Metrika (2021) 84:339–374
https://doi.org/10.1007/s00184-020-00781-3

A shrinkage approach to joint estimation of multiple covariance matrices

Zongliang Hu[1] · Zhishui Hu[2] · Kai Dong[3] · Tiejun Tong[3] · Yuedong Wang[4]

Received: 18 November 2018 / Published online: 19 June 2020

Abstract
In this paper, we propose a shrinkage framework for jointly estimating multiple covariance matrices by shrinking the sample covariance matrices towards the pooled sample covariance matrix. This framework allows us to borrow information across different groups. We derive the optimal shrinkage parameters under the Stein and quadratic loss functions, and prove that our derived estimators are asymptotically optimal when the sample size or the number of groups tends to infinity. Simulation studies demonstrate that our proposed shrinkage method performs favorably compared to the existing methods.

Keywords Covariance matrices · Joint estimation · Optimal estimator · Quadratic loss function · Shrinkage parameter · Stein loss function

1 Introduction

Estimation of the covariance matrices plays an important role in various areas such as principal component analysis, graphical models and outlier detection. In many applications the data may be comprised of several distinct groups. One such example

---

Zongliang Hu and Zhishui Hu are co-first authors.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00184-020-00781-3) contains supplementary material, which is available to authorized users.

✉ Tiejun Tong
tongt@hkbu.edu.hk

[1] College of Mathematics and Statistics, Shenzhen University, Shenzhen, China

[2] Department of Statistics and Finance, University of Science and Technology of China, Hefei, China

[3] Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[4] Department of Statistics and Applied Probability, University of California, Santa Barbara, USA

is gene expression data where researchers collect gene expression profiles for different cancer tissue samples. To analyze the co-expression network for each of the cancer samples, one needs to estimate several covariance matrices simultaneously. If the number of observations in each group is small, the estimates of the covariance matrices based on data from each individual group are not reliable. On the other hand, if we estimate the covariance matrices by the pooled sample covariance matrix, the differences among the tissues will be ignored leading to biased estimators when the covariance matrices are not all the same.

Various methods have been proposed to estimate the covariance matrices jointly. In general, certain decompositions were used to deal with positive semi-definite covariance matrices. For example, Boik (2002, 2003) used the spectral decomposition while Manly and Rayner (1987) and Barnard et al. (2000) used the variance-correlation decomposition. However, the estimates of the orthogonal and correlation matrices appeared in the spectral and variance-correlation decomposition often involve computationally challenging constrained optimization problems. Another constraint-free decomposition is the Cholesky decomposition. Pourahmadi et al. (2007) developed the maximum likelihood estimates with a pre-specified common structure. Daniels (2006), Hoff (2009) and Gaskins and Daniels (2016) proposed Bayesian priors on the Cholesky decomposed terms where the Markov chain Monte Carlo algorithm may become computationally intractable when the number of groups or the dimension of the covariance matrix is large. Recently, Guo et al. (2011), Danaher et al. (2014), Le and Hastie (2016) and Cai et al. (2016) proposed penalized likelihood methods that induce a sparsity structure in the precision matrices across groups. Price et al. (2015) suggested to add a ridge penalty and a ridge fusion penalty to the log-likelihood which yields another regularization for multiple precision matrices. Friedman (1989) and Ramey et al. (2016) proposed a regularized discriminant analysis, in which they replaced the unknown covariance matrices by a linear combination of sample covariances, the identity matrix, and the pooled sample covariance matrix.

In this paper, rather than decomposing and imposing structure assumption on the covariance matrices, we consider a shrinkage approach to jointly estimate the multiple covariance matrices. Unlike estimators under pre-specified structures where the imposed structures are usually hard to verify in practice, our new method can systematically borrow information across groups through a shrinkage framework. Specifically, let $\boldsymbol{X}_{gi} = (X_{1gi}, \ldots, X_{pgi})^T$ be independent and identically distributed (i.i.d.) random vectors from the multivariate normal distribution $N_p(\boldsymbol{\mu}_g, \Sigma_g)$, where $i = 1, \ldots, N$, $g = 1, \ldots, G$, $\boldsymbol{\mu}_g$ are the mean vectors, and Σ_g are the $p \times p$ covariance matrices. We propose the following shrinkage estimators:

$$\widehat{\Sigma}_g = \alpha \frac{S_g}{n} + (1-\alpha)\widehat{\Sigma}_{\text{pool}}, \quad 0 \le \alpha \le 1, \tag{1}$$

where $n = N - 1$, $\widehat{\Sigma}_{\text{pool}} = \sum_{g=1}^{G} S_g/(nG)$, $S_g = \sum_{i=1}^{N}(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)^T$, $\overline{\boldsymbol{X}}_g$ is the sample mean of the gth group, and α is a shrinkage parameter. When $\alpha = 1$, we estimate the covariance matrices by the sample covariance matrices. When $\alpha = 0$, we estimate all the covariance matrices by the pooled sample covariance matrix. The optimal shrinkage parameter is defined such that the average risk of the estimators is minimized under a specific loss function. We note that when $0 \leq \alpha < 1$, an optimal combination between the sample covariance matrices and the pooled covariance matrix will use all the information among the G groups rather than using only the information within the gth group, and hence reduces the total variation of the sample covariance matrices.

Since the first proposal by Stein (1956) for the estimation of means, the shrinkage methods have been extensively studied and extended to estimate other parameters including the variances (Tong and Wang 2007; Tong et al. 2012). Ledoit and Wolf (2004), Schäfer and Strimmer (2005) and Ledoit and Wolf (2012) have developed shrinkage estimators for a single high-dimensional covariance matrix. In this paper, we consider the shrinkage estimation of multiple covariance matrices from a decision-theoretic point of view that has not been studied before. It is well known that the conventional quadratic discriminant analysis (QDA) classifier computes the sample covariance matrices for each group. If the training set is not large, QDA may perform poorly due to a large variation of sample covariance matrices. To improve the performance, regularization was suggested for a proper compromise between the bias and variance of the classifier (Friedman 1989; Ramey et al. 2016; Le and Hastie 2016). Our proposed shrinkage estimators in this paper provide an explicit analytical solution for the optimal regularization in QDA.

The remainder of the paper is organized as follows. In Sect. 2, we derive the optimal shrinkage estimators for the covariance matrices under the Stein and quadratic loss functions, respectively. We also propose estimators for the optimal shrinkage parameters and investigate their asymptotic properties. We then conduct simulations in Sect. 3 to evaluate the finite sample performance of the proposed estimators and compare them with some existing methods. We conclude the paper in Sect. 4 with some discussions. Technical proofs are given in Sect. 5.

2 Joint estimation of the covariance matrices

As defined in Sect. 1, our proposed shrinkage estimators of the covariance matrices are

$$\begin{aligned}\widehat{\Sigma}_g &= \alpha\frac{S_g}{n} + (1-\alpha)\widehat{\Sigma}_{\text{pool}} \\ &= \left\{\frac{\alpha}{n} + \frac{(1-\alpha)}{nG}\right\}\sum_{i=1}^{N}(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)^T \\ &\quad + \frac{1-\alpha}{nG}\sum_{k\neq g}^{G}\sum_{j=1}^{N}(\boldsymbol{X}_{kj} - \overline{\boldsymbol{X}}_k)(\boldsymbol{X}_{kj} - \overline{\boldsymbol{X}}_k)^T,\end{aligned}$$

where α controls the level of shrinkage. In particular, the estimators only employ the observations within each group when $\alpha = 1$, and all estimates are shrunken to the pooled sample covariance matrix when $\alpha = 0$. If $\alpha \in [0, 1)$, the gth shrinkage estimator borrows information from other groups.

To find the optimal shrinkage parameter, we consider to minimize the average risk of the covariance matrix estimators under the Stein loss function

$$L_1(\widehat{\Sigma}, \Sigma) = \mathrm{tr}(\widehat{\Sigma}\Sigma^{-1}) - \log\det(\widehat{\Sigma}\Sigma^{-1}) - p,$$

and the quadratic loss function

$$L_2(\widehat{\Sigma}, \Sigma) = \mathrm{tr}(\widehat{\Sigma}\Sigma^{-1} - I)^2,$$

where I is the identity matrix, $\det(\cdot)$ and $\mathrm{tr}(\cdot)$ denote the determinant and trace of a covariance matrix, respectively. Note that these two loss functions are commonly used in the covariance matrix estimation Haff (1980, 1991); Yang and Berger (1994).

For ease of notation, let $A = [A_{ij}] \in \mathbb{R}^{p\times p}$ be a $p \times p$ matrix, where A_{ij} are the components of A. The Frobenius norm is defined as $\|A\| = \{\mathrm{tr}(AA^T)\}^{1/2} = (\sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}^2)^{1/2}$. For any random matrix A and positive integer k, $E\|A\|^k < \infty$ is equivalent to $E|A_{ij}|^k < \infty$ for any $i, j \in \{1, \ldots, p\}$. Throughout this paper, for any symmetric matrix A, $A > 0$ means that A is a positive definite matrix. In addition, we let $\Sigma = \{\Sigma_1, \ldots, \Sigma_G\}$ be a set of covariance matrices corresponding to the G groups, and $\widehat{\Sigma} = \{\widehat{\Sigma}_1, \ldots, \widehat{\Sigma}_G\}$ be an estimate of Σ.

2.1 Optimal estimator under the Stein loss function

Under the loss function L_1, the average risk is

$$\begin{aligned}
R_1(\alpha, \widehat{\Sigma}, \Sigma) &= \frac{1}{G}\sum_{g=1}^{G} EL_1(\widehat{\Sigma}_g, \Sigma_g)\\
&= \frac{1}{G}\sum_{g=1}^{G} E\mathrm{tr}\Big(\widehat{\Sigma}_g\Sigma_g^{-1}\Big) - \frac{1}{G}\sum_{g=1}^{G} E\Big\{\log\det(\widehat{\Sigma}_g\Sigma_g^{-1})\Big\} - p\\
&= \frac{1-\alpha}{G^2}\mathrm{tr}\Big\{\Big(\sum_{g=1}^{G}\Sigma_g\Big)\Big(\sum_{g=1}^{G}\Sigma_g^{-1}\Big)\Big\}\\
&\quad -\frac{1}{G}\sum_{g=1}^{G} E\log\det\Big[\Big\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\mathrm{pool}}\Big\}\Sigma_g^{-1}\Big] - (1-\alpha)p.
\end{aligned}$$

Let $R_1'(\alpha, \widehat{\Sigma}, \Sigma)$ and $R_1''(\alpha, \widehat{\Sigma}, \Sigma)$ be the first and second derivatives of $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ with respect to α. Let α_1^* be the optimal shrinkage parameter such that $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ achieves the minimum value for $\alpha \in [0, 1]$. We have the following two theorems.

Theorem 1 *For any fixed G, p and $n > p + 1$, $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ is a strictly convex function of α on $[0, 1]$ that satisfies*

(a) $R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} \le 0$, *where the equality holds if and only if* $\Sigma_1 = \cdots = \Sigma_G$, *and*
(b) $R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=1} > 0$.

Theorem 2 *For any fixed G and p, as* $n \to \infty$, *we have*

(a) $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ *tends to a constant function of* α *when* $\Sigma_1 = \cdots = \Sigma_G$, *and*
(b) $\alpha_1^* \to 1$ *when* Σ_g *are not all the same.*

The proofs of Theorems 1 and 2 are given in Sects. 5.1 and 5.2, respectively. Theorem 1 states that there exists a unique optimal shrinkage parameter $\alpha_1^* < 1$, which implies that the conventional sample covariance matrices are not admissible under the Stein loss function. In addition, if Σ_g are all the same, we have $\alpha_1^* = 0$ and hence the optimal shrinkage estimators for the covariance matrices are exactly the pooled sample covariance matrix. Theorem 2 implies that if the sample size is sufficiently large, it is no longer necessary to borrow information across other classes for unequal covariance matrices.

In practice, the optimal shrinkage parameter α_1^* is unknown since it depends on some unknown parameters. The following theorems provide an estimator of α_1^* and also derive its asymptotic properties. The proofs of Theorems 3 to 5 are given in Sects. 5.3 to 5.5, respectively.

Theorem 3 *For any fixed* G, p *and* $n > p + 1$, *let*

$$\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma) = p - \mathrm{tr}\Big[\Big(\frac{1}{G}\sum_{g=1}^{G} S_g/n\Big)\Big\{\frac{1}{G}\sum_{g=1}^{G}(n-p-1)S_g^{-1}\Big\}\Big]$$
$$-\frac{1}{G}\sum_{g=1}^{G}\mathrm{tr}\Big[\big(S_g/n - \widehat{\Sigma}_{\mathrm{pool}}\big)\big\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\mathrm{pool}}\big\}^{-1}\Big] - \frac{p(p+1)}{nG}.$$

Then

(a) $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)$ *is an unbiased estimator of* $R_1'(\alpha, \widehat{\Sigma}, \Sigma)$, *and is strictly increasing with respect to* α *on the interval* $[0, 1]$.
(b) *If* $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} \le 0$, *there exists a unique* α *that satisfies* $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma) = 0$ *and we denote the solution as* $\widehat{\alpha}_1^*$. *Otherwise, we let* $\widehat{\alpha}_1^* = 0$.

Theorem 4 *For any fixed G and p, as* $n \to \infty$, *we have* $\widehat{\alpha}_1^* \xrightarrow{a.s.} 1$ *when* Σ_g *are not all the same, where* $\xrightarrow{a.s.}$ *denotes the almost sure convergence.*

According to Theorem 4, if we plug $\widehat{\alpha}_1^*$ into (1), then for any fixed G and p, we have

$$\widehat{\Sigma}_g|_{\alpha=\widehat{\alpha}_1^*} = \widehat{\alpha}_1^* S_g/n + (1-\widehat{\alpha}_1^*)\widehat{\Sigma}_{\mathrm{pool}} \xrightarrow{a.s.} \Sigma_g \quad \text{as} \quad n \to \infty,$$

for $g = 1, \ldots, G$. Therefore, the optimal shrinkage estimators of the covariance matrices under the Stein loss remain to be consistent.

To investigate the asymptotic properties of $\widehat{\alpha}_1^*$ as $G \to \infty$, we assume that $\Sigma_g \overset{i.i.d.}{\sim} U$, where U is a probability measure supported on $\mathbb{H}^+ := \{A_{p\times p} : A > 0\}$.

Theorem 5 *For any fixed n and p with $n > p + 3$, assuming that $E\|S_1\|^7 < \infty$, $E\|S_1^{-1}\|^7 < \infty$, $E\|\Sigma_1\|^7 < \infty$ and $E\|\Sigma_1^{-1}\|^7 < \infty$, we have $R_1'(\alpha, \widehat{\Sigma}, \Sigma) - \widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma) \xrightarrow{a.s.} 0$ uniformly for $\alpha \in [0, 1]$ as $G \to \infty$. In addition, we have $\widehat{\alpha}_1^* - \alpha_1^* \xrightarrow{a.s.} 0$ as $G \to \infty$.*

According to Theorem 5, under some mild conditions, we can borrow information across groups to get a consistent estimator of the optimal shrinkage parameter.

