Large Random Matrices and Applications to Statistical Signal Processing

Jianfeng Yao

Department of Statistics & Act. Sci., The University of Hong Kong

2016 Conference on Applied Mathematics - August 2016 - HKU

(With contributions from Jamal Najim [CNRS, Paris])

Outline

Outline

Large Random Matrices

Random matrices

It is a $N \times N$ matrix

$$\mathbf{Y}_N = \begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & & \vdots \\ Y_{N1} & \cdots & Y_{NN} \end{bmatrix}$$

whose entries $(Y_{ij};\ 1 \leq i, j \leq N)$ are random variables.

Large Random Matrices

Random matrices

It is a $N \times N$ matrix

$$\mathbf{Y}_N = \begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & & \vdots \\ Y_{N1} & \cdots & Y_{NN} \end{bmatrix}$$

whose entries $(Y_{ij};\ 1 \leq i, j \leq N)$ are random variables.

Matrix features

Of interest are the following quantities

- $\mathbf{Y}_N$'s **spectrum** $(\lambda_i,\, 1 \leq i \leq N)$ in particular $\lambda_{\min}$ and $\lambda_{\max}$ (if real spectrum).

Large Random Matrices

Random matrices

It is a $N \times N$ matrix

$$\mathbf{Y}_N = \begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & & \vdots \\ Y_{N1} & \cdots & Y_{NN} \end{bmatrix}$$

whose entries $(Y_{ij};\ 1 \le i,j \le N)$ are random variables.

Matrix features

Of interest are the following quantities

- $\mathbf{Y}_N$'s **spectrum** $(\lambda_i,\, 1 \le i \le N)$ in particular $\lambda_{\min}$ and $\lambda_{\max}$ (if real spectrum).
- linear statistics

$$\operatorname{tr} f(\mathbf{Y}_N) = \sum_{i=1}^{N} f(\lambda_i)$$

Large Random Matrices

Random matrices

It is a $N \times N$ matrix

$$\mathbf{Y}_N = \begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & & \vdots \\ Y_{N1} & \cdots & Y_{NN} \end{bmatrix}$$

whose entries $(Y_{ij};\ 1 \leq i, j \leq N)$ are random variables.

Matrix features

Of interest are the following quantities

- $\mathbf{Y}_N$'s **spectrum** $(\lambda_i,\, 1 \leq i \leq N)$ in particular $\lambda_{\min}$ and $\lambda_{\max}$ (if real spectrum).
- linear statistics

$$\operatorname{tr} f(\mathbf{Y}_N) = \sum_{i=1}^{N} f(\lambda_i)$$

- eigenvectors, etc.

Large Random Matrices

Random matrices

It is a $N \times N$ matrix

$$\mathbf{Y}_N = \begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & & \vdots \\ Y_{N1} & \cdots & Y_{NN} \end{bmatrix}$$

whose entries $(Y_{ij};\ 1 \le i, j \le N)$ are random variables.

Matrix features

Of interest are the following quantities

- $\mathbf{Y}_N$'s **spectrum** $(\lambda_i,\ 1 \le i \le N)$ in particular $\lambda_{\min}$ and $\lambda_{\max}$ (if real spectrum).
- linear statistics

$$\mathrm{tr}\, f(\mathbf{Y}_N) = \sum_{i=1}^{N} f(\lambda_i)$$

- eigenvectors, etc.

Asymptotic regime

Often, the description of the previous features **takes a simplified form** as

$$\boxed{N \to \infty}$$

Large Random Matrices

Random matrices

It is a $N \times N$ matrix

$$\mathbf{Y}_N = \begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & & \vdots \\ Y_{N1} & \cdots & Y_{NN} \end{bmatrix}$$

whose entries $(Y_{ij};\ 1 \le i, j \le N)$ are random variables.

Matrix features

Of interest are the following quantities

- $\mathbf{Y}_N$'s **spectrum** $(\lambda_i,\ 1 \le i \le N)$ in particular $\lambda_{\min}$ and $\lambda_{\max}$ (if real spectrum).
- linear statistics

$$\operatorname{tr} f(\mathbf{Y}_N) = \sum_{i=1}^{N} f(\lambda_i)$$

- eigenvectors, etc.

Asymptotic regime

Often, the description of the previous features **takes a simplified form** as

$$\boxed{N \to \infty}$$

leading to "good enough" approximation in real applications with finite N.

Large Random Matrices: Wigner Matrices

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ a symmetric $N \times N$ matrix with i.i.d. entries **on and above** the diagonal with

$$\mathbb{E}X_{ij} = 0 \text{ and } \mathbb{E}\,|X_{ij}|^2 = 1$$

and $X_{ij} = X_{ji}$ (for symmetry).

- consider the spectrum of **Wigner matrix** $\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$

Large Random Matrices: Wigner Matrices

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ a symmetric $N \times N$ matrix with i.i.d. entries **on and above** the diagonal with

$$\mathbb{E}X_{ij} = 0 \text{ and } \mathbb{E}\,|X_{ij}|^2 = 1$$

and $X_{ij} = X_{ji}$ (for symmetry).

- consider the spectrum of **Wigner matrix** $\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$

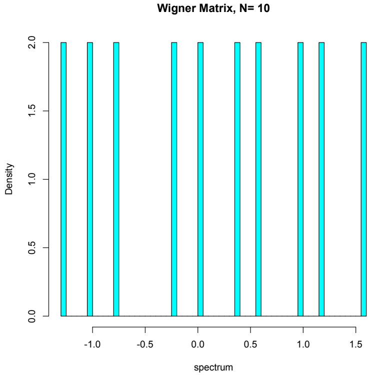


Figure : Histogram of the eigenvalues of $\mathbf{Y}_N$

Large Random Matrices: Wigner Matrices

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ a symmetric $N \times N$ matrix with i.i.d. entries **on and above** the diagonal with

$$\mathbb{E}X_{ij} = 0 \text{ and } \mathbb{E}\,|X_{ij}|^2 = 1$$

and $X_{ij} = X_{ji}$ (for symmetry).

- consider the spectrum of **Wigner matrix** $\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$

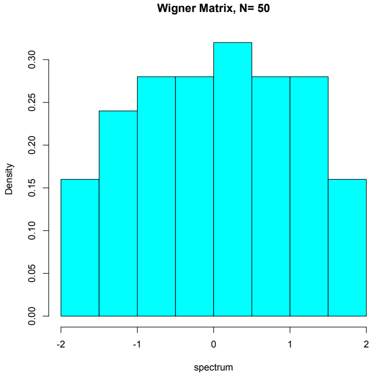


Figure : Histogram of the eigenvalues of $\mathbf{Y}_N$

Large Random Matrices: Wigner Matrices

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ a symmetric $N \times N$ matrix with i.i.d. entries **on and above** the diagonal with

$$\mathbb{E}X_{ij} = 0 \text{ and } \mathbb{E}\,|X_{ij}|^2 = 1$$

and $X_{ij} = X_{ji}$ (for symmetry).

- consider the spectrum of **Wigner matrix** $\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$

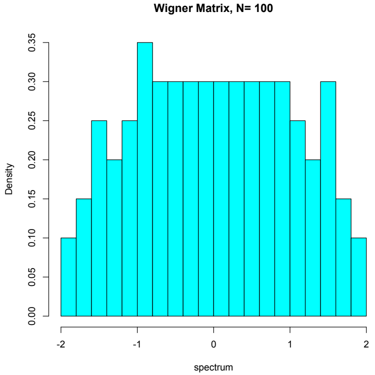


Figure : Histogram of the eigenvalues of $\mathbf{Y}_N$

Large Random Matrices: Wigner Matrices

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ a symmetric $N \times N$ matrix with i.i.d. entries **on and above** the diagonal with

$$\mathbb{E}X_{ij} = 0 \text{ and } \mathbb{E}\,|X_{ij}|^2 = 1$$

and $X_{ij} = X_{ji}$ (for symmetry).

- consider the spectrum of **Wigner matrix** $\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$

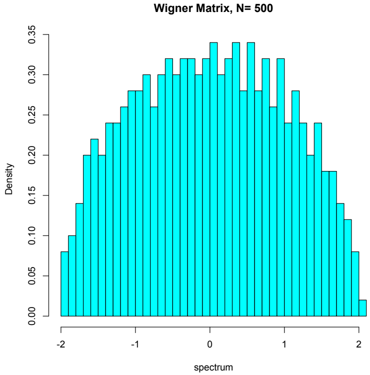


Figure : Histogram of the eigenvalues of $\mathbf{Y}_N$

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ a symmetric $N \times N$ matrix with i.i.d. entries **on and above** the diagonal with

$$\mathbb{E}X_{ij} = 0 \text{ and } \mathbb{E}\,|X_{ij}|^2 = 1$$

and $X_{ij} = X_{ji}$ (for symmetry).

- consider the spectrum of **Wigner matrix** $\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$

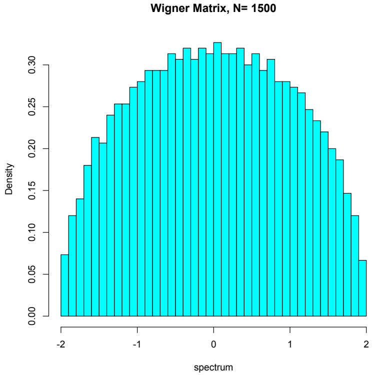


Figure : Histogram of the eigenvalues of $\mathbf{Y}_N$

Large Random Matrices: Wigner Matrices

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ a symmetric $N \times N$ matrix with i.i.d. entries **on and above** the diagonal with

$$\mathbb{E}X_{ij} = 0 \text{ and } \mathbb{E}\,|X_{ij}|^2 = 1$$

and $X_{ij} = X_{ji}$ (for symmetry).

- consider the spectrum of **Wigner matrix** $\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$

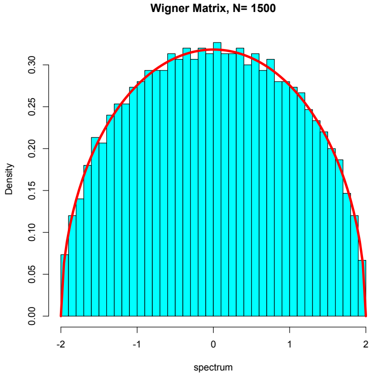


Figure : The semi-circular distribution (in red) with density $x \mapsto \frac{\sqrt{4-x^2}}{2\pi}$

Wigner's theorem (1948)

"The histogram of a Wigner matrix converges to the **semi-circular distribution**"

Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ in the regime where

$$N, n \to \infty \quad \text{and} \quad \frac{N}{n} \to c \in (0, \infty)$$

dimensions of matrix $\mathbf{X}_N$ of the **same order**

Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ in the regime where

$$N, n \to \infty \quad \text{and} \quad \frac{N}{n} \to c \in (0, \infty)$$

dimensions of matrix $\mathbf{X}_N$ of the **same order**

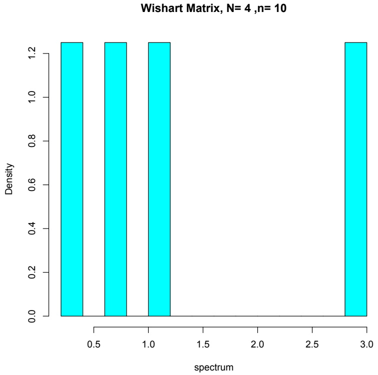


Figure : Spectrum's histogram - $\frac{N}{n} = 0.4$

Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ in the regime where

$$N, n \to \infty \quad \text{and} \quad \frac{N}{n} \to c \in (0, \infty)$$

dimensions of matrix $\mathbf{X}_N$ of the **same order**

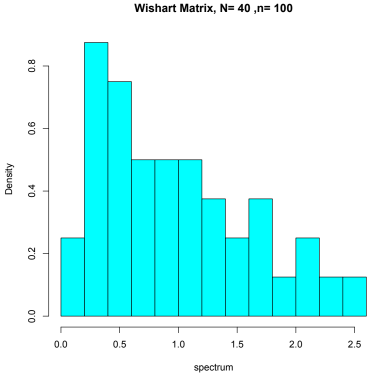


Figure : Spectrum's histogram - $\frac{N}{n} = 0.4$

Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ in the regime where

$$N, n \to \infty \quad \text{and} \quad \frac{N}{n} \to c \in (0, \infty)$$

dimensions of matrix $\mathbf{X}_N$ of the **same order**

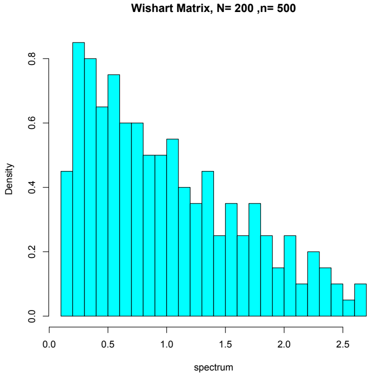


Figure : Spectrum's histogram - $\frac{N}{n} = 0.4$

Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ in the regime where

$$N, n \to \infty \quad \text{and} \quad \frac{N}{n} \to c \in (0, \infty)$$

dimensions of matrix $\mathbf{X}_N$ of the **same order**

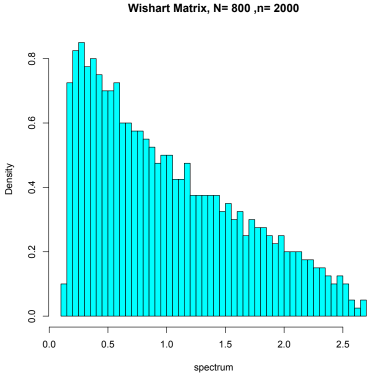


Figure : Spectrum's histogram - $\frac{N}{n} = 0.4$

Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ in the regime where

$$N, n \to \infty \quad \text{and} \quad \frac{N}{n} \to c \in (0, \infty)$$

dimensions of matrix $\mathbf{X}_N$ of the **same order**

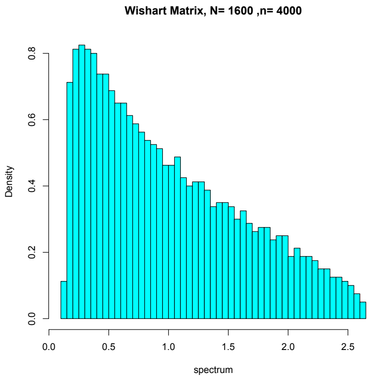


Figure : Spectrum's histogram - $\frac{N}{n} = 0.4$

Large Covariance Matrices : Marčenko-Pastur's theorem

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0\ ,\ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ in the regime where

$$N, n \to \infty \quad \text{and} \quad \frac{N}{n} \to c \in (0, \infty)$$

dimensions of matrix $\mathbf{X}_N$ of the **same order**

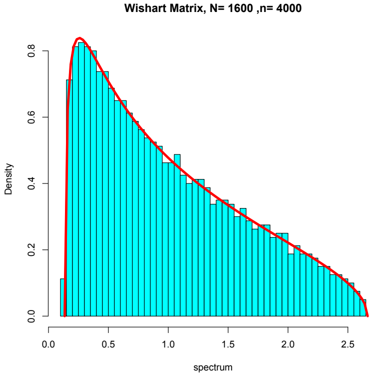


Figure : Marčenko-Pastur's distribution (in red)

Marčenko-Pastur's theorem (1967)

"The histogram of a **Large Covariance Matrix** converges to **Marčenko-Pastur distribution** with given parameter (here **0.4**)"

Matrix model

Let $\mathbf{X}_N$ be a $N \times N$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of matrix $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$ as $N \to \infty$

- In this case, the eigenvalues are **complex**!

Large Non-Hermitian Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times N$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of matrix $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$ as $N \to \infty$

- In this case, the eigenvalues are **complex**!

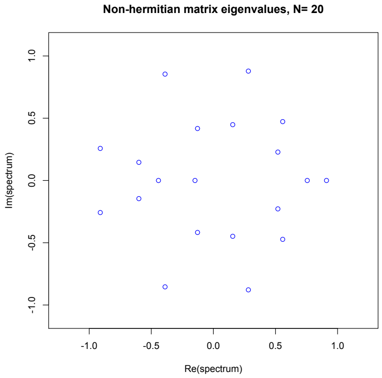


Figure : Distribution of $\mathbf{Y}_N$'s eigenvalues

Large Non-Hermitian Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times N$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of matrix $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$ as $N \to \infty$

- In this case, the eigenvalues are **complex**!

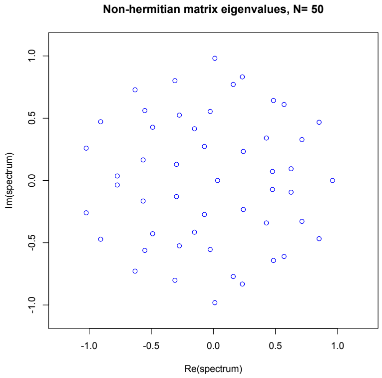


Figure : Distribution of $\mathbf{Y}_N$'s eigenvalues

Large Non-Hermitian Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times N$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of matrix $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$ as $N \to \infty$

- In this case, the eigenvalues are **complex**!

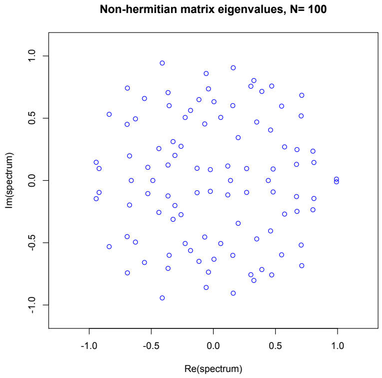


Figure : Distribution of $\mathbf{Y}_N$'s eigenvalues

Large Non-Hermitian Matrices : The Circular Law

Matrix model

Let $\mathbf{X}_N$ be a $N \times N$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of matrix $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$ as $N \to \infty$

- In this case, the eigenvalues are **complex**!

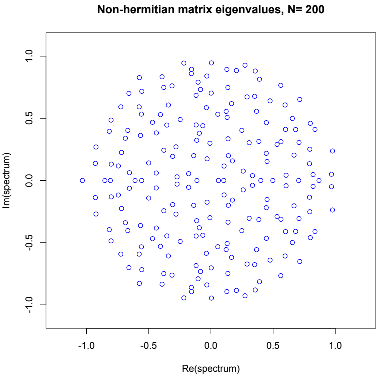


Figure : Distribution of $\mathbf{Y}_N$'s eigenvalues

Matrix model

Let $\mathbf{X}_N$ be a $N \times N$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of matrix $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$ as $N \to \infty$

- In this case, the eigenvalues are **complex**!

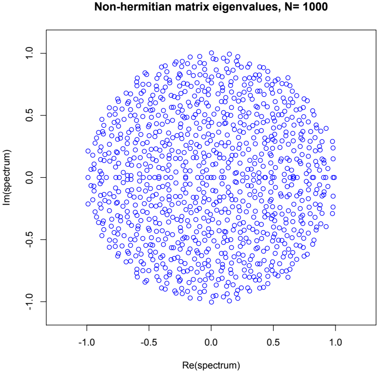


Figure : Distribution of $\mathbf{Y}_N$'s eigenvalues

Large Non-Hermitian Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times N$ matrix with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \ \mathbb{E}|X_{ij}|^2 = 1$$

and consider the spectrum of matrix $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$ as $N \to \infty$

- In this case, the eigenvalues are **complex**!

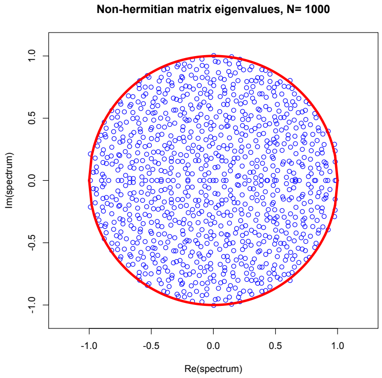


Figure : The circular law (in red)

Theorem: The Circular Law (Ginibre, Girko, Bai, Tao & Vu, etc.)

