Large deviations for the largest eigenvalue of sub-Gaussian Wigner matrices

Random matrices and related topics
Jeju Island
2024/05/09

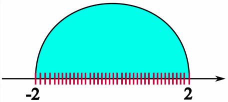
Nicholas Cook, Duke University

Based on joint work with Raphaël Ducatez and Alice Guionnet

The spectrum of Wigner matrices

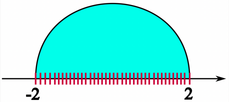
Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1 \leq i \leq j \leq N}$ iid copies of a standardized real sub-Gaussian distribution μ.

Eigenvalues $\lambda_N \leq \cdots \leq \lambda_1$.



The spectrum of Wigner matrices

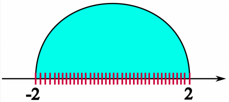
Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1 \le i \le j \le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.



Semicircle law: The ESD $\hat{\mu}_H = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i}$ concentrates around the semicircle measure $d\sigma(x) = \frac{1}{2\pi}(4 - x^2)_+^{1/2} dx$.

The spectrum of Wigner matrices

Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1\le i\le j\le N}$ iid copies of a standardized real sub-Gaussian distribution μ. Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.

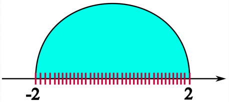


Semicircle law: The ESD $\hat{\mu}_H = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ concentrates around the semicircle measure $d\sigma(x) = \frac{1}{2\pi}(4-x^2)_+^{1/2}dx$.

* Quantitative: $\mathbb{P}(|\hat{\mu}_H(f) - \sigma(f)| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N^2)$ for f convex, 1-Lipschitz if μ has bounded support [Guionnet–Zeitouni '00]. Local law [Erdős–Schlein–Yau '08].

The spectrum of Wigner matrices

Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1\le i\le j\le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.

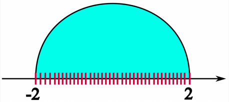


Semicircle law: The ESD $\hat{\mu}_H = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ concentrates around the semicircle measure $d\sigma(x) = \frac{1}{2\pi}(4-x^2)_+^{1/2}dx$.

* Quantitative: $\mathbb{P}(|\hat{\mu}_H(f) - \sigma(f)| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N^2)$ for f convex, 1-Lipschitz if μ has bounded support [Guionnet–Zeitouni '00].
  Local law [Erdős–Schlein–Yau '08].
* Fluctuations: $N\hat{\mu}_H(f)$ converges to a Gaussian for f smooth.
  (Optimal condition $f \in H^{1/2+\varepsilon}$ [Landon–Sosoe '22].)

The spectrum of Wigner matrices

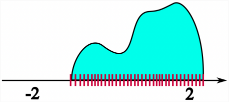
Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1\le i\le j\le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.



Semicircle law: The ESD $\hat{\mu}_H = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ concentrates around the semicircle measure $d\sigma(x) = \frac{1}{2\pi}(4-x^2)_+^{1/2}dx$.

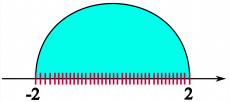
* Quantitative: $\mathbb{P}(|\hat{\mu}_H(f) - \sigma(f)| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N^2)$ for f convex, 1-Lipschitz if μ has bounded support [Guionnet–Zeitouni '00].
  Local law [Erdős–Schlein–Yau '08].
* Fluctuations: $N\hat{\mu}_H(f)$ converges to a Gaussian for f smooth.
  (Optimal condition $f \in H^{1/2+\varepsilon}$ [Landon–Sosoe '22].)
* LDP for GOE case [Ben Arous–Guionnet '97]:

$$-\frac{1}{N^2}\log \mathbb{P}(\hat{\mu}_H \sim \nu) \ \sim\ \mathcal{I}(\nu)$$



The spectrum of Wigner matrices

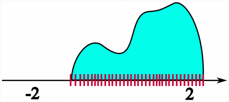
Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1 \le i \le j \le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.



Semicircle law: The ESD $\hat{\mu}_H = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ concentrates around the semicircle measure $d\sigma(x) = \frac{1}{2\pi}(4 - x^2)_+^{1/2} dx$.

* Quantitative: $\mathbb{P}(|\hat{\mu}_H(f) - \sigma(f)| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N^2)$ for f convex, 1-Lipschitz if μ has bounded support [Guionnet–Zeitouni '00]. Local law [Erdős–Schlein–Yau '08].
* Fluctuations: $N\hat{\mu}_H(f)$ converges to a Gaussian for f smooth. (Optimal condition $f \in H^{1/2+\varepsilon}$ [Landon–Sosoe '22].)

* LDP for GOE case [Ben Arous–Guionnet '97]:

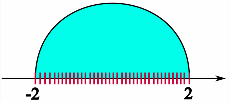


$$-\frac{1}{N^2} \log \mathbb{P}(\hat{\mu}_H \sim \nu) \ \sim \ \mathcal{I}(\nu)$$

$$= \ -\frac{1}{2}\int\int \log|x-y| d\nu(x)d\nu(y) + \frac{1}{4}\int x^2 d\nu(x) - c \, .$$

The spectrum of Wigner matrices

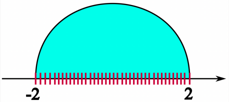
Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1 \le i \le j \le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.



Convergence at the edge: $\lambda_1 \to 2$ w.h.p. [Füredi–Komlós '81].

The spectrum of Wigner matrices

Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1 \le i \le j \le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.

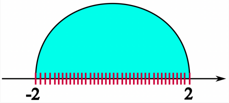


Convergence at the edge: $\lambda_1 \to 2$ w.h.p. [Füredi–Komlós '81].

* Quantitative: $\mathbb{P}(|\lambda_1 - 2| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N)$ if μ has bounded support.

The spectrum of Wigner matrices

Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1 \le i \le j \le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.

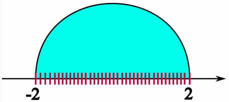


Convergence at the edge: $\lambda_1 \to 2$ w.h.p. [Füredi–Komlós '81].

- Quantitative: $\mathbb{P}(|\lambda_1 - 2| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N)$ if μ has bounded support.
- Fluctuations: $N^{2/3}(\lambda_1 - 2) \Rightarrow TW_1$. [Forrester, Tracy–Widom '94] (GOE), [Soshnikov '99]

The spectrum of Wigner matrices

Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1 \le i \le j \le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.



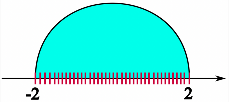
Convergence at the edge: $\lambda_1 \to 2$ w.h.p. [Füredi–Komlós '81].

* Quantitative: $\mathbb{P}(|\lambda_1 - 2| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N)$ if μ has bounded support.
* Fluctuations: $N^{2/3}(\lambda_1 - 2) \Rightarrow TW_1$. [Forrester, Tracy–Widom '94] (GOE), [Soshnikov '99]
* LDP for GOE case [Ben Arous–Dembo–Guionnet '99]:

$$-\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim \mathcal{I}^{\gamma}(x) := \begin{cases} \frac{1}{2} \int_2^x \sqrt{y^2 - 4}\, dy & x \ge 2 \\ +\infty & x < 2. \end{cases}$$

The spectrum of Wigner matrices

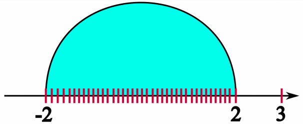
Let $H = (\frac{1}{\sqrt{N}} X_{ij})$ with $\{X_{ij}\}_{1\le i\le j\le N}$ iid copies of a standardized real sub-Gaussian distribution μ.
Eigenvalues $\lambda_N \le \cdots \le \lambda_1$.



