Topic 18: Eigenvalue and Spike Models

by Sai Zhang

Key points: A spike model can be used to determine the latent data structure.

Disclaimer: *The note is built on Prof. Jinchi Lv's lectures of the course at USC, DSO 607, High-Dimensional Statistics and Big Data Problems.*

18.1 Motivation

Consider n independent observations $\mathbf{X}_i \in \mathbb{R}^p$ drawn from a $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, then the covariance can be decomposed into 2 parts, white noise and low rank

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}_i) = \mathbf{I} + \sum_{k=1}^{M} \theta_k v_k v_k' = \boldsymbol{\Sigma}_0 + \boldsymbol{\Phi}$$

where M denotes the number of spikes in the distribution of eigenvalues. The idea is: spikes deviate from a reference model along a <u>small fixed number</u> of unknown directions. If $\boldsymbol{\Phi} = \mathbf{0}$, then none of the sample eigenvalues is separated from the bulk.

Why a spike model is interesting? A spike model can help determine the latent dimension of the data, some examples being

- Principal component analysis (PCA): spikes are related to the directions of the most variations of the data, i.e., the principal components
- Clustering model: M spikes is equivalent to $M + 1$ clusters
- Economic significance: M is related to the number of factor loadings

Then the question is threefold:

- How to determine M
- How to estimate v_k
- How to test θ_k

Under rank one alternative, we would like to test the hypothesis

$$H_1 : \boldsymbol{\Sigma} = \mathbf{I}_p + \theta \boldsymbol{v} \boldsymbol{v}', \theta > 0$$

against the null

$$H_0 : \boldsymbol{\Sigma} = \mathbf{I}_p$$

with the key assumptions:

A1 Gaussian error

A2 large p: $p \le n$ but allows $p/n \to \gamma \in (0, 1)$

Under these assumptions, for the $n \times p$ data matrix $\mathbf{X} = \left(\mathbf{X}_1' \quad \cdots \quad \mathbf{X}_n'\right)'$, $\mathbf{X}'\mathbf{X}$ has a $p-$dimensional **Wishart** distribution $W_p(n, \mathbf{\Sigma})$ with the degree of freedom n and covariance matrix $\mathbf{\Sigma}$, which is a *random matrix*.

If $\mathbf{Y} = \mathbf{M} + \mathbf{X}$, that is, the sum of the *random matrix* $\mathbf{X}$ and a *deterministic matrix* $\mathbf{M}$ (also $n \times p$), then $\mathbf{Y}'\mathbf{Y}$ has a $p-$dimensional Wishart distribution $W_p(n, \mathbf{\Sigma}, \mathbf{\Psi})$ with n degrees of freedom, covariance matrix $\mathbf{\Sigma}$ and non-centrality matrix $\mathbf{\Psi} = \mathbf{\Sigma}^{-1}\mathbf{M}'\mathbf{M}$.

Definition 18.1.1: Density of Wishart Distribution

The PDF of Wishart distribution is defined as

$$f(\mathbf{X}) = \frac{1}{2^{np/2}\Gamma_p\left(\frac{n}{2}\right)|\mathbf{\Sigma}|^{n/2}} \, |\mathbf{X}|^{(n-p-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{X}\right)\right)$$

where $\mathbf{X}$ is a symmetric positive semidefinite and $\Gamma_p\left(\frac{n}{2}\right)$ is a multivariate gamma function such that

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{\frac{p(p-1)}{4}}\prod_{j=1}^{p}\Gamma\left(\frac{n}{2} - \frac{j-1}{2}\right)$$

Notice that the sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}'\mathbf{X}$ is just a scaled version of Wishart distribution

$$n\mathbf{S} = \mathbf{X}'\mathbf{X} \sim W_p(n, \mathbf{\Sigma})$$

For $\mathbf{\Sigma} = \mathbf{I}_p$, the empirical distribution fo eigenvalues converges to Marcenko-Pastur distribution

$$f^{\mathrm{MP}}(x) = \frac{1}{2\pi\gamma x}\sqrt{(b_+ - x)(x - b_-)}$$

where $b_{\pm} = (1 \pm \sqrt{\gamma})^2$. Then:

- under $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$, we have

$$n^{2/3}\left(\frac{\lambda_1 - \mu(\gamma)}{\sigma(\gamma)}\right) \xrightarrow{d} \mathrm{TW}_1$$

  where TW_1 is the Tracy-Widom distribution

- under $H_1 : \mathbf{\Sigma} = \mathbf{I}_p + \theta\boldsymbol{v}\boldsymbol{v}'$, $\theta > 0$, if θ is strong ($\theta \gg \sqrt{\gamma}$), then

$$n^{1/2}\left(\frac{\lambda_1 - \rho(\theta, \gamma)}{\tau(\theta, \gamma)}\right) \xrightarrow{d} \mathcal{N}(0, 1)$$

Here, the largest eigenvalue test is the best test. **But** when the signal is weak ($0 \leq \theta < \sqrt{\gamma}$), the largest eigenvalue under the alternative converges to the same distribution as null:

$$n^{2/3}\left(\frac{\lambda_1 - \rho(\theta, \gamma)}{\tau(\theta, \gamma)}\right) \xrightarrow{d} \mathrm{TW}_1$$

which means that the largest eigenvalue test *fails*. On top of this, **resampling** also fails when p is large.

Next, we develop another test to cope with these problems.

Figure 18.1: Failure of Resampling Test ($n = p = 100$)

18.2 Johnstone and Onatski (2020)

Consider the basic equation of classical multivariate statistics:

$$\det(\mathbf{H} - \mathbf{x}\mathbf{E}) = 0 \tag{18.1}$$

with $p \times p$ matrices

$$n_1\mathbf{H} = \sum_{k=1}^{n_1} \mathbf{x}_k\mathbf{x}_k' \qquad\qquad \textit{hypothesis SS}$$

$$n_1\mathbf{E} = \sum_{k=1}^{n_1} \mathbf{z}_k\mathbf{z}_k' \qquad\qquad \textit{error SS}$$

The solution $\mathbf{x}$ is generalized eigenvalues $\{\lambda_i\}_{i=1}^p$, which are the eigenvalue of F-ratio $\mathbf{E}^{-1}\mathbf{H}$. Johnstone and Onatski (2020) summarized 5 topics using $\mathbf{E}^{-1}\mathbf{H}$ relying on the five most common hypergeometric functions[1] $_p\mathcal{F}_q$

---

[1]Hypergeometric functions are:
- scalar inputs

$$_p\mathcal{F}_q(a, b; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_p)_k} \frac{x^k}{k!}$$

where $(a_j)_k$ are generalized Pochhammer symbols
- single matrix inputs, where $\mathbf{S}$ is symmetric and usually diagonal

$$_p\mathcal{F}_q(a, b; \mathbf{S}) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa \cdots (a_p)_\kappa}{(b_1)_\kappa \cdots (b_p)_\kappa} \frac{C_\kappa(\mathbf{S})}{k!}$$

where C_k are the zonal polynomials. Easily, $_0\mathcal{F}_0(\mathbf{S}) = e^{\text{tr}(\mathbf{S})}$, $_1\mathcal{F}_0(a, \mathbf{S}) = |\mathbf{I} - \mathbf{S}|^{-a}$
- two matrix inputs, where $\mathbf{S}, \mathbf{T}$ are both symmetric

$$_p\mathcal{F}_q(a, b; \mathbf{S}, \mathbf{T}) = \int_{O(p)} {_p\mathcal{F}_q}(a, b; \mathbf{SUTU}')(d)\mathbf{U}$$

Table 18.1: 5 Statistical Methods

		Statistical method	$n_1 \mathbf{H}$	$n_2 \mathbf{E}$	Univariate Analog
$_0\mathcal{F}_0$	PCA	Principal components analysis	$W_p(n_1, \mathbf{\Sigma} + \mathbf{\Phi})$	$n_2\mathbf{\Sigma}$	χ^2
$_1\mathcal{F}_0$	SigD	Signal detection	$W_p(n_1, \mathbf{\Sigma} + \mathbf{\Phi})$	$W_p(n_2, \mathbf{\Sigma})$	non-central χ^2
$_0\mathcal{F}_1$	REG$_0$	Multivariate regression, with known error	$W_p(n_1, \mathbf{\Sigma}, n_1\mathbf{\Phi})$	$n_2\mathbf{\Sigma}$	F
$_1\mathcal{F}_1$	REG	Multivariate regression, with unknown error	$W_p(n_1, \mathbf{\Sigma}, n_1\mathbf{\Phi})$	$W_p(n_2, \mathbf{\Sigma})$	non-central F
$_2\mathcal{F}_1$	CCA	Canonical correlation analysis	$W_p(n_1, \mathbf{\Sigma}, \mathbf{\Phi}(\mathbf{Y}))$	$W_p(n_2, \mathbf{\Sigma})$	$\frac{r^2}{1-r^2}$

For $_0\mathcal{F}_0$ and $_0\mathcal{F}_1$, $\mathbf{E}$ is deterministic, $\mathbf{\Sigma}$ is known, n_2 disppears, otherwise $\mathbf{E}$ is independent of $\mathbf{H}$.

