Probability and Stochastic Process II: Random Matrix Theory and Applications
Lecture 3: From LLN to MP laws

Zhenyu Liao, Tiebin Mi, Caiming Qiu

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

March 1, 2023

Outline

SCM and MP law

Proof of Marčenko–Pastur law

RMT Basis

What we will have today

- sample covariance matrix and the limiting Marčenko–Pastur law
- Wigner matrix and the limiting semicircle law
- proof via Bai and Silverstein approach and/or Gaussian tool

What we will have today

- » sample covariance matrix and the limiting Marčenko–Pastur law
- » Wigner matrix and the limiting semicircle law
- » proof via Bai and Silverstein approach and/or Gaussian tool

What we will have today

» sample covariance matrix and the limiting Marčenko–Pastur law

» Wigner matrix and the limiting semicircle law

» proof via Bai and Silverstein approach and/or Gaussian tool

Outline

SCM and MP law

Proof of Marčenko–Pastur law

RMT Basis

Sample covariance matrix in the large n, p regime

» **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,

» Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

» In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

» due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p \|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \le \|\mathbf{A}\| \le p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

  almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").
- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \;\Rightarrow\; \text{eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\boxed{\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x}\sqrt{(x - E_-)^+(E_+ - x)^+}dx}$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1-\sqrt{\mathbf{c}})^2, E_+ = (1+\sqrt{\mathbf{c}})^2]$.
- for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm\mathbf{2}\sqrt{\mathbf{c}} = \pm\mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\boxed{\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x}\sqrt{(x - E_-)^+(E_+ - x)^+}dx}$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

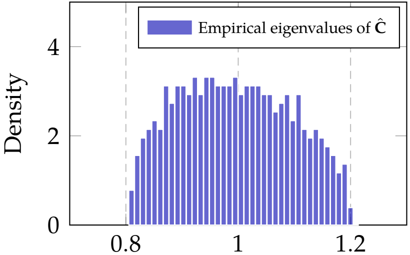

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

» eigenvalues span on $[E_- = (1-\sqrt{\mathbf{c}})^2, E_+ = (1+\sqrt{\mathbf{c}})^2]$.

» for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm 2\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\boxed{\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x}\sqrt{(x - E_-)^+(E_+ - x)^+}dx}$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

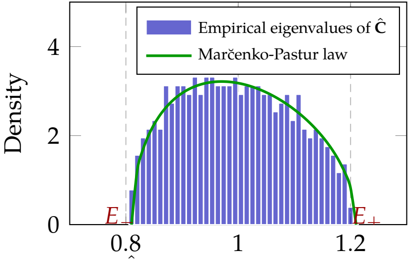

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1-\sqrt{c})^2, E_+ = (1+\sqrt{c})^2]$.
- for $n = 100p$, on a range of $\pm 2\sqrt{c} = \pm 0.2$ around the population eigenvalue 1.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\boxed{\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x}\sqrt{(x - E_-)^+ (E_+ - x)^+} dx}$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

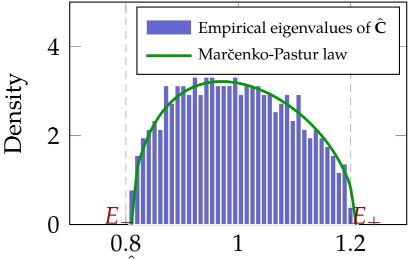

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1-\sqrt{\mathbf{c}})^2, E_+ = (1+\sqrt{\mathbf{c}})^2]$.
- for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm 2\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\boxed{\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x} \sqrt{(x - E_-)^+ (E_+ - x)^+} dx}$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

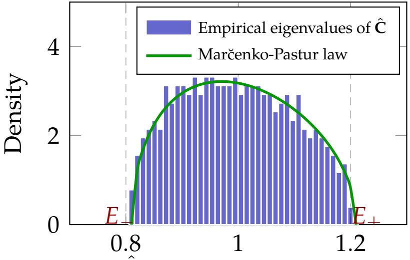

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

» eigenvalues span on $[E_- = (1 - \sqrt{\mathbf{c}})^2, E_+ = (1 + \sqrt{\mathbf{c}})^2]$.

