# Probability and Stochastic Process II: Random Matrix Theory and Applications
# Lecture 2: From Random Scalars to Random Matrices

**Zhenyu Liao**, Tiebin Mi, Caiming Qiu

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

March 1, 2023

# Outline

# What we will have today

» reminder on Law of Large Numbers (LLN) and Central Limit Theorem (CLT)

» from random scalars to random vectors and matrices

» RMT basic concepts: resolvent, spectral measure, and Stieltjes transform

» deterministic equivalent framework to RMT

# What we will have today

» reminder on Law of Large Numbers (LLN) and Central Limit Theorem (CLT)
» from random scalars to random vectors and matrices
» RMT basic concepts: resolvent, spectral measure, and Stieltjes transform
» deterministic equivalent framework to RMT

# What we will have today

- reminder on Law of Large Numbers (LLN) and Central Limit Theorem (CLT)
- from random scalars to random vectors and matrices
- RMT basic concepts: resolvent, spectral measure, and Stieltjes transform
- deterministic equivalent framework to RMT

# What we will have today

- reminder on Law of Large Numbers (LLN) and Central Limit Theorem (CLT)
- from random scalars to random vectors and matrices
- RMT basic concepts: resolvent, spectral measure, and Stieltjes transform
- deterministic equivalent framework to RMT

# Outline

# LLN and CLT

» (Strong) law of large numbers (LLN): for a sequence of i.i.d. random variables $x_1, \ldots, x_p$ with the same expectation $\mathbb{E}[x_i] = \mu$, we have

$$\frac{1}{p}\sum_{i=1}^{p} x_i \to \mu, \tag{1}$$

almost surely as $p \to \infty$.

» Central limit theorem (CLT, Lindeberg–Lévy tyep): for a sequence of i.i.d. random variables $x_1, \ldots, x_p$ with the same expectation $\mathbb{E}[x_i] = \mu$ and variance $\operatorname{Var}[x_i] = \sigma^2 < \infty$, we have

$$\sqrt{p}\left(\frac{1}{p}\sum_{i=1}^{p}(x_i - \mu)\right) \to \mathcal{N}(0, \sigma^2), \tag{2}$$

in distribution as $p \to \infty$.

# LLN and CLT

- (Strong) law of large numbers (LLN): for a sequence of i.i.d. random variables $x_1, \ldots, x_p$ with the same expectation $\mathbb{E}[x_i] = \mu$, we have

$$\frac{1}{p} \sum_{i=1}^{p} x_i \to \mu, \tag{1}$$

almost surely as $p \to \infty$.

- Central limit theorem (CLT, Lindeberg–Lévy tyep): for a sequence of i.i.d. random variables $x_1, \ldots, x_p$ with the same expectation $\mathbb{E}[x_i] = \mu$ and variance $\mathrm{Var}[x_i] = \sigma^2 < \infty$, we have

$$\sqrt{p} \left( \frac{1}{p} \sum_{i=1}^{p} (x_i - \mu) \right) \to \mathcal{N}(0, \sigma^2), \tag{2}$$

in distribution as $p \to \infty$.

# Outline

LLN and CLT

From Random Scalars to Random Matrices

RMT Basis

# OK with LLN and CLT, so what?

Different view of LLN and CLT: large-dimensional *deterministic* behavior and fluctuation.

**Single scalar random variables**

- *Scalar* random variable $x \in \mathbb{R}$, characterize its behavior distribution/law, characteristic function and/or successive moments, etc.
- $x$ in general *not* expected to establish some kind of "close-to-deterministic" behavior.
- True for a *single observation*, although certainly the sum of many such random variables may concentrate and exhibit a close-to-deterministic behavior.

**Random vectors: many scalar random variables**

Consider a set of size $p$ i.i.d. realizations/copies of such random variable. As a random vector $\mathbf{x} = [x_1, \ldots, x_p]^\top \in \mathbb{R}^p$, with $\mathbb{E}[x_i] = \mu$, $\mathrm{Var}[x_i] = 1$, $i \in \{1, \ldots, p\}$.

- as $p$ independent *scalar* random variables $x \in \mathbb{R}$; or
- as a single realization of a *random vector* $\mathbf{x} \in \mathbb{R}^p$, having independent entries.

# OK with LLN and CLT, so what?

Different view of LLN and CLT: large-dimensional *deterministic* behavior and fluctuation.

**Single scalar random variables**

- *Scalar* random variable $x \in \mathbb{R}$, characterize its behavior distribution/law, characteristic function and/or successive moments, etc.
- $x$ in general *not* expected to establish some kind of "close-to-deterministic" behavior.
- True for a *single observation*, although certainly the sum of many such random variables may concentrate and exhibit a close-to-deterministic behavior.

**Random vectors: many scalar random variables**

Consider a set of size $p$ i.i.d. realizations/copies of such random variable. As a random vector $\mathbf{x} = [x_1, \ldots, x_p]^\mathsf{T} \in \mathbb{R}^p$, with $\mathbb{E}[x_i] = \mu$, $\mathrm{Var}[x_i] = 1$, $i \in \{1, \ldots, p\}$.

- as $p$ independent *scalar* random variables $x \in \mathbb{R}$; or
- as a single realization of a *random vector* $\mathbf{x} \in \mathbb{R}^p$, having independent entries.

# OK with LLN and CLT, so what?

Different view of LLN and CLT: large-dimensional *deterministic* behavior and fluctuation.

**Single scalar random variables**

» *Scalar* random variable $x \in \mathbb{R}$, characterize its behavior distribution/law, characteristic function and/or successive moments, etc.

» $x$ in general *not* expected to establish some kind of "close-to-deterministic" behavior.

» True for a *single observation*, although certainly the sum of many such random variables may concentrate and exhibit a close-to-deterministic behavior.

**Random vectors: many scalar random variables**

Consider a set of size $p$ i.i.d. realizations/copies of such random variable. As a random vector $\mathbf{x} = [x_1, \ldots, x_p]^\top \in \mathbb{R}^p$, with $\mathbb{E}[x_i] = \mu$, $\mathrm{Var}[x_i] = 1$, $i \in \{1, \ldots, p\}$.

» as $p$ independent *scalar* random variables $x \in \mathbb{R}$; or

» as a single realization of a *random vector* $\mathbf{x} \in \mathbb{R}^p$, having independent entries.

# OK with LLN and CLT, so what?

Different view of LLN and CLT: large-dimensional *deterministic* behavior and fluctuation.

**Single scalar random variables**

- *Scalar* random variable $x \in \mathbb{R}$, characterize its behavior distribution/law, characteristic function and/or successive moments, etc.
- $x$ in general *not* expected to establish some kind of "close-to-deterministic" behavior.
- True for a *single observation*, although certainly the sum of many such random variables may concentrate and exhibit a close-to-deterministic behavior.

**Random vectors: many scalar random variables**

Consider a set of size $p$ i.i.d. realizations/copies of such random variable. As a random vector $\mathbf{x} = [x_1, \ldots, x_p]^\top \in \mathbb{R}^p$, with $\mathbb{E}[x_i] = \mu$, $\operatorname{Var}[x_i] = 1$, $i \in \{1, \ldots, p\}$.

- as $p$ independent *scalar* random variables $x \in \mathbb{R}$; or
- as a single realization of a *random vector* $\mathbf{x} \in \mathbb{R}^p$, having independent entries.

