Supplement to "High-dimensional quadratic classifiers in non-sparse settings"

Makoto Aoshima and Kazuyoshi Yata
Institute of Mathematics, University of Tsukuba, Ibaraki, Japan

Abstract

In this supplement, we give actual data analyses and proofs of the theoretical results in the main work in Aoshima and Yata [2] together with additional corollaries and proofs of the corollaries. The equation numbers and the mathematical symbols used in the supplement are the same as those which are made reference to in the main document.

Appendix A

In this appendix, we discuss the performance of the classifiers given by [2] in actual data analyses.

We first analyzed gene expression data given by Golub et al. [7] in which the data set consists of 7129 ($=p$) genes and 72 samples. We had two classes of leukemia subtypes, that is, π_1: acute lymphoblastic leukemia (ALL) (47 samples) and π_2: acute myeloid leukemia (AML) (25 samples). The data set consisted of two sets as 38 training samples (ALL: 27 samples and AML: 11 samples) and 34 test samples (ALL: 20 samples and AML: 14 samples). If each sample has unit variance, $\boldsymbol{S}_{1n_1(d)}=\boldsymbol{S}_{2n_2(d)}$. Thus, we did not standardize each sample so as to have unit variance.

First, we checked several sparsity conditions. We standardized each sample by $\boldsymbol{x}_{ik}/\{\sum_{l=1}^{2}\mathrm{tr}(\boldsymbol{S}_{ln_l})/(2p)\}^{1/2}$ for all i,k, so that $\mathrm{tr}(\boldsymbol{S}_{1n_1})/2+\mathrm{tr}(\boldsymbol{S}_{2n_2})/2=p$. By using all the samples (i.e., 72 samples), we calculated that

$$\hat{\Delta}_{(I)}=2060\ (=0.289p), \tag{A.1}$$

where $\hat{\Delta}_{(I)}$ is given in Section 4.2. Note that $E(\hat{\Delta}_{(I)})=||\boldsymbol{\mu}_{12}||^2$. From this observation, we concluded that $\boldsymbol{\mu}_{12}$ is non-sparse. Next, we considered an estimator of $||\boldsymbol{\Sigma}_{12}||_F^2=\sum_{i=1}^{2}\mathrm{tr}(\boldsymbol{\Sigma}_i^2)-2\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)$ by $\hat{\Delta}_{\Sigma}=\sum_{i=1}^{2}W_{in_i}-2\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2})$ having W_{in_i}s defined by (16) in Aoshima and Yata [1]. Here, W_{in_i} is an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$, so that $E(\hat{\Delta}_{\Sigma})=||\boldsymbol{\Sigma}_{12}||_F^2$. We calculated that

$$\hat{\Delta}_{\Sigma}=9.77\times 10^5\ (=137p). \tag{A.2}$$

Also, we estimated $\Delta_{i(III)}$ and $\Delta_{(III')}$ by $\hat{\Delta}_{i(III)}=\{\sum_{j=1}^{p}(\overline{x}_{1jn_1}-\overline{x}_{2jn_2})^2/s_{i'n_{i'}(j)}+s_{in_i(j)}/s_{i'n_{i'}(j)}-1+\log(s_{i'n_{i'}(j)}/s_{in_i(j)})\}$ and $\hat{\Delta}_{(III')}=\sum_{j=1}^{p}(\overline{x}_{1jn_1}-\overline{x}_{2jn_2})^2/s_{n(j)}$, where $i'\neq i$ and $s_{n(j)}$s are defined in Section 4.3. We calculated that

$$\begin{aligned}&\hat{\Delta}_{\min(II)}=\min\{\hat{\Delta}_{1(II)},\hat{\Delta}_{2(II)}\}=2037\ (=0.286p),\\&\hat{\Delta}_{\min(III)}=\min\{\hat{\Delta}_{1(III)},\hat{\Delta}_{2(III)}\}=16909\ (=2.371p)\ \text{ and }\ \hat{\Delta}_{(III')}=1709\ (=0.24p),\end{aligned} \tag{A.3}$$

where $\hat{\Delta}_{i(II)}$s are defined in Section 4.2. From (A.2), we concluded that $\boldsymbol{\Sigma}_{12}$ is non-sparse. In fact, from (A.3), the difference of diagonal elements between two covariance matrices must be very

Address correspondence to Makoto Aoshima, Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan; Fax: +81-298-53-6501; E-mail: aoshima@math.tsukuba.ac.jp

Table 1: Error rates of the classifiers for samples from Golub et al. [7].

Classifier	DBDA	GQDA	DLDA-bc	DQDA-bc	FS-DQDA with $\gamma = 1/6$ and $1/3$	FS-DQDA with $\gamma = 1/2, 2/3$ and $5/6$	HM-LSVM
				Test samples (ALL: 20 and AML: 14)			
Error rate	1/34	1/34	5/34	2/34	4/34	3/34	1/34
				LOOCV of samples (ALL: 47 and AML: 25)			
Error rate	3/72	6/72	11/72	1/72	0/72	0/72	2/72

large. Thus from Section 3.1, the classification rule by (3) with (III) has consistency (6), so that the Bayes error rates for this data set are probably close to 0. Also, we calculated

$$(\acute{\lambda}_{\max}(\boldsymbol{\Sigma}_1), \acute{\lambda}_{\max}(\boldsymbol{\Sigma}_2)) = (1223, 1457) \ (= (0.172p, 0.204p)), \tag{A.4}$$

where $\acute{\lambda}_{\max}(\boldsymbol{\Sigma}_i)$ is an estimate of the largest eigenvalue due to the noise-reduction methodology by Yata and Aoshima [12]. We concluded that "$\lambda(\boldsymbol{\Sigma}_i) \in (0, \infty)$ as $p \to \infty$" does not hold and $\boldsymbol{\Sigma}_i$s are not sparse because $\lambda_{\max}(\boldsymbol{\Sigma}_i)$s are very large. Therefore, we do not recommend to apply the classifier by thresholding (or sparse) estimation of $\boldsymbol{\Sigma}_i^{-1}$, such as $W_i(\{T_{\tau_{n_i}}(\boldsymbol{S}_{in_i})\}^{-1})$. See Section 5.2 for the details. Actually, we did not use any classifiers by thresholding estimation of $\boldsymbol{\Sigma}_i^{-1}$ in this section. Also, note that the computational cost for the thresholding estimation of $\boldsymbol{\Sigma}_i^{-1}$ is very high when p is large.

We constructed the classifiers: DBDA, GQDA, DLDA-bc, DQDA-bc and FS-DQDA, by using the training samples of sizes $n_1 = 27$ and $n_2 = 11$, and checked accuracy by using the test samples from each π_i. Throughout this section, we considered

$$\gamma = 1/6,\ 1/3,\ 1/2,\ 2/3,\ \text{and } 5/6 \ \text{ in (23) for FS-DQDA.}$$

We compared the classifiers with the hard-margin linear support vector machine (HM-LSVM). See Vapnik [11] for the details. Note that the data sets are linearly separable by a hyperplane because $p > n_1 + n_2$. We emphasize that the computational cost of DBDA, GQDA, DLDA-bc, DQDA-bc or FS-DQDA is as low as HM-LSVM even when $p \geq 10,000$. We summarized misclassification rates in the first block of Table 1. We observed that the error rates of FS-DQDA with $\gamma = 1/6$ and $1/3$ are 4/34, and the ones with $\gamma = 1/2$, $2/3$ and $5/6$ are 3/34. We note that $n_{\min} = 11$ and $n_{\min}^{-1} \log p = 0.81$, so that "$n_{\min}^{-1} \log p = o(1)$" does not hold. That is probably the reason why DLDA-bc, DQDA-bc and FS-DQDA lose the consistency property. See Sections 4 and 5 for the details. On the other hand, DBDA and GQDA gave reasonable performances even when n_is are small and seem to hold the consistency property. We calculated $\text{tr}(\boldsymbol{S}_{1n_1})/\text{tr}(\boldsymbol{S}_{2n_2}) = 0.989$ and $(\hat{\Delta}_{i(II)}\text{tr}(\boldsymbol{S}_{i'n_{i'}})/p)/\hat{\Delta}_{(I)} \approx 1$ for $i \neq i'$. The difference of the trace of the covariance matrices is small and this is probably the reason why DBDA gave a preferable performance. See Section 4.2 for the details. HM-LSVM also gave a preferable performance. See Hall et al. [8] for the consistency property of HM-LSVM. For this data set, Cai and Liu [5] summarized misclassification rates for several other classifiers including a sparse linear classifier called LPD. See Table 6 in [5] for the performance of the other classifiers. Note that LPD has the Bayes error rates asymptotically under several sparsity conditions. We observed that DBDA and GQDA gave the same accuracy as LPD. This is probably because the sparsity conditions do not hold for this data set, so that the Bayes error rates are almost 0. However, the computational cost for DBDA and GQDA is much lower than LPD.

Table 2: Estimates of $(||\boldsymbol{\mu}_{12}||^2,\ ||\boldsymbol{\Sigma}_{12}||_F^2,\ \Delta_{\min(II)},\ \Delta_{\min(III)},\ \Delta_{(III')})$ by $(\hat{\Delta}_{(I)},\ \hat{\Delta}_{\Sigma},\ \hat{\Delta}_{\min(II)},\ \hat{\Delta}_{\min(III)},\ \hat{\Delta}_{(III')})$ for Armstrong et al. [3].

Case	(a) ALL and MLL	(b) ALL and AML	(c) MLL and AML
$\|\|\boldsymbol{\mu}_{12}\|\|^2$	$4076\ (= 0.324p)$	$15050\ (= 1.2p)$	$8546\ (= 0.679p)$
$\|\|\boldsymbol{\Sigma}_{12}\|\|_F^2$	$1.12 \times 10^8\ (= 8863p)$	$5.49 \times 10^6\ (= 436p)$	$1.16 \times 10^8\ (= 9192p)$
$\Delta_{\min(II)}$	$4078\ (= 0.324p)$	$14212\ (= 1.13p)$	$7945\ (= 0.631p)$
$\Delta_{\min(III)}$	$42848\ (= 3.406p)$	$77701\ (= 6.176p)$	$27316\ (= 2.171p)$
$\Delta_{(III')}$	$3055\ (= 0.243p)$	$12187\ (= 0.969p)$	$7777\ (= 0.618p)$

Next, by using all the samples (i.e., 72 samples), we checked accuracy of the classifiers by the leave-one-out cross-validation (LOOCV). We summarized misclassification rates in the second block of Table 1. We note that $n_{\min} = 24$ and $n_{\min}^{-1} \log p = 0.37$ or $n_{\min} = 25$ and $n_{\min}^{-1} \log p = 0.35$ in this case, so that $n_{\min}^{-1} \log p$ is a little small. We observed that DQDA-bc and FS-DQDA give preferable performances. For $\gamma = 1/6,\ 1/3,\ 1/2,\ 2/3$ and $5/6$, all the error rates of FS-DQDA were 0/72. On the other hand, DLDA-bc gave a poor performance because it does not draw information about heteroscedasticity. See (A.3) and Section 4.2. For other classifiers, Tan et al. [10] summarized results of the LOOCV for this data set.

