Universal Spectral Laws I: Global Behavior

SDS 391P.6, Spring 2026

Pratik Patil

These notes are a work in progress and are provided as-is for instructional purposes only. They are not (yet) at the level of a scholarly document. In particular, the notes draw from various sources and do not (yet) have sufficient references to the original sources. Additionally, almost surely the notes have errors and they are only probably approximately correct. The notes will be updated regularly as the course progresses. Last updated: 2026-03-01.

1 Motivation

In the last lecture, we controlled the *operator norm* of random matrices by a three-step template:

$$\text{(fixed direction concentration)} \ + \ \text{(an } \varepsilon\text{-net)} \ + \ \text{(union bound)}.$$

For covariance estimation, this gave a non-asymptotic guarantee of the form

$$\|\widehat{\Sigma} - \Sigma\| \ \lesssim \ K^2\big(\sqrt{d/n} + d/n\big)\,\|\Sigma\| \qquad \text{with high probability,}$$

under sub-Gaussian assumptions.

But the operator norm is just one summary statistic. In high dimensions, a much richer (and often more informative) object is the *entire spectrum*. A striking phenomenon from random matrix theory is that when both dimension and sample size are large, the eigenvalue histogram often follows a universal law: it depends only on coarse distributional features (typically mean/variance), not on fine details.

Two canonical experiments:

- *Wigner noise.* Let W be an $n \times n$ symmetric matrix with independent mean-zero variance-one entries on and above the diagonal. The eigenvalues of $\frac{1}{\sqrt{n}}W$ form a compactly supported "semicircle" cloud.
- *White-noise covariance.* Let $X_1, \ldots, X_n \sim \mathcal{N}(0, I_d)$ and

  $$\Sigma_n := \frac{1}{n}\sum_{k=1}^{n} X_k X_k^\top.$$

  The eigenvalues of Σ_n do *not* concentrate at 1 when d is comparable to n. Instead, they spread across an interval whose endpoints are approximately

  $$(1-\sqrt{r})^2, \quad (1+\sqrt{r})^2, \qquad r := d/n,$$

  following the Marchenko–Pastur law.

This lecture explains these two global laws and a proof template based on the *resolvent* (matrix inverse) and the *Stieltjes transform*.

2 Resolvent and Stieltjes transform

Let $A \in \mathbb{R}^{n\times n}$ be symmetric with eigenvalues $\lambda_1(A), \dots, \lambda_n(A)$.

The empirical spectral distribution (ESD) of A is the probability measure

$$\mu_A := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}.$$

Informally, the eigenvalue histogram (with density normalization) is a picture of μ_A.

Rather than tracking eigenvalues directly, we probe μ_A via the resolvent

$$G_A(z) := (A - zI)^{-1}, \qquad z \in \mathbb{C}\backslash\mathbb{R}.$$

The key scalar quantity is the normalized trace

$$m_A(z) := \frac{1}{n}\,\mathrm{tr}\; G_A(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i(A) - z}. \tag{1}$$

This is exactly the Stieltjes transform of the ESD:

$$m_A(z) = \int \frac{1}{x - z}\, d\mu_A(x).$$

If μ has a density ρ (with respect to Lebesgue measure), then for continuity points x,

$$\rho(x) = \lim_{\eta\downarrow 0}\frac{1}{\pi}\,\mathrm{Im}\; m_\mu(x + i\eta), \qquad m_\mu(z) := \int \frac{1}{t - z}\, d\mu(t).$$

So if we can identify the limit of $m_{A_n}(z)$ for z in the upper half-plane, we can recover the limiting eigenvalue density.

The resolvent has two big advantages:

- It behaves smoothly as a function of z (analytic off the real axis).
- It admits algebraic identities (Schur complements, Sherman–Morrison) that let us isolate one row/column at a time and exploit independence.

3 The semicircle law

3.1 Model and statement

A *Wigner matrix* is a symmetric random matrix $W \in \mathbb{R}^{n\times n}$ such that

- $(W_{ij})_{1\leqslant i\leqslant j\leqslant n}$ are independent;
- $\mathbb{E}W_{ij} = 0$ for all $i \leqslant j$;
- $\mathbb{E}W_{ij}^2 = 1$ for $i < j$ (and the diagonal has uniformly bounded variance);
- typically we assume a mild moment condition (e.g. sub-Gaussian entries, or at least finite $4 + \epsilon$ moment).

