CIRJE-F-995

Linear Ridge Estimator of High-Dimensional Precision Matrix Using Random Matrix Theory

Tsubasa Ito
Graduate School of Economics, The University of Tokyo

Tatsuya Kubokawa
The University of Tokyo

November 2015

CIRJE Discussion Papers can be downloaded without charge from:

http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

Linear Ridge Estimator of High-Dimensional Precision Matrix Using Random Matrix Theory

Tsubasa Ito*and Tatsuya Kubokawa†

November 4, 2015

Abstract

In estimation of the large precision matrix, this paper suggests a new shrinkage estimator, called the linear ridge estimator. This estimator is motivated from a Bayesian aspect for a spike and slab prior distribution of the precision matrix, and has a form of convex combination of the ridge estimator and the identity matrix multiplied by scalar. The optimal parameters in the linear ridge estimator are derived in terms of minimizing a Frobenius loss function and estimated in closed forms based on the random matrix theory. Finally, the performance of the linear ridge estimator is numerically investigated and compared with some existing estimators.

Key words and phrases: Large-dimensional asymptotics, nonlinear shrinkage, precision matrix, random matrix theory, ridge type estimator, rotation-equivariant estimators.

1 Introduction

The estimation of a large covariance matrix has been actively studied in the literature in recent years. Analysis of high-dimensional data is an important and modern topic in genetics in bioscience, portfolio in financial economics and others. However, the sample covariance matrix is not invertible when the dimension p of the variables is larger than the sample size N. When p is large and close to N, the inverse of the sample covariance matrix may be ill-conditioned even if $N > p$. To address this problem, many approaches have been considered in the literature. Under some model structures such as sparsity or ordering, penalized methods have been proposed and the population covariance matrix can be estimated consistently. However, the true model structures are generally unknown and the estimates become inconsistent in the case that the model structure is misspecified. In the absence of such a prior information on the structure of the covariance matrix, the shrinkage method is a useful approach. Since the eigenvalues of the sample covariance matrix diverge more widely than the eigenvalues of the population covariance matrix, it is reasonable to shrink the sample eigenvalues in the direction of their center. Linear shrinkage estimators of the covariance matrix have been suggested in Ledoit and Wolf (2004), Srivastava and Kubokawa (2007), Touloumis (2015) and others. A nonlinear shrinkage approach

*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN. E-Mail: tsubasa_ito.0710@gmail.com

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN. E-Mail: tatsuya@e.u-tokyo.ac.jp

based on the random matrix theory is one of recent hot issues in statistics, and Ledoit and Wolf (2012, 15) developed some fundamental results and more efficient nonlinear shrinkage estimators. For the related works, see also Wang, Pan, Tong and Zhu (2015), Bodnar, Gupta and Parolya (2015) and the references therein.

Although the estimation of the covariance matrix has received more attention than that of the precision matrix, the estimation of the precision is more important in multivariate analysis, since for example, the precision matrix appears in the Fisher linear discriminant analysis, confidence intervals based on the Mahalanobis distance and generalized least squares estimators in multivariate linear regression models. As an approach, one can use the inverse matrix of estimators of the covariance matrix. However, the inverse of the covariance matrix estimator is not necessarily optimal in the estimation of the precision matrix. Further, two estimation problems of the covariance and precision matrices are drastically different when $p > N$. Thus, in this paper, we address the problem of estimating the precision matrix directly.

To explain the problem more specifically, let $\boldsymbol{\Sigma}_p$ be a $p \times p$ covariance matrix, and let $\boldsymbol{S}_p$ be the sample covariance matrix such that $E[\boldsymbol{S}_p] = \boldsymbol{\Sigma}_p$. Using the random matrix theory, Ledoit and Wolf (2012, 15) derived asymptotic optimal shrinkage coefficients in the general class of rotation-equivariant estimators, and proposed the nonlinear shrinkage estimator for the precision matrix. The nonlinear shrinkage estimator performs well for large N, but the performance dwindles when N or N/p is small, because the consistent estimators of the eigenvalues of $\boldsymbol{\Sigma}_p$ are not stable for small N or N/p. Bodnar, *et al.* (2015) suggested the linear shrinkage estimator

$$\boldsymbol{\Omega}_p^{linear} = \begin{cases} \alpha \boldsymbol{S}_p^{-1} + \beta \boldsymbol{I}_p & \text{if } N > p \\ \alpha \boldsymbol{S}_p^{+} + \beta \boldsymbol{I}_p & \text{if } N < p. \end{cases}$$

This estimator has an appealing form of shrinking $\boldsymbol{S}_p^{-1}$ or $\boldsymbol{S}_p^{+}$ towards $\beta \boldsymbol{I}_p$. However, a drawback is that under their settings the parameters α and β cannot be estimated in the case of $p > N$, without assuming that $\boldsymbol{\Sigma}_p = \sigma^2 \boldsymbol{I}_p$ for scalar σ^2. Wang, *et al.* (2015) considered the ridge-type estimator

$$\boldsymbol{\Omega}_p^{ridge} = \alpha(\boldsymbol{S}_p + \gamma \boldsymbol{I}_p)^{-1},$$

and provided estimators of α and γ using the random matrix theory. Although this estimator is stable for any p and N, it shrinks all the sample eigenvalues toward the same direction, which is not necessarily desirable in the estimation of the precision matrix.

In this paper, we suggest a new type of estimator given by

$$\boldsymbol{\Omega}_p^{LR} = \alpha(\boldsymbol{S}_p + \gamma \boldsymbol{I}_p)^{-1} + \beta \boldsymbol{I}_p,$$

and we call it the linear ridge estimator. This estimator is stable for any p and N, and re-shrinks the shrunk sample eigenvalues toward the constant β. By combining the linear and the ridge estimators, the linear ridge estimator can compensate their weakness. Using the random matrix theory, we can derive the limits of the optimal α and β for fixed γ. Based on the limits, we can provide the estimators of the optimal α and β in closed forms. It is also interesting to point out that the linear ridge estimator can be motivated from a Bayesian aspect for a spike and slab prior distribution of $\boldsymbol{\Sigma}_p^{-1}$. This motivation is useful when we need to restrict the estimators of α, β and γ.

The paper is organized as follows: The basic framework of the random matrix theory is described in Section 2. The loss function based on the Frobenius norm and the class of rotation-equivariant estimators are given. In Section 3, we provided the Ledoit-Wolf type estimator and the suggested linear ridge estimator. The optimal estimators of the parameters in these estimators are derived based on the random matrix theory. The Bayesian motivation of the linear ridge estimator is also given. In Section 4, we investigate the finite-sample performance of the proposed estimator through simulation and empirical studies. Concluding remarks are given in Section 5 and the technical proofs are given in the Appendix.

2 Basic Framework

2.1 Preliminary results in the random matrix theory

We begin by stating the basic assumptions which are common in estimation of the high-dimensional covariance matrix based on the random matrix theory. Throughout the paper, $\mathbb{R}$ and $\mathbb{C}$ denote the spaces of real and complex numbers, respectively. Also, $\mathbb{C}^+$ denotes the half-plane of complex numbers with strictly positive imaginary part. The real and imaginary parts of $z \in \mathbb{C}$ are denoted by $\Re(z)$ and $\Im(z)$, respectively.

(A1) Let N and $p \equiv p(N)$ denote the sample size and the number of variables respectively. In the large-dimensional asymptotics, p goes to infinity, as N does, but it is assumed that the ratio p/N converges to a limit $y \in (0,1) \cup (1,+\infty)$, as $N \to +\infty$. The case $y = 1$ is excluded for technical reason.

(A2) The population covariance matrix $\boldsymbol{\Sigma}_p$ is a non-random p-dimensional positive definite matrix, where the subscript p denotes the dimension of the matrix. Let $\boldsymbol{X}_p = (\boldsymbol{x}_{p,1}, \ldots, \boldsymbol{x}_{p,N})^{\mathrm{T}}$ be an $N \times p$ random matrix, where $\boldsymbol{x}_{p,1}, \ldots, \boldsymbol{x}_{p,N}$ are mutually independently and identically distributed as $E[\boldsymbol{x}_{p,j}] = \boldsymbol{0}$ and $\mathbf{Cov}\,(\boldsymbol{x}_{p,j}) = \boldsymbol{I}_p$. It is assumed that there exist fourth moments of all elements of $\boldsymbol{x}_{p,j}$. Let $\boldsymbol{Y}_p = (\boldsymbol{y}_{p,1}, \ldots, \boldsymbol{y}_{p,N})^{\mathrm{T}}$, where $\boldsymbol{y}_{p,j} = \boldsymbol{\Sigma}_p^{1/2} \boldsymbol{x}_{p,j}$ for the squared symmetric root matrix $\boldsymbol{\Sigma}^{1/2}$ such that $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}^{1/2})^2$. It is assumed that $\boldsymbol{Y}_p$ is observable.

(A3) Let $\boldsymbol{t}_p = (t_{p,1}, \cdots, t_{p,p})^{\mathrm{T}}$ be a system of eigenvalues of $\boldsymbol{\Sigma}_p$, sorted in decreasing order. The empirical spectral distribution (ESD) of the population eigenvalues is defined by

$$H_p(t) \equiv \frac{1}{p} \sum_{i=1}^{p} \mathbb{I}_{[t_{p,i},+\infty)}, \quad \forall t \in \mathbb{R},$$

where $\mathbb{I}_A$ denotes the indicator function of set A. It is assumed that $H_p(t)$ converges to some limit $H(t)$ at all points of continuity of H.

(A4) Supp(H), the support of H, is the union of a finite number of closed intervals, bounded away from zero and infinity. Furthermore, there exists a compact interval in $(0,+\infty)$ that contains Supp(H_p) for all N large enough.

Let $\boldsymbol{S}_p = N^{-1} \boldsymbol{Y}_p^{\mathrm{T}} \boldsymbol{Y}_p$. As noted in Section 5, all the results given in this paper still hold in the case of $E[\boldsymbol{y}_{p,j}] = \boldsymbol{\mu}$. A system of eigenvalues of $\boldsymbol{S}_p$, sorted in decreasing order, and the corresponding eigenvectors are denoted by $\boldsymbol{\ell}_p = (\ell_{p,1}, \cdots, \ell_{p,p})^{\mathrm{T}}$ and $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$, respectively. The empirical spectral distribution (ESD) of $\boldsymbol{S}_p$ is defined by

$$F_p(t) \equiv \frac{1}{p}\sum_{i=1}^{p} \mathbb{I}_{[\ell_{p,i},+\infty)}, \quad \forall t \in \mathbb{R}.$$

Marčenko and Pastur (1967) greatly contributed the research on large-dimensional asymptotic theories of eigenvalues of $\boldsymbol{S}_p$. This asymptotic theory is called the random matrix theory, and has been generalized by Silverstein (1995), Silverstein and Bai (1995), Silverstein and Choi (1995) among others. These literatures imply that under assumptions (A1)-(A4), there exists a distribution function F such that

$$F_p(x) \to F(x), \ \ \forall x \in \mathbb{R}\backslash\{0\},$$

which is called the limiting spectral distribution (LSD). Silverstein and Choi (1995) showed that F is everywhere continuous except at zero, and that the mass of F at zero is given by

$$F(0) = \max\{1 - y^{-1}, H(0)\}. \tag{2.1}$$

For a nondecreasing function G on the real line, the stieltjes transform m_G of G is defined by

$$m_G(z) \equiv \int \frac{1}{x - z} dG(x), \ \ \forall z \in \mathbb{C}^{+}.$$

The stieltjes transform has the well-known inversion formula

$$G\{[a, b]\} = \frac{1}{\pi}\lim_{\eta\to 0^{+}} \int_a^b \Im\left(m_G(\xi + i\eta)\right) d\xi, \tag{2.2}$$

if G is continuous at a and b. Silverstein (1995) developed the most general version of the equation which relates F to H and y, namely, $m \equiv m_F(z)$ is the unique solution in the set $\{m \in \mathbb{C} : -(1-y)/z + ym \in \mathbb{C}^{+}\}$ to the equation

$$m_F(z) = \int \frac{1}{t(1 - y - yzm_F(z)) - z} dH(t), \quad \forall z \in \mathbb{C}^{+}. \tag{2.3}$$

Silverstein and Choi (1995) showed that m_F is extended to $\mathbb{R}$, that is, $\lim_{z \in \mathbb{C}^{+} \to \lambda} m_F(z) \equiv \check{m}_F(\lambda)$ for $\lambda \in \mathbb{R}$. It is noted from the inversion formula that $F'(\lambda) = \pi^{-1}\Im[\check{m}_F(\lambda)]$ on $\mathbb{R}$.

