Statistica Sinica

Supplementary Materials to "Order Determination for Spiked Type Models"

Yicheng Zeng and Lixing Zhu

Department of Mathematics, Hong Kong Baptist University, Hong Kong

Center for Statistics and Data Science, Beijing Normal University at Zhuhai, China

Proof of Lemma 1. Firstly, we check the requirement (i). As f_n is differentiable, $\exists \xi_q \in (\hat{\lambda}_{q+1}/\sigma^2, \hat{\lambda}_q/\sigma^2),\ s.t.$

$$\hat{\delta}_q^* = f_n(\hat{\lambda}_q/\sigma^2) - f_n(\hat{\lambda}_{q+1}/\sigma^2) = f_n'(\xi_q)\hat{\delta}_q. \tag{1.1}$$

We then only need to check that

$$\mathbb{P}\{f_n'(\xi_q) \geq 1\} \longrightarrow 1, \quad as\ n \to \infty. \tag{1.2}$$

By conditions (a) and (b), it suffices to show that

$$\mathbb{P}\{\xi_q > b - \kappa_n\} \longrightarrow 1, \quad as\ n \to \infty. \tag{1.3}$$

On the other hand, from the definition of κ_n in condition (b), we have

$$\hat{\lambda}_{q+1}/\sigma^2 - b = o_P(\kappa_n), \tag{1.4}$$

which is equivalent to

$$\frac{\hat{\lambda}_{q+1}/\sigma^2 - b}{\kappa_n} = o_P(1). \tag{1.5}$$

Then we have

$$\mathbb{P}\{\xi_q > b - \kappa_n\} \geq \mathbb{P}\left\{\frac{\hat{\lambda}_{q+1}}{\sigma^2} > b - \kappa_n\right\} = \mathbb{P}\left\{\frac{\hat{\lambda}_{q+1}/\sigma^2 - b}{\kappa_n} > -1\right\} \to 1 \tag{1.6}$$

(i) is then verified.

Now we check (ii). Similarly, we have

$$\hat{\delta}_i^* = f'(\xi_i)\hat{\delta}_i/\sigma^2, \quad \text{for } q+1 \leq i \leq p-2, \tag{1.7}$$

where $\xi_i \in (\hat{\lambda}_{i+1}/\sigma^2, \hat{\lambda}_i/\sigma^2)$. Then it suffices to show that

$$\mathbb{P}\{\xi_i < b + \kappa_n\} \longrightarrow 1, \quad \text{for } q+1 \leq i \leq p-2. \tag{1.8}$$

Since $\xi_{q+1} > \cdots > \xi_{p-2}$, it is equivalent to

$$\mathbb{P}\{\xi_{q+1} < b + \kappa_n\} \longrightarrow 1, \tag{1.9}$$

whose proof is completely parallel to that of (i).

For (iii), we have

$$\frac{\hat{\delta}_{q+1}^*}{\hat{\delta}_q^*} = \frac{f_n'(\xi_{q+1})\hat{\delta}_{q+1}}{f_n'(\xi_q)\hat{\delta}_q} \tag{1.10}$$

Condition (b) yields

$$f_n'(\xi_{q+1}) \leq f_n'(\xi_q), \tag{1.11}$$

since $\xi_{q+1} < \hat{\lambda}_{q+1}/\sigma^2 < \xi_q$. Therefore,

$$\frac{\hat{\delta}_{q+1}^*}{\hat{\delta}_q^*} \leq \frac{\hat{\delta}_{q+1}}{\hat{\delta}_q}. \tag{1.12}$$

The requirement (iii) is then proved and the proof of the lemma is finished.

$\square$

Proof of Theorem 2. We only need to check that

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\hat{\delta}^*_{i+1}+c_n}{\hat{\delta}^*_i+c_n} > \tau\right\} = 1, \quad \text{for } q < i \le L-2 \tag{1.13}$$

and

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\hat{\delta}^*_{q+1}+c_n}{\hat{\delta}^*_q+c_n} \le \tau\right\} = 1. \tag{1.14}$$

On one hand, since Lemma 1 ensures the requirement (ii), for $q < i \le L-2$,

$$\hat{\delta}^*_i = o_p(c_n), \tag{1.15}$$

which leads to $\hat{\delta}^*_i c_n^{-1} = o_p(1)$. Then

$$\frac{\hat{\delta}^*_{i+1}+c_n}{\hat{\delta}^*_i+c_n} = \frac{\hat{\delta}^*_{i+1}c_n^{-1}+1}{\hat{\delta}^*_i c_n^{-1}+1} = \frac{o_p(1)+1}{o_p(1)+1} \xrightarrow{P} 1 > \tau. \tag{1.16}$$