The spectrum of $\mathbf{Y}_N$ converges to **the uniform probability on the disc**

Motivations

An old history

- Data Analysis (Wishart, 1928)
- Theoretical Physics (from the '50s - Wigner, Dyson, Pastur, etc.)
- Pure mathematics (from the late '80s - non-commutative probability, free probability, operator algebra - Voiculescu, etc.)
- Graph theory (spectrum of the Laplacian)
- Wireless communication (Telatar, 1995 - Verdú, Tse, Shamai, Lévêque, a Parisian group with Loubaton, Debbah, Najim, etc.)

Motivations

An old history

- Data Analysis (Wishart, 1928)
- Theoretical Physics (from the '50s - Wigner, Dyson, Pastur, etc.)
- Pure mathematics (from the late '80s - non-commutative probability, free probability, operator algebra - Voiculescu, etc.)
- Graph theory (spectrum of the Laplacian)
- Wireless communication (Telatar, 1995 - Verdú, Tse, Shamai, Lévêque, a Parisian group with Loubaton, Debbah, Najim, etc.)

Current trends

- Statistics in large dimension (Bai, Bickel & Levina, Ledoit and Wolf, etc.)
- Pure mathematics: universality questions, operator algebra (Tao, Vu, Erdös, Guionnet, etc.)
- Social networks, communication networks
- Neuroscience (non-hermitian models - G. Wainrib)

Outline

Empirical spectral distribution (ESD)

The spectral theorem

For a Hermitian (symmetric) matrix $\mathbf{A}$,

$$\mathbf{A} = \mathbf{U}^* \boldsymbol{\Lambda} \mathbf{U} = \sum_{j=1}^{N} \lambda_j \mathbf{u}_j \mathbf{u}_j^*$$

with its real eigenvalues $\{\lambda_j\}$ and orthonormalized eigenvectors $\{\mathbf{u}_j\}$.

Empirical spectral distribution (ESD)

The spectral theorem

For a Hermitian (symmetric) matrix $\mathbf{A}$,

$$\boxed{\mathbf{A} = \mathbf{U}^* \boldsymbol{\Lambda} \mathbf{U} = \sum_{j=1}^{N} \lambda_j \mathbf{u}_j \mathbf{u}_j^*}$$

with its real eigenvalues $\{\lambda_j\}$ and orthonormalized eigenvectors $\{\mathbf{u}_j\}$.

The ESD

The ESD of $\mathbf{A}$ is the normalized counting measure of the eigenvalues:

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} \qquad \text{that is,} \quad L_N(B) = \frac{1}{N} \#\{\lambda_i \in B\}.$$

Spectral analysis tool (i): by moment convergence

Example of the semi-circle law

- The Hermitian Wigner matrix is $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$;

Moment convergence method:

Note. Computation of the empirical moments $\{m_p(N)\}$ relies on heavy combinatorics.

Spectral analysis tool (i): by moment convergence

Example of the semi-circle law

- The Hermitian Wigner matrix is $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$;
- Moments of its ESD are

$$m_p(N) = \int x^p \, L_n(dx) = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^p = \frac{1}{N}\mathrm{tr}\, \mathbf{Y}_N^p.$$

Moment convergence method:

Note. Computation of the empirical moments $\{m_p(N)\}$ relies on heavy combinatorics.

Spectral analysis tool (i): by moment convergence

Example of the semi-circle law

- The Hermitian Wigner matrix is $\mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N$;
- Moments of its ESD are

$$m_p(N) = \int x^p \, L_n(dx) = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^p = \frac{1}{N}\operatorname{tr} \mathbf{Y}_N^p.$$

Moment convergence method:

1. Prove, in probability or almost surely, that

$$m_p(N) \xrightarrow[N\to\infty]{} \begin{cases} \frac{1}{k+1}\binom{2k}{k} & \text{if } p = 2k \ , \\ 0 & \text{if } p \text{ is odd} \end{cases}$$

Note. Computation of the empirical moments $\{m_p(N)\}$ relies on heavy combinatorics.

Spectral analysis tool (i): by moment convergence

Example of the semi-circle law

- The Hermitian Wigner matrix is $\mathbf{Y}_N = \frac{1}{\sqrt{N}} \mathbf{X}_N$;
- Moments of its ESD are

$$m_p(N) = \int x^p \, L_n(dx) = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^p = \frac{1}{N} \mathrm{tr}\, \mathbf{Y}_N^p .$$

Moment convergence method:

1. Prove, in probability or almost surely, that

$$m_p(N) \xrightarrow[N \to \infty]{} \begin{cases} \frac{1}{k+1}\binom{2k}{k} & \text{if } p = 2k \ , \\ 0 & \text{if } p \text{ is odd} \end{cases}$$

2. Figure out that these are exactly the moment sequence of the semi-circular law:

$$\int_{-2}^{2} x^k \, \mu_{sc}(dx) \quad = \begin{cases} \frac{1}{k+1}\binom{2k}{k} & \text{if } p = 2k \ , \\ 0 & \text{if } p \text{ is odd} \end{cases}$$

Note. Computation of the empirical moments $\{m_p(N)\}$ relies on heavy combinatorics.

Spectral analysis tool (i): by moment convergence

Example of the semi-circle law

- The Hermitian Wigner matrix is $\mathbf{Y}_N = \frac{1}{\sqrt{N}} \mathbf{X}_N$;
- Moments of its ESD are

$$m_p(N) = \int x^p \, L_n(dx) = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^p = \frac{1}{N} \mathrm{tr}\, \mathbf{Y}_N^p.$$

Moment convergence method:

1. Prove, in probability or almost surely, that

$$m_p(N) \xrightarrow[N\to\infty]{} \begin{cases} \frac{1}{k+1}\binom{2k}{k} & \text{if } p = 2k \ , \\ 0 & \text{if } p \text{ is odd} \end{cases}$$

2. Figure out that these are exactly the moment sequence of the semi-circular law:

$$\int_{-2}^{2} x^k \, \mu_{sc}(dx) \quad = \begin{cases} \frac{1}{k+1}\binom{2k}{k} & \text{if } p = 2k \ , \\ 0 & \text{if } p \text{ is odd} \end{cases}$$

3. Conclude, by Carleman's criterion, that $L_N \Longrightarrow \mu_{sc}$.

Note. Computation of the empirical moments $\{m_p(N)\}$ relies on heavy combinatorics.

Spectral analysis tool (ii): The Stieltjes Transform

- The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x - z} \mu(dx) \ , \quad z \in \mathbb{C}^+ \ ,$$

Spectral analysis tool (ii): The Stieltjes Transform

- The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} \mu(dx) \ , \quad z \in \mathbb{C}^+ \ ,$$

Spectral analysis tool (ii): The Stieltjes Transform

- The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z}\mu(dx) \ , \quad z \in \mathbb{C}^+ \ ,$$

- the transform characterize the measure through the inversion formula: for all continuity points a, b of μ,

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \int_a^b s_\mu(x + \mathbf{i}y)\, dx \ ,$$

Spectral analysis tool (ii): The Stieltjes Transform

- The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x - z} \mu(dx) \ , \quad z \in \mathbb{C}^+ \ ,$$

- the transform characterize the measure through the inversion formula: for all continuity points a, b of μ,

$$\mu([a, b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \int_a^b s_\mu(x + \mathbf{i}y) \, dx \ ,$$

Spectral analysis tool (ii): The Stieltjes Transform

- The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x - z} \mu(dx) \ , \quad z \in \mathbb{C}^+ \ ,$$

- the transform characterize the measure through the inversion formula: for all continuity points a, b of μ,

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \int_a^b s_\mu(x + \mathbf{i}y) \, dx \ ,$$

Examples

1. ESD of a Hermitian matrix A: $\quad s_{L_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z}$

(by convention, $\sqrt{z}$ has positive imaginary part for $z \in \mathbb{C}^+$)

Spectral analysis tool (ii): The Stieltjes Transform

- The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} \mu(dx) \ , \quad z \in \mathbb{C}^+ \ ,$$

- the transform characterize the measure through the inversion formula: for all continuity points a, b of μ,

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \int_a^b s_\mu(x + \mathbf{i}y) \, dx \ ,$$

Examples

1. ESD of a Hermitian matrix A: $\quad s_{L_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z}$
2. Semi-circle law: $\quad s_{\mu_{sc}}(z) = \int_{-2}^{2} \frac{1}{x-z} \frac{1}{2\pi} \sqrt{4-x^2} dx = \frac{1}{2} \left(-z + \sqrt{z^2 - 4} \right).$

(by convention, $\sqrt{z}$ has positive imaginary part for $z \in \mathbb{C}^+$)

Spectral analysis tool (ii): The Stieltjes Transform

- The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} \mu(dx) \;, \quad z \in \mathbb{C}^+ \;,$$

- the transform characterize the measure through the inversion formula: for all continuity points a, b of μ,

$$\mu([a,b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im \int_a^b s_\mu(x + \mathbf{i}y)\, dx \;,$$

Examples

1. ESD of a Hermitian matrix A: $\quad s_{L_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z}$
2. Semi-circle law: $\quad s_{\mu_{sc}}(z) = \int_{-2}^{2} \frac{1}{x-z} \frac{1}{2\pi} \sqrt{4-x^2} dx = \frac{1}{2} \left(-z + \sqrt{z^2 - 4} \right).$
3. Marčenko-Pastur Law:

$$s_{\mu_{MP}}(z) = \int_a^b \frac{1}{x-z} \frac{1}{2\pi c x} \sqrt{(b-x)(x-a)} dx = \frac{1 - c - z - \sqrt{(z-a)(z-b)}}{2cz}.$$

(by convention, $\sqrt{z}$ has positive imaginary part for $z \in \mathbb{C}^+$)

Why does RMT prefer Stieltjes transform ?

- For a Hermitian matrix $\mathbf{A}$,

$$
\begin{aligned}
s_{L_N}(z) &= \text{Stieltjes transform of } \left(\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}\right) \\
&= \frac{1}{N}\sum_{1}^{N}\frac{1}{\lambda_i - z}
\end{aligned}
$$

Why does RMT prefer Stieltjes transform ?

- For a Hermitian matrix $\mathbf{A}$,

$$
\begin{aligned}
s_{L_N}(z) &= \text{Stieltjes transform of } \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} \right) \\
&= \frac{1}{N} \sum_{1}^{N} \frac{1}{\lambda_i - z}
\end{aligned}
$$

Why does RMT prefer Stieltjes transform ?

- For a Hermitian matrix $\mathbf{A}$,

$$
\begin{aligned}
s_{L_N}(z) &= \text{Stieltjes transform of } \left(\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}\right) \\
&= \frac{1}{N}\sum_{1}^{N}\frac{1}{\lambda_i - z} \\
&= \frac{1}{N}\mathsf{tr}\,(\mathbf{A} - z\boldsymbol{I})^{-1}.
\end{aligned}
$$

- Write

$$
\mathbf{A} = \begin{pmatrix} a_{11} & \mathbf{a}_1^* \\ \mathbf{a}_1 & \mathbf{A}_1 \end{pmatrix},
$$

and similarly for the diagonal elements $a_{22}, \ldots, a_{NN}$ to get the sequence of $N-1$ dimensional vectors $\{\mathbf{a}_k\}$ and principal submatrices $\{\mathbf{A}_k\}$;

This shows how matrix algebra helps the study of the ESD of a large matrix $\mathbf{A}$.

Why does RMT prefer Stieltjes transform ?

- For a Hermitian matrix $\mathbf{A}$,

$$\begin{aligned} s_{L_N}(z) &= \text{Stieltjes transform of } \left(\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}\right) \\ &= \frac{1}{N}\sum_{1}^{N}\frac{1}{\lambda_i - z} \\ &= \frac{1}{N}\mathrm{tr}\,(\mathbf{A} - z\boldsymbol{I})^{-1}\,. \end{aligned}$$

- Write

$$\mathbf{A} = \begin{pmatrix} a_{11} & \mathbf{a}_1^* \\ \mathbf{a}_1 & \mathbf{A}_1 \end{pmatrix},$$

and similarly for the diagonal elements $a_{22}, \ldots, a_{NN}$ to get the sequence of $N-1$ dimensional vectors $\{\mathbf{a}_k\}$ and principal submatrices $\{\mathbf{A}_k\}$;

- By Schur complement

$$s_{L_N}(z) = \frac{1}{N}\mathrm{tr}\,(\mathbf{A} - z\boldsymbol{I})^{-1} = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{a_{kk} - z - \mathbf{a}_k^*(\mathbf{A}_k - z\mathbf{I})^{-1}\mathbf{a}_k}$$

This shows how matrix algebra helps the study of the ESD of a large matrix $\mathbf{A}$.

Sketched proof of Wigner's semi-circle law

- Now we let

$$\mathbf{A} = \mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N = \frac{1}{\sqrt{N}}\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{NN} \\ & x_{22} & \cdots & x_{2N} \\ & \vdots & \vdots & \vdots \\ & & & x_{NN} \end{pmatrix}$$

where $\{x_{ij} : i \leq j\}$ are i.i.d. with mean 0 and variance 1.

So,

$$a_{ij} = \frac{1}{\sqrt{N}} x_{ij}, \quad \mathbf{a}_k = \frac{1}{\sqrt{N}}\mathbf{x}_k, \quad \mathbf{A}_k = \frac{1}{\sqrt{N}}\mathbf{X}_k, \quad \text{etc.}$$

Sketched proof of Wigner's semi-circle law

- Now we let

$$\mathbf{A} = \mathbf{Y}_N = \frac{1}{\sqrt{N}}\mathbf{X}_N = \frac{1}{\sqrt{N}}\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{NN} \\ & x_{22} & \cdots & x_{2N} \\ & \vdots & \vdots & \vdots \\ & & & x_{NN} \end{pmatrix}$$

where $\{x_{ij} : \ i \leq j\}$ are i.i.d. with mean 0 and variance 1.
So,

$$a_{ij} = \frac{1}{\sqrt{N}}x_{ij}, \quad \mathbf{a}_k = \frac{1}{\sqrt{N}}\mathbf{x}_k, \quad \mathbf{A}_k = \frac{1}{\sqrt{N}}\mathbf{X}_k, \quad \text{etc.}$$

- We have

$$\begin{aligned} s_{L_N}(z) &= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{a_{kk} - z - \mathbf{a}_k^*(\mathbf{A}_k - z\mathbf{I})^{-1}\mathbf{a}_k} \\ &= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\frac{1}{\sqrt{N}}x_{kk} - z - \frac{1}{N}\mathbf{x}_k^*\left(\frac{1}{\sqrt{N}}\mathbf{X}_k - z\mathbf{I}\right)^{-1}\mathbf{x}_k} \end{aligned}$$

Sketched proof of Wigner's semi-circle law (cont.)

$$s_{L_N}(z) = \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_N - z\mathbf{I} \right)^{-1} = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{\frac{1}{\sqrt{N}} x_{kk} - z - \frac{1}{N} \mathbf{x}_k^* \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k}$$

- When $N \to \infty$, $\dfrac{1}{\sqrt{N}} x_{kk} \to 0$;

Sketched proof of Wigner's semi-circle law (cont.)

$$s_{L_N}(z) = \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_N - z\mathbf{I} \right)^{-1} = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{\frac{1}{\sqrt{N}} x_{kk} - z - \frac{1}{N} \mathbf{x}_k^* \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k}$$

- When $N \to \infty$, $\frac{1}{\sqrt{N}} x_{kk} \to 0$;
-

$$\begin{aligned}
\frac{1}{N} \mathbf{x}_k^* \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k &= \frac{1}{N} \operatorname{tr} \mathbf{x}_k^* \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k \\
&= \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k \mathbf{x}_k^* \\
&\simeq \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{I}_{N-1} \\
&= \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \\
&\simeq s_{L_N}(z).
\end{aligned}$$

Sketched proof of Wigner's semi-circle law (cont.)

$$s_{L_N}(z) = \frac{1}{N}\operatorname{tr}\left(\frac{1}{\sqrt{N}}\mathbf{X}_N - z\mathbf{I}\right)^{-1} = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\frac{1}{\sqrt{N}}x_{kk} - z - \frac{1}{N}\mathbf{x}_k^* \left(\frac{1}{\sqrt{N}}\mathbf{X}_k - z\mathbf{I}\right)^{-1}\mathbf{x}_k}$$

- When $N \to \infty$, $\frac{1}{\sqrt{N}}x_{kk} \to 0$;
- $$\begin{aligned}\frac{1}{N}\mathbf{x}_k^*\left(\frac{1}{\sqrt{N}}\mathbf{X}_k - z\mathbf{I}\right)^{-1}\mathbf{x}_k &= \frac{1}{N}\operatorname{tr}\,\mathbf{x}_k^*\left(\frac{1}{\sqrt{N}}\mathbf{X}_k - z\mathbf{I}\right)^{-1}\mathbf{x}_k \\ &= \frac{1}{N}\operatorname{tr}\left(\frac{1}{\sqrt{N}}\mathbf{X}_k - z\mathbf{I}\right)^{-1}\mathbf{x}_k\mathbf{x}_k^* \\ &\simeq \frac{1}{N}\operatorname{tr}\left(\frac{1}{\sqrt{N}}\mathbf{X}_k - z\mathbf{I}\right)^{-1}\mathbf{I}_{N-1} \\ &= \frac{1}{N}\operatorname{tr}\left(\frac{1}{\sqrt{N}}\mathbf{X}_k - z\mathbf{I}\right)^{-1} \\ &\simeq s_{L_N}(z).\end{aligned}$$
- So $s_{L_N}(z)$ does have a limit $s(z)$ satisfying

$$s = \frac{1}{-z-s}, \quad \text{that is,} \quad s^2 + zs + 1 = 0.$$

Sketched proof of Wigner's semi-circle law (cont.)

$$s_{L_N}(z) = \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_N - z\mathbf{I} \right)^{-1} = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{\frac{1}{\sqrt{N}} x_{kk} - z - \frac{1}{N} \mathbf{x}_k^* \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k}$$

- When $N \to \infty$, $\frac{1}{\sqrt{N}} x_{kk} \to 0$;
-

$$\begin{aligned} \frac{1}{N} \mathbf{x}_k^* \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k &= \frac{1}{N} \operatorname{tr} \mathbf{x}_k^* \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k \\ &= \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{x}_k \mathbf{x}_k^* \\ &\simeq \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \mathbf{I}_{N-1} \\ &= \frac{1}{N} \operatorname{tr} \left(\frac{1}{\sqrt{N}} \mathbf{X}_k - z\mathbf{I} \right)^{-1} \\ &\simeq s_{L_N}(z). \end{aligned}$$

- So $s_{L_N}(z)$ does have a limit $s(z)$ satisfying

$$s = \frac{1}{-z - s}, \quad \text{that is,} \quad s^2 + zs + 1 = 0.$$

- Solving the equation, we find $s(z) = \frac{1}{2}\left(-z + \sqrt{z^2 - 4}\right)$, i.e. $S_{\mu_{sc}}(z)$!

Outline

Wishart Matrices I

The model

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \;, \quad \mathbb{E}|X_{ij}|^2 = 1 \;.$$

Matrix $\mathbf{X}_N$ is a n-sample of N-dimensional vectors:

$$\mathbf{X}_N = [\mathbf{X}_{\cdot 1} \; \cdots \; \mathbf{X}_{\cdot n}] \quad \text{with} \quad \mathbb{E}\mathbf{X}_{\cdot 1}\mathbf{X}_{\cdot 1}^* = \mathbf{I}_N \;.$$

Wishart Matrices I

The model

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \;, \quad \mathbb{E}|X_{ij}|^2 = 1 \;.$$

Matrix $\mathbf{X}_N$ is a n-sample of N-dimensional vectors:

$$\mathbf{X}_N = [\mathbf{X}_{\cdot 1} \;\cdots\; \mathbf{X}_{\cdot n}] \quad \text{with} \quad \mathbb{E}\mathbf{X}_{\cdot 1}\mathbf{X}_{\cdot 1}^* = \mathbf{I}_N \;.$$

Objective

- to describe **the limiting spectrum** of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ as

$$\frac{N}{n} \xrightarrow[n\to\infty]{} c \in (0,\infty) \;.$$

i.e. dimensions of matrix $\mathbf{X}_N$ are **of the same order**.

Wishart Matrices II

The usual case $N << n$

Assume N fixed and $n \to \infty$.