We next consider the limiting spectral distribution (LSD) of the eigenvalues of $N^{-1}\boldsymbol{Y}_p\boldsymbol{Y}_p^{\mathrm{T}} = N^{-1}\boldsymbol{X}_p\boldsymbol{\Sigma}_p\boldsymbol{X}_p^{\mathrm{T}}$. Denote the LSD by $\underline{F}$. The eigenvalues of $N^{-1}\boldsymbol{Y}_p^{\mathrm{T}}\boldsymbol{Y}_p$ and $N^{-1}\boldsymbol{Y}_p\boldsymbol{Y}_p^{\mathrm{T}}$ only differs by $|N - p|$ zero eigenvalues. It then holds:

$$\begin{aligned}
\underline{F}(x) =& (1 - y)\mathbb{I}_{[0,+\infty)}(x) + yF(x), \quad \forall x \in \mathbb{R}, \\
F(x) =& \frac{y-1}{y}\mathbb{I}_{[0,+\infty)}(x) + \frac{1}{y}\underline{F}, \quad \forall x \in \mathbb{R}, \\
m_{\underline{F}}(z) =& \frac{y-1}{z} + ym_F(z), \quad \forall z \in \mathbb{C}^{+}, \\
m_F(z) =& \frac{1-y}{yz} + \frac{1}{y}m_{\underline{F}}(z), \quad \forall z \in \mathbb{C}^{+}.
\end{aligned}$$

2.2 Setup of the problem

Consider the problem of estimating $\boldsymbol{\Sigma}_p^{-1}$ by estimator $\boldsymbol{\Omega}_p$, where the estimator is evaluated by the risk function relative to the loss function

$$L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) \equiv \frac{1}{p}\mathrm{tr}\,(\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \boldsymbol{I}_p)(\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \boldsymbol{I}_p)^{\mathrm{T}}. \tag{2.4}$$

This loss function is based on the Frobenius norm $\|\boldsymbol{A}\|_F = \sqrt{\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})}$, namely, $L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) = p^{-1}\|\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \boldsymbol{I}_p\|_F^2$. We thus call it the Frobenius loss function. The loss function satisfies that $L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) \geq 0$ and $L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Sigma}_p^{-1}) = 0$. For remarks of other loss functions, see Section 5.

Ledoit and Wolf (2012, 15) considered a class of rotation-equivariant estimators in estimation of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$. An estimator $\widehat{\boldsymbol{\Sigma}}_p = \widehat{\boldsymbol{\Sigma}}_p(\boldsymbol{Y}_p)$ of $\boldsymbol{\Sigma}_p$ is called rotation-equivariant when for any p-dimensional orthogonal matrix $\boldsymbol{W}$, $\widehat{\boldsymbol{\Sigma}}_p(\boldsymbol{Y}_p\boldsymbol{W}) = \boldsymbol{W}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_p(\boldsymbol{Y}_p)\boldsymbol{W}$. For the sample covariance matrix, it is well known that eigenvalues of $\boldsymbol{S}_p$ diverge more widely than those of $\boldsymbol{\Sigma}_p$. Rotation-equivariant estimators have the same eigenvectors as $\boldsymbol{S}_p$ and replace eigenvalues of $\boldsymbol{S}_p$ with other values shrunk towards a center of the eigenvalues. In the estimation of the inverse $\boldsymbol{\Sigma}_p^{-1}$, rotation-equivariant estimators $\boldsymbol{\Omega}_p = \boldsymbol{\Omega}_p(\boldsymbol{Y}_p)$ satisfy $\boldsymbol{\Omega}_p(\boldsymbol{Y}_p\boldsymbol{W}) = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Omega}_p(\boldsymbol{Y}_p)\boldsymbol{W}$ for any orthogonal matrix $\boldsymbol{W}$. Thus, every rotation-equivariant estimator of $\boldsymbol{\Sigma}_p^{-1}$ is of the form

$$\boldsymbol{\Omega}_p(\boldsymbol{A}_p) = \boldsymbol{U}_p\boldsymbol{A}_p\boldsymbol{U}_p^{\mathrm{T}}, \quad \boldsymbol{A}_p = \mathrm{diag}(a_1, \cdots, a_p), \tag{2.5}$$

where $\boldsymbol{U}_p = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$, and a_i may depend on the eigenvalues $\boldsymbol{\ell}_p$ of $\boldsymbol{S}_p$.

The class of rotation-equivariant estimators include various estimators. The ridge estimator given by Wang, *et al.* (2014) corresponds to the case that $a_i = \alpha/(\ell_i + \gamma)$, and the linear estimator treated by Bodnar, *et al.* (2014) corresponds to $a_i = \alpha/\ell_i + \beta$ for $N > p$ and $a_i = \beta$ for $N < p$. Ledoit and Wolf (2015) treated the general rotation-equivariant estimator directly. The problem is how to estimate the parameters a_i's, α, γ or β. A reasonable way is the estimation of the optimal a_i's which minimize the risk function $R_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) \equiv E[L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p)]$. Instead of minimizing risk functions, Ledoit and Wolf (2012) obtained the optimal a_i's which minimize the loss function and estimated the asymptotic optimal a_i's using the random matrix theory. This argument is used in the next section to estimate the parameters in the estimation of the precision matrix $\boldsymbol{\Sigma}_p^{-1}$ under the Frobenius loss (2.4).

3 Non-linear Shrinkage Estimation of the Precision Matrix

In this paper, we treat two non-linear shrinkage estimators: the Ledoit-Wolf type estimator and the linear ridge estimator. The estimators of the optimal parameters are provided based on the random matrix theory.

3.1 Ledoit-Wolf type estimator

The Ledoit-Wolf type estimator can be derived based on the optimal a_i's which minimize the Frobenius loss (2.4) of the rotation-equivariant estimator $\boldsymbol{\Omega}_p(\boldsymbol{A}_p)$ in (2.5). In fact, the loss function is written as $L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p(\boldsymbol{A}_p)) = p^{-1}\sum_{i=1}^{p}\{a_i^2\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_i - 2a_i\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\boldsymbol{u}_i\} + 1$, which is minimized at

$$a_i^{oracle} = \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\boldsymbol{u}_i/\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_i, \tag{3.1}$$

for $i = 1, \ldots, p$. Since a_i^{oracle}'s depend on $\mathbf{\Sigma}_p$, we need to estimate them. Using the random matrix theory, Ledoit and Péché (2011) showed that under (A1)-(A4), $\boldsymbol{u}_i^{\mathrm{T}} \mathbf{\Sigma}_p \boldsymbol{u}_i$ can be approximated as $\delta(\ell_i)$, where

$$\delta(x) = \begin{cases} \dfrac{x}{|1-y-yx\breve{m}_F(x)|^2} & \text{if } x > 0, \\ \dfrac{1}{(y-1)\breve{m}_{\underline{F}}(0)} & \text{if } x = 0,\ y > 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Concerning the term $\boldsymbol{u}_i^{\mathrm{T}} \mathbf{\Sigma}_p^2 \boldsymbol{u}_i$, we can use a similar argument as in Ledoit and Péché (2011). In fact, the following theorem guarantees that it can be approximated by $\phi(\ell_i)$, where

$$\phi(x) = \begin{cases} \dfrac{x^2\{1-y^2-2y^2x\Re[\breve{m}_F(x)]-y^2x^2|\breve{m}_F(x)|^2\}}{|(1-y-yx\breve{m}_F(x))^2|^2} + \dfrac{\int_{-\infty}^{\infty} tdH(t)yx}{|1-y-yx\breve{m}_F(x)|^2} & \text{if } x > 0, \\ \dfrac{y}{y-1}\dfrac{1}{\breve{m}_{\underline{F}}(0)}\Big(\displaystyle\int_{-\infty}^{\infty} tdH(t) - \dfrac{1}{y\breve{m}_{\underline{F}}(0)}\Big) & \text{if } x = 0,\ y > 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

Theorem 1 *Under* (A1)-(A4)*, the rotation-equivariant estimator* $\mathbf{\Omega}_p(\boldsymbol{A}_p)$ *with* $a_i = a_i(\ell_i)$ *has the almost sure limit given by*

$$L_p(\mathbf{\Sigma}_p^{-1}, \mathbf{\Omega}_p(\boldsymbol{A}_p)) \to \int \big\{a^2(x)\phi(x) - 2a(x)\delta(x)\big\} dF(x) + 1. \tag{3.4}$$

The proof is given in the Appendix. Hence, one gets the Ledoit-Wolf type estimator

$$\begin{aligned} \mathbf{\Omega}_p^{LW*} &= \boldsymbol{U}_p \boldsymbol{A}_p^* \boldsymbol{U}_p^{\mathrm{T}}, \\ \boldsymbol{A}_p^* &= \mathrm{diag}(a_1^*, \cdots, a_p^*), \quad a_i^* = \delta(\ell_i)/\phi(\ell_i). \end{aligned} \tag{3.5}$$

Theorem 1 implies that the estimator $\mathbf{\Omega}_p^{LW*}$ attains the minimum of the limiting loss function, given by $-\int [\{\delta(x)\}^2/\phi(x)] dF(x) + 1$. Since y, F and H are unknown, however, $\mathbf{\Omega}_p^{LW*}$ is not feasible. The way for obtaining a *bona fide* estimator in the oracle estimator $\mathbf{\Omega}_p^{LW*}$ is provided by Ledoit and Wolf (2015). Here, the estimators of $\breve{m}_F(x)$ and $\breve{m}_{\underline{F}}(0)$ are denoted by

$$\widehat{\breve{m}_F}(x) \quad \text{and} \quad \widehat{\breve{m}_{\underline{F}}}(0), \tag{3.6}$$

respectively, where the detail for the derivation is given in the Appendix. Also $\int_{-\infty}^{\infty} tdH(t)$ in equation (3.3) is unknown, but this is the limit of $p^{-1}\mathrm{tr}\,[\mathbf{\Sigma}_p]$ and can be estimated by $p^{-1}\mathrm{tr}\,[\boldsymbol{S}_p]$. Replacing $\breve{m}_F(x)$, $\breve{m}_{\underline{F}}(0)$ and $\int_{-\infty}^{\infty} tdH(t)$ in equations (3.2) and (3.3) with their estimators, and replacing the limiting concentration ratio y with p/N , we have the consistent estimators

$$\widehat{\delta}(x) \quad \text{and} \quad \widehat{\phi}(x), \tag{3.7}$$

for $\delta(x)$ and $\phi(x)$. Then, we get the *bona fide* estimator

$$\begin{aligned} \mathbf{\Omega}_p^{LW} &= \boldsymbol{U}_p \widehat{\boldsymbol{A}}_p^* \boldsymbol{U}_p^{\mathrm{T}}, \\ \widehat{\boldsymbol{A}}_p^* &= \mathrm{diag}(\hat{a}_1^*, \cdots, \hat{a}_p^*), \quad \hat{a}_i^* = \widehat{\delta}(\ell_i)/\widehat{\phi}(\ell_i). \end{aligned} \tag{3.8}$$

