Electronic Journal of Statistics
Vol. 8 (2014) 130–158
ISSN: 1935-7524
DOI: 10.1214/14-EJS878

Estimation of covariance and precision matrices under scale-invariant quadratic loss in high dimension

Tatsuya Kubokawa

Department of Economics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
e-mail: tatsuya@e.u-tokyo.ac.jp

and

Akira Inoue

Graduate School of Economics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
e-mail: inoue-akira@hotmail.com

Abstract: The problem of estimating covariance and precision matrices of multivariate normal distributions is addressed when both the sample size and the dimension of variables are large. The estimation of the precision matrix is important in various statistical inference including the Fisher linear discriminant analysis, confidence region based on the Mahalanobis distance and others. A standard estimator is the inverse of the sample covariance matrix, but it may be instable or can not be defined in the high dimension. Although (adaptive) ridge type estimators are alternative procedures which are useful and stable for large dimension. However, we are faced with questions about how to choose ridge parameters and their estimators and how to set up asymptotic order in ridge functions in high dimensional cases. In this paper, we consider general types of ridge estimators for covariance and precision matrices, and derive asymptotic expansions of their risk functions. Then we suggest the ridge functions so that the second order terms of risks of ridge estimators are smaller than those of risks of the standard estimators.

MSC 2010 subject classifications: Primary 62F10, 62H12; secondary 62J07.
Keywords and phrases: Asymptotic expansion, covariance matrix, high dimension, Moore-Penrose inverse, multivariate normal distribution, point estimation, precision matrix, ridge estimator, risk comparison, scale-invariant quadratic loss, Stein-Haff identity, Wishart distribution.

Received August 2013.

Contents

1. Introduction

Statistical inference with high dimension has received much attention in recent years and has been actively studied from both theoretical and practical aspects in the literature. Of these, estimate of the precision matrix is required in many multivariate inference procedures including the Fisher linear discriminant analysis, confidence intervals based on the Mahalanobis distance and weighted least squares estimator in multivariate linear regression models. A standard estimator of the precision based on the sample covariance matrix is likely to be instable when the dimension p is large and close to the sample size N even if $N > p$. In the case of $p > N$, the inverse of the sample covariance matrix cannot be defined, and an estimator based on the Moore-Penrose generalized inverse of the sample covariance matrix has been used in Srivastava [11]. Another useful and stable estimator for the precision matrix is a ridge estimator, and its various variants have been used in literature. For example, see Ledoit and Wolf [8], [9], Fisher and Sun [3] and Bai and Shi [1]. However, superiority of the ridge-type estimators over the standard estimators have not been studied except Kubokawa and Srivastava [6], who obtained exact conditions for the ridge-type estimators to have uniformly smaller risks than the standard estimator. However, their results are limited to specific ridge functions and special loss functions.

To specify the problem considered here, let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ be independently and identically distributed (i.i.d.) as a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and $p \times p$ positive definite covariance matrix $\boldsymbol{\Sigma}$ denoted as $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$. Let $\overline{\boldsymbol{y}} = N^{-1} \sum_{i=1}^N \boldsymbol{y}_i$, $\boldsymbol{V} = \sum_{i=1}^N (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^t$ and $n = N-1$. Then, in the case of $n \geq p$, $\boldsymbol{V}$ has a Wishart distribution with mean $n\boldsymbol{\Sigma}$ and degrees of freedom n, denoted as $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$. When $n < p$, it is called a singular Wishart distribution, whose distribution has been recently studied by Srivastava [10]. In many inference procedures, an estimate of the precision matrix $\boldsymbol{\Sigma}^{-1}$ is required. In the case that $n > p$, the standard estimator of $\boldsymbol{\Sigma}^{-1}$ is $\widehat{\boldsymbol{\Sigma}}_0^{-1} = c\boldsymbol{V}^{-1}$ for a positive constant c, but it may not be stable when p is large and close to n. In the case of $p > n$, the estimator $c\boldsymbol{V}^{-1}$ cannot be defined. Srivastava [11] used the estimator $c\boldsymbol{V}^{+}$ based on the Moore-Penrose inverse $\boldsymbol{V}^{+}$ of $\boldsymbol{V}$.

In this paper, we address the problems of estimating both covariance matrix $\boldsymbol{\Sigma}$ and precision matrix $\boldsymbol{\Sigma}^{-1}$, and consider general ridge-type estimators, respectively given by

$$\widehat{\boldsymbol{\Sigma}}_{\Lambda} = c(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}}), \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1} = c(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}, \tag{1.1}$$

where c and d are positive constants based on (n, p), and $\widehat{\boldsymbol{\Lambda}}$ is a $p \times p$ positive definite matrix based on $\boldsymbol{V}$. Examples of the ridge function $\boldsymbol{\Lambda}$ include $\widehat{\boldsymbol{\Lambda}} = \hat{\lambda}\boldsymbol{I}$, $\widehat{\boldsymbol{\Lambda}} = \operatorname{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ and others, where $\hat{\lambda}$, $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ are functions of $\boldsymbol{V}$. We evaluate the difference of risk functions of the ridge-type and the standard estimators asymptotically for large n and p, where the risk functions are measured with respect to the scale-invariant quadratic loss functions. Then we derive conditions on d and $\widehat{\boldsymbol{\Lambda}}$ such that the ridge-type estimators improve on the standard estimators asymptotically.

The paper is organized as follows. Section 2 treats estimation of the covariance matrix $\boldsymbol{\Sigma}$, and gives asymptotic evaluations for risks of the ridge estimators when $(n, p) \to \infty$. The estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$ is dealt with in Section 3. For estimation of the covariance matrix relative to the scale-invariant quadratic loss, we can handle both cases of $n > p$ and $p > n$ in the unified framework. For the precision matrix, however, the ridge type estimator has different properties between the two cases, and the standard estimator is $c\boldsymbol{V}^{+}$ in the case of $p > n$, so that we need to treat the two cases separately. Some examples of ridge functions $\widehat{\boldsymbol{\Lambda}}$ are given in Section 4. Risk performances of the ridge-type estimators are investigated by simulation in Section 5. Concluding remarks are given in Section 6. Some technical tools and proofs are given in the appendix.

2. A unified result in estimation of covariance

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a $p \times n$ random matrix such that $\boldsymbol{x}_i \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ for $i = 1, \ldots, n$, where $\boldsymbol{\Sigma}$ is an unknown positive definite matrix. Let $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$. In the case of $n \geq p$, $\boldsymbol{V}$ is distributed as a Wishart distribution $\mathcal{W}_p(n, \boldsymbol{\Sigma})$ with n degrees of freedom. We first consider the estimation of the covariance matrix $\boldsymbol{\Sigma}$ in terms of the risk function $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = E[L(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}})]$, where $L(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}})$ is the scale-invariant quadratic loss

$$L(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \operatorname{tr}[(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2].$$

The loss function is invariant under the scale transformation $\boldsymbol{\Sigma} \to \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$ and $\widehat{\boldsymbol{\Sigma}} \to \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}\boldsymbol{A}'$ for any nonsingular matrix $\boldsymbol{A}$.

A standard estimator is of the form $c\boldsymbol{V}$ for $c \in \mathbb{R}_+$, where $\mathbb{R}_+$ is a set of real positive numbers, and the optimal c in terms of the risk is given by $c_1 = 1/(n + p + 1)$ and the risk of the estimator $\widehat{\boldsymbol{\Sigma}}_0 = c_1\boldsymbol{V}$ is $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0) = p(p+1)/(n+p+1)$. This can be easily seen for $n \geq p$ and it follows from Konno

[5] for $p > n$. To improve the estimator $c_1\boldsymbol{V}$, we consider a class of estimators given by

$$\widehat{\boldsymbol{\Sigma}}_\Lambda = c_1(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}}), \tag{2.1}$$

where $\widehat{\boldsymbol{\Lambda}}$ is a $p \times p$ positive definite matrix based on $\boldsymbol{V}$, and d is a positive constant. The risk difference between the two estimators $\widehat{\boldsymbol{\Sigma}}_\Lambda$ and $\widehat{\boldsymbol{\Sigma}}_0$ is denoted by

$$\Delta = R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda) - R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0).$$

To evaluate Δ asymptotically, we assume the following conditions:

(A1) Assume that $(n, p) \to \infty$. Throughout the paper, δ given in the following is a constant satisfying $0 < \delta \le 1$. Assume either (A1-1) or (A1-2) for order of (n, p), where

(A1-1) $p = O(n^\delta)$ for $0 < \delta \le 1$ in the case of $n \ge p$,
(A1-2) $n = O(p^\delta)$ for $0 < \delta \le 1$ in the case of $p > n$.

(A2) There exist limiting values

$$\lim_{p\to\infty} \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})^i]/p, \quad \lim_{p\to\infty} \mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^i]/p$$

for $i = 1, 2$, where $\boldsymbol{\Lambda}$ is a $p \times p$ symmetric matrix based on $\boldsymbol{\Sigma}$.

(A3) Assume that $\widehat{\boldsymbol{\Lambda}}$ is a $p \times p$ symmetric matrix based on $\boldsymbol{V}$ such that

$$\begin{aligned}
&\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]/p = O_p(n^{-1}),\\
&E[\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})]]/p = O(n^{-1}),\\
&E[\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})]]/p = O(n^{-1}).
\end{aligned}$$

Some examples of statistic $\widehat{\boldsymbol{\Lambda}}$ satisfying condition (A3) will be given in Section 4.

In this paper, we use the following notations:

$$\begin{aligned}
&\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}, \boldsymbol{\Gamma} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1/2}, \widehat{\boldsymbol{\Gamma}} = \boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1/2},\\
&m = n - p, \mathrm{Ch}_{max}(\boldsymbol{A}) = (\text{the largest eigenvalue of } \boldsymbol{A}).
\end{aligned}$$

Theorem 2.1. *Assume conditions* (A1)–(A3). *Then, the risk difference of the estimators* $\widehat{\boldsymbol{\Sigma}}_\Lambda = c_1(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})$ *and* $\widehat{\boldsymbol{\Sigma}}_0$ *is approximated as*

$$\begin{aligned}
\Delta = &\frac{pd}{(n+p)^2}\Big\{\frac{d}{p}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\}\\
&+ \frac{p}{n+p}\Big\{O\Big(d\frac{\sqrt{p/n}}{\sqrt{n+p}}\Big) + O\Big(\frac{d^2/n}{n+p}\Big)\Big\}.
\end{aligned} \tag{2.2}$$

Proof. The risk difference of the estimators $\widehat{\boldsymbol{\Sigma}}_\Lambda = c_1(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})$ and $\widehat{\boldsymbol{\Sigma}}_0$ is written as

$$\Delta = E\big[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I} + c_1 d\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}\}^2]\big] - R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0)$$

$$= E\big[2c_1 d\mathrm{tr}\,[(c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}] + c_1^2 d^2 \mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]\big]. \tag{2.3}$$

We shall evaluate each term in the r.h.s. of the last equality in (2.3). The first term is written as

$$\begin{aligned} &E\big[c_1 d\mathrm{tr}\,[(c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}]\big] \\ &= c_1 d(nc_1 - 1)\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + c_1 d E\big[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big] \\ &= -\frac{p(p+1)d}{(n+p+1)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]}{p} + c_1 d E\big[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big]. \end{aligned} \tag{2.4}$$