2.2 Optimal estimator under the quadratic loss function

Under the loss function L_2, the average risk is

$$R_2(\alpha, \widehat{\Sigma}, \Sigma) = \frac{1}{G}\sum_{g=1}^{G} EL_2(\widehat{\Sigma}_g, \Sigma_g) = \alpha^2 a_1 + 2\alpha a_2 + a_3,$$

where

$$\begin{aligned} a_1 &= \sum_{g=1}^{G} E\mathrm{tr}\Big\{(S_g/n - \widehat{\Sigma}_{\mathrm{pool}})\Sigma_g^{-1}\Big\}^2/G, \\ a_2 &= \sum_{g=1}^{G} E\mathrm{tr}\Big\{(S_g/n - \widehat{\Sigma}_{\mathrm{pool}})\Sigma_g^{-1}(\widehat{\Sigma}_{\mathrm{pool}}\Sigma_g^{-1} - I)\Big\}/G, \\ a_3 &= \sum_{g=1}^{G} E\mathrm{tr}\Big(\widehat{\Sigma}_{\mathrm{pool}}\Sigma_g^{-1} - I\Big)^2/G. \end{aligned} \tag{2}$$

According to Sect. 5.6, a_1 and a_2 can be rewritten as

$$\begin{aligned} a_1 &= \frac{1}{G}\sum_{g=1}^{G}\mathrm{tr}\Big\{(\Sigma_g - \overline{\Sigma})\Sigma_g^{-1}\Big\}^2 + \frac{(G-1)^2(p^2+p)}{nG^2} \\ &\quad + \frac{1}{nG^3}\sum_{g=1}^{G}\sum_{g'\neq g}\Big[\mathrm{tr}\Big\{(\Sigma_{g'}\Sigma_g^{-1})^2\Big\} + \Big\{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\Big\}^2\Big] \end{aligned} \tag{3}$$

and

$$a_2 = -a_1 + (G-1)(p^2+p)/n, \tag{4}$$

where $\overline{\Sigma} = \sum_{g=1}^{G}\Sigma_g/G$. Let α_2^* be the optimal shrinkage parameter such that $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ achieves the minimum value for $\alpha \in [0, 1]$. We have the following two theorems.

Theorem 6 *For any fixed* G, p *and* $n > p+1$, $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ *is a strictly convex function of* α *on* $[0, 1]$ *with the unique minimum point at* $\alpha_2^* = -a_2/a_1 \in [0, 1)$. *In the special case, when* Σ_g *are all the same, we have* $\alpha_2^* = 0$.

Theorem 7 *For any fixed* G *and* p, *as* $n \to \infty$, *we have*

(*a*) $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ *tends to a constant function of* α *when* $\Sigma_1 = \cdots = \Sigma_G$, *and*
(*b*) $\alpha_2^* \to 1$ *when* Σ_g *are not all the same.*

The proofs of Theorems 6 and 7 are given in Sects. 5.7 and 5.8, respectively. According to Theorem 6, the optimal shrinkage parameter α_2^* is less than 1, which implies that the conventional sample covariance matrices are not admissible under the quadratic loss function. In the special case when Σ_g are all the same, the optimal shrinkage estimators for the covariance matrices are given as the pooled sample covariance matrix. Theorem 7 indicates that if the sample size is sufficiently large, it is no longer necessary to borrow information across other classes for unequal covariance matrices.

To estimate the optimal shrinkage parameter α_2^*, we need an estimator of

$$b = \frac{1}{G}\sum_{g=1}^{G} \mathrm{tr}\Big\{(\Sigma_g - \overline{\Sigma})\Sigma_g^{-1}\Big\}^2. \tag{5}$$

The following theorem is to find an estimator of b.

Theorem 8 *Assume that* $n > p+3$, *for any fixed constant matrix* Σ_0, *we have*

$$E\Big[\mathrm{tr}\big(I - r\Sigma_0 S_g^{-1}\big)^2 - r\Big\{\mathrm{tr}(\Sigma_0 S_g^{-1})^2 + \big\{\mathrm{tr}(\Sigma_0 S_g^{-1})\big\}^2\Big\}\Big] = \mathrm{tr}\Big\{(\Sigma_g - \Sigma_0)\Sigma_g^{-1}\Big\}^2, \tag{6}$$

where $r = n - p - 1$. *Consequently,*

$$\mathrm{tr}\big(I - r\Sigma_0 S_g^{-1}\big)^2 - r\Big\{\mathrm{tr}(\Sigma_0 S_g^{-1})^2 + \big\{\mathrm{tr}(\Sigma_0 S_g^{-1})\big\}^2\Big\}$$

is an unbiased estimator of $\mathrm{tr}\big\{(\Sigma_g - \Sigma_0)\Sigma_g^{-1}\big\}^2$.

The proof of Theorem 8 is given in Sect. 5.9. According to this theorem, by using $\widehat{\Sigma}_{\text{pool}}$ to estimate $\overline{\Sigma}$ and noting that $b \geq 0$, we can define an estimator of b in (5) as

$$\widehat{b} = \max\left\{0,\ \frac{1}{G}\sum_{g=1}^{G}\Big[\mathrm{tr}(I - r\widehat{\Sigma}_{\text{pool}}S_g^{-1})^2 - r\Big\{\mathrm{tr}(\widehat{\Sigma}_{\text{pool}}S_g^{-1})^2 + \big\{\mathrm{tr}(\widehat{\Sigma}_{\text{pool}}S_g^{-1})\big\}^2\Big\}\Big]\right\}.$$

Then we can estimate a_1 and a_2, respectively, by

$$\widehat{a}_1 = \widehat{b} + \frac{(G-1)^2(p^2+p)}{nG^2} + \frac{r^2}{n^3G^3}\sum_{g=1}^{G}\sum_{g'\neq g}\Big\{\mathrm{tr}(S_{g'}S_g^{-1})^2 + \big\{\mathrm{tr}(S_{g'}S_g^{-1})\big\}^2\Big\},$$

$$\widehat{a}_2 = -\widehat{a}_1 + \frac{(G-1)(p^2+p)}{nG}. \tag{7}$$

It is easy to see that $-\widehat{a}_2 < \widehat{a}_1$. Therefore, the optimal shrinkage estimators can be given by $\widehat{\alpha}_2^* = \max\{0, -\widehat{a}_2/\widehat{a}_1\}$ and $\hat{\alpha}_2^* < 1$.

Theorem 9 *For any fixed G and p, as $n \to \infty$, we have $\widehat{\alpha}_2^* \xrightarrow{a.s.} 1$ when Σ_g are not all the same.*

The proofs of Theorem 9 is given in Sects. 5.10. According to this theorem, if we plug $\widehat{\alpha}_2^*$ into (1), then for any fixed G and p, we have

$$\widehat{\Sigma}_g|_{\alpha=\widehat{\alpha}_2^*} = \widehat{\alpha}_2^* S_g/n + (1-\widehat{\alpha}_2^*)\widehat{\Sigma}_{\mathrm{pool}} \xrightarrow{a.s.} \Sigma_g \ \text{ as } \ n \to \infty,$$

for $g = 1, \ldots, G$. In other words, the optimal shrinkage estimators of the covariance matrices under the quadratic loss remain to be consistent.

To investigate the asymptotic properties of $\widehat{\alpha}_2^*$ as $G \to \infty$, we assume that $\Sigma_g \overset{i.i.d.}{\sim} U$, where U is a probability measure supported on $\mathbb{H}^+ := \{A_{p\times p} : A > 0\}$.

Theorem 10 *For any fixed n and p with $n > p+3$, assuming that $E\|S_1\|^2 < \infty$, $E\|S_1^{-1}\|^2 < \infty$, $E\|\Sigma_1\|^2 < \infty$, and $E\|\Sigma_1^{-1}\|^2 < \infty$, we have $\widehat{\alpha}_2^* - \alpha_2^* \xrightarrow{a.s.} 0$ as $G \to \infty$.*

The proof of Theorem 10 is given in Sect. 5.11. Based on this theorem, under some mild conditions, we can borrow information across groups to get a consistent estimator of the optimal shrinkage parameter.

3 Numerical studies

In this section, we conduct simulations to assess the performance of the proposed shrinkage estimators. As the shrinkage parameter plays an important role in the shrinkage method, we first provide some insight into the limiting behaviors of $\widehat{\alpha}_1^*$ and $\widehat{\alpha}_2^*$ as the number of groups increases.

In the first simulation, we consider equal covariance matrices. We first generate a random matrix $W_1 \sim \text{Wishart}(df, W_0)$, where $df = 2(p+1)$ is the degrees of freedom, $W_0 = \text{diag}(1/df, \ldots, 1/df)$ is a diagonal matrix, and p is the dimension of W_0. We then let $R_1 = \{\text{diag}(W_1)\}^{-1/2} W_1 \{\text{diag}(W_1)\}^{-1/2}$ be the correlation matrix. Finally, we generate a diagonal matrix D_1 such that $D_1 = \text{diag}(\sigma_{1,11}, \ldots, \sigma_{1,pp})$ where $\sigma_{1,ii}^2 \overset{\text{i.i.d.}}{\sim} \text{Scale-inv-}\chi^2(\nu, \tau^2)$ with the degrees of freedom $\nu = 5$ and the scaling parameter $\tau^2 = 1$, and set $\Sigma_1 = \cdots = \Sigma_G = D_1 R_1 D_1$. In the second simulation, we consider unequal covariance matrices where $\Sigma_1, \ldots, \Sigma_G$ are generated to be different. Specifically, we first generate the correlation matrices $R_1, \ldots, R_G$ independently with the diagonal matrices $D_1, \ldots, D_G$ following the same algorithm as in the first study. We then let $\Sigma_g = D_g R_g D_g$ for $g = 1, \ldots, G$. In each scenario, we consider $p = 100$ and $N = 120$. Then for each group, we generate the data $\{\boldsymbol{X}_{gi}\}_{i=1}^N$ i.i.d. from $N_p(\mathbf{0}, \Sigma_g)$.

Figure 1 displays the average values of the shrinkage parameter estimates, together with the upper and lower (mean±1.96∗standard error) bounds, for different numbers of groups under the Stein and quadratic loss functions. With the results based on 500 simulations, we have the main findings as follows.

(I) Under the Stein loss function (see the first row of Fig. 1), when the covariance matrices are all equal, $\widehat{\alpha}_1^*$ tends to 0 as the number of groups tends to large. This phenomenon is consistent with the results in Theorems 1 and 5. As described in Theorem 1, if Σ_g are all equal, the optimal shrinkage parameter is $\alpha_1^* = 0$. Then according to Theorem 5, under some mild conditions, we have $\widehat{\alpha}_1^* - \alpha_1^* = 0$ as $G \to \infty$. This leads to $\widehat{\alpha}_1^* \to 0$ as $G \to \infty$. On the other hand, if the covariance matrices are not all equal, the optimal shrinkage parameter will be larger than 0. Consequently, $\widehat{\alpha}_1^*$ will become more stable and is close to α_1^* when the number of groups is large.

(II) Under the quadratic loss function (see the second row of Fig. 1), the optimal shrinkage parameter estimator, $\widehat{\alpha}_2^*$, behave similarly as $\widehat{\alpha}_1^*$. According to Theorem 6, if Σ_g are all equal, the optimal shrinkage parameter $\alpha_2^* = 0$; otherwise, we have $\alpha_2^* \in (0, 1)$. Also by Theorem 10, we have $\widehat{\alpha}_2^* \to \alpha_2^*$ as $G \to \infty$. This leads to $\widehat{\alpha}_2^* \to 0$ as $G \to \infty$. On the other hand, if the covariance matrices are not all equal, $\widehat{\alpha}_2^*$ will converge to $\alpha_2^* > 0$.

Next, to visualize the risk functions for more insights, we also plot the average values of $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ and $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ along with different values of p, G and N. Specifically in Fig. 2, the simulated curves of the risk functions are displayed with $p = G = 5$ and $N = 8$, 100 or 1000. It is evident that $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ and $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ are strictly convex functions, which are the same as described in Theorems 1 and 6, respectively. In addition, from the top two panels of Fig. 2 with equal covariance matrices, both $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ and $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ converge to a constant function of α when N is large, which coincides with part (a) in Theorems 2 and 7, respectively. Whereas from the bottom two panels of Fig. 2 with unequal covariance matrices, the optimal shrinkage parameters, α_1^* and α_2^*, both converge to 1 when N is large, which also coincides with part (b) in Theorems 2 and 7, respectively.

In what follows, we conduct simulations to evaluate the numerical performance of our new estimators and compare them with some existing methods. Specifically, we consider two linear shrinkage methods proposed by Ledoit and Wolf (2004) and Schäfer and Strimmer (2005), one nonlinear shrinkage method proposed by Ledoit and Wolf (2012), one joint estimation method based on the Cholesky decomposition (Pourahmadi et al. 2007), the method of estimating the covariance matrices all by the conventional pooled sample covariance matrix. Here, for the two linear and the one nonlinear shrinkage methods we perform shrinkage on each group individually. For convenience, we denote these five estimators as $\widehat{\Sigma}_{\text{ls}}$, $\widehat{\Sigma}_{\text{strim}}$, $\widehat{\Sigma}_{\text{nls}}$, $\widehat{\Sigma}_{\text{chol}}$ and $\widehat{\Sigma}_{\text{pool}}$, respectively. For the dimension and the sample size, we consider three scenarios: (i) p is

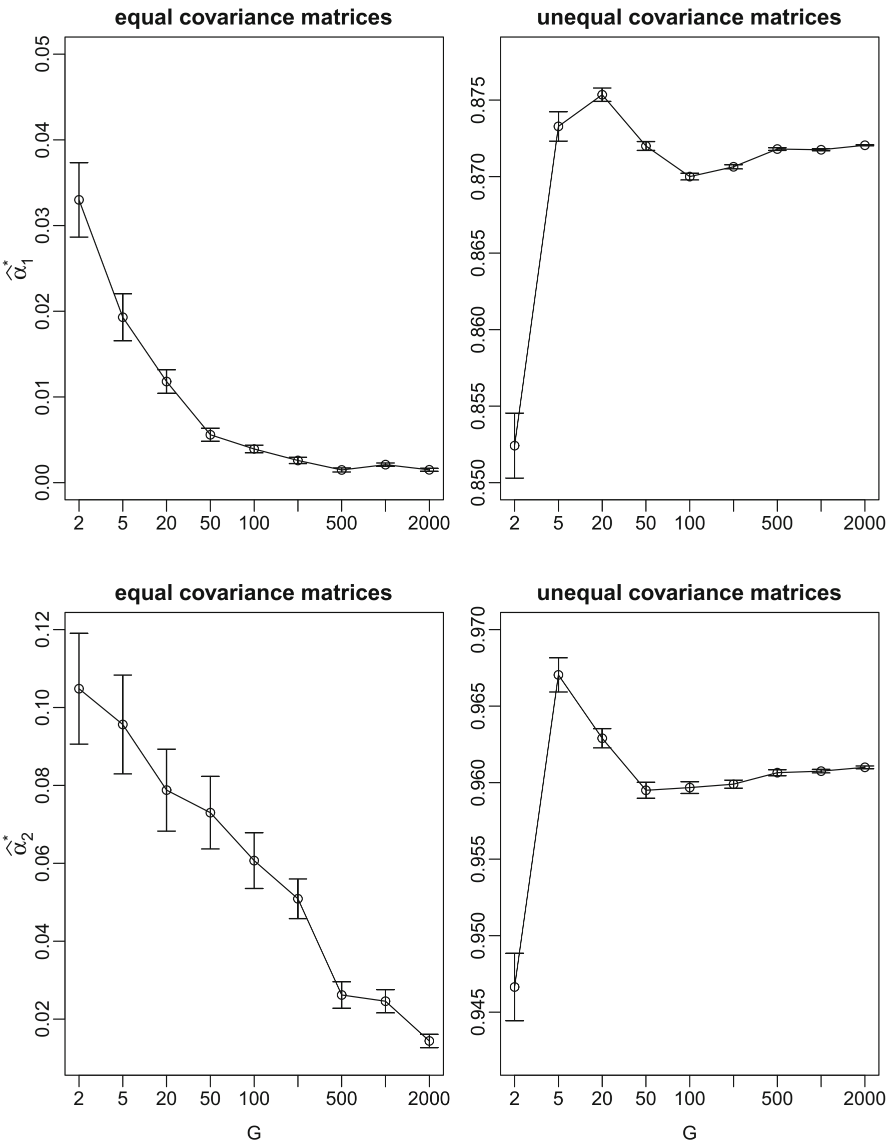

Fig. 1 Plots of the average values of $\widehat{\alpha}_1^*$ (top panels) and $\widehat{\alpha}_2^*$ (bottom panels), together with the upper and lower bounds under the settings of equal and unequal covariance matrices, where G ranges from 2 to 2000

much smaller than N with three combinations as ($p = 2$, $N = 20$), ($p = 2$, $N = 50$) and ($p = 2$, $N = 100$); (ii) p is half of N with three combinations as ($p = 20$, $N = 40$), ($p = 50$, $N = 100$) and ($p = 100$, $N = 200$); (iii) p is close to N with three combinations as ($p = 50$, $N = 70$), ($p = 100$, $N = 120$) and ($p = 300$, $N = 320$). In addition, we consider five numbers of groups as $G = 2, 5, 20, 100$ and 500. As in Ledoit and Wolf (2004), we consider the percentage relative improvement in average