3.2 Linear ridge estimator

In the estimation of the inverse $\mathbf{\Sigma}_p^{-1}$, we here consider a new type of the estimator

$$\mathbf{\Omega}_p(\alpha,\beta,\gamma)=\alpha(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}+\beta\boldsymbol{I}_p, \tag{3.9}$$

where α, β and γ are unknown parameters. This estimator not only belongs to the class of rotation-equivariant estimators (2.5) with $a_i=\alpha/(\ell_i+\gamma)+\beta$, but also corresponds to the ridge estimator for $\beta=0$ and the linear estimator for $\gamma=0$. We call $\mathbf{\Omega}_p(\alpha,\beta,\gamma)$ the linear ridge estimator. This estimator is well-conditioned and available when $p>N$. The use of the form $\beta_0\boldsymbol{I}_p$ for constant β_0 has been suggested as an estimator of $\mathbf{\Sigma}_p^{-1}$ in the literature in the ultra-high dimensional case. The ridge-type estimator $\alpha_0(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}$ for constant α_0 is reasonable unless p is ultra-high dimensional. Thus, the linear ridge estimator (3.9) is interpreted as a convex combination of $\alpha_0(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}$ and $\beta_0\boldsymbol{I}_p$, namely, for $0<w<1$, $(1-w)\alpha_0(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}+w\beta_0\boldsymbol{I}_p=\alpha(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}+\beta\boldsymbol{I}_p$ for $\alpha=(1-w)\alpha_0$ and $\beta=w\beta_0$. A related explanation is given in the next subsection from a Bayesian aspect.

The Frobenius loss function of $\mathbf{\Omega}_p(\alpha,\beta,\gamma)$ is written as $L_p(\mathbf{\Sigma}_p^{-1},\mathbf{\Omega}_p(\alpha,\beta,\gamma))=p^{-1}\mathrm{tr}\left[\{\alpha(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p+\beta\mathbf{\Sigma}_p-\boldsymbol{I}_p\}\{\alpha(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p+\beta\mathbf{\Sigma}_p-\boldsymbol{I}_p\}^{\mathrm{T}}\right]$. Given γ, the optimal α and β are given by

$$\begin{aligned}
\alpha^*(\gamma)=&\frac{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p]\mathrm{tr}\,[\mathbf{\Sigma}_p^2]-\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p^2]\mathrm{tr}\,[\mathbf{\Sigma}_p]}{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-2}\mathbf{\Sigma}_p^2]\mathrm{tr}\,[\mathbf{\Sigma}_p^2]-\{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p^2]\}^2},\\
\beta^*(\gamma)=&\frac{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-2}\mathbf{\Sigma}_p^2]\mathrm{tr}\,[\mathbf{\Sigma}_p]-\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p]\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p^2]}{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-2}\mathbf{\Sigma}_p^2]\mathrm{tr}\,[\mathbf{\Sigma}_p^2]-\{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p^2]\}^2}.
\end{aligned}$$

Substituting these optimal quantities into the loss function, we can get

$$\begin{aligned}
L_p^*(\gamma)=&L_p(\mathbf{\Sigma}_p^{-1},\mathbf{\Omega}_p^{LR}(\alpha^*(\gamma),\beta^*(\gamma),\gamma))\\
=&\frac{1}{p}\Big[\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-2}\mathbf{\Sigma}_p^2]\mathrm{tr}\,[\mathbf{\Sigma}_p^2]-\{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p^2]\}^2\Big]^{-1}\\
&\times\Big[-\{\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p]\}^2\mathrm{tr}\,[\mathbf{\Sigma}_p^2]-\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-2}\mathbf{\Sigma}_p^2](\mathrm{tr}\,[\mathbf{\Sigma}_p])^2\\
&\quad+2\,\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p]\mathrm{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\mathbf{\Sigma}_p^2]\mathrm{tr}\,[\mathbf{\Sigma}_p]\Big]+1.
\end{aligned}$$

The asymptotic behaviors of $\alpha^*(\gamma)$, $\beta^*(\gamma)$ and $L_p^*(\gamma)$ can easily be checked by applying the results of the literature. The results are given in the following theorem, which will be proved in the Appendix.

Theorem 2 *Let* $t_1=\int tdH(t)$, $t_2=\int t^2dH(t)$,

$$B(\gamma)=\frac{1-\gamma\breve{m}_F(-\gamma)}{1-y(1-\gamma\breve{m}_F(-\gamma))}\quad\text{and}\quad B'(\gamma)=\frac{d}{d\gamma}B(\gamma).$$

Under (A1)-(A4), $\alpha^*(\gamma)$, $\beta^*(\gamma)$ *and* $L_p^*(\gamma)$ *converge to* $\alpha(\gamma)$, $\beta(\gamma)$ *and* $L(\gamma)$, *respectively, where*

$$\begin{aligned}\alpha(\gamma) =&\frac{1}{c(\gamma)}\Big\{t_1 y\gamma B^2(\gamma) + (t_2 + t_1\gamma - t_1^2 y)B(\gamma) - t_1^2\Big\},\\ \beta(\gamma) =&\frac{1}{c(\gamma)}\Big\{y\gamma B^3(\gamma) + \gamma B^2(\gamma) + 2t_1 y\gamma B(\gamma)B'(\gamma) + t_1(\gamma - t_1)B'(\gamma)\Big\},\\ L(\gamma) =&1 + \frac{1}{c(\gamma)}\Big\{ - 2t_1 y\gamma B^3(\gamma) + (-t_2 + t_1^2 y - 2t_1\gamma)B^2(\gamma) + t_1^2 B(\gamma)\\ &\qquad - 2t_1^2 y\gamma B(\gamma)B'(\gamma) + t_1^2(t_1 - \gamma)B'(\gamma)\Big\},\end{aligned} \tag{3.10}$$

where

$$\begin{aligned}c(\gamma) =& - y^2\gamma^2 B^4(\gamma) + \{t_2 y\gamma + 2y\gamma(t_1 y - \gamma)\}B^3(\gamma)\\ &+ (t_2\gamma - t_1^2 y^2 + 4t_1 y\gamma - \gamma^2)B^2(\gamma) - 2t_1(yt_1 - \gamma)B(\gamma)\\ &+ 2t_1 t_2\gamma B(\gamma)B'(\gamma) + t_1 t_2(\gamma - t_1)B'(\gamma) - t_1^2.\end{aligned}$$

To estimate α, β and γ, we need to estimate $\alpha(\gamma)$, $\beta(\gamma)$ and $L(\gamma)$. As mentioned in Wang, *et al.* (2015), for $\gamma > 0$, $\breve{m}_F(-\gamma)$ can easily be estimated by $\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]/p$, since $\breve{m}_F(-\gamma)$ is the limit of $\breve{m}_{F_p}(-\gamma) = \mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]/p$. Thus, consistent estimators of $B(\gamma)$ and $B'(\gamma)$ given γ are

$$\begin{aligned}\widehat{B}(\gamma) =&\frac{1 - \gamma\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]/p}{1 - (p/N) + \gamma\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]/N},\\ \widehat{B}'(\gamma) =&\frac{p^{-1}\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]\{(p/N)\widehat{B}(\gamma) - 1\}}{1 - (p/N) + \gamma\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]/N}.\end{aligned} \tag{3.11}$$

Since $t_1 = \int t dH(t)$ is the limit of $a_1 = \mathrm{tr}\,[\boldsymbol{\Sigma}_p]/p$, it is estimated by $\hat{a}_1 = \mathrm{tr}\,[\boldsymbol{S}_p]/p$. Also, note that $t_2 = \int t^2 dH(t)$ is the limit of $a_2 = \mathrm{tr}\,[\boldsymbol{\Sigma}_p^2]/p$. Then we can use the consistent estimator

$$\hat{a}_2 = \frac{N-1}{N(N-2)(N-3)p}\Big\{(N-1)(N-2)\mathrm{tr}\,[\widetilde{\boldsymbol{S}}_p^2] + \{\mathrm{tr}\,[\widetilde{\boldsymbol{S}}_p]\}^2 - NQ\Big\},$$

where $\widetilde{\boldsymbol{S}}_p = (N-1)^{-1}\sum_{i=1}^N(\boldsymbol{y}_{p,i} - \overline{\boldsymbol{y}}_p)(\boldsymbol{y}_{p,i} - \overline{\boldsymbol{y}}_p)^{\mathrm{T}}$ and $Q = (N-1)^{-1}\sum_{i=1}^N\{(\boldsymbol{y}_{p,i} - \overline{\boldsymbol{y}}_p)^{\mathrm{T}}(\boldsymbol{y}_{p,i} - \overline{\boldsymbol{y}}_p)\}^2$ for $\overline{\boldsymbol{y}}_p = N^{-1}\sum_{i=1}^N \boldsymbol{y}_{p,i}$. For the details of the estimator $\hat{a}_2$, see Himeno and Yamada (2014). Thus, consistent estimators of t_1 and t_2 are given by $\hat{a}_1$ and $\hat{a}_2$.

Replacing $B(\gamma)$, $B'(\gamma)$, t_1, t_2 and y with $\widehat{B}(\gamma)$, $\widehat{B}'(\gamma)$, $\hat{a}_1$, $\hat{a}_2$ and p/N in (3.10), we have $\widetilde{\alpha}^*(\gamma)$, $\widetilde{\beta}^*(\gamma)$ and $\widetilde{L}^*(\gamma)$. Let $\widetilde{\gamma}$ be the solution of minimizing $\widetilde{L}^*(\gamma)$, namely $\widetilde{\gamma} = \mathrm{argmin}_{\gamma>0}\widetilde{L}^*(\gamma)$. Then, α and β are estimated by $\widetilde{\alpha} = \widetilde{\alpha}^*(\widetilde{\gamma})$ and $\widetilde{\beta} = \widetilde{\beta}^*(\widetilde{\gamma})$.

Although $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\gamma}$ are the consistent estimators of α, β and γ as $(N, p) \to \infty$ with $p/N \to y$, it is not guaranteed that $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\gamma}$ are inside reasonable ranges for finite p and N. For practical use, we need to adjust the values of $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\gamma}$ to be inside appropriate ranges. An approach to this end is the Bayesian argument given in the next subsection. We shall suggest the linear ridge estimator $\boldsymbol{\Omega}_p^{LR}$ in (3.16) through (3.13), (3.14) and (3.15).

3.3 Bayesian motivation and the suggested linear ridge estimator

We give a Bayesian motivation of the linear ridge estimator (3.9). This tells us about how the parameters α, β and γ should be restricted.

