2}\frac{\boldsymbol{z}_j^T}{\sqrt{n}}\hat{\boldsymbol{u}}_j+o_p(1)=o_p(1).$$

It concludes the result. □

Proof of Corollary 3. From Lemma 6, the result is obtained straightforwardly. □

Proof of Theorem 5. For each j $(=1,...,m)$, let us write that

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

With the help of Theorem 3 and Lemma 6, we have that $\lambda_j^{-1}\mathrm{MSE}(\tilde{s}_j)=o_p(1)$ under (i)-(ii) of Theorem 3. It concludes the result. □

Proof of Corollary 4 and Theorem 6. Let us write that $\boldsymbol{\Lambda}^{-1/2}\boldsymbol{H}^T\boldsymbol{x}_{new}=(z_{1(new)},...,z_{d(new)})^T$. We first consider the case that $\lambda_1>\cdots>\lambda_m$. In view of Remark 14, we have under (i)-(ii) of Theorem 3 that

$$\begin{aligned}\frac{\hat{\boldsymbol{u}}_j^T\boldsymbol{U}_{1i}\boldsymbol{U}_{1i}\hat{\boldsymbol{u}}_j}{\lambda_j^2}&=\sum_{s=1}^{s_i}(\lambda_j^{-1}\lambda_s n^{-1/2}\boldsymbol{z}_s^T\hat{\boldsymbol{u}}_j)^2+o_p(1)\sum_{s,s'}(\lambda_j^{-1}\lambda_s n^{-1/2}\boldsymbol{z}_s^T\hat{\boldsymbol{u}}_j)(\lambda_j^{-1}\lambda_{s'}n^{-1/2}\boldsymbol{z}_{s'}^T\hat{\boldsymbol{u}}_j)\\&=o_p(1)\quad\text{for } i\ (=1,...,l-1)\text{ and } j\ (=s_i+1,...,s_{i+1}).\end{aligned}$$

In a way similar to the proof of Theorem 3 in Yata and Aoshima [16], it holds under (i)-(ii) of Theorem 3 that

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

For the case when λ_j $(j\le m)$ has multiplicity one, we can claim the above result.

First, we consider Theorem 6. From Lemma 6, we have that

$$\frac{\tilde{\boldsymbol{h}}_j^T\boldsymbol{x}_{new}}{\lambda_j^{1/2}}=\sum_{s=1}^{d}\frac{\lambda_s z_{s(new)}\boldsymbol{z}_s^T\hat{\boldsymbol{u}}_j}{(n\tilde{\lambda}_j\lambda_j)^{1/2}}=z_{j(new)}+\sum_{s=m+1}^{d}\frac{\lambda_s z_{s(new)}\boldsymbol{z}_s^T\hat{\boldsymbol{u}}_j}{(n\tilde{\lambda}_j\lambda_j)^{1/2}}+o_p(1).$$

From the proof of Lemma 5, we have under (i)-(ii) of Theorem 4 that

$$\hat{\boldsymbol{u}}_j = (1 + o_p(n^{-1/2}))\tilde{\boldsymbol{z}}_j + o_p(n^{-1/4}) \times \boldsymbol{y}_j, \tag{33}$$

where $\boldsymbol{y}_j \in \boldsymbol{R}_n$ such that $\boldsymbol{y}_j^T \tilde{\boldsymbol{z}}_j = 0$. Note that $\sum_{s=m+1}^d \lambda_s z_{s(new)} n^{-1/2} \boldsymbol{z}_s^T \tilde{\boldsymbol{z}}_j / \lambda_j = o_p(1)$ and $||n^{-1/4} \sum_{s=m+1}^d \lambda_s z_{s(new)} n^{-1/2} \boldsymbol{z}_s^T / \lambda_j||^2 = o_p(1)$ under (i)-(ii) of Theorem 4. Then, it holds from (33) that $\sum_{s=m+1}^d \lambda_s z_{s(new)} \boldsymbol{z}_s^T \hat{\boldsymbol{u}}_j / (n\tilde{\lambda}_j \lambda_j)^{1/2} = o_p(1)$. Thus we have under (i)-(ii) of Theorem 4 that

$$\frac{\tilde{\boldsymbol{h}}_j^T \boldsymbol{x}_{new}}{\lambda_j^{1/2}} = z_{j(new)} + o_p(1).$$

By noting that $s_{j(new)} = \lambda_j^{1/2} z_{j(new)}$, it concludes the result in Theorem 6.