That is,

$$\lim_{n\to+\infty} \mathbb{P}\left\{\frac{\hat{\delta}^*_{i+1}+c_n}{\hat{\delta}^*_i+c_n} > \tau\right\} = 1, \quad \text{for } q < i \le L-2. \tag{1.17}$$

On the other hand, because of (i), (ii) and

$$\lim_{n\to+\infty} \mathbb{P}\left\{\frac{\hat{\delta}_{q+1}/\sigma^2+c_n}{\hat{\delta}_q/\sigma^2+c_n} \le \tau\right\} \longrightarrow 1, \tag{1.18}$$

we have

$$\lim_{n\to+\infty} \mathbb{P}\left\{\frac{\hat{\delta}^*_{q+1}+c_n}{\hat{\delta}^*_q+c_n} \le \tau\right\} \longrightarrow 1. \tag{1.19}$$

Thus, $\hat{q}_n^{TVACLE}$ is equal to q with a probability going to 1. The proof is concluded. $\square$

Proof of Proposition 3 and Theorem 3. When $\sigma^2 = 1$, Wang and Yao (2017) provided the limiting spectral distribution (LSD) of the matrix $\mathbf{F}_n = \mathbf{S}_1\mathbf{S}_2^{-1}$ and established the phase transition phenomenon for those extreme eigenvalues of $\mathbf{F}_n$. When $0 < c \leq 1$, the empirical spectral distribution (ESD) of $\mathbf{F}_n$ weakly converges to a distribution $F_{c,y}$ with the density function

$$f_{c,y}(x) = \frac{(1-y)\sqrt{(b_1 - x)(x - a_1)}}{2\pi x(c + xy)}, \quad a_2 \leq x \leq b_2, \tag{1.20}$$

where $a_2 = (\frac{1-\sqrt{c+y-cy}}{1-y})^2$ and $b_2 = (\frac{1+\sqrt{c+y-cy}}{1-y})^2$. Similarly as that of spiked population models, when $c > 1$, there is an additional probability measure of mass $1 - \frac{1}{c}$ for $F_{c,y}$. Further, they also proved a phase transition phenomenon that almost surely

$$\begin{aligned} \hat{\lambda}_i &\to \varphi(\lambda_i), \quad \lambda_i > \gamma(1 + \sqrt{c + y - cy}), \\ \hat{\lambda}_i &\to b_2, \quad 1 < \lambda_i \leq \gamma(1 + \sqrt{c + y - cy}), \end{aligned}$$

where $\gamma = \frac{1}{1-y} \in (1, +\infty)$ and $\varphi(x) = \frac{\gamma x(x-1+c)}{x-\gamma}, \ x \neq \gamma$.

Under the general Fisher matrix with the spiked structure

$$\mathrm{spec}(\Sigma_1\Sigma_2^{-1}) = \{\lambda_1, \lambda_2, \cdots, \lambda_{q_1}, \sigma^2, \cdots, \sigma^2\}. \tag{1.21}$$

Using the simple transformation $\hat{\lambda}_i \rightarrowtail (\sigma^2)^{-1}\hat{\lambda}_i$, we can similarly achieve the results in the case of $\sigma^2 = 1$. The empirical spectral distribution of $\mathbf{F}_n$ weakly converges to a distribution F_{c,y,σ^2} with the density function

$$f_{c,y,\sigma^2}(x) = \frac{1}{\sigma^2} f_{c,y}\left(\frac{x}{\sigma^2}\right), \quad \sigma^2 a_1 < x < \sigma^2 b_1, \tag{1.22}$$

and the additional point mass $1 - \frac{1}{c}$ at origin $x = 0$ also exists when $c > 1$. The phase transition phenomenon is modified as

$$\begin{aligned} \hat{\lambda}_i &\to \sigma^2 \varphi(\lambda_i/\sigma^2), \quad \lambda_i > \sigma^2\gamma(1 + \sqrt{c + y - cy}), \\ \hat{\lambda}_i &\to \sigma^2 b_2, \quad \sigma^2 < \lambda_i \le \sigma^2\gamma(1 + \sqrt{c + y - cy}), \end{aligned}$$

where the parameters b_2, γ and the function φ have the same definitions as those in the case with $\sigma^2 = 1$.

Recall that $q := \#\{\lambda_i : \lambda_i > \sigma^2\gamma(1+\sqrt{c + y - cy})\}$. According to these results, for any fixed L with $q + 3 < L < p$

$$\begin{aligned} \hat{\lambda}_i &\to \sigma^2 \varphi(\lambda_i/\sigma^2), \quad 1 \le i \le q, \\ \hat{\lambda}_i &\to \sigma^2 b_2, \quad q + 1 \le i \le L. \end{aligned} \tag{1.23}$$

That is, when i is larger than q, the estimated eigenvalue $\hat{\lambda}_i$ converges to the right edge $\sigma^2 b_2$ of the support of F_{c,y,σ^2}. This means that any eigenvalues such that $\sigma^2 < \lambda_i \le \sigma^2\gamma(1 + \sqrt{c + y - cy})$ cannot be identified through the estimated eigenvalues and then show the optimality of this lower bound. Thus, the Proposition 3 has been proved.