Let $\boldsymbol{V} = N\boldsymbol{S}_p$. Consider the case of $N > p$ and assume that $\boldsymbol{V}$ has a Wishart distribution $\mathcal{W}_p(N, \boldsymbol{\Sigma}_p)$. The density function of $\boldsymbol{V}$ is denoted by $f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1})$. Let η be a random variable having a Bernoulli distribution $Ber(\theta)$ with $P(\eta = 1) = \theta$. As a prior distribution of $\boldsymbol{\Sigma}_p^{-1}$, consider a spike and slab prior distribution: Given $\eta = 0$, $\boldsymbol{\Sigma}_p^{-1} \sim \mathcal{W}_p(k, \boldsymbol{\Lambda}_1)$, and given $\eta = 1$, $\boldsymbol{\Sigma}_p^{-1}$ has a one-point distribution $P(\boldsymbol{\Sigma}_p^{-1} = \boldsymbol{\Lambda}_0) = 1$. The density functions of $\boldsymbol{\Sigma}_p^{-1}$ given $\eta = 0$ and $\eta = 1$ are denoted by $\pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1)$ and $\delta_{\boldsymbol{\Lambda}_0}(\boldsymbol{\Sigma}_p^{-1})$, respectively, where $\delta_{\boldsymbol{\Lambda}_0}(\boldsymbol{\Sigma}_p^{-1})$ is assumed to satisfy that $\int \delta_{\boldsymbol{\Lambda}_0}(\boldsymbol{\Sigma}_p^{-1}) d\boldsymbol{\Sigma}_p^{-1} = 1$ and $\int \boldsymbol{\Sigma}_p^{-1} \delta_{\boldsymbol{\Lambda}_0}(\boldsymbol{\Sigma}_p^{-1}) d\boldsymbol{\Sigma}_p^{-1} = \boldsymbol{\Lambda}_0$. Then, the joint and marginal density functions of $(\boldsymbol{V}, \boldsymbol{\Sigma}_p^{-1})$ and $\boldsymbol{V}$ are written as, respectively,

$$\begin{aligned} f(\boldsymbol{V}, \boldsymbol{\Sigma}_p^{-1}) =& f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1})\{(1-\theta)\pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) + \theta \delta_{\boldsymbol{\Lambda}_0}(\boldsymbol{\Sigma}_p^{-1})\}, \\ f(\boldsymbol{V}) =& (1-\theta) \int f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1} + \theta f(\boldsymbol{V} \mid \boldsymbol{\Lambda}_0). \end{aligned}$$

Thus, the Bayes estimator, given by $\boldsymbol{\Omega}_p^{Bayes} = E[\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{V}]$, is expressed as

$$\begin{aligned} \boldsymbol{\Omega}_p^{Bayes} =& \int \boldsymbol{\Sigma}_p^{-1} f(\boldsymbol{V}, \boldsymbol{\Sigma}_p^{-1}) d\boldsymbol{\Sigma}_p^{-1} / f(\boldsymbol{V}) \\ =& \frac{(1-\theta)\int \boldsymbol{\Sigma}_p^{-1} f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1} + \theta \boldsymbol{\Lambda}_0 f(\boldsymbol{V} \mid \boldsymbol{\Lambda}_0)}{(1-\theta)\int f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1} + \theta f(\boldsymbol{V} \mid \boldsymbol{\Lambda}_0)} \\ =& (1-w_0) \frac{\int \boldsymbol{\Sigma}_p^{-1} f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1}}{\int f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1}} + w_0 \boldsymbol{\Lambda}_0, \end{aligned}$$

where

$$w_0 = \frac{\theta f(\boldsymbol{V} \mid \boldsymbol{\Lambda}_0)}{(1-\theta)\int f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1} + \theta f(\boldsymbol{V} \mid \boldsymbol{\Lambda}_0)}.$$

It can be easily seen that

$$\frac{\int \boldsymbol{\Sigma}_p^{-1} f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1}}{\int f(\boldsymbol{V} \mid \boldsymbol{\Sigma}_p^{-1}) \pi(\boldsymbol{\Sigma}_p^{-1} \mid \boldsymbol{\Lambda}_1) d\boldsymbol{\Sigma}_p^{-1}} = (N+k)(\boldsymbol{V} + \boldsymbol{\Lambda}_1)^{-1} = v_0(\boldsymbol{S}_p + N^{-1}\boldsymbol{\Lambda}_1)^{-1},$$

for $v_0 = (N+k)/N$. Thus, we get

$$\boldsymbol{\Omega}_p^{Bayes} = (1-w_0)v_0(\boldsymbol{S}_p + N^{-1}\boldsymbol{\Lambda}_1)^{-1} + w_0 \boldsymbol{\Lambda}_0,$$

where v_0 and w_0 satisfy that $v_0 > 1$ and $0 < w_0 < 1$.

Motivated from the Bayes estimator $\boldsymbol{\Omega}_p^{Bayes}$, we consider the following linear ridge estimator: Let $\boldsymbol{\Lambda}_1 = N\gamma \boldsymbol{I}_p$ and $\boldsymbol{\Lambda}_0 = (1/\bar{\ell})\boldsymbol{I}_p$ for $\bar{\ell} = \sum_{i=1}^p \ell_i / p = \mathrm{tr}\,[\boldsymbol{S}_p]/p$. Then the linear ridge estimator we treat is

$$\boldsymbol{\Omega}_p^{linear\,ridge} = v(1-w)(\boldsymbol{S}_p + \gamma \boldsymbol{I}_p)^{-1} + w(1/\bar{\ell})\boldsymbol{I}_p, \tag{3.12}$$

where v, w and γ are constants which satisfy the restrictions

$$v > 1, \quad 0 < w < 1, \quad \text{and} \quad \gamma > 0. \tag{3.13}$$

This estimator corresponds to (3.9) with $\alpha = v(1-w)$ and $\beta = w/\overline{\ell}$. On the other hand, from the results below (3.11), we have the consistent estimators $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\gamma}$ for α, β and γ. Thus, one gets the equations $\widetilde{\alpha} = \widetilde{v}(1-\widetilde{w})$ and $\widetilde{\beta} = \widetilde{w}/\overline{\ell}$, which yields that $\widetilde{w} = \overline{\ell}\widetilde{\beta}$ and $\widetilde{v} = \widetilde{\alpha}/(1-\widetilde{w}) = \widetilde{\alpha}/(1-\overline{\ell}\widetilde{\beta})$. Since $\widetilde{v}$, $\widetilde{w}$ and $\widetilde{\gamma}$ need to satisfy the restriction (3.13), the parameters v, w and γ are estimated by

$$\widehat{w} = 0 \vee \widetilde{w} \wedge 1 = 0 \vee (\overline{\ell}\widetilde{\beta}) \wedge 1, \quad \widehat{v} = \widetilde{v} \vee 1 = \widetilde{\alpha}/(1-\widetilde{w}) \vee 1, \quad \widehat{\gamma} = \widetilde{\gamma} \vee 0, \tag{3.14}$$

where $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$. This gives

$$\widehat{\alpha} = \widehat{v}(1-\widehat{w}) \quad \text{and} \quad \widehat{\beta} = \widehat{w}/\overline{\ell}, \tag{3.15}$$

and we suggest the use of the linear ridge estimator

$$\begin{aligned} \boldsymbol{\Omega}_p^{LR} =& \widehat{v}(1-\widehat{w})(\boldsymbol{S}_p + \widehat{\gamma}\boldsymbol{I}_p)^{-1} + \widehat{w}(1/\overline{\ell})\boldsymbol{I}_p \\ =& \widehat{\alpha}(\boldsymbol{S}_p + \widehat{\gamma}\boldsymbol{I}_p)^{-1} + \widehat{\beta}\boldsymbol{I}_p. \end{aligned} \tag{3.16}$$

Note that when $\widehat{w} = 0$ or $\widehat{\beta} = 0$, the linear ridge estimator becomes the ridge-type estimator $\widehat{\alpha}(\boldsymbol{S}_p + \widehat{\gamma}\boldsymbol{I}_p)^{-1} = \widehat{v}(\boldsymbol{S}_p + \widehat{\gamma}\boldsymbol{I}_p)^{-1}$. On the other hand, when $\widehat{w} = 1$ or $\widehat{\alpha} = 0$, the linear ridge estimator becomes $\widehat{\beta}\boldsymbol{I}_p = (1/\overline{\ell})\boldsymbol{I}_p$, which is proposed frequently as an estimator of an ultra-high dimensional precision matrix. Generally, in the case that $0 < \widehat{w} < 1$, the linear ridge estimator is a convex combination of the ridge estimator and $(1/\overline{\ell})\boldsymbol{I}_p$.

3.4 Ridge and linear shrinkage estimators for comparison

In the previous subsections, we have derived the Ledoit-Wolf type estimator (3.8) and the linear ridge estimator (3.16) relative to the Frobenius loss function (2.4). To compare them with the existing other estimators, we here look at two estimators: the ridge and the linear shrinkage estimators. Both estimators are special cases of the linear ridge estimator (3.9), and we can investigate how the linear ridge estimator is effective in comparison with them. Thus, we derive the optimal values of the parameters in the ridge and the linear shrinkage estimators with respect to the Frobenius loss.

[1] Ridge estimator. The ridge estimator is of the form

$$\boldsymbol{\Omega}_p^{ridge} = \alpha(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}, \tag{3.17}$$

which has been used and studies in the literature. For example, see Srivastava and Kubokawa (2007). Although most results were obtained under Gaussian distributions, Wang, *et al.* (2014) derived the estimators of the optimal α and γ using the random matrix theory without any assumption of underlying distributions, where they treated another quadratic loss function. Given γ, the optimal α relative to the Frobenius loss is

$$\alpha^{R*}(\gamma) = \frac{\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p]}{\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p^2(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]},$$

which leads to the reduced loss function

$$L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p^{ridge}(\alpha^{R*}(\gamma), \gamma)) = 1 - \frac{1}{p}\frac{\{\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p]\}^2}{\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p^2(\boldsymbol{S}_p + \gamma\boldsymbol{I}_p)^{-1}]}.$$

Then, it follows from the above results that $\alpha^{R*}(\gamma)$ and $L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p^{ridge}(\alpha^{R*}(\gamma), \gamma))$ converge to

$$\begin{aligned}\alpha^R(\gamma) =& \frac{B(\gamma)}{yB^2(\gamma) + B(\gamma) + 2y\gamma B(\gamma)B'(\gamma) + (\gamma - t_1)B'(\gamma)},\\ L^R(\gamma) =& 1 - \frac{B^2(\gamma)}{yB^2(\gamma) + B(\gamma) + 2y\gamma B(\gamma)B'(\gamma) + (\gamma - t_1)B'(\gamma)},\end{aligned}$$

for $B(\gamma)$ and $B'(\gamma)$ in Theorem 2. Replacing $B(\gamma)$, $B'(\gamma)$, t_1 and y with $\widehat{B}(\gamma)$, $\widehat{B}'(\gamma)$, $\hat{a}_1$ and p/N, we have $\widehat{\alpha}^{R*}$ and $\widehat{L}^{R*}(\gamma)$, where $\widehat{B}(\gamma)$ and $\widehat{B}'(\gamma)$ are given in (3.11). Let $\widehat{\gamma}^R$ be the solution of minimizing $\widehat{L}^{R*}(\gamma)$. Then, we get $\widehat{\alpha}^R = \widehat{\alpha}^{R*}(\widehat{\gamma}^R)$, which yields the ridge estimator.