18.2.1   Definitions and global assumptions

Let $\mathbf{Z}$ be an $n \times p$ data matrix with rows (observations) drawn **i.i.d.** from $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$, and a deterministic matrix $\mathbf{M}$ of $n \times p$, then for $\mathbf{Y} = \mathbf{M} + \mathbf{Z}$,

- $\mathbf{H} = \mathbf{Y}'\mathbf{Y}$ has a p dimensional Wishart distribution $W_p(n, \mathbf{\Sigma}, \mathbf{\Psi})$ with n degrees of freedom, covariance matrix $\mathbf{\Sigma}$ and non-centrality matrix $\mathbf{\Psi} = \mathbf{\Sigma}^{-1}\mathbf{M}'\mathbf{M}$
- the corresponding central Wishart distribution with $\mathbf{M} = \mathbf{0}$ is $W_p(n, \mathbf{\Sigma})$

Johnstone and Onatski (2020) assume a relative low dimensionality $p \leq \min\{n_1, n_2\}$ where n_1, n_2 are the degrees of freedom as in Table 18.1, where

- $p \leq n_2$ ensures almost sure invertibility of matrix $\mathbf{E}$ in Equation 18.1
- $p \leq n_1$ is not essential, but reduces the number of various situations of consideration.

18.2.2   5 classes of problems

With these assumptions, they established a unified statistical problem **symmetric matrix denoising (SMD)** that can be linked to the 5 classes of problems:

PCA n_1 i.i.d. observations drawn from $\mathcal{N}_p(\mathbf{0}, \mathbf{\Omega})$ to test the hull hypothesis that the population covariance $\mathbf{\Omega} = \mathbf{\Sigma}$, with the alternative of interest being

$$\mathbf{\Omega} = \mathbf{\Sigma} + \mathbf{\Phi}, \quad \text{with } \mathbf{\Phi} = \theta\boldsymbol{\phi}\boldsymbol{\phi}'$$

where $\theta > 0$, $\boldsymbol{\phi}$ are unknown, and $\boldsymbol{\phi}$ is normalized s.t. $\left\|\mathbf{\Sigma}^{-1/2}\boldsymbol{\phi}\right\| = 1$. W.L.O.G., assume $\mathbf{\Sigma} = \mathbf{I}_p$, then under the alternative, the first principal component explains a larger portion of the variation than the other principal components. Re-formulate the hypotheses in terms of the spectral *spike* parameter θ, we have

$$H_0 : \theta_0 = 0 \qquad\qquad H_1 : \theta_0 = \theta > 0 \qquad\qquad (18.2)$$

where θ_0 is the true value of the *spike*. A **maximal invariant statistic** consists of the solutions $\lambda_1 \geq \cdots \geq \lambda_p$ of Equation 18.1 with

- $n_1\mathbf{H}$ equal to the sample covariance matrix
- $\mathbf{E} = \mathbf{\Sigma}$

SigD   Now consider testing the **equality** of covariance matrices $\mathbf{\Omega}$ and $\mathbf{\Sigma}$, corresponding to 2 independent p−dimensional mean-zero Gaussian samples of size n_1 and n_2, with the alternative still

$$\mathbf{\Omega} = \mathbf{\Sigma} + \mathbf{\Phi}, \quad \text{with } \mathbf{\Phi} = \theta\boldsymbol{\phi}\boldsymbol{\phi}'$$

and again, assume $\Sigma = I_p$ (but NOT necessarily known), here, instead of Equation 18.1, consider

$$\det\left(\mathbf{H} - \lambda\left(\mathbf{E} + \frac{n_1}{n_2}\mathbf{H}\right)\right) = 0 \tag{18.3}$$

naturally, SigD reduces to PCA as $n_2 \to \infty$ while n_1 and p held constant.

REG$_0$ Next, consider a linear regression with multivariate response

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

with known covariance matrix Σ of the i.i.d. Gaussian rows of the error matrix $\boldsymbol{\epsilon}$. Here, to test linear restrictions on the matrix of coefficients $\boldsymbol{\beta}$, we can split the matrix of transformed response variables $\mathbf{Y}$ into 3 parts $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$, where

- $\mathbf{Y}_1$ is $n_1 \times p$ where p is the number of response variables, n_1 is the number of linear restrictions (per each of the p columns of matrix $\boldsymbol{\beta}$), under the null $H_0 : \mathbb{E}\mathbf{Y}_1 = 0$, versus the alternative

$$\mathbb{E}\mathbf{Y}_1 = \sqrt{n_1\theta}\,\psi\phi' \tag{18.4}$$

   where $\theta > 0$, $\left\|\Sigma^{-1/2}\phi\right\| = 1$ and $\|\psi\| = 1$
- $\mathbf{Y}_2$ is $(q - n_1) \times p$, where q is the number of regressors
- $\mathbf{Y}_3$ is $(T - q) \times p$, where T is the number of observations

In this case, tests can be based on the solutions $\lambda_1, \cdots, \lambda_p$ to

$$\det\left(\mathbf{H} - \lambda\mathbf{E}\right) = 0$$

where $\mathbf{H} = \mathbf{Y}_1'\mathbf{Y}_1/n_1$ and $\mathbf{E} = \Sigma$. The solutions represent a multivaraite analog of the difference between the sum of squared residuals in the restircted and unrestricted regressions. Under the null, $n_1\mathbf{H}$ is distributed as $W_p(n_1, \Sigma)$. Here,

$$n_1\mathbf{H} \sim W_p(n_1, \Sigma) \qquad\qquad\qquad \text{under } H_0$$
$$n_1\mathbf{H} \sim W_p(n_1, \Sigma, n_1\boldsymbol{\Phi}), \text{ where } \boldsymbol{\Phi} = \theta\Sigma^{-1}\phi\phi' \qquad\qquad \text{under } H_1$$

Again, W.L.O.G, assume $\Sigma = I_p$. This **canonical form** of REG$_0$ is essentially equivalent to the setting of **matrix denoising**

$$\mathbf{Y}_1 = \mathbf{M} + \mathbf{Z}$$

REG   Again, consider the linear regression

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

but **NOT** knowing the covariance matrix Σ of rows of $\boldsymbol{\epsilon}$. Here, the solutions again solve $\det\left(\mathbf{H} - \lambda\mathbf{E}\right) = 0$ with

$$\mathbf{H} = \mathbf{Y}_1'\mathbf{Y}_1/n_1, \quad \mathbf{E} = \mathbf{Y}_3'\mathbf{Y}_3/n_2$$

this represents a multivariate analog of the F ratio: the difference between the sum of squared residuals in the restricted and unrestricted regressions to the sum of squared residuals in the restricted regression. Again, as $n_2 \to \infty$, REG reduces to REG$_0$.

CCA   Consider testing for independence between Gaussian vectors $x_t \in \mathbb{R}^p$ and $y_t \in \mathbb{R}^{n_1}$, given zero-mean observations with $t = 1, \cdots, n_1 + n_2$. Partition the population and sample covariance matrices of the observations $(x_t', y_t')'$ into

$$\begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} \qquad\qquad \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} \end{pmatrix}$$

respsectively. Under $H_0 : \boldsymbol{\Sigma}_{xy} = \mathbf{0}$, while the alternative is

$$\boldsymbol{\Sigma}_{xy} = \sqrt{\frac{n_1 \theta}{n_1 \theta + n_1 + n_2}} \boldsymbol{\phi} \boldsymbol{\psi}' \tag{18.5}$$

where the nuisance parameters $\phi \in \mathbb{R}^p$ and $\psi \in \mathbb{R}^{n_1}$ are normalized s.t.

$$\left\| \boldsymbol{\Sigma}_{xx}^{-1/2} \boldsymbol{\phi} \right\| = \left\| \boldsymbol{\Sigma}_{yy}^{-1/2} \boldsymbol{\psi} \right\| = 1$$

And the test can be based on the squared sample canonical correlations $\lambda_1, \cdots, \lambda_p$ that solves

$$\det\left(\mathbf{H} - \lambda \mathbf{E} \right) = \mathbf{0}$$

with

$$\mathbf{H} = \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \qquad\qquad\qquad \mathbf{E} \mathbf{S}_{xx}$$

18.2.3   SMD

For a $\mathbf{X} = \boldsymbol{\Phi} + \mathbf{Z}/\sqrt{p}$ where $\mathbf{Z}$ is a noise matrix from the **Gaussian Orthogonal Ensemble** (GOE)[2] We seek to make inference about a symmetric rank-one *signal* matrix $\boldsymbol{\Phi} = \theta \boldsymbol{\phi} \boldsymbol{\phi}'$. The null and the alternative is again as in 18.2. The nuisance vector $\phi \in \mathbb{R}^p$ is normalized s.t. $\left\| \boldsymbol{\phi} \right\| = 1$.

