Lecture notes on random matrix theory

Charles Bordenave

January 11, 2019

Foreword

Some history

Wishart, Von Neumann and Goldstine, Wigner, Dyson, Pastur, Girko, Voiculescu, ...

Monographs

Here are recent monographs on different topics in random matrix theory.

[Tao12]: Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[AGZ10]: Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[BS10]: Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[EY17]: László Erdős and Horng-Tzer Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.

[For10]: P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[PS11]: Leonid Pastur and Mariya Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

Notation

The vector space of $n \times n$ matrices on the field $K \in \{\mathbb{R}, \mathbb{C}\}$ is denoted by $M_n(K)$. The vector space of hermitian matrices is denoted by $H_n(K)$.

We denote by $\mathbb{P}$ and $\mathbb{E}$ the probability and the expectation of our underlying random variables.

Lecture 1

Combinatorial proof of Wigner's semicircle law

1 Wigner's semicircle Theorem

1.1 Empirical distribution of eigenvalues

Let X be an hermitian matrix in $M_n(\mathbb{C})$. Counting multiplicities, its ordered eigenvalues are denoted as

$$\lambda_1(X) \geq \cdots \geq \lambda_n(X).$$

The *empirical spectral distribution (ESD)* is

$$\mu_X = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(X)}.$$

This is a global function of the spectrum. From the spectral theorem, for any function f,

$$\int f(\lambda) d\mu_X(\lambda) = \frac{1}{n} \mathrm{Tr} f(X).$$

1.2 Wigner matrix

Consider an infinite array of complex random variables (X_{ij}) where for $1 \leq i < j$

$$X_{ij} = \bar{X}_{ji}$$

are iid with law P on $\mathbb{C}$, independent of $X_{ii}, i \geq 1$ iid with common law Q on $\mathbb{R}$.

The random matrix $X = (X_{ij})_{1 \leq i,j \leq n}$ is hermitian. This matrix is called a *Wigner matrix*. There are some important cases:

- For $i > j$, $\sqrt{2}\Re(X_{ij})$, $\sqrt{2}\Im(X_{ij})$ and X_{ii} are independent standard Gaussian variables: *Gaussian Unitary Ensemble* (GUE).

- For $i > j$, X_{ij} and $X_{ii}/\sqrt{2}$ are independent standard Gaussian variables): *Gaussian Orthogonal Ensemble* (GOE).

- For $i > j$, X_{ij} and X_{ii} are independent $\{0,1\}$-Bernoulli distribution with parameters $0 \le p \le 1$ and $0 \le q \le 1$.

1.3 Weak Convergence

Let $\mu_n, \mu \in \mathcal{P}(\mathbb{R})$, we say that μ_n converges weakly to μ if for any bounded continuous function f,

$$\int f d\mu_n \to \int f\mu$$

This weak topology is metrizable with the Lévy distance d_L and $(\mathcal{P}(\mathbb{R}), d_L)$ is a Polish space.

If μ_n is a random measure, we say that μ_n converges weakly in probability to μ if for any bounded continuous function f, in probability.

$$\int f d\mu_n \to \int f\mu$$

Similarly, we say that μ_n converges weakly to μ almost surely if almost surely, for any bounded continuous function f,

$$\int f d\mu_n \to \int f\mu.$$

Exercise 1.1. *If for any bounded continuous function f, almost surely, $\int f d\mu_n \to \int f\mu$ then μ_n converges weakly to μ almost surely (Hint : Arzela-Ascoli's Theorem).*

1.4 Wigner's semicircle Theorem

In our setting, a general form of Wigner's semicircle law is the following result.

Theorem 1.1 (Semicircle Law)**.** *Let X be a Wigner matrix. Assume that $\mathbb{E}|X_{12} - \mathbb{E}X_{12}|^2 = 1$. Set $Y = X/\sqrt{n}$. Then a.s. weakly*

$$\mu_Y \to \mu_{sc}$$

where $\mu_{sc}(dx) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{|x|\le 2}dx$

2 Method of moments

Let Z real random variable with law μ and characteristic function $F(t) = \int e^{itx} d\mu(x)$. Assume that all moments are finite: for all integers $k \geq 0$, $\int x^k d\mu(x) = \mathbb{E}Z^k = m_k < \infty$.

Definition 1.1. *The probability measure μ is uniquely characterized by its moments if it is the unique measure on $\mathbb{R}$ with moments m_k for all integers $k \geq 0$.*

From Carleman's Theorem, μ is uniquely characterized by its moments if

$$\sum_{k \geq 1} m_{2k}^{-\frac{1}{2k}} = \infty.$$

A simpler sufficient condition follows from the observation that if F is analytic in a neighborhood of 0 then it is uniquely characterized by its moments. The next lemma is characterizes the latter condition.

Lemma 1.1. *The characteristic function F is analytic in a complex neighborhood of* 0 *(or in a complex neighborhood of* $\mathbb{R}$*) if and only if*

$$\limsup_k \left(\frac{m_{2k}}{(2k)!} \right)^{\frac{1}{2k}} < \infty.$$

Proof. Hadamard's formula: the radius of convergence of $\sum_n a_n z^n$ associated to a sequence of complex numbers (a_n) is $1/r$ with $r = \limsup_n |a_n|^{1/n}$. It gives the equivalence. To show analiticity on $\mathbb{R}$, we may use Taylor expansion

$$F(t+s) = \sum_{\ell=0}^{2k-1} \frac{(it)^\ell F^{(\ell)}(s)}{\ell!} + O\Big(\frac{m_{2k}|t|^{2k}}{(2k)!} \Big).$$

□

Recall Stirling's formula, $k! \sim \sqrt{2\pi k}(k/e)^k$. For example, if μ has bounded support we have $\mu_k \leq c^k$. If μ is a subGaussian law then $\mu_k \leq (c\sqrt{k})^k$. If μ is subexponential, $\mu_k \leq (ck)^k$. In these three cases, μ is uniquely characterized by its moments.

When a law is uniquely characterized by its moments, a commonly used method to prove that a sequence of real random variables $(Z_n)_{n \geq 1}$ converges weakly to a random variable Z is to show that for all integer $k \geq 1$, $\lim_n \mathbb{E}[Z_n^k] = \mathbb{E}[Z^k]$.

Lemma 1.2 (Method of moments)**.** *Assume that the law μ is uniquely determined by its moments. If for all integers $k \geq 1$, $\lim_n \int x^k d\mu_n(x) = \int x^k d\mu$ then $\mu_n \to \mu$ weakly.*

Proof. We have $\int x^2 d\mu(x) = m_2 + o(1)$, hence (μ_n) is relatively compact. Let ν be an accumulation point, since μ is uniquely determined by its moments, it is sufficient to check that for any integer $k \geq 1$, $\int x^k d\nu(x) = m_k$. This amounts to prove that $x \mapsto x^k$ is uniformly integrable for $(\mu_n)_{n\geq 1}$.

Let us check this by hand. Let Z_n be a random variable with law μ_n and W with law ν. Since $\mathbb{E}Z_n^{2k}$ is uniformly bounded, we have $\mathbb{P}(|Z_n| > t) \leq ct^{-2k}$ and from Portemanteau Theorem $\mathbb{P}(|W| > t) \leq Ct^{-2k}$. It follows that for any $\varepsilon > 0$, there is $t \geq 1$, such that $\mathbb{E}|W|^k \mathbf{1}_{|W|\geq t} < \varepsilon$ and $\mathbb{E}|Z_n|^k \mathbf{1}_{|Z_n|\geq t} < \varepsilon$. Consider f continuous $f(x) = x^k$ on $[-t, t]$ and $f(x) = t^k$ on $\mathbb{R}\backslash[-t, t]$. We get

$$\begin{aligned}\mathbb{E}W^k &= \mathbb{E}f(W) + \mathbb{E}W^k \mathbf{1}_{|W|\geq t} - \mathbb{E}f(W)\mathbf{1}_{|W|\geq t} = \mathbb{E}f(W) + \mathbb{E}W^k \mathbf{1}_{|W|\geq t} - t^k \mathbb{P}(|W| \geq t)\\ &= \mathbb{E}Z_n^k + \mathbb{E}f(W) - \mathbb{E}f(Z_n) + r,\end{aligned}$$

With $|r| \leq 4\varepsilon$. Then, since μ_n converges weakly to ν along a subsequence, along that subsequence for n large enough, we have $|\mathbb{E}f(Z_n) - \mathbb{E}f(W)| \leq \varepsilon$. □

There are many drawbacks to this method. First, the random variable Z_n needs to have finite moments of any order for all n large enough. Secondly, the computation of moments can more tedious than analytic proofs when they are available (see exercise below).

Exercise 1.2. *Let (X_i) be an iid sequence of real random variables. Assume that the X_i's have finite moment of any order, $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = 1$. Prove by the method of moments that $(X_1 + \cdots X_n)/\sqrt{n}$ converges weakly to a Gaussian random variable (Hint: check first that the $2k$-th moment of a standard Gaussian variable is equal to the number of pairings of $\{1, \ldots, 2k\}$).*

3 Moments of Wigner's semicircle law

$$d\mu_{sc}(x) = \frac{\mathbf{1}_{|x|\leq 2}}{2\pi}\sqrt{4 - x^2}dx$$

For integer $k \geq 0$,

$$m_{2k+1} = \int x^{2k+1} d\mu_{sc}(x) = 0 \quad \text{and} \quad m_{2k} = \int x^{2k} d\mu_{sc}(x) = c_k$$

where c_k is the k-th Catalan number

$$c_k = \frac{1}{k+1}\binom{2k}{k} = \binom{2k}{k} - \binom{2k}{k+1}.$$

4 Graphs, Catalan's number and Dyck paths

Let $G = (V, E)$ be a finite graph: V is finite set and E is a set of multisets of size 2 of V (G may have loops, edges of the type $\{x, x\}$ are called loops). A tree is a graph without cycles.

Lemma 1.3. *If $G = (V, E)$ is connected then $|E| - |V| + 1 \geq 0$ with equality if and only if G is a tree.*

Proof. A *spanning tree* of a connected graph $G = (V, E)$ is a subgraph of G which is a tree and has same vertices V. Let us consider the set of subgraphs of G which are trees. This set is not empty and is partially ordered by the inclusion. Since this set is finite, it contains a maximal element for the inclusion. Such a maximal element is a tree. Moreover it contains all the vertices of G, otherwise we could add an extra edge by using the connectivity of G.

Now, if G is a tree then $|E| = |V| - 1$ follows easily by induction on $|V|$. Moreover if $T = (V, E')$ with $E' \subset E$ is the spanning tree of a finite connected graph $G = (V, E)$ then $|E| \geq |E'| = |V| - 1$ from what precedes. Finally if G is such that $|E| = |V| - 1$ then $|E| = |E'|$ and thus $G = T$. $\square$

Fix an integer $k \geq 1$. We set of *Dyck paths* of length $2k$ is the set of paths of length $2k$ from 0 to 0 with ± 1 increments and which remain non-negative:

$$\mathcal{D}_k = \{(x_0, \ldots, x_{2k}) \in \mathbb{Z}^{2k+1} : x_t \geq 0, x_0 = x_{2k+1}, x_{t+1} - x_t \in \{-1, 1\} \text{ for } t = 1, \ldots, 2k\}.$$

$\mathcal{D}_k$ is an obvious bijection with sequence of proper sequence of $2k$ parenthesis: for $t = 1, \ldots, 2k$, if $x_t = x_{t-1} + 1$ place a left parenthesis '(' and right parenthesis ')' otherwise. For example, for $k = 3$, the path $(0, 1, 0, 1, 2, 1, 0)$ gives ()(()).

Lemma 1.4. *For any integer $k \geq 1$,*

$$|\mathcal{D}_k| = c_k.$$

Proof. Let $\mathcal{W}_k^s$ be set of walks of length $2k$ from 0 to even integer $-2s$ with increments $\{-1, +1\}$, $\mathcal{W}_k^s = \{(x_0, \ldots, x_{2k}) \in \mathbb{Z}^{2k+1} : x_0 = 0, x_{2k+1} = -2s, x_{t+1} - x_t \in \{-1, 1\} \text{ for } t = 0, \ldots, 2k\}$. We have $|\mathcal{W}_k^s| = \binom{2k}{k+s}$ since such walks are made of $k - s$ times t such that $x_{t+1} - x_t = 1$ and $k + s$ times such that $x_{t+1} - x_t = -1$. We check that $|\mathcal{D}_k| = |\mathcal{W}_k^0| - |\mathcal{W}_k^1|$. This is a consequence of the reflexion principle: paths in $\mathcal{W}_k^1$ are in bijections with paths in $\mathcal{W}_k$ such that $x_t = -1$ for some t (simply reflects x_{t+s} for $s \geq 0$ on x-axis). $\square$

Exercise 1.3. *Prove the recursive equation for the Catalan's number $c_0 = 1$ (convention) and for $k \geq 1$,*

$$c_k = \sum_{\ell=0}^{k-1} c_\ell c_{k-\ell-1}. \tag{1.1}$$

Deduce that the generating function of Catalan's number $S(z) = \sum_{k=0}^{\infty} c_k z^k$ satisfies

$$S(z) = 1 + zS(z)^2. \tag{1.2}$$

5 Trace method

Recall that, for a random probability measure $\mu \in \mathcal{P}(\mathcal{X})$, we may define its expectation $\mathbb{E}\mu \in \mathcal{P}(\mathcal{X})$, as for all Borel sets B,

$$(\mathbb{E}\mu)(B) = \mathbb{E}[\mu(B)].$$

Now, let X be Wigner matrix of size n and $Y = X/\sqrt{n}$. We will apply the method of moments to the deterministic probabilty measure $\mathbb{E}\mu_Y$.

We will prove the following statement

Theorem 1.2 (Semicircle law for averaged eigenvalues)**.** *Assume* $\mathbb{E}X_{11} = \mathbb{E}X_{12} = 0$, $\mathbb{E}|X_{12}|^2 = 1$ *and for any integer* $k \geq 1$, $\mathbb{E}|X_{12}|^k + \mathbb{E}|X_{11}|^k < \infty$*. Then, if* $Y = X/\sqrt{n}$*, weakly*

$$\mathbb{E}\mu_Y \to \mu_{sc}.$$

From Fubini's Theorem and the spectral theorem

$$\int \lambda^k d\mathbb{E}\mu_Y(\lambda) = \mathbb{E}\int \lambda^k d\mu_Y(\lambda) = \frac{1}{n}\mathbb{E}\mathrm{Tr}Y^k.$$

Hence, if (m_k) is the sequence moments of the semicircle law, in view of Lemma 1.2, it is sufficient to prove that

Lemma 1.5. *For each integer* $k \geq 1$,

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}Y^k = m_k + O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Expanding the trace, we have

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\mathrm{Tr}Y^k &= \frac{1}{n^{k/2+1}}\mathbb{E}\mathrm{Tr}X^k \\
&= \frac{1}{n^{k/2+1}}\mathbb{E}\sum_{(i_1,\cdots,i_k)}\prod_{\ell=1}^{k} X_{i_\ell i_{\ell+1}} \\
&= \frac{1}{n^{\frac{k}{2}+1}}\sum_{(i_1,\cdots,i_k)} P(\mathbf{i}),
\end{aligned}$$

where $i_{k+1} = i_1$ and

$$P(\mathbf{i}) = \mathbb{E}\prod_{\ell=1}^{k} X_{i_\ell i_{\ell+1}}.$$

We call such element $\mathbf{i}$ a path from i_1 to $i_{k+1} = i_1$. Let us say two paths $\mathbf{i} = (i_1, \cdots, i_k)$ are $\mathbf{j} = (j_1, \cdots, j_k)$ are equivalent and write $\mathbf{i} \sim \mathbf{j}$ if there exists a permutation σ of size n such that $\sigma(i_t) = j_t$ for $t = 1, \ldots, k$. This is relation of equivalence and from the invariance of the law of X

by permutation of the entries, $P(\mathbf{i})$ is constant on each equivalence class. Observe also that if we set $|\mathbf{i}| = |\{i_1, \cdots, i_k\}|$ there are

$$n(n-1)\dots(n-|\mathbf{i}|+1) \sim n^{|\mathbf{i}|}$$

paths in the equivalence class of $\mathbf{i}$. We may pick a canonical path in each class by setting $i_1 = 1$ and $i_{t+1} \leq \max_{s \leq t} i_s + 1$ (new elements are visited by increasing order). If $\mathcal{I}_k$ is the set of canonical paths, we may thus write

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}Y^k \quad = \quad \sum_{\mathbf{i} \in \mathcal{I}_k} \frac{n(n-1)\dots(n-|\mathbf{i}|+1)}{n^{k/2+1}} P(\mathbf{i}), \tag{1.3}$$

Fix a path $\mathbf{i}$. We first give a rough bound on $P(\mathbf{i})$. We define $G(\mathbf{i}) = (V, E)$ as the graph (with loops) obtained by setting $V = \{i_1, \cdots, i_k\}$ and $E = \{\{i_1, i_2\}, \cdots, \{i_k, i_1\}\}$ (we omit to write the dependency in $\mathbf{i}$ for brevity). We note that $G(\mathbf{i})$ is connected. For $e = \{i, j\} \in E$, with $i < j$, we define the multiplicities of the edge e as

$$m_e^+ = \sum_{\ell=1}^k \mathbf{1}((i_\ell, i_{\ell+1}) = (i,j)) \quad \text{ and } \quad m_e^- = \sum_{\ell=1}^k \mathbf{1}((i_\ell, i_{\ell+1}) = (j,i)).$$

If $i = j$, we set $2m_e^+ = 2m_e^- = \sum_{\ell=1}^k \mathbf{1}((i_\ell, i_{\ell+1}) = (i,i))$. We have

$$k = \sum_{e \in E} m_e^+ + m_e^-. \tag{1.4}$$

From the idependence of the $(X_{ij})_{i \geq j}$ and Hölder inequality, we get

$$P(\mathbf{i}) = \prod_{e \in E} (\mathbb{E} X_e^{m_e^+} \bar{X}_e^{m_e^-}) \quad \text{ and } \quad |P(\mathbf{i})| \leq \prod_{e \in E} (\mathbb{E}|X_e|^k)^{\frac{m_e^+ + m_e^-}{k}} \leq \max_{e \in E} \mathbb{E}|X_e|^k \leq \beta_k,$$

where $\beta_k = \max(\mathbb{E}|X_{11}|^k, \mathbb{E}|X_{12}|^k)$. Moreover, since $\mathbb{E}X_{12} = \mathbb{E}X_{11} = 0$, $P(\mathbf{i}) = 0$ unless the path $\mathbf{i}$ is such that for all $e \in E$, $m_e^+ + m_e^- \geq 2$. Therefore, the identity (1.4) yields

$$|E| \leq \left[\frac{k}{2}\right],$$

where $[x]$ is the integer part of x. By Lemma 1.3, we find

$$|V| \leq |E| + 1 \leq \left[\frac{k}{2}\right] + 1.$$

If α_k is the number of canonical paths which visit each edge at least twice, we get from (1.3)

$$\left| \frac{1}{n}\mathbb{E}\mathrm{Tr}Y^k \right| \leq \alpha_k \beta_k \frac{n^{\left[\frac{k}{2}\right]+1}}{n^{\frac{k}{2}+1}}.$$

In particular, if k is odd the above expression is $O(1/\sqrt{n})$. If k is even then

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}A^k = \sum_{\mathbf{i}\in\mathcal{I}_k^2} P(\mathbf{i}) + O\left(\frac{1}{n}\right),$$

where $\mathcal{I}_k^2$ is the set of of canonical paths $\mathbf{i} = (i_1, \cdots, i_k)$ such that $G(\mathbf{i})$ is a tree with ℓ edges and the multiplicity of each edge e has multiplicity $m^+(e) + m^-(e) = 2$. If $\mathbf{i} \in \mathcal{I}_k^2$, observe that $m^+(e) = m^-(e) = 1$. Indeed otherwise, if $m^+(e) = 2$ and $e = (i_s, i_{s+1}) = (i_t, i_{t+1})$ with $s < t$, the subpath of $\mathbf{i}$, $(i_s, i_{s+1}, \ldots, i_t, i_{t+1})$ must contain a cycle. The same argument implies also that for any $t = 1, \ldots, k$ with $i_{k+1} = 1$: either $i_{t+1} = \max_{s\le t} i_s + 1$ or $\{i_t, i_{t+1}\}$ is the last edge which has been visited only once.

