Principle components of spiked covariance matrices

Ke Wang

Hong Kong University of Science and Technology
(Joint work with Zhigang Bao & Xiucai Ding & Jingming Wang)

Sep 21, 2021

MSRI, Berkeley, USA

Estimate covariance matrices

$\mathbf{y} \in \mathbb{R}^m$: random vector with mean 0 and unknown covariance matrix Σ.

Empirically, collect n realizations $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ of $\mathbf{y} \in \mathbb{R}^m$.
Sample covariance matrix:

$$S = \frac{1}{n} X X^T$$

Main focus: the eigenstructure of S.

- m fixed, $n \to \infty$: Anderson '63, Muirhead '82, Tyler '83, etc.
- Both $m, n \to \infty$ and $\Sigma = I_m$ (null case): enormous progress recently.

Estimate covariance matrices

$\mathbf{y} \in \mathbb{R}^m$: random vector with mean 0 and unknown covariance matrix Σ.

Empirically, collect n realizations $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ of $\mathbf{y} \in \mathbb{R}^m$.
Sample covariance matrix:

$$S = \frac{1}{n} X X^T$$

Main focus: the eigenstructure of S.

- m fixed, $n \to \infty$: Anderson '63, Muirhead '82, Tyler '83, etc.
- Both $m, n \to \infty$ and $\Sigma = I_m$ (null case): enormous progress recently.

Spiked covariance matrix model

Proposed by Johnstone (2001).

Population covariance matrix:

$$\Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T,$$

$d_1 \geq d_2 \geq \cdots \geq d_r > 0, \quad r = O(1).$

Spiked (sample) covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T,$$

$\mu_1 \geq \cdots \geq \mu_m \geq 0$. Entries of $\sqrt{n}X$ i.i.d. with mean 0, var. 1.

- $m/n \to y \in (0, \infty)$.

Spiked covariance matrix model

Proposed by Johnstone (2001).

Population covariance matrix:

$$\Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T,$$

$d_1 \geq d_2 \geq \cdots \geq d_r > 0, \quad r = O(1).$

Spiked (sample) covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T,$$

$\mu_1 \geq \cdots \geq \mu_m \geq 0$. Entries of $\sqrt{n}X$ i.i.d. with mean 0, var. 1.

- $m/n \to y \in (0, \infty)$.

Spiked covariance matrix model

Proposed by Johnstone (2001).

Population covariance matrix:

$$\Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T,$$

$d_1 \geq d_2 \geq \cdots \geq d_r > 0, \quad r = O(1).$

Spiked (sample) covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T,$$

$\mu_1 \geq \cdots \geq \mu_m \geq 0$. Entries of $\sqrt{n}X$ i.i.d. with mean 0, var. 1.

- $m/n \to y \in (0, \infty)$.

Deformation of random matrices

Related deformation models:

▷ Spiked covariance matrix: Positive $\Sigma = I + S$, S fixed-rank

$$\Sigma^{1/2} X X^T \Sigma^{1/2} \quad \text{(multiplicative)}.$$

▷ Low-rank deformed Wigner: P fixed-rank Hermitian

$$W + P \quad \text{(additive)}.$$

▷ Matrix denoising model: $X + S$ with S fixed-rank

$$(X+S)^T (X+S) \quad \text{(additive \& multiplicative)}.$$

Related deformation models:

▷ Spiked covariance matrix: Positive $\Sigma = I + S$, S fixed-rank

$$\Sigma^{1/2} X X^T \Sigma^{1/2} \quad \text{(multiplicative)}.$$

▷ Low-rank deformed Wigner: P fixed-rank Hermitian

$$W + P \quad \text{(additive)}.$$

▷ Matrix denoising model: $X + S$ with S fixed-rank

$$(X+S)^T (X+S) \quad \text{(additive \& multiplicative)}.$$

Phase transition

BBP phase transition by Baik-Ben Arous-Péché '05 for extreme eigenvalues of spiked complex Gaussian covariance matrix. Phase transition for e.vector: Paul '07, Benaych-Georges-Nadakuditic '11 & '12.

General picture:

- (Supercritical): deformation strength bigger than a critical value c.

- (Subcritical): deformation strength less than a critical value c.

Picture source: *"The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices"*, Benaych-Georges-Nadakuditi, Advances in Mathematics, 227(1):494-521,2011.

Phase transition

BBP phase transition by Baik-Ben Arous-Péché '05 for extreme eigenvalues of spiked complex Gaussian covariance matrix. Phase transition for e.vector: Paul '07, Benaych-Georges-Nadakuditic '11 & '12.

General picture: $Q = \Sigma^{1/2} X X^T \Sigma^{1/2}, \quad \Sigma = I + d u u^T.$

- (Supercritical): deformation strength bigger than a critical value c.

- (Subcritical): deformation strength less than a critical value c.

Picture source: *"The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices"*, Benaych-Georges-Nadakuditi, Advances in Mathematics, 227(1):494-521,2011.

Previous results on deformation models

- ▷ Extreme eigenvalues
  - Convergent limit: Bai-Yao '12, Baik-Silverstein '06, Benaych-Georges- Nadakuditi '11 & '12, Capitaine-Donati-Martin '16, Ding '17, Knowles-Yin '13, Paul '07, etc.
  - Fluctuation: Bai-Yao '08, Bao-Pan-Zhou '15, Benaych-Georges-Guionnet-Maida '11, Bloemendal-Knowles-Yau-Yin '16, Bloemendal-Virág '13 & '16, Capitaine-Donati-Martin-Féral '09, Cai-Han-Pan '20, Knowles-Yin '13, Renfrew-Soshnikov '12, etc.
- ▷ Extreme eigenvectors
  - Convergent limit: Benaych-Georges-Nadakuditi '11 & '12, Capitaine '17, Ding '17, Paul '07.
  - Fluctuation: Paul '07, Bao-Ding-W. '20, Bao-Wang '21, Bloemendal-Knowles-Yau-Yin '16, Capitaine-Donati-Martin '18, Fan-Fan-Han-Lv '20.

