

Random Matrix Theory: A Numericist's Need-to-know

Common Ensembles and their Spectral Density

Mitchell Scott

Department of Mathematics, Emory University

February 27, 2026





Table of Contents

1 Introduction

2 RMT Distributions

Wigner Matrices

Wishart Matrices

Tracy - Widom Distribution

3 Analytical Tools

4 Applications

Dirichlet Random Matrices

Loss Surface Geometry





What is Random Matrix Theory (RMT)?

- Concerned with properties of matrices \mathbf{M} , where \mathbf{M}_{ij} are random variables
- The degree of independence and the manner in which the \mathbf{M}_{ij} are distributed determine the type of random matrix ensemble to which \mathbf{M} belongs.
- Originally used to model heavy metal nuclei, but has found uses in many disparate fields.
- Numerical Linear Algebraists and Random Matrix Theorists help each other out.
 - Alan Edelman (MIT) looks at numerically solving some of the difficult R - and S -transforms.
 - Michael Mahoney (UC Berkeley) connects RMT to machine learning models to understand the underlying structure.





Table of Contents

- 1 Introduction
- 2 RMT Distributions
 - Wigner Matrices
 - Wishart Matrices
 - Tracy - Widom Distribution
- 3 Analytical Tools
- 4 Applications
 - Dirichlet Random Matrices
 - Loss Surface Geometry





Wigner Ensemble and the Semicircle Law.

Definition (Wigner Ensemble)

The data for a *Wigner Ensemble* is a matrix of $N \times N$ data that are i.i.d.

Theorem (Wigner's Semicircle Law[1])

For a real Wigner matrix ensemble \mathbf{M} whose entries $\mathbf{M}_{ij} \sim \mathcal{N}(0, \sigma^2/2N)$, then

$$\rho_{\text{SC}}(x) = \begin{cases} \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2} & |x| \leq 2\sigma \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Remark

Data that is i.i.d. with some technical conditions (zero mean, finite moments, etc) also converge to ρ_{SC} [2].





Gaussian Ensembles

Remark

The most common random matrices are those with the Gaussian assumption, i.e., the elements are sampled from a variant of the standard normal distribution.

Example

- ($\beta = 1$) Gaussian Orthogonal Ensemble, then $\mathbf{X}_{i,j} \sim \mathcal{N}(0, 1)$,
- ($\beta = 2$) Gaussian Unitary Ensemble, then $\mathbf{X}_{i,j} \sim \mathcal{N}(0, 1/2) + i\mathcal{N}(0, 1/2)$,
- ($\beta = 4$) Gaussian Symplectic Ensemble, then $\mathbf{X}_{i,j} \sim \mathcal{N}(0, 1/4) + i\mathcal{N}(0, 1/4) + j\mathcal{N}(0, 1/4) + k\mathcal{N}(0, 1/4)$.

Let $\mathbf{M} = \frac{\sigma}{\sqrt{2N}} (\mathbf{X} + \mathbf{X}^*)$, then diagonalize \mathbf{M} to find the spectral properties.





Agreement of Wigner Matrices and Theory

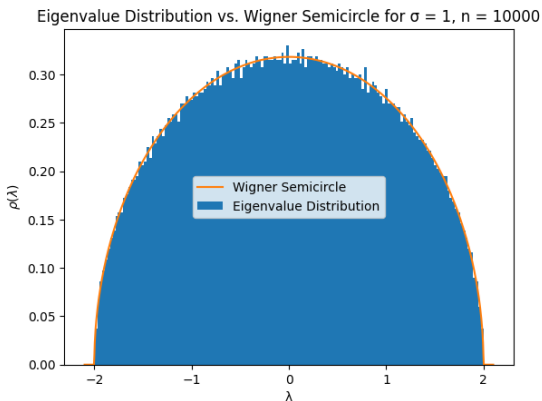


Figure 2: Random entries $\mathbf{M}_{ij} \sim \mathcal{N}(0, 1/2N)$, where $\mathbf{M} \in \mathbb{R}^{10,000 \times 10,000}$, numerically diagonalized and binned in 150 buckets.



Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$, where we have T observations and each observation contains N variables.



Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$, where we have T observations and each observation contains N variables.

Example

Wishart matrices can arise in many examples, such as:

- daily returns of N stocks over a certain time period,



Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$, where we have T observations and each observation contains N variables.

Example

Wishart matrices can arise in many examples, such as:

- daily returns of N stocks over a certain time period,
- number of spikes fired by N neurons during T consecutive intervals of Δt ,



Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$, where we have T observations and each observation contains N variables.

Example

Wishart matrices can arise in many examples, such as:

- daily returns of N stocks over a certain time period,
- number of spikes fired by N neurons during T consecutive intervals of Δt ,
- and so many more.



