Statistica Sinica Preprint No: SS-2020-0089	
Title	Order Determination for Spiked Type Models
Manuscript ID	SS-2020-0089
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0089
Complete List of Authors	Yicheng Zeng and Lixing Zhu
Corresponding Author	Lixing Zhu
E-mail	lzhu@hkbu.edu.hk
Notice: Accepted version subject to English editing.	

Statistica Sinica

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, China

Abstract: Motivated by dimension reduction in regression analysis and signal detection, we investigate the order determination for large dimensional matrices including spiked type models of which the numbers of covariates are proportional to the sample sizes for different models. Because the asymptotic behaviors of the estimated eigenvalues of the corresponding matrices differ completely from those in fixed dimension scenarios, we then discuss the largest possible number we can identify and introduce a "valley-cliff" criterion. We propose two versions of the criterion: one based on the original differences between eigenvalues and the other based on the transformed differences, which reduces the effect of ridge selection in the former one. This generic method is very easy to implement and computationally inexpensive, and can be applied to various matrices. As examples, we focus on spiked population models, spiked Fisher matrices and factor models with auto-covariance matrices. Numerical studies are conducted to examine the finite sample performances of the method and to compare it with existing methods.

Key words and phrases: Auto-covariance matrix, factor model, finite-rank perturbation, Fisher matrix, phase transition, ridge ratio, spiked population model.

1. Introduction

In many statistical methods, we need to determine how many eigenvalues of a matrix are important to estimations. This problem is called order determination. Examples include spiked population models proposed by Johnstone (2001); spiked Fisher matrices, which are motivated from signal detection and hypothesis testing for covariances; canonical correlation analysis; factor models; and some matrices in sufficient dimension reduction (see Li (1991); Zhu et al. (2010)). Luo and Li (2016) is useful literature on order determination and proposed a ladle estimation for several models. In this paper, we use spiked population models to set up the problem of interest and then introduce an estimation criterion that applies to more models including spiked Fisher matrices. The method is also applicable to sample auto-covariance matrices though they cannot be written as a spiked matrix at the population level. Therefore, we call them spiked type models.

The literature includes several proposals in the fixed dimension cases, such as the classic Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Some methods developed for sufficient dimension reduction can also be used in spiked type models. These include the sequential testing method (Li (1991)), the BIC-type criterion (Zhu et al. (2006)), ridge ratio estimation (Xia et al. (2015)) and ladle estimation (Luo and Li

(2016)), some of which can even handle cases with divergent dimension in the sense that $p/n \to 0$ at certain rate as $n \to \infty$. Here n denotes the sample size and p the dimension of the matrix.

However, when the dimension p is proportional to the sample size n such that $p/n \to c$ for some constant $c > 0$, the order determination becomes much more challenging. Some efforts have been devoted to this problem by using the large dimensional random matrix theory (see for example Kritchman and Nadler (2008); Onatski (2009)). Consider spiked population models. When $p/n \to c$, using the results derived by Baik and Silverstein (2006), Passemier and Yao (2012) introduced a criterion that counts the number of differences between consecutive eigenvalues above some predetermined threshold. However, if there exist equal spikes, the corresponding differences could be smaller than the designed threshold, then the criterion could easily define an estimator smaller than the true number. Passemier and Yao (2014) improved this method to accommodate cases with multiple spikes. The underestimation issue however remains when there are three or more equal spikes. Besides the spike multiplicity, the dominating effect by several largest eigenvalues also results in underestimation. That is, when a couple of eigenvalues are very large, those relatively small spikes would be ignored. For the number of factors in a factor model for high-dimensional

time series, Li et al. (2017) proposed a criterion similar to that in Passemier and Yao (2014). For spiked Fisher matrices, Wang and Yao (2017) used the classical scree plot to determine the number of spikes when a threshold is selected in a delicate manner. The underestimation is still an issue, which will be demonstrated in the numerical studies below. Relevant references include Lam and Yao (2012) and Xia et al. (2015).

Benefitting from existing asymptotic results for the estimated eigenvalues of large-dimensional random matrices in the literature, we introduce a novel and generic criterion in the high dimensional regime with $p/n \to c$. Our criterion relies on the eigenvalue difference-based ridge ratios with the following features. First, it can handle multiple spikes cases and alleviates the large eigenvalues dominance problem. Second, it has a nice "valley-cliff" pattern such that the consistent estimator is at the "valley bottom" facing the "cliff" upon which all the next ratios take the same values asymptotically and can then exceed a threshold. Third, adding ridge values is essential to make the ratios stable and create the "valley-cliff" pattern. Fourth, to reduce the sensitivity of the criterion to ridge selection, we suggest another version that uses transformed eigenvalues. Fifth, we discuss how to reduce the effect of model scale in the construction as well. As the new method completely avoids optimality procedure and thus is computationally effi-

cient.

The remainder of this paper is organized as follows. In Section 2, we focus on population spiked models, propose a VAlley-CLiff Estimation (VACLE) and provide an optimal lower bound to show what order can be identified. The VACLE is then improved when we use a transformation-based valley-cliff estimation (TVACLE) to alleviate the criterion's sensitivity to the designed ridge value. We also discuss the methods to select the transformation. In Section 3, we implement our method on two more examples: factor models with auto-covariance matrices and spiked Fisher matrices. We discuss its applicability in more general cases as well. Section 4 contains numerical studies and compares the VACLE and the TVACLE with some competitors. The analysis for a real data example is included in Section 5. Some concluding remarks are in Section 6, and the proofs of the theoretical results are included in the supplementary materials.

2. Order determination for population spiked models

In this section, we develop our method for the population spiked models introduced below with some important results for the estimated eigenvalues.

2.1 Spiked population models

Assume that a $p \times p$ non-negative definite matrix $\Sigma_p = \sigma^2 \mathbf{I}_p + \Delta_p$ has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{q_1} > \lambda_{q_1+1} = \cdots = \lambda_p = \sigma^2$ where q_1 is a

fixed number and the scale parameter σ^2 is either known or unknown. Let $Z \equiv (z_{ji})_{1\le j\le p, 1\le i\le n} \equiv (z_1, \cdots, z_n) \in \mathbb{R}^{p\times n}$ have i.i.d. entries each having zero mean and unit variance. Taking $x_i := \Sigma_p^{1/2} z_i$, Σ_p is the population covariance matrix of x_i and coincides with the spiked population model introduced in Johnstone (2001):

$$\text{spec}(\Sigma_p) = \{\lambda_1, \cdots, \lambda_{q_1}, \sigma^2, \cdots, \sigma^2\}, \tag{2.1}$$

where the eigenvalues λ_i, $1 \le i \le q_1$, are called spikes. Denote the corresponding sample covariance matrix by $\mathbf{S}_n := n^{-1}\sum x_i x_i^\top = n^{-1}\sum \Sigma_p^{1/2} z_i z_i^\top \Sigma_p^{1/2}$ and its eigenvalues by $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_p$, which can also be motivated from the signal detection problem (see, e.g. Nadler (2010)):

$$x_i = \mathbf{A} u_i + \varepsilon_i, \ 1 \le i \le n, \tag{2.2}$$

where $u_i \in \mathbb{R}^{q_1}$ is a random signal vector with zero mean components, $\varepsilon_i \in \mathbb{R}^p$ is a random vector with mean zero and covariance matrix $\sigma^2\mathbf{I}_p$, $\mathbf{A} \in \mathbb{R}^{p\times q_1}$ is the steering matrix whose q_1 columns are linearly independent of each other and $x_i \in \mathbb{R}^p$ is the observed vector on the p sensors.

Remark 2.1. Spiked population models allow small spikes (i.e. $\lambda_i < \sigma^2$), but we do not discuss this case because it is of less statistical significance.

2.2 Preliminary results for the estimated eigenvalues of $\mathbf{S}_n$

We use the following assumptions to specify the high dimensional framework and the moment conditions which will be used in different scenarios.

Assumption 2.1. p is proportional to n, i.e. $p/n \to c \in (0, +\infty)$.

Assumption 2.2. Z has i.i.d. entries z_{ji}, $1 \le j \le p$, $1 \le i \le n$ satisfying that $E(z_{11}) = 0$, $E(|z_{11}|^2) = 1$, $E(|z_{11}|^4) < \infty$.

Assumption 2.3. For any $k \in \mathbb{N}^+$, there exists a constant C_k s.t. $E(|z_{11}|^k) < C_k$.

Consider the sample covariance matrix $\mathbf{S}_n$ under Assumption 2.1 with a general σ^2. When $0 < c \le 1$, the empirical distribution of all the estimated eigenvalues $\hat{\lambda}_i$ almost surely converges to the well-known Marcenko-Pastur distribution with the support $(\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2) =: (\sigma^2 a, \sigma^2 b)$ (see, e.g. Theorem 2.14 in Yao et al. (2015)). To be specific, $\forall x \in \mathbb{R}$,

$$\frac{1}{p}\#\{\hat{\lambda}_i : \hat{\lambda}_i < x\} \to F_{c,\sigma^2}(x) \quad \text{a.s.} \tag{2.3}$$

with the density function

$$F'_{c,\sigma^2}(x) = \frac{1}{2\pi x c \sigma^2}\sqrt{(\sigma^2 b - x)(x - \sigma^2 a)}, \ \sigma^2 a < x < \sigma^2 b. \tag{2.4}$$

When $c > 1$, the integral of the above density function over the interval $(\sigma^2 a, \sigma^2 b)$ is equal to $1/c$, and there is an additional Dirac measure of mass $1 - 1/c$ at the origin $x = 0$.

For the extreme eigenvalues, Baik and Silverstein (2006) discovered the phase transition phenomenon in the case with $\sigma^2 = 1$ under Assumption 2.2. Slightly generalizing their result by a scale transformation $\hat{\lambda}_i \rightarrowtail \hat{\lambda}_i/\sigma^2$, we have that for any fixed L with $q + 1 < L < p$,

$$\hat{\lambda}_i \to \sigma^2\phi(\lambda_i/\sigma^2) \text{ a.s. for } i \le q, \quad \hat{\lambda}_i \to \sigma^2 b \text{ a.s. for } q+1 \le i \le L, \tag{2.5}$$

where $\phi(x) := x + cx(x-1)^{-1}$ is a strictly increasing function on $(1+\sqrt{c}, +\infty)$. Therefore, the number of identifiable spikes $q \leq q_1$ is defined as

$$q := \#\{\lambda_i : \lambda_i > \sigma^2(1+\sqrt{c})\}. \tag{2.6}$$

This is because there are only q extreme sample eigenvalues being outliers larger than $\sigma^2 b$ whenever the corresponding spikes exceed the value $\sigma^2(1+\sqrt{c})$ and any $\hat{\lambda}_i$ of λ_i with $\sigma^2 < \lambda_i \leq \sigma^2(1+\sqrt{c})$ for $q+1 \leq i \leq L$ converges in probability to the same upper bound $\sigma^2 b$.

