

Universal Spectral Laws II: Local 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

Last time, we discussed *bulk* spectral laws: for Wigner matrices, the empirical eigenvalue distribution converges to the semicircle law; for sample covariance matrices, it converges to the Marchenko–Pastur law. These results are striking because they are *universal* and *dimension-free in shape*: after the right normalization, the histogram of eigenvalues converges to a deterministic curve.

But bulk laws deliberately ignore the quantities that drive many statistical and algorithmic tasks:

- the spectral norm $\|A\| = \lambda_{\max}(A)$;
- the smallest eigenvalue (conditioning or invertibility);
- the top eigenspace (PCA and signal recovery);
- the microscopic spacing between eigenvalues (repulsion, local statistics).

All of these are governed by *edge* and *local* spectral statistics.

In this lecture we focus on three themes:

1. *Edge location*: the largest and smallest eigenvalues converge to the edges of the bulk support.
2. *Edge fluctuations*: after centering and scaling, extreme eigenvalues converge to the Tracy–Widom distribution (in the Gaussian case, and more generally by universality).
3. *Joint laws and repulsion*: in Gaussian ensembles the eigenvalues have an explicit joint density with a Vandermonde factor, revealing the “log-gas” structure and explaining repulsion.

We then briefly connect these phenomena to nuclear physics and analytic number theory.

2 Bulk versus edge

Let A be an $n \times n$ symmetric matrix with real eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_n(A)$.

The *empirical spectral distribution* (ESD) is

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

Bulk laws describe μ_A as $n \rightarrow \infty$. Edge laws instead describe

$$\lambda_{\max}(A) = \lambda_1(A), \quad \lambda_{\min}(A) = \lambda_n(A),$$

and (more delicately) their fluctuations around the deterministic edge.

A useful mental picture is:

bulk \sim “law of large numbers for all eigenvalues”, edge \sim “extreme value theory for a strongly dependent system”

The key twist is that eigenvalues are *not* independent; they interact strongly, and this interaction is what produces repulsion and Tracy–Widom fluctuations.

3 Extreme eigenvalues

3.1 Wigner matrices

A standard Wigner model is

$$H_n := \frac{1}{\sqrt{n}} W_n,$$

where W_n is symmetric, $(W_{ij})_{i < j}$ are independent with mean 0 and variance 1, and the diagonal has mean 0 and variance 2 (this normalization matches GOE in the Gaussian case). Then the semicircle law says $\mu_{H_n} \Rightarrow \rho_{sc}$ supported on $[-2, 2]$.

A bulk law does not automatically imply that $\lambda_{\max}(H_n) \rightarrow 2$, but in fact this is true under mild moment assumptions.

Theorem 3.1 (Bai–Yin law, informal version). *For a broad class of Wigner matrices (e.g. independent entries with mean 0, variance 1, and a finite fourth moment),*

$$\lambda_{\max}(H_n) \rightarrow 2 \quad a.s., \quad \lambda_{\min}(H_n) \rightarrow -2 \quad a.s.$$

3.2 Wishart matrices

Let $X \in \mathbb{R}^{d \times n}$ have i.i.d. entries with mean 0 and variance 1, and define the sample covariance

$$\Sigma_{d,n} := \frac{1}{n} X X^\top \in \mathbb{R}^{d \times d}.$$

Assume $d/n \rightarrow r \in (0, 1)$. The Marchenko–Pastur law says the bulk ESD converges to a compactly supported density on

$$[(1 - \sqrt{r})^2, (1 + \sqrt{r})^2].$$

The edge law says the extreme eigenvalues converge to these endpoints.

Theorem 3.2 (Wishart edge law, informal version). *Under standard moment assumptions (e.g. finite fourth moment),*

$$\lambda_{\max}(\Sigma_{d,n}) \rightarrow (1 + \sqrt{r})^2 \quad a.s., \quad \lambda_{\min}(\Sigma_{d,n}) \rightarrow (1 - \sqrt{r})^2 \quad a.s.$$

(When $r > 1$, the lower edge behavior changes because $\Sigma_{d,n}$ has rank $\leq n$ and develops zero eigenvalues.)

