Random Matrix theory and statistical learning

Alice Guionnet

August 4, 2020

Abstract These lecture notes (in progress) are intended for a course to be held in Les Houches in August 2020.

Contents

1 Limiting spectral measure

1.1 Method of moments for Wigner's theorem

We consider in this section an $N \times N$ matrix $\mathbb{X}^N$ with real or complex entries such that $\left(\mathbb{X}^N_{ij}, 1 \le i \le j \le N\right)$ are independent and $\mathbb{X}^N$ is self-adjoint; $\mathbb{X}^N_{ij} = \bar{\mathbb{X}}^N_{ji}$. We assume further that

$$\mathbb{E}[\mathbb{X}^N_{ij}] = 0, \lim_{N\to\infty} \max_{1\le i,j\le N} |N\mathbb{E}[|\mathbb{X}^N_{ij}|^2] - 1| = 0. \tag{1}$$

We shall show that, under some finite moments conditions on the entries, the eigenvalues $(\lambda_1, \cdots, \lambda_N)$ of $\mathbb{X}^N$ satisfy the almost sure convergence

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) = \int f(x) d\sigma(x) \tag{2}$$

where f is a bounded continuous function or a polynomial function. σ is the semi-circular law

$$\sigma(dx) = \frac{1}{2\pi} \sqrt{4 - x^2} 1_{|x|\le 2} dx. \tag{3}$$

We shall prove this convergence for polynomial functions and rely on the fact that for all $k \in \mathbb{N}$, $\int x^k d\sigma(x)$ is null when k is odd and given by the Catalan number

$$C_{k/2} = \frac{\begin{pmatrix} k \\ \frac{k}{2} \end{pmatrix}}{\frac{k}{2} + 1} \tag{4}$$

when k is even.

1.1.1 Wigner's theorem

In this section, we use the same notation for complex and for real entries since both cases will be treated at once and yield the same result. The aim of this section is to prove

Theorem 1.1. *[Wigner's theorem [24]] Assume that for all $k \in \mathbb{N}$,*

$$B_k := \sup_{N\in\mathbb{N}} \sup_{(i,j)\in\{1,\cdots,N\}^2} \mathbb{E}[|\sqrt{N}\mathbf{X}^N_{ij}|^k] < \infty. \tag{5}$$

Then,

$$\lim_{N\to\infty} \frac{1}{N}\; Tr\left((\mathbf{X}^N)^k\right) = \begin{cases} 0 & \text{if } k \text{ is odd}, \\ C_{\frac{k}{2}} & \text{otherwise}, \end{cases} \tag{6}$$

where the convergence holds in expectation and almost surely.

The Catalan number C_k will appear here as the number of non-crossing pair partitions of $2k$ elements. Namely, recall that a partition of the (ordered) set $S := \{1, \cdots, n\}$ is a decomposition

$$\pi = \{V_1, \cdots, V_r\} \tag{7}$$

such that $V_i \cap V_j = \emptyset$ if $i \neq j$ and $\cup V_i = S$. The $V_i, 1 \leq i \leq r$ are called the blocks of the partition and we say that $p \sim_\pi q$ if p, q belong to the same block of the partition π. A partition π of $\{1, \cdots, n\}$ is said to be *crossing* if there exist $1 \leq p_1 < q_1 < p_2 < q_2 \leq n$ with

$$p_1 \sim_\pi p_2 \not\sim_\pi q_1 \sim_\pi q_2. \tag{8}$$

It is *non-crossing* otherwise. We leave it as an exercise to the reader to prove that C_k as given in the theorem is exactly the number of non-crossing pair partitions of $\{1, 2, \cdots, 2k\}$.

Proof. We start the proof by showing the convergence in expectation, for which the strategy is simply to expand the trace over the matrix in terms of its entries. We then use some (easy) combinatorics on trees to find out the main contributing term in this expansion. The almost sure convergence is obtained by estimating the covariance of the considered random variables and apply Borel-Cantelli lemma..

- *Expanding the expectation.*

  Setting $\mathbf{Y}^N = \sqrt{N}\mathbf{X}^N$, we have

$$\mathbb{E}\left[\frac{1}{N} \operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right] = \sum_{i_1, \cdots, i_k = 1}^{N} N^{-\frac{k}{2}-1} \mathbb{E}[Y_{i_1 i_2} Y_{i_2 i_3} \cdots Y_{i_k i_1}] \tag{9}$$

  where $Y_{ij}, 1 \leq i, j \leq N$, denote the entries of $\mathbf{Y}^N$ (which may eventually depend on N). We denote $\mathbf{i} = (i_1, \cdots, i_k)$ and set

$$P^N(\mathbf{i}) := \mathbb{E}[Y_{i_1 i_2} Y_{i_2 i_3} \cdots Y_{i_k i_1}]. \tag{10}$$

  Note that P^N depends on N through $\mathbf{Y}^N$. By (5) and Hölder's inequality, $P(\mathbf{i})$ is bounded uniformly by B_k, independently of $\mathbf{i}$ and N. Since the random variables $(Y_{ij}, i \leq j)$ are independent and centered, $P(\mathbf{i})$ equals zero unless for any pair (i_p, i_{p+1}), $p \in \{1, \cdots, k\}$, there exists $l \neq p$ such that $(i_p, i_{p+1}) = (i_l, i_{l+1})$ or (i_{l+1}, i_l). Here, we used the convention $i_{k+1} = i_1$. To find more precisely which set of indices contribute to the first order in the right hand side of (9), we next provide some combinatorial insight into the sum over the indices.

- *Connected graphs and trees.*

  $V(\mathbf{i}) = \{i_1, \cdots, i_k\}$ will be called the vertices. We identify i_ℓ and i_p iff they are equal. An edge is a pair (i, j) with $i, j \in \{1, \cdots, N\}^2$. At this point, edges are directed in the sense that we distinguish (i, j) from (j, i) when $j \neq i$ and we shall point out later when we consider undirected edges. We denote by $E(\mathbf{i})$ the collection of the k edges $(e_p)_{p=1}^k = (i_p, i_{p+1})_{p=1}^k$.

  We consider the graph $G(\mathbf{i}) = (V(\mathbf{i}), E(\mathbf{i}))$. $G(\mathbf{i})$ is connected by construction. Note that $G(\mathbf{i})$ may contain loops (i.e cycles, for instance edges of type (i, i)) and multiple undirected edges.

The skeleton $\tilde{G}(\mathbf{i})$ of $G(\mathbf{i})$ is the graph $\tilde{G}(\mathbf{i}) = \left(\tilde{V}(\mathbf{i}), \tilde{E}(\mathbf{i})\right)$ where vertices in $\tilde{V}(\mathbf{i})$ appear only once, edges in $\tilde{E}(\mathbf{i})$ are undirected and appear only once.

In other words, $\tilde{G}(\mathbf{i})$ is the graph $G(\mathbf{i})$ where multiplicities and orientation have been erased. It is connected, as is $G(\mathbf{i})$.

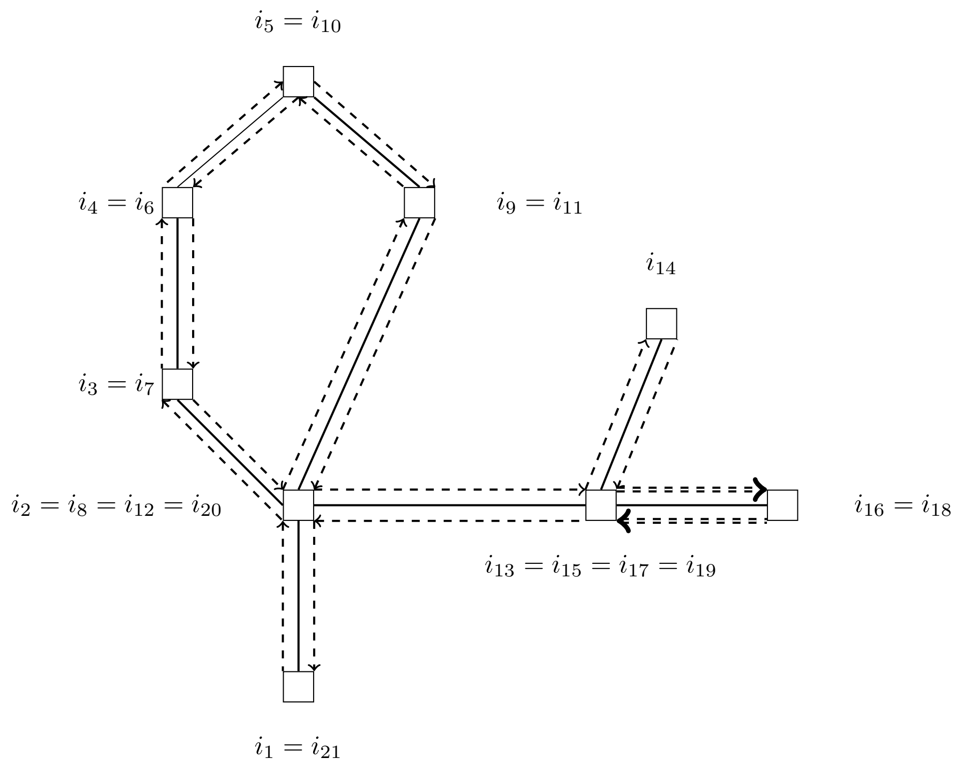

Figure 1: Figure of $G(\mathbf{i})$ (in dash) versus $\tilde{G}(\mathbf{i})$ (in bold), $|\tilde{E}(i)| = 9, |\tilde{V}(\mathbf{i})| = 9$

We now state and prove a well known inequality concerning undirected connected graphs $G = (V, E)$. If we let, for a discrete finite set A, $|A|$ be the number of its distinct elements, we have the following inequality

$$|V| \leq |E| + 1. \tag{11}$$

Let us prove this inequality and that equality holds only if G is a tree at the same time. This relation is straightforward when $|V| = 1$ and can be proven by induction as follows. Assume $|V| = n$ and consider one vertex v of V. This vertex is contained in l edges of E which we denote $(e_1, \cdots, e_l)$ and with $l \geq 1$ by connectedness. The graph G then decomposes into $(\{v\}, \{e_1, \cdots, e_l\})$ and $r \leq l$ undirected connected graphs $(G_1, \cdots, G_r)$.

We denote $G_j = (V_j, E_j)$, for $j \in \{1, \cdots, r\}$. We have

$$|V| - 1 = \sum_{j=1}^{r} |V_j|, \quad |E| - l = \sum_{j=1}^{r} |E_j|. \tag{12}$$

Applying the induction hypothesis to the graphs $(G_j)_{1\leq j\leq r}$ gives

$$\begin{aligned} |V| - 1 &\leq \sum_{i=1}^{r} (|E_j| + 1) \\ &= |E| + r - l \leq |E| \end{aligned} \tag{13}$$

which proves (11). In the case where $|V| = |E| + 1$, we claim that G is a tree, namely does not have a loop. In fact, for equality to hold, we need to have equalities when performing the previous decomposition of the graph, a decomposition which can be reproduced until all vertices have been considered. If the graph contains a loop, the first time that we erase a vertex of this loop when performing this decomposition, we will create one connected component less than the number of edges we erased and so a strict inequality occurs in the right hand side of (13) (i.e. $r < l$).

- *Convergence in expectation.*

  Since we noticed that $P(\mathbf{i})$ equals zero unless each edge in $E(\mathbf{i})$ is repeated at least twice, we have that

$$|\tilde{E}(\mathbf{i})| \leq 2^{-1}|E(\mathbf{i})| = \frac{k}{2}, \tag{14}$$

  and so by (11) applied to the skeleton $\tilde{G}(\mathbf{i})$ we find

$$|\tilde{V}(\mathbf{i})| \leq \left[\frac{k}{2}\right] + 1 \tag{15}$$

  where $[x]$ is the integer part of x. Thus, since the indices are chosen in $\{1, \cdots, N\}$, there are at most $N^{[\frac{k}{2}]+1}$ indices which contribute to the sum (9) and so we have

$$\left|\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right]\right| \leq B_k N^{[\frac{k}{2}]-\frac{k}{2}} \tag{16}$$

  where we used (5). In particular, if k is odd,

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N}\operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right] = 0. \tag{17}$$

  If k is even, the only indices which will contribute to the first order asymptotics in the sum are those such that

$$|\tilde{V}(\mathbf{i})| = \frac{k}{2} + 1, \tag{18}$$

since the other indices will be such that $|\tilde{V}(\mathbf{i})| \leq \frac{k}{2}$ and so will contribute at most by a term $N^{\frac{k}{2}} B_k N^{-\frac{k}{2}-1} = O(N^{-1})$. By the previous considerations, when $|\tilde{V}(\mathbf{i})| = \frac{k}{2}+1$, we have that

1. $\tilde{G}(\mathbf{i})$ is a tree,
2. $|\tilde{E}(\mathbf{i})| = 2^{-1}|E(\mathbf{i})| = \frac{k}{2}$ and so each edge in $E(\mathbf{i})$ appears exactly twice.

We can explore $G(\mathbf{i})$ by following the path P of edges $i_1 \to i_2 \to i_3 \cdots \to i_k \to i_1$. Since $\tilde{G}(\mathbf{i})$ is a tree, $G(\mathbf{i})$ appears as a fat tree where each edge of $\tilde{G}(\mathbf{i})$ is repeated exactly twice. We then see that each pair of directed edges corresponding to the same undirected edge in $\tilde{E}(\mathbf{i})$ is of the form $\{(i_p, i_{p+1}), (i_{p+1}, i_p)\}$ (since otherwise the path of edges has to form a loop to return to i_0). Therefore, for these indices, $\lim_N P^N(\mathbf{i}) = \lim_N E[|\sqrt{N} X^N_{ij}|^2]^{\frac{k}{2}} = 1$ does not depend on $\mathbf{i}$.

Finally, observe that $G(\mathbf{i})$ gives a pair partition of the edges of the path P (since each undirected edge has to appear exactly twice) and that this partition is non crossing (as can be seen by unfolding the path, keeping track of the pairing between edges by drawing an arc between paired edges). Therefore we have proved that

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N} \operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right] = \sharp\{ \text{ non-crossing pair partitions of } k \text{ edges } \}. \tag{19}$$

- *Almost sure convergence.* To prove the almost sure convergence, we estimate the variance and then use Borel Cantelli's lemma. The variance is given by

$$\begin{aligned} \operatorname{Var}((\mathbf{X}^N)^k) &:= \mathbb{E}\left[\frac{1}{N^2}\left(\operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right)^2\right] - \mathbb{E}\left[\frac{1}{N} \operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right]^2 \qquad (20)\\ &= \frac{1}{N^{2+k}} \sum_{\substack{i_1,\ldots,i_k=1\\ i'_1,\ldots,i'_k=1}}^{N} [P(\mathbf{i},\mathbf{i}') - P(\mathbf{i})P(\mathbf{i}')] \end{aligned}$$

with

$$P(\mathbf{i},\mathbf{i}') := \mathbb{E}[Y_{i_1 i_2} Y_{i_2 i_3} \cdots Y_{i_k i_1} Y_{i'_1 i'_2} \cdots Y_{i'_k i'_1}]. \tag{21}$$

We denote $G(\mathbf{i},\mathbf{i}')$ the graph with vertices $V(\mathbf{i},\mathbf{i}') = \{i_1, \cdots, i_k, i'_1, \cdots, i'_k\}$ and edges $E(\mathbf{i},\mathbf{i}') = \{(i_p, i_{p+1})_{1\leq p\leq k}, (i'_p, i'_{p+1})_{1\leq p\leq k}\}$. For $\mathbf{i},\mathbf{i}'$ to contribute to the sum, $G(\mathbf{i},\mathbf{i}')$ must be connected. Indeed, if $E(\mathbf{i}) \cap E(\mathbf{i}') = \emptyset$, $P(\mathbf{i},\mathbf{i}') = P(\mathbf{i})P(\mathbf{i}')$. Moreover, as before, each edge must appear at least twice to give a non zero contribution so that $|\tilde{E}(\mathbf{i},\mathbf{i}')| \leq k$. Therefore, we are in the same situation as before, and if $\tilde{G}(\mathbf{i},\mathbf{i}') = (\tilde{V}(\mathbf{i},\mathbf{i}'), \tilde{E}(\mathbf{i},\mathbf{i}'))$ denotes the skeleton of $G(\mathbf{i},\mathbf{i}')$, we have the relation

$$|\tilde{V}(\mathbf{i},\mathbf{i}')| \leq |\tilde{E}(\mathbf{i},\mathbf{i}')| + 1 \leq k+1. \tag{22}$$

