Large deviations for the largest eigenvalue of random matrices

Journée des doctorants GDR MEGA

Jonathan Husson

ENS de Lyon

23 novembre 2020

We consider Wigner matrices X_N defined by :

$$X_N = \frac{1}{\sqrt{N}} \begin{pmatrix} d_1 & a_{1,2} & \cdots & a_{1,N} \\ \overline{a_{1,2}} & d_2 & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \\ \overline{a_{1,N}} & \overline{a_{2,N}} & & d_N \end{pmatrix}.$$

Where the $(a_{i,j})_{i<j}$ and $(d_i)_{i\in\mathbb{N}}$ are two independent families of i.i.d. random variables centered with values in $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. We assume $\mathbb{E}[a_{1,2}^2] = 1$ and if $\mathbb{K} = \mathbb{R}$, $\mathbb{E}[d_1^2] = 2$ and if $\mathbb{K} = \mathbb{C}$, $\mathbb{E}[d_1^2] = 1$.

We have the two classical convergence results :

Theorem (Wigner 1955)

The sequence of empirical measures $\hat{\mu}_N := \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$ *converges weakly in probably toward the semi-circular measure* σ *defined by :*

$$d\sigma(x) = \frac{\sqrt{4-x^2}}{4\pi}dx$$

Theorem (Füredi and Komlòs 1981, Bai and Yin 1988)

With the preceding notations, if the $a_{i,j}$ *and the* d_i *have symmetric laws and* $\mathbb{E}[a_{1,2}^4], \mathbb{E}[d_1^4] < +\infty$*, almost surely* $\lambda_1(X_N)$ *et* $\lambda_N(X_N)$ *converge respectively toward* -2 *and* 2*.*

Random Matrices and large deviations

Definition

We say that a sequence of random variables Z_N with values in a topological space $\mathcal{E}$ follows a large deviation principle (LDP) with speed $v(N)$ and rate fonction $I : \mathcal{E} \to \mathbb{R}^+ \cup \{+\infty\}$ if I is lower semi-continuous and if for all measurable $A \subset \mathcal{E}$:

$$
\begin{aligned}
-\inf_{x \in \mathring{A}} I(x) \leq \liminf_{N \in \mathbb{N}} \frac{1}{v(N)} \log \mathbb{P}[Z_N \in A] \leq \\
\limsup_{N \in \mathbb{N}} \frac{1}{v(N)} \log \mathbb{P}[Z_N \in A] \leq -\inf_{x \in \overline{A}} I(x)
\end{aligned}
$$

Random Matrices and large deviations

Definition

We say that a sequence of random variables Z_N with values in a topological space $\mathcal{E}$ follows a large deviation principle (LDP) with speed $v(N)$ and rate fonction $I : \mathcal{E} \to \mathbb{R}^+ \cup \{+\infty\}$ if I is lower semi-continuous and if for all measurable $A \subset \mathcal{E}$:

$$
\begin{aligned}
-\inf_{x \in \mathring{A}} I(x) \leq \liminf_{N \in \mathbb{N}} \frac{1}{v(N)} \log \mathbb{P}[Z_N \in A] \leq \\
\limsup_{N \in \mathbb{N}} \frac{1}{v(N)} \log \mathbb{P}[Z_N \in A] \leq -\inf_{x \in \overline{A}} I(x)
\end{aligned}
$$

Intuitively, it means :

$$\mathbb{P}[Z_N \approx x] \approx e^{-v(N)I(x)}$$

Gaussian ensembles

Gaussian ensembles

In the case of Wigner matrices with Gaussian entries (called GOE and GUE), the joint distribution of the eigenvalues is known explicitly :

Theorem (Mehta 1991)

Let $\beta = 1$ or 2, X_N a matrix from the GOE if $\beta = 1$ or the GUE if $\beta = 2$ whose eigenvalues are $\lambda_1 < \ldots < \lambda_N$. The law of $(\lambda_1, \ldots, \lambda_N)$ has the following density on $\mathbb{R}^N$:

$$\frac{N!}{Z_N^\beta} \mathbb{1}_{x_1 \leq \ldots \leq x_N} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^\beta \exp\left(-\frac{\beta N}{4} \sum_{i=1}^{N} x_i^2 \right)$$

