Probability and Stochastic Process II: Random Matrix Theory and Applications Lecture 4: Large-dimensional Sample Covariance Matrix

Zhenyu Liao, Tiebin Mi, Caiming Qiu

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

March 15, 2023

Outline

What we will have today

» sample covariance matrix and the generalized Marčenko–Pastur law

» advanced topics for SCM: limiting spectrum, no eigenvalue outside the support

» statistical inference on SCM

» beyond SCM: bi-correlated model (separable covariance model), sample covariance of mixture models, and generalized semicircle law with a variance profile, etc

What we will have today

- sample covariance matrix and the generalized Marčenko–Pastur law
- advanced topics for SCM: limiting spectrum, no eigenvalue outside the support
- statistical inference on SCM
- beyond SCM: bi-correlated model (separable covariance model), sample covariance of mixture models, and generalized semicircle law with a variance profile, etc

What we will have today

- sample covariance matrix and the generalized Marčenko–Pastur law
- advanced topics for SCM: limiting spectrum, no eigenvalue outside the support
- statistical inference on SCM
- beyond SCM: bi-correlated model (separable covariance model), sample covariance of mixture models, and generalized semicircle law with a variance profile, etc

Outline

Generalized MP for SCM

Spectrum characterization

No eigenvalue outside the support

Statistical inference

SCM and generalized Marčenko–Pastur law

Let $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z} \in \mathbb{R}^{p\times n}$ with symmetric nonnegative definite $\mathbf{C} \in \mathbb{R}^{p\times p}$, $\mathbf{Z} \in \mathbb{R}^{p\times n}$ having independent zero mean and unit variance entries. Then, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, for $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p)^{-1}$ and $\tilde{\mathbf{Q}}(z) = (\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X} - z\mathbf{I}_n)^{-1}$,

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = -\frac{1}{z}\left(\mathbf{I}_p + \tilde{m}_p(z)\mathbf{C}\right)^{-1}, \quad \tilde{\mathbf{Q}}(z) \leftrightarrow \bar{\tilde{\mathbf{Q}}}(z) = \tilde{m}_p(z)\mathbf{I}_n,$$

with $\tilde{m}_p(z)$ unique solution to $\tilde{m}_p(z) = \left(-z + \frac{1}{n}\operatorname{tr}\mathbf{C}\left(\mathbf{I}_p + \tilde{m}_p(z)\mathbf{C}\right)^{-1}\right)^{-1}$.
If the empirical spectral measure of $\mathbf{C}$ converges $\mu_{\mathbf{C}} \to \nu$ as $p \to \infty$, then $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}} \to \mu$, $\mu_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}} \to \tilde{\mu}$ where $\mu, \tilde{\mu}$ admitting Stieltjes transforms $m(z)$ and $\tilde{m}(z)$ such that

$$m(z) = \frac{1}{c}\tilde{m}(z) + \frac{1-c}{cz}, \quad \tilde{m}(z) = \left(-z + c\int \frac{t\nu(dt)}{1+\tilde{m}(z)t}\right)^{-1}. \tag{1}$$

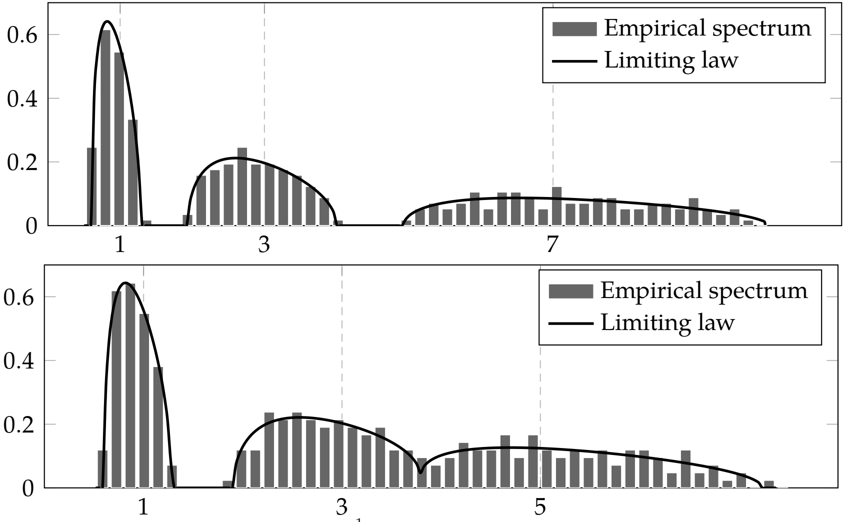


Figure: Histogram of the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{T}}$, $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z} \in \mathbb{R}^{p\times n}$, $[\mathbf{Z}]_{ij} \sim \mathcal{N}(0,1)$, $n = 3\,000$; for $p = 300$ and $\mathbf{C}$ having spectral measure $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_7)$ (**top**) and $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$ (**bottle**).

Outline

SCM: characterization of the limiting spectrum

We focus on the following equation that characterize the limiting spectrum of SCM

$$\boxed{\tilde{m}(z) = \left(-z + c \int \frac{t\nu(dt)}{1 + \tilde{m}(z)t} \right)^{-1}} \tag{2}$$

with ν the limiting spectral measure of $\mathbf{C}$.

- note that this is equivalent to $z = -\frac{1}{\tilde{m}(z)} + c \int \frac{t\nu(dt)}{1+t\tilde{m}(z)}$
- so $\tilde{m}(\cdot) : \mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}) \to \mathbb{C}, z \mapsto \tilde{m}(z)$ admits the functional inverse

$$z(\cdot) : \tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu})) \to \mathbb{C}$$
$$\tilde{m} \mapsto -\frac{1}{\tilde{m}} + c \int \frac{t\nu(dt)}{1 + t\tilde{m}}.$$

- only formally defined on the domain $\tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}))$, but can be *extended* to all values $\tilde{m} \in \mathbb{C}$ such that $0 \notin 1 + \tilde{m} \cdot \operatorname{supp}(\nu)$.

SCM: characterization of the limiting spectrum

We focus on the following equation that characterize the limiting spectrum of SCM

$$\boxed{\tilde{m}(z) = \left(-z + c \int \frac{t\nu(dt)}{1 + \tilde{m}(z)t} \right)^{-1}} \tag{2}$$

with ν the limiting spectral measure of $\mathbf{C}$.