This already shows that the variance is at most of order N^{-1} (since $P(\mathbf{i},\mathbf{i}') - P(\mathbf{i})P(\mathbf{i}')$ is bounded uniformly, independently of $(\mathbf{i},\mathbf{i}')$ and N), but we need a slightly better bound to prove the almost sure convergence. To improve our bound let us show that the case where $|\tilde{V}(\mathbf{i},\mathbf{i}')| = |\tilde{E}(\mathbf{i},\mathbf{i}')| + 1 = k+1$ can not occur. In this case, we have seen that $\tilde{G}(\mathbf{i},\mathbf{i}')$ must be a tree since then equality holds in (22). Also, $|\tilde{E}(\mathbf{i},\mathbf{i}')| = k$ implies that each edge appears with multiplicity exactly equal to 2. For any contributing set of indices $\mathbf{i},\mathbf{i}'$, $\tilde{G}(\mathbf{i},\mathbf{i}') \cap G(\mathbf{i})$ and $\tilde{G}(\mathbf{i},\mathbf{i}') \cap G(\mathbf{i}')$ must share at least one edge (i.e one edge must appear with multiplicity one in each subgraph) since otherwise $P(\mathbf{i},\mathbf{i}') = P(\mathbf{i})P(\mathbf{i}')$. This is a contradiction. Indeed, if we explore $\tilde{G}(\mathbf{i},\mathbf{i}')$ by following the path $i_1 \to i_2 \to \cdots \to i_1$, we see that each (non-oriented) visited edge appears twice or this path makes a loop. The first case is impossible since $\tilde{G}(\mathbf{i},\mathbf{i}') \cap G(\mathbf{i})$ and $\tilde{G}(\mathbf{i},\mathbf{i}') \cap G(\mathbf{i}')$ share one edge and each edge of $\tilde{G}(\mathbf{i},\mathbf{i}')$ has multiplicity 2, and the second case is also impossible since $\tilde{G}(\mathbf{i},\mathbf{i}')$ is a tree. Therefore, we conclude that for all contributing indices,

$$|\tilde{V}(\mathbf{i},\mathbf{i}')| \leq k \tag{23}$$

which implies

$$\mathrm{Var}((\mathbf{X}^N)^k) \leq p_k N^{-2} \tag{24}$$

with p_k a constant independent of N. Applying Chebychev's inequality gives for any $\delta > 0$

$$\mathbb{P}\left(\left| \frac{1}{N} \operatorname{Tr}\left((\mathbf{X}^N)^k\right) - \mathbb{E}\left[\frac{1}{N} \operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right] \right| > \delta \right) \leq \frac{p_k}{\delta^2 N^2}, \tag{25}$$

and so Borel-Cantelli's lemma implies

$$\lim_{N\to\infty} \left| \frac{1}{N} \operatorname{Tr}\left((\mathbf{X}^N)^k\right) - \mathbb{E}\left[\frac{1}{N} \operatorname{Tr}\left((\mathbf{X}^N)^k\right)\right] \right| = 0 \quad a.s. \tag{26}$$

The proof of the theorem is complete.

$\diamond$

Exercise 1.2. *Take for $L \in \mathbb{N}$, $\mathbf{X}^{N,L}$ the $N \times N$ self-adjoint matrix such that $\mathbf{X}^{N,L}_{ij} = (2L)^{-\frac{1}{2}} 1_{|i-j| \leq L} X_{ij}$ with $(X_{ij}, 1 \leq i \leq j \leq N)$ independent centered real random variables having all moments finite and $E[X_{ij}^2] = 1$. The purpose of this exercise is to show that for all $k \in \mathbb{N}$,*

$$\lim_{L\to\infty} \lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N}\ Tr((\mathbf{X}^{N,L})^k)\right] = C_{k/2} \tag{27}$$

with C_x null if x is not an integer. Moreover, if $L(N) \in \mathbb{N}$ is a sequence going to infinity with N so that $L(N)/N$ goes to zero, prove that

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N}\ Tr((\mathbf{X}^{N,L(N)})^k)\right] = C_{k/2}. \tag{28}$$

If $L(N) = [\alpha N]$, one can also prove the convergence of the moments of $\mathbf{X}^{N,L(N)}$. Show that this limit can not be given by the Catalan numbers $C_{k/2}$ by considering the case $k = 2$.

Hint: Show that for $k \geq 2$

$$\mathbb{E}\left[\frac{1}{N}\ Tr((\mathbf{X}^{N,L})^k)\right] = (2L)^{-k/2} \sum_{\substack{|i_2-\left[\frac{N}{2}\right]|\leq L,\\ |i_{p+1}-i_p|\leq L, p\geq 2}} \mathbb{E}[X_{[\frac{N}{2}]i_2}\cdots X_{i_k[\frac{N}{2}]}] + O(N^{-1}). \tag{29}$$

Then prove that the contributing indices to the above sum correspond to the case where $G(0, i_2, \cdot, i_k)$ is a tree with $k/2$ vertices and show that being given a tree there are approximately $(2L)^{\frac{k}{2}}$ possible choices of indices $i_2, \cdots, i_k$.

1.2 Stieltjes transform approach for Wigner's theorem

There is a second and very powerful approach to the study of the empirical measure of a random matrix through its Cauchy-Stieltjes transform. It is given, for $z \in \mathbb{C}\backslash\mathbb{R}$ by

$$G_N(z) = \frac{1}{N}\ \mathrm{Tr}(z - \mathbf{X})^{-1} = \frac{1}{N}\sum \frac{1}{z - \lambda_i}. \tag{30}$$

We shall see in this section how to prove law of large numbers by studying G_N, and in fact by showing that it is characterized by certain fixed points equations. The advantage of this approach is that it is well suited for more complicated matrices as we will see next. Also it can can be extended to study the global fluctuations of the eigenvectors. Finally, it can be refined sometimes to consider the case where the imaginary part of z goes to zero with N. This in turn will allow the study of more local properties of the spectrum, but also local delocalization.

1.2.1 Stieltjes transforms and Wigner theorem

We begin by recalling some classical results concerning the Stieltjes transform of a probability measure.

Definition 1.3. *Let μ be a positive, finite measure on the real line. The Stieltjes transform of μ is the function*

$$G_\mu(z) := \int_{\mathbb{R}} \frac{\mu(dx)}{z - x}\,,\ z \in \mathbb{C}\setminus\mathbb{R}\,. \tag{31}$$

Note that for $z \in \mathbb{C} \setminus \mathbb{R}$, both the real and imaginary parts of $1/(x-z)$ are continuous bounded functions of $x \in \mathbb{R}$, and further $|G_\mu(z)| \le \mu(\mathbb{R})/|\Im z|$. These crucial observations are used repeatedly in what follows.

Remark 1.4. *The generating function $\hat{\beta}(z)$ of moments of the semicircle distribution σ is closely related to its Stieltjes transform : for $|z| < 1/4$,*

$$\begin{aligned}
\hat{\beta}(z) &= \sum_{k=0}^{\infty} z^k \int x^{2k} \sigma(x) dx = \int \left(\sum_{k=0}^{\infty} \left(zx^2\right)^k \right) \sigma(x) dx \qquad (32) \\
&= \int \frac{1}{1 - zx^2} \sigma(x) dx \\
&= \int \frac{1}{1 - \sqrt{z}x} \sigma(x) dx = \frac{1}{\sqrt{z}} G_\sigma(1/\sqrt{z}),
\end{aligned}$$

where the third equality uses that the support of σ is the interval $[-2,2]$, and the fourth uses the symmetry of σ. Using the fact that $\int x^{2k} d\sigma(x) = C_k$ is the Catalan number which satisfies the induction relation

$$C_k = \sum_{\ell=0}^{k-1} C_\ell C_{k-\ell-1} \qquad (33)$$

it is not hard to deduce from the above that

$$G_\sigma(z) - \frac{1}{z} = \frac{1}{z} G_\sigma(z)^2. \qquad (34)$$

Stieltjes transforms can be inverted. In particular, one has

Theorem 1.5. *For any open interval I with neither endpoint on an atom of μ,*

$$\begin{aligned}
\mu(I) &= \lim_{\epsilon \to 0} \frac{1}{\pi} \int_I \frac{G_\mu(\lambda - i\epsilon) - G_\mu(\lambda + i\epsilon)}{2i} d\lambda \\
&= \lim_{\epsilon \to 0} \frac{1}{\pi} \int_I \Im G_\mu(\lambda - i\epsilon) d\lambda \,. \qquad (35)
\end{aligned}$$

Proof. Note first that because

$$\Im G_\mu(-i) = \int \frac{1}{1+x^2} \mu(dx) \,, \qquad (36)$$

we have that $G_\mu \equiv 0$ implies $\mu = 0$. So assume next that G_μ does not vanish identically. Then, since

$$\lim_{y \uparrow +\infty} y \Im G_\mu(-iy) = \lim_{y \uparrow +\infty} \int \frac{y^2}{x^2+y^2} \mu(dx) = \mu(\mathbb{R}) \qquad (37)$$

by bounded convergence, we may and will assume that $\mu(\mathbb{R}) = 1$, i.e. that μ is a probability measure.

Let X be distributed according to μ, and denote by C_ϵ a random variable, independent of X, Cauchy distributed with parameter ϵ, i.e. the law of C_ϵ has density

$$\frac{\epsilon dx}{\pi(x^2+\epsilon^2)}\,. \tag{38}$$

Then, $\Im G_\mu(\lambda - i\epsilon)/\pi$ is nothing but the density (with respect to Lebesgue measure) of the law of $X + C_\epsilon$ evaluated at $\lambda \in \mathbb{R}$. The convergence in (35) is then just a rewriting of the weak convergence of the law of $X + C_\epsilon$ to that of X, as $\epsilon \to 0$. $\diamond$The main tool in the analysis of the empirical measure through Stieltjes transform is the Schur complement formula that we present next.

Lemma 1.6. *Let W be a Hermitian matrix, and let w_i denote the i-th column of W with the entry $W(i,i)$ removed (i.e., w_i is an $N-1$-dimensional vector). Let $W^{(i)} \in \mathbb{H}^{(1)}_{N-1}$ ($\mathbb{H}^{(1)}_{N-1}$ being the set of $N-1 \times N-1$-hermitian matrices) denote the matrix obtained by erasing the i-th column and row from W. Then, for every $z \in \mathbb{C} \setminus \mathbb{R}$,*

$$(zI - W)^{-1}(i,i) = \frac{1}{z - W(i,i) - w_i^T (zI_{N-1} - W^{(i)})^{-1} w_i}\,. \tag{39}$$

Proof of Lemma 1.6 Note first that from Cramer's rule,

$$(zI_N - W)^{-1}(i,i) = \frac{\det(zI_{N-1} - W^{(i)})}{\det(zI - W)}\,. \tag{40}$$

Write next

$$zI_N - W = \begin{pmatrix} zI_{N-1} - W^{(N)} & -w_N \\ -w_N^T & z - W(N,N) \end{pmatrix}, \tag{41}$$

and use the matrix identity

$$\begin{aligned}\det\begin{pmatrix} A & B \\ C & D\end{pmatrix} &= \det\left(\begin{pmatrix} A & 0 \\ C & D - CA^{-1}B\end{pmatrix}\begin{pmatrix} 1 & A^{-1}B \\ 0 & 1\end{pmatrix}\right) \\ &= \det A \det(D - CA^{-1}B) \end{aligned} \tag{42}$$

with $A = zI_{N-1} - W^{(N)}$, $B = -w_N$, $C = -w_N^T$ and $D = z - W(N,N)$ to conclude that

$$\begin{aligned}&\det(zI_N - W) = \\ &\det(zI_{N-1} - W^{(N)}) \det\left[z - W(N,N) - w_N^T (zI_{N-1} - W^{(N)})^{-1} w_N\right]\,.\end{aligned} \tag{43}$$

The last formula holds in the same manner with $W^{(i)}$, w_i and $W(i,i)$ replacing $W^{(N)}, w_N$ and $W(N,N)$ respectively. Substituting in (40) completes the proof of Lemma 1.6. $\diamond$

With a truncation argument, we may and will assume in the sequel that $\mathbf{X}_N(i,i) = 0$ for all i and that for some constant C independent of N, it holds that $|\sqrt{N}\mathbf{X}_N(i,j)| \le C$ for all i,j.

Denote by $\mathbf{X}_N^{(k)} \in \mathbb{H}_N^{(1)}$ the matrix consisting of $\mathbf{X}_N$ with the kth row and column removed. By Lemma 1.6, one gets that

$$G_N(z) \equiv \frac{1}{N} \operatorname{Tr}(z - \mathbf{X}_N)^{-1} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z - \alpha_i^T (zI_{N-1} - \mathbf{X}_N^{(i)})^{-1} \alpha_i} \tag{44}$$

The main idea is to derive LLN from the convergence and fluctuations of $\alpha_i^T (zI_{N-1} - \mathbf{X}_N^{(i)})^{-1} \alpha_i$.

Theorem 1.7. *Assume that* $\sup_N \mathbb{E}[|\sqrt{N}X|^{2+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$. *Then* $G_N(z)$ *converges almost surely when* N *goes to infinity towards the unique solution of*

$$G(z) = \frac{1}{z - G(z)} \tag{45}$$

going to zero at infinity,

Proof. The first point in the proof is that $G_N(z)$, if it converges in expectation, will converge almost surely by concentration of measure.

1.2.2 Concentration of measure

Concentration of measure has played an important role in probability theory and random matrix theory. It turns out that general results of concentration of measure apply to this setting [15, 2]: such results demand in general strong assumptions over the entries. We first give a concentration result due to C. Bordenave, P. Caputo and D. Chafai which is based on Azuma-Hoeffding's inequality. The proof reveals that we only require the independence of the vectors $\{(X_{ij})_{j \leq i}, 1 \leq i \leq N\}$.

Lemma 1.8. *Let* $\|f\|_{TV}$ *be the total variation norm,*

$$\|f\|_{TV} = \sup_{x_1 < \cdots < x_n} \sum_{i=2}^{n} |f(x_i) - f(x_{i-1})| \tag{46}$$

Then, for any $\delta > 0$ *and any function* f *with finite total variation norm so that* $E\left[|\frac{1}{N} \sum_{i=1}^{N} f(\lambda_i)| \right] < \infty$,

$$P\left(\left| \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) - \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) \right] \right| \geq \delta \|f\|_{TV} \right) \leq 2e^{-\frac{N\delta^2}{8c_X}} \tag{47}$$

where $c_X = 1$ *for Wigner's matrices and* M/N *for Wishart matrices.*

Note that this lemma holds for real valued functions. For complex function it can be easily extended by changing the constant 8 by 16.