# OK with LLN and CLT, so what?

Different view of LLN and CLT: large-dimensional *deterministic* behavior and fluctuation.

**Single scalar random variables**

» *Scalar* random variable $x \in \mathbb{R}$, characterize its behavior distribution/law, characteristic function and/or successive moments, etc.

» $x$ in general *not* expected to establish some kind of "close-to-deterministic" behavior.

» True for a *single observation*, although certainly the sum of many such random variables may concentrate and exhibit a close-to-deterministic behavior.

**Random vectors: many scalar random variables**

Consider a set of size $p$ i.i.d. realizations/copies of such random variable. As a random vector $\mathbf{x} = [x_1, \ldots, x_p]^\mathsf{T} \in \mathbb{R}^p$, with $\mathbb{E}[x_i] = \mu$, $\mathrm{Var}[x_i] = 1$, $i \in \{1, \ldots, p\}$.

» as $p$ independent *scalar* random variables $x \in \mathbb{R}$; or

» as a single realization of a *random vector* $\mathbf{x} \in \mathbb{R}^p$, having independent entries.

# OK with LLN and CLT, so what?

Different view of LLN and CLT: large-dimensional *deterministic* behavior and fluctuation.

**Single scalar random variables**

» *Scalar* random variable $x \in \mathbb{R}$, characterize its behavior distribution/law, characteristic function and/or successive moments, etc.

» $x$ in general *not* expected to establish some kind of "close-to-deterministic" behavior.

» True for a *single observation*, although certainly the sum of many such random variables may concentrate and exhibit a close-to-deterministic behavior.

**Random vectors: many scalar random variables**

Consider a set of size $p$ i.i.d. realizations/copies of such random variable. As a random vector $\mathbf{x} = [x_1, \ldots, x_p]^\mathsf{T} \in \mathbb{R}^p$, with $\mathbb{E}[x_i] = \mu$, $\mathrm{Var}[x_i] = 1$, $i \in \{1, \ldots, p\}$.

» as $p$ independent *scalar* random variables $x \in \mathbb{R}$; or

» as a single realization of a *random vector* $\mathbf{x} \in \mathbb{R}^p$, having independent entries.

# OK with LLN and CLT, so what?

Different view of LLN and CLT: large-dimensional *deterministic* behavior and fluctuation.
**Single scalar random variables**

- *Scalar* random variable $x \in \mathbb{R}$, characterize its behavior distribution/law, characteristic function and/or successive moments, etc.
- $x$ in general *not* expected to establish some kind of "close-to-deterministic" behavior.
- True for a *single observation*, although certainly the sum of many such random variables may concentrate and exhibit a close-to-deterministic behavior.

**Random vectors: many scalar random variables**
Consider a set of size $p$ i.i.d. realizations/copies of such random variable. As a random vector $\mathbf{x} = [x_1, \ldots, x_p]^\mathsf{T} \in \mathbb{R}^p$, with $\mathbb{E}[x_i] = \mu$, $\mathrm{Var}[x_i] = 1$, $i \in \{1, \ldots, p\}$.

- as $p$ independent *scalar* random variables $x \in \mathbb{R}$; or
- as a single realization of a *random vector* $\mathbf{x} \in \mathbb{R}^p$, having independent entries.

# OK with LLN and CLT, so what?

(i) **Scalar**: nothing more can be said about each *individual* random variable:

- inappropriate to predict the behavior of $x_i$ with any *deterministic* value
- in general *incorrect* to say "the random $x_i$ is close to $\mu = \mathbb{E}[x_i]$", since, for $x_i$ with $\mathbb{E}[x] = \mu$ and $\mathrm{Var}[x] = 1$, by Chebyshev's inequality.

$$\mathbb{P}(|x - \mu| \geq t) \leq t^{-2}, \quad \forall t > 0. \tag{3}$$

- random fluctuation $x_i - \mathbb{E}[x_i]$ can be as large as $\mu = \mathbb{E}[x_i]$.

(ii) **Vector**: a different picture: single realization of random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$.

- cannot say anything in general about each individual vector $\mathbf{x}$.
- however, if we are interested in only the (scalar and linear) observations of the random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$ (with $\mathbb{E}[\mathbf{x}] = \mu \mathbf{1}_p/\sqrt{p}$), we known much more:

$$\frac{1}{p}\mathbf{x}^\top \mathbf{1}_p \xrightarrow{a.s.} \mathbb{E}[x_i] = \mu, \quad \frac{1}{\sqrt{p}}(\mathbf{x} - \mu \mathbf{1}_p)^\top \mathbf{1}_p \xrightarrow{d} \mathcal{N}(0,1), \quad p \to \infty. \tag{4}$$

# OK with LLN and CLT, so what?

(i) **Scalar**: nothing more can be said about each *individual* random variable:
- inappropriate to predict the behavior of $x_i$ with any *deterministic* value
- in general *incorrect* to say "the random $x_i$ is close to $\mu = \mathbb{E}[x_i]$", since, for $x_i$ with $\mathbb{E}[x] = \mu$ and $\mathrm{Var}[x] = 1$, by Chebyshev's inequality.

$$\mathbb{P}(|x - \mu| \geq t) \leq t^{-2}, \quad \forall t > 0. \tag{3}$$

- random fluctuation $x_i - \mathbb{E}[x_i]$ can be as large as $\mu = \mathbb{E}[x_i]$.

(ii) **Vector**: a different picture: single realization of random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$.
- cannot say anything in general about each individual vector $\mathbf{x}$.
- however, if we are interested in only the (scalar and linear) observations of the random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$ (with $\mathbb{E}[\mathbf{x}] = \mu \mathbf{1}_p/\sqrt{p}$), we known much more:

$$\frac{1}{p}\mathbf{x}^\top \mathbf{1}_p \xrightarrow{a.s.} \mathbb{E}[x_i] = \mu, \quad \frac{1}{\sqrt{p}}(\mathbf{x} - \mu \mathbf{1}_p)^\top \mathbf{1}_p \xrightarrow{d} \mathcal{N}(0,1), \quad p \to \infty. \tag{4}$$

# OK with LLN and CLT, so what?

(i) **Scalar**: nothing more can be said about each *individual* random variable:

- inappropriate to predict the behavior of $x_i$ with any *deterministic* value
- in general *incorrect* to say "the random $x_i$ is close to $\mu = \mathbb{E}[x_i]$", since, for $x_i$ with $\mathbb{E}[x] = \mu$ and $\operatorname{Var}[x] = 1$, by Chebyshev's inequality.

$$\mathbb{P}(|x - \mu| \geq t) \leq t^{-2}, \quad \forall t > 0. \tag{3}$$

- random fluctuation $x_i - \mathbb{E}[x_i]$ can be as large as $\mu = \mathbb{E}[x_i]$.

(ii) **Vector**: a different picture: single realization of random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$.