Previous results on spiked covariance matrices

Spiked covariance matrices:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T .$$

▷ **Fluctuation of e.vectors**

- (Subcritical) [Bloemendal-Knowles-Yau-Yin '16] Suppose $d_i < \sqrt{y}$,

$$(w^T \xi_i)^2 = \frac{1}{m} \vartheta(d_i, y, w, u_i) \cdot \Theta(i, w)$$

  where $\Theta(i, w) \to \chi_1^2$.

- (Critical) Still open.

  [Bao-Wang '21] for deformed GUE.

We study the outlier e.vectors in the supercritical regime in full generality.

Previous results on spiked covariance matrices

Spiked covariance matrices:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T .$$

▷ **Fluctuation of e.vectors**

- (Subcritical) [Bloemendal-Knowles-Yau-Yin '16] Suppose $d_i < \sqrt{y}$,

$$(w^T \xi_i)^2 = \frac{1}{m} \vartheta(d_i, y, w, u_i) \cdot \Theta(i, w)$$

  where $\Theta(i, w) \to \chi_1^2$.

- (Critical) Still open.

  [Bao-Wang '21] for deformed GUE.

We study the outlier e.vectors in the supercritical regime in full generality.

New results on spiked covariance matrix

Spiked covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2}, \text{ where } \Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T.$$

Assumptions:

- $m/n \to y \in (0, +\infty)$ as $n \to \infty$.
- Entries of $\sqrt{n}X$ i.i.d. with mean 0, var. 1, bounded high moments.

Almost minimal assumptions on Σ: Fix an index i.

- (No boundedness) d_i's could be n-dependent;
- (Multiplicity and minimal supercritical condition) A set $I(i)$ s.t. any $t \in I(i)$ satisfying $d_t = d_i$ and $d_i - \sqrt{y} > n^{-1/3+\epsilon}$.
- (Non-overlapping condition)

$$\delta_i := \min_{j \notin I(i)} |d_i - d_j| > d_i^{3/2} (d_i - y^{1/2})^{-1/2} n^{-1/2+\epsilon}.$$

New results on spiked covariance matrix

Spiked covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2}, \text{ where } \Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T .$$

Assumptions:

- $m/n \to y \in (0, +\infty)$ as $n \to \infty$.
- Entries of $\sqrt{n}X$ i.i.d. with mean 0, var. 1, bounded high moments.

Almost minimal assumptions on Σ: Fix an index i.

- (No boundedness) d_i's could be n-dependent;
- (Multiplicity and minimal supercritical condition) A set $I(i)$ s.t. any $t \in I(i)$ satisfying $d_t = d_i$ and $d_i - \sqrt{y} > n^{-1/3+\epsilon}$.
- (Non-overlapping condition)

$$\delta_i := \min_{j \notin I(i)} |d_i - d_j| > d_i^{3/2} (d_i - y^{1/2})^{-1/2} n^{-1/2+\epsilon}.$$

Spiked covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T, \quad \Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T.$$

For the $|I(i)|$-fold eigenvalue d_i, projection $Z_I = \sum_{i \in I(i)} u_i u_i^T$.
Random projection $P_I = \sum_{i \in I(i)} \xi_i \xi_i^T$.

Generalized component $w^T P_I w = \sum_{i \in I(i)} (\xi_i^T w)^2$ for any unit vector w.

New results:

- Limiting distribution of $w^T P_I w$:

$$w^T P_I w = \frac{d_i^2 - y}{d_i(d_i + y)} w^T Z_I w + \underbrace{\text{random fluctuation}}_{\text{sum of Gaussian and } \chi_1^2 \text{ r.v.'s}}$$

- Limiting joint distribution of μ_i $(i \in I(i))$ and $w^T P_I w$.

Spiked covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T, \quad \Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T.$$

For the $|I(i)|$-fold eigenvalue d_i, projection $Z_I = \sum_{i \in I(i)} u_i u_i^T$.

Random projection $P_I = \sum_{i \in I(i)} \xi_i \xi_i^T$.

Generalized component $w^T P_I w = \sum_{i \in I(i)} (\xi_i^T w)^2$ for any unit vector w.

New results:

- Limiting distribution of $w^T P_I w$:

$$w^T P_I w = \frac{d_i^2 - y}{d_i(d_i + y)} w^T Z_I w + \underbrace{\text{random fluctuation}}_{\text{sum of Gaussian and } \chi_1^2 \text{ r.v.'s}}$$

- Limiting joint distribution of μ_i $(i \in I(i))$ and $w^T P_I w$.

Spiked covariance matrix:

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T, \quad \Sigma = I_m + \sum_{i=1}^{r} d_i u_i u_i^T.$$

For the $|I(i)|$-fold eigenvalue d_i, projection $Z_I = \sum_{i \in I(i)} u_i u_i^T$.
Random projection $P_I = \sum_{i \in I(i)} \xi_i \xi_i^T$.

Generalized component $w^T P_I w = \sum_{i \in I(i)} (\xi_i^T w)^2$ for any unit vector w.