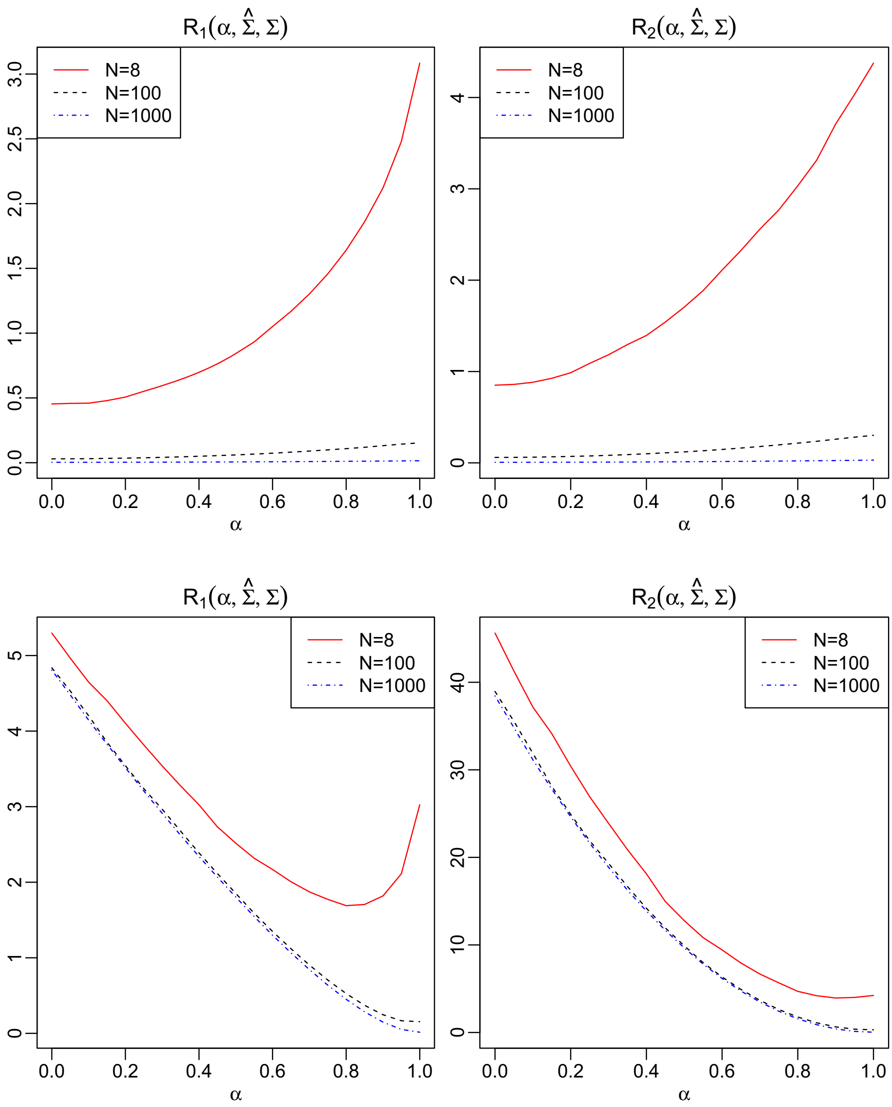

Fig. 2 Plots of the average values of $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ and $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ under the settings of equal (top panels) and unequal (bottom panels) covariance matrices, where α ranges from 0 to 1

loss (PRIAL) over the sample covariance matrices in our comparison. Specifically, for the loss function across G groups defined by $L_{\text{all}}(\widehat{\Sigma}, \Sigma) = \sum_{g=1}^{G} L(\widehat{\Sigma}_g, \Sigma_g)/G$, where $L(\cdot, \cdot)$ is a specified risk function such as $L_1(\cdot, \cdot)$ or $L_2(\cdot, \cdot)$, the PRIAL is given as

$$\text{PRIAL}(\widehat{\Sigma}, \widehat{\Sigma}_{\text{sam}}) = \left\{1 - \frac{\text{AL}(\widehat{\Sigma}, \Sigma)}{\text{AL}(\widehat{\Sigma}_{\text{sam}}, \Sigma)}\right\} \times 100\%,$$

where $\widehat{\Sigma}_{\text{sam}} = \{S_1/n, \ldots, S_G/n\}$, $\text{AL}(\widehat{\Sigma}, \Sigma) = \sum_{k=1}^{M} L_{\text{all}}^{(k)}/M$ is the average of L_{all} after M simulation repetitions. For simplicity, we denote $\text{PRIAL}(\widehat{\Sigma}, \widehat{\Sigma}_{\text{sam}})$ as $\text{PRIAL}(\widehat{\Sigma})$. Therefore, the PRIAL of $\widehat{\Sigma}_{\text{sam}}$ is 0%, meaning no improvement, by contrast, a positive PRIAL indicates that the evaluated estimator performs better than the sample covariance matrices. We set $M = 500$ throughout the simulations.

Figures 3 and 4 show the PRIALs for each of the three scenarios under the Stein and quadratic loss functions, respectively. First of all, we note that $\text{PRIAL}(\widehat{\Sigma}_{\text{pool}})$ and $\text{PRIAL}(\widehat{\Sigma}_{\text{chol}})$ are always far below 0, and hence do not present their simulation results in the figures (for details, see Tables 1 and 2 in online Appendix). From Fig. 3, it is evident that the new estimator outperforms the other five estimators in most settings. In particular, when p is relatively small and N is large, the new estimator is nearly as good as the sample covariance matrices in $\widehat{\Sigma}_{\text{sam}}$. Based on Theorem 2, when the sample size is large, there is no need to borrow information across other groups. When $p = N/2$, we note that the new estimator does borrow information across groups so that the loss can be reduced substantially across all different values of G. In the setting that p is close to N, the new proposed estimator can still consistently outperforms the other five estimators when $G > 5$. This is mainly because that the sample covariance matrices are not stable when p is close to N, and hence borrowing information across a large number of groups can reduce loss substantially.

Figure 4 shows the PRIALs relative to $\widehat{\Sigma}_{\text{sam}}$ under the quadratic loss function. Overall, the new method presents a favorable performance when compared with the existing approaches. In particular, when $p = N/2$, the new estimator, $\widehat{\Sigma}_{\text{new}}$, is the only one that provides a better performance than the sample covariance matrices. In addition, when the dimension is large but close to the sample size (e.g., $p = 300$, $N = 320$), our new estimator outperforms the other five estimators except for the sample covariance matrices. However, as the number of groups is large, the new estimator still has the best performance. One reason for this is that when getting an estimator of b in (5), we need to replace the unknown parameter $\overline{\Sigma}$ by $\widehat{\Sigma}_{\text{pool}}$. Therefore, as the dimension of covariance matrices goes to high, it is necessary to borrow information from a large number of groups. As we can see, when the number of groups goes to large, our optimal shrinkage estimator consistently has the best performance among the six estimators including $\widehat{\Sigma}_{\text{ls}}$, $\widehat{\Sigma}_{\text{strim}}$, $\widehat{\Sigma}_{\text{nls}}$, $\widehat{\Sigma}_{\text{chol}}$, $\widehat{\Sigma}_{\text{pool}}$ and $\widehat{\Sigma}_{\text{sam}}$.

4 Conclusion

In this paper, we develop a shrinkage framework for jointly estimating multiple covariance matrices across groups. We derive the optimal shrinkage parameters under the Stein and quadratic loss functions. We also propose estimators for the optimal shrinkage parameters, and study their asymptotic behaviors under different scenarios. Simulation results demonstrate that, when the number of groups is large, our proposed optimal shrinkage estimators perform better than the existing methods including the estimators of the covariance matrices based on the individual estimation as well as other existing joint estimators. In addition, unlike the structure specified estimators in which the imposed structure assumption is often hard to verify, our new method can automatically borrow information across groups through a shrinkage framework

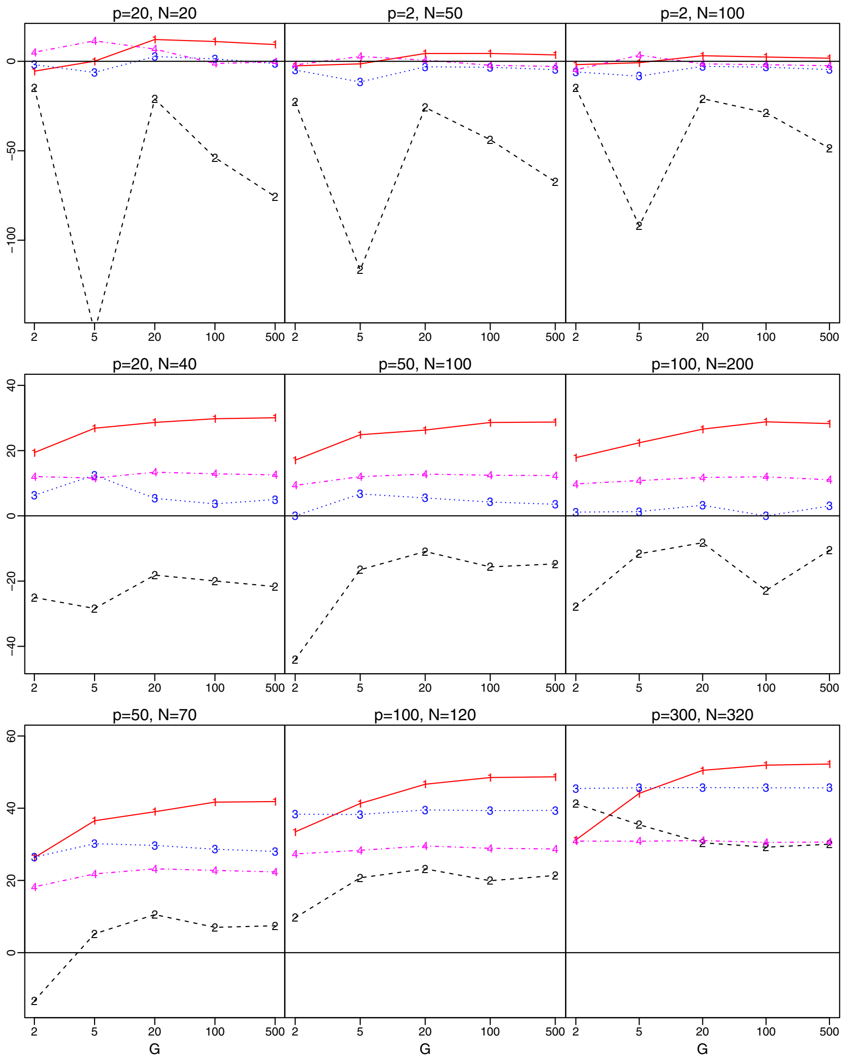

Fig. 3 Plots of PRIALs under the Stein loss function. Lines with "1" to "4" are the PRIALs of $\widehat{\Sigma}_{\text{new}}$, $\widehat{\Sigma}_{\text{ls}}$, $\widehat{\Sigma}_{\text{strim}}$, and $\widehat{\Sigma}_{\text{nls}}$, respectively, where G ranges from 2 to 500. Rows from top to bottom correspond to three scenarios: (i) p is much smaller than N; (ii) p is half of N; (iii) p is close to N. The horizontal black lines present PRIAL $= 0$

without any structure assumption. Our additional simulations in the online Appendix also demonstrate that the structure specified estimators may perform poorly when the imposed structure is incorrect.

In this paper, we have focused on a balanced sample size and a common shrinkage intensity for all groups. In practice, however, the sample size and the shrinkage intensity may vary across groups. For example, when the sample covariance matrix in the gth group has a higher estimation risk, a smaller shrinkage intensity α_g ought to

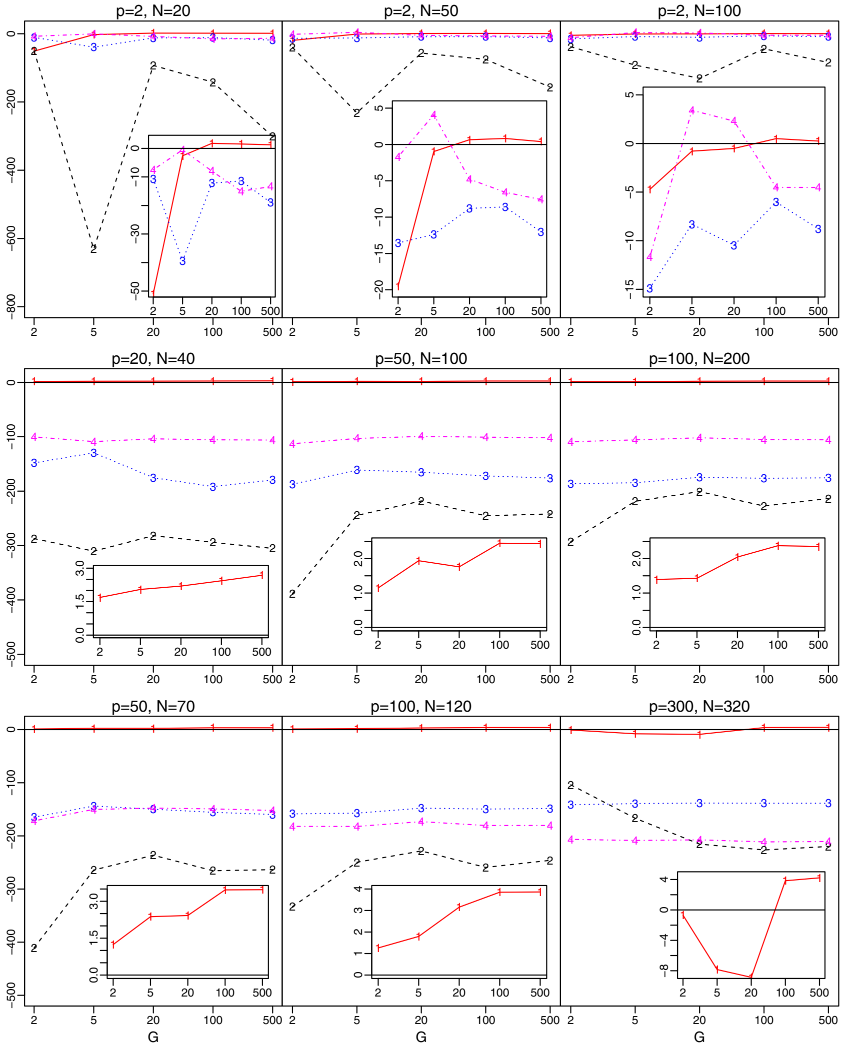

Fig. 4 Plots of PRIALs under the quadratic loss function. Lines with "1" to "4" are the PRIALs of $\widehat{\Sigma}_{\text{new}}$, $\widehat{\Sigma}_{\text{ls}}$, $\widehat{\Sigma}_{\text{strim}}$, and $\widehat{\Sigma}_{\text{nls}}$, respectively, where G ranges from 2 to 500. Rows from top to bottom correspond to three scenarios: (i) p is much smaller than N; (ii) p is half of N; (iii) p is close to N. The horizontal black line presents PRIAL $= 0$

be preferred. One possible direction may use Bayesian shrinkage and impose some priors for α_g, $g = 1, \ldots, G$. Consequently, the shrinkage intensities can be estimated via the posterior distribution. In addition, deriving the boundaries for the estimated shrinkage intensities may be an interesting future work. The derivation may involve the individual probability distribution under a finite sample size. Specifically, the probability distributions of the shrinkage intensities are determined by the sample size, the number of groups and the scalar functions of the true (yet unobservable) covariance matrices, $\Sigma_1, \ldots, \Sigma_G$, which are intractable under a finite sample size and groups. Future research is warranted.