Convergence at the edge: $\lambda_1 \to 2$ w.h.p. [Füredi–Komlós '81].

* Quantitative: $\mathbb{P}(|\lambda_1 - 2| > \varepsilon) \lesssim \exp(-c\varepsilon^2 N)$ if μ has bounded support.
* Fluctuations: $N^{2/3}(\lambda_1 - 2) \Rightarrow TW_1$. [Forrester, Tracy–Widom '94] (GOE), [Soshnikov '99]
* LDP for GOE case [Ben Arous–Dembo–Guionnet '99]:

$$-\frac{1}{N}\log \mathbb{P}(\lambda_1 \sim x) \sim \mathcal{I}^{\gamma}(x) := \begin{cases} \frac{1}{2}\int_2^x \sqrt{y^2-4}\,dy & x \ge 2 \\ +\infty & x < 2. \end{cases}$$



$\mathbb{P}(\lambda_1 \ge 3) \approx e^{-0.715N}$.

Large deviations for the spectrum: localization phenomena

Entries with stretched exponential tails:
Bordenave–Caputo '14 (ESDs), Augeri '16 (λ_1)

- Governed by the appearance of large entries.

Large deviations for the spectrum: localization phenomena

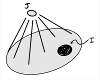
Entries with stretched exponential tails:
Bordenave–Caputo '14 (ESDs), Augeri '16 (λ_1)

- Governed by the appearance of large entries.

Bernoulli(p), $N^{-1} \ll p \ll 1$.

Augeri '18, C.–Dembo '18, Bhattacharya–Ganguly '18, Basak '21, C.–Dembo '22

- Governed by appearance of small dense structures (cliques and hubs) and high-degree vertices.
- LDP given by naïve mean-field optimization problem. Connections with extremal combinatorics (regularity method).
- Augeri–Basak '23: random sub-Gaussian weights
- Augeri '24. LDP for the ESD. Rate function supported on solutions of quadratic vector equations studied by Ajanki–Erdős–Kruger.



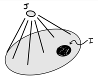
Large deviations for the spectrum: localization phenomena

Entries with stretched exponential tails:
Bordenave–Caputo '14 (ESDs), Augeri '16 (λ_1)

- Governed by the appearance of large entries.

Bernoulli(p), $N^{-1} \ll p \ll 1$.

Augeri '18, C.–Dembo '18, Bhattacharya–Ganguly '18, Basak '21, C.–Dembo '22



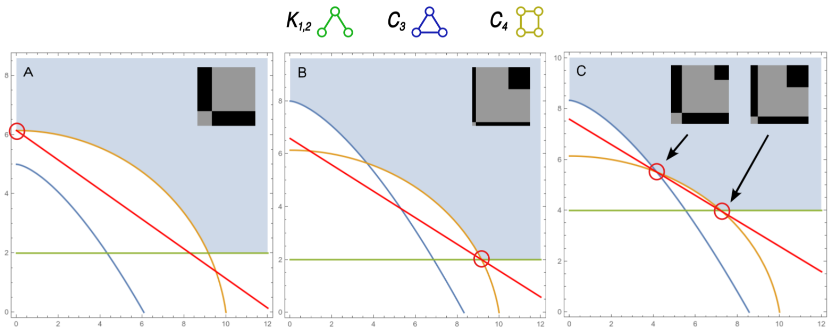
- Governed by appearance of small dense structures (cliques and hubs) and high-degree vertices.
- LDP given by naïve mean-field optimization problem. Connections with extremal combinatorics (regularity method).
- Augeri–Basak '23: random sub-Gaussian weights
- Augeri '24. LDP for the ESD. Rate function supported on solutions of quadratic vector equations studied by Ajanki–Erdős–Kruger.

Bernoulli(d/n), diluted networks.

- ESDs: Bordanave–Caputo '13
- Edge eigenvalues: Bhattacharya–Bhattacharya–Ganguly '20, Ganguly–Nam '21, Ganguly–Hiesmayr–Nam '22, Lee–Nam '23 (d-regular)
- GHN22: Wiebull-type tails $\exp(-cx^{\alpha})$. Universal Gaussian rate function for λ_1 when $\alpha > 2$, non-universal for $\alpha < 2$.

Conditional structure of sparse Bernoulli matrices (C.–Dembo ’22)

Joint upper tail events for subgraph counts (e.g. moments $\mathrm{Tr}(A^\ell)$), are dominated by the 2-parameter family of “clique-hub” matrices.



Size of clique (horizontal axis) and hub (vertical axis) are determined by a 2-dimensional relative entropy optimization problem.

Level lines for subgraph counts are green/blue/yellow curves.

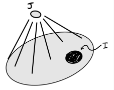
Level line for minimal relative entropy is red.

Conditional structure of sparse Bernoulli matrices on tail events

For $a, b \geq 0$, $\delta \in (0,1)$ let $\mathcal{E}_{a,b}(\delta)$ be the event that

$$\sum_{i,j \in I} A_{i,j} \geq (1-\delta)|I|^2, \qquad \sum_{i \in J, j \in J^c} A_{i,j} \geq (1-\delta)|J|(N-|J|)$$

for some $I, J \subset [N]$ with $|I| \sim \sqrt{a}pN$, $|J| \sim bp^2 N$.



Theorem (C.–Dembo '22)

For $N^{-1/3} \ll p \ll 1$ and fixed $\ell_1, \ldots, \ell_m \geq 3$, $s_1, \ldots, s_m > 0$,

$$\mathbb{P}\Bigg(\bigcup_{(a_*, b_*) \in \mathcal{O}(\underline{\ell}, \underline{s})} \mathcal{E}_{a,b}(\delta) \,\Bigg|\, \mathrm{Tr}(A^{\ell_k}) \geq (1+s_k)(Np)^{\ell_k}, \; k = 1, \ldots, m \Bigg) \geq 1 - p^{cN^2p^2}$$

for some $c(\underline{\ell}, \underline{s}, \delta) > 0$, where $\mathcal{O}(\underline{\ell}, \underline{s})$ is the set of minimizers for a non-convex linear optimization problem determined by $\underline{\ell}, \underline{s}$.

Special case of a result for any fixed collection of graphs. $\Rightarrow$ Typical structure of Exponential Random Graphs, extending [Chatterjee–Diaconis '12].

Harel–Mousset–Samotij '18: Case $m = 1$, $\ell_1 = 3$ (and general clique counts).
Basak–Basu '19: $m = 1$, general ℓ.

A universal (!) LDP (Guionnet–Husson '18)

For centered μ and $\Lambda_\mu(t) := \log \int_{\mathbb{R}} e^{tx} d\mu(x), \quad \mu$ is ...

$$\textit{sub-Gaussian (SG)} \quad \text{if} \quad \Lambda_\mu(t) \le Kt^2 \qquad \forall t \in \mathbb{R}, \text{ constant } K < \infty$$

A universal (!) LDP (Guionnet–Husson '18)

For centered μ and $\Lambda_\mu(t) := \log \int_{\mathbb{R}} e^{tx} d\mu(x)$, μ is ...

$$\textit{sub-Gaussian (SG)} \text{ if } \quad \Lambda_\mu(t) \leq Kt^2 \qquad \forall t \in \mathbb{R}, \text{ constant } K < \infty$$

$$\textit{sharp sub-Gaussian (SSG)} \text{ if } \quad \Lambda_\mu(t) \leq \tfrac{1}{2}t^2 = \Lambda_\gamma(t) \qquad \forall t \in \mathbb{R}$$

where $\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. SSG: Gaussian, Rademacher, Unif$[-\sqrt{3}, \sqrt{3}]$.