Set $k = 2\ell$. We claim that $\mathcal{I}_k^2$ is in bijection with the set of Dyck paths $\mathcal{D}_\ell$. Indeed, to any canonical path $\mathbf{i} \in \mathcal{I}_k^2$ associate the sequence $(x_0, \cdots, x_k)$ given by $x_0 = 0$ and for $t = 1, \ldots, k$ $x_t - x_{t-1} = 1$ if $i_t = \max_{s<t} i_s + 1$ and $x_t - x_{t-1} = -1$ otherwise. This is a Dyck path from what precedes. Conversely, to any Dyck path $(x_0, \cdots, x_k)$, we define $\mathbf{i} = (i_1, \cdots, i_k)$ by $i_1 = 0$ and the following iterative rule: if $x_t - x_{t-1} = 1$, $i_t = \max_{s<t} i_s + 1$ and if $x_t - x_{t-1} = -1$, $i_t = i_{s_t}$ where $s_t < t$ is the largest s such that $x_s = x_t$. Then $\mathbf{i}$ is a path in $\mathcal{I}_k^2$.

We find finally

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}Y^{2\ell} = |\mathcal{D}_\ell| + O\left(\frac{1}{n}\right).$$

It remains to use Lemma 1.4. □

6 Second moment method

Theorem 1.3 (Semicircle law with finite moments)**.** *Assume* $\mathbb{E}X_{11} = \mathbb{E}X_{12} = 0$, $\mathbb{E}|X_{12}|^2 = 1$ *and for any integer* $k \ge 1$, $\mathbb{E}|X_{12}|^k + \mathbb{E}|X_{11}|^k < \infty$. *Then, if* $Y = X/\sqrt{n}$, *a.s. weakly*

$$\mu_Y \to \mu_{sc}.$$

We prove the following lemma

Lemma 1.6. *For each integer* $k \ge 1$,

$$\mathrm{Var}\left(\frac{1}{n}\mathbb{E}\mathrm{Tr}Y^k\right) = O(n^{-2}).$$

Assuming Lemma 1.2, Theorem 1.3 is proved as follows. From the monotone convergence theorem

$$\mathbb{E}\sum_{n\ge 0}\left(\int x^k d\mu_Y - \mathbb{E}\int x^k d\mu_Y\right)^2 < \infty.$$

Hence a.s. $\sum_{n\ge 0}\left(\int x^k \mu_Y - \mathbb{E}\int x^k \mu_Y\right)^2 < \infty$ and, a.s. $\int x^k \mu_Y - \mathbb{E}\int x^k \mu_Y \to 0$. Then the theorem follows from Theorem 1.2.

Proof of Lemma. We use the notation of Lemma 1.5. We start with

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}Y^k\right) &= \mathbb{E}\left(\frac{1}{n^{k/2+1}}\sum_{(i_1,\cdots,i_k)}\prod_{\ell=1}^{k} X_{i_\ell i_{\ell+1}} - P(\mathbf{i})\right)^2 \\ &= \frac{1}{n^{k+2}}\sum_{(i_1,\cdots,i_k),(j_1,\cdots,j_k)} P(\mathbf{i},\mathbf{j}) - P(\mathbf{i})P(\mathbf{j}),\end{aligned}$$

where

$$P(\mathbf{i},\mathbf{j}) = \mathbb{E}\prod_{\ell=1}^{k} X_{i_\ell i_{\ell+1}} X_{j_\ell j_{\ell+1}}.$$

As in the proof of Lemma 1.5, we define $G(\mathbf{i},\mathbf{j}) = (V,E)$ as the corresponding weighted graph. Note that $P(\mathbf{i},\mathbf{j}) - P(\mathbf{i})P(\mathbf{j}) = 0$ if $\mathbf{i}$ and $\mathbf{j}$ do not have two successive indices in common. Hence we may restrict to $G = G(\mathbf{i},\mathbf{j})$ connected. We have $\sum_{e\in E} m_e^+ + m^- e = 2k$ and $m_e^+ + m_e^- \geq 2$. Hence $|E| \leq k$, $|V| \leq |E| + 1 \leq k+1$ and, arguing Lemma 1.5 we get that

$$\mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}Y^k\right) = O\left(\frac{n^{k+1}}{n^{k+2}}\right) = O\left(\frac{1}{n}\right).$$

It already implies that μ_Y converges to μ_{sc} weakly in probability. To improve the bound on the variance, we may restrict ourself to indices such that

$$|V| = |E| + 1 = k+1 \text{ and for all } e \text{ in } E \ m_e^+ + m_e^- = 2.$$

We may assume without loss of generality that $(i_1,i_2) = (j_1,j_2)$. Consider the path $\pi = (i_1,\cdots,i_k,i_1)$. Since $m_e^+ + m_e^- = 2$, we have $i_k \neq i_2$. Hence π is a closed path in G which contains a cycle. This contradicts the assumption that G is a tree. Therefore, since does not occur, we have $|E| \leq k-1$ and $|V| \leq k$. It follows that

$$\mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}Y^k\right) = O\left(\frac{n^{k}}{n^{k+2}}\right) = O\left(\frac{1}{n^2}\right).$$

It concludes the proof. $\square$

Lecture 2

Random matrices: a playground for concentration inequalities

It is possible to combine with a great effect classical pertubation inequalities and concentration inequalities. In this chapter, we will use two perturbations inequalities for eigenvalues and we review some classical concentration inequalities.

1 Two perturbation inequalities of eigenvalues

1.1 Courant-Fischer inequalities

Variational formula for the eigenvalues

We order the eigenvalues of $A \in H_n(\mathbb{C})$ non-increasingly

$$\lambda_n(A) \leq \cdots \leq \lambda_1(A). \tag{2.1}$$

Lemma 2.1 (Courant-Fischer min-max theorem)**.** *Let $A \in H_n(\mathbb{C})$. Then*

$$\lambda_k(A) = \max_{H:\dim(H)=k} \min_{x\in H, \|x\|_2=1} \langle Ax, x\rangle.$$

Proof. Let $(u_i)_i$ be an eigenvector basis of A associated to $\lambda_1, \cdots, \lambda_n$. We choose $H = \mathrm{span}(u_1, \cdots, u_k)$. We find

$$\max_{H:\dim(H)=k} \min_{x\in H, \|x\|_2=1} \langle Ax, x\rangle \geq \lambda_k.$$

On the other hand, let H be a vector space such that $\dim(H) = k$. Define $S = \mathrm{span}(u_n, \cdots, u_k)$ so that $\dim(S) = n - k + 1$. Since

$$n \geq \dim(H \cup S) = \dim(H) + \dim(S) - \dim(S \cap H)$$

we find $S \cap H \neq 0$. In particular,

$$\min_{x \in H, \|x\|_2 = 1} \langle Ax, x \rangle \leq \lambda_k.$$

as requested. $\square$

As an immediate corollary, we obtain the *subadditivity of the largest eigenvalue*, for any $A, B \in H_n(\mathbb{C})$,

$$\lambda_1(A + B) \leq \lambda_1(A) + \lambda_1(B). \tag{2.2}$$

Interlacing of eigenvalues

An important corollary of the Courant-Fischer min-max theorem is the *interlacing of eigenvalues*. The interlacing inequalities allow to derive deviation inequalities for eigenvalues where the deviation is measured through its *rank*. By convention if $A \in H_n(\mathbb{C})$, we set for integer $i \geq 1$,

$$\lambda_{n+i}(A) = -\infty \quad \text{ and } \quad \lambda_{1-i}(A) = +\infty \tag{2.3}$$

Lemma 2.2 (Weak interlacing)**.** *Let A, B in $H_n(\mathbb{C})$ and assume that* $\dim(A - B) = r$ *Then, for any* $1 \leq k \leq n$*,*

$$\lambda_{k+r}(A) \leq \lambda_k(B) \leq \lambda_{k-r}(A).$$

Proof. We prove $\lambda_{k+r}(A) \leq \lambda_k(B)$. We may assume that $k + r \leq n$. By definition, for some vector space H of dimension $k + r$,

$$\lambda_{k+r}(A) = \min_{x \in H, \|x\|_2 = 1} \langle Ax, x \rangle.$$

Take $H' = H \cap \ker(A - B)$. By construction

$$\lambda_{k+r}(A) \leq \min_{x \in H', \|x\|_2 = 1} \langle Ax, x \rangle = \min_{x \in H', \|x\|_2 = 1} \langle Bx, x \rangle \leq \lambda_{k'}(B)$$

where $k' = \dim(H')$. Now, the inequality,

$$n - \dim(H') \leq (n - \dim(H)) + \dim(\mathrm{im}(A - B))$$

yields $k' \geq k$. This concludes the proof of the inequality $\lambda_{k+r}(A) \leq \lambda_k(B)$. For the proof of $\lambda_k(B) \leq \lambda_{k-r}(A)$, we may assume that $k - r \geq 1$. Then, simply replace A and B in the above argument. $\square$

There are variants of the above interlacing inequality. The above argument gives also the following lemma.

Lemma 2.3 (Strong Interlacing)**.** *Let A, B in H_n and assume that $A = B + C$ with C positive semi-definite with rank r. Then*

$$\lambda_k(B) \leq \lambda_k(A) \leq \lambda_{k-r}(B).$$

Deviation in Kolmogorov-Smirnov distance

We now give a perturbation inequality which is a consequence of interlacing. For μ, μ' two real probability measure, we introduce the *Kolmogorov-Smirnov distance*

$$d_{\mathrm{KS}}(\mu,\mu') = \sup_{t\in\mathbb{R}} |\mu(-\infty,t] - \mu'(-\infty,t]| = \|F_\mu - F_{\mu'}\|_\infty,$$

where F_μ is the cumulative partition function of μ. The Kolmogorov-Smirnov distance is closely related to functions with bounded variations. More precisely, for $f:\mathbb{R}\to\mathbb{R}$ the bounded variation distance is defined as

$$\|f\|_{\mathrm{BV}} = \sup \sum_{k\in\mathbb{Z}} |f(x_{k+1}) - f(x_k)|,$$

where the supremum is over all sequence $(x_k)_{k\in\mathbb{Z}}$ with $x_n \le x_{n+1}$. If the $f = \mathbf{1}((-\infty,t))$ then $\|f\|_{\mathrm{BV}} = 1$ while if the derivative of f is in $L^1(\mathbb{R})$, we have

$$\|f\|_{\mathrm{BV}} = \int |f'| dt.$$

Functions with bounded variations have left and right limits at any points and have a countable number of discontinuity points. To avoid minor complications, we will denote by BV^+ the set of *right continuous* function with $\|f\|_{\mathrm{BV}} < \infty$.

We have a *variational formula* for the Kolomogorov-Smirnov distance.

Lemma 2.4. *For any real probability measures μ,μ',*

$$d_{\mathrm{KS}}(\mu,\mu') = \sup\left\{ \int f d\mu - \int f d\mu' : f\in \mathrm{BV}^+, \|f\|_{\mathrm{BV}} \le 1 \right\}.$$

Proof. Choosing $f = \mathbf{1}_{(-\infty,t]}$ gives a first upper bound on d_{KS}. The other way around, let $f\in\mathrm{BV}^+$. Then, a classical theorem (see [Roy63]) asserts that f can be written as $f(y) = f(x) + \int_x^y d\sigma(t)$ with σ signed measure on $\mathbb{R}$ and $\|f\|_{\mathrm{BV}} = \int d|\sigma|$.

Let τ is a continuity point of $F_\mu(t) = \mu(-\infty,t]$ and f, we have the integration by part formula

$$\int_{-\infty}^{\tau} f(t) d\mu(t) = -f(\tau)F_\mu(\tau) - \int_{-\infty}^{\tau} F_\mu(t) d\sigma(t).$$

This yields, letting τ tend to infinity,

$$\int f(t)d\mu(t) - \int f(t) d\mu'(t) = -\int \big(F_\mu(t) - F_{\mu'}(t)\big) d\sigma(t).$$

In particular,

$$\left| \int f d\mu - \int f d\mu' \right| \le \int \big|F_\mu(t) - F_{\mu'}(t)\big| d|\sigma|(t) \le \|f\|_{\mathrm{BV}} \|F_\mu - F_{\mu'}\|_\infty. \tag{2.4}$$

If gives the reverse inequality. $\square$

As a corollary from the interlacing inequality, we obtain the following deviation inequality for empirical spectral distributions.

Lemma 2.5 (Rank inequality for ESD)**.** *Let A, B in $H_n(\mathbb{C})$ and assume that* $\dim(A-B) = r$ *Then,*

$$d_{\mathrm{KS}}(\mu_A, \mu_B) \leq \frac{r}{n}.$$

and for any $f \in \mathrm{BV}^+$,

$$\left| \int f(\lambda) d\mu_A(\lambda) - \int f(\lambda) d\mu_B(\lambda) \right| \leq \left(\frac{r}{n} \right) \|f\|_{\mathrm{BV}}.$$

Proof. Fix $t \in \mathbb{R}$. Let k and k' be the smallest indices such that $\lambda_k(A) \leq t$ and $\lambda_{k'}(B) < t$ (recall our convention (2.3)). By lemma 2.2, we find

$$|k - k'| \leq r.$$

This yields

$$\left| F_\mu(t) - F_{\mu'}(t) \right| = \left| \frac{(n+1-k) - (n+1-k')}{n} \right| \leq \frac{r}{n}.$$

This gives the first statement. The second statements follows from (2.4). □

1.2 Hoffman-Wielandt inequality

We now present another matrix inequality which is particularly useful. It is an inequality in terms of perturbations in *Frobenius norm*, for $A \in H_n(\mathbb{C})$:

$$\|A\|_F = \sqrt{\mathrm{Tr} A^2}.$$

This norm is very natural: $(H_n(\mathbb{C}), \|\cdot\|_F)$ is isomorphic to the Euclidean space $(\mathbb{R}^{n^2}, \|\cdot\|_2)$ through the map $A \mapsto (A_{ii}, \sqrt{2}\Re(A_{ij}), \sqrt{2}\Im(A_{ij}))_{j>i}$. Recall the convention (2.1) to order eigenvalues of an hermitian matrix.

Lemma 2.6 (Hoffman-Wielandt inequality)**.** *Let A, B in $H_n(\mathbb{C})$,*

$$\sum_{i=1}^{n} (\lambda_i(A) - \lambda_i(B))^2 \leq \mathrm{Tr}(A-B)^2 = \|A - B\|_F^2.$$

Proof. From the septral theorem, we have $A = UCU^*$ and $B = VDV^*$, where U, V are unitary matrices and C, D are diagonal matrices with respective diagonals $(\lambda_1(A), \ldots, \lambda_n(A))$ and $(\lambda_1(B), \ldots, \lambda_n(B))$. Set $W = U^*V$, since $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, we find

$$\|A - B\|_F^2 = \|CW - WD\|_F.$$

Setting $P_{ij} = |W_{ij}|^2$ we get

$$\|A - B\|_F^2 = \sum_{i,j} P_{ij}(\lambda_i(A) - \lambda_j(B))^2.$$

Since W is unitary, $P = (P_{ij})$ is an $n \times n$ doubly stochastic matrix. If $\mathcal{S}_n$ denotes the set of $n \times n$ doubly stochastic matrices, we find

$$\|A - B\|_F^2 \geq \inf_{Q \in \mathcal{S}_n} \sum_{i,j} Q_{ij}(\lambda_i(A) - \lambda_j(B))^2.$$

The set $\mathcal{S}_n$ is convex and compact. Moreover, the above variational expression is linear in Q. It follows that the infimum is reached at an extremal point of $\mathcal{S}_n$. Now, a theorem due to Birkhoff and von Neumann states that the extremal points of $\mathcal{S}_n$ are the permutation matrices: that is matrices such that for some permutation $\sigma \in S_n$, $Q_{ij} = \mathbf{1}(\sigma(i) = j)$ for every $1 \leq, i, j \leq n$. It gives

$$\|A - B\|_F^2 \geq \inf_{\sigma \in S_n} \sum_i (\lambda_i(A) - \lambda_{\sigma(i)}(B))^2.$$

To conclude, we claim that if $x_1 \geq \ldots \geq x_n$ and $y_1 \geq \ldots \geq y_n$ then

$$\inf_{\sigma \in S_n} \sum_{i=1}^n (x_i - y_{\sigma(i)})^2 = \sum_{i=1}^n (x_i - y_i)^2. \tag{2.5}$$

Indeed, if $\sigma(j) < \sigma(i)$ for some $i < j$, we find that

$$\begin{aligned}(x_i - y_{\sigma(i)})^2 + (x_j - y_{\sigma(j)})^2 &= (x_i - y_{\sigma(j)})^2 + (x_j - y_{\sigma(i)})^2 + 2(x_i - x_j)(x_{\sigma(j)} - x_{\sigma(i)}) \\ &\geq (x_i - y_{\sigma(j)})^2 + (x_j - y_{\sigma(i)})^2.\end{aligned}$$

It implies easily (2.5). □

For $p \geq 1$, μ, μ' two real probability measure such that $\int |x|^p d\mu$ and $\int |x|^p d\mu'$ are finite. We define the L^p*-Wasserstein distance* as

$$W_p(\mu, \mu') = \left(\inf_\pi \int_{\mathbb{R} \times \mathbb{R}} |x - y|^p d\pi \right)^{\frac{1}{p}}$$

where the infimum is over all coupling π of μ and μ' (i.e. π is probability measure on $\mathbb{R} \times \mathbb{R}$ whose first marginal is equal to μ and second marginal is equal to μ'). Note that Hölder inequality gives for $1 \leq p \leq p'$,

$$W_p \leq W_{p'}.$$

For any $p \geq 1$, if $W_p(\mu_n, \mu)$ converges to 0 then $\mu_n \to \mu$ weakly. This follows for example from Kantorovich-Rubinstein duality

$$W_1(\mu, \mu') = \sup \left\{ \int f d\mu - \int f d\mu' : \|f\|_L \leq 1 \right\}.$$

where we have used the Lipschitz (semi)-norm

$$\|f\|_L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

We may then deduce a perturbation inequality for empirical spectral distribution in terms of Frobenius norm.

Corollary 2.1 (Trace inequality inequality for ESD)**.** *Let A, B in $H_n(\mathbb{C})$, then*

$$W_2(\mu_A, \mu_B) \leq \sqrt{\frac{1}{n}\mathrm{Tr}(A - B)^2} = \|A - B\|_F.$$

Proof. Consider the coupling π of (μ_A, μ_B) defined as

$$\pi = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A), \lambda_i(B)}.$$

We find

$$\int_{\mathbb{R}\times\mathbb{R}} |x - y|^2 d\pi \leq \frac{1}{n}\mathrm{Tr}(A - B)^2.$$

The left hand side is lower bounded $W_2^2(\mu_A, \mu_B)$ by construction (in fact it is even equal from (2.5)). □

We also get a Lipschitz bound for spectral statistics in terms of Frobenius norm.

Corollary 2.2 (Lipschitz continuity of spectral statistics)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipchitz function. Then the map $F : (H_n(\mathbb{C}), \|\cdot\|_F) \to \mathbb{R}$*

$$F(X) = f(\lambda_1(X), \cdots, \lambda_n(X)).$$

is Lispchitz with constant $\|f\|_L$. In particular, if $g : \mathbb{R} \to \mathbb{R}$ is Lipschitz,

$$G(X) = \frac{1}{n}\mathrm{Tr}\left\{ g\left(\frac{X}{\sqrt{n}}\right)\right\} = \int g(\lambda) d\mu_{\frac{X}{\sqrt{n}}}(\lambda)$$

is Lipschitz with constant $\|g\|_L/n$.

