

Topic 18: Eigenvalue and Spike Models

by Sai Zhang

Key points: .

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 $\mathcal{N}(\mathbf{0}, \Sigma)$, then the covariance can be decomposed into 2 parts, white noise and low rank

$$\Sigma = \text{Cov}(\mathbf{X}_i) = \mathbf{I} + \sum_{k=1}^M \theta_k \mathbf{v}_k \mathbf{v}_k' = \Sigma_0 + \Phi$$

where M denotes the **number of spikes** in the distribution of eigenvalues. The idea is: spikes deviate from a reference model along a **small fixed number** of unknown directions. If $\Phi = \mathbf{0}$, 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 \mathbf{v}_k
- How to test θ_k

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

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

against the null

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

with the key assumptions:

A1 Gaussian error

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

Under these assumptions, for the $n \times p$ data matrix $\mathbf{X} = (\mathbf{X}'_1 \cdots \mathbf{X}'_n)'$, $\mathbf{X}'\mathbf{X}$ has a p -dimensional **Wishart** distribution $W_p(n, \Sigma)$ with the degree of freedom n and covariance matrix Σ , 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 $W_p(n, \Sigma, \Psi)$ with n degrees of freedom, covariance matrix Σ and non-centrality matrix $\Psi = \Sigma^{-1}\mathbf{M}'\mathbf{M}$.

Definition 18.1.1: Density of Wishart Distribution

The PDF of Wishart distribution is defined as

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

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

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

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

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

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

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

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

where TW_1 is the Tracy-Widom distribution

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

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

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:

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

with $p \times p$ matrices

$$\begin{aligned} n_1 \mathbf{H} &= \sum_{k=1}^{n_1} \mathbf{x}_k \mathbf{x}_k' && \text{hypothesis SS} \\ n_1 \mathbf{E} &= \sum_{k=1}^{n_1} \mathbf{z}_k \mathbf{z}_k' && \text{error SS} \end{aligned}$$

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 functions¹

${}_p\mathcal{F}_q$

¹Hypergeometric 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 disappears, 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 null hypothesis that the population covariance $\mathbf{\Omega} = \mathbf{\Sigma}$, with the alternative of interest being

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

where $\theta > 0$, ϕ are unknown, and ϕ is normalized s.t. $\|\mathbf{\Sigma}^{-1/2}\phi\| = 1$. 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

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

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

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

and again, assume $\Sigma = \mathbf{I}_p$ (but NOT necessarily known), here, instead of Equation 18.1, consider

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

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

REG₀ Next, consider a linear regression with multivariate response

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

with known covariance matrix Σ of the i.i.d. Gaussian rows of the error matrix ϵ . Here, to test linear restrictions on the matrix of coefficients β , 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 β), under the null $H_0 : \mathbb{E}\mathbf{Y}_1 = 0$, versus the alternative

$$\mathbb{E}\mathbf{Y}_1 = \sqrt{n_1}\theta\psi\phi' \quad (18.4)$$

where $\theta > 0$, $\|\Sigma^{-1/2}\phi\| = 1$ 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 $\lambda_1, \dots, \lambda_p$ to

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

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

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

Again, W.L.O.G, assume $\Sigma = \mathbf{I}_p$. This **canonical form** of REG₀ is essentially equivalent to the setting of **matrix denoising**

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

REG Again, consider the linear regression

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

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

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

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 $n_2 \rightarrow \infty$, REG reduces to REG₀.

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 $t = 1, \dots, n_1 + n_2$. Partition the population and sample covariance matrices of the observations $(x'_t, y'_t)'$ into

$$\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \quad \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} \end{pmatrix}$$

respectively. Under $H_0 : \Sigma_{xy} = \mathbf{0}$, while the alternative is

$$\Sigma_{xy} = \sqrt{\frac{n_1 \theta}{n_1 \theta + n_1 + n_2}} \phi \psi' \quad (18.5)$$

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

$$\|\Sigma_{xx}^{-1/2} \phi\| = \|\Sigma_{yy}^{-1/2} \psi\| = 1$$

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

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

with

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

18.2.3 SMD

For a $\mathbf{X} = \mathbf{\Phi} + \mathbf{Z}/\sqrt{p}$ 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 $\mathbf{\Phi} = \theta \phi \phi'$. The null and the alternative is again as in 18.2. The nuisance vector $\phi \in \mathbb{R}^p$ is normalized s.t. $\|\phi\| = 1$.

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 $\lambda_1, \dots, \lambda_p$ to $\det(\mathbf{H} - \lambda \mathbf{E}) = 0$ 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**, **REG₀**: random \mathbf{H} and deterministic \mathbf{E}
- **PCA** and **SigD** are *parallel* to **REG₀**
- **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

$$\Lambda = \text{diag} \{ \lambda_1, \dots, \lambda_p \}$$

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

$$\mathbf{Z}_{ii} \sim \mathcal{N}(0, 2)$$

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

Figure 18.2: SMD and 5 Classes of Statistical Problems



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

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

where $\Phi = \Phi(\theta)$ is a p -dimensional matrix $\text{diag}\{\Phi_{11}, 0, \dots, 0\}$. Consider the hypergeometric functions of 2 matrix arguments Φ, Λ are defined as

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

where $\mathbf{a} = (a_1, \dots, a_p)$ and $\mathbf{b} = (b_1, \dots, b_q)$ 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	Φ_{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

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

which can be denoted as $\mathbf{n}, p \rightarrow_{\gamma} \infty$ where $\mathbf{n} = \{n_1, n_2\}$ and $\gamma = \{\gamma_1, \gamma_2\}$. Under the null (the true value of the spike $\theta_0 = 0$), $\lambda_1, \dots, \lambda_p$ are the eigenvalues of

- GOE/\sqrt{p} , for **SMD**
- $W_p(n_1, \mathbf{I}_p)/n_1$, for **PCA** and **REG₀**
- 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 $\lambda_1, \dots, \lambda_p$ follows

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

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

$$F_{\gamma}^W(\lambda)$$

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

Case	F_{γ}^{lim}	Density, $\lambda \in [\beta_-, \beta_+]$	β_{\pm}	Threshold $\hat{\theta}$
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	$\frac{(\gamma_1 + \gamma_2)R(\lambda)}{2\pi\gamma_1\lambda(\gamma_1 - \gamma_2\lambda)}$	$\gamma_1 \left(\frac{\rho \pm 1}{\rho \pm \gamma_2} \right)^2$	$\frac{\rho + \gamma_2}{1 - \gamma_2}$
where $R(\lambda) = \sqrt{(\beta_+ - \lambda)(\lambda - \beta_-)}$, $\rho = \sqrt{\gamma_1 + \gamma_2 - \gamma_1\gamma_2}$				

References

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