The problem remains **invariant** under the multiplication of $\mathbf{X}$ from the left by an orthogonal matrix, and from the right by its transpose. A maximal invariant statistic consists of the solutions $\lambda_1, \cdots, \lambda_p$ to $\det\left(\mathbf{H} - \lambda \mathbf{E} \right) = \mathbf{0}$ with $\mathbf{H} = \mathbf{X}$ and $\mathbf{E} = \mathbf{I}_p$.

SMD can be viewed as a degenerate version of the 5 classes of problems, as shown in Figure 18.2:

- **SMD**, **PCA**, **REG$_0$**: random $\mathbf{H}$ and deterministic $\mathbf{E}$
- **PCA** and **SigD** are *parallel* to **REG$_0$**
- **CCA** has a different structure of $\mathbf{H}$ and $\mathbf{E}$

18.2.4   The likelihood ratios

The goal is to study the asymptotic behavior of likelihood ratios based on the observed eigenvalues

$$\Lambda = \operatorname{diag}\left\{ \lambda_1, \cdots, \lambda_p \right\}$$

---

[2]$\mathbf{Z}$ is from the GOE that it is **symmetric** and

$$\mathbf{Z}_{ii} \sim \mathcal{N}(0, 2) \qquad\qquad\qquad \mathbf{Z}_{ij} \sim \mathcal{N}(0, 1) \text{ if } i > j$$

Figure 18.2: SMD and 5 Classes of Statistical Problems

SMD

$\mathbf{H} = \frac{GOE}{\sqrt{p}} + \boldsymbol{\Phi}, \mathbf{E} = \mathbf{I}_p$

$n_1 \to \infty, \theta \to \theta\sqrt{p/n_1}$ $\qquad$ $n_1 \to \infty, \theta \to \theta\sqrt{p/n_1}$

PCA

$n_1\mathbf{H} = W_p(n_1, \mathbf{I}_p + \boldsymbol{\Phi}), \mathbf{E} = \mathbf{I}_p$

REG$_0$

$n_1\mathbf{H} = W_p(n_1, \mathbf{I}_p, n_1\boldsymbol{\Phi}), \mathbf{E} = \mathbf{I}_p$

$n_2 \to \infty$ $\qquad$ $n_2 \to \infty$ $\qquad$ $n_2 \to \infty$

SigD

$n_1\mathbf{H} = W_p(n_1, \mathbf{I}_p + \boldsymbol{\Phi}), n_2\mathbf{E} = W_p(n_2, \mathbf{I}_p)$

CCA

$\mathbf{H} = \mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}, n_2\mathbf{E} = \mathbf{S}_{xx}$

REG

$n_1\mathbf{H} = W_p(n_1, \mathbf{I}_p, n_1\boldsymbol{\Phi}), n_2\mathbf{E} = W_p(n_2, \mathbf{I}_p)$

then the likelihood of the alternative versus the null is given by

$$\mathcal{L}(\theta, \Lambda) = \frac{p(\Lambda; \theta)}{p(\Lambda; 0)} = \alpha(\theta)_p\mathcal{D}_q(\mathbf{a}, \mathbf{b}; \boldsymbol{\Phi}, \Lambda) \tag{18.6}$$

where $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\theta)$ is a p-dimensional matrix diag $\{\boldsymbol{\Phi}_{11}, 0, \cdots, 0\}$. Consider the hypergeometric functions of 2 matrix arguments $\boldsymbol{\Phi}, \Lambda$ are defined as

$$_p\mathcal{F}_q(\mathbf{a}, \mathbf{b}; \boldsymbol{\Phi}, \Lambda) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\kappa - k} \frac{(a_1)_\kappa \cdots (a_p)_\kappa}{(b_1)_\kappa \cdots (b_1)_\kappa} \frac{C_\kappa(\boldsymbol{\Phi})C_\kappa(\Lambda)}{C_\kappa(\mathbf{I}_p)}$$

where $\mathbf{a} = (a_1, \cdots, a_p)$ and $\mathbf{b} = (b_1, \cdots, b_q)$ are parameters, κ are partitions of the integer k, $(a_j)_\kappa$ and $(b_i)_\kappa$ are the generalized Pochhammer symbols, C_κ are the zonal polynomials. For each of the 6 classes of problems, we have the parameters as in Table where $n = n_1 + n_2$.

Table 18.2: Parameters of the Likelihood Ratios in Eq.18.6

Classes	$_p\mathcal{F}_q$	$\alpha(\theta)$	a	b	$\boldsymbol{\Phi}_{11}$
SMD	$_0\mathcal{F}_0$	$\exp(-p\theta^2/4)$	-	-	$\theta p/2$
PCA	$_0\mathcal{F}_0$	$(1+\theta)^{-n_1/2}$	-	-	$\theta n_1/(2(1+\theta))$
SigD	$_1\mathcal{F}_0$	$(1+\theta)^{-n_1/2}$	-	-	$\theta n_1/(n_2(1+\theta))$
REG$_0$	$_0\mathcal{F}_1$	$\exp(-n_1\theta/2)$	-	$n_1/2$	$\theta n_1^2/4$
REG	$_1\mathcal{F}_1$	$\exp(-n_1\theta/2)$	$n/2$	$n_1/2$	$\theta n_1^2/(2n_2)$
CCA	$_2\mathcal{F}_1$	$(1+n_1\theta/n)^{-n/2}$	$(n/2, n/2)$	$n_1/2$	$\theta n_1^2/(n_2^2 + n_2 n_1(1+\theta))$

Some links in Fig.18.2 can also be established via asymptotic relations between hypergeometric functions.

Asymptotic behavior of the likelihood ratios consider that as n_1, n_2, p go to infinity so that

$$c_1 \equiv \frac{p}{n_1} \to \gamma_1 \in (0, 1) \qquad\qquad c_2 \equiv \frac{p}{n_2} \to \gamma_2 \in (0, 1] \tag{18.7}$$

which can be denoted as $\mathbf{n}, p \to_\gamma \infty$ where $\mathbf{n} = \{n_1, n_2\}$ and $\gamma = \{\gamma_1, \gamma_2\}$.

- **Under the null** (the true value of the spike $\theta_0 = 0$), $\lambda_1, \cdots, \lambda_p$ are the eigenvalues of
  - $GOE/\sqrt{p}$, for **SMD**
  - $W_p(n_1, \mathbf{I}_p)/n_1$, for **PCA** and **REG$_0$**
  - a $p-$dimensional multivariate beta matrix with parameters $n_1/2$ and $n_2/2$ (here scaled by a factor of n_2/n_1), for **SigD**, **REG**, **CCA**

  and the empirical distribution of $\lambda_1, \cdots, \lambda_p$ follows

$$\hat{F} = \frac{1}{p}\sum_{j=1}^{p} I\left\{\lambda_j \leq \lambda\right\} \xrightarrow{a.s.} F_\gamma = \begin{cases} F^{SC} & \text{semi-circle distribution, for SMD} \\ F^{MP} & \text{Marchenko-Pastur distribution, for PCA, REG}_0 \\ F^{W} & \text{Wachter distribution, for SigD, REG, CCA} \end{cases}$$

  A summary of the 3 classes of distributions is shown in Table 18.3. And the cumulative distribution functions $F_\gamma^{\lim}(\lambda)$ are linked in the sense that

$$F_\gamma^{W}(\lambda) \to F_{\gamma_1}^{MP}(\lambda) \qquad\qquad \gamma_2 \to 0$$
$$F_{\gamma_1}^{MP}(\sqrt{\gamma_1}\lambda + 1) \to F^{SC}(\lambda) \qquad\qquad \gamma_1 \to 0$$

  If φ is a *well-behaved* function, the centered **linear spectral statistic**

$$\sum_{j=1}^{p} \varphi(\lambda_j) - p \int \varphi(\lambda)(d)F_{\mathbf{c}}^{\lim}(\lambda) \tag{18.8}$$

  converges in distribtuion to a Gaussian random variable in each of the semicircle, Marchenko-Pastur and Wachter cases.[3]

