Huang–McKenzie–Yau talk #5: edge eigenvalues in the small-t regime

Zack Stier

April 1, 2025

We will discuss the implication from Proposition 3.13 to Proposition 3.11 in Huang, McKenzie, and Yau's new paper. Proposition 3.11 says that the joint cdfs of $H(0)$ and $H(t)$ agree up to vanishing factors in N, and it is essentially half of the input towards the edge universality result, Theorem 1.2.[*] Proposition 3.13 bounds the t-derivative of the expected product of differences of two Stieltjes transforms.[†] It is not at all obvious how to go from one to the other; the paper defers this to §17 of Erdos and Yau's book, and the goal of this talk is to describe how in that section, using a result closely resembling Proposition 3.13, a result closely resembling[‡] Proposition 3.11 is obtained. In both Erdos–Yau and Huang–McKenzie–Yau, the result in question is the main way in which different random matrix models are related; most other technical results focus on refining estimates for a single random matrix model (be it the adjacency matrix itself, GOE, or the Gaussian divisble ensemble $H(t)$).

Switch now to numbering from Erdos–Yau. We want to prove **Theorem 17.1**: for nice-enough $N \times N$ random matrices A and B with eigenvalues $\alpha_1 \leqslant \cdots \leqslant \alpha_N$ and $\beta_1 \leqslant \cdots \leqslant \beta_N$, there are $\varepsilon, \delta > 0$ such that for all N sufficiently large and $s \in \mathbb{R}$,

statement of EY's main result

$$\begin{aligned}
&\mathbb{P}\left[N^{2/3}(\alpha_N - 2) \leqslant s - N^{-\varepsilon}\right] - N^{-\delta} \\
\leqslant\ &\mathbb{P}\left[N^{2/3}(\beta_N - 2) \leqslant s\right] \\
\leqslant\ &\mathbb{P}\left[N^{2/3}(\alpha_N - 2) \leqslant s + N^{-\varepsilon}\right] + N^{-\delta}. \qquad (\star)
\end{aligned}$$

The main input is **Theorem 17.4**, a Green's function comparison theorem at the edge of the spectrum. Take $F : \mathbb{R} \longrightarrow \mathbb{R}$ nice-enough.[§] The result is that, for all

loose statement of EY's main input

[*]Recall that N is the size of the graph and $H(t) = H + \sqrt{t}Z$, where H is $\frac{1}{\sqrt{d-1}}$ times the adjacency matrix of a random d-regular graph on N vertices, and $Z \sim \mathrm{GOE}(N)$ conditioned to have 0 row/column sums.

[†]One is $m_t(\cdot)$, the Stieltjes transform of $H(t)$'s empirical eigenvalue distribution (EED). The other is $m_d(\cdot, t)$, a "limiting object," namely the Stieltjes transform of the t-weighted free convolution of the Kesten–McKay law (for the graph) with the semicircular law (for the GOE).

[‡]The main difference is that Huang–McKenzie–Yau study several (a fixed number of) top/bottom eigenvalues simultaneously; for simplicity we focus only on the very top eigenvalue.

[§]The main instance will be that F is a smooth indicator on a ray.

E_1 and E_2 close enough to 2 (the edge of the spectrum), if $\mathcal{N}'_M(E_1, E_2)$ is a smoothed eigenvalue count of M on the interval,* then

$$\left|\mathbb{E}\left[F(\mathcal{N}'_A(E_1, E_2))\right] - \mathbb{E}\left[F(\mathcal{N}'_B(E_1, E_2))\right]\right|$$

is also very small. This quantifies that very similar random matrices should have similar eigenvalue counts near the edge. The relevant aspect is that $\mathcal{N}'_M$ is actually written in terms of the Stieltjes transform of M's EED. Thus, before diving more into the details of the section, we first describe the plan of going from such a result to the main Theorem and why the level of control presented is used.

proof idea

Lemma 17.2 and **Corollary 17.3** study functionals of the Stieltjes transforms m_A and m_B and relate them to eigenvalue counts in intervals near the edge, such that if they are sufficiently close, then so too are $\alpha_N = \lambda_N(A)$ and $\beta_N = \lambda_N(B)$. In particular, for a given random matrix model, **Lemma 17.2** shows that the eigenvalue count is very close (wvhp) to a smoothed version which is an integral functional of the Stieltjes transform. **Corollary 17.3** then relates the eigenvalue count at the edge to smoothed versions at very small deviations from the edge. To make these results applicable will require control on functionals for already-close Stieltjes transforms of *different* random matrix models, which is where **Theorem 17.4** enters.