- cannot say anything in general about each individual vector $\mathbf{x}$.
- however, if we are interested in only the (scalar and linear) observations of the random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$ (with $\mathbb{E}[\mathbf{x}] = \mu \mathbf{1}_p/\sqrt{p}$), we known much more:

$$\frac{1}{p}\mathbf{x}^\top \mathbf{1}_p \xrightarrow{a.s.} \mathbb{E}[x_i] = \mu, \quad \frac{1}{\sqrt{p}}(\mathbf{x} - \mu \mathbf{1}_p)^\top \mathbf{1}_p \xrightarrow{d} \mathcal{N}(0,1), \quad p \to \infty. \tag{4}$$

# OK with LLN and CLT, so what?

(i) **Scalar**: nothing more can be said about each *individual* random variable:

- inappropriate to predict the behavior of $x_i$ with any *deterministic* value
- in general *incorrect* to say "the random $x_i$ is close to $\mu = \mathbb{E}[x_i]$", since, for $x_i$ with $\mathbb{E}[x] = \mu$ and $\mathrm{Var}[x] = 1$, by Chebyshev's inequality.

$$\mathbb{P}(|x - \mu| \geq t) \leq t^{-2}, \quad \forall t > 0. \tag{3}$$

- random fluctuation $x_i - \mathbb{E}[x_i]$ can be as large as $\mu = \mathbb{E}[x_i]$.

(ii) **Vector**: a different picture: single realization of random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$.

- cannot say anything in general about each individual vector $\mathbf{x}$.
- however, if we are interested in only the (scalar and linear) observations of the random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$ (with $\mathbb{E}[\mathbf{x}] = \mu \mathbf{1}_p/\sqrt{p}$), we known much more:

$$\frac{1}{p}\mathbf{x}^\top \mathbf{1}_p \xrightarrow{a.s.} \mathbb{E}[x_i] = \mu, \quad \frac{1}{\sqrt{p}}(\mathbf{x} - \mu \mathbf{1}_p)^\top \mathbf{1}_p \xrightarrow{d} \mathcal{N}(0,1), \quad p \to \infty. \tag{4}$$

# OK with LLN and CLT, so what?

(i) **Scalar**: nothing more can be said about each *individual* random variable:

- inappropriate to predict the behavior of $x_i$ with any *deterministic* value
- in general *incorrect* to say "the random $x_i$ is close to $\mu = \mathbb{E}[x_i]$", since, for $x_i$ with $\mathbb{E}[x] = \mu$ and $\operatorname{Var}[x] = 1$, by Chebyshev's inequality.

$$\mathbb{P}(|x - \mu| \geq t) \leq t^{-2}, \quad \forall t > 0. \tag{3}$$

- random fluctuation $x_i - \mathbb{E}[x_i]$ can be as large as $\mu = \mathbb{E}[x_i]$.

(ii) **Vector**: a different picture: single realization of random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$.

- cannot say anything in general about each individual vector $\mathbf{x}$.
- however, if we are interested in only the (scalar and linear) observations of the random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$ (with $\mathbb{E}[\mathbf{x}] = \mu \mathbf{1}_p/\sqrt{p}$), we known much more:

$$\frac{1}{p}\mathbf{x}^\top \mathbf{1}_p \xrightarrow{a.s.} \mathbb{E}[x_i] = \mu, \quad \frac{1}{\sqrt{p}}(\mathbf{x} - \mu \mathbf{1}_p)^\top \mathbf{1}_p \xrightarrow{d} \mathcal{N}(0,1), \quad p \to \infty. \tag{4}$$

# OK with LLN and CLT, so what?

(i) **Scalar**: nothing more can be said about each *individual* random variable:

- inappropriate to predict the behavior of $x_i$ with any *deterministic* value
- in general *incorrect* to say "the random $x_i$ is close to $\mu = \mathbb{E}[x_i]$", since, for $x_i$ with $\mathbb{E}[x] = \mu$ and $\mathrm{Var}[x] = 1$, by Chebyshev's inequality.

$$\mathbb{P}(|x - \mu| \geq t) \leq t^{-2}, \quad \forall t > 0. \tag{3}$$

- random fluctuation $x_i - \mathbb{E}[x_i]$ can be as large as $\mu = \mathbb{E}[x_i]$.

(ii) **Vector**: a different picture: single realization of random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$.

- cannot say anything in general about each individual vector $\mathbf{x}$.
- however, if we are interested in only the (scalar and linear) observations of the random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$ (with $\mathbb{E}[\mathbf{x}] = \mu \mathbf{1}_p/\sqrt{p}$), we known much more:

$$\frac{1}{p}\mathbf{x}^\top \mathbf{1}_p \xrightarrow{a.s.} \mathbb{E}[x_i] = \mu, \quad \frac{1}{\sqrt{p}}(\mathbf{x} - \mu\mathbf{1}_p)^\top \mathbf{1}_p \xrightarrow{d} \mathcal{N}(0,1), \quad p \to \infty. \tag{4}$$

# OK with LLN and CLT, so what?

(i) **Scalar**: nothing more can be said about each *individual* random variable:

- inappropriate to predict the behavior of $x_i$ with any *deterministic* value
- in general *incorrect* to say "the random $x_i$ is close to $\mu = \mathbb{E}[x_i]$", since, for $x_i$ with $\mathbb{E}[x] = \mu$ and $\mathrm{Var}[x] = 1$, by Chebyshev's inequality.

$$\mathbb{P}(|x - \mu| \geq t) \leq t^{-2}, \quad \forall t > 0. \tag{3}$$

- random fluctuation $x_i - \mathbb{E}[x_i]$ can be as large as $\mu = \mathbb{E}[x_i]$.

(ii) **Vector**: a different picture: single realization of random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$.

- cannot say anything in general about each individual vector $\mathbf{x}$.
- however, if we are interested in only the (scalar and linear) observations of the random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$ (with $\mathbb{E}[\mathbf{x}] = \mu \mathbf{1}_p/\sqrt{p}$), we known much more:

$$\frac{1}{p}\mathbf{x}^\mathsf{T}\mathbf{1}_p \xrightarrow{a.s.} \mathbb{E}[x_i] = \mu, \quad \frac{1}{\sqrt{p}}(\mathbf{x} - \mu\mathbf{1}_p)^\mathsf{T}\mathbf{1}_p \xrightarrow{d} \mathcal{N}(0,1), \quad p \to \infty. \tag{4}$$

# OK with LLN and CLT, so what?

This is

$$\frac{1}{p}\mathbf{x}^\mathsf{T}\mathbf{1}_p \simeq \underbrace{\mu}_{O(1)} + \underbrace{\frac{1}{\sqrt{p}}\mathcal{N}(0,1)}_{O(p^{-1/2})}.$$

» a large dimensional random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$, when "observed" via the linear map $\mathbf{1}_p^\mathsf{T}(\cdot)/\sqrt{p}$ of unit Euclidean norm (i.e., of "scale" independent of $p$);

» leads to $\mathbf{x}$ (when "observed" in this way) exhibiting the joint behavior of:

(i) approximately, in its first order, a *deterministic* quantity $\mu$; and

(ii) in its second-order, a **universal** Gaussian fluctuation that is **strongly concentrated** and **independent** of the specific law of $x_i$.

# OK with LLN and CLT, so what?

This is

$$\frac{1}{p}\mathbf{x}^\mathsf{T}\mathbf{1}_p \simeq \underbrace{\mu}_{O(1)} + \underbrace{\frac{1}{\sqrt{p}}\mathcal{N}(0,1)}_{O(p^{-1/2})}.$$

- » a large dimensional random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$, when "observed" via the linear map $\mathbf{1}_p^\mathsf{T}(\cdot)/\sqrt{p}$ of unit Euclidean norm (i.e., of "scale" independent of $p$);
- » leads to $\mathbf{x}$ (when "observed" in this way) exhibiting the joint behavior of:
  - (i) approximately, in its first order, a *deterministic* quantity $\mu$; and
  - (ii) in its second-order, a **universal** Gaussian fluctuation that is **strongly concentrated** and **independent** of the specific law of $x_i$.