Table 18.3: Semi-circle, Marchenko-Pastur, scaled Wachter distributions

Case	$F_\gamma^{\lim}$	Density, $\lambda \in [\beta_-, \beta_+]$	$\beta_\pm$	Threshold $\overline{\theta}$
SMD	SC	$\frac{R(\lambda)}{2\pi}$	± 2	1
PCA, REG$_0$	MP	$\frac{R(\lambda)}{2\pi\gamma_1\lambda}$	$(1 \pm \sqrt{\gamma_1})^2$	$\sqrt{\gamma_1}$
SigD, REG, CCA	W	$\frac{(\gamma_1+\gamma_2)R(\lambda)}{2\pi\gamma_1\lambda(\gamma_1-\gamma_2\lambda)}$	$\gamma_1\left(\frac{\rho\pm1}{\rho\pm\gamma_2}\right)^2$	$\frac{\rho+\gamma_2}{1-\gamma_2}$
where $R(\lambda) = \sqrt{(\beta_+ - \lambda)(\lambda - \beta_-)}$, $\rho = \sqrt{\gamma_1 + \gamma_2 - \gamma_1\gamma_2}$				

- **Under the alternative**
  - when $\theta \leq \overline{\theta}$ (in Tab.18.3), the top eigenvalue $\lambda_1 \to \beta_+$, the upper boundary of support of F_γ almost surely
  - when $\theta > \overline{\theta}$, λ_1 **separates** from *he bulk* of the other eigenvalues and a.s. converges to a point strictly above β_+

Hence,

- under the *super-critical* cases where $\theta > \overline{\theta}$, the likelihood ratio degenerates, the sequences of measures corresponding to the distributions of Λ under the null and under the *super-critical* alternatives are asymptotically mutually singular as $\mathbf{n}, p \xrightarrow{\gamma} \infty$ for SMD and PCA.
- under the *sub-critical* cases where $\theta < \overline{\theta}$, the likelihood ratio converges to a Guassian process, the sequences of measures corresponding to the distribution of Λ under the null and under the *sub-critical* alternatives are mutually contiguous.

---

[3]The centering constant is defined in terms of $F_{\mathbf{c}} = \{c_1, c_2\}$, that is, the *correct centering* can be computed using the densities in Tab.18.3, where γ_1, γ_2 are replaced by $c_1 \equiv p/n_1, c_2 \equiv p/n_2$ respectively.

Contour integral representation   The asymptotic behavioral of the likelihood ratios (Eq.18.6) depends on that of $_p\mathcal{F}_q(a, b; \mathbf{\Psi}, \mathbf{\Lambda})$, of which the asymptotics are well established when the dimension of the matrix arguments remain **fixed**. Now consider the case where $\mathbf{\Phi}, \mathbf{\Lambda}$ diverge to infinity. In single-spiked models, $\mathbf{\Phi}$ has rank one, then $_p\mathcal{F}_q(a, b; \mathbf{\Psi}, \mathbf{\Lambda})$ can be represented in the form of a **contour integral** of a hypergeometric function of a single scalar argument:

Lemma 18.2.1: Contour Integral Representations for Likelihood Ratios

Assume $p \leq \min\{n_1, n_2\}$, let $\mathcal{K}$ be a contour in the complex plan $\mathbb{C}$ taht starts at $-\infty$, encircles 0 and $\lambda_1, \cdots, \lambda_p$ counter-clockwise, returning to $-\infty$, then

$$\mathcal{L}(\theta; \mathbf{\Lambda}) = \frac{\Gamma(s+1)\alpha(\theta)q_s}{\mathbf{\Phi}_{11}^s 2\pi i} \int_{\mathcal{K}} {_p\mathcal{F}_q}(a - s, b - s; \mathbf{\Psi}_{11}z) \prod_{j=1}^{p}(z - \lambda_j)^{-1/2}dz \qquad (18.9)$$

where $s = p/2 - 1$, the values of $\alpha(\theta), \mathbf{\Phi}_{11}, a, b, \mathrm{p}, \mathrm{q}$ for difference cases are given in Tab.18.2, and $a - s, b - s$ are vectors with elements $a_j - s, b_j - s$ respectively,

$$q_s = \prod_{j=1}^{\mathrm{P}} \frac{\Gamma(a_j - s)}{\Gamma(a_j)} \prod_{i=1}^{\mathrm{q}} \frac{\Gamma(b_i)}{\Gamma(b_i - s)}$$

---

next, we want to approximate 18.9 in a Laplace form, that is, to make the right-hand side looks like

$$\mathcal{L}(\theta; \mathbf{\Lambda}) = \sqrt{\pi p}\frac{1}{2\pi i} \int_{\mathcal{K}} \exp\left\{-\left(\frac{p}{2}\right) f(z; \theta)\right\} g(z; \theta)dz \qquad (18.10)$$

The goal of the transformation is to have the function $f(\cdot), g(\cdot)$ will have the forms of a sum and a product:

$$f(z) = f_c + f_e(z) + f_h(z)$$
$$g(z) = g_c \times g_e(z) \times g_h(z)$$

where f_c and g_c do not depend on z.

The transformation from Eq.18.9 to Eq.18.10 is done in 3 steps:

- first

$$\frac{\Gamma(s+1)\alpha(\theta)q_s}{\mathbf{\Phi}_{11}^s 2\pi i} = \exp\left\{-\frac{p}{2}f_c\right\}g_c \qquad (18.11)$$

  where g_c remains bounded as $\mathbf{n}, p \xrightarrow{\gamma} \infty$, and the values of f_c and g_c are given as Tab.18.4[4]
- second, consider

$$\prod_{j=1}^{p}(z - \lambda_j)^{-1/2} = \exp\left\{-\frac{p}{2}f_e(z)\right\}g_e(z) \qquad (18.12)$$

  where

$$f_e(z) \int \ln(z - \lambda)dF_{\mathbf{c}}(\lambda) \qquad (18.13)$$

$$g_e(z) \exp\left\{-\frac{p}{2}\int \ln(z - \lambda)d\left(\hat{F}(\lambda) - F_{\mathbf{c}}(\lambda)\right)\right\} \xrightarrow{\mathbf{n},p \xrightarrow{\gamma} \infty} \text{Gaussian random variable} \qquad (18.14)$$

---

[4]In Tab.18.4 the terms $o(1)$ do **not** depend on θ. $l(\theta) = 1 + \frac{(1+\theta)c_2}{c_1}$, $r^2 = c_1 + c_2 - c_1c_2$, $f_{10} = -1 - \frac{r^2}{c_1c_2}\log\frac{r^2}{c_1+c_2} + \log\frac{c_1+c_2}{c_1}$, $\check{g}_{10} = c_1^{-1}r(c_1 + c_2)^{1/2}$, $f_{21} = -1 - \frac{\theta}{c_1} - \frac{r^2}{c_1c_2}\log\frac{r^2}{c_1 l(\theta)}$.

Table 18.4: Values of f_c and $\check{g}_c = \frac{g_c}{1+o(1)}$

Case	f_c	$\check{g}_c = \frac{g_c}{1+o(1)}$
SMD	$1 + \theta^2/2 + \log\theta$	θ
PCA	$1 + \frac{1-c_1}{c_1}\log(1+\theta) + \log\frac{\theta}{c_1}$	$\theta(1+\theta)^{-1}c_1^{-1}$
SigD	$f_c^{PCA} + f_{10}$	$\check{g}_c^{PCA}\check{g}_{10}$
REG$_0$	$1 + \frac{\theta+c_1}{c_1} + \log\frac{\theta}{c_1} + \frac{1-c_1}{c_1}\log(1-c_1)$	$\theta c_1^{-1}(1-c_1)^{-1/2}$
REG	$f_c^{REG_0} + f_{10}$	$\check{g}_c^{REG_0}\check{g}_{10}$
CCA	$f_c^{REG} + f_{21}$	$\frac{\check{g}_c^{REG}\check{g}_{10}}{l(\theta)}$

for $f_e(z)$ and $g_e(z)$ to be well-defined, $z \notin \text{supp}(F_{\mathbf{c}})$ and $z \notin \text{supp}(\hat{F})$.