Further, Bai and Yao (2008) built the Central Limit Theorem for the outliers $\hat{\lambda}_i$, $1 \leq i \leq q$, under Assumption 2.2 on the moments, which implies the $\sqrt{n}$-consistency of $\hat{\lambda}_i$ to $\sigma^2\phi(\lambda_i/\sigma^2)$. As for the eigenvalues $\hat{\lambda}_i$ sticking to the right edge $\sigma^2 b$ for $q+1 \leq i \leq L$, Theorem 2.5 in Cai et al. (2020) shows that $n^{2/3}(\hat{\lambda}_i - \sigma^2 b)$ has the limiting Type-1 Tracy-Widom distribution under Assumptions 2.2-2.3, which follows that $\hat{\lambda}_i - \sigma^2 b = O_{\mathrm{P}}(n^{-2/3})$.

Remark 2.2. These estimated eigenvalues $\hat{\lambda}_i$ corresponding to spikes $\lambda_i \leq \sigma^2(1+\sqrt{c})$ are not separated from those of $\lambda_i = \sigma^2$. Thus, we only estimate the number $q\,(\leq q_1)$ of spikes larger than $\sigma^2(1+\sqrt{c})$.

2.3 Valley-cliff criterion and estimation consistency

When p is proportional to n, estimated eigenvalues become much more dispersed as the Marcenko-Parstur law shown (see (2.3) and (2.4)). The estimation of $\lambda_i - \sigma^2$ is no longer consistent to 0, but that of $\lambda_i - \lambda_{i+1} =: \delta_i$

is still consistent for any $q_1 < i < \min\{n, p\}$. Thus, we do not directly use λ_i but rather δ_i in the criterion construction.

Define a sequence of ratios $r_i := \delta_{i+1}/\delta_i,\ 1 \le i \le p-2$. These ratios are scale-invariant and have the following property, when $i \le q_1$:

$$r_i = \frac{\delta_{i+1}}{\delta_i} = \frac{\delta_{i+1}/\sigma^2}{\delta_i/\sigma^2} = \begin{cases} \ge 0, & \text{for } i < q_1, \\ = 0, & \text{for } i = q_1. \end{cases} \tag{2.7}$$

For any $q_1 + 1 \le i \le p-2$, $r_i = 0/0$ is not well defined because of its instability which also occurs at the sample level. To alleviate its effect in constructing the criterion, we define a sequence by adding a ridge $c_n \to 0$ in both the numerator and the denominator:

$$r_i^{\mathrm{R}} := \frac{\delta_{i+1}/\sigma^2 + c_n}{\delta_i/\sigma^2 + c_n}, 1 \le i \le p-2, \tag{2.8}$$

where we use δ_i/σ^2 instead of δ_i in order to keep the selection of the ridge c_n independent of the scale parameter σ^2. Recalling the definition of δ_i and that $c_n \to 0$, these ratios have the following property:

$$r_i^{\mathrm{R}} = \frac{\delta_{i+1}/\sigma^2 + c_n}{\delta_i/\sigma^2 + c_n} = \begin{cases} \ge 0, & \text{for } i < q_1, \\ = c_n/(\delta_{q_1}/\sigma^2 + c_n) \to 0, & \text{for } i = q_1, \\ c_n/c_n = 1, & \text{for } q_1 + 1 \le i \le p-2, \end{cases}$$

They have a "valley-cliff" pattern, because q_1 should be the index of $r_{q_1}^{\mathrm{R}} \to 0$ at a "valley bottom" facing the "cliff" valued at 1 of all next ratios r_i^{R} for $i > q_1$. Define their sample versions $\hat{r}_i^{\mathrm{R}}$ with $\hat{\delta}_i := \hat{\lambda}_i - \hat{\lambda}_{i+1}$ as

$$\hat{r}_i^{\mathrm{R}} := \frac{\hat{\delta}_{i+1}/\sigma^2 + c_n}{\hat{\delta}_i/\sigma^2 + c_n},\ 1 \le i \le p-2, \tag{2.9}$$

where σ^2 is replaced by $\hat{\sigma}^2$ when σ^2 is unknown. Since $\hat{\lambda}_i$ is not consistent to λ_i, these ratios do not simply converge to their population counterparts, which makes the quantity q_1 generally unidentifiable. Hence, we estimate the largest possible order we can identify, namely q defined in (2.6).

According to the property of $\hat{\lambda}_i$, we have

$$\lim_{n\to\infty} \hat{\delta}_i = \begin{cases} \sigma^2\phi(\lambda_i/\sigma^2) - \sigma^2\phi(\lambda_{i+1}/\sigma^2) \text{ a.s.} & \text{for } 1 \le i \le q-1, \\ \sigma^2\phi(\lambda_q/\sigma^2) - \sigma^2 b > 0 \text{ a.s.} & \text{for } i = q, \\ 0, \text{ a.s.} & \text{for } q+1 \le i \le L-1. \end{cases}$$

More precisely, recalling the asymptotics of the extreme eigenvalues, we have that $\hat{\delta}_i = O_{\mathrm{P}}(n^{-2/3})$ for $q+1 \le i \le L-1$ and $\hat{\delta}_i$ for $1 \le q$, at the rate $O_{\mathrm{P}}(n^{-1/2})$, are either consistent to positive constants or to 0 when spikes are equal. When c_n is selected in the principle: $\hat{\delta}_i = o_{\mathrm{P}}(c_n)$ for $q+1 \le i \le p-1$, i.e. $c_n n^{2/3} \to \infty$, $\hat{r}_i^{\mathrm{R}}$ still have a nice "valley-cliff" pattern at $i = q$ as

$$\lim_{n\to\infty} \hat{r}_i^{\mathrm{R}} = \begin{cases} \ge 0, & i < q \\ 0, & i = q \\ 1, & q+1 \le i \le L-2 \end{cases} \tag{2.10}$$

with a probability going to one, where L is a prefixed upper bound for q. Taking this advantage, we define a thresholding VAlley-CLiff Estimator (VACLE) as, for a constant τ with $0 < \tau < 1$,

$$\hat{q}_n^{\mathrm{VACLE}} := \max_{1\le i\le L-2} \left\{ i : \hat{r}_i^{\mathrm{R}} \le \tau \right\}. \tag{2.11}$$

We state the estimation consistency as follows.

Theorem 2.1. *Suppose Assumptions 2.1-2.3 hold and that* $c_n \to 0$, $c_n n^{2/3} \to \infty$. *Then* $\mathbb{P}(\hat{q}_n^{\mathrm{VACLE}} = q) \to 1$ *as* $n \to \infty$.

Proof. The consistency of $\hat{q}_n^{\mathrm{VACLE}}$ is implied by (2.10).

Remark 2.3. Recalling $\hat{r}_i^{\mathrm{R}}$'s defined in (2.9), the value of $\hat{r}_i^{\mathrm{R}}$ depends on c_n and the estimator $\hat{\sigma}^2$ when σ^2 is unknown. Since the range of c_n can be rather wide, the criterion is not heavily affected when σ^2 is estimated, which is shown in the numerical studies we conduct later.

2.4 Modification of the VACLE

Selecting c_n plays an important role in the estimation efficiency of the VACLE. Although Theorem 2.1 provides estimation consistency, some numerical studies that are not presented in this paper indicate that the performance of $\hat{q}_n^{\mathrm{VACLE}}$ is sometimes and somehow sensitive to the value of the ridge c_n in finite sample cases. To be specific, when $\sigma^2\phi(\lambda_q/\sigma^2) - \sigma^2 b$ is small, the ratio at q could be close to 1, then we would easily achieve a smaller estimation of q. A small ridge c_n is therefore in demand. Meanwhile, a small c_n would result in the instability caused by 0/0 type ratios, and overestimation would be possible. There exists a trade-off between underestimation and overestimation in the choice of ridge c_n. We now alleviate this dilemma by using transformed eigenvalues.

Consider a transformation (depending on n) $f_n(\cdot)$ to define

$$\hat{\delta}_i^* := f_n(\hat{\lambda}_i/\sigma^2) - f_n(\hat{\lambda}_{i+1}/\sigma^2),\ i = 1, 2, \cdots, p-1. \tag{2.12}$$

The ratios are defined as

$$\hat{r}_i^{\mathrm{TR}} := \frac{\hat{\delta}_{i+1}^* + c_n}{\hat{\delta}_i^* + c_n},\ 1 \le i \le p-2, \tag{2.13}$$

and the estimator of q is defined as

$$\hat{q}_n^{\mathrm{TVACLE}} := \max_{1 \le i \le L-2} \left\{ i : \hat{r}_i^{\mathrm{TR}} \le \tau \right\}, \tag{2.14}$$

where c_n and τ have the same definitions as before. We call this criterion the transformation-based valley-cliff estimation(TVACLE).

For any transformation f_n, we wish that $\hat{r}_i^{\mathrm{TR}}$ remains close to 1 for $i > q$, and $\hat{r}_q^{\mathrm{TR}}$ is closer to zero than $\hat{r}_q^{\mathrm{R}}$. To this end, we use a transformation that satisfies conditions $(i) - (iii)$ below:

(i) $\mathbb{P}\{\hat{\delta}_q^* \ge \hat{\delta}_q/\sigma^2\} \to 1$; (ii) $\mathbb{P}\{\hat{\delta}_i^* \le \hat{\delta}_i/\sigma^2\} \to 1$, for $q+1 \le i \le p-2$;

(iii) $\hat{\delta}_{q+1}^*/\hat{\delta}_q^* \le \hat{\delta}_{q+1}/\hat{\delta}_q$.

Remark 2.4. Under conditions (i) and (ii), the transformation pulls the value of $\hat{\delta}_q$ up and presses that of $\hat{\delta}_i$, for $q+1 \le i \le p-2$ down. Condition (iii) is necessary to make the "valley" closer to its bottom "0" and then better separated from the "cliff" after the transformation, which is not implied by (i) and (ii) because (iii) holds in a deterministic way while (i) and (ii) hold with a probability going to one.

The following conditions (a) and (b) ensure that $f_n : \mathbb{R} \to \mathbb{R}$ satisfies the above conditions $(i) - (iii)$, letting $f_n'(x)$ be the derivative of $f_n(x)$:

(a) f_n is differentiable, and f_n' is increasing and nonnegative in $\mathbb{R}$;

(b) $\exists\ \kappa_n > 0$ s.t. $\kappa_n n^{2/3} \to \infty$ and $f_n'(x) = 1$, $\forall x \in (b - \kappa_n, b + \kappa_n)$.

