Reconstruction of a high-dimensional low-rank matrix

著者別名	矢田 和善, 青嶋 誠
journal or publication title	Electronic Journal of Statistics
volume	10
number	1
page range	895-917
year	2016
URL	http://hdl.handle.net/2241/00145091

doi: 10.1214/16-EJS1128

Electronic Journal of Statistics
Vol. 10 (2016) 895–917
ISSN: 1935-7524
DOI: 10.1214/16-EJS1128

Reconstruction of a high-dimensional low-rank matrix

Kazuyoshi Yata* **and Makoto Aoshima**†

Institute of Mathematics,
University of Tsukuba,
Ibaraki 305-8571, Japan
e-mail: yata@math.tsukuba.ac.jp*;* aoshima@math.tsukuba.ac.jp

Abstract: We consider the problem of recovering a low-rank signal matrix in high-dimensional situations. The main issue is how to estimate the signal matrix in the presence of huge noise. We introduce the power spiked model to describe the structure of singular values of a huge data matrix. We first consider the conventional PCA to recover the signal matrix and show that the estimation of the signal matrix holds consistency properties under severe conditions. The conventional PCA is heavily subjected to the noise. In order to reduce the noise we apply the noise-reduction (NR) methodology and propose a new estimation of the signal matrix. We show that the proposed estimation by the NR method holds the consistency properties under mild conditions and improves the error rate of the conventional PCA effectively. Finally, we demonstrate the reconstruction procedures by using a microarray data set.

MSC 2010 subject classifications: Primary 62H25; secondary 62F12.
Keywords and phrases: Eigenstructure, HDLSS, noise-reduction methodology, PCA, singular value decomposition.

Received October 2015.

Contents

*Supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 26800078.

†Supported by Grants-in-Aid for Scientific Research (A) and Challenging Exploratory Research, JSPS, under Contract Numbers 15H01678 and 26540010.

1. Introduction

In this paper, we address the problem of recovering an unknown $d \times n$ low-rank matrix, $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$. $\boldsymbol{A}$ is called the signal matrix. Let $r = \text{rank}(\boldsymbol{A})$, where r is unknown. We assume r $(< \min\{d, n\})$ is fixed. For high-dimensional data, the estimation of the low-rank matrix is quite important in many fields such as genomics, image denoising, recommendation systems and so on. Negahban and Wainwright [5] and Rohde and Tsybakov [6] considered the problem for high-dimensional regression models. Shabalin and Nobel [7] considered the estimation of $\boldsymbol{A}$ when observations have a Gaussian noise. In this paper, we consider the problem of recovering $\boldsymbol{A}$ when observations have a non-Gaussian noise.

Suppose we have a $d \times n$ data matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, where

$$\boldsymbol{X} = \sqrt{n}\boldsymbol{A} + \boldsymbol{W}. \tag{1.1}$$

Here, $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n]$ is a $d \times n$ noise matrix, where $\boldsymbol{w}_j,\ j = 1, \ldots, n$, are independent and identically distributed (i.i.d.) as a d-dimensional distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_W$ $(\neq \boldsymbol{O})$. Note that $\boldsymbol{x}_j - \sqrt{n}\boldsymbol{a}_j,\ j = 1, \ldots, n$, are i.i.d. Let $\boldsymbol{\Sigma}_A = \boldsymbol{A}\boldsymbol{A}^T$. Then, it holds that $E(\boldsymbol{X}\boldsymbol{X}^T)/n = \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_W$ $(= \boldsymbol{\Sigma}$, say). Shabalin and Nobel [7] considered (1.1) in a high-dimensional setting, where the data dimension d and the sample size n increase at the same rate, i.e. $n/d \to c > 0$. They assumed that the elements of $\boldsymbol{W}$ are i.i.d. normal random variables. We note that the conditions such as "$n/d \to c > 0$" and the Gaussianity of the noise are often strict in real high-dimensional analyses. In this paper, we consider (1.1) in high-dimensional settings without assuming those conditions. We assume the divergence condition for d and n such as $d \to \infty$ either when n is fixed or $n \to \infty$. The divergence condition includes both high-dimension, low-sample-size (HDLSS) settings such as "$n/d \to 0$" and high-dimension, large-sample-size settings such as "$n/d \to c > 0$" or "$n/d \to \infty$ as $d \to \infty$".

The eigen-decomposition of $\boldsymbol{\Sigma}_W$ is given by $\boldsymbol{\Sigma}_W = \boldsymbol{U}_W \boldsymbol{\Lambda}_W \boldsymbol{U}_W^T$, where $\boldsymbol{\Lambda}_W$ is a diagonal matrix of eigenvalues, $\lambda_{1(W)} \geq \cdots \geq \lambda_{d(W)} (\geq 0)$, and $\boldsymbol{U}_W$ is an orthogonal matrix of the corresponding eigenvectors. Let $\boldsymbol{W} = \boldsymbol{U}_W \boldsymbol{\Lambda}_W^{1/2} \boldsymbol{Z}$. Then, $\boldsymbol{Z}$ is a $d \times n$ sphered data matrix from a distribution with the identity covariance matrix. Here, we write $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_d]^T$ and $\boldsymbol{z}_j = (z_{j1}, \ldots, z_{jn})^T,\ j = 1, \ldots, d$. Note that $E(z_{jk} z_{j'k}) = 0$ $(j \neq j')$ and $\text{Var}(\boldsymbol{z}_j) = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the n-dimensional identity matrix. We assume that the fourth moments of each variable in $\boldsymbol{Z}$ are uniformly bounded. The singular value decomposition of $\boldsymbol{A}$ is given by $\boldsymbol{A} = \sum_{j=1}^r \lambda_{j(A)}^{1/2} \boldsymbol{u}_{j(A)} \boldsymbol{v}_{j(A)}^T$, where $\lambda_{1(A)}^{1/2} \geq \cdots \geq \lambda_{r(A)}^{1/2}$ (> 0) are singular values of $\boldsymbol{A}$ and $\boldsymbol{u}_{j(A)}$ (or $\boldsymbol{v}_{j(A)}$) denotes a unit left- (or right-) singular vector corresponding to $\lambda_{j(A)}^{1/2}$ $(j = 1, \dots, r)$. Note that $\boldsymbol{\Sigma}_A = \sum_{j=1}^{r} \lambda_{j(A)} \boldsymbol{u}_{j(A)} \boldsymbol{u}_{j(A)}^T$. Also, note that $\lambda_{j(A)}$s depend not only on d but also on n. When $r \geq 2$, we assume that $\lambda_{j(A)}$s are distinct in the sense that

$$\liminf_{d\to\infty} \frac{\lambda_{j(A)}}{\lambda_{j'(A)}} > 1 \quad \text{when } n \text{ is fixed or } n \to \infty \text{ for all } j < j' \ (\leq r).$$

In this paper, we consider the problem of recovering the signal matrix $\boldsymbol{A}$ in high-dimensional settings such as $d \to \infty$ either when n is fixed or $n \to \infty$. In Section 2, we introduce the power spiked model to describe the structure of the eigenvalues of $\boldsymbol{\Sigma}$. In Section 3, we consider using the conventional PCA to recover $\boldsymbol{A}$ and show that the estimation of $\boldsymbol{A}$ holds consistency properties under severe conditions. In Section 4, we consider the noise reduction (NR) methodology by Yata and Aoshima [11] in (1.1) and apply it to recovering $\boldsymbol{A}$. We show that the estimation of $\boldsymbol{A}$ by the NR method holds the consistency properties under mild conditions and improves the error rate of the conventional PCA. In Section 5, we discuss the choice of unknown rank r by using the consistency properties. In Section 6, we give several simulation results to recover signal matrices. Finally, in Section 7, we give an application of (1.1) and demonstrate reconstruction procedures by using a microarray data set.

2. PCA consistency for the power spiked model

In this section, we assume $\boldsymbol{A} = \boldsymbol{O}_{d,n}$ in (1.1), where $\boldsymbol{O}_{d,n}$ is the $d \times n$ zero matrix. The sample covariance matrix is given by $\boldsymbol{S} = n^{-1}\boldsymbol{X}\boldsymbol{X}^T$. We consider the dual sample covariance matrix defined by $\boldsymbol{S}_D = n^{-1}\boldsymbol{X}^T\boldsymbol{X}$. Let $m = \min\{d, n\}$. Note that $\boldsymbol{S}_D$ and $\boldsymbol{S}$ share non-zero eigenvalues and $\text{rank}(\boldsymbol{S}) = \text{rank}(\boldsymbol{S}_D) \leq m$. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_m \geq 0$ be the eigenvalues of $\boldsymbol{S}_D$. The eigen-decompositions of $\boldsymbol{S}$ and $\boldsymbol{S}_D$ are given by $\boldsymbol{S} = \sum_{j=1}^{m} \hat{\lambda}_j \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j^T$ and $\boldsymbol{S}_D = \sum_{j=1}^{m} \hat{\lambda}_j \hat{\boldsymbol{v}}_j \hat{\boldsymbol{v}}_j^T$, where $\hat{\boldsymbol{u}}_j$ (or $\hat{\boldsymbol{v}}_j$) denotes a unit left- (or right-) singular vector of $\boldsymbol{X}/n^{1/2}$ corresponding to $\hat{\lambda}_j^{1/2}$. Note that $\hat{\boldsymbol{u}}_j$ can be calculated by $\hat{\boldsymbol{u}}_j = (n\hat{\lambda}_j)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{v}}_j$ from the fact that $\boldsymbol{X}/n^{1/2} = \sum_{j=1}^{m} \hat{\lambda}_j^{1/2} \hat{\boldsymbol{u}}_j \hat{\boldsymbol{v}}_j^T$.

Jung and Marron [3] and Yata and Aoshima [10] investigated consistency properties of the conventional PCA for HDLSS data. Yata and Aoshima [11] gave consistent estimators both of the eigenvalues and eigenvectors together with the principal component (PC) scores by a method called the noise-reduction methodology. Shen et al. [8] gave a consistent estimator of the first eigenvector under a sparsity assumption. Zhou and Marron [13] investigated consistency properties of some estimators for the first eigenvector in outlier contaminated data.

Now, we consider the power spiked model in $\boldsymbol{\Sigma}$. The eigen-decomposition of $\boldsymbol{\Sigma}$ is written as $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues, $\lambda_1 \geq \cdots \geq \lambda_d (\geq 0)$, and $\boldsymbol{U} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_d]$ is an orthogonal matrix of the corresponding eigenvectors. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{(1)} + \boldsymbol{\Sigma}_{(2)}$, where $\boldsymbol{\Sigma}_{(1)} = \sum_{i=1}^{r_0} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T$ and $\boldsymbol{\Sigma}_{(2)} = \sum_{i=r_0+1}^{d} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T$ with some unknown and positive fixed integer

r_0 ($< d$). Here, $\boldsymbol{\Sigma}_{(1)}$ is regarded as an intrinsic part and $\boldsymbol{\Sigma}_{(2)}$ is regarded as a noise part. Then, if there exists a positive fixed integer k_{r_0} such that

$$\lim_{d\to\infty} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^{k_{r_0}})}{\lambda_{r_0}^{k_{r_0}}} = 0, \tag{2.1}$$

the eigenvalues are called the *power spiked model*. See Section 2 in Yata and Aoshima [12] for the details. When $r_0 \geq 2$, we assume that $\liminf_{d\to\infty}(\lambda_j/\lambda_{j'}) > 1$ for all $j < j'$ ($\leq r_0$). They gave the following results.