# OK with LLN and CLT, so what?

This is

$$\frac{1}{p}\mathbf{x}^\mathsf{T}\mathbf{1}_p \simeq \underbrace{\mu}_{O(1)} + \underbrace{\frac{1}{\sqrt{p}}\mathcal{N}(0,1)}_{O(p^{-1/2})}.$$

- » a large dimensional random vector $\mathbf{x}/\sqrt{p} \in \mathbb{R}^p$, when "observed" via the linear map $\mathbf{1}_p^\mathsf{T}(\cdot)/\sqrt{p}$ of unit Euclidean norm (i.e., of "scale" independent of $p$);
- » leads to $\mathbf{x}$ (when "observed" in this way) exhibiting the joint behavior of:

  (i) approximately, in its first order, a *deterministic* quantity $\mu$; and

  (ii) in its second-order, a **universal** Gaussian fluctuation that is **strongly concentrated** and **independent** of the specific law of $x_i$.

# What about random matrices?

- As in the case of (high-dimensional) random vectors, we should **NOT** expect random matrices themselves converge in any useful sense;
- e.g., there does **NOT** exist deterministic matrix $\bar{\mathbf{X}}$ so that the random matrix $\mathbf{X} \in \mathbb{R}^{p\times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{5}$$

in spectral norm as $p \to \infty$ (in probability or almost surely);
- nonetheless, "properly scaled" scalar observations $f : \mathbb{R}^{p\times p} \to \mathbb{R}$ of $\mathbf{X}$ **DO** converge, and there exists deterministic $\bar{\mathbf{X}}$ such that

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{6}$$

as $p \to \infty$. We say such $\bar{\mathbf{X}}$ is a **deterministic equivalent** of the random matrix $\mathbf{X}$.
- observation $f$ of interest in RMT include (empirical) eigenvalue distribution/measure, linear eigenvalue statistics, specific eigenvalue location, projection of eigenvectors, etc.

# What about random matrices?

- As in the case of (high-dimensional) random vectors, we should **NOT** expect random matrices themselves converge in any useful sense;
- e.g., there does **NOT** exist deterministic matrix $\bar{\mathbf{X}}$ so that the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{5}$$

  in spectral norm as $p \to \infty$ (in probability or almost surely);
- nonetheless, "properly scaled" scalar observations $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ of $\mathbf{X}$ **DO** converge, and there exists deterministic $\bar{\mathbf{X}}$ such that

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{6}$$

  as $p \to \infty$. We say such $\bar{\mathbf{X}}$ is a **deterministic equivalent** of the random matrix $\mathbf{X}$.
- observation $f$ of interest in RMT include (empirical) eigenvalue distribution/measure, linear eigenvalue statistics, specific eigenvalue location, projection of eigenvectors, etc.

# What about random matrices?

- As in the case of (high-dimensional) random vectors, we should **NOT** expect random matrices themselves converge in any useful sense;
- e.g., there does **NOT** exist deterministic matrix $\bar{\mathbf{X}}$ so that the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{5}$$

  in spectral norm as $p \to \infty$ (in probability or almost surely);
- nonetheless, "properly scaled" scalar observations $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ of $\mathbf{X}$ **DO** converge, and there exists deterministic $\bar{\mathbf{X}}$ such that

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{6}$$

  as $p \to \infty$. We say such $\bar{\mathbf{X}}$ is a **deterministic equivalent** of the random matrix $\mathbf{X}$.
- observation $f$ of interest in RMT include (empirical) eigenvalue distribution/measure, linear eigenvalue statistics, specific eigenvalue location, projection of eigenvectors, etc.

# What about random matrices?

- As in the case of (high-dimensional) random vectors, we should **NOT** expect random matrices themselves converge in any useful sense;
- e.g., there does **NOT** exist deterministic matrix $\bar{\mathbf{X}}$ so that the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{5}$$

  in spectral norm as $p \to \infty$ (in probability or almost surely);
- nonetheless, "properly scaled" scalar observations $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ of $\mathbf{X}$ **DO** converge, and there exists deterministic $\bar{\mathbf{X}}$ such that

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{6}$$

  as $p \to \infty$. We say such $\bar{\mathbf{X}}$ is a **deterministic equivalent** of the random matrix $\mathbf{X}$.
- observation $f$ of interest in RMT include (empirical) eigenvalue distribution/measure, linear eigenvalue statistics, specific eigenvalue location, projection of eigenvectors, etc.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - **however**, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \operatorname{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\operatorname{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - however, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - however, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - however, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - however, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

» while the random matrix $\mathbf{X} \in \mathbb{R}^{p\times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);

» scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p\times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.

» So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - however, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.

» We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- » while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- » scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- » So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - **however**, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- » We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - **however**, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

- while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);
- scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \mathrm{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\mathrm{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.
- So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - **however**, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.
- We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Deterministic equivalent for RMT: intuition and proof

What is actually happening with scalar observations of random matrices and the deterministic equivalent (DE)?

» while the random matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ remains random as the dimension $p$ grows (in fact even "more" random due to the growing degrees of freedom);

» scalar observation $f(\mathbf{X})$ of $\mathbf{X}$ becomes "more concentrated" as $p \to \infty$;
  - the random $f(\mathbf{X})$, if concentrates, must concentrated around its expectation $\mathbb{E}[f(\mathbf{X})]$;
  - in fact, as $p \to \infty$, more randomness in $\mathbf{X} \Rightarrow \operatorname{Var}[f(\mathbf{X})] \downarrow 0$, e.g., $\operatorname{Var}[f(\mathbf{X})] = p^{-4}$;
  - if the functional $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ is linear, then $\mathbb{E}[f(\mathbf{X})] = f(\mathbb{E}[\mathbf{X}])$.

» So, to propose a DE, it suffices to evaluate $\mathbb{E}[\mathbf{X}]$:
  - **however**, $\mathbb{E}[\mathbf{X}]$ may be hardly accessible (due to integration)
  - find a simple and more accessible deterministic $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}} \simeq \mathbb{E}[\mathbf{X}]$ in some sense for $p$ large, e.g., $\|\bar{\mathbf{X}} - \mathbb{E}[\mathbf{X}]\| \to 0$ as $p \to \infty$; and
  - show variance of $f(\mathbf{X})$ decay sufficiently fast as $p \to \infty$.

» We say $\bar{\mathbf{X}}$ is a DE for $\mathbf{X}$ when $f(\mathbf{X})$ is evaluated, and denote $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$.

# Outline

LLN and CLT

From Random Scalars to Random Matrices

RMT Basis

# Fundamental Objects

Core interest of RMT: evaluation of eigenvalues and eigenvectors of a random matrix.

For a symmetric/Hermitian matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$, the resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ of $\mathbf{X}$ is defined, for $z \in \mathbb{C}$ not an eigenvalue of $\mathbf{X}$, as $\mathbf{Q}_{\mathbf{X}}(z) \equiv \left(\mathbf{X} - z\mathbf{I}_p\right)^{-1}$.