From those formulas, one can deduce the following LDP for the empirical measure :

Theorem (Ben Arous and Guionnet, 1997)

If $(X_N)_{N\in\mathbb{N}}$ is a sequence of GOE or GUE matrices, their empirical measures $\hat{\mu}_N$ follow a LDP with speed N^2 and good rate function $I^{(\beta)}$ defined by :

$$I^{(\beta)}(\mu) = \frac{\beta}{2}\int \frac{x^2+y^2}{4} - \log|x-y|d\mu(x)d\mu(y) + C.$$

We also have a LDP with speed N for $\lambda_{\max}(X_N)$:

Theorem (Ben Arous, Dembo and Guionnet, 2001)

If $(X_N)_{N\in\mathbb{N}}$ is a sequence of GOE or GUE matrices, $\lambda_{\max}(X_N)$ follows a LDP with speed N and good rate function J_β defined as :

$$J_\beta(x) = \frac{\beta}{2}\int_2^x \sqrt{t^2-4}dt \ si\ x \geq 2$$
$$= +\infty \ if\ x < 2.$$

For tails heavier than Gaussian :

$$\log \mathbb{P}[|a_{i,j}| \geq t] \sim -at^{\alpha} \text{ with } a > 0 \text{ and } \alpha < 2$$

we also have LDPs for the empirical measures and the largest eigenvalue :

Theorem (Bordenave and Caputo, 2012)

The empirical measure of X_N satisfy a LDP with speed $N^{\alpha/2+1}$.

Theorem (Augeri, 2015)

The largest eigenvalue of X_N satisfy a LDP with speed $N^{\alpha/2}$.

In the general case, we just have concentration inequalities with speed N :

Theorem (Bordenave, Caputo and Chafaï 2014)

If X_N is an Hermitian random matrix such that $X_N(i) = (X_N(i,j))_{1 \leq j \leq i}$ are independent and $f : \mathbb{R} \to \mathbb{R}$ a function such that $||f||_{TV} \leq 1$ then :

$$\forall t \geq 0, \mathbb{P}\left[\frac{1}{N}\left|\sum_{i=1}^{N}(f(\lambda_i) - \mathbb{E}[f(\lambda_i)])\right| \geq t\right] \leq 2e^{-Nt^2/2}$$

With stronger hypothesis, we have the following concentration inequalities with speed N^2 for the empirical measure :

Theorem (Guionnet and Zeitouni, 2000)

If X_N is a Hermitian random matrix such that the $(\sqrt{N}X_N(i,j))_{1\leq i\leq j\leq N}$ are independent centered with bounded variance and satisfy log-Sobolev inequalities with the same constant or whose laws have their support in the same compact, we have for all f Lipschitz :

$$\mathbb{P}\left[\frac{1}{N}\left|\sum_{i=1}^{N}(f(\lambda_i)-\mathbb{E}[f(\lambda_i)])\right|\geq t\right]\leq 4\exp\left(-\frac{Ct^2N^2}{|f|_{\mathcal{L}}^2}\right)$$

Large deviations for the largest eigenvalue of Wigner matrices

Definition

If μ is a centered probability measure on $\mathbb{R}$ with finite variance we say that μ is sharp sub-Gaussian if for all $t \in \mathbb{R}$:

$$T_\mu(t) \leq \exp\left(\frac{\mu(x^2)t^2}{2}\right)$$

where $T_\mu(t) := \int_{\mathbb{R}} \exp(tx) d\mu(x)$.