A universal (!) LDP (Guionnet–Husson '18)

For centered μ and $\Lambda_\mu(t) := \log \int_{\mathbb{R}} e^{tx} d\mu(x)$, μ is ...

$$\textit{sub-Gaussian (SG)} \text{ if } \quad \Lambda_\mu(t) \le Kt^2 \qquad \forall t \in \mathbb{R}, \text{ constant } K < \infty$$

$$\textit{sharp sub-Gaussian (SSG)} \text{ if } \quad \Lambda_\mu(t) \le \tfrac{1}{2}t^2 = \Lambda_\gamma(t) \qquad \forall t \in \mathbb{R}$$

where $\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. SSG: Gaussian, Rademacher, Unif$[-\sqrt{3}, \sqrt{3}]$.

Theorem (Guionnet–Husson '18)

Assume μ is SSG. Then $\lambda_1(H)$ satisfies an LDP with speed N and good rate function $\mathcal{I}^\gamma$. In particular, for each fixed $x \in \mathbb{R}$,

$$\frac{1}{N} \log \mathbb{P}(|\lambda_1(H) - x| \le \delta) = -\mathcal{I}^\gamma(x) + o(1).$$

(In this talk we write "$o(1)$" for errors going to zero after $N \to \infty$ then $\delta \downarrow 0$.)

A universal (!) LDP (Guionnet–Husson '18)

For centered μ and $\Lambda_\mu(t) := \log \int_{\mathbb{R}} e^{tx} d\mu(x)$, μ is ...

$$\textit{sub-Gaussian (SG)} \text{ if } \quad \Lambda_\mu(t) \le Kt^2 \qquad \forall t \in \mathbb{R}, \text{ constant } K < \infty$$

$$\textit{sharp sub-Gaussian (SSG)} \text{ if } \quad \Lambda_\mu(t) \le \tfrac{1}{2}t^2 = \Lambda_\gamma(t) \qquad \forall t \in \mathbb{R}$$

where $\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. SSG: Gaussian, Rademacher, Unif$[-\sqrt{3}, \sqrt{3}]$.

Theorem (Guionnet–Husson '18)

Assume μ is SSG. Then $\lambda_1(H)$ satisfies an LDP with speed N and good rate function $\mathcal{I}^\gamma$. In particular, for each fixed $x \in \mathbb{R}$,

$$\frac{1}{N} \log \mathbb{P}(|\lambda_1(H) - x| \le \delta) = -\mathcal{I}^\gamma(x) + o(1).$$

(In this talk we write "$o(1)$" for errors going to zero after $N \to \infty$ then $\delta \downarrow 0$.)

How does the matrix "typically" achieve $\lambda_1 \sim x$?

Proof shows it is a BBP phenomenon: on $\{\lambda_1 \sim x\}$, the mean of H is shifted by a delocalized rank-1 matrix.

A universal (!) LDP (Guionnet–Husson '18)

For centered μ and $\Lambda_\mu(t) := \log \int_{\mathbb{R}} e^{tx} d\mu(x)$, μ is ...

$$\begin{aligned} &\textit{sub-Gaussian (SG)} \text{ if } \quad \Lambda_\mu(t) \le Kt^2 \qquad \forall t \in \mathbb{R}, \text{ constant } K < \infty \\ &\textit{sharp sub-Gaussian (SSG)} \text{ if } \quad \Lambda_\mu(t) \le \tfrac{1}{2}t^2 = \Lambda_\gamma(t) \qquad \forall t \in \mathbb{R} \end{aligned}$$

where $\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. SSG: Gaussian, Rademacher, Unif$[-\sqrt{3}, \sqrt{3}]$.

Theorem (Guionnet–Husson '18)

Assume μ is SSG. Then $\lambda_1(H)$ satisfies an LDP with speed N and good rate function $\mathcal{I}^\gamma$. In particular, for each fixed $x \in \mathbb{R}$,

$$\frac{1}{N} \log \mathbb{P}(|\lambda_1(H) - x| \le \delta) = -\mathcal{I}^\gamma(x) + o(1).$$

(In this talk we write "$o(1)$" for errors going to zero after $N \to \infty$ then $\delta \downarrow 0$.)

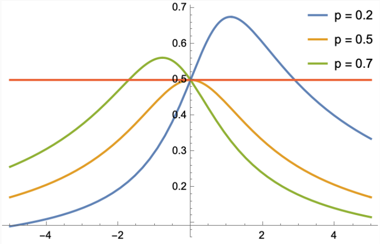
How does the matrix "typically" achieve $\lambda_1 \sim x$?

Proof shows it is a BBP phenomenon: on $\{\lambda_1 \sim x\}$, the mean of H is shifted by a delocalized rank-1 matrix.

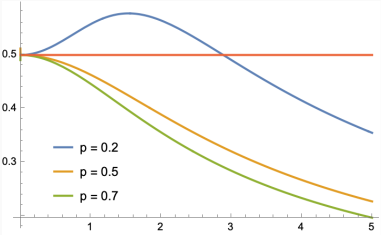
Does universal rate $\mathcal{I}^\gamma$ extend to general sub-Gaussian μ?

Partially addressed by Augeri–Guionnet–Husson '19.

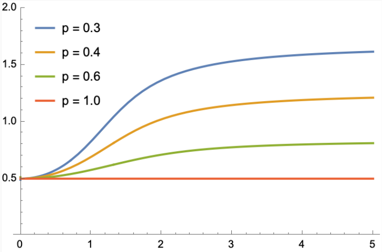
Plots of $t^{-2}\Lambda_\mu(t) = t^{-2}\log \mathbb{E}\exp(tX)$ for various sub-Gaussian μ.
$t^{-2}\Lambda_\gamma(t) \equiv \frac{1}{2}$ is in red.



Standardized Bernoulli(p)



Standardized p-sparse Rademacher

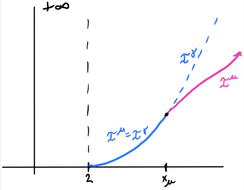


Standardized p-sparse Gaussian

General sub-Gaussian matrices: non-universal LDPs

Theorem (C.–Ducatez–Guionnet '23, Informal)

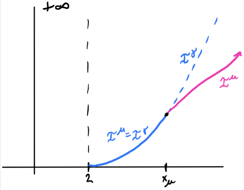
Assume μ is standardized and sub-Gaussian. Under some further technical assumptions, there exists a good rate function $\mathcal{I}^\mu$ on $\mathbb{R}$ that is infinite on $(-\infty, 2)$ and continuous and non-decreasing on $[2, \infty)$ such that $\lambda_1(H)$ satisfies an LDP with speed N and rate function $\mathcal{I}^\mu$.



General sub-Gaussian matrices: non-universal LDPs

Theorem (C.–Ducatez–Guionnet '23, Informal)

Assume μ is standardized and sub-Gaussian. Under some further technical assumptions, there exists a good rate function $\mathcal{I}^{\mu}$ on $\mathbb{R}$ that is infinite on $(-\infty, 2)$ and continuous and non-decreasing on $[2, \infty)$ such that $\lambda_1(H)$ satisfies an LDP with speed N and rate function $\mathcal{I}^{\mu}$.