Finally, we analyzed gene expression data given by Armstrong et al. [3] in which the data set consists of 12582 $(= p)$ genes and 72 samples. We had three classes of leukemia subtypes: acute lymphoblastic leukemia (ALL: 24 samples), mixed-lineage leukemia (MLL: 20 samples), and acute myeloid leukemia (AML: 28 samples). We considered three cases: (a) ALL and MLL, (b) ALL and AML, and (c) MLL and AML. We standardized each sample by $\boldsymbol{x}_{ik}/\{\sum_{l=1}^{3} \mathrm{tr}(\boldsymbol{S}_{ln_l})/(3p)\}^{1/2}$ for all i, k, as before. Then, we calculated $(\hat{\Delta}_{(I)},\ \hat{\Delta}_{\Sigma},\ \hat{\Delta}_{\min(II)},\ \hat{\Delta}_{\min(III)},\ \hat{\Delta}_{(III')})$ for the three cases. We summarized $(\hat{\Delta}_{(I)},\ \hat{\Delta}_{\Sigma},\ \hat{\Delta}_{\min(II)},\ \hat{\Delta}_{\min(III)},\ \hat{\Delta}_{(III')})$s in Table 2. From Table 2, we concluded that $\boldsymbol{\mu}_{12}$ and $\boldsymbol{\Sigma}_{12}$ are non-sparse for (a) to (c). Also, by using $\acute{\lambda}_{\max}(\boldsymbol{\Sigma}_i)$, we estimated the largest eigenvalues as 1896, 3206 and 2101 for ALL, MLL and AML, respectively. From this observation, we concluded that $\boldsymbol{\Sigma}_i$s are non-sparse. We estimated $\mathrm{tr}(\boldsymbol{\Sigma}_{\max}^2)/(n_{\min}\Delta_{(I)}^2)$ and $\lambda_{\max}/\Delta_{(I)}$ by $C_1 = \max\{W_{1n_1}, W_{2n_2}\}/(n_{\min}\hat{\Delta}_{(I)}^2)$ and $C_2 = \max\{\acute{\lambda}_{\max}(\boldsymbol{\Sigma}_1), \acute{\lambda}_{\max}(\boldsymbol{\Sigma}_2)\}/\hat{\Delta}_{(I)}$ in (C-i') and (C-ii'). Then, we had (C_1, C_2) as (0.362, 0.787) for (a), (0.001, 0.14) for (b), and (0.082, 0.375) for (c). Note that $\liminf_{p\to\infty} \Delta_{\min(II)}/\Delta_{(I)} > 0$ and $\liminf_{p\to\infty} \Delta_{\min(III)}/\Delta_{(I)} > 0$. See Table 2. From these observations, it is likely that the classifiers by (I) to (III) satisfy (C-i') and (C-ii') especially for (b) and hold consistency (6) from Proposition 2.

Based on all the samples, we checked accuracy of the classifiers by using the LOOCV for (a) to (c). In addition, we checked accuracy for 3-class classification by the multiclass classification rule given in Remark 1. In the 3-class classification, we used $\hat{\theta}_j$ given in Remark 4 for FS-DQDA and used the one-versus-one approach for HM-LSVM. We summarized misclassification rates in Table 3. We observed that FS-DQDA gives excellent performances for all γs. HM-LSVM also gave reasonable performances, however, it does not draw information about the difference of the covariance matrices. See Section 2.2 in [1] for such an example. As for (b), all the classifiers gave preferable performances. This is probably because the classifiers by (I) to (III) satisfy (C-i') and (C-ii') for (b). Throughout this section, we observed that FS-DQDA does not heavily depend on γ located around 1/2. Also, FS-DQDA with $\gamma = 1/2$ gave preferable performances in the simulation study of Section 5.3. Thus, for the choice of $\gamma \in (0, 1)$ in (23), one may apply the cross-validation

Table 3: Error rates of the classifiers for samples from Armstrong et al. [3].

Classifier	DBDA	GQDA	DLDA-bc	DQDA-bc	FS-DQDA with $\gamma = 1/6, 1/3$ and $1/2$	FS-DQDA with $\gamma = 2/3$ and $5/6$	HM-LSVM
	LOOCV of samples from (a) ALL: 24 and MLL: 20						
Error rate	1/44	2/44	6/44	1/44	0/44	0/44	0/44
	LOOCV of samples from (b) ALL: 24 and AML: 28						
Error rate	1/52	1/52	1/52	0/52	0/52	0/52	0/52
	LOOCV of samples from (c) MLL: 20 and AML: 28						
Error rate	4/48	4/48	1/48	3/48	3/48	3/48	3/48
	LOOCV of samples from ALL: 24, MLL: 20 and AML: 28						
Error rate	5/72	6/72	7/72	4/72	2/72	3/72	3/72

method or simply choose as $\gamma = 1/2$.

Appendix B

In this appendix, we give proofs of the theoretical results in the main work in [2]. Also, we give additional corollaries and proofs of the corollaries.

Let $\omega_1 = \{\mathrm{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{A}_1)^2\}/n_i + \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{A}_2\boldsymbol{\Sigma}_2\boldsymbol{A}_2)/n_2\}^{1/2}$. Let $\tilde{\boldsymbol{x}}_{1k} = \boldsymbol{A}_1^{1/2}(\boldsymbol{x}_{1k} - \boldsymbol{\mu}_1)$ and $\tilde{\boldsymbol{x}}_{2k} = \boldsymbol{A}_1^{-1/2}\boldsymbol{A}_2(\boldsymbol{x}_{2k} - \boldsymbol{\mu}_2)$ for $k = 1, ..., n_i$. Let $\tilde{\boldsymbol{\Sigma}}_1 = \boldsymbol{A}_1^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{A}_1^{1/2}$, $\tilde{\boldsymbol{\Sigma}}_2 = \boldsymbol{A}_1^{-1/2}\boldsymbol{A}_2\boldsymbol{\Sigma}_2\boldsymbol{A}_2\boldsymbol{A}_1^{-1/2}$, $\tilde{\boldsymbol{\Gamma}}_1 = [\tilde{\gamma}_{11}, ..., \tilde{\gamma}_{1q_1}] = \boldsymbol{A}_1^{1/2}\boldsymbol{\Gamma}_1$ and $\tilde{\boldsymbol{\Gamma}}_2 = [\tilde{\gamma}_{21}, ..., \tilde{\gamma}_{2q_2}] = \boldsymbol{A}_1^{-1/2}\boldsymbol{A}_2\boldsymbol{\Gamma}_2$. Note that $\mathrm{Var}(\tilde{\boldsymbol{x}}_{ij}) = \tilde{\boldsymbol{\Gamma}}_i\tilde{\boldsymbol{\Gamma}}_i^T = \sum_{j=1}^{q_i}\tilde{\gamma}_{ij}\tilde{\gamma}_{ij}^T = \tilde{\boldsymbol{\Sigma}}_i$, $i = 1, 2$. Let $\hat{\boldsymbol{B}}_i = \hat{\boldsymbol{A}}_i - \boldsymbol{A}_i$ for $i = 1, 2$. We consider the eigen-decomposition of $\boldsymbol{A}_i$ by $\boldsymbol{A}_i = \boldsymbol{H}_{i(A)}\boldsymbol{\Lambda}_{i(A)}\boldsymbol{H}_{i(A)}^T$ for $i = 1, 2$, where $\boldsymbol{\Lambda}_{i(A)} =\mathrm{diag}(\lambda_{i1(A)}, ..., \lambda_{ip(A)})$ having eigenvalues such as $\lambda_{i1(A)} \geq \cdots \geq \lambda_{ip(A)} > 0$ and $\boldsymbol{H}_{i(A)} = [\boldsymbol{h}_{i1(A)}, ..., \boldsymbol{h}_{ip(A)}]$ is an orthogonal matrix of the corresponding eigenvectors. Let $a_{i(j)}$ be the j-th diagonal element of $\boldsymbol{A}_i$ for $j = 1, ..., p$ $(i = 1, 2)$. Let $x_{oijk} = x_{ijk} - \mu_{ij}$ for $j = 1, ..., p$ $(i = 1, 2;\ k = 1, ..., n_i)$.

B.1 Lemmas

In this section, in order to prove the theoretical results in [2], we give the following lemmas.

Lemma B.1. *Under (A-i), (C-iv) and (C-vi), we have that*

$$(\boldsymbol{x}_0 - \boldsymbol{\mu}_i)^T\{\boldsymbol{A}_i(\overline{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i) - \boldsymbol{A}_{i'}(\overline{\boldsymbol{x}}_{i'n_{i'}} - \boldsymbol{\mu}_{i'})\}/\omega_i \Rightarrow N(0, 1) \quad \text{as } m \to \infty$$

when $\boldsymbol{x}_0 \in \pi_i$ *for* $i = 1, 2;\ i' \neq i$.

Proof of Lemma B.1. We consider the case when $i = 1$ $(i' = 2)$ and $\boldsymbol{x}_0 \in \pi_1$. Let $\tilde{\boldsymbol{x}}_0 = \boldsymbol{A}_1^{1/2}(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)$. Then, it holds that $\mathrm{Var}(\tilde{\boldsymbol{x}}_0|\boldsymbol{x}_0 \in \pi_1) = \mathrm{Var}(\tilde{\boldsymbol{x}}_{1k}) = \tilde{\boldsymbol{\Sigma}}_1$. Let

$$v_k = \tilde{\boldsymbol{x}}_0^T\tilde{\boldsymbol{x}}_{1k}/(n_1\omega_1),\ k = 1, ..., n_1, \text{ and } v_{n_1+k} = -\tilde{\boldsymbol{x}}_0^T\tilde{\boldsymbol{x}}_{2k}/(n_2\omega_1),\ k = 1, ..., n_2.$$

Note that $\sum_{k=1}^{n_1+n_2}E(v_k^2) = 1$ and $\sum_{k=1}^{n_1+n_2}v_k = (\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T\{\boldsymbol{A}_1(\overline{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1) - \boldsymbol{A}_2(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2)\}/\omega_1$. Then, it holds that $E(v_k|v_{k-1}, ..., v_1) = 0$ for $k = 2, ..., n_1 + n_2$. Under (A-i), we can write that $\tilde{\boldsymbol{x}}_{1l} = \tilde{\boldsymbol{\Gamma}}_1\boldsymbol{y}_{1l}$ and $\tilde{\boldsymbol{x}}_{2l} = \tilde{\boldsymbol{\Gamma}}_2\boldsymbol{y}_{2l}$. Note that $2\omega_1/\delta_1 = 1 + o(1)$ under (C-vi). Then, in a way similar to the proof of Theorem 3 in Aoshima and Yata [1], by applying the martingale central limit theorem given by McLeish [9], we can obtain the result. $\square$

Lemma B.2. *Under (A-ii), (C-iv) and (C-vii), we have that*

$$2(\boldsymbol{x}_0-\boldsymbol{\mu}_i)^T\big\{\boldsymbol{A}_i(\overline{\boldsymbol{x}}_{in_i}-\boldsymbol{\mu}_i)-\boldsymbol{A}_{i'}(\overline{\boldsymbol{x}}_{i'n_{i'}}-\boldsymbol{\mu}_{i'}+(-1)^i\boldsymbol{\mu}_{12})\big\}/\delta_i \Rightarrow N(0,1)$$

as $m\to\infty$ when $\boldsymbol{x}_0\in\pi_i$ for $i=1,2;\ i'\neq i$.