[2] Linear shrinkage estimator. The linear shrinkage estimator suggested by Bodnar, *et al*. (2014) is of the form

$$\boldsymbol{\Omega}_p^{linear} = \begin{cases} \alpha \boldsymbol{S}_p^{-1} + \beta\, \boldsymbol{I}_p & \text{if } N > p \\ \alpha \boldsymbol{S}_p^{+} + \beta\, \boldsymbol{I}_p & \text{if } N < p. \end{cases} \tag{3.18}$$

In the case of $N > p$, we can use the results given in the previous subsections and Theorem 3.2 in Bodnar, *et al*. (2014) to show that the Frobenius loss function of the linear shrinkage estimator has the limit

$$\begin{aligned}&\lim_{N,p\to\infty} L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p^{linear})\\ &= \lim_{N,p\to\infty} \frac{1}{p}\Big\{\alpha^2 \mathrm{tr}\,[\boldsymbol{S}_p^{-2}\boldsymbol{\Sigma}_p^2] + 2\alpha\beta \mathrm{tr}\,[\boldsymbol{S}_p^{-1}\boldsymbol{\Sigma}_p^2] + \beta^2 \mathrm{tr}\,[\boldsymbol{\Sigma}_p^2] - 2\alpha \mathrm{tr}\,[\boldsymbol{S}_p^{-1}\boldsymbol{\Sigma}_p] - 2\beta \mathrm{tr}\,[\boldsymbol{\Sigma}_p]\Big\} + 1\\ &= 1 + \Big(\frac{1}{(1-y)^2} + t_1\frac{1+\breve{m}_F(0)}{1-y}\Big)\alpha^2 + 2\frac{t_1}{1-y}\alpha\beta + t_2\beta^2 - 2\frac{1}{1-y}\alpha - 2t_1\beta,\end{aligned}$$

where t_1 and t_2 are given in Theorem 2. Then, the optimal α and β in terms of minimizing the limit of the loss function are given by

$$\begin{aligned}\alpha^L =& \frac{t_2 - t_1^2}{t_2\{1 + t_1(1+\breve{m}_F(0))(1-y)\}},\\ \beta^L =& \frac{t_1}{t_2}\Big(1 - \frac{\alpha^L}{1-y}\Big).\end{aligned}$$

Note that $\breve{m}_F(0)$ can be consistently estimated by $\mathrm{tr}\,[\boldsymbol{S}_p^{-1}]/p$ when $N > p$. Replacing t_1, t_2 and $\breve{m}_F(0)$ with their estimators and replacing y with p/N, one gets the estimators $\widehat{\alpha}^L$ and $\widehat{\beta}^L$, which yields the linear estimator.

In the case of $N < p$, Bodnar, *et al*. (2014) cannot provide estimators for general $\boldsymbol{\Sigma}_p$. This is because in their settings one needs the limit of $p^{-1}\mathrm{tr}\,[\boldsymbol{S}_p^{+}\boldsymbol{\Sigma}_p]$, but this cannot be obtained without assuming a structure such as $\boldsymbol{\Sigma}_p = \sigma^2 \boldsymbol{I}_p$ for scalar σ^2. Without assuming such a structure, however, we can obtain estimators of the optimal α and β using equation (3.3). In our setting, we can expand the loss function as

$$\begin{aligned}&L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p^{linear})\\ &= \frac{1}{p}\Big\{\alpha^2 \mathrm{tr}\,[(\boldsymbol{S}_p^{+})^2\boldsymbol{\Sigma}_p^2] + 2\alpha\beta \mathrm{tr}\,[\boldsymbol{S}_p^{+}\boldsymbol{\Sigma}_p^2] + \beta^2 \mathrm{tr}\,[\boldsymbol{\Sigma}_p^2] - 2\alpha \mathrm{tr}\,[\boldsymbol{S}_p^{+}\boldsymbol{\Sigma}_p] - 2\beta \mathrm{tr}\,[\boldsymbol{\Sigma}_p]\Big\} + 1,\end{aligned}$$

so that we need the limit of $p^{-1}\mathrm{tr}\,[(\boldsymbol{S}_p^+)^2\boldsymbol{\Sigma}_p^2]$, $p^{-1}\mathrm{tr}\,[\boldsymbol{S}_p^+\boldsymbol{\Sigma}_p^2]$ and $p^{-1}\mathrm{tr}\,[\boldsymbol{S}_p^+\boldsymbol{\Sigma}_p]$. Using equation (3.3) and Theorem 3.3 in Bodnar, *et al.* (2014), one gets

$$\begin{aligned}
&\lim_{N,p\to\infty} L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p^{linear}) \\
&= 1+\alpha^2 \lim_{N,p\to\infty} \frac{1}{p}\sum_{i=1}^{N}\frac{\phi(\ell_i)}{(\ell_i)^2} + 2\alpha\beta \lim_{N,p\to\infty}\frac{1}{p}\sum_{i=1}^{N}\frac{\phi(\ell_i)}{\ell_i} + t_2\beta^2 - 2\frac{1}{y-1}\alpha - 2t_1\beta \\
&= 1+\alpha^2\int\frac{\phi(x)}{x^2}d\underline{F}(x) + 2\alpha\beta\int\frac{\phi(x)}{x}d\underline{F}(x) + t_2\beta^2 - 2\frac{1}{y-1}\alpha - 2t_1\beta,
\end{aligned}$$

where $\phi(\cdot)$ is given in (3.3) and $\ell_i,\ i=1,\cdots,N$ are the eigenvalues of $\boldsymbol{S}_p$. Then, the optimal α and β in terms of minimizing the limit of the loss function are given by

$$\begin{aligned}
\alpha^L =& \frac{t_2/(y-1) - t_1\int\{\phi(x)/x\}d\underline{F}(x)}{t_2\int\{\phi(x)/x^2\}d\underline{F}(x) - [\int\{\phi(x)/x\}d\underline{F}(x)]^2}, \\
\beta^L =& \frac{t_1\int\{\phi(x)/x^2\}d\underline{F}(x) - \int\{\phi(x)/x\}d\underline{F}(x)/(y-1)}{t_2\int\{\phi(x)/x^2\}d\underline{F}(x) - [\int\{\phi(x)/x\}d\underline{F}(x)]^2}.
\end{aligned}$$

Note that $\int x^k\phi(x)d\underline{F}(x)$ for integer k can be estimated by $p^{-1}\sum_{i=1}^N \ell_i^k\widehat{\phi}(\ell_i)$ for $\widehat{\phi}(x)$ given in (3.7). Replacing t_1, t_2, $\int\{\phi(x)/x\}d\underline{F}(x)$ and $\int\{\phi(x)/x^2\}d\underline{F}(x)$ with their estimators and replacing y with p/N, we have the consitent estimators $\widehat{\alpha}^L$ and $\widehat{\beta}^L$, which produces the linear estimator for $p > N$.

4 Simulation and Real Data Analysis

4.1 Simulation study

In this section, we investigate performances of the procedures suggested in the previous sections through Monte Carlo simulation. In our simulation, eigenvalues of a population covariance matrix are fixed as follows: Let $F_{(a,b)}(x)$ be a cumulative distribution function of the beta distribution $beta(a,b)$, namely

$$F_{(a,b)}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^x t^{a-1}(1-t)^{b-1}dt,\ \ x\in[0,1].$$

Since the support of the beta distribution is $[0,1]$, we transform the interval linearly to $[1,10]$, and the population eigenvalues are given by

$$1+9F_{(a,b)}^{-1}\Big(\frac{i}{p}-\frac{1}{2p}\Big),\quad i=1,\cdots,p. \tag{4.1}$$

This setup is the same as in Ledoit and Wolf (2012).

We first treat the case of $(a,b)=(1,1)$, and investigate the numerical performances of the risk of $\boldsymbol{\Omega}_p^{oracle}$, $\boldsymbol{\Omega}_p^{LW}$, $\boldsymbol{\Omega}_p^{LR}$, $\boldsymbol{\Omega}_p^{ridge}$ and $\boldsymbol{\Omega}_p^{linear}$, which are given in (3.1), (3.8), (3.16), (3.17) and (3.18), respectively. These estimators are denoted by oracle, LW, LR, ridge and linear in the tables given below. For computation of $\boldsymbol{\Omega}_p^{LW}$ and $\boldsymbol{\Omega}_p^{linear}$, we need the QuEST function introduced in Ledoit and Wolf (2015). Using this package, we carried out the simulation experiments on the Matlab. Random observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ are generated as $\boldsymbol{y}_i = \boldsymbol{\Sigma}_p^{1/2}\boldsymbol{x}_i$ for $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ where $x_{i1}, \ldots, x_{ip}$ are mutually independently distributed as

(Case I) $x_{ij} \sim \mathcal{N}(0, 1)$,
(Case II) $x_{ij} \sim \sqrt{(m-2)/m}\, z_{i,j}$ for $z_{i,j} \sim t_m$,

where t_m is a t-distribution with m degrees of freedom. (Case II) is an example of a heavy-tailed distribution and we treat the case of $m = 10$. The simulation experiments are carried out under the above setup for $N = 50$, $p = 30$, 70, 150, 250, 500 and 700, and empirical risks of these estimators are calculated based on $1,000$ replications.

Table 1 provides values of empirical risks of the estimators in normal and $\sqrt{4/5}t_{10}$ distributions with $m = 10$, where eigenvalues of $\boldsymbol{\Sigma}$ are generated by (4.1) for $(a, b) = (1, 1)$. Definitely, the linear ridge estimator $\boldsymbol{\Omega}_p^{LR}$ performs better than the ridge and linear estimators $\boldsymbol{\Omega}_p^{ridge}$ and $\boldsymbol{\Omega}_p^{linear}$. Comparing $\boldsymbol{\Omega}_p^{LR}$ and $\boldsymbol{\Omega}_p^{LW}$, we can see that the linear ridge estimator $\boldsymbol{\Omega}_p^{LR}$ is better than the Ledoit-Wolf type estimator $\boldsymbol{\Omega}_p^{LW}$ except for the cases of $p = 70$ or 150. Especially, in the case of $p = 150$, the estimator $\boldsymbol{\Omega}_p^{LW}$ gives a smaller risk than $\boldsymbol{\Omega}_p^{LR}$ in the normal and t_{10} distributions, but the improvement is not significant. When p is small or p/N is comparatively large, $\boldsymbol{\Omega}_p^{LR}$ performs better than $\boldsymbol{\Omega}_p^{LW}$. It may be because in such cases, $\widehat{\delta}(\ell_i)$ and $\widehat{\phi}(\ell_i)$ in equation (3.8) give poor approximations for $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\boldsymbol{u}_i$ and $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_i$ in equation (3.1), respectively. The resulting estimate $\widehat{\delta}(\ell_i)/\widehat{\phi}(\ell_i)$ causes over-shrinkage or under-shrinkage, which leads to the increased risk of $\boldsymbol{\Omega}_p^{LW}$. For the linear ridge estimator, on the other hand, we only have to calculate $p^{-1}\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma \boldsymbol{I}_p)^{-1}]$, namely, we can use all the sample eigenvalues to estimate parameters. This leads to a stability of $\boldsymbol{\Omega}_p^{LR}$ in the cases that N and p are small or p/N is comparatively large. Comparing $\boldsymbol{\Omega}_p^{LW}$ and $\boldsymbol{\Omega}_p^{LR}$ with the oracle estimator, we observe that as p gets larger, the risks of $\boldsymbol{\Omega}^{LW}$ and $\boldsymbol{\Omega}_p^{LR}$ become more inflated in comparison with the oracle estimator, namely, the estimation error of $\hat{a}_i^*$ in (3.8) for estimating a_i in the Ledoit-Wolf type estimator increases for large p. The inflation in the risk of $\boldsymbol{\Omega}_p^{LR}$ seems relatively smaller than that of $\boldsymbol{\Omega}_p^{LW}$ for large p.

