Probability and Stochastic Process II: Random Matrix Theory and Applications Lecture 5: Spiked Model

Zhenyu Liao, Tiebin Mi, Caiming Qiu

School of Electronic Information and Communications (EIC)
Huazhong University of Science and Technology (HUST)

April 5, 2023

Outline

What we will have today

- sample covariance matrix $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ for some low-rank matrix $\mathbf{P}$
- extreme eigenvalues of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- extreme eigenvectors of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- phase transition behavior and debiasing

What we will have today

- » sample covariance matrix $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ for some low-rank matrix $\mathbf{P}$
- » extreme eigenvalues of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- » extreme eigenvectors of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- » phase transition behavior and debiasing

What we will have today

- sample covariance matrix $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ for some low-rank matrix $\mathbf{P}$
- extreme eigenvalues of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- extreme eigenvectors of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- phase transition behavior and debiasing

What we will have today

- sample covariance matrix $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ for some low-rank matrix $\mathbf{P}$
- extreme eigenvalues of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- extreme eigenvectors of $\hat{\mathbf{C}}$ and connection to those of the low-rank $\mathbf{P}$
- phase transition behavior and debiasing

Outline

Low-rank update of SCM: eigenvalues

Low-rank update of SCM: eigenvectors

Limiting fluctuation

Applications

Low-rank perturbation of SCM

We have studied:

- spectral behavior of the SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{C}^{\frac{1}{2}}$ for generic $\mathbf{C}$ and $\mathbf{Z}$ with i.i.d. entries
- in particular, how the eigenvalue distribution of $\hat{\mathbf{C}}$ (the previous μ) depends on that of $\mathbf{C}$ (denoted ν) and the dimension ratio $c = \lim p/n$
- characterization via implicit fixed point equation of the Stieltjes transform
- the behavior (e.g., location) of individual eigenvalue, however, remains unclear
- here, assess the behavior of individual eigenvalue and eigenvector via the **spiked model** analysis, in the *simple* setting of $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with low rank $\mathbf{P}$

Note that the limiting eigenvalue distribution of $\hat{\mathbf{C}} = \frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}$ is in fact the same as that of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}$, since the addition of low rank matrices asymptotically does not affect the **normalized** trace of the resolvent, and thus the Stieltjes transform.

Remark on the limiting eigenvalue distribution

Low-rank perturbation of SCM

We have studied:

- spectral behavior of the SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}\mathbf{C}^{\frac{1}{2}}$ for generic $\mathbf{C}$ and $\mathbf{Z}$ with i.i.d. entries
- in particular, how the eigenvalue distribution of $\hat{\mathbf{C}}$ (the previous μ) depends on that of $\mathbf{C}$ (denoted ν) and the dimension ratio $c = \lim p/n$
- characterization via implicit fixed point equation of the Stieltjes transform
- the behavior (e.g., location) of individual eigenvalue, however, remains unclear
- here, assess the behavior of individual eigenvalue and eigenvector via the **spiked model** analysis, in the *simple* setting of $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with low rank $\mathbf{P}$

Note that the limiting eigenvalue distribution of $\hat{\mathbf{C}} = \frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}$ is in fact the same as that of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T}$, since the addition of low rank matrices asymptotically does not affect the **normalized** trace of the resolvent, and thus the Stieltjes transform.

Remark on the limiting eigenvalue distribution

Low-rank perturbation of SCM

We have studied:

» spectral behavior of the SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}\mathbf{C}^{\frac{1}{2}}$ for generic $\mathbf{C}$ and $\mathbf{Z}$ with i.i.d. entries

» in particular, how the eigenvalue distribution of $\hat{\mathbf{C}}$ (the previous μ) depends on that of $\mathbf{C}$ (denoted ν) and the dimension ratio $c = \lim p/n$

» characterization via implicit fixed point equation of the Stieltjes transform

» the behavior (e.g., location) of individual eigenvalue, however, remains unclear

» here, assess the behavior of individual eigenvalue and eigenvector via the **spiked model** analysis, in the *simple* setting of $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with low rank $\mathbf{P}$

Note that the limiting eigenvalue distribution of $\hat{\mathbf{C}} = \frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}$ is in fact the same as that of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T}$, since the addition of low rank matrices asymptotically does not affect the **normalized** trace of the resolvent, and thus the Stieltjes transform.

Remark on the limiting eigenvalue distribution

Low-rank perturbation of SCM

We have studied:

» spectral behavior of the SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\top\mathbf{C}^{\frac{1}{2}}$ for generic $\mathbf{C}$ and $\mathbf{Z}$ with i.i.d. entries

» in particular, how the eigenvalue distribution of $\hat{\mathbf{C}}$ (the previous μ) depends on that of $\mathbf{C}$ (denoted ν) and the dimension ratio $c = \lim p/n$

» characterization via implicit fixed point equation of the Stieltjes transform

» the behavior (e.g., location) of individual eigenvalue, however, remains unclear

» here, assess the behavior of individual eigenvalue and eigenvector via the **spiked model** analysis, in the *simple* setting of $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with low rank $\mathbf{P}$

Note that the limiting eigenvalue distribution of $\hat{\mathbf{C}} = \frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\top(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}$ is in fact the same as that of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top$, since the addition of low rank matrices asymptotically does not affect the **normalized** trace of the resolvent, and thus the Stieltjes transform.

Remark on the limiting eigenvalue distribution

Low-rank perturbation of SCM

We have studied:

- spectral behavior of the SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}\mathbf{C}^{\frac{1}{2}}$ for generic $\mathbf{C}$ and $\mathbf{Z}$ with i.i.d. entries
- in particular, how the eigenvalue distribution of $\hat{\mathbf{C}}$ (the previous μ) depends on that of $\mathbf{C}$ (denoted ν) and the dimension ratio $c = \lim p/n$
- characterization via implicit fixed point equation of the Stieltjes transform
- the behavior (e.g., location) of individual eigenvalue, however, remains unclear
- here, assess the behavior of individual eigenvalue and eigenvector via the **spiked model** analysis, in the *simple* setting of $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with low rank $\mathbf{P}$

Note that the limiting eigenvalue distribution of $\hat{\mathbf{C}} = \frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}$ is in fact the same as that of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T}$, since the addition of low rank matrices asymptotically does not affect the **normalized** trace of the resolvent, and thus the Stieltjes transform.

Remark on the limiting eigenvalue distribution

Low-rank perturbation of SCM

We have studied:

» spectral behavior of the SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{C}^{\frac{1}{2}}$ for generic $\mathbf{C}$ and $\mathbf{Z}$ with i.i.d. entries

» in particular, how the eigenvalue distribution of $\hat{\mathbf{C}}$ (the previous μ) depends on that of $\mathbf{C}$ (denoted ν) and the dimension ratio $c = \lim p/n$

» characterization via implicit fixed point equation of the Stieltjes transform

» the behavior (e.g., location) of individual eigenvalue, however, remains unclear

» here, assess the behavior of individual eigenvalue and eigenvector via the **spiked model** analysis, in the *simple* setting of $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with low rank $\mathbf{P}$

Note that the limiting eigenvalue distribution of $\hat{\mathbf{C}} = \frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}$ is in fact the same as that of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}$, since the addition of low rank matrices asymptotically does not affect the **normalized** trace of the resolvent, and thus the Stieltjes transform.

Remark on the limiting eigenvalue distribution

Low-rank perturbation from the Marčenko-Pastur law

Consider $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$, $\mathbf{z}_i \in \mathbb{R}^p$ with standard i.i.d. entries and

$$\mathbf{C} = \mathbf{I}_p + \mathbf{P}, \quad \mathbf{P} = \sum_{i=1}^{k} \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$$

with k and $\ell_1 \geq \ldots \geq \ell_k > 0$ fixed with respect to n, p.

- note that here $\nu \equiv \lim_{p \to \infty} \mu_{\mathbf{C}} = \lim_{p \to \infty} \frac{p-k}{p} \delta_1 + \frac{1}{p} \sum_{i=1}^{k} \delta_{1+\ell_i} = \delta_1$
- so, while $\mathbf{C} \neq \mathbf{I}_p$, the limiting μ still follows the Marčenko-Pastur law
- however, we do not have "no eigenvalue outside the support," since the condition $\operatorname{dist}(1 + \ell_i, \operatorname{supp}(\nu)) \not\to 0$ for $i \in \{1, \ldots, k\}$ is violated
- and one may have some (order $O(1)$ in this setting) the eigenvalues of $\hat{\mathbf{C}}$ "jumping" out of the limiting support $\operatorname{supp}(\mu)$
- note for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so with its eigenvalues connected to those of $\mathbf{P}$

Low-rank perturbation from the Marčenko-Pastur law

Consider $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$, $\mathbf{z}_i \in \mathbb{R}^p$ with standard i.i.d. entries and

$$\mathbf{C} = \mathbf{I}_p + \mathbf{P}, \quad \mathbf{P} = \sum_{i=1}^{k} \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$$

with k and $\ell_1 \geq \ldots \geq \ell_k > 0$ fixed with respect to n, p.

» note that here $\nu \equiv \lim_{p\to\infty} \mu_{\mathbf{C}} = \lim_{p\to\infty} \frac{p-k}{p} \delta_1 + \frac{1}{p} \sum_{i=1}^k \delta_{1+\ell_i} = \delta_1$

» so, while $\mathbf{C} \neq \mathbf{I}_p$, the limiting μ still follows the Marčenko-Pastur law

» however, we do not have "no eigenvalue outside the support," since the condition $\operatorname{dist}(1 + \ell_i, \operatorname{supp}(\nu)) \not\to 0$ for $i \in \{1, \ldots, k\}$ is violated

» and one may have some (order $O(1)$ in this setting) the eigenvalues of $\hat{\mathbf{C}}$ "jumping" out of the limiting support $\operatorname{supp}(\mu)$

» note for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so with its eigenvalues connected to those of $\mathbf{P}$

Low-rank perturbation from the Marčenko-Pastur law

Consider $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$, $\mathbf{z}_i \in \mathbb{R}^p$ with standard i.i.d. entries and

$$\mathbf{C} = \mathbf{I}_p + \mathbf{P}, \quad \mathbf{P} = \sum_{i=1}^{k} \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$$

with k and $\ell_1 \geq \ldots \geq \ell_k > 0$ fixed with respect to n, p.

- note that here $\nu \equiv \lim_{p\to\infty} \mu_{\mathbf{C}} = \lim_{p\to\infty} \frac{p-k}{p} \delta_1 + \frac{1}{p} \sum_{i=1}^k \delta_{1+\ell_i} = \delta_1$
- so, while $\mathbf{C} \neq \mathbf{I}_p$, the limiting μ still follows the Marčenko-Pastur law
- however, we do not have "no eigenvalue outside the support," since the condition $\operatorname{dist}(1+\ell_i, \operatorname{supp}(\nu)) \not\to 0$ for $i \in \{1, \ldots, k\}$ is violated
- and one may have some (order $O(1)$ in this setting) the eigenvalues of $\hat{\mathbf{C}}$ "jumping" out of the limiting support $\operatorname{supp}(\mu)$
- note for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so with its eigenvalues connected to those of $\mathbf{P}$

Low-rank perturbation from the Marčenko-Pastur law

Consider $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$, $\mathbf{z}_i \in \mathbb{R}^p$ with standard i.i.d. entries and

$$\mathbf{C} = \mathbf{I}_p + \mathbf{P}, \quad \mathbf{P} = \sum_{i=1}^{k} \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$$

with k and $\ell_1 \geq \ldots \geq \ell_k > 0$ fixed with respect to n, p.

» note that here $\nu \equiv \lim_{p\to\infty} \mu_{\mathbf{C}} = \lim_{p\to\infty} \frac{p-k}{p} \delta_1 + \frac{1}{p} \sum_{i=1}^k \delta_{1+\ell_i} = \delta_1$

» so, while $\mathbf{C} \neq \mathbf{I}_p$, the limiting μ still follows the Marčenko-Pastur law

» however, we do not have "no eigenvalue outside the support," since the condition $\operatorname{dist}(1+\ell_i, \operatorname{supp}(\nu)) \not\to 0$ for $i \in \{1, \ldots, k\}$ is violated

» and one may have some (order $O(1)$ in this setting) the eigenvalues of $\hat{\mathbf{C}}$ "jumping" out of the limiting support $\operatorname{supp}(\mu)$

» note for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so with its eigenvalues connected to those of $\mathbf{P}$

Low-rank perturbation from the Marčenko-Pastur law

Consider $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}}\mathbf{z}_i$, $\mathbf{z}_i \in \mathbb{R}^p$ with standard i.i.d. entries and

$$\mathbf{C} = \mathbf{I}_p + \mathbf{P}, \quad \mathbf{P} = \sum_{i=1}^{k} \ell_i \mathbf{u}_i \mathbf{u}_i^\top$$

with k and $\ell_1 \geq \ldots \geq \ell_k > 0$ fixed with respect to n, p.