Lemma 2.1. *Conditions* (a) *and* (b) *imply conditions* $(i)-(iii)$ *for* $\{\hat{\delta}^*_{n,i}\}$ *and* $\{\hat{\delta}_{n,i}\}$ *defined as above.*

Remark 2.5. In condition (b), κ_n can take a wide range of values, as long as it satisfies that $\kappa_n n^{2/3} \to \infty$. We let f'_n take value 1 in $(b-\kappa_n, b+\kappa_n)$, so that all $\hat{\lambda}_i/\sigma^2$, for $q+1 \le i \le L-1$, fall into this interval. Thus, the ratios $\hat{r}_i^{\mathrm{TR}}$, for $q+1 \le i \le L-2$, remain unaffected by the transformation f_n. Besides, the selection of κ_n is independent of c_n.

We now give a piecewise quadratic function for this purpose as follows:

$$f_n(x) = \begin{cases} L_n - \frac{1}{2k_1}, & x < L_n - \frac{1}{k_1} \\ \frac{1}{2}k_1 x^2 + (1-k_1 L_n)x + \frac{1}{2}k_1 L_n^2, & L_n - \frac{1}{k_1} \le x < L_n \\ x, & L_n \le x < R_n \\ \frac{1}{2}k_2 x^2 + (1-k_2 R_n)x + \frac{1}{2}k_2 R_n^2, & x \ge R_n \end{cases} \tag{2.15}$$

where slopes k_1 and k_2 are to be determined, $L_n = b - \kappa_n$, $R_n = b + \kappa_n$. Obviously, the TVACLE degenerates to the VACLE when $k_1 = k_2 = 0$.

The consistency of $\hat{q}_n^{\mathrm{TVACLE}}$ is stated in the following theorem.

Theorem 2.2. *Under the same conditions of Theorem 2.1,* $\hat{q}_n^{\mathrm{TVACLE}}$ *with the above transformation* f_n *is equal to* q *with a probability going to 1.*

Remark 2.6. Although selecting an optimal transformation is desirable, we suspect the existence as there is a large class of functions satisfying the conditions. Thus, such an issue is beyond the scope of this paper.

3. More examples

We consider more examples with structures similar to spiked population models in this section.

3.1 Large dimensional auto-covariance matrix

The auto-covariance matrix has a complicated structure at the sample level, so we provide more discussion on it. Since the theory for the estimated matrix is not as complete as those for the spiked population models, we need to add an extra assumption on the convergence rate of the estimated eigenvalues, as reasonably conjectured by Li et al. (2017), to derive the estimation consistency. Although the assumption would be true, it requires a rigorous proof that is beyond the scope of this paper, and we leave it to a further study. In this section, we provide a proposition that assumes that the convergence rate can be achieved and use numerical studies to verify the usefulness of our method in practice.

Consider a factor model:

$$y_t = \mathbf{A}x_t + \varepsilon_t, \tag{3.16}$$

where for a fixed number q_0, $x_t \in \mathbb{R}^{q_0}$ is a common factor time series, $\mathbf{A}$ is the $p \times q_0$ factor loading matrix, $\{\varepsilon_t\}$ is a sequence of Gaussian noises independent of x_t, and y_t is the t-th column of the $p \times T$ observed matrix $\mathbf{Y}$. Let $\Sigma_y = \mathrm{Cov}(y_t, y_{t-1})$ be the lag-1 auto-covariance matrices of y_t. Then

$$\begin{aligned}\Sigma_y = \mathrm{Cov}(y_t, y_{t-1}) &= \mathrm{Cov}(\mathbf{A}x_t + \varepsilon_t, \mathbf{A}x_{t-1} + \varepsilon_{t-1}) \\ &= \mathbf{A}\mathrm{Cov}(x_t, x_{t-1})\mathbf{A}^\top + \mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1}) =: \Delta + \Sigma_\varepsilon,\end{aligned}$$

which is a finite-rank perturbation of Σ_ε. That is, Σ_y has a structure similar to the spiked popualtion model in (2.1) although Σ_y is not symmetric. The

order determination in this example is to estimate an identifiable quantity $q \leq q_0$ based on the singular values of $\hat{\Sigma}_y := T^{-1}\sum_{t=2}^{T+1} y_t y_{t-1}^\top$.

Let μ be a finite measure on the real line $\mathbb{R}$ with support denoted by $\text{supp}(\mu)$ and $\mathbb{C}\backslash\text{supp}(\mu)$ be a complex space $\mathbb{C}$ subtracting the set $\text{supp}(\mu)$. For any $z \in \mathbb{C}\backslash\text{supp}(\mu)$, the Stieltjes transformation and T-transformation of μ are respectively defined as

$$\mathcal{S}(z) := \int \frac{1}{t-z} d\mu(t), \ \mathcal{T}(z) := \int \frac{t}{z-t} d\mu(t). \tag{3.17}$$

When μ is supported on an interval, say $\text{supp}(\mu) = [A, B]$, and z is a real value, the T-transformation $\mathcal{T}(\cdot)$ is a decreasing homeomorphism from $(-\infty, A)$ onto $(\mathcal{T}(A-), 0)$ and from $(B, +\infty)$ onto $(0, \mathcal{T}(B+))$, where

$$\mathcal{T}(A-) := \lim_{z\in\mathbb{R}, z\to A-} \mathcal{T}(z), \ \mathcal{T}(B+) := \lim_{z\in\mathbb{R}, z\to B+} \mathcal{T}(z).$$

Give the assumptions on the time series $\{x_t\}_{1\leq t\leq T}$ and $\{\varepsilon_t\}_{1\leq t\leq T}$ (Li et al. (2017)) as follows.

Assumption 3.1. p is propotional to T, i.e. $p/T \to y \in (0, +\infty)$.

Assumption 3.2. $\{x_t\}_{1\leq t\leq T}$ is a q_0-dimensional stationary time series, where q_0 is a fixed number, with independent components and the following decomposition:

$$x_{i,t} = \sum_{l=0}^{\infty} \alpha_{i,l}\eta_{i,t-l}, \ i = 1, \cdots, q_0, \ t = 1, \cdots, T,$$

where $\{\eta_{i,k}\}$ is a real-valued and weakly stationary white noise with mean 0 and variance σ_i^2. Denote $\gamma_0(i)$ and $\gamma_1(i)$ as the variance and lag-1 auto-covariance of $\{x_{i,t}\}$, respectively.

Assumption 3.3. $\{\varepsilon_t\}$ is a p-dimensional real-valued random vector independent of $\{x_t\}$ and with independent components $\varepsilon_{i,t}$, satisfying $E(\varepsilon_{i,t}) = 0$, $E(\varepsilon_{i,t}^2) = \sigma^2$, and for any $\eta > 0$,

$$\frac{1}{\eta^4 p^T} \sum_{i=1}^{p} \sum_{t=1}^{T+1} E(|\varepsilon_{i,t}|^4 I_{(|\varepsilon_{i,t} \geq \eta T^{1/4}|)}) \longrightarrow 0 \text{ as } pT \to \infty.$$

We show the identifiability of q_0 by the following proposition.

Proposition 3.1. *Assume Assumptions 3.1-3.3 are satisfied. Denote $\mathcal{T}(\cdot)$ as the T-transformation of the limiting spectral distribution for matrix $\hat{\mathbf{M}}_y/\sigma^4 = \hat{\Sigma}_y \hat{\Sigma}_y^\top/\sigma^4$. Suppose that the above assumptions are satisfied. Let $q := \#\{i : 1 \leq i \leq q_0,\ \mathcal{T}_1(i) < \mathcal{T}(b_1+)\}$, where*

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

$$b_1 = (-1 + 20y + 8y^2 + (1+8y)^{3/2})/8,\ \mathcal{T}(b_1+) = \lim_{z \in \mathbb{R}, z \to b_1+} \mathcal{T}(z).$$

Then q is the largest number of identifiable common factors.

Remark 3.1. Although the constraint $\mathcal{T}_1(i) < \mathcal{T}(b_1+)$ does not have a simple formulation as presented in the spiked population models, it also provides the optimal bound.

Denoting $\hat{\lambda}_i$, $1 \leq i \leq p$, as the eigenvalues of $\hat{\mathbf{M}}_y$, we construct a VACLE and a TVACLE for the q defined above by replacing (σ^2, b) with (σ^4, b_1) in (2.11) and (2.14) respectively. Their consistencies are shown below.

Proposition 3.2. *If the estimated eigenvalues $\hat{\lambda}_i$ for $i > q$ have a convergence rate of order $O_{\mathrm{P}}(n^{-2/3})$ with the assumptions in Proposition 3.1, then $\mathbb{P}(\hat{q}_n^{\mathrm{VACLE}} = q) \to 1$ and $\mathbb{P}(\hat{q}_n^{\mathrm{TVACLE}} = q) \to 1$ as $n \to \infty$.*

Remark 3.2. As we commented above, Li et al. (2017) proposed a criterion with a reasonable conjecture on the convergence rate of order $O_{\mathrm{P}}(n^{-2/3})$ without rigorous proof. We have not proved this result either, and thus consider the above results to be propositions, rather than theorems. We will see that it works well in numerical studies.

3.2 Large-dimensional spiked Fisher matrix

Again consider the signal detection problem discussed above,

$$x_i = \mathbf{A}u_i + \varepsilon_i,\ 1 \le i \le n, \tag{3.18}$$

where x_i, $\mathbf{A}$ and u_i share the same settings of (2.2), whilst ε_i is a noise vector with a general covariance matrix Σ_2. Denote the population covariance matrix of x_i by Σ_1 such that $\Sigma_1 = \Sigma_2 + \Delta$, where $\Delta = \mathbf{A}\mathrm{Cov}(u_i)\mathbf{A}^T$ is a non-negative definite matrix with fixed rank q_1 provided that $\mathrm{Cov}(u_i)$ is of full rank. Then $\Sigma_1\Sigma_2^{-1}$ has a spiked structure as

$$\mathrm{spec}(\Sigma_1\Sigma_2^{-1}) = \{\lambda_1, \cdots, \lambda_{q_1}, 1, \cdots, 1\}, \tag{3.19}$$

where $\lambda_1 \ge \cdots \ge \lambda_{q_1} > 1$ and the number of spikes q_1 is fixed. When Σ_2 is known, the sample version of $\Sigma_1\Sigma_2^{-1}$ is $\mathbf{S}_n\Sigma_2^{-1}$ where $\mathbf{S}_n$ is the sample covariance matrix in the spiked population model. Otherwise, both of Σ_1 and Σ_2 need to be estimated. Let $\mathbf{S}_1 := n^{-1}\sum x_i x_i^\top$ and $\mathbf{S}_2 := T^{-1}\sum e_t e_t^\top$ corresponding to Σ_1 and Σ_2 with respective sample sizes of n and T, where the sample covariance matrix $\mathbf{S}_2$ comes from another sequence of pure noise observations, say $\{e_i\}_{1\le i\le T}$, with a different sample size T. When $\mathbf{S}_2$ is

invertible, the random matrix $\mathbf{F}_n := \mathbf{S}_1\mathbf{S}_2^{-1}$ is called a Fisher matrix, whose motivation comes from the following hypothesis testing problem:

$$H_0 : \ \Sigma_1 = \Sigma_2 \quad H_1 : \ \Sigma_1 = \Sigma_2 + \Delta. \tag{3.20}$$

See Wang and Yao (2017) as an example. Denote the eigenvalues of $\mathbf{F}_n$ as $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$. The difference between the two hypotheses relies upon those extreme eigenvalues of $\mathbf{F}_n$.