Table 1: Empirical Risks of $\boldsymbol{\Omega}_p^{oracle}$, $\boldsymbol{\Omega}_p^{LW}$, $\boldsymbol{\Omega}_p^{LR}$, $\boldsymbol{\Omega}_p^{ridge}$ and $\boldsymbol{\Omega}_p^{linear}$ with $N = 50$, $(a, b) = (1, 1)$

	p	oracle	LW	LR	ridge	linear
Normal (a,b)=(1,1)	30	0.1538	0.1710	0.1665	0.1681	0.1830
	70	0.1703	0.1782	0.1770	0.1813	0.8705
	150	0.1769	0.1854	0.1856	0.1901	0.8081
	250	0.1791	0.1933	0.1902	0.1951	0.9056
	500	0.1800	0.2342	0.1981	0.2209	0.9757
	700	0.1799	0.3789	0.2116	0.2561	0.9877
t_{10} (a,b)=(1,1)	30	0.1544	0.1704	0.1670	0.1689	0.1878
	70	0.1702	0.1786	0.1787	0.1823	0.8612
	150	0.1769	0.1849	0.1869	0.1896	0.8110
	250	0.1790	0.1911	0.1889	0.1932	0.9062
	500	0.1801	0.2189	0.2029	0.2117	0.9757
	700	0.1799	0.2632	0.2174	0.2464	0.9876

We next investigate the performances of the estimators for some variety of shapes of population spectral distribution. As the shape parameters (a, b) in the beta distribution, we deal with the cases of $(a, b) = (1.5, 1.5), (0.5, 0.5), (5, 5)$ and $(2, 5)$. For graphical illustration of the beta distribution under these parameters, see Ledoit and Wolf (2012). We carry out similar simulation experiments as described above for $N = 50$ and $p = 30$, 70, 150, 250 and 500. Population eigenvalues are generated by equation (4.1). The results are reported in Table 2. It is revealed that we have similar risk performances as mentioned in the case of $(a, b) = (1, 1)$. Thus, the results given in Tables 1 and 2 show that the suggested linear ridge estimator $\boldsymbol{\Omega}_p^{LR}$ gives a relatively good performance among the estimators compared in this simulation.

Table 2: Empirical Risks of $\boldsymbol{\Omega}_p^{oracle}$, $\boldsymbol{\Omega}_p^{LW}$, $\boldsymbol{\Omega}_p^{LR}$, $\boldsymbol{\Omega}_p^{ridge}$ and $\boldsymbol{\Omega}_p^{linear}$ with $N = 50$ under Normal Distribution

(a, b)	p	oracle	LW	LR	ridge	linear
	30	0.1216	0.1354	0.1327	0.1343	0.1437
	70	0.1335	0.1415	0.1416	0.1472	0.8494
(1.5,1.5)	150	0.1388	0.1468	0.1467	0.1534	0.8043
	250	0.1405	0.1551	0.1487	0.1580	0.9087
	500	0.1415	0.1887	0.1631	0.1813	0.9766
	30	0.2122	0.2288	0.2253	0.2304	0.2542
	70	0.2359	0.2441	0.2436	0.2463	0.8932
(0.5,0.5)	150	0.2444	0.2534	0.2565	0.2559	0.8279
	250	0.2468	0.2627	0.2547	0.2619	0.9008
	500	0.2480	0.2982	0.2629	0.2866	0.9738
	30	0.0496	0.0585	0.0570	0.0693	0.0595
	70	0.0536	0.0595	0.0604	0.0754	0.8357
(5,5)	150	0.0557	0.0624	0.0612	0.0757	0.7958
	250	0.0563	0.0688	0.0653	0.0769	0.9089
	500	0.0567	0.1072	0.0843	0.1015	0.9784
	30	0.1123	0.1277	0.1257	0.1268	0.1421
	70	0.1248	0.1340	0.1338	0.1376	0.8357
(2,5)	150	0.1323	0.1407	0.1416	0.1445	0.8042
	250	0.1350	0.1487	0.1443	0.1507	0.9100
	500	0.1364	0.1843	0.1589	0.1760	0.9769

4.2 Illustrative example

We now compare the performance of several estimators of the precision matrix through the quadratic discriminant analysis (QDA). In QDA, when $p \times 1$ observation $\boldsymbol{x}_{i,j}$, $i = 1, 2, j = 1, \ldots, N_i$ are obtained from the population group Π_i, $i = 1, 2$, a new observation $\boldsymbol{x}$ is classified into Π_1 or Π_2 via the classification rule

$$(\boldsymbol{x} - \overline{\boldsymbol{x}}_1)^{\mathrm{T}} \boldsymbol{\Omega}_{1,p} (\boldsymbol{x} - \overline{\boldsymbol{x}}_1) - (\boldsymbol{x} - \overline{\boldsymbol{x}}_2)^{\mathrm{T}} \boldsymbol{\Omega}_{2,p} (\boldsymbol{x} - \overline{\boldsymbol{x}}_2) < (\text{resp.} >)\, 0 \Longrightarrow \boldsymbol{x} \in \Pi_1 (\text{resp.} \Pi_2),$$

where $\overline{\boldsymbol{x}}_i = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{x}_{i,j}$, $i = 1, 2$ and $\boldsymbol{\Omega}_{i,p}$, $i = 1, 2$ is an estimator of the precision matrix of the population Π_i. The classification rules given by substituting the estimators $\boldsymbol{\Omega}_{1,p}$ and $\boldsymbol{\Omega}_{2,p}$ corresponding to $\boldsymbol{\Omega}_p^{LW}$, $\boldsymbol{\Omega}_p^{LR}$, $\boldsymbol{\Omega}_p^{ridge}$, $\boldsymbol{\Omega}_p^{linear}$, $\boldsymbol{\Omega}_p^{MP}$ and $\boldsymbol{\Omega}_p^{diag}$ are denoted by LW, LR, ridge, linear, MP and diag, where $\boldsymbol{\Omega}_p^{MP}$ is the Moore-Penrose generalized inverse matrix and $\boldsymbol{\Omega}_p^{diag}$ is the diagonal matrix $\boldsymbol{\Omega}_p^{diag} = (\mathrm{diag}(\boldsymbol{S}_p))^{-1}$. It is known in the literature that $\boldsymbol{\Omega}_p^{diag}$ has a good performance in discriminant analysis. To estimate these estimators, the sample covariance matrix of the population Π_i is estimated by

$$\boldsymbol{S}_{i,p} = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (\boldsymbol{x}_{i,j} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{i,j} - \overline{\boldsymbol{x}}_i)^{\mathrm{T}}.$$

We use the microarray data described in a colon cancer study by Alon, *et al.* (1999) where expression levels for 2000 genes were measured on 40 normal and on 22 colon tumor tissues. The dataset is available at http://genomics-pubs.princeton.edu/oncology/affydata. A base 10 logarithmic transformation is applied. The description of the above datasets and preprocessing are due to Dettling and Buhlmann (2003) except standardization is not followed. This dataset was used by Srivastava and Kubokawa (2007), Fisher and Sun (2011) and Touloumis (2015).

Using this dataset, we estimate the covariance matrices of the normal and colon cancer groups of the top p genes, where $p = 100, 250, 500$ and 900. The correct classification rates based on the estimators $\boldsymbol{\Omega}_p^{LW}$, $\boldsymbol{\Omega}_p^{LR}$, $\boldsymbol{\Omega}_p^{ridge}$, $\boldsymbol{\Omega}_p^{linear}$, $\boldsymbol{\Omega}_p^{MP}$ and $\boldsymbol{\Omega}_p^{diag}$ are estimated by a leave-one-out cross validation. The estimated correct classification rates are reported in Table 3. A relative gain in correct classification rate for our $\boldsymbol{\Omega}_p^{LR}$ is at least 28.7 % over $\boldsymbol{\Omega}_p^{LW}$ and is at most 28.7 % over $\boldsymbol{\Omega}_p^{ridge}$. Moreover, $\boldsymbol{\Omega}_p^{LR}$ shows equal or slightly higher correct classification rates than $\boldsymbol{\Omega}_p^{linear}$ and $\boldsymbol{\Omega}_p^{diag}$. Lastly we can see that the correct classification rate of $\boldsymbol{\Omega}_p^{linear}$ is higher than that of $\boldsymbol{\Omega}_p^{LW}$ and $\boldsymbol{\Omega}_p^{ridge}$, which seems to contradict the results in the previous simulation study. This is because this data set of the colon cancer has considerable sparsity in the covariance structure. Thus, we may improve the correct classification rate by assuming the sphericity or diagonality for the precision matrix.

Table 3: Correct Classification Rates in the Colon Cancer Dataset

p	LW	LR	ridge	linear	MP	diag
100	67.7 %	87.1 %	71.0 %	83.9 %	38.7 %	85.5 %
250	65.2 %	87.1 %	83.9 %	87.1 %	38.7 %	83.9 %
500	61.3 %	87.1 %	72.6 %	83.9 %	41.9 %	87.1 %
900	66.1 %	87.1 %	61.3 %	87.1 %	43.6 %	87.1 %

5 Concluding Remarks

In this paper, we have addressed the problem of estimating the high-dimensional precision matrix $\boldsymbol{\Sigma}_p^{-1}$ relative ot the Frobenius loss function $L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) = p^{-1}\mathrm{tr}\,(\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \boldsymbol{I}_p)(\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \boldsymbol{I}_p)^{\mathrm{T}}$. We have suggested the linear ridge estimator of the precision matrix motivated from the Bayesian aspect, and provided the estimators of the optimal parameters using the random matrix theory.

We also have derived the Ledoit-Wolf type estimators and compared it with the suggested linear ridge estimator and the other existing estimators. Some simulation results show that the linear ridge estimator performs better than the other existing estimators when N and p are small, or when p/N is large.

It is worth mentioning the estimation under other loss functions. A related loss function is the quadratic loss $p^{-1}\mathrm{tr}\,[(\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \boldsymbol{I}_p)^2]$. Under this loss fucntion, we can obtain the ridge, linear and linear ridge estimators. However, we could not derive a Ledoit-Wolf type nonlinear shrinkage estimator, because we need the asymptotic behavior of $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\boldsymbol{u}_j$ for $i \neq j$. In our numerical investigation, these quantities seem almost zero, but not just zero. This is left for future research. Ledoit and Wolf (2012) treated the loss functions $p^{-1}\mathrm{tr}\,[(\boldsymbol{\Omega}_p - \boldsymbol{\Sigma}^{-1})^2]$ and $p^{-1}\{\mathrm{tr}\,[\boldsymbol{\Omega}_p\boldsymbol{\Sigma}] - \log|\boldsymbol{\Omega}_p\boldsymbol{\Sigma}| - p$ and derived the nonlinear shrinkage estimators. We can derive the ridge, linear and linear ridge estimators relative to these loss functions, and we shall compare those estimators numerically under the loss function in a future study. To get estimators of the parameters a_i, α, β and γ under those loss functions, we have to solve equation (2.3) by resorting to the strong package QuEST run on Matlab. An advantage of the Frobenius loss function treated in this paper is that the estimators of α and β in the linear ridge estimator and the ridge estimator are explicitly expressed for fixed γ, and the resulting linear ridge and ridge estimators are written in closed forms. Since the other estimators can be derived by using the QuEST, we can compare the performance of the suggested linear ridge estimator with the existing other shrinkage ones.