- » note that here $\nu \equiv \lim_{p\to\infty} \mu_{\mathbf{C}} = \lim_{p\to\infty} \frac{p-k}{p}\delta_1 + \frac{1}{p}\sum_{i=1}^k \delta_{1+\ell_i} = \delta_1$
- » so, while $\mathbf{C} \neq \mathbf{I}_p$, the limiting μ still follows the Marčenko-Pastur law
- » however, we do not have "no eigenvalue outside the support," since the condition $\operatorname{dist}(1+\ell_i, \operatorname{supp}(\nu)) \not\to 0$ for $i \in \{1, \ldots, k\}$ is violated
- » and one may have some (order $O(1)$ in this setting) the eigenvalues of $\hat{\mathbf{C}}$ "jumping" out of the limiting support $\operatorname{supp}(\mu)$
- » note for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so with its eigenvalues connected to those of $\mathbf{P}$

Low-rank perturbation from the Marčenko-Pastur law

Consider $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$, $\mathbf{z}_i \in \mathbb{R}^p$ with standard i.i.d. entries and

$$\mathbf{C} = \mathbf{I}_p + \mathbf{P}, \quad \mathbf{P} = \sum_{i=1}^{k} \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$$

with k and $\ell_1 \geq \ldots \geq \ell_k > 0$ fixed with respect to n, p.

- note that here $\nu \equiv \lim_{p\to\infty} \mu_{\mathbf{C}} = \lim_{p\to\infty} \frac{p-k}{p}\delta_1 + \frac{1}{p}\sum_{i=1}^{k} \delta_{1+\ell_i} = \delta_1$
- so, while $\mathbf{C} \neq \mathbf{I}_p$, the limiting μ still follows the Marčenko-Pastur law
- however, we do not have "no eigenvalue outside the support," since the condition $\operatorname{dist}(1+\ell_i, \operatorname{supp}(\nu)) \not\to 0$ for $i \in \{1, \ldots, k\}$ is violated
- and one may have some (order $O(1)$ in this setting) the eigenvalues of $\hat{\mathbf{C}}$ "jumping" out of the limiting support $\operatorname{supp}(\mu)$
- note for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so with its eigenvalues connected to those of $\mathbf{P}$

Spiked eigenvalues and a phase transition

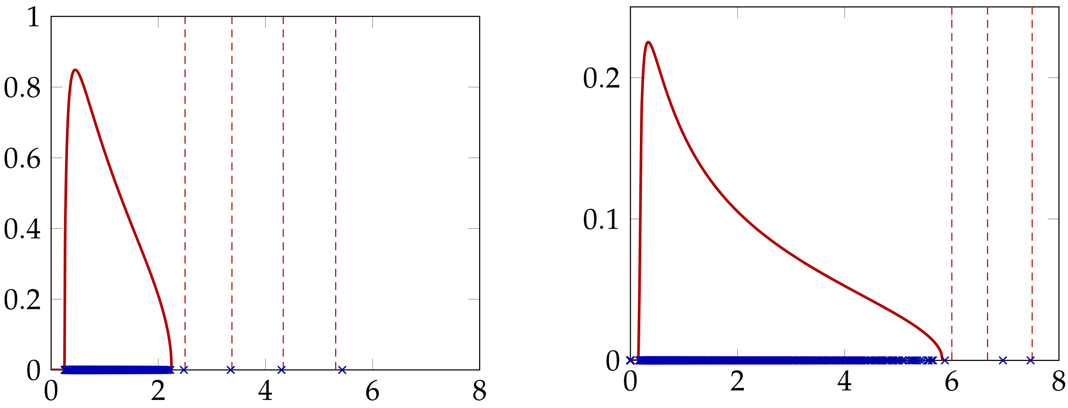
Here, depending on the values of ℓ_i and the ratio $c = \lim p/n$, the i-th largest eigenvalue $\hat{\lambda}_i$ of $\hat{\mathbf{C}}$ may indeed *isolate* from $\operatorname{supp}(\mu)$, due to [2].

For SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{C}^{\frac{1}{2}}$ with i.i.d. $\mathbb{E}[\mathbf{Z}_{ij}^4] < \infty$, let $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with $\mathbf{P} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^{\mathsf{T}}$ its spectral decomposition, where k and $\ell_1 \geq \ldots \geq \ell_k > 0$ are fixed with respect to n, p. Then, denoting $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p$ the eigenvalues of $\hat{\mathbf{C}}$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$,

$$\hat{\lambda}_i \xrightarrow{a.s.} \begin{cases} \lambda_i = 1 + \ell_i + c\frac{1+\ell_i}{\ell_i} > (1+\sqrt{c})^2 & , \ \ell_i > \sqrt{c} \\ (1+\sqrt{c})^2 & , \ \ell_i \leq \sqrt{c}. \end{cases}$$

Spiked eigenvalues

Jinho Baik and Jack W. Silverstein. “Eigenvalues of large sample covariance matrices of spiked population models”. In: *Journal of Multivariate Analysis* 97.6 (2006), 1382–1408. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2005.08.003



(a) $p/n = 1/4$

(b) $p/n = 2$

Figure: Eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ (**blue** crosses), the Marčenko-Pastur law (**red** solid line), and asymptotic spike locations (**red** dashed line), for $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$, $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with $\mu_{\mathbf{P}} = \frac{p-4}{p}\delta_0 + \frac{1}{p}(\delta_1 + \delta_2 + \delta_3 + \delta_4)$, for $p = 1024$ and different values of n.

Proof

- » solve the determinant equation $\det(\hat{\mathbf{C}} - \hat{\lambda}\mathbf{I}_p) = 0$ to find "isolated" eigenvalue $\hat{\lambda} \in \mathbb{R}$
- » use Sylvester's identity, $\det\left(\mathbf{AB} - \mathbf{I}_p\right) = \det\left(\mathbf{BA} - \mathbf{I}_k\right)$, to turn the p-dimensional equation into a k-dimensional one
- » solve the small-dimensional equation with the deterministic equivalent result

We write, with $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$,

$$0 = \det\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p\right) = \det\left(\frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} - \hat{\lambda}\mathbf{I}_p\right)$$
$$= \det\left(\mathbf{I}_p + \mathbf{P}\right)\det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\right) = \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\right),$$

since $\det(\mathbf{I}_p + \mathbf{P}) \neq 0$. Note from the resolvent identity ($\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}$)

$$(\mathbf{I}_p + \mathbf{P})^{-1} = \mathbf{I}_p - (\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}.$$

Proof

- solve the determinant equation $\det(\hat{\mathbf{C}} - \hat{\lambda}\mathbf{I}_p) = 0$ to find "isolated" eigenvalue $\hat{\lambda} \in \mathbb{R}$
- use Sylvester's identity, $\det\left(\mathbf{AB} - \mathbf{I}_p\right) = \det\left(\mathbf{BA} - \mathbf{I}_k\right)$, to turn the p-dimensional equation into a k-dimensional one
- solve the small-dimensional equation with the deterministic equivalent result

We write, with $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$,

$$0 = \det\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p\right) = \det\left(\frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} - \hat{\lambda}\mathbf{I}_p\right)$$
$$= \det\left(\mathbf{I}_p + \mathbf{P}\right)\det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\right) = \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\right),$$

since $\det(\mathbf{I}_p + \mathbf{P}) \neq 0$. Note from the resolvent identity $(\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1})$

$$(\mathbf{I}_p + \mathbf{P})^{-1} = \mathbf{I}_p - (\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}.$$

Proof

- solve the determinant equation $\det(\hat{\mathbf{C}} - \hat{\lambda}\mathbf{I}_p) = 0$ to find "isolated" eigenvalue $\hat{\lambda} \in \mathbb{R}$
- use Sylvester's identity, $\det\left(\mathbf{AB} - \mathbf{I}_p\right) = \det\left(\mathbf{BA} - \mathbf{I}_k\right)$, to turn the p-dimensional equation into a k-dimensional one
- solve the small-dimensional equation with the deterministic equivalent result

We write, with $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$,

$$0 = \det\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p\right) = \det\left(\frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} - \hat{\lambda}\mathbf{I}_p\right)$$

$$= \det\left(\mathbf{I}_p + \mathbf{P}\right)\det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\right) = \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\right),$$

since $\det(\mathbf{I}_p + \mathbf{P}) \neq 0$. Note from the resolvent identity $(\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1})$

$$(\mathbf{I}_p + \mathbf{P})^{-1} = \mathbf{I}_p - (\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}.$$

Proof (continue)

» We can then isolate the resolvent of the "whitened" model $\mathbf{Q}(\hat{\lambda}) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p)^{-1}$ and

$$0 = \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p + \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right) = \det \mathbf{Q}^{-1}(\hat{\lambda}) \cdot \det\left(\mathbf{I}_p + \hat{\lambda}\mathbf{Q}(\hat{\lambda})(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right).$$

» Using the "no eigenvalue of the support" result and the assumption $\mathbb{E}[\mathbf{Z}_{ij}^4] < \infty$, we are looking for isolated spiked eigenvalues such that $\hat{\lambda} > (1+\sqrt{c})^2$, so $\det \mathbf{Q}^{-1}(\hat{\lambda}) \neq 0$ with probability one as $n, p \to \infty$.

» Consider spectral decomposition $\mathbf{P} = \mathbf{U}\mathbf{L}\mathbf{U}^\mathsf{T} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$, then $(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P} = \sum_{i=1}^k \frac{\ell_i}{1+\ell_i}\mathbf{u}_i\mathbf{u}_i^\mathsf{T} = \mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\mathbf{U}^\mathsf{T}$ is also of rank k.

» With Sylvester's identity, we get $0 = \det\left(\mathbf{I}_k + \hat{\lambda}\mathbf{U}^\mathsf{T}\mathbf{Q}(\hat{\lambda})\mathbf{U} \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right)$.

Proof (continue)

» We can then isolate the resolvent of the "whitened" model $\mathbf{Q}(\hat{\lambda}) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p)^{-1}$ and

$$0 = \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p + \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right) = \det \mathbf{Q}^{-1}(\hat{\lambda}) \cdot \det\left(\mathbf{I}_p + \hat{\lambda}\mathbf{Q}(\hat{\lambda})(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right).$$

» Using the "no eigenvalue of the support" result and the assumption $\mathbb{E}[\mathbf{Z}_{ij}^4] < \infty$, we are looking for isolated spiked eigenvalues such that $\hat{\lambda} > (1+\sqrt{c})^2$, so $\det \mathbf{Q}^{-1}(\hat{\lambda}) \neq 0$ with probability one as $n, p \to \infty$.

» Consider spectral decomposition $\mathbf{P} = \mathbf{U}\mathbf{L}\mathbf{U}^\mathsf{T} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$, then $(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P} = \sum_{i=1}^k \frac{\ell_i}{1+\ell_i}\mathbf{u}_i\mathbf{u}_i^\mathsf{T} = \mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\mathbf{U}^\mathsf{T}$ is also of rank k.

» With Sylvester's identity, we get $0 = \det\left(\mathbf{I}_k + \hat{\lambda}\mathbf{U}^\mathsf{T}\mathbf{Q}(\hat{\lambda})\mathbf{U} \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right)$.

Proof (continue)

» We can then isolate the resolvent of the "whitened" model $\mathbf{Q}(\hat{\lambda}) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p)^{-1}$ and

$$0 = \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p + \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right) = \det \mathbf{Q}^{-1}(\hat{\lambda}) \cdot \det\left(\mathbf{I}_p + \hat{\lambda}\mathbf{Q}(\hat{\lambda})(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right).$$

» Using the "no eigenvalue of the support" result and the assumption $\mathbb{E}[\mathbf{Z}_{ij}^4] < \infty$, we are looking for isolated spiked eigenvalues such that $\hat{\lambda} > (1+\sqrt{c})^2$, so $\det \mathbf{Q}^{-1}(\hat{\lambda}) \neq 0$ with probability one as $n, p \to \infty$.

» Consider spectral decomposition $\mathbf{P} = \mathbf{U}\mathbf{L}\mathbf{U}^\mathsf{T} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$, then $(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P} = \sum_{i=1}^k \frac{\ell_i}{1+\ell_i}\mathbf{u}_i\mathbf{u}_i^\mathsf{T} = \mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\mathbf{U}^\mathsf{T}$ is also of rank k.

» With Sylvester's identity, we get $0 = \det\left(\mathbf{I}_k + \hat{\lambda}\mathbf{U}^\mathsf{T}\mathbf{Q}(\hat{\lambda})\mathbf{U} \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right)$.