Theorem 2.1 ([12])**.** *When* $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 = 0$ *for some* j ($\leq r_0$)*, it holds that as* $m \to \infty$

$$\frac{\hat{\lambda}_j}{\lambda_j} = 1 + o_p(1), \quad |\hat{\boldsymbol{u}}_j^T \boldsymbol{u}_j| = 1 + o_p(1) \quad \textit{and} \quad |\hat{\boldsymbol{v}}_j^T \boldsymbol{z}_j/n^{1/2}| = 1 + o_p(1) \tag{2.2}$$

under the conditions:

$$\frac{\sum_{s,t=r_0+1}^d \lambda_s\lambda_t E\{(z_{sk}^2-1)(z_{tk}^2-1)\}}{n\lambda_j^2} = o(1) \quad \textit{and} \quad \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})}{n\lambda_j} = o(1). \tag{2.3}$$

When $\limsup_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 > 0$ *for some* j ($\leq r_0$)*, (2.2) holds as* $m \to \infty$ *under the conditions in (2.3) and*

$$\frac{\sum_{p\neq q, s\neq t\geq r_0+1}^d \lambda_p\lambda_q\lambda_s\lambda_t\{E(z_{pk}z_{qk}z_{sk}z_{tk})\}^2}{n^2\lambda_j^4} = o(1) \quad \textit{and} \quad \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)^2}{n\lambda_j^4} = o(1).$$

Remark 2.1. A simple power spiked model is

$$\lambda_j = a_j d^{\alpha_j} \ (j = 1, \dots, r_0) \ \text{ and } \ \lambda_j = c_j \ (j = r_0+1, \dots, d),$$

where a_js, c_js and α_js are positive (fixed) constants. It should be noted that $\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 = 0$ when $\alpha_j > 1/2$ and $\limsup_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)/\lambda_j^2 > 0$ when $\alpha_j \leq 1/2$.

See [12] or Remark 3.1 for the details of Theorem 2.1. In (1.1), $\boldsymbol{\Sigma}_A$ is regarded as $\boldsymbol{\Sigma}_{(1)}$ and $\boldsymbol{\Sigma}_W$ is regarded as $\boldsymbol{\Sigma}_{(2)}$ in the power spiked model.

3. Reconstruction of the signal matrix by conventional PCA

In this section, we consider recovering the signal matrix $\boldsymbol{A}$ by using the conventional PCA in high-dimensional settings such as $d \to \infty$ either when n is fixed or $n \to \infty$. We reconstruct $\boldsymbol{A}$ by using $\hat{\lambda}_j$s, $\hat{\boldsymbol{u}}_j$s and $\hat{\boldsymbol{v}}_j$s. We assume $\hat{\boldsymbol{u}}_j^T \boldsymbol{u}_{j(A)} \geq 0$ and $\hat{\boldsymbol{v}}_j^T \boldsymbol{v}_{j(A)} \geq 0$ for all j ($\leq r$) without loss of generality.

We assume the power spiked model for (1.1) as follows: There exists a positive fixed integer k_r such that

$$\lim_{d\to\infty} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_W^{k_r})}{\lambda_{r(A)}^{k_r}} = 0 \quad \text{either when } n \text{ is fixed or } n \to \infty. \tag{3.1}$$

Under (3.1), it holds that

$$\frac{\lambda_{1(W)}}{\lambda_{r(A)}} = o(1),$$

so that $\lambda_{j(A)}$s are much larger than any eigenvalues of $\boldsymbol{\Sigma}_W$. We consider (3.1) for j $(\le r)$ either when n is fixed or $n \to \infty$ in the following two cases:

$$\textbf{(I)} \lim_{d\to\infty} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_W^2)}{\lambda_{j(A)}^2} = 0 \quad \text{and} \quad \textbf{(II)} \limsup_{d\to\infty} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_W^2)}{\lambda_{j(A)}^2} > 0.$$

We note that $\lambda_{j(A)}$ in (I) is larger than that in (II). See Remark 2.1 for the detail. Also, Murayama et al. [4] considered the estimation of $\boldsymbol{A}$ for a special case of (I). We consider the following conditions when $d \to \infty$ while n is fixed or $n \to \infty$:

$$\textbf{(C-i)} \ \frac{\sum_{s,t=1}^{d} \lambda_{s(W)}\lambda_{t(W)} E\{(z_{sk}^2-1)(z_{tk}^2-1)\}}{n\lambda_{j(A)}^2} = o(1);$$

$$\textbf{(C-ii)} \ \frac{\sum_{p\ne q, s\ne t}^{d} \ \lambda_{p(W)}\lambda_{q(W)}\lambda_{s(W)}\lambda_{t(W)}\{E(z_{pk}z_{qk}z_{sk}z_{tk})\}^2}{n^2\lambda_{j(A)}^4} = o(1);$$

$$\textbf{(C-iii)} \ \frac{\mathrm{tr}(\boldsymbol{\Sigma}_W^2)^2}{n\lambda_{j(A)}^4} = o(1); \quad \text{and} \quad \textbf{(C-iv)} \ \frac{\mathrm{tr}(\boldsymbol{\Sigma}_W)}{n\lambda_{j(A)}} = o(1).$$

Remark 3.1. We note that $z_{1k}, \dots, z_{dk}$ $(k = 1, \dots, n)$ are independent when $\boldsymbol{W}$ is Gaussian. Then, it holds that

$$\sum_{s,t=1}^{d} \lambda_{s(W)}\lambda_{t(W)} E\{(z_{sk}^2-1)(z_{tk}^2-1)\} = O\{\mathrm{tr}(\boldsymbol{\Sigma}_W^2)\} \ \text{ and} \tag{3.2}$$
$$\sum_{p\ne q, s\ne t}^{d} \lambda_{p(W)}\lambda_{q(W)}\lambda_{s(W)}\lambda_{t(W)}\{E(z_{pk}z_{qk}z_{sk}z_{tk})\}^2 = O\{\mathrm{tr}(\boldsymbol{\Sigma}_W^2)^2\},$$

so that (C-i) and (C-ii) hold under (C-iii) when $\boldsymbol{W}$ is Gaussian or $z_{1k}, \dots, z_{dk}$ $(k = 1, \dots, n)$ are independent.

Note that (C-iii) does not hold for (II) when n is fixed. If (3.2) holds, (C-i) is met even when n is fixed for j $(\le r)$ in (I). Let $\kappa_j = \mathrm{tr}(\boldsymbol{\Sigma}_W)/(n\lambda_{j(A)})$ for $j = 1, \dots, r$. We have the following results.

Theorem 3.1. *For j $(\le r)$, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\frac{\hat{\lambda}_j}{\lambda_{j(A)}} = 1 + \kappa_j + o_p(1), \quad \hat{\boldsymbol{u}}_j^T \boldsymbol{u}_{j(A)} = (1+\kappa_j)^{-1/2} + o_p(1)$$
$$\text{and} \quad \hat{\boldsymbol{v}}_j^T \boldsymbol{v}_{j(A)} = 1 + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Remark 3.2. If (3.2) holds, Theorem 3.1 is claimed even when n is fixed for j $(\le r)$ in (I).

Corollary 3.1. *For j $(\le r)$, under (C-i) and (C-iv) in (I) or under (C-i) to (C-iv) in (II), it holds that*

$$\frac{\hat{\lambda}_j}{\lambda_{j(A)}} = 1 + o_p(1) \quad \textit{and} \quad \hat{\boldsymbol{u}}_j^T \boldsymbol{u}_{j(A)} = 1 + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Note that $\hat{\boldsymbol{v}}_j$s hold the consistency property without (C-iv) contrary to $\hat{\lambda}_j$s and $\hat{\boldsymbol{u}}_j$s. Based on the theoretical background, we consider recovering the signal matrix $\boldsymbol{A}$ by $\widehat{\boldsymbol{A}}_r = \sum_{i=1}^r \hat{\lambda}_i^{1/2} \hat{\boldsymbol{u}}_i \hat{\boldsymbol{v}}_i^T$. In Section 5.1, we discuss the choice of r in $\widehat{\boldsymbol{A}}_r$. We define a loss function by

$$L(\widehat{\boldsymbol{A}}_r|\boldsymbol{A}) = ||\widehat{\boldsymbol{A}}_r - \boldsymbol{A}||_F^2,$$

where $||\cdot||_F$ denotes the Frobenius norm. Let $\psi = \text{tr}(\boldsymbol{\Sigma}_W)/n$. Then, we have the following results.

Theorem 3.2. *Under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$, it holds that*

$$L(\widehat{\boldsymbol{A}}_r|\boldsymbol{A}) = r\psi + o_p(\lambda_{r(A)})$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Remark 3.3. If (3.2) holds, Theorem 3.2 is claimed even when n is fixed under $\text{tr}(\boldsymbol{\Sigma}_W^2)/\lambda_{r(A)}^2 = o(1)$.

Corollary 3.2. *Under (C-i) and (C-iv) in (I) with $j = r$ or under (C-i) to (C-iv) in (II) with $j = r$, it holds that*

$$L(\widehat{\boldsymbol{A}}_r|\boldsymbol{A}) = o_p(\lambda_{r(A)})$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

From Theorem 3.2, if (C-iv) does not hold, the loss of $\widehat{\boldsymbol{A}}_r$ becomes $r\text{tr}(\boldsymbol{\Sigma}_W)/n$ asymptotically. In order to reduce the noise, we apply the NR method to recovering the signal matrix in Section 4.

4. Reconstruction of the signal matrix by NR method

We consider applying the *noise-reduction (NR) methodology* by Yata and Aoshima [11] to recover the signal matrix $\boldsymbol{A}$. By using the NR method, we obtain an estimator of $\lambda_{j(A)}$ as

$$\tilde{\lambda}_j = \hat{\lambda}_j - \frac{\text{tr}(\boldsymbol{S}_D) - \sum_{i=1}^j \hat{\lambda}_i}{n-j} \quad (j = 1, \dots, n-1). \tag{4.1}$$

Note that the second term in (4.1) is an estimator of ψ. See Lemma 5.1 in Section 5.1 for the details. Then, we have the following result.

Theorem 4.1. *For j $(\leq r)$, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\frac{\tilde{\lambda}_j}{\lambda_{j(A)}} = 1 + O_p\Big(\frac{\lambda_{j+1(A)}}{\lambda_{j(A)}n}\Big) + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$, where $\lambda_{r+1(A)} = 0$.

From Theorem 4.1, when $n \to \infty$ or $\lambda_{j+1(A)}/\lambda_{j(A)} = o(1)$, $\tilde{\lambda}_j$ holds the consistency property without (C-iv). Remember that $\hat{\lambda}_j$ requires (C-iv) to hold the consistency property.

Remark 4.1. For estimating eigenvalues, the NR method can improve the conventional PCA even when d is not sufficiently large (e.g. d is about 10). See Figure 1 in Ishii et al. [2] for example.

Now, we consider an adjustment of $\tilde{\lambda}_j$s as follows:

$$\tilde{\lambda}_{j(r)} = \hat{\lambda}_j - \frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^{r} \hat{\lambda}_i}{n-r} \quad (j = 1, \ldots, r). \tag{4.2}$$

Then, we have the following result.

Corollary 4.1. *For j $(\leq r)$, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\frac{\tilde{\lambda}_{j(r)}}{\lambda_{j(A)}} = 1 + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Remark 4.2. If (3.2) holds, Theorem 4.1 and Corollary 4.1 are claimed even when n is fixed for j $(\leq r)$ in (I).

We consider recovering $\boldsymbol{A}$ by $\widetilde{\boldsymbol{A}}_r = \sum_{i=1}^{r} \tilde{\lambda}_{i(r)}^{1/2} \hat{\boldsymbol{u}}_i \hat{\boldsymbol{v}}_i^T$. In Section 5.2, we discuss the choice of r in $\widetilde{\boldsymbol{A}}_r$. Let

$$\delta_i = \boldsymbol{u}_{i(A)}^T \boldsymbol{W} \boldsymbol{v}_{i(A)}/(n\lambda_{i(A)})^{1/2} \quad \text{for } i = 1, \ldots, r.$$

For the loss function by $L(\widetilde{\boldsymbol{A}}_r|\boldsymbol{A}) = ||\widetilde{\boldsymbol{A}}_r - \boldsymbol{A}||_F^2$, we have the following results.