- third, consider

$$_p\mathcal{F}_q(a-s, b-s, \Phi_{11}z) = \exp\left\{-\frac{p}{2}f_h(z)\right\} g_h(z) \tag{18.15}$$

where

$$f_h(z) = \begin{cases} -z\theta & \text{SMD} \\ -z\frac{\theta}{c_1(1+\theta)} & \text{PCA} \\ \ln\left[1 - \frac{c_2 z\theta}{c_1(1+\theta)}\right]\frac{r^2}{c_1 c_2} & \text{SigD} \end{cases} \tag{18.16}$$

$$g_h(z) = \begin{cases} 1 & \text{SMD, PCA} \\ \left[1 - \frac{c_2 z\theta}{c_1(1+\theta)}\right]^{-1} & \text{SigD} \end{cases} \tag{18.17}$$

- when q = 0, $_p\mathcal{F}_q$ can be expressed in terms of elementary functions: $_0\mathcal{F}_0(z) = e^z$, $_1\mathcal{F}_0(a;z) = (1-z)^{-a}$
- when q = 1, $_p\mathcal{F}_q$ can **NOT** be represented exactly in terms of elementary functions. Hence, consider the asymptotic approximations

$$_p\mathcal{F}_q = \begin{cases} _0\mathcal{F}_1(m+1; m^2\eta_0) \equiv F_0 & \text{REG}_0 \\ _1\mathcal{F}_1(m\kappa+1; m+1; m\eta_1) \equiv F_1 & \text{REG} \\ _2\mathcal{F}_1(m\kappa+1; m\kappa+1; m+1; \eta_2) \equiv F_2 & \text{CCA} \end{cases} \tag{18.18}$$

where $m = \frac{n_1-p}{2}$, $\kappa = \frac{n-p}{n_1-p}$, and

$$\eta_j \begin{cases} \frac{z\theta}{(1-c_1)^2} & j = 0 \\ \frac{z\theta c_2}{c_1(1-c_1)} & j = 1 \\ \frac{z\theta c_2^2}{c_1^2 l(\theta)} & j = 2, l(\theta) = 1 + \frac{(1+\theta)c_2}{c_1} \end{cases}$$

Johnstone and Onatski (2020) outlined the asymptotics of F_j, $j = 0, 1, 2$ as

* j=0: Let $\varphi_0(t) = \ln t - t - \eta_0/t + 1$ and $t_0 = (1 + \sqrt{1+t\eta_0})/2$, and $\forall\delta > 0$, let $\Omega_{0\delta}$ be the set of $\eta_0 \in \mathbb{C}$ s.t. $\left|\arg\eta_0\right| \leq \pi - \delta$, then as $m \to \infty$, we have

$$F_0 = (1 + 4\eta_0)^{-1/4}\exp\left\{-m\varphi_0(t_0)\right\}(1 + o(1))$$

* j=1,2: consider the contour integral representations

$$F_j = \frac{C_m}{2\pi i}\int_0^{(1+)}\exp\left\{-m\varphi_j(t)\right\}\psi_j(t)dt$$

where

$$C_m = \frac{\Gamma(m+1)\Gamma(m(\kappa-1)+1)}{\Gamma(m\kappa+1)}$$

and

$$\varphi_j(t) = \begin{cases} -\eta_j t - \kappa \ln t + (\kappa-1)\ln(t-1), & j=1 \\ -\kappa \ln(t/(1-\eta_j t)) + (\kappa-1)\ln(t-1), & j=2 \end{cases} \qquad \psi_j(t) = \begin{cases} (t-1)^{-1}, & j=1 \\ (t-1)^{-1}(1-\eta_j t)^{-1}, & j=2 \end{cases}$$

the relevant saddle points are given as

$$t_j = \begin{cases} \frac{1}{2\eta_j}\left\{\eta_j - 1 + \sqrt{(\eta_j-1)^2 + 4\kappa\eta_j}\right\}, & j=1 \\ \frac{1}{2\eta_j(\kappa-1)}\left\{-1 + \sqrt{1 + 4\kappa(\kappa-1)\eta_j}\right\}, & j=2 \end{cases}$$

then as $m \to \infty$, for $j = 1, 2$

$$F_j = C_m \psi_j(t_j) e^{-iw_j/2} \left|2\pi m \varphi_j''(t_j)\right|^{-1/2} \exp\left\{-m\varphi_j(t_j)\right\}(1+o(1))$$

now, we can set the components of the *Laplace form* of $_p\mathcal{F}_q$ for q = 1 as

$$f_h(z) \begin{cases} \frac{1-c_1}{c_1}\varphi_0(t_0) & \text{REG}_0 \\ \frac{1-c_1}{c_1}\left(\varphi_j(t_j) + \kappa \ln \kappa - (\kappa-1)\ln(\kappa-1)\right) & \text{REG,CCA} \end{cases} \tag{18.19}$$

$$g_h(z) \begin{cases} (1+4\eta_0)^{-1/4}(1+o(1)) & \text{REG}_0 \\ \sqrt{\frac{c_1}{r^2}} e^{-iw_j/2} \left|\varphi_j''(t_j)\right|^{-1/2} \varphi_j(t_j)(1+o(1)) & \text{REG,CCA} \end{cases} \tag{18.20}$$

Together, we have

Lemma 18.2.2: Saddle Points

The saddle points $z_0(\theta, \mathbf{c})$ of $f(z)$ satisfies

$$z_0(\theta, \mathbf{c}) = \begin{cases} \theta + 1/\theta & \text{SMD} \\ (1+\theta)(\theta+c_1)/\theta & \text{PCA, REG}_0 \\ (1+\theta)(\theta+c_1)/(\theta l(\theta)) & \text{SigD, REG, CCA} \end{cases} \tag{18.21}$$

for $\theta \in (0, \overline{\theta}_\mathbf{c})$, $z_0 > b_+$ where $\overline{\theta}_\mathbf{c}$ is the threshold corresponding to $F_\mathbf{c}$.

As $c_2 \to 0$, while c_1 stays constant, the value of z_0 for **SigD, REG, CCA** converges to **PCA** and **REG**$_0$, which converges to **SMD** as $c_1 \to 0$. Precisely, solving equation

$$\sqrt{c_1}z_0 + 1 = (1 + \sqrt{c_1}\theta)(\sqrt{c_1}\theta + c_1)/(\sqrt{c_1}\theta)$$

for z_0 and taking limit as $c_1 \to 0$ yields $z_0 = \theta + 1/\theta$.

Then, we have the deformed contour as $\mathcal{K} = \mathcal{K}_+ \cup \mathcal{K}_-$, with $\mathcal{K}_-$ is the complex conjugate of $\mathcal{K}_+$, and $\mathcal{K}_+ = \mathcal{K}_1 \cup \mathcal{K}_2$, where

- **SMD, PCA, SigD** (as in Fig.18.3)

$$\mathcal{K}_1 = \{z_0 + it : 0 \le t \le 2z_0\} \qquad\qquad \mathcal{K}_2 = \{x + i2z_0 : -\infty < x \le z_0\}$$

- **REG$_0$, CCA** (as in Fig.18.4)

$$\mathcal{K}_1 = \left\{z_1 + |z_0 - z_1| \exp\{i\gamma\} : \gamma \in [0, \pi/2]\right\} \qquad \mathcal{K}_2 = \left\{z_1 - x + |z_0 - z_1| \exp\{i\pi/2\} : x \geq 0\right\}$$

where

$$z_1 = \begin{cases} -(1 - c_1)^2/(4\theta) & \text{for REG}_0 \\ -c_1(1 - c_1)^2 l(\theta)/(4\theta r^2) & \text{for CCA} \end{cases}$$

- **REG**: it can be described as an image of a contour $\mathcal{C}$ in τ−plane where $\tau = \eta_1 t_1$ with $\eta_1 = z\theta c_2/[c_1(1-c_1)]$, see Johnstone and Onatski (2020, P.20-21) for details.

Figure 18.3: $\mathcal{K}$ for SMD, PCA, SigD

Figure 18.4: $\mathcal{K}$ for REG$_0$, CCA

Together, we have that for all 6 cases (SMD, PCA, SigD, REG$_0$, REG and CCA), we have

> **Lemma 18.2.3: $\mathcal{K}_1$ are of steep descent**
>
> As z moves along the corersponding $\mathcal{K}_1$ away from z_0, $-\text{Re} f(z)$ is **strictly decreasing**.