5 Proofs

To prove Theorems 1 to 10, we first show some results on matrix calculation, together with some key lemmas. For any $p \times p$ matrix A, when the eigenvalues of A are all reals, we denote them as

$$\lambda_1(A) \geq \cdots \geq \lambda_p(A).$$

It is well known that if A is a symmetric matrix, then all its eigenvalues are reals, and further more, we have $\lambda_p(A) > 0$ for any $A > 0$. By the Weilandt–Hoffman inequality (see (1.67) in Tao (2012), page 55),

$$|\lambda_p(A+B) - \lambda_p(A)|^2 \leq \sum_{i=1}^{p} |\lambda_i(A+B) - \lambda_i(A)|^2 \leq \|B\|^2 \tag{8}$$

holds for any symmetric matrices A and B.

Lemma 1 *For any $p \times p$ matrices A and B, we have*

$$|\mathrm{tr}(A)| \leq \sqrt{p}\,\|A\|, \quad \|AB\|^2 \leq \|A\|^2\|B\|^2.$$

Proof By the Cauchy–Schwarz inequality,

$$|\mathrm{tr}(A)| \leq \sum_{i=1}^{p} |a_{ii}| \leq \sqrt{p\sum_{i=1}^{p} a_{ii}^2} \leq \sqrt{p}\,\|A\|$$

and

$$\begin{aligned}\|AB\|^2 &= \sum_{i,j=1}^{p} (AB)_{ij}^2 = \sum_{i,j=1}^{p} \left(\sum_{k=1}^{p} A_{ik}B_{kj}\right)^2 \leq \sum_{i,j=1}^{p} \left(\sum_{k=1}^{p} A_{ik}^2 \sum_{l=1}^{p} B_{lj}^2\right) \\ &= \sum_{i=1}^{p}\sum_{k=1}^{p} A_{ik}^2 \sum_{j=1}^{p}\sum_{l=1}^{p} B_{lj}^2 = \|A\|^2\|B\|^2.\end{aligned}$$

The proof of Lemma 1 is complete. □

Lemma 2 *Let $A > 0$ and $B > 0$, then*

$$\mathrm{tr}\{(A-B)(B^{-1} - A^{-1})\} \geq 0, \tag{9}$$

where the equality holds if and only if $A = B$.

Proof Since $A > 0$ and $B > 0$, we have $\mathrm{tr}((A-B)(B^{-1}-A^{-1})) = \mathrm{tr}((A-B)B^{-1}(A-B)A^{-1}) = \mathrm{tr}(A^{-1/2}(A-B)B^{-1}(A-B)A^{-1/2}) > 0$.

Lemma 3 *Assume that* $\alpha A + B > 0$ *for any* $\alpha \in [0, 1]$ *and define*

$$f(\alpha) = \log\det\big(\alpha A + B\big).$$

Then for any $\alpha \in [0, 1]$*, we have* $f''(\alpha) < 0$ *and*

$$f'(\alpha) = \mathrm{tr}\big\{A(\alpha A + B)^{-1}\big\}. \tag{10}$$

Proof Let $\zeta_1, \ldots, \zeta_p$ be the eigenvalues of AB^{-1}. Then $I + B^{-1/2}AB^{-1/2}$ has eigenvalues $\{1+\zeta_1, \ldots, 1+\zeta_p\}$. Since $I + B^{-1/2}AB^{-1/2} > 0$, we have $1 + \zeta_i > 0$ for $i = 1, \ldots, p$. Let also

$$g(\alpha) = \log\det\big\{(\alpha A + B)B^{-1}\big\} = \log\det\big(I + \alpha AB^{-1}\big) = \sum_{i=1}^{p}\log\big(1+\alpha\zeta_i\big).$$

Then we have

$$\begin{aligned} f'(\alpha) = g'(\alpha) &= \sum_{i=1}^{p}\frac{\zeta_i}{1+\alpha\zeta_i} \\ &= \mathrm{tr}(B^{-1/2}AB^{-1/2}(I+\alpha B^{-1/2}AB^{-1/2})) = \mathrm{tr}(\mathrm{A}(\mathrm{A}+\mathrm{B})^{-1}), \\ f''(\alpha) = g''(\alpha) &= -\sum_{i=1}^{p}\frac{\zeta_i^2}{(1+\alpha\zeta_i)^2} \le 0. \end{aligned} \tag{11}$$

The proof of Lemma (3) is complete. □

Lemma 4 *For any* $A, B, C > 0$ *and* $0 \le \alpha \le 1$*, we have*

$$\begin{aligned} &(A-B)\big\{\alpha A + (1-\alpha)B\big\}^{-1} - (A-C)\big\{\alpha A + (1-\alpha)C\big\}^{-1} \\ &\quad = A\big\{\alpha A + (1-\alpha)B\big\}^{-1}(C-B)\big\{\alpha A + (1-\alpha)C\big\}^{-1}. \end{aligned} \tag{12}$$

Proof For $\alpha = 1$, (12) is clear. When $\alpha \ne 1$, by noting that

$$\begin{aligned} &(A-B)\{\alpha A + (1-\alpha)B\}^{-1} \\ &= \Big[(1-\alpha)^{-1}A - \frac{1}{1-\alpha}\{\alpha A + (1-\alpha)B\}\Big]\{\alpha A + (1-\alpha)B\}^{-1} \\ &= (1-\alpha)^{-1}A\{\alpha A + (1-\alpha)B\}^{-1} - (1-\alpha)^{-1}I, \end{aligned}$$

and, similarly, $(A-C)\{\alpha A + (1-\alpha)C\}^{-1} = (1-\alpha)^{-1}A\{\alpha A + (1-\alpha)C\}^{-1} - (1-\alpha)^{-1}I$, we have

$$(A-B)\{\alpha A + (1-\alpha)B\}^{-1} - (A-C)\{\alpha A + (1-\alpha)C\}^{-1}$$

$$
\begin{aligned}
&= \frac{1}{1-\alpha}A\{\alpha A+(1-\alpha)B\}^{-1} - \frac{1}{1-\alpha}A\{\alpha A+(1-\alpha)C\}^{-1} \\
&= \frac{1}{1-\alpha}A\{\alpha A+(1-\alpha)B\}^{-1}\{\alpha A+(1-\alpha)C\}\{\alpha A+(1-\alpha)C\}^{-1} \\
&\quad -\frac{1}{1-\alpha}A\{\alpha A+(1-\alpha)B\}^{-1}\{\alpha A+(1-\alpha)B\}\{\alpha A+(1-\alpha)C\}^{-1} \\
&= A\{\alpha A+(1-\alpha)B\}^{-1}(C-B)\{\alpha A+(1-\alpha)C\}^{-1}.
\end{aligned}
$$

□

Lemma 5 *For any $A > 0$ and $B > 0$, we have*

$$
\sup_{0\le\alpha\le 1} \left\|\{\alpha A+(1-\alpha)B\}^{-1}\right\| \le \sqrt{p}\left\|A^{-1}\right\|\left(1+\|A\|\left\|B^{-1}\right\|\right). \tag{13}
$$

Proof By the minimax formula for eigenvalues (see Theorem 1.3.2 in Tao (2012), page 49)

$$
\begin{aligned}
\left\|\{\alpha A+(1-\alpha)B\}^{-1}\right\|^2 &= \mathrm{tr}\left[\left\{\alpha A+(1-\alpha)B\right\}^{-2}\right] \\
&= \mathrm{tr}\left(A^{-1/2}WA^{-1}WA^{-1/2}\right) \\
&\le p \max_{\|v\|=1} v^T A^{-1/2}WA^{-1}WA^{-1/2}v \\
&\le p \max_{\|v\|=1} v^T A^{-1}v \max_{\|v\|=1} \frac{v^T A^{-1/2}W^2A^{-1/2}v}{v^T A^{-1}v} \\
&\quad \max_{\|v\|=1} \frac{v^T A^{-1/2}WA^{-1}WA^{-1/2}v}{v^T A^{-1/2}W^2A^{-1/2}v} \\
&\le p \max_{\|v\|=1} v^T A^{-1}v \max_{\|v\|=1} v^T W^2 v \max_{\|v\|=1} v^T A^{-1}v \\
&= p\left\{\lambda_1\left(A^{-1}\right)\right\}^2 \lambda_1\left(\left\{\alpha I+(1-\alpha)A^{-1/2}BA^{-1/2}\right\}^{-2}\right) \\
&= p\left\{\lambda_1\left(A^{-1}\right)\right\}^2 \left\{\alpha+(1-\alpha)\lambda_p\left(A^{-1/2}BA^{-1/2}\right)\right\}^{-2},
\end{aligned} \tag{14}
$$

where $W = W(\alpha, A, B) = \{\alpha I + (1-\alpha)A^{-1/2}BA^{-1/2}\}^{-1} > 0$, $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^p$, $\|v\| = (v_1^2 + \cdots + v_n^2)^{-1/2}$. By using the similar method as that in (14), we can get that $\lambda_p\left(A^{-1/2}BA^{-1/2}\right) = \min_{\|v\|=1} v^T A^{-1/2}BA^{-1/2}v \ge \min_{\|v\|=1} v^T Bv \min_{\|v\|=1} v^T A^{-1}v = \lambda_p(B)\lambda_p(A^{-1})$. Hence

$$
\begin{aligned}
\sup_{0\le a\le 1} \left\|\{\alpha A+(1-\alpha)B\}^{-1}\right\| &\le \sqrt{p}\lambda_1\left(A^{-1}\right) \sup_{0\le a\le 1} \{\alpha+(1-\alpha)\lambda_p\left(B\right)\lambda_p\left(A^{-1}\right)\}^{-1} \\
&= \sqrt{p}\lambda_1\left(A^{-1}\right)\max\left\{1, \frac{1}{\lambda_p(B)\lambda_p(A^{-1})}\right\} \\
&\le \sqrt{p}\lambda_1\left(A^{-1}\right)\left\{1+\lambda_1(A)\lambda_1\left(B^{-1}\right)\right\}
\end{aligned}
$$

$$\leq \sqrt{p}\,\|A^{-1}\|(1+\|A\|\,\|B^{-1}\|),$$

where we have used $\lambda_1(A) \leq \|A\|$ for any $A > 0$ since (8). The proof is complete. □

5.1 Proof of Theorem 1

Now for any $g \in \{1, \dots, G\}$, applying Lemma 3 with $A = S_g/n - \widehat{\Sigma}_{\text{pool}}$, $B = \widehat{\Sigma}_{\text{pool}}$, we have that for $\alpha \in [0, 1]$,

$$\begin{aligned}
&\Big[\log\det\big[\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\}\Sigma_g^{-1}\big]\Big]'_\alpha \\
&\quad = \big[\log\det\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\}\big]'_\alpha \\
&\quad = \text{tr}\big[(S_g/n - \widehat{\Sigma}_{\text{pool}})\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\}^{-1}\big],
\end{aligned}$$

and

$$\Big[\log\det\big[\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\}\Sigma_g^{-1}\big]\Big]''_\alpha = \big[\log\det\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\}\big]''_\alpha < 0.$$

Hence

$$R_1''(\alpha, \widehat{\Sigma}, \Sigma) = -\frac{1}{G}\sum_{g=1}^{G} E\Big[\log\det\big[\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\}\Sigma_g^{-1}\big]\Big]''_\alpha > 0,$$

which implies that $R_1(\alpha, \widehat{\Sigma}, \Sigma)$ is a strictly convex function on $[0, 1]$, and

$$\begin{aligned}
R_1'(\alpha, \widehat{\Sigma}, \Sigma) = p - \text{tr}&\left\{\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\Big)\right\} \\
&- \frac{1}{G}\sum_{g=1}^{G}\text{tr}E\big[(S_g/n - \widehat{\Sigma}_{\text{pool}})\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\}^{-1}\big].
\end{aligned}$$

Specially,

$$R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} = p - \text{tr}\Big\{\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\Big)\Big\}, \tag{15}$$

and

$$R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=1} = \text{tr}E\Big\{\Big(\frac{1}{G}\sum_{g=1}^{G}S_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}S_g^{-1}\Big)\Big\}$$

$$-\operatorname{tr}\left\{\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\right)\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\right)\right\}. \tag{16}$$

For any $g, g' \in \{1, \ldots, G\}$, by Lemma 2, we have $\operatorname{tr}[(\Sigma_g - \Sigma_{g'})(\Sigma_{g'}^{-1} - \Sigma_g^{-1})] \geq 0$, which implies $\operatorname{tr}(\Sigma_g \Sigma_{g'}^{-1} + \Sigma_{g'}\Sigma_g^{-1}) \geq 2p$. Thus

$$\operatorname{tr}\left\{\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\right)\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\right)\right\} \geq p \tag{17}$$

holds for any $\Sigma_g > 0$. This together with (15) implies $R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} \leq 0$. And according to Lemma 2, $R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} = 0$ if and only if $\Sigma_1 = \cdots = \Sigma_G$. This implies that for any fixed G, p and $n > p+1$, if Σ_g are all the same, we have $\alpha_1^* = 0$.

For any $g \in \{1, \ldots, G\}$, since $S_g \sim W_p(\Sigma_g, n)$, we have $ES_g = n\Sigma_g$, $S_g^{-1} \sim W_p^{-1}(\Sigma_g^{-1}, n)$, and $ES_g^{-1} = \Sigma_g^{-1}/(n-p-1)$. Hence, for any $g, g' \in \{1, \ldots, G\}$ with $g \neq g'$, we have $\operatorname{tr}\{E(S_{g'}S_g^{-1}) - \Sigma_{g'}\Sigma_g^{-1}\} = \operatorname{tr}\{E(S_{g'})E(S_g^{-1}) - \Sigma_{g'}\Sigma_g^{-1}\} = (p+1)\operatorname{tr}(\Sigma_g^{-1/2}\Sigma_{g'}\Sigma_g^{-1/2})/(n-p-1) > 0$ This, together with (16) and (17), implies (b).

5.2 Proof of Theorem 2

As shown in Theorem 1, $R_1''(\alpha, \hat{\Sigma}, \Sigma) > 0$ for any fixed n. This indicates that $R_1''(\alpha, \hat{\Sigma}, \Sigma) \geq 0$ as $n \to \infty$. For any $g = 1, \ldots, G$, when $n \to \infty$, we have

$$\begin{aligned} R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=1} &= \operatorname{tr}E\left\{\left(\frac{1}{G}\sum_{g=1}^{G}S_g\right)\left(\frac{1}{G}\sum_{g=1}^{G}S_g^{-1}\right)\right\} \\ &\quad - \operatorname{tr}\left\{\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\right)\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\right)\right\} \\ &= \frac{p+1}{G^2(n-p-1)}\sum_{g\neq g'}\operatorname{tr}\left(\Sigma_g\Sigma_{g'}^{-1}\right) \to 0. \end{aligned}$$

Note also that when $\Sigma_g = \Sigma_{g'}$ for any $g = g'$,

$$R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} = p - \operatorname{tr}\left\{\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\right)\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\right)\right\} = 0,$$

thus we have $R_1'(\alpha, \hat{\Sigma}, \Sigma) = 0$ as $n \to \infty$. Further by the Mean Value Theorem for Derivatives, for any $\alpha_1 \neq \alpha_2$, we have $R_1(\alpha_1, \hat{\Sigma}, \Sigma) - R_1(\alpha_2, \hat{\Sigma}, \Sigma) = (\alpha_1 -$

$\alpha_2)R_1^{'}(\xi, \hat{\Sigma}, \Sigma)$ with $\alpha_1 < \xi < \alpha_2$. This shows that $R_1(\alpha_1, \hat{\Sigma}, \Sigma) = R_1(\alpha_2, \hat{\Sigma}, \Sigma)$ as $n \to \infty$.