Even when the population covariance is I_d , the sample covariance has a nontrivial spectrum. The largest eigenvalue does not concentrate near 1; it concentrates near $(1 + \sqrt{r})^2$. This is the baseline “noise level” against which spikes must be detected.

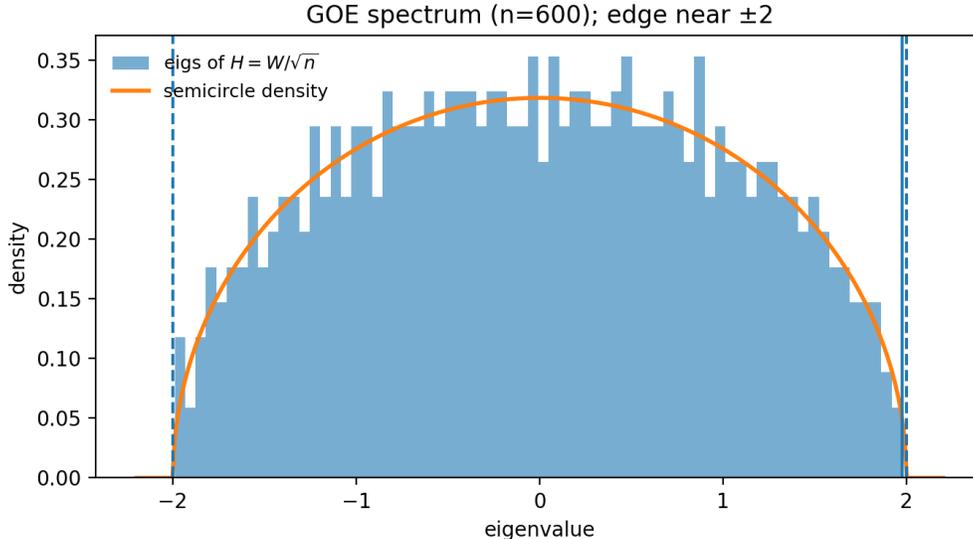


Figure 1: A GOE realization: histogram of eigenvalues of $H = W/\sqrt{n}$ (here $n = 600$), overlaid with the semicircle density. The dashed lines mark ± 2 ; the solid vertical line marks the realized λ_{\max} .

4 Fluctuations at the edge

Edge laws identify the deterministic limit of λ_{\max} , but the next question is: *on what scale does λ_{\max} fluctuate, and what is the limiting distribution?*

A crucial phenomenon is that the top eigenvalues are separated on a scale $n^{-2/3}$ (not n^{-1} and not $n^{-1/2}$). This is already hinted at by the semicircle density: it vanishes like a square root near the edge, and this forces a different local spacing scale.

4.1 Gaussian ensembles

In the Gaussian case one can identify the limiting distribution explicitly.

Theorem 4.1 (Tracy–Widom at the GOE edge, schematic form). *Let H_n be GOE normalized so that the semicircle support is $[-2, 2]$. Then*

$$n^{2/3}(\lambda_{\max}(H_n) - 2) \xrightarrow{d} \text{TW}_1,$$

where TW_1 is the Tracy–Widom distribution for the $\beta = 1$ (real symmetric) symmetry class.

For complex Hermitian (GUE), the same statement holds with TW_2 . More generally, the symmetry class is indexed by $\beta \in \{1, 2, 4\}$.