It is worth elaborating on the role of **Theorem 17.4** and why it is effective at the edge.† Let $z = E + i\eta$ where $E \approx 2$ and $\eta \ll \frac{1}{N^{2/3}}$. Consider $\operatorname{im} m_M(z)$, where m_M is the Stieltjes transform of M's EED. If M has an eigenvalue within η of E then

effectiveness at the edge

$$\operatorname{im} m_M(z) = \frac{1}{N} \sum_{\lambda \in \sigma(M)} \frac{\eta}{(\lambda - E)^2 + \eta^2} \geqslant \frac{1}{2N\eta} \gg \frac{1}{N^{1/3}}$$

whereas if there is no such eigenvalue then $\operatorname{im} m_M(z) \lesssim \frac{1}{N^{1/3}}$ since the denominator terms become too large—the position at the edge is critical here, by optimal rigidity.‡ So, we can use $\operatorname{im} m_M$ as a sort of indicator for the distribution of $\lambda_N(M)$, and if we can control this quantity to precision $\frac{1}{N^{1/3}}$ then we can identify the extremal eigenvalues.

We begin with some notation and the statements of the technical results (whose

technical background

---

*Properly defined in (†).

†It is also important here that optimal rigidity for the matrix model be known. Recall that optimal rigidity says that wvhp the eigenvalues $\lambda_k(M)$ of a random matrix M differ from the "expected" positions γ_k as

$$|\lambda_k(M) - \gamma_k| \leqslant \frac{1}{N^{-a+2/3} \min\{k, N+1-k\}^{1/3}},$$

where γ_k is the kth quantile of M's limiting object's spectral measure.

‡The idea is that for $|\lambda - E| = O(\eta)$, the terms in the sum are $O\left(\frac{1}{N\eta}\right)$ and there are constant-many such terms; if $|\lambda - E| = \Theta(1)$ then the terms are $O\left(\frac{\eta}{N}\right)$ and there are linearly-many such terms.

Compare this analysis with the situation in the bulk: the critical scale for η becomes $\frac{1}{N}$ since that is the eigenvalue spacing and the order of $\operatorname{im} m_M$ becomes 1. Thus it is easier to "flag" eigenvalues near the edge.

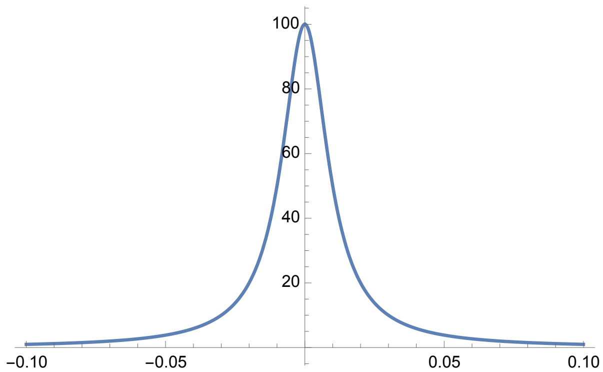

Figure 1: $\theta_{0.01}(x)$ for $-0.1 \leqslant x \leqslant 0.1$.