New results:

- Limiting distribution of $w^T P_I w$:

$$w^T P_I w = \frac{d_i^2 - y}{d_i(d_i + y)} w^T Z_I w + \underbrace{\text{random fluctuation}}_{\text{sum of Gaussian and } \chi_1^2 \text{ r.v.'s}}$$

- Limiting joint distribution of μ_i $(i \in I(i))$ and $w^T P_I w$.

New results on eigenvector

For any fixed unit vector w, define

$$w_I := Z_I w, \quad \varsigma_I := \sum_{j \in [m] \setminus I} \frac{d_i \sqrt{d_j + 1}}{d_i - d_j} \langle w, u_j \rangle u_j.$$

Theorem (Bao, Ding, Wang and W., 2020)

$$w^T P_I w = \frac{d_i^2 - y}{d_i(d_i + y)} w^T Z_I w + \frac{1}{\sqrt{n(d_i^2 - y)}} \Theta_{w_I} + \frac{\|\varsigma_I\| \sqrt{d_i - y^{1/2}}}{\sqrt{n} d_i} \Lambda_{\varsigma_I}$$
$$+ \frac{\|\varsigma_I\|^2}{n d_i} \sum_{t \in I} (\Delta_{u_t})^2 - \frac{1}{n} \sum_{j \in [r] \setminus I} \frac{d_i d_j}{(d_i - d_j)^2} (\Pi_{u_j})^2 + O_{\prec}(R),$$

where

$$\left(\Theta_{w_I}, \Lambda_{\varsigma_I}, \{\Delta_{u_t}\}_{t \in I}, \{\Pi_{u_j}\}_{j \in [r] \setminus I}\right) \simeq \mathcal{N}(\mathbf{0}, \mathbf{V}_{r+2}).$$

Notation: $A \prec B$ if $|A| \leq n^{\epsilon} B$ w.h.p. for any ϵ.

Definition of $\mathbf{V}_{r+2}$

$$\mathbf{V}_{r+2} = A_{\mathsf{I}}^{\mathbf{w}} + \kappa_4 \frac{d_i^2 - y}{d_i^2} B_{\mathsf{I}}^{\mathbf{w}}$$

$A_{\mathsf{I}}^{\mathbf{w}}$ and $B_{\mathsf{I}}^{\mathbf{w}}$ are explicit symmetric $(r+2) \times (r+2)$ matrices, indexed by $w_{\mathsf{I}}, \varsigma_{\mathsf{I}}, \{u_t\}_{t \in \mathsf{I}}$ and $\{u_j\}_{j \in [r] \setminus I}$.

For instance, the non-zero entries of A are given by

$$\begin{aligned}
&A_{\mathsf{I}}^{\mathbf{w}}(w_{\mathsf{I}}, w_{\mathsf{I}}) = 2y\mathtt{h}(d_i)^2(1 + y\mathtt{h}(d_i)^2)\|w_{\mathsf{I}}\|^4, \quad A_{\mathsf{I}}^{w}(\varsigma_{\mathsf{I}}, \varsigma_{\mathsf{I}}) = \mathtt{g}(d_i)^2\|w_{\mathsf{I}}\|^2, \\
&A_{\mathsf{I}}^{\mathbf{w}}(u_t, u_t) = \mathtt{h}(d_i), \quad A_{\mathsf{I}}^{\mathbf{w}}(u_j, u_j) = \mathtt{l}(d_i)^2\|w_{\mathsf{I}}\|^2, \\
&A_{\mathsf{I}}^{\mathbf{w}}(\varsigma_{\mathsf{I}}, u_t) = \mathtt{g}(d_i)\sqrt{\mathtt{h}(d_i)}\langle w_{\mathsf{I}}, u_t\rangle, \quad A_{\mathsf{I}}^{\mathbf{w}}(\varsigma_{\mathsf{I}}, u_j) = \mathtt{g}(d_i)\mathtt{l}(d_i)\langle \varsigma_{\mathsf{I}}^0, u_j\rangle\|w_{\mathsf{I}}\|^2, \\
&A_{\mathsf{I}}^{\mathbf{w}}(u_t, u_j) = \sqrt{\mathtt{h}(d_i)}\mathtt{l}(d_i)\langle w_{\mathsf{I}}, u_t\rangle\langle \varsigma_{\mathsf{I}}^0, u_j\rangle, \quad \text{for } t \in \mathsf{I}, j \in [r] \setminus I,
\end{aligned}$$

where

$$\begin{aligned}
&\mathtt{f}(d) := \frac{y(1+d)}{d(d+y)}\Big(1 + \frac{d(1+d)}{d+y}\Big), \qquad \mathtt{g}(d) := \frac{2\sqrt{(d+1)(d+\sqrt{y})}}{d+y}, \\
&\mathtt{h}(d) := \frac{d+1}{d+y}, \qquad \mathtt{l}(d) := \frac{1+d}{\sqrt{d(d+y)}}.
\end{aligned}$$

Special examples

One-spike: $\Sigma = I_m + duu^T$. Generalized component $w^T P_I w = (\xi_1^T w)^2$.

