数理解析研究所講究録
第 1999 巻 2016 年 86-94

Estimation of High Dimensional Precision Matrix using Random Matrix Theory

Tsubasa Ito
Graduate School of Economics, University of Tokyo

1 Introduction

About the problem of estimating the high-dimensional covariance matrix, it is well known that we cannot invert the standard sample covariance matrix $\mathbf{S}_p$ when $p > N$, and even if $N > p$ but p/N is relatively large it performs poorly. When we have no advance information about the structure of the population covariance matrix $\mathbf{\Sigma}_p$, shrinking $\mathbf{S}_p$ to some stable statistics improves the performance. There are little research on the direct estimation of the precision matrix and seems to be room for improvement over the estimators, $\mathbf{U}_p\mathbf{A}_p\mathbf{U}_p^{\mathrm{T}}$, $\alpha(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}$, $\alpha\mathbf{S}_p^{-1}+\beta\mathbf{I}_\mathbf{p}$ proposed in recent years. Then we propose $\alpha(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}+\beta\mathbf{I}_\mathbf{p}$.

2 Preliminaries

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, and denote the spaces of real and complex numbers, respectively. Also, $^+$ denotes the half-plane of complex numbers with strictly positive imaginary part. The real and imaginary parts of $z \in$ are denoted by $\Re(z)$ and $\Im(z)$, respectively.

(A1) $p/N \to y \in (0,1)\cup(1,+\infty)$ as $p, N \to +\infty$.

(A2) $\mathbf{\Sigma}_p$ is a non-random p-dimensional positive definite matrix. $\mathbf{X}_p = (\mathbf{x}_{p,1},\ldots,\mathbf{x}_{p,N})^{\mathrm{T}}$ is an $N\times p$ random matrix, where $\mathbf{x}_{p,1},\ldots,\mathbf{x}_{p,N}$ are mutually $i.i.d$ as $E[\mathbf{x}_{p,j}] = \mathbf{0}$ and $\mathbf{Cov}(\mathbf{x}_{p,j}) = \mathbf{I}_p$. $\mathbf{Y}_p = (\mathbf{y}_{p,1},\ldots,\mathbf{y}_{p,N})^{\mathrm{T}}$, where $\mathbf{y}_{p,j} = \mathbf{\Sigma}_p^{1/2}\mathbf{x}_{p,j}$.

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

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

$H_p(t)$ converges to 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.

Let $\mathbf{S}_p = N^{-1}\mathbf{Y}_p^{\mathrm{T}}\mathbf{Y}_p$. $\ell_p = (\ell_{p,1}, \cdots, \ell_{p,p})^{\mathrm{T}}$ and $(\mathbf{u}_1, \ldots, \mathbf{u}_p)$ are a system of eigenvalues sorted in decreasing order and eigenvectors of $\mathbf{S}_p$ The empirical spectral distribution (ESD) of $\mathbf{S}_p$ is defined by

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

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

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

where $\mathbb{C}^+$ denotes the half-plane of complex numbers with strictly positive imaginary part.

The stieltjes transform has the well-known inversion formula

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

if G is continuous at a and b. Stieltjes transform of $\mathbf{F}_p$ is

$$m_{F_p}(z) = \int \frac{1}{\lambda - z} dF_p(\lambda) = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{\ell_i - z} = \frac{1}{p}\mathrm{tr}(\mathbf{S}_p - z\mathbf{I}_p)^{-1}$$

Under (A1)-(A4) and assumption that entries of X_p are independent with common mean and variance and for any $\eta > 0$, as $p/N \to y$

$$\frac{1}{\eta^2 Np}\sum_{jk} E\left[|x_{jk}^{(p)}|^2 I(|x_{jk}^{(p)}| > \eta N^{1/2})\right] \to 0,$$

there exists a distribution function F (limiting spectral distribution (LSD)) such that

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

F is everywhere continuous except at zero, and that the mass of F at zero is

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

Under the same assumptions, $m \equiv m_F(z)$ is the unique solution to the equation (Silverstein (1995))