Consider a more general Fisher matrix with the spiked structure

$$\text{spec}(\Sigma_1\Sigma_2^{-1}) = \{\lambda_1, \cdots, \lambda_{q_1}, \sigma^2, \cdots, \sigma^2\}, \tag{3.21}$$

which is motivated by the hypothesis testing problem:

$$H_0 : \ \Sigma_1 = \sigma^2\Sigma_2 \quad H_1 : \ \Sigma_1 = \sigma^2\Sigma_2 + \Delta, \tag{3.22}$$

By using the simple transformation $\hat{\lambda}_i \rightarrowtail \hat{\lambda}_i/\sigma^2$, we can also achieve the results in the case of $\sigma^2 = 1$ in a similar manner.

Give the assumptions on the samples $\{x_i\}_{1\leq i\leq n}$ and $\{e_t\}_{1\leq t\leq T}$ as follows.

Assumption 3.4. $p/n \to c \in (0, \infty)$ and $p/T \to y \in (0, 1)$.

Assumption 3.5. Let $z_i := \Sigma_1^{-1/2}x_i$ and $w_t := \Sigma_2^{-1/2}e_t$ for $1 \leq i \leq n$ and $1 \leq t \leq T$. Assume that $\{z_i\}_{1\leq i\leq n}$ and $\{w_t\}_{1\leq t\leq T}$ are independent and satisfy moment conditions Assumptions 2.2 and 2.3.

The order determination in this example is to estimate the number q of spikes, whose identifiability is shown in the following proposition.

Proposition 3.3. *Suppose that Assumptions 3.4 and 3.5 are satisfied. Define* $q := \#\{i : \lambda_i > \sigma^2(1-y)^{-1}(1+\sqrt{c+y-cy})\}$. *Then* q *is the number of identifiable spikes.*

Let $b_2 := (1-y)^{-2}(1+\sqrt{c+y-cy})^2$ and construct a VACLE and a TVACLE for q by replacing b with b_2 in (2.11) and (2.14) respectively. We show consistencies of the VACLE and the TVACLE below.

Theorem 3.1. *Suppose that Assumptions 3.4 and 3.5 are satisfied. Then* $\mathbb{P}(\hat{q}_n^{\text{VACLE}} = q) \to 1$ *and* $\mathbb{P}(\hat{q}_n^{\text{TVACLE}} = q) \to 1$ *as* $n \to \infty$.

3.3 General cases

Beyond the three models studied above, we consider more general cases in this section. Suppose that $\mathbf{T}_n \in \mathbb{R}^{p\times p}$ is the matrix in the order determination problem of interest, for example, the sample covariance matrix $\mathbf{S}_n$ in Section 2, and $\hat{\lambda}_i$, $1 \le i \le p$, are its eigenvalues in descending order.

We assume the following model features for $\hat{\lambda}_i$, $1 \le i \le p$.

Model Feature 3.1. In the high dimensional regime with $p/n \to c \in (0,\infty)$, suppose that there exists a fixed constant $q \in \mathbb{N}^+$ satisfying:

(A1) there exists a constant d such that $\hat{\lambda}_q - d = o_{\text{P}}(1)$ as $n \to \infty$;

(A2) for a large fixed value L satisfying $q+1 < L < p$, there exist $e < d$ and a sequence $\tilde{c}_n \to 0$ such that $\hat{\lambda}_i - e = O_{\text{P}}(\tilde{c}_n)$, for $q+1 \le i \le L$.

Remark 3.3. Condition (A1) corresponds to the so-called *the phase transition phenomenon* for extreme eigenvalues. (A2) further focuses on the fluctuations of those eigenvalues sticking to the boundary of the bulk and the fluctuation is often of order $O_{\text{P}}(n^{-2/3})$, namely $\tilde{c}_n = n^{-2/3}$. General

theory for the phase transitions and fluctuations can be found, for example, in Péché (2006), Benaych-Georges et al. (2011), Benaych-Georges and Nadakuditi (2011) and Knowles and Yin (2017). The forementioned three models are typical examples.

Similar to spiked population models, the VACLE and the TVACLE for the q defined in this model feature can be constructed by replacing $\sigma^2 b$ with d and taking a ridge c_n satisfying $c_n/\tilde{c}_n \to \infty$ in (2.11) and (2.14).

4. Numerical Studies

4.1 Numerical studies on spiked population models

Consider the comparisons between the VACLE and the TVACLE written as $\hat{q}_n^{\text{VACLE}}$ and $\hat{q}_n^{\text{TVACLE}}$ and the estimator written as $\hat{q}_n^{\text{PY}}$ developed and refined by Passemier and Yao (2012) and Passemier and Yao (2014). Because estimating q is the main focus, we conduct simulations mainly with given σ^2. For the unknown σ^2 scenario, we give a simple one-step estimator of σ^2 and a brief discussion. In all simulations, we conduct 500 independent replications. We report the results, recalling $c = p/n$, with three scenarios: $c = .25, 1$ and 2 to represent the cases with different dimensions p smaller than and larger than the sample size n, respectively. $\hat{q}_n^{\text{PY}}$ is defined by

$$\hat{q}_n^{\text{PY}} := \min\{i \in \{1, \cdots, L\} : \hat{\delta}_{i+1} < d_n \text{ and } \hat{\delta}_{i+2} < d_n\}, \tag{4.23}$$

where $L > q$ is a prefixed bound large enough, $d_n = o(n^{-1/2})$ and $n^{2/3} d_n \to +\infty$.

Scale estimation. Passemier and Yao (2012) estimated σ^2 by simply taking the average over $\{\hat{\lambda}_i\}_{q+1\le i\le p}$ and Passemier et al. (2017) established its consistency and further introduced a refined version by subtracting the bias. But it involves an iteration procedure because the number q must be first estimated. To construct a robust estimator, Ulfarsson and Solo (2008) and Johnstone and Lu (2009) used the median of the sample eigenvalues $\{\hat{\lambda}_i : \hat{\lambda}_i \le b\}$ and the sample variances $\{n^{-1}\sum_{i=1}^n x_{ij}^2\}_{1\le j\le p}$, respectively. The former median still needs a crude estimation of the right edge $b = \sigma^2(1+\sqrt{c})^2$ in advance, which amounts to a rough initial estimation of σ^2.

We propose a one-step procedure that could be regarded as a simplified version of the method in Ulfarsson and Solo (2008). For spiked population models, the empirical spectral distribution of $\mathbf{S}_n$ almost surely converges to a Marcenko-Pastur distribution $F_{c,\sigma^2}(x)$ (see (2.3) and (2.4)). For $0 < \alpha < 1$, their α-quantiles are denoted by $\hat{\xi}^{(n)}_{c,\sigma^2}(\alpha)$ and $\xi_{c,\sigma^2}(\alpha)$, respectively:

$$\hat{\xi}^{(n)}_{c,\sigma^2}(\alpha) := \hat{\lambda}_{p-[p\alpha]}, \quad \xi_{c,\sigma^2}(\alpha) := \inf\{x : F_{c,\sigma^2}(x) \ge \alpha\}. \tag{4.24}$$

It then follows that $\hat{\xi}^{(n)}_{c,\sigma^2}(\alpha) \to \xi_{c,\sigma^2}(\alpha)$ as $n \to \infty$. Note that $\xi_{c,\sigma^2}(\alpha) = \sigma^2\xi_{c,1}(\alpha)$. Approximating a certain quantile, say $\xi_{c,\sigma^2}(\alpha)$, of the M-P distribution by its sample counterpart $\hat{\xi}^{(n)}_{c,\sigma^2}(\alpha)$, we obtain an estimator of σ^2,

$$\hat{\sigma}^2 = \hat{\xi}^{(n)}_{c,\sigma^2}(\alpha)/\xi_{c,1}(\alpha). \tag{4.25}$$

The consistency of $\hat{\sigma}^2$ is equivalent to that of $\hat{\xi}^{(n)}_{c,\sigma^2}(\alpha)$, which holds under Assumption 2.2. Practically, for simplicity and stability, let $\alpha = 0.5$ for

$0 < c < 1$; and $1 - (2c)^{-1}$ for $c \geq 1$. Then $\alpha = 1 - (2\max\{1, c\})^{-1}$. The sample quantile $\hat{\xi}^{(n)}_{c,\sigma^2}(\alpha)$ divides all positive eigenvalues of $\mathbf{S}_n$ into two equal parts. The estimator $\hat{\sigma}^2$ can be less sensitive to extreme eigenvalues of $\mathbf{S}_n$. Its performance is examined in the following numerical studies.

Remark 4.1. The rigidity of the eigenvalues of covariance matrix (see Theorem 3.3 in Pillai and Yin (2014)) implies that the convergence rate of $\hat{\sigma}^2$ is $o(n^{-1+\varepsilon})$ for any $\varepsilon > 0$. The consistencies still hold for the VACLE and the TVACLE with $\hat{\lambda}_i/\hat{\sigma}^2$, as $\hat{\sigma}^2$ has a higher convergence rate than extreme eigenvalues $\hat{\lambda}_i$, $1 \leq i \leq L$ for any fixed L. Repeating the construction of (4.25) for estimated eigenvalues of auto-covariance matrices and spiked Fisher matrices can lead to similar estimators for σ^2. Their consistencies are implied by the convergence of empirical spectral distributions of $\hat{\mathbf{M}}_y$ and $\mathbf{F}_n$ respectively (see Li et al. (2017) and Wang and Yao (2017)), but the convergence rates are still under study, we then do not give more discussion.

Models and parameters selections: the known σ^2 case.