» for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm 2\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\boxed{\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x} \sqrt{(x - E_-)^+ (E_+ - x)^+} dx}$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

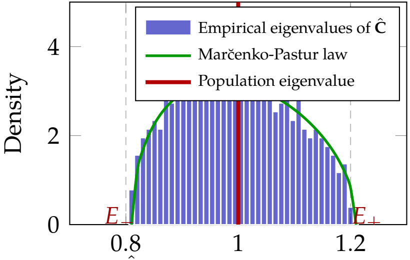

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1 - \sqrt{\mathbf{c}})^2, E_+ = (1 + \sqrt{\mathbf{c}})^2]$.
- for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm 2\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

Marčenko–Pastur law

Let $\mathbf{X} \in \mathbb{R}^{p\times n}$ be a random matrix with i.i.d. entries of zero mean and σ^2 variance. Then, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, with probability one, the empirical spectral measure $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}}$ of $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$ converges weakly to the probability measure μ

$$\mu(dx) = (1-c^{-1})^+\delta_0(x) + \frac{1}{2\pi c\sigma^2 x}\sqrt{(x-\sigma^2 E_-)^+(\sigma^2 E_+ - x)^+}\,dx, \tag{1}$$

where $E_\pm = (1\pm\sqrt{c})^2$ and $(x)^+ = \max(0,x)$. In particular, with $\sigma^2 = 1$,

$$\boxed{\mu(dx) = (1-c^{-1})^+\delta_0(x) + \frac{1}{2\pi c x}\sqrt{(x-E_-)^+\,(E_+ - x)^+}\,dx,} \tag{2}$$

which is known as the Marčenko-Pastur law.

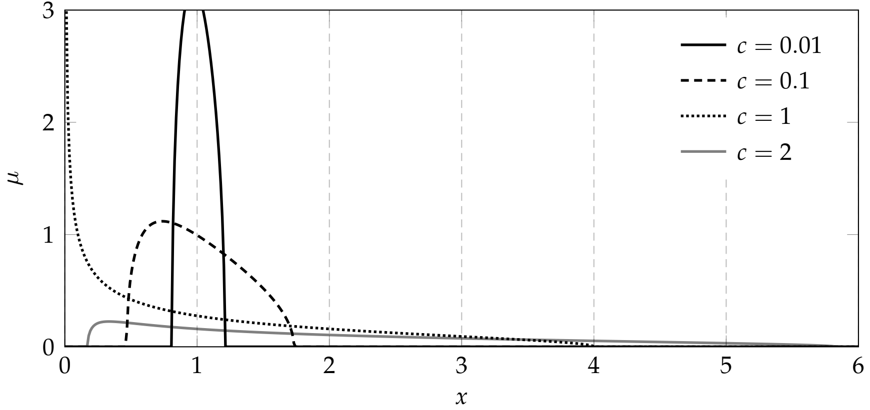

Figure: Marčenko-Pastur distribution for different values of c.

Outline

SCM and MP law

Proof of Marčenko–Pastur law

RMT Basis

Proof of Marčenko–Pastur law

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 $\mu_{\mathbf{X}} \to \mu$.

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 δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

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

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \quad (3)$$

Stieltjes transform

Proof of Marčenko–Pastur law

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 $\mu_{\mathbf{X}} \to \mu$.

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 δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

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

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{3}$$

Stieltjes transform

Proof of Marčenko–Pastur law

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 $\mu_{\mathbf{X}} \to \mu$.

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 δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

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

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{3}$$

Stieltjes transform

Proof of Marčenko–Pastur law

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 $\mu_\mathbf{X} \to \mu$.

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 δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

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

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{3}$$

Stieltjes transform

Heuristic proof of MP law via "leave-one-out"

» "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) = \mathbf{Q}(z) \left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\top \right) \bar{\mathbf{Q}}(z)$$
$$= \mathbf{Q}(z) \left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top \right) \bar{\mathbf{Q}}(z).$$

» for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\top \bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

» $\mathbf{x}_i^\top \bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.

» cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out"

- "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) = \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\bar{\mathbf{Q}}(z)$$
$$= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\right)\bar{\mathbf{Q}}(z).$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out"

- "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\begin{aligned}\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\bar{\mathbf{Q}}(z) \\ &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\right)\bar{\mathbf{Q}}(z).\end{aligned}$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out"

» "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\begin{aligned}\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\bar{\mathbf{Q}}(z) \\ &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\right)\bar{\mathbf{Q}}(z).\end{aligned}$$

» for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

» $\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.

» cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out"

- "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\begin{aligned}\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\top\right)\bar{\mathbf{Q}}(z) \\ &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)\bar{\mathbf{Q}}(z).\end{aligned}$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\top\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\top\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

- use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\top \mathbf{Q}_{-i}(z)\mathbf{x}_i}$,
- now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\top - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,
- quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\top \bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\top \bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\top \mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

- So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z + \frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.
- self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,

» now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,

» quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

» So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.

» self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,

» now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,

» quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

» So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.

» self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

- use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,
- now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,
- quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

- So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.
- self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,

» now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,

» quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

» So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.

» self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,

» now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,

» quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

» So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.