It is here noted from Cauchy-Shwartz' inequality that the inequality

$$(\mathrm{tr}\,[\boldsymbol{A}\boldsymbol{B}])^2 \le \mathrm{tr}\,[\boldsymbol{A}^2]\mathrm{tr}\,[\boldsymbol{B}^2] \tag{2.5}$$

holds for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. It is also noted that $\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}^2] = O_p(p^2(n+p)^{-1})$ since $E[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}^2]] = R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0) = p(p+1)/(n+p+1)$. Then,

$$\begin{aligned} &c_1 d\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}] \\ &\qquad \le c_1 d\big\{\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}^2]\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]\big\}^{1/2}, \end{aligned}$$

which is of order $O_p((n+p)^{-1}d[p^3 n^{-1}(n+p)^{-1}]^{1/2})$. Thus, from (2.4),

$$\begin{aligned} &E\big[c_1 d\mathrm{tr}\,[(c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}]\big] \\ &= -\frac{p(p+1)d}{(n+p+1)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]}{p} + O\Big(\frac{dp\sqrt{p/n}}{(n+p)^{3/2}}\Big). \end{aligned} \tag{2.6}$$

Finally, we estimate the term $c_1^2 d^2 E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]]$ in (2.3). Note that

$$\begin{aligned} c_1^2 d^2 E\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]\big] &= c_1^2 d^2 \mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + 2c_1^2 d^2 E\big[\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big] \\ &\quad + c_1^2 d^2 E\big[\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]\big]. \end{aligned} \tag{2.7}$$

Under condition (A3), it is observed that

$$c_1^2 d^2 E\big[\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big] = O\Big(\frac{d^2 p/n}{(n+p)^2}\Big),$$

$$c_1^2 d^2 E\big[\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]\big] = O\Big(\frac{d^2 p/n}{(n+p)^2}\Big),$$

so that

$$c_1^2 d^2 E\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]\big] = c_1^2 d^2 \mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + O\Big(\frac{d^2 p/n}{(n+p)^2}\Big). \tag{2.8}$$

Combining (2.6) and (2.8), we get

$$\Delta = -\frac{2p(p+1)d}{(n+p)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]}{p} + \frac{pd^2}{(n+p)^2}\frac{\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]}{p}$$

$$+ \frac{p}{n+p}\Big\{O\Big(\frac{d\sqrt{p/n}}{\sqrt{n+p}}\Big) + O\Big(\frac{d^2/n}{n+p}\Big)\Big\}, \tag{2.9}$$

which yields the approximation in Theorem 2.1. □

Since the leading term in (2.2) is a quadratic function of d, it can be minimized at

$$d = p\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]/\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]. \tag{2.10}$$

If we assume that there exist $\lim_{p\to\infty}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^j]/p$ for $j = 1, 2$, it is seen that the optimal d is of order p. When $\boldsymbol{\Lambda}$ is of the form $\boldsymbol{\Lambda} = \lambda\boldsymbol{I}$ for a positive parameter λ, the minimizing $d\lambda$ is

$$d\lambda = p\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]. \tag{2.11}$$

Thus, we can put $d = p$ and $\boldsymbol{\Lambda} = \{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\}\boldsymbol{I}$ as the optimal solution. In the case of $n > p$, a consistent estimator of $\boldsymbol{\Lambda}$ is given by $\widehat{\boldsymbol{\Lambda}}_3$ of Example 4.3 in Section 4. However, it is not easy to derive a consistent estimator for $\boldsymbol{\Lambda}$ when $p > n$. That is, we could not provide an estimator which minimizes the leading term in (2.2) when $d = p$ and $p > n$.

The approximation given in Theorem 2.1 shows that the leading term in (2.2) is negative when the order of d is less than p.

Corollary 2.1. *Assume conditions* (A1)–(A3) *with $d = o(p)$. Then, the estimator $\widehat{\boldsymbol{\Sigma}}_\Lambda = c_1(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})$ improves on $\widehat{\boldsymbol{\Sigma}}_0$ in terms of second order approximation of risk for any estimator $\widehat{\boldsymbol{\Lambda}}$ satisfying condition* (A3). *For instance, take $d = 1$ and $d = \max\{\sqrt{n}, \sqrt{p}\} \equiv d_{n,p}$. Then, from (2.2), it follows that for $d = 1$,*

$$\Delta = L_{n,p} + R_{n,p}, \quad \text{for} \quad L_{n,p} = -\frac{2p}{(n+p)^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}], \tag{2.12}$$

where $L_{n,p} = O(n^{2(\delta-1)})$, $R_{n,p} = O(n^{-2+3\delta/2})$ for $p = O(n^\delta)$, and $L_{n,p} = O(1)$, $R_{n,p} = O(p^{-\delta/2})$ for $n = O(p^\delta)$ for $0 < \delta \le 1$. Also, for $d = d_{n,p}$,

$$\Delta = L_{n,p} + R_{n,p}, \quad \text{for} \quad L_{n,p} = -2\frac{pd_{n,p}}{(n+p)^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}], \tag{2.13}$$

where $L_{n,p} = O(n^{2\delta-3/2})$, $R_{n,p} = O(n^{-1+\delta})$ for $p = O(n^\delta)$ and $1/2 < \delta \le 1$, and $L_{n,p} = O(p^{1/2})$, $R_{n,p} = O(p^{(1-\delta)/2})$ for $n = O(p^\delta)$ and $0 < \delta \le 1$.

Remark 2.1. For fixed $\boldsymbol{\Lambda}$ and constant c, the risk of estimator $c(\boldsymbol{V} + d\boldsymbol{\Lambda})$ is written as

$$\begin{aligned} R(\boldsymbol{\Sigma}, c(\boldsymbol{V} + d\boldsymbol{\Lambda})) = p\{n(n+p+1)c^2 - 2nc + 1\} + 2c(nc-1)d\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] \\ + c^2d^2\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]. \end{aligned}$$

When $c = 1/n$, $\boldsymbol{V}/n$ is an unbiased estimator of $\boldsymbol{\Sigma}$, and we have

$$R(\boldsymbol{\Sigma}, n^{-1}(\boldsymbol{V} + d\boldsymbol{\Lambda})) = p(p+1)/n + n^{-2}d^2\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2],$$

which is minimized at $d=0$. This means that the unbiased estimator cannot be improved on by the ridge-type estimator under the quadratic loss. The optimal c among estimators $c\boldsymbol{V}$ is $c_1 = 1/(n+p+1)$, and we have

$$R(\boldsymbol{\Sigma}, c_1(\boldsymbol{V}+d\boldsymbol{\Lambda})) \\ = \frac{p(p+1)}{n+p+1} + \frac{1}{(n+p+1)^2}\Big\{-2(p+1)d\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + d^2\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]\Big\}.$$

The second term of the above equality corresponds to the leading term in (2.2), and Theorem 2.1 guarantees that the risk of $c_1(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})$ with an estimator $\widehat{\boldsymbol{\Lambda}}$ of $\boldsymbol{\Lambda}$ can be approximated by the risk of $c_1(\boldsymbol{V}+d\boldsymbol{\Lambda})$ for fixed $\boldsymbol{\Lambda}$ under conditions (A1)–(A3). Noting that $c_1\boldsymbol{V}$ shrinks $n^{-1}\boldsymbol{V}$ toward zero, we can see from Theorem 2.1 that there is a room to improve on $c_1\boldsymbol{V}$ by expanding it with $c_1\boldsymbol{V} + c_1 d\widehat{\boldsymbol{\Lambda}}$. □

3. Estimation of precision

In this section we consider the estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$. For estimation of the covariance matrix, we have treated both cases of $n > p$ and $p > n$ in the unified framework. For the precision matrix, however, the ridge type estimator has different properties between the two cases, so that we need to treat the two cases separately.

3.1. Case of $n > p$

We begin by considering the case of $n > p$. The estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$ is addressed in terms of the risk function $R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = E[L^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1})]$ where $L^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = \mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]$, which is invariant under the scale transformation. A standard estimator of the form $c\boldsymbol{V}^{-1}$ for $c \in \mathbb{R}_+$ has the risk $R^*(\boldsymbol{\Sigma}, c\boldsymbol{V}^{-1}) = E[c^2\mathrm{tr}\,[\boldsymbol{W}^{-2}] - 2c\mathrm{tr}\,[\boldsymbol{W}^{-1}] + p]$, which is

$$R^*(\boldsymbol{\Sigma}, c\boldsymbol{V}^{-1}) = c^2\frac{p(m+p-1)}{m(m-1)(m-3)} - 2c\frac{p}{m-1} + p. \tag{3.1}$$

Thus, the best constant c is $c_2 = m(m-3)/(n-1)$, and the risk is $R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1}) = p(mp+2m-n+1)/\{(m-1)(n-1)\}$ for $\widehat{\boldsymbol{\Sigma}}_0^{-1} = c_2\boldsymbol{V}^{-1}$.

A drawback of $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ is that it may be close to be instable when p is large and $n-p$ is small. To modify the estimator $\widehat{\boldsymbol{\Sigma}}_0^{-1}$, we consider a class of estimators given by

$$\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1} = c_2(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}, \tag{3.2}$$

where $\widehat{\boldsymbol{\Lambda}}$ is a $p\times p$ positive definite matrix based on $\boldsymbol{V}$ satisfying condition (A3). Then, we investigate whether $\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1}$ improves $\widehat{\boldsymbol{\Sigma}}_0^{-1}$. Let $\Delta^* = R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda) - R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0)$. Assume the following condition:

(A4) There exist limiting values

$$\lim_{p\to\infty} \mathrm{tr}\,[\boldsymbol{\Sigma}^i]/p,\ i=1,2,$$
$$\lim_{p\to\infty} \mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^j]/p,\ j=1,2,3.$$

In the case of $\delta < 1$, we can show the following theorem which will be proved in the appendix.

Theorem 3.1. *Assume conditions* (A1-1)*,* (A3) *and* (A4)*. Also assume that* $n-p>7$ *and* $\delta<1$*, namely* $p=o(n)$*. Then, the risk difference of* $\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1}$ *and* $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ *can be approximated as*

$$\Delta^* = \frac{pd}{n^2}\Big\{\frac{d}{p}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\} + O(d^3n^{-3+\delta}) + O(d^2n^{-5/2+3\delta/2}) + O(dn^{-5/2+5\delta/2}) + O(dn^{-2+3\delta/2}). \tag{3.3}$$

In the case of $\delta=1$, we assume the condition given by

(A1-3) $(n,p)\to\infty$, $p/n\to\gamma$ and $n-p>7$ for $0<\gamma\le 1$.

Then, we can get the following result which will be shown in the appendix.

Theorem 3.2. *Assume conditions* (A1-3)*,* (A3) *and* (A4)*. Then,*

$$R^*(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}_\Lambda) = E[\mathrm{tr}\,[\{c_2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}\}^2]] + O(d^2n^{-2}) + O(dn^{-1/2}). \tag{3.4}$$

When $d=p$*, this expression can not be approximated anymore, so that we need to evaluate* $E[\mathrm{tr}\,[\{c_2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}\}^2]]$ *directly. When* $d=o(p)$*, however, the risk difference of* $\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1}$ *and* $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ *can be further approximated as*

$$\Delta^* = -2\frac{d}{n}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + O(d^3n^{-2}) + O(d^2n^{-1}) + O(dn^{-1/2}). \tag{3.5}$$

Since the leading term in (3.3) is a quadratic function of d, it can be minimized at $d = p\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]/\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]$, which is the same as in estimation of $\boldsymbol{\Sigma}$. This implies that the optimal d is of order p under condition (A4). When $\boldsymbol{\Lambda}$ is of the form $\boldsymbol{\Lambda}=\lambda\boldsymbol{I}$ for a positive parameter λ, the minimizing $d\lambda$ is $d\lambda = p\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]$, which is estimated by a consistent estimator given by $\widehat{\boldsymbol{\Lambda}}_3$ of Example 4.3 in Section 4 when $n>p$.