Proof (continue)

- We can then isolate the resolvent of the "whitened" model $\mathbf{Q}(\hat{\lambda}) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p)^{-1}$ and

$$0 = \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - \hat{\lambda}\mathbf{I}_p + \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right) = \det \mathbf{Q}^{-1}(\hat{\lambda}) \cdot \det\left(\mathbf{I}_p + \hat{\lambda}\mathbf{Q}(\hat{\lambda})(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right).$$

- Using the "no eigenvalue of the support" result and the assumption $\mathbb{E}[\mathbf{Z}_{ij}^4] < \infty$, we are looking for isolated spiked eigenvalues such that $\hat{\lambda} > (1+\sqrt{c})^2$, so $\det \mathbf{Q}^{-1}(\hat{\lambda}) \neq 0$ with probability one as $n, p \to \infty$.
- Consider spectral decomposition $\mathbf{P} = \mathbf{U}\mathbf{L}\mathbf{U}^\mathsf{T} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$, then $(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P} = \sum_{i=1}^k \frac{\ell_i}{1+\ell_i}\mathbf{u}_i\mathbf{u}_i^\mathsf{T} = \mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\mathbf{U}^\mathsf{T}$ is also of rank k.
- With Sylvester's identity, we get $0 = \det\left(\mathbf{I}_k + \hat{\lambda}\mathbf{U}^\mathsf{T}\mathbf{Q}(\hat{\lambda})\mathbf{U} \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right)$.

Proof (continue)

For $\mathbf{Z} \in \mathbb{R}^{p\times n}$ with independent zero mean and unit variance random variables and $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$, as $n, p \to \infty$ with $p/n \to (0,\infty)$, we have $\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$, with $m(z)$ the unique ST solution to $zcm^2(z) - (1-c-z)m(z) + 1 = 0$.

Deterministic equivalent result for SCM

» Looking for isolated spikes $\hat\lambda$ satisfying $0 = \det\left(\mathbf{I}_k + \hat\lambda\mathbf{U}^\mathsf{T}\mathbf{Q}(\hat\lambda)\mathbf{U}\cdot(\mathbf{I}_k+\mathbf{L})^{-1}\mathbf{L}\right)$

» With the deterministic equivalent result $\mathbf{Q}(\hat\lambda) \leftrightarrow m(\hat\lambda)\mathbf{I}_p$, leads to

$$0 = \det\left(\mathbf{I}_k + \hat\lambda m(\hat\lambda)\cdot(\mathbf{I}_k+\mathbf{L})^{-1}\mathbf{L}\right) = \prod_{i=1}^k\left(1+\hat\lambda m(\hat\lambda)\frac{\ell_i}{1+\ell_i}\right). \quad (1)$$

» If such $\hat\lambda$ exists, must satisfy $\boxed{\hat\lambda m(\hat\lambda) = -\frac{1+\ell_i}{\ell_i} + o(1)}$, for some $i \in \{1,\ldots,k\}$.

Proof (continue)

For $\mathbf{Z} \in \mathbb{R}^{p\times n}$ with independent zero mean and unit variance random variables and $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$, as $n, p \to \infty$ with $p/n \to (0, \infty)$, we have $\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$, with $m(z)$ the unique ST solution to $zcm^2(z) - (1-c-z)m(z) + 1 = 0$.

Deterministic equivalent result for SCM

- Looking for isolated spikes $\hat\lambda$ satisfying $0 = \det\left(\mathbf{I}_k + \hat\lambda \mathbf{U}^\mathsf{T}\mathbf{Q}(\hat\lambda)\mathbf{U} \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right)$
- With the deterministic equivalent result $\mathbf{Q}(\hat\lambda) \leftrightarrow m(\hat\lambda)\mathbf{I}_p$, leads to

$$0 = \det\left(\mathbf{I}_k + \hat\lambda m(\hat\lambda) \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right) = \prod_{i=1}^{k}\left(1 + \hat\lambda m(\hat\lambda)\frac{\ell_i}{1+\ell_i}\right). \tag{1}$$

- If such $\hat\lambda$ exists, must satisfy $\boxed{\hat\lambda m(\hat\lambda) = -\frac{1+\ell_i}{\ell_i} + o(1)}$, for some $i \in \{1, \ldots, k\}$.

Proof (continue)

For $\mathbf{Z} \in \mathbb{R}^{p\times n}$ with independent zero mean and unit variance random variables and $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$, as $n, p \to \infty$ with $p/n \to (0, \infty)$, we have $\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$, with $m(z)$ the unique ST solution to $zcm^2(z) - (1 - c - z)m(z) + 1 = 0$.

Deterministic equivalent result for SCM

» Looking for isolated spikes $\hat{\lambda}$ satisfying $0 = \det\left(\mathbf{I}_k + \hat{\lambda}\mathbf{U}^\mathsf{T}\mathbf{Q}(\hat{\lambda})\mathbf{U} \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right)$

» With the deterministic equivalent result $\mathbf{Q}(\hat{\lambda}) \leftrightarrow m(\hat{\lambda})\mathbf{I}_p$, leads to

$$0 = \det\left(\mathbf{I}_k + \hat{\lambda}m(\hat{\lambda}) \cdot (\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\right) = \prod_{i=1}^{k}\left(1 + \hat{\lambda}m(\hat{\lambda})\frac{\ell_i}{1+\ell_i}\right). \tag{1}$$

» If such $\hat{\lambda}$ exists, must satisfy $\boxed{\hat{\lambda}m(\hat{\lambda}) = -\frac{1+\ell_i}{\ell_i} + o(1)}$, for some $i \in \{1, \ldots, k\}$.

Proof (continue)

$$\boxed{\hat{\lambda} m(\hat{\lambda}) = -\frac{1+\ell_i}{\ell_i} + o(1)} \quad (2)$$

- For such a solution $\hat{\lambda}$ to exist, study the behavior of $xm(x) = \int \frac{x}{t-x}\mu(dt)$ which is increasing on its domain of definition with $xm(x) \to -1$ as $x \to \infty$.
- Using the Marčenko-Pastur equation $zcm^2(z) - (1-c-z)m(z) + 1 = 0 \Leftrightarrow zm(z) = -1 + \frac{1}{1-z-c\cdot zm(z)}$, so that $\lim_{x \downarrow (1+\sqrt{c})^2} = -\frac{1+\sqrt{c}}{\sqrt{c}}$.
- so the solution $\hat{\lambda}$ exists **if and only if** the corresponding $\ell_i > \sqrt{c}$, and

$$\hat{\lambda} \to \boxed{\lambda_i = 1 + \ell_i + c\frac{1+\ell_i}{\ell_i}}. \quad (3)$$

Proof (continue)

$$\boxed{\hat{\lambda} m(\hat{\lambda}) = -\frac{1+\ell_i}{\ell_i} + o(1)} \tag{2}$$

» For such a solution $\hat{\lambda}$ to exist, study the behavior of $xm(x) = \int \frac{x}{t-x}\mu(dt)$ which is increasing on its domain of definition with $xm(x) \to -1$ as $x \to \infty$.

» Using the Marčenko-Pastur equation $zcm^2(z) - (1-c-z)m(z) + 1 = 0 \Leftrightarrow zm(z) = -1 + \frac{1}{1-z-c\cdot zm(z)}$, so that $\lim_{x \downarrow (1+\sqrt{c})^2} = -\frac{1+\sqrt{c}}{\sqrt{c}}$.

» so the solution $\hat{\lambda}$ exists **if and only if** the corresponding $\ell_i > \sqrt{c}$, and

$$\hat{\lambda} \to \boxed{\lambda_i = 1 + \ell_i + c\frac{1+\ell_i}{\ell_i}}. \tag{3}$$

Proof (continue)

$$\boxed{\hat{\lambda} m(\hat{\lambda}) = -\frac{1+\ell_i}{\ell_i} + o(1)} \tag{2}$$

- For such a solution $\hat{\lambda}$ to exist, study the behavior of $xm(x) = \int \frac{x}{t-x}\mu(dt)$ which is increasing on its domain of definition with $xm(x) \to -1$ as $x \to \infty$.
- Using the Marčenko-Pastur equation $zcm^2(z) - (1-c-z)m(z) + 1 = 0 \Leftrightarrow zm(z) = -1 + \frac{1}{1-z-c\cdot zm(z)}$, so that $\lim_{x \downarrow (1+\sqrt{c})^2} = -\frac{1+\sqrt{c}}{\sqrt{c}}$.
- so the solution $\hat{\lambda}$ exists **if and only if** the corresponding $\ell_i > \sqrt{c}$, and

$$\hat{\lambda} \to \boxed{\lambda_i = 1 + \ell_i + c\frac{1+\ell_i}{\ell_i}}. \tag{3}$$

Proof (continue)

$$\boxed{\hat{\lambda} m(\hat{\lambda}) = -\frac{1+\ell_i}{\ell_i} + o(1)} \tag{2}$$

- For such a solution $\hat{\lambda}$ to exist, study the behavior of $xm(x) = \int \frac{x}{t-x}\mu(dt)$ which is increasing on its domain of definition with $xm(x) \to -1$ as $x \to \infty$.
- Using the Marčenko-Pastur equation $zcm^2(z) - (1-c-z)m(z) + 1 = 0 \Leftrightarrow zm(z) = -1 + \frac{1}{1-z-c\cdot zm(z)}$, so that $\lim_{x\downarrow(1+\sqrt{c})^2} = -\frac{1+\sqrt{c}}{\sqrt{c}}$.
- so the solution $\hat{\lambda}$ exists **if and only if** the corresponding $\ell_i > \sqrt{c}$, and

$$\hat{\lambda} \to \boxed{\lambda_i = 1 + \ell_i + c\frac{1+\ell_i}{\ell_i}}. \tag{3}$$

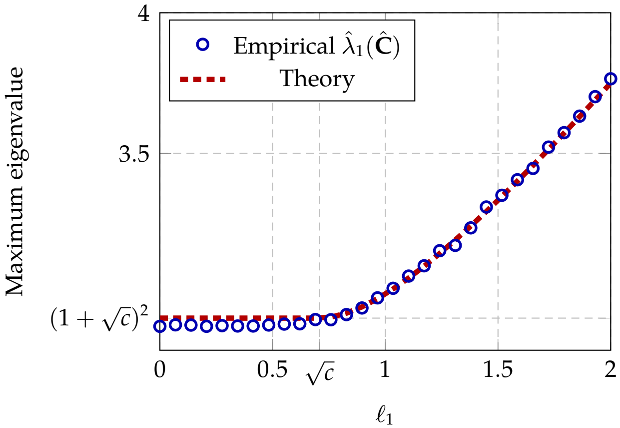


Figure: Phase transition behavior of the largest eigenvalue $\hat{\lambda}_1(\hat{\mathbf{C}})$ of $\hat{\mathbf{C}} = \frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}$ as a function of $\ell_1 = \|\mathbf{P}\|$ with rank one $\mathbf{P}$, for $p = 512$ and $n = 1\,024$. Empirical results obtained by averaging over 50 independent runs.

Outline

Spiked eigenvectors of SCM

We would also like to characterize the behavior of the **spiked eigenvectors**:

- it makes sense to believe that for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so its eigenvectors should "close to" those of $\mathbf{C}$ in some way and for sufficient large n/p
- formally, how the **top** eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_k$ of $\hat{\mathbf{C}}$ close to those $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ of $\mathbf{P}$
- some type of phase transition behavior is (again) expected.

In the absence of low-rank perturbation $\mathbf{P}$ and Gaussian $\mathbf{Z}$, it is known that the eigenvectors of the resulting Wishart matrix $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} \in \mathbb{R}^{p \times p}$ are *uniformly distributed on the unit sphere* $\mathbb{S}^{p-1}$ (also know as the p-dimensional Haar measure), which is close to, **for** p **large**, random vector with i.i.d. Gaussian entries.

Absence of $\mathbf{P}$

Spiked eigenvectors of SCM

We would also like to characterize the behavior of the **spiked eigenvectors**:

- it makes sense to believe that for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so its eigenvectors should "close to" those of $\mathbf{C}$ in some way and for sufficient large n/p
- formally, how the **top** eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_k$ of $\hat{\mathbf{C}}$ close to those $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ of $\mathbf{P}$
- some type of phase transition behavior is (again) expected.

In the absence of low-rank perturbation $\mathbf{P}$ and Gaussian $\mathbf{Z}$, it is known that the eigenvectors of the resulting Wishart matrix $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} \in \mathbb{R}^{p \times p}$ are *uniformly distributed on the unit sphere* $\mathbb{S}^{p-1}$ (also know as the p-dimensional Haar measure), which is close to, **for** p **large**, random vector with i.i.d. Gaussian entries.

Absence of $\mathbf{P}$

Spiked eigenvectors of SCM

We would also like to characterize the behavior of the **spiked eigenvectors**:

- it makes sense to believe that for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so its eigenvectors should "close to" those of $\mathbf{C}$ in some way and for sufficient large n/p
- formally, how the **top** eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_k$ of $\hat{\mathbf{C}}$ close to those $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ of $\mathbf{P}$
- some type of phase transition behavior is (again) expected.