» note that this is equivalent to $z = -\frac{1}{\tilde{m}(z)} + c \int \frac{t\nu(dt)}{1+t\tilde{m}(z)}$

» so $\tilde{m}(\cdot) : \mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}) \to \mathbb{C}, z \mapsto \tilde{m}(z)$ admits the functional inverse

$$z(\cdot) : \tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu})) \to \mathbb{C}$$
$$\tilde{m} \mapsto -\frac{1}{\tilde{m}} + c \int \frac{t\nu(dt)}{1 + t\tilde{m}}.$$

» only formally defined on the domain $\tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}))$, but can be *extended* to all values $\tilde{m} \in \mathbb{C}$ such that $0 \notin 1 + \tilde{m} \cdot \operatorname{supp}(\nu)$.

SCM: characterization of the limiting spectrum

We focus on the following equation that characterize the limiting spectrum of SCM

$$\boxed{\tilde{m}(z) = \left(-z + c \int \frac{t\nu(dt)}{1 + \tilde{m}(z)t} \right)^{-1}} \tag{2}$$

with ν the limiting spectral measure of $\mathbf{C}$.

» note that this is equivalent to $z = -\frac{1}{\tilde{m}(z)} + c \int \frac{t\nu(dt)}{1+t\tilde{m}(z)}$

» so $\tilde{m}(\cdot) : \mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}) \to \mathbb{C}$, $z \mapsto \tilde{m}(z)$ admits the functional inverse

$$z(\cdot) : \tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu})) \to \mathbb{C}$$
$$\tilde{m} \mapsto -\frac{1}{\tilde{m}} + c \int \frac{t\nu(dt)}{1 + t\tilde{m}}.$$

» only formally defined on the domain $\tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}))$, but can be *extended* to all values $\tilde{m} \in \mathbb{C}$ such that $0 \notin 1 + \tilde{m} \cdot \operatorname{supp}(\nu)$.

SCM: characterization of the limiting spectrum

We focus on the following equation that characterize the limiting spectrum of SCM

$$\boxed{\tilde{m}(z) = \left(-z + c \int \frac{t\nu(dt)}{1 + \tilde{m}(z)t} \right)^{-1}} \tag{2}$$

with ν the limiting spectral measure of $\mathbf{C}$.

» note that this is equivalent to $z = -\frac{1}{\tilde{m}(z)} + c \int \frac{t\nu(dt)}{1+t\tilde{m}(z)}$

» so $\tilde{m}(\cdot) : \mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}) \to \mathbb{C}, z \mapsto \tilde{m}(z)$ admits the functional inverse

$$z(\cdot) : \tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu})) \to \mathbb{C}$$
$$\tilde{m} \mapsto -\frac{1}{\tilde{m}} + c \int \frac{t\nu(dt)}{1 + t\tilde{m}}.$$

» only formally defined on the domain $\tilde{m}(\mathbb{C} \setminus \operatorname{supp}(\tilde{\mu}))$, but can be *extended* to all values $\tilde{m} \in \mathbb{C}$ such that $0 \notin 1 + \tilde{m} \cdot \operatorname{supp}(\nu)$.

Consider the extension of the functional inverse on $\mathbb{R}$

» **Outside the support:**

- the Stieltjes transform $m_\mu(x) = \int (t-x)^{-1}\mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \mathrm{supp}(\mu)$
- so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- if $x(\cdot)$ admits an extension to some domain $\mathcal{S}$ with $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- the complementary $\mathbb{R} \setminus \mathrm{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$ **[This is a non-trivial fact!]**
- this formally defines the support of the limiting measure μ (of $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}}$)

» **In the support:** to determine the density of μ, first prove the existence of $\tilde{m}^\circ(x) = \lim_{\epsilon \to 0} \tilde{m}(x + \imath\epsilon)$. Upon existence, since $\Im[\tilde{m}^\circ(x)] > 0$ for $x \in \mathrm{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde{m}(z)$ to find that $\tilde{m}^\circ(x)$ is a solution *with positive imaginary part* of $\tilde{m}^\circ(x) = \left(-x + c\int \frac{t\nu(dt)}{1+\tilde{m}^\circ(x)t}\right)^{-1}$.

Consider the extension of the functional inverse on $\mathbb{R}$

» **Outside the support:**

- the Stieltjes transform $m_\mu(x) = \int (t-x)^{-1}\mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \mathrm{supp}(\mu)$
- so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- if $x(\cdot)$ admits an extension to some domain $\mathcal{S}$ with $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- the complementary $\mathbb{R} \setminus \mathrm{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$ **[This is a non-trivial fact!]**
- this formally defines the support of the limiting measure μ (of $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}}$)

» **In the support:** to determine the density of μ, first prove the existence of $\tilde m^\circ(x) = \lim_{\epsilon\to 0} \tilde m(x + \imath\epsilon)$. Upon existence, since $\Im[\tilde m^\circ(x)] > 0$ for $x \in \mathrm{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde m(z)$ to find that $\tilde m^\circ(x)$ is a solution *with positive imaginary part* of $\tilde m^\circ(x) = \left(-x + c\int \frac{t\nu(dt)}{1+\tilde m^\circ(x)t}\right)^{-1}$.

Consider the extension of the functional inverse on $\mathbb{R}$

» **Outside the support:**
- the Stieltjes transform $m_\mu(x) = \int (t-x)^{-1}\mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \mathrm{supp}(\mu)$
- so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- if $x(\cdot)$ admits an extension to some domain $\mathcal{S}$ with $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- the complementary $\mathbb{R} \setminus \mathrm{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$ **[This is a non-trivial fact!]**
- this formally defines the support of the limiting measure μ (of $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}}$)

» **In the support:** to determine the density of μ, first prove the existence of $\tilde{m}^\circ(x) = \lim_{\epsilon \to 0} \tilde{m}(x + \imath\epsilon)$. Upon existence, since $\Im[\tilde{m}^\circ(x)] > 0$ for $x \in \mathrm{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde{m}(z)$ to find that $\tilde{m}^\circ(x)$ is a solution *with positive imaginary part* of $\tilde{m}^\circ(x) = \left(-x + c\int \frac{t\nu(dt)}{1+\tilde{m}^\circ(x)t}\right)^{-1}$.