The approximations given in Theorems 3.1 and 3.2 show that the leading terms in (3.3) and (3.5) are negative when $d=o(p)$.

Corollary 3.1. *Assume conditions* (A1-1) *or* (A1-3)*. Also assume* (A3) *and* (A4) *with* $d=o(p)$*. Then, the estimator* $\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1} = c_2(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}$ *improves on* $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ *in terms of second order approximation of risk for any estimator* $\widehat{\boldsymbol{\Lambda}}$ *satisfying condition* (A3)*. For instance, take* $d=\max\{\sqrt{n},\sqrt{p}\}\equiv d_{n,p}$*. Then, from* (*3.3*) *and* (*3.5*)*, it follows that for* $1/2<\delta<1$*,*

$$\Delta^* = -2\frac{p\sqrt{n}}{n^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + O(n^{-1+\delta}) + O(n^{-2+5\delta/2}),$$

and for $\delta = 1$, $\Delta^* = -2n^{-1/2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + O(1)$. *For* $d = 1$, *a similar expression can be given.*

3.2. Case of $p > n$

We next consider the case of $p > n$ in the estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$. In this case, $\boldsymbol{V}$ is singular, and there does not exist the inverse of $\boldsymbol{V}$. A feasible estimator is of the form

$$\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1} = c(\boldsymbol{V} + p\widehat{\boldsymbol{\Lambda}})^{-1}. \tag{3.6}$$

A loss function treated here is the scale-invariant quadratic loss $L^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = \mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]$. Relative to this loss function, an approximation of the risk is provided under the following conditions:

(A5) There exist the limiting values $\lim_{p\to\infty} \mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^i]/p$ for $i = 1, 2$.

(A6) Assume that $\widehat{\boldsymbol{\Lambda}}$ satisfies that

$$\begin{aligned}
&\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\}^2]/p = O_p(n^{-1}),\\
&E[\mathrm{tr}\,[\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})]]/p = O(n^{-1}),\\
&E[\mathrm{tr}\,[\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})]]/p = O(n^{-1}).
\end{aligned}$$

Theorem 3.3. *Assume conditions* (A1-2), (A5) *and* (A6) *with* $c = c_{n,p} = O(p)$. *Also assume that* $\widehat{\boldsymbol{\Lambda}}$ *satisfies the following condition:*

$$\mathrm{tr}\,[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}] = O_p(np) \quad \text{and} \quad \mathrm{tr}\,[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}] = O_p(np). \tag{3.7}$$

Then, the risk of the estimator $\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}$ *given in* (3.6) *is approximated as*

$$R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}) = p\Big\{1 - 2\frac{c}{p}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]}{p} + \frac{c^2}{p^2}\frac{\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2]}{p} + O(n^{-1})\Big\} + O(n). \tag{3.8}$$

Proof. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a $p \times n$ random matrix such that $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are i.i.d. as $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Note that

$$(\boldsymbol{V} + p\widehat{\boldsymbol{\Lambda}})^{-1} = p^{-1}\widehat{\boldsymbol{\Lambda}}^{-1} - p^{-2}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}.$$

The scale-invariant quadratic loss of $\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}$ is written as

$$\begin{aligned}
&\mathrm{tr}\,[\{c(\boldsymbol{V} + p\widehat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}^2]\\
&= \mathrm{tr}\,\Big[\Big\{\frac{c}{p}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I} - \frac{c}{p^2}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\Big\}^2\Big]\\
&= \mathrm{tr}\,\Big[\Big\{\frac{c}{p}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\Big\}^2\Big]
\end{aligned}$$

$$- 2\frac{c}{p^2}\mathrm{tr}\left[(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\{\frac{c}{p}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\right]$$
$$+ \frac{c^2}{p^4}\mathrm{tr}\left[\big\{(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\big\}^2\right],$$

where the second term in the r.h.s. of the last equality is of order $O_p(n)$ from condition (3.7). For the third term, it is observed that

$$\begin{aligned}
&p^{-2}\mathrm{tr}\left[\big\{(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\big\}^2\right]\\
&\quad\le p^{-1}\mathrm{tr}\left[(\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\big\{\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\big\}^2\right]\\
&\quad= p^{-1}\mathrm{tr}\left[\boldsymbol{P}\widehat{\boldsymbol{\Lambda}}^{-1/2}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1/2}\right]\\
&\quad\le p^{-1}\mathrm{tr}\left[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\right],
\end{aligned}$$

which is of order $O_p(n)$ from condition (3.7), where

$$\boldsymbol{P} = \widehat{\boldsymbol{\Lambda}}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1/2}.$$

Thus,

$$\mathrm{tr}\,[\{c(\boldsymbol{V} + p\widehat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}^2] = \mathrm{tr}\,[\{(c/p)\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}^2] + O_p(n). \tag{3.9}$$

We next evaluate the first term in the r.h.s. of (3.9) as

$$\begin{aligned}
&\frac{c^2}{p^2}\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma})^2] - 2\frac{c}{p}\mathrm{tr}\,[\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}]\\
&= \frac{c^2}{p^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2] - 2\frac{c}{p}\mathrm{tr}\,[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}] + 2\frac{c^2}{p^2}\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]\\
&\quad + \frac{c^2}{p^2}\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\}^2] - 2\frac{c}{p}\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}]\\
&= \frac{c^2}{p^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2] - 2\frac{c}{p}\mathrm{tr}\,[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}] + O_p(p/n),
\end{aligned}$$

from condition (A6). This shows (3.8). □

Concerning condition (3.7), it is seen that if $\widehat{\boldsymbol{\Lambda}}$ satisfies $\mathrm{Ch}_{max}(\widehat{\boldsymbol{\Lambda}}^{-1}) = O_p(1)$ for large (n, p) satisfying (A1-2), then condition (3.7) is satisfied under condition $\mathrm{tr}\,[\boldsymbol{\Sigma}^i]/p = O(1)$ for $i = 1, 2, 3$. In fact, since $\mathrm{Ch}_{max}(\widehat{\boldsymbol{\Lambda}}^{-1}) = O_p(1)$, it is sufficient to show that $E[\mathrm{tr}\,[\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\Sigma}^i]] = O(np)$, $i = 1, 2$, which can be easily verified if $\mathrm{tr}\,[\boldsymbol{\Sigma}^i]/p = O(1)$, $i = 1, 2, 3$.

As an example of $\widehat{\boldsymbol{\Lambda}}$, we consider the case of $\widehat{\boldsymbol{\Lambda}} = \hat{\lambda}\boldsymbol{I}_p$ for positive scalor function $\hat{\lambda}$ of $\boldsymbol{V}$, namely, the estimator given in (3.6) with $c = p$ is

$$\widehat{\boldsymbol{\Sigma}}_{\lambda}^{-1} = p(\boldsymbol{V} + p\hat{\lambda}\boldsymbol{I}_p)^{-1}. \tag{3.10}$$

Conditions on $\hat{\lambda}$ for the approximation of the risk given in (3.8) are provided from Theorem 3.3. For the ridge estimator (3.10), we can also use an approach based on the eigenvalue decomposition. Since this approach is useful for understanding ridge estimators in the case of $p > n$, we here describe the expressions based on the eigenvalue docomposition for the ridge-type estimator and the risk. Let $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$ be a $p \times p$ orthogonal matrix such that

$$\boldsymbol{V} = \boldsymbol{H}\begin{pmatrix} \boldsymbol{L} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}' = \boldsymbol{H}_1\boldsymbol{L}\boldsymbol{H}_1', \ \boldsymbol{L} = \mathrm{diag}\,(\ell_1, \ldots, \ell_n), \tag{3.11}$$

where $\ell_1 \geq \cdots \geq \ell_n$, and $\boldsymbol{H}_1$ is a $p \times n$ matrix satisfying $\boldsymbol{H}_1'\boldsymbol{H}_1 = \boldsymbol{I}_n$. Then, the estimator (3.10) is expressed as

$$\begin{aligned} p(\boldsymbol{V} + p\hat{\lambda}\boldsymbol{I}_p)^{-1} &= \boldsymbol{H}\begin{pmatrix} p(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\lambda}^{-1}\boldsymbol{I}_{p-n} \end{pmatrix}\boldsymbol{H}' \\ &= \boldsymbol{H}\begin{pmatrix} p(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1} - \hat{\lambda}^{-1}\boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}' + \hat{\lambda}^{-1}\boldsymbol{I}_p \\ &= \hat{\lambda}^{-1}\big(\boldsymbol{I}_p - \boldsymbol{H}_1\boldsymbol{L}(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1}\boldsymbol{H}_1'\big), \end{aligned} \tag{3.12}$$

and the risk function is written as

$$\begin{aligned} R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_{\lambda}^{-1}) &= E\big[\mathrm{tr}\,[\{\hat{\lambda}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}_p\}^2]\big] + E\big[\hat{\lambda}^{-2}\mathrm{tr}\,[\{\boldsymbol{L}(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1\}^2]\big] \\ &\quad - 2E\big[\hat{\lambda}^{-1}\mathrm{tr}\,[\{\hat{\lambda}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}_p\}\boldsymbol{H}_1\boldsymbol{L}(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}]\big]. \end{aligned} \tag{3.13}$$

Note that

$$\begin{aligned} &\mathrm{tr}\,[\{\boldsymbol{L}(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1\}^2] \leq \mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^2\boldsymbol{H}_1], \\ &\mathrm{tr}\,[\boldsymbol{\Sigma}\boldsymbol{H}_1\boldsymbol{L}(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}] \leq \mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^2\boldsymbol{H}_1], \\ &\mathrm{tr}\,[\boldsymbol{H}_1\boldsymbol{L}(\boldsymbol{L} + p\hat{\lambda}\boldsymbol{I}_n)^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}] \leq \mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1]. \end{aligned}$$

Then the risk expression (3.13) provides conditions on $\hat{\lambda}$ for the approximation of the risk.

Proposition 3.1. *Assume that* $E[\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1]] = O(n)$, $E[\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^2\boldsymbol{H}_1]] = O(n)$ *and* $\mathrm{tr}\,[\boldsymbol{\Sigma}^i]/p = O(1)$ *for* $i = 1, 2$. *Also assume that there exists a positive constant* λ *such that* $\hat{\lambda} - \lambda = O_p(n^{-1/2})$ *and* $E[\hat{\lambda} - \lambda] = O(n^{-1})$. *Under the condition* (A1-2), *the risk of the estimator (3.10) is approximated as*

$$R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_{\lambda}^{-1}) = p\big\{1 - 2\lambda^{-1}a_1 + \lambda^{-2}a_2 + O(n^{-1})\big\} + O(n), \tag{3.14}$$

where $a_1 = \mathrm{tr}\,[\boldsymbol{\Sigma}]/p$ *and* $a_2 = \mathrm{tr}\,[\boldsymbol{\Sigma}^2]/p$.

