The Dyson Equation for Random Matrices

Olivia Choi, Elliott Kobelansky (Mentor: Kevin Xiao)

The Matrix Dyson Equation, developed in recent years, provides a unified framework for studying the spectral properties of large random matrices. We introduce core tools of random matrix theory: the empirical spectral distribution, the Stieltjes transform, and the resolvent, and use them to derive the simplest instance of the Matrix Dyson Equation for Wigner matrices, recovering the Wigner semicircle law.

1 Introduction: Random Matrix Theory

Random matrices were first studied in the 1950s by Eugene Wigner to model the energy levels of heavy atomic nuclei. Over time, the theory expanded into pure mathematics, revealing connections to probability, combinatorics, and number theory, including connections to the zeros of the Riemann Zeta function. Today, random matrix theory continues to develop, with new applications appearing in physics, statistics, optimization, and machine learning.

Historically, random matrix theory has focused on understanding the asymptotic behavior of the eigenvalues of random matrices. While classical probability theory has well-established universal results such as the Law of Large Numbers and the Central Limit theorem, which apply to random scalars and vectors, these quickly break down in the non-commutative setting of matrices. Hence, a central question in random matrix theory is whether similar universality properties hold with matrices, and specifically their eigenvalues.

Additionally, analysis is often restricted to self-adjoint matrices. This restriction is motivated by a handful of factors, originating with applications to physics. This also ensures a certain level of mathematical tractability compared to the non-symmetric case. Most importantly, this ensures real eigenvalues and therefore probabilistic spectral analysis can be limited to the real numbers.

It is important to note that all of the theory that can be found in this overview is applicable to the complex Hermitian case up to minor modifications, but we restrict ourselves to the real symmetric case for simplicity.

Definition 1.1 (Symmetric Random Matrix). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. An $N \times N$ symmetric random matrix is a measurable mapping

$$H : \Omega \to \mathrm{Sym}_N(\mathbb{R}) := \{A \in \mathbb{R}^{N\times N} : A^T = A\}.$$

Equivalently, H may be specified by a collection of real-valued random variables $H_{ij} : \Omega \to \mathbb{R}$ for $1 \le i, j \le N$, such that each H_{ij} is measurable and $H_{ij} = H_{ji}$ almost surely for all $1 \le i, j \le N$.

2 Spectral Methods

In this section, we introduce the foundational mathematical tools to analyze the spectral properties of random matrices. These tools are necessary to link algebraic matrix properties to probability measures.

2.1 Empirical Spectral Distribution

Definition 2.1 (Empirical Spectral Distribution). Let $H \in \mathbb{R}^{N\times N}$ be a symmetric matrix with eigenvalues $\lambda_1, \dots, \lambda_N$, the *empirical spectral distribution (ESD)* of H is the probability measure

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i},$$

where $\delta_x(A) = \mathbf{1}_A(x)$ is the Dirac measure at x. If H is random, then μ_N is a random probability measure.

Of particular interest is the asymptotic behaviour of the sequence of random measures $(\mu_N)_{N\geq 1}$ associated with a given sequence of matrices $(H_N)_{N\geq 1}$, in particular whether μ_N converges weakly (in an appropriate probabilistic sense) to a deterministic limiting measure as $N \to \infty$.

2.2 The Stieltjes Transform

The Stieltjes transform provides an analytic encoding of a probability measure, allowing weak convergence of measures to be studied via convergence of holomorphic functions. It is the analogue in random matrix theory of the characteristic function in probability theory, in that it encodes a probability measure in a transform domain where convergence is more tractable.

Definition 2.2 (Stieltjes Transform). Let μ be a Borel probability measure on $\mathbb{R}$. The *Stieltjes transform* of μ is the function

$$m_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(x)}{x-z},$$

defined for $z \in \mathbb{H}$, where $\mathbb{H} = \{z \in \mathbb{C} : \text{Im } z > 0\}$. If μ is random, then m_μ is a random complex-valued function.

It can be shown that the Stieltjes transform is analytic, and has Im $m_\mu(z) > 0$ for all $z \in \mathbb{H}$.