Modifying the result of Wang and Yao (2017), we can show that those extreme eigenvalues $\hat{\lambda}_i$ corresponding to $\lambda_i > \sigma^2\gamma(1+\sqrt{c+y-cy})$ satisfy Central Limiting Theorem and thus have the convergence rate of order $1/\sqrt{n}$. For the fluctuation of those eigenvalues which stick to the bulk, Han et al. (2016) showed that $n^{2/3}(\hat{\lambda}_{q+1}-\sigma^2 b_2)$ is asymptotically Tracy-Widom distributed. Han et al. (2018) established an asymptotic joint distribution for $(n^{2/3}(\hat{\lambda}_{q+1}-\sigma^2 b_2), n^{2/3}(\hat{\lambda}_{q+2}-\sigma^2 b_2), \cdots, n^{2/3}(\hat{\lambda}_{q+k}-\sigma^2 b_2))$ for any fixed k. Thus, for any fixed $L > q$, $n^{2/3}(\hat{\lambda}_i - \sigma^2 b_2) = O_p(n^{-2/3})$ for $q+1 \le i \le L$.

We omit the remainder of the proof, since it is exactly the same with that of spiked population models. $\square$

Proof of Proposition 1. Let $\Sigma_y = \mathrm{Cov}(y_t, y_{t-1})$ be the lag-1 auto-covariance matrices of y_t and $\Sigma_x = \mathrm{Cov}(x_t, x_{t-1})$ the lag-1 auto-covariance matrix of x_t. As shown in Li et al. (2017), the sample auto-covariance matrix of y_t is

$$\begin{aligned}
\hat{\Sigma}_y &= \frac{1}{T}\sum_{t=2}^{T+1} y_t y_{t-1}' = \frac{1}{T}\sum_{t=2}^{T+1}(\mathbf{A}x_t+\varepsilon_t)(\mathbf{A}x_{t-1}+\varepsilon_{t-1})' \\
&= \frac{1}{T}\sum_{t=2}^{T+1}\mathbf{A}x_t x_{t-1}'\mathbf{A}' + \frac{1}{T}\sum_{t=2}^{T+1}(\mathbf{A}x_t\varepsilon_{t-1}' + \varepsilon_t x_{t-1}'\mathbf{A}') + \frac{1}{T}\sum_{t=2}^{T+1}\varepsilon_t\varepsilon_{t-1}' \\
&:= \mathbf{P_A} + \hat{\Sigma}_\varepsilon, \qquad (1.24)
\end{aligned}$$

where the matrix $\hat{\Sigma}_\varepsilon = \frac{1}{T}\sum_{t=2}^{T+1}\varepsilon_t\varepsilon_{t-1}'$ is the sample auto-covariance matrix of noise sequence $\{\varepsilon_t\}$. Notice that the matrix $\mathbf{P_A}$ is of finite rank, then the matrix $\hat{\Sigma}_y$ can be viewed as a finite-rank perturbation of $\hat{\Sigma}_\varepsilon$. Since both $\hat{\Sigma}_\varepsilon$

and $\hat{\Sigma}_y$ are asymmetric matrices, Li et al. (2017) considered their singular values. This is equivalent to considering the square root of the eigenvalues of the matrices $\hat{\mathbf{M}}_\varepsilon := \hat{\Sigma}_\varepsilon \hat{\Sigma}_\varepsilon'$ and $\hat{\mathbf{M}}_y := \hat{\Sigma}_y \hat{\Sigma}_y'$, respectively.

Define $\hat{\Sigma}_y/\sigma^2 = \mathbf{P}_\mathbf{A}/\sigma^2 + \hat{\Sigma}_\varepsilon/\sigma^2$, we can reduce the problem to the case with $\sigma^2 = 1$. When $p/T \to y > 0$, Li et al. (2015) proved that the empirical spectral distribution of $\hat{\mathbf{M}}_\varepsilon$ almost surely converges to a non-random limiting distribution, whose Stieltjes transformation $\mathcal{S}(z)$ defined in (4.10) satisfies the equation

$$z^2\mathcal{S}^3(z) - 2z(y-1)\mathcal{S}^2(z) + (y-1)^2\mathcal{S}(z) - z\mathcal{S}(z) - 1 = 0.$$