4.2 Wishart ensembles

For sample covariance matrices, the same phenomenon occurs: after centering at the upper MP edge $b = (1 + \sqrt{r})^2$ and scaling by $n^{-2/3}$ with an r -dependent prefactor, the largest eigenvalue converges to Tracy–Widom (again, TW_1 in the real case). We will not track the exact scaling

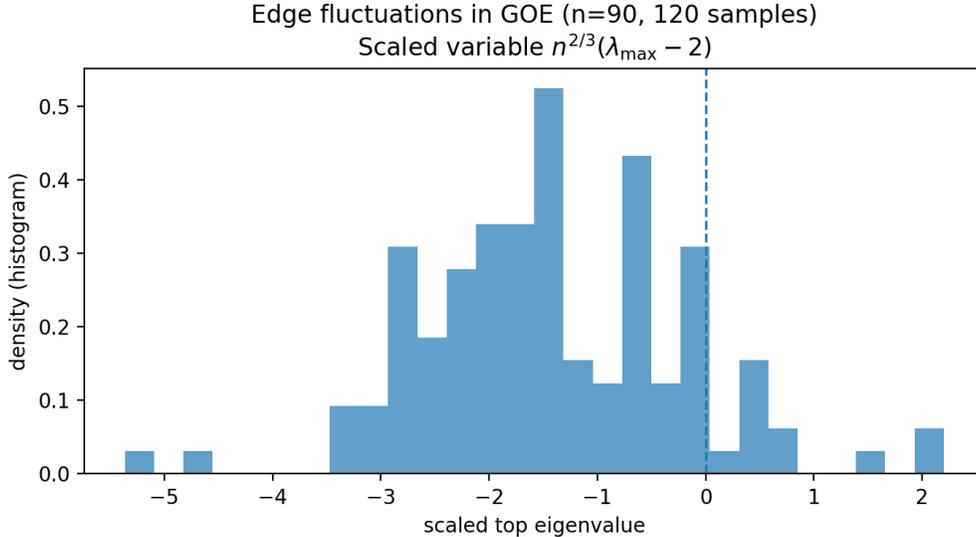


Figure 2: Simulation (GOE, $n = 90$, 120 samples): histogram of the scaled variable $n^{2/3}(\lambda_{\max} - 2)$. Theory predicts convergence in distribution to Tracy–Widom TW_1 as $n \rightarrow \infty$.

constant here; the key message is:

$$\lambda_{\max}(\Sigma_{d,n}) = (1 + \sqrt{r})^2 + \Theta(n^{-2/3}) \cdot (\text{Tracy–Widom fluctuations}).$$

In “white noise” covariance estimation, the scale of the top eigenvalue fluctuations is much smaller than the bulk width, but larger than classical $n^{-1/2}$ fluctuations. This is one reason why testing for weak spikes becomes delicate at high dimension.

5 Local statistics and eigenvalue repulsion

Bulk laws give a density curve; edge laws give the extremes. Local statistics ask: *how do eigenvalues look under a microscope?* For example, consider the gaps

$$\lambda_i - \lambda_{i+1}.$$

In the bulk, typical gaps are $\asymp 1/n$. At the edge, typical gaps are $\asymp n^{-2/3}$.

A signature of random matrix spectra is *repulsion*: eigenvalues avoid being too close, unlike i.i.d. samples where collisions are common.

A heuristic summary is:

$$\mathbb{P}\{\text{a gap is very small}\} \approx \text{const} \cdot s^\beta \quad \text{for small } s,$$

where β is the symmetry class ($\beta = 1$ GOE, $\beta = 2$ GUE, $\beta = 4$ GSE).

6 Joint eigenvalue distribution

The cleanest explanation for repulsion comes from the Gaussian ensembles, where one can write an explicit joint density for the eigenvalues.

