Shrinkage estimation of large dimensional precision matrix using random matrix theory

Cheng Wang[1,3], Guangming Pan[2], Tiejun Tong[3] and Lixing Zhu[3]

[1]*Shanghai Jiao Tong University*
[2]*Nanyang Technological University*
[3]*Hong Kong Baptist University*

Supplementary Material

A1. Proof of Theorem 1

By Corollary A.41 of Bai and Silverstein (2010), we have

$$L^3(F^{\frac{n\lambda}{p}\Sigma_p^{-1}}, F^{\frac{\lambda}{y}\Sigma_p^{-1}}) \le (\frac{n}{p} - \frac{1}{y})^2\lambda^2\frac{1}{p}tr(\Sigma_p^{-2}).$$

Condition S2 implies that

$$\frac{1}{p}tr(\Sigma_p^{-2}) \le C.$$

Noting $p/n \to y$, we have $L^3(F^{\frac{n\lambda}{p}\Sigma_p^{-1}}, F^{\frac{\lambda}{y}\Sigma_p^{-1}}) \to 0$. Therefore, the Stieltjes transform of the LSD of $F^{\frac{n\lambda}{p}\Sigma_p^{-1}}$ is

$$m_H(z) = \int \frac{1}{\frac{\lambda}{yt} - z} dH(t),$$

where $z \in \mathbb{C}^+$. Then by Bai and Silverstein (2010, chap. 4) or the main theorem in Pan (2010), as $n \to \infty$, $F^{\frac{1}{p}\mathbb{Y}\mathbb{Y}^T + \frac{n\lambda}{p}\Sigma_p^{-1}}$ converges almost surely to a non-random distribution F_1, whose Stieltjes transform $m_1(z)$ satisfies

$$m_1(z) = \int \frac{1}{\frac{\lambda}{ty} - z + \frac{1}{y(1+m_1(z))}} dH(t). \tag{1}$$

It is easy to verify that $F^{y(\frac{1}{p}\mathbb{Y}\mathbb{Y}^T + \frac{n\lambda}{p}\Sigma_p^{-1})}$ converges almost surely to a non-random distribution F_2, whose Stieltjes transform is $m_2(z) = \frac{1}{y}m_1(\frac{z}{y})$.

Similarly, by Corollary A.41 of Bai and Silverstein (2010), we can prove that $y(\frac{1}{p}\mathbb{Y}\mathbb{Y}^T + \frac{n\lambda}{p}\Sigma_p^{-1})$ and $\frac{1}{n}\mathbb{Y}\mathbb{Y}^T + \lambda\Sigma_p^{-1}$ have the same LSDs. Here we also use the fact that the support of F_1 or F_2 is bounded by Bai and Silverstein (1998) or the extreme eigenvalues of S_n is bounded by Pan and Zhou (2011).

Altogether, we have that, as $n \to \infty$, $F^{\frac{1}{n}\mathbb{Y}\mathbb{Y}^T+\lambda\Sigma_p^{-1}}$ converges almost surely to a non-random distribution F_2, whose Stieltjes transform $m_2(z)$ satisfies

$$m_2(z) = \int \frac{1}{\frac{\lambda}{t} - z + \frac{1}{1+ym_2(z)}} dH(t). \tag{2}$$

Finally, Theorem A.43 of Bai and Silverstein (2010) yields

$$\begin{aligned} & \|F^{\Sigma_p^{-1/2}(S_n+\lambda I_p)\Sigma_p^{-1/2}} - F^{\frac{1}{n}\mathbb{Y}\mathbb{Y}^T+\lambda\Sigma_p^{-1}}\| \\ \leq & \frac{1}{p} rank(\Sigma_p^{-1/2}(S_n+\lambda I_p)\Sigma_p^{-1/2} - (\frac{1}{n}\mathbb{Y}\mathbb{Y}^T + \lambda\Sigma_p^{-1})) \\ = & \frac{1}{p} rank(\bar{Y}\bar{Y}^T) \leq \frac{1}{p}, \end{aligned}$$

where $\|f\| = \sup_x |f(x)|$. The proof is completed.

A2. Proof of Lemma 1

First, $m_0(-\lambda)$ is the solution of Equation (2.7). Added to this, almost surely,

$$\frac{1}{p} tr(\frac{1}{\lambda} S_n + I_p)^{-1} \to \lambda m_0(-\lambda).$$

Then, $\lambda m_0(-\lambda) \geq \min(0, 1-\frac{1}{y})$. Hence, $1-y+y\lambda m_0(-\lambda) \geq 0$.