Theorem 4.2. *Under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$, it holds that*

$$L(\widetilde{\boldsymbol{A}}_r|\boldsymbol{A}) = 2\sum_{i=1}^{r} \lambda_{i(A)}(1+\delta_i)\Big(1 - \frac{1+\delta_i}{(1+\kappa_i+2\delta_i)^{1/2}}\Big) + o_p(\lambda_{r(A)})$$
$$\text{and} \quad \delta_i = o_p\{(\lambda_{r(A)}/\lambda_{i(A)})^{1/2}\} \quad \text{for } i = 1, \ldots, r$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Remark 4.3. If (3.2) holds, Theorem 4.2 is claimed even when n is fixed under $\mathrm{tr}(\boldsymbol{\Sigma}_W^2)/\lambda_{r(A)}^2 = o(1)$.

Corollary 4.2. *Under (C-i) and (C-iv) in (I) with $j = r$ or under (C-i) to (C-iv) in (II) with $j = r$, it holds that*

$$L(\widetilde{\boldsymbol{A}}_r|\boldsymbol{A}) = o_p(\lambda_{r(A)})$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

From Theorems 3.2 and 4.2, we compare $2\lambda_{i(A)}\{1-1/(1+\kappa_i)^{1/2}\}$ with ψ $(= \lambda_{i(A)}\kappa_i)$ by noting $\delta_i = o_p(1)$. It holds that $2\{1-1/(1+\kappa_i)^{1/2}\} < \kappa_i$ $(i = 1,\dots,r)$ for any $\kappa_i > 0$, so that $L(\widetilde{\boldsymbol{A}}_r|\boldsymbol{A})$ is smaller than $L(\widehat{\boldsymbol{A}}_r|\boldsymbol{A})$ asymptotically. Thus, $\widetilde{\boldsymbol{A}}_r$ improves the loss of $\widehat{\boldsymbol{A}}_r$.

5. Choice of the rank r

In this section, we discuss the choice of r in $\widehat{\boldsymbol{A}}_r$ and $\widetilde{\boldsymbol{A}}_r$.

5.1. Choice of r in $\widehat{\boldsymbol{A}}_r$

Let r_* (> 0) be a candidate (fixed) integer for r, where $r_* < \min\{d, n\}$. We write that $\widehat{\boldsymbol{A}}_{r_*} = \sum_{i=1}^{r_*} \hat{\lambda}_i^{1/2} \hat{\boldsymbol{u}}_i \hat{\boldsymbol{v}}_i^T$. Then, we have the following result.

Proposition 5.1. *Under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$, it holds that*

$$L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A}) = \begin{cases} r_*\psi + \sum_{i=r_*+1}^{r} \lambda_{i(A)} + o_p(\lambda_{r(A)}) & \text{when } r_* < r; \\ r_*\psi + o_p(\lambda_{r(A)}) & \text{when } r_* \geq r \end{cases}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

From Proposition 5.1, it is not always true that $r_* = r$ gives the smallest $L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A})$. In fact, for a power spiked model such as $(\lambda_{1(A)}, \lambda_{2(A)}) = (d, d^{2/3})$, $r = 2$ and $\mathrm{tr}(\boldsymbol{\Sigma}_W) = d$, $L(\widehat{\boldsymbol{A}}_1|\boldsymbol{A})$ is smaller than $L(\widehat{\boldsymbol{A}}_2|\boldsymbol{A})$ as $d \to \infty$ when n is fixed. From Proposition 5.1, one may choose r_* as the first integer i $(= r_1$, say) satisfying $\psi > \lambda_{i+1(A)}$ (i.e. $\kappa_{i+1} > 1$). Then, $r_* = r_1$ gives the smallest $L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A})$ asymptotically for candidate integers. Note that $r_1 \leq r$.

Now, we consider estimating ψ by

$$\hat{\psi}_j = \frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^{j} \hat{\lambda}_i}{n-j} \quad \text{for } j = 1,\dots,n-1.$$

Then, we have the following result.

Lemma 5.1. *Under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\hat{\psi}_j = \begin{cases} \psi + \dfrac{\sum_{i=j+1}^{r} \lambda_{i(A)}}{n-j} + o_p(\lambda_{j(A)}) & \text{when } j < r; \\ \psi + o_p(\lambda_{r(A)}) & \text{when } j = r \end{cases}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Let $\acute{r}_1$ be the first integer i satisfying $\hat{\psi}_{i+1} > \tilde{\lambda}_{i+1}$. Then, from Lemma 5.1 we have the following result.

Proposition 5.2. *Assume*

$$\limsup_{d\to\infty} \kappa_{r_1+1}^{-1} < 1 \quad \textit{when } r_1 < r; \quad \liminf_{d\to\infty} \kappa_{r_1} > 0 \quad \textit{when } r_1 = r;$$
$$\textit{and} \quad \limsup_{d\to\infty} \kappa_{r_1} < 1 - 2\frac{r - r_1}{n - r_1}$$

either when n is fixed or $n \to \infty$. Then, it holds that $P(\acute{r}_1 = r_1) \to 1$ as $d \to \infty$ either when n is fixed or $n \to \infty$ under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$.

If (C-iv) with $j = r$ holds, $\hat{\psi}_i/\tilde{\lambda}_i$ becomes small for a large integer i, that is, $\acute{r}_1$ becomes a large integer as $\acute{r}_1 = O(n)$. Hence, if one has an upper bound for r_* as $r_* \le r_u$ with integer r_u $(< \infty)$, one may use $\acute{r}_{1u} = \min\{\acute{r}_1, r_u\}$ instead of $\acute{r}_1$.

5.2. Choice of r in $\widetilde{\boldsymbol{A}}_r$

We write that $\widetilde{\boldsymbol{A}}_{r_*} = \sum_{i=1}^{r_*} \tilde{\lambda}_{i(r_*)}^{1/2} \hat{\boldsymbol{u}}_i \hat{\boldsymbol{v}}_i^T$. Let $\gamma_i = 1 - 1/(1+\kappa_i)^{1/2}$ for $i = 1, \dots, r$. We have the following result.

Proposition 5.3. *Assume $\lambda_{r_*+1(A)}/(\lambda_{r_*(A)} n) = o(1)$ when $r_* < r$. Then, under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$, it holds that*

$$L(\widetilde{\boldsymbol{A}}_{r_*}|\boldsymbol{A}) = \begin{cases} \sum_{i=1}^{r_*} \lambda_{i(A)}\Big(2\gamma_i + o_p(1)\Big) + \sum_{i=r_*+1}^{r} \lambda_{i(A)} + o_p(\lambda_{r(A)}) & \textit{when } r_* < r; \\ \sum_{i=1}^{r} \lambda_{i(A)}\Big(2\gamma_i + o_p(1)\Big) + o_p(\lambda_{r(A)}) & \textit{when } r_* \ge r \end{cases}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

It holds that $\lambda_{r_*+1(A)}/(\lambda_{r_*(A)} n) = o(1)$ when $n \to \infty$ or $\lambda_{r_*+1(A)}/\lambda_{r_*(A)} = o(1)$. From Propositions 5.1 and 5.3, one may use $\widetilde{\boldsymbol{A}}_{r_*}$ with $r_* = \acute{r}_{1u}$ because $L(\widetilde{\boldsymbol{A}}_{r_*}|\boldsymbol{A})$ with $r_* = \acute{r}_{1u}$ is smaller than $L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A})$ with $r_* = \acute{r}_{1u}$ asymptotically.

On the other hand, similar to Section 5.1, from Proposition 5.3, one may choose r_* as the first integer i $(= r_2$, say) satisfying $2\gamma_{i+1} > 1$ (i.e. $\kappa_{i+1} > 3$). Note that $r_1 \le r_2 \le r$. Let $\acute{r}_2$ be the first integer i satisfying $\hat{\psi}_{i+1} > 3\tilde{\lambda}_{i+1}$. Note that $\acute{r}_1 \le \acute{r}_2$. Then, from Lemma 5.1, we have the following result.

Proposition 5.4. *Assume*

$$\limsup_{d\to\infty} \kappa_{r_2+1}^{-1} < 1/3 \quad \textit{when } r_2 < r; \quad \liminf_{d\to\infty} \kappa_{r_2} > 0 \quad \textit{when } r_2 = r;$$
$$\textit{and} \quad \limsup_{d\to\infty} \kappa_{r_2} < 3 - 4\frac{r - r_2}{n - r_2}$$

either when n is fixed or $n \to \infty$. Then, it holds that $P(\acute{r}_2 = r_2) \to 1$ as $d \to \infty$ either when n is fixed or $n \to \infty$ under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$.

Hence, one may use $\widetilde{\boldsymbol{A}}_{r_*}$ with $r_* = \acute{r}_2$. Here, one should note that $\acute{r}_2$ tends to be large if (C-iv) with $j = r$ holds. Also, note that $r_2 \to r$ if (C-iv) with $j = r$ holds. From Proposition 5.3, for the loss function, $\widetilde{\boldsymbol{A}}_{r_*}$ with $r_* > r$ is asymptotically equivalent to $\widetilde{\boldsymbol{A}}_r$. Hence, when $r = r_2$, one may choose a relatively large r_* in $\widetilde{\boldsymbol{A}}_{r_*}$ as $r_* > r$. On the other hand, from Proposition 5.1, the loss of $\widehat{\boldsymbol{A}}_{r_*}$ with $r_* > r$ is larger than that of $\widehat{\boldsymbol{A}}_r$ asymptotically, so that one should not choose a relatively large r_* in $\widehat{\boldsymbol{A}}_{r_*}$. Let $\acute{r}_{2u} = \min\{\acute{r}_2, r_u\}$, where r_u is given in Section 5.1. When $r = r_2$ and $r \le r_u$, $r_* = \acute{r}_{2u}$ gives the smallest $L(\widetilde{\boldsymbol{A}}_{r_*}|\boldsymbol{A})$ for candidate integers. Hence, for a relatively large r_u, we recommend to use $r_* = \acute{r}_{2u}$ instead of $\acute{r}_2$ in $\widetilde{\boldsymbol{A}}_{r_*}$.

6. Simulations

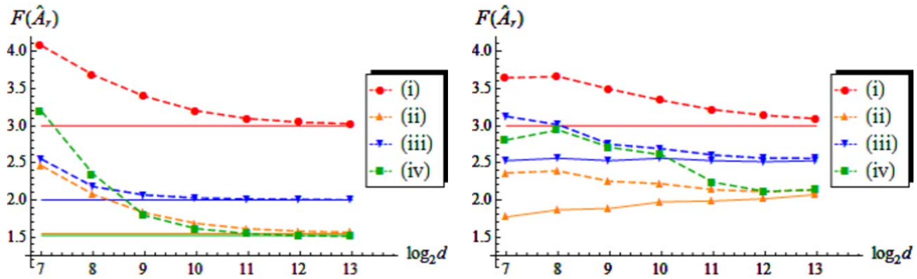
We used computer simulations to compare the performance of $\widetilde{\boldsymbol{A}}_{r_*}$ with $\widehat{\boldsymbol{A}}_{r_*}$. We set $r_u = 5$. We set $r_* = \acute{r}_{1u}$ for $\widehat{\boldsymbol{A}}_{r_*}$ and $r_* = \acute{r}_{2u}$ for $\widetilde{\boldsymbol{A}}_{r_*}$. See Section 5 for the details. We set $r = 3$, $\boldsymbol{\Sigma}_A = \mathrm{diag}(\lambda_{1(A)}, \lambda_{2(A)}, \lambda_{3(A)}, 0, \dots, 0)$ and $\boldsymbol{\Sigma}_W = (0.3^{|i-j|^{1/3}})$. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_W) = d$. We considered two cases:

(a) $\boldsymbol{w}_k$s are i.i.d. as d-variate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_W$, $(\lambda_{1(A)}, \lambda_{2(A)}, \lambda_{3(A)}) = (d/4, d/12, d/36)$, $d = 2^t$, $t = 7, \dots, 13$ and $n = 9$;

(b) $z_{sk} = (v_{sk} - 2)/2$ $(s = 1, \dots, d)$ in which v_{sk}s are i.i.d. as the chi-squared distribution with 2 degree of freedom, $(\lambda_{1(A)}, \lambda_{2(A)}, \lambda_{3(A)}) = (d^{3/4}, d^{2/3}, d^{1/2})$, $n = 3\lceil d^{1/2}/6 \rceil$ and $d = 2^t$, $t = 7, \dots, 13$, where $\lceil x \rceil$ denotes the smallest integer $\ge x$.