Remark 1.9. *Note that the above speed is not optimal for laws μ_N, ν_N which have sufficiently fast decaying tails as we will see below, in which case $\sum_{i=1}^N f(\lambda_i) - \mathbb{E}[\sum_{i=1}^N f(\lambda_i)]$ is of order one. However it is the optimal rate for instance for heavy tailed matrices where CLT holds for $N^{-1/2}\left(\sum_{i=1}^N f(\lambda_i) - \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N f(\lambda_i)\right]\right)$.*

Remark 1.10. *Note that we only required independence of the vectors of the random matrix, rather than the entries.*

Proof. Let us first recall Azuma-Hoeffding's inequality

Lemma 1.11. *(Azuma-Hoeffding's inequality) Suppose $M_k, k \geq 0$ is a martingale for the filtration $\mathcal{F}_k$ and $|M_k - M_{k-1}| \leq c_k$. Then for all $t \geq 0$*

$$P(M_n - M_0 \geq t) \leq \exp\{-\frac{t^2}{2\sum_{k=1}^n c_k^2}\}\,. \tag{48}$$

Proof. By Tchebychev's inequality for all $\lambda \geq 0$

$$P\left(M_n - M_0 \geq t\right) \leq e^{-\lambda t}\mathbb{E}[e^{\lambda\sum_{k=1}^n (M_k - M_{k-1})}]\,. \tag{49}$$

We first integrate conditionally to $\mathcal{F}_{n-1}$, that is control uniformly

$$f(\lambda) = \log \mathbb{E}[e^{\lambda(M_n - M_{n-1})}|\mathcal{F}_{n-1}]\,. \tag{50}$$

Clearly, $f(0) = f'(0) = 0$ whereas

$$f''(\lambda) \leq \frac{\mathbb{E}[(M_n - M_{n-1})^2 e^{\lambda(M_n - M_{n-1})}|\mathcal{F}_{n-1}]}{\mathbb{E}[e^{\lambda(M_n - M_{n-1})}|\mathcal{F}_{n-1}]} \leq c_n^2\,. \tag{51}$$

Therefore, we have the uniform bound

$$\mathbb{E}[e^{\lambda(M_n - M_{n-1})}|\mathcal{F}_{n-1}] \leq e^{\frac{1}{2}\lambda^2 c_n^2} \tag{52}$$

and proceeding by induction we deduce

$$\mathbb{E}[e^{\lambda\sum_{k=1}^n (M_k - M_{k-1})}] \leq e^{\frac{1}{2}\lambda^2\sum_{k=1}^n c_k^2}\,. \tag{53}$$

Plugging back this control into (49) and taking $\lambda = t/\sum c_k^2$ yields the lemma. $\diamond$

We finally prove Lemma 1.8 for a continuously differentiable function f, the generalization to all functions with finite variation norm then holds by density. We then have $\|f\|_{TV} = \int |f'(x)|dx$. We apply Azuma-Hoeffding's inequality to

$$M_k = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N f(\lambda_i)|\mathcal{F}_k\right] \tag{54}$$

where $\mathcal{F}_k$ is the filtration generated by $\{\mathbf{X}_N(i,j), 1 \leq i \leq j \leq k\}$. M_k is a martingale obviously and

$$M_N - M_0 = \frac{1}{N}\sum_{i=1}^N f(\lambda_i) - \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N f(\lambda_i)\right]\,. \tag{55}$$

Therefore we need to bound for each $k \in \{1, \cdots, N\}$

$$\begin{aligned} M_k - M_{k-1} &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N f(\lambda_i) - \mathbb{E}[\frac{1}{N}\sum_{i=1}^N f(\lambda_i)|\mathcal{F}_{k-1}]|\mathcal{F}_k\right] \\ &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N f(\lambda_i) - \frac{1}{N}\sum_{i=1}^N f(\tilde{\lambda}_i)|\mathcal{F}_k\right], \end{aligned} \tag{56}$$

where in the above expectation λ_i (resp $\tilde{\lambda}_i$) are the eigenvalues of $Z_N = \mathbf{X}_N(X_k, C_k, R_{k+1})$ (resp. $\tilde{Z}_N = \mathbf{X}_N(X_k, \tilde{C}_k, R_{k+1})$) where $\mathbf{X}_N(X, C, R)$ is the self-adjoint matrix with upper left $k \times k$ corner given by $X = (\mathbf{X}_N(X,C,R))_{i,j\le k}$, $k+1$th vector $C = (\mathbf{X}_N(X,C,R))_{k+1, j\le k+1}$ and rest $R = (\mathbf{X}_N(X,C,R))_{i,\text{ or } j\ge k+2}$. In the conditionnal expectation above we therfore take expectation over R_{k+1} and $\tilde{C}_k$. Hence the eigenvalues λ and $\tilde{\lambda}$ are the eigenvalues of two operators which have a same submatrix of dimension $N-1\times N-1$. Therefore the eigenvalues λ_i and $\tilde{\lambda}_i$ are interlaced, that is they are ordered

$$\tilde{\lambda}_{i-1} \le \lambda_i \le \tilde{\lambda}_{i+1}\,. \tag{57}$$

If g is increasing we deduce that

$$\sum_{i=1}^{N-2} g(\tilde{\lambda}_i) \le \sum_{i=2}^{N-1} g(\lambda_i) \le \sum_{i=3}^{N} g(\tilde{\lambda}_i) \tag{58}$$

which implies

$$|\sum_{i=1}^{N} g(\lambda_i) - \sum_{i=1}^{N} g(\tilde{\lambda}_i)| \le 2\|g\|_\infty. \tag{59}$$

Decomposing $f(x) - f(0)$ as the sum of the two increasing functions

$$f(x) - f(0) = \int_0^x f'(y)1_{f'(y)\ge 0}dy - \int_0^x (-f')(y)1_{f'(y)<0}dy \tag{60}$$

shows that

$$|M_k - M_{k-1}| \le \frac{2}{N}\|f\|_{TV} \tag{61}$$

which allows to conclude that for all $\delta > 0$

$$P\left(\frac{1}{N}\sum_{i=1}^N f(\lambda_i) - \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N f(\lambda_i)\right] \ge \delta\|f\|_{TV}\right) \le e^{-\frac{\delta^2 N}{8c_X}}\,. \tag{62}$$

The other bound is obtained by changing f into $-f$. $\diamond$Note that this estimate can be improved under stronger assumptions but it is optimal in general, for instance if the entries have heavy tails. If the entries satisfy log-Sobolev inequalities, that is there exists a finite constant c such that for any differentiable function $f : \mathbb{R}^p \to \mathbb{R}$ we have

$$\mu\left(f^2 \log \frac{f^2}{\mu(f^2)}\right) \le 2c\mu(\|\nabla f\|_2^2).$$

This is true for the Gaussian law, or for $e^{-V(x)}dx/Z$ with $\mathrm{Hess}(V) \geq I$, or the image $\varphi(g)$ of Gaussian variables by a Lipschitz map. Then a general result due to Herbst shows that

Lemma 1.12 (Herbst)**.** *Let $X_{i,j}$ be independent and with laws satisfying the log-Sobolev inequality with constant uniformly bounded by c. Let g be a Lipschitz function. For all $\delta \geq 0$*

$$\mathbb{P}\left(|g(X_{i,j}, i \leq j) - \mathbb{E}[g(X_{i,j}, i \leq j)]| \geq \delta \|g\|_L\right) \leq 2e^{-C\delta^2}$$

where C depends only on c.

Throughout these notes, we denote the Lipschitz constant of a function $G : \mathbb{R}^M \to \mathbb{R}$ by

$$|G|_L := \sup_{x \neq y \in \mathbb{R}^M} \frac{|G(x) - G(y)|}{\|x - y\|_2},$$

Similar results hold for bounded entries if one recenters the random variable with respect to its median thanks to Talagrands results:

Theorem 1.13 (Talagrand)**.** *Let K be a connected compact subset of $\mathbb{R}$ with diameter $|K| = \sup_{x,y \in K} |x - y|$. Consider a convex real-valued function f defined on K^N. Assume that f is Lipschitz on K^N, with constant $|f|_{\mathcal{L}}$. Let P be a probability measure on K and $X_1, \cdots, X_N$ be N independent copies with law P. Then, if M_f is the median of $f(X_1, \cdots, X_N)$, for all $\epsilon > 0$,*

$$P\left(|f(X_1, \cdots, X_N) - M_f| \geq \epsilon\right) \leq 4e^{-\frac{\epsilon^2}{16|K|^2|f|_{\mathcal{L}}^2}}.$$

These concentration statements apply to random matrices as soon as we can show that the functions we are interested in are Lipschitz. This is for instance the case of the eigenvalues. Indeed, thanks to Lidskii's Theorem **??** and for the spectral measure

Lemma 1.14. *We denote $\lambda_1(\mathbb{A}) \leq \lambda_2(\mathbb{A}) \leq \cdots \leq \lambda_N(\mathbb{A})$ the eigenvalues of $\mathbb{A} \in \mathbb{H}$. Then for all $k \in \{1, \cdots, N\}$, all $\mathbb{A}, \mathbb{B} \in \mathbb{H}$,*

$$|\lambda_k(\mathbb{A} + \mathbb{B}) - \lambda_k(\mathbb{A})| \leq \|\mathbb{B}\|_2.$$

In other words, for all $k \in \{1, \cdots, N\}$,

$$(A_{ij})_{1 \leq i \leq j \leq N} \in \mathbb{R}^{N(N+1)/2} \to \lambda_k(\mathbb{A})$$

is Lipschitz with constant one.

For all Lipschitz functions f with Lipschitz constant $|f|_L$, the function

$$(A_{ij})_{1 \leq i \leq j \leq N} \in \mathcal{E}_N^{(2)} \to \sum_{k=1}^{N} f(\lambda_k(\mathbb{A}))$$

is Lipschitz with respect to the Euclidean norm with a constant bounded above by $\sqrt{N}|f|_L$. When f is continuously differentiable we have

$$\lim_{\epsilon\to 0}\epsilon^{-1}\left(\sum_{k=1}^{N} f(\lambda_k(\mathbb{A}+\epsilon\mathbb{B}))-\sum_{k=1}^{N} f(\lambda_k(\mathbb{A}))\right)=Tr(f'(\mathbb{A})\mathbb{B}).$$

We can apply these results to Wigner matrices. For instance if its entries satisfy log-Sobolev inequalities we have

Corollary 1.15. *Let $\mathbf{X}_N(ij), 1\le i\le j\le N$ be independent, satisfying log-Sobolev inequality with uniformly bounded constant and denote by λ_k the eigenvalues of $\mathbf{X}_N$ in increasing order.Then there exists $C>0$ such that*

- *for any $\delta>0$ any Lipschitz function $f\mathbb{R}\to\mathbb{R}$*

$$\mathbb{P}\left(|\ Tr(f(\frac{\mathbf{X}_N}{\sqrt{N}}))-\mathbb{E}[\ Tr(f(\frac{\mathbf{X}_N}{\sqrt{N}}))]|>\delta\|f\|_L\right)\le 2e^{-C\delta^2}$$

- *for any $\delta>0$*

$$\mathbb{P}\left(|\lambda_k-\mathbb{E}[\lambda_k]|>\delta\right)\le 2e^{-C\delta^2}$$

- *For any fixed unit vector u, any $z\in\mathbb{C}\backslash\mathbb{R}$*

$$\mathbb{P}\left(|\langle u,(z-\frac{\mathbf{X}_N}{\sqrt{N}})^{-1}u\rangle-\mathbb{E}[\langle u,(z-\frac{\mathbf{X}_N}{\sqrt{N}})^{-1}u\rangle]|>\delta/|\Im z|^2\right)\le 2e^{-CN\delta^2}$$

The first two statements are clear consequences of the previous lemma, the last is by noticing that $X_{ij}\to\langle u,f(\frac{\mathbf{X}_N}{\sqrt{N}})u\rangle$ is Lipschitz with norm bounded by $1/\sqrt{N}|\mathrm{Im}z|^2$.

Another concentration of measure result which is useful concerns the case of a determinitic matrix evaluated in the direction of a random vector. For instance if $(X_i)_{1\le i\le N}$ have sub-Gaussian tails in the sense that for all $p\in\mathbb{N}^*$

$$\max_i \mathbb{E}[|X_i|^p]\le K^p p^{p/2}$$

and $E[X_i]=0$, Hanson-Wright inequalities provide exponential bounds :

$$\mathbb{P}\left(|\langle X,AX\rangle-\mathbb{E}_X[XAX^T]|>t\right)\le 2e^{-c\min\{\frac{t^2}{K^4\|A\|_F^2},\frac{t}{\|A\|K^2}\}}\qquad(63)$$

where $\|A\|_F^2=\sum|A_{ij}|^2$, $\|A\|=\sup_{\|x\|_2=1}|\langle Ax,x\rangle|$. This kind of control can be retrieved by using Herbst argument or Talagrand's result. Indeed, it is enough to prove it for $A\ge 0$ and then one can use that $X\to(XAX^T)^{1/2}$ is Lipschitz with norm $\|A\|^{1/2}$. Or that $X\to XAX^T$ is Lipschitz on $\{\|X\|_2^2/N\le K\}$ with norm $2\sqrt{N}\|A\|K$, use the fact that you can extend a Lipschitz function, see [16, section 5.4], to deduce that

$$\mathbb{P}\left(\{|\langle X,AX\rangle-\mathbb{E}_X[1_{\|X\|_2^2\le KN}XAX^T]|>\delta\sqrt{N}K\|A\|\}\cap\{\{\|X\|_2^2/N\le K\}\right)\le e^{-c\delta^2}$$

One then uses again these arguments to find that since $\mathbb{E}[\|X\|_2^2] = N$

$$\mathbb{P}\left(\|X\|_2^2 \geq (K+1)N\right) \leq e^{-cK^2 N}$$

The result follows by optimizing over the choice of K.

1.2.3 Convergence of the Stieltjes transform

We therefore from now on can prove the convergence of $\mathbb{E}[G_N(z)]$ instead of $G_N(z)$ since the function $(z - .)^{-1}$ has finite total variation.

The main idea in the proof is that the convergence and fluctuations of the term $\alpha_i^T(zI_{N-1} - \mathbf{X}_N^{(i)})^{-1}\alpha_i$ in terms of G_N will provide the convergence and fluctuations of $G_N(z)$. To this end let us write

$$\begin{aligned}\alpha_i^T(\mathbf{X}_N^{(i)} - zI_{N-1})^{-1}\alpha_i &= \sum_{j \neq k} \overline{\alpha_i(j)}\alpha_i(k)(\mathbf{X}_N^{(i)} - zI_{N-1})^{-1}_{jk} + \sum_j |\alpha_i(j)|^2 (\mathbf{X}_N^{(i)} - zI_{N-1})^{-1}_{jj} \quad \text{span} \\ &=: O(z) + D(z).\end{aligned} \tag{64}$$

We first observe that the off diagonal terms $O(z)$ will always be neglectable

Lemma 1.16. *Under the assumptions of the theorem, for all $\varepsilon > 0$, for any matrix C such that N^{-1} $TrCC^T$ is bounded independently of N*

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\sum_{j\neq k} \overline{\alpha_i(j)}\alpha_i(k)C_{jk}\right| \geq \varepsilon\right) = 0\,. \tag{65}$$

Proof. In fact, with probability going to 1 we may assume that all the A_{ij}, $1 \leq j \leq N$ in the decomposition $X = A + B$ are zero so that the variance of α_i can be assumed to be much smaller than $N^{-1/2}$. But then Tchebychev's inequality and independence yields

$$\mathbb{P}\left(\left|\sum_{j\neq k} \alpha_i(j)\alpha_i(k)C_{jk}\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sum_{j,k} \mathbb{E}[|\alpha_i(j)|^2]^2 C_{jk}^2 \ll \frac{1}{\varepsilon^2 N}\ \mathrm{Tr}(CC^T) \tag{66}$$

which proves the claim. ◇

We hence need only to focus on the diagonal term $D(z)$ and consider the equation

$$(z - \mathbf{X})^{-1}_{ii} = \frac{1}{z - \sum_{j\neq i} |\alpha_i(j)|^2 (z - \mathbf{X}^{(i)})^{-1}_{jj} + \varepsilon_i^N(z)} \tag{67}$$

with some $\varepsilon_i^N(z)$ going to zero in probability with N going to infinity. Recall that $\sqrt{N}\alpha_i(j)$ belongs to $L^{2+\varepsilon}$. In this case, LLN insures that

$$\lim_{N\to\infty}\left(\sum_j |\alpha_i(j)|^2(\mathbf{X}_N^{(i)} - zI_{N-1})^{-1}_{jj} - \frac{1}{N}\sum_j (\mathbf{X}_N^{(i)} - zI_{N-1})^{-1}_{jj}\right) = 0 \qquad a.s. \tag{68}$$

Hence, we see that

$$(z - \mathbf{X})^{-1}_{ii} = \frac{1}{z - \frac{1}{N}\ \mathrm{Tr}(z - \mathbf{X}^{(i)})^{-1} + \varepsilon_i^n(z)'} \tag{69}$$

where $\varepsilon_i^n(z)'$ goes to zero in probability. Finally, by (59), we see that

$$\left| \frac{1}{N}\ \mathrm{Tr}(z - \mathbf{X}^{(i)})^{-1} - \frac{1}{N}\ \mathrm{Tr}(z - \mathbf{X})^{-1} \right| \le \frac{2}{N \Im z}. \tag{70}$$

Therefore

$$\mathbb{E}[G_N(z)] \simeq \frac{1}{z - \mathbb{E}[G_N(z)]} \tag{71}$$

which implies, since we know that $\mathbb{E}[G_N(z)]$ goes to zero as $\Im z$ goes to infinity, that

$$\mathbb{E}[G_N(z)] \simeq \frac{z - \sqrt{z^2 - 4}}{2}. \tag{72}$$

This proves the claim.

1.3 Pastur-Marchenko's theorem and Stieltjes transform

Let $\mathbb{W}_N = \frac{1}{N}\mathbf{X}_N^T \mathbf{X}_N$ with $(X(i,j))_{\substack{1 \le i \le M \\ 1 \le j \le N}}$ independent centered entries with covariance 1. To simplify we assume they have the same distribution and are real valued (eventhough the proofs generalized readily to complex entries and different distributions provided they are centered and with covariance one). We assume throughout that M/N goes to α and assume $\alpha \le 1$ without loss of generality. Let $\hat{\mu}_N$ be the spectral measure of $\mathbb{W}_N$.