» self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via “leave-one-out”

Objective: “guess” the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» we have $\mathbf{F}(z) = \left(-z + \frac{1}{1+\frac{1}{n}\operatorname{tr}\bar{\mathbf{Q}}(z)} \right) \mathbf{I}_p$,

» and $\bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1+cm(z)} \right)^{-1}, \text{ or } zcm^2(z) - (1-c-z)m(z) + 1 = 0.$$

» has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho}e^{\imath\theta/2}\}$)

$$m(z) = \frac{1-c-z}{2cz} + \frac{\sqrt{((1+\sqrt{c})^2 - z)((1-\sqrt{c})^2 - z)}}{2cz},$$

only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.

» apply inverse Stieltjes transform we conclude the proof.

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» we have $\mathbf{F}(z) = \left(-z + \frac{1}{1+\frac{1}{n} \operatorname{tr} \bar{\mathbf{Q}}(z)} \right) \mathbf{I}_p$,

» and $\bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1 + cm(z)} \right)^{-1}, \text{ or } zcm^2(z) - (1 - c - z)m(z) + 1 = 0.$$

» has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho}e^{\imath\theta/2}\}$)

$$m(z) = \frac{1 - c - z}{2cz} + \frac{\sqrt{((1 + \sqrt{c})^2 - z)((1 - \sqrt{c})^2 - z)}}{2cz},$$

only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.

» apply inverse Stieltjes transform we conclude the proof.

Heuristic proof of MP law via “leave-one-out”

Objective: “guess” the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

- we have $\mathbf{F}(z) = \left(-z + \frac{1}{1+\frac{1}{n}\operatorname{tr}\bar{\mathbf{Q}}(z)} \right) \mathbf{I}_p$,
- and $\bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1+cm(z)} \right)^{-1}, \text{ or } zcm^2(z) - (1-c-z)m(z) + 1 = 0.$$

- has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho}e^{\imath\theta/2}\}$)

$$m(z) = \frac{1-c-z}{2cz} + \frac{\sqrt{((1+\sqrt{c})^2 - z)((1-\sqrt{c})^2 - z)}}{2cz},$$

  only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.
- apply inverse Stieltjes transform we conclude the proof.

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$.

» we have $\mathbf{F}(z) = \left(-z + \frac{1}{1+\frac{1}{n}\operatorname{tr}\bar{\mathbf{Q}}(z)} \right) \mathbf{I}_p$,

» and $\bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1+cm(z)} \right)^{-1}, \text{ or } zcm^2(z) - (1-c-z)m(z) + 1 = 0.$$

» has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho}e^{\imath\theta/2}\}$)

$$m(z) = \frac{1-c-z}{2cz} + \frac{\sqrt{((1+\sqrt{c})^2 - z)((1-\sqrt{c})^2 - z)}}{2cz},$$

only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.

» apply inverse Stieltjes transform we conclude the proof.

Some thoughts on the "leave-one-out" proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the "leave-one-out" proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the "leave-one-out" proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\xrightarrow{\text{RMT}}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the "leave-one-out" proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the “leave-one-out” proof

» in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)

» quadratic form close to the trace: anything more than LLN and concentration

» leave-one-out analysis of large-scale system: $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\mathbf{Q}_{-i}(z)$ for n, p large.

» low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation

» These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Proof of MP law with Gaussian method

Let $x \sim \mathcal{N}(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ a continuously differentiable function having at most polynomial growth and such that $\mathbb{E}[f'(x)] < \infty$. Then,

$$\mathbb{E}[xf(x)] = \mathbb{E}[f'(x)]. \tag{7}$$

In particular, for $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p\times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ a continuously differentiable function with derivatives having at most polynomial growth with respect to p,

$$\mathbb{E}[[\mathbf{x}]_i f(\mathbf{x})] = \sum_{j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j}\right], \tag{8}$$

where $\partial/\partial[\mathbf{x}]_i$ indicates differentiation with respect to the i-th entry of $\mathbf{x}$; or, in vector form $\mathbb{E}[\mathbf{x}f(\mathbf{x})] = \mathbf{C}\mathbb{E}[\nabla f(\mathbf{x})]$, with $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Stein's Lemma