Proof. From Hoffman-Wielandt inequality, we get

$$|F(X) - F(Y)|^2 \leq \|f\|_L^2 \sum_{k=1}^{n} |\lambda_k(X) - \lambda_k(Y)|^2 \leq \|f\|_L^2 \|X - Y\|_F^2.$$

The second statement follows from the first statement, since the map

$$(x_1, \cdots, x_n) \to \sum_{i=1}^{n} x_i$$

is Lipschitz with constant $\sqrt{n}$ (thanks to Cauchy-Schwartz inequality). □

Remark 2.1 (Frobenius versus Euclid)**.** *It will sometimes be more convenient to see a matrix* $X \in H_n(\mathbb{C})$ *has a vector in* $\mathbb{R}^{n^2}$ *defined by* $(X_{ii}, \Re(X_{ij}), \Im(X_{ij}))_{j>i}$. *Then, using that* $\|X\|_F \leq \sqrt{2}\|X\|_2$ *the Lipschitz bounds in* $(H_n(\mathbb{C}), \|\cdot\|_F)$ *of Corollary 2.2 can be replaced by Lipschitz bound in* $(\mathbb{R}^{n^2}, \|\cdot\|_2)$ *with constants multiplied by a factor* $\sqrt{2}$.

1.3 Deviation in bounded variation and Lipshitz norm

For real probability measures μ, μ', we define the distance

$$d(\mu, \mu') = \sup\left\{\int f d\mu - \int f d\mu' : \|f\|_L \leq 1, \|f\|_{\mathrm{BV}} \leq 1\right\},$$

so that we can apply both Corollary 2.1 and Lemma 2.5 to this distance :

$$d(\mu_A, \mu_B) \leq \min\left(\frac{\operatorname{rank}(A-B)}{n}, \sqrt{\frac{1}{n}\operatorname{Tr}(A-B)^2}\right). \tag{2.6}$$

This distance is stronger that the *Lévy distance.*

$$d_L(\mu, \mu') = \inf\{\varepsilon > 0 : \text{ for all real } t,\ F_\mu(t) \leq F_{\mu'}(t+\varepsilon) + \varepsilon, F_{\mu'}(t) \leq F_\mu(t+\varepsilon) + \varepsilon\}.$$

where $F_\mu(t) = \mu(-\infty, t]$ is the cumulative partition function. It follows that if $d(\mu_n, \mu) \to 0$ then μ_n converges weakly to μ.

Exercise 2.1. *Deduce Theorem 1.1 from Theorem 1.3 by a successive truncation argument relying on Equation* (2.6)*.*

2 Bounds on the variance

2.1 Efron-Stein inequality

We follow Boucheron, Lugosi, Massart [BLM13]. Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ be a product of measurable spaces and let $X = (X_1, \ldots, X_n) \in \mathcal{X}$ be a vector of independent random variables.

We denote by $\mathbb{E}_i$ the conditional expectation with respect to the σ-algebra generated by $(X_j)_{j\neq i}$ and, if $Z = f(X)$ is a measurable function of X, $\operatorname{Var}_i(Z) = \mathbb{E}_i(Z - \mathbb{E}_i Z)^2$ is the conditional variance. Efron-Inequality asserts that the variance is subadditive.

Theorem 2.1 (Efron-Stein Inequality)**.** *Let* $X = (X_1, \cdots, X_n)$ *be independent random variables and* $f : \mathcal{X} \to \mathbb{R}$ *such that* $\mathbb{E}f(X)^2 < \infty$ *then, if* $Z = f(X)$*,*

$$\operatorname{Var}(Z) \leq \sum_{i=1}^n \mathbb{E}\operatorname{Var}_i(Z) = \sum_{i=1}^n \mathbb{E}(Z - \mathbb{E}_i Z)^2.$$

Moreover, if $X' = (X'_1, \cdots, X'_n)$ is an copy of X and $Z'_i = f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)$,

$$\sum_{i=1}^n \mathbb{E}\mathrm{Var}_i(Z) = \frac{1}{2}\sum_{i=1}^n \mathbb{E}(Z - Z'_i)^2.$$

Proof. The second claim is easy as $\mathbb{E}_i(Z - \mathbb{E}_i Z)(Z'_i - \mathbb{E}_i Z) = 0$. We follow a beautiful argument of Chatterjee. We may assume that $\mathbb{E}f(X) = 0$. Let $X_{(i)} = (X'_1, \ldots, , X'_i, X_{i+1}, \ldots, X_n)$ and $X'_{(i)} = (X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)$. We have $X_{(0)} = X$, $X_{(n)} = X'$ and $Z'_i = f(X'_{(i)})$. We write, for any $f, g : \mathcal{X} \to \mathbb{R}$

$$\mathbb{E}g(X)f(X) = \sum_{i=1}^n \mathbb{E}g(X)(f(X_{(i-1)}) - f(X_{(i)}))$$

Switching X and X', we note that $g(X)(f(X_{(i-1)}) - f(X_{(i)})$ has the same distribution than $-g(X'_{(i)})(f(X_{(i-1)}) - f(X_{(i)})$. Hence,

$$\mathbb{E}g(X)f(X) = \frac{1}{2}\sum_{i=1}^n \mathbb{E}(g(X) - g(X'_{(i)}))(f(X_{(i-1)}) - f(X_{(i)}))$$

It remains to set $f = g$ and apply Cauchy-Schwartz inequality. □

A simple consequence of Efron-Stein inequality when f has *bounded differences*, that is f is Lipschitz for a weighted Hamming pseudo-distance, i.e. for every $x, y \in \mathcal{X}$,

$$|f(x) - f(y)| \leq \sum_{k=1}^n c_k \mathbf{1}_{x_k \neq y_k}. \tag{2.7}$$

for some $c = (c_1, \cdots, c_n) \in \mathbb{R}_+^n$. We denote by $\|y\|_2 = \sqrt{\sum_i y_i^2}$, the usual Euclidean norm. Theorem 2.1 implies the following.

Corollary 2.3. *Let $X = (X_1, \cdots, X_n)$ be independent random variables and $f : \mathcal{X} \to \mathbb{R}$ such that (2.7) holds then*

$$\mathrm{Var}(f(X)) \leq \frac{\|c\|_2^2}{2}.$$

We will however see a much more powerful concentration inequality for functions with bounded differences.

Random matrices with independent half-rows

We will obtain two applications for random matrices of Efron-Stein inequality, one for each pertubation inequality that we have seen (interlacing and Hoffman-Wielandt). The weakest bound applies to a large class of random matrices and it is obtained thanks to the interlacing inequality.

Theorem 2.2 (Variance of ESD with independent half-rows). *Let $X \in H_n(\mathbb{C})$ be an random hermitian matrix and for $1 \leq k \leq n$, define the variables $X_k = (X_{kj})_{1\leq j\leq k} \in \mathbb{C}^k$. If the variables $(X_k)_{1\leq k\leq n}$ are independent, then for every $f \in \mathrm{BV}^+$,*

$$\mathrm{Var}\left(\int f \, d\mu_X\right) \leq \frac{2\|f\|_{\mathrm{BV}}^2}{n}.$$

There are examples of random matrices such that the conclusion of Theorem 2.2 is sharp.

Proof. For any $x = (x_1, \ldots, x_n) \in \mathcal{X} = \left\{(x_i)_{1\leq i\leq n} : x_i \in \mathbb{C}^{i-1} \times \mathbb{R}\right\}$, let $H(x)$ be the $n\times n$ hermitian matrix given by $H(x)_{ij} = x_{ij}$ for $1 \leq j \leq i \leq n$. We thus have $X = H(X_1, \ldots, X_n)$. For all $x \in \mathcal{X}$ and $x_i' \in \mathbb{C}^{i-1} \times \mathbb{R}$, the matrix

$$H(x_1, \ldots, x_n) - H(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)$$

has only the i-th row and column possibly different from 0, and thus

$$\mathrm{rank}\left(H(x_1, \ldots, x_n) - H(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)\right) \leq 2.$$

Therefore from Lemma 2.5, we obtain,

$$\left|\int f d\mu_{H_{(x_1,\ldots,x_n)}} - \int f d\mu_{H_{(x_1,\ldots,x_{i-1},x_i',x_{i+1},\ldots,x_n)}}\right| \leq \frac{2\|f\|_{\mathrm{BV}}}{n}.$$

The desired result follows now from Corollary 2.3. $\square$

Convex Poincaré inequality

In Efron-Stein inequality, when $\mathcal{X}$ is $\mathbb{R}^m$ or $\mathbb{C}^m$, we may use differential calculus to estimate the upper bound in the inequality. If f is

Theorem 2.3 (Convex Poincaré for bounded variables). *Let $\mathcal{B}$ be the unit ball of $\mathbb{R}^k$ and $X = (X_1, \cdots, X_n) \in \mathcal{B}^n \subset \mathbb{R}^{kn}$ be a vector of independent random variables. If $f : \mathbb{R}^{kn} \to \mathbb{R}$ is convex and such that $\mathbb{E}f(X)^2 < \infty$ then,*

$$\mathrm{Var}(f(X)) \leq 2\mathbb{E}\|\nabla f(X)\|_2^2.$$

In particular, if f is Lipshitz with constant L, the variance is bounded by $2L^2$.

In the next subsection, we will introduce the probabilistic Poincaré inequality. Theorem 2.3 asserts that all vectors of independent bounded variables satisfy a Poincaré inequality restricted to convex functions.

Proof. Let X' be an independent copy of X and $X'_{(i)} = (X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)$. By symmetry and Theorem 2.1, we have

$$\mathrm{Var}(f(X)) \leq \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}(f(X) - f(X'_i))^2 = \sum_{i=1}^{n} \mathbb{E}(f(X) - f(X'_i))_+^2,$$

where $(a)_+^2 = \max(a,0)^2$. The convexity assumption implies that $f(x) - f(y) \leq \nabla f(x) \cdot (x-y)$ where $\cdot$ is the scalar product. Since $a \mapsto (a)_+^2$ is non-decreasing, we get from Cauchy-Schwartz inequality,

$$(f(X) - f(X'_i))_+^2 \leq \|\nabla_i f(X)\|_2^2 \|X_i - X'_i\|_2^2,$$

where $\nabla_i f(X) \in \mathbb{R}^k$ is the gradient of $X_i \mapsto f(X)$. By assumption $\|X_i - X'_i\|_2 \leq 2$ and we find

$$\sum_{i=1}^{n} (f(X) - f(X'_i))_+^2 \leq 2\sum_{i=1}^{n} \|\nabla_i f(X)\|_2^2 = 2\|\nabla f(X)\|_2^2$$

as requested. □

In some cases it is possible to remove the assumption that f is convex. If f is smooth, a possibility is to write a Taylor expansion to higher order. Another favorable case is when f is a finite weighted sum of convex functions (we will see these two examples appear in random matrices).

Variance of largest eigenvalue and linear statistics

A first easy consequence of the convex Poincaré inequality is the following.

Theorem 2.4 (Variance of largest eigenvalue with independent entries)**.** *Let $X \in H_n(\mathbb{C})$ be an random hermitian matrix such that the variables $(X_{ij})_{i \leq j}$ are independent and bounded by D. Then*

$$\mathrm{Var}\left(\lambda_1(X)\right) \leq 4D^2.$$

Proof. By Corollary 2.2, the real-valued function on $\mathbb{R}^n \times \mathbb{C}^{n(n-1)/2}$, $F : (X_{ii}, X_{ij})_{i<j} \mapsto \lambda_1(X)$ is convex with Lipschitz constant $\sqrt{2}$ (see Remark 2.1). Moreover it is convex by inequality (2.2). It remains to apply Theorem 2.3. □

This result is interesting but fails to capture the proper order of the fluctuations in n, the variance of $\lambda_1(X)$ is of order $O\left(\frac{1}{n^{1/3}}\right)$.

Another consequence of Theorem 2.3 is the following.

Theorem 2.5 (Variance of ESD with independent entries)**.** *Let $X \in H_n(\mathbb{C})$ be an random hermitian matrix such that the variables $(X_{ij})_{i \leq j}$ are independent and bounded by D. Then for every C^1-function $f : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant L and k inflection points,*

$$\mathrm{Var}\left(\int f \, d\mu_{\frac{X}{\sqrt{n}}}\right) \leq \frac{4(k+1)^2 D^2 L^2}{n^2}.$$

Note that if $(Y_1, \dots, Y_n)$ are iid real random variables and $f(Y_i)$ is square-integrable then the variance of

$$\mathrm{Var}\left(\frac{1}{n}\sum f(Y_i)\right) = O\left(\frac{1}{n}\right).$$

Hence, Theorem 2.5 implies that eigenvalues of random matrices are much more concentrated than iid samples. The first step of proof is the following lemma.

Lemma 2.7 (Peierls)**.** *Suppose that the function $f : \mathbb{R} \to \mathbb{R}$ is convex. The function $F : X \mapsto \mathrm{Tr} f(X)$ on $H_n(\mathbb{C})$ is convex.*

Proof. Let $\psi \in \mathbb{C}^n$ such that $\|\psi\|_2 = 1$. Let X be in $H_n(\mathbb{C})$, $(u_1, \dots, u_n)$ be an orthonormal basis of eigenvectors of X with eigenvalues $(\lambda_1, \cdots, \lambda_n)$. From the spectral theorem,

$$\begin{aligned} \langle \psi, f(X)\psi \rangle &= \sum_{k=1}^{n} f(\lambda_k)|\langle \psi, u_k \rangle|^2 \\ &\geq f\left(\sum_{k=1}^{n} \lambda_k |\langle \psi, u_k \rangle|^2\right), \end{aligned}$$

where the last step follows from Jensen inequality together with Pythagoras theorem: $\sum_k |\langle \psi, u_k \rangle|^2 = \|\psi\|_2^2 = 1$. Using again the spectral theorem, this proves the Peierls inequality

$$\langle \psi, f(X)\psi \rangle \geq f(\langle \psi, X\psi \rangle) \quad \text{when} \quad \|\psi\|_2 = 1. \tag{2.8}$$

Now, let X, Y be in $H_n(\mathbb{C})$, $t \in [0, 1]$ and $(v_1, \dots, v_n)$ an orthornormal basis of eigenvectors of $tX + (1-t)Y$. For all k, using that v_k is an eigenvector and the convexity of f, we have

$$\begin{aligned} \langle v_k, f(tX + (1-t)Y)v_k \rangle &= f(\langle v_k, (tX + (1-t)Y)v_k \rangle) \\ &= f(t\langle v_k, Xv_k \rangle + (1-t)\langle v_k, Yv_k \rangle) \\ &\leq tf(\langle v_k, Xv_k \rangle) + (1-t)f(\langle v_k, Yv_k \rangle) \\ &\leq t\langle v_k, f(X)v_k \rangle + (1-t)\langle v_k, f(Y)v_k \rangle, \end{aligned}$$

where at the last step, we have used (2.8). It remains to recall that for any matrix A in $M_n(\mathbb{C})$, $\mathrm{Tr}(A) = \sum_k \langle v_k, Av_k \rangle$. ☐

Lemma 2.8 (Weighted sum of convex functions)**.** *Suppose that the C^1-function $f : \mathbb{R} \to \mathbb{R}$ has k inflection points. Then, there exist $\varepsilon_i \in \{-1, 1\}$, f_i convex with $\|f_i\|_{\mathrm{L}} \leq \|f\|_{\mathrm{L}}$ such that*

$$f = \sum_{i=1}^{k+1} \varepsilon_i f_i.$$

Proof. By induction on k. Let $x_1 < \cdots < x_k$ be the inflection points of f. Up to considering $\pm(f(\cdot - x_k) - f(x_k))$, we may assume without loss of generality that $x_k = 0$, $f(0) = 0$ and $f''(x) > 0$ for $x > 0$. We decompose f as

$$\begin{aligned} f(x) &= f(x) - f'(0)x + f'(0)x \\ &= \big\{(f(x) - f'(0)x)\mathbf{1}(x < 0) + f'(0)x\big\} + (f(x) - f'(0)x)\mathbf{1}(x \geq 0) \\ &= g_1(x) + g_2(x). \end{aligned}$$

Notice that g_1 and g_2 are C^1-functions. Moreover, g_2 is convex and, for $x \geq 0$, $g_2'(x) = f'(x) - f'(0) \in [0, f'(x)]$ so that $\|g_2\|_{\mathrm{L}} \leq \|f\|_{\mathrm{L}}$. Similarly, g_1 has $k-1$ inflection points and, since $g_1(x) = f(x)$ on $(-\infty, 0]$ and $g_1(x) = f'(0)x$ on $[0, \infty)$ we find $\|g_1\|_{\mathrm{L}} \leq \|f\|_{\mathrm{L}}$. $\square$

Proof of Theorem 2.5. From Lemma 2.8, we have

$$\int f \, d\mu_{\frac{X}{\sqrt{n}}} = \sum_{t=1}^{k+1} \varepsilon_t \int f_t \, d\mu_{\frac{X}{\sqrt{n}}}.$$

for some signs ε_i and convex functions f_i. In particular, from Cauchy-Schwartz inequality,

$$\mathrm{Var}\left(\int f \, d\mu_{\frac{X}{\sqrt{n}}}\right) \leq (k+1) \sum_{t=1}^{k+1} \mathrm{Var}\left(\int f_t \, d\mu_{\frac{X}{\sqrt{n}}}\right).$$

By Corollary 2.2 and Lemma 2.7, the real-valued function on $\mathbb{R}^n \times \mathbb{C}^{n(n-1)/2}$, $F : (X_{ii}, X_{ij})_{i<j} \mapsto \int f_t \, d\mu_{\frac{X}{\sqrt{n}}}$ is convex with Lipschitz constant $\sqrt{2}L/n$ (see Remark 2.1). It remains to apply Theorem 2.3 to each term on the right hand side. $\square$

2.2 Poincaré Inequality

Definition and first properties

Definition 2.1 (Poincaré inequality)**.** *A random variable X on $\mathbb{R}^n$ satisfies Poincaré inequality with constant $c > 0$ if for any differentiable function with $\mathbb{E} f(X)^2 < \infty$,*

$$\mathrm{Var}(f(X)) \leq c \, \mathbb{E} \|\nabla f(X)\|_2^2$$

where $\|\nabla f\|_2^2 = \sum_{i=1}^n (\partial_i f)^2$.

We prove below that the standard Gaussian on $\mathbb{R}$ satisfies Poincaré with constant $c = 1$. More generally, let $V : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ such that $V(x) - c\|x\|_2^2/2$ is convex. Then the probability $P(dx) = Z^{-1} e^{-V(x)} dx$ satisfies Poincaré with constant c. We refer to Ledoux [Led01] for more properties on Poincaré inequalities.

We will use the following elementary properties of the Poincaré inequality.

Lemma 2.9 (Properties of Poincaré Inequality)**.** *Let X_1, X_2 be independent random variables in $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ which satisfy Poincaré with constants $c_1, c_2 > 0$.*

- Homogeneity *: for $a \in \mathbb{R}$, $b \in \mathbb{R}^{n_i}$, $aX_i + b$ satisfies Poincaré with constant $a^2 c_i$.*

- Tensorization *: (X_1, X_2) satisfies Poincaré in $\mathbb{R}^{n_1+n_2}$ with constant* $\max(c_1, c_2)$*.*

Proof. Homogeneity is obvious, the tensorization property is a direct consequence of Efron-Stein inequality, Theorem 2.1: if $X = (X_1, X_2)$,

$$\mathrm{Var}(f(X)) \leq \mathbb{E}(\mathrm{Var}_1 f(X)) + \mathbb{E}(\mathrm{Var}_2 f(X)) \leq c_1 \mathbb{E}\|\nabla_1 f(X)\|_2^2 + c_2 \mathbb{E}\|\nabla_2 f(X)\|_2^2.$$

where $\nabla_i f(X)$ is the gradient of $X_i \mapsto f(X)$. It concludes the proof as $\|\nabla_1 f(X)\|_2^2 + \|\nabla_2 f(X)\|_2^2 = \|\nabla f(X)\|_2^2$. □

Gaussian vectors satisfy Poincaré inequality.