Resolvent

For symmetric $\mathbf{X} \in \mathbb{R}^{p\times p}$, the *empirical spectral distribution* $(ESD)$ $\mu_{\mathbf{X}}$ of $\mathbf{X}$ is defined as the normalized counting measure of the eigenvalues $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ of $\mathbf{X}$, i.e., $\mu_{\mathbf{X}} \equiv \frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i(\mathbf{X})}$, where $\delta_x$ represents the Dirac measure at $x$.

Empirical Spectral Distribution (ESD)

# Resolvent as the core object

| Objects of interest | Functionals of resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ |
|---|---|
| Empirical Spectral Distribution (ESD) $\mu_{\mathbf{X}}$ of $\mathbf{X}$ | Stieltjes transform $m_{\mu_{\mathbf{X}}}(z) = \frac{1}{p} \operatorname{tr} \mathbf{Q}_{\mathbf{X}}(z)$ |
| Linear spectral statistics (LSS): $f(\mathbf{X}) \equiv \frac{1}{p} \sum_i f(\lambda_i(\mathbf{X}))$ | Integration of trace of $\mathbf{Q}_{\mathbf{X}}(z)$: $-\frac{1}{2\pi\imath} \oint_\Gamma f(z) \frac{1}{p} \operatorname{tr} \mathbf{Q}_{\mathbf{X}}(z)\, dz$ (via Cauchy's integral) |
| Projections of eigenvectors $\mathbf{v}^\mathsf{T}\mathbf{u}(\mathbf{X})$ and $\mathbf{v}^\mathsf{T}\mathbf{U}(\mathbf{X})$ onto some given vector $\mathbf{v} \in \mathbb{R}^p$ | Bilinear form $\mathbf{v}^\mathsf{T}\mathbf{Q}_{\mathbf{X}}(z)\mathbf{v}$ of $\mathbf{Q}_{\mathbf{X}}$ |
| General matrix functional $F(\mathbf{X}) = \sum_i f(\lambda_i(\mathbf{X}))\mathbf{v}_1^\mathsf{T}\mathbf{u}_i(\mathbf{X})\mathbf{u}_i(\mathbf{X})^\mathsf{T}\mathbf{v}_2$ involving both eigenvalues and eigenvectors | Integration of bilinear form of $\mathbf{Q}_{\mathbf{X}}(z)$: $-\frac{1}{2\pi\imath} \oint_\Gamma f(z)\mathbf{v}_1^\mathsf{T}\mathbf{Q}_{\mathbf{X}}(z)\mathbf{v}_2\, dz$ |

# Use resolvent for eigenvalue distribution

For a symmetric/Hermitian matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$, the resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ of $\mathbf{X}$ is defined, for $z \in \mathbb{C}$ not an eigenvalue of $\mathbf{X}$, as $\mathbf{Q}_{\mathbf{X}}(z) \equiv \left(\mathbf{X} - z\mathbf{I}_p\right)^{-1}$.

Resolvent

Let $\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\mathsf{T}$ be the spectral decomposition of $\mathbf{X}$, with $\boldsymbol{\Lambda} = \{\lambda_i(\mathbf{X})\}_{i=1}^p$ eigenvalues and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_p] \in \mathbb{R}^{p \times p}$ the associated eigenvectors. Then,

$$\mathbf{Q}(z) = \mathbf{U}(\boldsymbol{\Lambda} - z\mathbf{I}_p)^{-1}\mathbf{U}^\mathsf{T} = \sum_{i=1}^{p} \frac{\mathbf{u}_i\mathbf{u}_i^\mathsf{T}}{\lambda_i(\mathbf{X}) - z}. \tag{7}$$

Thus, for $\mu_{\mathbf{X}} \equiv \frac{1}{p}\sum_{i=1}^p \delta_{\lambda_i(\mathbf{X})}$ the ESD of $\mathbf{X}$,

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{\lambda_i(\mathbf{X}) - z} = \int \frac{\mu_{\mathbf{X}}(dt)}{t - z}. \tag{8}$$

# The Stieltjes transform

For a real probability measure $\mu$ with support $\operatorname{supp}(\mu)$, the *Stieltjes transform* $m_\mu(z)$ is defined, for all $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{9}$$

Stieltjes transform

For $m_\mu$ the Stieltjes transform of a probability measure $\mu$, then

- » $m_\mu$ is complex analytic on its domain of definition $\mathbb{C} \setminus \operatorname{supp}(\mu)$;
- » it is bounded $|m_\mu(z)| \leq 1/\operatorname{dist}(z, \operatorname{supp}(\mu))$;
- » it satisfies $m_\mu(z) > 0$ for $z < \inf \operatorname{supp}(\mu)$, $m_\mu(z) < 0$ for $z > \sup \operatorname{supp}(\mu)$ and $\Im[z] \cdot \Im[m_\mu(z)] > 0$ if $z \in \mathbb{C} \setminus \mathbb{R}$; and
- » it is an increasing function on all connected components of its restriction to $\mathbb{R} \setminus \operatorname{supp}(\mu)$ (since $m'_\mu(x) = \int (t-x)^{-2} \mu(dt) > 0$) with $\lim_{x \to \pm\infty} m_\mu(x) = 0$ if $\operatorname{supp}(\mu)$ is bounded.

# The Stieltjes transform

For a real probability measure $\mu$ with support $\operatorname{supp}(\mu)$, the *Stieltjes transform* $m_\mu(z)$ is defined, for all $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{9}$$

Stieltjes transform

For $m_\mu$ the Stieltjes transform of a probability measure $\mu$, then

» $m_\mu$ is complex analytic on its domain of definition $\mathbb{C} \setminus \operatorname{supp}(\mu)$;

» it is bounded $|m_\mu(z)| \leq 1/\operatorname{dist}(z, \operatorname{supp}(\mu))$;

» it satisfies $m_\mu(z) > 0$ for $z < \inf \operatorname{supp}(\mu)$, $m_\mu(z) < 0$ for $z > \sup \operatorname{supp}(\mu)$ and $\Im[z] \cdot \Im[m_\mu(z)] > 0$ if $z \in \mathbb{C} \setminus \mathbb{R}$; and

» it is an increasing function on all connected components of its restriction to $\mathbb{R} \setminus \operatorname{supp}(\mu)$ (since $m'_\mu(x) = \int (t - x)^{-2} \mu(dt) > 0$) with $\lim_{x \to \pm\infty} m_\mu(x) = 0$ if $\operatorname{supp}(\mu)$ is bounded.

# The inverse Stieltjes transform

For $a, b$ continuity points of the probability measure $\mu$, we have

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im \left[ m_\mu(x + \imath y) \right] dx. \tag{10}$$

Besides, if $\mu$ admits a density $f$ at $x$ (i.e., $\mu(x)$ is differentiable in a neighborhood of $x$ and $\lim_{\epsilon \to 0} (2\epsilon)^{-1} \mu([x - \epsilon, x + \epsilon]) = f(x)$),

$$f(x) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \left[ m_\mu(x + \imath y) \right]. \tag{11}$$

Inverse Stieltjes transform

**Workflow**: random matrix $\mathbf{X}$ of interest $\Rightarrow$ resolvent $\mathbf{Q}_\mathbf{X}(z)$ and ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}_\mathbf{X}(z) = m_\mathbf{X}(z)$ $\Rightarrow$ study the limiting ST $m_\mathbf{X}(z) \to m(z) \Rightarrow$ inverse ST to get limiting $\mu_\mathbf{X} \to \mu$.