Proof of MP law with Gaussian method

First observe that $\mathbf{Q} = \frac{1}{z}\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q} - \frac{1}{z}\mathbf{I}_p$, so that $\mathbb{E}[\mathbf{Q}_{ij}] = \frac{1}{zn}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] - \frac{1}{z}\delta_{ij}$, in which $\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}[xf(x)]$ for $x = \mathbf{X}_{ik}$ and $f(x) = [\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}$. Therefore, from Stein's lemma and the fact that $\partial\mathbf{Q} = -\frac{1}{n}\mathbf{Q}\partial(\mathbf{X}\mathbf{X}^\mathsf{T})\mathbf{Q}$,[a]

$$\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}\left[\frac{\partial[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}}{\partial\mathbf{X}_{ik}}\right] = \mathbb{E}[\mathbf{E}_{ik}^\mathsf{T}\mathbf{Q}]_{kj} - \mathbb{E}\left[\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{Q}(\mathbf{E}_{ik}\mathbf{X}^\mathsf{T} + \mathbf{X}\mathbf{E}_{ik}^\mathsf{T})\mathbf{Q}\right]_{kj}$$
$$= \mathbb{E}[\mathbf{Q}_{ij}] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ki}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}\right] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}\mathbf{X}]_{kk}\mathbf{Q}_{ij}\right]$$

for $\mathbf{E}_{ij}$ the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il}\delta_{jm}$, so that, summing over k,

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ij}.$$

[a] This is the matrix version of $d(1/x) = -dx/x^2$.

Proof of MP law with Gaussian method

First observe that $\mathbf{Q} = \frac{1}{z}\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q} - \frac{1}{z}\mathbf{I}_p$, so that $\mathbb{E}[\mathbf{Q}_{ij}] = \frac{1}{zn}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] - \frac{1}{z}\delta_{ij}$, in which $\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}[xf(x)]$ for $x = \mathbf{X}_{ik}$ and $f(x) = [\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}$. Therefore, from Stein's lemma and the fact that $\partial\mathbf{Q} = -\frac{1}{n}\mathbf{Q}\partial(\mathbf{X}\mathbf{X}^\mathsf{T})\mathbf{Q}$,[a]

$$\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}\left[\frac{\partial[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}}{\partial\mathbf{X}_{ik}}\right] = \mathbb{E}[\mathbf{E}_{ik}^\mathsf{T}\mathbf{Q}]_{kj} - \mathbb{E}\left[\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{Q}(\mathbf{E}_{ik}\mathbf{X}^\mathsf{T} + \mathbf{X}\mathbf{E}_{ik}^\mathsf{T})\mathbf{Q}\right]_{kj}$$
$$= \mathbb{E}[\mathbf{Q}_{ij}] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ki}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}\right] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}\mathbf{X}]_{kk}\mathbf{Q}_{ij}\right]$$

for $\mathbf{E}_{ij}$ the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il}\delta_{jm}$, so that, summing over k,

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ij}.$$

[a] This is the matrix version of $d(1/x) = -dx/x^2$.

Proof of MP law with Gaussian method

First observe that $\mathbf{Q} = \frac{1}{z}\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q} - \frac{1}{z}\mathbf{I}_p$, so that $\mathbb{E}[\mathbf{Q}_{ij}] = \frac{1}{zn}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] - \frac{1}{z}\delta_{ij}$, in which $\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}[xf(x)]$ for $x = \mathbf{X}_{ik}$ and $f(x) = [\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}$. Therefore, from Stein's lemma and the fact that $\partial\mathbf{Q} = -\frac{1}{n}\mathbf{Q}\partial(\mathbf{X}\mathbf{X}^\mathsf{T})\mathbf{Q}$,[a]

$$\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}\left[\frac{\partial[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}}{\partial\mathbf{X}_{ik}}\right] = \mathbb{E}[\mathbf{E}_{ik}^\mathsf{T}\mathbf{Q}]_{kj} - \mathbb{E}\left[\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{Q}(\mathbf{E}_{ik}\mathbf{X}^\mathsf{T} + \mathbf{X}\mathbf{E}_{ik}^\mathsf{T})\mathbf{Q}\right]_{kj}$$

$$= \mathbb{E}[\mathbf{Q}_{ij}] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ki}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}\right] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}\mathbf{X}]_{kk}\mathbf{Q}_{ij}\right]$$

for $\mathbf{E}_{ij}$ the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il}\delta_{jm}$, so that, summing over k,

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\,\mathrm{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ij}.$$

[a] This is the matrix version of $d(1/x) = -dx/x^2$.

Proof of MP law with Gaussian method

We have

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ij}.$$

The term in the second line has vanishing operator norm (of order $O(n^{-1})$) as $n, p \to \infty$. Also, $\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}) = np + zn\operatorname{tr}\mathbf{Q}$. As a result, matrix-wise, we obtain

$$\mathbb{E}[\mathbf{Q}] + \frac{1}{z}\mathbf{I}_p = \mathbb{E}[\mathbf{X}_{\cdot k}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{k\cdot}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}] - \frac{1}{z}\frac{1}{n}\mathbb{E}[\mathbf{Q}(p + z\operatorname{tr}\mathbf{Q})] + o_{\|\cdot\|}(1),$$

where $\mathbf{X}_{\cdot k}$ and $\mathbf{X}_{k\cdot}$ is the k-th column and row of $\mathbf{X}$, respectively. As the random $\frac{1}{p}\operatorname{tr}\mathbf{Q} \to m(z)$ as $n, p \to \infty$, take it out of the expectation in the limit and

$$\mathbb{E}[\mathbf{Q}](1 - p/n - z - p/n \cdot zm(z)) = \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which, taking the trace to identify $m(z)$, concludes the proof.