Lemma 2.10 (Gaussian Poincaré inequality)**.** *The standard Gaussian variable on $\mathbb{R}^n$ satisfies Poincaré with constant $c = 1$.*

Proof. From the tensorization property, it is sufficent to prove the claim for $n = 1$. There are many possible proofs, here we use the celebrated *Gaussian integration by parts formula*: for any-real valued function f with obvious integrability conditions

$$\mathbb{E} X f(X) = \mathbb{E} f'(X), \tag{2.9}$$

where X is a standard Gaussian variable.

Let X' be an independent copy of X and f a C^1-function such that $\mathbb{E}f(X) = 0$. For $t \in [0, 1]$, we define

$$X_t = \sqrt{t}X + \sqrt{1-t}X' \quad \text{ and } \quad Y_t = \sqrt{1-t}X - \sqrt{t}X'.$$

Note that X_t and Y_t are independent standard Gaussian variables, $X_1 = X$, $X_0 = X'$ and $X = \sqrt{t}X_t + \sqrt{1-t}Y_t$. Using Fubini's Theorem, we may write

$$\begin{aligned}
\mathbb{E}f(X)^2 &= \mathbb{E}f(X)(f(X) - f(X')) \\
&= \mathbb{E}\int_0^1 f(X)\partial_t f(X_t) dt \\
&= \mathbb{E}\int_0^1 f(X)\left(\frac{X}{2\sqrt{t}} - \frac{X'}{2\sqrt{t}}\right) f'(X_t) dt \\
&= \int_0^1 \frac{1}{2\sqrt{t}\sqrt{1-t}} \mathbb{E} f(\sqrt{t}X_t + \sqrt{1-t}Y_t) Y_t f'(X_t) dt.
\end{aligned}$$

Using the independence of X_t and Y_t, we may use (2.9) for Y_t. We get

$$\mathbb{E}f(X)^2 \quad = \quad \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E}f'(\sqrt{t}X_t + \sqrt{1-t}Y_t)f'(X_t)dt.$$

It remains to apply Cauchy-Schwartz inequality. □

Application to random matrices

An immediate consequence of Corollary 2.2 is the following.

Theorem 2.6 (Variance of ESD with Poincaré)**.** *Let $X \in H_n(\mathbb{C})$ be an random hermitian matrix such that the vector $(X_{ij})_{i \leq j}$ (seen as a vector in $\mathbb{R}^{n^2}$) satisfies Poincaré with constant c. Then for every C^1-function $f : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant L,*

$$\mathrm{Var}\left(\int f \, d\mu_{\frac{X}{\sqrt{n}}}\right) \leq \frac{2cL^2}{n^2}.$$

Compared to Theorem 2.5, Theorem has removed the convexity assumption. The variance bound in Theorem 2.6 has the optimal order of magnitude.

Using the first part of Corollary 2.2, we can also derive a result for the concentration of a single eigenvalue.

Theorem 2.7 (Variance of an eigenvalue with Poincaré)**.** *Let $X \in H_n(\mathbb{C})$ be an hermitian random matrix and assume that the vector $(X_{ij})_{i \leq j}$ satisfies Poincaré with constant $c > 0$ in $\mathbb{R}^{n^2}$. Then for any $1 \leq k \leq n$,*

$$\mathrm{Var}(\lambda_k(X)) \leq 2c.$$

This result does not capture the actual variance. For GUE matrices and $k = \lceil pn \rceil$ with $0 < p < 1$, the variance is of order $O\left(\frac{\log n}{n}\right)$. For $k \in \{1, n\}$, the variance is of order $O\left(\frac{1}{n^{1/3}}\right)$.

3 Exponential Tail Bounds

3.1 Bounded Martingale difference inequality

We now improve the variance bound of Corollary 2.3 to sharper exponential tail bound. Recall the setting of the previous section. Let $\mathcal{X}_1 \cdots \mathcal{X}_n$ be metric spaces and set $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$.

Theorem 2.8 (Azuma-Hoeffding's inequality)**.** *Let $X = (X_1, \cdots, X_n)$ be independent random variables and $f : \mathcal{X} \to \mathbb{R}$ such that (2.7) holds then for any $t \geq 0$,*

$$\mathbb{P}\left(f(X) - \mathbb{E}f(X) \geq t\right) \leq \exp\left(\frac{-t^2}{2\|c\|_2^2}\right).$$

This type of result is called a concentration inequality. It has found numerous applications in mathematics over the last decades. For more on concentration inequalities, we refer to [Led01, BLM13]. As a corollary, we deduce the Hoeffding's inequality.

Corollary 2.4 (Hoeffding's inequality)**.** *Let $(X_k)_{1\le k\le n}$ be an independent sequence of real random variables such that for all integer k, $X_k \in [a_k, b_k]$. Then,*

$$\mathbb{P}\left(\sum_{k=1}^n X_k - \mathbb{E}X_k \ge t\right) \le \exp\left(\frac{-t^2}{2\sum_{k=1}^n (b_k - a_k)^2}\right). \tag{2.10}$$

The proof of Theorem 2.8 will be based on a lemma due to Hoeffding.

Lemma 2.11. *Let X be real random variable in $[a,b]$ such that $\mathbb{E}X = 0$. Then, for all $\lambda \ge 0$,*

$$\mathbb{E}e^{\lambda X} \le e^{\frac{\lambda^2(b-a)^2}{8}}.$$

Proof. By the convexity of the exponential,

$$e^{\lambda X} \le \frac{b-X}{b-a}e^{\lambda a} + \frac{X-a}{b-a}e^{\lambda b}.$$

Taking expectation, we obtain, with $p = -a/(b-a)$,

$$\begin{aligned}
\mathbb{E}e^{\lambda X} &\le \frac{b}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b} \\
&= \left(1 - p + pe^{\lambda(b-a)}\right) e^{-p\lambda(b-a)} \\
&= e^{\varphi(\lambda(b-a))},
\end{aligned}$$

where $\varphi(x) = -px + \ln(1 - p + pe^x)$. The derivatives of φ are

$$\varphi'(x) = -p + \frac{pe^x}{(1-p)e^{-x} + p} \quad \text{and} \quad \varphi''(x) = \frac{p(1-p)}{((1-p)e^{-x} + p)^2} \le \frac{1}{4}.$$

Since $\varphi(0) = \varphi'(0) = 0$, we deduce from Taylor expansion that

$$\varphi(x) \le \varphi(0) + x\varphi'(0) + \frac{x^2}{2}\|\varphi''\|_\infty \le \frac{x^2}{8}.$$

□

Proof of Theorem 2.8. Let $(X_1, \cdots, X_n)$ be a random variable on $\mathcal{X}$ with distribution P. We shall prove that

$$\mathbb{P}(f(X_1, \cdots, X_n) - \mathbb{E}f(X_1, \cdots, X_n) \ge t) \le \exp\left(\frac{-t^2}{2\|c\|_2^2}\right).$$

For integer $1 \le k \le n$, let $\mathcal{F}_k = \sigma(X_1, \cdots, X_k)$, $Z_0 = \mathbb{E}f(X_1, \cdots, X_n)$, $Z_k = \mathbb{E}[F(X_1, \cdots, X_n)|\mathcal{F}_k]$, $Z_n = f(X_1, \cdots, X_n)$. We also define $Y_k = Z_k - Z_{k-1}$, so that $\mathbb{E}[Y_k|\mathcal{F}_{k-1}] = 0$. Finally, let

$(X_1', \cdots, X_n')$ be an independent copy of $(X_1, \cdots, X_n)$. If $\mathbb{E}'$ denote the expectation over $(X_1', \cdots, X_n')$, we have

$$Z_k = \mathbb{E}' f(X_1, \cdots, X_k, X_{k+1}', \cdots, X_n').$$

It follows by (2.7)

$$Y_k = \mathbb{E}' f(X_1, \cdots, X_k, X_{k+1}', \cdots, X_n') - \mathbb{E}' f(X_1, \cdots, X_{k-1}, X_k', \cdots, X_n') \in [-c_k, c_k].$$

Since $\mathbb{E}[Y_k | \mathcal{F}_{k-1}] = 0$, we may apply Lemma 2.11: for every $\lambda \geq 0$,

$$\mathbb{E}[e^{\lambda Y_k} | \mathcal{F}_{k-1}] \leq e^{\frac{\lambda^2 c_k^2}{2}}.$$

This estimates does not depend on $\mathcal{F}_{k-1}$, it follows that

$$\mathbb{E} e^{\lambda(Z_n - Z_0)} = \mathbb{E}[e^{\lambda \sum_{k=1}^n Y_k}] \leq e^{\frac{\lambda^2 \|c\|_2^2}{2}}.$$

From Chernov bound, for every $\lambda \geq 0$,

$$\mathbb{P}(f(X_1, \cdots, X_n) - \mathbb{E} f(X_1, \cdots, X_n) \geq t) \leq \exp\left(-\lambda t + \frac{\lambda^2 \|c\|_2^2}{2}\right).$$

Optimizing over the choice of λ, we choose $\lambda = t / \|c\|_2^2$. □

Random matrices with independent half-rows

The proof of Theorem 2.2 and Theorem 2.8 imply the following statement.

Theorem 2.9 (Concentration of ESD with independent half-rows)**.** *Let $X \in H_n(\mathbb{C})$ be an hermitian random matrix and let for $1 \leq k \leq n$, $X_k = (X_{kj})_{1 \leq j \leq k} \in \mathbb{C}^k$. If the variables $(X_k)_{1 \leq k \leq n}$ are independent, then for every $f \in \mathrm{BV}^+$ and $t \geq 0$,*

$$\mathbb{P}\left(\left|\int f \, d\mu_X - \mathbb{E} \int f \, d\mu_X\right| \geq t\right) \leq 2 \exp\left(-\frac{nt^2}{8\|f\|_{\mathrm{BV}}^2}\right).$$

3.2 Logarithmic Sobolev inequality

We are now going to derive optimal concentration inequalities. We follow Section 2.3.2 in [AGZ10].

Definition 2.2 (Logarithmic Sobolev inequality (LSI))**.** *A random variable X on $\mathbb{R}^n$ satisfies LSI with constant c if for any differentiable square integrable function f*

$$\mathrm{Ent}_X(f(X)^2) = \mathbb{E} f(X)^2 \log\left(\frac{f(X)^2}{\mathbb{E} f(X)^2}\right) \leq 2c \, \mathbb{E} \|\nabla f(X)\|_2^2,$$

where $\|\nabla f\|_2^2 = \sum_{i=1}^n (\partial_i f)^2$.

Recall that the entropy bound which is naturally related to transport inequalities. For example, the Pinsker's inequality relates the entropy to the total variation distance between two probability measures :

$$d_{\mathrm{TV}}(\mu,\mu') = \sup_A |\mu(A) - \mu(A')| \le \sqrt{2\mathrm{Ent}_\mu\left(\frac{\partial \mu'}{\partial \mu}\right)}.$$

The definition of LSI is due to Léonard Gross 1975. It bounds the entropy by an energy. It is closely related to hypercontractivity in semi-group theory. Refer to Ané et al. [ABC+00], to Ledoux [Led01]. For techniques to prove LSI, see also Villani Chap. 21-22 "optimal transport, old and new" and Guionnet and Zegarlinski [GZ03]. It is not difficult to check that a variable which satisfies LSI satisfies Poincaré with the same constant.

For us, it will be important that the standard Gaussian on $\mathbb{R}$ satisfies LSI(1). More generally, let $V : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ such that $V(x) - c\|x\|_2^2/2$ is convex. Then the probability $P(dx) = Z^{-1}e^{-V(x)}dx$ satisfies LSI(c), it is a consequence of Bakry-Émery criterion, see Bobkov-Ledoux 2000.

Lemma 2.12 (Properties of LSI)**.** *Let X_1, X_2 be independent random variables in $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ which satisfy* LSI *with constants c_1, c_2.*

- Homogeneity*: for $a \in \mathbb{R}$, $b \in \mathbb{R}^{n_i}$, $aX_i + b$ satisfies* LSI(a^2c)*.*

- Tensorization*: (X_1, X_2) in $\mathbb{R}^{n_1+n_2}$ satisfies* LSI($\max(c_1, c_2)$)*.*

Proof. Only the tensorization property deserves a proof. It is due to Han's inequality which implies the subbadivity of the relative entropy. Set $X = (X_1, X_2)$ and $Z = f^2(X)$ we denote by $\mathbb{E}_i$ the conditional expectation given X_j, $j \neq i$ and $\mathrm{Ent}_i(Z) = \mathbb{E}_i Z \log(Z/\mathbb{E}Z_i)$ the conditional entropy. We decompose the entropy as

$$\begin{aligned}
\mathrm{Ent}(Z) &= \mathbb{E}Z \log \frac{Z}{\mathbb{E}Z} \\
&= \mathbb{E}_2\left(\mathbb{E}_1\left(Z \log \frac{Z}{\mathbb{E}_1 Z}\right)\right) + \mathbb{E}_2\left((\mathbb{E}_1 Z) \log \frac{\mathbb{E}_1 Z}{\mathbb{E}Z}\right) \\
&= \mathbb{E}_2(\mathrm{Ent}_1(Z_1)) + \mathrm{Ent}_2(\mathbb{E}_1 Z).
\end{aligned}$$

Recall the variational formula for the entropy :

$$\mathrm{Ent}(Z) = \sup\left\{\mathbb{E}(Zh(X)) \,:\, \mathbb{E}\left(e^{h(X)}\right) = 1\right\},$$

(which follows from the inequality : $xy \le x\log x - x + e^y$, for any $x > 0$, $y \in \mathbb{R}$ which is applied to $x = Z/\mathbb{E}Z$ and $y = h(X)$)). This yields to, for any function $g : \mathbb{R}^{m_2} \to \mathbb{R}$ with $\mathbb{E}_2 e^{g(X_2)} = 1$,

$$\mathbb{E}_2((\mathbb{E}_1 Z)g(X_2)) = \mathbb{E}_1(\mathbb{E}_2(Zg(X_2))) \le \mathbb{E}_1(\mathrm{Ent}_2(Z)).$$

Taking the supremum over all g, yields the tensorization inequality for the entropy:

$$\mathrm{Ent}(Z) \leq \mathbb{E}_2(\mathrm{Ent}_1(Z)) + \mathbb{E}_1(\mathrm{Ent}_2(Z)) = \mathbb{E}(\mathrm{Ent}_1(Z)) + \mathbb{E}(\mathrm{Ent}_2(Z)).$$

We may now apply the LSI(c_i), for $i \in \{1,2\}$, and the statement of the lemma is straightforward. □

Lemma 2.13 (Herbst's argument). *Assume that X satisfies* LSI(c). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz with constant* 1. *Then for all $\lambda \in \mathbb{R}$,*

$$\mathbb{E}e^{\lambda(f(X)-\mathbb{E}f(X))} \leq e^{2c\lambda^2}.$$

and so for any $t \geq 0$,

$$\mathbb{P}\left(|f(X) - \mathbb{E}f(X)| \geq t\right) \leq 2e^{-\frac{t^2}{8c}}.$$

Proof. We use Chernoff bound, for $\lambda \in \mathbb{R}$,

$$\mathbb{P}\left(f(X) - \mathbb{E}f(X) \geq t\right) \leq e^{-\lambda t}\mathbb{E}e^{\lambda(f(X)-\mathbb{E}f(X))}.$$

Applied $\lambda = t/(4c)$, f and $-f$, we deduce that the second statement is a consequence of the first statement. By a density, we may assume that f continuously differentiable (refer to [ABC+00, lemma 7.3.3]). By homogeneity, we can assume $\mathbb{E}f(X) = 0$ and $\lambda > 0$. We set $Z = f(X)$. Consider the log-Laplace function

$$\Lambda(\lambda) = \log \mathbb{E}e^{2\lambda Z}.$$

Apply the definition of LSI to the function $f = e^{\lambda Z}$. We find

$$2\lambda\mathbb{E}f(X)e^{2\lambda Z} - \mathbb{E}e^{2\lambda Z}\log\mathbb{E}e^{2\lambda Z} \leq 2c\mathbb{E}\sum_{i=1}^{n}\left|(\partial_i Z)\lambda e^{\lambda Z}\right|^2 \leq 2cL^2\mathbb{E}e^{2\lambda Z},$$

where $L^2 = \max_{x\in\mathbb{R}^n}\|\nabla F(x)\|_2^2 = \max_{x\in\mathbb{R}^n}\sum_{i=1}^n|\partial_i F(x)|^2 \leq 1$. We observe that

$$\left(\frac{\Lambda(\lambda)}{\lambda}\right)' = 2\frac{\mathbb{E}Ze^{2\lambda Z}}{\lambda\mathbb{E}e^{2\lambda Z}} - \frac{\log\mathbb{E}e^{2\lambda Z}}{\lambda^2}.$$

It yields to

$$\left(\frac{\Lambda(\lambda)}{\lambda}\right)' \leq 2c.$$

Since $\mathbb{E}Z = 0$, $\Lambda(\lambda) = o(\lambda)$ as $\lambda \downarrow 0$, we find for all $\lambda > 0$,

$$\Lambda(\lambda) \leq 2c\lambda^2.$$

This proves the first statement. □

We can now derive powerful concentration inequalities for random matrices with independent entries. From Lemma 2.13 and Corollary 2.2, we find the following.

Theorem 2.10 (Concentration of ESD with LSI)**.** *Let $X \in H_n(\mathbb{C})$ be an hermitian random matrix and assume that the variable $(X_{ij})_{i\leq j}$ satisfies* LSI(c) *in $\mathbb{R}^{n^2}$. Then for any $f:\mathbb{R}\to\mathbb{R}$ such that $\|f\|_{\mathrm{L}} \leq 1$ and every $t \geq 0$,*

$$\mathbb{P}\left(\left|\int f\, d\mu_{X/\sqrt{n}} - \mathbb{E}\int f\, d\mu_{X/\sqrt{n}}\right| \geq t\right) \leq 2\exp\left(-\frac{n^2t^2}{16c}\right).$$

We can also derive a result for the concentration of single eigenvalues (the same comment below Theorem 2.7 applies here also).

Theorem 2.11 (Concentration of single eigenvalue with LSI)**.** *Let $X \in H_n(\mathbb{C})$ be an hermitian random matrix and assume that the random variable $(X_{ij})_{i\leq j}$ satisfies* LSI(c) *in $\mathbb{R}^{n^2}$. Then for any $1 \leq k \leq n$ and every $t \geq 0$,*

$$\mathbb{P}\left(|\lambda_k(X) - \mathbb{E}\lambda_k(X)| \geq t\right) \leq 2\exp\left(-\frac{t^2}{16c}\right).$$

3.3 With Talagrand's concentration inequality

We start by recalling Talagrand's concentration inequality.

Theorem 2.12 (Talagrand's concentration inequality)**.** *Let $\mathcal{B}$ be the unit ball of $\mathbb{C}$. Consider a convex Lipshitz real-valued function f defined on $\mathcal{B}^n$. Let $X = (X_1, \ldots, X_n) \in \mathcal{B}^n$ a vector of independent variables and let $m(f)$ be the median of $f(X)$. Then for any $t > 0$,*

$$\mathbb{P}(|f(X) - m(f)| \geq t) \leq 4\exp\left(-\frac{t^2}{8\|f\|_L^2}\right).$$

For a proof in the real case see Ledoux [Led01] (with constant 4 instead of 8). As in the proof of Theorem 2.3, we can use the subadditivity of relative entropy proved in Lemma 2.12 together with Herbst' argument Lemma 2.13 to obtain an analog bound for the upper tail (see [BLM13] for details). Oddly, this argument does not seem to give the bound for the lower tail (the assumption f convex is not symmetric).

With the assumption of the theorem, we find

$$\begin{aligned} |m(f) - \mathbb{E}(f(X))| &\leq \int_0^\infty \mathbb{P}(|f(X) - m(f)| \geq t)dt \\ &\leq 2\int_{-\infty}^\infty e^{-\frac{t^2}{8\|f\|_L^2}}\, dt \\ &= 4\sqrt{2\pi}\|f\|_L. \end{aligned}$$

At the cost of changing the constants, in Talagrand's Theorem 2.12, we may replace $m(f)$ by $\mathbb{E}f(X)$.