Consider the extension of the functional inverse on $\mathbb{R}$

» **Outside the support:**

- the Stieltjes transform $m_\mu(x) = \int (t-x)^{-1}\mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \mathrm{supp}(\mu)$
- so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- if $x(\cdot)$ admits an extension to some domain $\mathcal{S}$ with $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- the complementary $\mathbb{R} \setminus \mathrm{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$ **[This is a non-trivial fact!]**
- this formally defines the support of the limiting measure μ (of $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\top}$)

» **In the support:** to determine the density of μ, first prove the existence of $\tilde{m}^\circ(x) = \lim_{\epsilon \to 0} \tilde{m}(x + \imath\epsilon)$. Upon existence, since $\Im[\tilde{m}^\circ(x)] > 0$ for $x \in \mathrm{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde{m}(z)$ to find that $\tilde{m}^\circ(x)$ is a solution *with positive imaginary part* of $\tilde{m}^\circ(x) = \left(-x + c\int \frac{t\nu(dt)}{1+\tilde{m}^\circ(x)t}\right)^{-1}$.

Consider the extension of the functional inverse on $\mathbb{R}$

» **Outside the support:**

- the Stieltjes transform $m_\mu(x) = \int (t-x)^{-1}\mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \operatorname{supp}(\mu)$
- so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \operatorname{supp}(\mu))$
- if $x(\cdot)$ admits an extension to some domain $\mathcal{S}$ with $m_\mu(\mathbb{R} \setminus \operatorname{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \operatorname{supp}(\mu))$
- the complementary $\mathbb{R} \setminus \operatorname{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$ **[This is a non-trivial fact!]**
- this formally defines the support of the limiting measure μ (of $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\top}$)

» **In the support:** to determine the density of μ, first prove the existence of $\tilde{m}^\circ(x) = \lim_{\epsilon \to 0} \tilde{m}(x + \imath\epsilon)$. Upon existence, since $\Im[\tilde{m}^\circ(x)] > 0$ for $x \in \operatorname{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde{m}(z)$ to find that $\tilde{m}^\circ(x)$ is a solution *with positive imaginary part* of $\tilde{m}^\circ(x) = \left(-x + c\int \frac{t\nu(dt)}{1+\tilde{m}^\circ(x)t}\right)^{-1}$.

Consider the extension of the functional inverse on $\mathbb{R}$

» **Outside the support:**

- the Stieltjes transform $m_\mu(x) = \int (t-x)^{-1}\mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \operatorname{supp}(\mu)$
- so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \operatorname{supp}(\mu))$
- if $x(\cdot)$ admits an extension to some domain $\mathcal{S}$ with $m_\mu(\mathbb{R} \setminus \operatorname{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \operatorname{supp}(\mu))$
- the complementary $\mathbb{R} \setminus \operatorname{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$ **[This is a non-trivial fact!]**
- this formally defines the support of the limiting measure μ (of $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\top}$)

» **In the support:** to determine the density of μ, first prove the existence of $\tilde{m}^\circ(x) = \lim_{\epsilon \to 0} \tilde{m}(x + \imath\epsilon)$. Upon existence, since $\Im[\tilde{m}^\circ(x)] > 0$ for $x \in \operatorname{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde{m}(z)$ to find that $\tilde{m}^\circ(x)$ is a solution *with positive imaginary part* of $\tilde{m}^\circ(x) = \left(-x + c\int \frac{t\nu(dt)}{1+\tilde{m}^\circ(x)t}\right)^{-1}$.

Consider the extension of the functional inverse on $\mathbb{R}$

» **Outside the support:**
- the Stieltjes transform $m_\mu(x) = \int (t-x)^{-1}\mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \mathrm{supp}(\mu)$
- so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- if $x(\cdot)$ admits an extension to some domain $\mathcal{S}$ with $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \mathrm{supp}(\mu))$
- the complementary $\mathbb{R} \setminus \mathrm{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$ [**This is a non-trivial fact!**]
- this formally defines the support of the limiting measure μ (of $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}}$)

» **In the support:** to determine the density of μ, first prove the existence of $\tilde{m}^\circ(x) = \lim_{\epsilon \to 0} \tilde{m}(x + \imath\epsilon)$. Upon existence, since $\Im[\tilde{m}^\circ(x)] > 0$ for $x \in \mathrm{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde{m}(z)$ to find that $\tilde{m}^\circ(x)$ is a solution *with positive imaginary part* of $\tilde{m}^\circ(x) = \left(-x + c\int \frac{t\nu(dt)}{1+\tilde{m}^\circ(x)t}\right)^{-1}$.

Define the (extended) function

$$x(\cdot) : \mathbb{R} \setminus \{\tilde{m} \mid (-1/\tilde{m}) \in \operatorname{supp}(\nu)\} \to \mathbb{R}$$
$$\tilde{m} \mapsto -\frac{1}{\tilde{m}} + c \int \frac{t\nu(dt)}{1+\tilde{m}t}.$$

Then, $\tilde{\mu}$ has a density $\tilde{f}$ on $\mathbb{R} \setminus \{0\}$ and

- for $y \in \operatorname{supp}(\tilde{\mu})$, $\tilde{f}(y) = \frac{1}{\pi}\Im[\tilde{m}^\circ(y)]$ with $\tilde{m}^\circ(y)$ the unique solution with positive imaginary part of $x(\tilde{m}^\circ(y)) = y$;
- the support $\operatorname{supp}(\tilde{\mu}) \setminus \{0\}$, which coincides with $\operatorname{supp}(\mu) \setminus \{0\}$, is defined by

$$\operatorname{supp}(\mu) \setminus \{0\}$$
$$= \mathbb{R} \setminus \left\{ x(\tilde{m}) \mid (-1/\tilde{m}) \in \mathbb{R} \setminus \{\operatorname{supp}(\nu) \cup \{0\}\} \text{ and } \boxed{x'(\tilde{m}) > 0} \right\}.$$