In the absence of low-rank perturbation $\mathbf{P}$ and Gaussian $\mathbf{Z}$, it is known that the eigenvectors of the resulting Wishart matrix $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} \in \mathbb{R}^{p\times p}$ are *uniformly distributed on the unit sphere* $\mathbb{S}^{p-1}$ (also know as the p-dimensional Haar measure), which is close to, **for** p **large**, random vector with i.i.d. Gaussian entries.

Absence of $\mathbf{P}$

Spiked eigenvectors of SCM

We would also like to characterize the behavior of the **spiked eigenvectors**:

- it makes sense to believe that for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so its eigenvectors should "close to" those of $\mathbf{C}$ in some way and for sufficient large n/p
- formally, how the **top** eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_k$ of $\hat{\mathbf{C}}$ close to those $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ of $\mathbf{P}$
- some type of phase transition behavior is (again) expected.

In the absence of low-rank perturbation $\mathbf{P}$ and Gaussian $\mathbf{Z}$, it is known that the eigenvectors of the resulting Wishart matrix $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} \in \mathbb{R}^{p \times p}$ are *uniformly distributed on the unit sphere* $\mathbb{S}^{p-1}$ (also know as the p-dimensional Haar measure), which is close to, **for** p **large**, random vector with i.i.d. Gaussian entries.

Absence of $\mathbf{P}$

Spiked eigenvectors of SCM

We would also like to characterize the behavior of the **spiked eigenvectors**:

- it makes sense to believe that for $n \gg p$, $\hat{\mathbf{C}} \simeq \mathbf{C} = \mathbf{I}_p + \mathbf{P}$, so its eigenvectors should "close to" those of $\mathbf{C}$ in some way and for sufficient large n/p
- formally, how the **top** eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_k$ of $\hat{\mathbf{C}}$ close to those $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ of $\mathbf{P}$
- some type of phase transition behavior is (again) expected.

In the absence of low-rank perturbation $\mathbf{P}$ and Gaussian $\mathbf{Z}$, it is known that the eigenvectors of the resulting Wishart matrix $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} \in \mathbb{R}^{p \times p}$ are *uniformly distributed on the unit sphere* $\mathbb{S}^{p-1}$ (also know as the p-dimensional Haar measure), which is close to, **for** p **large**, random vector with i.i.d. Gaussian entries.

Absence of $\mathbf{P}$

Spiked eigenvector alignment

Let $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_k$ be the eigenvectors associated with the largest k eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_k$ of $\hat{\mathbf{C}}$. Further assume that $\ell_1 > \ldots > \ell_k > 0$ are all distinct. Then, for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$ unit norm deterministic vectors

$$\mathbf{a}^\top \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \mathbf{b} - \mathbf{a}^\top \mathbf{u}_i \mathbf{u}_i^\top \mathbf{b} \cdot \frac{1 - c\ell_i^{-2}}{1 + c\ell_i^{-1}} \cdot 1_{\ell_i > \sqrt{c}} \xrightarrow{a.s.} 0. \tag{4}$$

In particular, with $\mathbf{a} = \mathbf{b} = \mathbf{u}_i$ we obtain

$$(\mathbf{u}_i^\top \hat{\mathbf{u}}_i)^2 \xrightarrow{a.s.} \zeta_i \equiv \frac{1 - c\ell_i^{-2}}{1 + c\ell_i^{-1}} \cdot 1_{\ell_i > \sqrt{c}}. \tag{5}$$

Spiked eigenvector alignment

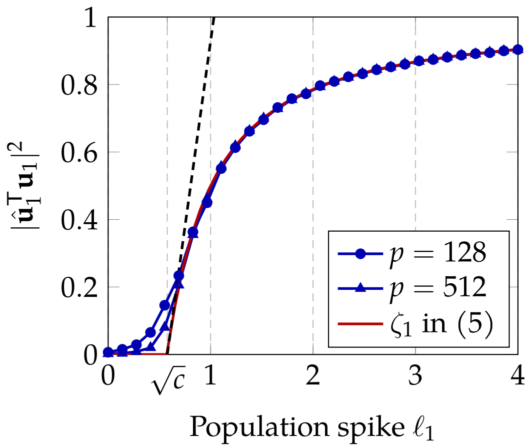


Figure: Empirical versus limiting $|\hat{\mathbf{u}}_1^\mathsf{T}\mathbf{u}_1|^2$ for $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$, $\mathbf{C} = \mathbf{I}_p + \ell_1\mathbf{u}_1\mathbf{u}_1^\mathsf{T}$ and standard Gaussian $\mathbf{Z}$, $p/n = 1/3$, for different values of ℓ_1. Results obtained by averaging over 200 runs. In **black** dashed line the local behavior around $\sqrt{c}$.

Proof

» First write that, for all large n, p almost surely and $\ell_i > \sqrt{c}$,

$$\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^\mathsf{T}\mathbf{b} = -\frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} \mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b}\, dz,$$

for Γ_{λ_i} a small contour enclosing only the almost sure limit $\lambda_i = 1 + \ell_i + c\frac{1+\ell_i}{\ell_i}$ of the eigenvalue $\hat{\lambda}_i$ of $\hat{\mathbf{C}}$ that we just determined.

$$\mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b} = \mathbf{a}^\mathsf{T}\left(\frac{1}{n}(\mathbf{I}_p+\mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{\frac{1}{2}} - z\mathbf{I}_p\right)^{-1}\mathbf{b}$$

$$= \mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p + z(\mathbf{I}_p+\mathbf{P})^{-1}\mathbf{P}\right)^{-1}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b}$$

Proof (continue)

Denote $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$, it follows from $(\mathbf{I}_p + \mathbf{P})^{-1} = \mathbf{I}_p - (\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}$ and the spectral decomposition $(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P} = \mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\mathbf{U}^\mathsf{T}$ for $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_k] \in \mathbb{R}^{p\times k}$ and $\mathbf{L} = \operatorname{diag}\{\ell_i\}_{i=1}^k$ that

$$
\begin{aligned}
&\mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b} \\
&= \mathbf{a}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{b} \\
&- z\mathbf{a}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)\mathbf{U}\left(\mathbf{I}_k + \mathbf{L}^{-1} + z\mathbf{U}^\mathsf{T}\mathbf{Q}(z)\mathbf{U}\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}(z)(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{b} \\
&= \mathbf{a}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{b} \\
&- z\mathbf{a}^\mathsf{T}(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)\mathbf{U}\left(\mathbf{L}^{-1} + (1 + zm(z))\mathbf{I}_k\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}(z)(\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}}\mathbf{b} + o(1),
\end{aligned}
$$

where we used Woodbury identity and $\mathbf{U}^\mathsf{T}\mathbf{Q}(z)\mathbf{U} = m(z)\mathbf{I}_k + o_{\|\cdot\|}(1)$.

Proof (continue)

Objective: $\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^\mathsf{T}\mathbf{b} = -\frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} \mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b}\,dz$, with $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$ and

$$\mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b} = \mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b}$$
$$- z\mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)\mathbf{U}\left(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b} + o(1),$$

- complex integration of first term vanishes (looking for spikes with well defined $\mathbf{Q}(z)$)
- complex integration of $\mathbf{Q}(z)$ on the contour Γ_{λ_i} only brings a non-trivial residue, due to the inverse $(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k)^{-1}$ which is singular at $z = \lambda_i$

$$\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{b} = \frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} zm^2(z)\mathbf{a}^\mathsf{T}\mathbf{U}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k)^{-1}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}\mathbf{U}^\mathsf{T}\mathbf{b}\,dz + o(1).$$

- can be evaluated by **residue calculus** at $z = \lambda_i$.

Proof (continue)

Objective: $\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^\mathsf{T}\mathbf{b} = -\frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} \mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b}\,dz$, with $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$ and

$$\mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b} = \mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b}$$
$$- z\mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)\mathbf{U}\left(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b} + o(1),$$

- complex integration of first term vanishes (looking for spikes with well defined $\mathbf{Q}(z)$)
- complex integration of $\mathbf{Q}(z)$ on the contour Γ_{λ_i} only brings a non-trivial residue, due to the inverse $(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k)^{-1}$ which is singular at $z = \lambda_i$

$$\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{b} = \frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} zm^2(z)\mathbf{a}^\mathsf{T}\mathbf{U}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k)^{-1}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}\mathbf{U}^\mathsf{T}\mathbf{b}\,dz + o(1).$$

- can be evaluated by **residue calculus** at $z = \lambda_i$.

Proof (continue)

Objective: $\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^\mathsf{T}\mathbf{b} = -\frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} \mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b}\,dz$, with $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$ and

$$\mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b} = \mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b}$$
$$- z\mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)\mathbf{U}\left(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b} + o(1),$$

» complex integration of first term vanishes (looking for spikes with well defined $\mathbf{Q}(z)$)

» complex integration of $\mathbf{Q}(z)$ on the contour Γ_{λ_i} only brings a non-trivial residue, due to the inverse $(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k)^{-1}$ which is singular at $z = \lambda_i$

$$\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{b} = \frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} zm^2(z)\mathbf{a}^\mathsf{T}\mathbf{U}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k)^{-1}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}\mathbf{U}^\mathsf{T}\mathbf{b}\,dz + o(1).$$

» can be evaluated by **residue calculus** at $z = \lambda_i$.

Proof (continue)

Objective: $\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^\mathsf{T}\mathbf{b} = -\frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} \mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b}\,dz$, with $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\mathsf{T} - z\mathbf{I}_p)^{-1}$ and

$$\mathbf{a}^\mathsf{T}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p\right)^{-1}\mathbf{b} = \mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b}$$
$$- z\mathbf{a}^\mathsf{T}(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{Q}(z)\mathbf{U}\left(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}(z)(\mathbf{I}_p+\mathbf{P})^{-\frac{1}{2}}\mathbf{b} + o(1),$$

- » complex integration of first term vanishes (looking for spikes with well defined $\mathbf{Q}(z)$)
- » complex integration of $\mathbf{Q}(z)$ on the contour Γ_{λ_i} only brings a non-trivial residue, due to the inverse $(\mathbf{L}^{-1} + (1+zm(z))\mathbf{I}_k)^{-1}$ which is singular at $z=\lambda_i$

$$\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{b} = \frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} zm^2(z)\mathbf{a}^\mathsf{T}\mathbf{U}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}(\mathbf{L}^{-1}+(1+zm(z))\mathbf{I}_k)^{-1}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}\mathbf{U}^\mathsf{T}\mathbf{b}\,dz + o(1).$$

- » can be evaluated by **residue calculus** at $z=\lambda_i$.

Proof (continue)

$$\boxed{\mathbf{a}^\top \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i \mathbf{b} = \frac{1}{2\pi\imath} \oint_{\Gamma_{\lambda_i}} z m^2(z) \mathbf{a}^\top \mathbf{U} (\mathbf{I}_k + \mathbf{L})^{-\frac{1}{2}} (\mathbf{L}^{-1} + (1 + z m(z)) \mathbf{I}_k)^{-1} (\mathbf{I}_k + \mathbf{L})^{-\frac{1}{2}} \mathbf{U}^\top \mathbf{b} \, dz + o(1)}.$$

» residue calculus:

$$\lim_{z \to \lambda_i} (z - \lambda_i)(\mathbf{L}^{-1} + (1 + z m(z)) \mathbf{I}_k)^{-1} = \frac{\mathbf{e}_i \mathbf{e}_i^\top}{m(\lambda_i) + \lambda_i m'(\lambda_i)}$$

with $\mathbf{e}_i \in \mathbb{R}^k$ canonical basis vector $[\mathbf{e}_i]_j = \delta_{ij}$.

» Using the Marčenko-Pastur equation $m(z) = \frac{1}{-z + \frac{1}{1 + c m(z)}}$, we get $m'(z) = \frac{m^2(z)}{1 - \frac{c m^2(z)}{(1 + c m(z))^2}}$, from which we have $m(\lambda_i) = -1/(\ell_i + c)$ and $m'(\lambda_i) = \ell_i^2 (\ell_i + c)^{-2} (\ell_i^2 - c)^{-1}$.

» We conclude that $\boxed{\mathbf{a}^\top \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \mathbf{b} = \mathbf{a}^\top \mathbf{u}_i \mathbf{u}_i^\top \mathbf{b} \cdot \frac{1 - c\ell_i^{-2}}{1 + c\ell_i^{-1}} + o(1)}$.

Proof (continue)

$$\boxed{\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{b} = \frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} zm^2(z)\mathbf{a}^\mathsf{T}\mathbf{U}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}(\mathbf{L}^{-1}+(1+zm(z))\mathbf{I}_k)^{-1}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}\mathbf{U}^\mathsf{T}\mathbf{b}\,dz + o(1)}.$$

» residue calculus:

$$\lim_{z\to\lambda_i}(z-\lambda_i)(\mathbf{L}^{-1}+(1+zm(z))\mathbf{I}_k)^{-1} = \frac{\mathbf{e}_i\mathbf{e}_i^\mathsf{T}}{m(\lambda_i)+\lambda_i m'(\lambda_i)}$$

with $\mathbf{e}_i \in \mathbb{R}^k$ canonical basis vector $[\mathbf{e}_i]_j = \delta_{ij}$.