The usual case $N << n$

Assume N fixed and $n \to \infty$. Since

$$\mathbb{E}\mathbf{X}_{\cdot 1}\mathbf{X}_{\cdot 1}^* = \mathbf{I}_N \ ,$$

L.L.N implies

$$\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{\cdot i}\mathbf{X}_{\cdot i}^* \quad \xrightarrow[n\to\infty]{a.s.} \quad \mathbf{I}_N$$

The usual case $N << n$

Assume N fixed and $n \to \infty$. Since

$$\mathbb{E}\mathbf{X}_{\cdot 1}\mathbf{X}_{\cdot 1}^* = \mathbf{I}_N \ ,$$

L.L.N implies

$$\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{\cdot i}\mathbf{X}_{\cdot i}^* \quad \xrightarrow[n\to\infty]{a.s.} \quad \mathbf{I}_N$$

In particular,

- all the eigenvalues of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ converge to 1,

Wishart Matrices II

The usual case $N << n$

Assume N fixed and $n \to \infty$. Since

$$\mathbb{E}\mathbf{X}_{\cdot 1}\mathbf{X}_{\cdot 1}^* = \mathbf{I}_N \ ,$$

L.L.N implies

$$\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{\cdot i}\mathbf{X}_{\cdot i}^* \quad \xrightarrow[n\to\infty]{a.s.} \quad \mathbf{I}_N$$

In particular,

- all the eigenvalues of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ converge to 1,
- equivalently, the spectral measure of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ converges to δ_1.

Wishart Matrices II

The usual case $N << n$

Assume N fixed and $n \to \infty$. Since

$$\mathbb{E}\mathbf{X}_{\cdot 1}\mathbf{X}_{\cdot 1}^* = \mathbf{I}_N \ ,$$

L.L.N implies

$$\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{\cdot i}\mathbf{X}_{\cdot i}^* \quad \xrightarrow[n\to\infty]{a.s.} \quad \mathbf{I}_N$$

In particular,

- all the eigenvalues of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ converge to 1,
- equivalently, the spectral measure of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ converges to δ_1.

A priori observation $\#$ 1

If the ratio of dimensions $c \searrow 0$, then the spectral measure should look like a Dirac measure at point 1.

Wishart Matrices III

The case where $c > 1$

Recall that $\mathbf{X}_N$ is $N \times n$ matrix and $c = \lim \frac{N}{n}$.

Wishart Matrices III

The case where $c > 1$

Recall that $\mathbf{X}_N$ is $N \times n$ matrix and $c = \lim \frac{N}{n}$.

If $N > n$, then $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ is **rank-defficient** and has rank n;

Wishart Matrices III

The case where $c > 1$

Recall that $\mathbf{X}_N$ is $N \times n$ matrix and $c = \lim \frac{N}{n}$.

If $N > n$, then $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ is **rank-defficient** and has rank n;

- in this case, eigenvalue 0 has multiplicity $N - n$ and the spectral measure writes:

$$L_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i} = \frac{1}{N}\sum_{i=1}^{n}\delta_{\lambda_i} + \frac{N-n}{N}\delta_0$$

Wishart Matrices III

The case where $c > 1$

Recall that $\mathbf{X}_N$ is $N \times n$ matrix and $c = \lim \frac{N}{n}$.

If $N > n$, then $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ is **rank-defficient** and has rank n;

- in this case, eigenvalue 0 has multiplicity $N - n$ and the spectral measure writes:

$$L_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i} = \frac{1}{N}\sum_{i=1}^{n}\delta_{\lambda_i} + \frac{N-n}{N}\delta_0$$

- The limiting spectral measure of L_N **necessarily features a Dirac measure at 0**:

$$\frac{N-n}{N}\delta_0 \longrightarrow \left(1 - \frac{1}{c}\right)\delta_0 \quad \text{as} \quad \frac{N}{n} \to c \; .$$

Wishart Matrices III

The case where $c > 1$

Recall that $\mathbf{X}_N$ is $N \times n$ matrix and $c = \lim \frac{N}{n}$.

If $N > n$, then $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$ is **rank-defficient** and has rank n;

- in this case, eigenvalue 0 has multiplicity $N - n$ and the spectral measure writes:

$$L_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i} = \frac{1}{N}\sum_{i=1}^{n}\delta_{\lambda_i} + \frac{N-n}{N}\delta_0$$

- The limiting spectral measure of L_N **necessarily features a Dirac measure at 0**:

$$\frac{N-n}{N}\delta_0 \longrightarrow \left(1 - \frac{1}{c}\right)\delta_0 \quad \text{as} \quad \frac{N}{n} \to c\ .$$

A priori observation #2

If $c > 1$, then the limiting spectral measure will feature a Dirac measure at 0 with weight $1 - \frac{1}{c}$.

Simulations

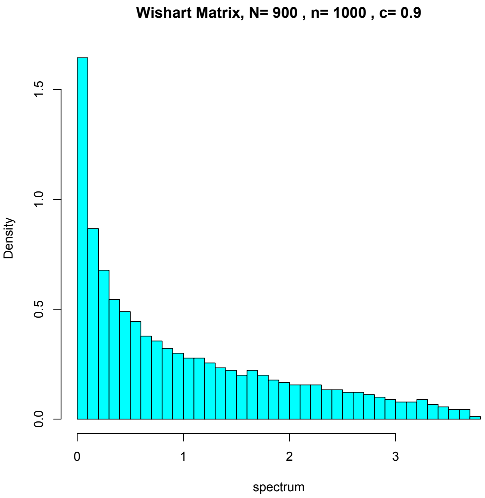


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Simulations

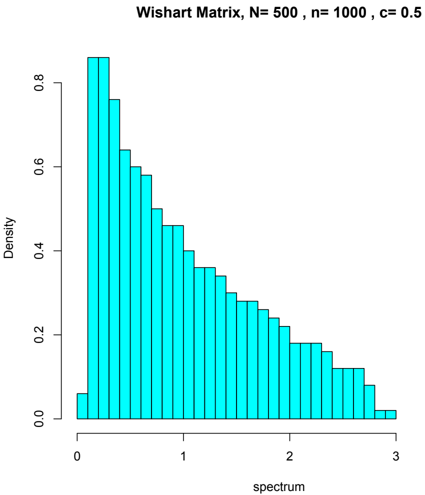


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Simulations

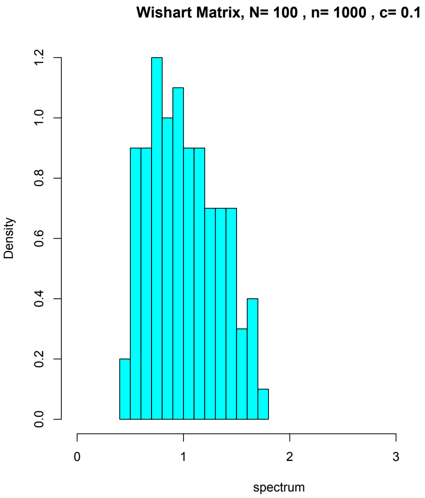


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Simulations

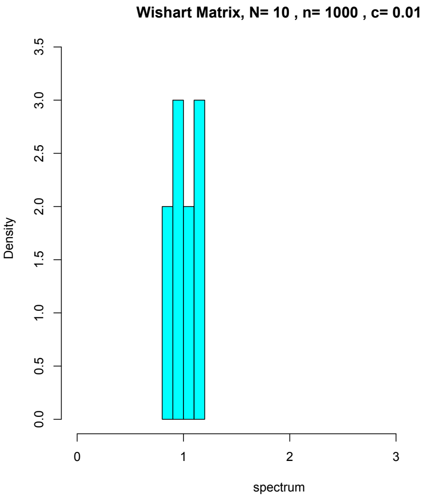


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Marčenko-Pastur theorem

Theorem

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \quad \mathbb{E}|X_{ij}|^2 = 1 \ .$$

with N and n **of the same order** and L_N its **spectral measure**:

Marčenko-Pastur theorem

Theorem

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \quad \mathbb{E}|X_{ij}|^2 = 1 \ .$$

with N and n **of the same order** and L_N its **spectral measure**:

$$c_n \stackrel{\triangle}{=} \frac{N}{n} \xrightarrow[n\to\infty]{} c \in (0,\infty) \ , \quad L_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} \ , \quad \lambda_i = \lambda_i\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right)$$

Marčenko-Pastur theorem

Theorem

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \; , \quad \mathbb{E}|X_{ij}|^2 = 1 \; .$$

with N and n **of the same order** and L_N its **spectral measure**:

$$c_n \triangleq \frac{N}{n} \xrightarrow[n\to\infty]{} c \in (0, \infty) \; , \quad L_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} \; , \quad \lambda_i = \lambda_i\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right)$$

- Then almost surely

$$L_N \xrightarrow[N,n\to\infty]{} \mu_{\mathrm{MP}} \quad \text{in distribution}$$

Marčenko-Pastur theorem

Theorem

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \ , \quad \mathbb{E}|X_{ij}|^2 = 1 \ .$$

with N and n **of the same order** and L_N its **spectral measure**:

$$c_n \triangleq \frac{N}{n} \xrightarrow[n\to\infty]{} c \in (0,\infty) \ , \quad L_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} \ , \quad \lambda_i = \lambda_i \left(\frac{1}{n} \mathbf{X}_N \mathbf{X}_N^* \right)$$

- Then almost surely

$$L_N \xrightarrow[N,n\to\infty]{} \mu_{\mathrm{MP}} \quad \text{in distribution}$$

where μ_{MP} is **Marčenko-Pastur** distribution:

Marčenko-Pastur theorem

Theorem

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0 \;, \quad \mathbb{E}|X_{ij}|^2 = 1 \;.$$

with N and n **of the same order** and L_N its **spectral measure**:

$$c_n \triangleq \frac{N}{n} \xrightarrow[n\to\infty]{} c \in (0,\infty) \;, \quad L_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} \;, \quad \lambda_i = \lambda_i\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right)$$

- Then almost surely

$$L_N \xrightarrow[N,n\to\infty]{} \mu_{\mathrm{MP}} \quad \text{in distribution}$$

where μ_{MP} is **Marčenko-Pastur** distribution:

$$\mu_{\mathrm{MP}}(dx) \quad = \quad \left(1-\frac{1}{c}\right)^{+}\delta_0(dx) + \frac{\sqrt{(b-x)(x-a)}}{2\pi c x} 1_{[a,b]}(x)\, dx$$

Marčenko-Pastur theorem

Theorem

- Consider a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries

$$\mathbb{E}X_{ij} = 0\ , \quad \mathbb{E}|X_{ij}|^2 = 1\ .$$

with N and n **of the same order** and L_N its **spectral measure**:

$$c_n \triangleq \frac{N}{n} \xrightarrow[n\to\infty]{} c \in (0,\infty)\ , \quad L_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i}\ , \quad \lambda_i = \lambda_i\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right)$$

- Then almost surely

$$L_N \xrightarrow[N,n\to\infty]{} \mu_{\mathrm{MP}} \quad \text{in distribution}$$

where μ_{MP} is **Marčenko-Pastur** distribution:

$$\begin{aligned} \mu_{\mathrm{MP}}(dx) \quad &= \quad \left(1-\frac{1}{c}\right)^+ \delta_0(dx) + \frac{\sqrt{(b-x)(x-a)}}{2\pi c x} 1_{[a,b]}(x)\, dx \\ \text{with} \quad & \quad \left\{ \begin{array}{lcl} a &=& (1-\sqrt{c})^2 \\ b &=& (1+\sqrt{c})^2 \end{array} \right. \end{aligned}$$

Simulations vs MP distribution

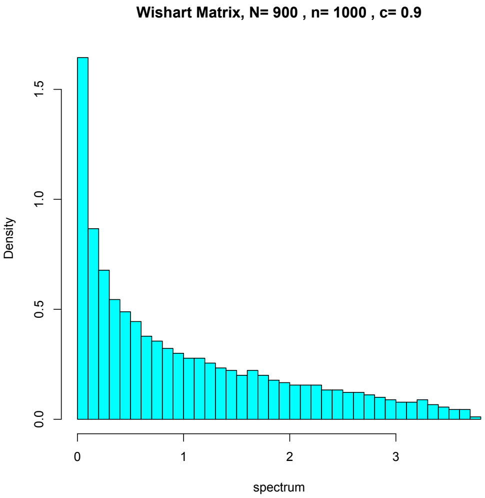


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Simulations vs MP distribution

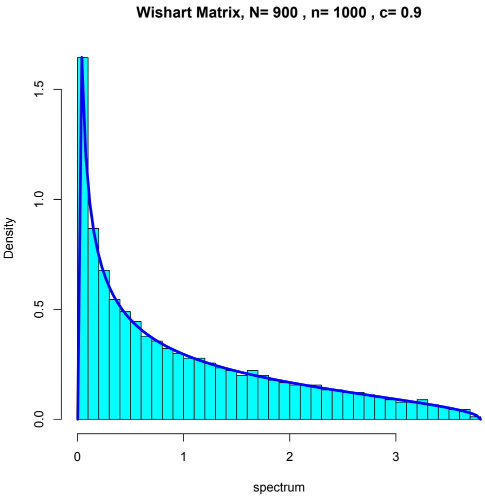


Figure : Marčenko-Pastur distribution for $c = 0.9$

Simulations vs MP distribution

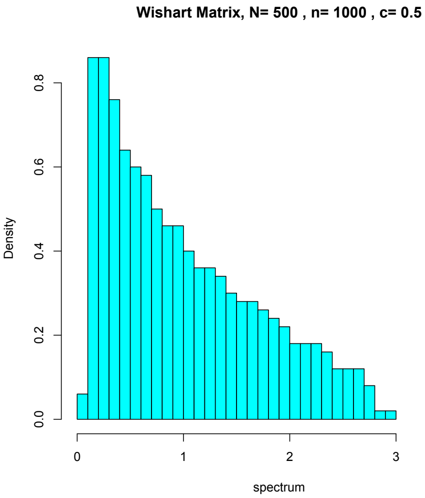


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Simulations vs M̌P distribution

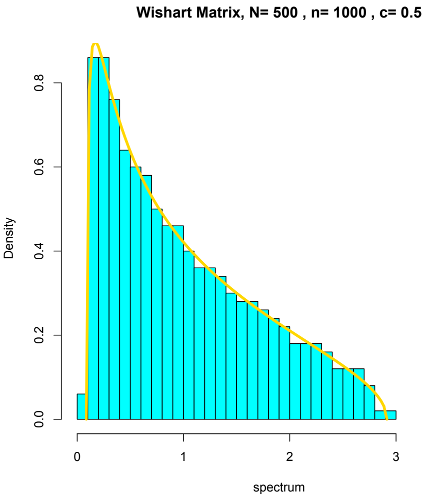


Figure : Marčenko-Pastur distribution for $c = 0.5$

Simulations vs MP distribution

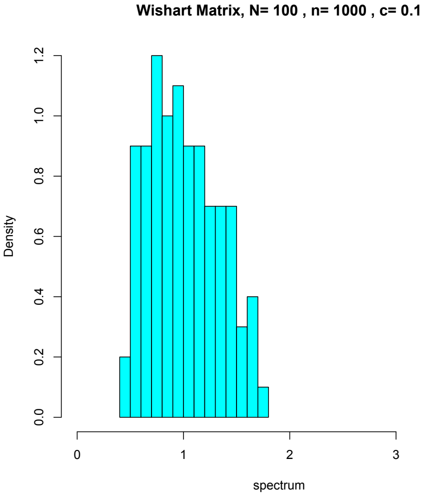


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Simulations vs MP distribution

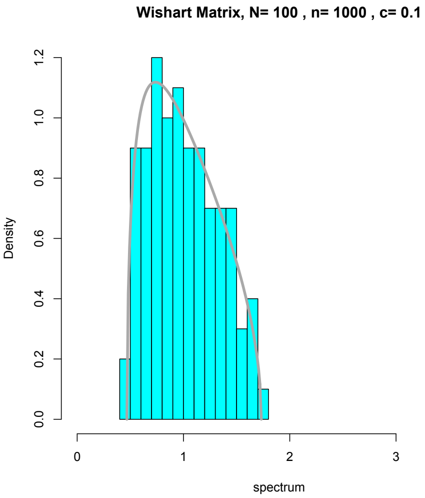


Figure : Marčenko-Pastur distribution for $c = 0.1$

Simulations vs MP distribution

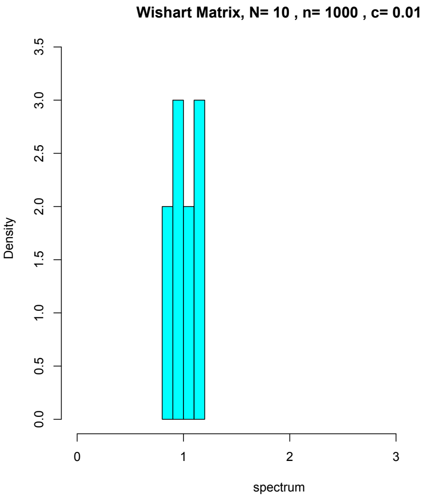


Figure : Histogram of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$, $\sigma^2 = 1$

Simulations vs M̌P distribution

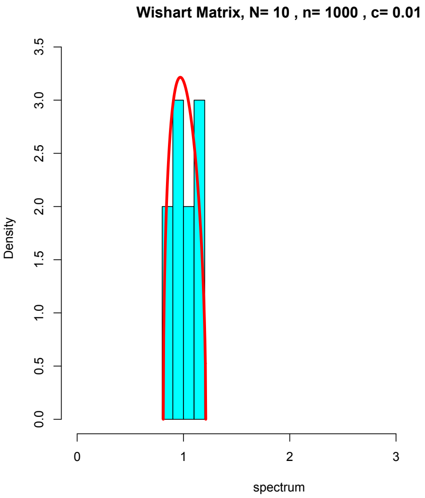


Figure : Marčenko-Pastur distribution for $c = 0.01$

Remarks I

- Marčenko-Pastur theorem describes the **global regime** of the spectrum.

Remarks I

- Marčenko-Pastur theorem describes the **global regime** of the spectrum.
- Convergence in distribution: For a given realization and every test function $\phi : \mathbb{R} \to \mathbb{R}$ (continuous and bounded), the theorem states:

$$\frac{1}{N} \sum_{i=1}^{N} \phi(\lambda_i) \xrightarrow[N,n\to\infty]{} \int \phi(x) \mu_{\mathrm{MP}}(dx) \ .$$

Remarks I

- Marčenko-Pastur theorem describes the **global regime** of the spectrum.
- Convergence in distribution: For a given realization and every test function $\phi : \mathbb{R} \to \mathbb{R}$ (continuous and bounded), the theorem states:

$$\frac{1}{N} \sum_{i=1}^{N} \phi(\lambda_i) \xrightarrow[N,n\to\infty]{} \int \phi(x) \mu_{\mathrm{MP}}(dx) \ .$$

- The Dirac measure at zero is **an artifact** due to the dimensions of the matrix if

$$N > n \quad \text{(cf. infra)} \ .$$

Remarks II

What if $c \searrow 0$?

- If $c \to 0$, that is $n >> N$, then typical from the usual regime **"small dimensional data vs large samples"**.

Remarks II

What if $c \searrow 0$?

- If $c \to 0$, that is $n >> N$, then typical from the usual regime **"small dimensional data vs large samples"**.
- the support of Marčenko-Pastur distribution

  $$[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$$

  concentrates around $\{1\}$ and

  $$\mu_{\text{MP}} \xrightarrow[c \to 0]{} \delta_1 \ .$$

Remarks II

What if $c \searrow 0$?

- If $c \to 0$, that is $n >> N$, then typical from the usual regime **"small dimensional data vs large samples"**.
- the support of Marčenko-Pastur distribution

$$[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$$

concentrates around $\{1\}$ and

$$\mu_{\text{MP}} \xrightarrow[c \to 0]{} \delta_1 \ .$$

- In accordance with **a priori information # 1**

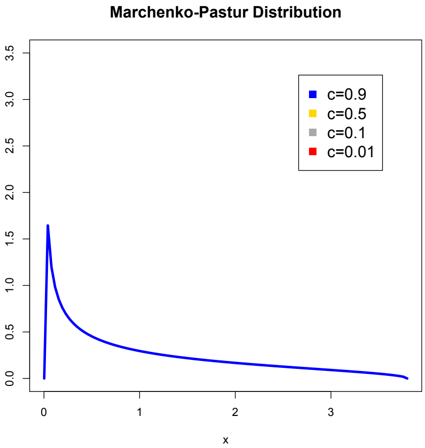


Figure : M̌P distribution as $c \searrow 0$

Remarks II

What if $c \searrow 0$?

