Théorie des matrices aléatoires et apprentissage

Romain Couillet & Jamal Najim

romain.couillet@gipsa-lab.grenoble-inp.fr,
najim@univ-mlv.fr

Master Mathématiques / **V**ision / **A**pprentissage - 2020

Random Matrix Theory (RMT) in an Nutshell

High dimensional statistics and ML applications

Outline of the course

Random Matrix Theory

The purpose of (Large) Random Matrix Theory is

Random Matrix Theory

The purpose of (Large) Random Matrix Theory is

- To describe the

$$\boxed{\textbf{eigenstructure} = \text{eigenvalues} + \text{eigenvectors}}$$

  of a matrix with **random entries**

Random Matrix Theory

The purpose of (Large) Random Matrix Theory is

- To describe the

  $$\boxed{\textbf{eigenstructure} = \text{eigenvalues} + \text{eigenvectors}}$$

  of a matrix with **random entries**

- Difficult task but

  $$\boxed{\textbf{tractable} \text{ if the dimensions of the matrix go to infinity}}$$

Random Matrix Theory

The purpose of (Large) Random Matrix Theory is

- To describe the

  $$\boxed{\textbf{eigenstructure} = \text{eigenvalues} + \text{eigenvectors}}$$

  of a matrix with **random entries**

- Difficult task but

  $$\boxed{\textbf{tractable} \text{ if the dimensions of the matrix go to infinity}}$$

- First result (1948) goes back to Nobel Laureate Eugene Wigner known as

  $$\boxed{\text{Wigner's semi-circle law}}$$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

- Define **Wigner matrix**

$$\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

- Define **Wigner matrix**

$$\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$$

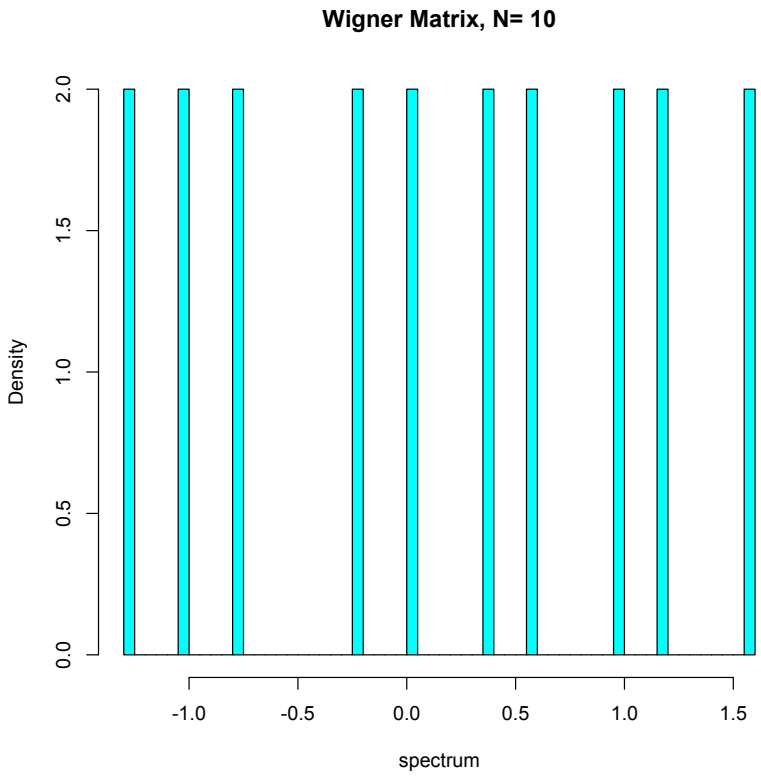

Figure: Histogram of the eigenvalues of $\mathbf{Y}_N$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