Sample Covariance Matrices

Definition (Sample Covariance Matrix)

Let $\mathbf{H} \in \mathbb{R}^{N \times T}$ with $\mathbf{H}_{it} = x_i^t$. The *sample covariances* of the data and the *sample covariance matrix* are given respectively by

$$\mathbf{E}_{ij} = \frac{1}{T} \sum_{t=1}^T x_i^t x_j^t, \quad \mathbf{E} = \frac{1}{T} \mathbf{H} \mathbf{H}^\top, \quad (2)$$

This results in an $N \times N$ matrix \mathbf{E} .

Theorem (Tao, Vu, 201[2])

For a matrix \mathbf{M} , whose entries \mathbf{M}_{ij} are i.i.d. random variables of mean zero, variance one, and having finite C_0^{th} moment for some sufficiently large constant C_0 , then the covariance matrix converges weakly to the Marčenko-Pastur distribution.





Convergence of Wishart Matrices' Spectrum

Remark

For this discussion, we will sample from the Gaussian β ensembles as discussed above.

Theorem (Marčenko-Pastur [3])

The full Marčenko-Pastur distribution can be written as such, let $\mathbf{M} \in \mathbb{R}^{N \times T}$, where $N, T \rightarrow \infty, N/T \rightarrow q \in (0, \infty)$. Now let $\lambda_{\pm} = \sigma^2 (1 \pm \sqrt{q})^2$. Then the density of the eigenvalues of $\mathbf{M}\mathbf{M}^T / T$ converges weakly to

$$\rho_{MP}(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{[(\lambda_+ - x)(x - \lambda_-)]_+}}{2\pi qx} \quad (3)$$

where $[a]_+ := \max\{a, 0\}$.



Convergence of Wishart Matrices' Spectrum

Remark

For this discussion, we will sample from the Gaussian β ensembles as discussed above.

Theorem (Marčenko-Pastur [3])

The full Marčenko-Pastur distribution can be written as such, let $\mathbf{M} \in \mathbb{R}^{N \times T}$, where $N, T \rightarrow \infty, N/T \rightarrow q \in (0, \infty)$. Now let $\lambda_{\pm} = \sigma^2 (1 \pm \sqrt{q})^2$. Then the density of the eigenvalues of $\mathbf{M}\mathbf{M}^T / T$ converges weakly to

$$\rho_{MP}(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{[(\lambda_+ - x)(x - \lambda_-)]_+}}{2\pi qx} + \left[\frac{q-1}{q} \right]_+ \delta(x) \quad (3)$$

where $[a]_+ := \max\{a, 0\}$.



Visualizing the Marčenko-Pastur Distribution

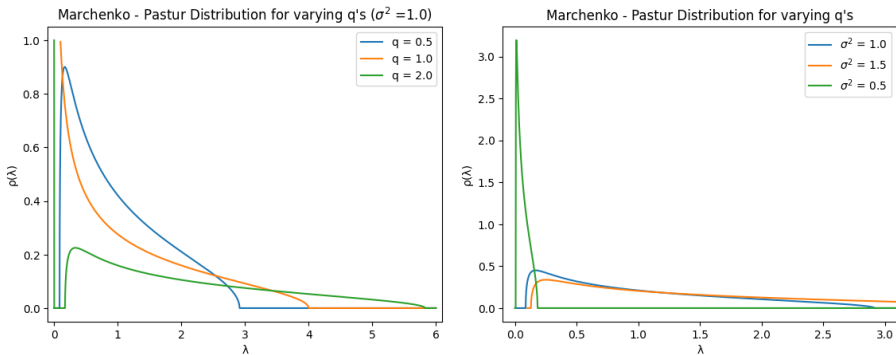


Figure 3: **Left:** The Marčenko-Pastur distribution for fixed variance ($\sigma^2 = 1$) with varying $q = N/T = m/n$, with one less than, equal to, and greater than 1. **Right:** The Marčenko-Pastur distribution for fixed aspect ratio ($q = 0.5$) with varying variances.





Agreement of Wishart Matrices and Theory

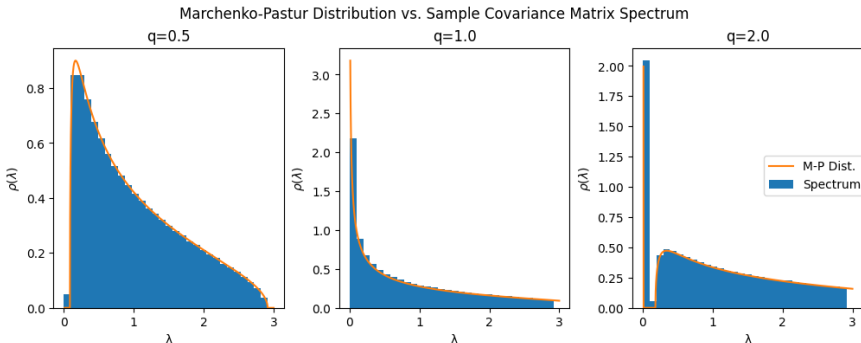


Figure 4: Random entries $\mathbf{M}_{ij} \sim \mathcal{N}(0, 1)$ where $\mathbf{M} \in \mathbb{R}^{10,000 \times T}$, where $T = [20,000, 10,000, 5,000]$, so $q = N/T = [0.5, 1, 2]$. Then $\mathbf{W} = \mathbf{M}\mathbf{M}^T/T$ numerically diagonalized and binned into 30 buckets.