For $\hat{q}_n^{\mathrm{PY}}$, the sequence $d_n = Cn^{-2/3}\sqrt{2\log\log n}$ with C being adjusted by an automatic procedure identical to that in Passemier and Yao (2014). For $\hat{q}_n^{\mathrm{VACLE}}$ and $\hat{q}_n^{\mathrm{TVACLE}}$, they share the same threshold $\tau = 0.5$ but have different ridges c_n. Theoretically, c_n can be selected flexibly on condition that $c_n \to 0$ and $n^{2/3}c_n \to +\infty$. Here, we give an automatic procedure for

ridge calibration by pure-noise simulations. For given (p, n), we conduct 500 independent pure-noise simulations and obtain the $\alpha-$quantile $\mathrm{q}_{p,n}(\alpha)$ and sample mean $\mathrm{m}_{p,n}$ of the difference $\{\tilde{\lambda}_1 - \tilde{\lambda}_2\}$, where $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are the two largest eigenvalues of the noise matrix. By results in Benaych-Georges et al. (2011), we can approximate $\hat{\delta}_{q+1}$ by $\{\tilde{\lambda}_1 - \tilde{\lambda}_2\}$:

$$\begin{aligned}&\mathbb{P}\{\mathrm{q}_{p,n}(0.01) - \mathrm{m}_{p,n} < \hat{\delta}_{q+1} - \mathrm{m}_{p,n} < \mathrm{q}_{p,n}(0.99) - \mathrm{m}_{p,n}\}\\ \approx&\mathbb{P}\{\mathrm{q}_{p,n}(0.01) - \mathrm{m}_{p,n} < \tilde{\lambda}_1 - \tilde{\lambda}_2 - \mathrm{m}_{p,n} < \mathrm{q}_{p,n}(0.99) - \mathrm{m}_{p,n}\} \approx 0.98.\end{aligned}$$

Thus, the value of $\{\hat{\delta}_{q+2} - \mathrm{m}_{p,n} + (\mathrm{q}_{p,n}(0.99) - \mathrm{q}_{p,n}(0.01))\}\{\hat{\delta}_{q+1} - \mathrm{m}_{p,n} + (\mathrm{q}_{p,n}(0.99) - \mathrm{q}_{p,n}(0.01))\}^{-1}$ would be dominated by the term $(\mathrm{q}_{p,n}(0.99) - \mathrm{q}_{p,n}(0.01) - \mathrm{m}_{p,n})$ and close to the "cliff" valued at 1 with a high probability. We use ridge $c_n^{(1)} = \log\log n(\mathrm{q}_{p,n}(0.95) - \mathrm{q}_{p,n}(0.05)) - \mathrm{m}_{p,n}$ for the VACLE and a smaller one $c_n^{(2)} = \sqrt{\log\log n}(\mathrm{q}_{p,n}(0.95) - \mathrm{q}_{p,n}(0.05)) - \mathrm{m}_{p,n}$ for the TVACLE. Note that $\mathrm{q}_{p,n}(\alpha)$ and $\mathrm{m}_{p,n}$ have the same convergence rate as $\hat{\lambda}_{q+1}$ which has a slightly faster rate to zero than $c_n^{(1)}$ and $c_n^{(2)}$. Also, we determine the sequence κ_n, the bound L, and the slopes k_1 and k_2 by the rule of thumb. We take $L = 20$ because it is much larger than the true value of q in the simulations and many practical scenarios, also large enough. Details in the selections of the parameters are reported in Table 1. Following the calibration procedure of Passemier and Yao (2014), we obtain the value of C for various $c = p/n$, as shown in Table 2.

Table 1: Parameters settings for the three methods

Method	d_n	τ	c_n	κ_n	k_1	k_2	L
PY	$Cn^{-2/3}\sqrt{2\log\log n}$	—	—	—	—	—	20
VACLE	—	0.5	$c_n^{(1)}$	—	—	—	20
TVACLE	—	0.5	$c_n^{(2)}$	$p^{-2/3}\log\log p$	5	5	20

Table 2: Values of C

c=p/n	0.25	1	2
C	5.5226	6.3424	7.6257

Remark 4.2. Note that we select different ridges c_n in $\hat{q}_n^{\mathrm{VACLE}}$ and $\hat{q}_n^{\mathrm{TVACLE}}$. As described above, we want a small ridge c_n to make $\hat{r}_{q+1}^{\mathrm{R}}$ well separated from $\hat{r}_q^{\mathrm{R}}$, but this may lead to instability of $\hat{r}_i^{\mathrm{R}}$ for $i > q+1$. Because $\hat{r}_i^{\mathrm{TR}}$ is less sensitive to the ridge than $\hat{r}_i^{\mathrm{R}}$, we can choose a smaller ridge for $\hat{q}_n^{\mathrm{TVACLE}}$. Besides, the ridges $c_n^{(1)}$ and $c_n^{(2)}$ are generated by an automatic procedure instead of manual selections. This calibration procedure only depends on (p, n). Overall, when the signals are stronger, the detection is easier.

Consider three models: for fair comparisons, Models 1 and 2 were used by Passemier and Yao (2012) with dispersed spikes and closely spaced but unequal spikes respectively, and Model 3 has two equal spikes:

Model 1. $q = 5$, $(\lambda_1, \cdots, \lambda_5) = (259.72, 17.97, 11.04, 7.88, 4.82)$,

Model 2. $q = 4$, $(\lambda_1, \cdots, \lambda_4) = (7, 6, 5, 4)$,

Model 3. $q = 4$, $(\lambda_1, \cdots, \lambda_4) = (5, 4, 3, 3)$.

Furthermore, we compare $\hat{q}_n^{\mathrm{TVACLE}}$ with $\hat{q}_n^{\mathrm{PY}}$ on a model with a greater multiplicity of spikes:

Model 4. $q = 6$, $(\lambda_1, \cdots, \lambda_6) = (5, 5, 5, 5, 5, 5)$.

Set $\sigma^2 = 1$. When σ^2 is regarded as unknown, use the one-step method in (4.25) to estimate it. We conduct the same simulations for $\hat{q}_n^{\text{VACLE}}$ and $\hat{q}_n^{\text{TVACLE}}$ as those with the known σ^2, but we do not report results for $\hat{q}_n^{\text{PY}}$ with the unknown σ^2 because the results and conclusions are very similar.

Table 3: Mean, mean square error and misestimation rates of $\hat{q}_n^{\text{PY}}$, $\hat{q}_n^{\text{VACLE}}$ and $\hat{q}_n^{\text{TVACLE}}$ over 500 independent replications for Models 1-3, with the known $\sigma^2 = 1$.

		$\hat{q}_n^{\text{PY}}$			$\hat{q}_n^{\text{VACLE}}$			$\hat{q}_n^{\text{TVACLE}}$		
	(p, n)	Mean	MSE	$\hat{q}_n^{\text{PY}} \neq q$	Mean	MSE	$\hat{q}_n^{\text{VACLE}} \neq q$	Mean	MSE	$\hat{q}_n^{\text{TVACLE}} \neq q$
Model 1	$(50, 200)$	5.022	0.022	0.022	5.004	0.004	0.004	5.024	0.024	0.024
	$(200, 800)$	5.012	0.012	0.012	5.002	0.002	0.002	5.016	0.016	0.016
	$(100, 100)$	5.016	0.02	0.02	4.97	0.046	0.046	4.998	0.002	0.002
	$(200, 200)$	5.026	0.03	0.024	5.01	0.01	0.01	5.004	0.004	0.004
	$(100, 50)$	4.846	0.218	0.212	4.484	1.296	0.41	4.782	0.222	0.216
	$(200, 100)$	4.99	0.074	0.074	4.758	0.486	0.194	4.954	0.046	0.046
Model 2	$(50, 200)$	4.018	0.058	0.028	4.006	0.006	0.006	4.016	0.016	0.016
	$(200, 800)$	4.016	0.02	0.014	4.004	0.004	0.004	4.032	0.04	0.028
	$(100, 100)$	3.922	0.246	0.074	3.416	2.112	0.22	3.968	0.036	0.036
	$(200, 200)$	4.014	0.014	0.014	3.92	0.304	0.048	4.006	0.006	0.006
	$(200, 100)$	3.558	0.83	0.342	2.452	5.144	0.584	3.712	0.304	0.28
	$(400, 200)$	3.906	0.162	0.118	3.046	3.138	0.364	3.958	0.05	0.044
Model 3	$(50, 200)$	3.994	0.118	0.032	3.772	0.804	0.08	4.024	0.024	0.024
	$(200, 800)$	4.018	0.018	0.018	4	0	0	4.036	0.036	0.036
	$(200, 200)$	3.456	0.92	0.414	1.94	6.684	0.734	3.614	0.518	0.326
	$(400, 400)$	3.904	0.18	0.122	2.7	4.152	0.478	3.898	0.142	0.112
	$(400, 200)$	2.222	3.81	0.952	1.08	9.736	0.968	2.648	2.296	0.91
	$(800, 400)$	2.626	2.482	0.844	1.588	7.104	0.954	3.022	1.558	0.7

From Table 3, we have the following observations. For Model 1, all three methods work well with high accuracies and small MSEs in the cases where the dimension p is smaller than n ($c = p/n = 0.25$). When either $c = 1$ or $c = 2$, $\hat{q}_n^{\text{TVACLE}}$ is the best, and $\hat{q}_n^{\text{PY}}$ also has smaller MSEs than $\hat{q}_n^{\text{VACLE}}$. In a word, all three methods perform in a satisfactory manner, but the performance of $\hat{q}_n^{\text{TVACLE}}$ is the stablest for various values of $c = p/n$.

Table 4: Mean, mean squared error and empirical distribution of $\hat{q}_n^{\mathrm{PY}}$ and $\hat{q}_n^{\mathrm{TVACLE}}$ over 500 independent replications for Model 4 ($q = 6$), with the known $\sigma^2 = 1$.

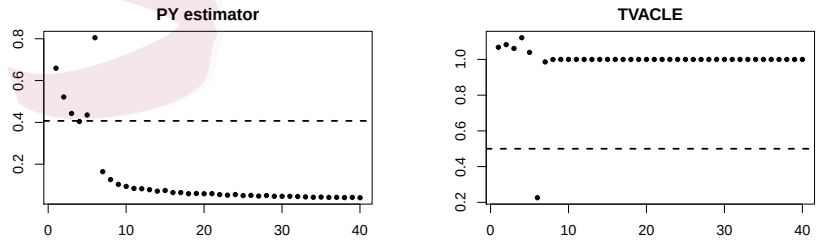
	(p, n)	Mean	MSE	$\hat{q}=0$	$\hat{q}=1$	$\hat{q}=2$	$\hat{q}=3$	$\hat{q}=4$	$\hat{q}=5$	$\hat{q}=6$	$\hat{q}\geq 7$
$\hat{q}_n^{\mathrm{PY}}$	$(50, 200)$	5.358	2.874	0.018	0.04	0.042	0.06	0	0	**0.826**	0.014
	$(200, 800)$	5.816	0.868	0.002	0.014	0.02	0.012	0	0	**0.94**	0.012
	$(100, 100)$	4.436	6.904	0.06	0.072	0.118	0.106	0.01	0.048	**0.572**	0.014
	$(200, 200)$	4.964	4.772	0.042	0.052	0.082	0.07	0	0	**0.742**	0.012
	$(400, 200)$	3.858	9.794	0.078	0.138	0.164	0.094	0.008	0.032	**0.484**	0.002
	$(800, 400)$	4.406	7.558	0.068	0.11	0.098	0.086	0	0	**0.626**	0.012
$\hat{q}_n^{\mathrm{TVACLE}}$	$(50, 200)$	6.006	0.006	0	0	0	0	0	0	**0.994**	0.006
	$(200, 800)$	6.024	0	0	0	0	0	0	0	**0.976**	0.024
	$(100, 100)$	5.886	0.122	0	0	0	0	0.004	0.106	**0.89**	0.11
	$(200, 200)$	6	0	0	0	0	0	0	0	**1**	0
	$(400, 200)$	5.952	0.06	0	0	0	0	0.004	0.042	**0.952**	0.002
	$(800, 400)$	6.002	0.002	0	0	0	0	0	0	**0.998**	0.002