Next, we consider Corollary 4. From (10), it holds that

$$\frac{\hat{\boldsymbol{h}}_j^T \boldsymbol{x}_{new}}{\lambda_j^{1/2}} = z_{j(new)} \sqrt{\frac{\lambda_j}{\hat{\lambda}_j}} + o_p(1) = z_{j(new)} + o_p(1)$$

under the conditions given by combining (YA-i)-(YA-ii) with (i)-(ii) of Theorem 4 (that is, the conditions (i), (ii) and (iii) of the present corollary). It concludes the result in Corollary 4. □

Proof of Theorem 7. We note that the conditions (i)-(ii) of Theorem 3 include the condition (i) of Theorem 7. From Lemma 5, we have under (i) of Theorem 7 that

$$\begin{aligned} \boldsymbol{h}_j^T \hat{\boldsymbol{h}}_j &= \big(||n^{-1/2}\boldsymbol{z}_j||^2 + \kappa/\lambda_j + O_p(n^{-1}) + O_p(d^{1-2\alpha_j} n^{-1}) \big)^{-1/2} ||n^{-1/2}\boldsymbol{z}_j|| \tilde{\boldsymbol{z}}_j^T \hat{\boldsymbol{u}}_j \\ &= \frac{||n^{-1/2}\boldsymbol{z}_j||}{(||n^{-1/2}\boldsymbol{z}_j||^2 + \kappa/\lambda_j)^{1/2}} + O_p(n^{-1}) + O_p(d^{1-2\alpha_j} n^{-1}). \end{aligned} \tag{34}$$

Here, by noting that $\alpha_1 < 1$, for any $\tau > 0$, we have as $d \to \infty$ that

$$\begin{aligned} &P\left(\kappa^{-1} \left| n^{-1} \sum_{s=1}^d \lambda_s ||n^{-1/2}\boldsymbol{z}_s||^2 - \kappa \right| > \tau \right) \\ &= P\left((n\kappa)^{-1} \left| \sum_{s=1}^m \lambda_s \sum_{k=1}^n \frac{z_{sk}^2}{n} + \sum_{s=m+1}^d \lambda_s n^{-1} \sum_{k=1}^n (z_{sk}^2 - 1) \right| > \tau \right) = O(d^{2\alpha_1 - 2}) + o(1) = o(1). \end{aligned}$$

Thus it holds that $\delta/\kappa = 1 + o_p(1)$. Then, we have from Lemma 5 and (34) that

$$\begin{aligned}
&\frac{\lambda_j}{\hat{\lambda}_j+\delta}\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_j\hat{\boldsymbol{h}}_j^T\boldsymbol{h}_j + \frac{\lambda_j}{\delta}(1-\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_j\hat{\boldsymbol{h}}_j^T\boldsymbol{h}_j)\\
&= \frac{\lambda_j^2}{(\lambda_j+\kappa)(\lambda_j+2\kappa)} + o_p(1) + \frac{1}{||n^{-1/2}\boldsymbol{z}_j||^2+\kappa/\lambda_j} + O_p((d^{\alpha_j}/\kappa)/n^{-1}) + O_p((d^{1-\alpha_j}/\kappa)/n^{-1})\\
&= \frac{\lambda_j^2}{(\lambda_j+\kappa)(\lambda_j+2\kappa)} + \frac{\lambda_j}{\lambda_j+\kappa} + o_p(1) = \frac{2\lambda_j}{\lambda_j+2\kappa} + o_p(1). \qquad (35)
\end{aligned}$$

Now, from Lemma 2, we have under (i) of Theorem 7 that

$$\kappa^{-1}\left(\boldsymbol{e}_{1n}^T\boldsymbol{U}_2\boldsymbol{e}_{1n}\right) = \kappa^{-1}\left(\boldsymbol{e}_{1n}^T\boldsymbol{U}_{22}\boldsymbol{e}_{1n} + \boldsymbol{e}_{1n}^T\mathrm{diag}(\kappa,...,\kappa)\boldsymbol{e}_{1n}\right) = 1 + o_p(1).$$

