Spectrum Analysis of Large Random Matrices: A Brief Introduction [2]

Robert C. Qiu, Ph.D

Professor, *IEEE Fellow*, rqiu@ieee.org

May 9, 2017

[2]This presentation note is based on a 6-week short course for PhD students at Shanghai Jiao Tong University, from May to June, 2017.

Outline for Lecture 1

- Semi-Circle Distribution
- Marchenko-Pasture Distribution
- Single-Ring Distribution
- Random SVD

Semi-Circle Distribution—A MATLAB Experimental Demonstration

- MATLAB implementation: Z=*randn*(n,m);
- A complex *matrix*.
- $\mathbf{Z} \in \mathbb{C}^{n \times m}$
- $z \in \mathbb{C}$. A complex number (scalar).
- This **Z** is arbitrary. The only restriction is that the entries $Z_{ij}, i = 1, ..., n, j = 1, ..., m$ are Gaussian random variables.
- A matrix has its eigenvalues and eigenvectors, as a complex number has its amplitude and phase.
- What can we say about the eigenvalues?
- What happens if we allow $n \to \infty, m \to \infty$ but $\frac{m}{n} \to c \in (0, \infty)$?

Semi-Circle Distribution—A MATLAB Experimental Demonstration

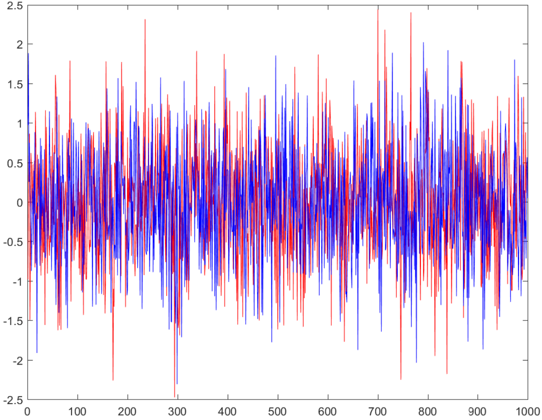


Figure: Comparing the random Gaussian vectors obtained from the first and second columns of $\mathbf{Z} \in \mathbb{C}^{n \times m}$. Z(:,1) and Z(:,2).

Semi-Circle Distribution—A MATLAB Experimental Demonstration

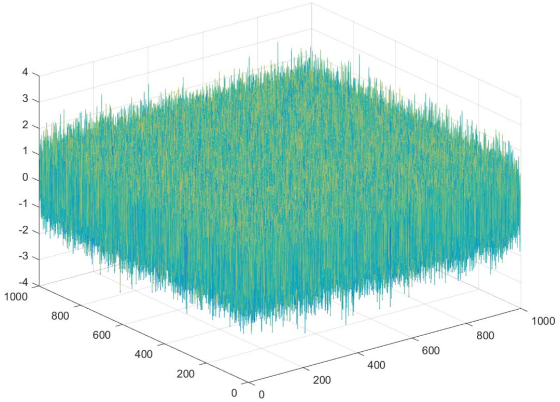
▶ 

Figure: A 3D view of a random Gaussian matrix $\mathbf{Z} \in \mathbb{C}^{n \times m}$.

Semi-Circle Distribution—A MATLAB Experimental Demonstration

- The eigenvalues of $\mathbf{Z} \in \mathbb{C}^{n \times m}$ are still too complicated!!! We will revisit this problem.
- Let us add more restrictions.
- For a complex number z, we study the real part $x = \frac{1}{2}\left(z + z*\right), \quad z \in \mathbb{C}$.
- In analogy, we study $\mathbf{X} = \frac{1}{2}\left(\mathbf{Z} + \mathbf{Z}^H\right), \quad \mathbf{Z} \in \mathbb{C}^{n \times m}$.

Semi-Circle Distribution—A MATLAB Experimental Demonstration

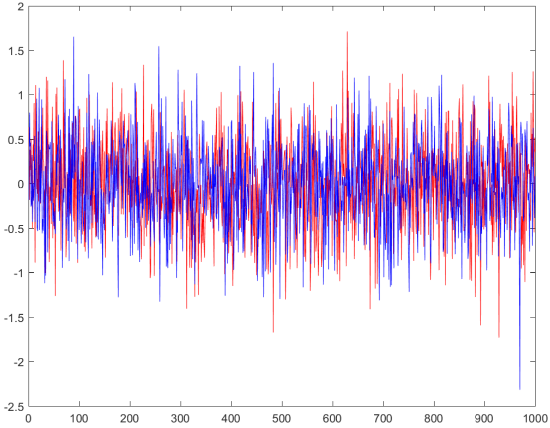


Figure: $\mathbf{X} = \frac{1}{2}\left(\mathbf{Z} + \mathbf{Z}^H\right), \quad \mathbf{Z} \in \mathbb{C}^{n \times m}$. First two columns.