Laplace approximation   Next, we can derive Laplace approximations to the integral (18.9)

$$\mathcal{L}(\theta; \Lambda) = \frac{\Gamma(s + 1)\alpha(\theta)q_s}{\Phi_{11}^s 2\pi i} \int_{\mathcal{K}} {}_p\mathcal{F}_q(a - s, b - s; \Psi_{11}z) \prod_{j=1}^{p}(z - \lambda_j)^{-1/2} dz$$

first, consider a general integral

$$I_{p,w} = \int_{\mathcal{K}_{p,\omega}} e^{-p\phi_{p,\omega}(z)} \chi_{p,\omega}(z) dz$$

where

- p is large, $\omega \in \Omega \subset \mathbb{R}^k$ is a k−dimensional parameter
- $\mathcal{K}_{p,\omega}$ is a path in $\mathbb{C}$ that starts at $a_{p,\omega}$ and ends at $b_{p,\omega}$
- $\phi_{p,\omega}, \chi_{p,\omega}(z)$ are single-valued holomorphic functions of z, in the case of $\chi_{p,\omega}$ with probability increasing to 1 (subscripts $_{p,\omega}$ are omitted hereafter)

Assuming that $\exists C_1, \cdots, C_4 > 0$ that do not depend on p, ω, s.t. $\forall \omega \in \Omega$ for sufficiently large p

A0 The length of the path $\mathcal{K}$ is bounded, uniformly over $\omega \in \Omega$ and all sufficiently large p,

$$\sup_{z \in (z_0, b)_{\mathcal{K}}} |z - z_0| > C_1 \qquad\qquad \sup_{z \in}$$

A1 $\phi(z)$ and $\chi(z)$ are holomorphic in the ball $|z - z_0| \leq C_1$

A2 ϕ_2 satisfies that $C_2 \leq |\phi_2| \leq C_3$

A3 The third derivative of $\phi(z)$ satisfies inequality

$$\sup_{|z - z_0| \leq C_1} \left| d^3 \phi(z)/dz^3 \right| \leq C_4$$

A4 $\forall 0 < \epsilon < C_1$ (not depending on p, ω), and $\forall z_i \in \mathcal{K}$ s.t. $|z_1 - z_0| = \epsilon$, $\exists C_5, C_6 > 0$ s.t.

$$\mathrm{Re}\left(\phi(z_1) - \phi_0 \right) C_5 \qquad\qquad \left| \mathrm{Im}\left(\phi(z_i) - \phi_0 \right) \right| < C_6$$

A5 For $\Theta \subset \mathbb{C}$ that consists of all points whose Euclidean distance from $\mathcal{K}$ is no larger than C_1

$$\sup_{z \in \Theta} |\chi(z)| = O_{\mathrm{p}}(1)$$

as $p \to \infty$, where $O_{\mathrm{p}}(1)$ is uniform in $\omega \in \Omega$

Under Assumption A0-A5, we have

> **Lemma 18.2.4: A General Integral and the Laplace Approximation**
>
> For any positive integer k as $p \to \infty$, we have
>
> $$I_{p,\omega} = 2 e^{-p\phi_0} \left[\sum_{s=0}^{k-1} \Gamma\left(s + \frac{1}{2} \right) \frac{a_{2s}}{p^{s+1/2}} + \frac{O_{\mathrm{p}}(1)}{p^{k+1/2}} \right]$$
>
> where
> - O_{p} is uniform in $\omega \in \Omega$
> - the coefficients a_{2s} can be expressed through ϕ_s and χ_s defined above:
>   - $a_0 = \phi / [2\phi_2^{1/2}]$ where $\phi_2^{1/2} = \exp\left\{ (\log |\phi_2| + \mathrm{i} \arg \phi_2)/2 \right\}$ with the branch of $\arg \phi_2$ chosen s.t. $\left| \arg \phi_2 + 2/\beta \right| \leq \pi/2$

We then use the lemma above to obtain the Laplace approximation to

$$\mathcal{L}_1(\theta, \Lambda) = \sqrt{\pi p}\, \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{K}_1 \cup \overline{\mathcal{K}}_1} e^{-(p/2)f(z)} g(z) dz$$

here, we must know the values of $f(z_0)$ and $d^2 f(z_0)/dz^2$:

- for all 6 cases, $f(z_0) = 0$

- for all 6 cases, $d^2 f(z_0)/dz^2 < 0$, its explicit form $D_2 \equiv \theta^2 \left(-d^2 f(z_0)/dz^2 \right)^{-1}$ is given in Tab.18.5

then, we can have the Laplace appoximation as

Table 18.5: Values of $D_2 \equiv \theta^2 \left(-\mathrm{d}^2 f(z_0)/\mathrm{d}z^2\right)^{-1}$

Case	Value of D_2	Case	Value of D_2
SMD	$1 - \theta^2$	PCA	$c_1\left(c_1 - \theta^2\right)(1 + \theta)^2$
REG_0	$c_1(1 + c_1 + 2\theta)\left(c_1 - \theta^2\right)$	REG	$c_1 h(c_1 + \theta + (1 + \theta)l)/l^4$
SigD	$r^2 h(1 + \theta)^2/l^4$	CCA	$c_1^2\left(2(c_1 + \theta) + l(1 - c_1)\right)/(l^3(c_1 + c_2))$

Theorem 18.2.5: Laplace Approximation

Suppose that the null hypothesis holds, i.e., $\theta_0 = 0$. Let $\overline{\theta}$ be the threshold corresponding to F_γ as given in Tab.18.3, and let ϵ be an arbitrarily small fixed positive number, then $\forall \theta \in (0, \overline{\theta} - \epsilon]$, as $\mathbf{n}, p \xrightarrow{\gamma} \infty$, we have

$$\mathcal{L}(\theta; \Lambda) = \frac{g(z)}{\sqrt{-\mathrm{d}^2 f(z_0)/\mathrm{d}z^2}} + O_{\mathrm{p}}(p^{-1}) \tag{18.22}$$

where $O_{\mathrm{p}}(p^{-1})$ is uniform in $\theta \in (0, \overline{\theta} - \epsilon]$ and the principal branch of the square root is taken.

Asymptotics of LR from Theorem 18.2.5, let

$$\Delta_p(\theta) = p \int \ln(z_0(\theta) - \lambda)\mathrm{d}\left(\hat{F}(\lambda) - F_{\mathbf{c}}(\lambda)\right)$$

where $\Delta_p(\theta)$ is defined as zero in the event of asymptotically negligible probability that $z_0 \leq \lambda_1$.

Theorem 18.2.6: Asymptotics of LR

Suppose that the null hypothesis holds, $\theta_0 = 0$. Let $\overline{\theta}$ be the threshold corresponding to F_γ as in Tab.18.3, let ϵ be an arbitrarily small fixed positive number, then $\forall \theta \in (0, \overline{\theta} - \epsilon]$, as $\mathbf{n}, p \xrightarrow{\gamma} \infty$, we have

$$\mathcal{L}(\theta, \Lambda) = \exp\left\{-\frac{1}{2}\Delta_p(\theta) + \frac{1}{2}\ln\left(1 - [\delta_p(\theta)]^2\right)\right\}(1 + o_{\mathrm{p}}(1))$$

where

$$\delta_p(\theta) = \begin{cases} \theta, & \text{SMD} \\ \theta/\sqrt{c_1}, & \text{PCA, } REG_0 \\ \theta r/(c_1 l(\theta)), & \text{SigD, REG, CCA} \end{cases}$$

and $r^2 = c_1 + c_2 - c_1 c_2$ and $o_{\mathrm{p}}(1)$ is uniform in $\theta \in (0, \overline{\theta} - \epsilon]$.

Here, statistic $\Delta_p(\theta)$ is a linear spectral statistic, weakly converging to a Gaussian process indexed by $\theta \in (0, \overline{\theta} - \epsilon]$. Next, we derive the asymptotic expectation and covariances of $\mathcal{L}(\theta, \Lambda)$:

Theorem 18.2.7: Asymptotic Moments of LR

Suppose that the null hypothesis holds, $\theta_0 = 0$. Let $\overline{\theta}$ be the threshold corresponding to F_γ as in Tab.18.3, let ϵ be an arbitrarily small fixed positive number and $C[0, \overline{\theta} - \epsilon]$ be the space of continuous functions on $[0, \overline{\theta} - \epsilon]$ equipped with the supremum norm. Then $\ln \mathcal{L}(\theta; \Lambda)$ viewed as random

elements of $C[0, \overline{\theta} - \epsilon]$ converge weakly to $\mathcal{L}(\theta)$ with Gaussian finite dimensional distributions such that

$$\mathbb{E}\mathcal{L}(\theta) = \frac{1}{4} \ln(1 - \delta^2(\theta))$$

$$\text{Cov}\left(\mathcal{L}(\theta_1), \mathcal{L}(\theta_2)\right) = -\frac{1}{2} \ln\left(1 - \delta(\theta_1)\delta(\theta_2)\right)$$

with

$$\delta(\theta) = \begin{cases} \theta, & \text{SMD} \\ \theta/\sqrt{\gamma_1}, & \text{PCA,REG}_0 \\ \theta\rho/(\gamma_1 + \gamma_2 + \theta\gamma_2), & \text{SigD,REG,CCA} \end{cases}$$

here, ρ, γ_1, γ_2 are the limits of r, c_1, c_2 as $\mathbf{n}, p \xrightarrow{\gamma} \infty$

Let $\{\mathbb{P}_{p,\theta}\}$ and $\{\mathbb{P}_{p,0}\}$ be the sequences of measures corresponding to the joint distributions of $\lambda_1, \cdots, \lambda_p$ when $\theta_0 = \theta$ and when $\theta_0 = 0$ respectively. Then, under Thm.18.2.7, the mutual contiguity of $\{\mathbb{P}_{p,\theta}\}$ and $\{\mathbb{P}_{p,0}\}$ as $\mathbf{n}, p \xrightarrow{\gamma} \infty$ for each $\theta < \overline{\theta}$. Hence, statistically, the phase transition thresholds are essentially the upper boundaries of the contiguity regions for spiked models.