Theorem 1.17. *Assume that the X_{ij} have finite $2 + \epsilon$ moment. Almost surely $\hat{\mu}_N$ converges towards μ_α given by*

$$\mu_\alpha(dx) = \frac{1}{2\pi\alpha x}\sqrt{(\lambda_+ - x)(x - \lambda_-)}dx$$

where $\lambda_\pm = \alpha(1 \pm \sqrt{\alpha}^{-1})^2$. μ_α is the unique measure with Stieltjes transform

$$G_{\mu_\alpha}(z) = \frac{1}{2z}(z + 1 - \alpha - \sqrt{(z + 1 - \alpha)^2 - 4z})\,.$$

We shall prove a generalization of this theorem for the matrix $P \times P$ matrix

$$\mathbb{V}_N = \frac{1}{P} S^T \mathbf{X}_N^T \mathbf{X}_N S$$

for a deterministic $N \times P$ matrix S. Observe that the eigenvalues of $\mathbb{V}_N$ are the same than those of the $N \times N$ matrix

$$\mathbb{W}_N = \frac{1}{P}\mathbf{X}_N S S^T \mathbf{X}_N^T$$

except for $|N-P|$ eigenvalues equal to zero. In particular we can replace S by the self-adjoint $N\times N$ matrix $(SS^T)^{1/2}$ without changing the spectrum except for some trivial null eigenvalues. This section is inspired from Bai-Silverstein [21]. In view of the next section where we generalize our approach to questions from Deep learning we stick with the slightly more complicated set up following Couillet-Louart-Liao [17] which give the conbvergence of the resolvent for the operator norm. We assume that the vectors $X_i=(X_{ij})_j$ are independent, centered with covariance one and such that we have the following concentration of measure: there exists $p>1$ for any deterministic matrix A

$$\mathbb{P}\left(|\langle X,AX\rangle-\mathbb{E}_X[XAX^T]|>\delta t\max\{\|A\|_F,\|A\|\}\right)\le\frac{1}{t^p}\tag{73}$$

This is slightly stronger than what we get when the X_{ij} have moments $2+\epsilon$. Finite and constant fourth moment is enough for $p=2$. This is true if X have the sub-Gaussian tails so that Hanson-Wright inequalities (63) apply. Let for $z\in\mathbb{C}^+$

$$R_N(z)=(z-\mathbb{V}_N)^{-1},\quad G_N(z)=\frac{1}{P}\operatorname{Tr}(R_N(z)),.$$

Theorem 1.18. *Under the above assumption, assume that $\Phi=S^TS$ is uniformly bounded and that P,N,M are of the same order. Then*

$$\lim_{N\to\infty}\|R_N(z)-(z-\Phi((\frac{M}{P}-1+zG_S(z))^{-1}\|_\infty=0\qquad a.s.$$

where G_S is the unique solution from $\mathbb{C}^+$ into $(1-\alpha)/z+\alpha G_S(z)\in\mathbb{C}^-$ of

$$G_S(z)=\frac{1}{P}Tr((z-\Phi(\frac{M}{P}-1+zG_S(z)))^{-1}).$$

Remark 1.19.
- *If $S=Id$, and M/N going to α we find that $G=G_{Id}$ is solution of*

  $$G(z)=(z-1+\alpha-\alpha zG(z))^{-1}$$

  yiielding

  $$G(z)=\frac{1}{2\alpha z}(z+\alpha-1-\sqrt{(z+\alpha-1)^2-4\alpha z})$$

  as announced

- *In [17], a precise speed of convergence of the operator norm is bounded by $N^{-1/2+\epsilon}$ for some $\epsilon>0$.*

Let $G_N(z)=\frac{1}{P}\operatorname{Tr}(z-\mathbb{V}_N)^{-1}$. First observe that the conclusion of Lemma 1.8 holds (has $\mathbb{V}_N$ is the sum of independent rank one projections since T is diagonal), so that since $\|(z-.)^{-1}\|_{TV}\le\frac{1}{|\Im z|}$, we know that

$$\mathbb{P}\left(|G_N(z)-\mathbb{E}[G_N(z)]|\ge\delta\frac{1}{|\Im z|}\right)\le e^{-Pc\delta^2}$$

Hence, we need only to show the convergence $\mathbb{E}[G_N(z)]$ to deduce by Borel-Cantelli Lemma the almost sure convergence of G_N. Moreover, notice that $G_N(z)$ is a sequence of analytic functions which is uniformly bounded on $\Im z \geq \epsilon$.

Denote by $R_N(z) = (z - \mathbb{V}_N)^{-1}$ the resolvent and write $\mathbb{V}_N = \frac{1}{P}\sum_{i=1}^M v_i^T v_i$ with $v_i = (WS)_i$. Observe that the v_i are independent. We want to approximate $\mathbb{E}[R_N(z)]$ by using the decomposition

$$R_N(z)\mathbb{V}_N = \frac{1}{P}\sum_{i=1}^M (z - \mathbb{V}_N)^{-1} v_i^T v_i$$

and use the identity

$$(A + tvv^T)^{-1}v = \frac{A^{-1}v}{1 + t\langle v, A^{-1}v\rangle} \tag{74}$$

with $A = z - \mathbb{V}_N^i, \mathbb{V}_N^i = \frac{1}{N}\sum_{j\neq i} v_j^T v_j$, $v = v_i$ and $t = -1/N$. This gives

$$-1 + zR_N(z) = R_N(z)\mathbb{V}_N = \frac{1}{P}\sum_{i=1}^M \frac{1}{1 - \frac{1}{N}\langle v_i^T, (z - \mathbb{V}_N^i)^{-1} v_i^T\rangle}(z - \mathbb{V}_N^i)^{-1} v_i^T v_i \tag{75}$$

Hereafter we denote by $\delta_i^z = \frac{1}{N}\langle v_i^T, (z - \mathbb{V}_N^i)^{-1} v_i^T\rangle$ and $\delta_z = \mathbb{E}[\frac{1}{N}\ \mathrm{Tr}(\Phi R_N(z))]$. We will use later that δ_i^z goes to δ_z by (73). Taking the the expectation we deduce that if $\alpha = M/P$

$$\begin{aligned}
& z\mathbb{E}[R_N(z)] - I = \mathbb{E}[\frac{1}{P}\sum_{i=1}^M \frac{1}{1 - \delta_i^z}(z - \mathbb{V}_N^i)^{-1} v_i^T v_i] \\
= \quad & \mathbb{E}[\frac{1}{P}\sum_{i=1}^M \frac{1}{1 - \delta_z}(z - \mathbb{V}_N^i)^{-1} v_i^T v_i] + \mathbb{E}[\frac{1}{P}\sum_{i=1}^M (\frac{1}{1 - \delta_z^i} - \frac{1}{1 - \delta_z})(z - \mathbb{V}_N^i)^{-1} v_i^T v_i] \\
= \quad & \mathbb{E}[\frac{1}{P}\sum_{i=1}^M \frac{1}{1 - \delta_z}\mathbb{E}[(z - \mathbb{V}_N^i)^{-1}]\Phi] + \mathbb{E}[\frac{1}{P}\sum_{i=1}^M (\frac{1}{1 - \delta_z^i} - \frac{1}{1 + \delta_z})(z - \mathbb{V}_N^i)^{-1} v_i^T v_i] \\
= \quad & \frac{\alpha}{1 - \delta_z}\mathbb{E}[R_N(z)]\Phi + \varepsilon_N(z)
\end{aligned} \tag{76}$$

where we used $\mathbb{E}[v_i^T v_i] = \Phi$ and that $\mathbb{V}_N^i$ is independent of v_i. We have set :

$$\varepsilon_N(z) = \varepsilon_N^1(z) + \varepsilon_N^2(z)$$

with

$$\begin{aligned}
\varepsilon_N^1(z) \quad &= \quad \mathbb{E}[\frac{1}{P}\sum_{i=1}^M \left(\frac{1}{1 - \delta_z^i} - \frac{1}{1 - \delta_z}\right)(z - \mathbb{V}_N^i)^{-1} v_i v_i^T] \\
\varepsilon_N^2(z) \quad &= \quad \mathbb{E}[\frac{1}{P}\sum_{i=1}^M \frac{1}{1 - \delta_z}((z - \mathbb{V}_N^i)^{-1} - (z - \mathbb{V}_N)^{-1})\Phi]
\end{aligned}$$

We will prove that the operator norm of $\varepsilon_N(z)$ is small at the end of the section. Then, solving the above equation we find that

$$(z - \frac{\alpha\Phi}{1-\delta_z})\mathbb{E}[R_N(z)] = I + \varepsilon_N(z)$$

so that

$$\mathbb{E}[R_N(z)] = (z - \frac{\alpha\Phi}{1-\delta_z})^{-1}(I + \varepsilon_N(z))$$

where we noticed that $\Im\delta(z) \leq 0$ implies that since Φ is non-negative $z - \frac{\alpha\Phi}{1+\delta_z}$ is invertible. This implies that

$$\mathbb{E}[G_N(z)] = \frac{1}{P}\ \mathrm{Tr}(\mathbb{E}[R_N(z)]) = \frac{1}{P}\ \mathrm{Tr}(z - \frac{\alpha\Phi}{1-\delta_z})^{-1} + \varepsilon_N^1(z)$$

for ε_N^1 going to zero. We finally can find that δ_z satisfies the approximate implicit equation:

$$\delta_z = \frac{1}{P}\ \mathrm{Tr}(\mathbb{E}[R_N(z)]\Phi) = \frac{1}{P}\ \mathrm{Tr}(\Phi(z - \alpha\frac{\Phi}{1-\delta_z})^{-1}(I + \varepsilon_N(z)))$$

which shows that δ_z approximately satisfies the equation

$$\begin{aligned}
\delta_z &= \frac{1}{P}\ \mathrm{Tr}(\Phi(z - \alpha\frac{\Phi}{1-\delta_z})^{-1}) + \varepsilon_N^2(z) \\
&= -\frac{1-\delta_z}{\alpha} + \frac{1-\delta_z}{\alpha} z \frac{1}{P}\ \mathrm{Tr}((z - \alpha\frac{\Phi}{1-\delta_z})^{-1}) + \varepsilon_N^2(z) \\
&= -\frac{1-\delta_z}{\alpha} + \frac{1-\delta_z}{\alpha} z\mathbb{E}[G_N(z)] + \varepsilon_N^2(z)
\end{aligned}$$

Solving this linear equation yields

$$\delta_z = \frac{-1 + z\mathbb{E}[G_N(z)]}{-1 + \alpha + z\mathbb{E}[G_N(z)]} + \varepsilon_N^3(z)\,.$$

Here we notice that $z\mathbb{E}[G_N(z)] - 1 = \mathbb{E}[\frac{1}{P}\mathrm{Tr}(\mathbb{V}_N(z - \mathbb{V}_N)^{-1}]$ belongs to $\mathbb{C}^-$ as $\mathbb{V}_N$ is non-negative and hence $-1 + \alpha + z\mathbb{E}[G_N(z)]$ is invertible. Therefore we arrive to the equation on $\mathbb{E}[G_N(z)]$:

$$\mathbb{E}[G_N(z)] = \frac{1}{P}\ \mathrm{Tr}(z - \Phi(-\alpha + 1 + z\mathbb{E}[G_N(z)]))^{-1} + \varepsilon_N^4(z)$$

with $\varepsilon^4(z)$ small if $\frac{1}{N}\ \mathrm{Tr}(\varepsilon_N(z)^2)$ goes to zero, and a fortiori if its operator norm goes to zero. Hence, we see

$$\mathbb{E}[G_N(z)] = \Psi(z, \mathbb{E}[G_N(z)])) + \varepsilon_N^4(z)$$

as the solution of an approximate fixed point equation. Because $\Im(z\mathbb{E}[G_N(z)]) \leq 0$ and Φ is non-negative and bounded we find

$$|\Psi(z,x) - \Psi(z,x')| \leq \frac{\|\Phi\|_\infty}{|\Im z|^2}|x - x'|$$

Insuring by the implicit function theorem that for $|\Im z|^2 > \|\Phi\|_\infty$ there exists a unique solution G_S going to zero at infinity to the equation $G_S(z) = \Psi(z, G_S(z))$. Moreover

$$|\mathbb{E}[G_N(z)] - G_S(z)| \leq (1 - \frac{\|\Phi\|_\infty}{|\Im z|^2})^{-1} \varepsilon_N^4(z)$$

goes to zero. This proves the convergence for large $\Im z$ as wished. But then G_S, as the limit of uniformly bounded analytic functions, is uniquely defined $\mathbb{C}^+$ by the above equation.

We finally show that ε_N is small in norm. By our assumption, we know that $|\delta_z^i - \delta_z$ is small uniformly with large probability.

For the first term in $\varepsilon_N(z)$ we apply backward (74) to find by symmetry over the indices that if $\delta_i(z) = \frac{1}{N}\langle v_i^T, (z - \mathbb{V}_N^i)^{-1} v_i^T \rangle$

$$\varepsilon_N^1(z) = \frac{1}{1+\delta_z} \frac{1}{N} \sum_{i=1}^N \mathbb{E}[(\delta_i(z) - \delta_z) R_N(z) v_i v_i^T] = \mathbb{E}[R_N(z) \frac{1}{N} \sum (\delta_z^i - \delta_z) v_i v_i^T]$$

By Hanson-Wright inequality

$$\max_i |\delta_z^i - \delta_z| \leq N^{-1/2+\epsilon}$$

with probability greater than $1 - e^{-N^{2\epsilon}}$. Since the norm of $\sum(\delta_z^i - \delta_z) v_i v_i^T$ is bounded by $\max |(\delta_z^i - \delta_z)| \sum_{i=1}^N \|v_i\|_2^2$ which is at most of order N^2 we can neglect what happens on the set $\max_i |\delta_z^i - \delta_z| \geq K N^{-1/2+\epsilon}$. On the complementary we claim that $K_N(z) = R_N(z) \frac{1}{N} \sum(\delta_z^i - \delta_z) v_i v_i^T$ is bounded by $N^{-1/2+\epsilon}$. Indeed for any unit vector u, using Cauchy-Schwartz inequality we find

$$\langle u, K_N(z) K_N(z)^T u \rangle = \frac{1}{N^2} \sum_{i,j} (\delta_z^i - \delta_z)(\delta_z^j - \delta_z) \langle u, v_i \rangle \langle u, v_j \rangle \langle v_i, R_N(z) R_N(z)^T v_j \rangle$$

$$\begin{aligned}
&\leq N^{-1+2\epsilon} \frac{1}{N^2} \left(\sum_{i,j} |\langle u, v_i \rangle|^2 |\langle u, v_j \rangle|^2 \sum_{i,j} \langle v_i, R_N(z) R_N(z)^T v_i \rangle \langle v_j, R_N(z) R_N(z)^T v_j \rangle \right)^2 \\
&\leq N^{-1+2\epsilon} \frac{1}{N^2} \sum_i \langle v_i, R_N(z) R_N(z)^T v_i \rangle \\
&= N^{-1+2\epsilon} \frac{1}{N} \operatorname{Tr}(R_N(z) R_N(z)^T \mathbb{V}_N) \leq N^{-1/2+\epsilon} \frac{1}{\Im z}(1 + \frac{|z|}{\Im z})
\end{aligned}$$

We bound ε_2 by using (74) again

$$\mathbb{E}[\frac{1}{N} \sum_{i=1}^N ((z - \mathbb{V}_N^i)^{-1} - (z - \mathbb{V}_N)^{-1}) \Phi] = \mathbb{E}[\frac{1}{N^2} \sum_{i=1}^N ((z - \mathbb{V}_N^i)^{-1} v_i v_i^T (z - \mathbb{V}_N)^{-1}) \Phi]$$

$$= \mathbb{E}[\frac{1}{N^2} \sum_{i=1}^N (1 + \delta_z^i)(z - \mathbb{V}_N)^{-1} v_i v_i^T (z - \mathbb{V}_N)^{-1} \Phi]$$

Using that $(z-\mathbb{V}_N)^{-1})\Phi$ is bounded as well as $\frac{1}{N}\sum_{i=1}^N(1+\delta_z^i)((z-\mathbb{V}_N)^{-1}v_iv_i^T$ completes the proof.

Notice that in fact uniqueness of the fixed point equation can be proven on the whole half plan $\mathbb{C}^+$ by a more careful analysis.