- If $c \to 0$, that is $n >> N$, then typical from the usual regime **"small dimensional data vs large samples"**.
- the support of Marčenko-Pastur distribution

$$[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$$

concentrates around $\{1\}$ and

$$\mu_{\mathrm{MP}} \xrightarrow[c \to 0]{} \delta_1 \ .$$

- In accordance with **a priori information # 1**

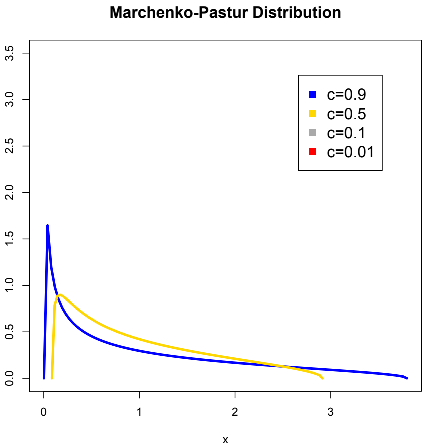


Figure : M̌P distribution as $c \searrow 0$

Remarks II

What if $c \searrow 0$?

- If $c \to 0$, that is $n >> N$, then typical from the usual regime **"small dimensional data vs large samples"**.
- the support of Marčenko-Pastur distribution

$$[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$$

concentrates around $\{1\}$ and

$$\mu_{\mathrm{MP}} \xrightarrow[c \to 0]{} \delta_1 \ .$$

- In accordance with **a priori information # 1**

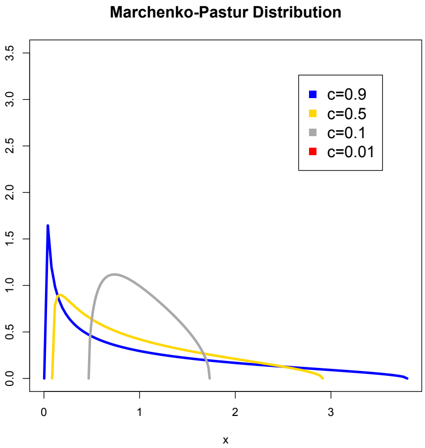


Figure : M̌P distribution as $c \searrow 0$

Remarks II

What if $c \searrow 0$?

- If $c \to 0$, that is $n >> N$, then typical from the usual regime **"small dimensional data vs large samples"**.
- the support of Marčenko-Pastur distribution

$$[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$$

concentrates around $\{1\}$ and

$$\mu_{\mathrm{MP}} \xrightarrow[c \to 0]{} \delta_1 \ .$$

- In accordance with **a priori information # 1**

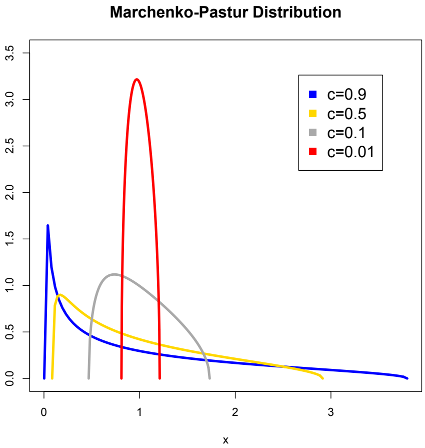


Figure : M̌P distribution as $c \searrow 0$

Extreme eigenvalues of Wishart matrices

Extreme eigenvalues of Wishart matrices

Convergence of extremal eigenvalues

Recall that $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ is the support of M̌P distribution, then:

$$\lambda_{\max}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{\text{almost surely}} (1+\sqrt{c})^2,$$

Convergence of extremal eigenvalues

Recall that $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ is the support of M̌P distribution, then:

$$\lambda_{\max}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{\text{almost surely}} (1+\sqrt{c})^2,$$
$$\lambda_{\min}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{\text{almost surely}} (1-\sqrt{c})^2,$$

under the 4th moment condition: $\mathbb{E}|X_{ij}|^4 < \infty$ (Bai and Yin, 1988).

Extreme eigenvalues of Wishart matrices

Convergence of extremal eigenvalues

Recall that $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ is the support of M̌P distribution, then:

$$\lambda_{\max}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{\text{almost surely}} (1+\sqrt{c})^2,$$
$$\lambda_{\min}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{\text{almost surely}} (1-\sqrt{c})^2,$$

under the 4th moment condition: $\mathbb{E}|X_{ij}|^4 < \infty$ (Bai and Yin, 1988).

Fluctuations of $\lambda_{\max}$: Tracy-Widom distribution

We can fully describe the fluctuations of $\lambda_{\max}$:

Extreme eigenvalues of Wishart matrices

Convergence of extremal eigenvalues

Recall that $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ is the support of M̌P distribution, then:

$$\lambda_{\max}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{\text{almost surely}} (1+\sqrt{c})^2,$$
$$\lambda_{\min}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{\text{almost surely}} (1-\sqrt{c})^2,$$

under the 4th moment condition: $\mathbb{E}|X_{ij}|^4 < \infty$ (Bai and Yin, 1988).

Fluctuations of $\lambda_{\max}$: Tracy-Widom distribution

We can fully describe the fluctuations of $\lambda_{\max}$:

$$\frac{N^{2/3}}{\Theta_N}\left\{\lambda_{\max}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) - (1+\sqrt{c_n})^2\right\} \xrightarrow[N,n\to\infty]{\mathcal{L}} \mu_{\mathrm{TW}}$$

where

$$c_n = \frac{N}{n} \quad \text{and} \quad \Theta_N = (1+\sqrt{c_n})\left(\frac{1}{\sqrt{c_n}}+1\right)^{1/3}$$

(Johnstone 2001).

Outline

Sketched proof of Marčenko-Pastur's theorem

Recall definition of the **Stieltjes transform** s_n:

$$s_n(z) \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} \quad = \quad \frac{1}{N}\mathsf{tr}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* - z\mathbf{I}_N\right)^{-1} \ .$$

Sketched proof of Marčenko-Pastur's theorem

Recall definition of the **Stieltjes transform** s_n:

$$s_n(z) \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} \quad = \quad \frac{1}{N}\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* - z\mathbf{I}_N\right)^{-1} .$$

1. As for the semi-circle law, similar steps lead to

$$s_n(z) \approx \frac{1}{(1-c_n) - z - zc_n s_n(z)}$$

Sketched proof of Marčenko-Pastur's theorem

Recall definition of the **Stieltjes transform** s_n:

$$s_n(z) \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} \quad = \quad \frac{1}{N}\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* - z\mathbf{I}_N\right)^{-1}.$$

1. As for the semi-circle law, similar steps lead to

$$s_n(z) \approx \frac{1}{(1-c_n) - z - zc_n s_n(z)}$$

2. Therefore, s_n does have a limit s, solution to **the fixed point equation**:

$$s(z) = \frac{1}{(1-c) - z - zcs(z)}.$$

$$\left[\text{semi-circle :} \quad s(z) = \frac{1}{-z - s(z)}\right]$$

Sketched proof of Marčenko-Pastur's theorem

Recall definition of the **Stieltjes transform** s_n:

$$s_n(z) \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} \quad = \quad \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* - z\mathbf{I}_N\right)^{-1}.$$

1. As for the semi-circle law, similar steps lead to

$$s_n(z) \approx \frac{1}{(1-c_n) - z - zc_n s_n(z)}$$

2. Therefore, s_n does have a limit s, solution to **the fixed point equation**:

$$s(z) = \frac{1}{(1-c) - z - zcs(z)}. \qquad \left[\text{semi-circle :} \quad s(z) = \frac{1}{-z - s(z)}\right]$$

3. An explicit solution is given by

$$s(z) = \frac{-(z + (c-1)) + \sqrt{(z-b)(z-a)}}{2cz}$$

which is $s_{\mu_{MP}}$!!

Sketched proof of Marčenko-Pastur's theorem

Recall definition of the **Stieltjes transform** s_n:

$$s_n(z) \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} \quad = \quad \frac{1}{N}\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^* - z\mathbf{I}_N\right)^{-1} .$$

1. As for the semi-circle law, similar steps lead to

$$s_n(z) \approx \frac{1}{(1-c_n) - z - zc_n s_n(z)}$$

2. Therefore, s_n does have a limit s, solution to **the fixed point equation**:

$$s(z) = \frac{1}{(1-c) - z - zcs(z)}. \qquad \left[\text{semi-circle : } \quad s(z) = \frac{1}{-z - s(z)}\right]$$

3. An explicit solution is given by

$$s(z) = \frac{-(z+(c-1)) + \sqrt{(z-b)(z-a)}}{2cz}$$

   which is $s_{\mu_{MP}}$!!

4. By the inversion formula, the density is found to be:

$$\mu_{\mathrm{MP}}(dx) \;=\; \left(1-\frac{1}{c}\right)^{+}\delta_0(dx) + \frac{\sqrt{(b-x)(x-a)}}{2\pi xc}1_{[a,b]}(x)\,dx$$

Outline

The largest eigenvalue in M̌P model

Given a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries $\mathbb{E}X_{ij} = 0$ and $\mathbb{E}|X_{ij}|^2 = 1$,

The largest eigenvalue in M̌P model

Given a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries $\mathbb{E}X_{ij} = 0$ and $\mathbb{E}|X_{ij}|^2 = 1$,

$$L_N\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{} \mu_{\mathrm{MP}}$$

where μ_{MP} has support

$$\mathcal{S}_{\check{\mathrm{M}}\mathrm{P}} = \{0\} \cup \underbrace{\left[(1-\sqrt{c})^2\ ,\ (1+\sqrt{c})^2\right]}_{\textbf{bulk}}$$

(remove the set $\{0\}$ if $c < 1$)

Introduction

The largest eigenvalue in M̌P model

Given a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries $\mathbb{E}X_{ij} = 0$ and $\mathbb{E}|X_{ij}|^2 = 1$,

$$L_N\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{} \mu_{\mathrm{MP}}$$

where μ_{MP} has support

$$\mathcal{S}_{\check{\mathrm{M}}\mathrm{P}} = \{0\} \cup \underbrace{\left[(1-\sqrt{c})^2\ ,\ (1+\sqrt{c})^2\right]}_{\textbf{bulk}}$$

(remove the set $\{0\}$ if $c < 1$)

Theorem

Introduction

The largest eigenvalue in M̌P model

Given a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries $\mathbb{E}X_{ij} = 0$ and $\mathbb{E}|X_{ij}|^2 = 1$,

$$L_N\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{} \mu_{\mathrm{MP}}$$

where μ_{MP} has support

$$\mathcal{S}_{\check{\mathrm{M}}\mathrm{P}} \;=\; \{0\} \;\cup\; \underbrace{\left[(1-\sqrt{c})^2\;,\;(1+\sqrt{c})^2\right]}_{\text{bulk}}$$

(remove the set $\{0\}$ if $c < 1$)

Theorem

- Let $\mathbb{E}|X_{ij}|^4 < \infty$, then:

$$\lambda_{\max}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{a.s.} (1+\sqrt{c})^2\ .$$

Introduction

The largest eigenvalue in $\check{\text{M}}$P model

Given a $N \times n$ matrix $\mathbf{X}_N$ with i.i.d. entries $\mathbb{E}X_{ij} = 0$ and $\mathbb{E}|X_{ij}|^2 = 1$,

$$L_N\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{} \mu_{\text{MP}}$$

where μ_{MP} has support

$$\mathcal{S}_{\check{\text{M}}\text{P}} \;=\; \{0\} \,\cup\, \underbrace{\left[(1-\sqrt{c})^2 \,,\, (1+\sqrt{c})^2\right]}_{\textbf{bulk}}$$

(remove the set $\{0\}$ if $c < 1$)

Theorem

- Let $\mathbb{E}|X_{ij}|^4 < \infty$, then:

$$\lambda_{\max}\left(\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*\right) \xrightarrow[N,n\to\infty]{a.s.} (1+\sqrt{c})^2 \;.$$

Message: The largest eigenvalue converges to the right edge of the bulk.

Spiked Models I

Definition

Let $\mathbf{\Pi}_N$ be a **small perturbation of the identity**:

Spiked Models I

Definition

Let $\mathbf{\Pi}_N$ be a **small perturbation of the identity**:

$$\mathbf{\Pi}_N = \mathbf{I}_N + \mathbf{P}_N \quad \text{where} \quad \mathbf{P}_N = \theta_1 \vec{\mathbf{u}}_1 \vec{\mathbf{u}}_1^* + \cdots + \theta_k \vec{\mathbf{u}}_k \vec{\mathbf{u}}_k^*$$

where k is **independent of the dimensions** N, n.

Spiked Models I

Definition

Let $\mathbf{\Pi}_N$ be a **small perturbation of the identity**:

$$\mathbf{\Pi}_N = \mathbf{I}_N + \mathbf{P}_N \quad \text{where} \quad \mathbf{P}_N = \theta_1 \vec{\mathbf{u}}_1 \vec{\mathbf{u}}_1^* + \cdots + \theta_k \vec{\mathbf{u}}_k \vec{\mathbf{u}}_k^*$$

where k is **independent of the dimensions** N, n.

Consider

$$\boxed{\tilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2} \mathbf{X}_N}$$

This model will be refered to as a (multiplicative) **spiked model**.

Spiked Models I

Definition

Let $\mathbf{\Pi}_N$ be a **small perturbation of the identity**:

$$\mathbf{\Pi}_N = \mathbf{I}_N + \mathbf{P}_N \quad \text{where} \quad \mathbf{P}_N = \theta_1 \vec{\mathbf{u}}_1 \vec{\mathbf{u}}_1^* + \cdots + \theta_k \vec{\mathbf{u}}_k \vec{\mathbf{u}}_k^*$$

where k is **independent of the dimensions** N, n.

Consider

$$\boxed{\tilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2} \mathbf{X}_N}$$

This model will be refered to as a (multiplicative) **spiked model**.

Think of $\mathbf{\Pi}_N$ as

$$\mathbf{\Pi}_N = \begin{pmatrix} 1+\theta_1 & & & & \\ & \ddots & & & \\ & & 1+\theta_k & & \\ & & & 1 & \\ & & & & \ddots \end{pmatrix}$$

Spiked Models I

Definition

Let $\mathbf{\Pi}_N$ be a **small perturbation of the identity**:

$$\mathbf{\Pi}_N = \mathbf{I}_N + \mathbf{P}_N \quad \text{where} \quad \mathbf{P}_N = \theta_1 \vec{\mathbf{u}}_1 \vec{\mathbf{u}}_1^* + \cdots + \theta_k \vec{\mathbf{u}}_k \vec{\mathbf{u}}_k^*$$

where k is **independent of the dimensions** N, n.

Consider

$$\boxed{\tilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2} \mathbf{X}_N}$$

This model will be refered to as a (multiplicative) **spiked model**.

Think of $\mathbf{\Pi}_N$ as

$$\mathbf{\Pi}_N = \begin{pmatrix} 1+\theta_1 & & & & \\ & \ddots & & & \\ & & 1+\theta_k & & \\ & & & 1 & \\ & & & & \ddots \end{pmatrix}$$

Very important: The rank k of perturbations is finite

Spiked Models II

Remarks

- The spiked model is a particular case of **large covariance matrix** model with

$$\mathbf{R}_N = \mathbf{I}_N + \sum_{\ell=1}^{k} \theta_\ell \vec{\mathbf{u}}_\ell \vec{\mathbf{u}}_\ell^*$$

Spiked Models II

Remarks

- The spiked model is a particular case of **large covariance matrix** model with

$$\mathbf{R}_N = \mathbf{I}_N + \sum_{\ell=1}^{k} \theta_\ell \vec{\mathbf{u}}_\ell \vec{\mathbf{u}}_\ell^*$$

- There are also additive spiked models: $\check{\mathbf{X}}_N = \mathbf{X}_N + \mathbf{A}_N$ where $\mathbf{A}_N$ is a matrix with finite rank.

Spiked Models II

Remarks

- The spiked model is a particular case of **large covariance matrix** model with

$$\mathbf{R}_N = \mathbf{I}_N + \sum_{\ell=1}^{k} \theta_\ell \vec{\mathbf{u}}_\ell \vec{\mathbf{u}}_\ell^*$$

- There are also additive spiked models: $\check{\mathbf{X}}_N = \mathbf{X}_N + \mathbf{A}_N$ where $\mathbf{A}_N$ is a matrix with finite rank.
- Spiked models have been introduced by Iain M. Johnstone in his 2001 paper in *Annals of Statistics* to take into account the fact that in many datasets, a small number of eigenvalues is "far away" the bulk of the other eigenvalues

Spiked Models II

Remarks

- The spiked model is a particular case of **large covariance matrix** model with

$$\mathbf{R}_N = \mathbf{I}_N + \sum_{\ell=1}^{k} \theta_\ell \vec{\mathbf{u}}_\ell \vec{\mathbf{u}}_\ell^*$$

- There are also additive spiked models: $\check{\mathbf{X}}_N = \mathbf{X}_N + \mathbf{A}_N$ where $\mathbf{A}_N$ is a matrix with finite rank.
- Spiked models have been introduced by Iain M. Johnstone in his 2001 paper in *Annals of Statistics* to take into account the fact that in many datasets, a small number of eigenvalues is "far away" the bulk of the other eigenvalues

Objective

- What is the influence of $\mathbf{\Pi}_N$ over $L_N\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right)$?
- What is the influence of $\mathbf{\Pi}_N$ over $\lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right)$?

Simulations I: Single spikes

Simulations I: Single spikes

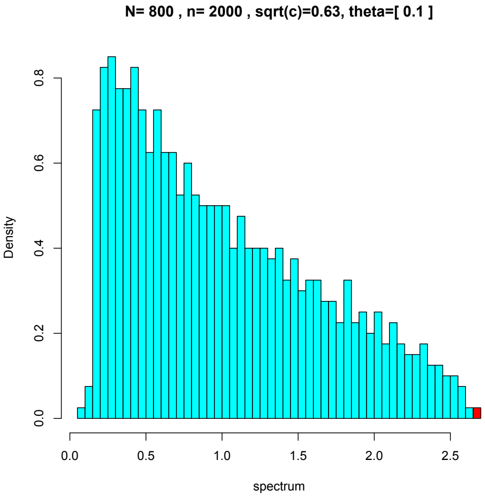


Figure : Spiked model - strength of the perturbation $\theta = 0.1$

Simulations I: Single spikes

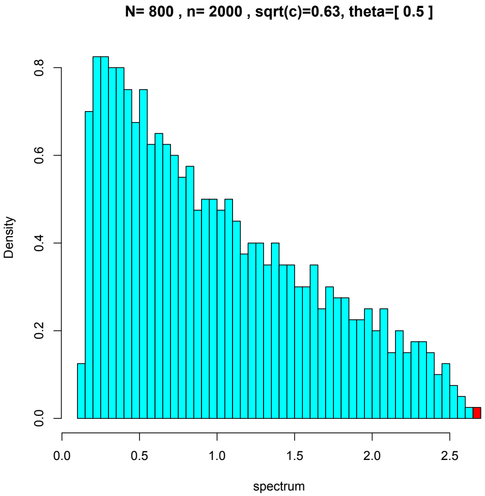


Figure : Spiked model - strength of the perturbation $\theta = 0.5$

Simulations I: Single spikes

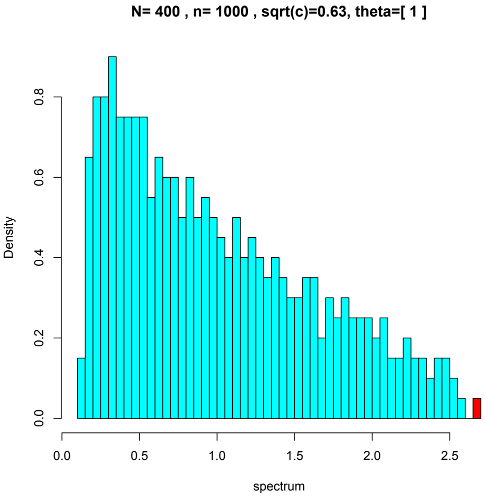


Figure : Spiked model - strength of the perturbation $\theta = 1$

Simulations I: Single spikes

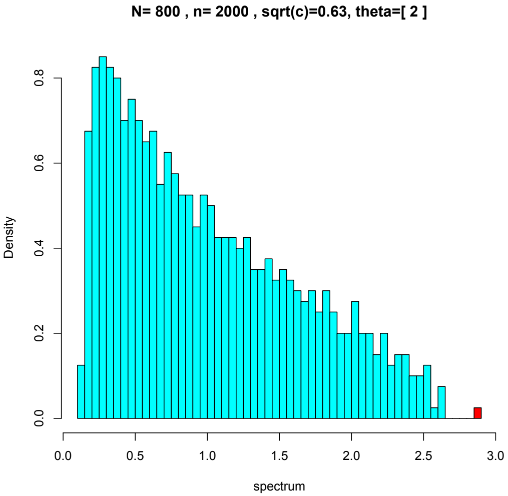


Figure : Spiked model - strength of the perturbation $\theta = 2$

Simulations I: Single spikes

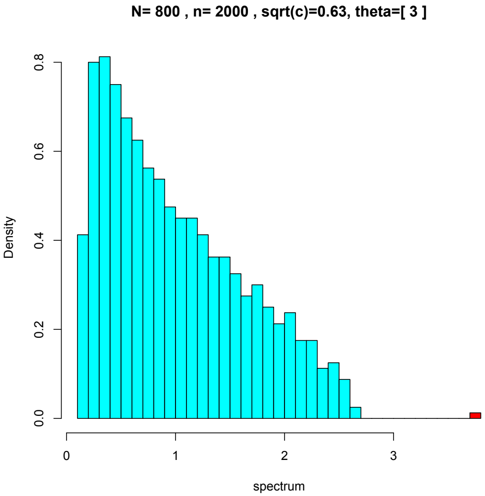


Figure : Spiked model - strength of the perturbation $\theta = 3$

Observation #1

If the **strength** θ of the perturbation $\mathbf{P}_N$ is large enough, then the limit of $\lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right)$ is **strictly larger** than the right edge of the bulk.