Next, suppose we have two solutions M_1, M_2 of Equation (2.7) and $1-y+y\lambda M_j \geq 0,\ j=1,2$. Then

$$\begin{aligned} M_1 &= \int \frac{dH(t)}{t(1-y+y\lambda M_1)+\lambda}, \\ M_2 &= \int \frac{dH(t)}{t(1-y+y\lambda M_2)+\lambda}. \end{aligned}$$

Hence,

$$M_1 - M_2 = (M_2 - M_1) \int \frac{ytdH(t)}{(t(1-y+y\lambda M_1)+\lambda)(t(1-y+y\lambda M_2)+\lambda)}.$$

If $M_1 \neq M_2$, we have

$$-1 = \int \frac{yt\lambda dH(t)}{(t(1-y+y\lambda M_1)+\lambda)(t(1-y+y\lambda M_2)+\lambda)},$$

which is in contradiction with $\frac{yt\lambda}{(t(1-y+y\lambda M_1)+\lambda)(t(1-y+y\lambda M_2)+\lambda)} \geq 0$. Therefore, Equation (2.7) has a unique solution.

A3. Proof of Theorem 2

To proof Theorem 2, we need the following lemma.

Lemma 1 *Under the conditions of Theorem 1, almost surely,*

$$\frac{1}{p}tr(\Sigma_p^{1/2}(S_n+\lambda I_p)^{-1}\Sigma_p^{1/2}) \to R_1(\lambda),$$
$$\frac{1}{p}tr(\Sigma_p^{1/2}(S_n+\lambda I_p)^{-1}\Sigma_p^{1/2})^2 \to R_2(\lambda),$$

where $R_1(\lambda)$ and $R_2(\lambda)$ satisfy

$$R_1(\lambda)=\int\frac{1}{\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)}}dH(t), \tag{3}$$
$$R_2(\lambda)=\int\frac{1+\frac{yR_2(\lambda)}{(1+yR_1(\lambda))^2}}{(\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)})^2}dH(t). \tag{4}$$

Proof: By the definition of ESD and Helly-Bray theorem,

$$\frac{1}{p}tr(\Sigma_p^{1/2}(S_n+\lambda I_p)^{-1}\Sigma_p^{1/2})=\int\frac{1}{x}dF^{\Sigma_p^{-1/2}(S_n+\lambda I_p)\Sigma_p^{-1/2}}(x)\xrightarrow{a.s.}\int\frac{1}{x}dF(x)=\lim_{z\to 0}m(z),$$
$$\frac{1}{p}tr(\Sigma_p^{1/2}(S_n+\lambda I_p)^{-1}\Sigma_p^{1/2})^2=\int\frac{1}{x^2}dF^{\Sigma_p^{-1/2}(S_n+\lambda I_p)\Sigma_p^{-1/2}}(x)\xrightarrow{a.s.}\int\frac{1}{x^2}dF(x)=\lim_{z\to 0}m'(z).$$

That is,

$$R_1(\lambda)=\int\frac{1}{x}dF(x)=\lim_{z\to 0}m(z),$$
$$R_2(\lambda)=\int\frac{1}{x^2}dF(x)=\lim_{z\to 0}m'(z).$$

Equation (2.5) yields

$$m'(z)=\int\frac{1+\frac{ym'(z)}{(1+ym(z))^2}}{(\frac{\lambda}{t}-z+\frac{1}{1+ym(z)})^2}dH(t). \tag{5}$$

For both sides of Equation (2.5) and (5), letting $z\to 0$, we can get

$$R_1(\lambda)=\int\frac{1}{\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)}}dH(t),$$
$$R_2(\lambda)=\int\frac{1+\frac{yR_2(\lambda)}{(1+yR_1(\lambda))^2}}{(\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)})^2}dH(t).$$

This finishes the proof of Lemma 1.

Now we prove Theorem 2. By Lemma 1, we have

$$m_0(-\lambda) = \int \frac{dH(t)}{t(1-y+y\lambda m_0(-\lambda))+\lambda}.$$

In Lemma 1, writing

$$v(\lambda) = \frac{1}{\lambda}(1-\frac{R_1(\lambda)}{1+yR_1(\lambda)}),$$

thus

$$R_1(\lambda) = \frac{1-\lambda v(\lambda)}{1-y(1-\lambda v(\lambda))},$$

and

$$\frac{1-\lambda v(\lambda)}{1-y(1-\lambda v(\lambda))} = \int \frac{1}{\frac{\lambda}{t}+\frac{1}{1+y\frac{1-\lambda v(\lambda)}{1-y(1-\lambda v(\lambda))}}} dH(t).$$