1.4 Applications to Deep learning

Take a model with to an input data $x\in\mathbb{R}^N$ one associates the vector $\sigma(W.x) = (\sigma(\langle W_i, x\rangle))_{1\le i\le M}$ for some weights $(W_{ij})_{\substack{1\le i\le M\\ 1\le j\le N}}$. The output is $\beta^T\sigma(W.x)$ Being given inputs $x_1,\dots,x_T$ and outputs $y_1,\dots y_T$ one wishes to find a matrix β minimizing

$$\frac{1}{T}\sum_{i=1}^T\|\beta^T\sigma(Wx_i)-y_i\|_2^2+\gamma\|\beta\|_F^2$$

where $\|\beta\|_F^2=\sum\beta_{ij}^2$ It is not hard to see that the minimizer is given if $\Sigma_{ij}=\sigma(W_i.x_j)$ by

$$\beta^T=\frac{1}{T}\Sigma(\frac{1}{T}\Sigma^T\Sigma+\gamma I)^{-1}Y^T$$

The mean-square error is then given by

$$E=\frac{1}{T}\|Y^T-\Sigma^T\beta\|_F^2=\frac{\gamma^2}{T}\mathrm{Tr}(Y^TY(\gamma+\frac{1}{T}\Sigma^T\Sigma)^{-2})=-\gamma^2\partial_\gamma\mathrm{Tr}(Y^TY(\gamma+\frac{1}{T}\Sigma^T\Sigma)^{-1}$$

We recognize in the right hand side the resolvant of $\Sigma\Sigma^T$. Imagine for a minute that σ is linear and W_{ij} are i.i.d. Then we are exactly in the previous configuration with P with column vectors $(x_k)_{1\le k\le T}$. Because we saw in the previous section that the resolvent is close to $(\frac{n}{T}\frac{\Phi}{1-\delta}+\gamma)^{-1}$ we can compute the asymptotics of E. The main point of [?] is to show that this result continues to hold even when σ is non-linear under appropriate assumptions that we now state.

- The W_{ij} are Gaussian, or image of Gaussians by Lipschitz functions (entries satisfying log-Sobolev inequalities should be fine)
- σ is Lipschitz
- $M/N, T/N$ are finite
- The norm of the matrix with entries $x_k(i)$ is bounded , the outputs $y_k(i)$ are uniformly bounded

Then the same result holds as follows

Theorem 1.20 (Couillet-Liao-Louart [?]). *Let*

$$\Phi=\mathbb{E}[\sigma(w^TX)^T\sigma(w^TX)]$$

. *Then*

$$\|\mathbb{E}([\frac{n}{T}\frac{\Phi}{1+\delta_\gamma}+\gamma I_T)^{-1}-Q\|\le n^{-1/2+\epsilon}$$

where

$$Q = (\frac{1}{T}\Sigma^T\Sigma + \gamma I_T)^{-1}$$

if δ_γ is, for $\gamma > 0$, the unique solution in $\mathbb{C}^+$ of

$$\delta_\gamma = \frac{1}{T}Tr\Phi(\frac{n}{T}\frac{\Phi}{1+\delta_\gamma} + \gamma I_T)^{-1}$$

A remarkable point of this theorem is that it does depend on σ only through Φ so that there is a kind of universality in this result. The proof is exactly the same as before since when we expend along line we still have independence as soon as we have the appropriate concentration of measure results. In fact, assuming that the W_{ij} are Gaussian, image of Gaussians by a Lipschitz function φ or satisfy Lo-Sobolev inequality, we have for all A with norm bounded by one

$$\mathbb{P}\left(\left|\langle\frac{1}{T}\sigma(w.X)A\sigma(wX)^T - \frac{1}{T}\mathrm{Tr}(\Phi A)\right| \geq t\right) \leq Ce^{-\frac{cT}{\|X\|^2\|\varphi\|_L^2\|\sigma\|_L^2}\min(t^2/t_0^2,t)}$$

This is due to the general concentration statement that if f is Lipschitz

$$\|P\left(|f(w) - \mathbb{E}(f(w))| \geq t\|f\|_L\right) \leq e^{-ct^2}$$

Here $\|f\|_L = \sup_{x\neq y}\|f(x)-f(y)\|_2/\|x-y\|_2$. It remains to show that $w \to \frac{1}{T}\sigma(w.X)A\sigma(wX)^T$ is Lipschitz. But

$$|\frac{1}{T}\sigma(w.X)A\sigma(wX)^T - \frac{1}{T}\sigma(w'.X)A\sigma(w'X)^T| \leq \frac{2}{\sqrt{T}}\frac{1}{\sqrt{T}}\|\sigma(w,X)\|_2\|X\|_2\|w-w'\|_2$$

is Lipschitz only on $\frac{1}{\sqrt{T}}\|\sigma(w.X)\|_2 \leq K$. But $\frac{1}{\sqrt{T}}\|\sigma(wX)\|_2$ is Lipschitz with constant $\|\sigma\|_L\|X\|$. Hence we get

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{T}}\|\sigma(wX)\|_2 - \mathbb{E}[\frac{1}{\sqrt{T}}\|\sigma(wX)\|_2]\right| \geq \delta\right) \leq e^{-c\delta^2 T}$$

We can then use concentration conditionnally to a set (see e.g [16, section 5.4]) to conclude. Note that similar results could be derived if the W_{ij} would be bounded by using Talagrand's concentration of measures.

2 Extreme eigenvalues and signal detection

Notice that the largest eigenvalue of a random matrix has to be asymptotically larger than its the right border of the limiting spectral measure. However it may a priori be much larger. We expect in general that matrices with enough randomness will have eigenvalues which stick to the bulk.

2.1 Eigenvalues stick to the bulk

This was first proven for Wigner's matrices by Kömlos and Furedi [12] by using the methods of moments. We sketch this approach briefly as later we will prove more precise large deviation results:

Theorem 2.1. *Take $(X_{ij})_{1\le i\le j\le N}$ be real independent, centered, with variance one. Assume that $\mathbb{E}[X_{ij}^{2k}]\le k^{ck}$ for all integer numbers. Let $\mathbf{X}_N$ be the $N\times N$ matrix with entries $X_{ij}/\sqrt{N}$ and denote by λ_{max} its largest eigenvalue. Then*

$$\lim_{N\to\infty}\lambda_{max}=2 \qquad a.s.$$

Proof. We have already seen that the empirical measure $\mu_{\mathbf{X}_N}$ converges almost surely towards the semi-circle law. Since this is not possible on $\lambda_{max}\le 2-\epsilon$ we deduce that

$$\liminf_{N\to\infty}\lambda_{max}\ge 2 \qquad a.s$$

Proving the upper bound is more tricky and the method of moments gives the easiest approach when possible. In fact we notice that

$$\mathbb{P}(\lambda_{max}\ge 2+\epsilon)\le\frac{(2+\epsilon)^{2k}}{\mathbb{E}}[\lambda_{max}^{2k}]\le\frac{1}{(2+\epsilon)^{2k}}\mathbb{E}[\ \mathrm{Tr}(\mathbf{X}_N^{2k})]$$

Assume we can show that

$$\mathbb{E}[\ \mathrm{Tr}(\mathbf{X}_N^{2k})]\le N4^k(1+(k^c/N)^k) \tag{77}$$

for some finite c. Then we deduce that

$$\mathbb{P}(\lambda_{max}\ge 2+\epsilon)\le N(\frac{2}{2+\epsilon})^{2k}(1+(k^c/N)^k)$$

If $k^c << N$ and $k\gg\ln N$ we see that the above right hand side goes to zero faster than $1/N^2$ hence yielding the announced convergence by Borel-Cantelli lemma. The main point is hence to show (77). We can again write

$$\mathbb{E}[\ \mathrm{Tr}(\mathbf{X}_N^{2k})]=\frac{1}{N^k}\sum_{i_1\ldots i_{2k}}\mathbb{E}[X_{i_1i_2}\cdots X_{i_{2k}i_1}]$$

We will show that the main order is given by indices whose corresponding graph is a rooted planar tree for k growing to infinity sufficiently slowly. This was shown in [22] for $k\le<<\sqrt{N}$ (which is enough for us), see also [2, 1]. We provide for completeness the proof of [22] which holds when the distribution of the entries is symmetric $P(-x\in .)=P(x\in .)$. We take the normalization $E[x^2]=1$. To prove the moment estimates we shall again expand the moments and count contributing paths, in particular estimate more precisely contributions from paths that are not trees. Yet, the central point of the proof is to show that these paths give a neglectable contribution. We follow the presentation of [22].

1. *Moments expansion.* As usual, we write

$$\mathbb{E}[\mathrm{Tr}\left((\mathbf{X}_N)^{2s}\right)] \quad = \quad \frac{1}{N^s}\sum_{i_0,\cdots,i_{2s-1}=1}^{N}\mathbb{E}[X_{i_0i_1}\cdots X_{i_{2s-1},i_0}]. \qquad (78)$$

   We let E denote the set of edges of the graph, i.e the undirected collection of couples $\{(i_p, i_{p+1}), p = 0, \cdots, 2s-1\}$. Because we assumed the law of the X_{ij}'s symmetric, only indices such that each edge in E appears an even number of times will contribute. We call a *closed path* the sequence $P : i_0 \to i_1 \to \cdots \to i_{2s-1} \to i_0$. An *even path* is a closed path where each edge appears with even multiplicity; they are the only contributing paths.

2. *Descriptions of paths.* We will say that the ℓ^{th} step $i_{\ell-1} \to i_\ell$ of a path P is *marked* if during the first ℓ steps of P, the edge $\{i_{\ell-1}, i_\ell\}$ appears an odd number of times (note here that the ℓ^{th} step is counted, and so a step is marked iff the edge $\{i_{\ell-1}, i_\ell\}$ appears an even number of times in the previous step, in particular if it does not appear). The step is *unmarked* otherwise. For even paths, the number of marked and unmarked edges is equal to s. The complete set of vertices $\mathcal{V}$ is the collection $\{1, \cdots, N\}$ of all possible values of the points $(i_k, 0 \le k \le 2s-1)$. We say that a vertex $i \in \mathcal{V}$ belongs to the subset $\mathcal{N}_k = \mathcal{N}_k(P)$ if the number of times we arrive at i via marked edges equals k. Note that no vertex of the path except i_0 can belong to $\mathcal{N}_0$. Moreover, $\mathcal{N}_p = 0$ for $p > s$ (since there are at most s edges). Note that if we let $n_k = \sharp\mathcal{N}_k$, since $(\mathcal{N}_0, \cdots, \mathcal{N}_s)$ is a partition of $\mathcal{V}$, $\sum_{k=0}^{s} n_k = N$. Moreover, $(\mathcal{N}_0, \cdots, \mathcal{N}_s)$ also induces a partition of the edges and hence

$$\sum_{k=0}^{s} k n_k = s.$$

   We say that P is of type $(n_0, n_1, \cdots, n_s)$ if $n_k = \sharp\mathcal{N}_k = \sharp\mathcal{N}_k(P)$ for all $k \in \{0, \cdots, s\}$. We finally say that a path is a *simple even path* if $i_0 \in \mathcal{N}_0$ and P is of type $(N-s, s, 0, \cdots, 0)$. Observe that in a simple even path, each edge appears only twice (since there are at most s different edges in P and here exactly s since there are s different vertices in $\mathcal{N}_1$). Also, we see that the graph corresponding to P has exactly s vertices in $\mathcal{N}_1$ plus $i_0 \in \mathcal{N}_0$ and so exactly $s+1$ vertices. Hence, the skeleton $(V, \tilde{E})$ of the graph drawn by P satisfies the relation $|V| = |\tilde{E}| + 1$ and hence is a tree. The strategy of the proof will be to show that simple even paths dominate the expectation when $s = o(\sqrt{N})$.

3. *Contribution of simple even paths.* Considering (78), we see that for simple even paths, $\mathbb{E}[X_{i_0i_1} \cdots X_{i_{2s-1}i_0}] = 1$. Moreover, given a simple even path, we have N possible choices for i_0, $N-1$ for the first new vertex encountered when following P , $N-2$ for the second new vertex encountered etc. Since we have $C_s = (2s)!/s!(s+1)!$ simple even paths (see Property **??**), we get

the contribution

$$C_1^N = \frac{1}{N^s} N(N-1)\cdots(N-s)\frac{(2s)!}{s!(s+1)!} = \frac{2^{2s}N}{\sqrt{\pi s^3}}(1+o_1(s,N))$$

where we have used Stirling's formula and found

$$o_1(s,N) = -\frac{1}{N}\sum_{k=1}^{s} k + \frac{1}{s} \approx \frac{s^2}{2N} + \frac{1}{s}.$$

In the case where $i_0 \notin \mathcal{N}_0$ but $n_1 = s, n_2 = 0 \cdots, n_s = 0$, we must have $i_0 \in \mathcal{N}_1$. This means that we have one cycle and one different vertex less in the graph of an even path. Note that if we split the vertex i_0 into two vertices, the new vertex being attached to the marked edge, then the old i_0 belongs to $\mathcal{N}_0$ and the new vertex to $\mathcal{N}_1$ and we are back to the previous setting.

There are s possibilities for the position of the marked edge incoming in i_0, but we are loosing $N-s$ possibilities to choose a different vertex. Hence, the contribution to this term is bounded by

$$C_2^N \leq \frac{s}{N-s} E[x^4] C_1^N$$

where the last term comes from the possibility that one edge attached to i_0 now has multiplicity 4.

4. *Contribution of paths that are not simple.* If a path is not as in the previous paragraph, there must be an $n_k \geq 1$ for $k \geq 2$. Let us count the number of these paths.

   Given $n_0, n_1, \cdots, n_s$, we have $\frac{N!}{n_0!n_1!\cdots n_s!}$ ways to choose the values of the vertices. Then, among the n_0 vertices in $\mathcal{N}_0$, we have at most n_0 ways to choose the vertex corresponding to i_0 (if $i_0 \in \mathcal{N}_0$).

   Being given the values of the vertices, a path is uniquely described if we know the order of appearance of the vertices at the marked steps, the times when the marked steps occur and the choice of end points of the unmarked steps. The moments of time when marked steps occur can be coded by a Dick path by adding $+1$ when the step is marked and -1 otherwise. Hence, there are $C_s = (2s)!/s!(s+1)!$ choices for the times of marked steps. Once we are given this path, we have s marked steps. The marked steps are partitioned into s sets corresponding to the $\mathcal{N}_k$, $1 \leq k \leq s$, with cardinality $n_k k$ each. Hence, we have $\frac{s!}{\prod_{k=1}^{s}(n_k k)!}$ possibilities to assign the sets into which the end points of the marked steps are. Finally, we have $(n_k k)!/(k!)^{n_k}$ ways to partition the set $\mathcal{N}_k$ into k copies of the same point of $\mathcal{N}_k$. So far, we have prescribed uniquely the marked steps and the set to which they belong.

   To prescribe the unmarked steps, we still have an indeterminate. In fact, let us follow the Dick path of the marked steps till the first decreasing part corresponding to unmarked steps. Let i_ℓ be the vertex assigned to the last step. Then, if i_ℓ appeared only once in the past path (in the edge $(i_{\ell-1}, i_\ell)$), we have no choice and the next vertex in the path has to be $i_{\ell-1}$. This is the case in particular if $i_\ell \in \mathcal{N}_1$. If now $i_\ell \in \mathcal{N}_k$ for $k \geq 2$, the undirected step (i_p, i_ℓ) for some i_p may have ocured already at most $2k$ times (since it could occur either as a step (i_p, i_ℓ) or a step (i_ℓ, i_p), the later happening also less than k times since it requires that a marked step arrived at i_ℓ before). We have thus at most $2k$ choices now for the next vertex; one of the i_p among the at most $2k$ vertices such that the step (i_p, i_ℓ) or (i_ℓ, i_p) were present in the past path. Once this choice has been made, we can proceed by induction since this choice comes with the prescription of the set $\mathcal{N}_l$ in which the vertex i_p belongs. Hence, since we have kn_k vertices in each set, we see that we have at most $\prod_{k=2}^s (2k)^{kn_k}$ choices for the end points of the unmarked steps.