Proof of MP law with Gaussian method

We have

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{ij}.$$

The term in the second line has vanishing operator norm (of order $O(n^{-1})$) as $n, p \to \infty$. Also, $\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}) = np + zn\operatorname{tr}\mathbf{Q}$. As a result, matrix-wise, we obtain

$$\mathbb{E}[\mathbf{Q}] + \frac{1}{z}\mathbf{I}_p = \mathbb{E}[\mathbf{X}_{\cdot k}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{k\cdot}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}] - \frac{1}{z}\frac{1}{n}\mathbb{E}[\mathbf{Q}(p + z\operatorname{tr}\mathbf{Q})] + o_{\|\cdot\|}(1),$$

where $\mathbf{X}_{\cdot k}$ and $\mathbf{X}_{k\cdot}$ is the k-th column and row of $\mathbf{X}$, respectively. As the random $\frac{1}{p}\operatorname{tr}\mathbf{Q} \to m(z)$ as $n, p \to \infty$, take it out of the expectation in the limit and

$$\mathbb{E}[\mathbf{Q}](1 - p/n - z - p/n \cdot zm(z)) = \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which, taking the trace to identify $m(z)$, concludes the proof.

Proof of MP law with Gaussian method

We have

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{ij}.$$

The term in the second line has vanishing operator norm (of order $O(n^{-1})$) as $n, p \to \infty$. Also, $\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}) = np + zn\operatorname{tr}\mathbf{Q}$. As a result, matrix-wise, we obtain

$$\mathbb{E}[\mathbf{Q}] + \frac{1}{z}\mathbf{I}_p = \mathbb{E}[\mathbf{X}_{\cdot k}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{k\cdot}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}] - \frac{1}{z}\frac{1}{n}\mathbb{E}[\mathbf{Q}(p + z\operatorname{tr}\mathbf{Q})] + o_{\|\cdot\|}(1),$$

where $\mathbf{X}_{\cdot k}$ and $\mathbf{X}_{k\cdot}$ is the k-th column and row of $\mathbf{X}$, respectively. As the random $\frac{1}{p}\operatorname{tr}\mathbf{Q} \to m(z)$ as $n, p \to \infty$, take it out of the expectation in the limit and

$$\mathbb{E}[\mathbf{Q}](1 - p/n - z - p/n \cdot zm(z)) = \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which, taking the trace to identify $m(z)$, concludes the proof.

For $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ continuously differentiable with derivatives having at most polynomial growth with respect to p,

$$\mathrm{Var}[f(\mathbf{x})] \leq \sum_{i,j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_i} \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j} \right] = \mathbb{E}\left[(\nabla f(\mathbf{x}))^\mathsf{T} \mathbf{C} \nabla f(\mathbf{x}) \right],$$

where we denote $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Nash–Poincaré inequality

» allow to bound the "fluctuation" of random functionals, e.g., the ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z)$, etc.

» to further establish stochastic convergence (in probability or almost surely) as $n, p \to \infty$.

For $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ continuously differentiable with derivatives having at most polynomial growth with respect to p,

$$\mathrm{Var}[f(\mathbf{x})] \leq \sum_{i,j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_i} \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j} \right] = \mathbb{E}\left[(\nabla f(\mathbf{x}))^\mathsf{T} \mathbf{C} \nabla f(\mathbf{x}) \right],$$

where we denote $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Nash–Poincaré inequality

- allow to bound the "fluctuation" of random functionals, e.g., the ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z)$, etc.
- to further establish stochastic convergence (in probability or almost surely) as $n, p \to \infty$.

For $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ continuously differentiable with derivatives having at most polynomial growth with respect to p,

$$\operatorname{Var}[f(\mathbf{x})] \leq \sum_{i,j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_i} \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j}\right] = \mathbb{E}\left[(\nabla f(\mathbf{x}))^{\mathsf{T}} \mathbf{C} \nabla f(\mathbf{x})\right],$$

where we denote $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Nash–Poincaré inequality

» allow to bound the "fluctuation" of random functionals, e.g., the ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z)$, etc.

» to further establish stochastic convergence (in probability or almost surely) as $n, p \to \infty$.