The approach based on the eigenvalue decomposition enables us to approximate the risk of the estimator $\widehat{\boldsymbol{\Sigma}}_0^{-1} = p\boldsymbol{V}^+$, where $\boldsymbol{V}^+$ is the Moore-Penrose generalized inverse of $\boldsymbol{V}$. Using the decomposition (3.11), we can rewrite $p\boldsymbol{V}^+$ as

$$\begin{aligned} p\boldsymbol{V}^+ &= \boldsymbol{H}\begin{pmatrix} p\boldsymbol{L}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}' \\ &= p\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1'. \end{aligned} \tag{3.15}$$

The risk function of $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ is

$$\begin{aligned}R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1}) &= p - 2E[\mathrm{tr}\,[p\boldsymbol{V}^+\boldsymbol{\Sigma}]] + E[\mathrm{tr}\,[(p\boldsymbol{V}^+\boldsymbol{\Sigma})^2]] \\ &= p\big\{1 - 2E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1]] + E[p\mathrm{tr}\,[(\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1)^2]]\big\}.\end{aligned}$$

It follows from Lemma A.1 that

$$\begin{aligned}E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1]] &\le \mathrm{Ch}_{max}(\boldsymbol{\Sigma})\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\frac{n}{p-n-1}, \\ E[p\mathrm{tr}\,[(\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1)^2]] &\le \frac{\{\mathrm{Ch}_{max}(\boldsymbol{\Sigma})\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\}^2 np(p-1)}{\{(p-n-1)(p-n-3)-2\}(p-n-1)},\end{aligned}$$

both of which are of order $O(p^{\delta-1})$ when $n = O(p^\delta)$ for $0 < \delta < 1$. Hence, we get the following proposition.

Proposition 3.2. *Assume that* $\mathrm{Ch}_{max}(\boldsymbol{\Sigma})$ *and* $\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})$ *are bounded for large* p, *and that* $p - n \ge 4$, $(n, p) \to \infty$ *and* $n = O(p^\delta)$ *for* $0 < \delta < 1$. *Then,* $R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1}) = p + O(p^\delta)$.

Combining Propositions 3.1 and 3.2 gives the following asymptotic approximation for $\Delta^* = R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\lambda^{-1}) - R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1})$.

Corollary 3.2. *Assume the conditions given in Propositions 3.1 and 3.2. Then,*

$$\frac{\Delta^*}{p} = -2\frac{a_1}{\lambda} + \frac{a_2}{\lambda^2} + O(p^{-\delta}) + O(p^{\delta-1}). \tag{3.16}$$

The leading term in (3.16) is minimized at $\lambda = \mathrm{tr}\,[\boldsymbol{\Sigma}^2]/\mathrm{tr}\,[\boldsymbol{\Sigma}] = a_2/a_1$. A consistent estimator of $\boldsymbol{\Lambda} = \lambda\boldsymbol{I}$ is given by $\widehat{\boldsymbol{\Lambda}}_2$ in Example 4.2 in the next section.

Remark 3.1. The expressions (3.12) and (3.13) of the ridge estimator $\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}$ and the risk function tell us about appropriate order of the ridge function $p\hat{\lambda}$. When we assume that $E[\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1]] = O(n)$, $E[\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^2\boldsymbol{H}_1]] = O(n)$ and $\mathrm{tr}\,[\boldsymbol{\Sigma}^i]/p = O(1)$ for $i = 1, 2$, we can evaluate the risk in (3.13) as

$$R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\lambda^{-1}) = E\Big[p - 2\frac{p}{\hat{\lambda}}a_1 + \frac{p}{\hat{\lambda}^2}a_2 + \frac{n}{\hat{\lambda}^2} \times O_p(1) + \frac{n}{\hat{\lambda}} \times O_p(1)\Big]. \tag{3.17}$$

When $1/\hat{\lambda} = O_p(1)$, we get the approximation given in Proposition 3.1. If $\hat{\lambda} = o_p(1)$ or $\lim_{p\to\infty} 1/\hat{\lambda} = \infty$, then $(p/\hat{\lambda}^2)a_2$ in (3.17) is the dominating term, which implies that $R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\lambda^{-1})/p \to \infty$ as $(p, n) \to \infty$. Hence, in this case, the ridge estimator with $\hat{\lambda} = o_p(1)$ has a larger risk than $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ asymptotically. This phenomenon is understandable when we compare the expressions in (3.12) and (3.15). When $\hat{\lambda} = o_p(1)$, the lower right matrix component $\hat{\lambda}^{-1}\boldsymbol{I}_{p-n}$ in (3.12) diverges while the lower right matrix component in (3.15) is zero. These arguments suggest that $\hat{\lambda}$ should be taken so as to satisfy $1/\hat{\lambda} = O_p(1)$.

Kubokawa and Srivastava [6] showed that $p\boldsymbol{V}^+$ is the best estimator among $a\boldsymbol{V}^+$ for $a \in \mathbb{R}_+$ relative to the loss $L_{KS}(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{V}) = \mathrm{tr}\,[\{\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\}^2\boldsymbol{V}^2]$, and established the exact dominance result that $p\boldsymbol{V}^+$ is dominated by

$$\widehat{\boldsymbol{\Sigma}}_{KS}^{-1} = p\Big(\boldsymbol{V} + \frac{c}{\mathrm{tr}\, \boldsymbol{V}^+}\boldsymbol{I}_p\Big)^{-1},$$

in terms of the loss $L_{KS}(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{V})$ if $0 < c \le 2(n-1)/p$. Since $\mathrm{tr}\,[\boldsymbol{V}^+] = O_p(n/p)$ from Lemma A.1, it is seen that $c/\mathrm{tr}\,[\boldsymbol{V}^+] = p\hat{\lambda}_{KS}$ for

$$\hat{\lambda}_{KS} = \frac{c/p}{\mathrm{tr}\,[\boldsymbol{V}^+]} = O_p(p^{-1}).$$

Since $\hat{\lambda}_{KS} = o_p(1)$, the estimator $\widehat{\boldsymbol{\Sigma}}_{KS}^{-1}$ has a larger risk than $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ asymptotically relative to the loss $L(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2]$. Why does the estimator $\widehat{\boldsymbol{\Sigma}}_{KS}^{-1}$ improve on $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ in terms of the loss $L_{KS}(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{V})$? The reason is that the lower right matrix component $\hat{\lambda}^{-1}\boldsymbol{I}_{p-n}$ in (3.12) disappears relative to the loss $L_{KS}(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{V})$ since $\boldsymbol{V}^2$ is incorporated in the loss function as seen from (3.11). □

4. Examples of statistic $\widehat{\boldsymbol{\Lambda}}$ for estimating $\boldsymbol{\Lambda}$

In this section, we provide some examples of statistic $\widehat{\boldsymbol{\Lambda}}$ satisfying conditions (A3) and/or (A6).

Example 4.1. Consider the statistic given by

$$\widehat{\boldsymbol{\Lambda}}_1 = \hat{a}_1\boldsymbol{I} \quad \text{for} \quad \hat{a}_1 = \mathrm{tr}\,[\boldsymbol{V}]/(np). \tag{4.1}$$

This is an unbiased estimator of $\boldsymbol{\Lambda}_1 = a_1\boldsymbol{I}$ for $a_1 = \mathrm{tr}\,[\boldsymbol{\Sigma}]/p$. Srivastava [11] showed that $\hat{a}_1 - a_1 = O_p((np)^{-1/2})$ under condition (A2) for large n or p. This shows that $\widehat{\boldsymbol{\Lambda}}_1 = \hat{a}_1\boldsymbol{I}$ satisfies condition (A3). Note that

$$\hat{a}_1^{-1} - a_1^{-1} = -a_1^{-2}(\hat{a}_1 - a_1) + a_1^{-3}(\hat{a}_1 - a_1)^2 + o_p((np)^{-1}).$$

Since $E[\hat{a}_1 - a_1] = 0$, $E[(\hat{a}_1 - a_1)^2] = O((np)^{-1})$ and $\hat{a}_1 = O_p(1)$, it is easily verified that condition (A6) is satisfied. □

Example 4.2. Consider the statistic given by

$$\widehat{\boldsymbol{\Lambda}}_2 = (\hat{a}_2/\hat{a}_1)\boldsymbol{I} \quad \text{for} \quad \hat{a}_2 = \frac{1}{(n-1)(n+2)p}\Big[\mathrm{tr}\,[\boldsymbol{V}^2] - (\mathrm{tr}\,[\boldsymbol{V}])^2/n\Big]. \tag{4.2}$$

Srivastava [11] showed that $\hat{a}_2$ is an unbiased estimator of $a_2 = \mathrm{tr}\,[\boldsymbol{\Sigma}^2]/p$ and $\hat{a}_2 - a_2 = O_p((np)^{-1/2}) + O_p(n^{-1})$ under condition (A2) for large n and p. Note that

$$\begin{aligned}\frac{\hat{a}_2}{\hat{a}_1} - \frac{a_2}{a_1} &= \frac{a_2}{a_1}\Big\{\frac{\hat{a}_2 - a_2}{a_2} - \frac{\hat{a}_1 - a_1}{a_1}\Big\} \\ &\quad - \frac{a_2}{a_1}\Big\{\frac{(\hat{a}_1 - a_1)^2}{a_1^2} + \frac{(\hat{a}_1 - a_1)(\hat{a}_2 - a_2)}{a_1 a_2}\Big\} + O_p(n^{-3/2}),\end{aligned} \tag{4.3}$$

which implies that $E[\hat{a}_2/\hat{a}_1 - a_2/a_1] = O(n^{-1})$ and $E[(\hat{a}_2/\hat{a}_1 - a_2/a_1)^2] = O(n^{-1})$. This shows that $\widehat{\boldsymbol{\Lambda}}_2$ satisfies condition (A3). Similar to (4.3), it can be verified that condition (A6) is satisfied, since $\hat{a}_2/\hat{a}_1 = O_p(1)$. It is noted that $p\widehat{\boldsymbol{\Lambda}}_2$ provides an optimal estimator which minimizes the leading term in (3.16) when $p > n$. □

Example 4.3. Consider the case of $n > p$ and $p = O(n^\delta)$ for $0 < \delta \le 1$. Then, Proposition A.1 proves that the estimators given by

$$\begin{aligned}\hat{b}_1 &= \frac{m}{p}\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}],\\ \hat{b}_2 &= \frac{m^2}{p}\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-2}] - \frac{p}{m}(\hat{b}_1)^2,\end{aligned} \tag{4.4}$$

are consistent estimators of b_1 and b_2, respectively, where $\hat{a}_1$ is given in (4.1) and $b_i = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-i}]/p$ for $i = 1, 2$. That is, $\hat{b}_1 - b_1 = O_p((np)^{-1/2})$ and $\hat{b}_2 - b_2 = O_p(n^{-1/2})$. Based on these statistics, we consider the statistic

$$\widehat{\boldsymbol{\Lambda}}_3 = (\hat{b}_1/\hat{b}_2)\boldsymbol{I}. \tag{4.5}$$

Similarly to (4.3), we can see that $E[\hat{b}_1/\hat{b}_2 - b_1/b_2] = O(n^{-1})$ and $E[(\hat{b}_1/\hat{b}_2 - b_1/b_2)^2] = O(n^{-1})$. This shows that $\widehat{\boldsymbol{\Lambda}}_3$ satisfies condition (A3). It is noted that $p\widehat{\boldsymbol{\Lambda}}_3$ provides an optimal estimator which minimizes the leading terms in (2.2) and (3.3) when $n > p$. □