If Σ_g are not all the same, then by using SLLN and the dominated convergence theorem, we have that, for any $1 \leq \alpha \leq 1$,

$$R_1''(\alpha, \widehat{\Sigma}, \Sigma) = \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} E\frac{\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}{\left\{1 + \alpha\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}$$
$$\geq \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} E\frac{\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}{2 + 2\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}$$
$$\to \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} \frac{\left\{\lambda_i\left(\overline{\Sigma}^{-1/2}(\Sigma_g - \overline{\Sigma})\overline{\Sigma}^{-1/2}\right)\right\}^2}{2 + 2\left\{\lambda_i\left(\overline{\Sigma}^{-1/2}(\Sigma_g - \overline{\Sigma})\overline{\Sigma}^{-1/2}\right)\right\}^2} > 0.$$

Hence, (b) follows by the fact that $R_1''(\alpha, \widehat{\Sigma}, \Sigma) > 0$ and $R_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=1} \to 0$.

5.3 Proof of Theorem 3

To get an estimator of α_1^*, we need to find an estimator of $R_1'(\alpha, \widehat{\Sigma}, \Sigma)$. From the proof of Theorem 1, $R_1'(\alpha, \widehat{\Sigma}, \Sigma)$ is given as

$$R_1'(\alpha, \widehat{\Sigma}, \Sigma) = p - \text{tr}\left\{\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\right)\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\right)\right\}$$
$$- \frac{1}{G}\sum_{g=1}^{G}\text{tr}E\left[\left(S_g/n - \widehat{\Sigma}_{\text{pool}}\right)\left\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\right\}^{-1}\right]. \quad (18)$$

We estimate Σ_g and Σ_g^{-1} by S_g/n and $(n-p-1)S_g^{-1}$, respectively. Then, an estimator of $R_1'(\alpha, \widehat{\Sigma}, \Sigma)$ is given as

$$\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma) = p - \text{tr}\left[\left(\frac{1}{G}\sum_{g=1}^{G}\frac{S_g}{n}\right)\left\{\frac{1}{G}\sum_{g=1}^{G}(n-p-1)S_g^{-1}\right\}\right]$$
$$- \frac{1}{G}\sum_{g=1}^{G}\text{tr}\left[\left(S_g/n - \widehat{\Sigma}_{\text{pool}}\right)\left\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\right\}^{-1}\right] - \frac{p(p+1)}{nG}. \quad (19)$$

We note that

$$E\{\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)\} - R_1'(\alpha, \widehat{\Sigma}, \Sigma)$$

$$= \mathrm{tr}\Big\{\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\Big)\Big\}$$
$$- E\Big[\mathrm{tr}\Big[\Big(\frac{1}{G}\sum_{g=1}^{G}\frac{S_g}{n}\Big)\Big\{\frac{1}{G}\sum_{g=1}^{G}(n-p-1)S_g^{-1}\Big\}\Big]\Big] - \frac{p(p+1)}{nG} = 0.$$

Hence, $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)$ is an unbiased estimator.

From the proof of Theorem 1, for any $\alpha \in [0, 1]$, we have

$$\widehat{R}_1''(\alpha, \widehat{\Sigma}, \Sigma) = -\frac{1}{G}\sum_{g=1}^{G}\Big[\log\det\big\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\mathrm{pool}}\big\}\Big]_\alpha'' > 0,$$

which implies that $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)$ is strictly increasing on $[0, 1]$.

In addition, we note that

$$\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} = p - \mathrm{tr}\Big[\Big(\frac{1}{G}\sum_{g=1}^{G}\frac{S_g}{n}\Big)\Big\{\frac{1}{G}\sum_{g=1}^{G}(n-p-1)S_g^{-1}\Big\}\Big] - \frac{p(p+1)}{nG},$$
$$\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=1} = \frac{p+1}{n}\mathrm{tr}\Big[\Big(\frac{1}{G}\sum_{g=1}^{G}S_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}S_g^{-1}\Big)\Big] - \frac{p(p+1)}{nG} > 0.$$

The last inequality is based on (17). Here, $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0}$ is not guaranteed to be negative. If $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} \le 0$, then there exists a unique α satisfies $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma) = 0$ and we denote the solution as $\widehat{\alpha}_1^*$. Otherwise, we set $\widehat{\alpha}_1^* = 0$.

5.4 Proof of Theorem 4

By SLLN, for any $g = 1, \ldots, G$ with fixed G and p, we have $S_g/n \xrightarrow{a.s.} \Sigma_g$ as $n \to \infty$, thus

$$\mathrm{tr}\Big\{\Big(\frac{1}{G}\sum_{g=1}^{G}S_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}S_g^{-1}\Big)\Big\} \xrightarrow{a.s.} \mathrm{tr}\Big\{\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\Big)\Big\},$$

and then

$$\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=1} = \frac{p+1}{n}\mathrm{tr}\Big\{\Big(\frac{1}{G}\sum_{g=1}^{G}S_g\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}S_g^{-1}\Big)\Big\} - \frac{p(p+1)}{nG} \xrightarrow{a.s.} 0.$$

By applying (11) and the strong law of large number (SLLN), and noting that $\lambda_i(\cdot)$ is a continuous function for any $i = 1, \ldots, p$, we get that

$$\widehat{R}_1''(\alpha, \widehat{\Sigma}, \Sigma) = \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} \frac{\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}{\left\{1 + \alpha\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}$$

$$\xrightarrow{a.s.} \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} \frac{\left\{\lambda_i\left(\overline{\Sigma}^{-1/2}(\Sigma_g - \overline{\Sigma})\overline{\Sigma}^{-1/2}\right)\right\}^2}{\left\{1 + \alpha\lambda_i\left(\overline{\Sigma}^{-1/2}(\Sigma_g - \overline{\Sigma})\overline{\Sigma}^{-1/2}\right)\right\}^2} > 0,$$

where $\overline{\Sigma} = \sum_{g=1}^{G} \Sigma_g/G$, and the equality holds if and only if $\Sigma_g = \Sigma_{g'}$ for any $g \neq g'$. Hence $\widehat{\alpha}_1^* \xrightarrow{a.s.} 1$ when Σ_g are not all the same.

5.5 Proof of Theorem 5

By (18) and (19), we have

$$\sup_{0\le\alpha\le 1} |R_1'(\alpha, \widehat{\Sigma}, \Sigma) - \widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)| \le \frac{p(p+1)}{nG} + \sum_{i=1}^{5} J_i,$$

where $p(p+1)/(nG) \to 0$ as $G \to \infty$, and

$$J_1 = \left|\text{tr}\left[\left(\frac{1}{G}\sum_{g=1}^{G}\frac{S_g}{n}\right)\left\{\frac{1}{G}\sum_{g=1}^{G}(n-p-1)S_g^{-1}\right\}\right] - \text{tr}\left(E\Sigma_1 E\Sigma_1^{-1}\right)\right|,$$

$$J_2 = \left|\text{tr}\left\{\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g\right)\left(\frac{1}{G}\sum_{g=1}^{G}\Sigma_g^{-1}\right)\right\} - \text{tr}(E\Sigma_1 E\Sigma_1^{-1})\right|,$$

$$J_3 = \frac{1}{G}\sum_{g=1}^{G}\sup_{\alpha\in[0,1]}\left|\text{tr}\left[\left(S_g/n - \widehat{\Sigma}_{\text{pool}}\right)\left\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\right\}^{-1}\right]\right.$$
$$\left. - \text{tr}\left[\left(S_g/n - E\Sigma_1\right)\left\{\alpha S_g/n + (1-\alpha)E\Sigma_1\right\}^{-1}\right]\right|,$$

$$J_4 = \sup_{\alpha\in[0,1]}\left|\frac{1}{G}\sum_{g=1}^{G}\text{tr}\left[\left(S_g/n - E\Sigma_1\right)\left\{\alpha S_g/n + (1-\alpha)E\Sigma_1\right\}^{-1}\right]\right.$$
$$\left. - E\text{tr}\left[\left(S_1/n - E\Sigma_1\right)\left\{\alpha S_1/n + (1-\alpha)E\Sigma_1\right\}^{-1}\right]\right|,$$

$$J_5 = \sup_{\alpha\in[0,1]}\left|\text{tr}E\left[\left(S_1/n - \widehat{\Sigma}_{\text{pool}}\right)\left\{\alpha S_1/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\right\}^{-1}\right]\right.$$
$$\left. - \text{tr}E\left[\left(S_1/n - E\Sigma_1\right)\left\{\alpha S_1/n + (1-\alpha)E\Sigma_1\right\}^{-1}\right]\right|.$$

For fixed n and p, as $G \to \infty$, we need to prove that, $J_i \xrightarrow{a.s.} 0$ for $i = 1, \ldots, 5$.

Since $ES_g = E\{E(S_g|\Sigma_g)\} = nE\Sigma_1$ and $(n-p-1)ES_g^{-1} = (n-p-1)E\{E(S_g^{-1}|\Sigma_g)\} = E\Sigma_1^{-1} > 0$, by the SLLN, we have

$$\widehat{\Sigma}_{\text{pool}} = \frac{1}{G}\sum_{g=1}^{G}\frac{S_g}{n} \xrightarrow{a.s.} E\Sigma_1, \quad \frac{1}{G}\sum_{g=1}^{G}(n-p-1)S_g^{-1} \xrightarrow{a.s.} E\Sigma_1^{-1},$$

which, together with the fact that tr$(\cdot)$ is a continuous function (by Lemma 1), implies that $J_1 \xrightarrow{a.s.} 0$. Similarly, $J_2 \xrightarrow{a.s.} 0$. By applying Lemma 4, we have

$$\begin{aligned}
&\big(S_g/n - \widehat{\Sigma}_{\text{pool}}\big)\big\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\big\}^{-1} - \big(S_g/n - E\Sigma_1\big)\big\{\alpha S_g/n + (1-\alpha)E\Sigma_1\big\}^{-1} \\
&\quad = (S_g/n)\big\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\big\}^{-1}(E\Sigma_1 - \widehat{\Sigma}_{\text{pool}})\big\{\alpha S_g/n + (1-\alpha)E\Sigma_1\big\}^{-1}.
\end{aligned} \tag{20}$$

Then, by Lemmas 1 and 5, we have

$$\begin{aligned}
J_3 &\le \frac{1}{G}\sum_{g=1}^{G}\big\|S_g/n\big\| \sup_{a\in[0,1]}\big\|\big\{\alpha S_g/n + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\big\}^{-1}\big\| \\
&\quad \times \sup_{a\in[0,1]}\big\|\big\{\alpha S_g/n + (1-\alpha)E\Sigma_1\big\}^{-1}\big\|\big\|E\Sigma_1 - \widehat{\Sigma}_{\text{pool}}\big\| \\
&\le \big\|E\Sigma_1 - \widehat{\Sigma}_{\text{pool}}\big\| \times \left\{\frac{1}{G}\sum_{g=1}^{G}T(S_g, \widehat{\Sigma}_{\text{pool}})\right\},
\end{aligned} \tag{21}$$

Here,

$$\begin{aligned}
T(S_g, \widehat{\Sigma}_{\text{pool}}) &= \left\|\frac{S_g}{n}\right\| \sup_{a\in[0,1]}\left\|\left\{\alpha\frac{S_g}{n} + (1-\alpha)\widehat{\Sigma}_{\text{pool}}\right\}^{-1}\right\| \\
&\quad \times \sup_{a\in[0,1]}\left\|\left\{\alpha\frac{S_g}{n} + (1-\alpha)E\Sigma_1\right\}^{-1}\right\| \\
&\le np\|S_g\|\,\|S_g^{-1}\|^2\left(1 + \big\|S_g/n\big\|\,\|(\widehat{\Sigma}_{\text{pool}})^{-1}\|\right)\left(1 + \big\|S_g/n\big\|\,\|(E\Sigma_1)^{-1}\|\right) \\
&= p\big(\|S_g\|\,\|S_g^{-1}\|^2\big)\big(n + \|S_g\|\,\|(E\Sigma_1)^{-1}\|\big) \\
&\quad + p\|(\widehat{\Sigma}_{\text{pool}})^{-1}\|\left(\|S_g\|^2\,\|S_g^{-1}\|^2\right)\left(1 + \|S_g/n\|\,\|(E\Sigma_1)^{-1}\|\right),
\end{aligned}$$

where the inequality is from Lemma 5. By noting that $\|(E\Sigma_1)^{-1}\|$ is a positive constant and recalling that $E\|S_1\|^7 < \infty,\ E\|S_1^{-1}\|^7 < \infty$, it is easy to verify that

$$\begin{aligned}
&E\big\{\|S_1\|^2\,\|S_1^{-1}\|^2\big(1 + \|S_1\|\,\|(E\Sigma_1)^{-1}\|\big)\big\} < \infty, \\
&E\big\{\|S_1\|\,\|S_1^{-1}\|^2\left(1 + \|S_1\|\,\|(E\Sigma_1)^{-1}\|\right)\big\} < \infty.
\end{aligned}$$

In addition, by the SLLN, we have $\|E\Sigma_1 - \widehat{\Sigma}_{\rm pool}\| \xrightarrow{a.s.} 0$, $\|(\widehat{\Sigma}_{\rm pool})^{-1}\| \xrightarrow{a.s.} \|(E\Sigma_1)^{-1}\|$, and

$$\begin{aligned}\frac{1}{G}\sum_{g=1}^{G} T(S_g, \widehat{\Sigma}_{\rm pool}) < \frac{p}{G}\sum_{g=1}^{G}\left(\|S_g\|\,\|S_g^{-1}\|^2\right)\left(n + \|S_g\|\,\|(E\Sigma_1)^{-1}\|\right) \\ + \frac{p}{G}\sum_{g=1}^{G}\|(\widehat{\Sigma}_{\rm pool})^{-1}\|\left(\|S_g\|^2\,\|S_g^{-1}\|^2\right)\left(1 + \|S_g\|\,\|(E\Sigma_1)^{-1}\|\right) \\ \xrightarrow{a.s.} pE\left\{\|S_1\|\|S_1^{-1}\|^2\left(n + \|S_1\|\|(E\Sigma_1)^{-1}\|\right)\right\} \\ + p(\|(E\Sigma_1)^{-1}\|)\, E\big\{\|S_1\|^2 \\ \|S_1^{-1}\|^2\left(1 + \|S_1\|\,\|(E\Sigma_1)^{-1}\|\right)\big\} < \infty.\end{aligned}$$

Consequently, we have $J_3 \xrightarrow{a.s.} 0$.

Note that by Lemmas 1 and 5, we have that for any $0 \le \alpha \le 1$,

$$\begin{aligned}&\left|\mathrm{tr}\left[(S_1/n - E\Sigma_1)\left\{\alpha S_1/n + (1-\alpha)E\Sigma_1\right\}^{-1}\right]\right| \\ &\le \sqrt{p}\,\left\|S_1/n - E\Sigma_1\right\| \sup_{0\le\alpha\le 1}\left\|\{\alpha S_1/n + (1-\alpha)E\Sigma_1\}^{-1}\right\| \\ &\le np\,(\|S_1/n\| + \|E\Sigma_1\|)\,\|S_1^{-1}\|\,\{1 + \|S_1/n\|\,\|(E\Sigma_1)^{-1}\|\} \\ &\le np\left(\|S_1\| + \|E\Sigma_1\|\right)\|S_1^{-1}\|\left(1 + \|S_1\|\,\|(E\Sigma_1)^{-1}\|\right).\end{aligned}$$

Since $E\left[(\|S_1\|+\|E\Sigma_1\|)\,\|S_1^{-1}\|\{1+\|S_1\|\,\|(E\Sigma_1)^{-1}\|\}\right] < \infty$, by applying a uniform SLLN (see Theorem 16(a) in Ferguson (1996)), we have $J_4 \xrightarrow{a.s.} 0$.

