# Eigenvalues and eigenvectors of some spiked covariance models

Debashis Paul

University of California, Davis

*SEA'06 Workshop, MIT*

## Various high-dimensional problems

- Functional data analysis (e.g. longitudinal studies)
- Array signal processing (e.g. snapshot model)
- Spatio-temporal models (e.g. climate studies)

Interest is in the behavior of the larger eigenvalues and corresponding eigenvectors of sample covariance matrix, to infer about their population counterparts.

- Larger eigen-elements convey much of “useful” information.
- “Signal” is low dimensional.

## Observation model

$N$ dimensional observations :

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} (0, \Sigma_N)$$

Factor analysis model :

$$\Sigma_N = \sum_{k=1}^{N} \ell_k \mathbf{b}_k \mathbf{b}_k^T$$

Eigenvalues :

$$\ell_1 \geq \ell_2 \geq \ldots \geq \ell_N$$

Eigenvectors (subject to *identifiability*) :

$$\mathbf{b}_1, \ldots, \mathbf{b}_N$$

## Classical asymptotic theory : fixed $N$

**Result (Anderson (1963)) :** If $\ell_k$ is of multiplicity 1, then as $n \to \infty$,

$$\begin{aligned} \sqrt{n}(\widehat{\ell}_k - \ell_k) &\implies N(0, 2\ell_k^2) \\ \sqrt{n}(\widehat{\mathbf{b}}_k - \mathbf{b}_k) &\implies N(0, \sum_{1 \le j \ne k \le M} \frac{\ell_k \ell_j}{(\ell_j - \ell_k)^2} \mathbf{b}_j \mathbf{b}_j^T) \end{aligned}$$

- **New asymptotic regime :** dimension $N = N(n)$ so that $\frac{N}{n} \to c \in (0, \infty]$.
- In this talk will focus on $c \in (0, 1)$ (can extend to $c \in (0, \infty)$).

## Simple spiked model : Eigenvalues

$$\ell_1 \geq \ldots \geq \ell_M > 1 = \ldots = 1$$

**Interpretation :** $M$-dimensional signal

**Result (Baik and Silverstein (2005)) :** If $\ell_k > 1 + \sqrt{c}$ then

$$\widehat{\ell}_k \overset{a.s.}{\to} \rho_k = \ell_k \left(1 + \frac{c}{\ell_k - 1}\right).$$

If $\ell_k \leq 1 + \sqrt{c}$ then

$$\widehat{\ell}_k \overset{a.s.}{\to} (1 + \sqrt{c})^2.$$

**Result (Paul (2005), Onatski (2005), Baik & Silverstein (?)) :** If $\ell_k > 1 + \sqrt{c}$, of multiplicity 1, and if $\sqrt{n}|\frac{N}{n} - c| \to 0$,

$$\sqrt{n}(\widehat{\ell}_k - \rho_k) \Longrightarrow N(0, 2\ell_k^2(1 - \frac{c}{(\ell_k - 1)^2}))$$

## Simple spiked model : Eigenvectors

$$\ell_1 \geq \ldots \geq \ell_M > 1 = \ldots = 1$$

Without loss of generality $\mathbf{b}_k = \mathbf{e}_k$ (canonical basis vector in $\mathbb{R}^N$).

**Result (Paul (2005), Hoyle & Rattray (2004)) :** If $\ell_k > 1 + \sqrt{c}$, of multiplicity 1, then

$$|\langle \widehat{\mathbf{b}}_k, \mathbf{e}_k \rangle| \quad \overset{a.s.}{\to} \quad \sqrt{\left(1 - \frac{c}{(\ell_k - 1)^2}\right) / \left(1 + \frac{c}{\ell_k - 1}\right)}$$

$$|\langle \widehat{\mathbf{b}}_k, \mathbf{e}_j \rangle| \quad \overset{a.s.}{\to} \quad 0 \quad \text{for } j \neq k$$

If $\ell_k \leq 1 + \sqrt{c}$, of multiplicity 1, then

$$|\langle \widehat{\mathbf{b}}_k, \mathbf{e}_k \rangle| \overset{a.s.}{\to} 0.$$

## Simple spiked model : Eigenvectors

Let $\overline{\mathbf{e}}_k$ be the $k$-th canonical basis vector in $\mathbb{R}^M$.