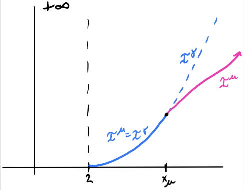
Furthermore:

- $\mathcal{I}^{\mu} \leq \mathcal{I}^{\gamma}$ *pointwise on* $\mathbb{R}$ *(large deviations are at least as likely as for GOE).*

General sub-Gaussian matrices: non-universal LDPs

Theorem (C.–Ducatez–Guionnet '23, Informal)

Assume μ is standardized and sub-Gaussian. Under some further technical assumptions, there exists a good rate function $\mathcal{I}^\mu$ on $\mathbb{R}$ that is infinite on $(-\infty, 2)$ and continuous and non-decreasing on $[2, \infty)$ such that $\lambda_1(H)$ satisfies an LDP with speed N and rate function $\mathcal{I}^\mu$.



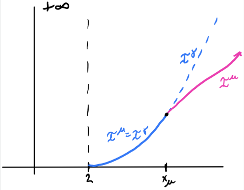
Furthermore:

- *$\mathcal{I}^\mu \leq \mathcal{I}^\gamma$ pointwise on $\mathbb{R}$ (large deviations are at least as likely as for GOE).*
- *(Universality phase). There exists $x_\mu > 2$ such that $\mathcal{I}^\mu = \mathcal{I}^\gamma$ on $(-\infty, x_\mu]$.*

General sub-Gaussian matrices: non-universal LDPs

Theorem (C.–Ducatez–Guionnet '23, Informal)

Assume μ is standardized and sub-Gaussian. Under some further technical assumptions, there exists a good rate function $\mathcal{I}^\mu$ on $\mathbb{R}$ that is infinite on $(-\infty, 2)$ and continuous and non-decreasing on $[2, \infty)$ such that $\lambda_1(H)$ satisfies an LDP with speed N and rate function $\mathcal{I}^\mu$.



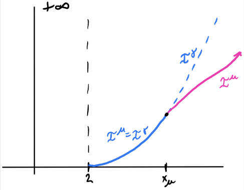
Furthermore:

- *$\mathcal{I}^\mu \leq \mathcal{I}^\gamma$ pointwise on $\mathbb{R}$ (large deviations are at least as likely as for GOE).*
- *(Universality phase). There exists $x_\mu > 2$ such that $\mathcal{I}^\mu = \mathcal{I}^\gamma$ on $(-\infty, x_\mu]$. Moreover, conditional on $\{\lambda_1 \approx x\}$ for fixed $x \in (2, x_\mu)$, with high probability the associated eigenvector v_1 is delocalized.*

General sub-Gaussian matrices: non-universal LDPs

Theorem (C.–Ducatez–Guionnet '23, Informal)

Assume μ is standardized and sub-Gaussian. Under some further technical assumptions, there exists a good rate function $\mathcal{I}^\mu$ on $\mathbb{R}$ that is infinite on $(-\infty, 2)$ and continuous and non-decreasing on $[2, \infty)$ such that $\lambda_1(H)$ satisfies an LDP with speed N and rate function $\mathcal{I}^\mu$.



Furthermore:

- *$\mathcal{I}^\mu \le \mathcal{I}^\gamma$ pointwise on $\mathbb{R}$ (large deviations are at least as likely as for GOE).*
- *(Universality phase). There exists $x_\mu > 2$ such that $\mathcal{I}^\mu = \mathcal{I}^\gamma$ on $(-\infty, x_\mu]$. Moreover, conditional on $\{\lambda_1 \approx x\}$ for fixed $x \in (2, x_\mu)$, with high probability the associated eigenvector v_1 is delocalized.*
- *(Non-universality). If μ is not SSG, then there exists $x'_\mu < \infty$ such that $\mathcal{I}^\mu < \mathcal{I}^\gamma$ on (x'_μ, ∞). Moreover, conditional on $\{\lambda_1 \approx x\}$ for fixed $x > x'_\mu$, with high probability v_1 has ℓ^2-mass $\gtrsim 1$ on coordinates of size $\ge N^{-1/4-\varepsilon}$.*

Rate function for the case $\Lambda_\mu(t)/\Lambda_\gamma(t)$ is increasing

For $\theta \geq 0, \alpha \in [0,1]$ and $x \geq 2$ let $\psi_\mu^{\text{lim}} = \lim_{t\to\infty} \Lambda_\mu(t)/t^2$,

$$\widehat{\varphi}(\theta,\alpha) := \theta^2\big[(1-\alpha)^2 + 2\psi_\mu^{\text{lim}}\alpha^2\big] + \sup_{\nu \in \mathcal{P}_{1-\alpha}(\mathbb{R})} \left\{ \int \Lambda_\mu(2\theta\alpha^{1/2}s)d\nu(s) - \mathsf{H}(\nu|\gamma) \right\} - \frac{\alpha}{2},$$

$$\widehat{\mathcal{J}}(x,\alpha) := \sup_{\theta\geq 0} \left\{ J(x,\theta) - \widehat{\varphi}\big(\theta, q_x(\theta)^2\alpha\big) \right\}.$$

Note: $\widehat{\varphi}(\theta,0) = \theta^2$ and $\widehat{\mathcal{J}}(x,0) = \sup_{\theta\geq 0}\{J(x,\theta) - \theta^2\} = \mathcal{I}^\gamma(x)$.

Rate function for the case $\Lambda_\mu(t)/\Lambda_\gamma(t)$ is increasing

For $\theta \geq 0, \alpha \in [0,1]$ and $x \geq 2$ let $\psi_\mu^{\lim} = \lim_{t\to\infty} \Lambda_\mu(t)/t^2$,

$$\widehat{\varphi}(\theta,\alpha) := \theta^2\left[(1-\alpha)^2 + 2\psi_\mu^{\lim}\alpha^2\right] + \sup_{\nu \in \mathcal{P}_{1-\alpha}(\mathbb{R})}\left\{\int \Lambda_\mu(2\theta\alpha^{1/2}s)d\nu(s) - \mathrm{H}(\nu|\gamma)\right\} - \frac{\alpha}{2},$$

$$\widehat{\mathcal{J}}(x,\alpha) := \sup_{\theta\geq 0}\left\{J(x,\theta) - \widehat{\varphi}\left(\theta, q_x(\theta)^2\alpha\right)\right\}.$$

Note: $\widehat{\varphi}(\theta,0) = \theta^2$ and $\widehat{\mathcal{J}}(x,0) = \sup_{\theta\geq 0}\{J(x,\theta) - \theta^2\} = \mathcal{I}^\gamma(x)$.

Theorem (C.–Ducatez–Guionnet '23)

Assume μ is symmetric, $\|\Lambda_\mu''\|_\infty < \infty$ and $t \mapsto \Lambda_\mu(t)/t^2$ is nondecreasing on $\mathbb{R}^+$.

(a) *λ_1 satisfies an LDP with speed N and good rate function $\mathcal{I}^\mu$ which is infinite on $(-\infty,2)$ and is otherwise given by*

$$\mathcal{I}^\mu(x) = \inf_{0\leq\alpha\leq 1-\rho_x} \widehat{\mathcal{J}}(x,\alpha)$$

with $\rho_x = c_\mu/x^4$. Moreover, for $x > 2$ the infimum is achieved on a closed nonempty set $A_x^* \subset [0, 1-\rho_x]$.

