Large Deviations for λ_{max} of Deformed Random Matrices

(I) Basics

GOE: IID entries up to symmetry
$W_{ij} \sim N\left(0, \frac{1+\delta_{ij}}{N}\right)$

Informal main thm (M. '19, Husson-M. '23): Consider either

- $GOE + D_N$, D_N deterministic ("deformed GOE")
- $Z_N D_N Z_N^T$, $Z_N \in \mathbb{R}^{M \times N}$, $\frac{M}{N} \to a$, $Z_{ij} \sim N\left(0, \frac{1}{\sqrt{M}}\right)$, D_N as above. ("deformed samp. cov.")

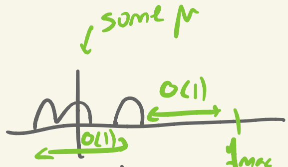
(Full rank, not BBP). In both cases, we compute explicit, simple $I : \mathbb{R} \to \mathbb{R}$ s.t.

$$\mathbb{P}(\lambda_{max} \approx x) \approx \exp(-N I(x))$$

and I is analytic except at explicit, simple x_c.

Goal today: Why is x_c interesting?

Recap on LDPs in RMT:

Given an RM whose evals typically look like 

with $\lambda_{max} \to r(\mu) :=$ right endpoint(μ) typically, ask either

① $\mathbb{P}\left(\hat{\mu}_N := \frac{1}{N} \sum \delta_{\lambda_i} \approx \nu\right) \approx \exp(-N^{\beta} J(\nu))$

speed rate fn

today

② $\mathbb{P}(\lambda_{max} \approx x) \approx \exp(-N^{\alpha} I(x))$.

(usually $\alpha < \beta$, so $I(x) = +\infty$ for $x < r(\mu)$).

Rmk: Different scaling from typical fluctuations $\mathbb{P}(\lambda_{max} \approx r(\mu) + sN^{-2/3}) \approx TW(s)$.

__Example thm__ (Ben Arous, Dembo, Guionnet '01) $\lambda_{max}(GOE)$ has $\alpha=1$,

$$I(x) = \begin{cases} +\infty & \text{if } x<2 \\ \frac{1}{2}\int_2^x \sqrt{y^2-4}\,dy & \text{if } x\geq 2 \end{cases}$$

__Rmks:__ ① "$x_c = +\infty$"

② Stieltjes transform $G(z) = \int \frac{\rho_{sc}(d\lambda)}{z-\lambda}$ satisfies $1-(z-G(z))G(z)=0$.

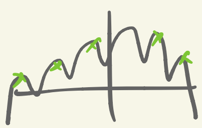
This equation has "second branch" $\overline{G}(z)$, and

$$I(x) = \frac{1}{2}\int_2^x \left[\overline{G}(y) - G(y)\right]dy$$ ← will see again later.

II Motivations

① Algorithmic motivations

Consider numerically maximizing a smooth, nonconvex, Gaussian $f_N : \mathbb{R}^N \to \mathbb{R}$.



Local minima are traps! Informally distinguish easy/hard regimes by counting them with __Kac-Rice formula__:

Need to understand Hessian at local max, specifically

$$\mathbb{E}\Big[\underbrace{|\det(\nabla^2 f_N)|}_{\text{fluctuates at same scale as LDP of } \lambda_{max}} \mathbb{1}\{\overbrace{\nabla^2 f_N \leq 0}^{\text{local max}}\} \,\Big|\, \nabla f_N = 0\Big]$$

So I can appear in count. eg. Ben Arous–Mei–Montanari–Nica '19 on "spiked tensor model" relies on Maïda '07 LDP for $\lambda_{max}(GOE + \theta e_1 e_1^T)$.

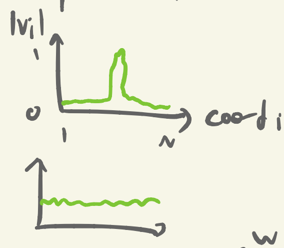
Ben Arous–Biroli–Maillard '19 (physics): Similar story, but for GLMs (some ML model), and for all crit pts instead of just local maxima.