Example 4.4. Consider the statistic given by

$$\widehat{\boldsymbol{\Lambda}}_4 = n^{-1}\mathrm{diag}\,(v_{11}, \ldots, v_{pp}), \tag{4.6}$$

where v_{ii} is the i-th diagonal element of $\boldsymbol{V}$. Then, $\widehat{\boldsymbol{\Lambda}}_4$ is an unbiased estimator of $\boldsymbol{\Lambda} = \mathrm{diag}\,(\sigma_{11}, \ldots, \sigma_{pp})$. We shall verify conditions (A3) and (A6). Note that $v_{ii}/\sigma_{ii} \sim \chi^2_n$ and that $E[(v_{ii}/n - \sigma_{ii})^2] = 2\sigma^2_{ii}/n$. For (A3), it is seen that for $\boldsymbol{\Sigma}^{-1} = (\sigma^{ij})$,

$$\begin{aligned}E[\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}_4 - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]] &= \sum_{i,j}\sigma^{ij}\sigma^{ji}E[(v_{ii}/n - \sigma_{ii})(v_{jj}/n - \sigma_{jj})]\\ &\le \sum_{i,j}\sigma^{ij}\sigma^{ji}\{E[(v_{ii}/n - \sigma_{ii})^2]E[(v_{jj}/n - \sigma_{jj})^2]\}^{1/2}\\ &= \frac{2}{n}\sum_{i,j}\sigma^{ij}\sigma^{ji}\sigma_{ii}\sigma_{jj} = \frac{2}{n}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2].\end{aligned} \tag{4.7}$$

Thus, condition (A3) holds if $\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]/p = O(1)$ for lage p. For (A6), it is seen that

$$\begin{aligned}&E[\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}_4^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\}^2]]\\ &\quad = \sum_{i,j}\frac{\sigma^{ij}\sigma^{ji}}{\sigma_{ii}\sigma_{jj}}E[(1 - n\sigma_{ii}/v_{ii})(1 - n\sigma_{jj}/v_{jj})]\end{aligned}$$

$$\le \sum_{i,j} \frac{\sigma^{ij}\sigma^{ji}}{\sigma_{ii}\sigma_{jj}} \{E[(1-n\sigma_{ii}/v_{ii})^2]E[(1-n\sigma_{jj}/v_{jj})^2]\}^{1/2}$$
$$= \frac{2(n+4)}{(n-2)(n-4)} \sum_{i,j} \frac{\sigma_{ij}\sigma_{ji}}{\sigma_{ii}\sigma_{jj}} = \frac{2(n+4)}{(n-2)(n-4)} \mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2],$$

which is of order $O(p/n)$ if $\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2]/p = O(1)$ for lage p. Similarly,

$$E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}_4^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}]] = 2(n-2)^{-1}p,$$
$$E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}_4^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]] = 2(n-2)^{-1}\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2].$$

Hence, conditions (A6) holds for $\widehat{\boldsymbol{\Lambda}}_4$. For condition (3.7), it is noted that if $\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]/p = O(1)$, then

$$\mathrm{tr}\,[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}] = \mathrm{Ch}_{max}(\boldsymbol{\Sigma})\mathrm{tr}\,[\widehat{\boldsymbol{\Lambda}}^{-2}\boldsymbol{X}\boldsymbol{X}'] = \mathrm{Ch}_{max}(\boldsymbol{\Sigma})\sum_{i=1}^{p} n^2/v_{ii},$$

where v_{ij} denotes the (i,j) element of $\boldsymbol{X}\boldsymbol{X}'$. Here,

$$\sum_{i=1}^{p} E[n^2/v_{ii}] = n^2 \sum_{i=1}^{p} E[(\sigma_{ii}\chi_n^2)^{-1}] = n^2(n-2)^{-1}\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}] = O(np),$$

so that $\mathrm{tr}\,[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}] = O_p(np)$ if $\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]/p = O(1)$ and $\mathrm{Ch}_{max}(\boldsymbol{\Sigma}) = O(1)$. Similarly, it can be seen that $E[\mathrm{tr}\,[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}]] = O(np)$ if $\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]/p = O(1)$. Thus, condition (3.7) is satisfied for $\widehat{\boldsymbol{\Lambda}}_4$ if $\mathrm{Ch}_{max}(\boldsymbol{\Sigma}) = O(1)$. □

5. Simulation studies

We now investigate the numerical performances of the risk functions of the ridge-type estimators through simulation.

As a structure of the covariance matrix, we consider a matrix of the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{pmatrix} \begin{pmatrix} \rho^{|1-1|/7} & \cdots & \rho^{|1-p|/7} \\ \vdots & \vdots & \vdots \\ \rho^{|p-1|/7} & \cdots & \rho^{|p-p|/7} \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{pmatrix},$$

for a constant ρ on the interval $(-1,1)$ and $\sigma_i = 3 + 0.2(-1)^{i-1}(p-i+1)/p$. As a model for simulation experiments, we treat the following three cases for random variables $\boldsymbol{x}_i$'s, $i = 1,\ldots,n$, where $\boldsymbol{x}_i$ is generated as $\boldsymbol{x}_i = \boldsymbol{\Sigma}^{1/2}\boldsymbol{z}_i$ for $\boldsymbol{z}_i = (z_{i1},\ldots,z_{pi})$ with $z_{i1},\ldots,z_{pi}$ being mutually independent.

(Case 1) $z_{ij} \sim \mathcal{N}(0,1)$,
(Case 2) $z_{ij} = (u_{ij} - m)/\sqrt{2m}$, $u_{ij} \sim \chi_m^2$ for $m = 8$,
(Case 3) $z_{ij} = (u_{ij} - m)/\sqrt{2m}$, $u_{ij} \sim \chi_m^2$ for $m = 2$.

The last two cases treat non-normal cases. Since the skewness and kurtosis (K_4+3) of χ^2_m is, repectively, $(8/m)^{1/2}$ and $3+12/m$, it is noted that χ^2_2 has higher skewness and kurtosis than χ^2_8.

Let $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$ for $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Then for estimation of $\boldsymbol{\Sigma}$, we can calculate the four kinds of ridge estimators $\widehat{\boldsymbol{\Sigma}}_{\Lambda,i} = c(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}}_i)$ for $\widehat{\boldsymbol{\Lambda}}_i$'s given in (4.1), (4.2), (4.5) and (4.6), which are denoted by Rid_1, Rid_2, Rid_3 and Rid_4. In the estimation of $\boldsymbol{\Sigma}$, c is given by $c = 1/(n+p+1)$. As values of d, we treat the three cases: $d = 1$, p and $d_{n,p}$ for $d_{n,p} = \max\{\sqrt{n}, \sqrt{p}\}$. We use these notations for estimation of $\boldsymbol{\Sigma}^{-1}$, where the constant c is $c = m(m-3)/(n-1)$ in the case of $n > p$ and $c = p$ for $p > n$. It is noted that $\widehat{\boldsymbol{\Lambda}}_3$ or Rid_3 is not available for $p > n$.

The simulation experiments are carried out under the above model for $(n,p) = (200, 20)$, $(100, 20)$, $(50, 100)$ and $(80, 100)$ and $\rho = 0.2$. Based on 10,000 replications, we calculate averages of the following *Relative Risk Gain* of the ridge estimators:

$$RRG_i = 100 \times \frac{E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2]] - E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_{\Lambda,i}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2]]}{E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2]]},$$

$$RRG_i^* = 100 \times \frac{E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]] - E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_{\Lambda,i}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]]}{E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]]},$$

where $\widehat{\boldsymbol{\Sigma}}_0^{-1} = p\boldsymbol{V}^+$ in the case of $p > n$.

The simulation results for estimation of $\boldsymbol{\Sigma}$ are reported in Table 1. In the cases of $d = 1$ and $d = d_{n,p}$, the ridge estimators are better than the standard estimator. This agrees with the analytical results given in Corollary 2.1, while the improvements for $d = 1$ are quite small. It is revealed from Table 1 that the performance of the ridge estimator Rid_1 is better than the others. Thus, the ridge estimator Rid_1 with $d = p$ is recommendable for estimating $\boldsymbol{\Sigma}$.

TABLE 1
Values of RRG_i in estimation of $\boldsymbol{\Sigma}$ for $\rho = 0.2$, where $d_{n,p} = \max(\sqrt{n}, \sqrt{p})$, and $\widehat{\boldsymbol{\Lambda}}_3$ is not available for $p > n$

			$d=1$				$d=p$				$d=d_{n,p}$			
Case	n	p	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4
	200	20	1.0	1.3	0.8	0.9	8.9	6.4	9.0	7.1	8.7	9.0	7.8	7.4
Case 1	100	20	1.8	2.4	1.0	1.7	16.1	11.3	14.1	12.8	13.3	15.5	8.6	11.7
	50	100	1.5	2.7	NA	1.5	62.5	−26.6	NA	56.7	14.2	24.4	NA	13.9
	80	100	1.3	2.3	NA	1.2	52.1	−18.6	NA	48.3	11.9	20.4	NA	11.6
	200	20	0.9	1.3	0.7	0.8	8.3	5.9	8.5	5.4	8.1	8.3	7.3	6.1
Case 2	100	20	1.7	2.3	1.0	1.4	15.1	10.2	13.5	9.6	12.5	14.6	8.2	9.9
	50	100	1.5	2.8	NA	1.4	62.1	−33.6	NA	52.3	14.2	24.7	NA	13.6
	80	100	1.2	2.3	NA	1.2	51.7	−21.6	NA	45.3	11.8	20.4	NA	11.4
	200	20	0.8	1.1	0.6	0.5	6.8	4.6	7.2	1.4	6.7	6.9	6.2	2.9
Case 3	100	20	1.4	2.0	0.9	1.0	12.7	7.5	12.0	2.0	10.6	12.4	7.4	5.5
	50	100	1.5	2.8	NA	1.4	61.0	−50.0	NA	39.5	14.0	25.3	NA	12.8
	80	100	1.2	2.3	NA	1.1	50.7	−29.4	NA	36.5	11.6	20.5	NA	10.7

Table 2
Values of RRG_i^* in estimation of $\boldsymbol{\Sigma}^{-1}$ for $\rho = 0.2$ in the case of $n > p$

			$d=1$				$d=p$				$d=d_{n,p}$			
Case	n	p	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4
Case 1	200	20	1.2	1.6	0.9	1.1	2.8	−3.3	5.4	2.0	5.7	3.4	6.3	4.9
	100	20	2.6	3.4	1.5	2.4	−5.1	−17.6	6.9	−5.2	7.3	3.9	8.4	6.7
Case 2	200	20	1.4	1.9	1.1	1.2	6.0	0.6	7.9	4.6	8.0	6.3	8.2	6.7
	100	20	3.1	4.1	1.8	2.7	−0.6	−12.8	10.0	−0.9	10.3	7.4	10.4	9.2
Case 3	200	20	1.9	2.6	1.5	1.5	14.3	11.3	14.5	11.2	14.1	14.3	13.0	11.3
	100	20	4.4	6.0	2.7	3.5	12.0	0.9	19.2	10.8	18.8	17.4	16.6	16.0

Table 3
Values of RRG_i^* in estimation of $\boldsymbol{\Sigma}^{-1}$ for $\rho = 0.2$ and $c = d = p$ in the case of $p > n$

Case	n	p	Rid_1	Rid_2	Rid_4
Case 1	50	100	82.4	67.9	80.3
	80	100	99.7	99.5	99.7
Case 2	50	100	82.8	68.0	80.0
	80	100	99.7	99.5	96.7
Case 3	50	100	84.0	69.0	78.5
	80	100	99.7	99.5	99.7

The simulation results for estimation of $\boldsymbol{\Sigma}^{-1}$ are reported in Table 2 for $n > p$ and Table 3 for $p > n$. In the case of $n > p$, the improvements of the ridge estiamtors with $d = 1$ over the standard estimator are small. The performances of the ridge estimators with $d = d_{n,p}$ are good. In the case of $p > n$, the ridge estimator Rid_1 with $d = p$ has a slightly better performance. Thus, we can use the ridge estimator Rid_1 where constant d is given by $d = d_{n,p}$ for $n > p$ and by $d = p$ for $p > n$.