Thus it holds that $\kappa^{-1}\hat{\lambda}_j = \kappa^{-1}\hat{\boldsymbol{u}}_j^T(\boldsymbol{U}_1+\boldsymbol{U}_2)\hat{\boldsymbol{u}}_j > 0$ w.p.1 for all $j = 1,...,n$. We can write that $\tilde{\boldsymbol{z}}_j = \sum_{k=1}^n b_{jk}\hat{\boldsymbol{u}}_k$ $(j = 1,...,d)$, where $\sum_{k=1}^n b_{jk}^2 = 1$. From Lemma 5, we have that $\tilde{\boldsymbol{z}}_j^T\hat{\boldsymbol{u}}_j = b_{jj} = 1 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j}n^{-1})$. Thus it holds that $\sum_{k(\backslash j)}^n b_{jk}^2 = O_p(n^{-1}) + O_p(d^{1-2\alpha_j}n^{-1})$, where $(\backslash j)$ excludes number j. Here, we have for j, j' that

$$\begin{aligned}
\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_i\hat{\boldsymbol{h}}_i^T\boldsymbol{h}_{j'} &= \hat{\lambda}_i^{-1}(\lambda_j\lambda_{j'})^{1/2}||n^{-1/2}\boldsymbol{z}_j||||n^{-1/2}\boldsymbol{z}_{j'}||(\tilde{\boldsymbol{z}}_j^T\hat{\boldsymbol{u}}_i)(\tilde{\boldsymbol{z}}_{j'}^T\hat{\boldsymbol{u}}_i)\\
&= (\lambda_j\lambda_{j'})^{1/2}b_{ji}b_{j'i}\times O_p(\kappa^{-1}).
\end{aligned}$$

Note that

$$\sum_{i(\backslash j,j')}^n |b_{ji}||b_{j'i}| \le \Big(\sum_{i(\backslash j,j')}^n b_{ji}^2\Big)^{1/2}\Big(\sum_{i(\backslash j,j')}^n b_{j'i}^2\Big)^{1/2} = O_p(n^{-1}) + O_p(d^{1/2-\alpha_{j'}}n^{-1}) + O_p(d^{1-\alpha_j-\alpha_{j'}}n^{-1})$$

for $j'(>j)$ satisfying (i) of Theorem 7, where $(\backslash j, j')$ excludes numbers j, j'. Then, by noting that $\alpha_1 + \gamma - 3/2 < 0$ when $\alpha_1 < 1 - \gamma/2$ and $\gamma < 1$, we claim that

$$\begin{aligned}
\frac{(\lambda_j\lambda_{j'})^{1/2}}{\delta}\sum_{i(\backslash j,j')}^n |\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_i\hat{\boldsymbol{h}}_i^T\boldsymbol{h}_{j'}| &\le \sum_{i(\backslash j,j')}^n \lambda_j\lambda_{j'}|b_{ji}||b_{j'i}|\times O_p(\kappa^{-2})\\
&= O_p(d^{\alpha_j+\alpha_{j'}+\gamma-2}) + O_p(d^{\alpha_j+\gamma-3/2}) + O_p(d^{\gamma-1}) = o_p(1). \qquad (36)
\end{aligned}$$

On the other hand, by noting that $b_{jj'} = O_p(n^{-1/2}) + O_p(d^{1/2-\alpha_j}n^{-1/2})$ and $b_{j'j} = O_p(n^{-1/2}) +$

$O_p(d^{1/2-\alpha_{j'}}n^{-1/2})$, we claim for $j \neq j'$ that

$$\frac{(\lambda_j\lambda_{j'})^{1/2}}{\delta}(\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_j\hat{\boldsymbol{h}}_j^T\boldsymbol{h}_{j'}+\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_{j'}\hat{\boldsymbol{h}}_{j'}^T\boldsymbol{h}_{j'}) = \frac{\lambda_j\lambda_{j'}||\boldsymbol{z}_j/n^{1/2}||||\boldsymbol{z}_{j'}/n^{1/2}||}{\delta}\left(\frac{b_{jj}b_{j'j}}{\hat{\lambda}_j}+\frac{b_{jj'}b_{j'j'}}{\hat{\lambda}_{j'}}\right)$$
$$= O_p(d^{\alpha_j-1+\gamma/2})+O_p(d^{\alpha_{j'}-1+\gamma/2})+o_p(1)=o_p(1). \tag{37}$$

When $\lambda_{j'}/\kappa \to 0$, it holds that $\sum_{i(\backslash j,j')}^n |b_{ji}||b_{j'i}| = O_p(n^{-1/2})+O_p(d^{1/2-\alpha_j}n^{-1/2})$ by noting that $(\sum_{i(\backslash j,\ j')}^n b_{j'i}^2) \le 1$. Then, we have that

$$\frac{(\lambda_j\lambda_{j'})^{1/2}}{\delta}\sum_{i(\backslash j,j')}^n|\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_i\hat{\boldsymbol{h}}_i^T\boldsymbol{h}_{j'}| = \sum_{i(\backslash j,j')}^n(\lambda_j\lambda_{j'})|b_{ji}||b_{j'i}|\times O_p(\kappa^{-2}) = O_p(d^{\alpha_1-1+\gamma/2})+o_p(1)$$
$$= o_p(1). \tag{38}$$