To get local maxima, would need LDP for $\lambda_{max}(Z_N D_N Z_N^T)$ D_N full rk __with__ outliers. Maillard '20 (physics): LDP for $\lambda_{max}(Z_N D_N Z_N^T)$, D_N full rk __without__ outliers. A special case of our result makes this rigorous.

② (De)localization of evecs

Classical pbl: Find some RMT model whose evecs interpolate b/w

- localized: most mass on few coords
- delocalized: mass spread more evenly



Most studied: Random band matrices: Real sym, ind$^+$ entries,

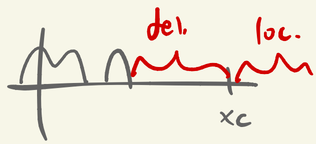
supposed (Fyodorov-Mirlin '91) to be

- localized for $W \ll N^{1/2}$ (SOTA: $W \ll N^{1/4}$, {Schenker / Cipolloni-Peled-Schenker-Shapiro '22})
- delocalized for $W \gg N^{1/2}$ (SOTA: $W \gg N^{3/4}$, Bourgade-Yang-Yau-Yin '19.

Emerging alternative: "Conditional evecs."

For some RMs, should happen that, conditioned on $\{\lambda_{max} \approx x\}$, v_{max} is

- delocalized for $x < x_c$
- localized for $x > x_c$



i.e. cheapest strategy to make $\lambda_{max} \approx x$ changes at x_c.

Examples:

① Sub-Gaussian Wigner matrices: Partial evec results ("loc. for x large", Cook-Ducatez-Guionnet '23), but limited info on what x_c should be.

② Our models: No evec results, but simple formula for candidate x_c.

①

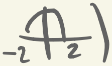
Def: Prob measure μ (centered, unit variance) is A-sub-Gaussian if

$$\int e^{tx}\,\mu(dt) \le e^{A\frac{x^2}{2}} \quad \forall x \in \mathbb{R}.$$

Taylor at zero $\Rightarrow$ must have $A \ge 1$. If $A=1$, "sharp sub-Gaussian" (SSG).

SSG includes $N(0,1)$, Bernoulli ± 1, Uniform.

Def: A Wigner(μ) matrix W_N is real sym, $W_{ij} \sim \frac{1}{\sqrt{N}}\mu$, IID up to symmetry.

whats known: (ρ_{sc} 

)

- $A=1$: LDP with same r.f. as GOE (Guionnet–Husson '18).
  "$x_c = +\infty$": All x are del. (Augeri–Guionnet–Husson '19)
- $A>1$: (certain μ) LDP with complicated I_μ (Cook–Ducatez–Guionnet '23, after AGH '19)
  - $I_\mu(x) \ge I_{GOE}(x) \quad \forall x$
  - $\exists x_\mu > 2$ s.t. $I_\mu(x) = I_{GOE}(x) \quad \forall x \in (2, x_\mu)$

    $\begin{cases} \text{For } A \in (1,2], \text{ can take } x_\mu = \sqrt{A-1} + \frac{1}{\sqrt{A-1}}. \\ \text{For } A > 2, \text{ no guess} \end{cases}$

    But $I_\mu(x) \overset{x\to\infty}{\sim} \frac{x^2}{4A}$, so $I_\mu \not\equiv I_{GOE}$
  - Loc. for x large enough.
- "$A=\infty$": For heavier tails ($\mu((s,\infty)) \sim e^{-s^\alpha}$, $0<\alpha<2$),
  "$x_c = 2$" always loc. (Augeri '16)

② Our work: Even when $A=1$, can make nontrivial x_c by deforming.