# The inverse Stieltjes transform

For $a, b$ continuity points of the probability measure $\mu$, we have

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im \left[ m_\mu(x + \imath y) \right] \, dx. \tag{10}$$

Besides, if $\mu$ admits a density $f$ at $x$ (i.e., $\mu(x)$ is differentiable in a neighborhood of $x$ and $\lim_{\epsilon \to 0} (2\epsilon)^{-1} \mu([x-\epsilon, x+\epsilon]) = f(x)$),

$$f(x) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \left[ m_\mu(x + \imath y) \right]. \tag{11}$$

Inverse Stieltjes transform

**Workflow**: random matrix $\mathbf{X}$ of interest $\Rightarrow$ resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ and ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}_{\mathbf{X}}(z) = m_{\mathbf{X}}(z)$ $\Rightarrow$ study the limiting ST $m_{\mathbf{X}}(z) \to m(z)$ $\Rightarrow$ inverse ST to get limiting $\mu_{\mathbf{X}} \to \mu$.

# The inverse Stieltjes transform

For $a, b$ continuity points of the probability measure $\mu$, we have

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im \left[ m_\mu(x + \imath y) \right] dx. \tag{10}$$

Besides, if $\mu$ admits a density $f$ at $x$ (i.e., $\mu(x)$ is differentiable in a neighborhood of $x$ and $\lim_{\epsilon \to 0} (2\epsilon)^{-1} \mu([x-\epsilon, x+\epsilon]) = f(x)$),

$$f(x) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \left[ m_\mu(x + \imath y) \right]. \tag{11}$$

Inverse Stieltjes transform

**Workflow**: random matrix $\mathbf{X}$ of interest $\Rightarrow$ resolvent $\mathbf{Q}_\mathbf{X}(z)$ and ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}_\mathbf{X}(z) = m_\mathbf{X}(z)$ $\Rightarrow$ study the limiting ST $m_\mathbf{X}(z) \to m(z)$ $\Rightarrow$ inverse ST to get limiting $\mu_\mathbf{X} \to \mu$.

# Use the resolvent for eigenvalue functionals

For a symmetric matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$, the *linear spectral statistics* (LSS) $f_{\mathbf{X}}$ of $\mathbf{X}$ is defined as the averaged statistics of the eigenvalues $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ of $\mathbf{X}$ via some function $f : \mathbb{R} \to \mathbb{R}$, that is

$$f(\mathbf{X}) = \frac{1}{p} \sum_{i=1}^{p} f(\lambda_i(\mathbf{X})) = \int f(t) \mu_{\mathbf{X}}(dt), \tag{12}$$

for $\mu_{\mathbf{X}}$ the ESD of $\mathbf{X}$.

Linear Spectral Statistics (LSS)

# Cauchy's integral formula

For $\Gamma \subset \mathbb{C}$ a positively (i.e., counterclockwise) oriented simple closed curve and a complex function $f(z)$ analytic in a region containing $\Gamma$ and its inside, then

(i) if $z_0 \in \mathbb{C}$ is enclosed by $\Gamma$, $f(z_0) = -\frac{1}{2\pi\imath} \oint_\Gamma \frac{f(z)}{z_0 - z}\, dz$;

(ii) if not, $\frac{1}{2\pi\imath} \oint_\Gamma \frac{f(z)}{z_0 - z}\, dz = 0$.

Cauchy's integral formula

**LSS via contour integration**: For $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ eigenvalues of a symmetric matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$, some function $f : \mathbb{R} \to \mathbb{R}$ that is complex analytic in a compact neighborhood of the support $\operatorname{supp}(\mu_\mathbf{X})$ (of the ESD $\mu_\mathbf{X}$ of $\mathbf{X}$), then

$$f(\mathbf{X}) = \int f(t)\mu_\mathbf{X}(dt) = -\int \frac{1}{2\pi\imath} \oint_\Gamma \frac{f(z)\, dz}{t - z} \mu_\mathbf{X}(dt) = -\frac{1}{2\pi\imath} \oint_\Gamma f(z) m_{\mu_\mathbf{X}}(z)\, dz, \qquad (13)$$

for *any* contour $\Gamma$ that encloses $\operatorname{supp}(\mu_\mathbf{X})$, i.e., all the eigenvalues $\lambda_i(\mathbf{X})$.

# Cauchy's integral formula

For $\Gamma \subset \mathbb{C}$ a positively (i.e., counterclockwise) oriented simple closed curve and a complex function $f(z)$ analytic in a region containing $\Gamma$ and its inside, then

(i) if $z_0 \in \mathbb{C}$ is enclosed by $\Gamma$, $f(z_0) = -\frac{1}{2\pi\imath} \oint_\Gamma \frac{f(z)}{z_0 - z} \, dz$;

(ii) if not, $\frac{1}{2\pi\imath} \oint_\Gamma \frac{f(z)}{z_0 - z} \, dz = 0$.

Cauchy's integral formula

**LSS via contour integration**: For $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ eigenvalues of a symmetric matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$, some function $f \colon \mathbb{R} \to \mathbb{R}$ that is complex analytic in a compact neighborhood of the support $\operatorname{supp}(\mu_{\mathbf{X}})$ (of the ESD $\mu_{\mathbf{X}}$ of $\mathbf{X}$), then

$$f(\mathbf{X}) = \int f(t) \mu_{\mathbf{X}}(dt) = -\int \frac{1}{2\pi\imath} \oint_\Gamma \frac{f(z) \, dz}{t - z} \mu_{\mathbf{X}}(dt) = -\frac{1}{2\pi\imath} \oint_\Gamma f(z) m_{\mu_{\mathbf{X}}}(z) \, dz, \tag{13}$$

for *any* contour $\Gamma$ that encloses $\operatorname{supp}(\mu_{\mathbf{X}})$, i.e., all the eigenvalues $\lambda_i(\mathbf{X})$.

# LSS to retrieve the inverse Stieltjes transform formula

$$\frac{1}{p}\sum_{\lambda_i(\mathbf{X})\in[a,b]}\delta_{\lambda_i(\mathbf{X})} = -\frac{1}{2\pi\imath}\oint_\Gamma 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz$$

$$= -\frac{1}{2\pi\imath}\int_{a-\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz - \frac{1}{2\pi\imath}\int_{b+\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz$$

$$- \frac{1}{2\pi\imath}\int_{a-\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz - \frac{1}{2\pi\imath}\int_{b+\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz.$$

- Since $\Re[m(x+\imath y)] = \Re[m(x-\imath y)], \Im[m(x+\imath y)] = -\Im[m(x-\imath y)]$;
- we have $\int_{a-\varepsilon_x}^{b+\varepsilon_x} m_{\mu_\mathbf{X}}(x-\imath\varepsilon_y)\,dx + \int_{b+\varepsilon_x}^{a-\varepsilon_x} m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)dx = -2\imath\int_{a-\varepsilon_x}^{b+\varepsilon_x}\Im[m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)]\,dx$;
- and consequently $\mu([a,b]) = \frac{1}{p}\sum_{\lambda_i(\mathbf{X})\in[a,b]}\lambda_i(\mathbf{X}) = \frac{1}{\pi}\lim_{\varepsilon_y\downarrow 0}\int_a^b \Im[m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)]\,dx$.