Example

1. *Gaussian laws.*
2. *Rademacher laws : $\frac{1}{2}(\delta_{-p} + \delta_p)$.*
3. *The uniform law on $[0, 1]$.*

Let $X_N = \frac{1}{\sqrt{N}}(a_{i,j})_{1\leq i,j\leq N}$ be a Wigner matrix :

Hypothesis

1. *The distribution of the $(a_{i,j})_{1\leq i,j\leq N}$ are sharp sub-Gaussian.*
2. *The distribution of the $(a_{i,j})_{1\leq i,j\leq N}$ either verify a log-Sobolev inequality or have their support in the same compact.*

With this hypothesis, we have :

Theorem (Guionnet et H., 2019)

$\lambda_{\max}(X_N)$ *satisfy a LDP with speed N and with the same rate function I as in the Gaussian case :*

$$I(x) = \frac{1}{2}\int_2^x \sqrt{t^2-4}dt \text{ pour } x \geq 2$$

Spherical Integrals

To prove this theorem, we would like to evaluate the Laplace transform of the largest eigenvalue :

$$t \mapsto \mathbb{E}[\exp(Nt\lambda_{\max}(X_N))]$$

But $\lambda_{\max}$ is a complicated function of X_N
As a proxy for the exponential of $\lambda_{\max}$ we will use the following spherical integral :

$$I_N(X,\theta) = \mathbb{E}_e[\exp(N\theta\langle Xe, e\rangle)]$$

where e is a unit vector taken uniformly on the sphere $\mathbb{S}^{N-1}$. This is a special case of Harris-Chandra-Itzikson-Zuber integral

$$HCIZ(A,B) = \int_{\mathcal{O}_N} \exp(Tr(AOBO^*))dO$$

for B of rank one.

When the empirical measure of X is near μ and $\lambda_{\max}(X)$ is near x :

$$I_N(X, \theta) \approx e^{-NJ(\theta,\mu,x)} \text{ (Guionnet et Maïda, 2004)}$$

Let's prove this result with $(X_N)_{N\in\mathbb{N}}$ of the form :

$$X_N = diag(\underbrace{\eta_1, ..., \eta_1}_{\alpha_1 N}, ..., \underbrace{\eta_p, ..., \eta_p}_{\alpha_p N}, \eta_{p+1})$$

with $\eta_1 < ... < \eta_p < \eta_{p+1}$ and $\alpha_1, ..., \alpha_p > 0$. The eigenvalue distribution converges toward $\mu = \frac{1}{p}\sum_{i=1}^{p} \delta_{\eta_i}$.

With $\gamma_i = \sum_{j,\lambda_j=\eta_i} e_j^2$, we have $\sum \gamma_i = 1$ and

$$\langle e, X_N e\rangle = \sum_{i=1}^{p+1} \eta_i \gamma_i = \eta_{p+1} + \sum_{i=1}^{p} (\eta_i - \eta_{p+1})\gamma_i$$

The $\gamma_1, \ldots, \gamma_p$ follow a Dirichlet law of density :

$$Z^{-1} \gamma_1^{\frac{\alpha_1 N}{2}-1} \ldots \gamma_p^{\frac{\alpha_p N}{2}-1} \left(1 - \sum_j \gamma_j\right)^{-1/2} \mathbb{1}_{\sum \gamma_j \leq 1} \prod_j d\gamma_j$$

With $\gamma_i = \sum_{j,\lambda_j=\eta_i} e_j^2$, we have $\sum \gamma_i = 1$ and

$$\langle e, X_N e\rangle = \sum_{i=1}^{p+1} \eta_i \gamma_i = \eta_{p+1} + \sum_{i=1}^{p} (\eta_i - \eta_{p+1})\gamma_i$$

The $\gamma_1, ..., \gamma_p$ follow a Dirichlet law of density :

$$Z^{-1} \gamma_1^{\frac{\alpha_1 N}{2}-1} ... \gamma_p^{\frac{\alpha_p N}{2}-1} \left(1 - \sum_j \gamma_j\right)^{-1/2} \mathbb{1}_{\sum \gamma_j \leq 1} \prod_j d\gamma_j$$

$$\approx \exp\left(\frac{N}{2}(\alpha_1 \log \gamma_1 + ... + \alpha_p \log \gamma_p + C + o(1))\right) \mathbb{1}_{\sum \gamma_j \leq 1} \prod_j d\gamma_j$$

where $C = -(\alpha_1 \log \alpha_1 + ... + \alpha_p \log \alpha_p)/2$

And so we have :

$$\mathbb{E}[\exp(N\theta\langle e, X_N, e\rangle)] = \int \exp\Big(N(\theta f(\gamma) + I(\gamma) + o(1))\Big)\mathbb{1}_{\sum \gamma_j \leq 1} d\gamma$$