- If $w = u$,

$$(\xi_1^T u)^2 = \frac{d^2 - y}{d(d+y)} + \frac{1}{\sqrt{n(d^2 - y)}}\Theta_u + O_\prec(R),$$

where

$$\Theta_u \simeq \mathcal{N}\left(0, f(d) + \kappa_4 g(d)\|u\|_{\ell_4}^4\right).$$

- If $w \in \{u\}^\perp$,

$$(\xi_1^T w)^2 = \frac{1}{nd}(\Delta_u)^2 + O_\prec(R),$$

where

$$\Delta_u \simeq \mathcal{N}\Big(0, h(d) + \kappa_4 h(d)\sum_{s=1}^m (u^s)^2 (w^s)^2\Big).$$

Special examples

One-spike: $\Sigma = I_m + duu^T$. Generalized component $w^T P_I w = (\xi_1^T w)^2$.

- If $w = u$,

$$(\xi_1^T u)^2 = \frac{d^2 - y}{d(d+y)} + \frac{1}{\sqrt{n(d^2 - y)}}\Theta_u + O_\prec(R),$$

where

$$\Theta_u \simeq \mathcal{N}\left(0, f(d) + \kappa_4 g(d)\|u\|_{\ell_4}^4\right).$$

- If $w \in \{u\}^\perp$,

$$(\xi_1^T w)^2 = \frac{1}{nd}(\Delta_u)^2 + O_\prec(R),$$

where

$$\Delta_u \simeq \mathcal{N}\Bigg(0, h(d) + \kappa_4 h(d) \sum_{s=1}^m (u^s)^2 (w^s)^2\Bigg).$$

Joint dist. of e.v. and e.vectors

Motivation:
One-spike: $\Sigma = I_m + duu^T$, d supercritical but unknown.

$$(\xi_1^T u)^2 = \frac{d^2 - y}{d(d+y)} + \frac{1}{\sqrt{n(d^2-y)}}\Theta_u + O_\prec(R),$$

Thus

$$\frac{\sqrt{n(d^2-y)}}{\sqrt{V(d)}}\left((\xi_1^T u)^2 - \frac{d^2-y}{d(d+y)}\right) \simeq \mathcal{N}(0,1).$$

Estimate d by $\hat{d} := \theta^{-1}(\mu_1)$. μ_1 the largest e.v. of Q.
However, $\mu_1 = \theta(d) + O(n^{-1/2})$ (Bloemendal-Knowles-Yau-Yin '16)

$$\implies \frac{(\hat{d})^2 - y}{\hat{d}(\hat{d}+y)} = \frac{d^2-y}{d(d+y)} + O(n^{-1/2}).$$

Motivation:
One-spike: $\Sigma = I_m + duu^T$, d supercritical but unknown.

$$(\xi_1^T u)^2 = \frac{d^2 - y}{d(d+y)} + \frac{1}{\sqrt{n(d^2 - y)}}\Theta_u + O_{\prec}(R),$$

Thus

$$\frac{\sqrt{n(d^2-y)}}{\sqrt{V(d)}}\left((\xi_1^T u)^2 - \frac{d^2-y}{d(d+y)}\right) \simeq \mathcal{N}(0,1).$$

Estimate d by $\hat{d} := \theta^{-1}(\mu_1)$. μ_1 the largest e.v. of Q.
However, $\mu_1 = \theta(d) + O(n^{-1/2})$ (Bloemendal-Knowles-Yau-Yin '16)

$$\Longrightarrow \frac{(\hat{d})^2 - y}{\hat{d}(\hat{d}+y)} = \frac{d^2-y}{d(d+y)} + O(n^{-1/2}).$$

Joint dist. of e.v. and e.vectors: simple d_i

Assume $I = \{i\}$. The generalized component

$$
\begin{aligned}
(\xi_i^T w)^2 = & \frac{d_i^2 - y}{d_i(d_i+y)}(u_i^T w)^2 + \frac{1}{\sqrt{n(d_i^2-y)}}\Theta_{w_i} + \frac{\|\varsigma_i\|\sqrt{d_i - y^{1/2}}}{\sqrt{n}d_i}\Lambda_{\varsigma_i} \\
& + \frac{\|\varsigma_i\|^2}{nd_i}(\Delta_{u_i})^2 - \frac{1}{n}\sum_{j\in[r]\setminus\{i\}}\frac{d_i d_j}{(d_i-d_j)^2}(\Pi_{u_j})^2 + O_\prec(R).
\end{aligned}
$$

Theorem (Bao, Ding, Wang and W., 2020)

For the e.v. μ_i,

$$
\mu_i = \underbrace{1 + d_i + y + \frac{y}{d_i}}_{\theta(d_i)} + \frac{\sqrt{d_i^2 - y}}{\sqrt{n}}\Phi_i + O_\prec(\mathcal{R})
$$

and

$$
\left(\Phi_i, \Theta_{w_i}, \Lambda_{\varsigma_i}, \Delta_{u_i}, \{\Pi_{u_j}\}_{j\in[r]\setminus\{i\}}\right) \simeq \mathcal{N}\left(0, C_{r+3}\right).
$$

Joint dist. of e.v. and e.vectors: multiple d_i

The generalized components

$$w^T P_I w = \frac{d_i^2 - y}{d_i(d_i+y)} w^T Z_I w + \frac{1}{\sqrt{n(d_i^2-y)}} \Theta_{w_I} + \frac{\|\varsigma_I\| \sqrt{d_i - y^{1/2}}}{\sqrt{n} d_i} \Lambda_{\varsigma_I}$$
$$+ \frac{\|\varsigma_I\|^2}{n d_i} \sum_{t \in I} (\Delta_{u_t})^2 - \frac{1}{n} \sum_{j \in [r] \setminus I} \frac{d_i d_j}{(d_i - d_j)^2} (\Pi_{u_j})^2 + O_{\prec}(R).$$