Simulations II: Spectral measure

Simulations II: Spectral measure

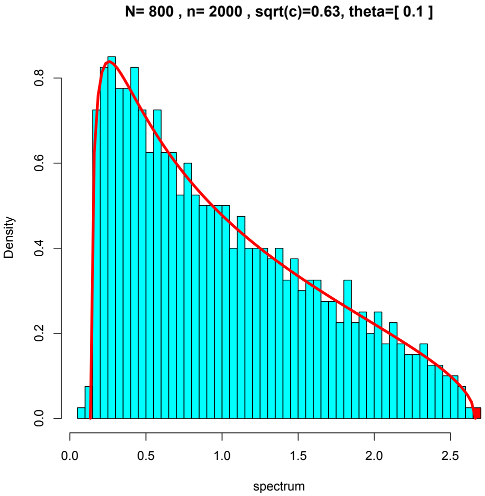


Figure : Spiked model - strength of the perturbation $\theta = 0.1$

Simulations II: Spectral measure

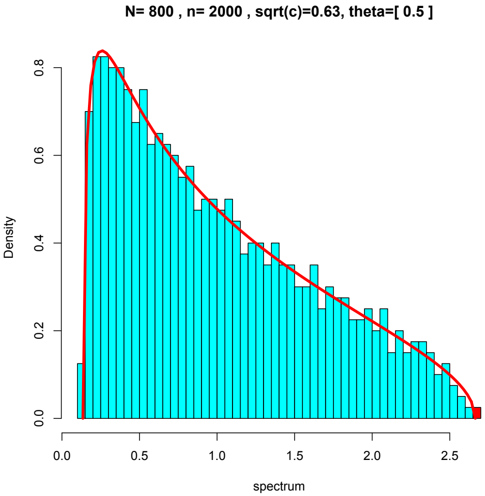


Figure : Spiked model - strength of the perturbation $\theta = 0.5$

Simulations II: Spectral measure

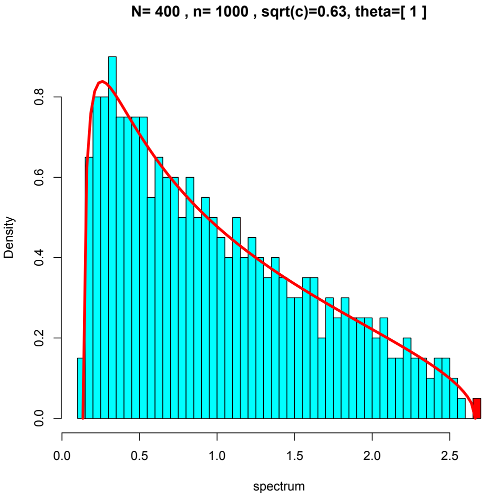


Figure : Spiked model - strength of the perturbation $\theta = 1$

Simulations II: Spectral measure

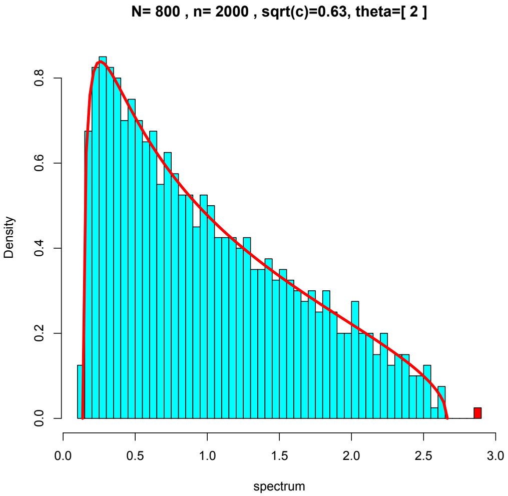


Figure : Spiked model - strength of the perturbation $\theta = 2$

Simulations II: Spectral measure

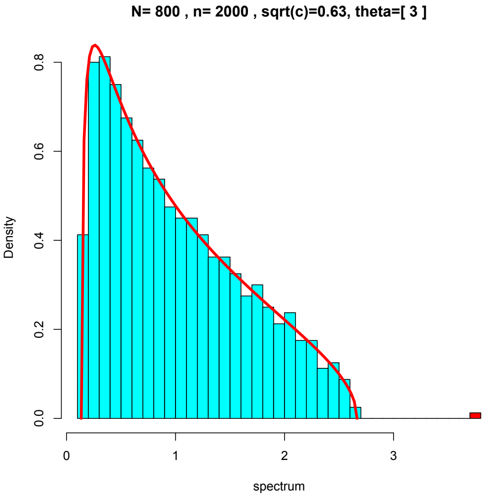


Figure : Spiked model - strength of the perturbation $\theta = 3$

Simulations III: Multiple Spikes

Simulations III: Multiple Spikes

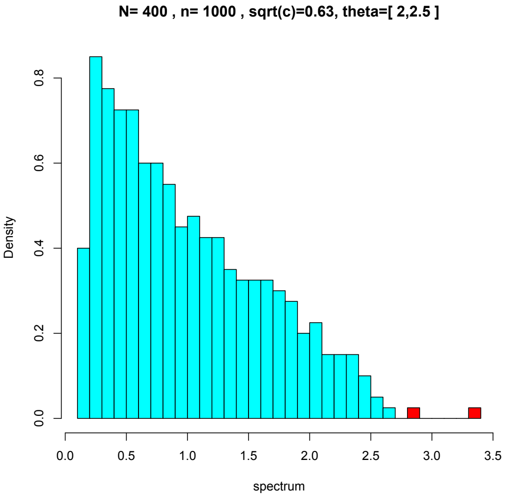


Figure : Spiked model - Two spikes

Simulations III: Multiple Spikes

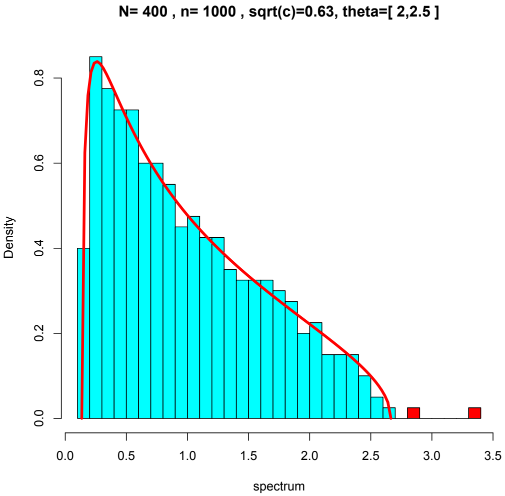


Figure : Spiked model - Two spikes

Simulations III: Multiple Spikes

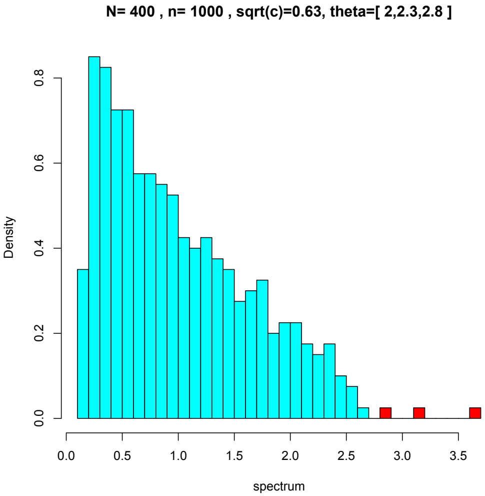


Figure : Spiked model - Three spikes

Simulations III: Multiple Spikes

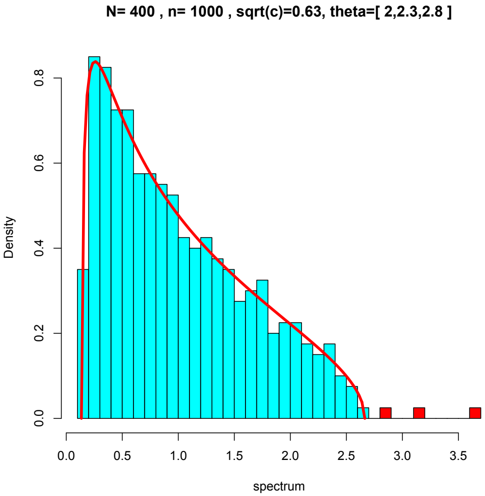


Figure : Spiked model - Three spikes

Simulations III: Multiple Spikes

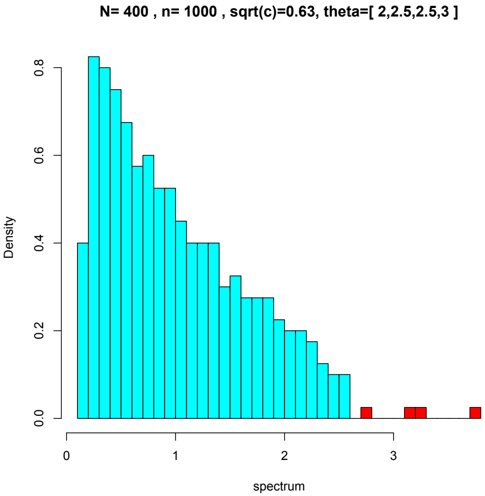


Figure : Spiked model - Multiple spikes

Simulations III: Multiple Spikes

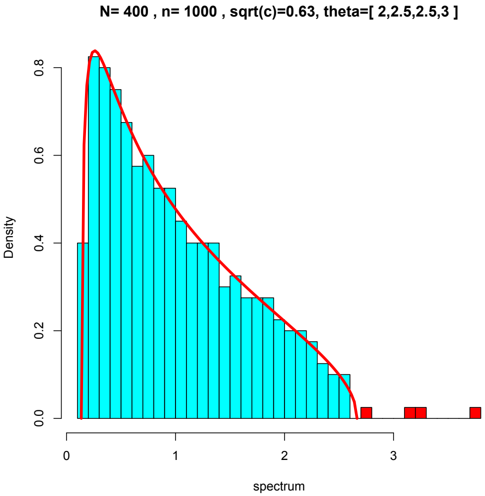


Figure : Spiked model - Multiple spikes

Observation # 2

Whathever the perturbations, the spectral measure converges toward Marčenko-Pastur distribution

Outline

The limiting spectral measure

Theorem

The following convergence holds true: $\boxed{L_N\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{a.s.} \mu_{\mathrm{MP}}\ .}$

The limiting spectral measure

Theorem

The following convergence holds true: $$L_N\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{a.s.} \mu_{\mathrm{MP}}\ .$$

The theorem is a simple consequence of the Cauchy (Weyl) interlacing theorem which states that the eigenvalues of a finite-rank perturbated Hermitian matrix (or a finite rank reduced submatrix) are interlaced with those of the original Hermitian matrix.

Remark

The limiting spectral measure is not sensitive to the presence of spikes

Outline

Behaviour of the largest eigenvalue

We consider the following spiked model:

$$\tilde{\mathbf{X}}_N = (\mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*)^{1/2}\,\mathbf{X}_N \quad \text{with} \quad \|\vec{\mathbf{u}}\| = 1 \ .$$

Behaviour of the largest eigenvalue

We consider the following spiked model:

$$\tilde{\mathbf{X}}_N = (\mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*)^{1/2}\,\mathbf{X}_N \quad \text{with} \quad \|\vec{\mathbf{u}}\| = 1 \ .$$

which corresponds to a **rank-one** perturbation.

Behaviour of the largest eigenvalue

We consider the following spiked model:

$$\tilde{\mathbf{X}}_N = (\mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*)^{1/2}\, \mathbf{X}_N \quad \text{with} \quad \|\vec{\mathbf{u}}\| = 1 \ .$$

which corresponds to a **rank-one** perturbation.

Theorem

Recall that $c = \lim_{N,n\to\infty} \frac{N}{n}$.

Behaviour of the largest eigenvalue

We consider the following spiked model:

$$\tilde{\mathbf{X}}_N = \left(\mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*\right)^{1/2} \mathbf{X}_N \quad \text{with} \quad \|\vec{\mathbf{u}}\| = 1 \ .$$

which corresponds to a **rank-one** perturbation.

Theorem

Recall that $c = \lim_{N,n\to\infty} \frac{N}{n}$.

- if $\boxed{\theta \leq \sqrt{c}}$ then

$$\lambda_{\max} = \lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N \tilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{a.s.} (1+\sqrt{c})^2$$

Behaviour of the largest eigenvalue

We consider the following spiked model:

$$\tilde{\mathbf{X}}_N = (\mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*)^{1/2}\, \mathbf{X}_N \quad \text{with} \quad \|\vec{\mathbf{u}}\| = 1 \; .$$

which corresponds to a **rank-one** perturbation.

Theorem

Recall that $c = \lim_{N,n\to\infty} \frac{N}{n}$.

- if $\boxed{\theta \leq \sqrt{c}}$ then

$$\lambda_{\max} = \lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N \tilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{a.s.} (1+\sqrt{c})^2$$

- if $\boxed{\theta > \sqrt{c}}$ then

$$\lambda_{\max} \xrightarrow[N,n\to\infty]{a.s.} (1+\theta)\left(1+\frac{c}{\theta}\right)$$

Behaviour of the largest eigenvalue

We consider the following spiked model:

$$\tilde{\mathbf{X}}_N = (\mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*)^{1/2}\, \mathbf{X}_N \quad \text{with} \quad \|\vec{\mathbf{u}}\| = 1\ .$$

which corresponds to a **rank-one** perturbation.

Theorem

Recall that $c = \lim_{N,n\to\infty} \frac{N}{n}$.

- if $\boxed{\theta \le \sqrt{c}}$ then

$$\lambda_{\max} = \lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N \tilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{a.s.} (1+\sqrt{c})^2$$

- if $\boxed{\theta > \sqrt{c}}$ then

$$\lambda_{\max} \xrightarrow[N,n\to\infty]{a.s.} (1+\theta)\left(1+\frac{c}{\theta}\right) > (1+\sqrt{c})^2$$

[Baik-Ben Arous-Péché (2005); Baik and Silverstein (2006)]

Phase transition Phenomenon

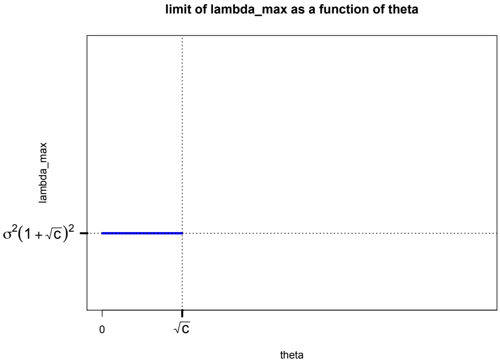


Figure : Limit of largest eigenvalue $\lambda_{\max}$ as a function of the perturbation θ

Phase transition Phenomenon

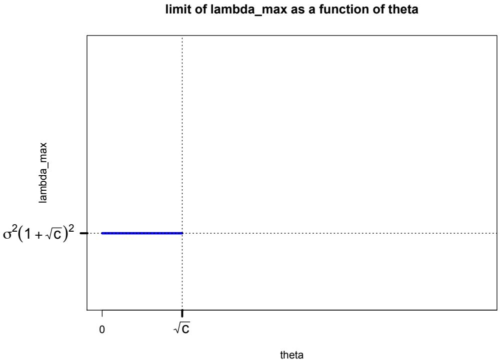


Figure : Limit of largest eigenvalue $\lambda_{\max}$ as a function of the perturbation θ

- If $\theta \leq \sqrt{c}$ then

$$\lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{} (1+\sqrt{c})^2 \ .$$

Phase transition Phenomenon

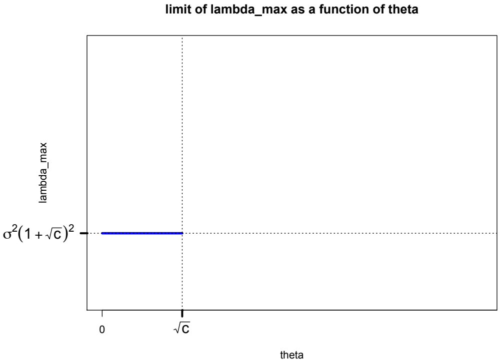


Figure : Limit of largest eigenvalue $\lambda_{\max}$ as a function of the perturbation θ

- If $\theta \leq \sqrt{c}$ then

$$\lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{} (1+\sqrt{c})^2 \ .$$

Below the threshold $\sqrt{c}$, $\lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right)$ asymptotically **sticks to the bulk**.

Phase transition Phenomenon

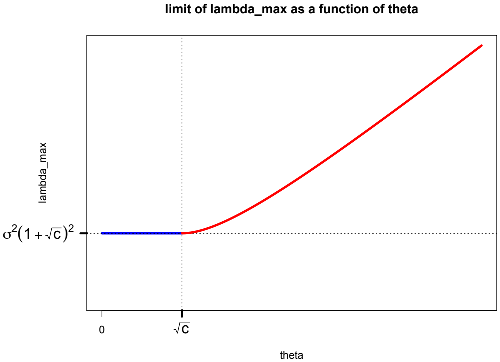


Figure : Limit of largest eigenvalue $\lambda_{\max}$ as a function of the perturbation θ

Phase transition Phenomenon

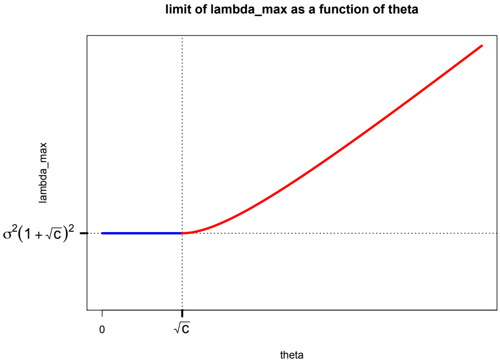


Figure : Limit of largest eigenvalue $\lambda_{\max}$ as a function of the perturbation θ

- if $\theta > \sqrt{c}$ then

$$\lim_{N,n} \lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right) \quad = \quad (1+\theta)\left(1+\frac{c}{\theta}\right)$$

Phase transition Phenomenon

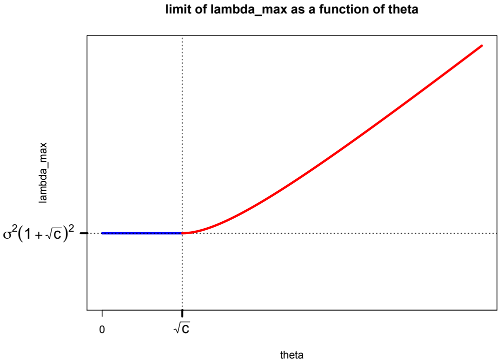


Figure : Limit of largest eigenvalue $\lambda_{\max}$ as a function of the perturbation θ

- if $\theta > \sqrt{c}$ then

$$\lim_{N,n} \lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right) \quad = \quad (1+\theta)\left(1+\frac{c}{\theta}\right) > \left(1+\sqrt{c}\right)^2$$

Above the threshold $\sqrt{c}$, $\lambda_{\max}\left(\frac{1}{n}\tilde{\mathbf{X}}_N\tilde{\mathbf{X}}_N^*\right)$ asymptotically **separates from the bulk**.