**Result (Paul (2005), Onatski (2005)):** If $\ell_k > 1 + \sqrt{c}$, of multiplicity 1, and if $\sqrt{n}|\frac{N}{n} - c| \to 0$,

$$\sqrt{n}\left(\frac{\mathbf{q}_k}{\| \mathbf{q}_k \|} - \overline{\mathbf{e}}_k\right) \Longrightarrow N(0, \Sigma(\ell_k))$$

where $\mathbf{q}_k = \sum_{j=1}^M \langle \widehat{\mathbf{b}}_k, \mathbf{e}_j \rangle \overline{\mathbf{e}}_j$ and

$$\Sigma(\ell_k) = \left(1 - \frac{c}{(\ell_k - 1)^2}\right)^{-1} \sum_{1 \le j \neq k \le M} \frac{\ell_k \ell_j}{(\ell_j - \ell_k)^2} \overline{\mathbf{e}}_j \overline{\mathbf{e}}_j^T$$

## Spiked model with "nicely behaved" eigenvalues

Basic model with the assumption that the *Empirical Spectral Distribution (ESD)* of the eigenvalues $\ell_{M+1} \geq \ldots \ell_N$ converges (as $N \to \infty$) to a distribution $H(\cdot)$ compactly supported on $\mathbb{R}^+$.

- Complex valued case studied in detail by **El Karoui (2005)**
- Condition : There is a $\tau_{c,N}$ solving the equation

$$\frac{n}{N-M} = \int \left( \frac{x}{\tau_{c,N} - x} \right)^2 dH_{N,M}(x)$$

  and satisfying $\limsup_{n \to \infty} \ell_{M+1}/\tau_{c,N} < 1$), where $H_{N,M}$ is the ESD of $\ell_{M+1}, \ldots, \ell_N$.

**Result :** There is a threshold $\tau_c (= \lim_{n\to\infty} \tau_{c,N})$, such that if $\ell_M > \tau_c$, then for $1 \leq k \leq M$,

$$\widehat{\ell}_k \stackrel{a.s.}{\to} \rho_k = \ell_k \left[ 1 + c \int \frac{x}{\ell_k - x} dH(x) \right]$$

## Asymptotic normality above threshold $\tau_c$

**Result :** If $\ell_k > \tau_c$, $(1 \leq k \leq M)$ of multiplicity one, $\sqrt{n}|\frac{N}{n} - c| \to 0$ and $\sqrt{n} \parallel H_{N,M} - H \parallel_\infty \to 0$, then

$$\sqrt{n}(\widehat{\ell}_k - \rho_k) \Longrightarrow N(0, \sigma^2(\ell_k))$$

where

$$\sigma^2(\ell_k) = 2\ell_k^2 \left[1 - c \int \frac{x^2}{(\ell_k - x)^2} dH(x)\right].$$

Analogous results for sample eigenvectors (a.s. limit, asymptotic normality) corresponding to eigenvalues above the critical point $\tau_c$ and with multiplicity 1.

## Spatio-temporal models

- Arise frequently in areas like atmospheric science, econometrics.
- Apart from spatial (coordinate-wise) dependence, there is also a "time"-dependence - observations $X_1, \ldots, X_n$ are **not** independent.
- A simple model : assume that spatial variability and temporal variability do not depend on each other - **separable spatio-temporal model**

  i.e., the $N \times n$ data matrix $\mathbf{X}_n = [X_1 : \ldots : X_n]$ has covariance $\Sigma_N \otimes \Delta_n$ where $\Sigma_N$ is $N \times N$ and $\Delta_n$ is $n \times n$.

## Separable "Spiked" model

$\Sigma_N$ has eigenvalues $\ell_1 \geq \ldots \geq \ell_M > 1 = \ldots = 1$

Empirical distribution of the eigenvalues of $\Delta_n$ converges to a distribution $G$ compactly supported on $\mathbb{R}^+$ (and satisfies some technical condition).

Let $F^{\Delta}$ denote the limiting ESD of the matrix $\frac{1}{n}\mathbf{Z}\Delta_n\mathbf{Z}^T$ where $\mathbf{Z}$ is $(N-M) \times n$ with i.i.d. $N(0,1)$ entries.

**Result :** There is a threshold $\tau_{c,\Delta}$ such that, if $\ell_k > \tau_{c,\Delta}$ then $\widehat{\ell}_k \stackrel{a.s.}{\to} \rho_k$ where $\rho_k$ satisfies

$$\rho_k = \ell_k \int \frac{t}{1 - t\alpha_1(\rho_k, c)} dG(t)$$

where

$$\alpha_j(\rho, c) = c \int \frac{1}{(\rho - x)^j} dF^{\Delta}(x), \quad j = 1, 2, \ldots.$$

for $\rho > \max\ supp(F^{\Delta})$.