6. Concluding remarks

In this paper, we have considered estimation of the covariance and precision matrices by the ridge-type estimators, and have derived asymptotic expansions of their risk functions relative to the scale-invariant quadratic loss functions when the sample size and the dimension are very large. These expansions clarify the conditions for the ridge-type estimators to have smaller risks than the standard estimators in terms of the second-order terms.

The conditions for the improvement depend on the choice of the ridge function $\widehat{\boldsymbol{\Lambda}}$ and the order of d, namely, in estimation of the covariance matrix, if the following inequality holds

$$(d/p)\mathrm{tr}\,[(\boldsymbol{\Lambda\Sigma}^{-1})^2] \le 2\mathrm{tr}\,[\boldsymbol{\Lambda\Sigma}^{-1}], \tag{6.1}$$

then the ridge-type estimators improve on the standard estimators asymptotically relative to the scale-invariant quadratic loss function in both cases of $n > p$ and $p > n$. It is interesting to note that in estimation of the precision matrix, under the same condition as in (6.1), the ridge-type estimator improves on the standard estimator asymptotically in the case of $n > p$. However, the condition for the improvement in estimation of the precision matrix in the case of $p > n$

is slightly different from (6.1). Although condition (6.1) always holds asymptotically when $d = o(p)$, it depends on $\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}$ in the case of $d = p$. Various variants of the ridge-type estimators have been investigated through the performances of the risk functions by simulation.

We would like to mention the loss functions for measuring estimation errors of estimators. In this paper we have used the scale-invariant quadratic loss functions

$$L(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \operatorname{tr}[(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2],$$
$$L^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = \operatorname{tr}[(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2],$$

which are used for estimating $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$, respectively. Another scale-invariant loss functions are the Stein loss functions given by

$$L_S(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \operatorname{tr}[\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}] - \log|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p,$$
$$L_S^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = \operatorname{tr}[\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}] - \log|\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}| - p.$$

Both scale-invariant loss functions have been used in most literature when $n > p$, since there exist minimax estimators with constant risks under the scale-invariant losses. Although the results under the Stein losses are omitted here, we can have similar properties as in the case of the scale-invariant quadratic losses. When $p > n$, however, the situation is different from the case of $n > p$. For instance, we can not employ the Stein loss for measuring the risk of the estimator $c\boldsymbol{V}$, since the determinants $|c\boldsymbol{V}\boldsymbol{\Sigma}^{-1}|$ and $|(c\boldsymbol{V})^{-1}\boldsymbol{\Sigma}|$ do not exist. In estimation of $\boldsymbol{\Sigma}$ in the case of $p > n$, Konno [5] used the scale-invariant quadratic loss $L(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}})$, and Ledoit and Wolf [8, 9] treated the non-scale-invariant quadratic loss function $\operatorname{tr}[(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})^2]$. No one knows which is appropriate for estimation of $\boldsymbol{\Sigma}$ in the case of $p > n$. In estimation of $\boldsymbol{\Sigma}^{-1}$, Kubokawa and Srivastava [6] used the loss functions $\operatorname{tr}[(\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1})^2\boldsymbol{V}^k]$ for $k = 0, 1, 2$. Although they obtained ridge-type estimators improving on the best multiple of $\boldsymbol{V}^+$ for finite n and p such as $p > n$, their ridge-type estimators are not necessarily stable as $(p, n) \to \infty$ as pointed out in Remark 3.1. Thus, possible loss functions for estimation of $\boldsymbol{\Sigma}^{-1}$ are the scale-invariant quadratic loss $L^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1})$ and the quadratic loss $\operatorname{tr}[(\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1})^2]$. It may be interesting to investigate which loss is appropriate for measuring estimators of $\boldsymbol{\Sigma}^{-1}$.

Finally, it is noted that the validity of the asymptotic expansions will not be discussed here. All the results in this paper are based on major terms obtained by Taylor series expansions. Although this paper provides the second order approximations without the validity, we need more conditions and many more steps for establishing the validity of the second-order approximations.

Acknowledgments

The authors are grateful to the Editor, the Associate Editor and two referees for their valuable comments and helpful suggestions. The research of the first author was supported in part by Grant-in-Aid for Scientific Research Nos. 21540114 and 23243039 from Japan Society for the Promotion of Science.

Appendix

A.1. Identities useful for evaluation of moments

The following identity derived by Konno [5] is useful. It is related to the Stein-Haff identity given by Stein [12] and Haff [4] for $n > p$, but it can be used in both cases of $n > p$ and $n \le p$. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a $p \times n$ random matrix such that $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are i.i.d. as $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Let $\boldsymbol{G}(\boldsymbol{V})$ be a $p \times p$ matrix of functions of $\boldsymbol{V}$. Then, Konno [5] derived the identity given by

$$E\big[\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{G}(\boldsymbol{V})]\big] = E\big[n\mathrm{tr}\,[\boldsymbol{G}(\boldsymbol{V})] + \mathrm{tr}\,[\boldsymbol{X}\boldsymbol{\nabla}'_X\boldsymbol{G}(\boldsymbol{V})']\big], \tag{A.1}$$

where $\boldsymbol{\nabla}_X = (\partial/\partial X_{ij})$ for $\boldsymbol{X} = (X_{ij})$.

In the case of $n > p$, we can use the Stein-Haff identity to evaluate higher moments of $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}$. Let $\boldsymbol{G}(\boldsymbol{W})$ be a $p \times p$ matrix such that the (i, j) element $g_{ij}(\boldsymbol{W})$ is a differentiable function of $\boldsymbol{W} = (w_{ij})$ and denote $\{\mathcal{D}_W\boldsymbol{G}(\boldsymbol{W})\}_{ac} = \sum_b d_{ab}g_{bc}(\boldsymbol{W})$, where $d_{ab} = 2^{-1}(1 + \delta_{ab})\partial/\partial w_{ab}$ with $\delta_{ab} = 1$ for $a = b$ and $\delta_{ab} = 0$ for $a \ne b$. In the case of $n > p$, Stein [12] and Haff [4] derived the Stein-Haff identity given by

$$E\left[\mathrm{tr}\,\{\boldsymbol{G}(\boldsymbol{W})\}\right] = E\left[(m-1)\mathrm{tr}\,\left\{\boldsymbol{G}(\boldsymbol{W})\boldsymbol{W}^{-1}\right\} + 2\mathrm{tr}\,\{\mathcal{D}_W\boldsymbol{G}(\boldsymbol{W})\}\right], \tag{A.2}$$

for $m = n - p$.

In the case of $p > n$, the corresponding identity was provided by Kubokawa and Srivastava [6]. This identity was also derived from (A.1) by Konno [5]. Let $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$ be a $p \times p$ orthogonal matrix such that

$$\boldsymbol{V} = \boldsymbol{H}\begin{pmatrix} \boldsymbol{L} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}' = \boldsymbol{H}_1\boldsymbol{L}\boldsymbol{H}_1', \quad \boldsymbol{L} = \mathrm{diag}\,(\ell_1, \ldots, \ell_n), \tag{A.3}$$

where $\ell_1 \ge \cdots \ge \ell_n$, and $\boldsymbol{H}_1$ is a $p \times n$ matrix satisfying $\boldsymbol{H}_1'\boldsymbol{H}_1 = \boldsymbol{I}_n$. Let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n)'$, and $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\phi_1(\boldsymbol{\ell}), \ldots, \phi_n(\boldsymbol{\ell}))$. In the case of $p > n$, Kubokawa and Srivastava [6] derived the Stein-Haff identity given by

$$E\left[\mathrm{tr}\,\{\boldsymbol{H}_1\boldsymbol{\Phi}(\boldsymbol{\ell})\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\}\right] = \sum_{i=1}^{n} E\Big[(p-n-1)\frac{\phi_i}{\ell_i} + 2\frac{\partial}{\partial \ell_i}\phi_i + 2\sum_{j>i}\frac{\phi_i - \phi_j}{\ell_i - \ell_j}\Big]. \tag{A.4}$$

Using (A.4), we can evaluate the moments of $\mathrm{tr}\,[\boldsymbol{L}^{-1}]$ and $\mathrm{tr}\,[\boldsymbol{L}^{-2}]$ from above.

Lemma A.1. *In the case of $p > n$, the following inequalities hold:*

$$E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]] \le \mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\frac{n}{p-n-1},$$

$$E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]] \le \{\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\}^2\frac{n(p-1)}{\{(p-n-1)(p-n-3)-2\}(p-n-1)}.$$

Proof. Putting $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \boldsymbol{I}$, $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \boldsymbol{L}^{-1}$ and $\boldsymbol{\Phi}(\boldsymbol{\ell}) = (\mathrm{tr}\,[\boldsymbol{L}^{-1}])\boldsymbol{I}$ in the identity (A.4), we get

$$E[\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] = (p-n-1)E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]], \tag{A.5}$$

$$E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] = (p-n-3)E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]] - E[(\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2], \tag{A.6}$$

$$E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] = (p-n-1)E[(\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2] - 2E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]], \tag{A.7}$$

respectively, where the second equality follows from the fact that

$$2\sum_{i=1}^{n}\sum_{j=i+1}^{p}(\ell_i\ell_j)^{-1} = (\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2.$$

The equality (A.5) yields the first inequality in Lemma A.1. Combining (A.6) and (A.7) gives the equality

$$E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]] = \frac{(p-n-1)E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] + E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]]}{(p-n-1)(p-n-3)-2},$$

which, together with (A.5), provides the second inequality in Lemma A.1. □

A.2. Evaluations of moments

Let $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}$, and $\boldsymbol{W}$ has $\mathcal{W}_p(n, \boldsymbol{I})$ for $n > p$. The following lemma provides exact moments of the inverted Wishart matrix $\boldsymbol{W}^{-1}$. For the proof, see Kubokawa, Hyodo and Srivastava [7]. Let $\alpha_2 = [m(m-1)(m-3)]^{-1}$, $\alpha_3 = \alpha_2[m(m-1)(m-3)]^{-1}$ and Let $\alpha_4 = \alpha_3[(m+2)(m-2)(m-7)]^{-1}$ for $m = n-p$.

Lemma A.2. *Assume that* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are any symmetric matrices. For* $m > 3$*,*

$$E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] = \alpha_2[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})],$$
$$E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})] = \alpha_2[2\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-2)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})].$$

For $m > 5$*,*

$$E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] = \alpha_3(n-1)[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})],$$
$$E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})] = \alpha_3(n-1)[4\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})].$$

For $m > 7$*,*

$$\begin{aligned} &E[\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] \\ &= \alpha_4(n-1)\Big\{\{(m-1)(n-2)-6\}[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})] \\ &\qquad + (2m+3p-2)[4\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]\Big\}, \\ &E[(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})] \\ &= \alpha_4(n-1)\Big\{2(2m+3p-2)[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})] \\ &\qquad + \{(m-4)(n-1)-6\}[4\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]\Big\}. \end{aligned}$$

Lemma A.3. *Let k_1, k_2, ℓ_1 and ℓ_2 be nonnegative integers satisfying $k_1\ell_1 + k_2\ell_2 = m$ for $m \le 4$. Assume that there exist limiting values of $\mathrm{tr}\,[\boldsymbol{A}^{k_1}]/p$ and $\mathrm{tr}\,[\boldsymbol{B}^{k_2}]/p$ for nonnegative definite matrices A and B. Also, assume that $\mathrm{tr}\,[\boldsymbol{W}^{-m}] < \infty$. Then, the moment*

$$M_{n,p} = E[\{\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{A})^{k_1}]/p\}^{\ell_1}\{\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{B})^{k_2}]/p\}^{\ell_2}]$$

is evaluated as $M_{n,p} = O(pn^{-m})$ for large n and p. In the special case of $p/n \to \gamma$ for $0 < \gamma < 1$, $M_{n,p}$ is of order $M_{n,p} = O(n^{-m})$.