Tracy-Widom

Definition

The rescaled distribution of $\lambda_{\max} - \lambda_+$ converges for $N \rightarrow \infty$ towards the Tracy-Widom Distribution F_β :

$$\mathbb{P}\left(\lambda_{\max} \leq \lambda_+ + \gamma N^{-2/3} u\right) = F_1(u)$$

$$\mathbb{P}\left(\lambda_{\max} \leq \sqrt{4N} + N^{-1/6} u\right) = F_2(u),$$

where γ is a problem dependent problem.

Example ($\beta = 1$ case)

- For Wigner problems $\lambda_+ = 2, \gamma = 1$.
- For Wishart problems, $\lambda_+ = (1 + \sqrt{q})^2, \gamma = \sqrt{q} \lambda_+^{2/3}$.



Remark (Asymptotics)

Let $f_1(u) = F'(u)$, then

$$\ln f_1(u) = \begin{cases} -2\beta/3|u|^{3/2} & u \rightarrow \infty \\ -\beta/24|u|^3 & u \rightarrow -\infty. \end{cases}$$

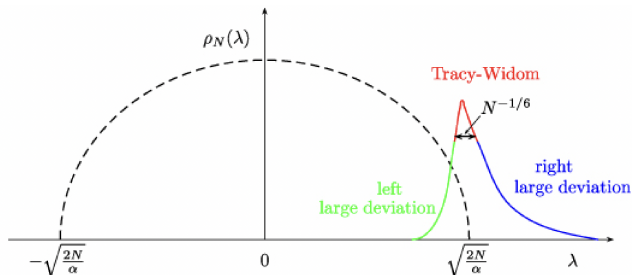


Figure 5: Depiction of the Wigner's Semicircle distribution overlaid by Tracy-Widom fluctuations on the edge.



Table of Contents

- 1 Introduction
- 2 RMT Distributions
 - Wigner Matrices
 - Wishart Matrices
 - Tracy - Widom Distribution
- 3 Analytical Tools
- 4 Applications
 - Dirichlet Random Matrices
 - Loss Surface Geometry



Definition

The *normalized trace operator* $\tau(\mathbf{A}) := \frac{1}{N} \mathbb{E} [\text{tr} \mathbf{A}]$ is the correct way to think of the k^{th} moments m_k of a random matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ as $N \rightarrow \infty$.

Definition (Resolvent, Stieltjes transform)

Given an $N \times N$ real symmetric matrix \mathbf{A} , its *resolvent* is given by

$$\mathbf{G}_{\mathbf{A}}(z) = (z\mathbf{I} - \mathbf{A})^{-1},$$

where z is a complex variable defined away from all the real eigenvalues of \mathbf{A} . Then the *Stieltjes transform* of \mathbf{A} is given by

$$g_N^{\mathbf{A}}(z) = \frac{1}{N} \text{tr} (\mathbf{G}_{\mathbf{A}}(z)) = \frac{1}{N} \sum_{k=1}^N \frac{1}{z - \lambda_k}.$$

For large enough z , $g(z) := \lim_{N \rightarrow \infty} \mathbb{E} [g_N(z)]$



Definition (Stieltjes Transform, revisited)

For $z \in \mathbb{C} \setminus \mathbb{R}$, we can also write the Stieltjes Transform as:

$$g_N^{\mathbf{A}}(z) = \int_{-\infty}^{+\infty} \frac{\rho_N^{\mathbf{A}}(\lambda)}{z - \lambda} d\lambda$$

Additionally,

$$g_N^{\mathbf{A}}(z) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \frac{1}{N} \operatorname{tr}(\mathbf{A}^k), \quad \text{as } N \rightarrow \infty \quad g(z) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \tau(\mathbf{A}^k).$$

Definition (R -transform)

The R -transform is defined as

$$R(g(z)) := \mathfrak{z}(g) - \frac{1}{g}$$

where $\mathfrak{z}(g) = z$ is the inverse Stieltjes transform.

Recipe to add two (free) random matrices

Example

Let $\mathbf{C} = \mathbf{A} + \mathbf{B}$, where \mathbf{A}, \mathbf{B} are independent (free), at least one is rotationally invariant.