Proof of Lemma B.2. We consider the case when $i=1$ $(i'=2)$ and $\boldsymbol{x}_0\in\pi_1$. Let $\boldsymbol{x}_0-\boldsymbol{\mu}_1=\boldsymbol{\Gamma}_1\boldsymbol{y}_0$ and $\boldsymbol{y}_0=(y_{01},...,y_{0q_1})^T$. Under (A-ii), $y_{0s},\ s=1,...,q_1$, are independent. Let $\acute{\boldsymbol{x}}_{ln_l}=\sum_{k=1}^{n_l}\tilde{\boldsymbol{x}}_{lk}/n_l$, $l=1,2$, $\tilde{\boldsymbol{\mu}}=\boldsymbol{A}_1^{-1/2}\boldsymbol{A}_2\boldsymbol{\mu}_{12}$ and

$$w_s=2y_{0j}\tilde{\boldsymbol{\gamma}}_{1s}^T(\acute{\boldsymbol{x}}_{1n_1}-\acute{\boldsymbol{x}}_{2n_2}+\tilde{\boldsymbol{\mu}})/\delta_1,\quad s=1,...,q_1.$$

Note that $q_1\geq p$, $E(w_s)=0$, $s=1,...,q_1$, $\sum_{s=1}^{q_1}E(w_s^2)=1$ and

$$\sum_{s=1}^{q_1}w_s=2(\boldsymbol{x}_0-\boldsymbol{\mu}_1)^T\{\boldsymbol{A}_1(\overline{\boldsymbol{x}}_{1n_1}-\boldsymbol{\mu}_1)-\boldsymbol{A}_2(\overline{\boldsymbol{x}}_{2n_2}-\boldsymbol{\mu}_2-\boldsymbol{\mu}_{12})\}/\delta_1.$$

Also, note that $E(w_s|w_{s-1},...,w_1)=0$ for $s=2,...,q_1$, under (A-ii). We consider applying the martingale central limit theorem. Let $M_{ls}=E(y_{lsk}^3)$ for all l,s. Note that $\limsup_{p\to\infty}|M_{ls}|<\infty$ for all l,s, under (A-ii) because $\limsup_{p\to\infty}E(y_{lsk}^4)<\infty$. Then, by using Schwarz's inequality and the arithmetic mean-geometric mean inequality, we can evaluate that under (A-ii)

$$\begin{aligned}
&E\{(\tilde{\boldsymbol{\gamma}}_{1s}^T\acute{\boldsymbol{x}}_{ln_l})^2(\tilde{\boldsymbol{\gamma}}_{1t}^T\acute{\boldsymbol{x}}_{ln_l})^2\}=\{1+o(1)\}\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1s}\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1t}/n_l^2+O\{(\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1t}/n_l)^2\};\ \text{and}\\
&|E\{(\tilde{\boldsymbol{\gamma}}_{1s}^T\acute{\boldsymbol{x}}_{ln_l})^2\tilde{\boldsymbol{\gamma}}_{1t}^T\acute{\boldsymbol{x}}_{ln_l}\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\mu}}\}|=\Big|\sum_{u=1}^{q_l}(\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\gamma}}_{lu})^2\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\gamma}}_{lu}\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\mu}}M_{lu}/n_l^2\Big|\\
&\leq\{E(\tilde{\boldsymbol{\gamma}}_{1s}^T\acute{\boldsymbol{x}}_{ln_l})^4\}^{1/2}\{E(\tilde{\boldsymbol{\gamma}}_{1t}^T\acute{\boldsymbol{x}}_{ln_l}\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\mu}})^2\}^{1/2}\\
&=O\{\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1s}(\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1t}/n_l)^{1/2}|\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\mu}}|/n_l\}\\
&=O[\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1s}\{\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1t}/n_l+(\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\mu}})^2\}/n_l]\\
&=O[\{\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1s}/n_l\}^2+\{\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1t}/n_l\}^2+(\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\mu}})^4],\ l=1,2
\end{aligned}$$

for all s,t. Then, we have that for all s,t

$$\delta_1^4E(w_s^4)=O\Big[\sum_{l=1}^{2}\{\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1s}/n_l\}^2+(\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\mu}})^4\Big]\quad\text{and}\tag{B.1}$$

$$\begin{aligned}
&(\delta_1/2)^4\frac{E(w_s^2w_t^2)}{E(y_{0s}^2y_{0t}^2)}-\tilde{\boldsymbol{\gamma}}_{1s}^T\Big(\sum_{l=1}^{2}\tilde{\boldsymbol{\Sigma}}_l/n_l+\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T\Big)\tilde{\boldsymbol{\gamma}}_{1s}\tilde{\boldsymbol{\gamma}}_{1t}^T\Big(\sum_{l=1}^{2}\tilde{\boldsymbol{\Sigma}}_l/n_l+\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T\Big)\tilde{\boldsymbol{\gamma}}_{1t}\\
&=2\sum_{l=1}^{2}(-1)^{l+1}\sum_{u=1}^{q_l}\{(\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\gamma}}_{lu})^2\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\gamma}}_{lu}\tilde{\boldsymbol{\gamma}}_{1t}^T+(\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\gamma}}_{lu})^2\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\gamma}}_{lu}\tilde{\boldsymbol{\gamma}}_{1s}^T\}\tilde{\boldsymbol{\mu}}M_{lu}/n_l^2\\
&\quad+o\Big[\sum_{l=1}^{2}\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1s}\tilde{\boldsymbol{\gamma}}_{1t}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1t}/n_l^2\Big]+O\Big[\sum_{l=1}^{2}(\tilde{\boldsymbol{\gamma}}_{1s}^T\tilde{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{\gamma}}_{1t}/n_l)^2\Big].
\end{aligned}\tag{B.2}$$

Here, under (C-iv), we can evaluate that

$$\sum_{s,t=1}^{q_1}\sum_{u=1}^{q_l}(\tilde{\gamma}_{1s}^T\tilde{\gamma}_{lu})^2\tilde{\gamma}_{1t}^T\tilde{\gamma}_{lu}\tilde{\gamma}_{1t}^T\tilde{\boldsymbol{\mu}}M_{lu}/n_l^2=\sum_{u=1}^{q_l}\tilde{\gamma}_{lu}^T\tilde{\boldsymbol{\Sigma}}_1\tilde{\gamma}_{lu}\tilde{\gamma}_{lu}^T\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\mu}}M_{lu}/n_l^2$$
$$=O\Big[||\tilde{\boldsymbol{\mu}}^T\tilde{\boldsymbol{\Sigma}}_1^{1/2}||\sum_{u=1}^{q_l}||\tilde{\gamma}_{lu}^T\tilde{\boldsymbol{\Sigma}}_1^{1/2}||\tilde{\gamma}_{lu}^T\tilde{\boldsymbol{\Sigma}}_1\tilde{\gamma}_{lu}/n_l^2\Big]$$
$$=O\Big[||\tilde{\boldsymbol{\mu}}^T\tilde{\boldsymbol{\Sigma}}_1^{1/2}||\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_l)^{1/2}\Big\{\sum_{u=1}^{q_l}(\tilde{\gamma}_{lu}^T\tilde{\boldsymbol{\Sigma}}_1\tilde{\gamma}_{lu})^2\Big\}^{1/2}/n_l^2\Big]$$
$$=O\big[\{\tilde{\boldsymbol{\mu}}^T\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\mu}}+\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_l)\}\mathrm{tr}\{(\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_l)^2\}^{1/2}/n_l^2\Big]=o(\delta_1^4),\quad l=1,2 \tag{B.3}$$

from the fact that $\sum_{u=1}^{q_l}(\tilde{\gamma}_{lu}^T\tilde{\boldsymbol{\Sigma}}_1\tilde{\gamma}_{lu})^2\le\sum_{u,u'=1}^{q_l}(\tilde{\gamma}_{lu}^T\tilde{\boldsymbol{\Sigma}}_1\tilde{\gamma}_{lu'})^2=\mathrm{tr}\{(\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_l)^2\}=o(n_l^2\delta_1^4)$ under (C-iv). Then, by combining (B.1) and (B.2) with (B.3), under (A-ii), (C-iv) and (C-vii), for any $\tau>0$, we have that as $m\to\infty$

$$\sum_{s=1}^{q_1}\frac{E(w_s^4)}{\tau}=O\Big[\frac{\sum_{l=1}^2\mathrm{tr}\{(\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_l)^2\}/n_l^2+\sum_{s=1}^{q_1}(\tilde{\gamma}_{1s}^T\tilde{\boldsymbol{\mu}})^4}{\delta_1^4}\Big]\to 0\ \text{ and}$$
$$P\Big(\Big|\sum_{s=1}^{q_1}w_s^2-1\Big|\ge\tau\Big)\le\frac{\sum_{s,t=1}^{q_1}E(w_s^2w_t^2)-1}{\tau^2}=O\Big[\sum_{s=1}^{q_1}E(w_s^4)\Big]+o(1)\to 0,$$

so that $\sum_{s=1}^{q_1}E\{w_s^2I(w_s^2\ge\tau)\}\le\sum_{s=1}^{q_1}E(w_s^4)/\tau\to 0$ and $\sum_{s=1}^{q_1}w_s^2=1+o_P(1)$. Hence, by using the martingale central limit theorem, we obtain that $\sum_{s=1}^{q_1}w_s\Rightarrow N(0,1)$ as $m\to\infty$ under (A-ii), (C-iv) and (C-vii). We conclude the result when $i=1$. For the case when $i=2$, we can have the same arguments. The proof is completed. □

Lemma B.3. *Assume that when $\boldsymbol{x}_0\in\pi_i$ for $i=1,2$,*

$$tr[\{(\boldsymbol{x}_0-\boldsymbol{\mu}_i)(\boldsymbol{x}_0-\boldsymbol{\mu}_i)^T-\boldsymbol{\Sigma}_i\}(\hat{\boldsymbol{B}}_1-\hat{\boldsymbol{B}}_2)]=o_P(\kappa); \tag{B.4}$$
$$tr\{\boldsymbol{\Sigma}_i(\hat{\boldsymbol{B}}_1-\hat{\boldsymbol{B}}_2)\}-\log|\hat{\boldsymbol{A}}_1\boldsymbol{A}_1^{-1}|+\log|\hat{\boldsymbol{A}}_2\boldsymbol{A}_2^{-1}|=o_P(\kappa);\ \ and \tag{B.5}$$
$$\{2(\boldsymbol{x}_0-\boldsymbol{\mu}_i)+(-1)^{i+1}\boldsymbol{\mu}_{12}\}^T\hat{\boldsymbol{B}}_{i'}\boldsymbol{\mu}_{12}=o_P(\kappa)\ (i'\neq i) \tag{B.6}$$
$$and\ \ (p/n_l^{1/2})||\hat{\boldsymbol{B}}_l||=o_P(\kappa),\quad l=1,2,$$

where $\kappa=\Delta_{\min}$ or $\kappa=\delta_{\min}$. Then, (15) holds.