Theorem (Bao, Ding, Wang and W., 2020)

Assume $|I| = |I(i)| > 1$. The eigenvalues have the expansion

$$\mu_t = \theta(d_i) + \frac{\sqrt{d_i^2 - y}}{\sqrt{n}} \lambda_t(\Phi) + O_{\prec}(R)$$

for $t \in I(i)$ and $\Phi = (\Phi_{st})_{s,t}$ is a $|I| \times |I|$ GOE. Then,

$$\left(\{\Phi_{st}\}_{t \in I, t \geq s}, \Theta_{w_I}, \Lambda_{\varsigma_I}, \{\Delta_{u_t}\}_{t \in I}, \{\Pi_{u_j}\}_{j \in [r] \setminus I} \right) \simeq \mathcal{N}(0, \hat{C}).$$

Statistical applications

Problem: hypothesis testing on the eigenspaces of covariance matrices.

Any set $\mathcal{I} \subset [r_0]$. Set

$$Z_{\mathcal{I}} = \sum_{t \in \mathcal{I}} u_t u_t^T.$$

Consider a statistical inference problem:

$$\mathbf{H}_0 : \; Z_{\mathcal{I}} = Z_0 \quad \text{vs} \quad \mathbf{H}_a : \; Z_{\mathcal{I}} \neq Z_0,$$

for a given projection Z_0.

Goal: construct a data-dependent test statistic for the inference.

Statistical applications

Problem: hypothesis testing on the eigenspaces of covariance matrices.

Any set $\mathcal{I} \subset [r_0]$. Set

$$Z_{\mathcal{I}} = \sum_{t \in \mathcal{I}} u_t u_t^T.$$

Consider a statistical inference problem:

$$\mathbf{H}_0: \; Z_{\mathcal{I}} = Z_0 \quad \text{vs} \quad \mathbf{H}_a: \; Z_{\mathcal{I}} \neq Z_0,$$

for a given projection Z_0.

Goal: construct a data-dependent test statistic for the inference.

Statistical applications

Suppose $Z_0 = \sum_{i \in \mathcal{I}} v_i v_i^T$. Consider

$$\mathcal{T} := \sum_{i \in \mathcal{I}} \left(v_i^T \mathrm{P}_{\mathcal{I}} v_i - \frac{(\hat{d}_i)^2 - y}{\hat{d}_i(\hat{d}_i + y)} \right),$$

where $\mathrm{P}_{\mathcal{I}} = \sum_{i \in \mathcal{I}} \xi_i \xi_i^T$ and

$$\hat{d}_i = \theta^{-1}(\mu_i) = \frac{1}{2}(-y + \mu_i - 1) + \frac{1}{2}\sqrt{(-y + \mu_i - 1)^2 - 4y}.$$

Suppose $\mathbf{H}_0: \ Z_{\mathcal{I}} = Z_0$ holds. Under certain assumptions,

$$\mathbb{T} := \frac{\sqrt{n}\mathcal{T}}{\sqrt{\mathbb{V}(\mathbf{d}_{\mathcal{I}})}} \simeq \mathcal{N}(0, 1).$$

Here $\mathbb{V}(\mathbf{d}_{\mathcal{I}})$ depends on $\mathbf{d}_{\mathcal{I}} = (\hat{d}_i)_{i \in \mathcal{I}}$.

Statistical applications

Suppose $Z_0 = \sum_{i\in\mathcal{I}} v_i v_i^T$. Consider

$$\mathcal{T} := \sum_{i\in\mathcal{I}} \Big(v_i^T \mathrm{P}_{\mathcal{I}} v_i - \frac{(\hat{d}_i)^2 - y}{\hat{d}_i(\hat{d}_i + y)} \Big),$$

where $\mathrm{P}_{\mathcal{I}} = \sum_{i\in\mathcal{I}} \xi_i \xi_i^T$ and

$$\hat{d}_i = \theta^{-1}(\mu_i) = \frac{1}{2}(-y + \mu_i - 1) + \frac{1}{2}\sqrt{(-y + \mu_i - 1)^2 - 4y}.$$

Suppose $\mathbf{H}_0: \; Z_{\mathcal{I}} = Z_0$ holds. Under certain assumptions,

$$\mathbb{T} := \frac{\sqrt{n}\mathcal{T}}{\sqrt{\mathtt{V}(\mathbf{d}_{\mathcal{I}})}} \simeq \mathcal{N}(0,1).$$

Here $\mathtt{V}(\mathbf{d}_{\mathcal{I}})$ depends on $\mathbf{d}_{\mathcal{I}} = (\hat{d}_i)_{i\in\mathcal{I}}$.

Sketch of proof for the e.vector

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T.$$

The empirical spectral distributions (ESD) of XX^T:

$$F_1(x) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\{\lambda_i(XX^T) \leq x\}}.$$

The Marchenko-Pastur (MP) law:

$$F_1(x) \to F_{MP,1}(x).$$

The Stieltjes's transform:

$$m_1(z) := \int \frac{1}{x-z} \, \mathrm{d}F_{MP,1}(x).$$

The Green function:

$$\mathcal{G}_1(z) = (XX^T - z)^{-1}.$$

$$Q = \Sigma^{1/2} X X^T \Sigma^{1/2} = \sum_{i=1}^{m} \mu_i \xi_i \xi_i^T .$$

The empirical spectral distributions (ESD) of XX^T:

$$F_1(x) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\{\lambda_i(XX^T) \leq x\}}.$$

The Marchenko-Pastur (MP) law:

$$F_1(x) \to F_{MP,1}(x).$$

The Stieltjes's transform:

$$m_1(z) := \int \frac{1}{x - z} \, \mathrm{d}F_{MP,1}(x).$$

The Green function:

$$\mathcal{G}_1(z) = (XX^T - z)^{-1}.$$

Sketch of proof: Green function representation

Generalized component $w^T P_I w = \sum_{t \in I} (\xi_t^T w)^2$.