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

3 Estimation of the precision matrix

We consider the following loss function $\mathrm{L}_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) \equiv \frac{1}{p}\mathrm{tr}(\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \mathbf{I}_p)(\boldsymbol{\Omega}_p\boldsymbol{\Sigma}_p - \mathbf{I}_p)^{\mathrm{T}}$ Instead of minimizing $R(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) \equiv E[L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p)]$, we minimize the limit of $L_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p)$ obtained from RMT. We consider rotation-equivariant estimator.

$$\boldsymbol{\Omega}_p = \mathbf{U}_p\mathbf{A}_p\mathbf{U}_p^{\mathrm{T}} \quad \text{where } \mathbf{A}_p \equiv \mathrm{Diag}(a_1, \cdots, a_p)$$

finite-sample optimal a_i is

$$a_i^* = \frac{\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\mathbf{u}_i}{\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\mathbf{u}_i}$$

Ledoit and Wolf (2012) consider the limit of $\tilde{a}_i = \mathbf{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\mathbf{u}_i$ under $\tilde{L}_p(\boldsymbol{\Sigma}_p^{-1}, \boldsymbol{\Omega}_p) = \frac{1}{p}\mathrm{tr}(\boldsymbol{\Sigma}_p^{-1} - \boldsymbol{\Omega}_p)^2$. $\delta(\ell_i)$, the limit of $\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\mathbf{u}_i$ is , (Ledoit and Peche (2011))

$$\delta(\ell_i) = \begin{cases} \frac{\ell_i}{|1-y-y\ell_i m_F(\ell_i)|^2} & \text{if } \ell_i > 0 \\ \frac{1}{(y-1)m_{\underline{F}}(0)} & \text{if } \ell_i = 0 \text{ and } y > 1 \\ 0 & \text{otherwise} \end{cases}$$

$\phi(\ell_i)$, the limit of $\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\mathbf{u}_i$ is

$$\phi(\ell_i) = \begin{cases} \frac{\ell_i^2\{1-y^2-2y^2\ell_i\Re[m_F(\ell_i)]-y^2\ell_i^2|m_F(\ell_i)|\}}{|(1-y-y\ell_i m_F(x))^2|^2} \\ \quad +\frac{\int_{-\infty}^{\infty} tdH(t)y\ell_i}{|1-y-y\ell_i m_F(\ell_i)|^2} & \text{if } \ell_i > 0 \\ \frac{y}{y-1}\frac{1}{m_{\underline{F}}(0)}(\int_{-\infty}^{\infty} tdH(t) - \frac{1}{ym_{\underline{F}}(0)}) & \text{if } \ell_i = 0 \text{ and } y > 1 \\ 0 & \text{otherwise} \end{cases}$$

$\underline{F}$ is LSD of $\frac{1}{N}\mathbf{Y}_p\mathbf{Y}_p^{\mathrm{T}} = \frac{1}{N}\mathbf{X}_p\boldsymbol{\Sigma}_p\mathbf{X}_p^{\mathrm{T}}$ and $m_{\underline{F}}(z)$ is the solution of $m = -[z - y\int \frac{t}{1+tm}dH(t)]^{-1}$. By replacing $m_F(\ell_i)$ and $m_{\underline{F}}(0)$ with their estimator $\hat{m}_F(\ell_i)$ and $\hat{m}_{\underline{F}}(0)$, we obtain $\hat{\boldsymbol{\Omega}}_p^{LW} = \mathbf{U}_p\hat{\mathbf{A}}_p\mathbf{U}_p^{\mathrm{T}}$ $\hat{a}_i^* = \hat{\delta}(\ell_i)/\hat{\phi}(\ell_i)$. We use a package QuEST on Matlab introduced in Ledoit and Wolf to estimate $\hat{m}_F(\ell_i)$. In this algorithm, we obtain $\hat{\mathbf{t}}_p$, the consistent estimator of eigenvalues of $\boldsymbol{\Sigma}_p$ and solve

$$m = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{\hat{t}_{i,p}(1-(p/N)-(p/N)\ell_i m) - \ell_i}$$

When N, p are relatively small, the approximations of $\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p\mathbf{u}_i$, $\mathbf{u}_i^{\mathrm{T}}\boldsymbol{\Sigma}_p^2\mathbf{u}_i$ by $\hat{\delta}(\ell_i)$, $\hat{\phi}(\ell_i)$ become bad, and $\hat{\boldsymbol{\Omega}}_p^{LW}$ performs poorly. We propose the following estimator of the precision matrix.