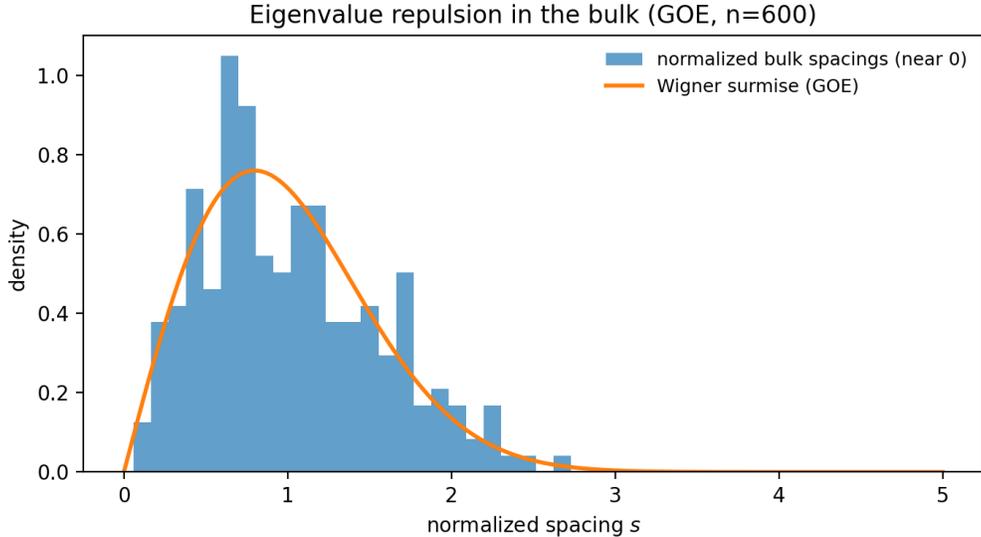


Figure 3: Bulk spacing statistics in GOE (single realization, $n = 600$). We collect eigenvalues near 0, form consecutive gaps, and normalize by the predicted mean spacing. The density is suppressed near 0, reflecting level repulsion. The curve is the classical *Wigner surmise* (an accurate approximation for GOE spacing).

In the Gaussian setting, the eigenvalues form a strongly interacting particle system.

Proposition 6.1 (Gaussian β -ensemble density, schematic). *Let H_n be a Gaussian β -ensemble normalized so that the bulk lives on $[-2, 2]$. Then the joint density of its (unordered) eigenvalues $(\lambda_1, \dots, \lambda_n)$ has the form*

$$p(\lambda_1, \dots, \lambda_n) \propto \exp\left(-\frac{\beta n}{4} \sum_{i=1}^n \lambda_i^2\right) \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta,$$

with $\beta = 1$ (GOE), $\beta = 2$ (GUE), $\beta = 4$ (GSE).

The factor

$$\prod_{i < j} |\lambda_i - \lambda_j|^\beta$$

is the Vandermonde determinant (to power β) and it *forces* eigenvalue repulsion: configurations with $\lambda_i \approx \lambda_j$ are strongly suppressed.

Taking logs suggests an energy landscape

$$\underbrace{\frac{n}{4} \sum_{i=1}^n \lambda_i^2}_{\text{confining quadratic potential}} - \underbrace{\sum_{i < j} \log |\lambda_i - \lambda_j|}_{\text{pairwise logarithmic repulsion}}.$$

So eigenvalues behave like charged particles in 1D with long-range repulsion, confined by a quadratic potential. From this viewpoint:

- the semicircle law is an “equilibrium density”;

- the sine-kernel statistics describe the local “crystal-like” structure in the bulk;
- the Airy kernel or Tracy–Widom law describe the soft edge of this gas.

7 Structured signals: BBP phase transitions

Bulk and edge laws describe “pure noise”. In statistics, we often have *signal + noise* and want to know when PCA detects signal.

7.1 Wigner matrices

Consider the spiked Wigner model

$$M_n := H_n + p u u^\top, \quad u \in S^{n-1} \text{ fixed, } p > 0,$$

where H_n is Wigner or GOE normalized to have semicircle support $[-2, 2]$.