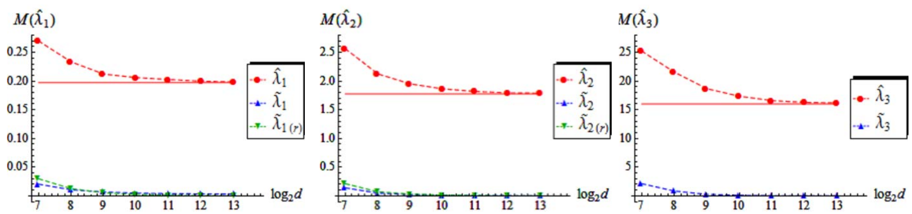
We considered the case when $d \to \infty$ while n is fixed in (a) and the case when $n \to \infty$ but $n/d \to 0$ in (b). Note that $(r_1, r_2) = (1, 2)$ in (a) and $(r_1, r_2) = (2, 3)$ in (b). From Remark 3.1, both in (a) and (b), (C-i) to (C-iii) with $j = r$ hold, but (C-iv) with $j = r$ does not hold.

Let $F(\boldsymbol{B}) = L(\boldsymbol{B}|\boldsymbol{A})/\psi$ for any $d \times n$ matrix, $\boldsymbol{B}$, and $M(b_j) = |b_j/\lambda_{j(A)} - 1|^2$ $(j = 1, \dots, r)$ for any constant, b_j. The findings were obtained by averaging the outcomes from 2000 independent replications. Figure 1 shows the averages of (i) $F(\widehat{\boldsymbol{A}}_r)$, (ii) $F(\widetilde{\boldsymbol{A}}_r)$, (iii) $F(\widehat{\boldsymbol{A}}_{r_*})$ with $r_* = \acute{r}_{1u}$ and (iv) $F(\widetilde{\boldsymbol{A}}_{r_*})$ with $r_* = \acute{r}_{2u}$ in (a) and (b). The dashed lines denote the simulation results. We gave the corresponding theoretical values by (i) r, (ii) $\sum_{i=1}^{r} 2\lambda_{i(A)}\gamma_i/\psi$, (iii) $r_1 + \sum_{i=r_1+1}^{r} \lambda_{i(A)}/\psi$ and (iv) $(\sum_{i=1}^{r_2} 2\lambda_{i(A)}\gamma_i + \sum_{i=r_2+1}^{r} \lambda_{i(A)})/\psi$, which were denoted by the solid lines in (a) and (b). See Theorems 3.2, 4.2, Propositions 5.1 and 5.3 for the details. The theoretical value by (iv) was not described for (b) because it is same as that of (iii). We also calculated the variances of simulation results by the 2000 replications. The variances of (i) to (iv) in (a) and (b) were quite small especially when d is large. For example, when $d = 2^t$ for $t \ge 11$,

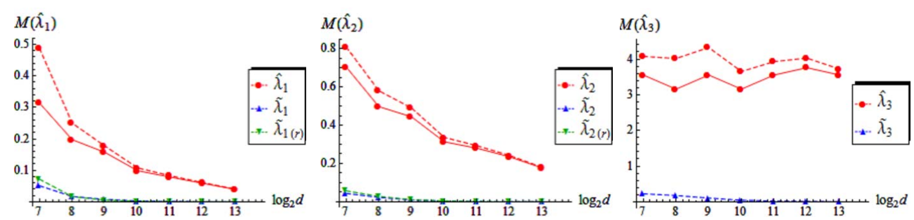


(a) $d \to \infty$ while n is fixed (b) $d \to \infty$ and $n \to \infty$, but $n/d \to 0$

FIG 1. *The averages of (i) $F(\widehat{\boldsymbol{A}}_r)$, (ii) $F(\widetilde{\boldsymbol{A}}_r)$, (iii) $F(\widehat{\boldsymbol{A}}_{r_*})$ with $r_* = \acute{r}_{1u}$ and (iv) $F(\widetilde{\boldsymbol{A}}_{r_*})$ with $r_* = \acute{r}_{2u}$ which are denoted by the dashed lines. The corresponding theoretical values are denoted by the solid lines. For (b), the theoretical value of (iv) was not described because it is same as that of (iii).*



(a) $d \to \infty$ while n is fixed



(b) $d \to \infty$ and $n \to \infty$, but $n/d \to 0$

FIG 2. *The averages of $M(\hat{\lambda}_j)$, $M(\tilde{\lambda}_j)$ and $M(\tilde{\lambda}_{j(r)})$ which are denoted by the dashed lines. The corresponding theoretical values are denoted by the solid lines. For the right panels, $M(\tilde{\lambda}_{3(r)})$ was not described because $\tilde{\lambda}_3 = \tilde{\lambda}_{3(r)}$.*

all the variances in (a) were smaller than 0.006. Figure 2 shows the averages of $M(\hat{\lambda}_j)$, $M(\tilde{\lambda}_j)$ and $M(\tilde{\lambda}_{j(r)})$ in (a) and (b). The dashed lines denote the simulation results. For $j = 3$ both in (a) and (b), the average of $M(\tilde{\lambda}_{j(r)})$ was not described because $\tilde{\lambda}_3$ is same as $\tilde{\lambda}_{3(r)}$. Note that the average of $M(\hat{\lambda}_j)$ is an estimated value of the mean square error (MSE), $E(|\hat{\lambda}_j/\lambda_{j(A)} - 1|^2)$. The averages of $M(\tilde{\lambda}_j)$ and $M(\tilde{\lambda}_{j(r)})$ are also the same as in $M(\hat{\lambda}_j)$. From Theorem 3.1, we gave the corresponding theoretical value, κ_j^2, for the MSE of $M(\hat{\lambda}_j)$. The theoretical values were denoted by the solid lines.

The simulation results appeared close to the theoretical values and it seemed to be good approximations when d is large. As expected theoretically, we observed that $\widetilde{\boldsymbol{A}}_r$ and $\widetilde{\boldsymbol{A}}_{r_*}$ with $r_* = \acute{r}_{2u}$ give more preferable performances compared to $\widehat{\boldsymbol{A}}_r$ and $\widehat{\boldsymbol{A}}_{r_*}$ with $r_* = \acute{r}_{1u}$ even when n is fixed. The main reason must be due to κ_j which is the bias of $\hat{\lambda}_j$. See Sections 4 and 5.2 for the details. In fact, from Figure 2, the MSE of $\hat{\lambda}_j$s were quite large especially when n is small because κ_js are large for the HDLSS settings. In contrast, the estimators by the NR method gave excellent performances even when n is small. For the estimation of $\boldsymbol{A}$ by the conventional PCA, $\widehat{\boldsymbol{A}}_{r_*}$ with $r_* = \acute{r}_{1u}$ gave a better performance compared to $\widehat{\boldsymbol{A}}_r$ because $r_1 < r$. See Section 5.1 for the details.

7. Example

In this section, we consider an application of (1.1) to a mixture model. We demonstrate the reconstruction procedures for the mixture model by using a microarray data set.

7.1. Application

We suppose that there are l classes, $\Pi_i,\ i = 1,\dots,l$, each having unknown mean vector, $\boldsymbol{\mu}_i$. We assume that an observation is sampled from one of Π_is and the label of the class is missing. Let $n_i = \#\{j|\boldsymbol{x}_j \in \Pi_i \text{ for } j = 1,\dots,n\}$ for $i = 1,\dots,k$, where $\#S$ denotes the number of elements in a set S. We define that $\boldsymbol{\mu}_{(j)} = \boldsymbol{\mu}_i$ according to $\boldsymbol{x}_j \in \Pi_i$ for $j = 1,\dots,n$. We consider the following mixture model.

$$\boldsymbol{x}_j = \boldsymbol{\mu}_{(j)} + \boldsymbol{w}_j \quad \text{for } j = 1,\dots,n. \tag{7.1}$$

Then, we can write that

$$\boldsymbol{A} = [\boldsymbol{\mu}_{(1)},\dots,\boldsymbol{\mu}_{(n)}]/n^{1/2}.$$

Note that $\sum_{i=1}^{r} \lambda_{i(A)} = ||\boldsymbol{A}||_F^2 = \sum_{i=1}^{l} (n_i/n)||\boldsymbol{\mu}_i||^2$, where $||\cdot||$ denotes the Euclidean norm. If $\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_l$ are linearly independent and $n_i > 0$ for all i, the rank of $\boldsymbol{A}$ becomes just l (i.e., $r = l$). Also, it is likely that $\lambda_{r(A)} \to \infty$ as $d \to \infty$ if $||\boldsymbol{\mu}_i|| \to \infty$ as $d \to \infty$ for all i.

7.2. Demonstration

We analyzed gene expression data by Bhattacharjee et al. [1] in which the data set consisted of five lung carcinomas types having 3312 genes ($d = 3312$). The data set is given in Yang et al. [9]. See [1] and [9] for details of the data set. We used four classes as Π_1 : adenocarcinomas (139 samples), Π_2 : normal lung (17 samples), Π_3 : squamous cell lung carcinomas (21 samples) and Π_4 : pulmonary carcinoids (20 samples). We consider the cases when $r_* = 1,\dots,7$ and $l = 1,\dots,4$. Here, from Section 7.1, l can be regarded as the rank, r. Note that

TABLE 1

Values of $(L(\widehat{\boldsymbol{A}}_{r_*}|\breve{\boldsymbol{A}}), L(\widetilde{\boldsymbol{A}}_{r_*}|\breve{\boldsymbol{A}}))$ *as estimates of* $(L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A}), L(\widetilde{\boldsymbol{A}}_{r_*}|\boldsymbol{A}))$ *for the microarray data set given by [1]. For each* l*, the values of* $(\acute{r}_{1u}, \acute{r}_{2u})$ *are also given in the bottom line. When* $r_* > n$ *for* $\widehat{\boldsymbol{A}}_{r_*}$ *or* $r_* > n-1$ *for* $\widetilde{\boldsymbol{A}}_{r_*}$*, those values were not available because* $\hat{\lambda}_j$ *or* $\tilde{\lambda}_j$ *is not available when* $j > n$ *or* $j > n-1$*.*

(a) $n_1 = \cdots = n_l = 5$

$r_*\backslash l$	1	2	3	4
1	(51.08, 51.12)	(71.26, 71.24)	(96.22, 96.23)	(167.99, 167.95)
2	(106.08, 79.63)	(53.11, 48.12)	(50.27, 48.1)	(88.97, 87.8)
3	(135.68, 85.94)	(81.06, 64.67)	(54.81, 47.64)	(55.35, 52.71)
4	(162.74, 94.51)	(99.24, 73.86)	(74.68, 58.78)	(62.61, 55.65)
5	(185.46, N/A)	(113.67, 79.75)	(91.44, 68.32)	(78.18, 65.01)
6	(N/A, N/A)	(127.08, 86.55)	(104.89, 75.24)	(92.13, 73.09)
7	(N/A, N/A)	(138.69, 93.16)	(116.21, 80.87)	(104.07, 80.15)
$(\acute{r}_{1u}, \acute{r}_{2u})$	(2, 2)	(3, 5)	(4, 5)	(5, 5)

(b) $n_1 = \cdots = n_l = 10$

$r_*\backslash l$	1	2	3	4
1	(41.16, 41.09)	(76.88, 76.82)	(106.79, 106.75)	(178.17,178.13)
2	(74.66, 60.23)	(45.84, 44.15)	(53.52, 52.98)	(99.81, 99.17)
3	(98.76, 72.74)	(63.74, 55.89)	(51.16, 48.51)	(60.9, 59.94)
4	(115.46, 78.36)	(78.11, 64.98)	(64.39, 57.19)	(60.55, 57.88)
5	(130.09, 82.46)	(89.18, 71.14)	(76.87, 65.62)	(71.66, 65.69)
6	(143.66, 86.21)	(99.01, 76.84)	(85.99, 71.1)	(82.5, 73.44)
7	(156.49, 90.28)	(107.43, 81.11)	(94.31, 76.04)	(90.81, 78.86)
$(\acute{r}_{1u}, \acute{r}_{2u})$	(2, 4)	(4, 5)	(5, 5)	(5, 5)

$\boldsymbol{\mu}_{(j)} = \boldsymbol{\mu}_1$ for all j in (7.1) when $l = 1$. We considered two cases: (a) $n_1 = \cdots = n_l = 5$ and (b) $n_1 = \cdots = n_l = 10$. Note that $n = 5l$ in (a) and $n = 10l$ in (b). We set $r_u = 5$.