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

In the case of $N > p$, consider the following hierarchical bayes model.

$$\begin{aligned}
&\mathbf{V}(= N\mathbf{S}_p) \mid \mathbf{\Sigma}_p \sim \mathcal{W}_p(N, \mathbf{\Sigma}_p)\\
&\mathbf{\Sigma}_p^{-1} \mid \eta \sim (1-\eta)\mathcal{W}_p(k, \mathbf{\Lambda}_1) + \eta\delta_{\mathbf{\Lambda}0}(\mathbf{\Sigma_p^{-1}})\\
&\eta \sim Ber(\theta)
\end{aligned}$$

Denote pdf of $\mathbf{V}$ and prior distribution of $\mathbf{\Sigma}_p^{-1}$ by

$$\begin{aligned}
&\mathbf{V} \mid \mathbf{\Sigma}_p^{-1} \sim f(\mathbf{V} \mid \mathbf{\Sigma}_p^{-1})\\
&\mathbf{\Sigma}_p^{-1} \mid \eta \sim (1-\eta)\pi(\mathbf{\Sigma}_p^{-1} \mid \mathbf{\Lambda}_1) + \eta\delta_{\mathbf{\Lambda}0}(\mathbf{\Sigma}_p^{-1})
\end{aligned}$$

The joint distribution of $(\mathbf{V}, \mathbf{\Sigma}_p^{-1})$ and marginal distribution of $\mathbf{V}$ are

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

$\mathbf{\Omega}_p^{Bayes} = E[\mathbf{\Sigma}_p^{-1} \mid \mathbf{V}]$ is

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

where

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

Let $v_0 = (N+k)/N$,

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

then, we get

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

where $v_0 > 1$, $0 < w_0 < 1$. Letting $\mathbf{\Lambda}_1 = N\gamma\mathbf{I}_p$, $\mathbf{\Lambda}_0 = (1/\bar{\ell})\mathbf{I}_p$, $\bar{\ell} = \sum_{i=1}^p \ell_i/p = \mathrm{tr}[\mathbf{S}_p]/p$, $\alpha = v_0(1-w_0)$, $\beta = w_0/\bar{\ell}$, we obtain

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

We estimate α, β, γ to satisfy $v_0 > 1$, $0 < w_0 < 1$.

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

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

Wang, et.al (2014) shows , for $\gamma > 0$

$$\frac{1}{p}\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}\boldsymbol{\Sigma}_p] \quad a.s. \quad \frac{1-\gamma m_F(-\gamma)}{1-y(1-\gamma m_F(-\gamma))}$$

Wang, et.al (2014) shows this by considering the limit of $F^{\boldsymbol{\Sigma}_p^{-1/2}(\mathbf{S}_p+\gamma\mathbf{I}_p)\boldsymbol{\Sigma}_p^{-1/2}}$. From slide 11, we know

$$\frac{1}{p}\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}\boldsymbol{\Sigma}_p^2] \quad a.s. \quad \frac{-\gamma+\gamma^2 m_F(-\gamma)}{(1-y(1-\gamma m_F(-\gamma)))^2} + \frac{\int t dH(t)}{1-y(1-\gamma m_F(-\gamma))}.$$

Since $p^{-1}\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-2}\boldsymbol{\Sigma}_p^2] = -(d/d\gamma)p^{-1}\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}\boldsymbol{\Sigma}_p^2]$,

$$\frac{1}{p}\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-2}\boldsymbol{\Sigma}_p^2] \;\rightarrow\; \frac{d}{d\gamma}\left\{\frac{-\gamma+\gamma^2 m_F(-\gamma)}{(1-y(1-\gamma m_F(-\gamma)))^2} + \frac{\int t dH(t)}{1-y(1-\gamma m_F(-\gamma))}\right\}$$