Theorem 7.1 (BBP transition for spiked Wigner, informal). *As $n \rightarrow \infty$, almost surely,*

$$\lambda_{\max}(M_n) \rightarrow \begin{cases} 2, & p \leq 1, \\ p + \frac{1}{p}, & p > 1. \end{cases}$$

Moreover, the top eigenvector becomes correlated with u if and only if $p > 1$, with limiting overlap

$$\langle v_1(M_n), u \rangle^2 \rightarrow \max \left\{ 0, 1 - \frac{1}{p^2} \right\}.$$

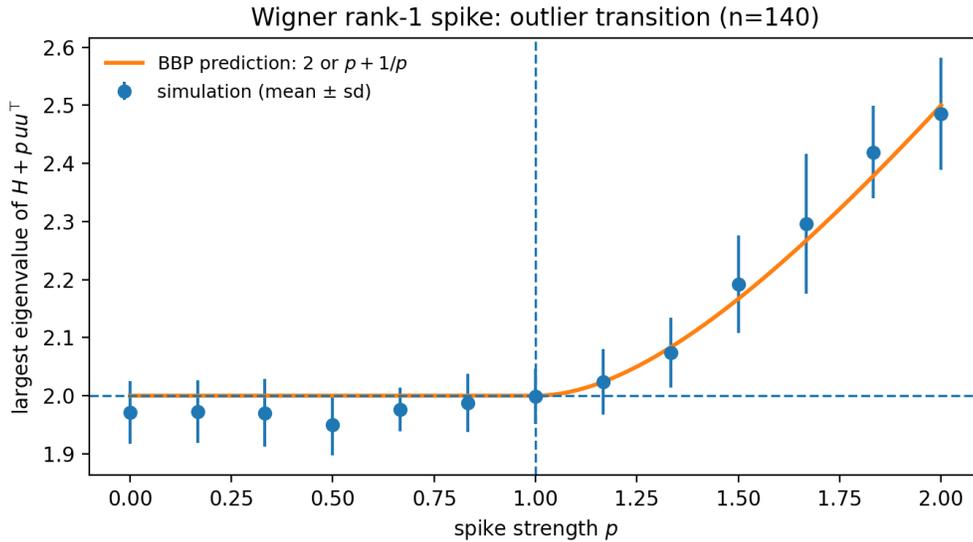


Figure 4: Spiked Wigner simulation ($n = 140$): the largest eigenvalue of $H + p u u^\top$ as a function of spike strength p , compared with the BBP prediction. The dashed line marks the threshold $p = 1$.

7.2 Wishart matrices

A parallel story holds for spiked covariance (Wishart) models. Assume the population covariance is

$$\Sigma = I_d + p u u^\top, \quad u \in S^{d-1}.$$

If $X_i \sim N(0, \Sigma)$ and $\Sigma_{d,n} = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top$ with $d/n \rightarrow r \in (0, 1)$, then there is a sharp detectability threshold:

the top eigenvalue separates from the MP bulk if and only if $p > \sqrt{r}$.

Above the threshold, the top eigenvector has nontrivial correlation with u ; below it, PCA is asymptotically uninformative.

This puts an asymptotic random-matrix-theory lens on covariance estimation: even when the covariance is exactly I_d , the top sample eigenvalue is at $(1 + \sqrt{r})^2$, and a spike must be strong enough to overcome this edge.

8 Universality

So far, we emphasized Gaussian ensembles because they admit closed-form densities. But a central phenomenon is that many of these limiting laws persist far beyond Gaussianity.

For broad families of Wigner and sample covariance matrices with independent entries and mild moment or tail assumptions, the following limits are the same as in the Gaussian case:

- bulk laws (semicircle, Marchenko–Pastur),
- edge laws (Bai–Yin type convergence of extremes),
- edge fluctuations (Tracy–Widom),
- local bulk statistics (sine-kernel or spacing laws, after unfolding).

Typically, the limit depends only on the symmetry class (β) and on coarse first or second-moment normalization, not on the fine details of the entry distribution.

A heuristic takeaway is:

Gaussian ensembles are exactly solvable; universality says they are also representative.

9 Look ahead

Bulk laws explain the global shape of the spectrum for “pure noise” matrices, but extreme eigenvalues and local spacing provide a much finer structure: edges have their own scaling, extremes fluctuate according to Tracy–Widom laws, and eigenvalues behave like a repulsive particle system. In the next lectures we will connect these asymptotic laws to non-asymptotic tools (matrix Bernstein/Chernoff and related inequalities) and to concrete statistical methods.

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.