Coming back to (78) we see that if the path is of type $(n_0, \cdots, n_s)$, entries appear at most n_k times with multiplicity $2k$ for $1 \leq k \leq s$. Thus Hölder's inequality gives

$$\mathbb{E}[X_{i_0 i_1} \cdots X_{i_{2s-1} i_0}] \leq \prod_{k=1}^{s} \mathbb{E}[x^{2k}]^{n_k} \leq \prod_{k=2}^{s} (ck)^{kn_k}$$

where we used that $\mathbb{E}[x^2] = 1$. This shows that the contribution of these paths can be bounded as follows.

$$\begin{aligned}
E_{n_0,\cdots,n_s} &= \sum_{i_0,\cdots i_{2s-1}: P \text{ of type}(n_0,\cdots,n_s)} \mathbb{E}[X_{i_0 i_1} \cdots X_{i_{2s-1} i_0}] \\
&\leq \frac{1}{N^s} n_0 \frac{N!}{n_0! n_1! \cdots n_s!} \frac{(2s)!}{s!(s+1)!} \frac{s!}{\prod_{k=1}^s (n_k k)!} \\
&\qquad \prod_{k=1}^{s} \frac{(n_k k)!}{(k!)^{n_k}} \prod_{k=2}^{s} (2k)^{kn_k} \prod_{k=2}^{s} (ck)^{kn_k} \\
&\leq n_0 \frac{N(N-1)\cdots(n_0+1)}{N^s} \frac{(2s)!}{s!(s+1)!} \frac{1}{n_1! \cdots n_s!} \\
&\qquad \frac{s!}{\prod_{k=1}^s (ke^{-1})^{n_k k}} \prod_{k=1}^{s} (2ck^2)^{kn_k} \\
&\leq N N^{N-n_0-s} \frac{(2s)!}{s!(s+1)!} \frac{s!}{n_1! \cdots n_s!} \prod_{k=2}^{s} (2cek)^{kn_k}
\end{aligned}$$

where we have used that $(k!)^{n_k} \geq (ke^{-1})^{kn_k}$. Since $s = \sum_{k=1}^s kn_k$ and $N = \sum_k n_k$, we have $N - n_0 - s = \sum_{k=2}^s (1-k) n_k$. Using $s! \leq (s)^s$, we obtain the bound

$$E_{n_0,\cdots,n_s} \leq N \frac{(2s)!}{s!(s+1)!} \prod_{k=2}^{s} \frac{1}{n_k!} (N^{1-k} (2ceks)^k)^{n_k}.$$

We next sum over all $n_i \geq 0$ so that at least one $n_i \geq 1$ for $i \in \{2, \cdots, s\}$. This gives, with $\gamma_k := N^{1-k}(2ceks)^k$,

$$
\begin{aligned}
\sum_{n_0,\cdots,n_s : \max_{j\geq 2} n_j \geq 1} E_{n_0,\cdots,n_s} &\leq N\frac{(2s)!}{s!(s+1)!}\sum_{k=2}^{s}(e^{\gamma_k}-1)\prod_{\ell\neq k} e^{\gamma_\ell} \\
&\leq N\frac{(2s)!}{s!(s+1)!} e^{\sum_{\ell\geq 2}\gamma_\ell}(\sum_{\ell\geq 2}\gamma_\ell)
\end{aligned}
$$

where we used that $e^x - 1 \leq xe^x$ for all $x \geq 0$. Note that in the range of s where $s^2 \leq N^{1-\epsilon}$, if we choose K big enough so that $K\epsilon \geq 1$,

$$
\begin{aligned}
\sum_{\ell}\gamma_\ell &= \sum_{2\leq \ell\leq s} N^{1-\ell}(2ce\ell s)^\ell \\
&\leq NK(2cesKN^{-1})^2 + N\sum_{K+1\leq\ell\leq s}(2ces^2N^{-1})^\ell \\
&\leq \text{constant}(N^{-1}K^2s^2 + N(2ceN^{-\epsilon})^{K+1}) \leq \text{constant}N^{-\epsilon}
\end{aligned}
$$

goes to zero as N goes to infinity. Thus, we conclude that

$$\sum_{n_0,\cdots,n_s} E_{n_0,\cdots,n_s} \leq CC_1^N N^{-\epsilon}.$$

Hence, in the regime s^2/N going to zero, the contribution of the indices $\{i_0, \cdots, i_{2s-1}\}$ associated with a path of type $(n_0, \cdots, n_s)$ with some $n_k \geq 1$ for some $k \geq 2$ is neglectable compared to the contribution of simple even paths. $\diamond$

$\diamond$

It is also possible to prove that eigenvalues stick to the bulk by using Stieltjes transform. We sketch the result of [4]:

Theorem 2.2. *Assume X_{ij} are iid centered with covariance one and finite fourth moment. Let $\mathbf{X}_N = (X_{ij})_{\substack{1\leq i\leq M\\ 1\leq j\leq N}}$ with M/N going to α Let T_n be a $M \times M$ non-negative Hermitian matrix with uniformly bounded norm and spectral distribution converging μ_T. Set*

$$\mathbb{V}_N = \frac{1}{N}T_M^{1/2}\mathbf{X}_N\mathbf{X}_N^T T_M^{1/2}$$

and let $\mu_{\alpha,T}$ be the distribution function with Stieltjes transform solution of

$$m = \int \frac{1}{t(1-\alpha-\alpha zm) - z} d\mu_T(t)$$

Let $[a,b]$ in the complement of the support of $[a,b]$. Then

$$\lim_{N\to\infty} \mathbb{P}(\textit{An eigenvalue of } \mathbb{V}_N \textit{belongs to [a,b]}) = 0$$

Proof. The idea of the proof is to improve upon the previous arguments (which we will not do here) to show that

$$\lim_{N\to\infty} \sup_{x\in[a,b],y\in[v_{N,1}]} Ny|\frac{1}{N}\operatorname{Tr}((x+iy-\mathbb{V}_N)^{-1})-G_T(x+iy)|=0$$

with $v_N = N^{-1/68}$. This is indeed sufficient for our purpose since Green's theorem shows (see e.g [2, Proof of Lemma 5.5.5]) that

$$\int \Psi(t,0)d\mu(t) = \Re\left(\int_0^\infty dy \int_{-\infty}^{+\infty} dx \int\int \frac{\bar\partial\Psi(x,y)}{t-x-iy}\mu(dt)\right)$$

provided Ψ is smooth, compactly supported, so that $\Im\Psi(x,0)=0$ and such that

$$\bar\partial\Psi(x,y) = \frac{1}{\pi}(\partial_x + i\partial_y)\Psi(x,y)$$

is such that $|\bar\partial\Psi(x,y)|/|y|$ is uniformly bounded, and $\bar\partial\Psi(x,0) = 0$. When $\psi(x) = \Psi(x,0)$ is C^ℓ we can even take Ψ such that $|\bar\partial\Psi(x,y)|/|y|^\ell$ is bounded by taking $\Psi(x,y) = \sum_{i=0}^{\ell} i^\ell y^\ell \psi^{(i)}(x)\chi(y)$ with χ smooth, compactly supported on $[0,c_0]$, equal to one in the neighborhood of the origin. Thus, if $[a,b]$ is outside the support of μ_S and ψ is smooth supported on $[a,b]$

$$N\hat\mu_{\mathbb{V}_N}(\psi) = \Re\left(\int_0^\infty dy \int_{-\infty}^{+\infty} dx \int \bar\partial\Psi(x,y)N(G_N(x+iy)-G_T(x+iy))\mu(dt)\right)$$

Cutting the integral on $y\in[0,v_N]$ and $[v_N,c_0]$ and using that on $[0,v_N]$ $|N(G_N(x+iy)-G_T(x+iy))|\le 2N/y$, we get

$$|N\hat\mu_{\mathbb{V}_N}(\psi)| \le CN\int_0^{v_N} y^{\ell-1}dy + o(1)$$

which goes to zero as soon as $v_N^\ell N$ goes to zero. Since we can take ψ non negative and equal to one on $[a+\epsilon, b-\epsilon]$ we deduce that there are no eigenvalues almost surely in $[a+\epsilon, b-\epsilon]$ as N goes to infinity. $\diamond$

2.2 BBP transition

As we have just seen, showing that eigenvalues stick to the bulk is a quite delicate property to prove. In fact it can easily be destroyed by small perturbation, for instance by adding a rank one perturbation. This question was studied first by Baik-Ben Arous-Péché who studied the eigenvalues of

$$\mathbb{V}_N^\gamma = \frac{1}{N}(1+\gamma uu^T)^{1/2}\mathbf{X}_N\mathbf{X}_N^T(1+\gamma uu^T)^{1/2}$$

for a unit vector u and standard Gaussian entries $(X_{ij})_{\substack{1\le i\le M\\ 1\le j\le N}}$. They proved

Theorem 2.3 (BBP transition)**.** *[5] Assume M/N goes to $\alpha < 1$.*

- *If $\gamma < \sqrt{\alpha}$, the largest eigenvalue of $\mathbb{V}_N^\gamma$ sticks to the bulk and converges to $(1+\sqrt{\alpha})^2$ almost surely.*
- *If $\gamma > \sqrt{\alpha}$, the largest eigenvalue of $\mathbb{V}_N^\gamma$ converges towards $(1+\gamma)(1+\frac{\alpha}{\gamma})$.*

.The most general way to prove the theorem follows the following ideas. We first look for an eigenvalue λ of $\mathbb{V}_N^\gamma$ outside the spectrum of $\mathbb{V}_N^0$. It must satisfy

$$\begin{aligned} 0 &= \det(\lambda - \mathbb{V}_N^\gamma) \\ &= \det(\lambda - \frac{1}{N}\mathbf{X}_N\mathbf{X}_N^T(I+\gamma uu^T)) \\ &= \det(\lambda - \frac{1}{N}\mathbf{X}_N\mathbf{X}_N^T)\det(1-\gamma(\lambda - \frac{1}{N}\mathbf{X}_N\mathbf{X}_N^T)^{-1}\mathbf{X}_N\mathbf{X}_N^T uu^T)) \end{aligned}$$

Since by assumption the first term is non vanishing, we deduce that the second must vanish which is equivalent to

$$1 = \gamma\langle u, (\lambda - \frac{1}{N}\mathbf{X}_N\mathbf{X}_N^T)^{-1}\frac{1}{N}\mathbf{X}_N\mathbf{X}_N^T u\rangle$$

For Gaussian entries u can be chosen uniformly distributed on the sphere since $\mathbb{V}_N^0$ is invariant under orthogonal conjugation. Then again concentration of measure shows that

$$\langle u, (\lambda - \mathbf{X}_N^T\mathbf{X}_N)^{-1}\mathbf{X}_N^T\mathbf{X}_N u\rangle \simeq \frac{1}{N}\,\mathrm{Tr}((\lambda - \mathbf{X}_N^T\mathbf{X}_N)^{-1}\mathbf{X}_N^T\mathbf{X}_N)$$

This is true in greater generality and if u is deterministic and the entries of $\mathbf{X}_N$ have enough moments, we expect to have such an isotropic law. . From there, we can use the convergence of the spectral measure of $\mathbb{V}_N^0$ to see that λ must satisfy

$$\frac{1}{\gamma} = \int \frac{x}{\lambda - x} d\mu_\alpha = -1 + \lambda G(\lambda)$$

The result follows by inverting $\lambda \to \lambda G(\lambda)$.

By a similar argument we find that if $\mathbf{X}_N$ is a Wigner matrix.

Theorem 2.4. *Let $\mathbf{X}_N^\theta = \mathbf{X}_N + \theta uu^T$ with a fixed unit vector u. Then the largest eigenvalue converges towards 2 if $\theta \le 1$ and towards $\theta + \theta^{-1}$ if $\theta \ge 1$.*

The proof is left to the reader. In fact, not only a strong enough finite perturbation can be detected by outliers but also we can measure the scalar product of the eigenvectors of these outliers with the original data [11]. We state the result below for Wigner's matrices but similar results hold for Wishart

Theorem 2.5. *[11] Assume that $\gamma > 1$ and let v be the unit eigenvector of $\mathbf{X}_N^\gamma$ corresponding to the largest eigemvalue λ_1 converging towards $\lambda^* = \gamma + \gamma^{-1}$. Let G be the Stieljes transform of the semi-circle law. Then*

$$\lim_{N\to\infty}\langle u, v\rangle^2 = -\frac{G(\lambda^*)^2}{G'(\lambda)}$$

The proof easily follows from noticing that

$$\lambda_1 v = \mathbf{X}_N^\gamma v \Leftrightarrow (\lambda_1 - \mathbf{X}_N)v = \gamma u \langle u, v\rangle$$

SInce v is a unit vector, this is equivalent to

$$v = \frac{(\lambda_1 - \mathbf{X}_N)^{-1}u}{\|(\lambda_1 - \mathbf{X}_N)^{-1}v\|_2}$$

and hence

$$\langle v, u\rangle^2 = \frac{\langle u, (\lambda_1 - \mathbf{X}_N)^{-1}u\rangle^2}{\langle u, (\lambda_1 - \mathbf{X}_N)^{-2}u\rangle}$$

Again, one can check that since λ_1 lies away from the eigenvalues of $\mathbf{X}_N$, we have

$$\langle u, (\lambda_1 - \mathbf{X}_N)^{-k}u\rangle \simeq \frac{1}{N}\operatorname{Tr}(\lambda_1 - \mathbf{X}_N)^{-k}$$

for $k = 1, 2$. The result follows.

2.3 Applications to signal detection

BBP transition phenomenon was shown to be universal in the sense that it does not depend much on the distribution of the entries. It shows that data can be retrieved if it is in some sense stronger than the noise

- If $\gamma > \sqrt{\alpha}$, signal can be detected and recovered by Principal Component Analysis.
- If $\gamma < \sqrt{\alpha}$ and the noise is Gaussian,
  - No test based on the eigenvalues can reliably detect the signal (Montanari, Reichman, Zeitouni '17)
  - No test can reliably detect the signal (El Alaoui, Krzakala, Jordan '18)

2.4 Generalizations

- The BBP transition generalizes to $\mathbf{X}_N \mathbf{X}_N^T + \sum \gamma_i u_i u_i^T$
- It generalizes to Wigner matrices.
- The fluctuations around the outliers away from the bulk are Gaussian if $\mathbf{X}_N$ has Gaussian entries but non universality can occur [20].
- It generalizes for much more structured random matrices (see Belinschi,Bercovici, Capitaine, Capitaine-Donati Martin
- It generalizes to non-Hermitian matrices (Tao, Bordenave-Capitaine)

3 Large deviations

We have seen in section 1.2.2 that the eigenvalues and the empirical measure of the eigenvalues concentrate. This section deals with the question of estimating exactly the probability of deviations and we will concentrate on Wigner matrices even though our statements generalize to Wishart matrices. We first outline some motivations.

3.1 Motivations from signal detection and landscape of random functions

- Large deviations estimates can be interesting to test signal detection (see the article by P. Bianchi, M. Debbah, M. Maida, J.Najim 11') Assume that the signa
$$\Sigma = (E(Y_i Y_j))_{i,j} = \sigma^2 Id + uu^T$$
with $u = 0$ under (H_0), and $u \neq 0$ under (H_1).
The Generalized Likelyhood Ratio is given by
$$L_N = \frac{\sup_{\sigma,u} P_1(Y|u,\sigma)}{\sup_\sigma P_0(Y|u,\sigma)}, P_t(Y|u,\sigma) \propto \frac{e^{-N\mathrm{Tr}(Y(\sigma^2+tuu^T)^{-1}Y^T)}}{\det(\sigma^2+tuu^T)^{N/2}}$$
and one rejects (H_0) iff $L_N > \zeta_N$ with $P_0(L_N > \zeta_N) \leq \alpha$. $1 - P_1(L_N < \zeta_N)$ is the power of the test. One finds that
$$L_N = \phi(\frac{\lambda_K}{\frac{1}{N}\,\mathrm{Tr}(YY^T)})$$
so that the test is based on precise estimates of :
$$P_0(\frac{\lambda_K}{\frac{1}{N}\,\mathrm{Tr}(YY^T)} > \zeta_N), P_1(\frac{\lambda_K}{\frac{1}{N}\,\mathrm{Tr}(YY^T)} < \zeta_N)$$
which are large deviations estimates.
- Landscape of a spiked tensor (Ben Arous,Song Mei, Andrea Montanari, Mihai Nica [**?**] Let
$$\mathbf{Y}_N = \lambda u^{\otimes k} + \mathbf{X}_N$$
with a Wigner matrix $\mathbf{X}_N$. We would like to recover u from the observation of $\mathbf{Y}_N$. We can expect (and was shown) that if $\lambda > \lambda_c(k)$ it is possible to weak;y recover the signal (i.e find a vector whose scalar product with the signal does not go to zero) by the eigenvector corresponding to the outlier of $\mathbf{Y}_N$ whereas it is impossible if λ is too small. However, the maximum likelihood estimator requires to maximize $f(\sigma) = \langle \mathbf{Y}_N, \sigma^{\otimes k}\rangle$ over unit vectors σ. This is NP-hard for $k \geq 3$. To explore the landscape of f, [**?**] analyze its volume of critical points:
$$\mathrm{Crt}_{n,*}(M,E) = \sum_{\sigma:\mathrm{grad} f=0} 1_{\langle sigma,u\rangle \in M} 1_{f(\sigma)\in E}\,.$$

They get a formula for $\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Crt}_{n,*}(M,E)$ thanks to Kac-Rice formula which implies

$$\mathbb{E}[\mathrm{Crt}_{n,*}(M,E)] = \int_M V_n(m)\mathbb{E}[|\det(H)|1_{H\leq 0}1_{f\in E}\phi_\sigma(0)(1-m^2)^{-1/2}dm$$

where $V_n(m)$ is the area of the (n-1) dimensional sphere with radius $(1-m^2)^{1/2}$, ϕ_σ is the density of g at 0 where $g = k\lambda m^{k-1}\sqrt{1-m^2}e_1 + \sqrt{k/2n}\tilde{g}_{n-1}$ and manly

$$H = k(k-1)\lambda m^{k-2}(1-m^2)e_1e_1^T + \sqrt{(k(k-1)(n-1)/2n}W - k(\lambda m_+^k \frac{1}{\sqrt{2n}}Z)I$$

with a GOE matrix W. Using the large deviation principle of M. Maida [18] for perturbed GOE allows to estimate the RHS.