By (20), we have

$$\begin{aligned}&E\left[\left(\frac{S_1}{n} - \widehat{\Sigma}_{\rm pool}\right)\left\{\alpha\frac{S_1}{n} + (1-\alpha)\widehat{\Sigma}_{\rm pool}\right\}^{-1}\right] \\ &\quad - E\left[\left(\frac{S_1}{n} - E\Sigma_1\right)\left\{\alpha\frac{S_1}{n} + (1-\alpha)E\Sigma_1\right\}^{-1}\right] \\ &= E\Big[(S_1/n)\left\{\alpha S_1/n + (1-\alpha)\widehat{\Sigma}_{\rm pool}\right\}^{-1} \\ &\quad \left(E\Sigma_1 - \widehat{\Sigma}_{\rm pool}\right)\left\{\alpha S_1/n + (1-\alpha)E\Sigma_1\right\}^{-1}\Big],\end{aligned}$$

and, similarly to (21), we have

$$\begin{aligned}J_5 &\le \sqrt{p}E\left\{T(S_1, \widehat{\Sigma}_{\rm pool})\,\|E\Sigma_1 - \widehat{\Sigma}_{\rm pool}\|\right\} \\ &\le p^{3/2}E\Big[\|S_1\|^2\,\|S_1^{-1}\|^2\,\{\|(\Sigma_{\rm pool})^{-1}\|\} \\ &\quad \{1 + \|S_1\|\,\|(E\Sigma_1)^{-1}\|\}\left(\|E\Sigma_1 - \widehat{\Sigma}_{\rm pool}\|\right)\Big] \\ &\quad + p^{3/2}E\left[\|S_1\|\|S_1^{-1}\|^2\,\{n + \|S_1\|\|(E\Sigma_1)^{-1}\|\}\left(\|E\Sigma_1 - \widehat{\Sigma}_{\rm pool}\|\right)\right]\end{aligned}$$

$$\leq p^{3/2}\left\{1+\|(E\Sigma_1)^{-1}\|\right\}(J_{51}+n\times J_{52}).$$

where

$$J_{51}=E\left\{\|S_1\|\,\|S_1^{-1}\|^2\,(1+\|S_1\|)^2\,\|E\Sigma_1-\widehat{\Sigma}_{\text{pool}}\|\right\},$$
$$J_{52}=E\left\{\|S_1\|^2\,\|S_1^{-1}\|^2\,(1+\|S_1\|)\,\|E\Sigma_1-\widehat{\Sigma}_{\text{pool}}\|\,\|(\widehat{\Sigma}_{\text{pool}})^{-1}\|\right\}.$$

By Hölder's inequality, we have

$$J_{52}\leq\left[E\left\{\|S_1\|^2(1+\|S_1\|)\right\}^{7/3}\right]^{3/7}\left(E\|S_1^{-1}\|^7\right)^{2/7}$$
$$\left(E\|E\Sigma_1-\widehat{\Sigma}_{\text{pool}}\|^7\right)^{1/7}\left\{E\left\|(\widehat{\Sigma}_{\text{pool}})^{-1}\right\|^7\right\}^{1/7}.$$

By the assumption $E\|S_1\|^7<\infty$, we have $E\{\|S_1\|^2(1+\|S_1\|)\}^{7/3}<\infty$, and by the L^p convergence theorem, $E\|E\Sigma_1-\widehat{\Sigma}_{\text{pool}}\|^7=E\|\sum_{g=1}^{G}\{S_g/n-E(S_g/n)\}/G\|^7\to 0$.

Define a function $f:\mathbb{H}^+\to\mathbb{R}$ with

$$f(A)=\|A^{-1}\|^2=\text{tr}(A^{-2})=\sum_{i=1}^{p}1/(\lambda_i(A))^2,\quad A>0.$$

Since $x\mapsto 1/x^2$ is a convex function on $(0,\infty)$, by Klein's lemma (See, for instance, Guionnet (2009), page 78), f is a convex function on $\mathbb{H}^+$. Thus

$$\|(\widehat{\Sigma}_{\text{pool}})^{-1}\|^2=\left\|\left(\frac{1}{G}\sum_{g=1}^{G}\frac{S_g}{n}\right)^{-1}\right\|^2\leq\frac{1}{G}\sum_{g=1}^{G}\left\|\left(\frac{S_g}{n}\right)^{-1}\right\|^2.$$

Then, by Jensen's inequality,

$$E\|(\widehat{\Sigma}_{\text{pool}})^{-1}\|^7\leq E\left(\frac{1}{G}\sum_{g=1}^{G}\|nS_g^{-1}\|^2\right)^{7/2}\leq E\left(\frac{1}{G}\sum_{g=1}^{G}\|nS_g^{-1}\|^7\right)=E\|nS_1^{-1}\|^7<\infty.$$

Combing the above facts, we get that $J_{52}\to 0$. Similarly, $J_{51}\to 0$. Thus $J_5\to 0$.

Finally, we have

$$\sup_{0\leq\alpha\leq 1}\left|R_1'(\alpha,\widehat{\Sigma},\Sigma)-\widehat{R}_1'(\alpha,\widehat{\Sigma},\Sigma)\right|\leq\frac{p(p+1)}{nG}+\sum_{i=1}^{5}J_i\xrightarrow{a.s.}0$$

as $G\to\infty$, for fixed n and p.

As follows, we show that $\widehat{\alpha}_1^*-\alpha_1^*\xrightarrow{a.s.}0$. Noting that

$$\widehat{R}_1'(\alpha,\widehat{\Sigma},\Sigma)|_{\alpha=0}=p-\text{tr}\left[\left(\frac{1}{G}\sum_{g=1}^{G}S_g/n\right)\left\{\frac{1}{G}\sum_{g=1}^{G}(n-p-1)S_g^{-1}\right\}\right]$$

$$\xrightarrow{a.s.} p - \text{tr}\{E(\Sigma_1)E(\Sigma_1^{-1})\}.$$

First of all, we show that the convergence holds if Σ_g are not all the same. Then by Jensen's inequality, we have $\text{tr}\{E(\Sigma_1)E(\Sigma_1^{-1})\} > p$. As a consequence, $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} < 0$ a.s. as $G \to \infty$. Note also that $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=1} > 0$, we have $\widehat{\alpha}_1^* \in (0, 1)$ a.s. as $G \to \infty$.

Since $\widehat{R}_1'(\widehat{\alpha}_1^*, \widehat{\Sigma}, \Sigma) = R_1'(\alpha_1^*, \widehat{\Sigma}, \Sigma) = 0$, we have $|\widehat{R}_1'(\alpha_1^*, \widehat{\Sigma}, \Sigma) - \widehat{R}_1'(\widehat{\alpha}_1^*, \widehat{\Sigma}, \Sigma)| = |\widehat{R}_1'(\alpha_1^*, \widehat{\Sigma}, \Sigma) - R_1'(\alpha_1^*, \widehat{\Sigma}, \Sigma)|$. By the mean value theorem,

$$\begin{aligned} |\alpha_1^* - \widehat{\alpha}_1^*| &\le \left|\widehat{R}_1'(\alpha_1^*, \widehat{\Sigma}, \Sigma) - \widehat{R}_1'(\widehat{\alpha}_1^*, \widehat{\Sigma}, \Sigma)\right| \Big/ \inf_{0\le\alpha\le1} \widehat{R}_1''(\alpha, \widehat{\Sigma}, \Sigma) \\ &= \left|\widehat{R}_1'(\alpha_1^*, \widehat{\Sigma}, \Sigma) - R_1'(\alpha_1^*, \widehat{\Sigma}, \Sigma)\right| \Big/ \inf_{0\le\alpha\le1} \widehat{R}_1''(\alpha, \widehat{\Sigma}, \Sigma). \end{aligned} \tag{22}$$

Note that $R_1'(\alpha, \widehat{\Sigma}, \Sigma) - \widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma) \xrightarrow{a.s.} 0$ uniformly for $\alpha \in [0, 1]$ as $G \to \infty$, we only need to verify that $\inf_{0\le\alpha\le1} \widehat{R}_1''(\alpha, \widehat{\Sigma}, \Sigma) > 0$. By applying (11) and using the similar method as that in (14), we have

$$\begin{aligned} \inf_{0\le\alpha\le1} \widehat{R}_1''(\alpha, \widehat{\Sigma}, \Sigma) &= \inf_{0\le\alpha\le1} \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} \frac{\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}{\left\{1 + \alpha\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2} \\ &\ge \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} \frac{\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2}{2 + 2\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)\right\}^2} \\ &= \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{p} \frac{\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)^2\right\}}{2 + 2\left\{\lambda_i\left(\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\right)^2\right\}} \\ &\ge \frac{1}{4}\frac{1}{G}\sum_{g=1}^{G}\min\left\{1, \lambda_p\{\widehat{\Sigma}_{\text{pool}}^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1}(S_g/n - \widehat{\Sigma}_{\text{pool}})\widehat{\Sigma}_{\text{pool}}^{-1/2}\}\right\} \\ &\ge \frac{1}{4}\frac{1}{G}\sum_{g=1}^{G}\min\left\{1, \lambda_p\{(S_g/n - \widehat{\Sigma}_{\text{pool}})^2\}\{\lambda_p(\widehat{\Sigma}_{\text{pool}}^{-1})\}^2\right\} \\ &\ge \frac{1}{4}\min\left\{1, \{\lambda_p(\widehat{\Sigma}_{\text{pool}}^{-1})\}^2\right\}\frac{1}{G}\sum_{g=1}^{G}\min\left\{1, \lambda_p\{(S_g/n - \widehat{\Sigma}_{\text{pool}})^2\}\right\}. \end{aligned}$$

Note that for any y_1 and y_2, we have $|\min\{1, y_1\} - \min\{1, y_2\}| \le |y_1 - y_2|$. Combing this inequality with (8), we have

$$\left|\frac{1}{G}\sum_{g=1}^{G}\min\left\{1, \lambda_p\{(S_g/n - \widehat{\Sigma}_{\text{pool}})^2\}\right\} - \frac{1}{G}\sum_{g=1}^{G}\min\left\{1, \lambda_p\{(S_g/n - E\Sigma_1)^2\}\right\}\right|$$

$$\leq \frac{1}{G}\sum_{g=1}^{G}\Big|\min\Big\{1, \lambda_p\big\{\big(S_g/n - \widehat{\Sigma}_{\text{pool}}\big)^2\big\}\Big\} - \min\Big\{1, \lambda_p\big\{\big(S_g/n - E\Sigma_1\big)^2\big\}\Big\}\Big|$$

$$\leq \frac{1}{G}\sum_{g=1}^{G}\Big|\lambda_p\big\{\big(S_g/n - \widehat{\Sigma}_{\text{pool}}\big)^2\big\} - \lambda_p\big\{\big(S_g/n - E\Sigma_1\big)^2\big\}\Big|$$

$$\leq \frac{1}{G}\sum_{g=1}^{G}\Big\|\big(S_g/n - \widehat{\Sigma}_{\text{pool}}\big)^2 - \big(S_g/n - E\Sigma_1\big)^2\Big\|.$$

Note that

$$\begin{aligned}(A+B)^2-(A+C)^2 &= (A+B)^2-(A+B)(A+C)\\ &\quad+(A+B)(A+C)-(A+C)^2\\ &=(A+B)(B-C)+(B-C)(A+C)\\ &=(A+C)(B-C)+(B-C)^2+(B-C)(A+C)\end{aligned}$$

holds for any $A, B, C > 0$. This, together with Lemma 1, SLLN and the fact that $E(\|S_1/n - E\Sigma_1\|) \leq E\|S_1\|/n + E\|E\Sigma_1\| < \infty$, yields that

$$\begin{aligned}&\frac{1}{G}\sum_{g=1}^{G}\Big\|\big(S_g/n - \widehat{\Sigma}_{\text{pool}}\big)^2 - \big(S_g/n - E\Sigma_1\big)^2\Big\|\\ &\quad\leq \Big\|E\Sigma_1 - \widehat{\Sigma}_{\text{pool}}\Big\|^2 + \frac{2}{G}\sum_{g=1}^{G}\Big(\big\|E\Sigma_1 - \widehat{\Sigma}_{\text{pool}}\big\|\Big)\Big(\big\|S_g/n - E\Sigma_1\big\|\Big) \xrightarrow{a.s.} 0.\end{aligned}$$

Thus, by noting that $\lambda_p(\cdot)$ is a continuous function (since (8)) and applying SLLN,

$$\begin{aligned}&\liminf_{G\to\infty}\inf_{0\leq\alpha\leq1}\widehat{R}_1''(\alpha, \widehat{\Sigma}, \Sigma)\\ &\quad\geq \liminf_{G\to\infty}\frac{1}{4}\min\Big\{1, \{\lambda_p(\widehat{\Sigma}_{\text{pool}}^{-1})\}^2\Big\}\frac{1}{G}\sum_{g=1}^{G}\min\Big\{1, \lambda_p\{(S_g/n - \widehat{\Sigma}_{\text{pool}})^2\}\Big\}\\ &\quad= \frac{1}{4}\min\Big\{1, \big[\lambda_p\{(E\Sigma_1)^{-1}\}\big]^2\Big\}E\Big[\min\Big\{1, \lambda_p\{(S_1/n - E\Sigma_1)^2\}\Big\}\Big] > 0 \ \ a.s.\end{aligned}$$

Now, by (22), we have $\widehat{\alpha}_1^* - \alpha_1^* \xrightarrow{a.s.} 0$.

On the other hand, if Σ_g are all the same, we can not guarantee $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} < 0\, a.s.$ as $G \to \infty$. We consider the following two cases. If $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} < 0$, then as the same proof in the case that Σ_g are not all the same, we can show that (22) holds. If $\widehat{R}_1'(\alpha, \widehat{\Sigma}, \Sigma)|_{\alpha=0} \geq 0$, then by Theorem 3, we have $\widehat{\alpha}_1^* = 0$. Note that according to Theorem 1, when Σ_g are all the same, we have $\alpha_1^* = 0$, and hence $\widehat{\alpha}_1^* - \alpha_1^* = 0$. Therefore, the inequality (22) still holds. The rest of the proof is the same as the case that Σ_g are not all the same, we can get $\widehat{\alpha}_1^* - \alpha_1^* \xrightarrow{a.s.} 0$.