Define the scaled matrix

$$A_n := \frac{1}{\sqrt{n}} W.$$

The semicircle law describes the limiting ESD of A_n.

Theorem 3.1 (Semicircle law, global form)**.** *Let* $A_n = \frac{1}{\sqrt{n}}W$ *be a Wigner matrix as above. Then* μ_{A_n} *converges weakly (in probability, and under standard assumptions almost surely) to the semicircle distribution* μ_{sc} *on* $[-2, 2]$ *with density*

$$\rho_{\mathrm{sc}}(x) := \frac{1}{2\pi}\sqrt{4 - x^2}\,\mathbf{1}_{|x|\leqslant 2}.$$

Equivalently, for every bounded continuous $f : \mathbb{R} \to \mathbb{R}$*,*

$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i(A_n)) \longrightarrow \int f(x)\,\rho_{\mathrm{sc}}(x)\,dx.$$

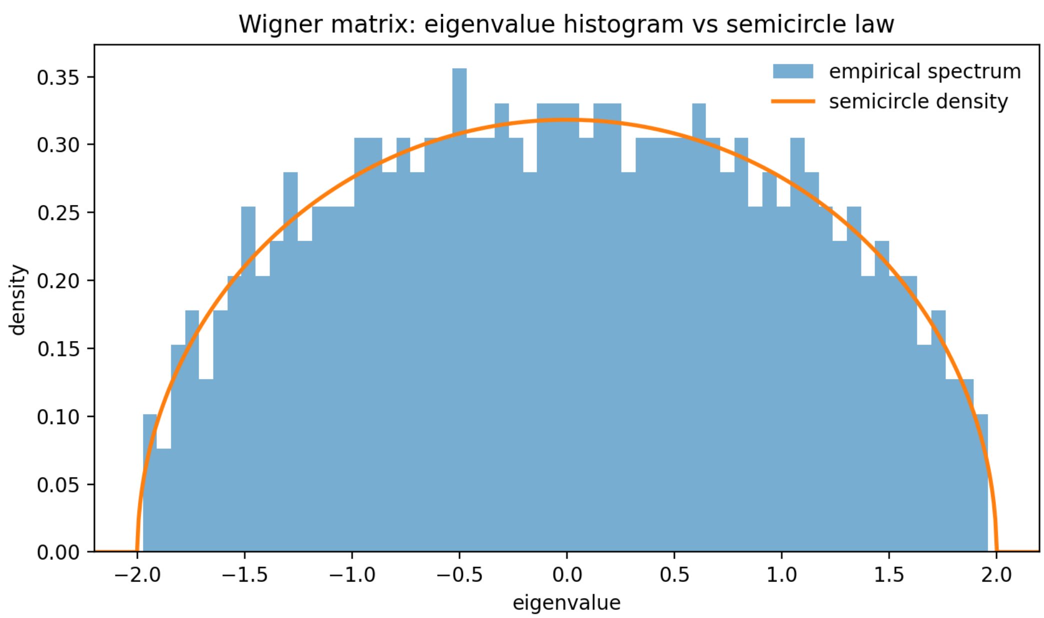

Figure 1: Eigenvalue histogram of a simulated Wigner matrix $\frac{1}{\sqrt{n}}W$ (here $n \approx 600$), overlaid with the semicircle density ρ_{sc}.

The semicircle density has support $[-2, 2]$, suggesting $\|A_n\| \approx 2$ and hence $\|W\| \approx 2\sqrt{n}$, consistent with the non-asymptotic bounds we proved earlier (up to constants).

3.2 Proof sketch

Fix $z \in \mathbb{C}$ with $\operatorname{Im} z > 0$ and set

$$G(z) := (A_n - zI)^{-1}, \qquad m_n(z) := \frac{1}{n}\operatorname{tr} G(z).$$

The goal is to show $m_n(z)$ concentrates around a deterministic limit $m(z)$ that solves a self-consistent equation; then identify $m(z)$ as the Stieltjes transform of μ_{sc}.