Semi-Circle Distribution—A MATLAB Experimental Demonstration

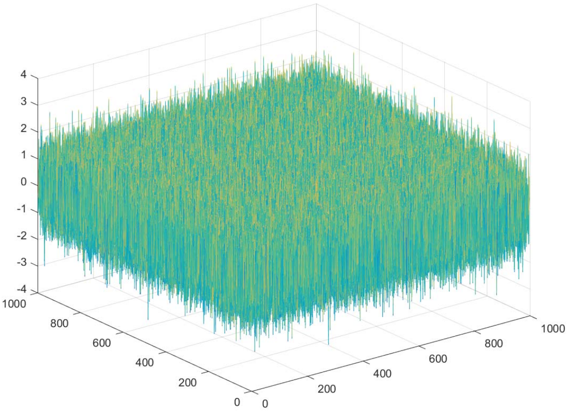


Figure: $\mathbf{X} = \frac{1}{2}\left(\mathbf{Z} + \mathbf{Z}^H\right), \quad \mathbf{Z} \in \mathbb{C}^{n \times m}$. 3D view.

Semi-Circle Distribution—A MATLAB Experimental Demonstration

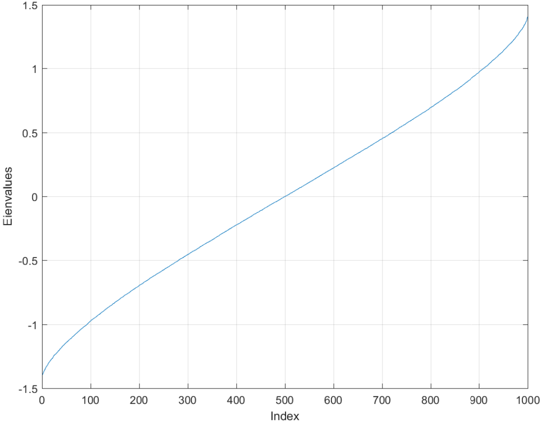


Figure: Eigenvalues of $\mathbf{X}$ are real, since it is a complex Hermitian matrix.

Semi-Circle Distribution—A MATLAB Experimental Demonstration

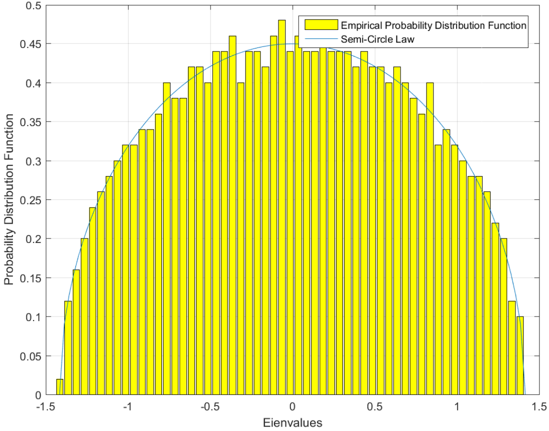


Figure: The probability distribution function of all the n eigenvalues follows a semi-circle distribution. Parameters: $n = 1000$, $m = 1000$.

Semi-Circle Distribution—A MATLAB Experimental Demonstration

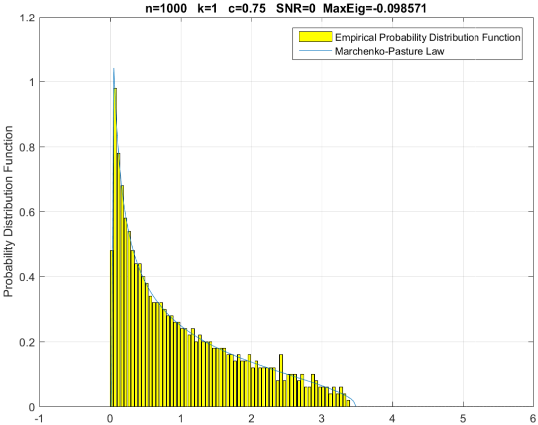
Figure: The probability distribution function of all the n eigenvalues follows a semi-circle distribution. Parameters: $n = 2000$, $m = 2000$.