Further, we can show that

$$v(\lambda) = \int \frac{dH(t)}{t(1-y+y\lambda v(\lambda))+\lambda},$$

which is the same as Euqation (2.7). In addition,

$$1-y+y\lambda v(\lambda) = 1-y+y(1-\frac{R_1(\lambda)}{1+yR_1(\lambda)}) = \frac{1}{1+yR_1(\lambda)} \geq 0.$$

Hence, $v(\lambda) = m_0(-\lambda)$ and

$$R_1(\lambda) = \frac{1-\lambda m_0(-\lambda)}{1-y(1-\lambda m_0(-\lambda))}. \tag{6}$$

Further,

$$R_1'(\lambda) = \frac{dR_1(\lambda)}{d\lambda} = -\frac{m_0(-\lambda)-\lambda m_0'(-\lambda)}{(1-y(1-\lambda m_0(-\lambda)))^2}. \tag{7}$$

By the formula of (3),

$$R_1'(\lambda) = -\int \frac{\frac{1}{t}-\frac{yR_1'(\lambda)}{(1+yR_1(\lambda))^2}}{(\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)})^2} dH(t),$$

then

$$\int \frac{1}{(\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)})^2}dH(t)=\frac{(1+yR_1(\lambda))^2(R_1(\lambda)+\lambda R_1'(\lambda))}{1+yR_1(\lambda)+y\lambda R_1'(\lambda)}.$$

By (4), we have

$$\begin{aligned} R_2(\lambda) &= \frac{\int \frac{1}{(\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)})^2}dH(t)}{1-\frac{y}{(1+yR_1(\lambda))^2}\int \frac{1}{(\frac{\lambda}{t}+\frac{1}{1+yR_1(\lambda)})^2}dH(t)} \\ &= (1+yR_1(\lambda))^2(R_1(\lambda)+\lambda R_1'(\lambda)). \end{aligned}$$

Finally, by (6) and (7) we have

$$R_2(\lambda)=\frac{1-\lambda m_0(-\lambda)}{(1-y(1-\lambda m_0(-\lambda)))^3}-\frac{\lambda m_0(-\lambda)-\lambda^2 m_0'(-\lambda)}{(1-y(1-\lambda m_0(-\lambda)))^4}.$$

This finishes the proof of Theorem 2.

A4. Proof of Theorem 3

Lemma 1 implies that

$$\begin{aligned} R_1(\beta) &= \int \frac{1}{\frac{\beta}{t}+\frac{1}{1+yR_1(\beta)}}dH(t), \\ R_2(\beta) &= \int \frac{1+\frac{yR_2(\beta)}{(1+yR_1(\beta))^2}}{(\frac{\beta}{t}+\frac{1}{1+yR_1(\beta)})^2}dH(t). \end{aligned}$$

Moreover, $\frac{1}{1+yR_1(\beta)}=1-y(1-\beta m_0(-\beta))$ where $m_0(-z)$ is the Stieltjes transform of LSD of S_n. Denoting the LSD of S_n as $F^{(0)}(x)$, then

$$m_0(-\beta)=\int \frac{1}{x+\beta}dF^{(0)}(x). \tag{8}$$

Further, if we define another distribution function as

$$F^{(1)}(x)=(1-y)I_{(0,\infty)}(x)+yF^{(0)}(x), \tag{9}$$

and

$$m_1(-\beta)=\int \frac{1}{x+\beta}dF^{(1)}(x), \tag{10}$$

we have $1-y(1-\beta m_0(-\beta))=\beta m_1(-\beta)$. Writing $\gamma=\gamma(\beta)=1/m_1(-\beta)$, we have

$$R_1(\beta)=\frac{\gamma}{\beta}\int\frac{t}{t+\gamma}dH(t),$$
$$R_2(\beta)=\frac{\frac{\gamma^2}{\beta^2}\int(\frac{t}{t+\gamma})^2dH(t)}{1-y\int(\frac{t}{t+\gamma})^2dH(t)}.$$

Therefore,

$$\begin{aligned} L(\beta)=1-\frac{(R_1(\beta))^2}{R_2(\beta)} \\ &= 1-(\int\frac{t}{t+\gamma}dH(t))^2(\frac{1}{\int\frac{t^2}{(t+\gamma)^2}dH(t)}-y) \\ &= L_H(\gamma). \end{aligned}$$