Rate function for the case $\Lambda_\mu(t)/\Lambda_\gamma(t)$ is increasing

For $\theta \geq 0, \alpha \in [0,1]$ and $x \geq 2$ let $\psi_\mu^{\lim} = \lim_{t\to\infty} \Lambda_\mu(t)/t^2$,

$$\widehat{\varphi}(\theta,\alpha) := \theta^2\left[(1-\alpha)^2 + 2\psi_\mu^{\lim}\alpha^2\right] + \sup_{\nu\in\mathcal{P}_{1-\alpha}(\mathbb{R})}\left\{\int \Lambda_\mu(2\theta\alpha^{1/2}s)d\nu(s) - \mathrm{H}(\nu|\gamma)\right\} - \frac{\alpha}{2},$$

$$\widehat{\mathcal{J}}(x,\alpha) := \sup_{\theta\geq 0}\left\{J(x,\theta) - \widehat{\varphi}\left(\theta, q_x(\theta)^2\alpha\right)\right\}.$$

Note: $\widehat{\varphi}(\theta,0) = \theta^2$ and $\widehat{\mathcal{J}}(x,0) = \sup_{\theta\geq 0}\{J(x,\theta)-\theta^2\} = \mathcal{I}^\gamma(x)$.

Theorem (C.–Ducatez–Guionnet '23)

Assume μ is symmetric, $\|\Lambda_\mu''\|_\infty < \infty$ and $t \mapsto \Lambda_\mu(t)/t^2$ is nondecreasing on $\mathbb{R}^+$.

(a) *λ_1 satisfies an LDP with speed N and good rate function $\mathcal{I}^\mu$ which is infinite on $(-\infty,2)$ and is otherwise given by*

$$\mathcal{I}^\mu(x) = \inf_{0\leq\alpha\leq 1-\rho_x} \widehat{\mathcal{J}}(x,\alpha)$$

with $\rho_x = c_\mu/x^4$. Moreover, for $x>2$ the infimum is achieved on a closed nonempty set $A_x^* \subset [0,1-\rho_x]$.

(b) *If $\Lambda_\mu(t)/t^2$ is strictly increasing on $\mathbb{R}^+$, then for any $x>2$ and $\eta,\varepsilon \in (0,\frac{1}{10})$, conditional on $|\lambda_1 - x| < \delta$ we have that with probability $1-o(1)$, v_1 is within distance ε of a vector with one entry of magnitude at least $\sqrt{\inf A_x^*}$ and all other entries bounded by $N^{-\frac{1}{2}+\eta}$.*

Joint large deviations for λ_1, v_1

Non-universal rate function comes from emergence of localized large deviation mechanisms for λ_1.

Localization will be reflected by large entries of the associated eigenvector v_1. The key idea is to get a joint LDP for λ_1 and the large entries of v_1, then contract to LDP for λ_1.

Joint large deviations for λ_1, v_1

Non-universal rate function comes from emergence of localized large deviation mechanisms for λ_1.

Localization will be reflected by large entries of the associated eigenvector v_1. The key idea is to get a joint LDP for λ_1 and the large entries of v_1, then contract to LDP for λ_1.

Roughly speaking, we get an asymptotic

$$\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim -\mathcal{I}_N^{\mu}(x)$$

where

$$\mathcal{I}_N^{\mu}(x) = \inf_z \mathcal{J}_N(x, z)$$

with the infimum taken over sparse vectors z, and $\mathcal{J}_N(x, z)$ is a joint rate function for $\{\lambda_1 \sim x, v_1^{large} \sim z\}$.

By analyzing $\mathcal{J}_N(x, z)$ around minimizing z, we obtain structure of the localized part of v_1 conditional on $\{\lambda_1 \sim x\}$.

Classical tilting: Cramér LDP

For the sample mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ for iid $X_i \sim \mu$, we have

$$\frac{1}{N}\log \mathbb{P}(|\overline{X} - x| \leq \delta) = -\Lambda_\mu^*(x) + o(1)$$

where $\Lambda_\mu^*(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_\mu(\theta)\}$.

Classical tilting: Cramér LDP

For the sample mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ for iid $X_i \sim \mu$, we have

$$\frac{1}{N}\log \mathbb{P}(|\overline{X} - x| \leq \delta) = -\Lambda_\mu^*(x) + o(1)$$

where $\Lambda_\mu^*(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_\mu(\theta)\}$.

Proof:

Classical tilting: Cramér LDP

For the sample mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ for iid $X_i \sim \mu$, we have

$$\frac{1}{N}\log \mathbb{P}(|\overline{X} - x| \leq \delta) = -\Lambda_\mu^*(x) + o(1)$$

where $\Lambda_\mu^*(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_\mu(\theta)\}$.

Proof: Defining a one-parameter family of tilted measures:

$$\mathbb{P}^{(\theta)}(\cdot) := e^{-N\Lambda_\mu(\theta)}\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(\cdot)\,, \qquad \theta \in \mathbb{R}$$

we can reexpress

$$\mathbb{P}(|\overline{X} - x| \leq \delta) = e^{-N(\theta x + o(1))}\,\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(|\overline{X} - x| \leq \delta)$$

Classical tilting: Cramér LDP

For the sample mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ for iid $X_i \sim \mu$, we have

$$\frac{1}{N}\log \mathbb{P}(|\overline{X} - x| \leq \delta) = -\Lambda_\mu^*(x) + o(1)$$

where $\Lambda_\mu^*(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_\mu(\theta)\}$.

Proof: Defining a one-parameter family of tilted measures:

$$\mathbb{P}^{(\theta)}(\cdot) := e^{-N\Lambda_\mu(\theta)}\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(\cdot)\,, \qquad \theta \in \mathbb{R}$$

we can reexpress

$$\begin{aligned}\mathbb{P}(|\overline{X} - x| \leq \delta) &= e^{-N(\theta x + o(1))}\,\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(|\overline{X} - x| \leq \delta) \\ &= e^{N(\Lambda_\mu(\theta) - \theta x + o(1))}\,\mathbb{P}^{(\theta)}(|\overline{X} - x| \leq \delta)\,.\end{aligned}$$

Classical tilting: Cramér LDP

For the sample mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ for iid $X_i \sim \mu$, we have

$$\frac{1}{N}\log \mathbb{P}(|\overline{X} - x| \leq \delta) = -\Lambda_\mu^*(x) + o(1)$$

where $\Lambda_\mu^*(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_\mu(\theta)\}$.

Proof: Defining a one-parameter family of tilted measures:

$$\mathbb{P}^{(\theta)}(\cdot) := e^{-N\Lambda_\mu(\theta)}\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(\cdot)\,, \qquad \theta \in \mathbb{R}$$

we can reexpress

$$\begin{aligned}\mathbb{P}(|\overline{X} - x| \leq \delta) &= e^{-N(\theta x + o(1))}\,\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(|\overline{X} - x| \leq \delta) \\ &= e^{N(\Lambda_\mu(\theta) - \theta x + o(1))}\,\mathbb{P}^{(\theta)}(|\overline{X} - x| \leq \delta)\,.\end{aligned}$$

Upper bound: trivially bound $\mathbb{P}^{(\theta)}(|\overline{X} - x| \leq \delta) \leq 1$ and optimize θ.

Classical tilting: Cramér LDP

For the sample mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ for iid $X_i \sim \mu$, we have

$$\frac{1}{N}\log \mathbb{P}(|\overline{X} - x| \leq \delta) = -\Lambda_\mu^*(x) + o(1)$$

where $\Lambda_\mu^*(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda_\mu(\theta)\}$.

Proof: Defining a one-parameter family of tilted measures:

$$\mathbb{P}^{(\theta)}(\cdot) := e^{-N\Lambda_\mu(\theta)}\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(\cdot)\,, \qquad \theta \in \mathbb{R}$$

we can reexpress

$$\begin{aligned}\mathbb{P}(|\overline{X} - x| \leq \delta) &= e^{-N(\theta x + o(1))}\,\mathbb{E}e^{\theta N\overline{X}}\,\mathbb{1}(|\overline{X} - x| \leq \delta)\\ &= e^{N(\Lambda_\mu(\theta) - \theta x + o(1))}\,\mathbb{P}^{(\theta)}(|\overline{X} - x| \leq \delta)\,.\end{aligned}$$

Upper bound: trivially bound $\mathbb{P}^{(\theta)}(|\overline{X} - x| \leq \delta) \leq 1$ and optimize θ.