Marchenk-Pasture Distribution—A MATLAB Experimental Demonstration

- $a = zz^*, \quad z \in \mathbb{C}.$
- $\mathbf{S} = \frac{1}{n}\mathbf{Z}\mathbf{Z}^H, \quad \mathbf{Z} \in \mathbb{C}^{n\times m}, \quad \mathbf{S} \in \mathbb{C}^{n\times n}.$
- As multiplication is much more complicated than addition, this problem, using matrix multication, is much harder!
-
$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_m \end{bmatrix}_{n\times m} \in \mathbb{C}^{n\times m}$$
$$\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_m \in \mathbb{C}^n$$
$$\mathbf{Z}\mathbf{Z}^H = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_m \end{bmatrix} \begin{bmatrix} \mathbf{z}_1^H \\ \mathbf{z}_2^H \\ \vdots \\ \mathbf{z}_m^H \end{bmatrix} = \sum_{i=1}^{m} \mathbf{z}_i\mathbf{z}_i^H$$
$$\mathbf{z}_i\mathbf{z}_i^H \in \mathbb{C}^{n\times n}, \text{ rank - one matrix}$$

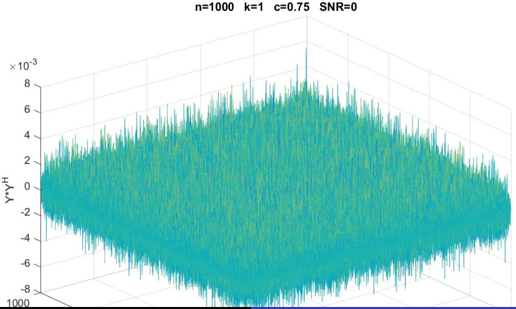
Marchenko-Pasture Distribution—A MATLAB Experimental Demonstration

- For $n \to \infty, m \to \infty$ but $\frac{m}{n} \to c \in (0, \infty)$, we have a convergent distribution, called Marchenko-Pasture Distribution.



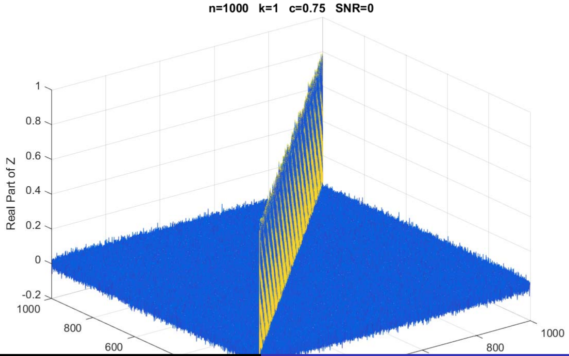
Marchenko-Pasture Distribution—A MATLAB Experimental Demonstration

- For random vectors, we certainly have the intuition.
- The rank-one matrix $\mathbf{z}_i\mathbf{z}_i^H$ is the building block for the whole sample covariance matrix.
- what is interesting lies in that the random vector is very long! Here the length is $n = 1000$. This kind of intuition is not common, before the popularity of big data.



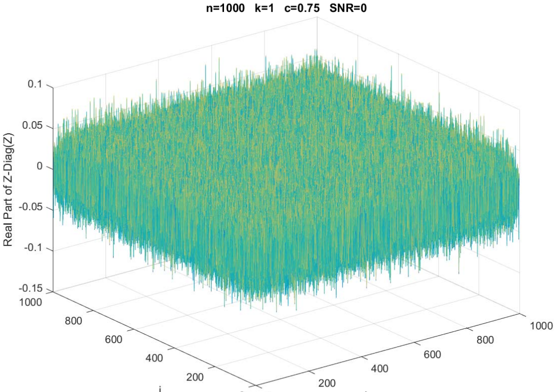
Marchenko-Pasture Distribution—A MATLAB Experimental Demonstration

- What we are interested in is the sum of the rank-one matrix.
- $\frac{1}{n}\mathbf{Z}\mathbf{Z}^H = \frac{1}{n}\sum_{i=1}^{m}\mathbf{z}_i\mathbf{z}_i^H \in \mathbb{C}^{n\times n}$, here $n = 1000, m = 750$.
- Diagonal entries are fundamentally different from the non-diagonal ones; they add coherently.