- 1 Find $\mathfrak{g}_{\mathbf{B}}(z)$ and $\mathfrak{g}_{\mathbf{A}}(z)$
- 2 Invert $\mathfrak{g}_{\mathbf{B}}(z)$ and $\mathfrak{g}_{\mathbf{A}}(z)$ to get $\mathfrak{z}_{\mathbf{B}}(g)$ and $\mathfrak{z}_{\mathbf{A}}(g)$, hence $R_{\mathbf{B}}(x)$ and $R_{\mathbf{A}}(x)$.
- 3 $R_{\mathbf{C}(x)} = R_{\mathbf{B}}(x) + R_{\mathbf{A}}(x)$, which gives $\mathfrak{z}_{\mathbf{C}}(g) = R_{\mathbf{C}}(g) - g^{-1}$.
- 4 Solve $\mathfrak{z}_{\mathbf{C}}(g) = z$ for $\mathfrak{g}_{\mathbf{C}}(z)$.
- 5 Use Sokhotski-Plemelj formula to find the density:

$$\rho_{\mathbf{C}}(\lambda) = \frac{\lim_{\eta \rightarrow 0^+} \operatorname{Im} \mathfrak{g}_{\mathbf{C}}(\lambda - i\eta)}{\pi}$$

Wishart + Wigner Example

Example

Let $\mathbf{X} = \mathbf{W} + \mathbf{G}$, where \mathbf{W} be a Wishart matrix with $\sigma^2 = 1$, $\phi = q = N/T$, and \mathbf{G} be a Wigner matrix with variance $\sigma^2 = 2\epsilon$. We are interested in the spectrum of \mathbf{X} . The Stieltjes Transform can be found by integrating over the measure for \mathbf{G} :

$$g_{\mathbf{G}}(z) = \int_{\text{supp}\{\rho\}} \frac{\rho_{\text{SC}}(x)}{z-x} dx, \quad \text{OR} \quad \frac{1}{g_{\mathbf{G}}(z)} = z - g_{\mathbf{G}}(z)$$

Then the R -transform can be found as $R(z) = 2\epsilon z$. Additionally, you can carry this out for the Wishart matrix \mathbf{W} as well.

$$g_{\mathbf{W}}(z) = \int_{\text{supp}\{\rho\}} \frac{\rho_{\text{MP}}(x)}{z-x} dx, \quad \text{OR} \quad g_{\mathbf{W}}(z) = \frac{z - (1 - \phi) - \sqrt{z - \lambda_+} \sqrt{z - \lambda_-}}{2\phi z}.$$

This means that the R -transform for the Wishart matrix \mathbf{W} is $R_{\mathbf{W}}(z) = 1/(1 - \phi z)$.



Example

Now $R_{\mathbf{X}}(z) = R_{\mathbf{W}}(z) + R_{\mathbf{G}}(z) = 1/(1 - z\phi) + 2\epsilon z$. Now we find the Stieltjes transform of \mathbf{X} is

$$2\epsilon\phi g_{\mathbf{X}}(z)^3 - (2\epsilon + z\phi)g_{\mathbf{X}}(z)^2 + (z - \phi - 1)g_{\mathbf{X}}(z) - 1 = 0.$$

Lastly, we find the root which $\sim 1/z$ as $z \rightarrow \infty$, and use the Sokhotski-Plemelj formula to find the spectral density.

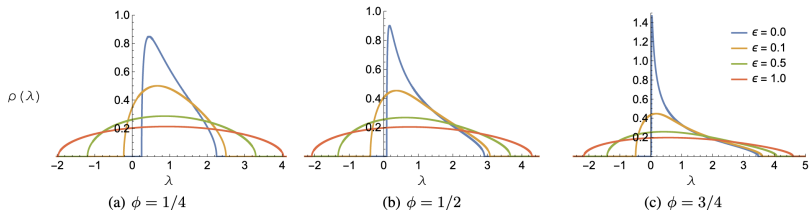


Figure 7: Spectral distributions of $\mathbf{X} = \mathbf{W} + \mathbf{G}$ with different $\phi = N/T$ ratios and different variances.



Table of Contents

- 1 Introduction
- 2 RMT Distributions
 - Wigner Matrices
 - Wishart Matrices
 - Tracy - Widom Distribution
- 3 Analytical Tools
- 4 Applications
 - Dirichlet Random Matrices
 - Loss Surface Geometry





Dirichlet Distribution

Definition (Dirichlet Distribution PDF)

Also known as the multivariate Beta distribution, the *Dirichlet Distribution PDF* of order $K \geq 2$ with parameters $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^\top$ is given by

$$f(\mathbf{x}; \boldsymbol{\beta}) = \frac{1}{B(\boldsymbol{\beta})} \prod_{i=1}^K x_i^{\beta_i - 1}, \quad \text{where} \quad (4)$$

$$B(\boldsymbol{\beta}) = \frac{\prod_{i=1}^K \Gamma(\beta_i)}{\Gamma\left(\sum_{i=1}^K \beta_i\right)}, \quad (5)$$

and $x_i \in [0, 1], \forall i \in \{1, 2, \dots, K\}$, subject to $\sum_{i=1}^K x_i = 1$.