# LSS to retrieve the inverse Stieltjes transform formula

$$
\begin{aligned}
\frac{1}{p}\sum_{\lambda_i(\mathbf{X})\in[a,b]}\delta_{\lambda_i(\mathbf{X})} &= -\frac{1}{2\pi\imath}\oint_{\Gamma} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_{\mathbf{X}}}(z)\,dz \\
&= -\frac{1}{2\pi\imath}\int_{a-\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_{\mathbf{X}}}(z)\,dz - \frac{1}{2\pi\imath}\int_{b+\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_{\mathbf{X}}}(z)\,dz \\
&- \frac{1}{2\pi\imath}\int_{a-\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_{\mathbf{X}}}(z)\,dz - \frac{1}{2\pi\imath}\int_{b+\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_{\mathbf{X}}}(z)\,dz.
\end{aligned}
$$

» Since $\Re[m(x+\imath y)] = \Re[m(x-\imath y)]$, $\Im[m(x+\imath y)] = -\Im[m(x-\imath y)]$;

» we have $\int_{a-\varepsilon_x}^{b+\varepsilon_x} m_{\mu_{\mathbf{X}}}(x-\imath\varepsilon_y)\,dx + \int_{b+\varepsilon_x}^{a-\varepsilon_x} m_{\mu_{\mathbf{X}}}(x+\imath\varepsilon_y)dx = -2\imath\int_{a-\varepsilon_x}^{b+\varepsilon_x} \Im[m_{\mu_{\mathbf{X}}}(x+\imath\varepsilon_y)]\,dx$;

» and consequently $\mu([a,b]) = \frac{1}{p}\sum_{\lambda_i(\mathbf{X})\in[a,b]}\lambda_i(\mathbf{X}) = \frac{1}{\pi}\lim_{\varepsilon_y\downarrow 0}\int_a^b \Im[m_{\mu_{\mathbf{X}}}(x+\imath\varepsilon_y)]\,dx$.

# LSS to retrieve the inverse Stieltjes transform formula

$$\frac{1}{p}\sum_{\lambda_i(\mathbf{X})\in[a,b]}\delta_{\lambda_i(\mathbf{X})} = -\frac{1}{2\pi\imath}\oint_\Gamma 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz$$

$$= -\frac{1}{2\pi\imath}\int_{a-\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz - \frac{1}{2\pi\imath}\int_{b+\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz$$

$$-\frac{1}{2\pi\imath}\int_{a-\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz - \frac{1}{2\pi\imath}\int_{b+\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z)m_{\mu_\mathbf{X}}(z)\,dz.$$

- Since $\Re[m(x+\imath y)] = \Re[m(x-\imath y)], \Im[m(x+\imath y)] = -\Im[m(x-\imath y)]$;
- we have $\int_{a-\varepsilon_x}^{b+\varepsilon_x} m_{\mu_\mathbf{X}}(x-\imath\varepsilon_y)\,dx + \int_{b+\varepsilon_x}^{a-\varepsilon_x} m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)dx = -2\imath\int_{a-\varepsilon_x}^{b+\varepsilon_x} \Im[m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)]\,dx$;
- and consequently $\mu([a,b]) = \frac{1}{p}\sum_{\lambda_i(\mathbf{X})\in[a,b]}\lambda_i(\mathbf{X}) = \frac{1}{\pi}\lim_{\varepsilon_y\downarrow 0}\int_a^b \Im[m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)]\,dx$.

# LSS to retrieve the inverse Stieltjes transform formula

$$
\begin{aligned}
&\frac{1}{p} \sum_{\lambda_i(\mathbf{X}) \in [a,b]} \delta_{\lambda_i(\mathbf{X})} = -\frac{1}{2\pi\imath} \oint_\Gamma 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_\mathbf{X}}(z)\, dz \\
&= -\frac{1}{2\pi\imath} \int_{a-\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_\mathbf{X}}(z)\, dz - \frac{1}{2\pi\imath} \int_{b+\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_\mathbf{X}}(z)\, dz \\
&- \frac{1}{2\pi\imath} \int_{a-\varepsilon_x+\imath\varepsilon_y}^{a-\varepsilon_x-\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_\mathbf{X}}(z)\, dz - \frac{1}{2\pi\imath} \int_{b+\varepsilon_x-\imath\varepsilon_y}^{b+\varepsilon_x+\imath\varepsilon_y} 1_{\Re[z]\in[a-\varepsilon,b+\varepsilon]}(z) m_{\mu_\mathbf{X}}(z)\, dz.
\end{aligned}
$$

» Since $\Re[m(x+\imath y)] = \Re[m(x-\imath y)], \Im[m(x+\imath y)] = -\Im[m(x-\imath y)]$;

» we have $\int_{a-\varepsilon_x}^{b+\varepsilon_x} m_{\mu_\mathbf{X}}(x-\imath\varepsilon_y)\, dx + \int_{b+\varepsilon_x}^{a-\varepsilon_x} m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y) dx = -2\imath \int_{a-\varepsilon_x}^{b+\varepsilon_x} \Im[m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)]\, dx$;

» and consequently $\mu([a,b]) = \frac{1}{p}\sum_{\lambda_i(\mathbf{X})\in[a,b]} \lambda_i(\mathbf{X}) = \frac{1}{\pi} \lim_{\varepsilon_y \downarrow 0} \int_a^b \Im[m_{\mu_\mathbf{X}}(x+\imath\varepsilon_y)]\, dx$.

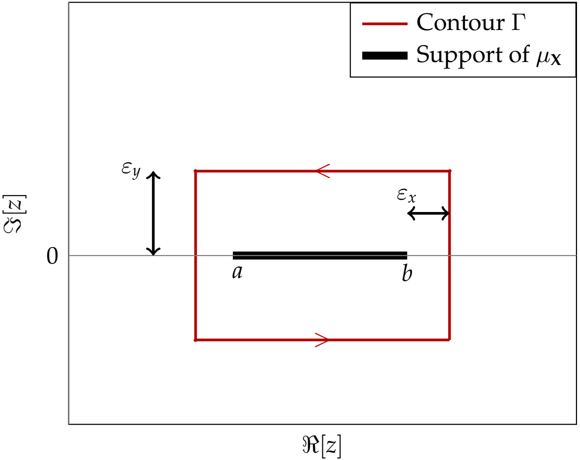


Figure: Illustration of a rectangular contour $\Gamma$ and support of $\mu_X$ on the complex plane.

# Use resolvent for eigenvectors and eigenspace

Resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ contains eigenvector information about $\mathbf{X}$, recall

$$\mathbf{Q}_{\mathbf{X}}(z) = \sum_{i=1}^{p} \frac{\mathbf{u}_i \mathbf{u}_i^{\mathsf{T}}}{\lambda_i(\mathbf{X}) - z},$$

and that we have direct access to the $i$-th eigenvector $\mathbf{u}_i$ of $\mathbf{X}$ through

$$\mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} = -\frac{1}{2\pi \imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{Q}_{\mathbf{X}}(z)\, dz, \tag{14}$$

for $\Gamma_{\lambda_i(\mathbf{X})}$ a contour circling around $\lambda_i(\mathbf{X})$ only.