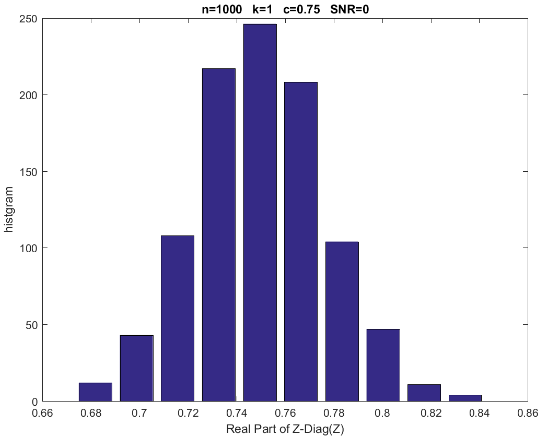
Marchenko-Pasture Distribution—A MATLAB Experimental Demonstration

- Removing the diagonal entries.
- All non-diagonal entries behave similarly. They are bounded in magnitude.



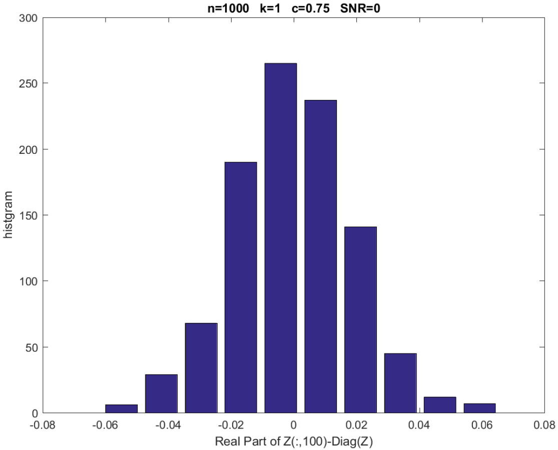
Marchenko-Pasture Distribution—A MATLAB Experimental Demonstration

- The distribution of diagognal entries is Gaussian.



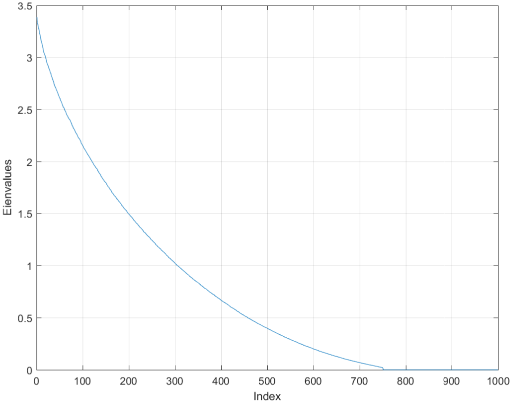
Marchenko-Pasture Distribution—A MATLAB Experimental Demonstration

- The non-diagonal entries are also Gaussian.



Marchenko-Pasture Distribution—A MATLAB Experimental Demonstration

- The eigenvalues vary from zero to a finite upper bound.



Single-Ring Distribution—A MATLAB Experimental Demonstration

- Let us get back the most general Gaussian matrix $\mathbf{Z} \in \mathbb{C}^{n \times m}$, where the entries $Z_{ij}, i = 1, ..., n, j = 1, ..., m$ are complex Gaussian random variables.
- What happens if

$$n \to \infty, m \to \infty \text{ but } \frac{m}{n} \to c \in (0, \infty)?$$

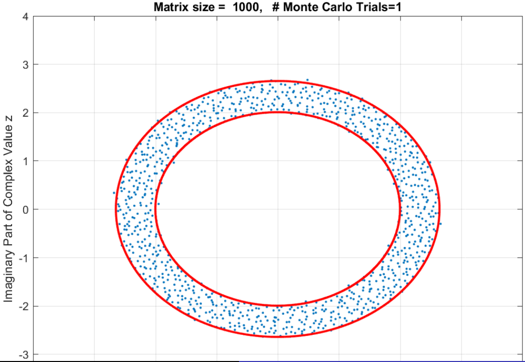
- The distribution of the complex eigenvalues converges to a *single ring* on the complex plane.

Single-Ring Distribution—A MATLAB Experimental Demonstration

- Consider

$$\frac{1}{n}\sqrt{\mathbf{Z}\mathbf{Z}^H}\mathbf{H},$$

where $\mathbf{H} \in \mathbb{C}^{n\times n}$ is a Haar matrix whose measure is a unit circle.