Proof. It is noted that

$$\begin{aligned} M_{n,p} &\le n^{-m}E[\{\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1})\}^m]\{\mathrm{tr}\,[\boldsymbol{A}^{k_1}]/p\}^{\ell_1}\{\mathrm{tr}\,[\boldsymbol{B}^{k_2}]/p\}^{\ell_2} \\ &\le pn^{-m}E[\mathrm{tr}\,[(n\boldsymbol{W}^{-1})^m]]\{\mathrm{tr}\,[\boldsymbol{A}^{k_1}]/p\}^{\ell_1}\{\mathrm{tr}\,[\boldsymbol{B}^{k_2}]/p\}^{\ell_2} \\ &= O(pn^{-m}), \end{aligned} \tag{A.8}$$

from Lemma A.2. In the case that $p/n \to \gamma$ for $0 < \gamma < 1$, we can use the result of Bai and Yin [2], namely, $\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1}) = O_p(1)$. Thus, from (A.8), it can be seen that $M_{n,p} = O(n^{-m})$. □

Lemma A.4. *Assume that $m > 5$. Then,*

$$\begin{aligned} E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}])^3] &= \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^3}{m^3} + O(p^3n^{-4}), \\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]] &= \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]}{m^3} + \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^3}{m^4} + O(p^2n^{-4}), \\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-2}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]] &= \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}])^2}{m^3} + \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{m^4} \\ &\quad + O(p^2n^{-4}), \\ E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^3]] &= \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}]}{m^3} + O(p^3n^{-4}). \end{aligned}$$

Proof. Let $\boldsymbol{D}$ be a $p \times p$ diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}^{-1}$. Letting $\boldsymbol{G} = \boldsymbol{D}(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2$, $\boldsymbol{G} = \boldsymbol{D}\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$, $\boldsymbol{G} = \boldsymbol{D}^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$ and $\boldsymbol{G} = \boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^2$ in (A.2), we have

$$\begin{aligned} \mathrm{tr}\,[\boldsymbol{D}]E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] &= (m-1)E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] \\ &\quad - 4E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]], \\ \mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= (m-1)E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\ &\quad - 4E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]], \\ \mathrm{tr}\,[\boldsymbol{D}^2]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= (m-1)E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\ &\quad - 4E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^3]], \\ E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= (m-3)E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]] \\ &\quad - 2E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]], \end{aligned}$$

respectively. These can be rewritten as

$$\begin{aligned}
E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] &= \frac{1}{m-1}\{\mathrm{tr}\,[\boldsymbol{D}]E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] \\
&\quad + 4E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]]\}, \\
E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= \frac{1}{m-1}\{\mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\quad + 4E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]]\}, \\
E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= \frac{1}{m-1}\{\mathrm{tr}\,[\boldsymbol{D}^2]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\quad + 4E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^3]]\}, \\
E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]] &= \frac{1}{m-3}\{E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\quad + 2E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]]\}.
\end{aligned}$$

Further, from Lemmas A.2 and A.3, these third-order terms can be evaluated as

$$\begin{aligned}
E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] &= \frac{(\mathrm{tr}\,[\boldsymbol{D}])^3}{m^3} + O(p^3n^{-4}), \\
E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= \frac{\mathrm{tr}\,[\boldsymbol{D}]\mathrm{tr}\,[\boldsymbol{D}^2]}{m^3} + \frac{(\mathrm{tr}\,[\boldsymbol{D}])^3}{m^4} + O(p^2n^{-4}), \\
E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= \frac{(\mathrm{tr}\,[\boldsymbol{D}^2])^2}{m^3} + \frac{\mathrm{tr}\,[\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{D}])^2}{m^4} + O(p^2n^{-4}), \\
E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]] &= \frac{\mathrm{tr}\,[\boldsymbol{D}^3]}{m^3} + O(p^3n^{-4}),
\end{aligned}$$

which yields the evaluations in Lemma A.4. □

Lemma A.5. *Assume that $m > 7$. Then,*

$$\begin{aligned}
\frac{m^2}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}])^4] &= \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^4}{p^2m^2} + O(p^2n^{-3}) + O(pn^{-2}), \\
\frac{m^3}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}])^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]] &= \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{p^2m} \\
&\quad + \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^4}{p^2m^2} + O(pn^{-2}), \\
\frac{m^4}{p^2}E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2])^2] &= \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}])^2}{p^2} + 2\frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{p^2m} \\
&\quad + \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^4}{p^2m^2} + O(n^{-1}).
\end{aligned}$$

Proof. It is hard to obtain exact expressions of the requested expectations in Lemma A.5. Instead of that, we derive the leading terms and orders of the remainder terms using the same arguments as in the proof of Lemma A.4. Letting

$\boldsymbol{G} = \boldsymbol{D}(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3$, $\boldsymbol{G} = \boldsymbol{D}\boldsymbol{W}^{-1}\boldsymbol{D}\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$ and

$$\boldsymbol{G} = \boldsymbol{D}\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$$

in (A.2) gives, respectively,

$$\begin{aligned}
&E[\mathrm{tr}\,[\boldsymbol{D}](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] \\
&\quad = (m-1)E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^4] - 6E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2], \\
&E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] \\
&\quad = (m-2)E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] \\
&\qquad - E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2]] - 4E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^4]], \\
&E[\mathrm{tr}\,[\boldsymbol{D}]\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\quad = (m-1)E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2])^2] - 2E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2]] \\
&\qquad - 4E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]].
\end{aligned}$$

Then, from Lemma A.3, the fourth-order moments can be evaluated as

$$\begin{aligned}
&\frac{m^2}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^4] = \frac{m}{p^2}\mathrm{tr}\,[\boldsymbol{D}]E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] + O(p^2n^{-3}), \\
&\frac{m^3}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\quad = \frac{m^2}{p^2}\mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] + Op(n^{-2}), \\
&\frac{m^4}{p^2}E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2])^2] \\
&\quad = \frac{m^3}{p^2}E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\qquad + \frac{m^2}{p^2}\mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] + O(n^{-1}).
\end{aligned}$$

Hence, from Lemma A.4, we have

$$\begin{aligned}
&\frac{m^2}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^4] = \frac{(\mathrm{tr}\,[\boldsymbol{D}])^4}{p^2m^2} + O(p^2n^{-3}) + O(pn^{-2}), \\
&\frac{m^3}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] = \frac{\mathrm{tr}\,[\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{D}])^2}{p^2m} + \frac{(\mathrm{tr}\,[\boldsymbol{D}])^4}{p^2m^2} + O(pn^{-2}), \\
&\frac{m^4}{p^2}E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2])^2] = \frac{(\mathrm{tr}\,[\boldsymbol{D}^2])^2}{p^2} + 2\frac{\mathrm{tr}\,[\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{D}])^2}{p^2m} + \frac{(\mathrm{tr}\,[\boldsymbol{D}])^4}{p^2m^2} + O(n^{-1}),
\end{aligned}$$

which yields the results in Lemma A.5. □

A.3. Asymptotic properties of $\hat{b}_i$

Proposition A.1. *$E[\hat{b}_1 - b_1] = O(n^{-1})$, $E[\hat{b}_2 - b_2] = O(n^{-1})$, $Var[\hat{b}_1] = O((np)^{-1})$ and $Var[\hat{b}_2] = O(n^{-1})$ for large n and p satisfying $n > p + 3$.*

Proof. It follows from (A.10) and Lemma A.2 that

$$E[\hat{b}_1] = p^{-1}E[m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - m\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]] = b_1 + O(n^{-1}).$$

For $Var[\hat{b}_1]$, it is written as

$$\begin{aligned}Var[\hat{b}_1] &= p^{-2}E[\{m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}] - m\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2] + O(n^{-2})\\ &= p^{-2}E[\{m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\}^2]\\ &\quad - 2p^{-2}mE[\{m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\}\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]]\\ &\quad + p^{-2}m^2E[\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2] + O(n^{-2})\\ &= J_1 - 2J_2 + J_3 + O(n^{-2}).\end{aligned}$$

It can be seen from Lemma A.2 that

$$\begin{aligned}J_1 &= p^{-2}E\Big[m^2(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - 2m\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\mathrm{tr}\,[\boldsymbol{V}^{-1}] + (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2\Big]\\ &= p^{-2}\Big\{m^2\alpha_2\big[2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] + (m-2)(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2\big]\\ &\quad - 2\frac{m}{m-1}(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2 + (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2\Big\}\\ &= 2\frac{m}{p^2(m-1)(m-3)}\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] + \frac{3}{p^2(m-1)(m-3)}(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2,\end{aligned}$$

which is of order $O((np)^{-1})$. For J_3, it is noted that

$$\frac{m^2}{p^2}\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2 \le \frac{m^2}{p^2}\hat{a}_1^2\{\mathrm{tr}\,[\boldsymbol{V}^{-2}]\}^2 = \frac{m^2}{p^2}\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]\}^2,$$

which is of order $O_p(n^{-2})$ as seen from Lemma A.4. Since $J_2 = O((pn^3)^{-1/2})$, it is seen that $Var(\hat{b}_1) = O((np)^{-1})$, which implies that $\hat{b}_1 - b_1 = O_p((np)^{-1/2})$.