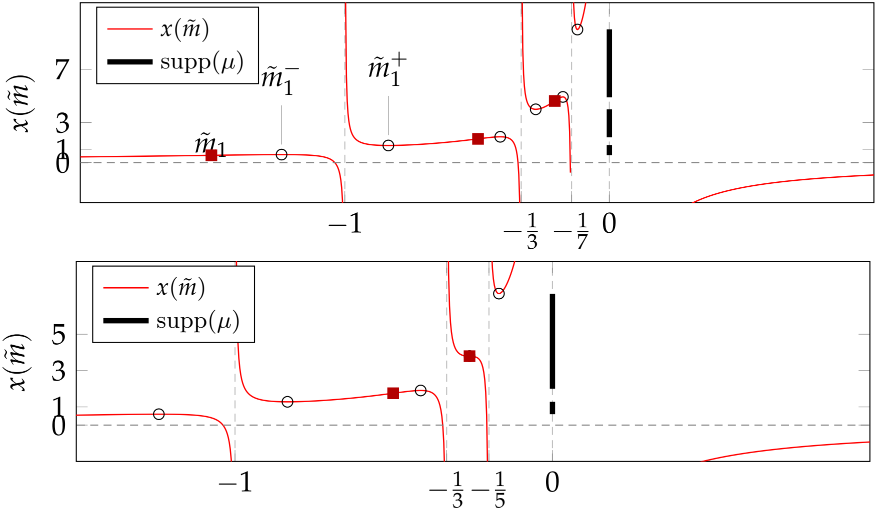


Figure: $x(\tilde{m})$ for $-1/\tilde{m} \in \mathbb{R} \setminus \text{supp}(\nu)$, with $\nu = \frac{1}{3}(\delta_1 + \delta_3 + \delta_7)$ (**top**) and $\nu = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$ (**bottle**), with $c = 1/10$ in both cases. Local extrema are marked by circles, inflexion points by squares. supp μ can be read on the vertical axes.

A few comments on the theorem and figure

- the restriction of $x(\cdot)$, as the functional inverse of $\tilde{m}(\cdot)$, to its growing sections, is a growing function
- in the figure, since ν is discrete, $x(\cdot)$ presents asymptotes at each $-1/t$, $t \in \operatorname{supp}(\nu)$. Thus, from the previous item, $\operatorname{supp}(\mu)$ is here determined by the union $\cup_k [\tilde{m}_k^-, \tilde{m}_k^+]$ for $\tilde{m}_1^- < \tilde{m}_1^+ < \tilde{m}_2^- < \ldots$ the successive values of $\tilde{m}$ such that $x'(\tilde{m}) = 0$. This remark may however not hold for ν with continuous support.
- the derivative of $x(\cdot)$ is given by

$$x'(\tilde{m}) = \frac{1}{\tilde{m}^2} - c \int \frac{t^2 \nu(dt)}{(1 + t\tilde{m})^2}$$

  and thus $\tilde{m}^2 x'(\tilde{m})$ converges to $1 - c$ as $|\tilde{m}| \to \infty$, while $x(\tilde{m}) \to 0$. Thus $x(\cdot)$ is either decreasing or increasing at $\pm\infty$ depending on whether $c < 1$ or $c > 1$.

A few comments on the theorem and figure

- the restriction of $x(\cdot)$, as the functional inverse of $\tilde{m}(\cdot)$, to its growing sections, is a growing function
- in the figure, since ν is discrete, $x(\cdot)$ presents asymptotes at each $-1/t$, $t \in \operatorname{supp}(\nu)$. Thus, from the previous item, $\operatorname{supp}(\mu)$ is here determined by the union $\cup_k [\tilde{m}_k^-, \tilde{m}_k^+]$ for $\tilde{m}_1^- < \tilde{m}_1^+ < \tilde{m}_2^- < \ldots$ the successive values of $\tilde{m}$ such that $x'(\tilde{m}) = 0$. This remark may however not hold for ν with continuous support.
- the derivative of $x(\cdot)$ is given by

$$x'(\tilde{m}) = \frac{1}{\tilde{m}^2} - c \int \frac{t^2 \nu(dt)}{(1 + t\tilde{m})^2}$$

and thus $\tilde{m}^2 x'(\tilde{m})$ converges to $1 - c$ as $|\tilde{m}| \to \infty$, while $x(\tilde{m}) \to 0$. Thus $x(\cdot)$ is either decreasing or increasing at $\pm\infty$ depending on whether $c < 1$ or $c > 1$.

A few comments on the theorem and figure

- the restriction of $x(\cdot)$, as the functional inverse of $\tilde{m}(\cdot)$, to its growing sections, is a growing function
- in the figure, since ν is discrete, $x(\cdot)$ presents asymptotes at each $-1/t$, $t \in \operatorname{supp}(\nu)$. Thus, from the previous item, $\operatorname{supp}(\mu)$ is here determined by the union $\cup_k [\tilde{m}_k^-, \tilde{m}_k^+]$ for $\tilde{m}_1^- < \tilde{m}_1^+ < \tilde{m}_2^- < \ldots$ the successive values of $\tilde{m}$ such that $x'(\tilde{m}) = 0$. This remark may however not hold for ν with continuous support.
- the derivative of $x(\cdot)$ is given by

$$x'(\tilde{m}) = \frac{1}{\tilde{m}^2} - c \int \frac{t^2 \nu(dt)}{(1 + t\tilde{m})^2}$$

and thus $\tilde{m}^2 x'(\tilde{m})$ converges to $1 - c$ as $|\tilde{m}| \to \infty$, while $x(\tilde{m}) \to 0$. Thus $x(\cdot)$ is either decreasing or increasing at $\pm\infty$ depending on whether $c < 1$ or $c > 1$.

Change of variable on the complex plane

» An important consequence of the study above of $z(\cdot)$ (and its restriction $x(\cdot)$ to the real axis) is that the function

$$\gamma : \mathbb{C} \setminus \{\operatorname{supp}(\mu) \cup \{0\}\} \to \mathbb{C}$$
$$z = z(\tilde{m}) \mapsto -\frac{1}{\tilde{m}} \tag{3}$$

provides an injective mapping between points outside the support of μ and points outside the support of ν with the property that

$$\gamma(\mathbb{C} \setminus \mathbb{R}) \subset \mathbb{C} \setminus \mathbb{R} \quad \text{and} \quad \gamma(\mathbb{R} \setminus \operatorname{supp}(\mu)) \subset \mathbb{R} \setminus \operatorname{supp}(\nu)$$

but where the inclusion is strict in general.

» This, as we shall see, will play a crucial role in **statistical inference** (from the empirical observation μ to "infer" ν).