Proof of Lemma B.3. We consider the case when $\boldsymbol{x}_0\in\pi_1$. We have that

$$\begin{aligned}&W_1(\hat{\boldsymbol{A}}_1)-W_1(\boldsymbol{A}_1)-W_2(\hat{\boldsymbol{A}}_2)+W_2(\boldsymbol{A}_2)\\&=\mathrm{tr}[\{(\boldsymbol{x}_0-\boldsymbol{\mu}_1)(\boldsymbol{x}_0-\boldsymbol{\mu}_1)^T-\boldsymbol{\Sigma}_1\}(\hat{\boldsymbol{B}}_1-\hat{\boldsymbol{B}}_2)]\\&\quad+\mathrm{tr}\{\boldsymbol{\Sigma}_1(\hat{\boldsymbol{B}}_1-\hat{\boldsymbol{B}}_2)\}-\log|\hat{\boldsymbol{A}}_1\boldsymbol{A}_1^{-1}|+\log|\hat{\boldsymbol{A}}_2\boldsymbol{A}_2^{-1}|\\&\quad+\sum_{l=1}^2(-1)^{l+1}\mathrm{tr}[\{2(\boldsymbol{x}_0-\boldsymbol{\mu}_1-(\overline{\boldsymbol{x}}_{ln_l}-\boldsymbol{\mu}_1)/2)(\boldsymbol{\mu}_1-\overline{\boldsymbol{x}}_{ln_l})^T-\boldsymbol{S}_{ln_l}/n_l\}\hat{\boldsymbol{B}}_l].\end{aligned}$$

Note that $\mathrm{tr}(\boldsymbol{S}_{ln_l})=O_P(p)$, $||\overline{\boldsymbol{x}}_{ln_l}-\boldsymbol{\mu}_1||^2\le||\overline{\boldsymbol{x}}_{ln_l}-\boldsymbol{\mu}_l||^2+||\boldsymbol{\mu}_l-\boldsymbol{\mu}_1||^2=||\boldsymbol{\mu}_l-\boldsymbol{\mu}_1||^2+O_P(p/n_l)$ and $||\boldsymbol{x}_0-\boldsymbol{\mu}_1-(\overline{\boldsymbol{x}}_{ln_l}-\boldsymbol{\mu}_1)/2||^2\le||\boldsymbol{x}_0-\boldsymbol{\mu}_1||^2+||\overline{\boldsymbol{x}}_{ln_l}-\boldsymbol{\mu}_l||^2+||\boldsymbol{\mu}_1-\boldsymbol{\mu}_l||^2=O_P(p)$, $l=1,2$, from the facts that $E(||\boldsymbol{x}_0 - \boldsymbol{\mu}_1||^2) = \mathrm{tr}(\boldsymbol{\Sigma}_1)$, $E\{\mathrm{tr}(\boldsymbol{S}_{ln_l})\} = \mathrm{tr}(\boldsymbol{\Sigma}_l)$, $E(||\overline{\boldsymbol{x}}_{ln_l} - \boldsymbol{\mu}_l||^2) = \mathrm{tr}(\boldsymbol{\Sigma}_l)/n_l$, $\mathrm{tr}(\boldsymbol{\Sigma}_i) = O(p)$, $i = 1, 2$, and $||\boldsymbol{\mu}_{12}||^2 = O(p)$. Then, we have that for $l = 1, 2$

$$
\begin{aligned}
&|\mathrm{tr}[\{2(\boldsymbol{x}_0 - \boldsymbol{\mu}_1 - (\overline{\boldsymbol{x}}_{ln_l} - \boldsymbol{\mu}_1)/2)(\boldsymbol{\mu}_l - \overline{\boldsymbol{x}}_{ln_l})^T - \boldsymbol{S}_{ln_l}/n_l\}\hat{\boldsymbol{B}}_l]| \\
&\le 2||\boldsymbol{x}_0 - \boldsymbol{\mu}_1 - (\overline{\boldsymbol{x}}_{ln_l} - \boldsymbol{\mu}_1)/2|| \cdot ||\overline{\boldsymbol{x}}_{ln_l} - \boldsymbol{\mu}_l|| \cdot ||\hat{\boldsymbol{B}}_l|| + \mathrm{tr}(\boldsymbol{S}_{ln_l})||\hat{\boldsymbol{B}}_l||/n_l = O_P\{(p/n_l^{1/2})||\hat{\boldsymbol{B}}_l||\}.
\end{aligned}
$$

Also, we have that $|(\overline{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2)^T \hat{\boldsymbol{B}}_2 \boldsymbol{\mu}_{12}| = O_P\{(p/n_2^{1/2})||\hat{\boldsymbol{B}}_2||\}$. Thus it holds that

$$
\begin{aligned}
&\sum_{l=1}^{2} (-1)^{l+1} \mathrm{tr}[\{2(\boldsymbol{x}_0 - \boldsymbol{\mu}_1 - (\overline{\boldsymbol{x}}_{ln_l} - \boldsymbol{\mu}_1)/2)(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}}_{ln_l})^T - \boldsymbol{S}_{ln_l}/n_l\}\hat{\boldsymbol{B}}_l] \\
&= -\{2(\boldsymbol{x}_0 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_{12}\}^T \hat{\boldsymbol{B}}_2 \boldsymbol{\mu}_{12} + O_P\{(p/n_1^{1/2})||\hat{\boldsymbol{B}}_1|| + (p/n_2^{1/2})||\hat{\boldsymbol{B}}_2||\}.
\end{aligned}
$$

Hence, it concludes the result when $\boldsymbol{x}_0 \in \pi_1$. For the case when $\boldsymbol{x}_0 \in \pi_2$, we can have the same arguments. The proof is completed. □

B.2 Proofs of the theoretical results

In this section, we give proofs of the theoretical results in the main work in [2].

Proof of Proposition 1. We can write that $\mathrm{tr}(\boldsymbol{A}_i^{-1}\boldsymbol{A}_{i'}) = \sum_{j=1}^{p} \boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)}/\lambda_{ij(A)}$. Note that $\sum_{j=1}^{p} \boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)} = \mathrm{tr}(\boldsymbol{A}_{i'})$ and $\sum_{j=1}^{t} (\boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)} - \lambda_{i'j(A)}) \le 0$ for any $t \in \{1, ..., p\}$. Then, by noting that $\lambda_{i1(A)} \ge \cdots \ge \lambda_{ip(A)} > 0$, we have that

$$
\begin{aligned}
\mathrm{tr}(\boldsymbol{A}_i^{-1}\boldsymbol{A}_{i'}) &= \frac{\lambda_{i'1(A)}}{\lambda_{i1(A)}} + \frac{\boldsymbol{h}_{i1(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{i1(A)} - \lambda_{i'1(A)}}{\lambda_{i1(A)}} + \sum_{j=2}^{p} \frac{\boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)}}{\lambda_{ij(A)}} \\
&\ge \sum_{j=1}^{2} \frac{\lambda_{i'j(A)}}{\lambda_{ij(A)}} + \sum_{j=1}^{2} \frac{\boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)} - \lambda_{i'j(A)}}{\lambda_{i2(A)}} + \sum_{j=3}^{p} \frac{\boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)}}{\lambda_{ij(A)}} \\
&\ \vdots \\
&\ge \sum_{j=1}^{p} \frac{\lambda_{i'j(A)}}{\lambda_{ij(A)}} + \sum_{j=1}^{p} \frac{\boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)} - \lambda_{i'j(A)}}{\lambda_{ip(A)}} = \sum_{j=1}^{p} \frac{\lambda_{i'j(A)}}{\lambda_{ij(A)}}. \qquad \text{(B.7)}
\end{aligned}
$$

Thus, when $\mathrm{tr}\{\boldsymbol{\Sigma}_i(\boldsymbol{A}_{i'} - \boldsymbol{A}_i)\} = \mathrm{tr}(\boldsymbol{A}_i^{-1}\boldsymbol{A}_{i'}) - p$, it holds that

$$
\Delta_i \ge \sum_{j=1}^{p} \{\lambda_{i'j(A)}/\lambda_{ij(A)} - 1 + \log(\lambda_{ij(A)}/\lambda_{i'j(A)})\} \ge 0
$$

from the fact that $c - 1 + \log c^{-1} \ge 0$ for any positive constant c. Note that $\lambda_{1j(A)} \ne \lambda_{2j(A)}$ or $\boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)} < \lambda_{i'j(A)}$ for some j when $\boldsymbol{A}_1 \ne \boldsymbol{A}_2$. Since $c - 1 + \log c^{-1} > 0$ when $c \ne 1$, it holds that $\Delta_i > 0$ when $\lambda_{1j(A)} \ne \lambda_{2j(A)}$ for some j. From (B.7), if $\boldsymbol{h}_{ij(A)}^T \boldsymbol{A}_{i'} \boldsymbol{h}_{ij(A)} < \lambda_{i'j(A)}$ for some j, it follows that $\mathrm{tr}(\boldsymbol{A}_i^{-1}\boldsymbol{A}_{i'}) > \sum_{j=1}^{p} (\lambda_{i'j(A)}/\lambda_{ij(A)})$, so that $\Delta_i > 0$. When $\boldsymbol{\mu}_1 \ne \boldsymbol{\mu}_2$, it holds that $\Delta_i \ge \boldsymbol{\mu}_{12}^T \boldsymbol{A}_{i'} \boldsymbol{\mu}_{12} > 0$. Hence, it concludes the results. □

Proof of Theorem 1. We consider the case when $\boldsymbol{x}_0 \in \pi_1$. Under (C-i) and (C-ii), it holds that for $i = 1, 2$

$$\begin{aligned} &\mathrm{Var}\{(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T \boldsymbol{A}_i(\overline{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i)\} = \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{A}_i \boldsymbol{\Sigma}_1 \boldsymbol{A}_i)/n_i = o(\Delta_1^2) \\ &\text{and } \mathrm{Var}\{(\boldsymbol{x}_0 - \boldsymbol{\mu}_1 - \overline{\boldsymbol{x}}_{2n_2} + \boldsymbol{\mu}_2)^T \boldsymbol{A}_2 \boldsymbol{\mu}_{12}\} = \boldsymbol{\mu}_{12}^T \boldsymbol{A}_2(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2/n_2)\boldsymbol{A}_2 \boldsymbol{\mu}_{12} = o(\Delta_1^2) \end{aligned} \tag{B.8}$$

from the fact that

$$\boldsymbol{\mu}_{12}^T \boldsymbol{A}_2 \boldsymbol{\Sigma}_2 \boldsymbol{A}_2 \boldsymbol{\mu}_{12} \le \boldsymbol{\mu}_{12}^T \boldsymbol{A}_2 \boldsymbol{\mu}_{12} \lambda_{\max}(\boldsymbol{A}_2^{1/2} \boldsymbol{\Sigma}_2 \boldsymbol{A}_2^{1/2}) \le \Delta_1 \mathrm{tr}\{(\boldsymbol{\Sigma}_2 \boldsymbol{A}_2)^2\}^{1/2} = o(n_2 \Delta_1^2)$$

under (C-i). Note that $(\overline{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i)^T \boldsymbol{A}_i (\overline{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i) - \mathrm{tr}(\boldsymbol{A}_i \boldsymbol{S}_{in_i})/n_i = \sum_{k \neq k'}^{n_i} (\boldsymbol{x}_{ik} - \boldsymbol{\mu}_i)^T \boldsymbol{A}_i (\boldsymbol{x}_{ik'} - \boldsymbol{\mu}_i)/\{n_i(n_i - 1)\}$. Then, under (C-i) it follows that for $i = 1, 2$

$$\mathrm{Var}\{(\overline{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i)^T \boldsymbol{A}_i (\overline{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i) - \mathrm{tr}(\boldsymbol{A}_i \boldsymbol{S}_{in_i})/n_i\} = O[\mathrm{tr}\{(\boldsymbol{\Sigma}_i \boldsymbol{A}_i)^2\}/n_i^2] = o(\Delta_1^2). \tag{B.9}$$

Then, by using Chebyshev's inequality, from (B.8) and (B.9), we find that

$$W_2(\boldsymbol{A}_2) - W_1(\boldsymbol{A}_1) = \mathrm{tr}[\{(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T - \boldsymbol{\Sigma}_1\}(\boldsymbol{A}_2 - \boldsymbol{A}_1)] + \Delta_1 + o_P(\Delta_1). \tag{B.10}$$