» Using the Marčenko-Pastur equation $m(z) = \frac{1}{-z+\frac{1}{1+cm(z)}}$, we get $m'(z) = \frac{m^2(z)}{1-\frac{cm^2(z)}{(1+cm(z))^2}}$, from which we have $m(\lambda_i) = -1/(\ell_i+c)$ and $m'(\lambda_i) = \ell_i^2(\ell_i+c)^{-2}(\ell_i^2-c)^{-1}$.

» We conclude that $\boxed{\mathbf{a}^\mathsf{T}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^\mathsf{T}\mathbf{b} = \mathbf{a}^\mathsf{T}\mathbf{u}_i\mathbf{u}_i^\mathsf{T}\mathbf{b}\cdot\frac{1-c\ell_i^{-2}}{1+c\ell_i^{-1}} + o(1)}$.

Proof (continue)

$$\boxed{\mathbf{a}^{\mathsf{T}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{b} = \frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} zm^2(z)\mathbf{a}^{\mathsf{T}}\mathbf{U}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}(\mathbf{L}^{-1}+(1+zm(z))\mathbf{I}_k)^{-1}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}\mathbf{U}^{\mathsf{T}}\mathbf{b}\,dz + o(1)}.$$

» residue calculus:

$$\lim_{z\to\lambda_i}(z-\lambda_i)(\mathbf{L}^{-1}+(1+zm(z))\mathbf{I}_k)^{-1} = \frac{\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}}{m(\lambda_i)+\lambda_i m'(\lambda_i)}$$

with $\mathbf{e}_i \in \mathbb{R}^k$ canonical basis vector $[\mathbf{e}_i]_j = \delta_{ij}$.

» Using the Marčenko-Pastur equation $m(z) = \frac{1}{-z+\frac{1}{1+cm(z)}}$, we get $m'(z) = \frac{m^2(z)}{1-\frac{cm^2(z)}{(1+cm(z))^2}}$, from which we have $m(\lambda_i) = -1/(\ell_i + c)$ and $m'(\lambda_i) = \ell_i^2(\ell_i+c)^{-2}(\ell_i^2-c)^{-1}$.

» We conclude that $\boxed{\mathbf{a}^{\mathsf{T}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{T}}\mathbf{b} = \mathbf{a}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbf{b}\cdot\frac{1-c\ell_i^{-2}}{1+c\ell_i^{-1}} + o(1)}$.

Proof (continue)

$$\boxed{\mathbf{a}^{\mathsf{T}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{b} = \frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i}} zm^2(z)\mathbf{a}^{\mathsf{T}}\mathbf{U}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}(\mathbf{L}^{-1}+(1+zm(z))\mathbf{I}_k)^{-1}(\mathbf{I}_k+\mathbf{L})^{-\frac{1}{2}}\mathbf{U}^{\mathsf{T}}\mathbf{b}\,dz + o(1)}.$$

» residue calculus:

$$\lim_{z\to\lambda_i}(z-\lambda_i)(\mathbf{L}^{-1}+(1+zm(z))\mathbf{I}_k)^{-1} = \frac{\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}}{m(\lambda_i)+\lambda_i m'(\lambda_i)}$$

with $\mathbf{e}_i \in \mathbb{R}^k$ canonical basis vector $[\mathbf{e}_i]_j = \delta_{ij}$.

» Using the Marčenko-Pastur equation $m(z) = \frac{1}{-z+\frac{1}{1+cm(z)}}$, we get $m'(z) = \frac{m^2(z)}{1-\frac{cm^2(z)}{(1+cm(z))^2}}$, from which we have $m(\lambda_i) = -1/(\ell_i+c)$ and $m'(\lambda_i) = \ell_i^2(\ell_i+c)^{-2}(\ell_i^2-c)^{-1}$.

» We conclude that $\boxed{\mathbf{a}^{\mathsf{T}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{T}}\mathbf{b} = \mathbf{a}^{\mathsf{T}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}\mathbf{b}\cdot\frac{1-c\ell_i^{-2}}{1+c\ell_i^{-1}} + o(1)}$.

Outline

Limiting fluctuations

- we have seen that the (asymptotic) **location** of the largest eigenvalue $\hat{\lambda}_1(\hat{\mathbf{C}})$ establishes a phase transition behavior if the corresponding population $\ell_i > \sqrt{c}$
- so below the threshold $\hat{\lambda}_1(\hat{\mathbf{C}}) = (1+\sqrt{c})^2 + o(1)$ almost surely as $n, p \to \infty$
- we want to understand more on this $o(1)$ local behavior

Under the same setting, assume $0 \le \ell_k < \ldots < \ell_1 < \sqrt{c}$. Then,

$$n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}} \to \mathrm{TW}_1$$

in law, where TW_1 is the (real) Tracy-Widom distribution.

Fluctuation of the largest eigenvalue

- below phase transition: $\hat{\lambda}_1 = (1+\sqrt{c})^2 + n^{-\frac{2}{3}} T$ where T is a (scaled) Tracy-Widom RV

Limiting fluctuations

- we have seen that the (asymptotic) **location** of the largest eigenvalue $\hat{\lambda}_1(\hat{\mathbf{C}})$ establishes a phase transition behavior if the corresponding population $\ell_i > \sqrt{c}$
- so below the threshold $\hat{\lambda}_1(\hat{\mathbf{C}}) = (1+\sqrt{c})^2 + o(1)$ almost surely as $n, p \to \infty$
- we want to understand more on this $o(1)$ local behavior

Under the same setting, assume $0 \leq \ell_k < \ldots < \ell_1 < \sqrt{c}$. Then,

$$n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}} \to \mathrm{TW}_1$$

in law, where TW_1 is the (real) Tracy-Widom distribution.

Fluctuation of the largest eigenvalue

- below phase transition: $\hat{\lambda}_1 = (1+\sqrt{c})^2 + n^{-\frac{2}{3}} T$ where T is a (scaled) Tracy-Widom RV

Limiting fluctuations

- we have seen that the (asymptotic) **location** of the largest eigenvalue $\hat{\lambda}_1(\hat{\mathbf{C}})$ establishes a phase transition behavior if the corresponding population $\ell_i > \sqrt{c}$
- so below the threshold $\hat{\lambda}_1(\hat{\mathbf{C}}) = (1+\sqrt{c})^2 + o(1)$ almost surely as $n, p \to \infty$
- we want to understand more on this $o(1)$ local behavior

Under the same setting, assume $0 \leq \ell_k < \ldots < \ell_1 < \sqrt{c}$. Then,

$$n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}} \to \mathrm{TW}_1$$

in law, where TW_1 is the (real) Tracy-Widom distribution.

Fluctuation of the largest eigenvalue

- below phase transition: $\hat{\lambda}_1 = (1+\sqrt{c})^2 + n^{-\frac{2}{3}} T$ where T is a (scaled) Tracy-Widom RV

Limiting fluctuations

- we have seen that the (asymptotic) **location** of the largest eigenvalue $\hat{\lambda}_1(\hat{\mathbf{C}})$ establishes a phase transition behavior if the corresponding population $\ell_i > \sqrt{c}$
- so below the threshold $\hat{\lambda}_1(\hat{\mathbf{C}}) = (1+\sqrt{c})^2 + o(1)$ almost surely as $n, p \to \infty$
- we want to understand more on this $o(1)$ local behavior

Under the same setting, assume $0 \le \ell_k < \ldots < \ell_1 < \sqrt{c}$. Then,

$$n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}} \to \mathrm{TW}_1$$

in law, where TW_1 is the (real) Tracy-Widom distribution.

Fluctuation of the largest eigenvalue

- below phase transition: $\hat{\lambda}_1 = (1+\sqrt{c})^2 + n^{-\frac{2}{3}} T$ where T is a (scaled) Tracy-Widom RV

Limiting fluctuations

- we have seen that the (asymptotic) **location** of the largest eigenvalue $\hat{\lambda}_1(\hat{\mathbf{C}})$ establishes a phase transition behavior if the corresponding population $\ell_i > \sqrt{c}$
- so below the threshold $\hat{\lambda}_1(\hat{\mathbf{C}}) = (1+\sqrt{c})^2 + o(1)$ almost surely as $n, p \to \infty$
- we want to understand more on this $o(1)$ local behavior

Under the same setting, assume $0 \leq \ell_k < \ldots < \ell_1 < \sqrt{c}$. Then,

$$n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}} \to \mathrm{TW}_1$$

in law, where TW_1 is the (real) Tracy-Widom distribution.

Fluctuation of the largest eigenvalue

- below phase transition: $\hat{\lambda}_1 = (1+\sqrt{c})^2 + n^{-\frac{2}{3}} T$ where T is a (scaled) Tracy-Widom RV

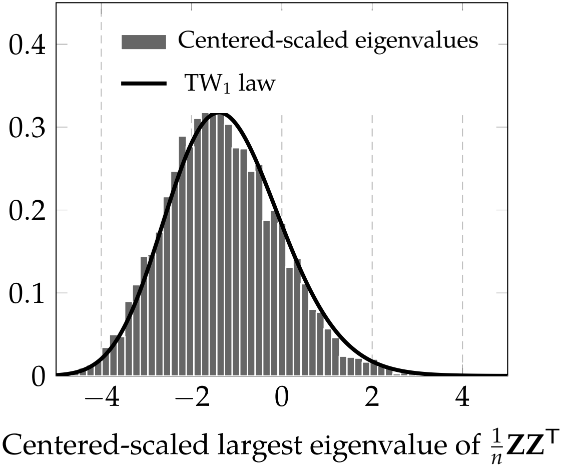


Figure: Empirical histogram of $n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}}$ for $p = 256$, $n = 512$ and standard Gaussian $\mathbf{Z}$, versus the real Tracy-Widom law TW_1. Histogram obtained over 5 000 independent runs.

Remarks on Tracy-Widom law

- somewhat surprising: limiting fluctuation of $\hat{\lambda}_1$ is not Gaussian but follow the Tracy-Widom distribution and of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$)
- rate related to the following observation:
  - Marčenko-Pastur law: $\mu(dx) = \frac{1}{2\pi c x}\sqrt{(x-E_-)^+(E_+-x)^+}\,dx, \quad E_\pm = (1\pm\sqrt{c})^2.$
  - so near the right edge E_+: $\mu(dx) \simeq_{x\uparrow E_+} \frac{c^{1/4}}{\pi c(1+\sqrt{c})^2}\sqrt{|E_+-x|}$
  - so a typical number of eigenvalues in a space of size ϵ near the edge is

$$\int_{(1+\sqrt{c})^2-\epsilon}^{(1+\sqrt{c})^2} \sqrt{(1+\sqrt{c})^2 - x}\,dx \propto \epsilon^{\frac{3}{2}} \tag{6}$$

  - to have $O(1)$ eigenvalues within $[E_+-\epsilon, E_+]$ needs $\epsilon = O(n^{-\frac{2}{3}})$ (this is in fact the "spacing" between eigenvalues, which is of order $O(n^{-1})$ away from the edge)
- **Question**: hard-edge setting with $c=1$, what happens?

Remarks on Tracy-Widom law

- somewhat surprising: limiting fluctuation of $\hat{\lambda}_1$ is not Gaussian but follow the Tracy-Widom distribution and of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$)
- rate related to the following observation:
  - Marčenko-Pastur law: $\mu(dx) = \frac{1}{2\pi c x}\sqrt{(x-E_-)^+(E_+-x)^+}\,dx, \quad E_\pm = (1\pm\sqrt{c})^2$.
  - so near the right edge E_+: $\mu(dx) \simeq_{x\uparrow E_+} \frac{c^{1/4}}{\pi c(1+\sqrt{c})^2}\sqrt{|E_+ - x|}$
  - so a typical number of eigenvalues in a space of size ϵ near the edge is

$$\int_{(1+\sqrt{c})^2-\epsilon}^{(1+\sqrt{c})^2} \sqrt{(1+\sqrt{c})^2 - x}\,dx \propto \epsilon^{\frac{3}{2}} \tag{6}$$

  - to have $O(1)$ eigenvalues within $[E_+ - \epsilon, E_+]$ needs $\epsilon = O(n^{-\frac{2}{3}})$ (this is in fact the "spacing" between eigenvalues, which is of order $O(n^{-1})$ away from the edge)
- **Question**: hard-edge setting with $c = 1$, what happens?