Application to random matrices

Talagrand inequality readily implies the following concentration inequalities.

Theorem 2.13. *Let $X \in H_n(\mathbb{C})$ be an random hermitian matrix such that the variables $(X_{ij})_{i\leq j}$ are independent and bounded by D. Then, for any $t \geq 0$,*

$$\mathbb{P}\left(|\lambda_1(X) - m| \geq t\right) \leq 4\exp\left(-\frac{t^2}{16D^2}\right),$$

where m is the median of $\lambda_1(X)$.

As in the proof of Theorem 2.5, we get the following consequence on linear statistics of the eigenvalues.

Theorem 2.14. *Let $X \in H_n(\mathbb{C})$ be an random hermitian matrix such that the variables $(X_{ij})_{i\leq j}$ are independent and bounded by D. Then for any C^1-function $f : \mathbb{R} \to \mathbb{R}$ with k inflection points, for every $t \geq 0$,*

$$\mathbb{P}\left(\left|\int f\, d\mu_{X/\sqrt{n}} - \mathbb{E}\int f\, d\mu_{X/\sqrt{n}}\right| \geq t\right) \leq c(k+1)\exp\left(-\frac{n^2t^2}{c^2(k+1)^2D^2\|f\|_{\mathrm{L}}^2}\right),$$

where $c > 0$ is a universal constant.

Proof. First observe from Lemma 2.8 and the fact:

$$\{x_1 + \cdots + x_\ell \geq t\} \subset \bigcup_{1\leq i\leq \ell}\left\{x_i \geq \frac{t}{\ell}\right\},$$

that it is sufficient to prove the statement for f convex. The latter is a consequence of Theorem 2.12, Corollary 2.2 and Lemma 2.7. □

Lecture 3

Resolvent of random matrices

In Lecture 1, we have seen that the even moments of Wigner's semi-circular law are given by the Catalan number. The generating function of the Catalan's number satisfies a very simple fixed point equation (1.2). This hints that the generating function of moments of ESD of random matrices could be easier to compute than the actual moments. The *resolvent method* formalizes this ideas.

1 Cauchy-Stieltjes transform

1.1 Definition and properties

Let μ be a finite measure on $\mathbb{R}$. Define its *Cauchy-Stieltjes transform* as for all $z \in \mathbb{C}_+ = \{z \in \mathbb{C} : \Im(z) > 0\}$,

$$g_\mu(z) = \int \frac{1}{\lambda - z} d\mu(\lambda).$$

Note that if μ has bounded support we have

$$g_\mu(z) = -\sum_{k \geq 0} \frac{1}{z^{k+1}} \int \lambda^k d\mu(\lambda).$$

The Cauchy-Stieltjes transform is thus essentially the generating function of the moments of the measure μ.

Lemma 3.1 (Properties of Cauchy-Stieltjes transform)**.** *Let μ be a finite measure on $\mathbb{R}$ with mass $\mu(\mathbb{R}) \leq 1$.*

(i) Analytic : *the function g_μ is an analytic function from $\mathbb{C}_+ \to \mathbb{C}_+$.*

(ii) Bounded : *for any $z \in \mathbb{C}_+$, $|g_\mu(z)| \leq (\Im(z))^{-1}$.*

The Cauchy-Stieltjes transform characterizes the measure. More precisely, the following holds.

Lemma 3.2 (Inversion of Cauchy-Stieltjes transform)**.** *Let μ be a finite measure on $\mathbb{R}$.*

(i) For any bounded continous f,

$$\int f d\mu = \lim_{t \downarrow 0} \frac{1}{\pi} \int f(x) \Im g_\mu(x+it) dx.$$

(ii) For any $x \in \mathbb{R}$,

$$\mu(\{x\}) = \lim_{t \downarrow 0} t \Im g_\mu(x+it).$$

(iii) For almost all $x \in \mathbb{R}$, the density of μ at x is equal to

$$\lim_{t \downarrow 0} \frac{1}{\pi} \Im g_\mu(x+it).$$

Proof. By linearity, we can assume that μ is probability measure. We have the identity

$$\Im g(x+it) = \int \frac{t}{(\lambda - x)^2 + t^2} d\mu(\lambda).$$

Hence $\frac{1}{\pi} \Im g(x+it)$ is the equal to density at x of the distribution $(\mu * P_t)$, P_t is a Cauchy distribution with density

$$P_t(x) = \frac{t}{\pi(x^2+t^2)}.$$

In other words,

$$\frac{1}{\pi} \int f(x) \Im g_\mu(x+it) dx = \mathbb{E} f(X + tY),$$

where X has law μ and is independent of Y with distribution P_1. The statements follow easily. □

1.2 Cauchy-Stieltjes transform and weak convergence

The convergence of Cauchy-Stieltjes transform is equivalent to the weak convergence.

Corollary 3.1. *Let μ and $(\mu_n)_{n \geq 1}$ be a sequence of real probability measures. The following are equivalent*

(i) As $n \to \infty$, weakly $\mu_n \to \mu$.

(ii) For all $z \in \mathbb{C}_+$, as $n \to \infty$, $g_{\mu_n}(z) \to g_\mu(z)$.

(iii) There exists a set $D \subset \mathbb{C}_+$ with an accumulation point in $\mathbb{C}_+$ such that for all $z \in D$, as $n \to \infty$, $g_{\mu_n}(z) \to g_\mu(z)$.

Proof. Statement "(i) implies (ii)" follows from the definition of weak convergence applied to the real and imaginary part of $f(\lambda) = (\lambda - z)^{-1}$. Statement "(ii) implies (iii)" is trivial. For statement "(iii) implies (i)", from Helly selection theorem, the sequence (μ_n) is relatively compact for the vague convergence. Let ν be such vague limit, it is a finite measure with mass at most 1. By assumption, for any $z \in D$, $g_\mu(z) = g_\nu(z)$. Two analytic functions equal on a set with an accumulation point are equal on their domain (principle of analytic extension). Hence, by lemma 3.1, for any $z \in \mathbb{C}_+$, $g_\mu(z) = g_\nu(z)$. By lemma 3.2, we deduce that ν is a probability measure and $\nu = \mu$. □

1.3 Cauchy-Stieltjes transform of the semicircle law

The Cauchy-Stieltjes semi-circular distribution μ_{sc} satisfies the fixed point for all $z \in \mathbb{C}_+$,

$$g_{\mu_{sc}}(z) = -\frac{1}{z + g_{\mu_{sc}}(z)} \quad \text{or} \quad g_{\mu_{sc}}(z)^2 + z g_{\mu_{sc}}(z) + 1 = 0. \tag{3.1}$$

Let $z \mapsto \sqrt{z}$ be the analytical continuation of $x \mapsto \sqrt{x}$ on $\mathbb{C} \backslash \mathbb{R}_-$ with a positive imaginary part. We find

$$g_{\mu_{sc}}(z) = \frac{-1 + \sqrt{1 - 4z}}{2}.$$

2 Resolvent

2.1 Spectral measure at a vector

Let $A \in H_n(\mathbb{C})$ and $\psi \in \mathbb{C}^n$ be a vector with unit ℓ_2-norm, $\|\psi\|_2 = 1$.

The spectral theorem guarantees the existence of $(v_1, \cdots, v_n)$, an orthonormal basis of $\mathbb{C}^n$ of *eigenvectors* of A, that is, for any $1 \leq i \leq n$, $Av_i = \lambda_i(A) v_i$. The *spectral measure with vector* ψ is the real probability measure defined by

$$\mu_A^\psi = \sum_{k=1}^n |\langle v_k, \psi \rangle|^2 \delta_{\lambda_k(A)}. \tag{3.2}$$

It may also be defined as the unique probability measure μ_A^ϕ such that

$$\int \lambda^k d\mu_A^\phi(\lambda) = \langle \psi, A^k \psi \rangle \qquad \text{for all integers } k \geq 1. \tag{3.3}$$

If $(e_1, \cdots, e_n)$ is the canonical basis of $\mathbb{C}^n$, summing (3.2), we find, with $\psi_j = \langle \psi, e_j \rangle$,

$$\mu_A^\psi = \sum_{j=1}^n |\psi_j|^2 \mu_A^{e_j} \quad \text{and} \quad \mu_A = \frac{1}{n} \sum_{j=1}^n \mu_A^{e_j}. \tag{3.4}$$

2.2 Resolvent matrix

If $A \in H_n(\mathbb{C})$ and $z \in \mathbb{C}_+ = \{z \in \mathbb{C} : \Im(z) > 0\}$, then $A - zI$ is invertible. We define *the resolvent* of A as the function $R : \mathbb{C}_+ \mapsto M_n(\mathbb{C})$,

$$R(z) = (A - zI)^{-1}.$$

We have the identity

$$\langle \psi, R(z)\psi \rangle = \int \frac{d\mu_A^\psi(\lambda)}{\lambda - z} = g_{\mu_A^\psi}(z), \tag{3.5}$$

where μ_A^ψ is the spectral measure with vector ψ. Also,

$$g_{\mu_A}(z) = \frac{1}{n}\mathrm{Tr}(R(z)).$$

Lemma 3.3 (Basic properties of the resolvent matrix)**.** *Let $A \in H_n(\mathbb{C})$ and $R(z) = (A - zI)^{-1}$ be its resolvent. For any $z \in \mathbb{C}_+$, $1 \leq i, j \leq n$,*

(i) Analytic : $z \mapsto R(z)_{ij}$ *is an analytic function on* $\mathbb{C}_+ \to \mathbb{C}_+$.

(ii) Bounded : $\|R(z)\| \leq \Im(z)^{-1}$.

(iii) Normal : $R(z)R(z)^* = R(z)^*R(z)$.

Proof. All properties come from the decomposition

$$R(z) = \sum_{k=1}^{n} \frac{v_k v_k^*}{\lambda_k(A) - z},$$

where $(v_1, \ldots, v_n)$ is an orthogonal basis of eigenvectors of A. □

2.3 Resolvent identity and perturbation inequalities

Let A, B in $H_n(\mathbb{C})$. For $z \in \mathbb{C}_+$, we denote their resolvent by $R_A(z) = (A - zI_n)^{-1}$ and $R_B(z) = (B - zI_n)^{-1}$. For invertible matrices M, N, the identity

$$M^{-1} = N^{-1} + M^{-1}(N - M)N^{-1} = N^{-1} + N^{-1}(N - M)M^{-1}$$

implies the *resolvent identity*:

$$R_A = R_B + R_A(B - A)R_B = R_B + R_B(B - A)R_A. \tag{3.6}$$

The following lemma strengthens Equation (2.6).

Lemma 3.4 (Pertubation of resolvent)**.** *Let A, B in $H_n(\mathbb{C})$. Then, if $z \in \mathbb{C}_+$, $R_A(z) = (A - zI_n)^{-1}$ and $R_B(z) = (B - zI_n)^{-1}$,*

$$\sum_{k=1}^{n} |R_A(z)_{kk} - R_B(z)_{kk}| \leq 2\frac{\operatorname{rank}(A-B)}{\Im z}$$

and

$$\|R_A - R_B\| \leq \frac{\|A-B\|}{\Im(z)^2}.$$

Proof. The second statement is an obvious consequence of the resolvent identity. For the first statement, the resolvent identity asserts that

$$M = R_A - R_B = R_A(B-A)R_B.$$

It follows that $r = \operatorname{rank}(M) \leq \operatorname{rank}(A-B)$. We notice also that $\|M\| \leq 2\Im(z)^{-1}$. Hence, in the singular value decomposition of $M = UDV$, at most r entries of $D = \operatorname{diag}(s_1, \cdots, s_n)$ are non zero and they are bounded by $\|M\|$. We denote by $u_1, \cdots, u_r$ and $v_1, \cdots, v_r$ the associated orthonormal vectors so that

$$M = \sum_{i=1}^{r} s_i u_i v_i^*,$$

and

$$|R_A(z)_{kk} - R_B(z)_{kk}| = |M_{kk}| = \left| \sum_{i=1}^{r} s_i \langle u_i, e_k \rangle \langle v_i, e_k \rangle \right| \leq \|M\| \sum_{i=1}^{r} |\langle u_i, e_k \rangle| |\langle v_i, e_k \rangle|.$$

We obtain from Cauchy-Schwarz,

$$\begin{aligned}\sum_{k=1}^{n} |R_A(z)_{kk} - R_B(z)_{kk}| &\leq \|M\| \sum_{i=1}^{r} \sqrt{\sum_{k=1}^{n} |\langle u_i, e_k \rangle|^2} \sqrt{\sum_{k=1}^{n} |\langle v_i, e_k \rangle|^2} \\ &= r\|M\|.\end{aligned}$$

Its proves the first claim. □

As a corollary, we have a concentration inequality for the resolvent.

Corollary 3.2 (Concentration of diagonal resolvent entries)**.** *Let $X \in H_n(\mathbb{C})$ be an hermitian random matrix and let for $1 \leq k \leq n$, $X_k = (X_{kj})_{1 \leq j \leq k} \in \mathbb{C}^k$. If the variables $(X_k)_{1 \leq k \leq n}$ are independent, then for every complex valued functions $(f_1, \ldots, f_k)$ with Lipschitz constants 1 and $t \geq 0$,*

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{k=1}^{n} f_k(R_{kk}) - \mathbb{E}\frac{1}{n} \sum_{k=1}^{n} f_k(R_{kk}) \right| \geq t \right) \leq 2\exp\left(-\frac{n\Im(z)^2 t^2}{32} \right).$$

where $R = (X - zI)^{-1}$ is the resolvent of X at $z \in \mathbb{C}_+$.

Proof. By Lemma 3.4, if A and B differs only the i-th row and column, we have the inequality

$$\left|\frac{1}{n}\sum_{k=1}^{n} f_k((R_A)_{kk}) - \frac{1}{n}\sum_{k=1}^{n} f_k((R_B)_{kk})\right| \leq \frac{1}{n}\sum_{k=1}^{n} |(R_A)_{kk} - (R_B)_{kk}| \leq \frac{4}{n\Im z}.$$

It remains to argue as in the proof of Theorem 2.2 and use Theorem 2.8. □

2.4 Resolvent complement formula

The Schur complement is simply a block inversion by part of an invertible matrix.

Lemma 3.5 (Schur's complement formula)**.** *Let $A \in M_n(\mathbb{C})$ be an invertible matrix. Set*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

where $A_{11}, B_{11} \in M_p(\mathbb{C})$. Then, if A_{22} and B_{11} are invertible, we have

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}.$$

An immediate consequence is the following formula:

Corollary 3.3 (Resolvent complement formula)**.** *Let $n \geq 2$, $A = (A_{ij})_{1\leq i,j\leq n} \in H_n(\mathbb{C})$, $z \in \mathbb{C}_+$ and $R = (A - zI_n)^{-1}$. For any $1 \leq i \leq n$,*

$$R_{ii} = -(z - A_{ii} + \langle A_i, R_i A_i\rangle)^{-1},$$

where $A_i = (A_{ij})_{j\neq i} \in \mathbb{C}^{n-1}$, $R_i = (A^i - zI_{n-1})^{-1}$ and $A^i \in H_{n-1}(\mathbb{C})$ is the principal minor of A where the i-th row and column have been removed.

3 Resolvent method for random matrices

In this section, we will present on an exemple the resolvent method for random matrix. This method will be based on two components : a probabilistic component, the concentration of bilinear forms and a linear algebra component, the Schur complement formula.

3.1 Concentration for bilinear forms

Lemma 3.6 (Variance of Bilinear form of independent vectors)**.** *Let $A \in M_n(\mathbb{C})$ and $X = (X_1, \cdots, X_n) \in \mathbb{C}^n$ be a vector of centered and independent variables with $\mathbb{E}|X_i|^2 \leq 1$ and $\mathrm{Var}(|X_i|^2) \leq K$ for $1 \leq i \leq n$. Then*

$$\mathbb{E}\langle X, AX\rangle = \sum_{i=1}^{n} \mathbb{E}|X_i|^2 A_{ii} \quad \text{and} \quad \mathrm{Var}\langle X, AX\rangle \leq 2\mathrm{Tr}AA^* + K\sum_{i=1}^{n} |A_{ii}|^2.$$

Proof. We have

$$\langle X, AX\rangle = \sum_{1\le i,j\le n} \bar{X}_i A_{ij} X_j.$$

This yields to

$$\mathbb{E}\langle X, AX\rangle = \sum_{i=1}^{n} \mathbb{E}|X_i|^2 A_{ii}$$

and

$$\mathrm{Var}\langle X, AX\rangle = \sum_{i_1,j_1,i_2,j_2} A_{i_1j_1}\bar{A}_{i_2j_2}\mathbb{E}\bar{X}_{i_1}X_{j_2}X_{i_2}\bar{X}_{j_2} - \sum_{i,j}\mathbb{E}|X_i|^2\mathbb{E}|X_j|^2 A_{ii}\bar{A}_{jj}.$$

The first sum is non zero only if $(i_1, i_2) = (j_1, j_2)$, $(i_1, j_1) = (i_2, j_2)$ or $(i_1, j_1) = (j_2, i_2)$. We get

$$\begin{aligned}\mathrm{Var}\langle X, AX\rangle &= \sum_i \mathrm{Var}(|X_i|^2)|A_{ii}|^2 + \sum_{i\ne j}|A_{ij}|^2\mathbb{E}|X_i|^2\mathbb{E}|X_j|^2 + \sum_{i\ne j} A_{ij}\bar{A}_{ji}\mathbb{E}\bar{X}_i^2\mathbb{E}X_j^2 \\ &\le K\sum_{i=1}^{n}|A_{ii}|^2 + \sum_{1\le i,j\le n}|A_{ij}|^2 + \sum_{1\le i,j\le n}|A_{ij}||A_{ji}|.\end{aligned}$$

The second term is equal $\mathrm{Tr}(AA^*)$, while the third term is upper bounded by $\mathrm{Tr}(AA^*)$ from Cauchy-Schwarz inequality. □

With more moments assumption, it is of course possible to strengthen lemma 3.6. For example, for entries with sub-Gaussian tail, this is topic of the Hanson-Wright theorem.

3.2 Random matrices with variance profile

We illustrate the resolvent method for random matrices with inhomogeneous variances. For each integer $n \ge 1$, we assume that $(Y_{ij})_{1\le i\le j}$ are independent centered variables with variance

$$\mathbb{E}|Y_{ij}|^2 = \frac{1}{n}\left(\int_{Q_{ij}} \frac{\sigma^2(x,y)}{|Q_{ij}|}dxdy + \delta_{ij}(n)\right), \tag{3.7}$$

where $Q_{ij} = [(i-1)/n, i/n] \times [(j-1)/n, j/n]$, $\sigma : [0,1]^2 \to [0,1]$ is a measurable function such that $\sigma(x, y) = \sigma(y, x)$ and $\delta_{ij}(n)$ is a vanishing sequence. The function σ is fixed but the law of Y_{ij} depends n. We set $Y_{ji} = \bar{Y}_{ij}$ and we consider the hermitian matrix

$$Y_n = (Y_{ij})_{1\le i,j\le n}.$$

If σ is continuous, Equation (3.7) asserts that the variance of $\sqrt{n}Y_{ij}$ is roughly $\sigma^2(i/n, j/n)$. We assume that all these matrices are defined on a common probability space.