proofs we do not give). Let $\mathcal{N}_M(E_1, E_2)$ count M's eigenvalues lying in $[E_1, E_2]$. Let $E_+ = 2 + \frac{1}{N^{\varepsilon+2/3}}$. If M obeys optimal rigidity, then wvhp

$$\mathcal{N}_M(E, \infty) = \mathcal{N}_M(E, E_+)$$

so letting $\mathcal{X}_E$ be the indicator on $[E, E+]$ (for $E \approx 2$) we are interested in $\operatorname{tr} \mathcal{X}_E(M)$ to count eigenvalues at the upper edge. However, $\mathcal{X}_E$ is not smooth, so letting

$$\theta_\eta(x) := \operatorname{im} \frac{1}{x - i\eta}$$

we prefer $\mathcal{X}_E * \theta_\eta$; see Figure 1. We first must check that this smoothing does not substantially change the eigenvalue count:

Lemma 17.2

Let $\ell = \frac{1}{N^{3\varepsilon+2/3}}$ and $\eta = \frac{1}{N^{9\varepsilon+2/3}}$. Then for all N sufficiently large and E within $\frac{3}{2}\frac{1}{N^{-\varepsilon+2/3}}$ of 2,

$$\mathbb{P}\left[\left|\mathcal{N}_M(E, \infty) - \operatorname{tr}(\mathcal{X}_E * \theta_\eta)(M)\right| \lesssim \left(\frac{1}{N^{2\varepsilon}} + \mathcal{N}_M(E-\ell, E+\ell)\right)\right] \geqslant 1 - \frac{1}{\operatorname{poly} N}$$

This is complemented by studying $\mathcal{N}_M(E-\ell, E+\ell)$. Here we work with any smooth function F_0 which is 1 on $(-\infty, 1/9]$ and 0 on $[2/9, \infty)$, the idea being that $\frac{1}{9}$ is "close enough" to 0 for the smoothed eigenvalue count.

Corollary 17.3

Let $\ell' = \frac{1}{2}\frac{1}{N^{\varepsilon+2/3}}$. Then for all N sufficiently large and E within $\frac{1}{N^{-\varepsilon+2/3}}$ of 2,

$$\mathbb{P}\left[\operatorname{tr}(\mathcal{X}_{E+\ell'} * \theta_\eta)(M) - \frac{1}{N^\varepsilon} \leqslant \mathcal{N}_M(E,\infty) \leqslant \operatorname{tr}(\mathcal{X}_{E-\ell'} * \theta_\eta)(M) + \frac{1}{N^\varepsilon}\right] \geqslant 1 - \frac{1}{\operatorname{poly} N}.$$

Moreover,

$$\begin{aligned}
&\mathbb{E}\left[F_0\left(\operatorname{tr}(\mathcal{X}_{E-\ell'} * \theta_\eta)(M)\right)\right] \\
&\leqslant \mathbb{P}[\mathcal{N}_M(E,\infty) = 0] \\
&\leqslant \mathbb{E}\left[F_0\left(\operatorname{tr}(\mathcal{X}_{E+\ell'} * \theta_\eta)(M)\right)\right] + O\left(\frac{1}{\operatorname{poly} N}\right)
\end{aligned} \tag{17.6}$$

The proof of this result essentially goes by relating the exact eigenvalue count to the smoothed one by **Lemma 17.2**, using optimal rigidity, and the structure of F_0 as an indicator on all but a small, bounded window.

We then use the exact form

$$\operatorname{tr}(\mathcal{X}_E * \theta_\eta)(M) = N \int_E^{E_+} \operatorname{im} m_M(y + i\eta)\, \mathrm{d}y \tag{†}$$

to relate the distribution of $\lambda_N(M)$ to m_M:

Theorem 17.4

Let $\eta = \frac{1}{N^{\varepsilon+2/3}}$. For F smooth with bounded first three derivatives, and $\varepsilon > 0$ sufficiently small, there is a constant C independent of N such that if E_1 and E_2 are both within $\frac{C}{N^{-\varepsilon+2/3}}$ of 2, we have that

$$\left|\mathbb{E}\left[F\left(N\int_{E_1}^{E_2} \operatorname{im} m_A(y+i\eta)\,\mathrm{d}y\right)\right] - \mathbb{E}\left[F\left(N\int_{E_1}^{E_2} \operatorname{im} m_B(y+i\eta)\,\mathrm{d}y\right)\right]\right| \leqslant \frac{1}{\operatorname{poly} N}. \tag{17.10}$$

There are a few comments to make about this fact.