Proposition 2.3 (Inverse Stieltjes Transform). *Let μ be a Borel probability measure, $a, b \in \mathbb{R}$ with $a < b$. Then,*

$$\lim_{\eta\to 0^+} \frac{1}{\pi}\int_a^b \operatorname{Im} m_\mu(x+i\eta)\, dx = \mu((a,b)) + \frac{1}{2}\big(\mu(\{a\}) + \mu(\{b\})\big).$$

Proof. First, observe that

$$\operatorname{Im} m_\mu(x+i\eta) = \int_{\mathbb{R}} \operatorname{Im}\left(\frac{1}{t-(x+i\eta)}\right) d\mu(t) = \int_{\mathbb{R}} \frac{\eta}{(t-x)^2+\eta^2}\, d\mu(t).$$

Since the integrand is non-negative, Tonelli's theorem gives

$$\int_a^b \operatorname{Im} m_\mu(x+i\eta)\,dx = \int_{\mathbb{R}} \int_a^b \frac{\eta}{(t-x)^2+\eta^2}\,dx\,d\mu(t).$$

Substituting $u=(t-x)/\eta$ into the inner integral yields

$$\begin{aligned}\int_a^b \frac{\eta}{(t-x)^2+\eta^2}\,dx &= \int_{(a-t)/\eta}^{(b-t)/\eta} \frac{1}{1+u^2}\,du \\ &= \arctan\left(\frac{b-t}{\eta}\right) - \arctan\left(\frac{a-t}{\eta}\right).\end{aligned}$$

Thus, we obtain:

$$\int_a^b \operatorname{Im} m_\mu(x+i\eta)\,dx = \int_{\mathbb{R}} \left[\arctan\left(\frac{b-t}{\eta}\right) - \arctan\left(\frac{a-t}{\eta}\right)\right] d\mu(t).$$

Now, notice that the integrand is bounded in absolute value by π, which is integrable with respect to the probability measure μ. Moreover, as $\eta \to 0^+$, pointwise convergence holds:

$$\arctan\left(\frac{b-t}{\eta}\right) - \arctan\left(\frac{a-t}{\eta}\right) \to \pi \mathbf{1}_{(a,b)}(t) + \frac{\pi}{2}\big(\mathbf{1}_{\{a\}}(t) + \mathbf{1}_{\{b\}}(t)\big).$$

Finally, applying the dominated convergence theorem, we obtain the desired limit:

$$\lim_{\eta\to 0^+} \frac{1}{\pi} \int_{\mathbb{R}} \left[\arctan\left(\frac{b-t}{\eta}\right) - \arctan\left(\frac{a-t}{\eta}\right)\right] d\mu(t) = \mu((a,b)) + \frac{1}{2}\big(\mu(\{a\}) + \mu(\{b\})\big),$$

which completes the proof. □

In particular, observe that if μ is absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}$, the inverse Stieltjes transform of m_μ recovers exactly $\mu(a,b)$. Furthermore, the density $\rho = \frac{d\mu}{d\lambda}$ has the property

$$\rho(x) = \lim_{\eta\to 0^+} \frac{1}{\pi} \operatorname{Im}\, m_\mu(x+i\eta),$$

almost everywhere, since $\operatorname{Im}\, m_\mu(x+i\eta) = (P_\eta * \rho)(x)$, where P_η is the Poisson kernel. Since $(P_\eta)_{\eta>0}$ is an approximation to the identity, a result in [7] implies that $(P_\eta * \rho)(x) \to \rho(x)$ for almost every $x \in \mathbb{R}$.

Proposition 2.4 (Equivalence of Convergence). *Let $(\mu_N)_{N\geq 1}$ be a sequence of Borel probability measures on $\mathbb{R}$, and let m_{μ_N} denote their Stieltjes transforms. Then $\mu_N \implies \mu$ weakly for some probability measure μ if and only if*

$$m_{\mu_N}(z) \to m_\mu(z) \quad \text{for all } z \in \mathbb{H}.$$

The proof of this result is found in [1] and requires several technical lemmas that are not pertinent to the scope of this report. Nonetheless, the result is fundamental as it reduces questions of convergence of measures to convergence of analytic functions.

In particular, this equivalence extends naturally to random probability measures and preserves the mode of convergence. Specifically, if a sequence of random probability measures (μ_N) has Stieltjes transforms that converge almost surely to the transform of a limit measure μ for all $z \in \mathbb{H}$, then the measures are guaranteed to converge almost surely to μ. The same holds true in probability: if the transforms converge in probability, the measures also converge in probability.

2.3 The Resolvent

Finally, we introduce the resolvent matrix (G(z)), which provides a convenient analytic representation of spectral information. Rather than working directly with the empirical eigenvalue distribution, which is difficult to access through the characteristic polynomial, we instead study the resolvent whose analytic structure allows access to the empirical spectral distribution through algebraic manipulation.

Definition 2.5 (Resolvent). Let $H \in \mathbb{R}^{N\times N}$ be a symmetric matrix. The resolvent of H is the matrix-valued function

$$G(z) = (H - zI)^{-1},$$

defined for all $z \in \mathbb{C}$ such that z is not an eigenvalue of H. If H is a random, then $G(z)$ is a random complex matrix-valued function.