This limiting spectral distribution is continuous with a compact support $[a_1\mathbf{1}_{\{y\geq 1\}}, b_1]$, where

$$a_1 = (-1 + 20y + 8y^2 - (1+8y)^{3/2})/8,$$

$$b_1 = (-1 + 20y + 8y^2 + (1+8y)^{3/2})/8$$

From Wang and Yao (2016), the largest eigenvalue $\hat{\lambda}_{\varepsilon,1}$ of $\hat{\mathbf{M}}_\varepsilon$ almost surely converges to the right edge b_1. Like the previous models, for any fixed $L > q_0 + 1$, and any $1 \leq i \leq L$ the largest eigenvalues $\hat{\lambda}_{\varepsilon,i}$ of $\hat{\mathbf{M}}_\varepsilon$ converge to the same value b_1. Further, for general σ^2, the result of Li et al. (2017) implies that the limiting spectral distribution of the perturbed matrix $\hat{\mathbf{M}}_y$ is identical to that of $\hat{\mathbf{M}}_\varepsilon$. They also built a phase transition phenomenon for those extreme eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_q$. The following proposition confirms the optimality of the bound restriction $\mathcal{T}_1(i) < \mathcal{T}(b_1+)$ such that the corresponding q factors in $\mathbf{P}_\mathbf{A}$ can be identified.

Lemma A. (Li et al. (2017)) *Denote $\mathcal{T}(\cdot)$ as the T-transformation of the Limiting Spectral Distribution (LSD) for matrix $\hat{\mathbf{M}}_y/\sigma^4$. Suppose that the model (3.17) satisfies Assumptions 3.1-3.3, $\{\varepsilon_t\}$ are normally distributed and the loading matrix $\mathbf{A}$ is standardized as $\mathbf{A}'\mathbf{A} = \mathbf{I}_k$. Let $\hat{\lambda}_i,\ 1 \leq i \leq q_0$ denote the q_0 largest eigenvalues of $\hat{\mathbf{M}}_y$. Then for each $1 \leq i \leq q_0$, $\hat{\lambda}_i/\sigma^4$ converges almost surely to a limit β_i. Moreover,*

$$\beta_i > b_1 \text{ when } \mathcal{T}_1(i) < \mathcal{T}(b_1+),$$

and

$$\beta_i = b_1 \text{ when } \mathcal{T}_1(i) \geq \mathcal{T}(b_1+)$$

where

$$\mathcal{T}_1(i) = \frac{2y\sigma^2\gamma_0(i) + \gamma_1(i)^2 - \sqrt{(2y\sigma^2\gamma_0(i) + \gamma_1(i)^2)^2 - 4y^2\sigma^4(\gamma_0(i)^2 - \gamma_1(i))^2}}{2\gamma_0(i)^2 - 2\gamma_1(i)^2}.$$

From this lemma, we can see that the bound for the number of common factors determined by the constraint $\mathcal{T}_1(i) < \mathcal{T}(b_1+)$ is optimal. That is, only q common factors in $\mathbf{P}_\mathbf{A}$ can be well separated from the noise ε_t's theoretically. This is because $\hat{\lambda}_{q+1}$ will converge to b_1 and thus cannot be well separated from those large estimated eigenvalues of $\hat{\Sigma}_\varepsilon$ that tend to the right edge b_1 as well. □

A justification of Proposition 2. By the results of Wang and Yao (2016), the phase transitions hold. Further, under the assumption that the estimated eigenvalues $\hat{\lambda}_i$ for $i > q$ have the convergence rate of order $O_p(n^{-2/3})$, the results hold by following the arguments used in spiked population models. □

References

Han, X., Pan, G. and Zhang, B. (2016). The Tracy–Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44**, 1564–1592.

Han, X., Pan, G. and Yang, Q. (2018). A unified matrix model including both CCA and F matrices in multivariate analysis: the largest eigenvalue and its applications. *Bernoulli* **24**, 3447–3468.

Li, Z., Pan, G. and Yao, J. (2015). On singular value distribution of large-dimensional autocovariance matrices. *J. Multivariate Anal.* **137**, 119–140.

Li, Z., Wang, Q. and Yao, J. (2017). Identifying the number of factors from singular values of a large sample auto-covariance matrix. *Ann. Statist.* **45**, 257-288.

Wang, Q. and Yao, J. (2016). Moment approach for singular values distribution of a large auto-covariance matrix. *Ann. Inst. H. Poincaré Probab. Statist.* **52**, 1641–1666.

Wang, Q. and Yao, J. (2017). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *Ann. Statist.* **45**, 415-460.

Yicheng Zeng, Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, China

E-mail: statzyc@gmail.com

Lixing Zhu, Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, China

Center for Statistics and Data Science, Beijing Normal University at Zhuhai, Zhuhai, Guangdong, China

E-mail: lzhu@hkbu.edu.hk