- Define **Wigner matrix**

$$\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$$

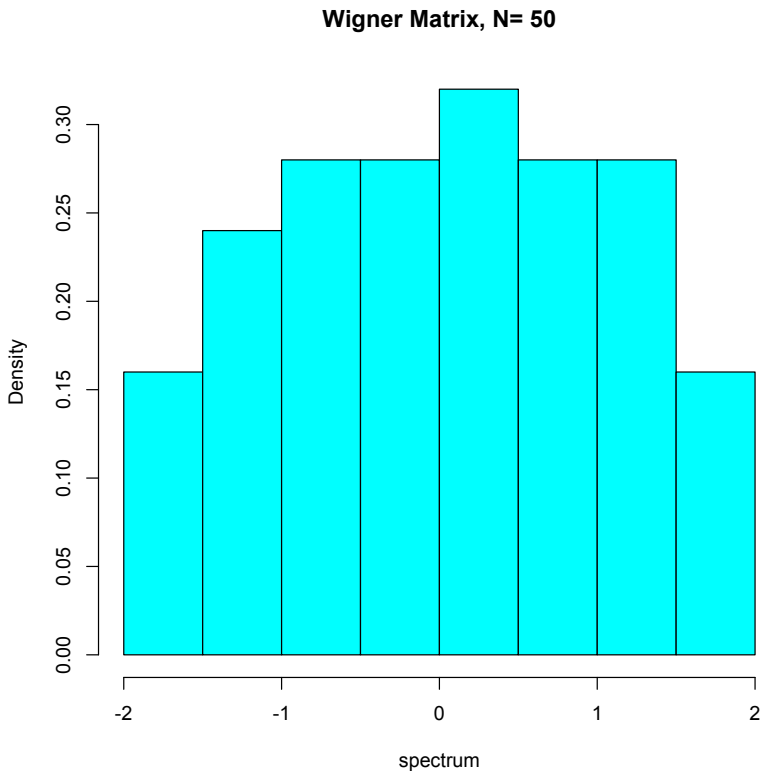

Figure: Histogram of the eigenvalues of $\mathbf{Y}_N$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

- Define **Wigner matrix**

$$\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$$

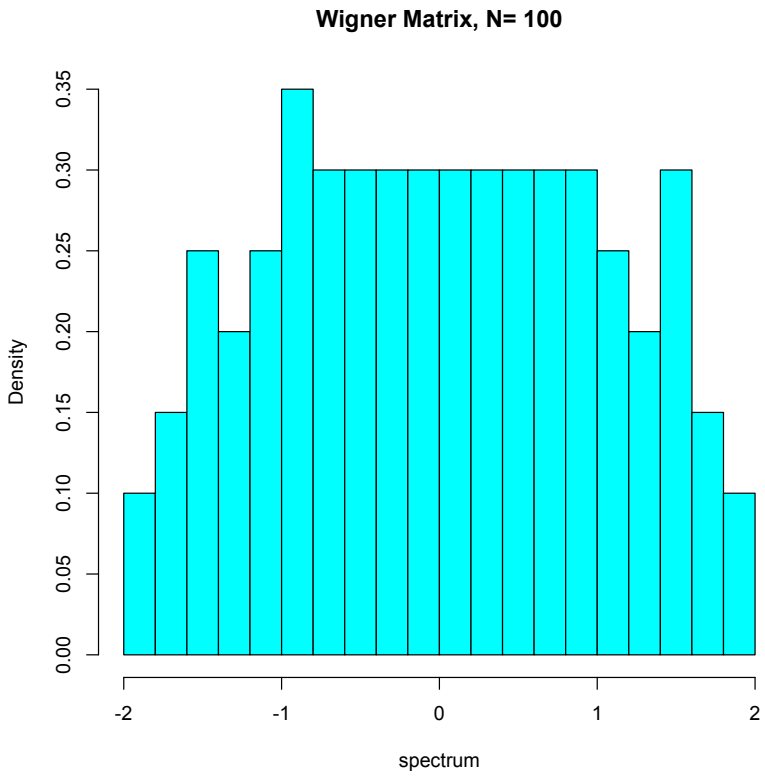

Figure: Histogram of the eigenvalues of $\mathbf{Y}_N$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