Here, under (A-i) and (C-iii), it follows that

$$\mathrm{Var}\big(\mathrm{tr}[\{(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T - \boldsymbol{\Sigma}_1\}(\boldsymbol{A}_2 - \boldsymbol{A}_1)]\big) = O\big(\mathrm{tr}[\{\boldsymbol{\Sigma}_1(\boldsymbol{A}_2 - \boldsymbol{A}_1)\}^2]\big) = o(\Delta_1^2). \tag{B.11}$$

Thus by combining (B.10) with (B.11), under (A-i) and (C-i) to (C-iii), we obtain that $\{W_2(\boldsymbol{A}_2) - W_1(\boldsymbol{A}_1)\}/\Delta_1 = 1 + o_P(1)$, so that $P\{W_2(\boldsymbol{A}_2) - W_1(\boldsymbol{A}_1) > 0\} \to 1$. When $\boldsymbol{x}_0 \in \pi_2$, we have the same arguments. The proof is completed. □

Proof of Proposition 2. We note that

$$\begin{aligned} &\Delta_{iA} \le \boldsymbol{\mu}_{12}^T \boldsymbol{A}_{i'} \boldsymbol{\mu}_{12} \lambda_{\max}(\boldsymbol{A}_{i'}^{1/2} \boldsymbol{\Sigma}_i \boldsymbol{A}_{i'}^{1/2}) \le \Delta_i \lambda_{\max}(\boldsymbol{A}_{i'}^{1/2} \boldsymbol{\Sigma}_i \boldsymbol{A}_{i'}^{1/2}) \\ &\text{and } \ \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{A}_{i'} \boldsymbol{\Sigma}_{i'} \boldsymbol{A}_{i'}) \le \mathrm{tr}\{(\boldsymbol{\Sigma}_i \boldsymbol{A}_{i'})^2\}^{1/2} \mathrm{tr}\{(\boldsymbol{\Sigma}_{i'} \boldsymbol{A}_{i'})^2\}^{1/2}. \end{aligned} \tag{B.12}$$

When $\limsup_{p \to \infty} \lambda_{i1(A)} < \infty$, $i = 1, 2$, it holds that

$$\begin{aligned} &\lambda_{\max}(\boldsymbol{A}_{i'}^{1/2} \boldsymbol{\Sigma}_i \boldsymbol{A}_{i'}^{1/2}) \le \lambda_{i1} \lambda_{\max}(\boldsymbol{A}_{i'}) = \lambda_{i1} \lambda_{i'1(A)} = O(\lambda_{i1}) \ \text{ and} \\ &\mathrm{tr}\{(\boldsymbol{\Sigma}_l \boldsymbol{A}_{l'})^2\} \le \mathrm{tr}(\boldsymbol{\Sigma}_l \boldsymbol{A}_{l'} \boldsymbol{\Sigma}_l) \lambda_{l'1(A)} \le \mathrm{tr}(\boldsymbol{\Sigma}_l^2) \lambda_{l'1(A)}^2 = O\{\mathrm{tr}(\boldsymbol{\Sigma}_l^2)\} \end{aligned} \tag{B.13}$$

for all l, l'. By combining (B.12) with (B.13), (C-i') and (C-ii') imply (C-i) and (C-ii).

Next, for (C-iii), it holds that $\mathrm{tr}[\{\boldsymbol{\Sigma}_i(\boldsymbol{A}_1 - \boldsymbol{A}_2)\}^2] \le \lambda_{i1} \mathrm{tr}\{(\boldsymbol{A}_1 - \boldsymbol{A}_2)\boldsymbol{\Sigma}_i(\boldsymbol{A}_1 - \boldsymbol{A}_2)\}$. When $\boldsymbol{A}_i$s are diagonal matrices such as $\boldsymbol{A}_i = \mathrm{diag}(a_{i(1)}, ..., a_{i(p)})$, $i = 1, 2$, it holds that $\Delta_i \ge \sum_{j=1}^p \{a_{i'(j)}/a_{i(j)} - 1 - \log(a_{i'(j)}/a_{i(j)})\}$ and $\mathrm{tr}\{(\boldsymbol{A}_1 - \boldsymbol{A}_2)\boldsymbol{\Sigma}_i(\boldsymbol{A}_1 - \boldsymbol{A}_2)\} = \sum_{j=1}^p \sigma_{i(j)}(a_{1(j)} - a_{2(j)})^2$. Note that $a_{i(j)} \in (0, \infty)$ as $p \to \infty$ for all i, j, under $\lambda(\boldsymbol{A}_i) \in (0, \infty)$ as $p \to \infty$ for $i = 1, 2$. By Taylor expansion, we claim that

$$a_{i'(j)}/a_{i(j)} - 1 - \log(a_{i'(j)}/a_{i(j)}) \ge a_{i(j)}^{-2}(a_{1(j)} - a_{2(j)})^2/(2\max\{1, a_{i'(j)}^2/a_{i(j)}^2\}).$$

Then, it follows that $\sum_{j=1}^p \sigma_{i(j)}(a_{1(j)} - a_{2(j)})^2 = O(\Delta_i)$ because $\sigma_{i(j)} \in (0, \infty)$ as $p \to \infty$ for all i, j. Thus we have that $\mathrm{tr}[\{\boldsymbol{\Sigma}_i(\boldsymbol{A}_1 - \boldsymbol{A}_2)\}^2] = O(\Delta_i \lambda_{i1})$. It concludes the results. □

Proof of Corollary 1. From Theorem 1 and Proposition 2, we can claim Corollary 1 straightforwardly. □

Proof of Proposition 3. We first consider the case when $\liminf_{p\to\infty}\sum_{j=1}^{p}|\lambda_{ij}/\lambda_{i'j}-1|/p>0$. When $c_{1j}<|\lambda_{ij}/\lambda_{i'j}-1|<c_{2j}$ for some constants c_{1j} (>0) and c_{2j} $(<\infty)$, by Taylor expansion, it holds that

$$\lambda_{ij}/\lambda_{i'j}-1-\log(\lambda_{ij}/\lambda_{i'j})\geq\frac{(\lambda_{ij}/\lambda_{i'j}-1)^2}{2\max\{1,\lambda_{ij}^2/\lambda_{i'j}^2\}}\geq\frac{c_{1j}|\lambda_{ij}/\lambda_{i'j}-1|}{2(c_{2j}+1)^2}.$$

When $\lambda_{ij}/\lambda_{i'j}\to\infty$ as $p\to\infty$, it holds that for sufficiently large p

$$\lambda_{ij}/\lambda_{i'j}-1-\log(\lambda_{ij}/\lambda_{i'j})>|\lambda_{ij}/\lambda_{i'j}-1|/2.$$

Thus, when $\liminf_{p\to\infty}\sum_{j=1}^{p}|\lambda_{ij}/\lambda_{i'j}-1|/p>0$, it follows that

$$\liminf_{p\to\infty}\Delta_{i(IV)}/p\geq\liminf_{p\to\infty}\sum_{j=1}^{p}\{\lambda_{ij}/\lambda_{i'j}-1-\log(\lambda_{ij}/\lambda_{i'j})\}/p>0$$

from (B.7).

Next, we consider the case when $\liminf_{p\to\infty}|\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})/p-1|>0$. We note that $\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})\geq\sum_{j=1}^{p}\lambda_{ij}/\lambda_{i'j}$ from (B.7). When $\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})/(\sum_{j=1}^{p}\lambda_{ij}/\lambda_{i'j})\to 1$ as $p\to\infty$, it holds that

$$\liminf_{p\to\infty}\Big|\sum_{j=1}^{p}(\lambda_{ij}/\lambda_{i'j})/p-1\Big|>0$$

under $\liminf_{p\to\infty}|\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})/p-1|>0$. It follows that $\liminf_{p\to\infty}\Delta_{i(IV)}/p>0$ from the fact that $\sum_{j=1}^{p}|\lambda_{ij}/\lambda_{i'j}-1|/p\geq|\sum_{j=1}^{p}(\lambda_{ij}/\lambda_{i'j})/p-1|$. On the other hand, we note that

$$\Delta_{i(IV)}\geq\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})-p-\sum_{j=1}^{p}\log(\lambda_{ij}/\lambda_{i'j})\geq\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})-\sum_{j=1}^{p}(\lambda_{ij}/\lambda_{i'j})$$

because $\sum_{j=1}^{p}\{\lambda_{ij}/\lambda_{i'j}-1-\log(\lambda_{ij}/\lambda_{i'j})\}\geq 0$. Thus, when $\sum_{j=1}^{p}(\lambda_{ij}/\lambda_{i'j})/p-1\to 0$ as $p\to\infty$ and $\liminf_{p\to\infty}\{\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})/(\sum_{j=1}^{p}\lambda_{ij}/\lambda_{i'j})\}>1$, we have that $\liminf_{p\to\infty}\Delta_{i(IV)}/p>0$. Hence, it concludes the results. □

Proof of Theorem 2. Note that $\mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{A}_i)^2\}/n_i^2=o(\delta_i^2)$, $i=1,2$. Then, similar to (B.8) to (B.11), under (A-i) and (C-iv) to (C-vi), we have that as $m\to\infty$

$$W_{i'}(\boldsymbol{A}_{i'})-W_i(\boldsymbol{A}_i)-\Delta_i=2(\boldsymbol{x}_0-\boldsymbol{\mu}_i)^T\{\boldsymbol{A}_i(\overline{\boldsymbol{x}}_{in_i}-\boldsymbol{\mu}_i)-\boldsymbol{A}_{i'}(\overline{\boldsymbol{x}}_{i'n_{i'}}-\boldsymbol{\mu}_{i'})\}+o_P(\delta_i) \tag{B.14}$$

when $\boldsymbol{x}_0\in\pi_i$ $(i'\neq i)$. Note that $2\omega_i/\delta_i\to 1$ as $m\to\infty$ for $i=1,2$, under (C-vi). Then, by combining Lemma B.1 with (B.14), we conclude the results. □

Proof of Theorem 3. Similar to (B.14), under (A-i), (C-iv) and (C-v), we have that as $m\to\infty$

$$\begin{aligned}W_{i'}(\boldsymbol{A}_{i'})-W_i(\boldsymbol{A}_i)-\Delta_i=&2(\boldsymbol{x}_0-\boldsymbol{\mu}_i)^T\{\boldsymbol{A}_i(\overline{\boldsymbol{x}}_{in_i}-\boldsymbol{\mu}_i)\\&-\boldsymbol{A}_{i'}(\overline{\boldsymbol{x}}_{i'n_{i'}}-\boldsymbol{\mu}_{i'}+(-1)^i\boldsymbol{\mu}_{12})\}+o_P(\delta_i)\end{aligned} \tag{B.15}$$

when $\boldsymbol{x}_0\in\pi_i$ $(i'\neq i)$. Then, by combining Lemma B.2 with (B.15), we conclude the results. □

Proof of Corollary 2. From Theorems 2 and 3, we can claim Corollary 2 straightforwardly. □

Proofs of Propositions 4 and 5. We consider the case when $\boldsymbol{x}_0 \in \pi_1$. Similar to Proof of Lemma B.3, we can claim that $|\{2(\boldsymbol{x}_0 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_{12}\}^T \hat{\boldsymbol{B}}_2 \boldsymbol{\mu}_{12}| \le ||2(\boldsymbol{x}_0 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_{12}|| \cdot ||\boldsymbol{\mu}_{12}|| \cdot ||\hat{\boldsymbol{B}}_2|| = O_P(p^{1/2}||\boldsymbol{\mu}_{12}|| \cdot ||\hat{\boldsymbol{B}}_2||) = O_P(p||\hat{\boldsymbol{B}}_2||)$ because $||\boldsymbol{\mu}_{12}||^2 = O(p)$ and $||2(\boldsymbol{x}_0 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_{12}||^2 = O_P(p)$. Thus, (B.6) holds under (C-viii) or (C-ix). Note that (B.4) and (B.5) naturally hold when $\hat{\boldsymbol{A}}_1 = \hat{\boldsymbol{A}}_2$ and $\boldsymbol{A}_1 = \boldsymbol{A}_2$. Hence, from Lemma B.3, it concludes the result of Proposition 5 when $\boldsymbol{x}_0 \in \pi_1$.