Remarks on Tracy-Widom law

- somewhat surprising: limiting fluctuation of $\hat{\lambda}_1$ is not Gaussian but follow the Tracy-Widom distribution and of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$)
- rate related to the following observation:
  - Marčenko-Pastur law: $\mu(dx) = \frac{1}{2\pi c x}\sqrt{(x-E_-)^+(E_+-x)^+}\,dx, \quad E_\pm = (1 \pm \sqrt{c})^2.$
  - so near the right edge E_+: $\mu(dx) \simeq_{x \uparrow E_+} \frac{c^{1/4}}{\pi c(1+\sqrt{c})^2}\sqrt{|E_+ - x|}$
  - so a typical number of eigenvalues in a space of size ϵ near the edge is

$$\int_{(1+\sqrt{c})^2-\epsilon}^{(1+\sqrt{c})^2} \sqrt{(1+\sqrt{c})^2 - x}\, dx \propto \epsilon^{\frac{3}{2}} \tag{6}$$

  - to have $O(1)$ eigenvalues within $[E_+ - \epsilon, E_+]$ needs $\epsilon = O(n^{-\frac{2}{3}})$ (this is in fact the "spacing" between eigenvalues, which is of order $O(n^{-1})$ away from the edge)
- **Question**: hard-edge setting with $c = 1$, what happens?

Remarks on Tracy-Widom law

- somewhat surprising: limiting fluctuation of $\hat{\lambda}_1$ is not Gaussian but follow the Tracy-Widom distribution and of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$)
- rate related to the following observation:
  - Marčenko-Pastur law: $\mu(dx) = \frac{1}{2\pi c x}\sqrt{(x-E_-)^+(E_+-x)^+}\,dx, \quad E_\pm = (1\pm\sqrt{c})^2$.
  - so near the right edge E_+: $\mu(dx) \simeq_{x\uparrow E_+} \frac{c^{1/4}}{\pi c(1+\sqrt{c})^2}\sqrt{|E_+-x|}$
  - so a typical number of eigenvalues in a space of size ϵ near the edge is

$$\int_{(1+\sqrt{c})^2-\epsilon}^{(1+\sqrt{c})^2} \sqrt{(1+\sqrt{c})^2 - x}\,dx \propto \epsilon^{\frac{3}{2}} \tag{6}$$

  - to have $O(1)$ eigenvalues within $[E_+ - \epsilon, E_+]$ needs $\epsilon = O(n^{-\frac{2}{3}})$ (this is in fact the "spacing" between eigenvalues, which is of order $O(n^{-1})$ away from the edge)
- **Question**: hard-edge setting with $c=1$, what happens?

Remarks on Tracy-Widom law

- somewhat surprising: limiting fluctuation of $\hat{\lambda}_1$ is not Gaussian but follow the Tracy-Widom distribution and of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$)
- rate related to the following observation:
  - Marčenko-Pastur law: $\mu(dx) = \frac{1}{2\pi c x}\sqrt{(x-E_-)^+(E_+-x)^+}\,dx, \quad E_\pm = (1\pm\sqrt{c})^2$.
  - so near the right edge E_+: $\mu(dx) \simeq_{x\uparrow E_+} \frac{c^{1/4}}{\pi c(1+\sqrt{c})^2}\sqrt{|E_+-x|}$
  - so a typical number of eigenvalues in a space of size ϵ near the edge is

$$\int_{(1+\sqrt{c})^2-\epsilon}^{(1+\sqrt{c})^2} \sqrt{(1+\sqrt{c})^2 - x}\,dx \propto \epsilon^{\frac{3}{2}} \tag{6}$$

  - to have $O(1)$ eigenvalues within $[E_+-\epsilon, E_+]$ needs $\epsilon = O(n^{-\frac{2}{3}})$ (this is in fact the "spacing" between eigenvalues, which is of order $O(n^{-1})$ away from the edge)
- **Question**: hard-edge setting with $c=1$, what happens?

Remarks on Tracy-Widom law

- somewhat surprising: limiting fluctuation of $\hat{\lambda}_1$ is not Gaussian but follow the Tracy-Widom distribution and of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$)
- rate related to the following observation:
  - Marčenko-Pastur law: $\mu(dx) = \frac{1}{2\pi c x}\sqrt{(x-E_-)^+(E_+-x)^+}\,dx, \quad E_\pm = (1\pm\sqrt{c})^2.$
  - so near the right edge E_+: $\mu(dx) \simeq_{x\uparrow E_+} \frac{c^{1/4}}{\pi c(1+\sqrt{c})^2}\sqrt{|E_+-x|}$
  - so a typical number of eigenvalues in a space of size ϵ near the edge is

$$\int_{(1+\sqrt{c})^2-\epsilon}^{(1+\sqrt{c})^2} \sqrt{(1+\sqrt{c})^2 - x}\,dx \propto \epsilon^{\frac{3}{2}} \tag{6}$$

  - to have $O(1)$ eigenvalues within $[E_+-\epsilon, E_+]$ needs $\epsilon = O(n^{-\frac{2}{3}})$ (this is in fact the "spacing" between eigenvalues, which is of order $O(n^{-1})$ away from the edge)
- **Question**: hard-edge setting with $c = 1$, what happens?

Remarks on Tracy-Widom law

- somewhat surprising: limiting fluctuation of $\hat{\lambda}_1$ is not Gaussian but follow the Tracy-Widom distribution and of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$)
- rate related to the following observation:
  - Marčenko-Pastur law: $\mu(dx) = \frac{1}{2\pi c x}\sqrt{(x-E_-)^+(E_+-x)^+}\,dx, \quad E_\pm = (1 \pm \sqrt{c})^2$.
  - so near the right edge E_+: $\mu(dx) \simeq_{x \uparrow E_+} \frac{c^{1/4}}{\pi c (1+\sqrt{c})^2}\sqrt{|E_+ - x|}$
  - so a typical number of eigenvalues in a space of size ϵ near the edge is

$$\int_{(1+\sqrt{c})^2-\epsilon}^{(1+\sqrt{c})^2} \sqrt{(1+\sqrt{c})^2 - x}\,dx \propto \epsilon^{\frac{3}{2}} \tag{6}$$

  - to have $O(1)$ eigenvalues within $[E_+ - \epsilon, E_+]$ needs $\epsilon = O(n^{-\frac{2}{3}})$ (this is in fact the "spacing" between eigenvalues, which is of order $O(n^{-1})$ away from the edge)
- **Question**: hard-edge setting with $c = 1$, what happens?

Remarks on Fluctuation of the largest eigenvalue

- **BBP phase transition** [1]: named after the authors Jinho Baik, Gerard Ben Arous, Sandrine Peche, says that beyond the phase transition threshold, the fluctuation becomes a standard CLT type of order $O(n^{-1/2})$, from TW law of order $O(n^{-2/3})$
- universality for Tracy-Widom, real TW_1, complex TW_2, and quaternionic TW_4, for Wishart and Wigner matrix models, smallest and largest eigenvalues
- Tracy–Widom distributions connected in the asymptotics of a few growth models in the Kardar–Parisi–Zhang (KPZ) universality class

More spiked models:

- information-plus-noise model of the type $\frac{1}{n}(\mathbf{Z}+\mathbf{P})(\mathbf{Z}+\mathbf{P})^{\mathsf{T}}$
- additive $\mathbf{X}+\mathbf{P}$ for Wishart or Wigner type $\mathbf{X}$

Jinho Baik, Gérard Ben Arous and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *The Annals of Probability* 33.5 (2005), 1643–1697. ISSN: 0091-1798. DOI: 10.1214/009117905000000233

Remarks on Fluctuation of the largest eigenvalue

- **BBP phase transition** [1]: named after the authors Jinho Baik, Gerard Ben Arous, Sandrine Peche, says that beyond the phase transition threshold, the fluctuation becomes a standard CLT type of order $O(n^{-1/2})$, from TW law of order $O(n^{-2/3})$
- universality for Tracy-Widom, real TW_1, complex TW_2, and quaternionic TW_4, for Wishart and Wigner matrix models, smallest and largest eigenvalues
- Tracy–Widom distributions connected in the asymptotics of a few growth models in the Kardar–Parisi–Zhang (KPZ) universality class

More spiked models:

- information-plus-noise model of the type $\frac{1}{n}(\mathbf{Z}+\mathbf{P})(\mathbf{Z}+\mathbf{P})^{\mathsf{T}}$
- additive $\mathbf{X}+\mathbf{P}$ for Wishart or Wigner type $\mathbf{X}$

Jinho Baik, Gérard Ben Arous and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *The Annals of Probability* 33.5 (2005), 1643–1697. ISSN: 0091-1798. DOI: 10.1214/009117905000000233

Remarks on Fluctuation of the largest eigenvalue

- **BBP phase transition** [1]: named after the authors Jinho Baik, Gerard Ben Arous, Sandrine Peche, says that beyond the phase transition threshold, the fluctuation becomes a standard CLT type of order $O(n^{-1/2})$, from TW law of order $O(n^{-2/3})$
- universality for Tracy-Widom, real TW_1, complex TW_2, and quaternionic TW_4, for Wishart and Wigner matrix models, smallest and largest eigenvalues
- Tracy–Widom distributions connected in the asymptotics of a few growth models in the Kardar–Parisi–Zhang (KPZ) universality class

More spiked models:

- information-plus-noise model of the type $\frac{1}{n}(\mathbf{Z}+\mathbf{P})(\mathbf{Z}+\mathbf{P})^{\mathsf{T}}$
- additive $\mathbf{X}+\mathbf{P}$ for Wishart or Wigner type $\mathbf{X}$

Jinho Baik, Gérard Ben Arous and Sandrine Péché. “Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices”. In: *The Annals of Probability* 33.5 (2005), 1643–1697. ISSN: 0091-1798. DOI: 10.1214/009117905000000233

Remarks on Fluctuation of the largest eigenvalue

- **BBP phase transition** [1]: named after the authors Jinho Baik, Gerard Ben Arous, Sandrine Peche, says that beyond the phase transition threshold, the fluctuation becomes a standard CLT type of order $O(n^{-1/2})$, from TW law of order $O(n^{-2/3})$
- universality for Tracy-Widom, real TW_1, complex TW_2, and quaternionic TW_4, for Wishart and Wigner matrix models, smallest and largest eigenvalues
- Tracy–Widom distributions connected in the asymptotics of a few growth models in the Kardar–Parisi–Zhang (KPZ) universality class

More spiked models:

- information-plus-noise model of the type $\frac{1}{n}(\mathbf{Z}+\mathbf{P})(\mathbf{Z}+\mathbf{P})^\mathsf{T}$
- additive $\mathbf{X}+\mathbf{P}$ for Wishart or Wigner type $\mathbf{X}$

Jinho Baik, Gérard Ben Arous and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *The Annals of Probability* 33.5 (2005), 1643–1697. ISSN: 0091-1798. DOI: 10.1214/009117905000000233

Remarks on Fluctuation of the largest eigenvalue

» **BBP phase transition** [1]: named after the authors Jinho Baik, Gerard Ben Arous, Sandrine Peche, says that beyond the phase transition threshold, the fluctuation becomes a standard CLT type of order $O(n^{-1/2})$, from TW law of order $O(n^{-2/3})$

» universality for Tracy-Widom, real TW_1, complex TW_2, and quaternionic TW_4, for Wishart and Wigner matrix models, smallest and largest eigenvalues

» Tracy–Widom distributions connected in the asymptotics of a few growth models in the Kardar–Parisi–Zhang (KPZ) universality class

More spiked models:

» information-plus-noise model of the type $\frac{1}{n}(\mathbf{Z}+\mathbf{P})(\mathbf{Z}+\mathbf{P})^{\mathsf{T}}$

» additive $\mathbf{X}+\mathbf{P}$ for Wishart or Wigner type $\mathbf{X}$

Jinho Baik, Gérard Ben Arous and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *The Annals of Probability* 33.5 (2005), 1643–1697. ISSN: 0091-1798. DOI: 10.1214/009117905000000233

Remarks on Fluctuation of the largest eigenvalue

- **BBP phase transition** [1]: named after the authors Jinho Baik, Gerard Ben Arous, Sandrine Peche, says that beyond the phase transition threshold, the fluctuation becomes a standard CLT type of order $O(n^{-1/2})$, from TW law of order $O(n^{-2/3})$
- universality for Tracy-Widom, real TW_1, complex TW_2, and quaternionic TW_4, for Wishart and Wigner matrix models, smallest and largest eigenvalues
- Tracy–Widom distributions connected in the asymptotics of a few growth models in the Kardar–Parisi–Zhang (KPZ) universality class

More spiked models:

- information-plus-noise model of the type $\frac{1}{n}(\mathbf{Z}+\mathbf{P})(\mathbf{Z}+\mathbf{P})^\mathsf{T}$
- additive $\mathbf{X}+\mathbf{P}$ for Wishart or Wigner type $\mathbf{X}$

Jinho Baik, Gérard Ben Arous and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *The Annals of Probability* 33.5 (2005), 1643–1697. ISSN: 0091-1798. DOI: 10.1214/009117905000000233

Outline

Hypothesis testing in a signal-plus-noise model for cognitive radios

System model: let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with i.i.d. columns $\mathbf{x}_i \in \mathbb{R}^p$ received by array of p sensors, signal decision as the following binary hypothesis test:

$$\mathbf{X} = \begin{cases} \sigma \mathbf{Z}, & \mathcal{H}_0 \\ \mathbf{a}\mathbf{s}^\mathsf{T} + \sigma \mathbf{Z}, & \mathcal{H}_1 \end{cases}$$

where $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{p \times n}$, $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{a} \in \mathbb{R}^p$ deterministic of unit norm $\|\mathbf{a}\| = 1$, signal $\mathbf{s} = [s_1, \ldots, s_n]^\mathsf{T} \in \mathbb{R}^n$ with s_i i.i.d. random, and $\sigma > 0$. Denote $c = p/n > 0$.