- Define **Wigner matrix**

$$\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$$

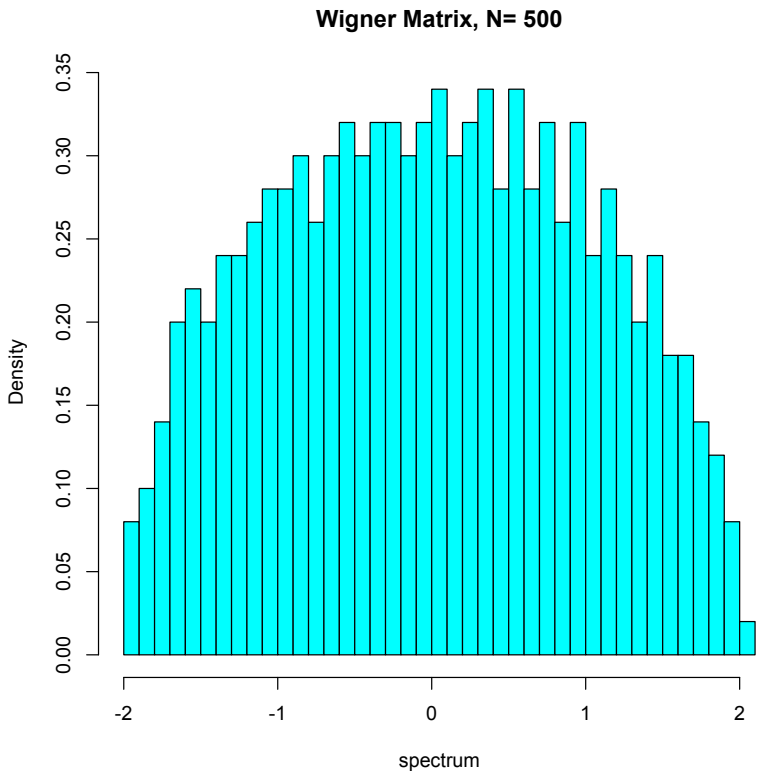

Figure: Histogram of the eigenvalues of $\mathbf{Y}_N$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

- Define **Wigner matrix**

$$\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$$

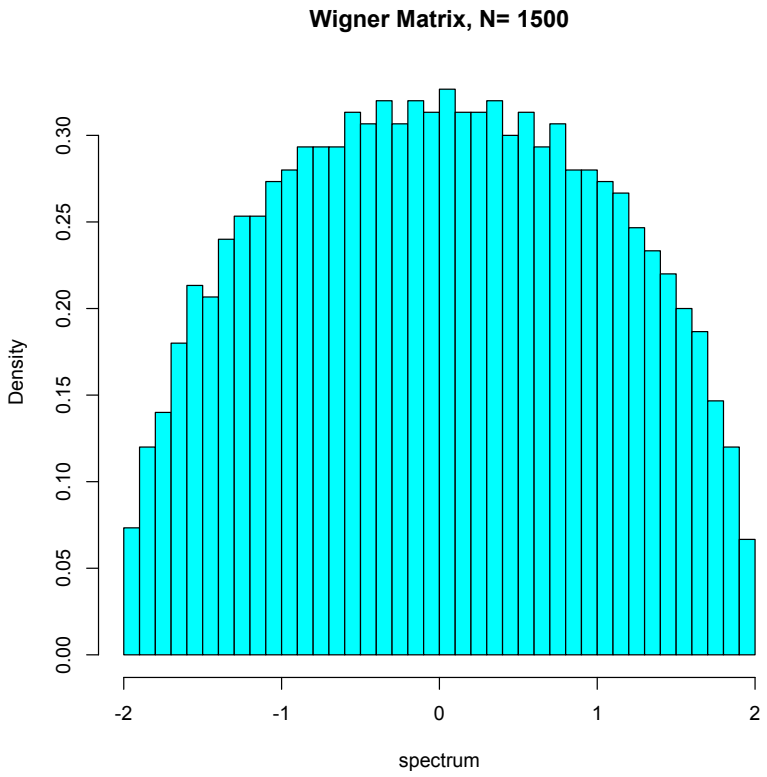

Figure: Histogram of the eigenvalues of $\mathbf{Y}_N$

Example 1: Wigner Model

Matrix model

Let $\mathbf{X}_N = (X_{ij})$ symmetric $N \times N$

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{12} & \ddots & & \\ \vdots & & \ddots & \\ X_{1N} & & & X_{NN} \end{pmatrix}$$