Next, we consider (B.4) and the first term of (B.5). We have that for $l = 1, 2$

$$\begin{aligned}
&|\mathrm{tr}(\boldsymbol{\Sigma}_1 \hat{\boldsymbol{B}}_l)| \le \mathrm{tr}(\boldsymbol{\Sigma}_1)||\hat{\boldsymbol{B}}_l|| = O_P(p||\hat{\boldsymbol{B}}_l||) \text{ and} \\
&|\mathrm{tr}[\{(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)(\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T - \boldsymbol{\Sigma}_1\}\hat{\boldsymbol{B}}_l]| \le ||\boldsymbol{x}_0 - \boldsymbol{\mu}_1||^2||\hat{\boldsymbol{B}}_l|| + \mathrm{tr}(\boldsymbol{\Sigma}_1)||\hat{\boldsymbol{B}}_l|| = O_P(p||\hat{\boldsymbol{B}}_l||).
\end{aligned}$$

Finally, we consider $\log|\hat{\boldsymbol{A}}_l \boldsymbol{A}_l^{-1}|, l = 1, 2$, in (B.5). Let $\boldsymbol{e}_p$ be an arbitrary (random) p-vector such that $||\boldsymbol{e}_p|| = 1$. Note that $||\boldsymbol{e}_p^T \boldsymbol{A}_l^{-1/2}|| \in (0, \infty)$ as $p \to \infty$ under $\lambda(\boldsymbol{A}_l) \in (0, \infty)$ as $p \to \infty$. Thus we have that

$$\boldsymbol{e}_p^T \boldsymbol{A}_l^{-1/2} \hat{\boldsymbol{B}}_l \boldsymbol{A}_l^{-1/2} \boldsymbol{e}_p = \boldsymbol{e}_p^T \boldsymbol{A}_l^{-1/2} \hat{\boldsymbol{A}}_l \boldsymbol{A}_l^{-1/2} \boldsymbol{e}_p - 1 = O_P(||\hat{\boldsymbol{B}}_l||),$$

so that $\lambda_{\min}(\boldsymbol{A}_l^{-1/2} \hat{\boldsymbol{A}}_l \boldsymbol{A}_l^{-1/2}) - 1 = O_P(||\hat{\boldsymbol{B}}_l||)$ and $\lambda_{\max}(\boldsymbol{A}_l^{-1/2} \hat{\boldsymbol{A}}_l \boldsymbol{A}_l^{-1/2}) - 1 = O_P(||\hat{\boldsymbol{B}}_l||)$. Hence, under $||\hat{\boldsymbol{B}}_l|| = o_P(1)$, it holds that for $l = 1, 2$

$$\log|\hat{\boldsymbol{A}}_l \boldsymbol{A}_l^{-1}| = \log|\boldsymbol{A}_l^{-1/2} \hat{\boldsymbol{A}}_l \boldsymbol{A}_l^{-1/2}| = O_P(p||\hat{\boldsymbol{B}}_l||).$$

Note that $\Delta_{\min} = O(p)$ and $\delta_{\min} = O(p)$ under $\lambda(\boldsymbol{A}_i) \in (0, \infty)$ as $p \to \infty$ for $i = 1, 2$. Then, under (C-viii), it holds that $||\hat{\boldsymbol{B}}_l|| = o_P(1)$ for $l = 1, 2$. Hence, (C-viii) implies (B.4) and (B.5). It concludes the result of Proposition 4 when $\boldsymbol{x}_0 \in \pi_1$. For the case when $\boldsymbol{x}_0 \in \pi_2$, we can have the same arguments. The proof is completed. □

Proof of Corollary 3. Under (A-i) we have that $\mathrm{Var}\{\mathrm{tr}(\boldsymbol{S}_{in_i})\} = O(\mathrm{tr}(\boldsymbol{\Sigma}_i^2)/n_i)$, $i = 1, 2$, so that $\mathrm{tr}(\boldsymbol{S}_{in_i}) = \mathrm{tr}(\boldsymbol{\Sigma}_i) + O_P\{(\mathrm{tr}(\boldsymbol{\Sigma}_i^2)/n_i)^{1/2}\}$. Then, under (C-i') it holds that $\mathrm{tr}(\boldsymbol{S}_{in_i}) = \mathrm{tr}(\boldsymbol{\Sigma}_i) + o_P(\Delta_{\min(II)}) = \mathrm{tr}(\boldsymbol{\Sigma}_i)\{1 + o_P(1)\}$ and $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)/(n_i p^2) = o(\Delta_{\min(II)}^2/p^2) = o(1)$ for $i = 1, 2$, because $\Delta_{\min(II)} = O(p)$. Thus, we have that under (A-i) and (C-i')

$$\begin{aligned}
||\hat{\boldsymbol{B}}_i|| = ||\{p/\mathrm{tr}(\boldsymbol{S}_{in_i}) - p/\mathrm{tr}(\boldsymbol{\Sigma}_i)\}\boldsymbol{I}_p|| &= \frac{p|\mathrm{tr}(\boldsymbol{S}_{in_i}) - \mathrm{tr}(\boldsymbol{\Sigma}_i)|}{\mathrm{tr}(\boldsymbol{S}_{in_i})\mathrm{tr}(\boldsymbol{\Sigma}_i)} \\
&= O_P\{(\mathrm{tr}(\boldsymbol{\Sigma}_i^2)/n_i)^{1/2}/\mathrm{tr}(\boldsymbol{S}_{in_i})\} = o_P\{\Delta_{\min(II)}/p\} = o_P(1),
\end{aligned} \tag{B.16}$$

so that $p||\hat{\boldsymbol{B}}_i|| = o_P(\Delta_{\min(II)})$. Note that $\lambda_{\max}(\boldsymbol{A}_i) = \lambda_{\min}(\boldsymbol{A}_i) = \mathrm{tr}(\boldsymbol{\Sigma}_i)/p \in (0, \infty)$ as $p \to \infty$. Thus, from Theorem 1, Propositions 2 and 4, it concludes the result. □

Proof of Corollary 4. We can write that

$$s_{in_i(j)} = n_i s_{oin_i(j)}/(n_i - 1) - n_i(\overline{x}_{ijn_i} - \mu_{ij})^2/(n_i - 1), \tag{B.17}$$

where $s_{oin_i(j)} = \sum_{k=1}^{n_i}(x_{ijk} - \mu_{ij})^2/n_i$. Note that $\limsup_{p\to\infty} E\{\exp(t_{ij}|(x_{ijk} - \mu_{ij})^2 - \sigma_{i(j)}|/\eta_{i(j)}^{1/2})\} \le \limsup_{p\to\infty}[E\{\exp(t_{ij}|x_{ijk} - \mu_{ij}|^2/\eta_{i(j)}^{1/2})\} + \exp(t_{ij}\sigma_{i(j)}/\eta_{i(j)}^{1/2})] < \infty$ under (A-iii). Then, under (A-iii), for any x satisfying $x \to \infty$ and $x = o(n_i^{1/2})$ as $n_i \to \infty$, we have that as $n_i \to \infty$

$$P(n_i^{1/2}|s_{oin_i(j)} - \sigma_{i(j)}|/\eta_{i(j)}^{1/2} \ge x) = \exp\Big(-\frac{x^2}{2}\{1 + o(1)\}\Big).$$

Refer to Chapter 6 in de la Peña et al. [6] for the details of this result. Let $\tau_{1j} = M(\eta_{i(j)} n_i^{-1} \log p)^{1/2}$ for $j = 1, ..., p$, where $M > 2^{1/2}$. Then, under $n_i^{-1} \log p = o(1)$, it holds that as $p \to \infty$

$$\sum_{j=1}^{p} P(|s_{oin_i(j)} - \sigma_{i(j)}| \geq \tau_{1j}) = \sum_{j=1}^{p} P(n_i^{1/2} |s_{oin_i(j)} - \sigma_{i(j)}| / \eta_{i(j)}^{1/2} \geq M(\log p)^{1/2})$$

$$= \sum_{j=1}^{p} \exp\Big(- \frac{M^2 \log p}{2} \{1 + o(1)\}\Big) \to 0. \quad \text{(B.18)}$$

Next, we consider the second term of (B.17). Let $u_{ij} = t_{ij}(\sigma_{i(j)}/\eta_{i(j)})^{1/2}$ for $j = 1, ..., p$. Then, we have that for $j = 1, ..., p$

$$\begin{aligned}
&E\{\exp(u_{ij}|x_{oijk}|/\sigma_{i(j)}^{1/2})\} \\
&= E\{\exp(u_{ij}|x_{oijk}|/\sigma_{i(j)}^{1/2}) I(|x_{oijk}| \leq 1)\} + E\{\exp(u_{ij}|x_{oijk}|/\sigma_{i(j)}^{1/2}) I(|x_{oijk}| > 1)\} \\
&\leq \exp(u_{ij}/\sigma_{i(j)}^{1/2}) + E\{\exp(u_{ij} x_{oijk}^2/\sigma_{i(j)}^{1/2})\} \leq \exp(u_{ij}/\sigma_{i(j)}^{1/2}) + E\{\exp(t_{is} x_{oijk}^2/\eta_{i(j)}^{1/2})\},
\end{aligned}$$

so that $\limsup_{p\to\infty} E\{\exp(u_{ij}|x_{oijk}|/\sigma_{i(j)}^{1/2})\} < \infty$ under (A-iii). Thus, in a way similar to (B.18), we have that

$$\sum_{j=1}^{p} P(|\overline{x}_{ijn_i} - \mu_{ij}| \geq \tau_{2j}) = \sum_{j=1}^{p} P(n_i^{1/2} |\overline{x}_{ijn_i} - \mu_{ij}| / \sigma_{i(j)}^{1/2} \geq M(\log p)^{1/2}) \to 0 \quad \text{(B.19)}$$

for $\tau_{2j} = M(\sigma_{i(j)} n_i^{-1} \log p)^{1/2}$, $j = 1, ..., p$. By combining (B.18) and (B.19) with (B.17), under $n_i^{-1} \log p = o(1)$ and (A-iii), we have that

$$\begin{aligned}
&\sum_{j=1}^{p} P\{|s_{in_i(j)} - n_i \sigma_{i(j)}/(n_i - 1)| \geq n_i(\tau_{1j} + \tau_{2j}^2)/(n_i - 1)\} \\
&\leq \sum_{j=1}^{p} P(|s_{oin_i(j)} - \sigma_{i(j)}| + |\overline{x}_{ijn_i} - \mu_{ij}|^2 \geq \tau_{1j} + \tau_{2j}^2) \\
&\leq \sum_{j=1}^{p} P(|s_{oin_i(j)} - \sigma_{i(j)}| \geq \tau_{1j}) + \sum_{j=1}^{p} P(|\overline{x}_{ijn_i} - \mu_{ij}|^2 \geq \tau_{2j}^2) \to 0.
\end{aligned}$$