For each r_* and l, we constructed $\widehat{\boldsymbol{A}}_{r_*}$ and $\widetilde{\boldsymbol{A}}_{r_*}$ by using the first 5 samples in (a) or 10 samples in (b) from each class. We investigated their accuracies by using the remaining samples of each class as a test data set. We defined that $\breve{\boldsymbol{\mu}}_{(j)} = \breve{\boldsymbol{\mu}}_i$ according to $\boldsymbol{x}_j \in \Pi_i$ for $j = 1, \dots, n$, where $\breve{\boldsymbol{\mu}}_i$ is the sample mean vector of the test data set for each i. For each r_* and l, we constructed $\breve{\boldsymbol{A}} = [\breve{\boldsymbol{\mu}}_{(1)}, \dots, \breve{\boldsymbol{\mu}}_{(n)}]/n^{1/2}$ as an estimator of $\boldsymbol{A}$ when the labels of the data set are known. Hence, $L(\widehat{\boldsymbol{A}}_{r_*}|\breve{\boldsymbol{A}}) = ||\widehat{\boldsymbol{A}}_{r_*} - \breve{\boldsymbol{A}}||_F^2$ and $L(\widetilde{\boldsymbol{A}}_{r_*}|\breve{\boldsymbol{A}}) = ||\widetilde{\boldsymbol{A}}_{r_*} - \breve{\boldsymbol{A}}||_F^2$ can be regarded as estimators of $L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A})$ and $L(\widetilde{\boldsymbol{A}}_{r_*}|\boldsymbol{A})$. We gave the values of $(L(\widehat{\boldsymbol{A}}_{r_*}|\breve{\boldsymbol{A}}),\ L(\widetilde{\boldsymbol{A}}_{r_*}|\breve{\boldsymbol{A}}))$ for $r_* = 1, \dots, 7$ and $l = 1, \dots, 4$ in Table 1. We also gave the values of $(\acute{r}_{1u}, \acute{r}_{2u})$ for each l.

As expected theoretically, we observed that $\widetilde{\boldsymbol{A}}_{r_*}$ gave more preferable performances than $\widehat{\boldsymbol{A}}_{r_*}$ for most cases of (r_*, l) in (a) and (b). Also, for each r_* and l, when $r_* \ge l\ (= r)$, most values in (b) are smaller than those in (a). This is probably because the sample size in (b) is larger than that in (a). See Propositions 5.1 and 5.3 for the details. On the other hand, we observed that $\acute{r}_{1u}$ and $\acute{r}_{2u}$ were larger than r. However, $\widetilde{\boldsymbol{A}}_{r_*}$ gave adequate performances even when $r_* > r$. See Section 5.2 for the theoretical reason. Hence, we recommend to use $\widetilde{\boldsymbol{A}}_{r_*}$ with $r_* = \acute{r}_{1u}$ or $\acute{r}_{2u}$.

Appendix A

In this section, we give several lemmas, proofs of the lemmas, and proof of Lemma 5.1.

Throughout, let $\boldsymbol{e}_{in} = (e_{i1}, \ldots, e_{in})^T$, $i = 1, 2$, be arbitrary unit random vectors.

Lemma A.1. *For j ($\le r$), under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\boldsymbol{e}_{1n}^T \frac{\boldsymbol{W}^T \boldsymbol{W}}{n\lambda_{j(A)}} \boldsymbol{e}_{2n} = \kappa_j \boldsymbol{e}_{1n}^T \boldsymbol{e}_{2n} + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Proof. We write that

$$\boldsymbol{e}_{1n}^T \frac{\boldsymbol{W}^T \boldsymbol{W}}{n} \boldsymbol{e}_{2n} = \boldsymbol{e}_{1n}^T \sum_{s=1}^{d} \frac{\lambda_{s(W)} \boldsymbol{z}_s \boldsymbol{z}_s^T}{n} \boldsymbol{e}_{2n} = \boldsymbol{e}_{1n}^T \sum_{s=1}^{d} \frac{\lambda_{s(W)} (\boldsymbol{z}_s \boldsymbol{z}_s^T - \boldsymbol{I}_n)}{n} \boldsymbol{e}_{2n} + \psi \boldsymbol{e}_{1n}^T \boldsymbol{e}_{2n}.$$

When $n \to \infty$, from Lemma 5 given in Yata and Aoshima [12], it holds that as $d \to \infty$

$$\boldsymbol{e}_{1n}^T \frac{\sum_{s=1}^{d} \lambda_{s(W)} (\boldsymbol{z}_s \boldsymbol{z}_s^T - \boldsymbol{I}_n)}{n\lambda_{j(A)}} \boldsymbol{e}_{2n} = o_p(1) \tag{A.1}$$

for j ($\le r$) under (C-i) in (I) or under (C-i) to (C-iii) in (II). On the other hand, when n is fixed, (C-iii) does not hold in (II). Hence, we consider only the case of (I) when n is fixed. By using Markov's inequality, for any $\tau > 0$ and j ($\le r$), under (C-i) in (I), we have that

$$\sum_{k=1}^{n} P\Big\{\Big(\sum_{s=1}^{d} \frac{\lambda_{s(W)}(z_{sk}^2 - 1)}{n\lambda_{j(A)}}\Big)^2 \ge \tau\Big\} \le \sum_{s,t=1}^{d} \frac{\lambda_{s(W)}\lambda_{t(W)} E\{(z_{sk}^2 - 1)(z_{tk}^2 - 1)\}}{\tau n \lambda_{j(A)}^2} = o(1)$$

$$\text{and} \quad \sum_{k \ne k'}^{n} P\Big\{\Big(\sum_{s=1}^{d} \frac{\lambda_{s(W)} z_{sk} z_{sk'}}{n\lambda_{j(A)}}\Big)^2 \ge \tau\Big\} \le \frac{\mathrm{tr}(\boldsymbol{\Sigma}_W^2)}{\tau \lambda_{j(A)}^2} = o(1)$$

as $d \to \infty$ when n is fixed, so that (A.1) holds in (I) when n is fixed. Thus it concludes the result. □

Lemma A.2. *It holds that under (3.1)*

$$\frac{\boldsymbol{u}_{i(A)}^T \boldsymbol{W} \boldsymbol{e}_{1n}}{n^{1/2}} = o_p(\lambda_{r(A)}^{1/2}), \ i = 1, \ldots, r$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Proof. We write that $\boldsymbol{u}_{i(A)}^T \boldsymbol{W}\boldsymbol{e}_{1n} = \sum_{k=1}^n e_{1k}\boldsymbol{w}_k^T \boldsymbol{u}_{i(A)}$. Note that $\lambda_{1(W)} = o(\lambda_{r(A)})$ under (3.1). Also, note that $\boldsymbol{u}_{i(A)}^T \boldsymbol{\Sigma}_W \boldsymbol{u}_{i(A)} \le \lambda_{1(W)}$ for $i = 1, \dots, r$. By using Markov's inequality, for any $\tau > 0$ and $i = 1, \dots, r$, under (3.1), we have that

$$P\Big(\sum_{k=1}^n (\boldsymbol{w}_k^T \boldsymbol{u}_{i(A)})^2/n \ge \tau\lambda_{r(A)}\Big) \le \frac{E\{\sum_{k=1}^n (\boldsymbol{w}_k^T \boldsymbol{u}_{i(A)})^2\}}{\tau n \lambda_{r(A)}} = \frac{\boldsymbol{u}_{i(A)}^T \boldsymbol{\Sigma}_W \boldsymbol{u}_{i(A)}}{\tau\lambda_{r(A)}} \le \frac{\lambda_{1(W)}}{\tau\lambda_{r(A)}} = o(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$, so that $\sum_{k=1}^n (\boldsymbol{w}_k^T \boldsymbol{u}_{i(A)})^2/n = o_p(\lambda_{r(A)})$. Then, by noting that

$$\Big|\sum_{k=1}^n e_{1k}(\boldsymbol{w}_k^T \boldsymbol{u}_{i(A)})/n^{1/2}\Big| \le \Big\{\sum_{k=1}^n e_{1k}^2\Big\}^{1/2}\Big\{\sum_{k=1}^n (\boldsymbol{w}_k^T \boldsymbol{u}_{i(A)})^2/n\Big\}^{1/2} = \Big\{\sum_{k=1}^n (\boldsymbol{w}_k^T \boldsymbol{u}_{i(A)})^2/n\Big\}^{1/2},$$

we can conclude the result. □

Lemma A.3. *For j $(\le r)$, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\frac{\hat{\lambda}_j}{\lambda_{j(A)}} = 1 + \kappa_j + o_p(1) \quad \textit{and} \quad \hat{\boldsymbol{v}}_j^T \boldsymbol{v}_{j(A)} = 1 + o_p(1);$$
$$\boldsymbol{v}_{i'(A)}^T \hat{\boldsymbol{v}}_i = o_p\{(\lambda_{j(A)}/\lambda_{i'(A)})^{1/2}\} \quad \textit{for } i' < j \le i; \quad \textit{and}$$
$$\hat{\boldsymbol{v}}_i^T \boldsymbol{A}^T \boldsymbol{W} \hat{\boldsymbol{v}}_{i'}/n^{1/2} = o_p\{(\lambda_{r(A)}\lambda_{j(A)})^{1/2}\} \quad \textit{for } i \ge j;\ i' = 1, \dots, r$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Proof. We write that

$$\frac{\hat{\lambda}_j}{\lambda_{j(A)}} = \hat{\boldsymbol{v}}_j^T \frac{\boldsymbol{S}_D}{\lambda_{j(A)}}\hat{\boldsymbol{v}}_j = \hat{\boldsymbol{v}}_j^T \frac{(\boldsymbol{A} + \boldsymbol{W}/n^{1/2})^T(\boldsymbol{A} + \boldsymbol{W}/n^{1/2})}{\lambda_{j(A)}}\hat{\boldsymbol{v}}_j = \hat{\boldsymbol{v}}_j^T \frac{\boldsymbol{A}^T\boldsymbol{A}}{\lambda_{j(A)}}\hat{\boldsymbol{v}}_j + 2\hat{\boldsymbol{v}}_j^T \frac{\boldsymbol{A}^T\boldsymbol{W}}{n^{1/2}\lambda_{j(A)}}\hat{\boldsymbol{v}}_j + \hat{\boldsymbol{v}}_j^T \frac{\boldsymbol{W}^T\boldsymbol{W}}{n\lambda_{j(A)}}\hat{\boldsymbol{v}}_j. \tag{A.2}$$

When $j = 1$, we note that $|\hat{\boldsymbol{v}}_i^T \boldsymbol{A}^T \boldsymbol{W}\hat{\boldsymbol{v}}_{i'}| \le \sum_{s=1}^r \lambda_{s(A)}^{1/2} |\boldsymbol{u}_{s(A)}^T \boldsymbol{W}\hat{\boldsymbol{v}}_{i'}|$ for all i, i'. From Lemma A.2, under (3.1), it holds that for all i, i'

$$\hat{\boldsymbol{v}}_i^T \boldsymbol{A}^T \boldsymbol{W}\hat{\boldsymbol{v}}_{i'}/n^{1/2} = o_p\{(\lambda_{r(A)}\lambda_{1(A)})^{1/2}\} \tag{A.3}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. By combining (A.2) with Lemma A.1 and (A.3), it holds that

$$\frac{\hat{\lambda}_1}{\lambda_{1(A)}} = \hat{\boldsymbol{v}}_1^T \frac{\sum_{s=1}^r \lambda_{s(A)}\boldsymbol{v}_{s(A)}\boldsymbol{v}_{s(A)}^T}{\lambda_{1(A)}}\hat{\boldsymbol{v}}_1 + \kappa_1 + o_p(1) = 1 + \kappa_1 + o_p(1)$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Thus, we have that $\hat{\boldsymbol{v}}_1^T\boldsymbol{v}_{1(A)} = 1+o_p(1)$ from the assumption that $\lambda_{j(A)}$s are distinct.