Outline

Summary I

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]
- $\mathbf{X}_N$ a $N \times n$ matrix with i.i.d. entries

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]
- $\mathbf{X}_N$ a $N \times n$ matrix with i.i.d. entries

then $\boxed{\widetilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2}\mathbf{X}_N}$ is a (multiplicative) spiked model

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]
- $\mathbf{X}_N$ a $N \times n$ matrix with i.i.d. entries

then $\boxed{\widetilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2} \mathbf{X}_N}$ is a (multiplicative) spiked model

Global regime

The spectral measure $L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N \widetilde{\mathbf{X}}_N^*\right)$ converges to Marčenko-Pastur distribution:

$$\text{almost surely,} \quad L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N \widetilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{\mathcal{L}} \mu_{\mathrm{MP}}$$

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]
- $\mathbf{X}_N$ a $N \times n$ matrix with i.i.d. entries

then $\boxed{\widetilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2}\mathbf{X}_N}$ is a (multiplicative) spiked model

Global regime

The spectral measure $L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to Marčenko-Pastur distribution:

$$\text{almost surely,}\ \ L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{\mathcal{L}} \mu_{\mathrm{MP}}$$

Largest eigenvalue

- if $\boxed{\theta \leq \sqrt{c}}$, then $\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to the right edge of the bulk

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]
- $\mathbf{X}_N$ a $N \times n$ matrix with i.i.d. entries

then $\boxed{\widetilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2}\mathbf{X}_N}$ is a (multiplicative) spiked model

Global regime

The spectral measure $L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to Marčenko-Pastur distribution:

$$\text{almost surely, } L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{\mathcal{L}} \mu_{\text{MP}}$$

Largest eigenvalue

- if $\boxed{\theta \leq \sqrt{c}}$, then $\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to the right edge of the bulk
- if $\boxed{\theta > \sqrt{c}}$, then $\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ separates from the bulk

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]
- $\mathbf{X}_N$ a $N \times n$ matrix with i.i.d. entries

then $\boxed{\widetilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2} \mathbf{X}_N}$ is a (multiplicative) spiked model

Global regime

The spectral measure $L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to Marčenko-Pastur distribution:

$$\text{almost surely,}\ \ L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{\mathcal{L}} \mu_{\mathrm{MP}}$$

Largest eigenvalue

- if $\boxed{\theta \leq \sqrt{c}}$, then $\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to the right edge of the bulk
- if $\boxed{\theta > \sqrt{c}}$, then $\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ separates from the bulk

$$\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right) \to (1+\theta)\left(1+\frac{c}{\theta}\right)$$

Summary I

Spiked model

Let

- $\mathbf{\Pi}_N$ a small perturbation of the identity [Example: $\mathbf{\Pi}_N = \mathbf{I}_N + \theta \vec{\mathbf{u}}\vec{\mathbf{u}}^*$]
- $\mathbf{X}_N$ a $N \times n$ matrix with i.i.d. entries

then $\boxed{\widetilde{\mathbf{X}}_N = \mathbf{\Pi}_N^{1/2} \mathbf{X}_N}$ is a (multiplicative) spiked model

Global regime

The spectral measure $L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to Marčenko-Pastur distribution:

$$\text{almost surely, } L_N\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right) \xrightarrow[N,n\to\infty]{\mathcal{L}} \mu_{\text{MP}}$$

Largest eigenvalue

- if $\boxed{\theta \leq \sqrt{c}}$, then $\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ converges to the right edge of the bulk
- if $\boxed{\theta > \sqrt{c}}$, then $\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right)$ separates from the bulk

$$\lambda_{\max}\left(\frac{1}{N}\widetilde{\mathbf{X}}_N\widetilde{\mathbf{X}}_N^*\right) \to (1+\theta)\left(1+\frac{c}{\theta}\right) > (1+\sqrt{c})^2$$

Outline

The hypothesis testing problem

Statistical Setup

let

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

The $\vec{\mathbf{y}}(k)$'s are n observations all either drawn under H_0 or H_1. Here,

The hypothesis testing problem

Statistical Setup

let

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

The $\vec{\mathbf{y}}(k)$'s are n observations all either drawn under H_0 or H_1. Here,

- $\vec{\mathbf{w}}(k)$ is a $N \times 1$ complex gaussian white noise process:

$$\vec{\mathbf{w}}(k) \;\sim\; \mathcal{C}N(0, \mathbf{I}_N)$$

The hypothesis testing problem

Statistical Setup

let

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma\vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma\vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \quad \text{for } k = 1:n$$

The $\vec{\mathbf{y}}(k)$'s are n observations all either drawn under H_0 or H_1. Here,

- $\vec{\mathbf{w}}(k)$ is a $N \times 1$ complex gaussian white noise process:

$$\vec{\mathbf{w}}(k) \sim \mathcal{C}N(0, \mathbf{I}_N)$$

- $\vec{\mathbf{h}}$ is a $N \times 1$ **deterministic** and **unknown** vector representing the characteristics of the propagation channel

The hypothesis testing problem

Statistical Setup

let

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

The $\vec{\mathbf{y}}(k)$'s are n observations all either drawn under H_0 or H_1. Here,

- $\vec{\mathbf{w}}(k)$ is a $N \times 1$ complex gaussian white noise process:

$$\vec{\mathbf{w}}(k) \;\sim\; \mathcal{C}N(0, \mathbf{I}_N)$$

- $\vec{\mathbf{h}}$ is a $N \times 1$ **deterministic** and **unknown** vector representing the characteristics of the propagation channel
- $s(k)$ represent the signal; it is a scalar complex Gaussian i.i.d. process

The hypothesis testing problem

Statistical Setup

let

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1:n$$

The $\vec{\mathbf{y}}(k)$'s are n observations all either drawn under H_0 or H_1. Here,

- $\vec{\mathbf{w}}(k)$ is a $N \times 1$ complex gaussian white noise process:

$$\vec{\mathbf{w}}(k) \;\sim\; \mathcal{C}N(0, \mathbf{I}_N)$$

- $\vec{\mathbf{h}}$ is a $N \times 1$ **deterministic** and **unknown** vector representing the characteristics of the propagation channel
- $s(k)$ represent the signal; it is a scalar complex Gaussian i.i.d. process

Objective

Given n observations $(\vec{\mathbf{y}}(k), 1 \le k \le n)$, and the associated **sample covariance matrix**

$$\hat{\mathbf{R}}_n = \frac{1}{n}\mathbf{Y}_n \mathbf{Y}_n^* \quad \text{where} \quad \mathbf{Y}_n = [\vec{\mathbf{y}}(1), \cdots, \vec{\mathbf{y}}(n)] \quad \text{is } N \times n \;,$$

The hypothesis testing problem

Statistical Setup

let

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

The $\vec{\mathbf{y}}(k)$'s are n observations all either drawn under H_0 or H_1. Here,

- $\vec{\mathbf{w}}(k)$ is a $N \times 1$ complex gaussian white noise process:

$$\vec{\mathbf{w}}(k) \;\sim\; \mathcal{C}N(0, \mathbf{I}_N)$$

- $\vec{\mathbf{h}}$ is a $N \times 1$ **deterministic** and **unknown** vector representing the characteristics of the propagation channel
- $s(k)$ represent the signal; it is a scalar complex Gaussian i.i.d. process

Objective

Given n observations $(\vec{\mathbf{y}}(k), 1 \le k \le n)$, and the associated **sample covariance matrix**

$$\hat{\mathbf{R}}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^* \quad \text{where} \quad \mathbf{Y}_n = [\vec{\mathbf{y}}(1), \cdots, \vec{\mathbf{y}}(n)] \quad \text{is } N \times n\ ,$$

the aim is to decide H_0 (no signal) or H_1 (single-source detection) in the case where

$$\frac{N}{n} \to c \in (0,1) \quad \text{i.e.}$$

The hypothesis testing problem

Statistical Setup

let

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma\vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma\vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

The $\vec{\mathbf{y}}(k)$'s are n observations all either drawn under H_0 or H_1. Here,

- $\vec{\mathbf{w}}(k)$ is a $N \times 1$ complex gaussian white noise process:

$$\vec{\mathbf{w}}(k) \sim \mathcal{C}N(0, \mathbf{I}_N)$$

- $\vec{\mathbf{h}}$ is a $N \times 1$ **deterministic** and **unknown** vector representing the characteristics of the propagation channel
- $s(k)$ represent the signal; it is a scalar complex Gaussian i.i.d. process

Objective

Given n observations $(\vec{\mathbf{y}}(k), 1 \le k \le n)$, and the associated **sample covariance matrix**

$$\hat{\mathbf{R}}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^* \quad \text{where} \quad \mathbf{Y}_n = [\vec{\mathbf{y}}(1), \cdots, \vec{\mathbf{y}}(n)] \quad \text{is } N \times n \;,$$

the aim is to decide H_0 (no signal) or H_1 (single-source detection) in the case where

$$\frac{N}{n} \to c \in (0,1) \quad \text{i.e.} \quad \boxed{\text{Dimension } N \text{ of observations} \propto \text{size } n \text{ of sample}}$$

Neyman-Pearson procedure

Likelihood functions

Neyman-Pearson procedure

Likelihood functions

Notice that $\mathbf{Y}_n$ is a $N \times n$ matrix whose columns are i.i.d. vectors with covariance matrix defined by

$$\mathbf{\Sigma}_N = \begin{cases} \mathbf{I}_N & \text{under } H_0 \\ \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N & \text{under } H_1 \end{cases},$$

Neyman-Pearson procedure

Likelihood functions

Notice that $\mathbf{Y}_n$ is a $N \times n$ matrix whose columns are i.i.d. vectors with covariance matrix defined by

$$\boldsymbol{\Sigma}_N = \begin{cases} \mathbf{I}_N & \text{under } H_0 \\ \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N & \text{under } H_1 \end{cases},$$

hence the likelihood functions write

$$\begin{aligned} p_0(\mathbf{Y}_N; \sigma^2) &= \frac{1}{(\pi\sigma^2)^{Nn}} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N\right) \\ p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2) &= \frac{1}{\left[\pi^N \det\left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)\right]^n} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)^{-1}\right) \end{aligned}$$

Neyman-Pearson procedure

Likelihood functions

Notice that $\mathbf{Y}_n$ is a $N \times n$ matrix whose columns are i.i.d. vectors with covariance matrix defined by

$$\mathbf{\Sigma}_N = \begin{cases} \mathbf{I}_N & \text{under } H_0 \ , \\ \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N & \text{under } H_1 \end{cases}$$

hence the likelihood functions write

$$\begin{aligned} p_0(\mathbf{Y}_N; \sigma^2) &= \frac{1}{(\pi\sigma^2)^{Nn}} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N\right) \\ p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2) &= \frac{1}{\left[\pi^N \det\left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)\right]^n} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)^{-1}\right) \end{aligned}$$

Neyman-Pearson

In case where σ^2 and $\vec{\mathbf{h}}$ are **known**, the **Likelihood Ratio Statistics**

$$\frac{p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2)}{p_0(\mathbf{Y}_N; \sigma^2)}$$

provides a **uniformly most powerful test**:

Neyman-Pearson procedure

Likelihood functions

Notice that $\mathbf{Y}_n$ is a $N \times n$ matrix whose columns are i.i.d. vectors with covariance matrix defined by

$$\boldsymbol{\Sigma}_N = \begin{cases} \mathbf{I}_N & \text{under } H_0 \\ \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N & \text{under } H_1 \end{cases},$$

hence the likelihood functions write

$$p_0(\mathbf{Y}_N; \sigma^2) = \frac{1}{(\pi\sigma^2)^{Nn}} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N\right)$$

$$p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2) = \frac{1}{\left[\pi^N \det\left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)\right]^n} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)^{-1}\right)$$

Neyman-Pearson

In case where σ^2 and $\vec{\mathbf{h}}$ are **known**, the **Likelihood Ratio Statistics**

$$\frac{p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2)}{p_0(\mathbf{Y}_N; \sigma^2)}$$

provides a **uniformly most powerful test**:

- Fix a given level $\alpha \in (0, 1)$

Neyman-Pearson procedure

Likelihood functions

Notice that $\mathbf{Y}_n$ is a $N \times n$ matrix whose columns are i.i.d. vectors with covariance matrix defined by

$$\mathbf{\Sigma}_N = \begin{cases} \mathbf{I}_N & \text{under } H_0 \ , \\ \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N & \text{under } H_1 \end{cases}$$

hence the likelihood functions write

$$p_0(\mathbf{Y}_N; \sigma^2) = \frac{1}{(\pi\sigma^2)^{Nn}} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N\right)$$

$$p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2) = \frac{1}{\left[\pi^N \det\left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)\right]^n} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)^{-1}\right)$$

Neyman-Pearson

In case where σ^2 and $\vec{\mathbf{h}}$ are **known**, the **Likelihood Ratio Statistics**

$$\frac{p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2)}{p_0(\mathbf{Y}_N; \sigma^2)}$$

provides a **uniformly most powerful test**:

- Fix a given level $\alpha \in (0, 1)$
- The condition over the **Probability of False Alarm** $\mathbb{P}(H_1 \mid H_0) \leq \alpha$ sets the threshold

Neyman-Pearson procedure

Likelihood functions

Notice that $\mathbf{Y}_n$ is a $N \times n$ matrix whose columns are i.i.d. vectors with covariance matrix defined by

$$\mathbf{\Sigma}_N = \begin{cases} \mathbf{I}_N & \text{under } H_0 \ , \\ \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N & \text{under } H_1 \end{cases}$$

hence the likelihood functions write

$$\begin{aligned} p_0(\mathbf{Y}_N; \sigma^2) &= \frac{1}{(\pi\sigma^2)^{Nn}} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N\right) \\ p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2) &= \frac{1}{\left[\pi^N \det\left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)\right]^n} \exp\left(-\frac{n}{\sigma^2} \operatorname{tr} \hat{\mathbf{R}}_N \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2 \mathbf{I}_N\right)^{-1}\right) \end{aligned}$$

Neyman-Pearson

In case where σ^2 and $\vec{\mathbf{h}}$ are **known**, the **Likelihood Ratio Statistics**

$$\frac{p_1(\mathbf{Y}_N; \vec{\mathbf{h}}; \sigma^2)}{p_0(\mathbf{Y}_N; \sigma^2)}$$

provides a **uniformly most powerful test**:

- Fix a given level $\alpha \in (0, 1)$
- The condition over the **Probability of False Alarm** $\mathbb{P}(H_1 \mid H_0) \leq \alpha$ sets the threshold
- the **maximum achievable power**

$$1 - \mathbb{P}(H_0 \mid H_1)$$

is guaranteed by Neyman-Pearson.

The GLRT

The Generalized Likelihood Ratio Test

In the case where $\vec{\mathbf{h}}$ and σ^2 are unknown, we use instead:

$$L_n = \frac{\sup_{\sigma^2, \vec{\mathbf{h}}} \; p_1(\mathbf{Y}_n, \sigma^2, \vec{\mathbf{h}})}{\sup_{\sigma^2} \; p_0(\mathbf{Y}_n, \sigma^2)}$$

which is no longer uniformily most powerful.

The GLRT

The Generalized Likelihood Ratio Test

In the case where $\vec{\mathbf{h}}$ and σ^2 are unknown, we use instead:

$$L_n = \frac{\sup_{\sigma^2, \vec{\mathbf{h}}} \; p_1(\mathbf{Y}_n, \sigma^2, \vec{\mathbf{h}})}{\sup_{\sigma^2} \; p_0(\mathbf{Y}_n, \sigma^2)}$$

which is no longer uniformily most powerful.

Expression of the GLRT

The GLRT statistics writes

$$L_n = \frac{\left(1 - \frac{1}{N}\right)^{(1-N)n}}{\left(\frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}\right)^n \left(1 - \frac{1}{N}\frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}\right)^{(N-1)n}}$$

and is a **deterministic function** of $\boxed{T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}}$

Outline

Limit of the test statistics T_n - I

Under H_0

Recall $T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$.

Limit of the test statistics T_n - I

Under H_0

Recall $T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$. We have:

$$\lambda_{\max}(\hat{\mathbf{R}}_n) \quad \xrightarrow[N,n\to\infty]{a.s.} \quad \sigma^2(1+\sqrt{c})^2$$

Limit of the test statistics T_n - I

Under H_0

Recall $T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$. We have:

$$\begin{aligned} \lambda_{\max}(\hat{\mathbf{R}}_n) &\xrightarrow[N,n\to\infty]{a.s.} \sigma^2(1+\sqrt{c})^2 \\ \frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n = \frac{1}{Nn}\sum_{i,j}|Y_{ij}|^2 &\xrightarrow[N,n\to\infty]{a.s.} \sigma^2 \end{aligned}$$

Limit of the test statistics T_n - I

Under H_0

Recall $T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$. We have:

$$\lambda_{\max}(\hat{\mathbf{R}}_n) \xrightarrow[N,n\to\infty]{a.s.} \sigma^2(1+\sqrt{c})^2$$

$$\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n = \frac{1}{Nn}\sum_{i,j}|Y_{ij}|^2 \xrightarrow[N,n\to\infty]{a.s.} \sigma^2$$

hence

$$\boxed{T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n} \xrightarrow[N,n\to\infty]{a.s.} (1+\sqrt{c})^2}$$

Limit of the test statistics T_n - II

Under H_1

Let

$$\boxed{\mathbf{snr} = \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2}}$$

the **Signal-to-Noise (SNR)** ratio.

Limit of the test statistics T_n - II

Under H_1

Let

$$\boxed{\mathbf{snr} = \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2}}$$

the **Signal-to-Noise (SNR)** ratio.

- if $\mathbf{snr} > \sqrt{c}$ then

$$T_n \xrightarrow[N,n\to\infty]{a.s.} (1 + \mathbf{snr})\left(1 + \frac{c}{\mathbf{snr}}\right)$$

Limit of the test statistics T_n - II

Under H_1

Let

$$\boxed{\mathbf{snr} = \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2}}$$

the **Signal-to-Noise (SNR)** ratio.

- if $\mathbf{snr} > \sqrt{c}$ then

$$T_n \xrightarrow[N,n\to\infty]{a.s.} (1+\mathbf{snr})\left(1+\frac{c}{\mathbf{snr}}\right) > (1+\sqrt{c})^2$$

Limit of the test statistics T_n - II

Under H_1

Let

$$\boxed{\mathbf{snr} = \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2}}$$

the **Signal-to-Noise (SNR)** ratio.