Green function representation of $w^T P_I w$: From

$$w^T (Q - z)^{-1} w = \sum_{i=1}^{n} \frac{(\xi_i^T w)^2}{\mu_i - z},$$

select a contour $\theta(\Gamma_i)$ that encloses exactly $|I|$ e.v. of Q, i.e. μ_t $(t \in I)$, by residue theorem,

$$w^T P_I w = \sum_{i \in I} (\xi_i^T w)^2 = -\frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} w^T (Q - z)^{-1} w \, \mathrm{d}z.$$

Set $\Sigma = I + S = I + VDV^T$. $\quad \mathcal{G}_1(z) = (XX^T - z)^{-1}$.

$$(Q - z)^{-1} = \left(\Sigma^{\frac{1}{2}} XX^T \Sigma^{\frac{1}{2}} - zI\right)^{-1} = \Sigma^{-\frac{1}{2}} \left(\mathcal{G}_1^{-1}(z) + zVDV^T\right)^{-1} \Sigma^{-\frac{1}{2}}$$

$$= \Sigma^{-\frac{1}{2}} \mathcal{G}_1(z) \Sigma^{-\frac{1}{2}} - z\Sigma^{-\frac{1}{2}} \mathcal{G}_1(z) V \left(D^{-1} + zV^T \mathcal{G}_1(z) V\right)^{-1} V^T \mathcal{G}_1(z) \Sigma^{-\frac{1}{2}}.$$

Sketch of proof: Green function representation

Generalized component $w^T P_I w = \sum_{t\in I}(\xi_t^T w)^2$.

Green function representation of $w^T P_I w$: From

$$w^T(Q-z)^{-1}w = \sum_{i=1}^{n} \frac{(\xi_i^T w)^2}{\mu_i - z},$$

select a contour $\theta(\Gamma_i)$ that encloses exactly $|I|$ e.v. of Q, i.e. μ_t $(t \in I)$, by residue theorem,

$$w^T P_I w = \sum_{i\in I}(\xi_i^T w)^2 = -\frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} w^T(Q-z)^{-1}w \, \mathrm{d}z.$$

Set $\Sigma = I + S = I + VDV^T$. $\quad \mathcal{G}_1(z) = (XX^T - z)^{-1}$.

$$(Q-z)^{-1} = \left(\Sigma^{\frac{1}{2}}XX^T\Sigma^{\frac{1}{2}} - zI\right)^{-1} = \Sigma^{-\frac{1}{2}}\left(\mathcal{G}_1^{-1}(z) + zVDV^T\right)^{-1}\Sigma^{-\frac{1}{2}}$$

$$= \Sigma^{-\frac{1}{2}}\mathcal{G}_1(z)\Sigma^{-\frac{1}{2}} - z\Sigma^{-\frac{1}{2}}\mathcal{G}_1(z)V\left(D^{-1} + zV^T\mathcal{G}_1(z)V\right)^{-1}V^T\mathcal{G}_1(z)\Sigma^{-\frac{1}{2}}.$$

Sketch of proof: Green function representation

Generalized component $w^T P_I w = \sum_{t \in I} (\xi_t^T w)^2$.

Green function representation of $w^T P_I w$: From

$$w^T (Q - z)^{-1} w = \sum_{i=1}^{n} \frac{(\xi_i^T w)^2}{\mu_i - z},$$

select a contour $\theta(\Gamma_i)$ that encloses exactly $|I|$ e.v. of Q, i.e. μ_t $(t \in I)$, by residue theorem,

$$w^T P_I w = \sum_{i \in I} (\xi_i^T w)^2 = -\frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} w^T (Q - z)^{-1} w \, \mathrm{d}z.$$

Set $\Sigma = I + S = I + VDV^T$. $\mathcal{G}_1(z) = (XX^T - z)^{-1}$.

$$(Q - z)^{-1} = \left(\Sigma^{\frac{1}{2}} XX^T \Sigma^{\frac{1}{2}} - zI \right)^{-1} = \Sigma^{-\frac{1}{2}} \left(\mathcal{G}_1^{-1}(z) + zVDV^T \right)^{-1} \Sigma^{-\frac{1}{2}}$$

$$= \Sigma^{-\frac{1}{2}} \mathcal{G}_1(z) \Sigma^{-\frac{1}{2}} - z \Sigma^{-\frac{1}{2}} \mathcal{G}_1(z) V \left(D^{-1} + zV^T \mathcal{G}_1(z) V \right)^{-1} V^T \mathcal{G}_1(z) \Sigma^{-\frac{1}{2}}.$$

Sketch of proof: Green function representation

Generalized component $w^T P_I w = \sum_{t\in I}(\xi_t^T w)^2$.