Sampling from Dirichlet Distribution

Example

While sampling from the Beta distribution is available in `numpy`, Dirichlet distributions are not. However, we do have access to the Gamma distribution. This means to sample a random vector $\mathbf{x} = [x_1, x_2, \dots, x_K]^\top$ from the K -dimensional Dirichlet distribution with parameters $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_K]^\top$, we draw K independent samples from Gamma distribution and normalize to sum to 1.

$$\text{Gamma}(\beta_i, 1) = \frac{y_i^{\beta_i-1} e^{-y_i}}{\Gamma(\beta_i)} \quad (6)$$

$$x_i = \frac{y_i}{\sum_{j=1}^K y_j} \quad (7)$$



Visualizing these Vectors

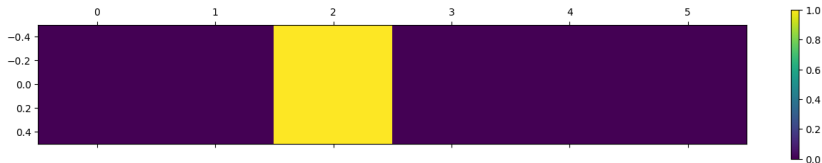


Figure 8: Small values of $\beta = [0.0001, 0.0001]^T$ promote sparsity.

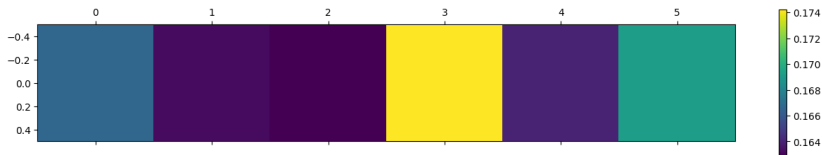


Figure 9: Large values of $\beta = [1000, 1000]^T$ promote uniformity ($1/N$).



Theoretical properties of Dirichlet Random Matrices

Conjecture

The spectrum of the Dirichlet random covariance matrix, DRCM, is of Marčenko-Pastur type in the bulk with Tracy-Widom fluctuations on the upper edge.



Numerical Results in the Bulk

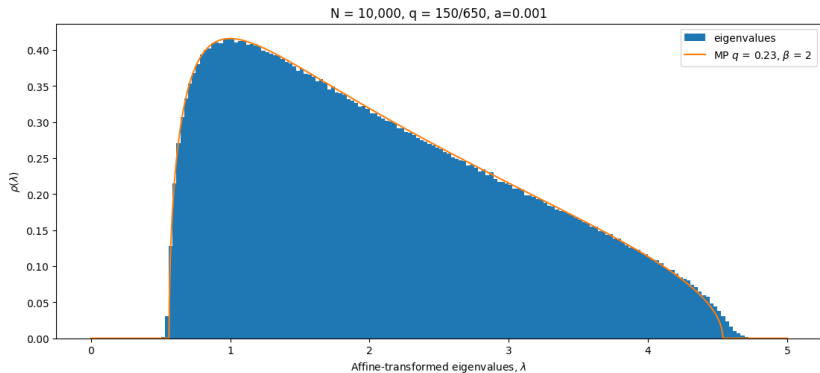


Figure 10: DRCM's spectrum converge to Marčenko-Pastur with hand tuned σ .



Numerical Results on the Edge

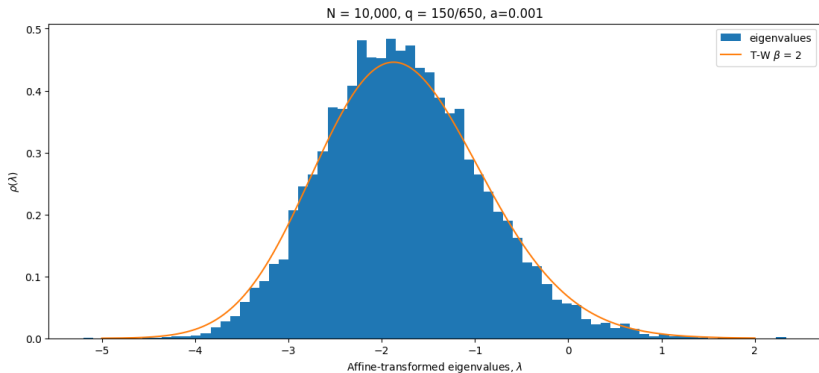


Figure 11: DRCM appears to have a Tracy-Widom like decay on the edge.



Number of Negative eigenvalues

Remark

The number of negative eigenvalues of the Hessian measures the number of descent directions.

Definition (Index)

The fraction of the number of negative eigenvalues, α , is the *index* of a critical point.

$$\alpha(\sigma, q) := \int_{-\infty}^0 \rho(\lambda; \sigma^2, q) \, d\lambda = 1 - \int_0^{\infty} \rho(\lambda; \sigma^2, q) \, d\lambda$$

Remark

The Wigner + Wishart matrix example was actually used in [4] to approximate the Hessian as $\mathbf{H} = \mathbf{J}\mathbf{J}^\top / T + \mathbf{R}$, or the sum of a GGN and random residual noise.