- if $\mathbf{snr} > \sqrt{c}$ then

$$T_n \xrightarrow[N,n\to\infty]{a.s.} (1+\mathbf{snr})\left(1+\frac{c}{\mathbf{snr}}\right) > (1+\sqrt{c})^2$$

- if $\mathbf{snr} \leq \sqrt{c}$ then

$$T_n \xrightarrow[N,n\to\infty]{a.s.} (1+\sqrt{c})^2$$

(Phase transition)

Limit of the test statistics T_n - III

Remarks

- Condition $\boxed{\mathbf{snr} > \sqrt{c}}$ is **automatically fulfilled** in the classical regime where

$$N \text{ fixed} \quad \text{and} \quad n \to \infty \quad \text{as} \quad c = \lim_{n \to \infty} \frac{N}{n} = 0 \ .$$

Limit of the test statistics T_n - III

Remarks

- Condition $\boxed{\mathbf{snr} > \sqrt{c}}$ is **automatically fulfilled** in the classical regime where

$$N \text{ fixed} \quad \text{and} \quad n \to \infty \quad \text{as} \quad c = \lim_{n \to \infty} \frac{N}{n} = 0 \ .$$

- In the case $N, n \to \infty$, recall that the support of Marčenko-Pastur distribution is

$$[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2] \ ,$$

Limit of the test statistics T_n - III

Remarks

- Condition $\boxed{\mathbf{snr} > \sqrt{c}}$ is **automatically fulfilled** in the classical regime where

$$N \text{ fixed} \quad \text{and} \quad n \to \infty \quad \text{as} \quad c = \lim_{n\to\infty} \frac{N}{n} = 0 \ .$$

- In the case $N, n \to \infty$, recall that the support of Marčenko-Pastur distribution is

$$[(1-\sqrt{c})^2, (1+\sqrt{c})^2] \ ,$$

i.e.

The higher $\sqrt{c}$, the larger the support

Limit of the test statistics T_n - III

Remarks

- Condition $\boxed{\mathbf{snr} > \sqrt{c}}$ is **automatically fulfilled** in the classical regime where

$$N \text{ fixed} \quad \text{and} \quad n \to \infty \quad \text{as} \quad c = \lim_{n\to\infty} \frac{N}{n} = 0 \; .$$

- In the case $N, n \to \infty$, recall that the support of Marčenko-Pastur distribution is

$$[(1-\sqrt{c})^2, (1+\sqrt{c})^2] \; ,$$

  i.e.

  $\boxed{\text{The higher } \sqrt{c}\text{, the larger the support}}$

  One can interpret $\sqrt{c}$ as a level of the **asymptotic noise induced by the data dimension** (=asymptotic data noise).

Limit of the test statistics T_n - III

Remarks

- Condition $\boxed{\mathbf{snr} > \sqrt{c}}$ is **automatically fulfilled** in the classical regime where

$$N \text{ fixed} \quad \text{and} \quad n \to \infty \quad \text{as} \quad c = \lim_{n\to\infty} \frac{N}{n} = 0 \ .$$

- In the case $N, n \to \infty$, recall that the support of Marčenko-Pastur distribution is

$$[(1-\sqrt{c})^2, (1+\sqrt{c})^2] \ ,$$

i.e.

The higher $\sqrt{c}$, the larger the support

One can interpret $\sqrt{c}$ as a level of the **asymptotic noise induced by the data dimension** (=asymptotic data noise).

Hence the rule of thumb

Detection occurs if $\mathbf{snr}$ higher than **asymptotic data noise**.

Simulations

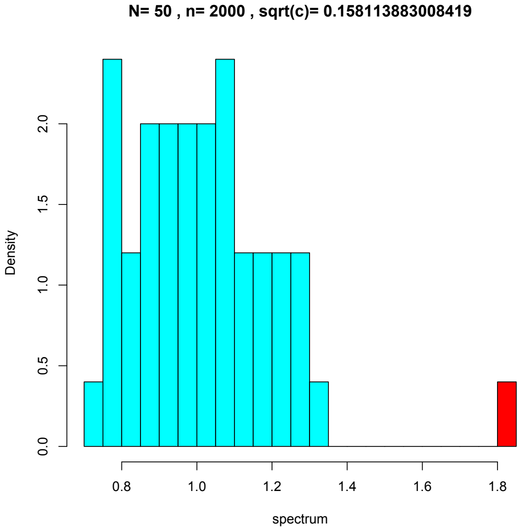


Figure : Influence of **asymptotic data noise** as $\sqrt{c}$ increases

Simulations

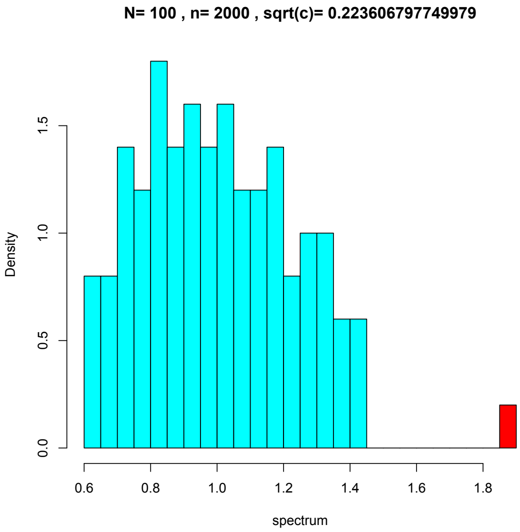


Figure : Influence of **asymptotic data noise** as $\sqrt{c}$ increases

Simulations

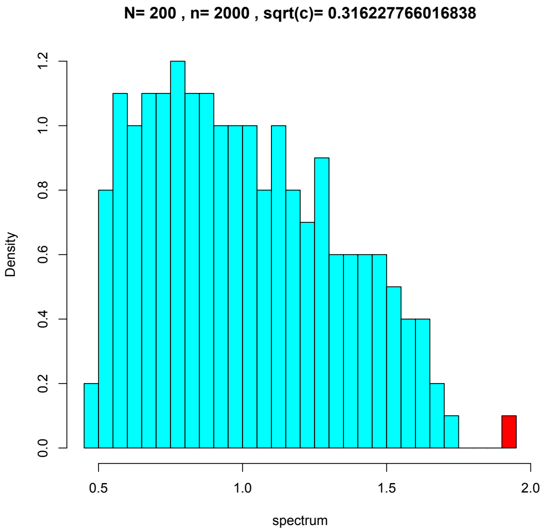


Figure : Influence of **asymptotic data noise** as $\sqrt{c}$ increases

Simulations

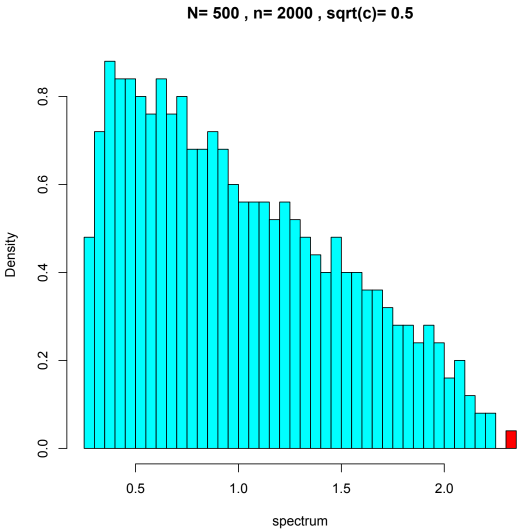


Figure : Influence of **asymptotic data noise** as $\sqrt{c}$ increases

Simulations

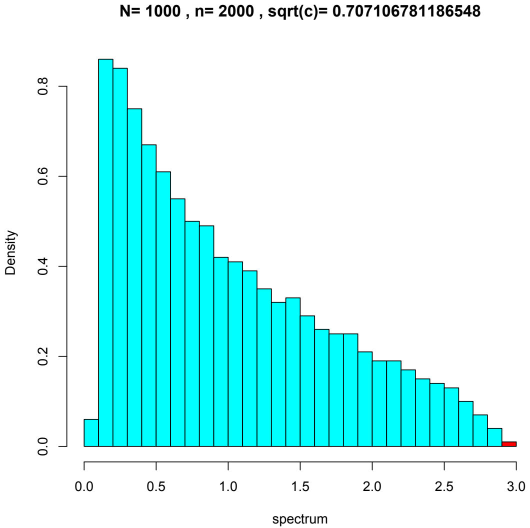


Figure : Influence of **asymptotic data noise** as $\sqrt{c}$ increases

Sketched proof - I

- We are interested in the largest eigenvalue of the matrix model

$$\frac{\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*}{\frac{1}{N}\operatorname{tr}(\hat{\mathbf{R}}_n)}$$

Sketched proof - I

- We are interested in the largest eigenvalue of the matrix model

$$\frac{\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*}{\frac{1}{N}\mathsf{tr}(\hat{\mathbf{R}}_n)}$$

asymptotically equivalent to

$$\frac{1}{n}\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{\sigma^2} \quad \text{as} \quad \frac{1}{N}\mathsf{tr}(\hat{\mathbf{R}}_n) \xrightarrow[N,n\to\infty]{a.s.} \sigma^2$$

Sketched proof - I

- We are interested in the largest eigenvalue of the matrix model

$$\frac{\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*}{\frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n)}$$

asymptotically equivalent to

$$\frac{1}{n}\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{\sigma^2} \quad \text{as} \quad \frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n) \xrightarrow[N,n\to\infty]{a.s.} \sigma^2$$

- Notice that

$$\mathbf{Y}_n = [\vec{\mathbf{y}}_1, \cdots, \vec{\mathbf{y}}_n] \quad \text{with} \quad \vec{\mathbf{y}}_i \sim \mathcal{C}N(0, \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N)$$

Sketched proof - I

- We are interested in the largest eigenvalue of the matrix model

$$\frac{\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*}{\frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n)}$$

asymptotically equivalent to

$$\frac{1}{n}\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{\sigma^2} \quad \text{as} \quad \frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n) \xrightarrow[N,n\to\infty]{a.s.} \sigma^2$$

- Notice that

$$\mathbf{Y}_n = [\vec{\mathbf{y}}_1, \cdots, \vec{\mathbf{y}}_n] \quad \text{with} \quad \vec{\mathbf{y}}_i \sim \mathcal{C}N(0, \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N)$$

Hence

$$\mathbf{Y}_N = \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N\right)^{1/2}\mathbf{X}_N$$

Sketched proof - I

- We are interested in the largest eigenvalue of the matrix model

$$\frac{\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*}{\frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n)}$$

asymptotically equivalent to

$$\frac{1}{n}\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{\sigma^2} \quad \text{as} \quad \frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n) \xrightarrow[N,n\to\infty]{a.s.} \sigma^2$$

- Notice that

$$\mathbf{Y}_n = [\vec{\mathbf{y}}_1, \cdots, \vec{\mathbf{y}}_n] \quad \text{with} \quad \vec{\mathbf{y}}_i \sim \mathcal{C}N(0, \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N)$$

Hence

$$\mathbf{Y}_N = \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N\right)^{1/2}\mathbf{X}_N \quad \Rightarrow \quad \frac{\mathbf{Y}_N}{\sigma} \quad = \quad \left(\mathbf{I}_N + \frac{\vec{\mathbf{h}}\vec{\mathbf{h}}^*}{\sigma^2}\right)^{1/2}\mathbf{X}_N$$

Sketched proof - I

- We are interested in the largest eigenvalue of the matrix model

$$\frac{\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*}{\frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n)}$$

asymptotically equivalent to

$$\frac{1}{n}\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{\sigma^2} \quad \text{as} \quad \frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n) \xrightarrow[N,n\to\infty]{a.s.} \sigma^2$$

- Notice that

$$\mathbf{Y}_n = [\vec{\mathbf{y}}_1, \cdots, \vec{\mathbf{y}}_n] \quad \text{with} \quad \vec{\mathbf{y}}_i \sim \mathcal{C}N(0, \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N)$$

Hence

$$\mathbf{Y}_N = \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N\right)^{1/2}\mathbf{X}_N \quad \Rightarrow \quad \frac{\mathbf{Y}_N}{\sigma} = \left(\mathbf{I}_N + \frac{\vec{\mathbf{h}}\vec{\mathbf{h}}^*}{\sigma^2}\right)^{1/2}\mathbf{X}_N$$

$$= \left(\mathbf{I}_N + \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2}\vec{\mathbf{u}}\vec{\mathbf{u}}^*\right)^{1/2}\mathbf{X}_N$$

with $\mathbf{X}_N$ a $N \times n$ matrix having i.i.d. entries $\mathcal{C}N(0,1)$ and $\vec{\mathbf{u}} = \frac{\vec{\mathbf{h}}}{\|\vec{\mathbf{h}}\|}$

Sketched proof - I

- We are interested in the largest eigenvalue of the matrix model

$$\frac{\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*}{\frac{1}{N}\text{tr}(\hat{\mathbf{R}}_n)}$$

asymptotically equivalent to

$$\frac{1}{n}\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{\sigma^2} \quad \text{as} \quad \frac{1}{N}\text{tr}(\hat{\mathbf{R}}_n) \xrightarrow[N,n\to\infty]{a.s.} \sigma^2$$

- Notice that

$$\mathbf{Y}_n = [\vec{\mathbf{y}}_1, \cdots, \vec{\mathbf{y}}_n] \quad \text{with} \quad \vec{\mathbf{y}}_i \sim \mathcal{C}N(0, \vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N)$$

Hence

$$\mathbf{Y}_N = \left(\vec{\mathbf{h}}\vec{\mathbf{h}}^* + \sigma^2\mathbf{I}_N\right)^{1/2}\mathbf{X}_N \quad \Rightarrow \quad \frac{\mathbf{Y}_N}{\sigma} = \left(\mathbf{I}_N + \frac{\vec{\mathbf{h}}\vec{\mathbf{h}}^*}{\sigma^2}\right)^{1/2}\mathbf{X}_N$$

$$= \left(\mathbf{I}_N + \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2}\vec{\mathbf{u}}\vec{\mathbf{u}}^*\right)^{1/2}\mathbf{X}_N$$

with $\mathbf{X}_N$ a $N \times n$ matrix having i.i.d. entries $\mathcal{C}N(0,1)$ and $\vec{\mathbf{u}} = \frac{\vec{\mathbf{h}}}{\|\vec{\mathbf{h}}\|}$

Conclusion

Spectrum of $\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*$ follows a spiked model with **rank-one perturbation**

Elements of proof - II

We can now conclude:

Elements of proof - II

We can now conclude:

- If $\boxed{\mathbf{snr} > \sqrt{c}}$ then

$$\frac{\lambda_{\max}\left(\hat{\mathbf{R}}_n\right)}{\frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n)} \xrightarrow[N,n\to\infty]{(H_1)} (1+\mathbf{snr})\left(1+\frac{c}{\mathbf{snr}}\right) > (1+\sqrt{c})^2$$

and the test statistics discriminates between the hypotheses H_0 and H_1.

Elements of proof - II

We can now conclude:

- If $\boxed{\mathbf{snr} > \sqrt{c}}$ then

$$\frac{\lambda_{\max}\left(\hat{\mathbf{R}}_n\right)}{\frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n)} \xrightarrow[N,n\to\infty]{(H_1)} (1+\mathbf{snr})\left(1+\frac{c}{\mathbf{snr}}\right) > (1+\sqrt{c})^2$$

  and the test statistics discriminates between the hypotheses H_0 and H_1.

- If $\boxed{\mathbf{snr} \leq \sqrt{c}}$ then

$$\frac{\lambda_{\max}\left(\hat{\mathbf{R}}_n\right)}{\frac{1}{N}\mathrm{tr}(\hat{\mathbf{R}}_n)} \xrightarrow[N,n\to\infty]{(H_1)} (1+\sqrt{c})^2$$

  Same limit as under H_0. The test statistics does not discriminate between the two hypotheses.

Outline

Fluctuations of the GLRT under H_0 - I

- The exact distribution of the statistics

$$L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$$

is needed to set the threshold of the test fo a given confidence level $\alpha \in (0,1)$:

$$\mathbb{P}_{H_0}\left(L_N > \boldsymbol{t_\alpha}\right) = \alpha \ ,$$

Fluctuations of the GLRT under H_0 - I

- The exact distribution of the statistics

$$L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\mathrm{tr}\,\hat{\mathbf{R}}_n}$$

is needed to set the threshold of the test fo a given confidence level $\alpha \in (0,1)$:

$$\mathbb{P}_{H_0}\left(L_N > \boldsymbol{t_\alpha}\right) = \alpha\ ,$$

but **hard to obtain**.

Fluctuations of the GLRT under H_0 - I

- The exact distribution of the statistics

$$L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$$

  is needed to set the threshold of the test fo a given confidence level $\alpha \in (0,1)$:

$$\mathbb{P}_{H_0}\left(L_N > \boldsymbol{t_\alpha}\right) = \alpha \ ,$$

  but **hard to obtain**.
- We rather study the asymptotic fluctuations of L_n under the regime

$$N, n \to \infty \ , \quad \frac{N}{n} \to c \in (0,1) \ .$$

Fluctuations of the GLRT under H_0 - I

- The exact distribution of the statistics

$$L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$$

is needed to set the threshold of the test fo a given confidence level $\alpha \in (0,1)$:

$$\mathbb{P}_{H_0}\left(L_N > \boldsymbol{t_\alpha}\right) = \alpha \ ,$$

but **hard to obtain**.

- We rather study the asymptotic fluctuations of L_n under the regime

$$N, n \to \infty \ , \quad \frac{N}{n} \to c \in (0,1) \ .$$

- L_N is the ratio of **two random variables**. We need to understand

Fluctuations of the GLRT under H_0 - I

- The exact distribution of the statistics

$$L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$$

is needed to set the threshold of the test fo a given confidence level $\alpha \in (0,1)$:

$$\mathbb{P}_{H_0}\left(L_N > \boldsymbol{t_\alpha}\right) = \alpha \ ,$$

but **hard to obtain**.

- We rather study the asymptotic fluctuations of L_n under the regime

$$N, n \to \infty \ , \quad \frac{N}{n} \to c \in (0,1) \ .$$

- L_N is the ratio of **two random variables**. We need to understand
  - the fluctuations of $\lambda_{\max}(\hat{\mathbf{R}}_n)$ under H_0,

Fluctuations of the GLRT under H_0 - I

- The exact distribution of the statistics

$$L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$$

  is needed to set the threshold of the test fo a given confidence level $\alpha \in (0,1)$:

$$\mathbb{P}_{H_0}\left(L_N > \boldsymbol{t_\alpha}\right) = \alpha \ ,$$

  but **hard to obtain**.
- We rather study the asymptotic fluctuations of L_n under the regime

$$N, n \to \infty \ , \quad \frac{N}{n} \to c \in (0,1) \ .$$

- L_N is the ratio of **two random variables**. We need to understand
  - the fluctuations of $\lambda_{\max}(\hat{\mathbf{R}}_n)$ under H_0,
  - the fluctuations of $\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n$ under H_0.

Fluctuations of the GLRT under H_0 - II

Fluctuations of $\lambda_{\max}(\hat{\mathbf{R}}_n)$: Tracy-Widom distribution **at rate $N^{2/3}$**

Fluctuations of the GLRT under H_0 - II

Fluctuations of $\lambda_{\max}(\hat{\mathbf{R}}_n)$: Tracy-Widom distribution **at rate $N^{2/3}$**

$$\frac{N^{2/3}}{\Theta_N}\ \left\{\lambda_{\max}\left(\hat{\mathbf{R}}_n\right) - \sigma^2(1+\sqrt{c_n})^2\right\} \xrightarrow[N,n\to\infty]{\mathcal{L}} \mathbb{P}_{\mathrm{TW}}$$

Fluctuations of the GLRT under H_0 - II

Fluctuations of $\lambda_{\max}(\hat{\mathbf{R}}_n)$: Tracy-Widom distribution **at rate $N^{2/3}$**

$$\frac{N^{2/3}}{\Theta_N} \left\{ \lambda_{\max}\left(\hat{\mathbf{R}}_n\right) - \sigma^2(1+\sqrt{c_n})^2 \right\} \xrightarrow[N,n\to\infty]{\mathcal{L}} \mathbb{P}_{\mathrm{TW}}$$

where

$$c_n = \frac{N}{n} \quad \text{and} \quad \Theta_N = \sigma^2(1+\sqrt{c_n})\left(\frac{1}{\sqrt{c_n}}+1\right)^{1/3}$$

Fluctuations of the GLRT under H_0 - II

Fluctuations of $\lambda_{\max}(\hat{\mathbf{R}}_n)$: Tracy-Widom distribution **at rate $N^{2/3}$**

$$\frac{N^{2/3}}{\Theta_N} \left\{ \lambda_{\max}\left(\hat{\mathbf{R}}_n\right) - \sigma^2(1+\sqrt{c_n})^2 \right\} \xrightarrow[N,n\to\infty]{\mathcal{L}} \mathbb{P}_{\mathrm{TW}}$$

where

$$c_n = \frac{N}{n} \quad \text{and} \quad \Theta_N = \sigma^2(1+\sqrt{c_n})\left(\frac{1}{\sqrt{c_n}}+1\right)^{1/3}$$

Otherwise stated,

$$\boxed{\lambda_{\max}\left(\hat{\mathbf{R}}_n\right) = \sigma^2(1+\sqrt{c_n})^2 + \frac{\Theta_N}{N^{2/3}}\boldsymbol{X}_{TW} + o_P(N^{-2/3})}$$

where $\boldsymbol{X}_{TW}$ is a random variable with Tracy-Widom distribution.