- Define **Wigner matrix**

$$\boxed{\mathbf{Y}_N = \frac{\mathbf{X}_N}{\sqrt{N}}}$$

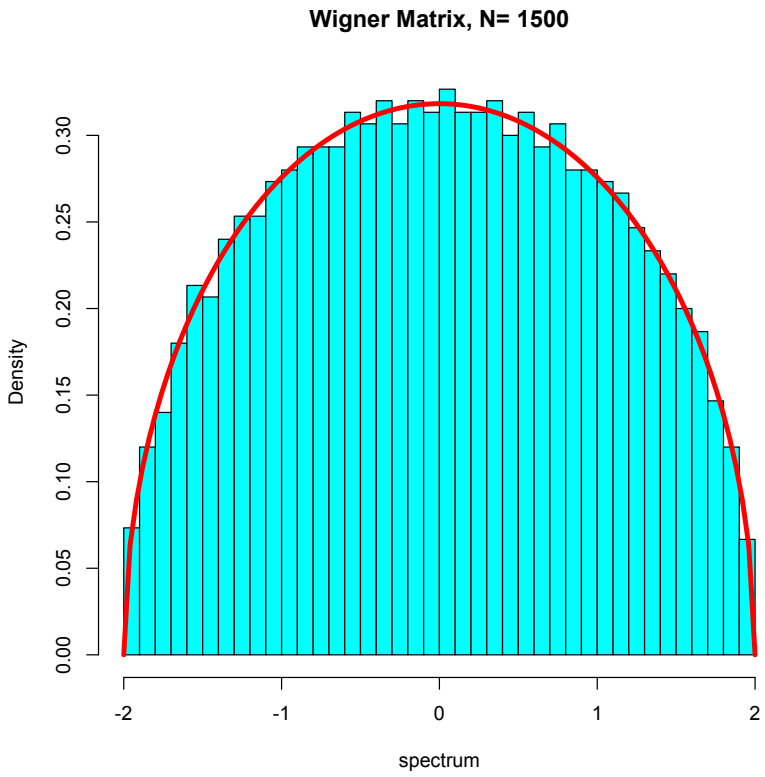

Figure: The semi-circular distribution (in red) with density $x \mapsto \frac{\sqrt{4-x^2}}{2\pi}$

Wigner's theorem (1948)

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

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

- Consider the **empirical covariance matrix**

$$\boxed{\hat{\mathbf{R}}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*}$$

with dimensions of **same order**:

$$\boxed{\frac{N}{n} \approx c\,, \quad c > 0}$$

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

- Consider the **empirical covariance matrix**

$$\boxed{\hat{\mathbf{R}}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*}$$

with dimensions of **same order**:

$$\boxed{\frac{N}{n} \approx c\,, \quad c > 0}$$

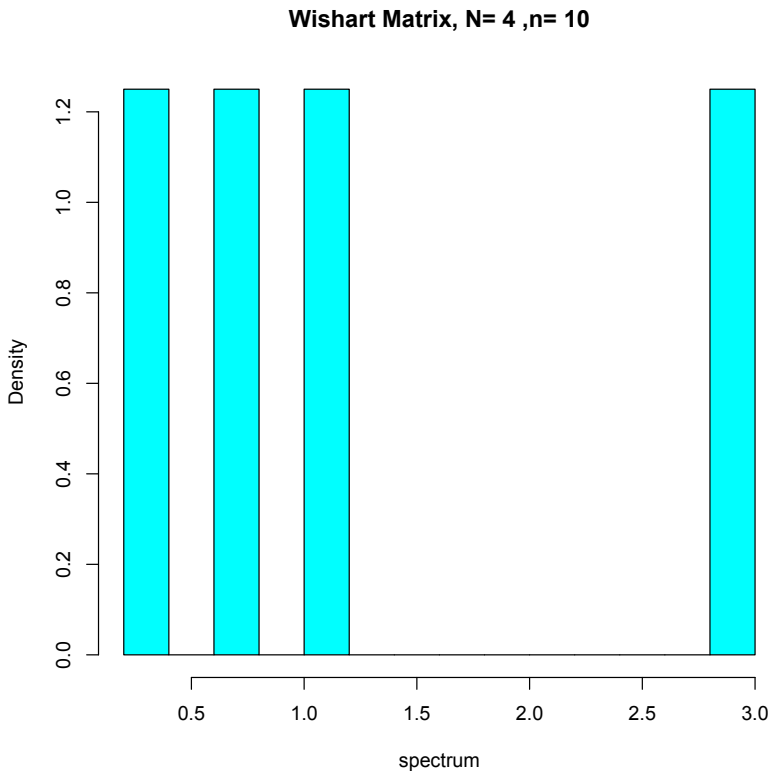

Figure: Spectrum's histogram - $\frac{N}{n} = 0.4$

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

- Consider the **empirical covariance matrix**

$$\boxed{\hat{\mathbf{R}}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*}$$

with dimensions of **same order**:

$$\boxed{\frac{N}{n} \approx c\,, \quad c > 0}$$

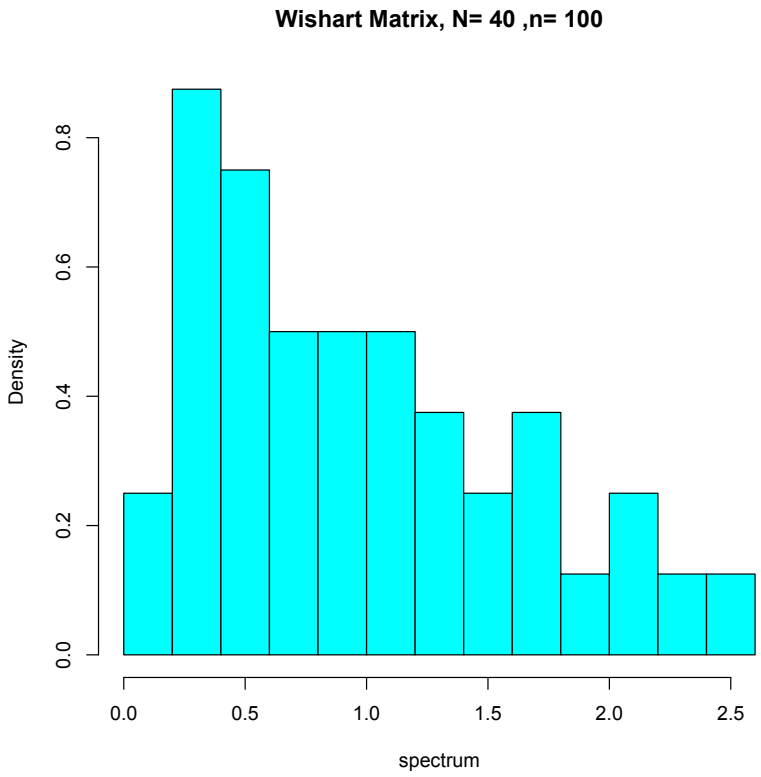

Figure: Spectrum's histogram - $\frac{N}{n} = 0.4$

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

- Consider the **empirical covariance matrix**

$$\boxed{\hat{\mathbf{R}}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*}$$

with dimensions of **same order**:

$$\boxed{\frac{N}{n} \approx c\,, \quad c > 0}$$

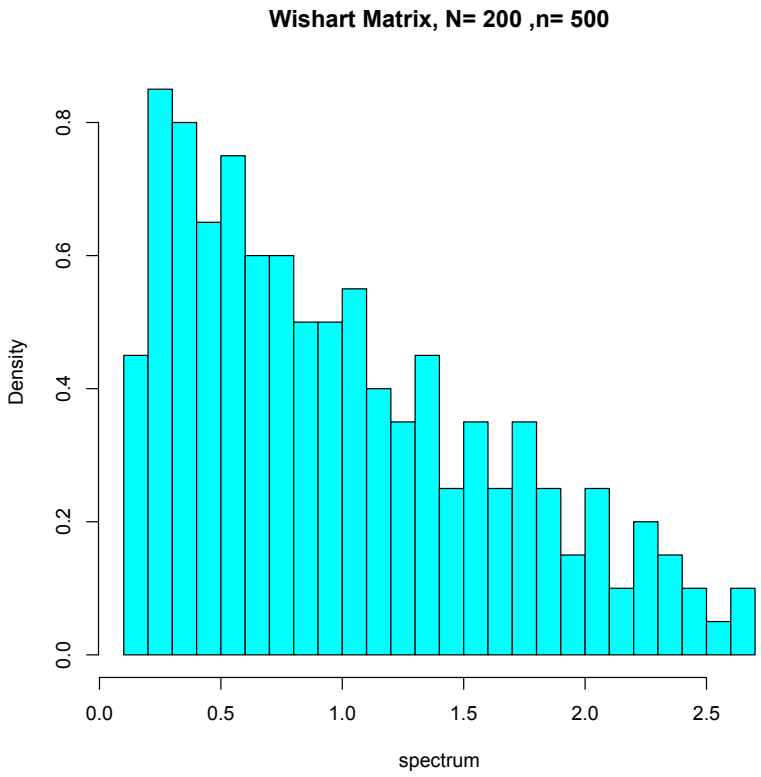

Figure: Spectrum's histogram - $\frac{N}{n} = 0.4$

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

- Consider the **empirical covariance matrix**

$$\boxed{\hat{\mathbf{R}}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*}$$

with dimensions of **same order**:

$$\boxed{\frac{N}{n} \approx c\,, \quad c > 0}$$

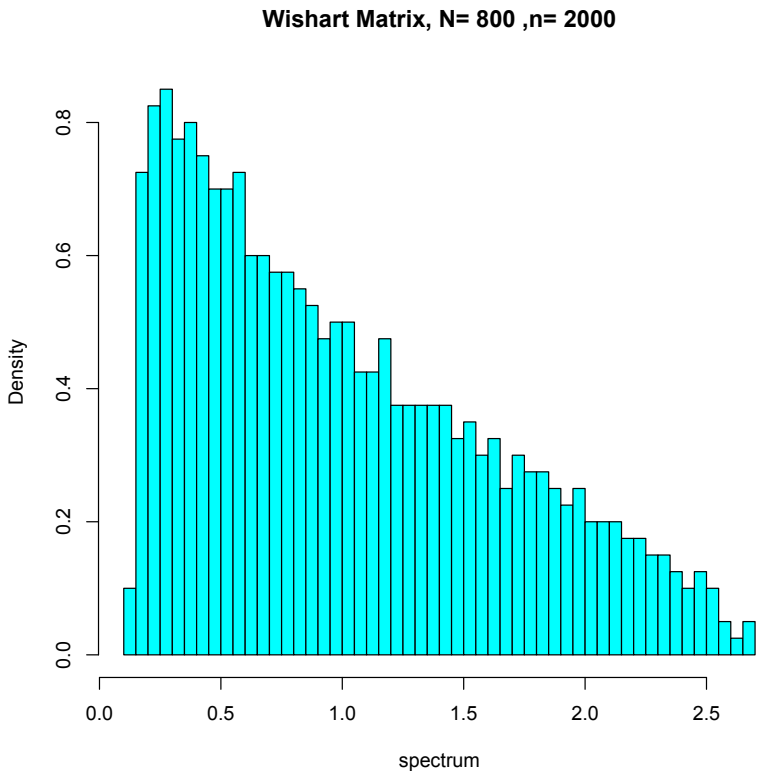

Figure: Spectrum's histogram - $\frac{N}{n} = 0.4$

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

- Consider the **empirical covariance matrix**

$$\boxed{\hat{\mathbf{R}}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*}$$

with dimensions of **same order**:

$$\boxed{\frac{N}{n} \approx c\,, \quad c > 0}$$

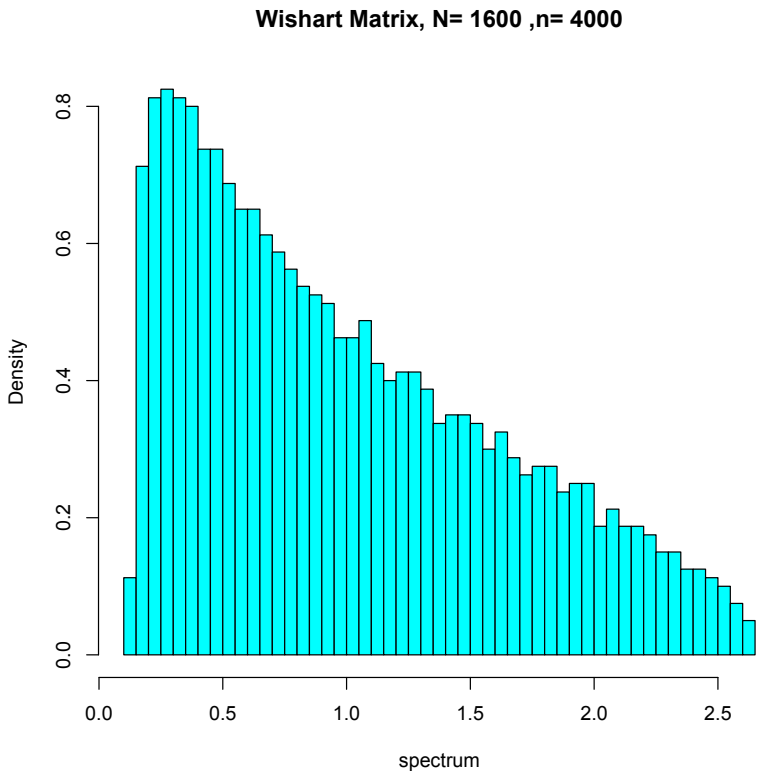

Figure: Spectrum's histogram - $\frac{N}{n} = 0.4$

Example 2: Large Covariance Matrices

Matrix model

Let $\mathbf{X}_N$ be a $N \times n$ rectangular matrix

$$\mathbf{X}_N = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{N1} & \cdots & X_{Nn} \end{pmatrix}$$

- Consider the **empirical covariance matrix**

$$\boxed{\hat{\mathbf{R}}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*}$$

with dimensions of **same order**:

$$\boxed{\frac{N}{n} \approx c\,, \quad c > 0}$$

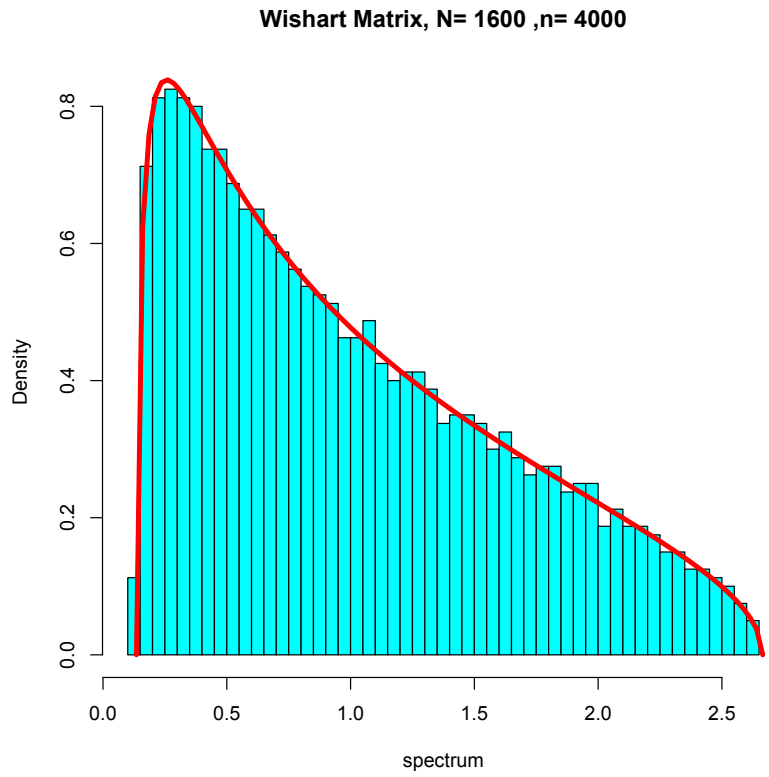

Figure: Marčenko-Pastur's distribution (in red)

Marčenko-Pastur's theorem (1967)

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

RMT and its relation with other subjects

- [maths] combinatorics, free probability, quantum information, operator theory
- Theoretical physics
- Wireless communication
- Random graphs
- **High Dimensional Statistics**
- **Machine Learning**
- Theoretical Ecology

Random Matrix Theory (RMT) in an Nutshell

High dimensional statistics and ML applications

Outline of the course

High dimensional statistics and RMT

The context of the high dimension

- Consider a sample of P-dimensional observations

$$\underbrace{\vec{\boldsymbol{x}}_{\mathbf{1}}, \ \cdots, \vec{\boldsymbol{x}}_{\boldsymbol{N}}}_{N-\text{sample}} \quad \text{with} \quad \vec{\boldsymbol{x}}_{\boldsymbol{i}} \in \mathbb{R}^P$$

- If the size of the data is of the same order as the sample size, i.e.

$$\boxed{P \propto N}$$

- Then standard (empirically-based) estimators **perform poorly** or **may even be false**

RMT offers a sound theoretical framework for ...

- Statistical analysis
- Performance analysis
- algorithm design
- practical insights (heuristics, etc.)

Example 3: Kernel-based Machine Learning in Large Dimensions

- RMT helps understanding **and improving** kernels, SVMs, neural nets, etc. **Even for real data!**

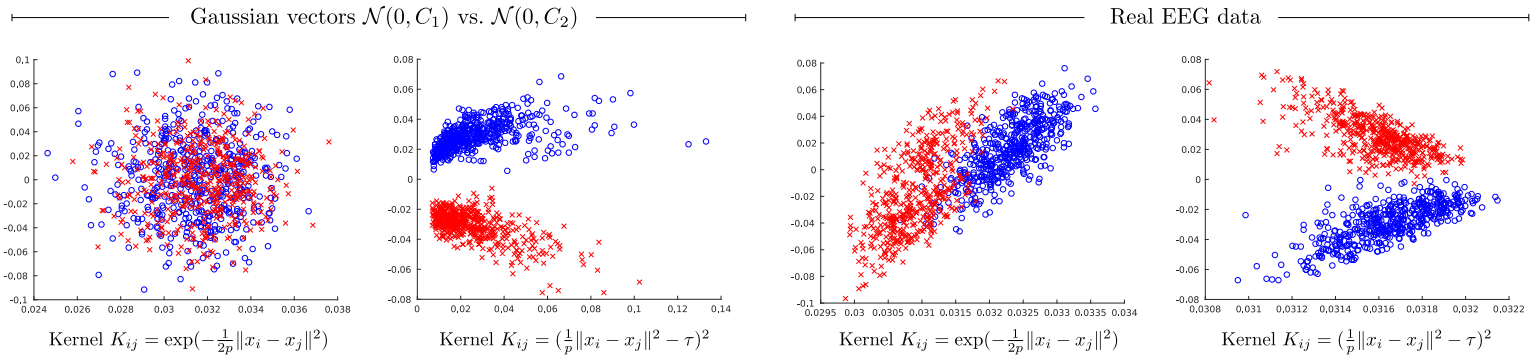

Figure: The Gaussian kernel (of constant use) is sometimes **not** good in large dimensions.