Proposition 2.6 (Resolvent-Stieltjes Relation). *Let $H \in \mathbb{R}^{N\times N}$ be a symmetric matrix. Then, for all $z \in C$ for which $G(z)$ is defined,*

$$\frac{1}{N}\mathbf{tr}\, G(z) = m_N(z),$$

where m_N is the Stieltjes transform of the empirical spectral distribution of H.

Proof. Since H is real symmetric, it admits an orthogonal diagonalization $H = Q\Lambda Q^T$, where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ and $QQ^T = I$. For all $z \in C$ for which $G(z)$ is defined, we have

$$G(z) = (H - zI)^{-1} = (Q\Lambda Q^T - zQIQ^T)^{-1} = Q(\Lambda - zI)^{-1}Q^T.$$

Using the cyclic invariance of the trace,

$$\frac{1}{N}\mathbf{tr}\, G(z) = \frac{1}{N}\mathbf{tr}[(\Lambda - zI)^{-1}Q^TQ] = \frac{1}{N}\mathbf{tr}\,(\Lambda - zI)^{-1} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z}.$$

By definition of the empirical spectral distribution $\mu_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$, its Stieltjes transform is

$$m_N(z) = \int_{\mathbb{R}} \frac{1}{x - z}\, d\mu_N(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} = \frac{1}{N}\mathbf{tr}\, G.$$

□

This connection allows spectral problems to be framed as analytical matrix problems involving $G(z)$, which is especially useful for determining the limiting spectral distribution of a sequence of matrices $(H_N)_{N\geq 1}$. By Proposition 2.4, it suffices to study the convergence of the normalized trace of the resolvent. If this limiting quantity exists and is deterministic, it can be inverted using Proposition 2.3 to recover the limiting spectral distribution.

At this point, we have reframed the problem of determining the limiting spectral distribution of $(H_N)_{N\geq 1}$ in terms of the normalized trace of the resolvent $G(z)$. Depending on the class of matrices (H_N), a variety of methods exist to establish convergence of the normalized trace of the resolvent. In many classical random matrix ensembles, the limiting Stieltjes transform is characterized as the solution of a deterministic self-consistent equation, commonly referred to as the Dyson equation. We turn to this formulation next.

3 The Dyson Equation for Wigner Matrices

Wigner matrices are a fundamental starting point in random matrix theory and arise as a simplified description of large complex systems with many weakly interacting components. Their importance stems from the fact that, despite the simplicity of the independence assumptions on the entries, they capture universal features of large random systems.

Definition 3.1 (Wigner Matrix). A *Wigner matrix* H is a symmetric random matrix $(H_{ij})_{i,j=1}^N$ where the entries H_{ij} are independent and identically distributed with

$$\mathbb{E}[H_{ij}] = 0, \quad \mathbb{E}[H_{ij}^2] = \frac{1}{N}$$

for $i \leq j$, and $H_{ji} = H_{ij}$ for $i > j$.

No further assumptions are imposed on the distribution of H_{ij}, it may be discrete or continuous, and need not satisfy any smoothness or structural conditions beyond the condition on the first and second moments. For examples of various Wigner matrices as well as simulations of their asymptotic distribution, see Section 4.1. These experiments suggest that the asymptotic distribution of (H_N) does not depend on the specific distribution of the entries, and is given by the semicircle distribution. The following theorem gives a precise statement of this convergence, via the Dyson equation for Wigner matrices. We expand the argument presented in [3] by filling in several intermediate steps.

Theorem 3.2 (Wigner Semicircle Law via the Dyson Equation). *Let $(H_N)_{N\geq 1}$ be a sequence of Wigner matrices. There exists a deterministic function $m : \mathbb{H} \to \mathbb{H}$ such that $m_N(z) \xrightarrow{p} m(z)$ as $N \to \infty$ pointwise, where m_N is the Stieltjes transform of the empirical spectral distribution μ_N of H_N. Moreover, m is the unique solution to the Dyson equation*

$$m(z) = -\frac{1}{z + m(z)}.$$

Let μ be the probability measure whose Stieltjes transform is m. Then $\mu_N \implies \mu$ in probability, and μ is absolutely continuous with respect to the Lebesgue measure with density

$$\rho(x) = \frac{1}{2\pi}\sqrt{(4 - x^2)_+}, \qquad x \in \mathbb{R}.$$

3.1 Probabilistic Step

The first objects of interest are the diagonal elements G_{ii} of the resolvent. For a fixed i, we may assume without loss of generality that $i = 1$, since the entries are i.i.d (up to symmetry), and relabeling the indices does not affect any relevant quantities. For convenience, we will often write $G(z)$ simply as G.