» observation of either zero-mean Gaussian **noise** $\sigma \mathbf{z}_i$ of power σ^2, or deterministic **information** vector $\mathbf{a}$ modulated by an added scalar (random) **signal** s_i (e.g., ± 1).

» If $\mathbf{a}$, σ, and statistics of s_i are known, the decision-optimal Neyman-Pearson () test:

$$\frac{\mathbb{P}(\mathbf{X} \mid \mathcal{H}_1)}{\mathbb{P}(\mathbf{X} \mid \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha \tag{7}$$

for some $\alpha > 0$ controlling the Type I and II error rates.

Hypothesis testing in a signal-plus-noise model for cognitive radios

System model: let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with i.i.d. columns $\mathbf{x}_i \in \mathbb{R}^p$ received by array of p sensors, signal decision as the following binary hypothesis test:

$$\mathbf{X} = \begin{cases} \sigma \mathbf{Z}, & \mathcal{H}_0 \\ \mathbf{a}\mathbf{s}^\mathsf{T} + \sigma \mathbf{Z}, & \mathcal{H}_1 \end{cases}$$

where $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{p \times n}$, $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{a} \in \mathbb{R}^p$ deterministic of unit norm $\|\mathbf{a}\| = 1$, signal $\mathbf{s} = [s_1, \ldots, s_n]^\mathsf{T} \in \mathbb{R}^n$ with s_i i.i.d. random, and $\sigma > 0$. Denote $c = p/n > 0$.

» observation of either zero-mean Gaussian **noise** $\sigma \mathbf{z}_i$ of power σ^2, or deterministic **information** vector $\mathbf{a}$ modulated by an added scalar (random) **signal** s_i (e.g., ± 1).

» If $\mathbf{a}$, σ, and statistics of s_i are known, the decision-optimal Neyman-Pearson () test:

$$\frac{\mathbb{P}(\mathbf{X} \mid \mathcal{H}_1)}{\mathbb{P}(\mathbf{X} \mid \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha \tag{7}$$

for some $\alpha > 0$ controlling the Type I and II error rates.

Hypothesis testing in a signal-plus-noise model for cognitive radios

System model: let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with i.i.d. columns $\mathbf{x}_i \in \mathbb{R}^p$ received by array of p sensors, signal decision as the following binary hypothesis test:

$$\mathbf{X} = \begin{cases} \sigma \mathbf{Z}, & \mathcal{H}_0 \\ \mathbf{a}\mathbf{s}^\mathsf{T} + \sigma \mathbf{Z}, & \mathcal{H}_1 \end{cases}$$

where $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{p \times n}$, $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{a} \in \mathbb{R}^p$ deterministic of unit norm $\|\mathbf{a}\| = 1$, signal $\mathbf{s} = [s_1, \ldots, s_n]^\mathsf{T} \in \mathbb{R}^n$ with s_i i.i.d. random, and $\sigma > 0$. Denote $c = p/n > 0$.

» observation of either zero-mean Gaussian **noise** $\sigma \mathbf{z}_i$ of power σ^2, or deterministic **information** vector $\mathbf{a}$ modulated by an added scalar (random) **signal** s_i (e.g., ± 1).

» If $\mathbf{a}$, σ, and statistics of s_i are known, the decision-optimal Neyman-Pearson () test:

$$\frac{\mathbb{P}(\mathbf{X} \mid \mathcal{H}_1)}{\mathbb{P}(\mathbf{X} \mid \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha \tag{7}$$

for some $\alpha > 0$ controlling the Type I and II error rates.

Hypothesis testing via GLRT

- However, in practice, we do not know σ, nor the information vector $\mathbf{a} \in \mathbb{R}^p$ (to be recovered)
- in the case of $\mathbf{a}$ fully unknown, one may resort to a **generalized likelihood ratio test** (GLRT) defined as

$$\frac{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathbf{a}, \mathcal{H}_1)}{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha.$$

- Gaussian noise and signal s_i, GLRT has an explicit expression as a monotonous increasing function of $\|\mathbf{X}\mathbf{X}^\mathsf{T}\| / \operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})$, test equivalent to, for some known f,

$$T_p \equiv \frac{\|\mathbf{X}\mathbf{X}^\mathsf{T}\|}{\operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} f(\alpha).$$

- to evaluate the power of GLRT above, we need to assess the max and mean eigenvalues of SCM $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$

Hypothesis testing via GLRT

- However, in practice, we do not know σ, nor the information vector $\mathbf{a} \in \mathbb{R}^p$ (to be recovered)
- in the case of $\mathbf{a}$ fully unknown, one may resort to a **generalized likelihood ratio test** (GLRT) defined as

$$\frac{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathbf{a}, \mathcal{H}_1)}{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathcal{H}_0)} \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \alpha.$$

- Gaussian noise and signal s_i, GLRT has an explicit expression as a monotonous increasing function of $\|\mathbf{X}\mathbf{X}^\mathsf{T}\| / \operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})$, test equivalent to, for some known f,

$$T_p \equiv \frac{\|\mathbf{X}\mathbf{X}^\mathsf{T}\|}{\operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})} \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} f(\alpha).$$

- to evaluate the power of GLRT above, we need to assess the max and mean eigenvalues of SCM $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$

Hypothesis testing via GLRT

- However, in practice, we do not know σ, nor the information vector $\mathbf{a} \in \mathbb{R}^p$ (to be recovered)
- in the case of $\mathbf{a}$ fully unknown, one may resort to a **generalized likelihood ratio test** (GLRT) defined as

$$\frac{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathbf{a}, \mathcal{H}_1)}{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha.$$

- Gaussian noise and signal s_i, GLRT has an explicit expression as a monotonous increasing function of $\|\mathbf{X}\mathbf{X}^\mathsf{T}\|/\operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})$, test equivalent to, for some known f,

$$T_p \equiv \frac{\|\mathbf{X}\mathbf{X}^\mathsf{T}\|}{\operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} f(\alpha).$$

- to evaluate the power of GLRT above, we need to assess the max and mean eigenvalues of SCM $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$

Hypothesis testing via GLRT

- However, in practice, we do not know σ, nor the information vector $\mathbf{a} \in \mathbb{R}^p$ (to be recovered)
- in the case of $\mathbf{a}$ fully unknown, one may resort to a **generalized likelihood ratio test** (GLRT) defined as

$$\frac{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathbf{a}, \mathcal{H}_1)}{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha.$$

- Gaussian noise and signal s_i, GLRT has an explicit expression as a monotonous increasing function of $\|\mathbf{X}\mathbf{X}^\mathsf{T}\| / \operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})$, test equivalent to, for some known f,

$$T_p \equiv \frac{\|\mathbf{X}\mathbf{X}^\mathsf{T}\|}{\operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} f(\alpha).$$

- to evaluate the power of GLRT above, we need to assess the max and mean eigenvalues of SCM $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$

Hypothesis testing via GLRT

- However, in practice, we do not know σ, nor the information vector $\mathbf{a} \in \mathbb{R}^p$ (to be recovered)
- in the case of $\mathbf{a}$ fully unknown, one may resort to a **generalized likelihood ratio test** (GLRT) defined as

$$\frac{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathbf{a}, \mathcal{H}_1)}{\sup_{\sigma,\mathbf{a}} \mathbb{P}(\mathbf{X} \mid \sigma, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \alpha.$$

- Gaussian noise and signal s_i, GLRT has an explicit expression as a monotonous increasing function of $\|\mathbf{X}\mathbf{X}^\mathsf{T}\| / \operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})$, test equivalent to, for some known f,

$$T_p \equiv \frac{\|\mathbf{X}\mathbf{X}^\mathsf{T}\|}{\operatorname{tr}(\mathbf{X}\mathbf{X}^\mathsf{T})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} f(\alpha).$$

- to evaluate the power of GLRT above, we need to assess the max and mean eigenvalues of SCM $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$

Hypothesis testing in a signal-plus-noise model via GLRT

To set a maximum false alarm rate (or Type I error) of $r > 0$ for large n, p, according to RMT, one must choose a threshold $f(\alpha)$ for T_p:

$$\mathbb{P}(T_p \geq f(\alpha)) = r \Leftrightarrow \mu_{\mathrm{TW}_1}([A_p, +\infty)) = r, \quad A_p = (f(\alpha) - (1+\sqrt{c})^2)(1+\sqrt{c})^{-\frac{4}{3}} c^{\frac{1}{6}} n^{\frac{2}{3}} \quad (8)$$

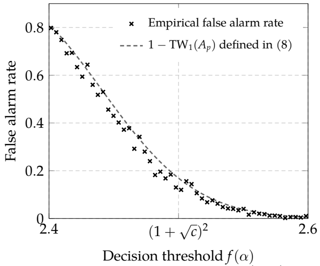
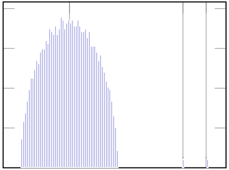


Figure: Comparison between empirical false alarm rates and $1 - \mathrm{TW}_1(A_p)$ for A_p of the form in (8), as a function of the threshold $f(\alpha) \in [(1+\sqrt{c})^2 - 5n^{-2/3}, (1+\sqrt{c})^2 + 5n^{-2/3}]$, for $p = 256$, $n = 1\,024$ and $\sigma = 1$.

Reminder on kernel spectral clustering

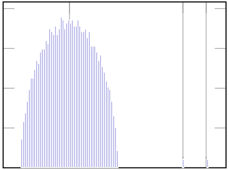
Two-step classification of n data points with distance kernel $\mathbf{K} \equiv \{f(\|\mathbf{x}_i - \mathbf{x}_j\|^2/p)\}_{i,j=1}^n$:



⇓ **Top eigenvectors** ⇓

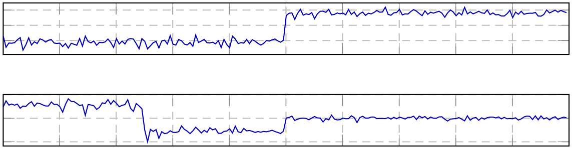
Reminder on kernel spectral clustering

Two-step classification of n data points with distance kernel $\mathbf{K} \equiv \{f(\|\mathbf{x}_i - \mathbf{x}_j\|^2/p)\}_{i,j=1}^n$:



⇓ **Top eigenvectors** ⇓

Reminder on kernel spectral clustering

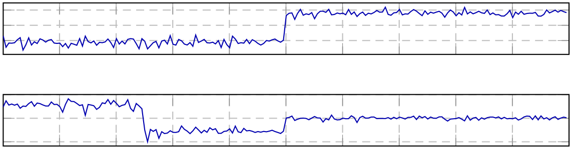


⇓ K-dimensional representation ⇓

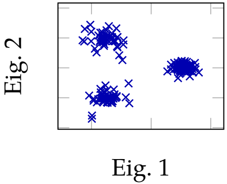
⇓

EM or k-means clustering

Reminder on kernel spectral clustering



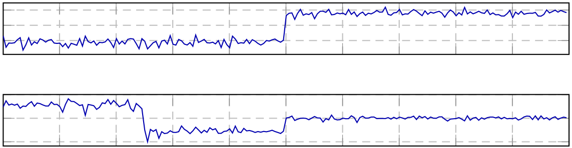
⇓ K-dimensional representation ⇓



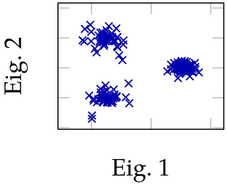
⇓

EM or k-means clustering

Reminder on kernel spectral clustering



⇓ K-dimensional representation ⇓



⇓

EM or k-means clustering

Application to "compressed" spectral clustering

Entry-wise *nonlinear* transformation of $\mathbf{X}^\mathsf{T}\mathbf{X}$: $\mathbf{K} = \{f(\mathbf{x}_i^\mathsf{T}\mathbf{x}_j/\sqrt{p})/\sqrt{p}\}_{i,j=1}^n$, with