The S-transform and recipe for multiplication of random matrices

Example

Let $\mathbf{C} = \mathbf{A}^{1/2} \mathbf{B} \mathbf{A}^{1/2}$.

- 1 Find $t_{\mathbf{B}}(\zeta)$ and $t_{\mathbf{A}}(\zeta)$.
- 2 Invert $t_{\mathbf{B}}(\zeta)$ and $t_{\mathbf{A}}(\zeta)$ to get $\zeta_{\mathbf{B}}(t)$ and $\zeta_{\mathbf{A}}(t)$, and hence $S_{\mathbf{B}}(t)$ and $S_{\mathbf{A}}(t)$.
- 3 $S_{\mathbf{C}}(t) = S_{\mathbf{B}}(t) S_{\mathbf{A}}(t)$, which gives $\zeta_{\mathbf{C}}(t) S_{\mathbf{C}}(t) t = t + 1$
- 4 Solve $\zeta_{\mathbf{C}}(t) = \zeta$ for $t_{\mathbf{C}}(\zeta)$.
- 5 Use Sokhotski-Plemelj formula for $g_{\mathbf{C}}(z) = (t_{\mathbf{C}}(z) + 1)/z$ is equivalent to:

$$\rho_{\mathbf{C}}(\lambda) = \frac{\lim_{\eta \rightarrow 0^+} \operatorname{Im} t_{\mathbf{C}}(\lambda - i\eta)}{\pi \lambda}$$

Remark

Useful for comparing the spectrum of \mathbf{B} and \mathbf{C} , the split preconditioned Hessian.



Smallest positive eigenvalue

Remark

The “spectral gap”, which is the distance between the smallest positive eigenvalue and zero, can be well approximated by the Tracy-Widom distribution.

Theorem (S., et al., 2025 (informally))

The convergence of preconditioned SGD in a locally convex basin, converges to the local minimizer with high probability. This probability depends on the largest eigenvalue of the preconditioned Hessian, and the μ_{PL} constant. Additionally, knowing this μ_{PL} informs the rate of convergence, the noise floor, and the radius of the basin.

Remark

This could be a useful metric to see if your preconditioner will be effective a priori.

Questions?

Thank You!



References

- [1] Eugene P. Wigner. “Characteristic Vectors of Bordered Matrices With Infinite Dimensions”. In: *Annals of Mathematics* 62.3 (1955). Publisher: [Annals of Mathematics, Trustees of Princeton University on Behalf of the Annals of Mathematics, Mathematics Department, Princeton University], pp. 548–564. ISSN: 0003-486X. DOI: 10.2307/1970079. URL: <https://www.jstor.org/stable/1970079> (visited on 11/06/2025).
- [2] Terence Tao and Van Vu. “Random covariance matrices: Universality of local statistics of eigenvalues”. In: *The Annals of Probability* 40.3 (May 2012). arXiv:0912.0966 [math]. ISSN: 0091-1798. DOI: 10.1214/11-AOP648. URL: <http://arxiv.org/abs/0912.0966> (visited on 10/26/2025).
- [3] V A Marčenko and L A Pastur. “DISTRIBUTION OF EIGENVALUES FOR SOME SETS OF RANDOM MATRICES”. en. In: *Mathematics of the USSR-Sbornik* 1.4 (Apr. 1967), pp. 457–483. ISSN: 0025-5734. DOI: 10.1070/SM1967v001n04ABEH001994. URL: <https://www.mathnet.ru/eng/sm4101> (visited on 11/06/2025).





References (cont.)

- [4] Jeffrey Pennington and Yasaman Bahri. “Geometry of neural network loss surfaces via random matrix theory”. In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. ICML’17. Sydney, NSW, Australia: JMLR.org, Aug. 2017, pp. 2798–2806. (Visited on 11/06/2025).





Table of Contents

5 Proof

Proof of Wigner's Semicircle

Proof of Marčenko-Pastur's Distribution





Proof of Wigner's Semicircle law

Proof.