Note that $n_i \sigma_{i(j)}/(n_i - 1) = \sigma_{i(j)} + o(n_i^{-1/2})$ and $\tau_{2j}^2 = o(\tau_{1j})$ under $n_i^{-1} \log p = o(1)$. Thus we have that $\max_{j=1,...,p}\{|s_{in_i(j)} - \sigma_{i(j)}|\} = O_P(\max_{j=1,...,p} \tau_{1j})$ under $n_i^{-1} \log p = o(1)$ and (A-iii), so that

$$\max_{j=1,...,p} \{|s_{in_i(j)} - \sigma_{i(j)}|\} = O_P\{(n_i^{-1} \log p)^{1/2}\}. \quad \text{(B.20)}$$

Then, for $i = 1, 2$, it holds that under $n_i^{-1} \log p = o(1)$

$$\begin{aligned}
||\hat{\boldsymbol{B}}_i|| = ||\boldsymbol{S}_{i(d)}^{-1} - \boldsymbol{\Sigma}_{i(d)}^{-1}|| &= \max_{j=1,...,p} \{|s_{in_i(j)} - \sigma_{i(j)}|/(s_{in_i(j)} \sigma_{i(j)})\} \\
&= O_P\{(n_i^{-1} \log p)^{1/2}\} = o_P(1). \quad \text{(B.21)}
\end{aligned}$$

Then, it follows that (C-i') holds under (18). From the fact that $\Delta_{\min(III)} = O(p)$, note that $n_{\min}^{-1} \log p = o(1)$ under (18). Then, by combining (B.21) with Theorem 1, Propositions 2 and 4, we can claim the result of Corollary 4. □

Proofs of Corollary 5. First, note that $s_{n(j)}-\sigma_{(j)}=\sum_{i=1}^{2}(n_i-1)(s_{in_i(j)}-\sigma_{i(j)})/(\sum_{i=1}^{2}n_i-2)$. From (B.20), we can claim that $\max_{j=1,...,p}\{|s_{n(j)}-\sigma_{(j)}|\}=O_P\{(n_{\min}^{-1}\log p)^{1/2}\}$ under $n_{\min}^{-1}\log p=o(1)$ and (A-iii). Thus it follows that $||\boldsymbol{S}_{n(d)}^{-1}-\boldsymbol{\Sigma}_{(d)}^{-1}||=O_P\{(n_{\min}^{-1}\log p)^{1/2}\}$. Note that $\Delta_{(III')}/||\boldsymbol{\mu}_{12}||^2\in(0,\infty)$ as $p\to\infty$. Then, by combining Theorem 1 with Propositions 2 and 5, we can claim the result of Corollary 5. □

Proof of Theorem 4. By using (B.19) and (B.20), we claim the result. □

Proof of Corollary 6. By using Theorem 4, we can claim the result straightforwardly. □

Proof of Corollary 7. Let us write that for $i=1,2$

$$W_i(\boldsymbol{\Sigma}_{i(d)}^{-1})_{FS}=\sum_{j\in\boldsymbol{D}}\{(x_{0j}-\overline{x}_{ijn_i})^2/\sigma_{i(j)}-s_{in_i(j)}/(\sigma_{i(j)}n_i)+\log\sigma_{i(j)}\}.$$

Note that

$$E\{W_{i'}(\boldsymbol{\Sigma}_{i'(d)}^{-1})_{FS}\}-E\{W_i(\boldsymbol{\Sigma}_{i(d)}^{-1})_{FS}\}=\Delta_{i(III)}\ (i'\neq i)\text{ when }\boldsymbol{x}_0\in\pi_i.$$

Also note that $\liminf_{p\to\infty}\Delta_{\min(III)}/p_*>0$ under $\liminf_{p\to\infty}\theta_j>0$ for all $j\in\boldsymbol{D}$. If $\lambda_{\max}(\boldsymbol{\Sigma}_{i*})=o(p_*)$, (C-i') and (C-ii') hold for $\boldsymbol{\Sigma}_{i*}$, $i=1,2$. Here, let us write that $\boldsymbol{\Sigma}_{i(d)*}=\text{diag}(\sigma_{i(j_1)},...,\sigma_{i(j_{p_*})})$ and $\boldsymbol{S}_{i(d)*}=\text{diag}(s_{in_i(j_1)},...,s_{in_i(j_{p_*})})$ for $i=1,2$, where $\boldsymbol{D}=\{j_1,....,j_{p_*}\}$. Then, in a way similar to (B.21), under $n_i^{-1}\log p=o(1)$ and (A-iii), it holds that $||\boldsymbol{S}_{i(d)*}^{-1}-\boldsymbol{\Sigma}_{i(d)*}^{-1}||=O_P\{(n_i^{-1}\log p)^{1/2}\}$. Hence, we have that $p_*||\boldsymbol{S}_{j(d)*}^{-1}-\boldsymbol{\Sigma}_{j(d)*}^{-1}||=o_P(\Delta_{\min(III)})$ under $\liminf_{p\to\infty}\theta_j>0$ for all $j\in\boldsymbol{D}$. By combining Corollary 6 with Propositions 2 and 4, we can claim the result. □

B.3 Additional corollaries

In this section, we give two corollaries and proofs of the corollaries.

Corollary B.1. *Assume either (A-i) and (C-vi) or (A-ii) and (C-vii). Then, for the classification rule by (3) with (16), we have (12) under*

$$\frac{\lambda_{\max}(||\boldsymbol{\mu}_{12}||^2+\lambda_{\max})}{n_{\min}\delta_{\min(II)}^2}=o(1)\quad\textit{and}\quad\frac{tr(\boldsymbol{\Sigma}_{\max}^2)\{(tr(\boldsymbol{\Sigma}_1)/tr(\boldsymbol{\Sigma}_2)-1)^2+1/n_{\min}^2\}}{\delta_{\min(II)}^2}=o(1),\qquad\text{(B.22)}$$

where $\delta_{\min(II)}=\min\{\delta_{1(II)},\delta_{2(II)}\}$.

Proof of Corollary B.1. We consider the case when $\boldsymbol{x}_0\in\pi_i$. Note that $\text{tr}(\boldsymbol{S}_{ln_l})/\text{tr}(\boldsymbol{\Sigma}_l)=1+O_P\{(\text{tr}(\boldsymbol{\Sigma}_l^2)/n_l)^{1/2}/p\}=1+o_P(1)$, $l=1,2$, and $\text{tr}\{(\boldsymbol{x}_0-\boldsymbol{\mu}_i)(\boldsymbol{x}_0-\boldsymbol{\mu}_i)^T-\boldsymbol{\Sigma}_i\}=O_P(\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2})$ under (A-i). Also, note that

$$\text{tr}(\boldsymbol{\Sigma}_i^2)\text{tr}(\boldsymbol{\Sigma}_l^2)\le\lambda_{i1}\lambda_{l1}\text{tr}(\boldsymbol{\Sigma}_i)\text{tr}(\boldsymbol{\Sigma}_l)=o(n_{\min}\delta_{\min(II)}^2p^2),\ l=1,2$$

under (B.22). Then, from (B.16), it holds that for $l=1,2$

$$\begin{aligned}&\text{tr}[\{(\boldsymbol{x}_0-\boldsymbol{\mu}_i)(\boldsymbol{x}_0-\boldsymbol{\mu}_i)^T-\boldsymbol{\Sigma}_i\}\hat{\boldsymbol{B}}_l]\\&=p\frac{\text{tr}(\boldsymbol{\Sigma}_l)-\text{tr}(\boldsymbol{S}_{ln_l})}{\text{tr}(\boldsymbol{\Sigma}_l)\text{tr}(\boldsymbol{S}_{ln_l})}\text{tr}\{(\boldsymbol{x}_0-\boldsymbol{\mu}_i)(\boldsymbol{x}_0-\boldsymbol{\mu}_i)^T-\boldsymbol{\Sigma}_i\}\\&=O_P\{(\text{tr}(\boldsymbol{\Sigma}_i^2)\text{tr}(\boldsymbol{\Sigma}_l^2)/n_l)^{1/2}/p\}=o_P(\delta_{\min(II)}),\quad\text{and}\qquad\text{(B.23)}\\&p||\hat{\boldsymbol{B}}_l||/n_l^{1/2}=O_P\{\text{tr}(\boldsymbol{\Sigma}_l^2)^{1/2}/n_l\}=o_P(\delta_{\min(II)})\end{aligned}$$

under (A-i) and (B.22). Similarly, from (B.16), under (A-i) and (B.22), we have that for $i' \neq i$

$$\begin{aligned}
&\{2(\boldsymbol{x}_0-\boldsymbol{\mu}_i)+(-1)^{i+1}\boldsymbol{\mu}_{12}\}^T\hat{\boldsymbol{B}}_{i'}\boldsymbol{\mu}_{12}\\
&=O_P\{(\boldsymbol{\mu}_{12}^T\boldsymbol{\Sigma}_i\boldsymbol{\mu}_{12}/n_{i'})^{1/2}\}+O_P\{(\mathrm{tr}(\boldsymbol{\Sigma}_{i'}^2)/n_{i'})^{1/2}||\boldsymbol{\mu}_{12}||^2/p\}\\
&=O_P\{(\lambda_{i1}||\boldsymbol{\mu}_{12}||^2/n_{i'})^{1/2}\}+O_P\{(\lambda_{i'1}||\boldsymbol{\mu}_{12}||^2/n_{i'})^{1/2}\}=o_P(\delta_{\min(II)})
\end{aligned}$$

from the facts that $\boldsymbol{\mu}_{12}^T\boldsymbol{\Sigma}_i\boldsymbol{\mu}_{12}\le\lambda_{i1}||\boldsymbol{\mu}_{12}||^2$ and $\mathrm{tr}(\boldsymbol{\Sigma}_{i'}^2)=O(\lambda_{i'1}p)$. On the other hand, under (A-i) and (B.22), from (B.16), we have that for $l=1,2$

$$\begin{aligned}
\log\{\mathrm{tr}(\boldsymbol{\Sigma}_l)/\mathrm{tr}(\boldsymbol{S}_{ln_l})\}&=(\mathrm{tr}(\boldsymbol{\Sigma}_l)/\mathrm{tr}(\boldsymbol{S}_{ln_l})-1)+O_P\{(\mathrm{tr}(\boldsymbol{\Sigma}_l)/\mathrm{tr}(\boldsymbol{S}_{ln_l})-1)^2\}\\
&=(\mathrm{tr}(\boldsymbol{\Sigma}_l)/\mathrm{tr}(\boldsymbol{S}_{ln_l})-1)+O_P\{\mathrm{tr}(\boldsymbol{\Sigma}_l^2)/(n_lp^2)\}\\
&=(\mathrm{tr}(\boldsymbol{\Sigma}_l)/\mathrm{tr}(\boldsymbol{S}_{ln_l})-1)+o_P(\delta_{\min(II)}/p)
\end{aligned}$$

from the facts that $\mathrm{tr}(\boldsymbol{\Sigma}_l^2)/p=O(\lambda_{l1})$ and $\mathrm{tr}(\boldsymbol{\Sigma}_l)/\mathrm{tr}(\boldsymbol{S}_{ln_l})=1+o_P(1)$. Then, under (A-i) and (B.22), it holds that

$$\mathrm{tr}(\boldsymbol{\Sigma}_i\hat{\boldsymbol{B}}_i)-\log|\hat{\boldsymbol{A}}_i\boldsymbol{A}_i^{-1}|=p(\mathrm{tr}(\boldsymbol{\Sigma}_i)/\mathrm{tr}(\boldsymbol{S}_{in_i})-1)-p\log\{\mathrm{tr}(\boldsymbol{\Sigma}_i)/\mathrm{tr}(\boldsymbol{S}_{in_i})\}=o_P(\delta_{\min(II)}).$$