And so we have :

$$\begin{aligned}\mathbb{E}[\exp(N\theta\langle e, X_N, e\rangle)] &= \int \exp\Big(N(\theta f(\gamma) + I(\gamma) + o(1))\Big) \mathbb{1}_{\sum \gamma_j \leq 1} d\gamma \\ &\approx \exp\Big(N(\max_{\sum \gamma_i \leq 1} (\theta f(\gamma) + I(\gamma)) + o(1))\Big)\end{aligned}$$

where

$$f(\gamma) = \eta_{p+1} + \sum_{i=1}^{p} (\eta_i - \eta_{p+1})\gamma_i \text{ and } I(\gamma) = \frac{1}{2}\sum_{i=1}^{p} \alpha_i \log \gamma_i + C$$

And so we have :

$$\begin{aligned}\mathbb{E}[\exp(N\theta\langle e, X_N, e\rangle)] &= \int \exp\Big(N(\theta f(\gamma) + I(\gamma) + o(1))\Big)\mathbb{1}_{\sum \gamma_j \le 1} d\gamma \\ &\approx \exp\Big(N(\max_{\sum \gamma_i \le 1}(\theta f(\gamma) + I(\gamma)) + o(1))\Big)\end{aligned}$$

where

$$f(\gamma) = \eta_{p+1} + \sum_{i=1}^{p}(\eta_i - \eta_{p+1})\gamma_i \text{ and } I(\gamma) = \frac{1}{2}\sum_{i=1}^{p}\alpha_i \log \gamma_i + C$$

so :

$$\mathbb{E}[\exp(N\theta\langle e, X_N, e\rangle)] \approx \exp\Big(N(\max_{\sum \gamma_i \le 1}(\theta f(\gamma) + I(\gamma)) + o(1))\Big)$$

If we denote

$$G(z) = \sum_{1}^{p} \alpha_i (z - \eta_i)^{-1} = \int_{\mathbb{R}} \frac{1}{z-t} d\mu(t)$$

on $]\eta_p, +\infty[$.

If we denote

$$G(z) = \sum_{1}^{p} \alpha_i (z - \eta_i)^{-1} = \int_{\mathbb{R}} \frac{1}{z - t} d\mu(t)$$

on $]\eta_p, +\infty[$.
If $G(\eta_{p+1}) \leq 2\theta$ the limit of $N^{-1} \log \mathbb{E}[\exp(N\theta\langle e, X_N, e\rangle)]$ is :

$$2\theta\eta_{p+1} + \log 2\theta - \sum_{1}^{p} \alpha_i \log(\eta_{p+1} - \eta_i) - 1$$

If $G(\eta_{p+1}) > 2\theta$, it is :

$$2\theta G^{-1}(2\theta) + \log 2\theta - \sum_{1}^{p} \alpha_i \log(G^{-1}(2\theta) - \eta_i) - 1$$

Let us sketch the proof in the Wigner case, $\beta = 1$ and $a_{i,j}, a_{i,i} = d_i/\sqrt{2}$ have the same law. First we begin by examining the behavior of :

$$\mathbb{E}[I_N(X_N, \theta)] = \mathbb{E}_{X,e}[\exp(N\theta\langle X_N e, e\rangle)]$$

We want to determine the limit $F(\theta)$ of $N^{-1} \log \mathbb{E}[I_N(X_N, \theta)]$ if it exists.