Lemma 3.3 (Schur Formula). *Let $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{m\times n}$, and $C \in \mathbb{C}^{m\times m}$, and define*

$$D = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix}.$$

Assume C is invertible and define the Schur complement $D' = A - B^T C^{-1} B$. If D is invertible, then D' is invertible, and $(D^{-1})_{ij} = (D')^{-1}_{ij}$ for $i, j \leq n$.

Proof. We factor the block matrix D using block Gaussian elimination,

$$\begin{pmatrix} I_n & -B^T C^{-1} \\ 0 & I_m \end{pmatrix} \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} = \begin{pmatrix} A - B^T C^{-1} B & 0 \\ B & C \end{pmatrix} = \begin{pmatrix} D' & 0 \\ B & C \end{pmatrix}.$$

Factor once more on the right,

$$\begin{pmatrix} D' & 0 \\ B & C \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -C^{-1}B & I_m \end{pmatrix} = \begin{pmatrix} D' & 0 \\ 0 & C \end{pmatrix}.$$

Combining the two factorizations yields

$$D = \begin{pmatrix} I_n & -B^T C^{-1} \\ 0 & I_m \end{pmatrix}^{-1} \begin{pmatrix} D' & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -C^{-1}B & I_m \end{pmatrix}^{-1}.$$

Since the outer factors are invertible whenever C is invertible, we see that D is invertible if and only if D' is invertible. Furthermore,

$$D^{-1} = \begin{pmatrix} I_n & 0 \\ -C^{-1}B & I_m \end{pmatrix} \begin{pmatrix} D'^{-1} & 0 \\ 0 & C^{-1} \end{pmatrix} \begin{pmatrix} I_n & -B^T C^{-1} \\ 0 & I_m \end{pmatrix},$$

which has upper left block D'^{-1}. □

Let $H^{[i]}$ denote the (i,i) minor of H. Set $G^{[i]} = (H^{[i]} - zI)^{-1}$. Note that this is not the same as $((H - zI)^{-1})^{[i]}$. For the case of $i = 1$, we have

$$H - zI = \begin{pmatrix} H_{11} - z & a^T \\ a & H^{[1]} - zI_{N-1} \end{pmatrix},$$

where $a = (H_{12}, \ldots, H_{1N})^T$. Applying Lemma 3.3 gives

$$G_{11} = \left[(H - zI)^{-1}\right]_{11} = \frac{1}{H_{11} - z - a^T G^{[1]} a}.$$

Since a and $G^{[1]}$ are independent and $\mathbb{E}[H_{1j}H_{1k}] = \delta_{jk}$, we have that

$$\mathbb{E}[a^T G^{[1]} a \,|\, G^{[1]}] = \sum_{j,k>1} \mathbb{E}[H_{1j}H_{1k}] G^{[1]}_{jk} = \frac{1}{N-1} \operatorname{tr} G^{[1]}.$$

It is also possible to show that $\operatorname{Var}(a^T G^{[1]} a) = O(1/N)$, although its proof is technical and not essential to the main structure of the derivation and has therefore been omitted. Along with the fact that $\operatorname{Var}(H_{11}) = 1/N$, we have that

$$G_{11} = -\frac{1}{z + \frac{1}{N-1} \mathbf{tr}\, G^{[i]} + o_p(1)}.$$

Lemma 3.4. *Let $H \in \mathbb{R}^{N\times N}$ be symmetric with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N$. Let $\gamma_1 \geq \ldots \geq \gamma_{N-1}$ be the eigenvalues of the (i,i) minor $H^{[i]}$. Then,*

$$\lambda_1 \geq \gamma_1 \geq \lambda_2 \geq \gamma_2 \geq \ldots \geq \gamma_{N-1} \geq \lambda_N.$$

As a consequence, for any $z \in \mathbb{H}$,

$$\frac{1}{N-1} \operatorname{Tr} G^{[i]}(z) = \frac{1}{N} \operatorname{Tr} G(z) + o_p(1).$$

Proof. The first part is the Cauchy Interlace Theorem, a standard result in matrix analysis, see [5]. Let $f(x) = \frac{1}{x-z}$ for $x \in \mathbb{R}$. Then,

$$\frac{1}{N} \operatorname{Tr} G(z) - \frac{1}{N-1} \operatorname{Tr} G^{[i]}(z) = \frac{f(\lambda_N)}{N} - \frac{1}{N} \sum_{i=1}^{N-1} (f(\lambda_i) - f(\gamma_i)) + \left(\frac{1}{N} - \frac{1}{N-1}\right) \sum_{j=1}^{N-1} f(\gamma_j).$$