For $\gamma(\beta)$, we have

$$\gamma(\beta)=\frac{1}{\int\frac{1}{x+\beta}dF^{(1)}(x)},$$

which is a strictly increasing function on β. Therefore, γ and β are one-to-one mapping. Specially, when $y\le 1$ that is $F^{(1)}(x)$ has a point mass $1-y$ at the origin, the function $\gamma(\beta):\ (0,\infty)\longmapsto(0,\infty)$. When $y\ge 1$, the function $\gamma(\beta):\ (0,\infty)\longmapsto(\gamma_0,\infty)$ where $\gamma_0\int 1/xdF^{(1)}(x)=1$. Altogether, we have

$$\min_{\beta>0}L(\beta)=\min_{\gamma>0}L_H(\gamma),\ y\le 1,$$
$$\min_{\beta>0}L(\beta)=\min_{\gamma>\gamma_0}L_H(\gamma),\ y\ge 1.$$

When $H(x)$ is a degenerate distribution at σ^2,

$$L_H(\gamma)=y(\frac{\sigma^2}{\sigma^2+\gamma})^2.$$

Obviously, $L_H(\gamma)$ achieves its minimum value $L_0=0$ at $\gamma^*=\infty$. Moreover, $\beta_{\text{opt}}\to\infty$ and

$$\frac{1}{\alpha_{\text{opt}}}=\frac{R_2(\beta_{\text{opt}})}{R_1(\beta_{\text{opt}})}=\lim_{\gamma\to\infty}\frac{\frac{\sigma^4}{(\sigma^2+\gamma)^2}}{\frac{\sigma^2}{\sigma^2+\gamma}}=0,$$
$$\frac{\beta_{\text{opt}}}{\alpha_{\text{opt}}}=\lim_{\gamma\to\infty}\frac{1}{\gamma}\frac{\frac{\sigma^4}{(\sigma^2+\gamma)^2}}{\frac{\sigma^2}{\sigma^2+\gamma}}=\sigma^2,$$

which means the theoretical optimal estimator is $\hat{\Omega}_p = \sigma^{-2} I_p$.

For general distribution $H(x)$, denoting $f_k(x) = \int (\frac{t}{t+x})^k dH(t), k = 1,2,3$, we have

$$\begin{aligned}
\frac{dL_H(x)}{dx} &= \frac{f_1(x)}{(f_2(x))^2}(f_1(x)f_2'(x) - 2(1 - yf_2(x))f_1'(x)f_2(x)) \\
&= \frac{2f_1(x)}{x(f_2(x))^2}(f_1(x)(f_3(x) - f_2(x)) - (1 - yf_2(x))(f_2(x) - f_1(x))f_2(x)) \\
&= \frac{2f_1(x)(f_2(x) - f_1(x))}{x}(y - \frac{f_1(x)f_3(x) - f_2(x)f_2(x)}{f_2(x)f_2(x)(f_1(x) - f_2(x))}),
\end{aligned}$$

where we use the facts that $f_k'(x) = -k\int \frac{t^k}{(t+x)^{k+1}} dH(t)$ and $xf_k'(x) = k(f_{k+1}(x) - f_k(x))$.

Writing $g(x) = \frac{f_1(x)f_3(x) - f_2(x)f_2(x)}{f_2(x)f_2(x)(f_1(x) - f_2(x))}$, it is easy to show $\lim_{x\to 0^+} g(x) = 0$ and $\lim_{x\to+\infty} g(x) = +\infty$. Therefore, $L_H(\gamma)$ can achieve its global minimum value at γ^* which satisfies

$$\frac{f_1(\gamma^*)f_3(\gamma^*) - f_2(\gamma^*)f_2(\gamma^*)}{f_2(\gamma^*)f_2(\gamma^*)(f_1(\gamma^*) - f_2(\gamma^*))} = y.$$

Thus, by the definition of $\gamma(\beta)$, when $y \leq 1$, β_{opt} satisfies the equation $\gamma^* = \frac{\beta_{\text{opt}}}{1 - y(1 - \beta_{\text{opt}} m_0(-\beta_{\text{opt}}))}$. The proof is finished.

A5. Proof of Theorem 4

By Theorem 3, almost surely, as $n \to \infty$,

$$\begin{aligned}
\hat{R}_1(\lambda) &\to R_1(\lambda), \\
\hat{R}_2(\lambda) &\to R_2(\lambda).
\end{aligned}$$

By the continuous mapping theorem, almost surely, we have

$$L_n(\lambda) = 1 - \frac{(\hat{R}_1(\lambda))^2}{\hat{R}_2(\lambda)} \to L(\lambda).$$

By the definition of β_n^*, we have

$$L_n(\beta_n^*) \leq L_n(\beta_{\text{opt}}) \xrightarrow{a.s.} L(\beta_{\text{opt}}) = L_0. \tag{11}$$