Next, derive the asymptotic power envelopes for tests of the null hypothesis $\theta_0 = 0$ against the alternative $\theta_0 > 0$.

Theorem 18.2.8: Asymptotic Power Envelope

Let $\overline{\theta}$ be the threshold corresponding to F_γ as given in Tab.18.3. $\forall \theta \in [0, \overline{\theta})$, the value of the asymptotic power envelope for the tests of the null $\theta_0 = 0$ against the alternative $\theta_0 > 0$ which are based on $\lambda_1, \cdots, \lambda_p$ and have asymptotic size α is given by

$$\text{PE}(\theta) = 1 - \Phi\left[\Phi^{-1}(1 - \alpha) - \sigma(\theta)\right] \qquad\qquad \sigma(\theta) = \sqrt{-\frac{1}{2}\ln\left(1 - \delta^2(\theta)\right)}$$

where Φ is the standard normal cumulative distribution function. For $\theta \geq \overline{\theta}$ the value of the asymptotic power envelope equals one.

Remarks

- Thm. 18.2.7 establishes the weak convergence of the log likelihood ratio viewed as a random element of the space of continuous functions. This is **stronger** than simply the cnovergence of the finite dimensional distributions of the log likelihood process. The theorem can be used to find the asymptotic distribution of the supremum of the likelihood ratio, and thus, to find the asymptotic critical values of the likelihood ratio test. It can also be used to construct asymptotic confidence intervals for a sub-critical spike as well as to describe the asymptotic properties of its maximum likelihood estimator.

  Consider the log likelihood ratio

  $$\ln \mathcal{L}(\theta; \Lambda) - \ln \mathcal{L}(\theta_0; \Lambda)$$

  Thm.18.2.7 gives that this ratio converges to $X(\theta) \equiv \mathcal{L}(\theta) - \mathcal{L}(\theta_0)$. Under the null hypothesis that the

true value of the spike equals θ_0, $X(\theta)$ is a Gaussian process where

$$\mathbb{E}X(\theta) = \frac{1}{4}\ln\frac{\left(1-\delta^2(\theta)\right)\left(1-\delta^2(\theta_0)\right)}{(1-\delta(\theta)\delta(\theta_0))^2}$$

$$\text{Cov}\left(X(\theta_1),X(\theta_2)\right) = -\frac{1}{2}\ln\frac{(1-\delta(\theta_1)\delta(\theta_2))\left(1-\delta^2(\theta)\right)}{(1-\delta(\theta_1)\delta(\theta_0))\left(1-\delta(\theta_2)\delta(\theta_0)\right)}$$

An approximation can be obtained via simulation.

- The maximum likelihood estimator for the spike $\hat\theta_{ML}$. The limiting process $X(\theta)$ achieves maximum at a unique point with probability one. Therefore by the argmax continuous mapping theorem, $\hat\theta_{ML}$ converges in distribution to the argmax of $X(\theta)$. The distribution of such an argmax can also be approximated via simulations. But **the quality of the estimator $\hat\theta_{ML}$ cannot be good**.

- **Multiple spikes**: the asymptotic analysis can be extended to a **fixed** number of sub-critical spikes. In the case of slowly increasing number of spikes, relatively little is known about the phase transition phenomenon. It is not clear whether the experiments of observing sample covariance eigenvalues corresponding to the null case and an alternative with a growing number of sub-critical eigenvalues remain mutually contiguous.

  Intuitively, the asymptotic power of the likelihood ratio test of the null hypothesis of no spikes against the alternative of one spike should not decrease if the rank-one assumption on the alternative is wrong and there are additional spikes.

- **Relaxing** the assumption $p \leq n_1$:

  - **SMD**: the assumption is irrelevant
  - **PCA**: already covered by previous work
  - **REG$_0$**: the assumption can be relaxed using the symmetry of the problem
  - **CCA**: the sample canonical correlations are not well defined for $p > n_1$
  - **SigD** and **REG**: unclear under the current results

18.3   Fan et al. (2022)

Building on the framework set up by Johnstone and Onatski (2020), Fan et al. (2022) discuss the asymptotics theory of eigenvectors for random matrices with diverging spikes. Consider an $n \times n$ random matrix

$$\mathbf{X} = \mathbf{H} + \mathbf{W} \tag{18.23}$$

where

- $\mathbf{H} = \mathbf{VDV}'$ is a deterministic latent mean matrix of low rank structure, where $K \ll n$

  - $\mathbf{V} = (\mathbf{v_1}, \cdots, \mathbf{v_K})$ is $n \times K$ orthonormal matrix of population eigenvectors
  - $\mathbf{D} = \text{diag}(d_1, \cdots, d_K)$ is a diagonal matrix of population eigenvalues, with $|d_1| \geq \cdots \geq |d_K|$
- $\mathbf{W} = \left(w_{ij}\right)_{1 \leq i,j \leq n}$ is a symmetric random matrix of independent noise on and above the diagonal with $\mathbb{E}w_{ij} = 0$, $\mathbb{E}w_{ij}^2 = \sigma_{ij}^2$, and $\max_{1 \leq i,j \leq n}\left|w_{ij}\right| \leq 1$

The goal is to, with

- the **latent eigenstructure** $(\mathbf{D}, \mathbf{V})$
- the **data matrix** $\mathbf{X}$

conduct a spectral decomposition of the observed data matrix. Denote its eigenvalues by $\lambda_1, \cdots, \lambda_n$, s.t. $|\lambda_1| \geq \cdots \geq |\lambda_n|$, and $\hat{\mathbf{v}}_1, \cdots, \hat{\mathbf{v}}_n$ the corresponding eigenvectors, what's the asymptotic behavior of the **spiked** empirical eigenvalues and eigenvectors?

Technical conditions   Under the conditions

- **Cond. 1**: Assume that

$$\alpha_n = \left\| \mathbb{E}\left(\mathbf{W} - \mathbb{E}\mathbf{W}\right)^2 \right\|^{1/2} \xrightarrow{n \to \infty} \infty$$

- **Cond. 2**: $\exists c_0 < 1$, s.t.

$$\min\left\{ \frac{|d_i|}{|d_j|} : 1 \le i \le j \le K+1, d_i \ne d_j \right\} \ge 1 + c_0$$

  in addition, either of the following 2 conditions holds

  - $|d_K| / (n^\epsilon \alpha_n) \to \infty$ with some small $\epsilon > 0$
  - $\max_{i,j} \operatorname{var}(w_{ij}) \le \left(c_1^2 \alpha_n^2\right)/n$ and $|d_K| > c\alpha_n \log n$ with $c_1 > 1, c > 4c_1(1 + c_0/2)$

- **Cond. 3**: It holds that $|d_1| = O(|d_K|)$, $|d_K| \sigma_{\min}/\alpha_n \to \infty$, $\|\mathbf{v}_k\|_\infty^2 / \sigma_{\min} \to 0$, $\alpha_n^4 \|\mathbf{v}_k\|_\infty^4 / (\sqrt{n} \sigma_{\min}^2) \to 0$, and $\sigma_{\min}^2 n \to \infty$, where $\sigma_{\min} = \left\{ \min_{1 \le i,j, \le n, i \ne j} \mathbb{E}w_{ij}^2 \right\}^{1/2}$

Asymptotic distributions of spiked eigenvalues   For each $1 \le k \le K$, denote the solution to the equation

$$f_k(z) = 1 + d_k \left\{ \mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, z) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z) \times \left[(\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \right\} = 0 \qquad (18.24)$$

when restricted to the interval $z \in [a_k, b_k]$, where

$$a_k = \begin{cases} d_k/(1 + c_0/2) & d_k > 0 \\ d_k(1 + c_0/2) & d_k < 0 \end{cases} \qquad\qquad b_k \begin{cases} d_k(1 + c_0/2) & d_k > 0 \\ d_k/(1 + c_0/2) & d_k < 0 \end{cases}$$

Equation 18.24 has a **unique** solution in the interval $z \in [a_k, b_k]$ and thus t_k's are well defined. Moreover, for each $1 \le k \le K$, we have

$$t_k/d_k \xrightarrow{n \to \infty} 1$$

and the following theorem characterizes the asymptotic distribution of spiked eigenvalues and eigenvectors

---

Theorem 18.3.1: Asymptotic Distribution of Spiked Eigenvalues and Eigenvectors

Under Condition 1 and 2, for each $1 \le k \le K$, we have

$$\lambda_k - t_k = \mathbf{v}_k' \mathbf{W} \mathbf{v}_k + O_p(\alpha_n d_k^{-1})$$