Since $f(x) \leq \frac{1}{\operatorname{Im} z}$, we have that

$$\left| \left(\frac{1}{N} - \frac{1}{N-1}\right) \sum_{j=1}^{N-1} f(\mu_j) \right| \leq \left| \frac{1}{N} - \frac{1}{N-1} \right| \cdot \frac{N-1}{\operatorname{Im} z} = \frac{1}{N \operatorname{Im} z},$$

which goes to 0 as $N \to \infty$. As for the second term, we first note that f is Lipschitz continuous with constant $L = (\operatorname{Im} z)^{-2}$ since for any $x, y \in \mathbb{R}$,

$$|f(x) - f(y)| = \left| \frac{1}{x-z} - \frac{1}{y-z} \right| = \frac{|x-y|}{|x-z||y-z|} \leq \frac{|x-y|}{(\operatorname{Im} z)^2}.$$

Since $\gamma_i \in [\lambda_{i+1}, \lambda_i]$,

$$|f(\gamma_i) - f(\lambda_i)| \leq L(\lambda_i - \gamma_i) \leq L(\lambda_i - \lambda_{i+1}),$$

we have

$$\frac{1}{N} \sum_{i=1}^{N-1} f(\lambda_i) - f(\gamma_i) \leq \frac{L}{N} \sum_{i=1}^{N-1} (\lambda_i - \lambda_{i+1}) = \frac{L}{N} (\lambda_N - \lambda_1).$$

Since $\lambda_N = ||H||_{\text{op}}$ and $\lambda_1 = -||-H||_{\text{op}}$, we have that $\lambda_N - \lambda_1 \leq 2||H||_{\text{op}}$, and it suffices to bound the operator norm. This probabilistic bound is shown in greater generality in [4], and gives that $\lambda_N - \lambda_1 \xrightarrow{p} 4$, hence this second term is $o_p(1)$.

Finally, the first term has $\frac{f(\lambda_N)}{N} \leq \frac{1}{N \operatorname{Im} z}$, and thus goes to 0. Since all terms involved have limit 0 in probability as $N \to \infty$, the proof is complete.

□

Updating the previous equation using Lemma 3.4,

$$G_{11} = -\frac{1}{z + \frac{1}{N}\mathbf{tr}\, G + o_p(1)}.$$

Since we assumed $i = 1$ without loss of generality, we obtain the normalized trace

$$\frac{1}{N}\mathbf{tr}\, G = -\frac{1}{z + \frac{1}{N}\mathbf{tr}\, G + o_p(1)}.$$

By Proposition 2.6, we have the asymptotic self-consistent equation

$$m_N(z) = -\frac{1}{z + m_N(z) + o_p(1)}.$$

For any fixed z, we have that $|m_N(z)| \leq (\operatorname{Im} z)^{-1}$ since $|\lambda_i - z| \geq \operatorname{Im} z$ for all eigenvalues. This bound is constant with respect to N, hence rearranging the above equation we obtain

$$1 + m_N(z)(z + m_N(z)) \xrightarrow{p} 0,$$

pointwise. It remains to show that this asymptotic self-consistent equation is stable, in the sense that any sequence (m_N) satisfying $1 + m_N(z)(z + m_N(z)) \xrightarrow{p} 0$ must converge to the unique solution m of the equation $1 + m(z)(z + m(z)) = 0$.

3.2 Stability Step

Let m satisfy the exact equation $1 + m(z)(z + m(z)) = 0$.

Lemma 3.5 (Exact Solution). *For all $z \in \mathbb{H}$, the unique solution to $1 + m(z)(z + m(z)) = 0$ is given by*

$$m(z) = \frac{-z + \sqrt{z^2 - 4}}{2},$$

whose corresponding inverse Stieltjes transform has density

$$\rho(z) = \frac{1}{2\pi}\sqrt{(4 - x^2)_+}, \quad x \in \mathbb{R}.$$

Proof. Solving the exact equation using the quadratic formula yields two solutions

$$m(z) = \frac{-z \pm \sqrt{z^2 - 4}}{2}.$$

First, consider the positive branch. By using the Taylor expansion of the square root,

$$m_+(z) = \frac{-z + z - 2/z + O(z^{-3})}{2} = -\frac{1}{z} + O(|z|^{-3}).$$

Letting $z = x + i\eta$ with $x > 0$ and $\eta > 0$,

$$\operatorname{Im}\left(-\frac{1}{z}\right) = -\operatorname{Im}\left(\frac{x - i\eta}{x^2 + \eta^2}\right) = \frac{\eta}{x^2 + \eta^2} > 0.$$