## Asymptotic normality

**Result :** Under the stated assumptions (plus a few regularity conditions), if $\ell_k$ is of multiplicity one, then

$$\sqrt{n}(\widehat{\ell}_k - \rho_k) \Longrightarrow N(0, \sigma^2(\ell_k))$$

where

$$\sigma^2(\ell_k) = 2\ell_k^2 \left[ \frac{\beta(\rho_k, c)(1 + \alpha_2(\rho_k, c)\beta(\rho_k, c))}{(1 + \ell_k \alpha_2(\rho_k, c)\beta(\rho_k, c))^2} \right]$$

with

$$\beta(\rho_k, c) := \int \frac{t^2}{(1 - t\alpha_1(\rho_k, c))^2} dG(t).$$

Analogous results for sample eigenvectors (a.s. limit, asymptotic normality) corresponding to eigenvalues above a threshold $\tau_c$ and with multiplicity 1.

## Basic structure

W.l.o.g. assume that $\Sigma_N$ is diagonal with *decreasing* diagonal elements. Express the data matrix $\mathbf{X}_n$ and sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^T$ as

$$\mathbf{X}_n = \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} \mathbf{S}_{AA} & \mathbf{S}_{AB} \\ \mathbf{S}_{BA} & \mathbf{S}_{BB} \end{bmatrix}$$

where $\mathbf{X}_A$ is $M \times n$, $\mathbf{X}_B$ is $(N-M) \times n$, $\mathbf{S}_{AA} = \frac{1}{n}\mathbf{X}_A\mathbf{X}_A^T$, $\mathbf{S}_{AB} = \frac{1}{n}\mathbf{X}_A\mathbf{X}_B^T$ and $\mathbf{S}_{BB} = \frac{1}{n}\mathbf{X}_B\mathbf{X}_B^T$.

Following relationship holds a.s.

$$(\mathbf{S}_{AA} + \mathbf{S}_{AB}(\widehat{\ell}_k I - \mathbf{S}_{BB})^{-1}\mathbf{S}_{BA})\frac{\mathbf{q}_k}{\| \mathbf{q}_k \|} = \widehat{\ell}_k \frac{\mathbf{q}_k}{\| \mathbf{q}_k \|} \tag{1}$$

## Implication of equation (1)

For eigenvalues bigger than a threshold, determination of asymptotic limits boils down to the computation of first two limiting spectral moments of the $n \times n$ matrix (where $\mathbf{Z}$ an $(N-M) \times n$ matrix with i.i.d. $N(0,1)$ entries)

$$K(\rho) = \Delta_n + \frac{1}{n}\Delta_n \mathbf{Z}^T \Gamma_N^{1/2}(\rho I - \frac{1}{n}\Gamma_N^{1/2}\mathbf{Z}\Delta_n\mathbf{Z}^T\Gamma_N^{1/2})^{-1}\Gamma_N^{1/2}\mathbf{Z}\Delta_n$$

- In the simple spiked model case, take $\Delta_n = I_n$, $\Gamma_N = I_{N-M}$.
- In the general spiked model case (with i.i.d. observations), $\Gamma_N =$ *diag* $(\ell_{M+1}, \ldots, \ell_N)$, $\Delta_n = I_n$.
- In the spatio-temporal case, take $\Gamma_N = I_{N-M}$.

## Some interesting issues

- Precise determination of the threshold at which phase transition occurs (cf. Baik and Silverstein (2004)).
- Behavior near the threshold.
- Rates of convergence.
- Behavior of the projection of the leading sample eigenvectors onto the "noise subspace" (spanned by eigenvectors with smaller eigenvalues)

## Concluding remarks

- Behavior of the larger eigenvalues and corresponding eigenvectors in the spiked models has a common pattern (bias, asymptotic normality above a threshold).
- A possible way of getting better (less biased) estimates of the larger eigenvalues and eigenvectors when some information is available on the smaller (noise) eigenvalues.
- Further generalizations (particularly in spatio-temporal and/or signal + noise settings).

**Acknowledgement**

Iain M. Johnstone

Jack Silverstein, Jinho Baik, Noureddine El Karoui