1. leave-one-out. Write A_n in block form:

$$A_n = \begin{bmatrix} a_{11} & a^\top \\ a & \widetilde{A} \end{bmatrix}, \qquad A_n - zI = \begin{bmatrix} a_{11} - z & a^\top \\ a & \widetilde{A} - zI \end{bmatrix}.$$

By the Schur complement formula,

$$G_{11}(z) = \big[(A_n - zI)^{-1}\big]_{11} = \frac{1}{(a_{11} - z) - a^\top(\widetilde{A} - zI)^{-1}a}. \tag{2}$$

2. quadratic form concentration. The vector a consists of the first column (excluding the diagonal), scaled by $1/\sqrt{n}$. Conditionally on $\widetilde{A}$, the quadratic form

$$a^\top(\widetilde{A} - zI)^{-1}a$$

concentrates around its conditional expectation. By independence and isotropy, the conditional expectation is essentially a trace:

$$\mathbb{E}\Big[a^\top(\widetilde{A} - zI)^{-1}a \,\Big|\, \widetilde{A}\Big] \approx \frac{1}{n}\operatorname{tr}(\widetilde{A} - zI)^{-1} \approx m_n(z).$$

Also $a_{11} = W_{11}/\sqrt{n}$ is negligible at the global scale.

3. self-consistent equation. Plugging these heuristics into (2) suggests that *typically*

$$G_{11}(z) \approx \frac{1}{-z - m_n(z)}.$$

By exchangeability, the same holds for most diagonal entries $G_{ii}(z)$. Averaging over i then gives the self-consistent relation

$$m_n(z) \approx \frac{1}{-z - m_n(z)}.$$

Passing to the limit suggests $m(z)$ solves

$$m(z)^2 + zm(z) + 1 = 0, \qquad \operatorname{Im} z > 0, \tag{3}$$

with the branch condition $m(z) \sim -1/z$ as $|z| \to \infty$.

Solving (3) yields

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

where the square root is chosen so that $\operatorname{Im}\sqrt{z^2 - 4} > 0$ when $\operatorname{Im} z > 0$. By Stieltjes inversion, this m_{sc} corresponds exactly to the semicircle density.

For a full proof, one needs to justify: (i) concentration of the quadratic form uniformly in i, (ii) that $m_n(z)$ concentrates around a deterministic limit, and (iii) stability of the fixed-point equation. These can be done under mild moment conditions; the modern theory proves far stronger *local* versions (down to spectral windows of size $n^{-1+\epsilon}$).

4 The Marchenko–Pastur law

4.1 Model and statement

Let $X_1, \dots, X_n \in \mathbb{R}^d$ be i.i.d. with mean zero and covariance I_d. (For concreteness you can think $X_k \sim \mathcal{N}(0, I_d)$; many non-Gaussian models behave the same at the global level.) Define the sample covariance matrix

$$\Sigma_n := \frac{1}{n} \sum_{k=1}^{n} X_k X_k^\top \in \mathbb{R}^{d \times d}.$$

Assume the aspect ratio converges:

$$r := \frac{d}{n} \to r_\infty \in (0, \infty).$$

To avoid extra case distinctions, we state the law for general r and mention the atom at zero.

Theorem 4.1 (Marchenko–Pastur law, global form)**.** *Assume $d/n \to r \in (0, \infty)$ and the entries of the data vectors have mean 0, variance 1, and a mild moment condition (e.g. sub-Gaussian). Then the ESD μ_{Σ_n} converges weakly (in probability, and under standard assumptions almost surely) to the Marchenko–Pastur distribution $\mu_{\mathrm{MP},r}$ given by:*

$$\mu_{\mathrm{MP},r} = \left(1 - \frac{1}{r}\right)_+ \delta_0 \; + \; \rho_{\mathrm{MP},r}(x)\, dx,$$

where $(u)_+ := \max\{u, 0\}$ and the density is

$$\rho_{\mathrm{MP},r}(x) = \frac{1}{2\pi r x} \sqrt{(b - x)(x - a)}\, \mathbf{1}_{a \leqslant x \leqslant b}, \qquad a = (1 - \sqrt{r})^2, \;\; b = (1 + \sqrt{r})^2.$$

In particular, when $r < 1$ there is no atom at 0, while when $r > 1$ a fraction $1 - 1/r$ of the eigenvalues are exactly zero (rank deficiency).