Hence, we have $\operatorname{Im} m_+(z) > 0$ for large enough z. For the negative branch, a similar analysis yields

$$m_-(z) = -z + O(z^{-1}),$$

so $\operatorname{Im} m_-(z) < 0$ for large enough z. Since each branch is clearly analytic, if $\operatorname{Im} m(z)$ changed sign inside $\mathbb{H}$, we would necessarily have $\operatorname{Im} m(z) = 0$ for some $z \in \mathbb{H}$. Let $y \in \mathbb{R}$ be the value of $m(z)$ at this point. From the exact equation, we have

$$m(z) = -\frac{1}{z + m(z)},$$

which would imply that

$$\operatorname{Im} m(z) = -\operatorname{Im}\left(\frac{1}{x + i\eta + y}\right) = -\operatorname{Im}\left(\frac{x + y - i\eta}{(x+y)^2 + \eta^2}\right) = \frac{\eta}{(x+y)^2 + \eta^2} > 0,$$

which contradicts that $\operatorname{Im} m(z) = 0$. Thus, $m(z)$ does not change signs. Since we must have $\operatorname{Im} m(z) > 0$ for all $z \in \mathbb{H}$ by a basic property of the Stieltjes transform, the positive branch is therefore the unique solution.

Using Proposition 2.3,

$$\frac{1}{\pi} \lim_{\eta \to 0^+} \int_a^b \operatorname{Im}\left(\frac{-(x + i\eta) + \sqrt{(x + i\eta)^2 - 4}}{2}\right) dx = \frac{1}{2\pi} \int_a^b \sqrt{(4 - x^2)_+}\, dx = \mu((a, b)).$$

Since a finite Borel measure on $\mathbb{R}$ is uniquely determined by its values on intervals, it follows that μ as precisely the probability measure with density $\rho(x) = \frac{1}{2\pi}\sqrt{(4 - x^2)_+}$. $\square$

Let $\varepsilon_N(z) = 1 + m_N(z)(z + m_N(z))$, where $\varepsilon_N(z) \xrightarrow{p} 0$ pointwise, as we have shown. Let $\delta_N(z) = m_N(z) - m(z)$. For simplicity, we drop the parameter z from relevant quantities as well as the mode of convergence when obvious by context. We wish to show that for any $z \in \mathbb{H}$, if $\varepsilon_N \to 0$ then $\delta_N \to 0$.

By subtracting the exact equation from this equation, we obtain

$$\varepsilon_N = m_N(z + m_N) - m(z + m) = z\delta_N + \delta_N(\delta_N + 2m) = (z + 2m)\delta_N + \delta_N^2.$$

Furthermore, rearranging the exact equation gives $z = -\frac{m^2+1}{m}$, and hence

$$\varepsilon_N = \frac{m^2 - 1}{m}\delta_N + \delta_N^2.$$

In order to establish convergence, we first show a bound $|\delta_N| \leq C\,|\varepsilon_N|$ for some $C > 0$ potentially depending on

z for a deterministic sequence (m_N).

Lemma 3.6. *For any fixed $z \in \mathbb{H}$, there exists a $c(z) > 0$ such that*

$$\left| \frac{(m(z))^2 - 1}{m(z)} \right| \geq c(z) > 0.$$

Furthermore, if (δ_N) is a deterministic sequence, there exists an $N_0 \geq 1$ such that for all $N \geq N_0$,

$$|\delta_N(z)| \leq \frac{c(z)}{2} \quad \text{and} \quad |\delta_N(z)| \leq \frac{2}{c(z)} |\varepsilon_N(z)|.$$

Proof. Let $s(z) = \sqrt{z^2 - 4}$. Rearranging the exact solution for m found in Lemma 3.5, we have $2m + z = s$ From the exact Dyson equation $1 + m(z + m) = 0$, we also have $m + \frac{1}{m} = -z$. Thus,

$$\frac{m^2 - 1}{m} = 2m - \left(m + \frac{1}{m}\right) = 2m + z = s(z).$$

Let $c(z) = \min\{|z - 2|, |z + 2|\}$. For $c(z) > 0$,

$$|s(z)|^2 = |z^2 - 4| = |z - 2||z + 2| \geq c(z)^2.$$

Thus, $|s(z)| \geq c(z) > 0$. The case $\tau = 0$ is more involved and is treated in greater generality in [2].