Extension to non-Gaussian case

For $x \in \mathbb{R}$ a random variable with zero mean and unit variance, $y \sim \mathcal{N}(0,1)$, and f a $(k+2)$-times differentiable function with bounded derivatives,

$$\mathbb{E}[f(x)] - \mathbb{E}[f(y)] = \sum_{\ell=2}^{k} \frac{\kappa_{\ell+1}}{2\ell!} \int_0^1 \mathbb{E}[f^{(\ell+1)} x(t)] t^{(\ell-1)/2} dt + \epsilon_k,$$

where κ_ℓ is the ℓ^{th} cumulant of x, $x(t) = \sqrt{t}x + (1-\sqrt{t})y$, and $|\epsilon_k| \leq C_k \mathbb{E}[|x|^{k+2}] \cdot \sup_t |f^{(k+2)}(t)|$ for some constant C_k only dependent on k.

Interpolation trick

Wigner semicircle law

Let $\mathbf{X} \in \mathbb{R}^{n\times n}$ be symmetric and such that the $\mathbf{X}_{ij} \in \mathbb{R}$, $j \geq i$, are independent zero mean and unit variance random variables. Then, for $\mathbf{Q}(z) = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$, as $n \to \infty$,

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z), \quad \bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_n, \tag{9}$$

with $m(z)$ the unique ST solution to

$$m^2(z) + zm(z) + 1 = 0. \tag{10}$$

The function $m(z)$ is the Stieltjes transform of the probability measure

$$\mu(dx) = \frac{1}{2\pi}\sqrt{(4 - x^2)^+}\, dx, \tag{11}$$

known as the *Wigner semicircle law*.

Proof of semicircle law: leave one out heuristic

Let $\mathbf{Q} = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$ be the resolvent, by diagonal entries of matrix inverse lemma,

$$\mathbf{Q}_{ii} = \left(\mathbf{X}_{ii}/\sqrt{n} - z - \mathbf{x}_i^\mathsf{T} \mathbf{Q}_{-i} \mathbf{x}_i / n \right)^{-1},$$

with $[\mathbf{Q}]_{-i} = (\mathbf{X}_{-i}/\sqrt{n} - z\mathbf{I}_{n-1})^{-1}$, $\mathbf{X}_{-i} \in \mathbb{R}^{(n-1)\times(n-1)}$ the matrix obtained by deleting the i-th row and column from $\mathbf{X}$, and $\mathbf{x}_i \in \mathbb{R}^{n-1}$ the i-th column/row of $\mathbf{X}$ with its i-th entry removed. Summing over i,

$$\frac{1}{n} \operatorname{tr} \mathbf{Q} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\frac{1}{\sqrt{n}} \mathbf{X}_{ii} - z - \frac{1}{n} \mathbf{x}_i^\mathsf{T} \mathbf{Q}_{-i} \mathbf{x}_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{-z - \frac{1}{n} \mathbf{x}_i^\mathsf{T} \mathbf{Q}_{-i} \mathbf{x}_i} + o(1),$$

since $\frac{1}{\sqrt{n}} \mathbf{X}_{ii}$ vanishes as $n \to \infty$. By quadratic form close to the trace, for large n,

$$(\operatorname{tr} \mathbf{Q}/n)^2 + z \operatorname{tr} \mathbf{Q}/n + 1 \simeq o(1),$$

that is $m^2(z) + zm(z) + 1 = 0$ and thus the conclusion.

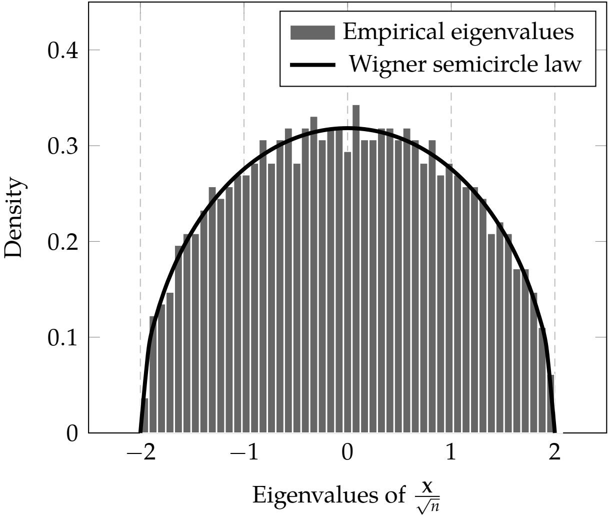

Figure: Histogram of the eigenvalues of $\mathbf{X}/\sqrt{n}$ versus Wigner semicircle law, for $\mathbf{X}$ having standard Gaussian entries and $n = 1\,000$.