Change of variable on the complex plane

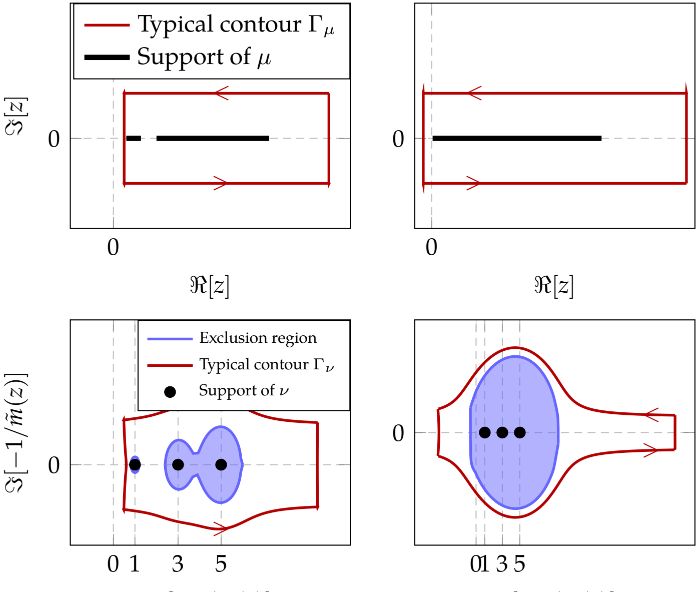
- An important consequence of the study above of $z(\cdot)$ (and its restriction $x(\cdot)$ to the real axis) is that the function

$$\begin{aligned} \gamma : \mathbb{C} \setminus \{\mathrm{supp}(\mu) \cup \{0\}\} &\to \mathbb{C} \\ z = z(\tilde{m}) &\mapsto -\frac{1}{\tilde{m}} \end{aligned} \tag{3}$$

  provides an injective mapping between points outside the support of μ and points outside the support of ν with the property that

$$\gamma(\mathbb{C} \setminus \mathbb{R}) \subset \mathbb{C} \setminus \mathbb{R} \quad \text{and} \quad \gamma(\mathbb{R} \setminus \mathrm{supp}(\mu)) \subset \mathbb{R} \setminus \mathrm{supp}(\nu)$$

  but where the inclusion is strict in general.
- This, as we shall see, will play a crucial role in **statistical inference** (from the empirical observation μ to "infer" ν).


Typical contour Γ_μ
Support of μ
$\Im[z]$
0
$\Re[z]$
Exclusion region
Typical contour Γ_ν
Support of ν
$\Im[-1/\tilde{m}(z)]$
0 1 3 5
01 3 5

Outline

Generalized MP for SCM

Spectrum characterization

No eigenvalue outside the support

Statistical inference

What can we say about the largest eigenvalue of SCM?

Let $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z} \in \mathbb{R}^{p\times n}$ with nonnegative definite $\mathbf{C} \in \mathbb{R}^{p\times p}$, $\mathbf{Z} \in \mathbb{R}^{p\times n}$ having independent zero mean and unit variance entries. Then, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, if the empirical spectral measure of $\mathbf{C}$ converges $\mu_{\mathbf{C}} \to \nu$ as $p \to \infty$, then $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}} \to \mu$, $\mu_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}} \to \tilde{\mu}$ where $\mu, \tilde{\mu}$ admitting Stieltjes transforms $m(z)$ and $\tilde{m}(z)$ such that $m(z) = \frac{1}{c}\tilde{m}(z) + \frac{1-c}{cz}$, $\quad \tilde{m}(z) = \left(-z + c\int \frac{t\nu(dt)}{1+\tilde{m}(z)t}\right)^{-1}$.

Generalized Marčenko–Pastur law

- » **weak convergences** for the *normalized* counting measure $\frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T})}$
- » by definition: for every continuous bounded f,

$$\frac{1}{p}\sum_{i=1}^{p} f\left(\lambda_i\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\right) - \int f(t)\mu(dt) \to 0.$$

What about the largest eigenvalue?

$$\frac{1}{p}\sum_{i=1}^{p} f\left(\lambda_i\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\right) - \int f(t)\mu(dt) \to 0.$$

» let f be a smoothed version of the indicator $1_{[a,b]}$ for $a, b \in \operatorname{supp}(\mu)$, only says that the averaged number of eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$ within $[a, b]$ converges to $\mu([a, b])$;

» in fact, only guarantees that, for $[a, b]$ a connected component of $\mathbb{R} \setminus \operatorname{supp}(\mu)$, the number of eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$ inside $[a, b]$ is asymptotically of **order** $o(p)$;

» $[a, b]$ may never be empty, even for arbitrarily large n, p (it can contain a fixed finite number of eigenvalues or even a growing number of eigenvalues, so long that this number is much less than $O(p)$).

» in particular, does not prevent a few eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$ from “leaking” from the limiting support of μ, which, e.g., may cause problems in statistical inference.

No eigenvalue outside the support

Let $\|\mathbf{C}\|$ be bounded with $\mu_{\mathbf{C}} \to \nu$ and $\boxed{\max_{1\le i\le p} \operatorname{dist}(\lambda_i(\mathbf{C}), \operatorname{supp}(\nu)) \to 0, \text{ as } p \to \infty}$. Consider $-\infty \le a < b \le \infty$ such that $a, b \in \mathbb{R}^+ \setminus \operatorname{supp}(\mu)$. Then,

- if $\mathbb{E}[|\mathbf{Z}_{ij}|^4] < \infty$, then, for $|\mathcal{A}|$ the cardinality of set $\mathcal{A}$ and $\gamma(\cdot)$ the change-of-variable function,

$$\left| \left\{ \lambda_i \left(\frac{1}{n} \mathbf{X}\mathbf{X}^\mathsf{T} \right) \in [a, b] \right\} \right| - |\{\lambda_i(\mathbf{C}) \in [\gamma(a), \gamma(b)]\}| \xrightarrow{a.s.} 0 \tag{4}$$

  If $[a, b]$ is a connected component of $\mathbb{R}^+ \setminus \operatorname{supp}(\mu)$, then $\left|\left\{ \lambda_i \left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} \right) \in [a, b] \right\}\right| \xrightarrow{a.s.} 0$.

- if $\mathbb{E}[\mathbf{Z}_{ij}^4] = \infty$, then $\max_{1\le i\le p} \lambda_i \left(\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} \right) \xrightarrow{a.s.} \infty$.

No eigenvalue outside the support

Outline

Generalized MP for SCM

Spectrum characterization

No eigenvalue outside the support

Statistical inference

Inference of the LSS of population covariance

Estimate the population linear eigenvalue statistics of the form $\frac{1}{p}\sum_{i=1}^{p} f(\lambda_i(\mathbf{C}))$ from sample observations $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}}\mathbf{z}_i$ and $\mathbf{z}_i$ with standard i.i.d. entries.