For Model 2 $\hat{q}_n^{\mathrm{VACLE}}$ is sensitive to the ratio c, particularly its MSE. Although when $c = 2$, $\hat{q}_n^{\mathrm{TVACLE}}$ may sometimes slightly underestimate the true number, it is less serious than $\hat{q}_n^{\mathrm{PY}}$. For Model 3 with two equal spikes, $\hat{q}_n^{\mathrm{TVACLE}}$ works much better than both $\hat{q}_n^{\mathrm{PY}}$ and $\hat{q}_n^{\mathrm{VACLE}}$ that underestimate q significantly. To further confirm this phenomenon, we report the results for Model 4 with more equal spikes. The results in Table 4 suggest that $\hat{q}_n^{\mathrm{TVACLE}}$ overall performs better than $\hat{q}_n^{\mathrm{PY}}$ in terms of estimation accuracy and MSE. It has an underestimation problem as its searching procedure stops earlier once the difference between consecutive eigenvalues corresponding to equal spikes is below the threshold d_n. This conclusion can be made after observing its empirical distributions in Table 4. In contrast, $\hat{q}_n^{\mathrm{TVACLE}}$ largely avoids this problem. To better illustrate this fact, we plot in Figure 1 the first 40 differences $\hat{\delta}_i$ for $\hat{q}_n^{\mathrm{PY}}$ and the first 40 ratios of $\hat{r}_i^{\mathrm{TR}}$ for $\hat{q}_n^{\mathrm{TVACLE}}$. The left subfigure shows that there are three $\hat{\delta}_i$, $i = 3, 4, 5$, being very close to

the threshold line $y = d_n$, which causes the underestimation problem shown in Table 4. In contrast, the right subfigure shows that the "valley" $\hat{r}_q^{\mathrm{TR}}$ and the "cliff" $\hat{r}_{q+1}^{\mathrm{TR}}$ are well separated by the threshold line $\tau = 0.5$.

As we claimed in Sections 2.3 and 2.4, the VACLE could be somehow sensitive to the ridge selection. The results reported in Table 3 confirm this claim. To explore how the ridge c_n affects both the VACLE and the TVACLE, Figures 2 presents, for Model 2 with $(p, n) = (400, 200)$, the boxplots of the first 7 ratios without ridge $\hat{r}_i$; the first 7 ridge ratios $\hat{r}_i^{\mathrm{R}}$; and the first 7 transformed ridge ratios $\hat{r}_i^{\mathrm{TR}}$. From the left to right subfigure of Figure 2, we can see that $\hat{r}_i$ fluctuates much more than $\hat{r}_i^{\mathrm{R}}$ for $i > q = 4$, and that $\hat{r}_4^{\mathrm{TR}}$ and $\hat{r}_i^{\mathrm{TR}}$, $i > 4$ are separated more significantly. This confirms the necessity of using a ridge with a stable ratio $\hat{r}_i^{\mathrm{R}}$ and transformation can enhance the estimation accuracy.

Figure 1: Plots of the first 40 differences and ratios: the left is for differences $\hat{\delta}_i$, $1 \le i \le 40$, in $\hat{q}_n^{\mathrm{PY}}$; the right is for ratios $\hat{r}_i^{\mathrm{TR}}$, $1 \le i \le 40$, in $\hat{q}_n^{\mathrm{TVACLE}}$. The results are based on simulations for Model 4 with 500 independent replications, and $(p, n) = (400, 200)$.



The unknown σ^2 Case. Use Models 2 and 4 and regard σ^2 as an un-

known value. These two models represent the cases with and without equal spikes. Furthermore, because the conclusions are very similar to those with known σ^2, we then report only the results for $\hat{q}_n^{\mathrm{VACLE}}$ and $\hat{q}_n^{\mathrm{TVACLE}}$ to further confirm the advantages of $\hat{q}_n^{\mathrm{TVACLE}}$. The numerical results are shown in Table 5. The results in the last two columns show that the one-step estimation $\hat{\sigma}^2$ has good performance in terms of accuracy and robustness.

Figure 2: Boxplots of the first 7 ratios: the left is for ratios without ridge, $\hat{r}_i$; the middle is for ratios with ridge $\hat{r}_i^{\mathrm{R}}$; the right is for transformed ratios with ridge $\hat{r}_i^{\mathrm{TR}}$.

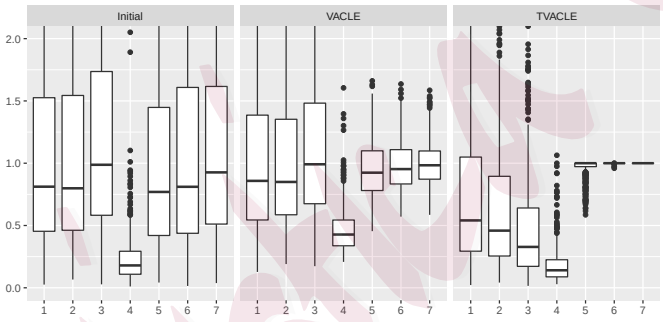


Table 5: Mean and mean square error of $\hat{q}_n^{\mathrm{VACLE}}$, $\hat{q}_n^{\mathrm{TVACLE}}$ and $\hat{\sigma}^2$, and the misestimation rates of $\hat{q}_n^{\mathrm{VACLE}}$ and $\hat{q}_n^{\mathrm{TVACLE}}$ over 500 independent replications for Model 2 and 4, with the unknown σ^2 whose true value is 1.

		$\hat{q}_n^{\mathrm{VACLE}}$			$\hat{q}_n^{\mathrm{TVACLE}}$			$\hat{\sigma}^2$	
	(p,n)	Mean	MSE	$\hat{q}_n^{\mathrm{VACLE}} \neq q$	Mean	MSE	$\hat{q}_n^{\mathrm{TVACLE}} \neq q$	Mean	MSE
Model 2	$(50,200)$	4.002	0.002	0.002	4.012	0.012	0.012	1.0513	0.0033
	$(200,800)$	4.002	0.002	0.002	4.014	0.014	0.014	1.0119	0.0002
	$(100,100)$	3.326	2.346	0.258	3.966	0.038	0.038	1.0326	0.0022
	$(200,200)$	3.96	0.176	0.04	4.006	0.006	0.006	1.0169	0.0006
	$(200,100)$	2.334	5.726	0.616	3.71	0.306	0.282	1.0205	0.0008
	$(400,200)$	3.266	2.454	0.292	3.962	0.038	0.038	1.0094	0.0002
Model 4	$(50,200)$	6.01	0.01	0.01	6.01	0.01	0.01	1.0788	0.0069
	$(200,800)$	6.002	0.002	0.002	6.022	0.022	0.022	1.0181	0.0004
	$(100,100)$	4.082	10.362	0.388	5.878	0.142	0.112	1.0555	0.0042
	$(200,200)$	5.846	0.938	0.034	6	0	0	1.0256	0.0009
	$(400,200)$	4.524	8.064	0.306	5.958	0.042	0.042	1.0165	0.0004
	$(800,400)$	5.822	0.966	0.034	6.01	0.01	0.01	1.0079	9×10^{-5}

4.2 Numerical studies on large dimensional auto-covariance matrices

To estimate the number of factors in Model (3.16), Li et al. (2017) introduced the following ratio-based estimator,

$$\hat{q}_T^{\mathrm{LWY}} := \min\{i \geq 1 : \ \hat{\lambda}_{i+1}/\hat{\lambda}_i > 1 - d_T \text{ and } \hat{\lambda}_{i+2}/\hat{\lambda}_{i+1} > 1 - d_T\} - 1, \quad (4.26)$$

where $\hat{\lambda}_i, 1 \leq i \leq p$, are in descending order and d_T is a tuning parameter selected as that in Section 3.1 of Li et al. (2017). We use $\hat{q}_T^{\mathrm{LWY}}$ as the competitor to examine the performance of $\hat{q}_n^{\mathrm{TVACLE}}$. For the ratio $p/T = y$, we consider two values $y = 0.5$ and $y = 2$. The dimension $p = 100, 200, 300, 400$ and 500. In each case, we repeat the experiment 500 times. To be fair and concise, we conduct the simulation with two models as follows. The model structure is the same as in Lam and Yao (2012) and Li et al. (2017): for $1 \leq t \leq T$

$$y_t = \mathbf{A}x_t + \varepsilon_t, \ \varepsilon_t \sim \mathbb{N}_p(\mathbf{0}, \mathbf{I}_p), \ x_t = \Theta x_{t-1} + e_t, \ e_t \sim \mathbb{N}_k(\mathbf{0}, \Gamma), \quad (4.27)$$

where $\mathbf{A} \in \mathbb{R}^{p \times q}$ is the factor loading matrix and $\{\varepsilon_t\}$ is a white noise sequence with unit variance $\sigma^2 = 1$. As in Li et al. (2017), $\mathbf{A}$ and Γ take the forms as $\mathbf{A} = (\mathbf{I}_q, \mathbf{O}_{(p-q)\times q})^\top$, $\Gamma = \mathrm{diag}(2, 2, \cdots, 2)$. We manipulate the strength of factors by adjusting the matrix Θ in different models as follows:

Model 5. This model is the same as *Scenario III* in Li et al. (2017). There are $q = 3$ factors whose theoretical limits equal $(7.726, 5.496, 3.613)$ in the case of $y = 0.5$ and $(23.744, 20.464, 17.970)$ in the case of $y = 2$. The upper

edge b_1 of the supports in these two cases are respectively 2.773 and 17.637. $q = 3$ factors are identifiable, and $\Theta = \text{diag}(0.6, -0.5, 0.3)$.

Model 6. This model has more factors. There are $q = 6$ factors with identical strength, and their theoretical limits are 5.496 in the case of $y = 0.5$ and 20.464 in the case of $y = 2$. Because these limits exceed their corresponding upper edge b_1, all $q = 6$ factors are identifiable in theory with $\Theta = \text{diag}(0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$.

All parameters in the simulations share the same settings of parameters in Section 4.1 where we conduct numerical studies for spiked population models. These parameters in the TVACLE are shown in Table 6.

Table 6: Parameters in the TVACLE.