We proceed by finding a relationship between the Stieltjes transform of a Wigner matrix of size N and one of size $N - 1$ (cavity method). Find the 1,1 element of inverse of $\mathbf{M} = z\mathbf{I} - \mathbf{W}$.

$$\frac{1}{(\mathbf{G}_W)_{11}} = \mathbf{M}_{11} - \sum_{k,\ell=2}^N \mathbf{M}_{1k} (\mathbf{M}_{22})_{k\ell}^{-1} \mathbf{M}_{\ell 1}. \quad (8)$$

Observe that $\mathbb{E}[\mathbf{M}_{11}] = z$, and entries in \mathbf{M}_{22} are independent of $\mathbf{M}_{1i} = -\mathbf{W}_{1i}$, so

$$\begin{aligned} \mathbb{E}_{\{\mathbf{w}_{1i}\}} \left[\mathbf{M}_{1i} (\mathbf{M}_{22})_{ij}^{-1} \mathbf{M}_{1j} \right] &= \frac{\sigma^2}{N} (\mathbf{M}_{22})_{ii}^{-1} \delta_{ij} \\ \mathbb{E}_{\{\mathbf{w}_{1i}\}} \left[\sum_{k,\ell=2}^N \mathbf{M}_{1k} (\mathbf{M}_{22})_{k\ell}^{-1} \mathbf{M}_{\ell 1} \right] &= \frac{\sigma^2}{N} \text{tr} \left(\mathbf{M}_{22}^{-1} \right). \end{aligned}$$

Recall from earlier that $1/(N - 1) \text{tr} \left(\mathbf{M}_{22}^{-1} \right)$ is the Stieltjes transform of an $N - 1$ matrix with variance $\sigma^2(N - 1)/N$.



Proof of Wigner's Semicircle law (cont.)

Proof.

As $N \rightarrow \infty$, the difference is negligible, so

$$\mathbb{E} \left[\frac{1}{N} \operatorname{tr} (\mathbf{M}_{22}^{-1}) \right] \rightarrow \mathfrak{g}(z).$$

This means that we have a deterministic limit (negligible fluctuations) so

$$\mathbb{E} \left[\frac{1}{(\mathbf{G}_w)_{11}} \right] = \frac{1}{\mathbb{E}[(\mathbf{G}_w)_{11}]}.$$

Since \mathbf{W} is rotationally invariant, so is \mathbf{G}_w , which means all diagonal entries of \mathbf{G}_w must have the same expectation value:

$$\mathbb{E}[(\mathbf{G}_w)_{11}] = \frac{1}{N} \mathbb{E}[\operatorname{tr}(\mathbf{G}_w)] = \mathbb{E}[g_N] \rightarrow \mathfrak{g}.$$

Proof of Wigner's Semicircle law (cont.)

Proof.

Putting all our observations into Eq. 8, we have

$$\frac{1}{g(z)} = z - \sigma^2 g(z) \implies g = \frac{z \pm z\sqrt{1 - 4\sigma^2/z^2}}{2\sigma^2}.$$

Since $g(z)$ has to be analytic for large complex z , so the “+” sign gives $g(z) \sim z/\sigma^2$, which isn't analytic, but the “-” sign gives $g(z) \sim 1/z$ for any large, complex z , so

$$g(z) = \frac{z - z\sqrt{1 - 4\sigma^2/z^2}}{2\sigma^2}.$$

Now using the Sokhotski-Plemelj formula: $\lim_{\eta \rightarrow 0^+} \text{Im } g(z - i\eta) = \pi\rho(x)$, $g(z)$ can only have an imaginary part if $\sqrt{x^2 - 4\sigma^2}$ is imaginary. This lets us arrive at

$$\rho_{\text{SC}}(x) = \frac{1}{\pi} \lim_{\eta \rightarrow 0^+} \text{Im } g(z - i\eta) = \frac{\sqrt{4\sigma^2 - x^2}}{2\pi\sigma^2}, \quad |x| \leq 2\sigma \quad (9)$$





Proof of Marčenko-Pastur's law

Proof.

We derive a self-consistent equation satisfied by the Wishart Stieltjes transform:

$$g_{\mathbf{W}}(z) = \tau(\mathbf{G}_{\mathbf{W}}(z)), \quad \mathbf{G}_{\mathbf{W}}(z) := (\mathbf{I} - \mathbf{W})^{-1}.$$

Using the Schur complement formula, we once again have

$$\frac{1}{(\mathbf{G}(z))_{11}} = \mathbf{M}_{11} - \mathbf{M}_{12} \mathbf{M}_{22}^{-1} \mathbf{M}_{21}, \quad \mathbf{M} := z\mathbf{I} - \mathbf{W}.$$

Since $\mathbf{W} = \mathbf{H}\mathbf{H}^T/T$, and \mathbf{M}_{22} , \mathbf{H}_{jt} ($j \geq 2$), \mathbf{H}_{ks} ($k \geq 2$) are independent of the \mathbf{H}_{it} , $\forall t$, so

$$\begin{aligned} \frac{1}{(\mathbf{G}(z))_{11}} &= z - \mathbf{W}_{11} - \frac{1}{T^2} \sum_{t,s=1}^T \sum_{j,k=2}^N \mathbf{H}_{1t} \mathbf{H}_{jt} (\mathbf{M}_{22})_{jk}^{-1} \mathbf{H}_{ks} \mathbf{H}_{1s} \\ &= z - \mathbf{W}_{11} - \frac{1}{T} \sum_{t,s=1}^T \Omega_{ts} \mathbf{H}_{1s} \quad \text{with} \quad \Omega_{ts} := \frac{1}{T} \sum_{j,k=2}^N \mathbf{H}_{jt} (\mathbf{M}_{22})_{jk}^{-1} \mathbf{H}_{ks}. \end{aligned}$$



Proof of Marčenko-Pastur's law (cont.)