LSS Inference

Needs to "invert" the following characterization of SCM $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ from $\mathbf{C}$,

$$\boxed{\tilde{m}(z) = \left(-z + c\int \frac{t\nu(dt)}{1+\tilde{m}(z)t}\right)^{-1}} \tag{5}$$

with ν the limiting spectral measure of $\mathbf{C}$, equivalent to

$$m_\nu\left(-\frac{1}{\tilde{m}(z)}\right) = -zm(z)\tilde{m}(z). \tag{6}$$

Eigen-inference

For $f : \mathbb{C} \to \mathbb{C}$ analytic in a neighborhood of the eigenvalues of $\mathbf{C}$, by Cauchy's integral formula, the LSS $\frac{1}{p}\sum_{i=1}^{p} f(\lambda_i(\mathbf{C}))$ of population covariance $\mathbf{C}$ writes

$$\frac{1}{p}\sum_{i=1}^{p} f(\lambda_i(\mathbf{C})) \simeq \int f(t)\nu(dt) = \int \left[\frac{1}{2\pi\imath}\oint_{\Gamma_\nu} \frac{f(z)\,dz}{z-t}\right]\nu(dt)$$

$$= -\frac{1}{2\pi\imath}\oint_{\Gamma_\nu} f(z)\left[\int \frac{\nu(dt)}{t-z}\right]dz = -\frac{1}{2\pi\imath}\oint_{\Gamma_\nu} f(z) m_\nu(z)\,dz$$

where $\Gamma_\nu \subset \mathbb{C}$ is a positive contour encircling the support of ν but no singularity of f, which we would like to further relate to the observable $\tilde{m}(z)$ using the (asymptotic) relation $m_\nu\left(-\frac{1}{\tilde{m}(z)}\right) = -zm(z)\tilde{m}(z)$ via the change of variable $z \mapsto -1/\tilde{m}(z)$.

Is the change of variable $z \mapsto -1/\tilde{m}(z)$ allowed throughout the Cauchy's integral? Only possible if there exists a $\Gamma_\nu \subset \mathbb{C}$ such that $\Gamma_\nu = -1/\tilde{m}(\Gamma_\mu)$ for some well defined Γ_μ.

Licit change of variable?

Assume Γ_ν is indeed well defined as $\Gamma_\nu = -1/\tilde{m}(\Gamma_\mu)$ for some valid Γ_μ. Then,

$$\begin{aligned}\int f(t)\nu(dt) &= -\frac{1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right) m_\nu\left(-\frac{1}{\tilde{m}(\omega)}\right)\frac{\tilde{m}'(\omega)}{\tilde{m}^2(\omega)}d\omega \\ &= \frac{1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right)\omega\frac{m(\omega)\tilde{m}'(\omega)}{\tilde{m}(\omega)}d\omega\end{aligned}$$

where we wrote $z = -1/\tilde{m}(\omega)$. With $m(\omega) = \frac{1}{c}\tilde{m}(\omega) + (1-c)/(c\omega)$,

$$\begin{aligned}\int f(t)\nu(dt) &= \frac{1}{2c\pi\imath}\oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right)\frac{(\omega\tilde{m}(\omega) + (1-c))\,\tilde{m}'(\omega)}{\tilde{m}(\omega)}d\omega \\ &= \frac{1}{2c\pi\imath}\oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right)\omega\tilde{m}'(\omega)d\omega - \frac{1-c}{c}f(0)\cdot 1_{\{0\in\Gamma_\nu^\circ\}}\end{aligned}$$

For $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$ with $\mathbf{Z} \in \mathbb{R}^{p\times n}$ having i.i.d. entries with $\mathbb{E}[|\mathbf{Z}_{ij}|^4] < \infty$ and $\max_{1\le i\le p} \operatorname{dist}(\lambda_i(\mathbf{C}), \operatorname{supp}(\nu)) \to 0$, let $f : \mathbb{C} \to \mathbb{C}$ be a complex function analytic on the complement of $\gamma(\mathbb{C} \setminus \operatorname{supp}(\mu))$ in $\mathbb{C}$ with γ defined in (3). Then,

$$\frac{1}{p}\sum_{i=1}^{p} f(\lambda_i(\mathbf{C})) - \frac{1}{2c\pi\imath}\oint_{\Gamma_\mu} f\left(\frac{-1}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}\right)\omega m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)d\omega \xrightarrow{a.s.} 0,$$

for some complex positively oriented contour $\Gamma_\mu \subset \mathbb{C}$ surrounding $\operatorname{supp}(\mu) \setminus \{0\}$. In particular, if $c < 1$, the result holds for any f analytic on $\{z \in \mathbb{C},\ \Re[z] > 0\}$ with Γ_μ chosen as any such contour within $\{z \in \mathbb{C},\ \Re[z] > 0\}$.

Theorem: LSS inference

To estimate population eigenvalues of large multiplicity, use $f(z) = z$ and change Γ_μ into $\Gamma_\mu^{(a)}$, a contour circling around the a-th connected component of $\operatorname{supp}(\mu)$ only.

Application: estimate population eigenvalues of large multiplicity

Consider then the following setting of SCM inference,

$$\nu_{\mathbf{C}} = \frac{1}{p} \sum_{i=1}^{k} p_i \delta_{\ell_i} \to \sum_{i=1}^{k} c_i \delta_{\ell_i}$$

for $\ell_1 > \ldots > \ell_k > 0$, k fixed with respect to n, p, and $p_i/p \to c_i > 0$ as $p \to \infty$ (i.e., each eigenvalue has a large multiplicity of order $O(p)$). Consider the fully separable case and each eigenvalue of $\mathbf{C}$ is "mapped" to a single connected component of $\operatorname{supp}(\mu)$, then

$$\ell_a - \hat{\ell}_a \xrightarrow{a.s.} 0, \quad \hat{\ell}_a = -\frac{n}{p_a} \frac{1}{2\pi\imath} \oint_{\Gamma_\mu^{(a)}} \omega \frac{m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)} d\omega. \tag{7}$$

$$\ell_a - \hat{\ell}_a \xrightarrow{a.s.} 0, \quad \hat{\ell}_a = -\frac{n}{p_a} \frac{1}{2\pi\imath} \oint_{\Gamma_\mu^{(a)}} \omega \frac{m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)} d\omega. \tag{8}$$