5.6 Proofs of (3) and (4)

We have

$$
\begin{aligned}
& E\mathrm{tr}\Big\{(S_g/n - \widehat{\Sigma}_{\mathrm{pool}})\Sigma_g^{-1}\Big\}^2 - \mathrm{tr}\Big\{(\Sigma_g - \overline{\Sigma})\Sigma_g^{-1}\Big\}^2 \\
&= E\mathrm{tr}\Big\{\Sigma_g^{-1/2}(S_g/n - \widehat{\Sigma}_{\mathrm{pool}})\Sigma_g^{-1/2}\Big\}^2 - \mathrm{tr}\Big\{\Sigma_g^{-1/2}(\Sigma_g - \overline{\Sigma})\Sigma_g^{-1/2}\Big\}^2 \\
&= \mathrm{tr}E\Big[\Sigma_g^{-1/2}\{S_g/n - \widehat{\Sigma}_{\mathrm{pool}} - (\Sigma_g - \overline{\Sigma})\}\Sigma_g^{-1/2}\Big]^2 \\
&= \sum_{i,j=1}^{p} E\Big[\Big\{\Sigma_g^{-1/2}\{S_g/n - \widehat{\Sigma}_{\mathrm{pool}} - (\Sigma_g - \overline{\Sigma})\}\Sigma_g^{-1/2}\Big\}_{ij}^2\Big].
\end{aligned}
$$

Let $S^{g,g'} = \Sigma_g^{-1/2} S_{g'} \Sigma_g^{-1/2}$, $\Sigma^{g,g'} = \Sigma_g^{-1/2}\Sigma_{g'}\Sigma_g^{-1/2}$, then $S^{g,g'} \sim W_p(\Sigma^{g,g'}, n)$. Note that $\mathrm{Var}(S_{ij}^{g,g'}) = n\{(\Sigma_{ij}^{g,g'})^2 + \Sigma_{ii}^{g,g'}\Sigma_{jj}^{g,g'}\}$, we have

$$
\begin{aligned}
& E\Big[\Sigma_g^{-1/2}\{S_g/n - \widehat{\Sigma}_{\mathrm{pool}} - (\Sigma_g - \overline{\Sigma})\}\Sigma_g^{-1/2}\Big]_{ij}^2 \\
&= \mathrm{Var}\Big(\frac{G-1}{nG} S_{ij}^{g,g} - \frac{1}{nG}\sum_{g' \neq g} S_{ij}^{g,g'}\Big) \\
&= \frac{(G-1)^2}{n^2G^2}\mathrm{Var}\big(S_{ij}^{g,g}\big) + \frac{1}{n^2G^2}\sum_{g'\neq g}\mathrm{Var}\big(S_{ij}^{g,g'}\big) \\
&= \frac{(G-1)^2}{nG^2}(\delta_{ij}+1) + \frac{1}{G^2}\sum_{g'\neq g}\frac{1}{n}\Big\{(\Sigma_{ij}^{g,g'})^2 + \Sigma_{ii}^{g,g'}\Sigma_{jj}^{g,g'}\Big\},
\end{aligned}
$$

where $\delta_{ij} = 1$ if $i = j$, otherwise, $\delta_{ij} = 0$, $S_{ij}^{g,g'}$ and $\Sigma_{ij}^{g,g'}$ are the (i, j)th components of $S^{g,g'}$ and $\Sigma^{g,g'}$, respectively. Thus

$$
\begin{aligned}
E\mathrm{tr}\Big\{\big(S_g/n - \widehat{\Sigma}_{\mathrm{pool}}\big)\Sigma_g^{-1}\Big\}^2 &= \mathrm{tr}\Big\{(\Sigma_g - \overline{\Sigma})\Sigma_g^{-1}\Big\}^2 + \frac{(G-1)^2(p^2+p)}{nG^2} \\
&\quad + \frac{1}{nG^2}\sum_{g'\neq g}\Big[\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})^2 + \big\{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\big\}^2\Big].
\end{aligned}
$$

Then we get (3) from (2).

For a_2, we have

$$
a_2 = -a_1 + \frac{1}{G}\sum_{g=1}^{G} E\mathrm{tr}\Big\{\big(S_g/n - \widehat{\Sigma}_{\mathrm{pool}}\big)\Sigma_g^{-1}\big(S_g\Sigma_g^{-1}/n - I\big)\Big\}.
$$

Since $ES_g = n\Sigma_g$ and S_g, $g = 1, \ldots, G$ are independent, we have

$$\begin{aligned} E\mathrm{tr}\Big\{\big(S_g/n - \widehat{\Sigma}_{\text{pool}}\big) \\ \Sigma_g^{-1}\big(S_g\Sigma_g^{-1}/n - I\big)\Big\} &= \frac{G-1}{G}E\mathrm{tr}\Big\{\big(S_g\Sigma_g^{-1}/n\big)\big(S_g\Sigma_g^{-1}/n - I\big)\Big\} \\ &= \frac{G-1}{n^2G}E\mathrm{tr}(S^{g,g})^2 - \frac{(G-1)p}{G} \\ &= \frac{(G-1)(p^2+p)}{nG}, \end{aligned}$$

where we have used $S^{g,g} \sim W_p(I, n)$ and $E\mathrm{tr}(S^{g,g})^2 = \sum_{i,j=1}^{p} E(S_{i,j}^{g,g})^2 = \sum_{i,j=1}^{p}\left\{\mathrm{Var}(S_{i,j}^{g,g}) + n^2\delta_{ij}\right\} = (n^2+n)p + np^2$. Hence

$$a_2 = -a_1 + \frac{(G-1)(p^2+p)}{nG},$$

and we get (4).

5.7 Proof of Theorem 6

First of all, we show that $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ is a strictly convex function of α on [0, 1], which is equivalent to verify that $a_1 > 0$.

Define $V_g := \Sigma_g^{-1/2}(S_g/n - \widehat{\Sigma}_{\text{pool}})\,\Sigma_g^{-1}\,(S_g/n - \widehat{\Sigma}_{\text{pool}})\,\Sigma_g^{-1/2}$. Since $V_g \geq 0$, we have

$$E\mathrm{tr}\Big\{(S_g/n - \widehat{\Sigma}_{\text{pool}})\Sigma_g^{-1}\Big\}^2 = E\mathrm{tr}(V_g) \geq 0,$$

and the equality holds if and only if $V_g = 0$ $a.s.$ Note that $V_g = 0$ $a.s.$ implies $S_g/n = \widehat{\Sigma}_{\text{pool}}$ $a.s.$, which is impossible. Hence $a_1 = E\mathrm{tr}(V_g)/G > 0$.

Secondly, we show that $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ has unique minimum point at $\alpha_2^* = -a_2/a_1$.

As $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ is a quadratic form, it is easy to verify that the unique minimum value is attained at $\alpha_2^* = -a_2/a_1$. The remainder is to verify that $\alpha_2^* \in [0, 1]$. It follows from (4) that $a_1 > -a_2$. We only need to verify $-a_2 \geq 0$. By Lemma 2, we know $\mathrm{tr}(AB^{-1} + BA^{-1}) \geq 2p$ holds for any $A > 0$, $B > 0$. Thus, for any $g \neq g'$,

$$\begin{aligned} &\frac{1}{n}\Big\{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})^2 + \mathrm{tr}(\Sigma_g\Sigma_{g'}^{-1})^2\Big\} = \frac{1}{n}\Big[\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})^2 + \mathrm{tr}\big\{(\Sigma_{g'}\Sigma_g^{-1})^2\big\}^{-1}\Big] \geq 2p/n, \\ &\frac{1}{n}\Big[\big\{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\big\}^2 + \big\{\mathrm{tr}(\Sigma_g\Sigma_{g'}^{-1})\big\}^2\Big] \geq \frac{1}{2n}\Big\{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1}) + \mathrm{tr}(\Sigma_g\Sigma_{g'}^{-1})\Big\}^2 \geq 2p^2/n. \end{aligned}$$

Then, by (3) and (4),

$$-a_2 \geq \frac{(G-1)^2(p^2+p)}{nG^2} - \frac{(G-1)(p^2+p)}{nG}$$

$$+\frac{1}{nG^3}\sum_{g=1}^{G}\sum_{g'\neq g}\Big[\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})^2$$
$$+\left\{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\right\}^2\Big]$$
$$\geq \frac{1}{n}\sum_{g=1}^{G}\left\{\frac{(G-1)^2(p^2+p)}{G^2}-\frac{(G-1)(p^2+p)}{G}\right\}$$
$$+\frac{p^2+p}{G^3}\sum_{g=1}^{G}\sum_{g'>g}^{G}\frac{2}{n}$$
$$\geq -\frac{(p^2+p)(G-1)}{nG^2}+\frac{p^2+p}{G^3}\frac{G(G-1)}{n}=0.$$

Note that when $\Sigma_g = \Sigma_g'$ for any $g \neq g'$, we have these equalities hold. As a consequence, we have $a_2 = 0$, and hence $\alpha_2^* = 0$. The proof of Theorem 6 is complete.

5.8 Proof of Theorem 7

For any $g = 1, \ldots, G$, we have that, $S_g/n \xrightarrow{a.s.} \Sigma_g$. Consequently, as $n \to \infty$, we have

$$a_1 \xrightarrow{a.s.} \sum_{g=1}^{G}\mathrm{tr}\left\{(\Sigma_g-\overline{\Sigma})\Sigma_g^{-1}(\Sigma_g-\overline{\Sigma})\Sigma_g^{-1}\right\}/G=\sum_{g=1}^{G}\mathrm{tr}\left\{(I-\overline{\Sigma}\Sigma_g^{-1})^2\right\}/G,$$

where $\overline{\Sigma} = \sum_{g=1}^{G}\Sigma_g/G$. Similarly, as $n \to \infty$,

$$a_2 \xrightarrow{a.s.} \sum_{g=1}^{G}\mathrm{tr}\left\{(\Sigma_g-\overline{\Sigma})\Sigma_g^{-1}(\overline{\Sigma}\Sigma_g^{-1}-I)\right\}/G=-\sum_{g=1}^{G}\mathrm{tr}\left\{(I-\overline{\Sigma}\Sigma_g^{-1})^2\right\}/G,$$

and $a_3 \xrightarrow{a.s.} \sum_{g=1}^{G}\mathrm{tr}\big(\overline{\Sigma}\Sigma_g^{-1}-I\big)^2/G$.

When Σ_g are not all the same, we have $\lim\limits_{n\to\infty} a_1 > 0$ and then $\alpha_2^* = -a_2/a_1 \xrightarrow{a.s.} 1$.

When $\Sigma_g = \Sigma_{g'}$ for any $g \neq g'$, we have $\lim_{n\to\infty} a_1 = -\lim_{n\to\infty} a_2 = \sum_{g=1}^{G}\mathrm{tr}\{(I-\overline{\Sigma}\Sigma_g^{-1})^2\}/G = 0$. Hence $\lim_{n\to\infty} R_2(\alpha, \widehat{\Sigma}, \Sigma) = \lim_{n\to\infty} a_3 = \sum_{g=1}^{G}\mathrm{tr}\big(\overline{\Sigma}\Sigma_g^{-1}-I\big)^2/G$ $a.s$. Therefore, $R_2(\alpha, \widehat{\Sigma}, \Sigma)$ is a constant function of α.

5.9 Proof of Theorem 8

Define $\widetilde{S} := (\Sigma_0)^{-1/2}S_g(\Sigma_0)^{-1/2}$, $\widetilde{\Sigma} := (\Sigma_0)^{-1/2}\Sigma_g(\Sigma_0)^{-1/2}$, then $\widetilde{S} \sim W_p(\widetilde{\Sigma}, n)$ and $\widetilde{S}^{-1}$ has an inverse Wishart distribution, i.e. $\widetilde{S}^{-1} \sim W_p^{-1}(\widetilde{\Sigma}^{-1}, n)$.

Note that (see, for instance, Letac and Massam (2004), page 308)

$$E\widetilde{S}^{-2} = \frac{(n-p-1)\widetilde{\Sigma}^{-2} + \widetilde{\Sigma}^{-1}\mathrm{tr}(\widetilde{\Sigma}^{-1})}{(n-p)(n-p-1)(n-p-3)},$$
$$E\left\{\widetilde{S}^{-1}\mathrm{tr}(\widetilde{S}^{-1})\right\} = \frac{2\widetilde{\Sigma}^{-2} + (n-p-2)\widetilde{\Sigma}^{-1}\mathrm{tr}(\widetilde{\Sigma}^{-1})}{(n-p)(n-p-1)(n-p-3)},$$

we have $(n-p-2)E\widetilde{S}^{-2} - E\{\widetilde{S}^{-1}\mathrm{tr}(\widetilde{S}^{-1})\} = \widetilde{\Sigma}^{-2}/(n-p-1)$.

Then, by noting that $E\left(\widetilde{S}^{-1}\right) = \widetilde{\Sigma}^{-1}/(n-p-1)$,

$$\begin{aligned}
&E\mathrm{tr}\Big[\left\{I-(n-p-1)\Sigma_0 S_g^{-1}\right\}^2 - (n-p-1)\\
&\quad\left\{\mathrm{tr}(\Sigma_0 S_g^{-1})^2 + \left\{\mathrm{tr}(\Sigma_0 S_g^{-1})\right\}^2\right\}\Big]\\
&= E\Big[p - 2(n-p-1)\mathrm{tr}(\Sigma_0 S_g^{-1}) + (n-p-1)\Big\{(n-p-2)\mathrm{tr}(\Sigma_0 S_g^{-1})^2\\
&\quad -\left\{\mathrm{tr}(\Sigma_0 S_g^{-1})\right\}^2\Big\}\Big]\\
&= p - 2(n-p-1)\mathrm{tr}E(\widetilde{S}^{-1})\\
&\quad +(n-p-1)\left[(n-p-2)\mathrm{tr}E(\widetilde{S}^{-2}) - \left\{\mathrm{tr}E(\widetilde{S}^{-1})\right\}^2\right]\\
&= \mathrm{tr}\left\{(\Sigma_g - \Sigma_0)\Sigma_g^{-1}\right\}^2.
\end{aligned}$$

where we have used that $\mathrm{tr}\{(\Sigma_g - \Sigma_0)\Sigma_g^{-1}\}^2 = \mathrm{tr}(I - \widetilde{\Sigma}^{-1})^2 = p - 2\mathrm{tr}(\widetilde{\Sigma}^{-1}) + \mathrm{tr}(\widetilde{\Sigma}^{-2})$. Hence we get (6).

5.10 Proof of Theorem 9

For any $g = 1, \ldots, G$, we have that, $S_g/n \xrightarrow{a.s.} \Sigma_g$ as $n \to \infty$. First of all, we proof that $\widehat{b} \xrightarrow{a.s.} b = \sum_{g=1}^{G} \mathrm{tr}(I - \overline{\Sigma}\Sigma_g^{-1})^2/G > 0$.

As $n \to \infty$,

$$\mathrm{tr}(I - r\widehat{\Sigma}_{\mathrm{pool}} S_g^{-1})^2 \xrightarrow{a.s.} \mathrm{tr}(I - \overline{\Sigma}\Sigma_g^{-1})^2 > 0,$$
$$r^2\left[\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}} S_g^{-1})^2 + \left\{\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}} S_g^{-1})\right\}^2\right] \xrightarrow{a.s.} \mathrm{tr}(\overline{\Sigma}\Sigma_g^{-1})^2 + \left\{\mathrm{tr}(\overline{\Sigma}\Sigma_g^{-1})\right\}^2.$$

Hence, $\mathrm{tr}(I - r\widehat{\Sigma}_{\mathrm{pool}} S_g^{-1})^2 - r\left[\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}} S_g^{-1})^2 + \left\{\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}} S_g^{-1})\right\}^2\right] \xrightarrow{a.s.} \mathrm{tr}(I - \overline{\Sigma}\Sigma_g^{-1})^2 > 0$. As a consequence, we have $\widehat{b} \xrightarrow{a.s.} b$.

Since $r^2\mathrm{tr}\{(S_{g'}S_g^{-1})^2\}/n^2 + r^2\{\mathrm{tr}(S_{g'}S_g^{-1})\}^2/n^2 \xrightarrow{a.s.} \mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})^2 + \{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\}^2$, we have

$$\frac{(G-1)^2(p^2+p)}{nG^2} + \frac{r^2}{n^3G^3}\sum_{g=1}^{G}\sum_{g'\neq g}\left[\mathrm{tr}(S_{g'}S_g^{-1})^2 + \{\mathrm{tr}(S_{g'}S_g^{-1})\}^2\right] \xrightarrow{a.s.} 0.$$

Therefore, as $n \to \infty$, $\widehat{a}_1 \xrightarrow{a.s.} \sum_{g=1}^{G} \mathrm{tr}(I - \overline{\Sigma}\Sigma_g^{-1})^2/G > 0$. By (7), as $n \to \infty$, $\lim_{n\to\infty} \widehat{a}_2 = -\lim_{n\to\infty}\widehat{a}_1 = -b \;\; a.s.$. Finally, we have $\widehat{\alpha}_2^* \xrightarrow{a.s.} 1$.