Green function representation of $w^T P_I w$: From

$$w^T(Q-z)^{-1}w = \sum_{i=1}^{n} \frac{(\xi_i^T w)^2}{\mu_i - z},$$

select a contour $\theta(\Gamma_i)$ that encloses exactly $|I|$ e.v. of Q, i.e. μ_t $(t \in I)$, by residue theorem,

$$w^T P_I w = \sum_{i\in I}(\xi_i^T w)^2 = -\frac{1}{2\pi \mathtt{i}} \oint_{\theta(\Gamma_i)} w^T(Q-z)^{-1}w \, \mathrm{d}z.$$

Set $\Sigma = I + S = I + VDV^T$. $\quad \mathcal{G}_1(z) = (XX^T - z)^{-1}.$

$$(Q-z)^{-1} = \left(\Sigma^{\frac{1}{2}} XX^T \Sigma^{\frac{1}{2}} - zI\right)^{-1} = \Sigma^{-\frac{1}{2}} \left(\mathcal{G}_1^{-1}(z) + zVDV^T\right)^{-1} \Sigma^{-\frac{1}{2}}$$
$$= \Sigma^{-\frac{1}{2}} \mathcal{G}_1(z) \Sigma^{-\frac{1}{2}} - z\Sigma^{-\frac{1}{2}} \mathcal{G}_1(z) V \left(D^{-1} + zV^T \mathcal{G}_1(z) V\right)^{-1} V^T \mathcal{G}_1(z) \Sigma^{-\frac{1}{2}}.$$

Sketch of proof

Set $\widetilde{w} = \Sigma^{-\frac{1}{2}} w$.

$$w^T P_I w = \frac{1}{2\pi \mathtt{i}} \oint_{\theta(\Gamma_i)} z \widetilde{w}^T \mathcal{G}_1(z) V \underbrace{\left(D^{-1} + z V^T \mathcal{G}_1(z) V\right)^{-1}}_{\text{Apply resolvent expansion \& } \mathcal{G}_1(z) \approx m_1(z) I} V^T \mathcal{G}_1(z) \widetilde{w} \, \mathrm{d}z.$$

Denote $\Xi(z) = \mathcal{G}_1(z) - m_1(z) I$.

$$\begin{aligned} w^T P_I w = {} & \frac{d_i^2 - y}{d_i(d_i + y)} w^T Z_I w + f_1(d_i) w_I^T \Xi(z) w_I + f_2(d_i) \varsigma_I^T \Xi(z) w_I \\ & + f_3(d_i) w_I^T \Xi'(z) w_I + f_4(d_i) \sum_{t \in I} \left(u_t^T \Xi(z) \varsigma_I \right)^2 \\ & + \sum_{j \in [r] \setminus I} f_5(d_i, d_j) \left(u_j^T \Xi(z) w_I \right)^2 + O_{\prec}(R) \qquad \text{at } z = \theta(d_i). \end{aligned}$$

Sketch of proof

Set $\widetilde{w} = \Sigma^{-\frac{1}{2}} w$.

$$w^T P_I w = \frac{1}{2\pi \mathrm{i}} \oint_{\theta(\Gamma_i)} z \widetilde{w}^T \mathcal{G}_1(z) V \underbrace{\left(D^{-1} + zV^T \mathcal{G}_1(z) V\right)^{-1}}_{\text{Apply resolvent expansion \& } \mathcal{G}_1(z) \approx m_1(z) I} V^T \mathcal{G}_1(z) \widetilde{w} \, \mathrm{d}z.$$

Denote $\Xi(z) = \mathcal{G}_1(z) - m_1(z) I$.

$$\begin{aligned} w^T P_I w = & \frac{d_i^2 - y}{d_i(d_i + y)} w^T Z_I w + f_1(d_i) w_I^T \Xi(z) w_I + f_2(d_i) \varsigma_I^T \Xi(z) w_I \\ & + f_3(d_i) w_I^T \Xi'(z) w_I + f_4(d_i) \sum_{t \in I} \left(u_t^T \Xi(z) \varsigma_I \right)^2 \\ & + \sum_{j \in [r] \setminus I} f_5(d_i, d_j) \left(u_j^T \Xi(z) w_I \right)^2 + O_\prec(R) \qquad \text{at } z = \theta(d_i). \end{aligned}$$

Sketch of proof: further reduction

Derive the law for the random vector

$$\mathcal{Q} := \left(w_I^T \Xi'(z) w_I, w_I^T \Xi(z) w_I, \varsigma_I^T \Xi(z) w_I, \{u_t^T \Xi(z) \varsigma_I\}_{t \in I}, \{u_j^T \Xi(z) w_I\}_{j \in [r] \setminus I} \right)$$

at $z = \theta(d_i)$.

Goal: Show $\mathcal{Q}$ is multivariate Gaussian.

Let $\mathcal{P}$ be a linear combination of the components of $\mathcal{Q}$ with any appropriately scaled deterministic coefficients.

Sketch of proof: further reduction

Derive the law for the random vector

$$\mathcal{Q} := \left(w_I^T \Xi'(z) w_I, w_I^T \Xi(z) w_I, \varsigma_I^T \Xi(z) w_I, \{u_t^T \Xi(z) \varsigma_I\}_{t \in I}, \{u_j^T \Xi(z) w_I\}_{j \in [r] \setminus I} \right)$$

at $z = \theta(d_i)$.

Goal: Show $\mathcal{Q}$ is multivariate Gaussian.

Let $\mathcal{P}$ be a linear combination of the components of $\mathcal{Q}$ with any appropriately scaled deterministic coefficients.

Sketch of proof: further reduction

Derive the law for the random vector

$$\mathcal{Q} := \left(w_I^T \Xi'(z) w_I, w_I^T \Xi(z) w_I, \varsigma_I^T \Xi(z) w_I, \{u_t^T \Xi(z) \varsigma_I\}_{t \in I}, \{u_j^T \Xi(z) w_I\}_{j \in [r] \setminus I} \right)$$

at $z = \theta(d_i)$.