Similarly, under (A-i) and (B.22), we have that

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{\Sigma}_i\hat{\boldsymbol{B}}_{i'})-\log|\hat{\boldsymbol{A}}_{i'}\boldsymbol{A}_{i'}^{-1}|&=p(\mathrm{tr}(\boldsymbol{\Sigma}_i)/\mathrm{tr}(\boldsymbol{\Sigma}_{i'})-1)(\mathrm{tr}(\boldsymbol{\Sigma}_{i'})/\mathrm{tr}(\boldsymbol{S}_{i'n_{i'}})-1)+o_P(\delta_{\min(II)})\\
&=O_P(|\mathrm{tr}(\boldsymbol{\Sigma}_i)/\mathrm{tr}(\boldsymbol{\Sigma}_{i'})-1|(\mathrm{tr}(\boldsymbol{\Sigma}_{i'}^2)/n_{i'})^{1/2})+o_P(\delta_{\min(II)})=o_P(\delta_{\min(II)}).
\end{aligned} \tag{B.24}$$

Here, by noting that $\mathrm{tr}(\boldsymbol{\Sigma}_l)/p\in(0,\infty)$ as $p\to\infty$ for $l=1,2$, we have that

$$\begin{aligned}
&\mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_l)^2\}=\mathrm{tr}\{(\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\Sigma}_l\boldsymbol{\Sigma}_i^{1/2})^2\}\\
&\le\lambda_{\max}(\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\Sigma}_l\boldsymbol{\Sigma}_i^{1/2})\mathrm{tr}(\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\Sigma}_l\boldsymbol{\Sigma}_i^{1/2})\le\lambda_{\max}(\boldsymbol{\Sigma}_i)\lambda_{l1}\delta_{i(II)}^2n_l=O(\lambda_{i1}\lambda_{l1}\delta_{i(II)}^2n_l);\quad\text{and}\\
&\boldsymbol{\mu}_{12}^T\boldsymbol{\Sigma}_l\boldsymbol{\mu}_{12}\le||\boldsymbol{\mu}_{12}||^2\lambda_{\max}(\boldsymbol{\Sigma}_l)=O(||\boldsymbol{\mu}_{12}||^2\lambda_{l1})\quad\text{for } l=1,2.
\end{aligned}$$

Thus, (C-iv) holds under (B.22). Also, (C-v) holds under (B.22). Hence, by combining (B.23) to (B.24) with Lemma B.3, Theorems 2 and 3, we can claim the result. □

Corollary B.2. *Let $\eta_{i(rs)}=Var(x_{oirk}x_{oisk})$ for $i=1,2,$ and $r,s=1,...,p$ $(k=1,...,n_i)$. Assume (A-i) and (A-iii). Assume also $\lambda(\boldsymbol{\Sigma}_i)\in(0,\infty)$ as $p\to\infty$ and $\liminf_{p\to\infty}\eta_{i(rs)}>0$ for all $r,s;$ $i=1,2$. Then, for the classification rule by (3) with (21), we have (6) under the conditions that $p^{1/2}/\Delta_{\min(IV)}=o(1)$ and $p^4\log p/(n_{\min}\Delta_{\min(IV)}^2)=o(1)$.*

Proofs of Corollary B.2. Let $\boldsymbol{S}_{oin_i}=\sum_{k=1}^{n_i}(\boldsymbol{x}_{ik}-\boldsymbol{\mu}_i)(\boldsymbol{x}_{ik}-\boldsymbol{\mu}_i)^T/n_i$ and denote its (r,s) element by $s_{oin_i(rs)}$ for $r,s=1,...,p$. Let $u_{i(rs)}=\min\{t_{ir}/\eta_{i(r)}^{1/2},t_{is}/\eta_{i(s)}^{1/2}\}\eta_{i(rs)}^{1/2}$ for $r,s=1,...,p$. Then, we have that for $r,s=1,...,p$

$$\begin{aligned}
&E\{\exp(u_{i(rs)}|x_{oirk}x_{oisk}-\sigma_{i(rs)}|/\eta_{i(rs)}^{1/2})\}\\
&\le E[\exp\{u_{i(rs)}(x_{oirk}^2/2+x_{oisk}^2/2+\sigma_{i(rs)})/\eta_{i(rs)}^{1/2}\}]\\
&\le\exp(u_{i(rs)}\sigma_{i(rs)}/\eta_{i(rs)}^{1/2})E[\exp\{t_{ir}x_{oirk}^2/(2\eta_{i(r)}^{1/2})\}\exp\{t_{is}x_{oisk}^2/(2\eta_{i(s)}^{1/2})\}]\\
&\le\exp(u_{i(rs)}\sigma_{i(rs)}/\eta_{i(rs)}^{1/2})[E\{\exp(t_{ir}x_{oirk}^2/\eta_{i(r)}^{1/2})\}E\{\exp(t_{is}x_{oisk}^2/\eta_{i(s)}^{1/2})\}]^{1/2},
\end{aligned}$$

so that $\limsup_{p\to\infty} E\{\exp(u_{i(rs)}|x_{oirk}x_{oisk}-\sigma_{i(rs)}|/\eta_{i(rs)}^{1/2})\}<\infty$ under (A-iii). Note that $s_{in_i(rs)} = n_i s_{oin_i(rs)}/(n_i-1) - n_i(\overline{x}_{irn_i}-\mu_{ir})(\overline{x}_{isn_i}-\mu_{is})/(n_i-1)$, where $s_{in_i(rs)}$ is the (r,s) element of $\boldsymbol{S}_{in_i}$. Also, note that $\eta_{i(rs)}\in(0,\infty)$ as $p\to\infty$ under (A-iii) and $\liminf_{p\to\infty}\eta_{i(rs)}>0$ for all r,s, from the fact that $\eta_{i(rs)}\le\{(\eta_{i(r)}+\sigma_{i(r)}^2)(\eta_{i(s)}+\sigma_{i(s)}^2)\}^{1/2}$. In a way similar to (B.18) and (B.19), under $n_i^{-1}\log p=o(1)$, (A-iii) and $\liminf_{p\to\infty}\eta_{i(rs)}>0$ for all r,s, we have that

$$\begin{aligned}&\sum_{r,s=1}^{p} P\{|s_{in_i(rs)}-n_i\sigma_{i(rs)}/(n_i-1)|\ge n_i(\tau_{1(rs)}+\tau_{2(rs)})/(n_i-1)\}\\ &\le \sum_{r,s=1}^{p}\{P(|s_{oin_i(rs)}-\sigma_{i(rs)}|\ge\tau_{1(rs)})+P(|\overline{x}_{irn_i}-\mu_{ir}||\overline{x}_{isn_i}-\mu_{is}|\ge\tau_{2(rs)})\}\\ &\le \sum_{r,s=1}^{p} P(|\overline{x}_{irn_i}-\mu_{ir}|^2+|\overline{x}_{isn_i}-\mu_{sr}|^2\ge\tau_{2(rs)})\}+o(1)\to 0\end{aligned}$$

for $\tau_{1(rs)}=M(\eta_{i(rs)}n_i^{-1}\log p)^{1/2}$ and $\tau_{2(rs)}=M^2\{(\sigma_{i(r)}+\sigma_{i(s)})n_i^{-1}\log p\}$, $r,s=1,...,p$, where $M>2$. Thus it holds that $\max_{r,s=1,...,p}\{|s_{in_i(rs)}-\sigma_{i(rs)}|\}=O_P(\max_{r,s=1,...,p}\tau_{1(rs)})$ because $\tau_{2(rs)}=o(\tau_{1(rs)})$, so that

$$\max_{r,s=1,...,p}\{|s_{in_i(rs)}-\sigma_{i(rs)}\}=O_P\{(n_i^{-1}\log p)^{1/2}\}. \tag{B.25}$$

Here, from the equations (A1) and (A2) in Bickel and Levina [4], we have that $||\boldsymbol{M}||\le\max_{s=1,...,p}\sum_{t=1}^p|m_{st}|$ for any symmetric matrix $\boldsymbol{M}$, where m_{st} is the (s,t) element of $\boldsymbol{M}$. From (B.25), we have that

$$||\boldsymbol{S}_{in_i}-\boldsymbol{\Sigma}_i||=O_P\{p(n_i^{-1}\log p)^{1/2}\}=o_P(1) \tag{B.26}$$

under $n_i^{-1}p^2\log p=o(1)$, (A-iii) and $\liminf_{p\to\infty}\eta_{i(rs)}>0$ for all r,s. Then, under $\lambda(\boldsymbol{\Sigma}_i)\in(0,\infty)$ as $p\to\infty$, we can claim that $\lambda(\boldsymbol{S}_{in_i})\in(0,\infty)$ in probability. Thus it holds that $||\boldsymbol{e}_p^T\boldsymbol{\Sigma}_i^{-1}||\in(0,\infty)$ and $||\boldsymbol{e}_p^T\boldsymbol{S}_{in_i}^{-1}||\in(0,\infty)$ in probability, where $\boldsymbol{e}_p$ is an arbitrary (random) p-vector such that $||\boldsymbol{e}_p||=1$. Then, from (B.26), we have that $\boldsymbol{e}_p^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{S}_{in_i}-\boldsymbol{\Sigma}_i)\boldsymbol{S}_{in_i}^{-1}\boldsymbol{e}_p=\boldsymbol{e}_p^T(\boldsymbol{\Sigma}_i^{-1}-\boldsymbol{S}_{in_i}^{-1})\boldsymbol{e}_p=O_P\{p(n_i^{-1}\log p)^{1/2}\}$ under $n_i^{-1}p^2\log p=o(1)$, (A-iii) and $\liminf_{p\to\infty}\eta_{i(rs)}>0$ for all r,s, so that $||\hat{\boldsymbol{B}}_i||=O_P\{p(n_i^{-1}\log p)^{1/2}\}=o_P(1)$. Note that (C-i') and (C-ii') hold under the conditions of Corollary B.2. Also, note that $\mathrm{tr}\{(\boldsymbol{I}_p-\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}^{-1})^2\}=O(p)$ $(i'\ne i)$ under $\lambda(\boldsymbol{\Sigma}_i)\in(0,\infty)$ as $p\to\infty$. By combining Theorem 1 with Propositions 2 and 4, we can claim the result. □

References

[1] M. Aoshima, K. Yata, A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data, Ann. Inst. Statist. Math. 66 (2014) 983-1010.

[2] M. Aoshima, K. Yata, High-dimensional quadratic classifiers in non-sparse settings.

[3] S.A. Armstrong, J.E. Staunton, L.B. Silverman, R. Pieters, M.L. den Boer, M.D. Minden, S.E. Sallan, E.S. Lander, T.R. Golub, S.J. Korsmeyer, MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia, Nature Genetics 30 (2002) 41-47.

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

[5] T.T. Cai, W. Liu, A direct estimation approach to sparse linear discriminant analysis, J. Amer. Statist. Assoc. 106 (2011) 1566-1577.

[6] V.H. de la Peña, T.L. Lai, Q.M. Shao, Self-Normalized Processes, Springer-Verlag, Berlin, 2009.

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

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

[9] D.L. McLeish, Dependent central limit theorems and invariance principles, Ann. Probab. 2 (1974) 620-628.

[10] A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, Simple decision rules for classifying human cancers from gene expression profiles, Bioinformatics 21 (2005) 3896-3904.

[11] V.N. Vapnik, The Nature of Statistical Learning Theory (second ed.), Springer-Verlag, New York, 2000.

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