Similarly, we claim (37) when $\lambda_{j'}/\kappa \to 0$. Then, by combining (35)-(38), we obtain that

$$v_{jj(\delta)} = \frac{2\lambda_j}{\lambda_j+2\kappa}+o_p(1)+\sum_{i(\backslash j)}^n\left(\frac{\lambda_j}{\hat{\lambda}_i+\delta}-\frac{\lambda_j}{\delta}\right)\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_i\hat{\boldsymbol{h}}_i^T\boldsymbol{h}_j = \frac{2\lambda_j}{\lambda_j+2\kappa}+o_p(1),$$
$$v_{jj'(\delta)} = o_p(1), \quad j'=j+1,...,d.$$

Next, we consider the case that $\lambda_j/\kappa \to 0$ as $d\to\infty$. Note that $\sum_{i=1}^n|b_{ji}||b_{j'i}| \le 1$. Then, it holds that

$$\frac{(\lambda_j\lambda_{j'})^{1/2}}{\kappa}\sum_{i=1}^n|\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_i\hat{\boldsymbol{h}}_i^T\boldsymbol{h}_{j'}| \le \frac{\lambda_j\lambda_{j'}}{\kappa}\sum_{i=1}^n|b_{ij}||b_{ij'}|\times O_p(\kappa^{-1})$$
$$= O_p(\lambda_j\lambda_{j'}\kappa^{-2}) = o_p(\lambda_j\kappa^{-1}), \quad j'=j,...,d.$$

Thus we have that

$$v_{jj(\delta)} = \lambda_j\boldsymbol{h}_j^T\boldsymbol{S}_\delta^{-1}\boldsymbol{h}_j = \frac{\lambda_j}{\delta}+\sum_{i=1}^n\left(\frac{\lambda_j}{\hat{\lambda}_i+\delta}-\frac{\lambda_j}{\delta}\right)\ \boldsymbol{h}_j^T\hat{\boldsymbol{h}}_i\hat{\boldsymbol{h}}_i^T\boldsymbol{h}_j = \frac{\lambda_j}{\kappa}+o_p(\lambda_j\kappa^{-1}),$$
$$v_{jj'(\delta)} = o_p(\lambda_j\kappa^{-1}), \quad j'=j+1,...,d.$$

It concludes the result. □

Proof of Theorem 8. Let $R_\omega = \{j \in [1, ..., n-1] | \acute{\lambda}_j > \omega\}$. Then, we have that

$$\boldsymbol{S}_\omega^{-1} = \sum_{j \in R_\omega} \tilde{\lambda}_j^{-1} \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T + \omega^{-1} \left(\boldsymbol{I}_d - \sum_{j \in R_\omega} \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T \right).$$

We note that the conditions (i)-(ii) of Theorem 3 include the conditions (i)-(ii) of Theorem 8. Note that $\acute{\lambda}_j > \omega$ for $j \in R_\omega$ and $\omega/\psi = 1 + o_p(1)$ under (i)-(ii) of Theorem 8. We first consider the case when $\gamma < 2/3$. Then, similarly to the proof of Theorem 7, we have for $j \neq j' \in R_\omega$ that

$$\frac{(\lambda_j \lambda_{j'})^{1/2}}{\omega} \sum_{i(\backslash j, j') \in R_\omega}^{n} |\boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_i \tilde{\boldsymbol{h}}_i^T \boldsymbol{h}_{j'}| = O_p(d^{\alpha_j + \alpha_{j'} - 1 - \gamma/2}) + O_p(d^{-\gamma/2}) + o_p(1) = o_p(1). \tag{39}$$

Similarly, we claim for $j \neq j' \in R_\omega$ that

$$\frac{(\lambda_j \lambda_{j'})^{1/2}}{\omega} (\boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_{j'} + \boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_{j'} \tilde{\boldsymbol{h}}_{j'}^T \boldsymbol{h}_{j'}) = o_p(1). \tag{40}$$

Here, from Lemma 5 and (31), we have under (i)-(ii) of Theorem 3 that

$$\boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_j = \left(\frac{\lambda_j}{\tilde{\lambda}_j} \right)^{1/2} \frac{\boldsymbol{z}_j^T}{\sqrt{n}} \hat{\boldsymbol{u}}_j = 1 + O_p(n^{-1}) + O_p(d^{1-2\alpha_j} n^{-1}).$$