3.2 Gaussian ensembles

In this part, we first consider the case where the entries of the matrix $\mathbf{X}^N$ are Gaussian, that is the GOE. We denote P_N the law of $\mathbf{X}^{N,\beta}$.

The main advantage of the Gaussian ensembles is that the law of the eigenvalues of these matrices is explicit and rather simple. Namely, we now discuss the following lemma.

Lemma 3.1. *Let $\mathbf{X} \in \mathbb{H}_N$ be random with law P_N. The joint distribution of the eigenvalues $\lambda_1(X) \leq \cdots \leq \lambda_N(X)$, has density proportional to*

$$1_{x_1\leq\cdots\leq x_N} \prod_{1\leq i<j\leq N} |x_i - x_j| \prod_{i=1}^{N} e^{-x_i^2/4}. \tag{79}$$

The proof is standard, see e.g [2]. Let us however emphasize the ideas. It is simply to write the decomposition $X = ODO^*$, with the eigenvalues matrix D that is diagonal and with real entries, and with the eigenvectors matrix O (that is orthogonal). Suppose this map was a bijection (which it is not, at least at the matrices X that do not possess all distinct eigenvalues) and that one can parametrize the eigenvectors by $N(N-1)/2$ parameters in a smooth way (which one cannot in general). Then, it is easy to deduce from the formula $X = ODO^*$ that the Jacobian of this change of variables depends polynomially on the entries of D and is of degree $N(N-1)/2$ in these variables. Since the bijection must break down when $D_{ii} = D_{jj}$ for some $i \neq j$, the Jacobian must vanish on that set. When $\beta = 1$, this imposes that the polynomial must be proportional to $\prod_{1\leq i<j\leq N}(x_i - x_j)$. Further degree and symmetry considerations allow to generalize this to $\beta = 2$. We refer the reader to [2] for a full proof, that shows that the set of matrices for which the above manipulations are not permitted has Lebesgue measure zero.

We have a similar result for Wishart matrices which gives a joint distribution of $\lambda_1 \leq \cdots \leq \lambda_N$ if $N \leq M$ given by

$$P_W^N(d\ell_1, \cdots, d\ell_N) = (Z_W^N)^{-1}|\mathbf{D}(\ell)| \prod_{i=1}^N \lambda_i^{M-N} e^{-N\sum_{i=1}^N \ell_i} \prod_{i=1}^N d\ell_i$$

Here, $\mathbf{D}(\ell) = \prod_{1\le i<j\le N}(\ell_i - \ell_j)$. Hence, we will consider the more general type of distribution

$$P_{V,\beta}^N(d\ell_1, \cdots, d\ell_N) = (Z_{V,\beta}^N)^{-1}|\mathbf{D}(\ell)|^\beta e^{-N\sum_{i=1}^N V(\ell_i)} \prod_{i=1}^N d\ell_i, \tag{80}$$

for a continuous function $V : \mathbb{R} \to \mathbb{R}$ such that

$$\liminf_{|x|\to\infty} \frac{V(x)}{\beta \log|x|} > 1 \tag{81}$$

and a positive real number β.

3.3 Large deviations for the law of the spectral measure of Gaussian Wigner's matrices

In this section, we consider the law of N random variables $(\ell_1, \cdots, \ell_N)$ with law

When $V(x) = 4^{-1}\beta x^2$, we have seen in Lemma 3.1 that $P_{4^{-1}\beta x^2,\beta}^N$ is the law of the eigenvalues of a $N\times N$ GOE matrix when $\beta = 1$, and of a GUE matrix when $\beta = 2$. The case $\beta = 4$ corresponds to another matrix ensemble, namely the GSE. In view of this remarks and other applications discussed in Chapter **??**, we consider in this section the slightly more general model with a potential V. We emphasize however that the distribution (80) precludes us from considering random matrices with independent non Gaussian entries.

We have proved already at the beginning of these notes that the empirical measure

$$L_N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i^N}$$

converges almost surely towards the semi-circular law. Moreover, we studied its fluctuations around its mean, both by central limit theorem and by concentration inequalities. Such results did not depend much on the Gaussian nature of the entries.

We address here a different type of question. Namely, we study the probability that L_N takes a very unlikely value. This was already considered in our discussion of concentration inequalities, c.f. Section **??**, where the emphasis was put on obtaining upper bounds on the probability of deviation. In contrast, the purpose of the analysis here is to exhibit a precise estimate on these probabilities, or at least on their logarithmic asymptotics. The appropriate tool for handling such questions is large deviations theory, and we bring in Appendix **??** a concise introduction to that theory and related definitions and references.

Endow $\mathcal{P}(\mathbb{R})$ with the usual weak topology. Our goal is to estimate the probability $P_{V,\beta}^N(L_N \in A)$, for measurable sets $A \subset \mathcal{P}(\mathbb{R})$. Of particular interest is the case where A does not contain the limiting distribution of L_N.

Define the *non-commutative entropy* $\Sigma : \mathcal{P}(\mathbb{R}) \to [-\infty, \infty]$, as

$$\Sigma(\mu) = \int\int \log|x-y| d\mu(x) d\mu(y) \,. \tag{82}$$

Set next

$$I^V_\beta(\mu) = \begin{cases} \int V(x)d\mu(x) - \frac{\beta}{2}\Sigma(\mu) - c^V_\beta \,, & \text{if } \int V(x)d\mu(x) < \infty \\ \infty, & \text{otherwise}\,, \end{cases} \tag{83}$$

with $c^V_\beta = \inf_{\nu\in\mathbf{m}1}\{\int V(x)d\nu(x) - \frac{\beta}{2}\Sigma(\nu)\}$.

Theorem 3.2. *Let $L_N = N^{-1}\sum_{i=1}^N \delta_{\lambda_i^N}$ where the random variables $\{\lambda_i^N\}_{i=1}^N$ are distributed according to the law $P^N_{V,\beta}$, see (80). Then, the family of random measures L_N satisfies, in $\mathcal{P}(\mathbb{R})$ equipped with the weak topology, a full large deviation principle with good rate function I^V_β in the scale N^2. That is, I^V_β : $\mathcal{P}(\mathbb{R}) \to [0,\infty]$ possesses compact level sets $\{\nu : I^V_\beta(\nu) \le M\}$ for all $M \in \mathbb{R}_+$, and*

$$\begin{gathered}\text{For any open set } O \subset \mathbf{m}1\,, \\ \liminf_{N\to\infty} \frac{1}{N^2}\log P^N_{\beta,V}(L_N \in O) \ge -\inf_O I^V_\beta\,, \end{gathered} \tag{84}$$

and

$$\begin{gathered}\text{For any closed set } F \subset \mathbf{m}1\,, \\ \limsup_{N\to\infty} \frac{1}{N^2}\log P^N_{\beta,V}(L_N \in F) \le -\inf_F I^V_\beta\,. \end{gathered} \tag{85}$$

The proof of Theorem 3.2 relies on the properties of the function I^V_β collected in Lemma 3.3 below. Define the *logarithmic capacity* of a measurable set $A \subset \mathbb{R}$ as

$$\gamma(A) := \exp\left\{-\inf_{\nu\in M_1(A)}\int\int \log\frac{1}{|x-y|} d\nu(x)d\nu(y)\right\} .$$

Lemma 3.3.

a. I^V_β *is well defined on* $\mathbf{m}1$ *and takes its values in* $[0, +\infty]$.

b. $I^V_\beta(\mu)$ *is infinite as soon as* μ *satisfies one of the following conditions*

b.1 $\int V(x)d\mu(x) = +\infty$.

b.2 There exists a set $A \subset \mathbb{R}$ *of positive* μ *mass but null logarithmic capacity, i.e. a set* A *such that* $\mu(A) > 0$ *but* $\gamma(A) = 0$.

c. I^V_β *is a good rate function.*

d. I^V_β *is a strictly convex function on* $\mathbf{m}1$.

e. I^V_β *achieves its minimum value at a unique probability measure* σ^V_β *on* $\mathbb{R}$ *characterized by*

$$V(x) - \beta\int \log|y-x| d\sigma^V_\beta(y) = \inf_{\nu\in\mathbf{m}1}\left(\int V d\nu - \beta\Sigma(\nu)\right), \ \sigma^V_\beta \ a.s., \tag{86}$$

and, for all x except possibly on a set with null logarithmic capacity,

$$V(x) - \beta \int \log |y - x| d\sigma_\beta^V(y) \geq \inf_{\nu \in \mathbf{m1}} \left(\int V d\nu - \beta \Sigma(\nu) \right). \tag{87}$$

As an immediate corollary of Theorem 3.2 and of part e. of Lemma 3.3 we have the following.

Corollary 3.4 (Second proof of Wigner's theorem)**.** *Under $P^N_{V,\beta}$, L_N converges almost surely towards σ^V_β.*

Proof of Lemma 3.3 If $I^V_\beta(\mu) < \infty$, since V is bounded below by assumption (81), $\Sigma(\mu) > -\infty$ and therefore also $\int V d\mu < \infty$. This proves that $I^V_\beta(\mu)$ is well defined (and by definition non negative), yielding point a.

Set

$$f(x, y) = \frac{1}{2} V(x) + \frac{1}{2} V(y) - \frac{\beta}{2} \log |x - y|. \tag{88}$$

Note that $f(x, y)$ goes to $+\infty$ when x, y do by (81). Indeed, $\log |x - y| \leq \log(|x| + 1) + \log(|y| + 1)$ implies

$$f(x, y) \geq \frac{1}{2}(V(x) - \beta \log(|x| + 1)) + \frac{1}{2}(V(y) - \beta \log(|y| + 1)) \tag{89}$$

as well as when x, y approach the diagonal $\{x = y\}$; for all $L > 0$, there exist constants $K(L)$ (going to infinity with L) such that

$$\begin{aligned}
&\{(x, y) : \ f(x, y) \geq K(L)\} \subset B_L \, , \\
&B_L := \{(x, y) : \ |x - y| < L^{-1}\} \cup \{(x, y) : \ |x| > L\} \cup \{(x, y) : \ |y| > L\}.
\end{aligned} \tag{90}$$

Since f is continuous on the compact set B^c_L, we conclude that f is bounded below, and denote $b_f > -\infty$ a lower bound. Therefore, since for any measurable subset A of $\mathbb{R}$,

$$\begin{aligned}
I^V_\beta(\mu) &= \int \int (f(x, y) - b_f) d\mu(x) d\mu(y) + b_f - c^V_\beta \\
&\geq \int_A \int_A (f(x, y) - b_f) d\mu(x) d\mu(y) + b_f - c^V_\beta \\
&\geq \frac{\beta}{2} \int_A \int_A \log |x - y|^{-1} d\mu(x) d\mu(y) + \inf_{x \in \mathbb{R}} V(x) \mu(A)^2 - |b_f| - c^V_\beta \\
&\geq -\frac{\beta}{2} \mu(A)^2 \log(\gamma(A)) - |b_f| - c^V_\beta + \inf_{x \in \mathbb{R}} V(x) \mu(A)^2
\end{aligned}$$

one concludes that if $I^V_\beta(\mu) < \infty$, and A is a measurable set with $\mu(A) > 0$, then $\gamma(A) > 0$. This completes the proof of point b.

We now show that I_V^β is a good rate function, and first that its level sets $\{I_V^\beta \le M\}$ are closed, that is that I_V^β is lower semi-continuous. Indeed, by the monotone convergence theorem,

$$\begin{aligned} I_\beta^V(\mu) &= \int\int f(x,y)d\mu(x)d\mu(y) - c_\beta^V \\ &= \sup_{M\ge 0}\int\int (f(x,y)\wedge M)d\mu(x)d\mu(y) - c_\beta^V \end{aligned}$$

But $f^M = f\wedge M$ is bounded continuous and so for $M<\infty$,

$$I_\beta^{V,M}(\mu) = \int\int (f(x,y)\wedge M)d\mu(x)d\mu(y)$$

is bounded continuous on $\mathbf{m}1$. As a supremum of the continuous functions $I_\beta^{V,M}$, I_β^V is lower semi-continuous. Hence, by Theorem **??**, to prove that $\{I_\beta^V \le L\}$ is compact, it is enough to show that $\{I_\beta^V \le L\}$ is included in a compact subset of $\mathbf{m}1$ of the form

$$K_\epsilon = \cap_{B\in\mathbb{N}}\{\mu\in\mathbf{m}1 : \mu([-B,B]^c)\le \epsilon(B)\}$$

with a sequence $\epsilon(B)$ going to zero as B goes to infinity.

Arguing as in (90), there exist constants $K'(L)$ going to infinity as L goes to infinity, such that

$$\{(x,y):\ |x|>L, |y|>L\}\subset\{(x,y):\ f(x,y)\ge K'(L)\}\,. \tag{91}$$

Hence, for any $L>0$ large,

$$\begin{aligned} \mu(|x|>L)^2 &= \mu\otimes\mu(|x|>L,|y|>L) \\ &\le \mu\otimes\mu(f(x,y)\ge K'(L)) \\ &\le \frac{1}{K'(L)-b_f}\int\int(f(x,y)-b_f)d\mu(x)d\mu(y) \\ &= \frac{1}{K'(L)-b_f}(I_\beta^V(\mu)+c_\beta^V-b_f) \end{aligned}$$

Hence, with $\epsilon(B) = [\sqrt{(M+c_\beta^V-b_f)_+}/\sqrt{(K'(B)-b_f)_+}]\wedge 1$ going to zero when B goes to infinity, one has that $\{I_\beta^V\le M\}\subset K_\epsilon$. This completes the proof of point c.

Since I_β^V is a good rate function, it achieves its minimal value. Let σ_β^V be a minimizer. Then, for any signed measure $\bar\nu(dx) = \phi(x)\sigma_\beta^V(dx)+\psi(x)dx$ with two bounded measurable compactly supported functions (ϕ,ψ) such that $\psi\ge 0$ and $\bar\nu(\mathbb{R})=0$, for $\epsilon>0$ small enough, $\sigma_\beta^V+\epsilon\bar\nu$ is a probability measure so that

$$I_\beta^V(\sigma_\beta^V+\epsilon\bar\nu)\ge I_\beta^V(\sigma_\beta^V)$$

which implies

$$\int \left(V(x) - \beta \int \log|x-y| d\sigma_\beta^V(y) \right) d\bar{\nu}(x) \geq 0.$$

Taking $\psi = 0$, we deduce by symmetry that there is a constant C_β^V such that

$$V(x) - \beta \int \log|x-y| d\sigma_\beta^V(y) = C_\beta^V\,, \quad \sigma_\beta^V \text{ a.s.}, \tag{92}$$

which implies that σ_β^V is compactly supported (as $V(x) - \beta \int \log|x-y| d\sigma_\beta^V(y)$ goes to infinity when x does). Taking $\phi(x) = -\int \psi(y)dy$, we then find that

$$V(x) - \beta \int \log|x-y| d\sigma_\beta^V(y) \geq C_\beta^V \tag{93}$$

Lebesgue almost surely, and then everywhere outside of the support of σ_β^V by continuity. By (92) and (93) we deduce that

$$C_\beta^V = \inf_{\nu \in \mathbf{m1}} \{ \int (V(x) - \beta \int \log|x-y| d\sigma_\beta^V(y)) d\nu(x) \}\,.$$

This completes the proof of (86) and (87). The claimed uniqueness of σ_β^V, and hence the completion of the proof of part e., will then follow from the strict convexity claim (point d. of the lemma), which we turn to next.