Theorem 3.1 (ESD of matrices with variance profile)**.** *Assume that for all i, j, $\mathbb{E}Y_{ij} = 0$, $\mathbb{E}|Y_{ij}|^2$ as in* (3.7) *with $|\delta_{ij}(n)| \leq \delta(n)$ and $\mathbb{E}|Y_{ij}|^4 \leq \delta(n)/n$ for some sequence $\delta(n)$ going to* 0*. Then, there exists a probability measure μ_σ depending on σ such that a.s. weakly*

$$\mu_Y \to \mu_\sigma.$$

The Cauchy-Stieltjes transform g_{μ_σ} of μ_σ is given by the formula

$$g_{\mu_\sigma}(z) = \int_0^1 g(x, z)dx,$$

where the $[0, 1] \times \mathbb{C}_+ \to \mathbb{C}_+$ map, $g : (x, z) \mapsto g(x, z)$ satisfies : for a.a. $x \in [0, 1]$, $z \mapsto g(x, z)$ analytic on $\mathbb{C}_+$ and for each $z \in \mathbb{C}_+$ with $\Im(z) > 1$, $x \mapsto g(x, z)$ is the unique function in $L^1([0, 1]; \mathbb{C}_+)$ solution of the equation, for a.a. $x \in [0, 1]$,

$$g(x, z) = -\left(z + \int_0^1 \sigma^2(x, y)g(y, z)dy\right)^{-1}. \tag{3.8}$$

Note that by analyticity, for $0 < \Im(z) \leq 1$, $x \mapsto g(x, z)$ is also a solution of (3.8) (which may however a priori not be unique in $L^1([0, 1]; \mathbb{C}_+)$). A typical application of Theorem 3.1 is the following.

Corollary 3.4 (ESD of generalized Wigner matrices)**.** *Assume that for a.a. $x \in [0, 1]$, $\int_0^1 \sigma^2(x, y)dx = 1$. Then, a.s. weakly*

$$\mu_Y \to \mu_{sc}.$$

The above corollary follows from noticing that the semicircle law satisfies the fixed point equation (3.1) and the unicity statement in Theorem 3.1. The following exercise is an easy Corollary of 3.4.

Exercise 3.1 (Adjacency matrix of Erdős-Rényi graphs)**.** *Consider the adjacency matrix A of $G(n, d/n)$ with $0 \leq d(n) \leq n$ and $d(n) \to \infty$. Namely $(A_{ij})_{1\leq i<j\leq n}$ are i.i.d. $\{0, 1\}$-Bernoulli random variables with mean d/n. Prove that a.s. weakly $\mu_{A/\sqrt{d}} \to \mu_{sc}$.*

Let $1 \leq p \leq n$ and define the matrix in $M_{p,n}(\mathbb{R})$,

$$X = (X_{ij})_{1\leq i\leq p, 1\leq j\leq n}.$$

where (X_{ij}) are iid random variables.

Corollary 3.5 (Marcenko-Pastur law)**.** *Assume $p(n)/n \to c \in (0, 1]$, $\mathbb{E}X_{11} = 0$ and $\mathbb{E}|X_{11}|^2 = 1$, then a.s weakly*

$$\mu_{XX^*/n} \to \mu_c.$$

where, $b_- = (1-\sqrt{c})^2$*,* $b_+ = (1+\sqrt{c})^2$ *and*

$$\mu_c(dx) = \frac{1}{2\pi x}\sqrt{(x-b_-)(b_+-x)}\mathbf{1}_{b_-\leq x\leq b_+}dx$$

Note that from the above corollary, it is also possible to deal with case $c > 1$. It suffices to reverse the role of p and n and notice that, for $1 \leq p \leq n$

$$n\mu_{X^*X} = p\mu_{XX^*} + (n-p)\delta_0.$$

Proof of Corollary 3.5. By a trunction argument, we may assume that the variables are uniformly bounded. Consider the block matrix in $H_{n+p}(\mathbb{C})$,

$$Z = \begin{pmatrix} 0_n & X^* \\ X & 0_p \end{pmatrix}. \tag{3.9}$$

If $0 \leq \lambda_1 \leq \cdots \leq \lambda_p$ are the eigenvalues of XX^* with $\lambda_1 = \cdots = \lambda_m = 0$, $\lambda_{m+1} > 0$, then the non-zero eigenvalues of Z are

$$\pm\sqrt{\lambda_{m+1}}, \cdots, \pm\sqrt{\lambda_p}.$$

In particular, if $\mu_{Z/\sqrt{n+p}}$ converges weakly toward a limit measure ν with

$$\nu = \frac{1-c}{1+c}\delta_0 + \frac{2c}{1+c}\hat{\nu},$$

where $\hat{\nu}$ is a symmetric probability measure on $\mathbb{R}$ with density f, then $\mu_{XX^*/n}$ converges weakly to μ with density on $(0,\infty)$ given by

$$d\mu(x) = \frac{f(\sqrt{x})}{\sqrt{x}}dx.$$

Now, coming back to (3.9), we introduce the $[0,1]^2 \to [0,1]$ function

$$\sigma(x,y) = \mathbf{1}\Big(0 < x < \frac{1}{1+c}\Big)\mathbf{1}\Big(\frac{1}{1+c} < y < 1\Big) + \mathbf{1}\Big(\frac{1}{1+c} < x < 1\Big)\mathbf{1}\Big(0 < y < \frac{1}{1+c}\Big).$$

Note that $\sigma(x,y) = \sigma(y,x)$ and $Y = Z/\sqrt{n+p}$ satisfies the assumptions of Theorem 3.1 with σ and $\delta = O(1/n)$. It is an exercise to compute explicitly μ_σ in this case. $\square$

3.3 Proof of Theorem 3.1

The proof of Theorem 3.1 is a typical instance of the resolvent method. In the first step of the proof, we check tightness of μ_Y and that for each $z \in \mathbb{C}_+$, a.s.

$$\frac{1}{n}\sum_{i=1}^{n} R_{ii}(z) \to \int_0^1 g(x,z)dx, \tag{3.10}$$

where, R is the resolvent of Y and for each $z \in \mathbb{C}_+$, $g : x \mapsto g(x,z)$ is a fixed point of the $L^1([0,1];\mathbb{C}_+) \to L^1([0,1];\mathbb{C}_+)$ map

$$F_{z,\sigma}(g)(x) = -\left(z + \int \sigma^2(x,y)g(y)dy\right)^{-1}.$$

Observe that (3.10) implies that, if D is countable dense set in $\mathbb{C}_+$, a.s. for all $z \in D$, (3.10) holds. Then, in a second step, we prove the uniqueness of the solution of $F_{z,\sigma}(g) = g$ for $\Im(z)$ large enough. Now, consider a converging subsequence of $(\mu_Y)_{n\geq 1}$ to μ. Invoking (3.5) and Corollary 3.1, we conclude that a.s. for $z \in \mathbb{C}_+$, $g_\mu(z) = \int_0^1 g(x,z)dx$. By the unicity of the limit and a new application of Corollary 3.1, it will conclude the proof of Theorem 3.1. In the sequel, the parameter $z \in \mathbb{C}_+$ is fixed and, for ease of notation, we will often omit it.

Tightness and concentration

We write

$$\mathbb{E}\int \lambda^2 d\mu_Y = \mathbb{E}\frac{1}{n}\mathrm{Tr}Y^2 \leq \frac{1}{n^2}\sum_{i,j}\left(\int_{Q_{ij}} \frac{\sigma^2(x,y)}{|Q_{ij}|}dxdy + \delta_{ij}(n)\right) = O(1).$$

Hence the sequence of probability measures $(\mathbb{E}\mu_Y)_{n\geq 1}$ is tight. By Theorem 2.9 and Borel-Cantelli Lemma, it implies that a.s., the sequence of probability measures $(\mu_Y)_{n\geq 1}$ is tight.

Moreover, Theorem 2.9 implies that it is sufficient to prove that

$$\mathbb{E}\frac{1}{n}\sum_{i=1}^{n} R_{ii} \to \int_0^1 g(x,z)dx, \tag{3.11}$$

and (3.10) will follow.

Approximation of the variance profile

Let L be a integer and let $(P_{k\ell})_{1\leq k,\ell\leq L}$ be the usual partition of $[0,1]^2$ into squares of size $1/L^2$. Define

$$\rho = \sum_{1\leq k,\ell\leq L} \rho_{k\ell}\mathbf{1}_{P_{k\ell}},$$

where $\rho_{k\ell} = L^2\int_{P_{k\ell}} \sigma(x,y)dxdy$. We define the hermitian matrix Z whose entries are, if $\mathrm{Var}(Y_{ij}) \neq 0$,

$$Z_{ij} = \frac{Y_{ij}}{\sqrt{n\mathrm{Var}(Y_{ij})}}\rho\left(\frac{i}{n},\frac{j}{n}\right), \tag{3.12}$$

and if $\mathrm{Var}(Y_{ij}) = 0$, we set $Z_{ij} = 0$. We write

$$\begin{aligned}
\mathbb{E}\frac{1}{n}\mathrm{Tr}(Z-Y)^2 &= \frac{1}{n}\mathbb{E}\sum_{i,j}|Z_{ij}-Y_{ij}|^2 \\
&= \frac{1}{n^2}\sum_{i,j}|\rho\left(\frac{i}{n},\frac{j}{n}\right) - \sqrt{n\mathrm{Var}(Y_{ij})}|^2 \\
&\le \frac{2}{n^2}\sum_{i,j}|\rho^2\left(\frac{i}{n},\frac{j}{n}\right) - n\mathrm{Var}(Y_{ij})| \\
&\le 2\int_{[0,1]^2}|\sigma^2(x,y)-\rho^2(x,y)|dxdy + O\left(\delta(n)+\frac{1}{L^2}\right),
\end{aligned}$$

From Lebesgue' Theorem, for a.a (x,y) as $L\to\infty$, $\sigma(x,y)-\rho(x,y)\to 0$. Hence, by dominated convergence we deduce that

$$\|\sigma^2-\rho^2\|_1 \to_{L\to\infty} 0$$

In particular,

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}\mathrm{Tr}(Z-Y)^2 \le \varepsilon(L), \tag{3.13}$$

for some function ε going to 0 as L goes to infinity.

Approximate fixed point equation

Consider the matrix Z given by (3.12) and denote by R its resolvent. The objective is to prove that the resolvent of Z satisfies nearly a fixed point equation. To this end, we use Schur complement formula, corollary 3.3,

$$R_{ii} = -\left(z - Z_{ii} + \langle Z_i, R^{(i)}Z_i\rangle\right)^{-1},$$

where $Z_i = (Z_{ij})_{j\neq i}$ and $R^{(i)} = (Z^{(i)}-zI)^{-1}$ is the resolvent of the minor matrix $Z^{(i)}$ obtained from Z where the i-th row and column have been removed.

Notice that if $z, w, w' \in \mathbb{C}_+$, then

$$\left|\frac{1}{z+w} - \frac{1}{z+w'}\right| \le \frac{|w-w'|}{\Im(z)^2}. \tag{3.14}$$

Since $\langle Z_i, R^{(i)}Z_i\rangle \in \mathbb{C}_+$, $R^{(i)}_{jj}\in\mathbb{C}_+$ and $n\mathbb{E}|Z_{ij}|^2 = \rho\left(\frac{i}{n},\frac{j}{n}\right)^2$, we find

$$\left|R_{ii} + \left(z + \frac{1}{n}\sum_{j\neq i}\rho\left(\frac{i}{n},\frac{j}{n}\right)^2 R^{(i)}_{jj}\right)^{-1}\right| \le \frac{1}{\Im(z)^2}\left(|Z_{ii}| + \left|\langle Z_i, R^{(i)}Z_i\rangle - \sum_{j\neq i}(\mathbb{E}|Z_{ij}|^2)R^{(i)}_{jj}\right|\right).$$

Now, by construction, the vector $(Z_{ij})_j$ is independent of $R^{(i)}$. We condition on $R^{(i)}$ and use Lemma 3.6, we deduce that in $L^2(\mathbb{P})$,

$$R_{ii} + \left(z + \frac{1}{n}\sum_{j\neq i}\rho\left(\frac{i}{n},\frac{j}{n}\right)^2 R^{(i)}_{jj}\right)^{-1} \to 0.$$

We define $N_k = \{1 \leq i \leq n : (k-1)/L < i/n \leq k/L\}$ and $N_k^{(i)} = N_k \backslash \{i\}$. So that $|N_k|/n \to 1/L$. If $i \in N_k$, it yields to, in $L^2(\mathbb{P})$,

$$R_{ii} + \left(z + \frac{1}{L} \sum_{\ell=1}^{L} \rho_{k\ell}^2 G_\ell^{(i)} \right)^{-1} \to 0,$$

where we have defined

$$G_k^{(i)} = \frac{1}{|N_k|} \sum_{j \in N_k^{(i)}} R_{jj}^{(i)} \quad \text{and} \quad G_k = \frac{1}{|N_k|} \sum_{j \in N_k} R_{jj}$$

Now, by Lemma 3.4,

$$|G_k - G_k^{(i)}| \leq \frac{2}{\Im(z)|N_k|},$$

and Corollary 3.2 gives that

$$\mathbb{E}|G_k - \mathbb{E}G_k|^2 = O\left(\frac{n}{\Im(z)^2 |N_k|^2} \right) = O\left(\frac{L^2}{\Im(z)^2 n} \right).$$

It conclusion, using again (3.14), for any $1 \leq k \leq L$,

$$\mathbb{E}G_k + \left(z + \frac{1}{L} \sum_{\ell=1}^{L} \rho_{k\ell}^2 \mathbb{E}G_\ell \right)^{-1} \to 0, \tag{3.15}$$

and

$$\mathbb{E}\frac{1}{L} \sum_{k=1}^{L} G_k = \mathbb{E}\frac{1}{n} \sum_{i=1}^{n} R_{ii}.$$

Unicity of fixed point equation

Consider the function $\bar{G} : [0,1] \to \mathbb{C}_+$ given

$$\bar{G}(x) = \frac{1}{L} \sum_{k=1}^{L} \mathbf{1}_{\frac{k-1}{L} < x \leq \frac{k}{L}} \mathbb{E}G_k.$$

Consider an accumulation point of the vector $(\mathbb{E}G_1, \cdots, \mathbb{E}G_L)$, say $(g_1, \cdots, g_L)$. Then $\bar{G}$ converges in L^∞-norm to

$$g_\rho(x) = \frac{1}{L} \sum_{k=1}^{L} \mathbf{1}_{\frac{k-1}{L} < x \leq \frac{k}{L}} g_k.$$

By (3.15), g_ρ satisfies the fixed point equation, for all $x \in [0,1]$,

$$g = F_{z,\rho}(g),$$

with

$$F_{z,\rho}(g)(x) = -\left(z + \int \rho(x,y)^2 g(y) dy \right)^{-1}.$$

If $g, h \in L^1([0,1]; \mathbb{C}_+)$, we find,

$$|F_{z,\rho}(g)(x) - F_{z,\rho}(h)(x)| \leq \frac{\int \rho^2(x,y)|g(y) - h(y)|dy}{\Im(z)^2} \leq \frac{\|g - h\|_1}{\Im(z)^2},$$

where we have used again (3.14) and $\rho(x,y) \leq 1$. In particular

$$\|F_{z,\rho}(g) - F_{z,\rho}(h)\|_1 \leq \frac{\|g - h\|_1}{\Im(z)^2}.$$

Hence for $\Im(z) > 1$, $F_{z,\rho}$ is a contraction on the Banach space $L^1([0,1]; \mathbb{C}_+)$. Hence there is a unique solution of the fixed point

$$g = F_{z,\rho}(g).$$

The same argument works for the functions σ and its associated map $F_{z,\rho}$. Now similarly, we have, if $g \in L^\infty([0,1]; \mathbb{C}_+)$,

$$\|F_{z,\sigma}(g) - F_{z,\rho}(g)\|_1 \leq \frac{\int |\rho^2(x,y) - \sigma^2(x,y)||g(y)|dy}{\Im(z)^2} \leq \frac{\|g\|_\infty \|\sigma^2 - \rho^2\|_1}{\Im(z)^2},$$

In particular, if g_σ is the unique fixed point $g = F_{z,\sigma}(g)$, since $\|g_\rho\|_\infty \leq 1/\Im(z)$, we deduce

$$\begin{aligned}\|g_\sigma - g_\rho\|_1 = \|F_{z,\sigma}(g_\sigma) - F_{z,\rho}(g_\rho)\|_1 &\leq \|F_{z,\sigma}(g_\sigma) - F_{z,\sigma}(g_\rho)\|_1 + \|F_{z,\sigma}(g_\rho) - F_{z,\rho} g_\rho\|_1 \\ &\leq \frac{\|g_\sigma - g_\rho\|_1}{\Im(z)^2} + \frac{\|\sigma^2 - \rho^2\|_1}{\Im(z)^3},\end{aligned}$$

This gives

$$\|g_\sigma - g_\rho\|_1 \leq \frac{\|\sigma^2 - \rho^2\|_1}{\Im(z)^2(\Im(z) - 1)}, \tag{3.16}$$

As already pointed, $\|\rho^2 - \sigma^2\|_1 \to 0$ as $L \to \infty$.

End of proof

In summary, we have proved the following, fix $\varepsilon > 0$, $z \in \mathbb{C}_+$ with $\Im(z) > 1$. From Theorem 2.9, we have, a.s., for all n large enough,

$$|g_{\mu_Y}(z) - \mathbb{E} g_{\mu_Y}(z)| \leq \varepsilon.$$

By (3.16), we may fix L large enough so that

$$|g_\rho(z) - g_\sigma(z)| \leq \varepsilon.$$

Then, by(3.13), for all n large enough,

$$|\mathbb{E} g_{\mu_Z}(z) - \mathbb{E} g_{\mu_Y}(z)| \leq \varepsilon,$$

and by (3.15)

$$|\mathbb{E}g_{\mu_Z}(z) - g_\rho(z)| \leq \varepsilon.$$

This concludes the proof of Theorem 3.1 (since $\mathbb{E}\mu_Z$ and $\mathbb{E}\mu_Y$ are tight, g_ρ and g_σ are necessarily Cauchy-Stieltjes transforms of probability measures). $\square$

Lecture 4

Gaussian Wigner matrices

1 Matrix differentiation formulas and Gaussian integration by parts

1.1 Derivative of resolvent

We identify $H_n(\mathbb{C})$ with $\mathbb{R}^{n^2}$. Then, if $\Phi : H_n(\mathbb{C}) \mapsto \mathbb{C}$ is a continuously differentiable function, we define $\partial_{\Re(jk)}\Phi(X)$ as the derivative with respect to $\Re(X_{jk})$, and for $1 \le j \ne k \le n$, $\partial_{\Im(jk)}\Phi(X)$ as the derivative with respect to $\Im(X_{jk})$.

Define the resolvent $R_A = (A - z)^{-1}$, $z \in \mathbb{C}_+$. From the resolvent identity (3.6), a simple computation shows that for any integers $1 \le j, k \le n$, and $1 \le a \ne b \le n$,

$$\partial_{\Re(ab)} R_{jk} = -(R_{ja}R_{bk} + R_{jb}R_{ak}) \quad \text{and} \quad \partial_{\Im(ab)} R_{jk} = -i(R_{ja}R_{bk} - R_{jb}R_{ak}),$$

while if $1 \le a \le n$, then

$$\partial_{\Re(aa)} R_{jk} = -R_{ja}R_{ak}.$$

1.2 Gaussian differentiation formulas

We consider a Wigner matrix $X = (X_{ij})_{1 \le i,j \le n}$. We assume that

(A1) $(\Re(X_{12}), \Im(X_{12}))$ is a centered Gaussian vector in $\mathbb{R}^2$ with covariance $K \in H_2(\mathbb{R})$, $\mathrm{Tr}(K) = 1$.

(A2) X_{11} is a centered Gaussian in $\mathbb{R}$ with variance σ^2.

We recall the Gaussian integration by part formula which we have already use in (2.9).