- $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)$ (and its derivative) are rational functions, leads to simple residue calculus.
- the integrand in $\hat{\ell}_a$ has two types of poles: (i) the $\lambda_i = \lambda_i(\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X})$ falling inside the surface described by $\Gamma_\mu^{(a)}$, since in the neighborhood of λ_i,

$$-\frac{n}{p_a}\omega \frac{m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)} = -\frac{n}{p_a}\omega \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{(\lambda_i-\omega)^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i-\omega}} \sim_{\omega\sim\lambda_i} -\frac{n}{p_a}\frac{\omega}{\lambda_i-\omega}$$

  and (ii) the zeros of $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}$ falling within $\Gamma_\mu^{(a)}$.
- sort the eigenvalues of $\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}$ as $\lambda_1 \geq \ldots \geq \lambda_n$, the first type of poles is easy: the λ_i falling within $\Gamma_\mu^{(1)}$ are precisely the p_1 largest, within $\Gamma_\mu^{(2)}$ the next p_2 largest, etc.,

$$\lim_{\omega\to\lambda_i} (\omega-\lambda_i)\frac{n}{p_a}\frac{-\omega}{\lambda_i-\omega} = \frac{n}{p_a}\lambda_i.$$

$$\ell_a - \hat{\ell}_a \xrightarrow{a.s.} 0, \quad \hat{\ell}_a = -\frac{n}{p_a}\frac{1}{2\pi\imath}\oint_{\Gamma_\mu^{(a)}} \omega \frac{m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)} d\omega. \tag{8}$$

» $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)$ (and its derivative) are rational functions, leads to simple residue calculus.

» the integrand in $\hat{\ell}_a$ has two types of poles: (i) the $\lambda_i = \lambda_i(\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X})$ falling inside the surface described by $\Gamma_\mu^{(a)}$, since in the neighborhood of λ_i,

$$-\frac{n}{p_a}\omega \frac{m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)} = -\frac{n}{p_a}\omega \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{(\lambda_i-\omega)^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i-\omega}} \sim_{\omega\sim\lambda_i} -\frac{n}{p_a}\frac{\omega}{\lambda_i-\omega}$$

and (ii) the zeros of $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}$ falling within $\Gamma_\mu^{(a)}$.

» sort the eigenvalues of $\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}$ as $\lambda_1 \geq \ldots \geq \lambda_n$, the first type of poles is easy: the λ_i falling within $\Gamma_\mu^{(1)}$ are precisely the p_1 largest, within $\Gamma_\mu^{(2)}$ the next p_2 largest, etc.,

$$\lim_{\omega\to\lambda_i} (\omega-\lambda_i)\frac{n}{p_a}\frac{-\omega}{\lambda_i-\omega} = \frac{n}{p_a}\lambda_i.$$

$$\ell_a - \hat{\ell}_a \xrightarrow{a.s.} 0, \quad \hat{\ell}_a = -\frac{n}{p_a} \frac{1}{2\pi\imath} \oint_{\Gamma_\mu^{(a)}} \omega \frac{m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)} d\omega. \tag{8}$$

- $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)$ (and its derivative) are rational functions, leads to simple residue calculus.
- the integrand in $\hat{\ell}_a$ has two types of poles: (i) the $\lambda_i = \lambda_i(\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X})$ falling inside the surface described by $\Gamma_\mu^{(a)}$, since in the neighborhood of λ_i,

$$-\frac{n}{p_a}\omega \frac{m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)} = -\frac{n}{p_a}\omega \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{(\lambda_i-\omega)^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i-\omega}} \sim_{\omega\sim\lambda_i} -\frac{n}{p_a}\frac{\omega}{\lambda_i-\omega}$$

  and (ii) the zeros of $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}$ falling within $\Gamma_\mu^{(a)}$.
- sort the eigenvalues of $\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}$ as $\lambda_1 \geq \ldots \geq \lambda_n$, the first type of poles is easy: the λ_i falling within $\Gamma_\mu^{(1)}$ are precisely the p_1 largest, within $\Gamma_\mu^{(2)}$ the next p_2 largest, etc.,

$$\lim_{\omega\to\lambda_i} (\omega - \lambda_i)\frac{n}{p_a}\frac{-\omega}{\lambda_i - \omega} = \frac{n}{p_a}\lambda_i.$$

» The second set of poles is less immediate to retrieve.

» Remark that the zeros η_j (sorted as $\eta_1 \geq \eta_2 \geq \ldots$) of $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)$ are real and satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - \eta_j} = 0.$$

» Since the function $x \mapsto \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - x}$ is increasing and has ∞ and $-\infty$ asymptotes at $x = \lambda_i - 0$ and $x = \lambda_i + 0$, respectively, each η_j falls exactly in one of the intervals $[\lambda_i, \lambda_{i+1}]$ and thus each λ_i pole is accompanied by its η_i pole (if sorted similarly). The residue calculus then gives, by Taylor expanding the denominator,

$$\lim_{\omega \to \eta_j}(\omega - \eta_j)\frac{n}{p_a}\frac{-\omega m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{0 + m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\eta_j)(\omega - \eta_j)} = -\frac{n}{p_a}\eta_j.$$

» we finally have the estimator $\hat{\ell}_a = \frac{n}{p_a}\sum_{i=p_1+\ldots+p_{a-1}+1}^{p_1+\ldots+p_a}\lambda_i - \eta_i$.

- The second set of poles is less immediate to retrieve.
- Remark that the zeros η_j (sorted as $\eta_1 \geq \eta_2 \geq \ldots$) of $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)$ are real and satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - \eta_j} = 0.$$

- Since the function $x \mapsto \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - x}$ is increasing and has ∞ and $-\infty$ asymptotes at $x = \lambda_i - 0$ and $x = \lambda_i + 0$, respectively, each η_j falls exactly in one of the intervals $[\lambda_i, \lambda_{i+1}]$ and thus each λ_i pole is accompanied by its η_i pole (if sorted similarly). The residue calculus then gives, by Taylor expanding the denominator,

$$\lim_{\omega \to \eta_j}(\omega - \eta_j)\frac{n}{p_a}\frac{-\omega m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{0 + m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\eta_j)(\omega - \eta_j)} = -\frac{n}{p_a}\eta_j.$$

- we finally have the estimator $\hat{\ell}_a = \frac{n}{p_a}\sum_{i=p_1+\ldots+p_{a-1}+1}^{p_1+\ldots+p_a}\lambda_i - \eta_i$.