τ	c_T	κ_T	k_1	k_2	L
0.5	$\sqrt{\log\log T}[\mathrm{q}_{p,T}(0.95) - \mathrm{q}_{p,T}(0.05)] - \mathrm{m}_{p,T}$	$p^{-2/3}\log\log p$	5	5	20

Table 7: Mean, mean squared error and empirical distribution of $\hat{q}_T^{\text{LWY}}$ and $\hat{q}_T^{\text{TVACLE}}$ over 500 independent replications for Model 5.

	p	100	200	300	400	500	p	100	200	300	400	500
	$T = 2p$	200	400	600	800	1000	$T = 0.5p$	50	100	150	200	250
$\hat{q}_T^{\text{LWY}}$	$\hat{q} = 0$	0.024	0.002	0	0	0	$\hat{q} = 0$	0.53	0.238	0.234	0.138	0.054
	$\hat{q} = 1$	0.028	0	0	0	0	$\hat{q} = 1$	0.326	0.412	0.38	0.36	0.282
	$\hat{q} = 2$	0.384	0.138	0.05	0.014	0.008	$\hat{q} = 2$	**0.136**	**0.32**	**0.356**	**0.464**	**0.572**
	$\hat{q} = 3$	**0.544**	**0.85**	**0.948**	**0.976**	**0.986**	$\hat{q} = 3$	**0.008**	**0.03**	**0.03**	**0.036**	**0.092**
	$\hat{q} \geq 4$	0.02	0.01	0.002	0.01	0.006	$\hat{q} \geq 4$	0	0	0	0.002	0
	Mean	2.508	2.866	2.952	2.996	2.998	Mean	0.622	1.142	1.182	1.404	1.702
	MSE	0.732	0.166	0.052	0.024	0.014	MSE	6.21	4.11	3.982	3.148	2.186
$\hat{q}_T^{\text{TVACLE}}$	$\hat{q} = 0$	0	0	0	0	0	$\hat{q} = 0$	0.02	0.002	0	0.002	0
	$\hat{q} = 1$	0	0	0	0	0	$\hat{q} = 1$	0.332	0.182	0.116	0.104	0.054
	$\hat{q} = 2$	0.196	0.02	0.014	0.008	0.002	$\hat{q} = 2$	**0.584**	**0.698**	**0.688**	**0.73**	**0.676**
	$\hat{q} = 3$	**0.782**	**0.948**	**0.974**	**0.964**	**0.974**	$\hat{q} = 3$	**0.062**	**0.116**	**0.196**	**0.16**	**0.268**
	$\hat{q} \geq 4$	0.022	0.032	0.012	0.028	0.024	$\hat{q} \geq 4$	0.002	0.002	0	0.004	0.002
	Mean	2.826	3.012	2.998	3.02	3.022	Mean	1.694	1.934	2.08	2.06	2.218
	MSE	0.218	0.052	0.026	0.036	0.026	MSE	2.094	1.446	1.152	1.168	0.894

4.2 Numerical studies on large dimensional auto-covariance matrices

Table 8: Mean, mean squared error and empirical distribution of $\hat{q}_T^{\mathrm{LWY}}$ and $\hat{q}_T^{\mathrm{TVACLE}}$ over 500 independent replications for Model 6.

	p	100	200	300	400	500	p	100	200	300	400	500
	$T=2p$	200	400	600	800	1000	$T=0.5p$	50	100	150	200	250
$\hat{q}_T^{\mathrm{LWY}}$	$\hat{q}=0$	0.156	0.104	0.072	0.098	0.054	$\hat{q}=0$	0.226	0.202	0.26	0.24	0.134
	$\hat{q}=1$	0.178	0.154	0.124	0.146	0.076	$\hat{q}=1$	0.418	0.35	0.326	0.304	0.28
	$\hat{q}=2$	0.19	0.154	0.114	0.104	0.062	$\hat{q}=2$	0.296	0.314	0.262	0.236	0.312
	$\hat{q}=3$	0.162	0.112	0.068	0.106	0.044	$\hat{q}=3$	0.06	0.122	0.134	0.17	0.188
	$\hat{q}=4$	0.13	0.006	0	0	0	$\hat{q}=4$	0	0.012	0.016	0.05	0.07
	$\hat{q}=5$	0.112	0.072	0	0	0	$\hat{q}=5$	0	0	0.002	0	0.014
	$\hat{q}=6$	**0.072**	**0.394**	**0.62**	**0.542**	**0.754**	$\hat{q}=6$	**0**	**0**	**0**	**0**	**0.002**
	$\hat{q}\geq 7$	0	0.004	0.002	0.004	0.01	$\hat{q}\geq 7$	0	0	0	0	0
	Mean	2.556	3.574	4.29	3.954	4.926	Mean	1.19	1.392	1.326	1.486	1.83
	MSE	15.196	11.166	8.13	9.806	5.242	MSE	23.862	22.192	22.974	21.746	18.802
$\hat{q}_T^{\mathrm{TVACLE}}$	$\hat{q}=0$	0	0	0	0	0	$\hat{q}=0$	0	0	0	0	0
	$\hat{q}=1$	0	0	0	0	0	$\hat{q}=1$	0.01	0	0	0	0
	$\hat{q}=2$	0	0	0	0	0	$\hat{q}=2$	0.206	0.07	0.03	0.008	0.008
	$\hat{q}=3$	0.004	0	0	0	0	$\hat{q}=3$	0.586	0.496	0.33	0.224	0.13
	$\hat{q}=4$	0.066	0.002	0	0	0	$\hat{q}=4$	0.19	0.414	0.546	0.574	0.554
	$\hat{q}=5$	0.418	0.038	0	0	0	$\hat{q}=5$	0.008	0.02	0.094	0.188	0.294
	$\hat{q}=6$	**0.51**	**0.946**	**0.99**	**0.984**	**0.97**	$\hat{q}=6$	**0**	**0**	**0**	**0.006**	**0.014**
	$\hat{q}\geq 7$	0.002	0.014	0.01	0.016	0.03	$\hat{q}\geq 7$	0	0	0	0	0
	Mean	5.44	5.972	6.01	6.016	6.03	Mean	2.98	3.384	3.704	3.96	4.176
	MSE	0.72	0.06	0.01	0.016	0.03	MSE	9.588	7.26	5.728	4.628	3.808

From Table 7, we can see that when $T=2p$, $\hat{q}_T^{\mathrm{LWY}}$ works well. That is, when T is large, $\hat{q}_T^{\mathrm{LWY}}$ shows good performance, whilst when T is not large, it tends to underestimate the true number q. Our method outperforms $\hat{q}_T^{\mathrm{LWY}}$. Although when T is small, the true value is somewhat underestimated, but still, with a high proportion, to be two or greater. Table 8 shows that for Model 6 with equal spikes, when $T=2p$, the performance of $\hat{q}_T^{\mathrm{LWY}}$ is not encouraging, and when $T=0.5p$, the underestimation problem becomes very serious, with a very high proportion having $\hat{q}_T^{\mathrm{LWY}} \leq 2$. In contrast, our method performs well when $T=2p$ and when $T=0.5p$; underestimation still occurs, but it is much less serious than $\hat{q}_T^{\mathrm{LWY}}$ in the sense that $\hat{q}>2$ with high proportion. Overall, our estimator $\hat{q}_T^{\mathrm{TVACLE}}$ is superior to $\hat{q}_T^{\mathrm{LWY}}$

in these limited simulations.

4.3 Numerical studies on large dimensional spiked Fisher matrices

Because the TVACLE has been demonstrated to outperform the VACLE overall, we only consider the comparison between $\hat{q}_n^{\mathrm{TVACLE}}$ and the estimator $\hat{q}_n^{\mathrm{WY}}$ introduced by Wang and Yao (2017). Sharing notations in Section 3.2, the estimator $\hat{q}_n^{\mathrm{WY}}$ can be written as

$$\hat{q}_n^{\mathrm{WY}} := \max\{i : \hat{\lambda}_i \geq b_2 + d_n\}, \tag{4.28}$$

where d_n was recommended to be $(\log\log p)p^{-2/3}$ in their paper.

As a Fisher matrix $\mathbf{F}_n = \mathbf{S}_1\mathbf{S}_2^{-1}$ involves two random matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, its eigenvalues are more dispersed, with wider range of the support, than the spiked sample covariance matrices and auto-covariance matrices. The aforementioned automatic procedure for ridge selection would then generate a larger c_n, and this in turn increases the value at the "valley". Hence, we use a larger threshold $\tau = 0.8$ to avoid underestimation. Further, in the following Model 7, we set the ridge $c_n^{(3)} = \sqrt{\log\log p}[\mathrm{q}_{p,n}(0.95) - \mathrm{q}_{p,n}(0.05)] - \mathrm{m}_{p,n}$, whilst for Model 8 with dramatically-fluctuated extreme eigenvalues, we need to set $c_n^{(3)} = \sqrt{\log\log p}[\mathrm{q}_{p,n}(0.8) - \mathrm{q}_{p,n}(0.05)] - \mathrm{m}_{p,n}$ to avoid too large ridge. Other parameters in $\hat{q}_n^{\mathrm{TVACLE}}$ share the same settings with that of the spikd population models, which are shown in Table 9.

Table 9: Parameters in the TVACLE.

τ	c_n	κ_n	k_1	k_2	L
0.8	$c_n^{(3)}$	$p^{-2/3}\log\log p$	5	5	20

Again, for a fair comparison, we design two models, one was used by Wang and Yao (2017) and the other is with weaker spikes. For $y = p/T$ and $c = p/n$, we set $(0.5, 0.2)$ and $(0.2, 0.5)$ for the respective models. The dimension p takes values of $50, 100, 150, 200$ and 250. For each combination (p, T, n), the experiment is repeated 500 times. Consider the number of spikes to be $q = 3$ and $\mathbf{A}$ to be a $p \times 3$ matrix as:

$$\begin{pmatrix} \sqrt{\alpha_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\frac{\alpha_2}{2}} & \sqrt{\frac{\alpha_2}{2}} & 0 & \cdots & 0 \\ 0 & \sqrt{\frac{\alpha_3}{2}} & -\sqrt{\frac{\alpha_3}{2}} & 0 & \cdots & 0 \end{pmatrix}^{\top}_{3\times p}, \tag{4.29}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ assumes different values in two models. Assume the covariance matrix $\mathrm{Cov}(u_i) = \mathbf{I}_3$ and $\Sigma_2 = \mathrm{diag}(1, \cdots, 1, 2, \cdots, 2)$, where "1" and "2" both have multiplicity $p/2$. The two models are:

Model 7. Let $\alpha = (10, 5, 5)$, $(y, c) = (0.5, 0.2)$, which is Model 1 in Wang and Yao (2017). The matrix $\Sigma_1\Sigma_2^{-1}$ has three spikes $\lambda_1 = 11$ and $\lambda_2 = \lambda_3 = 6$ that are all significantly larger than the identifiability bound $\sqrt{b_2} = (1-y)^{-1}(1+\sqrt{c+y-cy}) \approx 3.55$.