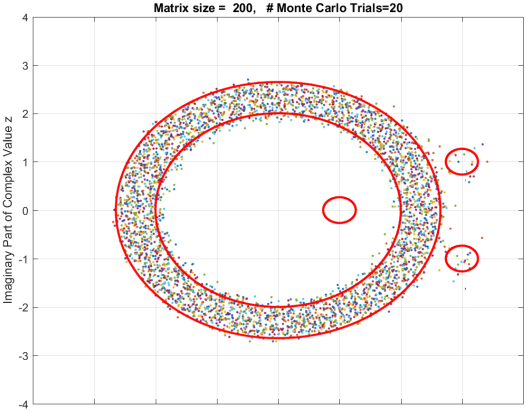


Single-Ring Distribution with Outliers—A MATLAB Experimental Demonstration

- Consider

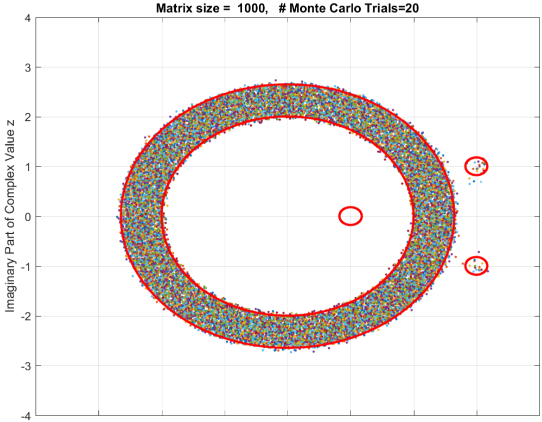
$$\frac{1}{n}\sqrt{\mathbf{Z}\mathbf{Z}^H}\mathbf{H} + \mathbf{A},$$

where $\mathbf{A}$ is a matrix of low rank.



Single-Ring Distribution with Outliers—A MATLAB Experimental Demonstration

- Larger matrix implies better resolution.
- Outlier detection. Anomaly detection. PCA.
- Power grid, wireless network, stocks market



Random SVD—A MATLAB Experimental Demonstration

- Consider

$$\mathbf{Z} = \mathbf{U} \begin{pmatrix} s_1 \cdots 0 \\ \vdots \ddots \vdots \\ 0 \cdots s_n \end{pmatrix} \mathbf{V},$$

  where **U** and **V** are Haar-distributed unitary random matrices and the s_i's are positive numbers that are independent of **U** and **V**.

- If **Z** is a random matrix with i.i.d. entries. The singular values s_i are compactly supported on $\mathbb{R}^+$. The Marchenko-Pastur quarter circular distribution is defined as

$$\mu_s(dx) = \frac{1}{\pi}\sqrt{4 - x^2}\mathbf{1}_{[0,2]}(x)\,dx,$$

  where $\mathbf{1}_{[c,d]}(x)$ is the indicator function that is one on the interval $[c, d]$ and zero outside the interval.

Random SVD—A MATLAB Experimental Demonstration

- The eigenvalue distribution converges, in probability, weakly to a deterministic probability measure whose support on the complex plane is a single ring defined as $\{z \in \mathbb{C} : a \leqslant |z| \leqslant b\}$.
- The inner radius a and outer radius b of the ring is

$$a = \left(\int_0^\infty x^{-2} d\mu_s(x) \right)^{-1/2}, \quad b = \left(\int_0^\infty x^2 d\mu_s(x) \right)^{1/2}$$

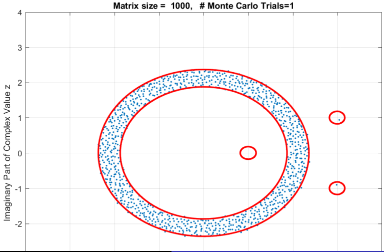
Random SVD—A MATLAB Experimental Demonstration

- Singular values have a uniform distribution on the interval $[\alpha, \beta]$, for $\beta > \alpha$.
- The inner radius a and outer radius b of the ring is

$$a = \left(\int_0^\infty x^{-2} d\mu_s(x) \right)^{-1/2} = \sqrt{\alpha\beta},$$

$$b = \left(\int_0^\infty x^{2} d\mu_s(x) \right)^{1/2} = \frac{1}{\sqrt{3}} \sqrt{\alpha^2 + \alpha\beta + \beta^2}.$$

- Example: $a = 1, b = 3.5$



Lecture 2

- Matrix Concentration Inequities for Sample Covariance Matrix
- Linear Eigenvalue Statistics
- Free Probability
-

Reference

- R. C. Qiu and P. Antonik, Smart Grid using Big Data Analytics: A Random Matrix Theory Approach, John Wiley, 2017.
- R. C. Qiu, Random Matrix Theory and Big Data Analysis, 252 pages, manuscript, May 1, 2017.