The MP support $[a, b]$ shows the typical spread of eigenvalues around 1 is on the order of $\sqrt{r} = \sqrt{d/n}$. This matches the leading term in the non-asymptotic operator-norm bound $\|\Sigma_n - I\| \lesssim \sqrt{d/n}$.

4.2 Proof sketch

Fix $z \in \mathbb{C}$ with $\operatorname{Im} z > 0$ and define the resolvent and its trace

$$G(z) := (\Sigma_n - zI)^{-1}, \qquad m_n(z) := \frac{1}{d} \operatorname{tr} G(z).$$

As for the semicircle law, we aim to identify the limiting $m(z)$.

It is convenient to clear the $1/n$ scaling. Let

$$A := \sum_{k=1}^{n} X_k X_k^\top - zn\, I_d, \qquad \text{so that} \qquad \Sigma_n - zI = \frac{1}{n} A,$$

and therefore

$$G(z) = (\Sigma_n - zI)^{-1} = nA^{-1}, \qquad m_n(z) = \frac{n}{d} \operatorname{tr}(A^{-1}) = \frac{1}{r} \operatorname{tr}(A^{-1}), \quad r = \frac{d}{n}.$$

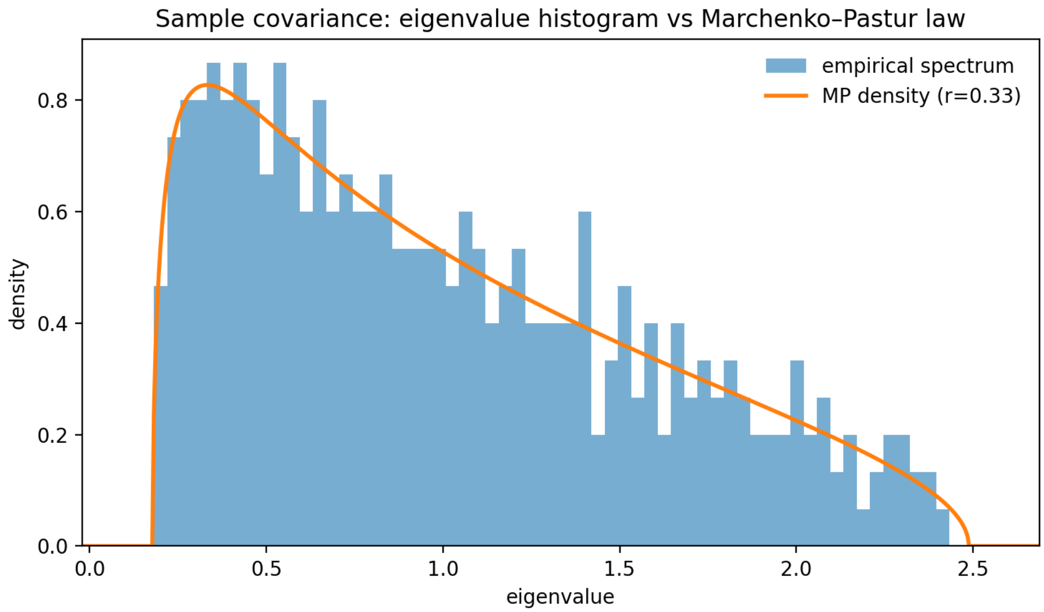

Figure 2: Eigenvalue histogram of a Gaussian sample covariance matrix Σ_n overlaid with the MP density $\rho_{\mathrm{MP},r}$ (here $r = d/n \approx 0.33$).