Lower bound: show that for the optimizer θ_x, $\mathbb{P}^{(\theta_x)}(|\overline{X} - x| \leq \delta) \geq e^{-o(N)}$. □

Tilting by spherical integrals (Guionnet–Husson '18)

For $N \times N$ symmetric M, $\theta \geq 0$ and P the uniform surface measure on $\mathbb{S}^{N-1}$,

$$I(M, \theta) := \int_{\mathbb{S}^{N-1}} e^{\theta N \langle u, Mu \rangle} dP(u)$$

Tilting by spherical integrals (Guionnet–Husson '18)

For $N \times N$ symmetric M, $\theta \geq 0$ and P the uniform surface measure on $\mathbb{S}^{N-1}$,

$$I(M, \theta) := \int_{\mathbb{S}^{N-1}} e^{\theta N \langle u, Mu \rangle} dP(u)$$

Quenched free energy: (Guionnet–Maida '05). On $\mathcal{E}_x := \{\lambda_1 \sim x, \hat{\mu}_H \sim \sigma\}$,

$$\frac{1}{N} \log I(H, \theta) \sim J(x, \theta) := \begin{cases} \theta^2 & \theta \leq \frac{1}{2} G_\sigma(x) \\ \theta x - \frac{1}{2} \int \log(x - \lambda) d\sigma(\lambda) - \frac{1}{2} \log(2e\theta) & \theta \geq \frac{1}{2} G_\sigma(x) \end{cases}$$

where $G_\sigma(x) = \frac{1}{2}(x - \sqrt{x^2 - 4})$ is the Stieltjes transform of σ at $x \geq 2$.

Tilting by spherical integrals (Guionnet–Husson '18)

For $N \times N$ symmetric M, $\theta \geq 0$ and P the uniform surface measure on $\mathbb{S}^{N-1}$,

$$I(M,\theta) := \int_{\mathbb{S}^{N-1}} e^{\theta N \langle u, Mu \rangle} dP(u)$$

Quenched free energy: (Guionnet–Maida '05). On $\mathcal{E}_x := \{\lambda_1 \sim x, \hat{\mu}_H \sim \sigma\}$,

$$\frac{1}{N} \log I(H,\theta) \sim J(x,\theta) := \begin{cases} \theta^2 & \theta \leq \frac{1}{2} G_\sigma(x) \\ \theta x - \frac{1}{2} \int \log(x-\lambda) d\sigma(\lambda) - \frac{1}{2}\log(2e\theta) & \theta \geq \frac{1}{2} G_\sigma(x) \end{cases}$$

where $G_\sigma(x) = \frac{1}{2}(x - \sqrt{x^2-4})$ is the Stieltjes transform of σ at $x \geq 2$.

Annealed free energy:

$$F_N(\theta) := \frac{1}{N} \log \mathbb{E} I(H,\theta) = \frac{1}{N} \log \int_{\mathbb{S}^{N-1}} \exp\Big(\sum_{i \leq j} \Lambda_\mu(2\theta\sqrt{N} u_i u_j) \Big) dP(u)\,.$$

Defining $d\mathbb{P}^{(\theta,u)} \propto e^{\theta N \langle u, Hu \rangle} d\mathbb{P}$ and $dQ^{(\theta)}(u) \propto \mathbb{E} e^{\theta N \langle u, Hu \rangle} dP(u)$,

$$\mathbb{P}(\mathcal{E}_x) = e^{-N(J(x,\theta)+o(1))} \mathbb{E} I(H,\theta)\, \mathbb{1}(\mathcal{E}_x) = e^{N(F_N(\theta) - J(x,\theta) + o(1))} \int_{\mathbb{S}^{N-1}} \mathbb{P}^{(\theta,u)}(\mathcal{E}_x) dQ^{(\theta)}(u).$$

Tilting by spherical integrals (Guionnet–Husson '18)

For $N \times N$ symmetric M, $\theta \geq 0$ and P the uniform surface measure on $\mathbb{S}^{N-1}$,

$$I(M,\theta) := \int_{\mathbb{S}^{N-1}} e^{\theta N \langle u, Mu \rangle} dP(u)$$

Quenched free energy: (Guionnet–Maida '05). On $\mathcal{E}_x := \{\lambda_1 \sim x, \hat{\mu}_H \sim \sigma\}$,

$$\frac{1}{N} \log I(H,\theta) \sim J(x,\theta) := \begin{cases} \theta^2 & \theta \leq \frac{1}{2} G_\sigma(x) \\ \theta x - \frac{1}{2} \int \log(x-\lambda) d\sigma(\lambda) - \frac{1}{2}\log(2e\theta) & \theta \geq \frac{1}{2} G_\sigma(x) \end{cases}$$

where $G_\sigma(x) = \frac{1}{2}(x - \sqrt{x^2-4})$ is the Stieltjes transform of σ at $x \geq 2$.

Annealed free energy:

$$F_N(\theta) := \frac{1}{N} \log \mathbb{E} I(H,\theta) = \frac{1}{N} \log \int_{\mathbb{S}^{N-1}} \exp\Big(\sum_{i \leq j} \Lambda_\mu(2\theta\sqrt{N} u_i u_j)\Big) dP(u).$$

Defining $d\mathbb{P}^{(\theta,u)} \propto e^{\theta N \langle u, Hu \rangle} d\mathbb{P}$ and $dQ^{(\theta)}(u) \propto \mathbb{E} e^{\theta N \langle u, Hu\rangle} dP(u)$,

$$\mathbb{P}(\mathcal{E}_x) = e^{-N(J(x,\theta)+o(1))} \mathbb{E} I(H,\theta)\, \mathbb{1}(\mathcal{E}_x) = e^{N(F_N(\theta) - J(x,\theta) + o(1))} \int_{\mathbb{S}^{N-1}} \mathbb{P}^{(\theta,u)}(\mathcal{E}_x) dQ^{(\theta)}(u).$$

Bounding integral by 1, get $\frac{1}{N} \log \mathbb{P}(\mathcal{E}_x) \leq -\sup_{\theta \geq 0}\{J(x,\theta) - F_N(\theta)\} + o(1)$.

Tilting by spherical integrals: Annealed – Quenched

Lower bound:

Showed $$\mathbb{P}(\mathcal{E}_x) = e^{N(F_N(\theta) - J(x,\theta)) + o(N)} \int_{\mathbb{S}^{N-1}} \mathbb{P}^{(\theta,u)}(\mathcal{E}_x) dQ^{(\theta)}(u)$$

where

$$\mathcal{E}_x = \{\lambda_1 \sim x, \hat{\mu}_H \sim \sigma\}, \qquad \frac{d\mathbb{P}^{(\theta,u)}}{d\mathbb{P}} \propto e^{\theta N \langle u, Hu \rangle}, \qquad \frac{dQ^{(\theta)}}{dP}(u) \propto \mathbb{E} e^{\theta N \langle u, Hu \rangle}.$$

With θ_x the optimizing choice of θ from the upper bound, only remains to show $\{\lambda_1 \sim x\}$ is likely under $\mathbb{P}^{(\theta_x,u)}$, at least for all u in some $\mathcal{D} \subset \mathbb{S}^{N-1}$ such that $Q^{(\theta_x)}(\mathcal{D}) \geq e^{-o(N)}$.