» seen as a matrix-version of LSS formula

» with the Stieltjes transform $m_{\mu_{\mathbf{X}}}(z)$ replaced by the associated resolvent $\mathbf{Q}_{\mathbf{X}}(z)$

# Use resolvent for eigenvectors and eigenspace

Resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ contains eigenvector information about $\mathbf{X}$, recall

$$\mathbf{Q}_{\mathbf{X}}(z) = \sum_{i=1}^{p} \frac{\mathbf{u}_i \mathbf{u}_i^{\mathsf{T}}}{\lambda_i(\mathbf{X}) - z},$$

and that we have direct access to the $i$-th eigenvector $\mathbf{u}_i$ of $\mathbf{X}$ through

$$\mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} = -\frac{1}{2\pi\imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{Q}_{\mathbf{X}}(z)\, dz, \tag{14}$$

for $\Gamma_{\lambda_i(\mathbf{X})}$ a contour circling around $\lambda_i(\mathbf{X})$ only.

» seen as a matrix-version of LSS formula

» with the Stieltjes transform $m_{\mu_{\mathbf{X}}}(z)$ replaced by the associated resolvent $\mathbf{Q}_{\mathbf{X}}(z)$

# Use resolvent for eigenvectors and eigenspace

Resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ contains eigenvector information about $\mathbf{X}$, recall

$$\mathbf{Q}_{\mathbf{X}}(z) = \sum_{i=1}^{p} \frac{\mathbf{u}_i \mathbf{u}_i^\mathsf{T}}{\lambda_i(\mathbf{X}) - z},$$

and that we have direct access to the $i$-th eigenvector $\mathbf{u}_i$ of $\mathbf{X}$ through

$$\mathbf{u}_i \mathbf{u}_i^\mathsf{T} = -\frac{1}{2\pi\imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{Q}_{\mathbf{X}}(z)\, dz, \tag{14}$$

for $\Gamma_{\lambda_i(\mathbf{X})}$ a contour circling around $\lambda_i(\mathbf{X})$ only.

» seen as a matrix-version of LSS formula

» with the Stieltjes transform $m_{\mu_{\mathbf{X}}}(z)$ replaced by the associated resolvent $\mathbf{Q}_{\mathbf{X}}(z)$

# Spectral functionals via resolvent

For a symmetric matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$, we say $F \colon \mathbb{R}^{p \times p} \to \mathbb{R}^{p \times p}$ is a (matrix) spectral functional of $\mathbf{X}$,

$$F(\mathbf{X}) = \sum_{i \in \mathcal{I} \subseteq \{1,\ldots,p\}} f(\lambda_i(\mathbf{X}))\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}, \quad \mathbf{X} = \sum_{i=1}^{p} \lambda_i(\mathbf{X})\mathbf{u}_i\mathbf{u}_i^{\mathsf{T}}. \tag{15}$$

Matrix spectral functionals

**Spectral functional via contour integration**: For $\mathbf{X} \in \mathbb{R}^{p \times p}$, resolvent $\mathbf{Q}_{\mathbf{X}}(z) = (\mathbf{X} - z\mathbf{I}_p)^{-1}$, $z \in \mathbb{C}$, and $f : \mathbb{R} \to \mathbb{R}$ analytic in a neighborhood of the contour $\Gamma_{\mathcal{I}}$ that circles around the eigenvalues $\lambda_i(\mathbf{X})$ of $\mathbf{X}$ with their indices in the set $\mathcal{I} \subseteq \{1, \ldots, p\}$,

$$F(\mathbf{X}) = -\frac{1}{2\pi \imath} \oint_{\Gamma_{\mathcal{I}}} f(z)\mathbf{Q}_{\mathbf{X}}(z)\, dz. \tag{16}$$

**Example**: eigenvector projection $(\mathbf{v}^{\mathsf{T}}\mathbf{u}_i)^2 = -\frac{1}{2\pi \imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{v}^{\mathsf{T}}\mathbf{Q}_{\mathbf{X}}(z)\mathbf{v}\, dz$.

# Spectral functionals via resolvent

For a symmetric matrix $\mathbf{X} \in \mathbb{R}^{p\times p}$, we say $F\colon \mathbb{R}^{p\times p} \to \mathbb{R}^{p\times p}$ is a (matrix) spectral functional of $\mathbf{X}$,

$$F(\mathbf{X}) = \sum_{i\in\mathcal{I}\subseteq\{1,\ldots,p\}} f(\lambda_i(\mathbf{X}))\mathbf{u}_i\mathbf{u}_i^\mathsf{T}, \quad \mathbf{X} = \sum_{i=1}^{p} \lambda_i(\mathbf{X})\mathbf{u}_i\mathbf{u}_i^\mathsf{T}. \tag{15}$$

Matrix spectral functionals

**Spectral functional via contour integration**: For $\mathbf{X} \in \mathbb{R}^{p\times p}$, resolvent $\mathbf{Q}_\mathbf{X}(z) = (\mathbf{X} - z\mathbf{I}_p)^{-1}$, $z \in \mathbb{C}$, and $f\colon \mathbb{R} \to \mathbb{R}$ analytic in a neighborhood of the contour $\Gamma_\mathcal{I}$ that circles around the eigenvalues $\lambda_i(\mathbf{X})$ of $\mathbf{X}$ with their indices in the set $\mathcal{I} \subseteq \{1,\ldots,p\}$,

$$F(\mathbf{X}) = -\frac{1}{2\pi\imath}\oint_{\Gamma_\mathcal{I}} f(z)\mathbf{Q}_\mathbf{X}(z)\,dz. \tag{16}$$

**Example**: eigenvector projection $(\mathbf{v}^\mathsf{T}\mathbf{u}_i)^2 = -\frac{1}{2\pi\imath}\oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{v}^\mathsf{T}\mathbf{Q}_\mathbf{X}(z)\mathbf{v}\,dz$.

# Spectral functionals via resolvent

For a symmetric matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$, we say $F \colon \mathbb{R}^{p \times p} \to \mathbb{R}^{p \times p}$ is a (matrix) spectral functional of $\mathbf{X}$,

$$F(\mathbf{X}) = \sum_{i \in \mathcal{I} \subseteq \{1,\ldots,p\}} f(\lambda_i(\mathbf{X})) \mathbf{u}_i \mathbf{u}_i^\mathsf{T}, \quad \mathbf{X} = \sum_{i=1}^{p} \lambda_i(\mathbf{X}) \mathbf{u}_i \mathbf{u}_i^\mathsf{T}. \tag{15}$$

Matrix spectral functionals

**Spectral functional via contour integration**: For $\mathbf{X} \in \mathbb{R}^{p \times p}$, resolvent $\mathbf{Q}_\mathbf{X}(z) = (\mathbf{X} - z\mathbf{I}_p)^{-1}$, $z \in \mathbb{C}$, and $f \colon \mathbb{R} \to \mathbb{R}$ analytic in a neighborhood of the contour $\Gamma_\mathcal{I}$ that circles around the eigenvalues $\lambda_i(\mathbf{X})$ of $\mathbf{X}$ with their indices in the set $\mathcal{I} \subseteq \{1, \ldots, p\}$,

$$F(\mathbf{X}) = -\frac{1}{2\pi \imath} \oint_{\Gamma_\mathcal{I}} f(z) \mathbf{Q}_\mathbf{X}(z) \, dz. \tag{16}$$

**Example**: eigenvector projection $(\mathbf{v}^\mathsf{T} \mathbf{u}_i)^2 = -\frac{1}{2\pi\imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{v}^\mathsf{T} \mathbf{Q}_\mathbf{X}(z) \mathbf{v} \, dz$.