- Observe that in the argument of F above, "$N\int$" provides a factor of order $N^{\varepsilon+1/3}$ which compensates for the size of the Stieltjes transforms, i.e. the arguments of F are approximately constant-sized.
- The result is stated for a very general class of F, but we only need it for the specific choice used in **Corollary 17.3**.
- We remark that this result is crucial since it is essentially the only time in this proof that A and B will "meet"—everything else is focused on the study of a single random matrix model, not two distinct but similar models.

- While **Theorem 17.4** is proved in Erdos–Yau, we take care to omit its proof as subsequent talks in the seminar will focus on the techniques needed in Huang–McKenzie–Yau for the analogous result (in contrast to **Lemma 17.2** and **Corollary 17.3**, which are omitted for pedagogical reasons).*

With these three preliminaries in hand, we prove ($\star$).

proof of EY's main result

Proof (Theorem 17.1; ($\star$)). From optimal rigidity at the edge, wvhp, α_N and β_N are within $\frac{1}{N^{-\varepsilon+2/3}}$ of 2, and $\mathcal{N}\left(2-\frac{C}{N^{-\varepsilon+2/3}}, 2+\frac{C}{N^{-\varepsilon+2/3}}\right) \leqslant CN^{\varepsilon}$. Thus, if $|s| > N^{\varepsilon}$, ($\star$) consists of three trivial quantities. Suppose not, then, and let $E = 2 + \frac{s}{N^{2/3}}$. We begin by proving the lower bound of ($\star$).

To do this, we simply chain three inequalities:

$$\begin{aligned}
\mathbb{P}[\mathcal{N}_B(E,\infty)=0] &\geqslant \mathbb{E}\left[F_0\big(\mathrm{tr}(\chi_{E-\ell} * \theta_\eta)(B)\big)\right] && \text{by (17.6) lower} \\
&\geqslant \mathbb{E}\left[F_0\big(\mathrm{tr}(\chi_{E-\ell} * \theta_\eta)(A)\big)\right] - \frac{1}{\mathrm{poly}\, N} && \text{by (17.10)} \\
&\geqslant \mathbb{P}[\mathcal{N}_A(E-2\ell,\infty)=0] - \frac{1}{\mathrm{poly}\, N} && \text{by (17.6) upper.}
\end{aligned}$$

The other inequality follows by symmetry of the models. ■

Let's revisit Huang–McKenzie–Yau in light of this. Proposition 3.13, in the case $p = 1$, says that†

$$\mathbb{E}\left[N^{1/3}(m_t(z) - m_d(z,t))\right]\Bigg|_{t=0}^{t=\frac{1}{N^{a+1/3}}} \leqslant \frac{1}{\mathrm{poly}\, N}$$

and substituting in the bounds for t and letting $T = \frac{1}{N^{a+1/3}}$ we find that the LHS is

$$\mathbb{E}\left[N^{1/3}((m_d(z,0) - m_d(z,T)) + (m_T(z) - m_0(z)))\right]$$

and the term $m_d(z,0) - m_d(z,T)$ can be readily understood since the corresponding measure for $m_d(\cdot,t)$ is a free convolution of two well-understood measures, so that the Proposition really is a statement about expected differences of Stieltjes transforms, in analogy to the case $F(x) = x$ in **Theorem 17.4**; the case of higher p corresponds to more complicated F.

This Proposition arises from showing that the derivative in T is bounded (and subsequently integrating), so we will discuss starting next time how to reach this result by controlling the leading order terms; this will simultaneously shed light on the techniques used for Stieltjes transform results required for the large-t regime presented by Izzy.

*The Erdos–Yau book focuses on *generalized Wigner matrices*, whose entries are independent. The model of interest in Huang–McKenzie–Yau importantly does not obey this.

†Assuming the random graph obeys certain overwhelmingly-high-probability properties about uniform tight bounds on Green's function entries and Stieltjes transforms; this is given as Theorem 2.14 in the paper, and is a reproduction of Theorem 4.2 of Huang and Yau.