Model 8. Let $\alpha = (10, 2, 2)$, $(y, c) = (0.2, 0.5)$. The matrix $\Sigma_1\Sigma_2^{-1}$ then also has three spikes $\lambda_1 = 11$ and $\lambda_2 = \lambda_2 = 3$ larger than the identifiability

bound $\sqrt{b}_2 = (1-y)^{-1}(1+\sqrt{c+y-cy}) \approx 2.22$. Then $\lambda_2 = \lambda_2 = 3$ are relatively more difficult to detect.

Table 10: Mean, mean squared error and empirical distribution of $\hat{q}_n^{\mathrm{WY}}$ and $\hat{q}_n^{\mathrm{TVACLE}}$ for Model 7.

	(p,T,n)	Mean	MSE	$\hat{q}=0$	$\hat{q}=1$	$\hat{q}=2$	$\hat{q}=3$	$\hat{q}=4$
$\hat{q}_n^{\mathrm{WY}}$	$(50,100,250)$	2.344	0.732	0	0.034	0.592	**0.37**	0.004
	$(100,200,500)$	2.672	0.352	0	0.004	0.328	**0.66**	0.008
	$(150,300,750)$	2.822	0.194	0	0	0.186	**0.806**	0.008
	$(200,400,1000)$	2.964	0.092	0	0	0.064	**0.908**	0.028
	$(250,500,1250)$	2.96	0.068	0	0	0.054	**0.932**	0.014
$\hat{q}_n^{\mathrm{TVACLE}}$	$(50,100,250)$	2.364	0.7	0	0.028	0.584	**0.384**	0.004
	$(100,200,500)$	2.688	0.336	0	0.004	0.312	**0.676**	0.008
	$(150,300,750)$	2.842	0.182	0	0	0.17	**0.818**	0.012
	$(200,400,1000)$	2.974	0.082	0	0	0.054	**0.918**	0.028
	$(250,500,1250)$	2.964	0.064	0	0	0.05	**0.936**	0.014

Table 11: Mean, mean squared error and empirical distribution of $\hat{q}_n^{\mathrm{WY}}$ and $\hat{q}_n^{\mathrm{TVACLE}}$ for Model 8.

	(p,T,n)	Mean	MSE	$\hat{q}=0$	$\hat{q}=1$	$\hat{q}=2$	$\hat{q}=3$	$\hat{q}=4$
$\hat{q}_n^{\mathrm{WY}}$	$(50,250,100)$	2.114	1.07	0	0.09	0.708	**0.2**	0.002
	$(100,500,200)$	2.302	0.79	0	0.046	0.606	**0.348**	0
	$(150,750,300)$	2.498	0.538	0	0.018	0.466	**0.516**	0
	$(200,1000,400)$	2.622	0.394	0	0.006	0.368	**0.624**	0.002
	$(250,1250,500)$	2.692	0.324	0	0.004	0.304	**0.688**	0.004
$\hat{q}_n^{\mathrm{TVACLE}}$	$(50,250,100)$	2.238	0.898	0	0.064	0.638	**0.294**	0.004
	$(100,500,200)$	2.462	0.602	0	0.03	0.48	**0.488**	0.002
	$(150,750,300)$	2.71	0.314	0	0	0.302	**0.686**	0.012
	$(200,1000,400)$	2.82	0.232	0	0.002	0.2	**0.774**	0.024
	$(250,1250,500)$	2.904	0.164	0	0	0.13	**0.836**	0.034

The results reported in Tables 10 and 11 show that $\hat{q}_n^{\mathrm{TVACLE}}$ shows overall better performance than $\hat{q}_n^{\mathrm{WY}}$. For Model 8, $\hat{q}_n^{\mathrm{TVACLE}}$ is superior to $\hat{q}_n^{\mathrm{WY}}$ when the signals are relatively weak.

5. A real data example

Consider a data set of the daily prices of 100 stocks (see Li et al. (2017)). This dataset includes the stock prices of the S&P500 from 2005-01-03 to

2006-12-29. Except for incomplete data, every stock has 502 observations of log returns. Thus, $T = 502$, $p = 100$, and then $c = p/T \approx 0.2$.

Denote $y_t \in \mathbb{R}^p$, $1 \le t \le T$, as the t-th observation of the log return of these 100 stocks, and we then obtain its lag-1 sample auto-covariance matrix $\hat{\Sigma}_y$ and the matrix $\hat{\mathbf{M}}_y = \hat{\Sigma}_y \hat{\Sigma}_y^\top$ as formulated in Section 3.1. Use $\hat{q}^{\mathrm{TVACLE}}$ and $\hat{q}^{\mathrm{LWY}}$ in Li et al. (2017) to determine the number of factors. All parameters in these two methods share the same settings with the simulation parts. Besides, the unknown σ^2 in $\hat{q}^{\mathrm{TVACLE}}$ is estimated by the method (4.25) after necessary modification as we commented in Remark 4.1. We can see that the two largest eigenvalues of $\hat{\mathbf{M}}_y$ are 7.17×10^{-7} and 2.01×10^{-7}, and the third to the 40-th eigenvalues are shown in Figure 3.

Figure 3: Eigenvalues of $\hat{\mathbf{M}}_y$ from $\hat{\lambda}_3$ to $\hat{\lambda}_{40}$

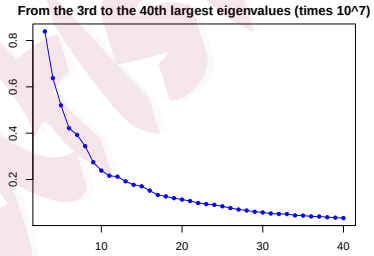
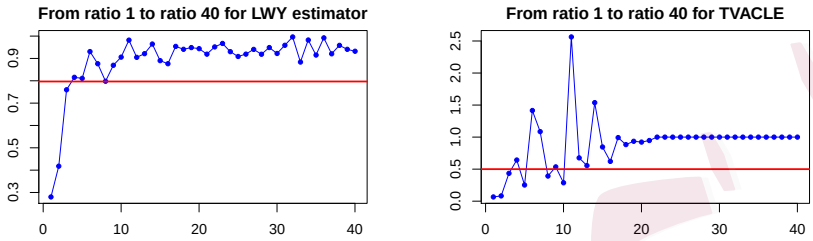


Figure 4 shows that $\hat{q}^{\mathrm{LWY}} = 5$. However, as shown in Figure 3, the gap between the 5th eigenvalue and several following eigenvalues is evidently insignificant. As $\hat{q}^{\mathrm{LWY}}$ is based on the magnitudes of the next two consecutive ratios. If eigenvalue multiplicity occurs, $\hat{q}^{\mathrm{LWY}}$ could likely select a value smaller than the true number.

Figure 4: Ratios $\hat{\lambda}_{i+1}/\hat{\lambda}_i$ in Li et al. (2017) and Ratios $\hat{r}_i^{\mathrm{TR}}$ in the TVACLE, $1 \le i \le 40$.



When the TVACLE is used, $\hat{q}^{\mathrm{TVACLE}} = 10$. Figure 4 shows that the 11th ratio is much larger than the 10th ratio, although some values get smaller. Note that in this example, $c \sim 0.2$ and the ridge is relatively small, which would not very much dominate the difference between the eigenvalues and thus some oscillating values remain after the 10th ratio.

It is considered that $\hat{q}^{\mathrm{LWY}}$ would neglect several factors and likely result in an underestimation. For a real data example, we usually cannot give a definitive answer. However, our method could provide an estimation that would be relatively conservative but necessary, particularly in the initial stage of data analysis; otherwise, an excessively parsimonious model would cause misleading conclusions.

6. Concluding remarks

In this paper, we propose a valley-cliff criterion for spiked models, and the method can be applied to other order determination problems when the dimension is proportional to the sample size, such as those in sufficient

dimension reduction if the corresponding asymptotics can be well investigated. The method is for the case with a fixed order q. An extension to the case with diverging q will be proposed in our future work. Besides, our method applies to general Σ cases provided that we have the true value or a reliable estimation of the right edge corresponding to Σ. But it is generally hard to have a good estimation of the right edge in real data analysis when Σ is unknown.

Supplementary Materials

Proofs and technical details are contained in the supplementary materials.

Acknowledgements

The authors gratefully acknowledge the support from a grant from the University Grants Council of Hong Kong, Hong Kong, and an SNFC grant (NSFC11671042) from the National Natural Science Foundation of China.

References

Antoine, B. and Renault, E. (2012). Efficient minimum distance estimation with multiple rates of convergence. *J. Economet.* **170**, 350-367.

Bai, Z. and Yao, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincaré Probab. Statist.* **44**, 447-474.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97**, 1382-1408.

Benaych-Georges, F., Guionnet, A. and Maida, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16**, 1621-1662.

Benaych-Georges, F. and Nadakuditi, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227**, 494-521.

Cai, T. T., Han, X. and Pan, G. (2020). Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Ann. Statist.* **48**, 1255-1280.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295-327.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104**, 682-693.

Knowles, A. and Yin, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169**, 257-352.

Kritchman, S. and Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometr. Intell. Lab. Syst.* **94**, 19-32.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.* **40**, 694-726.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86**,

316-327.

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.

Luo, W. and Li, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika* **103**, 875-887.

Nadler, B. (2010). Nonparametric detection of signals by information theoretic criteria: performance analysis and an improved estimator. *IEEE Trans. Signal Process.* **58**, 2746-2756.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* **77**, 1447-1479.

Passemier, D., Li, Z. and Yao, J. (2017). On estimation of the noise variance in high dimensional probabilistic principal component analysis. *J. Roy. Statist. Soc. Ser. B* **79**, 51-67.

Passemier, D. and Yao, J. (2012). On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices: Theory and Applications* **1**, 1150002.

Passemier, D. and Yao, J. (2014). Estimation of the number of spikes, possibly equal, in the high-dimensional case. *J. Multivariate Anal.* **127**, 173-183.

Péché, S. (2006). The largest eigenvalue of small rank perturbations of hermitian random matrices. *Probab. Theory Related Fields* **134**, 127-173.

Pillai, N. S. and Yin, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24**, 935-1001.

Ulfarsson, M. O. and Solo, V. (2008). Dimension estimation in noisy PCA with SURE and random matrix theory. *IEEE Trans. Signal Process.* **56**, 5804-5816.

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

Xia, Q., Xu, W. and Zhu, L. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statist. Sinica* **25**, 1025-1044.

Yao, J., Zheng, S. and Bai, Z. (2015). *Large sample covariance matrices and high-dimensional data analysis*. Cambridge University Press.

Zhu, L., Miao, B. and Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101**, 630-643.

Zhu, L., Wang, T., Zhu, L. and Ferré, L. (2010). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika* **97**, 295-304.

Yicheng Zeng, Department of Mathematics, Hong Kong Baptist University, Hong Kong

E-mail: statzyc@gmail.com

Lixing Zhu, Department of Mathematics, Hong Kong Baptist University, Hong Kong

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

E-mail: lzhu@hkbu.edu.hk