Goal: Show $\mathcal{Q}$ is multivariate Gaussian.

Let $\mathcal{P}$ be a linear combination of the components of $\mathcal{Q}$ with any appropriately scaled deterministic coefficients.

Sketch of proof: recursive estimates

Goal: Show $\mathcal{P}$ is Gaussian.

Our strategy is to establish the recursive estimates

- $\mathbb{E}\mathcal{P} = o(1)$;
- $\mathbb{E}\mathcal{P}^l = (l-1)V \cdot \mathbb{E}\mathcal{P}^{l-2} + o(1)$ for $l \geq 2$.

Key ingredients in the proof of recursive estimates:

- **Cumulant expansion formula:** For $f \in C^{\ell+1}(\mathbb{R})$ and ξ a centered random variable with finite $l+2$ moments,

$$\mathbb{E}(\xi f(\xi)) = \sum_{k=1}^{\ell} \frac{\kappa_{k+1}(\xi)}{k!} \mathbb{E}(f^{(k)}(\xi)) + \mathbb{E}(\epsilon_\ell(\xi f(\xi))).$$

  Applications in RMT: Khorunzhy-Khoruzhenko-Pastur '96, Lytova-Pastur 09', Lee-Schnelli '16, He-Knowles '16.

Sketch of proof: recursive estimates

Goal: Show $\mathcal{P}$ is Gaussian.

Our strategy is to establish the recursive estimates

- $\mathbb{E}\mathcal{P} = o(1)$;
- $\mathbb{E}\mathcal{P}^l = (l-1)V \cdot \mathbb{E}\mathcal{P}^{l-2} + o(1) \quad$ for $l \geq 2$.

Key ingredients in the proof of recursive estimates:

- **Cumulant expansion formula:** For $f \in C^{\ell+1}(\mathbb{R})$ and ξ a centered random variable with finite $l+2$ moments,

$$\mathbb{E}(\xi f(\xi)) = \sum_{k=1}^{\ell} \frac{\kappa_{k+1}(\xi)}{k!} \mathbb{E}(f^{(k)}(\xi)) + \mathbb{E}(\epsilon_\ell(\xi f(\xi))).$$

  Applications in RMT: Khorunzhy-Khoruzhenko-Pastur '96, Lytova-Pastur 09', Lee-Schnelli '16, He-Knowles '16.

Sketch of proof: key technical inputs

Key ingredients in the proof of recursive estimates.

- **Isotropic local laws:** large deviation bounds of

$$\langle u, (\mathcal{G}_1^{(s)}(z) - m_1^{(s)}(z) I) v \rangle \quad \text{for } s \in \mathbb{N}.$$

  Established in Bloemendal-Erdős-Knowles-Yau-Yin '16, Knowles-Yin '17.

- **Convergence rate of VESD**: Denote $H = XX^T$. $\lambda_i(H)$ the i-th largest e.v. and ϕ_i the associated unit e. vector. For a fixed unit vector $\mathbf{v}$, the *eigenvector empirical spectral distribution* (VESD) $F_{1n}^{\mathbf{v}}(x) = \sum_{i=1}^{m} \left| \langle \phi_i, \mathbf{v} \rangle \right|^2 \mathbf{1}(\lambda_i(H) \le x)$.

$$\sup_x |F_{1n}^{\mathbf{v}}(x) - F_{MP,1}(x)| \prec n^{-\frac{1}{2}}.$$

  Established in Xi-Yang-Yin '20.

Sketch of proof: key technical inputs

Key ingredients in the proof of recursive estimates.

- **Isotropic local laws:** large deviation bounds of

$$\langle u, (\mathcal{G}_1^{(s)}(z) - m_1^{(s)}(z) I) v \rangle \quad \text{for } s \in \mathbb{N}.$$

  Established in Bloemendal-Erdős-Knowles-Yau-Yin '16, Knowles-Yin '17.

- **Convergence rate of VESD**: Denote $H = XX^T$. $\lambda_i(H)$ the i-th largest e.v. and ϕ_i the associated unit e. vector. For a fixed unit vector $\mathbf{v}$, the *eigenvector empirical spectral distribution* (VESD) $F_{1n}^{\mathbf{v}}(x) = \sum_{i=1}^{m} \left|\langle \phi_i, \mathbf{v} \rangle\right|^2 \mathbf{1}(\lambda_i(H) \le x)$.

$$\sup_x |F_{1n}^{\mathbf{v}}(x) - F_{MP,1}(x)| \prec n^{-\frac{1}{2}}.$$

  Established in Xi-Yang-Yin '20.

Related results

Supercritical regime: fluctuation of outlier e.v. and e.vectors for deformation models.

- Knowles-Yin '13 (deformed Wigner): limiting dist. of the outlier eigenvalues in full generality. Proof based on a "two-step" comparison method.
- Capitaine-Donati-Martin '18 (deformed Wigner): fluctuation of outlier eigenvectors where the deformation is diagonal and entries of Wigner have symmetric dist. and satisfy Poincaré inequality.
- Fan-Fan-Han-Lv '20 (deformed Wigner): fluctuation of outlier e.vectors assuming the spikes are diverging sufficiently fast.
- Bao-Ding-W. '20 (matrix denoising model): limiting dist. of the outlier singular vector in full generality.

THANK YOU!

[1] Research supported by Hong Kong RGC grant GRF 16301618 and GRF 16308219 and ECS 26304920.