Assume for contradiction that for all N_0, there exists an $N_* > N_0$ such that $|\delta_N| > \frac{c}{2}$. By the reverse triangle inequality,

$$|\varepsilon_{N_*}| \geq \left| \frac{m^2 - 1}{m} \right| |\delta_{N_*}| - |\delta_{N_*}|^2 \geq |\delta_{N_*}|(c - |\delta_{N_*}|) > \frac{c^2}{4}.$$

By assumption, $\varepsilon_N \to 0$, which contradicts the previous claim. Now that we have established that there is some N_0 such that $|\delta_N| \leq c/2$ for $N \geq N_0$, a slight modification to the valid part of the previous inequality gives

$$|\delta_N| \leq \frac{|\varepsilon_N|}{c - |\delta_N|} \leq \frac{2}{c} |\varepsilon_N|,$$

for all $N \geq N_0$. $\square$

Since $c(z) > 0$ is independent of N and $\varepsilon_N(z) \to 0$, it follows that $\delta_N(z) \to 0$, and thus $m_N \to m$ pointwise. While this was done for a deterministic sequence $m_N(z)$, the same analysis can be carried out in the probabilistic setting, up to minor modifications, yielding $m_N \xrightarrow{p} m$ pointwise. By Proposition 2.4 applied to the exact solution found in Lemma 3.5, we have that the sequence of empirical spectral distributions μ_N converges weakly to the measure μ with density ρ. This concludes the proof of Theorem 3.2.

4 Numerical Experiments

4.1 Empirical Spectral Distribution

We consider the following sequences of Wigner matrices:

1. (A_N) with entries $A_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/N)$,
2. (B_N) with i.i.d. entries $B_{ij} = \pm 1/\sqrt{N}$, each with probability 1/2,
3. (C_N) with entries $C_{ij} = (Y_{ij} - 1)/\sqrt{N}$, where $Y_{ij} \overset{\text{i.i.d.}}{\sim} \text{Exp}(1)$.

Observe that the distributions of the entries of A_N, B_N, and C_N, are continuous, discrete, and non-symmetric, respectively. We define the histogram density approximation of the empirical spectral measure, given by

$$\hat{\rho}(B_k) = \frac{\mu_N(B_k)}{|B_k|},$$

where $(B_k)_k$ is a partition of $\mathbb{R}$ into bins. Using appropriately chosen bin sizes, Figure 1 displays $\hat{\rho}$ for a single realization of A_N (blue), B_N (red), and C_N (green), for $N = 50$, 500, 5000. These results provide strong evidence for the semicircle law.

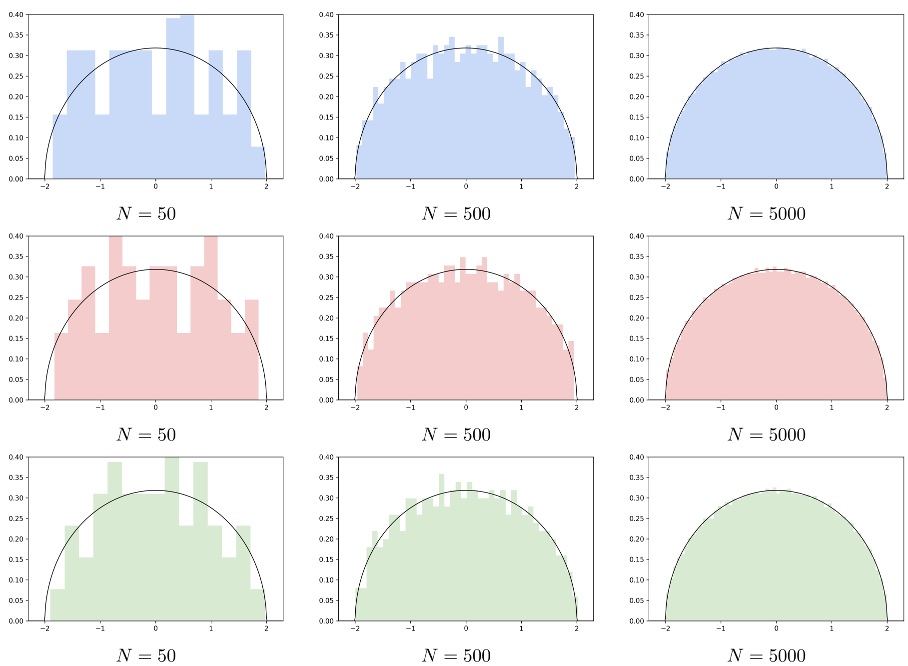

Figure 1: Convergence of $\hat{\rho}$ to the semicircle law.

4.2 Stieltjes Transform

Let $z = 1 + i\eta$, where $\eta = 0.1$, 0.25, 0.5, 1.0, 2.0. For evenly spaced values of N between 50 to 2000, we compute $|\hat{\delta}_N|$, the sample mean of $|\delta_N| = |m_N(z) - m(z)|$, calculated using 10 realizations of A_N. Figure 2 plots $\log(\delta_N)$ as a function of N for each choice of η.