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 $\mathbf{X} \in \mathbb{R}^{p\times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{12}$$

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

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{13}$$

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 $\mathbf{X} \in \mathbb{R}^{p \times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{12}$$

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

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{13}$$

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 $\mathbf{X} \in \mathbb{R}^{p \times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{12}$$

  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

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{13}$$

  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 $\mathbf{X} \in \mathbb{R}^{p \times p}$

$$\|\mathbf{X} - \bar{\mathbf{X}}\| \to 0, \tag{12}$$

  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

$$f(\mathbf{X}) - f(\bar{\mathbf{X}}) \to 0, \tag{13}$$

  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 $\mathbf{X} \Rightarrow \operatorname{Var}[f(\mathbf{X})] \downarrow 0$, 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 $\mathbf{X} \Rightarrow \operatorname{Var}[f(\mathbf{X})] \downarrow 0$, 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 $\mathbf{X} \Rightarrow \operatorname{Var}[f(\mathbf{X})] \downarrow 0$, 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 $\mathbf{X} \Rightarrow \operatorname{Var}[f(\mathbf{X})] \downarrow 0$, 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

SCM and MP law

Proof of Marčenko–Pastur law

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 δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

Resolvent as the core object

Objects of interest	Functionals of resolvent $\mathbf{Q_X}(z)$
Empirical Spectral Distribution (ESD) $\mu_{\mathbf{X}}$ of $\mathbf{X}$	Stieltjes transform $m_{\mu_{\mathbf{X}}}(z) = \frac{1}{p} \operatorname{tr} \mathbf{Q_X}(z)$
Linear spectral statistics (LSS): $f(\mathbf{X}) \equiv \frac{1}{p} \sum_i f(\lambda_i(\mathbf{X}))$	Integration of trace of $\mathbf{Q_X}(z)$: $-\frac{1}{2\pi\imath} \oint_\Gamma f(z) \frac{1}{p} \operatorname{tr} \mathbf{Q_X}(z)\, dz$ (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 $\mathbf{v}^\mathsf{T}\mathbf{Q_X}(z)\mathbf{v}$ of $\mathbf{Q_X}$
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 $\mathbf{Q_X}(z)$: $-\frac{1}{2\pi\imath} \oint_\Gamma f(z)\mathbf{v}_1^\mathsf{T}\mathbf{Q_X}(z)\mathbf{v}_2\, dz$

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

$$\mathbf{Q}(z) = \mathbf{U}(\mathbf{\Lambda} - z\mathbf{I}_p)^{-1}\mathbf{U}^\mathsf{T} = \sum_{i=1}^p \frac{\mathbf{u}_i\mathbf{u}_i^\mathsf{T}}{\lambda_i(\mathbf{X}) - z}. \tag{14}$$

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

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) = \frac{1}{p}\sum_{i=1}^p \frac{1}{\lambda_i(\mathbf{X}) - z} = \int \frac{\mu_{\mathbf{X}}(dt)}{t - z}. \tag{15}$$

The Stieltjes transform

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

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{16}$$

Stieltjes transform

For m_μ the Stieltjes transform of a probability measure μ, then

- m_μ 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 μ with support $\operatorname{supp}(\mu)$, the *Stieltjes transform* $m_\mu(z)$ is defined, for all $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t-z}. \tag{16}$$

Stieltjes transform

For m_μ the Stieltjes transform of a probability measure μ, then

- » m_μ 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 μ, we have

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im \left[m_\mu(x + \imath y) \right] \, dx. \tag{17}$$

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

$$f(x) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \left[m_\mu(x + \imath y) \right]. \tag{18}$$

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 μ, we have

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im \left[m_\mu(x + \imath y) \right] dx. \tag{17}$$

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

$$f(x) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \left[m_\mu(x + \imath y) \right]. \tag{18}$$

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 μ, we have

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im\left[m_\mu(x + \imath y)\right] dx. \tag{17}$$

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

$$f(x) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im\left[m_\mu(x + \imath y)\right]. \tag{18}$$

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

$$f(\mathbf{X}) = \frac{1}{p} \sum_{i=1}^{p} f(\lambda_i(\mathbf{X})) = \int f(t) \mu_{\mathbf{X}}(dt), \tag{19}$$

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 Γ and its inside, then

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

(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 $\mathrm{supp}(\mu_{\mathbf{X}})$ (of the ESD $\mu_{\mathbf{X}}$ of $\mathbf{X}$), then

$$f(\mathbf{X}) = \int f(t)\mu_{\mathbf{X}}(dt) = -\int \frac{1}{2\pi\imath} \oint_\Gamma \frac{f(z)\, dz}{t - z} \mu_{\mathbf{X}}(dt) = -\frac{1}{2\pi\imath} \oint_\Gamma f(z) m_{\mu_{\mathbf{X}}}(z)\, dz, \tag{20}$$

for *any* contour Γ that encloses $\mathrm{supp}(\mu_{\mathbf{X}})$, 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 Γ and its inside, then