Tilting by spherical integrals: Annealed – Quenched

Lower bound:

Showed $\quad \mathbb{P}(\mathcal{E}_x) = e^{N(F_N(\theta)-J(x,\theta))+o(N)} \int_{\mathbb{S}^{N-1}} \mathbb{P}^{(\theta,u)}(\mathcal{E}_x) dQ^{(\theta)}(u)$

where

$$\mathcal{E}_x = \{\lambda_1 \sim x, \hat{\mu}_H \sim \sigma\}, \qquad \frac{d\mathbb{P}^{(\theta,u)}}{d\mathbb{P}} \propto e^{\theta N \langle u, Hu \rangle}, \qquad \frac{dQ^{(\theta)}}{dP}(u) \propto \mathbb{E} e^{\theta N \langle u, Hu \rangle}.$$

With θ_x the optimizing choice of θ from the upper bound, only remains to show $\{\lambda_1 \sim x\}$ is likely under $\mathbb{P}^{(\theta_x,u)}$, at least for all u in some $\mathcal{D} \subset \mathbb{S}^{N-1}$ such that $Q^{(\theta_x)}(\mathcal{D}) \geq e^{-o(N)}$.

Take $\mathcal{D} = \{u \in \mathbb{S}^{N-1} : \|u\|_\infty \leq N^{-\frac{1}{4}-\varepsilon}\}$ set of delocalized unit vectors. Then

(1) $Q^{(\theta)}(\mathcal{D}) \geq e^{-o(N)}$ (easy).

Tilting by spherical integrals: Annealed – Quenched

Lower bound:

Showed $\quad \mathbb{P}(\mathcal{E}_x) = e^{N(F_N(\theta) - J(x,\theta)) + o(N)} \int_{\mathbb{S}^{N-1}} \mathbb{P}^{(\theta,u)}(\mathcal{E}_x) dQ^{(\theta)}(u)$

where

$$\mathcal{E}_x = \{\lambda_1 \sim x, \hat{\mu}_H \sim \sigma\}, \qquad \frac{d\mathbb{P}^{(\theta,u)}}{d\mathbb{P}} \propto e^{\theta N \langle u, Hu \rangle}, \qquad \frac{dQ^{(\theta)}}{dP}(u) \propto \mathbb{E} e^{\theta N \langle u, Hu \rangle}.$$

With θ_x the optimizing choice of θ from the upper bound, only remains to show $\{\lambda_1 \sim x\}$ is likely under $\mathbb{P}^{(\theta_x,u)}$, at least for all u in some $\mathcal{D} \subset \mathbb{S}^{N-1}$ such that $Q^{(\theta_x)}(\mathcal{D}) \geq e^{-o(N)}$.

Take $\mathcal{D} = \{u \in \mathbb{S}^{N-1} : \|u\|_\infty \leq N^{-\frac{1}{4}-\varepsilon}\}$ set of delocalized unit vectors. Then

(1) $Q^{(\theta)}(\mathcal{D}) \geq e^{-o(N)}$ (easy).
(2) For any $u \in \mathcal{D}$, $H \overset{d}{\approx} 2\theta uu^\mathsf{T} + \widetilde{H}$ under $\mathbb{P}^{(\theta,u)}$ for a Wigner matrix $\widetilde{H}$.

(Note that $\mathbb{E}^{(\theta,u)} H_{ij} = \frac{1}{\sqrt{N}} \Lambda'_\mu(2\theta\sqrt{N} u_i u_j) \sim 2\theta u_i u_j$.)

Lower bound:

Showed $\mathbb{P}(\mathcal{E}_x) = e^{N(F_N(\theta) - J(x,\theta)) + o(N)} \int_{\mathbb{S}^{N-1}} \mathbb{P}^{(\theta,u)}(\mathcal{E}_x) dQ^{(\theta)}(u)$

where

$$\mathcal{E}_x = \{\lambda_1 \sim x, \hat{\mu}_H \sim \sigma\}, \qquad \frac{d\mathbb{P}^{(\theta,u)}}{d\mathbb{P}} \propto e^{\theta N \langle u, Hu \rangle}, \qquad \frac{dQ^{(\theta)}}{dP}(u) \propto \mathbb{E} e^{\theta N \langle u, Hu \rangle}.$$

With θ_x the optimizing choice of θ from the upper bound, only remains to show $\{\lambda_1 \sim x\}$ is likely under $\mathbb{P}^{(\theta_x,u)}$, at least for all u in some $\mathcal{D} \subset \mathbb{S}^{N-1}$ such that $Q^{(\theta_x)}(\mathcal{D}) \geq e^{-o(N)}$.

Take $\mathcal{D} = \{u \in \mathbb{S}^{N-1} : \|u\|_\infty \leq N^{-\frac{1}{4}-\varepsilon}\}$ set of delocalized unit vectors. Then

(1) $Q^{(\theta)}(\mathcal{D}) \geq e^{-o(N)}$ (easy).

(2) For any $u \in \mathcal{D}$, $H \overset{d}{\approx} 2\theta uu^{\mathsf{T}} + \widetilde{H}$ under $\mathbb{P}^{(\theta,u)}$ for a Wigner matrix $\widetilde{H}$.

(Note that $\mathbb{E}^{(\theta,u)} H_{ij} = \frac{1}{\sqrt{N}} \Lambda'_\mu(2\theta\sqrt{N} u_i u_j) \sim 2\theta u_i u_j$.)

By the BBP transition we get $\lambda_1 \sim x$ w.h.p. under $\mathbb{P}^{(\theta_x,u)}$ for any $u \in \mathcal{D}$. □

New ideas to capture localization

For μ SSG, get: **(A)** $\frac{1}{N}\log \mathbb{P}(\lambda_1 \sim x) \sim \sup_{\theta \geq 0}\{F_N(\theta) - J(x,\theta)\}$,
(B) $F_N(\theta) \to \theta^2$.

New ideas to capture localization

For μ SSG, get: **(A)** $\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim \sup_{\theta \geq 0}\{F_N(\theta) - J(x, \theta)\}$,
(B) $F_N(\theta) \to \theta^2$.

When μ is not SSG, both (A) and (B) are false.

Heavier tails open up non-universal localization strategies that compete with delocalized tilt. Large deviations of λ_1 result from a combination of the two!

New ideas to capture localization

For μ SSG, get: **(A)** $\frac{1}{N}\log\mathbb{P}(\lambda_1 \sim x) \sim \sup_{\theta\geq 0}\{F_N(\theta) - J(x,\theta)\}$,
(B) $F_N(\theta) \to \theta^2$.

When μ is not SSG, both (A) and (B) are false.

Heavier tails open up non-universal localization strategies that compete with delocalized tilt. Large deviations of λ_1 result from a combination of the two!

The idea is to get a joint LDP for λ_1 and the large entries of v_1, then contract to LDP for λ_1.

For a sparse vector w let

$$\mathcal{U}_w := \left\{u \in \mathbb{S}^{N-1} : u^{large} \approx w,\ \|u|_{\mathrm{supp}(z)^c}\|_\infty \leq N^{-1/2+\eta}\right\}$$

and denote the *restricted annealed free energy*

$$F_N(\theta; w) := \frac{1}{N}\log\mathbb{E}\int_{\mathcal{U}_w} e^{\theta N\langle u, Hu\rangle}dP(u).$$

We obtain an explicit (but complicated) asymptotic $F_N(\theta, w) \sim \varphi_N(\theta, w)$.