Example 4: Performance analysis for deep neural networks

- RMT can go way beyond... to understand **deep neural network behaviors**, here with GANs (Generative Adversarial Nets)

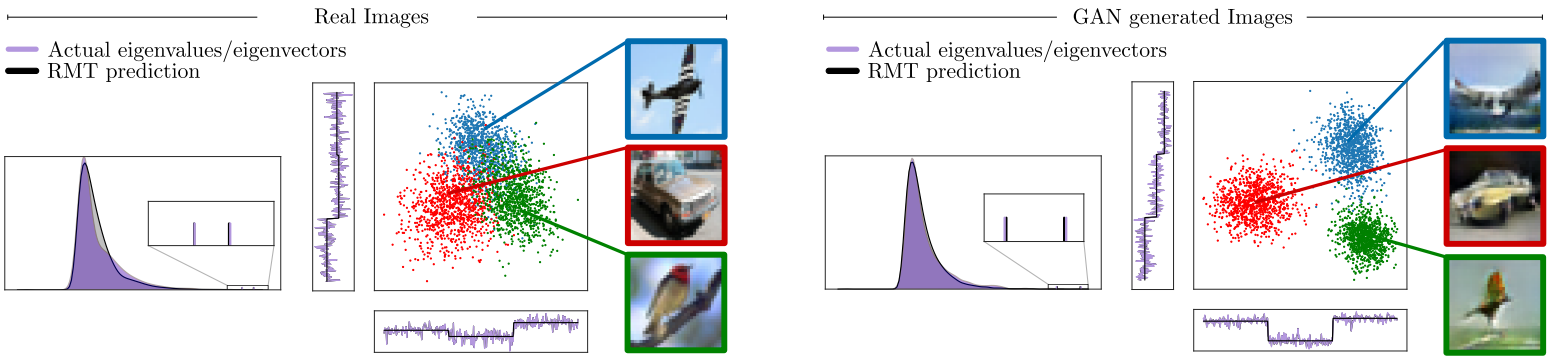

Figure: CNN (Convolutional Neural Network) clustering of real images (left) and GAN-generated images (right). RMT theoretically predicts the right. . . and works great on the left!

Random Matrix Theory (RMT) in an Nutshell

High dimensional statistics and ML applications

Outline of the course

To summarise

Random Matrices are versatile in modelling

- A random matrix (sparse, structured, with variance profile, dependent entries, etc.) is an ubiquitous object in mathematical modelization.

To summarise

Random Matrices are versatile in modelling

- A random matrix (sparse, structured, with variance profile, dependent entries, etc.) is an ubiquitous object in mathematical modelization.

Aim of the course

- Introduce **mathematical methods** to describe the eigenstructure of large random matrices.
- Specific emphasis on **high dimensional statistics and machine learning applications**.

To summarise

Random Matrices are versatile in modelling

- A random matrix (sparse, structured, with variance profile, dependent entries, etc.) is an ubiquitous object in mathematical modelization.

Aim of the course

- Introduce **mathematical methods** to describe the eigenstructure of large random matrices.
- Specific emphasis on **high dimensional statistics and machine learning applications**.

Outline of the course

- 8 lectures
- 2 lab works
- exam

To summarise

Random Matrices are versatile in modelling

- A random matrix (sparse, structured, with variance profile, dependent entries, etc.) is an ubiquitous object in mathematical modelization.

Aim of the course

- Introduce **mathematical methods** to describe the eigenstructure of large random matrices.
- Specific emphasis on **high dimensional statistics and machine learning applications**.

Outline of the course

- 8 lectures
- 2 lab works
- exam

Grading

- Final Exam, lab work (TP) $\times 2$, maybe homework (DM) $\times 1$

Internships

- Internships are available:
  - "A Random Matrix Analysis for Natural Language Processing" (Couillet)
  - "A Random Matrix Analysis of Neural Network Dynamics" (Couillet)
  - "A Random Matrix Analysis of Implicit Optimization in Machine Learning" (Couillet)
  - "A Random Matrix Approach to Theoretical Ecology" (Najim)
- **Possibility to customize the subject**: theoretical versus practical balance following your own taste!
- Possibility of PhD openings.
- Full description available on the MVA website.

Contact

- romain.couillet@gipsa-lab.grenoble-inp.fr
- jamal.najim@univ-mlv.fr