Moreover, if $\operatorname{var}\left(\mathbf{v}_k' \mathbf{W} \mathbf{v}_k \right) \gg \alpha_n^2 d_k^{-2}$, then

$$\frac{\lambda_k - t_k - \mathbb{E}\left(\mathbf{v}_k' \mathbf{W} \mathbf{v}_k \right)}{\left[\operatorname{var}\left(\mathbf{v}_k' \mathbf{W} \mathbf{v}_k \right) \right]^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

and
- If the unit vector $\mathbf{u}$ satisfies that $|\mathbf{u}'\mathbf{v}_k| \in [0, 1)$, $\alpha_n^2 d_k^2 \operatorname{var}\left[\left(\mathbf{b}_{\mathbf{u},k,t_k}' - \mathbf{u}'\mathbf{v}_k \mathbf{v}_k' \right) \mathbf{W} \mathbf{v}_k \right] \to \infty$ and

$\left(\mathbf{b}'_{\mathbf{u},k,t_k} - \mathbf{u}'\mathbf{v}_k\mathbf{v}'_k, \mathbf{v}_k\right)$ satisfies that $\mathbf{W}^1$-CLT condition, then it holds that

$$\frac{t_k\left(\mathbf{u}'\hat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k}\tilde{\mathcal{P}}^{1/2}_{k,t_k}\right) - \mathbb{E}\left[\left(\mathbf{b}'_{\mathbf{u},k,t_k} - \mathbf{u}'\mathbf{v}_k\mathbf{v}'_k\right)\mathbf{W}\mathbf{v}_k\right]}{\left\{\mathrm{var}\left[\left(\mathbf{b}'_{\mathbf{u},k,t_k} - \mathbf{u}'\mathbf{v}_k\mathbf{v}'_k\right)\mathbf{W}\mathbf{v}_k\right]\right\}^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $A_{\mathbf{u},k,t_k}\tilde{\mathcal{P}}^{1/2}_{k,t_k} = -\mathbf{u}'\mathbf{v}_k + O(\alpha_n^2 d_k^{-2})$

- If $\left(\alpha_n^{-4}d_k^2 + 1\right)\mathrm{var}\left(\mathbf{v}'_k\mathbf{W}^2\mathbf{v}_k\right) \to \infty$ and $\left(\mathbf{b}'_{\mathbf{u},k,t_k} - \mathbf{u}'\mathbf{v}_k\mathbf{v}'_k, \mathbf{v}_k\right)$ satisfies that $\mathbf{W}^2$-CLT condition, then it holds that

$$\frac{2t_k^2\left(\mathbf{v}'_k\hat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k}\tilde{\mathcal{P}}^{1/2}_{k,t_k}\right)}{\left[\mathrm{var}\left(\mathbf{v}'_k\mathbf{W}^2\mathbf{v}_k\right)\right]^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $A_{\mathbf{u},k,t_k}\tilde{\mathcal{P}}^{1/2}_{k,t_k} = -1 + 2^{-1}t_k^{-2}\mathbf{v}'_k\mathbb{E}\mathbf{W}^2\mathbf{v}_k + O(\alpha_n^3 d_k^{-3})$.

and under Conditions 1-3, for each $1 \leq k \leq K$, we have
- **Eigenvalues**

$$\frac{\lambda_k - t_k - \mathbb{E}\left(\mathbf{v}'_k\mathbf{W}\mathbf{v}_k\right)}{\left[\mathrm{var}\left(\mathbf{v}'_k\mathbf{W}\mathbf{v}_k\right)\right]^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)$$

- **Eigenvectors**: if the unit vector $\mathbf{u}$ satisfies that $|\mathbf{u}'\mathbf{v}_k| \in [0, 1-\epsilon]$ for some $\epsilon > 0$, and $\sigma_{\min}^{-1}\left\|\mathbf{v}_k\left(\mathbf{b}'_{\mathbf{u},\mathbf{k},\mathbf{t}_k} - \mathbf{u}'\mathbf{v}_k\mathbf{v}'_k\right)\right\|_\infty \to 0$, then

$$\frac{t_k\left(\mathbf{u}'\hat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k}\tilde{\mathcal{P}}^{1/2}_{k,t_k}\right) - \mathbb{E}\left[\left(\mathbf{b}'_{\mathbf{u},k,t_k} - \mathbf{u}'\mathbf{v}_k\mathbf{v}'_k\right)\mathbf{W}\mathbf{v}_k\right]}{\left\{\mathrm{var}\left[\left(\mathbf{b}'_{\mathbf{u},k,t_k} - \mathbf{u}'\mathbf{v}_k\mathbf{v}'_k\right)\mathbf{W}\mathbf{v}_k\right]\right\}^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)$$

$$\frac{2t_k^2\left(\mathbf{v}'_k\hat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k}\tilde{\mathcal{P}}^{1/2}_{k,t_k}\right)}{\left[\mathrm{var}\left(\mathbf{v}'_k\mathbf{W}^2\mathbf{v}_k\right)\right]^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)$$

Clustering power detection   Now consider

$$\mathbb{E}\mathbf{X} = \boldsymbol{\Theta}\mathbf{B}\boldsymbol{\Theta}' := \mathbf{H}$$

where $\boldsymbol{\Theta}$ is an $n \times K$ community membership matrix, and $\mathbf{B}$ is a $K \times K$ symmetric matrix. Then, based on Thm. 18.3.1, we can apply the $K-$means clustering algorithm to K spiked eigenvectors and recover the memberships of nodes in network models:

$$H_0 : \mathbf{v}_k = \mathbf{1}/\sqrt{n} \qquad\qquad H_1 : \mathbf{v}_k \neq \mathbf{1}/\sqrt{n}$$

using the test statistic $\hat{\mathbf{v}}'_k\mathbf{1}$. If we assume $n = n_1 + n_2, k = 2$ and the 2 communities are $\{1, \cdots, n_1\}$,

$\{n_1 + 1, \cdots, n\}$. For $0 < p < q < 1$,

$$\mathbf{B} = \begin{bmatrix} q & p \\ p & p \end{bmatrix}$$

then

$$\mathbf{H} = \left[\begin{array}{ccc:ccc} q & \cdots & q & p & \cdots & p \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ q & \cdots & q & p & \cdots & p \\ \hdashline p & \cdots & p & p & \cdots & p \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p & \cdots & p & p & \cdots & p \end{array}\right]_{n \times n}$$

Consider the null and alternative:

$$H_0 : \mathbf{H} = p\mathbf{1}\mathbf{1}' \qquad\qquad H_1 : \mathbf{H} = p\mathbf{1}\mathbf{1}' + (q - p)\mathbf{l}\mathbf{l}'$$

where $\mathbf{l}$ is a vector with the first n_1 entries being 1 and rest 0, $q \in (p, 1]$, and

- under the null, the top eigenvector is $\mathbf{v}_1 = \mathbf{1}/\sqrt{n}$ and thus

$$\mathbb{E}\left(\mathbf{v}_1'\mathbf{W}^2\mathbf{v}_1\right) = np(1 - p)$$
$$\mathrm{var}\left(\mathbf{v}_1'\mathbf{W}^2\mathbf{v}_1\right) = p(1 - p)\left[2(n - 1) + p^3 + (1 - p)^3\right]$$

- the natural test statistic based on Thm.18.3.1 is given by

$$T_n := \frac{2\lambda_1^2\left(\mathbf{1}'\mathbf{v}_1/\sqrt{n} - 1\right) + n\hat{p}(1 - \hat{p})}{\sqrt{\hat{p}(1 - \hat{p})\left[2(n - 1) + \hat{p}^3 + (1 - \hat{p})^3\right]}} \xrightarrow{d} \mathcal{N}(0, 1)$$

where $\hat{p} = \frac{1}{n(n-1)}\sum_{1 \leq i \leq j \leq n} x_{ij}$

Rank inference   Consider the hypothesis

$$H_0 : K = K_0 \qquad\qquad H_1 : K > K_0$$

where K is the true **rank** of the mean matrix $\mathbf{H}$ and K_0 is some pre-specified positive integer satisfying $K_0 \leq K$, define

$$\hat{w}_{ij} = x_{ij} - \sum_{i=1}^{K_0} \lambda_k \mathbf{e}'i\hat{\mathbf{v}}_k\hat{\mathbf{v}}_k'\mathbf{e}_j$$

under the null and some additional regularity conditions, $\hat{w}_{ij}$ is expected to be close to w_{ij}, hence the test statistic can be given by

$$T_n := \frac{\sum_{i=1}^n \hat{w}_{ii}}{\sqrt{\sum_{i=1}^n \hat{w}_{ii}^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

References

Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. Asymptotic theory of eigenvectors for random matrices with diverging spikes. *Journal of the American Statistical Association*, 117(538):996–1009, 2022.

Iain M Johnstone and Alexei Onatski. Testing in high-dimensional spiked models. *The Annals of Statistics*, 48(3), 2020.