When $j=2$, we note that $\hat{\boldsymbol{v}}_1^T\boldsymbol{v}_{i(A)} = o_p(1)$ for $i=2,\dots,r$, because $\hat{\boldsymbol{v}}_1^T\boldsymbol{v}_{1(A)} = 1+o_p(1)$. From Lemma A.1 and (A.3), we have that for $i\ge 2$

$$\begin{aligned}0 = \hat{\boldsymbol{v}}_1^T\frac{\boldsymbol{S}_D}{\lambda_2}\hat{\boldsymbol{v}}_i &= \frac{\lambda_{1(A)}}{\lambda_{2(A)}}\{1+o_p(1)\}\boldsymbol{v}_{1(A)}^T\hat{\boldsymbol{v}}_i + \hat{\boldsymbol{v}}_1^T\frac{\boldsymbol{A}^T\boldsymbol{W}+\boldsymbol{W}^T\boldsymbol{A}}{n^{1/2}\lambda_{2(A)}}\hat{\boldsymbol{v}}_i + o_p(1)\\ &= \frac{\lambda_{1(A)}}{\lambda_{2(A)}}\{1+o_p(1)\}\boldsymbol{v}_{1(A)}^T\hat{\boldsymbol{v}}_i + o_p\{(\lambda_{1(A)}/\lambda_{2(A)})^{1/2}\} + o_p(1)\end{aligned}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Thus, it follows that for $i\ge 2$

$$\boldsymbol{v}_{1(A)}^T\hat{\boldsymbol{v}}_i = o_p\{(\lambda_{2(A)}/\lambda_{1(A)})^{1/2}\}. \tag{A.4}$$

It holds that for $i\ge 2$

$$\boldsymbol{A}\hat{\boldsymbol{v}}_i = \sum_{j=1}^r \lambda_{j(A)}^{1/2}\boldsymbol{u}_{j(A)}\boldsymbol{v}_{j(A)}^T\hat{\boldsymbol{v}}_i = \sum_{j=2}^r \lambda_{j(A)}^{1/2}\boldsymbol{u}_{j(A)}\boldsymbol{v}_{j(A)}^T\hat{\boldsymbol{v}}_i + \lambda_{2(A)}^{1/2}\boldsymbol{u}_{1(A)}\times o_p(1),$$

so that $|\hat{\boldsymbol{v}}_i^T\boldsymbol{A}^T\boldsymbol{W}\hat{\boldsymbol{v}}_{i'}| \le \sum_{j=2}^r\lambda_{j(A)}^{1/2}|\boldsymbol{u}_{j(A)}^T\boldsymbol{W}\hat{\boldsymbol{v}}_{i'}| + \lambda_{2(A)}^{1/2}|\boldsymbol{u}_{1(A)}^T\boldsymbol{W}\hat{\boldsymbol{v}}_{i'}\times o_p(1)|$ for $i'=1,\dots,r$. Hence, from Lemma A.2 it holds that

$$\hat{\boldsymbol{v}}_i^T\boldsymbol{A}^T\boldsymbol{W}\hat{\boldsymbol{v}}_{i'}/n^{1/2} = o_p\{(\lambda_{r(A)}\lambda_{2(A)})^{1/2}\}\ \text{ for } i\ge 2;\ i'=1,\dots,r \tag{A.5}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). By combining (A.2) with Lemma A.1, (A.4) and (A.5), we have that

$$\begin{aligned}\frac{\hat{\lambda}_2}{\lambda_{2(A)}} &= \hat{\boldsymbol{v}}_2^T\frac{\sum_{s=1}^r\lambda_{s(A)}\boldsymbol{v}_{s(A)}\boldsymbol{v}_{s(A)}^T}{\lambda_{2(A)}}\hat{\boldsymbol{v}}_2 + \kappa_2 + o_p(1)\\ &= \hat{\boldsymbol{v}}_2^T\frac{\sum_{s=2}^r\lambda_{s(A)}\boldsymbol{v}_{s(A)}\boldsymbol{v}_{s(A)}^T}{\lambda_{2(A)}}\hat{\boldsymbol{v}}_2 + \kappa_2 + o_p(1) = 1+\kappa_2+o_p(1)\end{aligned} \tag{A.6}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Thus, we have that $\hat{\boldsymbol{v}}_2^T\boldsymbol{v}_{2(A)} = 1+o_p(1)$.

When $j=3$, we note that $\hat{\boldsymbol{v}}_{i'}^T\boldsymbol{v}_{i(A)} = o_p(1)$ for $i=i'+1,\dots,r;\ i'=1,2$, because $\hat{\boldsymbol{v}}_{i'}^T\boldsymbol{v}_{i'(A)} = 1+o_p(1)$, $i'=1,2$. From Lemma A.1, (A.3), (A.4) and (A.5), we have that for $i\ge 3$

$$\begin{aligned}0 = \hat{\boldsymbol{v}}_1^T\frac{\boldsymbol{S}_D}{\lambda_3}\hat{\boldsymbol{v}}_i =& \frac{\lambda_{1(A)}}{\lambda_{3(A)}}\{1+o_p(1)\}\boldsymbol{v}_{1(A)}^T\hat{\boldsymbol{v}}_i + \boldsymbol{v}_{2(A)}^T\hat{\boldsymbol{v}}_i\times o_p\{(\lambda_{2(A)}/\lambda_{3(A)})\}\\ &+ o_p\{(\lambda_{1(A)}/\lambda_{3(A)})^{1/2}\} + o_p(1);\\ 0 = \hat{\boldsymbol{v}}_2^T\frac{\boldsymbol{S}_D}{\lambda_3}\hat{\boldsymbol{v}}_i =& \frac{\lambda_{2(A)}}{\lambda_{3(A)}}\{1+o_p(1)\}\boldsymbol{v}_{2(A)}^T\hat{\boldsymbol{v}}_i + \boldsymbol{v}_{1(A)}^T\hat{\boldsymbol{v}}_i\times o_p\{(\lambda_{1(A)}^{1/2}\lambda_{2(A)}^{1/2}/\lambda_{3(A)})\}\\ &+ o_p\{(\lambda_{2(A)}/\lambda_{3(A)})^{1/2}\} + o_p(1)\end{aligned}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Thus, it follows that $\boldsymbol{v}_{i'(A)}^T \hat{\boldsymbol{v}}_i = o_p\{(\lambda_{3(A)}/\lambda_{i'(A)})^{1/2}\}$ for $i' = 1, 2;\ i \geq 3$. Similar to (A.5) and (A.6), it holds that $\hat{\lambda}_3/\lambda_{3(A)} = 1 + \kappa_3 + o_p(1)$ and

$$\hat{\boldsymbol{v}}_i^T \boldsymbol{A}^T \boldsymbol{W} \hat{\boldsymbol{v}}_{i'}/n^{1/2} = o_p\{(\lambda_{r(A)}\lambda_{3(A)})^{1/2}\} \text{ for } i \geq 3;\ i' = 1, \dots, r$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Hence, we have that $\hat{\boldsymbol{v}}_3^T \boldsymbol{v}_{3(A)} = 1 + o_p(1)$. In a way similar to the case of $\lambda_{3(A)}$, we have the results for $j = 4, \dots, r$, as well. It concludes the results. □

Lemma A.4. *For j $(\leq r)$, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\hat{\boldsymbol{u}}_j^T \boldsymbol{u}_{j(A)} = (1 + \kappa_j)^{-1/2} + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Proof. Note that $\hat{\boldsymbol{u}}_j = (n\hat{\lambda}_j)^{-1/2} \boldsymbol{X} \hat{\boldsymbol{v}}_j$. From Lemmas A.2 and A.3, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that for j $(\leq r)$

$$\boldsymbol{u}_{j(A)}^T \hat{\boldsymbol{u}}_j = \hat{\lambda}_j^{-1/2} \lambda_{j(A)}^{1/2} \boldsymbol{v}_{j(A)}^T \hat{\boldsymbol{v}}_j + (n\hat{\lambda}_j)^{-1/2} \boldsymbol{u}_{j(A)}^T \boldsymbol{W} \hat{\boldsymbol{v}}_j = (1 + \kappa_j)^{-1/2} + o_p(1)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. It concludes the result. □

Lemma A.5. *For j $(\leq r)$, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\begin{aligned} &\frac{\hat{\lambda}_i}{\lambda_{i(A)}} = 1 + \kappa_i + 2\delta_i + o_p\Big\{\Big(\frac{\lambda_{j(A)}}{\lambda_{i(A)}}\Big)\Big\} \\ &\textit{and} \quad \hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i(A)} = 1 + o_p\Big(\frac{\lambda_{j(A)}}{\lambda_{i(A)}}\Big) \quad \textit{for } i = 1, \dots, j; \\ &\hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i'(A)} = o_p\Big\{\Big(\frac{\lambda_{j(A)}}{\max\{\lambda_{i(A)}, \lambda_{i'(A)}\}}\Big)^{1/2}\Big\} \quad \textit{for all } i \neq i' \ (\leq j) \end{aligned}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Proof. From Lemma A.1, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that for j $(\leq r)$

$$\boldsymbol{e}_{1n}^T (\boldsymbol{W}^T \boldsymbol{W}/n) \boldsymbol{e}_{2n} = \boldsymbol{e}_{1n}^T \boldsymbol{e}_{2n} \psi + o_p(\lambda_{j(A)}) \tag{A.7}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. Then, from Lemmas A.2 and A.3, we have that for all $i \neq i'$ $(\leq j)$

$$\begin{aligned} \hat{\boldsymbol{v}}_i^T (\boldsymbol{S}_D/\lambda_{i(A)}) \boldsymbol{v}_{i'(A)} &= (\hat{\lambda}_i/\lambda_{i(A)}) \hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i'(A)} = \{1 + \kappa_i + o_p(1)\} \hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i'(A)} &\text{(A.8)} \\ &= (\lambda_{i'}/\lambda_{i(A)}) \hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i'(A)} + \kappa_i \hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i'(A)} + o_p(\lambda_{j(A)}/\lambda_{i(A)}) \\ &\quad + o_p\{\lambda_{r(A)}^{1/2} (\lambda_{i(A)}^{1/2} + \lambda_{i'(A)}^{1/2})/\lambda_{i(A)}\}. &\text{(A.9)} \end{aligned}$$

By combining (A.8) and (A.9), we can claim that for all $i \neq i'$ $(\leq j)$

$$\hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i'(A)} = o_p(\lambda_{j(A)}^{1/2} / \max\{\lambda_{i(A)}^{1/2}, \lambda_{i'(A)}^{1/2}\}). \tag{A.10}$$