Details on Tracy-Widom distribution

Tracy-Widom distribution is defined by

- its cumulative distribution function

$$F_{TW}(x) = \exp\left\{-\int_x^\infty (u-x)^2 q^2(u)\,du\right\}$$

- where

$$q''(x) = xq(x) + 2q^3(x) \quad \text{and} \quad q(x) \sim \mathrm{Ai}(x) \text{ as } x \to \infty \,.$$

  $x \mapsto \mathrm{Ai}(x)$ being the Airy function.

Details on Tracy-Widom distribution

Tracy-Widom distribution is defined by

- its cumulative distribution function

$$F_{TW}(x) = \exp\left\{-\int_x^\infty (u-x)^2 q^2(u)\,du\right\}$$

- where

$$q''(x) = xq(x) + 2q^3(x) \quad \text{and} \quad q(x) \sim \text{Ai}(x) \text{ as } x \to \infty\ .$$

  $x \mapsto \text{Ai}(x)$ being the Airy function.

Don't bother .. just download it

- For simulations, cf. `R Package 'RMTstat'`, by Johnstone et al.
- Also, Folkmar Bornemann (TU München) has developed fast `matlab` code

Tracy-Widom curve

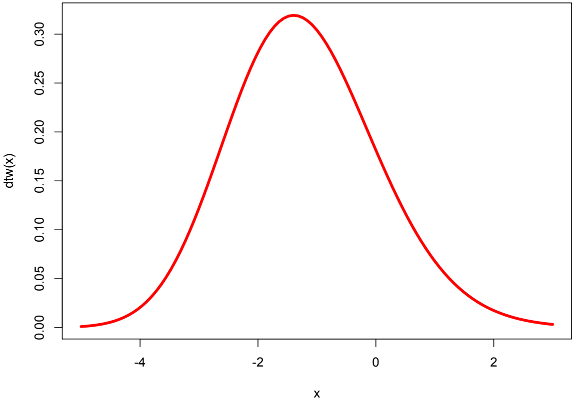


Figure : Tracy-Widom density

Tracy-Widom curve

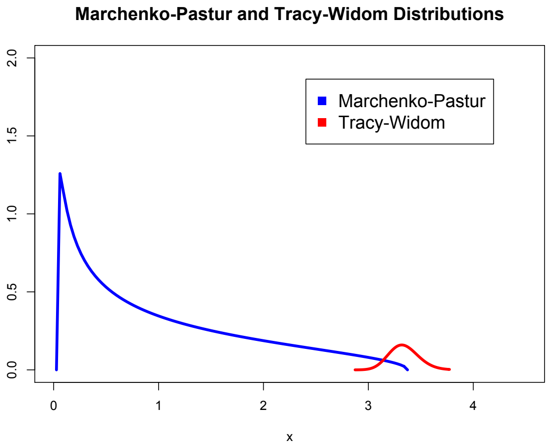


Figure : Fluctuations of the largest eigenvalue $\lambda_{\max}(\hat{\mathbf{R}}_n)$ under H_0

Fluctuations of the GLRT under H_0 - III

Fluctuations of $\frac{1}{N}\mathrm{tr}\,(\hat{\mathbf{R}}_n)$: Gaussian distributions **at rate** $\boldsymbol{N}$

$$N\left\{\frac{1}{N}\sum_{i=1}^{N}\lambda_i(\hat{\mathbf{R}}_n)-\sigma^2\right\}\xrightarrow[N,n\to\infty]{\mathcal{L}}\mathcal{N}(0,\Gamma)\ ,$$

Fluctuations of the GLRT under H_0 - III

Fluctuations of $\frac{1}{N}\text{tr}\,(\hat{\mathbf{R}}_n)$: Gaussian distributions **at rate N**

$$N\left\{\frac{1}{N}\sum_{i=1}^{N}\lambda_i(\hat{\mathbf{R}}_n)-\sigma^2\right\}\xrightarrow[N,n\to\infty]{\mathcal{L}}\mathcal{N}(0,\Gamma)\ ,$$

Otherwise stated:

$$\boxed{\frac{1}{N}\text{tr}\,(\hat{\mathbf{R}}_n)=\frac{1}{N}\sum_{i=1}^{N}\lambda_i(\hat{\mathbf{R}}_n)=\sigma^2+\frac{\sqrt{\Gamma}}{N}\boldsymbol{Z}+o_P(N^{-1})}$$

where $\boldsymbol{Z}$ is a random variable with distribution $\mathcal{N}(0,1)$.

Fluctuations of the GLRT under H_0 - IV

Conclusion

- Fluctuations of $L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$ are driven by $\lambda_{\max}(\hat{\mathbf{R}}_n)$:

$$\frac{N^{2/3}}{\widetilde{\Theta}_N}\left\{L_N - (1+\sqrt{c_n})^2\right\} \xrightarrow[N,n\to\infty]{\mathcal{L}} \mathbb{P}_{\mathrm{TW}} \quad \text{with} \quad \widetilde{\Theta}_N = (1+\sqrt{c_n})\left(\frac{1}{\sqrt{c_n}}+1\right)^{1/3}$$

Fluctuations of the GLRT under H_0 - IV

Conclusion

- Fluctuations of $L_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}$ are driven by $\lambda_{\max}(\hat{\mathbf{R}}_n)$:

$$\frac{N^{2/3}}{\widetilde{\Theta}_N}\left\{L_N - (1+\sqrt{c_n})^2\right\} \xrightarrow[N,n\to\infty]{\mathcal{L}} \mathbb{P}_{\mathrm{TW}} \quad \text{with} \quad \widetilde{\Theta}_N = (1+\sqrt{c_n})\left(\frac{1}{\sqrt{c_n}}+1\right)^{1/3}$$

- In order to set the threshold $\boldsymbol{\alpha}$, we choose $\boldsymbol{t_\alpha^n}$ as

$$\boldsymbol{t_\alpha^n} = (1+\sqrt{c_n})^2 + \frac{\widetilde{\Theta}_N}{N^{2/3}}\boldsymbol{t_\alpha}^{\textbf{Tracy-Widom}}$$

  where $\boldsymbol{t_\alpha}^{\textbf{Tracy-Widom}}$ is the corresponding quantile for a Tracy-Widom random variable:

$$\mathbb{P}\{\boldsymbol{X}_{TW} > \boldsymbol{t_\alpha}^{\textbf{Tracy-Widom}}\} \le \alpha.$$

Outline

Summary

- Consider the following hypothesis

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

Summary

- Consider the following hypothesis

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1:n$$

then the GLRT amounts to study

$$\boxed{T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}}$$

Summary

- Consider the following hypothesis

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1:n$$

then the GLRT amounts to study

$$\boxed{T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\operatorname{tr}\hat{\mathbf{R}}_n}}$$

- The test statistics T_n discriminates between H_0 and H_1 if $\boxed{\mathbf{snr} = \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2} > \sqrt{c}}$

Summary

- Consider the following hypothesis

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma \vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma \vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

  then the GLRT amounts to study

$$\boxed{T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N} \operatorname{tr} \hat{\mathbf{R}}_n}}$$

- The test statistics T_n discriminates between H_0 and H_1 if $\boxed{\mathbf{snr} = \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2} > \sqrt{c}}$
- The threshold can be asymptotically determined by Tracy-Widom quantiles.

Summary

- Consider the following hypothesis

$$\vec{\mathbf{y}}(k) = \begin{cases} \sigma\vec{\mathbf{w}}(k) & \text{under } H_0 \\ \vec{\mathbf{h}}\, s(k) + \sigma\vec{\mathbf{w}}(k) & \text{under } H_1 \end{cases} \qquad \text{for } k = 1 : n$$

then the GLRT amounts to study

$$\boxed{T_n = \frac{\lambda_{\max}(\hat{\mathbf{R}}_n)}{\frac{1}{N}\mathrm{tr}\,\hat{\mathbf{R}}_n}}$$

- The test statistics T_n discriminates between H_0 and H_1 if $\boxed{\mathbf{snr} = \frac{\|\vec{\mathbf{h}}\|^2}{\sigma^2} > \sqrt{c}}$
- The threshold can be asymptotically determined by Tracy-Widom quantiles.
- The type II error (equivalentlty power of the test) can be analyzed via the error exponent of the test

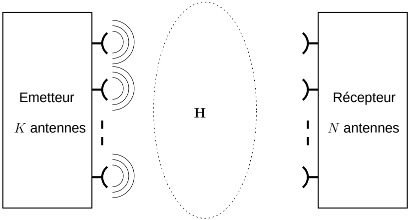
$$\boldsymbol{\mathcal{E}} = \lim_{N,n\to\infty} -\frac{1}{n}\log \mathbb{P}_{H_1}(L_N < \boldsymbol{t_\alpha})\ .$$

Outline

MIMO channel

MIMO = Multiple Input Multiple Output

It is a channel with multiple antennas at the emission and reception



- The received signal writes: $\boxed{\vec{\mathbf{y}} = \mathbf{H}\vec{\mathbf{x}} + \vec{\mathbf{v}}}$ where
  - $\vec{\mathbf{x}}$ is the signal that is sent,
  - $\vec{\mathbf{v}}$ is an additive gaussian white noise with variance σ^2,
  - $\mathbf{H}$ is the random gain matrix. Its distribution is associated to the features of the channel.
  - $\vec{\mathbf{y}}$ is the received signal.

Features of the Gain matrix $\mathbf{H}$

- The entry $[\mathbf{H}]_{ij}$ represents the gain between emitting antenna j and receiving antenna i.
- The gain matrix $\mathbf{H}$ is random.
- The distribuon of $\mathbf{H}$ depends on the nature of the channel:
  - ▷ Absence of correlation between antennas

$$\mathbf{H} \quad = \quad \frac{1}{\sqrt{K}}\mathbf{X} \qquad [\mathbf{X}]_{ij} \text{ à entrées i.i.d., variance } \theta^2$$

  - ▷ Correlation between emitting antennas ($\tilde{\mathbf{D}}^{1/2}$) and receiving antennas ($\mathbf{D}^{1/2}$)

$$\mathbf{H} \quad = \quad \frac{1}{\sqrt{K}}\mathbf{D}^{1/2}\mathbf{X}\tilde{\mathbf{D}}^{1/2} \qquad \text{(Rayleigh channel)}$$

  - ▷ Existence of a line-of-sight component (matrix $\mathbf{A}$ deterministic) + correlations

$$\mathbf{H} \quad = \quad \frac{1}{\sqrt{K}}\mathbf{D}^{1/2}\mathbf{X}\tilde{\mathbf{D}}^{1/2} + \mathbf{A} \qquad \text{(Rice channel)}$$

Performances

- Shannon's **mutual information** (per antenna)

$$\boxed{\mathcal{I} = \frac{1}{N}\log\det\left(\mathbf{I} + \frac{\mathbf{HH}^*}{\sigma^2}\right)} = \frac{1}{N}\sum_{i=1}^{N}\log\left(1 + \frac{\lambda_i(\mathbf{HH}^*)}{\sigma^2}\right)$$

$\Rightarrow$ depends on the spectrum of matrix $\mathbf{HH}^*$.

- **Ergodic Mutual Information**:

$$\boxed{\mathcal{I}^{\mathbf{e}} = \mathbb{E}\,\mathcal{I}\ .}$$

- **Ergodic capacity**:

$$\boxed{\sup_{\mathbf{Q}\geq 0,\, \frac{1}{K}\operatorname{tr}\mathbf{Q}\leq 1} \mathbb{E}\log\det\left(I + \frac{\mathbf{HQH}^*}{\sigma^2}\right)}$$

- Regime of interest:

{ # emitting antennas} $\propto$ { # receiving antennas}

Questions

- Behaviour of the empirical measure of the eigenvalues:

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(\mathbf{HH}^*)}$$

- Explicit expression for the **logdet**:

$$\frac{1}{N} \log\det\left(I + \frac{\mathbf{HH}^*}{\sigma^2}\right) = \frac{1}{N} \sum_{i=1}^{N} \log\left(1 + \frac{\lambda_i \mathbf{HH}^*}{\sigma^2}\right)$$

- Fluctuations?
- Ergodic capacity $\Rightarrow$ Optimisation?
- Asymptotic regime: $\boxed{N \propto K}$. Formally

$$N, K \to \infty, \qquad \frac{N}{K} \to c \in (0, \infty)$$

It's the asymptotic regime of large random matrices.

Empirical measure: the white case

Channel $\mathbf{H}$ with i.i.d. entries

- Marčenko-Pastur Stieltjes transform $g(z) = \int \frac{\mu_{\mathrm{MP}}(d\lambda)}{\lambda - z}$ satisfies:

$$zc\theta^2 g^2(z) + (z + (c-1)\theta^2) g(z) + 1 = 0 \ .$$

- Convergence of the mutual information:

$$\begin{aligned} \mathcal{I} \quad &= \quad \frac{1}{N} \log\det\left(I + \frac{\mathbf{H}\mathbf{H}^*}{\sigma^2}\right) = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + \frac{\lambda_i(\mathbf{H}\mathbf{H}^*)}{\sigma^2}\right) \\ \longrightarrow \quad & \mathcal{I}_{\mathrm{approx}} \triangleq \int \log\left(1 + \frac{x}{\sigma^2}\right) \mu_{\mathrm{MP}}(dx) \\ &= \int_{\sigma^2}^{\infty} \left(\frac{1}{w} - g(-w)\right) dw \end{aligned}$$

- Explicit formula for the limit:

$$\mathcal{I}_{\mathrm{approx}} = -\log \sigma^2 g(-\sigma^2) + \frac{1}{c} \log\left(\frac{1 + c\theta^2 g(-\sigma^2)}{\sigma^2}\right) - \frac{\theta^2 g(-\sigma^2)}{1 + c\theta^2 g(-\sigma^2)}$$

- Important results:
  1. $\boxed{\mathbb{E} \log\det\left(I + \frac{\mathbf{H}\mathbf{H}^*}{\sigma^2}\right) \propto \min(N, K)}$
  2. Speed of convergence [for Gaussian entries]: $\boxed{\mathcal{I}^{\mathbf{e}} - \mathcal{I}_{\mathrm{approx}} = \mathcal{O}\left(\frac{1}{N^2}\right)}$

Rice channel

The gain matrix writes in this case:

$$\mathbf{H} = \frac{1}{\sqrt{K}}\mathbf{D}^{1/2}\mathbf{X}\tilde{\mathbf{D}}^{1/2} + \mathbf{A}$$

- We have again $\mathcal{I}^{\mathbf{e}} - \mathcal{I}^{\mathbf{e}}_{\mathtt{approx}} \to 0$ where

$$\mathcal{I}^{\mathbf{e}}_{\mathtt{approx}} = \frac{1}{N}\log\det\left[\mathbf{I} + \tilde{\delta}\mathbf{D} + \frac{\mathbf{1}}{\sigma^{\mathbf{2}}}\mathbf{A}(\mathbf{I}+\delta\tilde{\mathbf{D}})^{-\mathbf{1}}\mathbf{A}^*\right] + \frac{1}{N}\log\det\left(\mathbf{I}+\delta\tilde{\mathbf{D}}\right) - \frac{\sigma^2 n}{N}\delta\tilde{\delta}$$

  and $(\delta_n, \tilde{\delta}_n)$ unique solutions of the system:

$$\delta = \frac{1}{n}\mathrm{tr}\left[\mathbf{D}\left(-z(\mathbf{I}+\tilde{\delta}\mathbf{D}) + \mathbf{A}(\mathbf{I}+\delta\tilde{\mathbf{D}})^{-\mathbf{1}}\mathbf{A}^*\right)^{-1}\right]$$

$$\tilde{\delta} = \frac{1}{n}\mathrm{tr}\left[\tilde{\mathbf{D}}\left(-z(\mathbf{I}+\delta\tilde{\mathbf{D}}) + \mathbf{A}^*(\mathbf{I}+\tilde{\delta}\mathbf{D})^{-\mathbf{1}}\mathbf{A}\right)^{-1}\right]$$

- moreover, $\mathcal{I} - \mathcal{I}_{\mathtt{approx}} = \mathcal{O}\left(\frac{1}{N^2}\right)$ for Gaussian entries

Ergodic capacity and precoding

MIMO channel with precoding

- The channel becomes $\mathbf{HQ}^{1/2}$, mutual information becomes

$$\mathcal{I}^{\mathbf{e}}(\mathbf{Q}) = \frac{1}{N}\mathbb{E}\log\det\left(\mathbf{I}_N + \frac{\mathbf{HQH}^*}{\sigma^2}\right)$$

- We can still compute a "large random matrix" approximation

$$\begin{aligned}
\mathcal{I}^{\mathbf{e}}_{\mathtt{approx}} &= \mathcal{I}^{\mathbf{e}}_{\mathtt{approx}}(\mathbf{Q}) \\
&= \frac{1}{N}\log\det\left[\mathbf{I} + \tilde{\delta}\mathbf{D} + \frac{\mathbf{1}}{\boldsymbol{\sigma}^{\mathbf{2}}}\mathbf{AQ}^{\mathbf{1/2}}(\mathbf{I} + \delta\tilde{\mathbf{D}}\mathbf{Q})^{-\mathbf{1}}\mathbf{Q}^{\mathbf{1/2}}\mathbf{A}^*\right] \\
&\quad + \frac{1}{N}\log\det\left(\mathbf{I} + \delta\tilde{\mathbf{D}}\mathbf{Q}\right) - \frac{\sigma^2 n}{N}\delta\tilde{\delta}
\end{aligned}$$

Ergodic capacity

The **ergodic capacity** is obtained by optimizing the mutual information with respect to linear precoders $\mathbf{Q}^{1/2}$ with finite energy:

$$C = \sup_{\mathbf{Q}\geq 0;\, \frac{1}{K}\operatorname{Tr}\mathbf{Q}\leq 1} \frac{1}{K}\mathbb{E}\log\det\left(\mathbf{I}_N + \frac{\mathbf{HQH}^*}{\sigma^2}\right)$$

Approximating problem

Consider the following approximating problem:

$$C_{\texttt{approx}} = \sup_{\mathbf{Q}\geq 0;\, \frac{1}{K}\operatorname{Tr}\mathbf{Q}\leq 1} \mathcal{I}^{\mathbf{e}}_{\texttt{approx}}(\mathbf{Q})$$

Results

1. We have $C - C_{\texttt{approx}} \to 0$
2. $\mathbf{Q}^* = \arg\max \mathcal{I}^{\mathbf{e}}(\mathbf{Q})$ close to ${\mathbf{Q}^*}_{\texttt{approx}} = \arg\max \mathcal{I}^{\mathbf{e}}_{\texttt{approx}}(\mathbf{Q})$
3. Exists an iterative algorithm (i.e. quick) to compute $C_{\texttt{approx}}$ and ${\mathbf{Q}^*}_{\texttt{approx}}$

Simulations

- The iterative algorithm outperforms Paulraj & Vu algorithm with respect to the complexity (average time per iterations - in s):

	$N = n = 2$	$N = n = 4$	$N = n = 8$
Paulraj-Vu	0.75	8.2	138
iterative algo.	10^{-2}	3.10^{-2}	7.10^{-2}

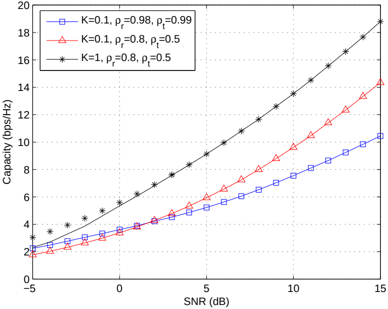


Figure : Comparing with Vu & Paulraj algorithm