Concerning the mean $\boldsymbol{\mu} = E[\boldsymbol{x}_{p,i}]$, we have treated the case of $\boldsymbol{\mu} = \boldsymbol{0}$ in this paper. In the case of $\boldsymbol{\mu} \neq \boldsymbol{0}$, however, it is noted that all the results in this paper still hold by replacing $\boldsymbol{y}_{p,i}$ and N with $\boldsymbol{y}_{p,i} - \overline{\boldsymbol{y}}_p$ and $n = N - 1$ for $\overline{\boldsymbol{y}}_p = N^{-1}\sum_{i=1}^N \boldsymbol{y}_i$ as long as we consider the asymptotics on N and p. That is, the sample covariance matrix is

$$\boldsymbol{S}_p = \frac{1}{n}\sum_{i=1}^N (\boldsymbol{y}_{p,i} - \overline{\boldsymbol{y}}_p)(\boldsymbol{y}_{p,i} - \overline{\boldsymbol{y}}_p)^{\mathrm{T}},$$

which is rewritten as

$$\boldsymbol{S}_p = \frac{N}{n}\frac{1}{N}\sum_{i=1}^N (\boldsymbol{y}_{p,i} - \boldsymbol{\mu})(\boldsymbol{y}_{p,i} - \boldsymbol{\mu})^{\mathrm{T}} - \frac{N}{n}(\overline{\boldsymbol{y}}_p - \boldsymbol{\mu})(\overline{\boldsymbol{y}} - \boldsymbol{\mu})^{\mathrm{T}}.$$

Leaving the second term out of consideration does not affect the LSD of $\boldsymbol{S}_p$ since $\mathrm{rank}((\overline{\boldsymbol{y}}_p - \boldsymbol{\mu})(\overline{\boldsymbol{y}}_p - \boldsymbol{\mu})^{\mathrm{T}}) = 1$ and is negligible in our arguments for large N and p.

It would be interesting to use the suggested estimator for testing the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$. Reasonable test statistics are functions of $\overline{\boldsymbol{y}}_p^{\mathrm{T}}\boldsymbol{\Omega}_p\overline{\boldsymbol{y}}_p$. For example, test statistics of Bai and Saradanasa (1996) and Srivastava (2007) are based on $\overline{\boldsymbol{y}}_p^{\mathrm{T}}\overline{\boldsymbol{y}}_p/(p^{-1}\mathrm{tr}\,[\boldsymbol{S}_p])$ and $\overline{\boldsymbol{y}}_p^{\mathrm{T}}\boldsymbol{S}^{+}\overline{\boldsymbol{y}}_p$, respectively, which are correspond to $\boldsymbol{\Omega}_p = (p/\mathrm{tr}\,\boldsymbol{S})\boldsymbol{I}_p = (1/\overline{\ell})\boldsymbol{I}_p$ and $\boldsymbol{\Omega}_p = \boldsymbol{S}^{+}$. Using the linear ridge estimator for $\boldsymbol{\Omega}_p$, we can suggest a test statistic based on

$$\begin{aligned}\overline{\boldsymbol{y}}_p^{\mathrm{T}}\boldsymbol{\Omega}_p^{LR}\overline{\boldsymbol{y}}_p =& \overline{\boldsymbol{y}}_p^{\mathrm{T}}\Big((1-w)\alpha_0(\boldsymbol{S} + \gamma\boldsymbol{I}_p)^{-1} + w(1/\overline{\ell})\boldsymbol{I}_p\Big)\overline{\boldsymbol{y}} \\ =& (1-w)\alpha_0\overline{\boldsymbol{y}}_p^{\mathrm{T}}(\boldsymbol{S} + \gamma\boldsymbol{I}_p)^{-1}\overline{\boldsymbol{y}} + w(1/\overline{\ell})\overline{\boldsymbol{y}}^{\mathrm{T}}\overline{\boldsymbol{y}},\end{aligned}$$

which is a convex combination of the Bai-Saranadasa statistic and the Hotelling T-statistic replacing $\boldsymbol{S}^{-1}$ or $\boldsymbol{S}^{+}$ with the ridge estimator. We will study the performance of a test statistic based $\overline{\boldsymbol{y}}_p^{\mathrm{T}}\boldsymbol{\Omega}_p^{LR}\overline{\boldsymbol{y}}_p$ as a future project.

A Appendix

A.1 Proof of Theorem 1

The Frobenius loss function of the rotation-equivariant estimator is

$$L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p(\boldsymbol{A}_p)) = \frac{1}{p}\sum_{i=1}^{p}\{a^2(\ell_i)\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_i - 2a(\ell_i)\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\boldsymbol{u}_i\} + 1.$$

Theorem 4 in Ledoit and Péché (2011) shows that

$$\frac{1}{p}\sum_{i=1}^{p} a(\ell_i)\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\boldsymbol{u}_i \to \int a(x)\delta(x)dF(x),$$

where $\delta(x)$ is given in (3.2). Thus, we shall evaluate the asymptotic quantity of $p^{-1}\sum_{i=1}^{p} a(\ell_i)\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_i$. The same arguments as used in the proof of Theorem 4 in Ledoit and Péché (2011) are heavily exploited for the evaluation. Let $\Delta_p^{(2)}(x)$ be the nondecreasing function defined by

$$\Delta_p^{(2)}(x) = \frac{1}{p}\sum_{j=1}^{p}\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_j \times \mathbb{I}_{[\ell_j,+\infty)}(x),\ \forall x \in \mathbb{R}.$$

When all the sample eigenvalues are distinct, it is noted that $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_i\,(i = 1, \cdots, p)$ can be recovered from $\Delta_p^{(2)}$ as

$$\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\boldsymbol{u}_i = \lim_{\epsilon\to 0^+}\frac{\Delta_p^{(2)}(\ell_i+\epsilon) - \Delta_p^{(2)}(\ell_i-\epsilon)}{F_p(\ell_i+\epsilon) - F_p(\ell_i-\epsilon)}.$$

Thus, it suffices to examine the asymptotic behavior of $\Delta_p^{(2)}(x)$. Note that the stieltjes transform of $\Delta_p^{(2)}(x)$ is the same with Euation (3) of Ledoit and Péché (2011) for $g(\tau_j) = \tau_j^2$. Let $\Theta_p^{(2)}(z)$ be the stieltjes transform of $\Delta_p^{(2)}(x)$. Then, $\Theta_p^{(2)}(z)$ converges $a.s.$ to $\Theta^{(2)}(z)$ for all $z \in \mathbb{C}^+$. Moreover, it follows from Lemmas 2 and 3 in Ledoit and Péché (2011) that

$$\Theta^{(2)}(z) = \frac{z + z^2 m_F(z)}{(1 - y - yzm_F(z))^2} + \frac{\int t dH(t)}{1 - y - yzm_F(z)},\ \forall z \in \mathbb{C}^+. \tag{A.1}$$

Using Lemma 6 in Ledoit and Péché (2011), we can see that $\lim_{p\to\infty}\Delta_p^{(2)}(x) \equiv \Delta^{(2)}(x)$ exists and is equal to

$$\Delta^{(2)}(x) = \lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\infty}^{x}\Im[\Theta^{(2)}(\lambda + i\eta)]d\lambda. \tag{A.2}$$

for every continuous point $x \in \mathbb{R}$ of $\Delta^{(2)}(x)$. Plugging (A.1) into (A.2) yields

$$\begin{aligned}
\Delta^{(2)}(x) &= \lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\infty}^{x}\Im\Big[\frac{(\lambda+i\eta) + (\lambda+i\eta)^2 m_F(\lambda+i\eta)}{(1 - y - y(\lambda+i\eta)m_F(\lambda+i\eta))^2} + \frac{\int t dH(t)}{1 - y - y(\lambda+i\eta)m_F(\lambda+i\eta)}\Big]d\lambda \\
&= \lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\infty}^{x}\Im\Big[\frac{(\lambda+i\eta) + (\lambda+i\eta)^2 m_F(\lambda+i\eta)}{(1 - y - y(\lambda+i\eta)m_F(\lambda+i\eta))^2}\Big]d\lambda \\
&\quad + \lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\infty}^{x}\Im\Big[\frac{\int t dH(t)}{1 - y - y(\lambda+i\eta)m_F(\lambda+i\eta)}\Big]d\lambda.
\end{aligned} \tag{A.3}$$

We begin by evaluating $\Delta^{(2)}(x)$ in the case of $y \in (0,1)$. The first term in RHS of (A.3) is

$$\lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\infty}^{x}\Im\Big[\frac{(\lambda+i\eta)+(\lambda+i\eta)^2 m_F(\lambda+i\eta)}{(1-y-y(\lambda+i\eta)m_F(\lambda+i\eta))^2}\Big]d\lambda$$

$$= \lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\infty}^{x}\Im\Big[(\lambda+i\eta)\Big\{1+(\lambda+i\eta)m_F(\lambda+i\eta)\Big\}$$

$$\times\Big\{(1-y-y\Re[(\lambda+i\eta)m_F(\lambda+i\eta)])^2-y^2(\Im[(\lambda+i\eta)m_F(\lambda+i\eta)])^2$$

$$+2iy\Im[(\lambda+i\eta)m_F(\lambda+i\eta)](1-y-y\Re[(\lambda+i\eta)m_F(\lambda+i\eta)])\Big\}\Big]$$

$$\times(|(1-y-y(\lambda+i\eta)m_F(\lambda+i\eta))^2|^2)^{-1}d\lambda,$$

which is rewritten as

$$\frac{1}{\pi}\int_{-\infty}^{x}\frac{\lambda\Im[\lambda m_F(\lambda)]\big\{(1+y+y\Re[\lambda m_F(\lambda)])(1-y-y\Re[\lambda m_F(\lambda)])-y^2(\Im[\lambda m_F(\lambda)])^2\big\}}{|(1-y-y\lambda m_F(\lambda))^2|^2}d\lambda$$

$$=\frac{1}{\pi}\int_{-\infty}^{x}\frac{\Im[m_F(\lambda)]\lambda^2\{1-y^2-2y^2\lambda\Re[m_F(\lambda)]-y^2\lambda^2|m_F(\lambda)|^2\}}{|(1-y-y\lambda m_F(\lambda))^2|^2}d\lambda$$

$$=\int_{-\infty}^{x}\frac{\lambda^2\{1-y^2-2y^2\lambda\Re[m_F(\lambda)]-y^2\lambda^2|m_F(\lambda)|^2\}}{|(1-y-y\lambda m_F(\lambda))^2|^2}dF(\lambda).$$

The second term in RHS of (A.3) is

$$\lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\infty}^{x}\Im[\frac{\int tdH(t)}{1-y-y(\lambda+i\eta)m_F(\lambda+i\eta)}]d\lambda$$

$$=\int tdH(t)\times\frac{1}{\pi}\lim_{\eta\to 0^+}\int_{-\infty}^{x}\frac{y\Im[(\lambda+i\eta)m_F(\lambda+i\eta)]}{|1-y-y(\lambda+i\eta)m_F(\lambda+i\eta)|^2}d\lambda,$$

which is equal to

$$\int tdH(t)\times\frac{1}{\pi}\int_{-\infty}^{x}\frac{y\Im[\lambda m_F(\lambda)]}{|1-y-y\lambda m_F(\lambda)|^2}d\lambda$$

$$=\int tdH(t)\times\int_{-\infty}^{x}\frac{y\lambda F'(\lambda)}{|1-y-y\lambda m_F(\lambda)|^2}d\lambda=\int_{-\infty}^{x}\frac{\int tdH(t)y\lambda}{|1-y-y\lambda m_F(\lambda)|^2}dF(\lambda).$$

Hence, for $x>0$, it is concluded that

$$\Delta^{(2)}(x)=\int_{-\infty}^{x}\phi(\lambda)dF(\lambda),$$

for $\phi(\cdot)$ given in (3.3). This shows Theorem 1 in the case of $y \in (0,1)$.