Here the sharp sub-Gaussian character of the entries is crucial. Letting $L(t) = \log \mathbb{E}[\exp(ta_{1,2})]$, we have by fixing $e \in \mathbb{S}^{N-1}$:

Here the sharp sub-Gaussian character of the entries is crucial. Letting $L(t) = \log \mathbb{E}[\exp(ta_{1,2})]$, we have by fixing $e \in \mathbb{S}^{N-1}$:

$$
\begin{aligned}
\mathbb{E}_X[\exp(N\theta\langle e, X_N e\rangle)] &= \mathbb{E}_X[\prod_{1\le i\le j\le N} \exp(\sqrt{2}^{\mathbb{1}_{i\neq j}}\sqrt{2N}\theta a_{i,j}e_ie_j)] \\
&= \prod_{1\le i\le j\le N} \mathbb{E}[\exp(\sqrt{2}^{\mathbb{1}_{i\neq j}}\sqrt{2N}\theta a_{i,j}e_ie_j)] \\
&= \exp(\sum_{1\le i\le j\le N} L(\sqrt{2}^{\mathbb{1}_{i\neq j}}\sqrt{2N}\theta e_ie_j)) \\
&\le \exp(\sum_{1\le i\le j\le N} 2^{\mathbb{1}_{i\neq j}} N\theta^2|e_ie_j|^2) \\
&\le \exp(N\theta^2)
\end{aligned}
$$

If $|e_i| \leq N^{-1/4-\epsilon}$, we have $\sqrt{N}|e_i e_j| \leq N^{-2\epsilon}$ and

$$\mathbb{E}_X[\exp(N\theta\langle e, X_N e\rangle)] = \mathbb{E}_X[\prod_{1\leq i\leq j\leq N} \exp(\sqrt{2}^{\mathbb{1}_{i\neq j}}\sqrt{2N}\theta a_{i,j} e_i e_j)]$$
$$= \exp(\sum_{1\leq i\leq j\leq N} 2^{\mathbb{1}_{i\neq j}} N\theta^2 |e_i e_j|^2 + o(N^{-2\epsilon}))$$

For $0 < \epsilon < 1/4$, $\mathbb{P}_e[\forall i, |e_i| \leq N^{-1/4-\epsilon}] \sim 1$ and so :

$$F(\theta) = \lim_N \frac{1}{N} \log \mathbb{E}_X[I_N(X_N, \theta)] = \theta^2$$

Large deviation upper bound

$$
\begin{aligned}
\mathbb{P}[\lambda_{\max}(X_N) \approx x] &\leq \mathbb{E}[\mathbb{1}_{\lambda^N_{\max} \approx x} \frac{I_N(X_N, \theta)}{I_N(X_N, \theta)}] \\
&= \mathbb{E}[\mathbb{1}_{\lambda^N_{\max} \approx x} I_N(X_N, \theta)] e^{-N(J(\theta, \sigma, x) + o(1))} \\
&\leq e^{-N(J(\theta, \sigma, x) - F(\theta) + o(1))} \\
&\leq e^{-N(I(x) + o(1))}
\end{aligned}
$$

Large deviation upper bound

$$
\begin{aligned}
\mathbb{P}[\lambda_{\max}(X_N) \approx x] &\leq \mathbb{E}[\mathbb{1}_{\lambda_{\max}^N \approx x} \frac{I_N(X_N, \theta)}{I_N(X_N, \theta)}] \\
&= \mathbb{E}[\mathbb{1}_{\lambda_{\max}^N \approx x} I_N(X_N, \theta)] e^{-N(J(\theta,\sigma,x)+o(1))} \\
&\leq e^{-N(J(\theta,\sigma,x)-F(\theta)+o(1))} \\
&\leq e^{-N(I(x)+o(1))}
\end{aligned}
$$

where $I(x) = \sup_{\theta>0}(J(\theta, \sigma, x) - F(\theta))$.