Let $\boldsymbol{A}_{(i)} = \lambda_{i(A)}^{1/2} \boldsymbol{u}_{i(A)} \boldsymbol{v}_{i(A)}^T$ for $i = 1, \ldots, j$. From Lemmas A.2, A.3, (A.2), (A.7) and (A.10), we have that for $i = 1, \ldots, j$

$$\begin{aligned}\frac{\hat{\lambda}_i}{\lambda_{i(A)}} - \kappa_i =& (\hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i(A)})^2 + \hat{\boldsymbol{v}}_i^T \frac{\boldsymbol{A}^T \boldsymbol{W} + \boldsymbol{W}^T \boldsymbol{A}}{n^{1/2} \lambda_{i(A)}} \hat{\boldsymbol{v}}_i + o_p(\lambda_{j(A)} / \lambda_{i(A)}) \\ =& (\hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i(A)})^2 + \hat{\boldsymbol{v}}_i^T \frac{\boldsymbol{A}_{(i)}^T \boldsymbol{W} + \boldsymbol{W}^T \boldsymbol{A}_{(i)}}{n^{1/2} \lambda_{i(A)}} \hat{\boldsymbol{v}}_i + o_p(\lambda_{j(A)} / \lambda_{i(A)}) \quad \text{(A.11)}\end{aligned}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Here, there exist a random variable $\varepsilon_i \in [-1, 1]$ and a random unit vector $\boldsymbol{y}_i$ such that

$$\hat{\boldsymbol{v}}_i = (1 - \varepsilon_i^2)^{1/2} \boldsymbol{v}_{i(A)} + \varepsilon_i \boldsymbol{y}_i \quad \text{and} \quad \boldsymbol{v}_{i(A)}^T \boldsymbol{y}_i = 0.$$

Note that $\varepsilon_i = o_p(1)$ from $\hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i(A)} = 1 + o_p(1)$. By combining (A.11) with Lemma A.2, we have that for $i = 1, \ldots, j$

$$\begin{aligned}&\frac{\hat{\lambda}_i}{\lambda_{i(A)}} - \kappa_i - 2\boldsymbol{v}_{i(A)}^T \frac{\boldsymbol{A}_{(i)}^T \boldsymbol{W}}{n^{1/2} \lambda_{i(A)}} \boldsymbol{v}_{i(A)} \\ &= 1 - \varepsilon_i^2 + \hat{\boldsymbol{v}}_i^T \frac{\boldsymbol{A}_{(i)}^T \boldsymbol{W} + \boldsymbol{W}^T \boldsymbol{A}_{(i)}}{n^{1/2} \lambda_{i(A)}} \hat{\boldsymbol{v}}_i - 2\boldsymbol{v}_{i(A)}^T \frac{\boldsymbol{A}_{(i)}^T \boldsymbol{W}}{n^{1/2} \lambda_{i(A)}} \boldsymbol{v}_{i(A)} + o_p(\lambda_{j(A)} / \lambda_{i(A)}) \\ &= 1 - \varepsilon_i^2 + o_p(\varepsilon_i^2) + 2\varepsilon_i \boldsymbol{y}_i^T \frac{\boldsymbol{A}_{(i)}^T \boldsymbol{W} + \boldsymbol{W}^T \boldsymbol{A}_{(i)}}{n^{1/2} \lambda_{i(A)}} \boldsymbol{v}_{i(A)} + o_p(\lambda_{j(A)} / \lambda_{i(A)}) \\ &= 1 - \varepsilon_i^2 \{1 + o_p(1)\} + o_p\{\varepsilon_i (\lambda_{r(A)} / \lambda_{i(A)})^{1/2}\} + o_p(\lambda_{j(A)} / \lambda_{i(A)}) \\ &= 1 + o_p(\lambda_{j(A)} / \lambda_{i(A)}) \qquad \text{(A.12)}\end{aligned}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Thus, it follows that for $i = 1, \ldots, j$

$$\varepsilon_i = o_p\{(\lambda_{j(A)} / \lambda_{i(A)})^{1/2}\}. \tag{A.13}$$

Hence, from $\hat{\boldsymbol{v}}_i = (1 - \varepsilon_i^2)^{1/2} \boldsymbol{v}_{i(A)} + \varepsilon_i \boldsymbol{y}_i$, it holds that for $i = 1, \ldots, j$

$$\hat{\boldsymbol{v}}_i^T \boldsymbol{v}_{i(A)} = (1 - \varepsilon_i^2)^{1/2} = 1 + o_p(\lambda_{j(A)} / \lambda_{i(A)}).$$

On the other hand, from (A.12), for $i = 1, \ldots, j$, it holds $\hat{\lambda}_i / \lambda_{i(A)} = 1 + \kappa_i + 2\delta_i + o_p(\lambda_{j(A)} / \lambda_{i(A)})$ under (C-i) in (I) or under (C-i) to (C-iii) in (II). It concludes the results. □

Lemma A.6. *For j $(\leq r)$, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that*

$$\hat{\boldsymbol{u}}_i^T \boldsymbol{u}_{i(A)} = (1 + \kappa_i + 2\delta_i)^{-1/2} \{1 + \delta_i + o_p(\lambda_{j(A)} / \lambda_{i(A)})\} \quad \textit{for } i = 1, \ldots, j;$$
$$\hat{\boldsymbol{u}}_i^T \boldsymbol{u}_{i'(A)} = (1 + \kappa_i + 2\delta_i)^{-1/2} \times o_p\{(\lambda_{j(A)} / \lambda_{i(A)})^{1/2}\} \quad \textit{for all } i \neq i' \ (\leq j)$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Proof. Note that $\hat{\boldsymbol{u}}_i = (n\hat{\lambda}_i)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{v}}_i$. From (A.13), Lemmas A.2 and A.5, under (C-i) in (I) or under (C-i) to (C-iii) in (II), we have that for j $(\le r)$

$$\begin{aligned}\frac{\hat{\lambda}_i^{1/2}}{\lambda_{i(A)}^{1/2}}\boldsymbol{u}_{i(A)}^T\hat{\boldsymbol{u}}_i &= \boldsymbol{v}_{i(A)}^T\hat{\boldsymbol{v}}_i + \frac{\boldsymbol{u}_{i(A)}^T\boldsymbol{W}\hat{\boldsymbol{v}}_i}{n^{1/2}\lambda_{i(A)}^{1/2}} \\ &= 1+\delta_i + o_p(\lambda_{j(A)}/\lambda_{i(A)}) \quad \text{for } i=1,\dots,j; \\ \frac{\hat{\lambda}_i^{1/2}}{\lambda_{i(A)}^{1/2}}\boldsymbol{u}_{i'(A)}^T\hat{\boldsymbol{u}}_i &= \frac{\lambda_{i'(A)}^{1/2}}{\lambda_{i(A)}^{1/2}}\boldsymbol{v}_{i'(A)}^T\hat{\boldsymbol{v}}_i + \frac{\boldsymbol{u}_{i'(A)}^T\boldsymbol{W}\hat{\boldsymbol{v}}_i}{n^{1/2}\lambda_{i(A)}^{1/2}} \\ &= o_p\{(\lambda_{j(A)}/\lambda_{i(A)})^{1/2}\} \quad \text{for all } i\ne i' \ (\le j)\end{aligned}$$

as $d\to\infty$ either when n is fixed or $n\to\infty$. From Lemma A.5, we can conclude the results. □

Proof of Lemma 5.1. We write that $\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W}/n) = \sum_{s=1}^d \lambda_{s(W)}\sum_{k=1}^n (z_{sk}^2 - 1)/n + \mathrm{tr}(\boldsymbol{\Sigma}_W)$. Under (C-i), it holds that for j $(\le r)$

$$\frac{E[\{\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W}/n) - \mathrm{tr}(\boldsymbol{\Sigma}_W)\}^2]}{\lambda_{j(A)}^2} = \sum_{s,t=1}^d \frac{\lambda_{s(W)}\lambda_{t(W)}E\{(z_{sk}^2-1)(z_{tk}^2-1)\}}{n\lambda_{j(A)}^2} = o(1)$$

as $d\to\infty$ either when n is fixed or $n\to\infty$. By using Chebyshev's inequality, for any $\tau>0$, it holds that

$$P\Big(|\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W}/n) - \mathrm{tr}(\boldsymbol{\Sigma}_W)| \ge \tau\lambda_{j(A)}\Big) \le \frac{E[\{\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W}/n) - \mathrm{tr}(\boldsymbol{\Sigma}_W)\}^2]}{\tau^2\lambda_{j(A)}^2} = o(1)$$

for j $(\le r)$. It follows that $\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W}/n) = \mathrm{tr}(\boldsymbol{\Sigma}_W) + o_p(\lambda_{j(A)})$. We write that

$$\begin{aligned}&\mathrm{tr}(\boldsymbol{S}_D) \\ &= \mathrm{tr}(\boldsymbol{A}^T\boldsymbol{A}) + 2\frac{\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{W})}{n^{1/2}} + \frac{\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W})}{n} = \sum_{i=1}^r \lambda_{i(A)}(1+2\delta_i) + \frac{\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W})}{n}.\end{aligned}$$

Since $\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W}/n) = \mathrm{tr}(\boldsymbol{\Sigma}_W) + o_p(\lambda_{j(A)})$, it holds that for j $(\le r)$

$$\mathrm{tr}(\boldsymbol{S}_D) = \sum_{i=1}^r \lambda_{i(A)}(1+2\delta_i) + \mathrm{tr}(\boldsymbol{\Sigma}_W) + o_p(\lambda_{j(A)}) \tag{A.14}$$

under (C-i). Note that

$$\delta_i = o_p\{(\lambda_{r(A)}/\lambda_{i(A)})^{1/2}\} \quad \text{for } i=1,\dots,r \tag{A.15}$$

from Lemma A.2. Then, from Lemma A.5 and (A.14), it holds that for j $(\le r)$

$$\frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{i=1}^j \hat{\lambda}_i}{n-j} = \begin{cases}\psi + \sum_{i=j+1}^r \lambda_{i(A)}/(n-j) + o_p(\lambda_{j(A)}) & \text{when } j<r; \\ \psi + o_p(\lambda_{r(A)}) & \text{when } j=r\end{cases}$$

under (C-i) in (I) or under (C-i), (C-iii) and (C-iv) in (II). It concludes the result. □

Appendix B

In this section, we give proofs of the theorems, corollaries and propositions in Sections 3 to 5.

Proofs of Theorem 3.1 and Corollary 3.1. By noting that $\kappa_j = o(1)$ for $j\ (\le r)$ under (C-iv), the results are obtained straightforwardly from Lemmas A.3 and A.4. □

Proofs of Theorem 3.2 and Corollary 3.2. Note that $\delta_i = o_p(1)$ for $i = 1, \dots, r$, from Lemma A.2. Then, from Lemmas A.5 and A.6, under (C-i) in (I) or under (C-i) to (C-iii) in (II), we have that for $j\ (\le r)$

$$\begin{aligned}
&||\hat{\lambda}_i^{1/2}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}_i^T - \lambda_{i(A)}^{1/2}\boldsymbol{u}_{i(A)}\boldsymbol{v}_{i(A)}^T||_F^2 \\
&= \hat{\lambda}_i + \lambda_{i(A)} - 2\hat{\lambda}_i^{1/2}\lambda_{i(A)}^{1/2}\hat{\boldsymbol{u}}_i^T\boldsymbol{u}_{i(A)}\{1 + o_p(\lambda_{j(A)}/\lambda_{i(A)})\} \\
&= \psi + o_p(\lambda_{j(A)}) \quad \text{for } i = 1, \dots, j; \\
&\text{tr}\{(\hat{\lambda}_i^{1/2}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}_i^T)(\lambda_{i'(A)}^{1/2}\boldsymbol{v}_{i'(A)}\boldsymbol{u}_{i'(A)}^T)\} = o_p(\lambda_{j(A)}) \quad \text{for all } i \ne i'\ (\le j)
\end{aligned} \tag{B.1}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. Thus, it holds that

$$L(\widehat{\boldsymbol{A}}_r|\boldsymbol{A}) = \Big\|\sum_{i=1}^r(\hat{\lambda}_i^{1/2}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}_i^T - \lambda_{i(A)}^{1/2}\boldsymbol{u}_{i(A)}\boldsymbol{v}_{i(A)}^T)\Big\|_F^2 = r\psi + o_p(\lambda_{r(A)})$$

under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$. It concludes the result of Theorem 3.2. By noting that $\psi = o(\lambda_r)$ under (C-iv) with $j = r$, the result of Corollary 3.2 is obtained straightforwardly from Theorem 3.2. □