1. leave-one-out. Let $B := A - X_n X_n^\top = \sum_{k=1}^{n-1} X_k X_k^\top - zn\, I_d$. Then $A = B + X_n X_n^\top$. Sherman–Morrison gives the identity

$$X_n^\top A^{-1} X_n = 1 - \frac{1}{1 + X_n^\top B^{-1} X_n}. \tag{4}$$

2. quadratic form concentration. Conditionally on B, the vector X_n is independent and isotropic, so

$$X_n^\top B^{-1} X_n \approx \mathrm{tr}(B^{-1}).$$

Also, removing one sample should not change the normalized trace much, so $\mathrm{tr}(B^{-1}) \approx \mathrm{tr}(A^{-1})$. Plugging these approximations into (4) suggests

$$X_n^\top A^{-1} X_n \approx 1 - \frac{1}{1 + \mathrm{tr}(A^{-1})}.$$

3. self-consistent equation. Summing (4)-type expressions over $k = 1, \ldots, n$ and using cyclicity,

$$\begin{aligned} \sum_{k=1}^{n} X_k^\top A^{-1} X_k &= \mathrm{tr}\left(\sum_{k=1}^{n} X_k X_k^\top A^{-1}\right) = \mathrm{tr}\left((A + znI)A^{-1}\right) \\ &= \mathrm{tr}(I) + zn\ \mathrm{tr}(A^{-1}) = d + zn\ \mathrm{tr}(A^{-1}). \end{aligned}$$

On the other hand, the heuristic above suggests the left-hand side is close to

$$n\left(1 - \frac{1}{1 + \mathrm{tr}(A^{-1})}\right) = n \cdot \frac{\mathrm{tr}(A^{-1})}{1 + \mathrm{tr}(A^{-1})}.$$

Equating the two and writing $T := \mathrm{tr}(A^{-1})$ gives the approximate relation

$$d + zn\,T \approx n\,\frac{T}{1+T}.$$

Divide by n and use $r = d/n$:

$$r + zT \approx \frac{T}{1+T}.$$

Substitute $T = r\,m_n(z)$ (since $m_n = (1/r)T$) and simplify to obtain the MP self-consistent equation

$$rz\,m(z)^2 + (z + r - 1)\,m(z) + 1 = 0, \qquad \mathrm{Im}\,z > 0, \tag{5}$$

together with the boundary condition $m(z) \sim -1/z$ as $|z| \to \infty$.

Solving (5) yields the Stieltjes transform of $\mu_{\mathrm{MP},r}$:

$$m_{\mathrm{MP},r}(z) = \frac{1 - z - r + \sqrt{(1-z-r)^2 - 4rz}}{2rz},$$

with the branch chosen so that $\mathrm{Im}\,m_{\mathrm{MP},r}(z) > 0$ for $\mathrm{Im}\,z > 0$. Stieltjes inversion then recovers the MP density.

Just as for semicircle, for a full proof, one needs a careful concentration argument for the quadratic forms and a stability argument for the fixed-point equation. Under mild moment assumptions, the global statement follows, and much stronger *local* versions are known.

5 Universality

Both laws above are called *universal* in the following sense: once we normalize correctly, the limiting eigenvalue distribution depends primarily on *low-order moments* (typically mean and variance) and not on the precise entry distribution. Gaussianity is not essential; it is mainly a convenient reference model.

This universality theme will reappear repeatedly:

- Global laws (semicircle/MP) describe the macroscopic eigenvalue histogram.
- Edge behavior refines the location and fluctuations of the extreme eigenvalues (e.g. $\lambda_{\max} \approx 2$ for Wigner and $\lambda_{\max} \approx (1 + \sqrt{r})^2$ for MP).
- Local statistics describe spacing and fine-scale fluctuations in the bulk and at the edge, often leading to Tracy–Widom and sine-kernel phenomena.

6 Look ahead

In this lecture we moved from non-asymptotic operator norm bounds to a global description of the entire spectrum. The semicircle law explains why Wigner noise has a compact spectral cloud after $1/\sqrt{n}$ scaling, and the Marchenko–Pastur law explains why sample covariance eigenvalues spread nontrivially when d/n is not tiny. Next, we will sharpen this picture by studying *edge statistics* and extreme eigenvalues, and we will connect these spectral laws back to statistical questions (PCA, signal-plus-noise models, and when spikes separate from the bulk).

Source material

Parts of this lecture are based on references: Vershynin (2018), in addition to the author's accumulated experience working on related topics.

References

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.