We estimate $m_F(-\gamma)$ と $m_F'(-\gamma)$ by $p^{-1}\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}]$, $p^{-1}\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-2}]$. Consistent estimator of $p^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_p) \to \int t dH(t)$ is $p^{-1}\mathrm{tr}(\mathbf{S}_p)$. For $p^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_p^2) \to \int t^2 dH(t)$, $\hat{a}_2 = (N-1)N^{-1}(N-2)^{-1}(N-3)^{-1}p^{-1}[(N-1)(N-2)\mathrm{tr}(\mathbf{S}_p)^2+\{\mathrm{tr}(\mathbf{S}_p)\}^2-NQ]$, where, $Q = (N-1)^{-1}\sum_{i=1}^N\{(\mathbf{y}_i-\bar{\mathbf{y}})^{\mathrm{T}}(\mathbf{y}_i-\bar{\mathbf{y}})\}^2$ is a consistent estimator which proposed by Himeno and Yamada (2014). We look at two estimators: the ridge and the linear shrinkage estimators and check the optimal values of the parameters in these estimators with respect to our loss function.

[1] Ridge estimator. (Wang, et.al (2014))

The ridge estimator is of the form $\Omega_p^{ridge} = \alpha(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}$ Given γ, the optimal α is

$\alpha^{ridge*}(\gamma) = \frac{\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}\mathbf{\Sigma}_p]}{\mathrm{tr}[(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}\mathbf{\Sigma}_p^2(\mathbf{S}_p+\gamma\mathbf{I}_p)^{-1}]}$, which leads to the reduced loss function

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

[2] Linear shrinkage estimator. (Bodnar, et.al (2014))

The linear shrinkage estimator is of the form $\Omega_p^{linear} = \begin{cases} \alpha\mathbf{S}_p^{-1} + \beta\,\mathbf{I}_p & \text{if } N > p \\ \alpha\mathbf{S}_p^{+} + \beta\,\mathbf{I}_p & \text{if } N < p. \end{cases}$ In the case of $N > p$,

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

In the case of $N < p$, Bodnar, (2014) cannot provide estimators for general $\mathbf{\Sigma}_p$, because the limit of $p^{-1}\mathrm{tr}[\mathbf{S}_p^{+}\mathbf{\Sigma}_p^{-1}]$ is needed, which cannot be obtained without assuming a structure such as $\mathbf{\Sigma}_p = \sigma^2\mathbf{I}_p$. Without assuming such a structure, however, we can obtain estimators of the optimal α and β in our situation. The loss function is

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

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

$$\lim_{N,p\to\infty} p^{-1}\mathrm{tr}[(\mathbf{S}_p^{+})^2\mathbf{\Sigma}_p^2] = \lim_{N,p\to\infty} p^{-1}\sum_{i=1}^{N}\frac{\phi(\ell_i)}{\ell_i^2} = \int\frac{\phi(x)}{x^2}d\underline{F}(x)$$
$$\lim_{N,p\to\infty} p^{-1}\mathrm{tr}[\mathbf{S}_p^{+}\mathbf{\Sigma}_p^2] = \lim_{N,p\to\infty} p^{-1}\sum_{i=1}^{N}\frac{\phi(\ell_i)}{\ell_i} = \int\frac{\phi(x)}{x}d\underline{F}(x)$$
$$\lim_{N,p\to\infty} p^{-1}\mathrm{tr}[\mathbf{S}_p^{+}\mathbf{\Sigma}_p] = \frac{1}{y-1}$$

$p^{-1}\sum_{i=1}^{N}\frac{\hat{\phi}(\ell_i)}{\ell_i^2}$ is the estimator of $p^{-1}\mathrm{tr}[(\mathbf{S}_p^{+})^2\mathbf{\Sigma}_p^2]$.