Lemma 4.1. *Let X be a centered Gaussian vector in $\mathbb{R}^n$ with covariance matrix $K = \mathbb{E}XX^*$. For any continuously differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$, with $\mathbb{E}\|\nabla f(X)\|_2 < \infty$,*

$$\mathbb{E}f(X)X = K\mathbb{E}\nabla f(X). \tag{4.1}$$

The use of Gaussian integration by part in random matrix theory was initiated by Khorunzhy, Khoruzhenko and Pastur [KKP96]. Introduce the resolvent of $X/\sqrt{n}$ at $z \in \mathbb{C}_+$,

$$R = (X/\sqrt{n} - z)^{-1}.$$

(For ease of notation, we do not write explicitly the dependency in z). Using (4.1) we get, for $0 \le a \neq b \le n$, and all j, k:

$$\begin{aligned}
\mathbb{E} R_{jk} X_{ab} &= \frac{1}{\sqrt{n}} \mathbb{E}\left[K_{11} \partial_{\Re(ab)} R_{jk} + K_{12} \partial_{\Im(ab)} R_{jk} + i K_{21} \partial_{\Re(ab)} R_{jk} + i K_{22} \partial_{\Im(ab)} R_{jk}\right] \\
&= -\frac{1}{\sqrt{n}} \mathbb{E}\left[(K_{11} - K_{22} + iK_{12} + iK_{21}) R_{ja} R_{bk} + (K_{11} + K_{22} - iK_{12} + iK_{21}) R_{jb} R_{ak}\right] \\
&= -\frac{1}{\sqrt{n}} \mathbb{E}(\gamma R_{ja} R_{bk} + R_{jb} R_{ak}), \qquad (4.2)
\end{aligned}$$

where at the last line, we have used the symmetry of K and $\mathrm{Tr}(K) = 1$, together with the notation

$$\gamma = K_{11} - K_{22} + 2iK_{12} = \mathbb{E} X_{12}^2.$$

Notice that $|\gamma| \le 1$. Similarly, for $a = b$ one has

$$\mathbb{E} R_{jk} X_{aa} = -\frac{\sigma^2}{\sqrt{n}} \mathbb{E} R_{ja} R_{ak}. \qquad (4.3)$$

For further use, we also set

$$\kappa = \sigma^2 - 1 - \gamma.$$

In the GUE and GOE case $\kappa = 0$ while γ is equal respectively to 0 and 1.

2 Semicircle law for Gaussian random matrices

The resolvent identity gives

$$-zR = R \cdot \left(\frac{X}{\sqrt{n}} - zI - \frac{X}{\sqrt{n}}\right) = I - \frac{1}{\sqrt{n}} RX.$$

Hence, for $1 \le j, k \le n$, using (4.2)-(4.3),

$$\begin{aligned}
-z\mathbb{E} R_{jk} &= \mathbf{1}_{j=k} - \frac{1}{\sqrt{n}} \sum_{1 \le a \le n} \mathbb{E}[R_{ja} X_{ak}] \\
&= \mathbf{1}_{j=k} + \frac{1}{n} \sum_{1 \le a \le n} \mathbb{E}[R_{jk} R_{aa}] + \frac{\gamma}{n} \sum_{1 \le a \neq k \le n} \mathbb{E}[R_{ja} R_{ka}] + \frac{(\sigma^2 - 1)}{n} \mathbb{E}[R_{jk} R_{kk}].
\end{aligned}$$

We set

$$g = g_{\mu_{X/\sqrt{n}}}(z) = \frac{1}{n} \mathrm{Tr}(R), \quad \overline{g} = \mathbb{E} g, \quad \underline{g} = g - \mathbb{E} g,$$

and consider the diagonal matrix D with $D_{jk} = \mathbf{1}_{j=k} R_{jk}$. We find

$$-z\mathbb{E}R = I + \mathbb{E}[gR] + \frac{1}{n}\mathbb{E}[R(\kappa D + \gamma R^\top)].$$

Subtracting $\overline{g}R$ one has

$$-\overline{g}\mathbb{E}R - z\mathbb{E}R = I + \mathbb{E}\underline{g}R + \frac{1}{n}\mathbb{E}R(\kappa D + \gamma R^\top).$$

Finally, multiplying by $-\frac{1}{n}$ and taking the trace,

$$\overline{g}^2 + z\overline{g} + 1 = -\mathbb{E}\underline{g}g - \frac{1}{n^2}\mathbb{E}\mathrm{Tr}[R(\kappa D + \gamma R^\top)] = -\mathbb{E}\underline{g}^2 - \frac{1}{n^2}\mathbb{E}\mathrm{Tr}[R(\kappa D + \gamma R^\top)]. \tag{4.4}$$

As a function of the entries of X, g has Lipschitz constant $O(n^{-1}\Im(z)^{-2})$. This fact follows from Corollary 2.2 applied to $f(x) = 1/(x-z)$. Since the entries of X satisfy a Poincaré inequality, by Theorem 2.6

$$\mathbb{E}|\underline{g}|^2 = O(n^{-2}\Im(z)^{-4}). \tag{4.5}$$

Also, since $|\mathrm{Tr}(AB)| \leq n\|A\|\|B\|$, we find

$$\left|\mathrm{Tr}R(\kappa D + \gamma R^\top)\right| = O(n\Im(z)^{-2}).$$

We deduce

$$\mathbb{E}\underline{g}^2 = O(n^{-2}\Im(z)^{-4}) \quad \text{and} \quad \frac{1}{n^2}\mathbb{E}\mathrm{Tr}[R(\kappa D + \gamma R^\top)] = O(n^{-1}\Im(z)^{-2}).$$

We thus have proved that

$$\overline{g}^2 + z\overline{g} + 1 = O_z(n^{-1}). \tag{4.6}$$

Lemma 4.2. *Let $\delta \in \mathbb{C}$ and $z \in \mathbb{C}_+$. If $x \in \mathbb{C}_+$ satisfies $x^2 + zx + 1 = \delta$, then,*

$$|x - g_{sc}(z)| \leq \frac{|\delta|}{\Im(z)}.$$

Proof. Recall that $g_{sc}^2(z) + zg_{sc}(z) + 1 = 0$. It follows that

$$\left(g_{sc}(z) + \frac{z}{2}\right)^2 = -1 + \frac{z^2}{4} \quad \text{and} \quad \left(x + \frac{z}{2}\right)^2 = -1 + \frac{z^2}{4} + \delta.$$

Hence

$$\delta = \left(\left(x + \frac{z}{2}\right)^2 - \left(g_{sc}(z) + \frac{z}{2}\right)^2\right) = (x - g_{sc}(z))(x + g_{sc}(z) + z).$$

It yields,

$$|x - g_{sc}(z)| = \frac{|\delta|}{|x + g_{sc}(z) + z|}.$$

Since $x, g_{sc}(z) \in \mathbb{C}_+$, $|x + g_{sc}(z) + z| \geq \Im(x + g_{sc}(z) + z) \geq \Im(z)$. $\square$

From (4.6) and lemma 4.2, we deduce a new proof of the semicircle law for Gaussian Wigner matrices.

3 Convergence of edge eigenvalues for Gaussian matrices

We pursue the analysis of Gaussian Wigner matrices to the study of extremal eigenvalues of $X/\sqrt{n}$. Our aim is to prove

Theorem 4.1. *Let X be a Gaussian Wigner matrix satisfying assumptions (A). We have a.s.*

$$\lim_{n\to\infty} \lambda_1\left(\frac{X}{\sqrt{n}}\right) = -\lim_{n\to\infty} \lambda_n\left(\frac{X}{\sqrt{n}}\right) = 2.$$

For simplicity, we set $\lambda_k = \lambda_k(X/\sqrt{n})$. Note that X and $-X$ have the same law. Hence, by symmetry, we may restrict to λ_1. Note also that Wigner semicircle law implies that a.s.

$$\liminf_{n\to\infty} \lambda_1 \geq 2.$$

We first observe that the method of moments used in the proof of Theorem 1.2 implies that $X \in H_n(\mathbb{C})$ is an hermitian random matrix such that $(X_{ij})_{i\geq j}$ are independent centered variables with k-th moment bounded by $c > 0$, then there exists a constant K (depending on k) such that

$$\mathbb{E}\frac{1}{n}\mathrm{Tr}\left(\frac{X}{\sqrt{n}}\right)^k \leq cK.$$

Since, for *even* k, $\lambda_1^k \leq \mathrm{Tr}(X/\sqrt{n})^k$, we deduce that

$$\mathbb{E}\lambda_1 \leq (\mathbb{E}\lambda_1^k)^{1/k} \leq n^{1/k}(cK)^k.$$

This rough bound can be improved by a net argument. A *centered* complex variable Y is subgaussian with constant c if for any complex λ,

$$\mathbb{E}\exp\left(\Re(\bar{\lambda}Y)\right) \leq \exp\left(c^2|\lambda|^2\right).$$

(the Laplace transform of a centered subgaussian variable is dominated by the Laplace transform a Gaussian variable).

Lemma 4.3. *Let $X = (X_{ij})_{1\leq i,j\leq n} \in H_n(\mathbb{C})$ be an hermitian random matrix such $(X_{ij})_{i\geq j}$ are independent centered and subgaussian random variables with common constant $c > 0$.*

$$\mathbb{E}\|X\| \leq cK\sqrt{n},$$

where K is a universal constant.

The lemma is a consequence of the following classical statement. In a metric space X, for $\varepsilon > 0$, an *ε-net* is a subset Y of X such that for any $x \in X$ there exists a $y \in Y$ at distance at most ε.

Lemma 4.4 (Net of the sphere). *For any integer $n \geq 1$, there exists an ε-net of the sphere S^{n-1} of cardinality at most $(1+2/\varepsilon)^n$ (for the Euclidean distance in $\mathbb{R}^n$).*

Proof. There is a simple volumetric argument. Let N be a maximal ε-separated set (that is all pairs $x \neq y$ in N are at distance larger than ε and it is not possible to increase N without breaking this property). Then N is also an ε-net by the maximality assumption. We will prove that $|N| \leq (1+\varepsilon/2)^n$. Observe that the set $\cup_{x \in N} B(x, \varepsilon/2)$ where $B(x,r)$ is the open ball of radius r and center x is a disjoint union and it is contained in $B(0, 1+\varepsilon/2)$. Its volume is thus $N(\varepsilon/2)^n v_n$ where v_n is the volume of the unit ball in $\mathbb{R}^n$. Since the volume of $B(0,1+\varepsilon/2)$ is $(1+\varepsilon/2)^n v_n$, it follows that $|N|(\varepsilon/2)^n \leq (1+\varepsilon/2)^n$ as required. □

Proof of Lemma 4.3. We may assume $c = 1$. Let N be an $1/4$-net of S^{n-1} of cardinality at most 5^n (guaranteed by Lemma 4.4). Let $\|X\|_N = \max_{u \in N} |\langle u, Xu \rangle|$. We claim that

$$\|X\| \leq 4\|X\|_N. \tag{4.7}$$

Indeed, Let $v \in S^{n-1}$ be such that $|\langle v, Xv \rangle| = \|X\|$ (recall that $\|X\| = \max(\lambda_1(X), -\lambda_n(X)) = \max_{v \in S^{n-1}} |\langle v, Xv \rangle|$). There exists $u \in N$, such that $\|u - v\|_2 \leq 1/4$. In particular, from the triangle inequality

$$\begin{aligned} \|X\| = |\langle v, Xv \rangle| &= |\langle u + v - u, X(u+v-u) \rangle| \\ &\leq |\langle u, Xu \rangle| + 2|\langle u, X(v-u) \rangle| + |\langle (v-u), X(v-u) \rangle| \\ &\leq \|X\|_N + \frac{1}{2}\|X\| + \frac{1}{16}\|X\|. \end{aligned}$$

It proves (4.7). It follows from the union bound that for any $t > 0$,

$$\mathbb{P}(\|X\| \geq 4t) \leq \mathbb{P}(\|X\|_N \geq t) \leq |N| \max_{u \in N} \mathbb{P}(|\langle u, Xu \rangle| \geq t).$$

Now, for $u \in S^{n-1}$ and real λ, we claim that

$$\mathbb{E} e^{\lambda \langle u, Xu \rangle} \leq e^{2\lambda^2}.$$

Indeed, since $\langle u, Xu \rangle = \sum_i |u_i|^2 X_{ii} + \sum_{i>j} 2\Re(u_i X_{ij} u_j)$, by independence and the subgaussian assumption,

$$\mathbb{E} e^{\lambda \langle u, Xu \rangle} = \prod_i \mathbb{E} e^{\lambda |u_i|^2 X_{ii}} \prod_{i>j} \mathbb{E} e^{2\lambda \Re(u_i X_{ij} u_j)} \leq \mathbb{E} e^{\lambda^2 \sum_i |u_i|^2 + 4\lambda^2 \sum_{i>j} |u_i|^2 |u_j|^2}$$

as claimed. It follows from Chernov' bound, with $\lambda = t\sqrt{n}/4$,

$$\mathbb{P}(|\langle u, Xu \rangle| \geq t\sqrt{n}) \leq 2e^{-\lambda t \sqrt{n}} e^{2\lambda^2} \leq 2e^{-\frac{nt^2}{8}}.$$

We write $\mathbb{E}\|X\|/\sqrt{n} \leq a + \int_a^\infty \mathbb{P}(\|X\| \geq t\sqrt{n}) dt$ and use $|N| \leq 5^n$, the conclusion follows easily by choosing a large enough. □

We now prove Theorem 4.1. A proof of this result can be proved by using the moment method, see Fűredi and Komlós [FK81]. Here, we will instead use the resolvent differentiation formulas. In the GUE case, this approach was initiated by Haagerup and Thorbjørnsen [HT05].

By Lemma 2.11 and Borel-Cantelli Lemma, a.s.

$$\lim_{n\to\infty} |\lambda_1 - \mathbb{E}\lambda_1| = 0.$$

Hence, in view of Lemma 4.3, it is thus sufficient to prove that, for any $\varepsilon > 0$, a.s. for all $n \gg 1$,

$$\mathbf{1}(\lambda_1 \in [2+2\varepsilon, K]) = 0.$$

where K is a large constant. To this end, we fix $\varepsilon > 0$ and set

$$\Delta = [2+2\varepsilon, K]. \tag{4.8}$$

Consider a smooth function $\varphi : \mathbb{R} \mapsto [0,1]$ with support $[2+\varepsilon, 2K]$ such that $\varphi(x) = 1$ on $[2+2\varepsilon, K]$. By Lemma 2.11 and Borel-Cantelli Lemma, a.s.

$$\lim_{n\to\infty} |\varphi(\lambda_1) - \mathbb{E}\varphi(\lambda_1)| = 0.$$

Assume that we manage to prove that

$$\mathbb{E}\frac{1}{n}\sum_{k=1}^{n}\varphi(\lambda_k) = \mathbb{E}\int \varphi d\mu_{X/\sqrt{n}} \leq \frac{1}{2n}. \tag{4.9}$$

Then, using $\mathbf{1}(\lambda \in \Delta) \leq \varphi(\lambda)$, we would deduce, that

$$\mathbb{P}(\lambda_1 \in \Delta) \leq \mathbb{E}\varphi(\lambda_1) \leq n\mathbb{E}\int \varphi d\mu_{X/\sqrt{n}} \leq \frac{1}{2}. \tag{4.10}$$

And it would yields to a.s. for $n \gg 1$,

$$\mathbf{1}(\lambda_1 \in \Delta) \leq 1/3.$$

Hence, the indicator function is equal to 0 and $\lambda_1 \in \Delta$. It follows that if (4.9) holds for any $\varepsilon > 0$, our Theorem 4.1 is proved.

The first step of proof is to relate $\int \varphi d\mu$ to an integral over $g_\mu(z)$. For C^1 functions $f : \mathbb{C} \mapsto \mathbb{C}$, we set $\bar{\partial} f(z) = \partial_x f(z) + i\partial_y f(z)$, where $z = x+iy$. In the next lemma $\chi : \mathbb{R} \to \mathbb{R}$ is a compactly supported smooth function such that $\chi(y) = 1$ in a neighbourhood of 0.

Lemma 4.5 (Helfer and Sjöstrand)**.** *Let $k \geq 1$ be an integer and $\varphi : \mathbb{R} \mapsto \mathbb{R}$ a compactly supported C^{k+1}-function, then for any $\mu \in \mathcal{P}(\mathbb{R})$,*

$$\int \varphi d\mu = \frac{1}{\pi}\Re\int_{\mathbb{C}_+} \bar{\partial}\Phi(x+iy) g_\mu(x+iy) dx dy,$$

where $\Phi(x+iy) = \sum_{\ell=0}^{k} \frac{(iy)^\ell}{\ell!}\varphi^{(\ell)}(x)\chi(y)$.

Proof. Observe that in a neighbourhood of $\mathbb{R}$,

$$\bar{\partial}\Phi(x+iy) = \varphi^{(k+1)}(x)(iy)^k/k!. \tag{4.11}$$

It follows that for any $\lambda \in \mathbb{R}$, $\bar{\partial}\Phi(z)/(z-\lambda)$ is integrable. Now, from Fubini's Theorem, it suffices to check this for $\mu = \delta_0$:

$$\varphi(0) = \Phi(0) = -\frac{1}{\pi}\Re\int_{\mathbb{C}_+} \frac{\bar{\partial}\Phi(x+iy)}{x+iy}dxdy,$$

If $z = x+iy$, we have

$$\Re\left(\frac{\bar{\partial}\Phi(z)}{z}\right) = \frac{x\partial_x\Phi(z)}{x^2+y^2} + \frac{y\partial_y\Phi(z)}{x^2+y^2} = \frac{1}{r}\partial_r\Phi(z),$$

where $r = |z|$ and ∂_r is the radial derivative in polar coordinates. We thus find

$$\Re\int_{\mathbb{C}_+} \frac{\bar{\partial}\Phi(x+iy)}{x+iy}dxdy = \int_0^\infty\int_0^\pi \partial_r\Phi(re^{i\theta})drd\theta = -\pi\Phi(0).$$

as requested. □

Proof of Theorem 4.1: GUE case. we can now prove Theorem 4.1 for GUE matrices. Indeed, in this case, $\gamma = 0$ and $\kappa = 0$. In particular, from Equations (4.4) and (4.5), we find

$$\overline{g}^2 + z\overline{g} + 1 = O(n^{-2}\Im(z)^{-4}).$$

From Lemma 4.2 we find

$$\overline{g} - g_{sc} = \delta(z) = O(n^{-2}\Im(z)^{-5}). \tag{4.12}$$

We may apply Lemma 4.5 with $k = 6$ to our smooth function φ with support $[2+\varepsilon, 2K]$. We find

$$\begin{aligned}
\mathbb{E}\int \varphi d\mu_{X/\sqrt{n}} &= \frac{1}{\pi}\Re\mathbb{E}\int_{2+\varepsilon}^\infty\int_0^\infty \bar{\partial}\Phi(x+iy)\overline{g}(x+iy)dydx \\
&= \frac{1}{\pi}\Re\mathbb{E}\int_{2+\varepsilon}^\infty\int_0^\infty \bar{\partial}\Phi(x+iy)(g_{sc}(x+iy)+\delta(x,y))dydx \\
&= \int \varphi d\mu_{sc} + \frac{1}{\pi}\Re\mathbb{E}\int_{2+\varepsilon}^{2K}\int_0^\infty \bar{\partial}\Phi(x+iy)\delta(x,y)dydx
\end{aligned}$$

Now, the support of μ_{sc} is $[-2,2]$. Hence $\int\varphi d\mu_{sc} = 0$. Also, from (4.11)

$$\bar{\partial}\Phi(x+iy)\delta(x,y) = O(n^{-2}).$$

and is compactly supported from (4.8) and the definition of φ. We thus have proved that

$$\mathbb{E}\int\varphi d\mu_{X/\sqrt{n}} = O(n^{-2}).$$

It concludes the proof of (4.9) in the GUE case.

Proof of Theorem 4.1: general case. The above argument cannot work directly if $\gamma \neq 0$ or $\kappa \neq 0$. Indeed, Equation (4.4) gives only

$$\overline{g}^2 + z\overline{g} + 1 = O(n^{-1}\Im(z)^{-2} + n^{-2}\Im(z)^{-4}).$$

and from Lemma 4.2,

$$\overline{g} = g_{sc} + O(n^{-1}(\Im(z)^{-5} \wedge 1)). \tag{4.13}$$

We thus have to study more precisely

$$\frac{1}{n^2}\mathrm{Tr}(R(\kappa D + \gamma R^\top)).$$

We first need a lemma

Lemma 4.6. *For any $\varepsilon > 0$, we have*

$$\inf\{|g_{sc}(z)| : z \in \mathbb{C},\, d(z, [-2, 2]) \geq \varepsilon\} < 1,$$

where $d(z, A) = \inf\{|w - z| : w \in A\}$.