We next consider the case of $y \in (1,\infty)$. Rewriting $\Theta^{(2)}(z)$ based on the relation between $m_F(z)$ and $m_{\underline{F}}(z)$ given by $y+yzm_F(z)=1+zm_{\underline{F}}(z)$, we have

$$\Theta^{(2)}(z)=\frac{1}{-zm_{\underline{F}}(z)}\Big(z\frac{\frac{1}{y}+\frac{1}{y}zm_{\underline{F}}(z)}{-zm_{\underline{F}}(z)}+\int tdH(t)\Big)$$

$$=\frac{1}{zm_{\underline{F}}(z)}\Big(\frac{1+zm_{\underline{F}}(z)}{ym_{\underline{F}}(z)}-\int tdH(t)\Big)=-\frac{\mu(z)}{z}$$

where

$$\mu(z) = \frac{1}{m_{\underline{F}}(z)}\Big(- \frac{1+zm_{\underline{F}}(z)}{ym_{\underline{F}}(z)} + \int tdH(t)\Big).$$

The inversion formula for the stieltjes transform implies that:

$$\begin{aligned}\lim_{\epsilon\to 0^+}(\Delta^{(2)}(\epsilon) - \Delta^{(2)}(-\epsilon)) &= \lim_{\epsilon\to 0^+}\lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\epsilon}^{\epsilon}\Im[\Theta^{(2)}(\lambda+i\eta)]d\lambda \\ &= \lim_{\epsilon\to 0^+}\lim_{\eta\to 0^+}\frac{1}{\pi}\int_{-\epsilon}^{\epsilon}\Im\Big[-\frac{\mu(\lambda+i\eta)}{\lambda+i\eta}\Big]d\lambda \\ &=\mu(0) \\ &=\frac{1}{m_{\underline{F}}(0)}\Big(-\frac{1}{ym_{\underline{F}}(0)} + \int tdH(t)\Big), \end{aligned} \tag{A.4}$$

The third equality follows from Lemma 9 in Ledoit and Péché (2011), since μ is a complex holomorphic function and $\mu(0)\in\mathbb{R}$. Also from Lemma 8 in Ledoit and Péché (2011)], it is noted that $F(\lambda) = (1 - y^{-1})\mathbb{I}_{[0,+\infty)}(\lambda)$ for λ in a neighborhood of zero. For x in a neighborhood of zero, from (A.4), it can be seen that

$$\Delta^{(2)}(x) = \int_{-\infty}^{x}\frac{1}{m_{\underline{F}}(0)}\Big(-\frac{1}{ym_{\underline{F}}(0)} + \int tdH(t)\Big)d\mathbb{I}_{[0,+\infty)}(\lambda)\lambda.$$

Comparing the two expressions, we can see that for x in a neighborhood of zero,

$$\Delta^{(2)}(x) = \int_{-\infty}^{x}\frac{y}{y-1}\frac{1}{m_{\underline{F}}(0)}\Big(-\frac{1}{ym_{\underline{F}}(0)} + \int tdH(t)\Big)dF(\lambda)$$

Therefore, for x in a neighborhood of zero,

$$\Delta^{(2)}(x) = \int_{-\infty}^{x}\phi(\lambda)dF(\lambda),$$

which proves Theorem 1. □

A.2 Proof of Theorem 2

We begin by providing the proof for the limit of $\alpha^*(\gamma)$. Given $\gamma > 0$, the optimal α is

$$\alpha^*(\gamma) = \frac{\frac{1}{p}\text{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p]\frac{1}{p}\text{tr}\,[\boldsymbol{\Sigma}_p^2] - \frac{1}{p}\text{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p^2]\frac{1}{p}\text{tr}\,[\boldsymbol{\Sigma}_p]}{\frac{1}{p}\text{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-2}\boldsymbol{\Sigma}_p^2]\frac{1}{p}\text{tr}\,[\boldsymbol{\Sigma}_p^2] - \{\frac{1}{p}\text{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p^2]\}^2}$$

It follows from Theorem 2 in Wang, *et al.* (2014) that when $N\to\infty$ and $p/N\to y\in(0,\infty)$,

$$\frac{1}{p}\text{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p] \to \frac{1-\gamma\breve{m}_F(-\gamma)}{1-y(1-\gamma\breve{m}_F(-\gamma))}.$$

Let $B(\gamma) = \{1-\gamma\breve{m}_F(-\gamma)\}/\{1-y(1-\gamma\breve{m}_F(-\gamma))\}$. Since $p^{-1}\text{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p^2]$ is $\Theta_p^{(2)}(-\gamma)$ in the proof of Theorem 1, it can be seen from the equation (A.1) that

$$\begin{aligned}\frac{1}{p}\text{tr}\,[(\boldsymbol{S}_p+\gamma\boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p^2] &\to \frac{-\gamma+\gamma^2\breve{m}_F(-\gamma)}{(1-y(1-y\gamma\breve{m}_F(-\gamma)))^2} + \frac{\int tdH(t)}{1-y(1-y\gamma\breve{m}_F(-\gamma))} \\ &=\Big(\int tdH(t) - \gamma B(\gamma)\Big)\big(1+yB(\gamma)\big).\end{aligned}$$

Since $p^{-1}\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma \boldsymbol{I}_p)^{-2}\boldsymbol{\Sigma}_p^2] = -\frac{d}{d\gamma}p^{-1}\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma \boldsymbol{I}_p)^{-1}\boldsymbol{\Sigma}_p^2]$, the limit of $p^{-1}\mathrm{tr}\,[(\boldsymbol{S}_p + \gamma \boldsymbol{I}_p)^{-2}\boldsymbol{\Sigma}_p^2]$ is

$$\begin{aligned}
&-\frac{d}{d\gamma}\Big(\int tdH(t) - \gamma B(\gamma)\Big)\big(1 + yB(\gamma)\big) \\
&\quad = yB^2(\gamma) + B(\gamma) + 2y\gamma B(\gamma)B'(\gamma) + \Big(\gamma - \int tdH(t)\Big)B'(\gamma).
\end{aligned}$$

For $\beta^*(\gamma)$ and $L_p^*(\gamma)$, we can prove their limits using the same arguments, and the proof of Theorem 2 is complete. □

A.3 Consistent estimator of the stieltjes transform $\breve{m}_F(x)$

This section explains briefly the method of Ledoit and Wolf (2015) for estimating $\breve{m}_F(x)$. Ledoit and Wolf (2015) provided the nonrandom multivariate function called the Quantized Eigenvalues Sampling Transforms or QuEST for short. For any positive integers N and p, the QuEST function denoted by $Q_{N,p}$ is defined as

$$Q_{N,p} : [0,\infty)^p \to [0,\infty)^p \tag{A.5}$$

$$\mathbf{t}_p \equiv (t_1, \cdots, t_p)^{\mathrm{T}} \mapsto Q_{N,p}(\mathbf{t}) \equiv (q^1_{N,p}(\mathbf{t}), \cdots, q^p_{N,p}(\mathbf{t}))^{\mathrm{T}}. \tag{A.6}$$

For $\forall z \in \mathbb{C}^+$, $m \equiv m^{\mathbf{t}}_{N,p}(z)$ is the unique solution in the set $\{m \in \mathbb{C} : -(N-p)/Nz + pm/N \in \mathbb{C}^+\}$ to the equation

$$m = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{t_i\{1-(p/N)-(p/N)zm\} - z}, \tag{A.7}$$

$$\forall x \in \mathbb{R}, \quad F^{\mathbf{t}}_{N,p}(x) \equiv \begin{cases} \max\Big\{1 - \dfrac{p}{N}, \dfrac{1}{p}\displaystyle\sum_{i=1}^{p}\mathbb{I}_{(t_i=0)}\Big\} & \text{if } x = 0, \\ \lim_{\eta \to 0^+}\dfrac{1}{\pi}\displaystyle\int_{-\infty}^{x}\Im[m^{\mathbf{t}}_{N,p}(\xi + i\eta)]d\xi & \text{otherwise,} \end{cases} \tag{A.8}$$

$$\forall u \in [0,1], \quad (F^{\mathbf{t}}_{N,p})^{-1}(u) \equiv \sup\{x \in \mathbb{R} : F^{\mathbf{t}}_{N,p}(x) \le u\}, \tag{A.9}$$

and

$$\forall i = 1, \cdots, p, \quad q^i_{N,p}(\mathbf{t}) \equiv p\int_{(i-1)/p}^{i/p}(F^{\mathbf{t}}_{N,p})^{-1}(u)du. \tag{A.10}$$

It can be seen that equation (A.7) quantizes equation (2.3), and that equation (A.8) quantizes equations (2.1) and (2.2). $F^{\mathbf{t}}_{N,p}$ is the limiting distribution of sample eigenvalues corresponding to the population spectral distribution $p^{-1}\sum_{i=1}^{p}\mathbb{I}_{(t_i=0)}$. By equation (A.9), $(F^{\mathbf{t}}_{N,p})^{-1}$ represents the inverse spectral distribution function, or the quantile function. By equation (A.10), $q^i_{N,p}(\mathbf{t})$

can be interpreted as a smoothed version of the $(i-0.5)/p$ quantile of $F^{\mathbf{t}}_{N,p}$. Ledoit and Wolf (2015) estimates the eigenvalues of the population covariance matrix by

$$\hat{\mathbf{t}}_p = \mathrm{argmin}_{\mathbf{t}\,\in\,[0,\infty)^p} \frac{1}{p}\sum_{i=1}^{p}[q^i_{N,p}(\mathbf{t}) - \ell_i]^2,$$

where $\boldsymbol{\ell}_p = (\ell_1, \ldots, \ell_p)^{\mathrm{T}}$ are eigenvalues of the sample covariance matrix, and showed that

$$\frac{1}{p}\sum_{i=1}^{p}[\hat{t}_i - t_i]^2 \to 0.$$

Lastly, the estimator of $\breve{m}_F(x)$ is obtained as the unique solution $\hat{m} \in \mathbb{R}\cup\mathbb{C}^+$ to the equation

$$m = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{\hat{t}_i\{1-(p/N)-(p/N)xm\} - x}.$$

Acknowledgments.

First of all, we are really grateful to Professor Olivier Ledoit for his kindness of giving us his programs for computation. Research of the second author was supported in part by Grant-in-Aid for Scientific Research (15H01943 and 26330036) from Japan Society for the Promotion of Science.

References

[1] Alon, U., Barkai, N., Motterman, D., Gish, K., Mack, S., and Levine, J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceeding of the National Academy of Science of the United States of America*, **96**, 6745-6750.

[2] Bai, Z., and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica*, **6**, 311-329.

[3] Bai, J., and Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. *Ann. Economics and finance*, **12**, 199-215.

[4] Bodnar, T., Gupta, A.K., and Parolya, N. (2015). Direct shrinkage estimation of large dimensional precision matrix. *J. Multivariate Analysis*, to appear.

[5] Dettling, M., and Buhlmann, P. (2002). Boosting for tumor classification with gene expression data. *Bioinfomatics*, **19**, 1061-1069.

[6] 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.

[7] Himeno, T. and Yamada, T. (2014). Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. *J. Multivariate Analysis*, **130**, 27-44.

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

[9] Ledoit, O., and Péché, S. (2011). Eigenvectors of some large sample covariance matrix. *Prob. Theory Relat. Fields*, **152**, 233-264.

[10] Ledoit, O., and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.*, **40**, 1024-1060.

[11] Ledoit, O., and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivariate Anal.*, **139**, 360-384.

[12] Marčenko, V.A., and Patur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb.*, **1**, 457-486.

[13] Silverstein, J. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.*, **55**, 331-339.

[14] Silverstein, J., and Bai, Z.D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, **54**, 175-192.

[15] Silverstein, J., and Choi, S. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.*, **54**, 295-309.

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

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

[18] Wang, C., Pan, G., Tong, T., and Zhu, L. (2015). Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statistica Sinica*, **25**, 993-1008.

[19] Touloumis, A. (2015). Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional setting. *Comp. Statist. Data Analysis*, **83**, 251-261.