4 Numerical Results

We compare estimators with $\alpha(\mathbf{S}_p + \gamma\mathbf{I}_p)^{-1}$(Wang, et.al (2014)), $\alpha\mathbf{S}_p^{-1} + \beta\mathbf{I}_p$(Bodnar, et.al (2014)). Data is as follows. $\mathbf{y}_i = \mathbf{\Sigma}_p^{1/2}\mathbf{x}_i$

(D1) $x_{ij}\, i.i.d \sim N(0,1),\ i=1,\cdots,N,\ j=1,\cdots,p$
(D2) $x_{ij}=\sqrt{(m-2)/m}z_{ij},\ z_{ij}\, i.i.d \sim t_m,\ i=1,\cdots,N,\ j=1,\cdots,p,\ m=10$
L.S.D of $\boldsymbol{\Sigma}_p$ is based on Beta distribution

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

and the population eigenvalues are generated by

$$1+9H^{-1}_{(a,b)}\Big(\frac{i}{p}-\frac{1}{2p}\Big),\ i=1,\cdots,p.$$

Risk is evaluated by the averaging the empirical losses from 1000 times simulation.

表 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
	30	0.1538	0.1710	0.1665	0.1681	0.1830
	70	0.1703	0.1782	0.1770	0.1813	0.8705
Normal	150	0.1769	0.1854	0.1856	0.1901	0.8081
(a,b)=(1,1)	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
	30	0.1544	0.1704	0.1670	0.1689	0.1878
	70	0.1702	0.1786	0.1787	0.1823	0.8612
t_{10}	150	0.1769	0.1849	0.1869	0.1896	0.8110
(a,b)=(1,1)	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 conduct Quadratic Discriminant Analysis using the microarray data where expression levels for 2000 genes were measured on 22 normal and on 40 colon tumor tissues. Discriminat rule is

$$\frac{N_1}{N_1+1}(\mathbf{x}-\bar{\mathbf{x}}_1)^{\mathrm{T}}\boldsymbol{\Omega}_p^{(1)}(\mathbf{x}-\bar{\mathbf{x}}_1)-\frac{N_2}{N_2+1}(\mathbf{x}-\bar{\mathbf{x}}_2)^{\mathrm{T}}\boldsymbol{\Omega}_p^{(2)}(\mathbf{x}-\bar{\mathbf{x}}_2)<0\Rightarrow\mathbf{x}\in\Pi_1$$

where $\boldsymbol{\Omega}_p^{LR}$, $\boldsymbol{\Omega}_p^{LW}$, $\boldsymbol{\Omega}_p^{ridge}$, $\boldsymbol{\Omega}_p^{linear}$, $\boldsymbol{\Omega}_p^{MP}$, $\boldsymbol{\Omega}_p^{diag}$ are used. Correct classification rates are evaluated by leave-one-out cross-validation.

References

[1] Bodnar, T., Gupta, A. K., and Parolya, N. (2015), Optimal linear shrinkage estimator for large dimensional precision matrix. *J. Multivariate Analysis*, **132**, 215-228.

表 2: Empirical Risks of Ω_p^{oracle}, Ω_p^{LW}, Ω_p^{LR}, Ω_p^{ridge} and Ω_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

表 3: Empirical Risks of Ω_p^{oracle}, Ω_p^{LW}, Ω_p^{LR}, Ω_p^{ridge} and Ω_p^{linear} with $N = 50$ under Normal Distribution

(a, b)	p	oracle	LW	LR	ridge	linear
	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

表 4: 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 %

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

[3] Ledoit, O., and Peche, S. (2015), Spectram estimation: A unified framework for covarianc matrix estimation and PCA in large dimensions. *J. Multivariate Analysis*, **88**, 365-411.

[4] Silverstein, J. W., (1995), Strong convergence of the empirical distribution of the eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, **54**, 295-309.

[5] Wang, C., Pan, G., Tong, T., and Zhu, L. (2014), Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statistica Sinica*, **25**, 993-1008.

Graduate School of Economics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033
JAPAN
E-mail address: `tsubasa.ito710@gmail.com`

東京大学大学院経済学研究科　伊藤　翼