Then, it holds that

$$\frac{\lambda_j}{\omega} (1 - \boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j) = O_p(d^{\alpha_j - 1/2 - (3/4)\gamma}) + O_p(d^{1/2 - \alpha_j - (3/4)\gamma}) = o_p(1). \tag{41}$$

Note that $\psi/\lambda_j = O(d^{1/2 - \gamma/4 - \alpha_j})$. Thus it holds that $\acute{\lambda}_j/\lambda_j = \max(\tilde{\lambda}_j, \omega)/\lambda_j = \max(1, \psi/\lambda_j) + o_p(1)$ under (i) of Theorem 8. We have that

$$\frac{\lambda_j}{\acute{\lambda}_j} \boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j = \frac{1}{\max(1 + \psi/\lambda_j)} + o_p(1). \tag{42}$$

Next, we consider case when $\gamma \in [2/3, 1)$. Similarly to the proof of Theorem 7, we claim (39)-(42). Then, by combining (39)-(42), we obtain under (i)-(ii) of Theorem 8 that

$$\begin{aligned} v_{jj(\omega)} &= \frac{\lambda_j}{\acute{\lambda}_j} \boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j + \frac{\lambda_j}{\omega} (1 - \boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j) + \sum_{i(\backslash j) \in R_\omega}^{n} \left(\frac{\lambda_j}{\acute{\lambda}_j} - \frac{\lambda_j}{\omega} \right) \boldsymbol{h}_j^T \tilde{\boldsymbol{h}}_j \tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j \\ &= \frac{1}{\max(1, \psi/\lambda_j)} + o_p(1), \\ v_{jj'(\delta)} &= o_p(1), \quad j' = j+1, ..., d. \end{aligned}$$

Finally, we consider the case that $\lambda_j/\psi \to 0$ as $d \to \infty$. Note that $\sum_{i\in R_\omega}^{n} |b_{ji}||b_{j'i}| \le 1$. Then, we have that

$$\frac{(\lambda_j\lambda_{j'})^{1/2}}{\omega}\sum_{i\in R_\omega}^{n} |\boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^T\boldsymbol{h}_{j'}| \le \frac{\lambda_j\lambda_{j'}}{\omega}\sum_{i\in R_\omega}^{n}|b_{ji}||b_{j'i}| \times O_p(\psi^{-1})$$
$$= O_p(\lambda_j\lambda_{j'}\psi^{-2}) = o_p(\lambda_j\psi^{-1}), \quad j' = j, ..., d.$$

Thus it holds that

$$v_{jj(\delta)} = \frac{\lambda_j}{\omega} + \sum_{i\in R_\omega}^{n}\left(\frac{\lambda_j}{\tilde{\lambda}_j} - \frac{\lambda_j}{\omega}\right)\boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^T\boldsymbol{h}_j = \frac{\lambda_j}{\psi} + o_p(\lambda_j\psi^{-1}),$$
$$v_{jj'(\delta)} = o_p(\lambda_j\psi^{-1}), \quad j' = j+1, ..., d.$$

It concludes the result. □

Acknowledgements

The authors would like to thank the anonymous referees for their constructive comments that made the exposition of the paper more accessible. Research of the second author was partially supported by Grant-in-Aid for Scientific Research (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 22300094.

References

[1] J. Ahn, J. S. Marron, K. M. Muller, Y. -Y. Chi, The high-dimension, low-sample-size geometric representation holds under mild conditions, Biometrika 94 (2007) 760-766.

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

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

[4] R. C. Bradley, Basic properties of strong mixing conditions. A survey and some open questions, Probab. Surv. 2 (2005) 107-144 (electronic).

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

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

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

[8] A. N. Kolmogorov, Y. A. Rozanov, On strong mixing conditions for stationary Gaussian processes, Theory Probab. Appl. 5 (1960) 204-208.

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

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

[11] M. S. Srivastava, Minimum distance classification rules for high dimensional data, J. Mult. Anal. 97 (2006) 2057-2070.

[12] M. S. Srivastava, Multivariate theory for analyzing high dimensional data, J. Japan Statist. Soc. 37 (2007) 53-86.

[13] M. S. Srivastava, T. Kubokawa, Comparison of discrimination methods for high dimensional data, J. Japan Statist. Soc. 37 (2007) 123-134.

[14] K. Yata, M. Aoshima, PCA consistency for non-gaussian data in high dimension, low sample size context, Commun. Statist.-Theory Meth. 38 (2009) 2634-2652.

[15] K. Yata, M. Aoshima, Intrinsic dimensionality estimation of high dimension, low sample size data with d-asymptotics, Commun. Statist.-Theory Meth. 39 (2010) 1511-1521.

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