New ideas to capture localization

With $\mathcal{E}_{x,z} = \{\lambda_1 \sim x, v_1^{large} \sim z\}$ and $q = q_x(\theta) = (1 - \frac{G_\sigma(x)}{2\theta})_+^{1/2}$, we can show

$$\mathbb{P}(\mathcal{E}_{x,z}) = e^{N(F_N(\theta;qz) - J(x,\theta) + o(1))} \int_{\mathcal{U}_{qz}} \mathbb{P}^{(\theta,u)}(\lambda_1 \sim x) \, \frac{dQ^{(\theta)}(u)}{Q^{(\theta)}(\mathcal{U}_{qz})} \, .$$

If we can show integral $= e^{o(N)}$ at optimizer $\theta = \theta_{x,z}$ for main term, then putting everything together,

$$\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim -\inf_z \mathcal{J}_N(x,z) \qquad \mathcal{J}_N(x,z) := \sup_{\theta \geq 0} \{J(x,\theta) - \varphi_N(\theta, qz)\}$$

New ideas to capture localization

With $\mathcal{E}_{x,z} = \{\lambda_1 \sim x, v_1^{large} \sim z\}$ and $q = q_x(\theta) = (1 - \frac{G_\sigma(x)}{2\theta})_+^{1/2}$, we can show

$$\mathbb{P}(\mathcal{E}_{x,z}) = e^{N(F_N(\theta;qz) - J(x,\theta) + o(1))} \int_{\mathcal{U}_{qz}} \mathbb{P}^{(\theta,u)}(\lambda_1 \sim x) \, \frac{dQ^{(\theta)}(u)}{Q^{(\theta)}(\mathcal{U}_{qz})} \, .$$

If we can show integral $= e^{o(N)}$ at optimizer $\theta = \theta_{x,z}$ for main term, then putting everything together,

$$\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim - \inf_z \mathcal{J}_N(x,z) \qquad \mathcal{J}_N(x,z) := \sup_{\theta \geq 0} \{J(x,\theta) - \varphi_N(\theta, qz)\}$$

Problem: As $u \in \mathcal{U}_{qz}$ are not delocalized, we can't get $\mathbb{E}^{(\theta,u)}\lambda_1$ by a BBP computation.

New ideas to capture localization

With $\mathcal{E}_{x,z} = \{\lambda_1 \sim x, v_1^{large} \sim z\}$ and $q = q_x(\theta) = (1 - \frac{G_\sigma(x)}{2\theta})_+^{1/2}$, we can show

$$\mathbb{P}(\mathcal{E}_{x,z}) = e^{N(F_N(\theta;qz) - J(x,\theta) + o(1))} \int_{\mathcal{U}_{qz}} \mathbb{P}^{(\theta,u)}(\lambda_1 \sim x) \frac{dQ^{(\theta)}(u)}{Q^{(\theta)}(\mathcal{U}_{qz})} .$$

If we can show integral $= e^{o(N)}$ at optimizer $\theta = \theta_{x,z}$ for main term, then putting everything together,

$$\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim -\inf_z \mathcal{J}_N(x,z) \qquad \mathcal{J}_N(x,z) := \sup_{\theta \geq 0}\{J(x,\theta) - \varphi_N(\theta, qz)\}$$

Problem: As $u \in \mathcal{U}_{qz}$ are not delocalized, we can't get $\mathbb{E}^{(\theta,u)}\lambda_1$ by a BBP computation.

Solution: Can show λ_1 concentrates under $\mathbb{P}^{(\theta,u)}$, with mean $\approx$ continuous in θ and u (under the ℓ^2 metric). $\mathbb{E}^{(\theta,u)}\lambda_1 \sim 2$ for small θ, $\mathbb{E}^{(\theta,u)}\lambda_1 \to \infty$ as $\theta \to \infty$.

New ideas to capture localization

With $\mathcal{E}_{x,z} = \{\lambda_1 \sim x, v_1^{large} \sim z\}$ and $q = q_x(\theta) = (1 - \frac{G_\sigma(x)}{2\theta})_+^{1/2}$, we can show

$$\mathbb{P}(\mathcal{E}_{x,z}) = e^{N(F_N(\theta;qz) - J(x,\theta) + o(1))} \int_{\mathcal{U}_{qz}} \mathbb{P}^{(\theta,u)}(\lambda_1 \sim x) \frac{dQ^{(\theta)}(u)}{Q^{(\theta)}(\mathcal{U}_{qz})}.$$

If we can show integral $= e^{o(N)}$ at optimizer $\theta = \theta_{x,z}$ for main term, then putting everything together,

$$\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim -\inf_z \mathcal{J}_N(x,z) \qquad \mathcal{J}_N(x,z) := \sup_{\theta \geq 0} \{J(x,\theta) - \varphi_N(\theta, qz)\}$$

Problem: As $u \in \mathcal{U}_{qz}$ are not delocalized, we can't get $\mathbb{E}^{(\theta,u)}\lambda_1$ by a BBP computation.

Solution: Can show λ_1 concentrates under $\mathbb{P}^{(\theta,u)}$, with mean $\approx$ continuous in θ and u (under the ℓ^2 metric). $\mathbb{E}^{(\theta,u)}\lambda_1 \sim 2$ for small θ, $\mathbb{E}^{(\theta,u)}\lambda_1 \to \infty$ as $\theta \to \infty$.

Moreover, the measures $Q^{(\theta)} 1_{\mathcal{U}_{qz}}$ concentrate on a small ball in the 2-Wasserstein metric, with center $v_{\theta,z} \in \mathbb{S}^{N-1}$ that varies continuously with θ.

New ideas to capture localization

With $\mathcal{E}_{x,z} = \{\lambda_1 \sim x, v_1^{large} \sim z\}$ and $q = q_x(\theta) = (1 - \frac{G_\sigma(x)}{2\theta})_+^{1/2}$, we can show

$$\mathbb{P}(\mathcal{E}_{x,z}) = e^{N(F_N(\theta;qz) - J(x,\theta) + o(1))} \int_{\mathcal{U}_{qz}} \mathbb{P}^{(\theta,u)}(\lambda_1 \sim x) \frac{dQ^{(\theta)}(u)}{Q^{(\theta)}(\mathcal{U}_{qz})}.$$

If we can show integral $= e^{o(N)}$ at optimizer $\theta = \theta_{x,z}$ for main term, then putting everything together,

$$\frac{1}{N} \log \mathbb{P}(\lambda_1 \sim x) \sim - \inf_z \mathcal{J}_N(x,z) \qquad \mathcal{J}_N(x,z) := \sup_{\theta \geq 0} \{J(x,\theta) - \varphi_N(\theta, qz)\}$$

Problem: As $u \in \mathcal{U}_{qz}$ are not delocalized, we can't get $\mathbb{E}^{(\theta,u)}\lambda_1$ by a BBP computation.

Solution: Can show λ_1 concentrates under $\mathbb{P}^{(\theta,u)}$, with mean $\approx$ continuous in θ and u (under the ℓ^2 metric). $\mathbb{E}^{(\theta,u)}\lambda_1 \sim 2$ for small θ, $\mathbb{E}^{(\theta,u)}\lambda_1 \to \infty$ as $\theta \to \infty$.

Moreover, the measures $Q^{(\theta)} 1_{\mathcal{U}_{qz}}$ concentrate on a small ball in the 2-Wasserstein metric, with center $v_{\theta,z} \in \mathbb{S}^{N-1}$ that varies continuously with θ.

Intermediate Value Theorem yields $\theta = \theta_{x,z}$ such that $\mathbb{E}^{(\theta, v_{\theta,z})}\lambda_1 \sim x$. □

Thanks for your attention!