Although not shown directly, several calculations in the preceding analysis suggest that δ_N should scale $O\left(\frac{1}{N\eta}\right)$. As a consequence, we should expect that $\log \delta_N \approx -\log N - \log \eta + C$, whose behaviour is consistent with the results in Figure 2. We expect similar behaviour for other choices of $\text{Re}\, z$ as η varies.

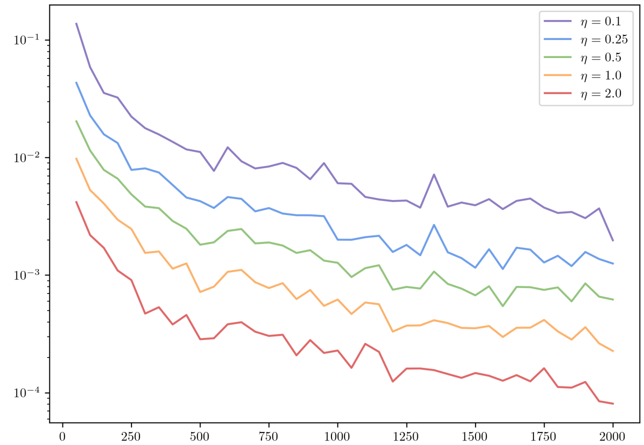

Figure 2: Convergence of $|\hat{\delta}_N|$.

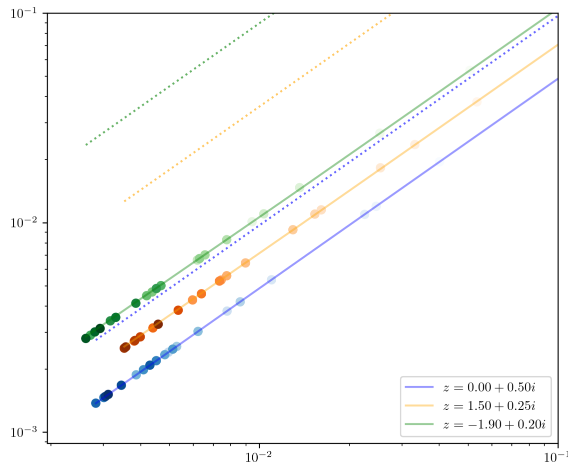

Figure 3: Convergence and Boundedness of $|\hat{\delta}_N|$ with respect to $|\hat{\varepsilon}_N|$ on a log-log scale. Each point $(|\hat{\varepsilon}_N|, |\hat{\delta}_N|)$ corresponds to a given N, with opacity increasing as N grows. Dotted lines indicate the bound shown in Lemma 3.6.

4.3 Dyson Equation

Lemma 3.6 establishes a bound of the form $|\delta_N| \leq C|\varepsilon_N|$ for sufficiently large N, where $C = 2/\text{dist}(z, \{-2, 2\})$. In Figure 3, we illustrate the log-log relationship between $|\hat{\varepsilon}_N|$ and $|\hat{\delta}_N|$ for three values of z alongside the corresponding theoretical bounds, over a range of N values from 50 to 1000 with equal spacing.

The seemingly linear relationship is likely due to the fact that we have the exact equality

$$\varepsilon_N = \frac{m^2 - 1}{m}\delta_N + \delta_N^2,$$

which is roughly linear for $\delta_N \approx 0$.

5 Conclusion

While we presented the Dyson equation on the classical Wigner ensemble, the most general framework extends well beyond this case. Our analysis is primarily based on the initial parts of [3], and the remainder of that work develops this more general approach to deriving and using Dyson equations, notably for the case of correlated entries with potentially non-uniform variance profiles. The Dyson equation can also be used to derive the Marchenko–Pastur law for eigenvalues of sample covariance matrices, as shown in [6].

Acknowledgments

We would like to thank our mentor Kevin Xiao for his guidance, insight, and expertise throughout this project. We are especially grateful for his thoughtful advice and perspectives on graduate-level mathematics throughout the course of this work. We also appreciate his sense of humor, which made the learning process a lot of fun.

References

[1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2010.

[2] László Erdős and Horng-Tzer Yau. *A Dynamical Approach to Random Matrix Theory*, volume 28 of *Courant Lecture Notes in Mathematics*. American Mathematical Society, 2017.

[3] László Erdős. The matrix dyson equation and its applications for random matrices. *arXiv preprint arXiv:1903.10060*, 2019.

[4] Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[5] Suk-Geun Hwang. Cauchy's interlace theorem for eigenvalues of hermitian matrices. *The American Mathematical Monthly*, 111(2):157–159, 2004.

[6] Elliot Paquette. Random matrix theory and optimization. Lecture notes, CRM-PIMS Summer School, 2024. Available at https://elliotpaquette.github.io/notes/RMTandOptimization-20240701.pdf.

[7] Elias M. Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2005.