Noting that $\hat{R}_k, k = 1,2$ are decreasing functions, it is straightforward to show that $\hat{R}_1, \hat{R}_2$ are uniformly convergent on the bounded interval $[C_1, C_2]$. That is, for any $\varepsilon > 0$, when n is large enough, for all $\beta \in [C_1, C_2]$, we have,

$$\begin{aligned}
|\hat{R}_1(\beta) - R_1(\beta)| &\leq \varepsilon, \ a.s. \\
|\hat{R}_2(\beta) - R_2(\beta)| &\leq \varepsilon, \ a.s.
\end{aligned}$$

which can guarantee the uniformly convergence of $L_n(\beta)$. Therefore, we can claim for any $\varepsilon > 0$, when n is large enough, almost surely,

$$|L_n(\beta) - L(\beta)| \leq \varepsilon, \quad \text{for any } \beta \in [C_1, C_2]. \tag{12}$$

Specially, we have, almost surely,

$$L_n(\beta_n^*) \geq L(\beta_n^*) - \varepsilon \geq L_0 - \varepsilon. \tag{13}$$

Together with (11), we get $L_n(\beta_n^*) \xrightarrow{a.s.} L_0$.

Similarly, denoting

$$R_{1n}(\beta) = \frac{1}{p} tr(\Sigma_p^{1/2}(S_n + \beta I_p)^{-1}\Sigma_p^{1/2}),$$
$$R_{2n}(\beta) = \frac{1}{p} tr(\Sigma_p^{1/2}(S_n + \lambda I_p)^{-1}\Sigma_p^{1/2})^2,$$

we have, for any $\varepsilon > 0$, when n is large enough, for all $\beta \in [C_1, C_2]$,

$$|R_{1n}(\beta) - R_1(\beta)| \leq \varepsilon, \ a.s.$$
$$|R_{2n}(\beta) - R_2(\beta)| \leq \varepsilon, \ a.s,$$

and

$$|\hat{R}_1(\beta) - R_{1n}(\beta)| \leq 2\varepsilon, \ a.s.$$
$$|\hat{R}_2(\beta) - R_{2n}(\beta)| \leq 2\varepsilon, \ a.s..$$

Then, we have

$$\begin{aligned} \frac{1}{p} tr(\alpha_n^*(S_n + \beta_n^* I_p)^{-1}\Sigma_p - I_p)^2 &= (\alpha_n^*)^2 R_{2n}(\beta_n^*) - 2\alpha_n^* R_{1n}(\beta_n^*) + 1 \\ &= R_{2n}(\beta_n^*)(\alpha_n^* - \frac{R_{1n}(\beta_n^*)}{R_{2n}(\beta_n^*)})^2 + 1 - \frac{R_{1n}^2(\beta_n^*)}{R_{2n}(\beta_n^*)} \\ &= R_{2n}(\beta_n^*)(\frac{\hat{R}_1(\beta_n^*)}{\hat{R}_2(\beta_n^*)} - \frac{R_{1n}(\beta_n^*)}{R_{2n}(\beta_n^*)})^2 + 1 - \frac{R_{1n}^2(\beta_n^*)}{R_{2n}(\beta_n^*)}. \end{aligned}$$

Moreover, for any $\beta \in [C_1, C_2]$, almost surely, $R_1(\beta)$ and $R_2(\beta)$ are bounded which ensure there are constants C_3, C_4 that $0 < C_3 < \hat{R}_1(\beta), \hat{R}_1(\beta), R_{1n}(\beta), R_{2n}(\beta) < C_4$. Then, almost surely,

$$|\frac{1}{p} tr(\alpha_n^*(S_n + \beta_n^* I_p)^{-1}\Sigma_p - I_p)^2 - L_n(\beta_n^*)| < C_0\varepsilon. \tag{14}$$

Together with $L_n(\beta_n^*) \xrightarrow{a.s.} L_0$, we have

$$\frac{1}{p} tr(\alpha_n^*(S_n + \beta_n^* I_p)^{-1}\Sigma_p - I_p)^2 \xrightarrow{a.s.} L_0. \quad (15)$$

The proof is completed.

References

Bai, Z. and Silverstein, J. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of Probability* **26**, 316–345.

Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer, New York.

Pan, G. (2010). Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix. *Journal of Multivariate Analysis* **101**, 1330–1338.

Pan, G. and Zhou, W. (2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *Annals of Applied Probability* **21**, 1860–1910.

Cheng Wang

Department of Mathematics, Shanghai Jiao Tong University, Shanghai, China

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

E-mail: cescwang@gmail.com

Guangming Pan

School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

E-mail: gmpan@ntu.edu.sg

Tiejun Tong

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

E-mail: tongt@hkbu.edu.hk

Lixing Zhu

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong
E-mail: lzhu@hkbu.edu.hk