For the lower bound, we will tilt the measure and find $\theta > 0$ such that :

$$\lim_N \frac{1}{N} \log \frac{\mathbb{E}[\mathbb{1}_{\lambda_{\max} \approx x} I_N(X_N, \theta)]}{\mathbb{E}[I_N(X_N, \theta)]} = 0$$

We can restrict ourselves to the study of the tilted measures $\mathbb{P}^{(\theta,e)}$:

$$\begin{aligned} d\mathbb{P}^{(\theta,e)}(X) &= \frac{\exp(N\theta\langle e, X_N e\rangle)}{\mathbb{E}_X[\exp(N\theta\langle e, X_N e\rangle)]} d\mathbb{P}(X) \\ &= \prod_{1 \le i \le j \le N} \exp(\sqrt{2}^{\mathbb{1}_{i \neq j}} \theta \sqrt{2N} e_i e_j a_{i,j} - L(\sqrt{2}^{\mathbb{1}_{i \neq j}} \theta \sqrt{2N} e_i e_j)) \end{aligned}$$

avec $\forall i, |e_i| \le N^{-1/4-\epsilon}$.

Under $\mathbb{P}^{(\theta,e)}$, the entries of X_N have the following expectations :

$$\mathbb{E}^{(e,\theta)}[(X_N)_{i,j}] = \sqrt{2}^{\mathbb{1}_{i=j}} \frac{L'(\sqrt{2}^{\mathbb{1}_{i\neq j}}\theta\sqrt{2N}e_i e_j)}{\sqrt{N}} = (2\theta ee^*)_{i,j} + O(N^{-2\epsilon}).$$

Under $\mathbb{P}^{(\theta,e)}$, the entries of X_N have the following expectations :

$$\mathbb{E}^{(e,\theta)}[(X_N)_{i,j}] = \sqrt{2}^{\mathbb{1}_{i=j}} \frac{L'(\sqrt{2}^{\mathbb{1}_{i\neq j}}\theta\sqrt{2N}e_i e_j)}{\sqrt{N}} = (2\theta ee^*)_{i,j} + O(N^{-2\epsilon}).$$

$$\mathrm{Var}^{e,\theta}[(X_N)_{i,j}] = 2^{\mathbb{1}_{i=j}} \frac{L''(\sqrt{2}^{\mathbb{1}_{i\neq j}}\theta\sqrt{2N}e_i e_j)}{\sqrt{N}} = \frac{2^{\mathbb{1}_{i=j}}}{\sqrt{N}} + O(N^{-2\epsilon})$$

Under $\mathbb{P}^{(\theta,e)}$, the entries of X_N have the following expectations :

$$\mathbb{E}^{(e,\theta)}[(X_N)_{i,j}] = \sqrt{2}^{\mathbb{1}_{i=j}} \frac{L'(\sqrt{2}^{\mathbb{1}_{i\neq j}}\theta\sqrt{2N}e_i e_j)}{\sqrt{N}} = (2\theta ee^*)_{i,j} + O(N^{-2\epsilon}).$$

$$\mathrm{Var}^{e,\theta}[(X_N)_{i,j}] = 2^{\mathbb{1}_{i=j}} \frac{L''(\sqrt{2}^{\mathbb{1}_{i\neq j}}\theta\sqrt{2N}e_i e_j)}{\sqrt{N}} = \frac{2^{\mathbb{1}_{i=j}}}{\sqrt{N}} + O(N^{-2\epsilon})$$

We have then $X_N = \tilde{X}_N + 2\theta ee^*$ where $\tilde{X}_N$ is a Wigner matrix.

The behavior of the largest eigenvalue of this sum is well-known (BBP transition). If $z > 2$ is an eigenvalue of $\tilde{X}_N + 2\theta ee^*$ outside the bulk then :

$$\begin{aligned} \det(\tilde{X}_N + 2\theta ee^* - z) &= 0 \\ \det(I_N + 2\theta(\tilde{X}_N - z)^{-1} ee^*) &= 0 \\ 1 + 2\theta \langle e, (\tilde{X}_N - z)^{-1} e \rangle &= 0 \\ G_\sigma(z) &= \frac{1}{2\theta} \end{aligned}$$

In order to have the lower bound for an $x > 2$, we have to choose θ such that $G_\sigma(z) = \frac{1}{2\theta}$.

With the same sharp sub-Gaussian assumptions on the entries, we can generalize the previous result to other models :

- Wishart matrices.
- Matrices with variance profiles (with some conditions on the variance profile).