Proof. Let $r(z) = |g_{sc}(z)|$ and $t = \varepsilon/\sqrt{2}$. If $d(z, [-2, 2]) \geq \varepsilon$ then either $|\Im(z)| \geq t$ or $|\Re(z)| \geq 2 + t$. We first assume that $\Im(z) \geq t$. Note that if $z = E + i\eta$ and ξ has distribution μ_{sc}, then $r(z) = \mathbb{E}|(\xi - E) - i\eta|^{-1}$. By symmetry and monotony, $r(z) \geq r(i\eta) \geq r(it)$. We find

$$r(it) = \frac{1}{2\pi}\int_{-2}^{2}\frac{t\sqrt{4 - x^2}}{t^2 + x^2}dx = \frac{1}{2\pi}\int_{-2/t}^{2/t}\frac{\sqrt{4 - (tx)^2}}{1 + x^2}dx < \frac{1}{2\pi}\int_{\infty}^{\infty}\frac{2}{1 + x^2}dx = 1.$$

It remains to deal with $z = E + i\eta$ and $|E| \geq 2 + t$. We have $r(z) \geq r(E) = r(|E|) \geq r(2 + t)$ and

$$r(2 + t) = \frac{1}{2\pi}\int_{-2}^{2}\frac{\sqrt{4 - x^2}}{2 + t - x}dx < \frac{1}{2\pi}\int_{-2}^{2}\frac{\sqrt{4 - x^2}}{2 - x}dx = 1.$$

□

Lemma 4.7. *Let $f(z) = \gamma g_{sc}^2(z)/(\gamma g_{sc}^2(z) + 1) + \kappa g_{sc}^2(z)$, we have*

$$\mathbb{E}\frac{1}{n}\mathrm{Tr}(R(\kappa D + \gamma R^\top)) = f(z) + O(n^{-1}(1 - |g_{sc}(z)|^2)^{-1}(1 \wedge \Im(z)^{-5})).$$

Proof. We may again use the Gaussian integration by part formula. Using (4.1)-(4.2)-(4.3), we get, for $0 \leq a \neq b \leq n$, and all j, k, ℓ, m:

$$\mathbb{E}R_{jk}R_{\ell m}X_{ab} = -\frac{1}{\sqrt{n}}(\mathbb{E}(\gamma R_{ja}R_{bk} + R_{jb}R_{ak})R_{\ell m} + \mathbb{E}R_{jk}(\gamma R_{\ell a}R_{bm} + R_{\ell b}R_{am})), \tag{4.14}$$

and

$$\mathbb{E}R_{jk}R_{\ell m}X_{aa} = -\frac{\sigma^2}{\sqrt{n}}(\mathbb{E}R_{ja}R_{ak}R_{\ell m} + \mathbb{E}R_{jk}R_{\ell a}R_{am}), \tag{4.15}$$

We use again the identity $-zR = I - \frac{1}{\sqrt{n}}RX$. Taking conjugate and composing by R yields to $-zRR^\top = R - \frac{1}{\sqrt{n}}RX^\top R^\top$, we find

$$-z(RR^\top)_{kk} = R_{kk} - \frac{1}{\sqrt{n}}\sum_{a,b} R_{kb}R_{ka}X_{ab}.$$

We now take expectation and use (4.14)-(4.15),

$$\begin{aligned}\mathbb{E}(RR^T)_{kk} =& \mathbb{E}R_{kk} + \frac{\gamma}{n}\sum_{a\neq b}\mathbb{E}R_{ka}^2R_{bb} + \frac{1}{n}\sum_{a\neq b}\mathbb{E}R_{kb}R_{ab}R_{ka} + \frac{\gamma}{n}\sum_{a\neq b}\mathbb{E}R_{kb}R_{ka}R_{ba}\\ &+ \frac{1}{n}\sum_{a\neq b}\mathbb{E}R_{kb}^2R_{aa} + \frac{2\sigma^2}{n}\sum_{a}\mathbb{E}R_{ka}^2R_{aa}\end{aligned}$$

We set

$$m = \frac{1}{n}\mathrm{Tr}(RR^\top), \quad \overline{m} = \mathbb{E}m \quad \text{and} \quad \underline{m} = m - \mathbb{E}m.$$

Recall that $D_{kk} = R_{kk}$. Taking Tr and dividing by n, in the above expression we obtain

$$-z\overline{m} = \overline{g} + (\gamma+1)\mathbb{E}gm + \frac{1}{n^2}\mathrm{Tr}(R(R^T)^2 + \gamma R^2R^T + 2\kappa RR^TD).$$

We deduce that

$$-(z + (\gamma+1)\overline{g})\overline{m} = \overline{g} + \mathbb{E}\underline{g}\underline{m} + \frac{1}{n^2}\mathrm{Tr}(RR^T((1+\gamma)R^T + 2\kappa D)). \tag{4.16}$$

Using (4.5), $|m| \leq \Im(z)^{-2}$ and $|\mathrm{Tr}(A)| \leq n\|A\|$, we find

$$-(z + (\gamma+1)\overline{g})\overline{m} = \overline{g} + O(n^{-1}(\Im(z)^{-4}\wedge 1)).$$

We deduce from (4.13)

$$-(z + (\gamma+1)g_{sc})\overline{m} = g_{sc} + O(n^{-1}(\Im(z)^{-5}\wedge 1)).$$

We multiply by $g_{sc} = O(1)$ and use that $g_{sc}^2 + zg_{sc} + 1 = 0$. From Lemma 4.6 and $|\gamma| \leq 1$, $|\gamma g_{sc}^2 + 1| \geq 1 - |g_{sc}|^2 > 0$. We find

$$\overline{m} = \frac{g_{sc}^2}{\gamma g_{sc}^2 + 1} + O(n^{-1}(1 - |g_{sc}(z)|^2)^{-1}(\Im(z)^{-5}\wedge 1)).$$

we have $(z + (\gamma+1)g_{sc})$.

We set similarly

$$m' = \frac{1}{n}\sum_{k=1}^{n}(R_{kk})^2, \quad \overline{m}' = \mathbb{E}m' \quad \text{and} \quad \underline{m}' = m' - \mathbb{E}m',$$

so that $\frac{1}{n}\mathrm{Tr}(RD) = m'$. From $-zR = I - \frac{1}{\sqrt{n}}RX$, multiplying by R_{kk}, we obtain

$$-z(R_{kk})^2 = R_{kk} - \frac{1}{\sqrt{n}}\sum_a R_{ka}R_{kk}X_{ak}.$$

We find analogously

$$-(z+\overline{g})\overline{m}' = \overline{g} + O(n^{-1}\Im(z)^{-3}). \tag{4.17}$$

and, from (4.13), $g_{sc}^2 + zg_{sc} + 1 = 0$ and $|g_{sc}| = O(1)$.

$$\overline{m}' = g_{sc}^2 + O(n^{-1}(\Im(z)^{-5} \wedge 1)).$$

This concludes the proof. $\square$

We may now conclude the proof of Theorem (4.9). Let $S = \{z \in \mathbb{C} : \Re(z) \geq 2+\varepsilon\}$ and $S_+ = S \cap \mathbb{C}_+$. We also set

$$L = \frac{1}{zn^2}\mathbb{E}\mathrm{Tr}[R(\kappa D + \gamma R^\top)].$$

On S_+, we have $L = O(n^{-1}\Im(z)^{-2})$. From Equations (4.4) and (4.5), we find

$$(\overline{g}+L)^2 + z(\overline{g}+L) + 1 = O(n^{-2}\Im(z)^{-4}).$$

For n large enough, $\overline{g}+L \in \mathbb{C}_+$. Hence, from Lemma 4.2, we find

$$\overline{g} + L - g_{sc} = \delta(z) = O(n^{-2}\Im(z)^{-5}).$$

So finally, from (4.13) and Lemma 4.7, for all $z \in K_+$,

$$\overline{g} = g_{sc} - \frac{f(z)}{n} + O(n^{-2}(\Im(z)^{-5} \wedge 1)), \tag{4.18}$$

where the $O(\cdot)$ depends on ε.

As above, for our smooth function φ with support $[2+\varepsilon, 2K)$ we may apply Lemma 4.5 with $k = 5$. We find

$$\begin{aligned}
\mathbb{E}\int \varphi d\mu_{X/\sqrt{n}} &= \frac{1}{\pi}\Re\int_{S_+}\bar{\partial}\Phi(x+iy)\overline{g}(x+iy)dxdy \\
&= \frac{1}{\pi}\Re\int_{S_+}\bar{\partial}\Phi(x+iy)(g_{sc}(x+iy) - \frac{f(x+iy)}{n} + \delta(x,y))dxdy \\
&= \frac{1}{\pi}\Re\int_{S_+}\bar{\partial}\Phi(x+iy)(g_{sc}(x+iy) - \frac{f(x+iy)}{n})dxdy \\
&\quad \frac{1}{\pi}\Re\int_{S_+}\bar{\partial}\Phi(x+iy)\delta(x,y)dxdy
\end{aligned}$$

Now, notice that $g_{sc}(z) - f(z)/n$ is analytic on an open neighbourhood of S_+. In particular, $\bar{\partial}(g_{sc} - f/n) = 0$ on this neighbourhood. Hence, by integration by part, the first integral of the above expression is 0. Also, from (4.11)

$$\bar{\partial}\Phi(x+iy)\delta(x,y) = O(n^{-2}).$$

and is compactly supported. We thus have proved that

$$\mathbb{E}\int \varphi d\mu_{X/\sqrt{n}} = O(n^{-2}).$$

It concludes the proof of Theorem 4.1. $\square$

4 Variance of linear statistics

Our goal is to prove the following theorem.

Theorem 4.2. *Let be real-valued function $f \in C^1$ and X be a GUE matrix. We have*

$$\lim_{n\to\infty} \operatorname{Var}\left(\sum_{k=1}^{n} f\left(\lambda_k\left(\frac{X}{\sqrt{n}}\right)\right)\right) = \frac{1}{4\pi^2}\int_{[-2,2]^2}\left(\frac{f(\lambda_1)-f(\lambda_2)}{\lambda_1-\lambda_2}\right)^2 \frac{4-\lambda_1\lambda_2}{\sqrt{4-\lambda_1^2}\sqrt{4-\lambda_2^2}} d\lambda_1 d\lambda_2.$$

such that

We essentially follow Pastur and Scherbina [PS11, Chapter 3]. Using Lemma 4.5, the proof is a consequence of the following lemma (we use the notation of the previous section)

Lemma 4.8. *For GUE matrices, for any $z_1, z_2 \in \mathbb{C}$,*

$$n^2\operatorname{Cov}(g(z_1), g(z_2)) = n^2\mathbb{E}\underline{g}(z_1)\underline{g}(z_2) = \frac{1}{2(z_1-z_2)^2}\left(\frac{z_1z_2-1}{\sqrt{z_1^2-4}\sqrt{z_2^2-4}} - 1\right) + O\left(\frac{1}{ny^8}\right),$$

where $y = \min(|\Im(z_1)|, |\Im(z_2)|)$.

Proof. Since $g(\bar{z}) = \bar{g}(z)$, we may assume without loss of generality that $\Im(z_1), \Im(z_2)$ are positive. We have

$$\operatorname{Cov}(g(z_1), g(z_2)) = \mathbb{E}\underline{g}(z_1)\underline{g}(z_2) = \mathbb{E}g(z_1)\underline{g}(z_2).$$

Using as always the resolvent identity, we write

$$-z_1 g(z_1) = 1 - \frac{1}{n^{3/2}}\sum_{j,k} R_{jk}(z_1)X_{kj}.$$

Hence,

$$-z_1\mathbb{E}\underline{g}(z_1)\underline{g}(z_2) = -z_1\mathbb{E}g(z_1)\underline{g}(z_2) = -\frac{1}{n^{5/2}}\sum_{j,k,\ell}\mathbb{E}R_{jk}(z_1)\underline{R_{\ell\ell}}(z_2)X_{kj}.$$

Using (4)From (4.14)-(4.15), we get

$$\begin{aligned}
-z_1\mathbb{E}\underline{g}(z_1)\underline{g}(z_2) &= \mathbb{E}\frac{1}{n^3}\sum_{j,k,\ell}\gamma R_{jk}(z_1)R_{kj}(z_1)\underline{R}_{\ell\ell}(z_2)+R_{jj}(z_1)R_{kk}(z_1)\underline{R}_{\ell\ell}(z_2)\\
&\quad+\mathbb{E}\frac{1}{n^3}\sum_{j,k,\ell}\gamma R_{jk}(z_1)R_{\ell k}(z_2)R_{j\ell}(z_2)+R_{jk}(z_1)R_{\ell j}(z_2)R_{k\ell}(z_2)\\
&\quad+\mathbb{E}\frac{\kappa}{n^3}\sum_{j,\ell}(R_{jj}(z_1))^2\underline{R}_{\ell\ell}(z_2)+R_{jj}(z_1)R_{\ell j}(z_2)R_{j\ell}(z_2)\\
&= \frac{\gamma}{n^2}\mathbb{E}\mathrm{Tr}R^2(z_1)\underline{g}(z_2)+\mathbb{E}g(z_1)^2\underline{g}(z_2)+\frac{\gamma}{n^3}\mathbb{E}\mathrm{Tr}R(z_1)^\top R^2(z_2)\\
&\quad+\frac{1}{n^3}\mathbb{E}\mathrm{Tr}R(z_1)R^2(z_2)+\frac{\kappa}{n^2}\mathbb{E}\mathrm{Tr}D^2(z_1)\underline{g}(z_2)+\frac{\kappa}{n^3}\mathbb{E}\mathrm{Tr}D(z_1)R^2(z_2).
\end{aligned}$$

We write that

$$\mathbb{E}g(z_1)^2\underline{g}(z_2)=2\bar{g}(z_1)\mathbb{E}\underline{g}(z_1)\underline{g}(z_2)+\mathbb{E}\underline{g}(z_1)^2\underline{g}(z_2).$$

Besides, from Theorem 2.10 and Cauchy-Schwartz inequality,

$$\mathbb{E}\underline{g}(z_1)^2\underline{g}(z_2)\leq\sqrt{\mathbb{E}|\underline{g}(z_1)|^4\mathbb{E}|\underline{g}(z_2)|^2}=O\left(\frac{1}{n^3y^6}\right)$$

In the GUE case, $\kappa=\gamma=0$, we deduce that

$$n^2\mathrm{Cov}(g(z_1),g(z_2))=-\frac{1}{z_1+2\bar{g}(z_1)}\left(\frac{1}{n}\mathbb{E}\mathrm{Tr}R(z_1)R^2(z_2)+O\left(\frac{1}{ny^6}\right)\right).$$

We may linearize the expression $R(z_1)R^(z_2)$ as follows. We use that,

$$R(z_1)R(z_2)=\frac{R(z_1)-R(z_2)}{z_1-z_2}$$

and

$$R^2=\partial_z R.$$

We find

$$n^2\mathrm{Cov}(g(z_1),g(z_2))=-\frac{1}{z_1+2\bar{g}(z_1)}\left(\partial_{z_2}\frac{\bar{g}(z_1)-\bar{g}(z_2)}{z_1-z_2}+O\left(\frac{1}{ny^6}\right)\right).$$

Now, in order to use (4.12), we will benefit from the analyticity of g. We write

$$\partial_{z_2}\frac{\bar{g}(z_1)-\bar{g}(z_2)}{z_1-z_2}=\int_0^1\bar{g}''(z_2+t(z_1-z_2))tdt$$

and, if $z_0=z_2+t(z_1-z_2)$ from the residue theorem

$$\bar{g}''(z_0)=\frac{1}{i\pi}\oint\frac{\bar{g}(z)}{(z-z_0)^3}dz,$$

where we take a contour on a disc around z_0 with radius $y/2$ (which stays at distance at least $y/2$ from the real axis since $\Im(z_1), \Im(z_2) \geq y$). We get from (4.12)

$$\partial_{z_2} \frac{\bar{g}(z_1) - \bar{g}(z_2)}{z_1 - z_2} = \partial_{z_2} \frac{\bar{g}_{sc}(z_1) - \bar{g}_{sc}(z_2)}{z_1 - z_2} + O\left(\frac{1}{n^2 y^8}\right).$$

Similarly, since $\Im(z_1 + 2\bar{g}(z_1)) \geq y$,

$$\frac{1}{z_1 + 2\bar{g}(z_1)} = \frac{1}{z_1 + 2g_{sc}(z_1)} + O\left(\frac{1}{n^2 y^6}\right).$$

It remains to use the explicit formula of the Cauchy-Stieltjes transform of the semicircle law. $\square$

Proof of Theorem 4.2. By a Theorem 2.6, we already know that the variance of $\sum_{k=1}^n f(\lambda_k)$ is of order $\|f\|_L^2$. By concentration and Theorem 4.1, with exponentially large probability, all eigenvalues are in a neighborhood of $[-2, 2]$. By a density argument, we may thus assume that f is analytic In this case, the residue theorem implies that

$$f(x) = \frac{1}{2\pi i} \oint \frac{f(z)}{x - z} dz$$

where the direct contour is around x. It follows that if all eigenvalues are in $[-2 - \varepsilon, 2 + \varepsilon]$ that

$$\sum_{k=1}^n f(\lambda_k) = \frac{n}{2\pi i} \oint f(z) g(z) dz,$$

where the contour is around $[-2 - \varepsilon, 2 + \varepsilon]$. We get

$$\mathrm{Var}\left(\sum_{k=1}^n f\left(\lambda_k\left(\frac{X}{\sqrt{n}}\right)\right)\right) = -\frac{n^2}{4\pi^2} \oint \oint f(z_1) f(z_2) \mathrm{Cov}(g(z_1), g_(z_2)) dz_1 dz_2,$$

It remains to use Lemma 4.8 and let the contour go to $[-2, 2]$ ... $\square$

5 Beyond Gaussian Wigner matrices

It is possible to extend the Gaussian methods developped in the previous section to more general distributions. By a truncation step, it is often possible to reduce ourselves to bounded variables. Then Talagrand's concentration inequality for bounded variables is available in place of Gaussian concentration inequalities. The Gaussian integration by part is no longer exactly available. This issue can be addressed with the following lemma due to Khorunzhy, Khoruzhenko and Pastur [KKP96].

Lemma 4.9. *Let ξ be a real-valued random variable such that $\mathbb{E}|\xi|^{k+2} < \infty$. Let $f : \mathbb{R} \to \mathbb{C}$ be a C^{k+1} function such that the $(k+1)$-th derivative is uniformly bounded. Then,*

$$\mathbb{E}\xi f(\xi) = \sum_{\ell=0}^{k} \frac{\kappa_{\ell+1}}{\ell!} \mathbb{E} f^{(\ell)}(\xi) + O(\|f^{(k+1)}\|_\infty \mathbb{E}|\xi|^{k+2}),$$

where κ_ℓ is the ℓ-th cumulant of ξ and $O(\cdot)$ depends only on k.

Hence, at the cost of taking more derivatives, we may also use differential calculus to compute the expectation of expressions with resolvent entries. For example, to prove the analog of Theorem 4.1 for a random real symmetric matrix with bounded entries above the diagonal, we need to take $k = 3$ in the above lemma. We will however not pursue this method any further here.

Bibliography

[ABC$^+$00] Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[BS10] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[EY17] László Erdős and Horng-Tzer Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.

[FK81] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[For10] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[GZ03] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

[HT05] Uffe Haagerup and Steen Thorbjørnsen. A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)*, 162(2):711–775, 2005.

[KKP96] Alexei M. Khorunzhy, Boris A. Khoruzhenko, and Leonid A. Pastur. Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.*, 37(10):5033–5060, 1996.

[Led01] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[PS11] Leonid Pastur and Mariya Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

[Roy63] H. L. Royden. *Real analysis*. The Macmillan Co., New York; Collier-Macmillan Ltd., London, 1963.

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.