For $\hat{b}_2$, it is noted that

$$\begin{aligned}\hat{b}_2 &= p^{-1}\Big\{m^2\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-2}] - m\{\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}]\}^2\Big\}\\ &= p^{-1}\Big\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m^2\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-2}] - m^2\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-2}\boldsymbol{V}^{-1}]\\ &\quad - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 + 2m\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\mathrm{tr}\,[\boldsymbol{V}^{-1}]\\ &\quad - m\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2\Big\}.\end{aligned}$$

It here follows from Lemma A.4 that $\mathrm{tr}\,[(\boldsymbol{V}+\hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-2}] \le \mathrm{tr}\,[\boldsymbol{V}^{-3}] = O_p(n^{-1})$. The same arguments can be used to approximate $\hat{b}_2$ as

$$\hat{b}_2 = \frac{m^2}{p}\mathrm{tr}\,[\boldsymbol{V}^{-2}] - \frac{m}{p}(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 + O_p(n^{-1}).$$

Using Lemma A.2, we can verify that $E[\hat{b}_2 - b_2] = O(n^{-1})$. Also,

$$Var(\hat{b}_2) = \frac{1}{p^2}E[\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\}^2]$$

$$+ 2E[\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\} \times O_p(n^{-1})] + O(n^{-2}).$$

Here, using Lemma A.5, we can show that

$$\begin{aligned}
&\frac{1}{p^2}E[\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\}^2] \\
&\quad = \frac{1}{p^2}E\Big[m^4(\mathrm{tr}\,[\boldsymbol{V}^{-2}])^2 + m^2(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^4 + (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2})^2 \\
&\quad - 2m^3\mathrm{tr}\,[\boldsymbol{V}^{-2}](\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - 2m^2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\mathrm{tr}\,[\boldsymbol{V}^{-2}] + 2m\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2\Big],
\end{aligned}$$

which is of order $O(n^{-1})$. Therefore, the proof of Proposition A.1 is complete. □

A.4. Proofs of Theorems 3.1 and 3.2

Proof of Theorem 3.1. For evaluating the risk $R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda^{-1})$, it is noted that

$$\begin{aligned}
(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1} &= (\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1} + \{(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1} - (\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\} \\
&= (\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1} - d(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1},
\end{aligned}$$

so that the risk of the estimator $\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1} = c_2(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}$ is written as

$$\begin{aligned}
&R^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda^{-1}) \\
&= E\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}^2]\big] \\
&= E\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I} + c_2 d(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}\}^2]\big] \\
&= E\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}^2]\big] \\
&\quad + c_2^2 d^2 E\big[\mathrm{tr}\,[\{(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}\}^2]\big] \\
&\quad - 2c_2 dE\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}]\big] \\
&= I_1 + I_2 + I_3. \qquad (A.9)
\end{aligned}$$

We shall evaluate each term in (A.9). We begin by estimating I_1. Since

$$(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1} = \boldsymbol{V}^{-1} - (\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\boldsymbol{V}^{-1}, \qquad (A.10)$$

the term $c_2(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}$ is rewritten as

$$\begin{aligned}
&c_2(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I} \\
&= c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma} - c_2(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I} \\
&= (c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}) - (\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}) - (\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda},
\end{aligned}$$

so that the first term I_1 is expressed as

$$I_1 = E\Big[\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2] + \mathrm{tr}\,[\{(\boldsymbol{V} + d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})\}^2]$$

$$
\begin{aligned}
&+ \mathrm{tr}\,[\{(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\}^2] - 2\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})^2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}]\\
&- 2\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}]\\
&+ 2\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\{(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\}^2]\Big]\\
={}& R^*(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}_0^{-1}) + I_{11} + I_{12} - 2I_{13} - 2I_{14} + 2I_{15}. \qquad \text{(A.11)}
\end{aligned}
$$

We shall evaluate each term in (A.11). Using Lemma A.2, we can evaluate I_{11} as

$$
\begin{aligned}
I_{11} &\le d^2E\big[\mathrm{tr}\,[\{(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{V}^{-1}\boldsymbol{\Lambda}\}^2] = d^2E\big[\mathrm{tr}\,[\{(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-1}\boldsymbol{\Gamma}\}^2]\\
&= d^2E\big[\mathrm{tr}\,[c_2^2\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-2}\boldsymbol{\Gamma} - 2c_2\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma} + \boldsymbol{W}^{-1}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big]\\
&= O(d^2n^{-3+2\delta}).
\end{aligned}
$$

Similarly, from (A.10), Lemma A.4 and condition (A4),

$$
\begin{aligned}
I_{12} &= d^2E\big[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2] - 2d\mathrm{tr}\,[(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(d\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2]\\
&\quad + d^2\mathrm{tr}\,[\{(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}d\boldsymbol{W}^{-1}\boldsymbol{\Gamma}\}^2]\big]\\
&= \frac{d^2}{n^2}\mathrm{tr}\,[\boldsymbol{\Gamma}^2] + O(d^3n^{-3+\delta}),
\end{aligned}
$$

since $d^2\mathrm{tr}\,[\{(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}d\boldsymbol{W}^{-1}\boldsymbol{\Gamma}\}^2 \le d^3\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^3]$ and $\mathrm{tr}\,[\boldsymbol{\Gamma}^3]/p = O(1)$. For I_{13}, from (A.10),

$$
\begin{aligned}
I_{13} &= dE\big[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2(\boldsymbol{W}^{-1} - (\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}d\boldsymbol{\Gamma}\boldsymbol{W}^{-1})\boldsymbol{\Gamma}]\big]\\
&= dE\big[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big]\\
&\quad - d^2E\big[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big]\\
&= I_{131} - I_{132}.
\end{aligned}
$$

It can be seen that $I_{131} = pdn^{-2}\mathrm{tr}\,[\boldsymbol{\Gamma}] + O(dn^{-3+2\delta})$. Also, it is observed that

$$
\begin{aligned}
I_{132} &\le d^2E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2]]\\
&= d^2c_2E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]] - E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2]]\\
&\le d^2c_2E[\mathrm{tr}\,[\{(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-1}\}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^4]] + O(d^2n^{-3+2\delta}),
\end{aligned}
$$

where the first term in the last equality is estimated as $O(d^2n^{-5/2+3\delta/2})$. Thus,

$$
I_{13} = \frac{pd}{n^2}\mathrm{tr}\,[\boldsymbol{\Gamma}] + O(dn^{-3+2\delta}) + O(d^2n^{-5/2+3\delta/2}).
$$

The term I_{14} is evaluated as

$$
\begin{aligned}
I_{14} &= dE[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})(\boldsymbol{W}^{-1} - (\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}d\boldsymbol{\Gamma}\boldsymbol{W}^{-1})\boldsymbol{\Gamma}]]\\
&= dE[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]]\\
&\quad - d^2E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\boldsymbol{W}^{-1})\boldsymbol{\Gamma}]]\\
&= I_{141} + I_{142}.
\end{aligned}
$$

It can be seen that $I_{141}=0$ and $I_{142}=O(d^2n^{-5/2+3\delta/2})$. Thus,

$$I_{14}=O(d^2n^{-5/2+3\delta/2}).$$

For I_{15}, since $\mathrm{tr}\,[\{\boldsymbol{\Gamma}(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\}^2]=O_p(n^{-1+2\delta})$, it is noted that

$$\begin{aligned}&\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\{(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\}^2]\\&\le d^2\big[\mathrm{tr}\,[\{\boldsymbol{\Gamma}(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-2}\boldsymbol{\Gamma})^2]\big]^{1/2}\\&=O_p(d^2n^{-5/2+3\delta/2}).\end{aligned}$$

Combining these evaluations gives that

$$\begin{aligned}I_1=&R^*(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}_0^{-1})+\frac{pd}{n^2}\Big\{\frac{d}{p}\mathrm{tr}\,[\boldsymbol{\Gamma}^2]-2\mathrm{tr}\,[\boldsymbol{\Gamma}]\Big\}\\&+O(d^3n^{-3+\delta})+O(d^2n^{-5/2+3\delta/2})+O(dn^{-3+2\delta}).\end{aligned}\tag{A.12}$$

Concerning I_2 in (A.9), it is estimated as

$$\begin{aligned}&c_2^2d^2\mathrm{tr}\,[\{(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}\}^2\\&\quad\le c_2^2d^2\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1}]\\&\quad=c_2^2d^2\mathrm{tr}\,[(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{W}^{-2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{W}^{-2}]\\&\quad=\frac{d^2}{(m-1)^2}\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I}+\boldsymbol{I})(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2].\end{aligned}\tag{A.13}$$

for $\beta=m(m-1)(m-3)/(n-1)$. Since $E[\boldsymbol{W}^{-2}]=(1/\beta)\boldsymbol{I}$, it is noted that $E[\beta\boldsymbol{W}^{-2}]=\boldsymbol{I}$. Thus, I_2 is evaluated from above as

$$I_2\le 2\frac{d^2}{(m-1)^2}E\big[\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2]\big]+2\frac{d^2}{(m-1)^2}E\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})^2]\big].\tag{A.14}$$

Using Lemma A.2 and condition (A3), we can demonstrate that $\mathrm{tr}\,[(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})^2]=O_p(p^2/n)$, so that the first term in (A.14) is evaluated as

$$\begin{aligned}&d^2n^{-2}E[\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2]]\\&\qquad\le d^2n^{-2}E[\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})\}^2]\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2]],\end{aligned}$$

which is of order $O(d^2n^{-2-1+2\delta-1+\delta})$, or $O(d^2n^{-4+3\delta})$. Since the third term is of order $O(d^2n^{-3+\delta})$, it is observed that $I_2=O(d^2n^{-4+3\delta})+O(d^2n^{-3+\delta})$.

Concerning I_3 in (A.9), since $I_1=O(n^{-1+2\delta})$, we have

$$\begin{aligned}&c_2d\mathrm{tr}\,[\{c_2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}\}(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}]\\&\le\Big[O_p(n^{-1+2\delta})\times\{O_p(d^2n^{-4+3\delta})+O_p(d^2n^{-3+\delta})\}\Big]^{1/2}\\&=O_p(dn^{-5/2+5\delta/2})+O_p(dn^{-2+3\delta/2}),\end{aligned}\tag{A.15}$$

so that

$$I_2 + I_3 = O(d^2 n^{-4+3\delta}) + O(d^2 n^{-3+\delta}) + O_p(dn^{-5/2+5\delta/2}) + O_p(dn^{-2+3\delta/2}). \tag{A.16}$$

Combining (A.12) and (A.16), we get Theorem 3.1. □

Proof of Theorem 3.2. We need to check each step in the proof of Theorem 3.1. It follows from (A.9) that $R^*(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}_\Lambda^{-1}) = I_1 + I_2 + I_3$. Under the condition that $p/n \to \gamma$ for $0 < \gamma < 1$, we can use the result of Bai and Yin [2], namely, $\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1}) = O_p(1)$. Then,

$$\begin{aligned} I_2 &\le \frac{c_2^2 d^2}{n^4} E[\mathrm{tr}\,[(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})(\boldsymbol{W}/n)^{-2}(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})(\boldsymbol{W}/n)^{-2}]] \\ &\le \frac{c_2^2 d^2}{n^2} E[\{\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1})\}^{-4} \mathrm{tr}\,[(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\}^2]], \end{aligned}$$

which means that $I_2 = O(d^2/n^2)$. Since $I_1 = O(n^{-1+2\delta})$, it follows from (A.15) that $I_3 \le \{O(d^2/n^2) \times O(n)\}^{1/2} = O(d/\sqrt{n})$, so that

$$I_2 + I_3 = O(d^2 n^{-2}) + O_p(dn^{-1/2}). \tag{A.17}$$

Thus, we get the approximation given in (3.4). When $d = o(p)$, we get (3.5) by combining (A.12) and (A.17). □

References

[1] Bai, J. and Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. *Ann. Econ. Finance*, **12**, 199–215.

[2] Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Statist.*, **21**, 1275–1294. MR1235416

[3] Fisher, T.J. and Sun, X. (2011). Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Comp. Statist. Data Analysis*, **55**, 1909–1918. MR2765053

[4] Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531–542. MR0556910

[5] Konno, Y. (2009). Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss. *J. Multivariate Analysis*, **100**, 2237–2253. MR2560366

[6] Kubokawa, T. and Srivastava, M.S. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *J. Multivariate Analysis*, **99**, 1906–1928. MR2466543

[7] Kubokawa, T., Hyodo, M., and Srivastava, M.S. (2013). Asymptotic expansion and estimation of EPMC for linear classification rules in high dimension. *J. Multivariate Analysis*, **115**, 496–515. MR3004572

[8] Ledoit, O., and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance*, **10**, 603–621.

[9] Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Analysis*, **88**, 365–411. MR2026339

[10] Srivastava, M.S. (2003). Singular Wishart and multivariate beta distributions. *Ann. Statist.*, **31**, 1537–1560. MR2012825

[11] Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high-dimensional data. *J. Japan Statist. Soc.*, **35**, 251–272. MR2328427

[12] Stein, C. (1977). Lectures on multivariate estimation theory. (In Russian.) In *Investigation on Statistical Estimation Theory* I, 4–65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad. MR0501548