We have the same universality phenomenon holds (the rate function does not depend on the law of the entries).
If we relax the sharpness of the sub-Gaussian assumption, we end up with large deviations upper and lower bounds that coincide near ∞ and near 2.

For these other results, the two crucial aspects of the proof that differ are :

For these other results, the two crucial aspects of the proof that differ are :

1. The convergence of $\frac{1}{N} \log \mathbb{E}_{X,e}[\exp(\theta N \langle X_N e, e \rangle)]$ to some $F(\theta)$.

For these other results, the two crucial aspects of the proof that differ are :

1. The convergence of $\frac{1}{N} \log \mathbb{E}_{X,e}[\exp(\theta N \langle X_N e, e \rangle)]$ to some $F(\theta)$.
2. The large deviations lower bound.

Sub-Gaussian matrices

We now relax the sub-Gaussian hypothesis and we only assume that :

$$\sup_{x \in \mathbb{R}} \frac{L(x)}{x^2} = A \in]\frac{1}{2}, +\infty[$$

The large deviation upper bound still holds :

$$\mathbb{P}[\lambda_{\max}(X_N) \approx x] \leq e^{-N\bar{I}(x)}$$

were $\bar{I}(x) = \sup(J(\theta, \sigma, x) - \bar{F}(\theta))$ with

$$\bar{F}(\theta) = \limsup_N \frac{1}{N} \mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]$$

$F(\theta)$ and large deviation lower bound

For the lower bound, we need to prove $F(\theta)$ actually exist sand is differentiable.

1. This is true for small enough θ if $A < 2$. Then $F(\theta) = \theta^2$ and we have the same lower bound than in the GOE case near $x = 2$.
2. This is true for large enough θ if $\psi : x \mapsto L(x)/x^2$ is either increasing on $\mathbb{R}^+$ or attains its maximum at unique point m_* such that $\psi''(m_*) < 0$. Then we have a lower bound for large x but $I(x) \approx x^2/4A$.

Other instances of uses of spherical integral for LDPs :

1. Largest eigenvalue of $A_N + UB_NU^*$ [Guionnet and Maïda 2019]
2. Largest eigenvalue of $X_N + D_N$ [McKenna 2019]
3. Empirical measure of the diagonal entries of UB_NU^* [Belinschi, Guionnet and Huang 2020]
4. The couple $(\lambda_{\max}, |v_1|^2)$ (where v is the eigenvector associated to $\lambda_{\max}$) for a spiked Gaussian matrix $X_N + \theta ww^*$ [Guionnet and Biroli 2019]

We can generalize the sharp sub-Gaussian results to the k largest eigenvalue by generalizing the result on the spherical integrals.

Theorem (Guionnet and H., 2020)

Let $(X_N)_N$ be a sequence of deterministic self-adjoint matrices such that, $||X|| \leq M$ the i-th largest eigenvalue converges toward λ_i for $i \leq k$ and the eigenvalue distribution converge toward μ
If $e_1, ..., e_k$ are k random vectors taken uniformly on the unit sphere and conditioned to be orthogonal and $\theta_1 \geq \theta_2 \geq .... \geq \theta_k \geq 0$, then :

$$\lim_{N \to \infty} \frac{1}{N} \log \mathbb{E}[\exp(N \sum_{i=1}^{k} \theta_i \langle e_i, X_N e_i \rangle)] = \sum_{i=1}^{k} J(\theta_i, \mu, \lambda_i)$$

As a consequence we have :

Theorem (Guionnet and H., 2020)

With the same assumptions as in the first theorem, if $\lambda_1 \geq ... \geq \lambda_k$ *are the largest eigenvalue of* X_N*, the* k*-uplet* $(\lambda_1, ..., \lambda_k)$ *satisfy a LDP with rate function* $I^{(k)}$ *defined as :*

$$I(x_1, ..., x_k) = \mathbb{1}_{2 \leq x_k \leq ... \leq x_1} \sum_{i=1}^{k} I(x)$$

Thanks for your attention.