III Our Results

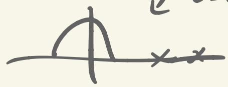
Today: Just deformed $\beta=1$ Wigner W_N (def. sample cov. analogous).

Recall: Let D_N be deterministic, diagonal. How do evals of

$$H_N = W_N + D_N$$

behave?

← BBP, not today.

- If D_N is finite rk, semicircle w/ outliers 
- If D_N is full rk, eg. $\hat{\mu}_{D_N} = \frac{1}{N}\sum \delta_{\lambda_i(D_N)} \to$ some μ_D, then not semicircle: instead "free convolution" $\rho_{sc} \boxplus \mu_D$.

Example: $D_N = \mathrm{diag}(a, \dots, a, -a, \dots, -a)$:

$a<1$: 

$a=1$: 

$a>1$: 

$\mu_D = \frac{\delta_a + \delta_{-a}}{2}$.

We assume D_N has no outliers

$\Rightarrow \lambda_{max}(H_N) \to r(\rho_{sc} \boxplus \mu_D)$ typically.

Thm: (M.'19, Husson–M. '23) $\lambda_{max}(H_N)$ satisfies an LDP, speed N,

$$I(x) = \begin{cases} +\infty \\ \frac{1}{2}\int_{r(\rho_{sc}\boxplus\mu_D)}^{x} \left[\overline{G}(y) - G_{\rho_{sc}\boxplus\mu_D}(y)\right] dy \end{cases}$$

second branch

Stieltjes

analytic except at

$$x_c = x_c(\mu_D) = \begin{cases} r(\mu_D) + G_{\mu_D}(r(\mu_D)) & \text{if finite} \\ +\infty & \text{o.w.} \end{cases}$$

Rmk: $x_c(\delta_0) = +\infty$

Pf sketch:

Classical idea: Tilt the measure

Here, good tilt (Guionnet-Husson '18) is "(rk-one) spherical integral":

For $\theta \geq 0$, determ. $B \in \mathbb{R}^{N \times N}$ sym,

exp.

$$\mathbb{E}_{e \sim \mathrm{unif}(\mathbb{S}^{N-1})}\left[\exp(N\theta\langle e, Be\rangle)\right] \overset{N\to\infty}{\sim} \exp(N J(\hat{\mu}_B, \theta, \lambda_{max}(B)))$$

Guionnet-Maïda '05

which is a special case of "HCIZ integral"

$$\int_{O_N} d\mathrm{Haar}(O) \exp(N \operatorname{Tr}(AOBO^T))$$

when $A = \mathrm{diag}(\theta, 0, \dots, 0)$.

Tilt

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}}(B) = \frac{1}{Z_\theta} \mathbb{E}_e[\exp(N\theta\langle e, Be\rangle)]$$

favors increasing λ_{max}, esp. for large θ.

We show:
- $\forall x \in (2, x_c)\ \exists \theta_x$ s.t. $\mathbb{P}^{\theta_x}(\lambda_{max} \approx x) \approx 1$.
- Estimating Z_θ: For $x < x_c$, $\mathbb{E}_e$ dominated by delocalized e.
- For $x > x_c$, no longer true.
  - We approx by models with $x_c = +\infty$.
  - CDG '23: Try to understand exactly how localized e contribute.

(IV) Open Problems

(1) Guess for undeformed x_c?

(2) Evec results for us?

(3) Extension to case where D_N has outliers (needed for ML directions)

(4) Rk-one HCIZ scales e^{cN}, gives LDP for λ_{max}, speed N.

RK-N HCIZ scales e^{cN^2}. Gives LDP for $\hat{\mu}_N$, speed N^2?

(Belinschi-Guionnet-Huang '20).

(V) Summary

- LDPs for λ_{max} of
    - $GOE + D_N$
    - $Z_N D_N Z_N^T$

  via HCIZ tilt.

- Latter confirms/extends Maillard '20.

- Simple rate fns, analytic except at simple x_c, conjecturally related to evec (de)localization.