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

(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

$$f(\mathbf{X}) = \int f(t)\mu_{\mathbf{X}}(dt) = -\int \frac{1}{2\pi\imath}\oint_\Gamma \frac{f(z)\,dz}{t - z}\mu_{\mathbf{X}}(dt) = -\frac{1}{2\pi\imath}\oint_\Gamma f(z)m_{\mu_{\mathbf{X}}}(z)\,dz, \tag{20}$$

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

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

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

- 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

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

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

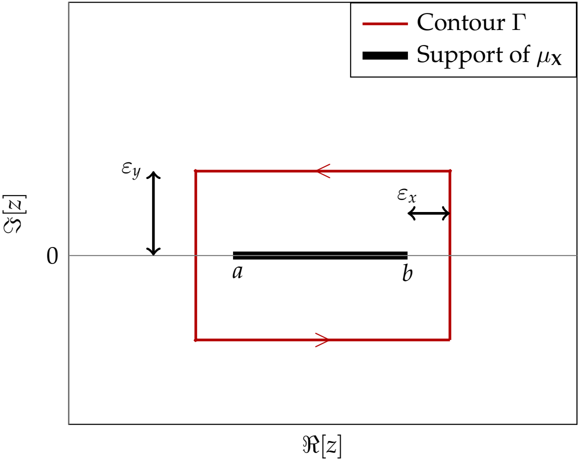

Figure: Illustration of a rectangular contour Γ and support of $\mu_{\mathbf{X}}$ on the complex plane.

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

$$\mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} = -\frac{1}{2\pi\imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{Q}_{\mathbf{X}}(z)\, dz, \tag{21}$$

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

$$\mathbf{u}_i \mathbf{u}_i^{\mathsf{T}} = -\frac{1}{2\pi\imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{Q}_{\mathbf{X}}(z)\, dz, \tag{21}$$

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

$$\mathbf{u}_i \mathbf{u}_i^\mathsf{T} = -\frac{1}{2\pi\imath} \oint_{\Gamma_{\lambda_i(\mathbf{X})}} \mathbf{Q}_{\mathbf{X}}(z)\, dz, \tag{21}$$

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 $\mathbf{X}$,

$$F(\mathbf{X}) = \sum_{i\in\mathcal{I}\subseteq\{1,\ldots,p\}} f(\lambda_i(\mathbf{X}))\mathbf{u}_i\mathbf{u}_i^\mathsf{T}, \quad \mathbf{X} = \sum_{i=1}^{p} \lambda_i(\mathbf{X})\mathbf{u}_i\mathbf{u}_i^\mathsf{T}. \tag{22}$$

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 $\mathcal{I} \subseteq \{1,\ldots,p\}$,

$$F(\mathbf{X}) = -\frac{1}{2\pi\imath}\oint_{\Gamma_\mathcal{I}} f(z)\mathbf{Q}_\mathbf{X}(z)\,dz. \tag{23}$$

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 $\mathbf{X}$,

$$F(\mathbf{X}) = \sum_{i\in\mathcal{I}\subseteq\{1,\ldots,p\}} f(\lambda_i(\mathbf{X}))\mathbf{u}_i\mathbf{u}_i^\mathsf{T}, \quad \mathbf{X} = \sum_{i=1}^{p} \lambda_i(\mathbf{X})\mathbf{u}_i\mathbf{u}_i^\mathsf{T}. \tag{22}$$

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 $\mathcal{I} \subseteq \{1,\ldots,p\}$,

$$F(\mathbf{X}) = -\frac{1}{2\pi\imath}\oint_{\Gamma_\mathcal{I}} f(z)\mathbf{Q}_\mathbf{X}(z)\,dz. \tag{23}$$

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 $\mathbf{X}$,

$$F(\mathbf{X}) = \sum_{i\in\mathcal{I}\subseteq\{1,\ldots,p\}} f(\lambda_i(\mathbf{X}))\mathbf{u}_i\mathbf{u}_i^\mathsf{T}, \quad \mathbf{X} = \sum_{i=1}^{p} \lambda_i(\mathbf{X})\mathbf{u}_i\mathbf{u}_i^\mathsf{T}. \tag{22}$$

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 $\mathcal{I} \subseteq \{1,\ldots,p\}$,

$$F(\mathbf{X}) = -\frac{1}{2\pi\imath}\oint_{\Gamma_\mathcal{I}} f(z)\mathbf{Q}_\mathbf{X}(z)\,dz. \tag{23}$$

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