5.11 Proof of Theorem 10

Note that

$$a_1 = \frac{1}{G}\sum_{g=1}^{G}\mathrm{tr}\{(\Sigma_g - \overline{\Sigma})\Sigma_g^{-1}\}^2 + \frac{(G-1)^2(p^2+p)}{nG^2}$$
$$+ \frac{1}{nG^3}\sum_{g=1}^{G}\sum_{g'\neq g}\left[\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})^2 + \{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\}^2\right],$$
$$\widehat{a}_1 = \widehat{b} + \frac{(G-1)^2(p^2+p)}{nG^2} + \frac{r^2}{n^3G^3}\sum_{g=1}^{G}\sum_{g'\neq g}\left[\mathrm{tr}(S_{g'}S_g^{-1})^2 + \{\mathrm{tr}(S_{g'}S_g^{-1})\}^2\right].$$

First of all, we show that, for fixed n and p, $\widehat{a}_1 \xrightarrow{a.s.} a_1$ as $G \to \infty$, which is equivalent to prove that

$$\widehat{b} \xrightarrow{a.s.} \mathrm{tr}E\{I - (E\Sigma_1)\Sigma_1^{-1}\}^2 \geq 0, \tag{23}$$

$$\frac{1}{G}\sum_{g=1}^{G}\mathrm{tr}\{(\Sigma_g - \overline{\Sigma})\Sigma_g^{-1}\}^2 \xrightarrow{a.s.} \mathrm{tr}E\{I - (E\Sigma_1)\Sigma_1^{-1}\}^2, \tag{24}$$

$$\frac{1}{nG^3}\sum_{g=1}^{G}\sum_{g'\neq g}\left[\mathrm{tr}(S_{g'}S_g^{-1})^2 + \{\mathrm{tr}(S_{g'}S_g^{-1})\}^2\right] \xrightarrow{a.s.} 0, \tag{25}$$

$$\frac{1}{nG^3}\sum_{g=1}^{G}\sum_{g'\neq g}\left[\{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\}^2 + \{\mathrm{tr}(\Sigma_{g'}\Sigma_g^{-1})\}^2\right] \xrightarrow{a.s.} 0 \tag{26}$$

In the following, we first prove that (23) holds. By (6), we have

$$E\mathrm{tr}\left[\{I - r(E\Sigma_1)S_g^{-1}\}^2 - r\left[\mathrm{tr}\{(E\Sigma_1)S_g^{-1}\}^2 + \left[\mathrm{tr}\{(E\Sigma_1)S_g^{-1}\}\right]^2\right]\middle|\Sigma_g\right] = \mathrm{tr}\{I - (E\Sigma_1)\Sigma_g^{-1}\}^2.$$

Thus, by SLLN,

$$\frac{1}{G}\sum_{g=1}^{G}\left[\mathrm{tr}\{I-r(E\Sigma_1)S_g^{-1}\}^2-r\left[\mathrm{tr}\{(E\Sigma_1)S_g^{-1}\}^2+\left\{\mathrm{tr}\left\{(E\Sigma_1)S_g^{-1}\right\}\right\}^2\right]\right]$$
$$\xrightarrow{a.s.} \mathrm{tr}E\{I-(E\Sigma_1)\Sigma_1^{-1}\}^2 \geq 0.$$

Noting that

$$\widehat{b}=\max\left\{0,\ \frac{1}{G}\sum_{g=1}^{G}\left[\mathrm{tr}(I-r\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})^2-r\left\{\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})^2+\{\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})\}^2\right\}\right]\right\}.$$

In order to show that $\widehat{b}\xrightarrow{a.s.}\mathrm{tr}E\{I-(E\Sigma_1)\Sigma_1^{-1}\}^2\geq 0$, we need to prove that

$$\frac{1}{G}\sum_{g=1}^{G}\left[\mathrm{tr}(I-r\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})^2-r\left\{\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})^2+\{\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})\}^2\right\}\right]$$
$$-\frac{1}{G}\sum_{g=1}^{G}\left[\mathrm{tr}\{I-r(E\Sigma_1)S_g^{-1}\}^2-r\left[\mathrm{tr}\left\{(E\Sigma_1)S_g^{-1}\right\}^2+\left\{\mathrm{tr}\left\{(E\Sigma_1)S_g^{-1}\right\}\right\}^2\right]\right]\xrightarrow{a.s.}0.$$

It is sufficient to prove that

$$\frac{1}{G}\sum_{g=1}^{G}\left[\mathrm{tr}(I-r\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})^2-\mathrm{tr}\left\{I-r(E\Sigma_1)S_g^{-1}\right\}^2\right]\xrightarrow{a.s.}0, \tag{27}$$

$$\frac{1}{G}\sum_{g=1}^{G}\left[\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})^2-\mathrm{tr}\left\{(E\Sigma_1)S_g^{-1}\right\}^2\right]\xrightarrow{a.s.}0, \tag{28}$$

$$\frac{1}{G}\sum_{g=1}^{G}\left[\{\mathrm{tr}(\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1})\}^2-\left\{\mathrm{tr}\left\{(E\Sigma_1)S_g^{-1}\right\}\right\}^2\right]\xrightarrow{a.s.}0. \tag{29}$$

Note that $\mathrm{tr}(A^2)-\mathrm{tr}(B^2)=\mathrm{tr}(A-B)(A+B)$ and apply SLLN and Lemma 1, we have

$$\left|\frac{1}{G}\sum_{g=1}^{G}\left[\mathrm{tr}\left(I-r\widehat{\Sigma}_{\mathrm{pool}}S_g^{-1}\right)^2-\mathrm{tr}\left\{I-r(E\Sigma_1)S_g^{-1}\right\}^2\right]\right|$$
$$\leq\frac{r}{G}\sum_{g=1}^{G}\left|\mathrm{tr}\left[\left\{(\widehat{\Sigma}_{\mathrm{pool}}-E\Sigma_1)S_g^{-1}\right\}\left\{2I-r(\widehat{\Sigma}_{\mathrm{pool}}+E\Sigma_1)S_g^{-1}\right\}\right]\right|$$
$$\leq\frac{r\sqrt{p}}{G}\sum_{g=1}^{G}\left(\|\widehat{\Sigma}_{\mathrm{pool}}-E\Sigma_1\|\right)\|S_g^{-1}\|\left\{2p+r\left(\|\widehat{\Sigma}_{\mathrm{pool}}\|+\|E\Sigma_1\|\right)\|S_g^{-1}\|\right\}$$

$$= 2rp^{\frac{3}{2}}\left(\|\widehat{\Sigma}_{\text{pool}} - E\Sigma_1\|\right)\left(\frac{1}{G}\sum_{g=1}^{G}\|S_g^{-1}\|\right)$$

$$+r^2p^{\frac{1}{2}}\left(\|\widehat{\Sigma}_{\text{pool}}\| + \|E\Sigma_1\|\right)\left(\|\widehat{\Sigma}_{\text{pool}} - E\Sigma_1\|\right)\left(\frac{1}{G}\sum_{g=1}^{G}\|S_g^{-1}\|^2\right)$$

$$= 2rp^{\frac{3}{2}}\left(\|\widehat{\Sigma}_{\text{pool}} - E\Sigma_1\|\right)\left(E\|S_1^{-1}\|\right) + r^2p^{\frac{1}{2}}\left(\|\widehat{\Sigma}_{\text{pool}}\|\right.$$
$$\left.+\|E\Sigma_1\|\right)\left(\|\widehat{\Sigma}_{\text{pool}} - E\Sigma_1\|\right)\left(E\|S_g^{-1}\|^2\right).$$

This proves (27). And follow the same procedure, we can get (28).

To verify (29), we note that

$$\left|\frac{1}{G}\sum_{g=1}^{G}\left[\left\{\text{tr}(\widehat{\Sigma}_{\text{pool}}S_g^{-1})\right\}^2 - \left\{\text{tr}\left((E\Sigma_1)S_g^{-1}\right)\right\}^2\right]\right|$$

$$\leq \frac{1}{G}\sum_{g=1}^{G}\left\{\left|\text{tr}\left\{(\widehat{\Sigma}_{\text{pool}} - E\Sigma_1)S_g^{-1}\right\}\right|\right\}\left\{\left|\text{tr}\left\{(\widehat{\Sigma}_{\text{pool}} + E\Sigma_1)S_g^{-1}\right\}\right|\right\}$$

$$\leq p\left(\|\widehat{\Sigma}_{\text{pool}} - E\Sigma_1\|\right)\left(\|\widehat{\Sigma}_{\text{pool}} + E\Sigma_1\|\right)\left(\frac{1}{G}\sum_{g=1}^{G}\|S_g^{-1}\|^2\right) \xrightarrow{a.s.} 0.$$

where we have used $\sum_{g=1}^{G}\|S_g^{-1}\|^2/G \xrightarrow{a.s.} E\|S_g^{-1}\|^2 < \infty$, and $\widehat{\Sigma}_{\text{pool}} \xrightarrow{a.s.} E\Sigma_1$. Therefore we get (23).

In order to prove (24), we have

$$\left|\frac{1}{G}\sum_{g=1}^{G}\left[\text{tr}(I - \overline{\Sigma}\Sigma_g^{-1})^2 - \text{tr}\{I - (E\Sigma_1)\Sigma_g^{-1}\}^2\right]\right|$$

$$\leq \frac{1}{G}\sum_{g=1}^{G}\left|\text{tr}\left[\{(\widehat{\Sigma} - E\Sigma_1)\Sigma_g^{-1}\}\{2I - (\overline{\Sigma} + E\Sigma_1)\Sigma_g^{-1}\}\right]\right|$$

$$\leq \frac{\sqrt{p}}{G}\sum_{g=1}^{G}\left(\|\overline{\Sigma} - E\Sigma_1\|\right)\left(\|\Sigma_g^{-1}\|\right)\left\{2p + \left(\|\overline{\Sigma}\| + \|E\Sigma_1\|\right)\|\Sigma_g^{-1}\|\right\}$$

$$= 2p^{\frac{3}{2}}\left(\|\overline{\Sigma} - E\Sigma_1\|\right)\left(\frac{1}{G}\sum_{g=1}^{G}\|\Sigma_g^{-1}\|\right)$$

$$+\sqrt{p}\left(\|\overline{\Sigma}\| + \|E\Sigma_1\|\right)\left(\|\overline{\Sigma} - E\Sigma_1\|\right)\left(\frac{1}{G}\sum_{g=1}^{G}\|\Sigma_g^{-1}\|^2\right)$$

$$\xrightarrow{a.s.} 0.$$

where we have used $\overline{\Sigma} \xrightarrow{a.s.} E\Sigma_1$, and $\sum_{g=1}^{G}\|\Sigma_g^{-1}\|^2/G \xrightarrow{a.s.} E\|\Sigma_1^{-1}\|^2 < \infty$.

The proofs of (25) and (26) are similar, so we only prove (25). By Lemma 1 and SLLN, as $G \to \infty$, we have

$$\begin{aligned}
&\frac{1}{G^3}\sum_{g=1}^{G}\sum_{g'\neq g}\left[\mathrm{tr}(S_{g'}S_g^{-1})^2 + \left\{\mathrm{tr}(S_{g'}S_g^{-1})\right\}^2\right] \\
&\leq \frac{1}{G^3}\sum_{g=1}^{G}\sum_{g'\neq g}\left\{\sqrt{p}\|(S_{g'}S_g^{-1})^2\| + p\|S_{g'}\|^2\,\|S_g^{-1}\|^2\right\} \\
&\leq \frac{2p}{G^3}\sum_{g=1}^{G}\sum_{g'\neq g}\|S_{g'}\|^2\,\|S_g^{-1}\|^2 \leq \frac{2p}{G}\Big(\frac{1}{G}\sum_{g=1}^{G}\|S_g\|^2\Big)\Big(\frac{1}{G}\sum_{g=1}^{G}\|S_g^{-1}\|^2\Big) \xrightarrow{a.s.} 0.
\end{aligned}$$

Hence we get (25). Therefore, for fixed n and p, we have $\widehat{a}_1 \xrightarrow{a.s.} a_1$ as $G \to \infty$. On the other hand, according to (4) and (7), we have $|\widehat{a}_2 - a_2| = |\widehat{a}_1 - a_1|$, and hence $\widehat{a}_2 \xrightarrow{a.s.} a_2$ as $G \to \infty$. By the fact that $\alpha_2^* = -a_2/a_1 \geq 0$ and $\widehat{\alpha}_2^* = \max\{0, -\widehat{a}_2/\widehat{a}_1\}$, we have $\widehat{\alpha}_2^* - \alpha_2^* \xrightarrow{a.s.} 0$. The proof of Theorem 10 is complete.

Acknowledgements Zongliang Hu's research was supported by the Natural Science Foundation grant (No. 2019083) of Shenzhen University, and the Foundation and Applied Basic Research Programs (No. 2019A1515110449) of Guangdong Province of China. Zhishui Hu's research was partially supported by the National Natural Science Foundation of China grant (No. 11671373). Tiejun Tong's research was supported by the General Research Fund (No. HKBU12303918), the National Natural Science Foundation of China (No. 11671338), and the Initiation Grant for Faculty Niche Research Areas (No. RC-IG-FNRA/17-18/13) of Hong Kong Baptist University. Yuedong Wang's research was partially supported by the National Science Foundation grant (No. DMS-1507620). The authors also thank the editor, the associate editor, and two reviewers for their constructive comments that have led to a substantial improvement of this paper.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Barnard J, McCulloch R, Meng X (2000) Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica 10:1281–1311

Boik RJ (2002) Spectral models for covariance matrices. Biometrika 89:159–182

Boik RJ (2003) Principal component models for correlation matrices. Biometrika 90:679–701

Cai T, Li H, Liu W, Xie J (2016) Joint estimation of multiple high-dimensional precision matrices. Statistica Sinica 26:445–464

Danaher P, Wang P, Witten DM (2014) The joint graphical lasso for inverse covariance estimation across multiple classes. J R Stat Soc Ser B 76:373–397

Daniels MJ (2006) Bayesian modeling of several covariance matrices and some results on propriety of the posterior for linear regression with correlated or heterogeneous errors. J Multivar Anal 97:1185–1207

Ferguson TS (1996) A course in large sample theory. Chapman & Hall, London

Friedman JH (1989) Regularized discriminant analysis. J Am Stat Assoc 84:165–175

Gaskins J, Daniels M (2016) Covariance partition priors: a Bayesian approach to simultaneous covariance estimation for longitudinal data. J Comput Gr Stat 25:167–186

Guionnet A (2009) Large random matrices: lectures on macroscopic asymptotics. Springer, Berlin

Guo J, Levina E, Michailidis G, Zhu J (2011) Joint estimation of multiple graphical models. Biometrika 98:1–15

Haff L (1980) Empirical Bayes estimation of the multivariate normal covariance matrix. Ann Stat 8:586–597

Haff L (1991) The variational form of certain Bayes estimators. Ann Stat 19:1163–1190

Hoff PD (2009) A hierarchical eigenmodel for pooled covariance estimation. J R Stat Soc Ser B 71:971–992

Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. J Multivar Anal 88:365–411

Ledoit O, Wolf M (2012) Nonlinear shrinkage estimation of large-dimensional covariance matrices. Ann Stat 40:1024–1060

Le Y, Hastie T (2016) Sparse quadratic discriminant analysis and community Bayes, *arXiv preprint*arXiv:1407.4543

Letac G, Massam H (2004) All invariant moments of the Wishart distribution. Scand J Stat 31:295–318

Manly BF, Rayner J (1987) The comparison of sample covariance matrices using likelihood ratio tests. Biometrika 74:841–847

Pourahmadi M, Daniels MJ, Park T (2007) Simultaneous modelling of the Cholesky decomposition of several covariance matrices. J Multivar Anal 98:568–587

Price BS, Geyer CJ, Rothman AJ (2015) Ridge fusion in statistical learning. J Comput Gr Stat 24:439–454

Ramey JA, Stein CK, Young PD, Young DM (2016) High-dimensional regularized discriminant analysis, *arXiv* preprint arXiv:1602.01182

Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mol Biol 4:1175–1189

Stein C (1956) Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. Proc Third Berkeley Symp Math Stat Prob 1:197–206

Tao T (2012) Topics in random matrix theory. American Mathematical Society, Providence

Tong T, Wang Y (2007) Optimal shrinkage estimation of variances with applications to microarray data analysis. J Am Stat Assoc 102:113–122

Tong T, Jang H, Wang Y (2012) James–Stein type estimators of variances. J Multivar Anal 107:232–243

Yang R, Berger JO (1994) Estimation of a covariance matrix using the reference prior. Ann Stat 22:1195–1211