Proof.

Since $\gamma^2 := 1/T \operatorname{tr} \Omega^2$ converges to a finite limit when $T \rightarrow \infty$, we see that the above sum converges to $1/T \operatorname{tr} \Omega$ with $\mathcal{O}(\gamma 1/\sqrt{T})$ fluctuations. So

$$\begin{aligned} \frac{1}{(\mathbf{G}(z))_{11}} &= z - \mathbf{W}_{11} - \frac{1}{T} \sum_{2 \leq j, k \leq N} \frac{\sum_t \mathbf{H}_{kt} \mathbf{H}_{jt}}{T} (\mathbf{M}_{22})_{jk}^{-1} + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right) \\ &= z - \mathbf{W}_{11} - \frac{1}{T} \sum_{2 \leq j, k \leq N} \mathbf{W}_{kj} (\mathbf{M}_{22})_{jk}^{-1} + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right) \\ &= z - 1 - \frac{1}{T} \operatorname{tr}(\mathbf{W}_2 \mathbf{G}_2(z)) + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right), \end{aligned}$$

since $\mathbf{W}_{11} = 1 + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right)$, and $\mathbf{W}_2, \mathbf{G}_2(z)$ are the $N - 1$ variable SCM and resolvent, respectively.



Proof of Marčenko-Pastur's law (cont.)

Proof.

Rewriting the trace term,

$$\begin{aligned}
 \operatorname{tr}(\mathbf{W}_2 \mathbf{G}_2(z)) &= \operatorname{tr}(\mathbf{W}_2 (z\mathbf{I} - \mathbf{W}_2)^{-1}) \\
 &= -\operatorname{tr}(\mathbf{I}) + z \operatorname{tr}((z\mathbf{I} - \mathbf{W})^{-1}) \\
 &= -\operatorname{tr}(\mathbf{I}) + z \operatorname{tr}(\mathbf{G}_2(z)).
 \end{aligned}$$

Wherever $\tau(\mathbf{G}(z))$ converges, so too must $\tau(\mathbf{G}_2(z))$, where both converge to $g(z)$, so in this region of convergence, we have

$$\frac{1}{(\mathbf{G}(z))_{11}} = z - 1 + q - qzg(z) + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right),$$

where $q = N/T = \mathcal{O}(1)$ and $N^{-1/2}$ and $T^{-1/2}$ are of the same magnitude.

Proof of Marčenko-Pastur's law (cont.)

Proof.

Also observe that $1/(\mathbf{G}(z))_{11}$ has negligible fluctuations so

$$\frac{1}{(\mathbf{G}(z))_{11}} = \mathbb{E} \left[\frac{1}{(\mathbf{G}(z))_{11}} \right] + \mathcal{O} \left(\frac{1}{\sqrt{N}} \right) = \mathbb{E} \left[\frac{1}{(\mathbf{G}(z))_{11}} \right] + \mathcal{O} \left(\frac{1}{\sqrt{N}} \right)$$

By rotational invariance of \mathbf{W} , we have

$$\mathbb{E} [(\mathbf{G}(z))_{11}] = \frac{1}{N} \mathbb{E} [\text{tr } \mathbf{G}(z)] \rightarrow \mathfrak{g}(z).$$

In the large N limit, we have the self-consistent equation,

$$\frac{1}{\mathfrak{g}(z)} = z - 1 + q - qz\mathfrak{g}(z) \quad \implies \quad \mathfrak{g}(z) = \frac{z + q - 1 \pm \sqrt{(z + q - 1)^2 - 4qz}}{2qz}.$$

The roots of the argument of the square root are $\lambda_{\pm} = (1 \pm \sqrt{q})^2$.



Proof of Marčenko-Pastur's law (cont.)

Proof.

We choose the correct branch to see

$$g(z) = \frac{z - (1 - q) - \sqrt{z - \lambda_+} \sqrt{z - \lambda_-}}{2qz}$$

has the correct analytical properties. For $z = x - i\eta$, where $x \neq 0, \eta \rightarrow 0$, $g(z)$ can only have an imaginary part if $\sqrt{(x - \lambda_+)(x - \lambda_-)}$ is imaginary. Lastly,

$$\rho(x) = \frac{1}{\pi} \lim_{\eta \rightarrow 0^+} \operatorname{Im} g(x - i\eta) = \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi qx}, \quad \lambda_- < x < \lambda_+.$$

Studying g as $z \rightarrow 0$, we see there is a pole at 0 when $q > 1$, which gives a delta mass as $z \rightarrow 0$. These correspond to the $N - T$ trivial zero eigenvalues of \mathbf{E} if $N > T$. \square