Sparsification: $f_1(t) = t \cdot 1_{|t|>\sqrt{2}s}$

Quantization: $f_2(t) = 2^{2-M}(\lfloor t \cdot 2^{M-2}/\sqrt{2}s\rfloor + 1/2) \cdot 1_{|t|\le\sqrt{2}s} + \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Binarization: $f_3(t) = \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Application to "compressed" spectral clustering

Entry-wise *nonlinear* transformation of $\mathbf{X}^\mathsf{T}\mathbf{X}$: $\mathbf{K} = \{f(\mathbf{x}_i^\mathsf{T}\mathbf{x}_j/\sqrt{p})/\sqrt{p}\}_{i,j=1}^n$, with

Sparsification: $f_1(t) = t \cdot 1_{|t|>\sqrt{2}s}$

Quantization: $f_2(t) = 2^{2-M}(\lfloor t \cdot 2^{M-2}/\sqrt{2}s \rfloor + 1/2) \cdot 1_{|t|\leq\sqrt{2}s} + \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Binarization: $f_3(t) = \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Application to "compressed" spectral clustering

Entry-wise *nonlinear* transformation of $\mathbf{X}^\mathsf{T}\mathbf{X}$: $\mathbf{K} = \{f(\mathbf{x}_i^\mathsf{T}\mathbf{x}_j/\sqrt{p})/\sqrt{p}\}_{i,j=1}^n$, with

Sparsification: $f_1(t) = t \cdot 1_{|t|>\sqrt{2}s}$

Quantization: $f_2(t) = 2^{2-M}(\lfloor t \cdot 2^{M-2}/\sqrt{2}s \rfloor + 1/2) \cdot 1_{|t|\le\sqrt{2}s} + \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Binarization: $f_3(t) = \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Application to "compressed" spectral clustering

Entry-wise *nonlinear* transformation of $\mathbf{X}^\top\mathbf{X}$: $\mathbf{K} = \{f(\mathbf{x}_i^\top \mathbf{x}_j/\sqrt{p})/\sqrt{p}\}_{i,j=1}^n$, with

Sparsification: $f_1(t) = t \cdot 1_{|t|>\sqrt{2}s}$

Quantization: $f_2(t) = 2^{2-M}(\lfloor t \cdot 2^{M-2}/\sqrt{2}s \rfloor + 1/2) \cdot 1_{|t|\le\sqrt{2}s} + \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Binarization: $f_3(t) = \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Application to "compressed" spectral clustering

Entry-wise *nonlinear* transformation of $\mathbf{X}^\mathsf{T}\mathbf{X}$: $\mathbf{K} = \{f(\mathbf{x}_i^\mathsf{T}\mathbf{x}_j/\sqrt{p})/\sqrt{p}\}_{i,j=1}^n$, with

Sparsification: $f_1(t) = t \cdot 1_{|t|>\sqrt{2}s}$

Quantization: $f_2(t) = 2^{2-M}(\lfloor t \cdot 2^{M-2}/\sqrt{2}s \rfloor + 1/2) \cdot 1_{|t|\leq\sqrt{2}s} + \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Binarization: $f_3(t) = \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

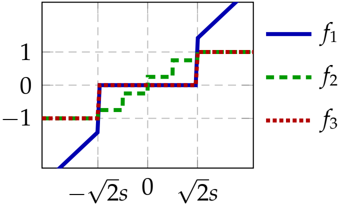
Application to "compressed" spectral clustering

Entry-wise *nonlinear* transformation of $\mathbf{X}^\mathsf{T}\mathbf{X}$: $\mathbf{K} = \{f(\mathbf{x}_i^\mathsf{T}\mathbf{x}_j/\sqrt{p})/\sqrt{p}\}_{i,j=1}^n$, with

Sparsification: $f_1(t) = t \cdot 1_{|t|>\sqrt{2}s}$

Quantization: $f_2(t) = 2^{2-M}(\lfloor t \cdot 2^{M-2}/\sqrt{2}s \rfloor + 1/2) \cdot 1_{|t|\le\sqrt{2}s} + \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$

Binarization: $f_3(t) = \mathrm{sign}(t) \cdot 1_{|t|>\sqrt{2}s}$



Notations: For each f and $\xi \sim \mathcal{N}(0,1)$, define the (generalized) moments

$$a_0 = \mathbb{E}[f(\xi)] = 0, \quad \mathbf{a_1} = \mathbb{E}[\xi f(\xi)], \quad \mathbf{a_2} = \mathbb{E}[\xi^2 f(\xi)]/\sqrt{2}, \quad \boldsymbol{\nu} = \mathbb{E}[f^2(\xi)] \geq a_1^2 + a_2^2. \tag{9}$$

f	$\mathbf{a_1}$	$\boldsymbol{\nu}$
f_1	$\operatorname{erfc}(s) + 2se^{-s^2}/\sqrt{\pi}$	$\operatorname{erfc}(s) + 2se^{-s^2}/\sqrt{\pi}$
f_2	$\sqrt{\frac{2}{\pi}} \cdot 2^{1-M}(1 + e^{-s^2} + \sum_{k=1}^{2^{M-2}-1} 2e^{-\frac{k^2 s^2}{4^{M-2}}})$	$1 - \frac{2^M-1}{4^{M-1}} \operatorname{erf}(s) - \sum_{k=1}^{2^{M-2}-1} \frac{k \operatorname{erf}(ks \cdot 2^{2-M})}{2^{2M-5}}$
f_3	$e^{-s^2}\sqrt{2/\pi}$	$\operatorname{erfc}(s)$

with $\mathbf{a_2} = \mathbf{0}$, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$, $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ error/complementary error function.

As $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, the empirical spectral measure $\omega_{\mathbf{K}} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(\mathbf{K})}$ can be asymptotically determined by $m(z) = \int (t-z)^{-1} \omega(dt)$ solution to

$$z = -\frac{1}{m(z)} - \frac{\nu - {a_1}^2}{c} m(z) - \frac{{a_1}^2 m(z)}{c + a_1 m(z)}. \tag{10}$$

Limiting spectral measure

Notations: For each f and $\xi \sim \mathcal{N}(0,1)$, define the (generalized) moments

$$a_0 = \mathbb{E}[f(\xi)] = 0, \quad \mathbf{a_1} = \mathbb{E}[\xi f(\xi)], \quad \mathbf{a_2} = \mathbb{E}[\xi^2 f(\xi)]/\sqrt{2}, \quad \boldsymbol{\nu} = \mathbb{E}[f^2(\xi)] \geq a_1^2 + a_2^2. \tag{9}$$

f	$\mathbf{a_1}$	$\boldsymbol{\nu}$
f_1	$\operatorname{erfc}(s) + 2se^{-s^2}/\sqrt{\pi}$	$\operatorname{erfc}(s) + 2se^{-s^2}/\sqrt{\pi}$
f_2	$\sqrt{\frac{2}{\pi}} \cdot 2^{1-M}(1 + e^{-s^2} + \sum_{k=1}^{2^{M-2}-1} 2e^{-\frac{k^2 s^2}{4^{M-2}}})$	$1 - \frac{2^M-1}{4^{M-1}} \operatorname{erf}(s) - \sum_{k=1}^{2^{M-2}-1} \frac{k \operatorname{erf}(ks \cdot 2^{2-M})}{2^{2M-5}}$
f_3	$e^{-s^2}\sqrt{2/\pi}$	$\operatorname{erfc}(s)$

with $\mathbf{a_2} = \mathbf{0}$, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$, $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ error/complementary error function.

As $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, the empirical spectral measure $\omega_{\mathbf{K}} = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(\mathbf{K})}$ can be asymptotically determined by $m(z) = \int (t - z)^{-1} \omega(dt)$ solution to

$$z = -\frac{1}{m(z)} - \frac{\nu - {a_1}^2}{c} m(z) - \frac{{a_1}^2 m(z)}{c + a_1 m(z)}. \tag{10}$$

Limiting spectral measure

Notations: For each f and $\xi \sim \mathcal{N}(0,1)$, define the (generalized) moments

$$a_0 = \mathbb{E}[f(\xi)] = 0, \quad \mathbf{a_1} = \mathbb{E}[\xi f(\xi)], \quad \mathbf{a_2} = \mathbb{E}[\xi^2 f(\xi)]/\sqrt{2}, \quad \boldsymbol{\nu} = \mathbb{E}[f^2(\xi)] \geq a_1^2 + a_2^2. \tag{9}$$

f	$\mathbf{a_1}$	$\boldsymbol{\nu}$
f_1	$\operatorname{erfc}(s) + 2se^{-s^2}/\sqrt{\pi}$	$\operatorname{erfc}(s) + 2se^{-s^2}/\sqrt{\pi}$
f_2	$\sqrt{\frac{2}{\pi}} \cdot 2^{1-M}(1 + e^{-s^2} + \sum_{k=1}^{2^{M-2}-1} 2e^{-\frac{k^2s^2}{4^{M-2}}})$	$1 - \frac{2^M-1}{4^{M-1}} \operatorname{erf}(s) - \sum_{k=1}^{2^{M-2}-1} \frac{k \operatorname{erf}(ks \cdot 2^{2-M})}{2^{2M-5}}$
f_3	$e^{-s^2}\sqrt{2/\pi}$	$\operatorname{erfc}(s)$

with $\mathbf{a_2} = \mathbf{0}$, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$, $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ error/complementary error function.

As $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, the empirical spectral measure $\omega_{\mathbf{K}} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(\mathbf{K})}$ can be asymptotically determined by $m(z) = \int (t-z)^{-1} \omega(dt)$ solution to

$$z = -\frac{1}{m(z)} - \frac{\nu - {a_1}^2}{c} m(z) - \frac{{a_1}^2 m(z)}{c + a_1 m(z)}. \tag{10}$$

Limiting spectral measure

For $a_1 > 0$ and $\mathbf{a_2 = 0}$, similarly define $F(x) = x^4 + 2x^3 + \left(1 - \frac{c\nu}{{a_1}^2}\right)x^2 - 2cx - c$ and $G(x) = \frac{a_1}{c}(1+x) + \frac{a_1}{x} + \frac{\nu - {a_1}^2}{a_1}\frac{1}{1+x}$ and let γ be the largest real solution to $F(\gamma) = 0$. Then,

$$\hat{\lambda} \to \lambda = \begin{cases} G(\rho), & \rho > \gamma \\ G(\gamma), & \rho \leq \gamma \end{cases}, \quad \frac{1}{n}|\hat{\mathbf{v}}^\mathsf{T}\mathbf{v}|^2 \to \alpha = \begin{cases} \frac{F(\rho)}{\rho(1+\rho)^3}, & \rho > \gamma \\ 0, & \rho \leq \gamma \end{cases} \tag{11}$$

as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, for SNR $\rho = \lim \|\boldsymbol{\mu}\|^2$.

Informative spike and a phase transition

"Compressed" spectral clustering: practical implications

Let $a_1 > 0, a_2 = 0$, and $\hat{\mathcal{C}}_i = \operatorname{sign}([\hat{\mathbf{v}}]_i)$ be the estimate of class $\mathcal{C}_i$ of the datum $\mathbf{x}_i$, with $\hat{\mathbf{v}}^\mathsf{T}\mathbf{v} \geq 0$ for $\hat{\mathbf{v}}$ the top eigenvector of $\mathbf{K}$. Then, the misclassification rate satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{\mathcal{C}}_i \neq \mathcal{C}_i} \to \frac{1}{2}\operatorname{erfc}(\sqrt{\alpha/(2-2\alpha)}), \quad \alpha = \lim_{n,p\to\infty} \frac{1}{n}|\hat{\mathbf{v}}^\mathsf{T}\mathbf{v}|^2.$$

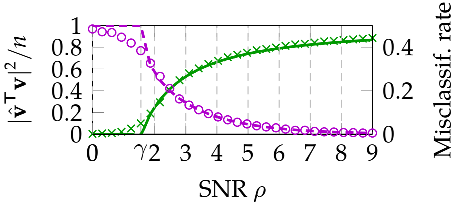
Performance of spectral clustering

"Compressed" spectral clustering: practical implications

Let $a_1 > 0, a_2 = 0$, and $\hat{\mathcal{C}}_i = \operatorname{sign}([\hat{\mathbf{v}}]_i)$ be the estimate of class $\mathcal{C}_i$ of the datum $\mathbf{x}_i$, with $\hat{\mathbf{v}}^\mathsf{T}\mathbf{v} \geq 0$ for $\hat{\mathbf{v}}$ the top eigenvector of $\mathbf{K}$. Then, the misclassification rate satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{\mathcal{C}}_i \neq \mathcal{C}_i} \to \frac{1}{2}\operatorname{erfc}(\sqrt{\alpha/(2-2\alpha)}), \quad \alpha = \lim_{n,p\to\infty} \frac{1}{n}|\hat{\mathbf{v}}^\mathsf{T}\mathbf{v}|^2.$$

Performance of spectral clustering



Consequence: optimal quantization/binarization threshold