- The second set of poles is less immediate to retrieve.
- Remark that the zeros η_j (sorted as $\eta_1 \geq \eta_2 \geq \ldots$) of $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)$ are real and satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - \eta_j} = 0.$$

- Since the function $x \mapsto \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - x}$ is increasing and has ∞ and $-\infty$ asymptotes at $x = \lambda_i - 0$ and $x = \lambda_i + 0$, respectively, each η_j falls exactly in one of the intervals $[\lambda_i, \lambda_{i+1}]$ and thus each λ_i pole is accompanied by its η_i pole (if sorted similarly). The residue calculus then gives, by Taylor expanding the denominator,

$$\lim_{\omega\to\eta_j}(\omega - \eta_j)\frac{n}{p_a}\frac{-\omega m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{0 + m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\eta_j)(\omega - \eta_j)} = -\frac{n}{p_a}\eta_j.$$

- we finally have the estimator $\hat{\ell}_a = \frac{n}{p_a}\sum_{i=p_1+\ldots+p_{a-1}+1}^{p_1+\ldots+p_a}\lambda_i - \eta_i$.

» The second set of poles is less immediate to retrieve.

» Remark that the zeros η_j (sorted as $\eta_1 \geq \eta_2 \geq \ldots$) of $m_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)$ are real and satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - \eta_j} = 0.$$

» Since the function $x \mapsto \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - x}$ is increasing and has ∞ and $-\infty$ asymptotes at $x = \lambda_i - 0$ and $x = \lambda_i + 0$, respectively, each η_j falls exactly in one of the intervals $[\lambda_i, \lambda_{i+1}]$ and thus each λ_i pole is accompanied by its η_i pole (if sorted similarly). The residue calculus then gives, by Taylor expanding the denominator,

$$\lim_{\omega \to \eta_j} (\omega - \eta_j)\frac{n}{p_a}\frac{-\omega m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\omega)}{0 + m'_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}}(\eta_j)(\omega - \eta_j)} = -\frac{n}{p_a}\eta_j.$$

» we finally have the estimator $\hat{\ell}_a = \frac{n}{p_a}\sum_{i=p_1+\ldots+p_{a-1}+1}^{p_1+\ldots+p_a} \lambda_i - \eta_i$.

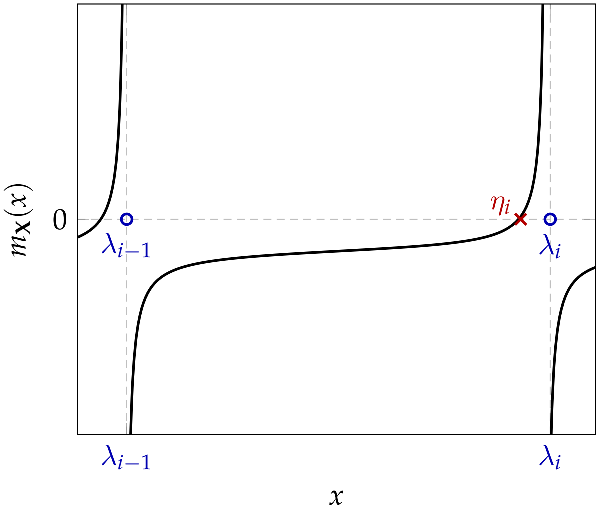


Figure: Illustration of the zeros (η_i) and poles (λ_i) of the (restriction to the real axis of the) Stieltjes transform $m_{\mathbf{X}}(x)$.

For $\mathbf{X} \in \mathbb{R}^{n\times n}$ symmetric with eigenvalues $\lambda_1 > \ldots > \lambda_n$, the zeros $\eta_1 > \eta_2 > \ldots$ of $m_{\mathbf{X}}(z)$ satisfy the following equivalence relations

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - \eta_j} = 0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}\frac{-\eta_j}{\lambda_i - \eta_j} = 0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda_i}{\lambda_i - \eta_j} - 1 = 0$$

$$\Leftrightarrow \frac{1}{n}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}(\boldsymbol{\Lambda} - \eta_j\mathbf{I}_n)^{-1}\sqrt{\boldsymbol{\lambda}} - 1 = 0 = \det\left(\frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}(\boldsymbol{\Lambda} - \eta_j\mathbf{I}_n)^{-1} - \mathbf{I}_n\right)$$

$$\Leftrightarrow \det\left(\frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}} - \boldsymbol{\Lambda} + \eta_j\mathbf{I}_n\right) = 0$$

where we denoted $\sqrt{\boldsymbol{\lambda}} \in \mathbb{R}^p$ the vector of $\sqrt{\lambda_i}$'s and $\boldsymbol{\Lambda} \equiv \operatorname{diag}\{\lambda_i\}_{i=1}^{p}$, and used the fact that $\det(\boldsymbol{\Lambda} - \eta_j\mathbf{I}_n) \neq 0$.The zeros of $m_{\mathbf{X}}$ are exactly the eigenvalues of

$$\boldsymbol{\Lambda} - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}.$$

Explicit expression for the zeros of $m_{\mathbf{X}}(z)$

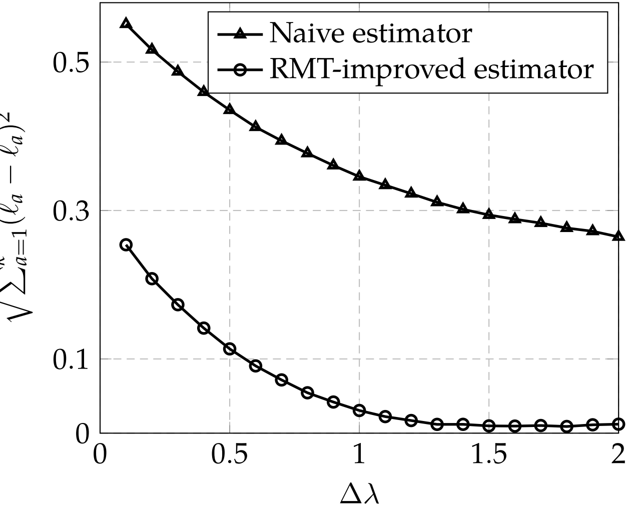


Figure: Eigenvalue estimation errors with naive and RMT-improved approach, as a function of $\Delta\lambda$, for $\ell_1 = 1$, $\ell_2 = 1 + \Delta\lambda$, $p = 256$ and $n = 1\,024$. Results averaged over 30 runs.