Proof of Theorem 4.1. By combining Lemma A.3 with Lemma 5.1, under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that for $j\ (\le r)$

$$\tilde{\lambda}_j = \begin{cases} \lambda_{j(A)} - \sum_{i=j+1}^r \lambda_{i(A)}/(n-j) + o_p(\lambda_{j(A)}) & \text{when } j < r; \\ \lambda_{r(A)} + o_p(\lambda_{r(A)}) & \text{when } j = r \end{cases} \tag{B.2}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. It concludes the result. □

Proof of Corollary 4.1. For $j\ (< r)$, we first consider the case when $\liminf_{d\to\infty} \lambda_{i(A)}/\lambda_{j(A)} > 0$ for $i > j$ either when n is fixed or $n \to \infty$. From Lemma A.3, under (C-i) in (I) or under (C-i) to (C-iii) in (II), we can claim that

$$\frac{\hat{\lambda}_i}{\lambda_{j(A)}} = \frac{\lambda_{i(A)}}{\lambda_{j(A)}} + \kappa_j + o_p(1) \tag{B.3}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Next, we consider the case when $\lambda_{i(A)}/\lambda_{j(A)} = o(1)$ for $i > j$. From Lemma A.3, we obtain that for $j\ (\le r)$

$$\boldsymbol{v}_{i'(A)}^T\hat{\boldsymbol{v}}_i = o_p\{(\lambda_{j(A)}/\lambda_{i'(A)})^{1/2}\} \quad \text{for } i' = 1, \dots, j;\ i > j \tag{B.4}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Then, from Lemmas A.1, A.2 and (A.2), we have that for j $(\le r)$ and $i > j$

$$\frac{\hat{\lambda}_i}{\lambda_{j(A)}} = \hat{\boldsymbol{v}}_i^T \frac{\boldsymbol{A}^T\boldsymbol{A}}{\lambda_{j(A)}}\hat{\boldsymbol{v}}_i + 2\hat{\boldsymbol{v}}_i^T \frac{\boldsymbol{A}^T\boldsymbol{W}}{n^{1/2}\lambda_{j(A)}}\hat{\boldsymbol{v}}_i + \hat{\boldsymbol{v}}_i^T \frac{\boldsymbol{W}^T\boldsymbol{W}}{n\lambda_{j(A)}}\hat{\boldsymbol{v}}_i = \kappa_j + o_p(1) \qquad \text{(B.5)}$$

because $\lambda_{i(A)}/\lambda_{j(A)} = o(1)$. By using the convergent subsequence of $\lambda_{i(A)}/\lambda_{j(A)}$ for $i > j$, from Lemma A.5, (A.14), (A.15), (B.3) and (B.5), it holds that for j $(\le r)$

$$\hat{\psi}_r = \psi + o_p(\lambda_{j(A)}) \qquad \text{(B.6)}$$

under (C-i) in (I) or under (C-i) to (C-iii) in (II). Hence, from Lemma A.3, we can conclude the result. □

Proofs of Theorem 4.2 and Corollary 4.2. From Lemma A.5 and (B.6), under (C-i) in (I) or under (C-i) to (C-iii) in (II), it holds that for j $(\le r)$

$$\frac{\tilde{\lambda}_{i(r)}}{\lambda_{i(A)}} = 1 + 2\delta_i + o_p(\lambda_{j(A)}/\lambda_{i(A)}) \quad \text{for } i = 1, \dots, j$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. Then, from Lemmas A.5, A.6 and (A.15), we have that

$$\begin{aligned}
&||\tilde{\lambda}_{i(r)}^{1/2}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}_i^T - \lambda_{i(A)}^{1/2}\boldsymbol{u}_{i(A)}\boldsymbol{v}_{i(A)}^T||_F^2 \\
&= 2\lambda_{i(A)}\{1 + \delta_i - (1 + 2\delta_i)^{1/2}\hat{\boldsymbol{u}}_i^T\boldsymbol{u}_{i(A)}\} + o_p(\lambda_{j(A)}) \\
&= 2\lambda_{i(A)}(1 + \delta_i)(1 - \hat{\boldsymbol{u}}_i^T\boldsymbol{u}_{i(A)}) + o_p(\lambda_{j(A)}) \quad \text{for } i = 1, \dots, j; \\
&\mathrm{tr}\{(\tilde{\lambda}_{i(r)}^{1/2}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}_i^T)(\lambda_{i'(A)}^{1/2}\boldsymbol{v}_{i'(A)}\boldsymbol{u}_{i'(A)}^T)\} = o_p(\lambda_{j(A)}) \quad \text{for all } i \ne i' \ (\le j).
\end{aligned}$$

Hence, it holds that

$$L(\widetilde{\boldsymbol{A}}_r|\boldsymbol{A}) = 2\sum_{i=1}^{r}\lambda_{i(A)}(1+\delta_i)\Big(1 - \frac{1+\delta_i}{(1+\kappa_i+2\delta_i)^{1/2}}\Big) + o_p(\lambda_{r(A)}) \qquad \text{(B.7)}$$

under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$. From (A.15) and (B.7), it concludes the results of Theorem 4.2. By noting that $\psi = o(\lambda_r)$ under (C-iv) with $j = r$, the result of Corollary 4.2 is obtained straightforwardly from Theorem 4.2. □

Proof of Proposition 5.1. We first consider the case when $r_* < r$. From (B.1), under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$, it holds that

$$L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A}) = r_*\psi + \sum_{i=r_*+1}^{r}\lambda_{i(A)} + o_p(\lambda_{r(A)})$$

as $d \to \infty$ either when n is fixed or $n \to \infty$.

Next, we consider the case when $r_* > r$. From (B.5), it holds that

$$\hat{\lambda}_i = \psi + o_p(\lambda_{r(A)}) \quad \text{for } i > r \qquad \text{(B.8)}$$

under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$. Note that $\hat{\boldsymbol{u}}_i = (n\hat{\lambda}_i)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{v}}_i$. From Lemma A.2 and (B.4), it holds that for $i > r$ and $i' = 1, \dots, r$

$$\hat{\lambda}_i^{1/2}\boldsymbol{u}_{i'(A)}^T\hat{\boldsymbol{u}}_i = \lambda_{i'(A)}^{1/2}\boldsymbol{v}_{i'(A)}^T\hat{\boldsymbol{v}}_i + \boldsymbol{u}_{i'(A)}^T\boldsymbol{W}\hat{\boldsymbol{v}}_i/n^{1/2} = o_p(\lambda_{r(A)}^{1/2}) \tag{B.9}$$

under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$. By combining Theorem 3.2 with (B.4), (B.8) and (B.9), it holds that $L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A}) = r_*\psi + o_p(\lambda_{r(A)})$. It concludes the result. □

Proof of Proposition 5.3. We first consider the case when $r_* \geq r$. From Lemma A.5, (A.14) and (B.5), under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$, we can claim that for $i > r$

$$\hat{\psi}_i = \psi + o_p(\lambda_{r(A)}) \tag{B.10}$$

as $d \to \infty$ either when n is fixed or $n \to \infty$. Thus, from (B.5) and (B.10), it follows that

$$\tilde{\lambda}_{i(r_*)}/\lambda_{r(A)} = o_p(1) \quad \text{for } r < i \leq r_* \tag{B.11}$$

under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$. Note that $\tilde{\lambda}_{i(r_*)}\boldsymbol{u}_{i'(A)}^T\hat{\boldsymbol{u}}_i = o_p(\lambda_{r(A)}^{1/2})$ for $i > r$ and $i' = 1, \dots, r$, from (B.9) and $\tilde{\lambda}_{i(r_*)} \leq \hat{\lambda}_i$ for all i. Then, by combining Theorem 4.2 with (A.15), (B.4) and (B.11), it holds that

$$L(\widehat{\boldsymbol{A}}_{r_*}|\boldsymbol{A}) = \sum_{i=1}^{r}\lambda_{i(A)}\Big(2\gamma_i + o_p(1)\Big) + o_p(\lambda_{r(A)}).$$

It concludes the result when $r_* \geq r$.

Next, we consider the case when $r_* < r$. Assume $\lambda_{r_*+1(A)}/(\lambda_{r_*(A)}n) = o(1)$. In a way similar to (B.2), we obtain that $\tilde{\lambda}_{i(r_*)} = \lambda_{i(A)} - \sum_{i'=r_*+1}^{r}\lambda_{i'(A)}/(n - r_*) + o_p(\lambda_{i(A)}) = \lambda_{i(A)}\{1 + o_p(1)\}$ for $i \leq r_*$ under (C-i) in (I) with $j = r$ or under (C-i) to (C-iii) in (II) with $j = r$. Then, from Lemmas A.5, A.6 and (A.15), we have that for $i = 1, \dots, r_*$

$$\begin{aligned}
&||\tilde{\lambda}_{i(r_*)}^{1/2}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}_i^T - \lambda_{i(A)}^{1/2}\boldsymbol{u}_{i(A)}\boldsymbol{v}_{i(A)}^T||_F^2 \\
&= 2\lambda_{i(A)}\{1 - \hat{\boldsymbol{u}}_i^T\boldsymbol{u}_{i(A)} + o_p(1)\} = 2\lambda_{i(A)}\{\gamma_i + o_p(1)\}; \\
&\text{and} \ \ \mathrm{tr}\{(\tilde{\lambda}_{i(r_*)}^{1/2}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}_i^T)(\lambda_{i'(A)}^{1/2}\boldsymbol{v}_{i'(A)}\boldsymbol{u}_{i'(A)}^T)\} = o_p(\lambda_{r(A)}) \quad \text{for } i'(\neq i) = 1, \dots, r.
\end{aligned}$$

Hence, we can conclude the result when $r_* < r$. It concludes the results of Proposition 5.3. □

Proofs of Propositions 5.2 and 5.4. Note that $\sum_{i=j+1}^{r}\lambda_{i(A)}/\{(n-j)\lambda_{j(A)}\} \leq (r-j)/(n-j)$ for $j < r$. From Lemma 5.1, (B.2), (B.10) and (B.11), we can conclude the results. □

Acknowledgements

The authors would like to thank a referee for his/her kind comments.

References

[1] Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J. and Meyerson, M. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **98** 13790–13795.

[2] Ishii, A., Yata, K. and Aoshima, M. (2016). Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context. *J. Statist. Plann. Inference* **170** 186–199. MR3427869

[3] Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454

[4] Murayama, W., Yata, K. and Aoshima, M. (2015). Reconstruction of a signal matrix for high-dimension, low-sample-size data. *RIMS Koukyuroku* **1954** 23–31.

[5] Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348

[6] Rohde, A. and Tsybakov, A. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342

[7] Shabalin, A. and Nobel, A. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *J. Multivariate Anal.* **118** 67–76. MR3054091

[8] Shen, D., Shen, H. and Marron, J. S. (2013). Consistency of sparse PCA in high dimension, low sample size contexts. *J. Multivariate Anal.* **115** 317–333. MR3004561

[9] Yang, K., Cai, Z., Li, J. and Lin, G. (2006). A stable gene selection in microarray data analysis. *BMC Bioinformatics* **7** 228.

[10] Yata, K. and Aoshima, M. (2009). PCA consistency for non-Gaussian data in high dimension, low sample size context. *Comm. Statist. Theory Methods, Special Issue Honoring Zacks, S. (ed. Mukhopadhyay, N.)* **38** 2634–2652. MR2568176

[11] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105** 193–215. MR2877512

[12] Yata, K. and Aoshima, M. (2013). PCA consistency for the power spiked model in high-dimensional settings. *J. Multivariate Anal.* **122** 334–354. MR3189327

[13] Zhou, Y.-H. and Marron, J. S. (2015). High dimension low sample size asymptotics of robust PCA. *Electron. J. Stat.* **9** 204–218. MR3312407