Note first that we can rewrite I_β^V as

$$I_\beta^V(\mu) = -\frac{\beta}{2}\Sigma(\mu - \sigma_\beta^V) + \int \left(V - \beta \int \log|x-y| d\sigma_\beta^V(y) - C_\beta^V \right) d\mu(x).$$

The fact that I_β^V is strictly convex comes from the observation that Σ is strictly concave, as can be checked from the formula

$$\log|x-y| = \int_0^\infty \frac{1}{2t} \left(\exp\{-\frac{1}{2t}\} - \exp\{-\frac{|x-y|^2}{2t}\} \right) dt \tag{94}$$

which entails that for any $\mu \in \mathbf{m1}$,

$$\Sigma(\mu - \sigma_\beta^V) = -\int_0^\infty \frac{1}{2t} \left(\int \int \exp\{-\frac{|x-y|^2}{2t}\} d(\mu - \sigma_\beta^V)(x) d(\mu - \sigma_\beta^V)(y) \right) dt.$$

Indeed, one may apply Fubini's theorem when μ_1, μ_2 are supported in $[-\frac{1}{2}, \frac{1}{2}]$ since then $\mu_1 \otimes \mu_2(\exp\{-\frac{1}{2t}\} - \exp\{-\frac{|x-y|^2}{2t}\} \leq 0) = 1$. One then deduces the claim for any compactly supported probability measures by scaling and finally for all probability measures by approximations. The fact that for all $t \geq 0$,

$$\int \int \exp\{-\frac{|x-y|^2}{2t}\} d(\mu - \sigma_\beta^V)(x) d(\mu - \sigma_\beta^V)(y)$$

$$= \sqrt{\frac{t}{2\pi}} \int_{-\infty}^{+\infty} \left| \int \exp\{i\lambda x\} d(\mu - \sigma_\beta^V)(x) \right|^2 \exp\{-\frac{t\lambda^2}{2}\} d\lambda$$

therefore entails that Σ is concave since $\mu \rightarrow \left| \int \exp\{i\lambda x\} d(\mu - \sigma_\beta^V)(x) \right|^2$ is convex for all $\ell \in \mathbb{R}$. Strict convexity comes from the fact by the Cauchy-Schwarz inequality, $\Sigma(\alpha\mu + (1-\alpha)\nu) = \alpha\Sigma(\mu) + (1-\alpha)\Sigma(\nu)$ if and only if $\Sigma(\nu - \mu) = 0$ which implies that all the Fourier transforms of $\nu - \mu$ are null, and hence $\mu = \nu$. This completes the proof of part d and hence of the lemma. $\diamond$

Proof of Theorem 3.2: To begin, let us remark that with f as in (88),

$$P_{V,\beta}^N(d\ell_1, \cdots, d\ell_N) = (Z_N^{\beta,V})^{-1} e^{-N^2 \int_{x \neq y} f(x,y) dL_N(x) dL_N(y)} \prod_{i=1}^N e^{-V(\ell_i)} d\ell_i.$$

Hence, if $\mu \rightarrow \int_{x \neq y} f(x,y) d\mu(x) d\mu(y)$ was a bounded continuous function, the proof would follow from a standard Laplace method (see Theorem **??** in the appendix). The main point will be therefore to overcome the singularity of this function, with the most delicate part being overcoming the singularity of the logarithm.

Following Appendix **??** (see Corollary **??** and Definition **??**), a full large deviation principle can be proved by proving that exponential tightness holds, as well as estimating the probability of small balls. We follow these steps below.

- *Exponential tightness* Observe that by Jensen's inequality,

$$\begin{aligned} \log Z_N^{\beta,V} \quad &\geq \quad N \log \int e^{-V(x)} dx \\ &-N^2 \int \left(\int_{x \neq y} f(x,y) dL_N(x) dL_N(y) \right) \prod_{i=1}^N \frac{e^{-V(\ell_i)} d\ell_i}{\int e^{-V(x)} dx} \geq -CN^2 \end{aligned}$$

with some finite constant C. Moreover, by (89) and (81), there exist constants $a > 0$ and $c > -\infty$ so that

$$f(x,y) \geq a|V(x)| + a|V(y)| + c$$

from which one concludes that for all $M \geq 0$,

$$P_{V,\beta}^N \left(\int |V(x)| dL_N \geq M \right) \leq e^{-2aN^2 M + (C-c)N^2} \left(\int e^{-V(x)} dx \right)^N.$$

Since V goes to infinity at infinity, $K_M = \{\mu \in \mathbf{m}1 : \int |V| d\mu \leq M\}$ is a compact set for all $M < \infty$, so that we have proved that the law of L_N under $P_{V,\beta}^N$ is exponentially tight.

-*Large deviation upper bound* d will denote the Dudley metric, see (**??**). We prove here that for any $\mu \in \mathbf{m}1$, if we set $\bar{P}_{V,\beta}^N = Z_N^{\beta,V} P_{V,\beta}^N$

$$\lim_{\epsilon \to 0} \limsup_{N \to \infty} \frac{1}{N^2} \log \bar{P}_{V,\beta}^N (d(L_N, \mu) \leq \epsilon) \leq -\int f(x,y) d\mu(x) d\mu(y). \tag{95}$$

For any $M \geq 0$, the following bound holds

$$\bar{P}^N_{V,\beta}\left(d(L_N,\mu)\leq\epsilon\right)$$

$$\leq \int_{d(L_N,\mu)\leq\epsilon} e^{-N^2\int_{x\neq y} f(x,y)\wedge M dL_N(x)dL_N(y)} \prod_{i=1}^N e^{-V(\ell_i)}d\ell_i.$$

Since under the product Lebesgue measure, the ℓ_i's are almost surely distinct, it holds that $L_N\otimes L_N(x=y)=N^{-1}$, $\bar{P}^N_{V,\beta}$ almost surely. Thus, we deduce for all $M\geq 0$, with $f_M(x,y)=f(x,y)\wedge M$,

$$\int f_M(x,y)dL_N(x)dL_N(y)=\int_{x\neq y} f_M(x,y)dL_N(x)dL_N(y)+MN^{-1},$$

and so

$$\bar{P}^N_{V,\beta}\left(d(L_N,\mu)\leq\epsilon\right)$$

$$\leq e^{MN}\int_{d(L_N,\mu)\leq\epsilon} e^{-N^2\int f_M(x,y)dL_N(x)dL_N(y)} \prod_{i=1}^N e^{-V(\ell_i)}d\ell_i.$$

Since $I^{V,M}_\beta(\nu)=\int f_M(x,y)d\nu(x)d\nu(y)$ is bounded continuous, we deduce that

$$\lim_{\epsilon\to 0}\limsup_{N\to\infty}\frac{1}{N^2}\log\bar{P}^N_{V,\beta}\left(d(L_N,\mu)\leq\epsilon\right)\leq -I^{V,M}_\beta(\mu).$$

We finally let M go to infinity and conclude by the monotone convergence theorem. Note that the same argument shows that

$$\limsup_{N\to\infty}\frac{1}{N^2}\log Z^{\beta,V}_N\leq -\inf_{\mu\in\mathbf{m1}}\int f(x,y)d\mu(x)d\mu(y). \tag{96}$$

- *Large deviation lower bound.* We prove here that for any $\mu\in\mathbf{m1}$

$$\lim_{\epsilon\to 0}\liminf_{N\to\infty}\frac{1}{N^2}\log\bar{P}^N_{V,\beta}\left(d(L_N,\mu)\leq\epsilon\right)\geq -\int f(x,y)d\mu(x)d\mu(y). \tag{97}$$

Note that we can assume without loss of generality that $I^V_\beta(\mu)<\infty$, since otherwise the bound in trivial, and so in particular, we may and will assume that μ has no atoms. We can also assume that μ is compactly supported since if we consider $\mu_M=\mu([-M,M])^{-1}1_{|x|\leq M}d\mu(x)$, clearly μ_M converges towards μ and by the monotone convergence theorem, one checks that, since f is bounded below,

$$\lim_{M\uparrow\infty}\int f(x,y)d\mu_M(x)d\mu_M(y)=\int f(x,y)d\mu(x)d\mu(y)$$

which insures that it is enough to prove the lower bound for $(\mu_M, M\in\mathbb{R}, I^V_\beta(\mu)<\infty)$, and so for compactly supported probability measures with finite entropy.

The idea is to localize the eigenvalues $(\lambda_i)_{1\le i\le N}$ in small sets and to take advantage of the fast speed N^2 of the large deviations to neglect the small volume of these sets. To do so, we first remark that for any $\nu\in\mathbf{m}1$ with no atoms if we set

$$\begin{aligned}
x^{1,N} &= \inf\left\{x|\ \nu\left(]-\infty,x]\right)\ge\frac{1}{N+1}\right\}\\
x^{i+1,N} &= \inf\left\{x\ge x^{i,N}|\ \nu\left(]x^{i,N},x]\right)\ge\frac{1}{N+1}\right\}\qquad 1\le i\le N-1,
\end{aligned}$$

for any real number η, there exists an integer number $N(\eta)$ such that, for any N larger than $N(\eta)$,

$$d\left(\nu,\frac{1}{N}\sum_{i=1}^N\delta_{x^{i,N}}\right)<\eta.$$

In particular, for $N\ge N(\frac{\mathrm{d}}{2})$,

$$\left\{(\lambda_i)_{1\le i\le N}\ |\ |\lambda_i-x^{i,N}|<\frac{\mathrm{d}}{2}\ \forall i\in[1,N]\right\}\subset\{(\lambda_i)_{1\le i\le N}\ |\ d(L_N,\nu)<\delta\}$$

so that we have the lower bound

$$\begin{aligned}
&\bar{P}^N_{V,\beta}\left(d(L_N,\mu)\le\epsilon\right)\\
\ge\ & \int_{\cap_i\{|\lambda_i-x^{i,N}|<\frac{\mathrm{d}}{2}\ \}} e^{-N^2\int_{x\ne y}f(x,y)dL_N(x)dL_N(y)}\prod_{i=1}^N e^{-V(\lambda_i)}d\lambda_i\\
=\ & \int_{\cap_i\{|\lambda_i|<\frac{\mathrm{d}}{2}\}}\prod_{i<j}|x^{i,N}-x^{j,N}+\lambda_i-\lambda_j|^\beta e^{-N\sum_{i=1}^N V(x^{i,N}+\lambda_i)}\prod_{i=1}^N d\lambda_i\\
\ge\ & \left(\prod_{i+1<j}|x^{i,N}-x^{j,N}|^\beta\prod_i|x^{i,N}-x^{i+1,N}|^{\frac{\beta}{2}}e^{-N\sum_{i=1}^N V(x^{i,N})}\right)\\
&\times\left(\int_{\substack{\cap_i\{|\lambda_i|<\frac{\mathrm{d}}{2}\}\\ \lambda_i<\lambda_{i+1}}}\prod_i|\ell_i-\ell_{i+1}|^{\frac{\beta}{2}}e^{-N\sum_{i=1}^N[V(x^{i,N}+\lambda_i)-V(x^{i,N})]}\prod_{i=1}^N d\lambda_i\right)\\
=:\ & P_{N,1}\times P_{N,2}
\end{aligned}\tag{98}$$

where we used that $|x^{i,N}-x^{j,N}+\lambda_i-\lambda_j|\ge|x^{i,N}-x^{j,N}|\vee|\ell_i-\ell_j|$ when $\ell_i\ge\ell_j$ and $x^{i,N}\ge x^{j,N}$. To estimate $P_{N,2}$, note that since we assumed that μ is compactly supported, the $(x^{i,N},1\le i\le N)_{N\in\mathbb{N}}$ are uniformly bounded and so by continuity of V

$$\lim_{N\to\infty}\sup_{N\in\mathbb{N}}\sup_{1\le i\le N}\sup_{|x|\le\mathrm{d}}|V(x^{i,N}+x)-V(x^{i,N})|=0.$$

Moreover, writing $u_1 = \ell_1$, $u_{i+1} = \ell_{i+1} - \ell_i$,

$$\int_{\substack{|\lambda_i|<\frac{\mathrm{d}}{2}\ \forall i\\ \lambda_i<\lambda_{i-1}}} \prod_i |\ell_i - \ell_{i+1}|^{\frac{\beta}{2}} \prod_{i=1}^N d\ell_i \geq \int_{0<u_i<\frac{\mathrm{d}}{2N}} \prod_{i=2}^N u_i^{\frac{\beta}{2}} \prod_{i=1}^N du_i$$

$$\geq \left(\frac{\mathrm{d}}{(\beta+2)N}\right)^{N(\frac{\beta}{2}+1)}.$$

Therefore,

$$\lim_{\mathrm{d}\to 0} \liminf_{N\to\infty} \frac{1}{N^2} \log P_{N,2} \geq 0. \tag{99}$$

To handle the term $P_{N,1}$, the uniform boundness of the $x^{i,N}$'s and the convergence of their empirical measure towards μ imply that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N V(x^{i,N}) = \int V(x) d\mu(x). \tag{100}$$

Finally since $x \to \log(x)$ increases on $\mathbb{R}^+$, we notice that

$$\int_{x^{1,N}\leq x<y\leq x^{N,N}} \log(y-x) d\mu(x) d\mu(y)$$

$$\leq \sum_{1\leq i\leq j\leq N-1} \log(x^{j+1,N} - x^{i,N}) \int_{\substack{x\in[x^{i,N},x^{i+1,N}]\\ y\in[x^{j,N},x^{j+1,N}]}} 1_{x<y} d\mu(x) d\mu(y)$$

$$= \frac{1}{(N+1)^2} \sum_{i<j} \log|x^{i,N} - x^{j+1,N}| + \frac{1}{2(N+1)^2} \sum_{i=1}^{N-1} \log|x^{i+1,N} - x^{i,N}|.$$

Since $\log|x-y|$ is bounded when x, y are in the support of the compactly supported measure μ, the monotone convergence theorem implies that the left side in the last display converges towards $\int\int \log|x-y| d\mu(x) d\mu(x)$. Thus, with (100), we have proved

$$\liminf_{N\to\infty} \frac{1}{N^2} \log P_{N,1} \geq \int_{x<y} \log(y-x) d\mu(x) d\mu(y) - \int V(x) d\mu(x)$$

which concludes, with (98) and (99), the proof of (97).

-*Conclusion* By (97), for all $\mu \in \mathbf{m1}$,

$$\liminf_{N\to\infty} \frac{1}{N^2} \log Z_{\beta,V}^N \geq \lim_{\epsilon\to 0} \liminf_{N\to\infty} \frac{1}{N^2} \log \bar{P}_{V,\beta}^N (d(L_N, \mu) \leq \epsilon)$$

$$\geq -\int f(x,y) d\mu(x) d\mu(y)$$

and so optimizing with respect to $\mu \in \mathbf{m1}$ and with (96),

$$\lim_{N\to\infty} \frac{1}{N^2} \log Z_{\beta,V}^N = -\inf_{\mu\in\mathbf{m1}} \{\int f(x,y) d\mu(x) d\mu(y)\} = -c_\beta^V.$$

Thus, (97) and (95) imply the weak large deviation principle, i.e. that for all $\mu \in \mathbf{m1}$,

$$\lim_{\epsilon\to 0}\liminf_{N\to\infty}\frac{1}{N^2}\log P^N_{V,\beta}\left(d(L_N,\mu)\le\epsilon\right)$$
$$=\lim_{\epsilon\to 0}\limsup_{N\to\infty}\frac{1}{N^2}\log P^N_{V,\beta}\left(d(L_N,\mu)\le\epsilon\right)=-I^V_\beta(\mu).$$

This, together with the exponential tightness property proved above completes the proof of the full large deviation principle stated in Theorem 3.2. $\diamond$

Bibliographical Notes The proof of Theorem 3.2 is a slight generalization of the techniques introduced in [7] to more general potentials.

3.4 Large deviations of the maximum eigenvalue

We here restrict ourselve to the case where $V(x)=\beta x^2/4$ and denote in short P^N_β the law of the eigenvalues $(\lambda_i)_{1\le i\le N}$;

$$P^N_\beta(d\lambda_1,\cdots,d\lambda_N)=\frac{1}{Z^N_\beta}\prod_{1\le i<j\le N}|\lambda_i-\lambda_j|^\beta\prod_{1\le i\le N}e^{-\frac{\beta N\lambda_i^2}{4}}d\lambda_i$$

with

$$Z^N_\beta=\int\prod_{1\le i<j\le N}|\lambda_i-\lambda_j|^\beta\prod_{1\le i\le N}e^{-\frac{\beta N\lambda_i^2}{4}}d\lambda_i.$$

Selberg (c.f. [19, Theorem 4.1.1] or [2]) found the explicit formula for Z^N_β for any $\beta\ge 0$;

$$Z^N_\beta=(2\pi)^{\frac{N}{2}}\left(\frac{\beta N}{2}\right)^{-\beta N(N-1)/4-\frac{N}{2}}\prod_{j=1}^N\frac{\Gamma(\frac{j\beta}{2})}{\Gamma(\frac{\beta}{2})}. \tag{101}$$

The knowledge of Z^N_β up to the second order will be crucial below, reason why we will restrict ourselves to quadratic potentials in this section. We could however generalize our result by using the expansion of Z^N_β that we will derive in chapter **??**.

We prove here that (this theorem is taken from [6])

Theorem 3.5. *The maximal eigenvalue $\ell^*_N=\max_{i=1}^N\ell_i$ under P^N_β, with $\beta\ge 0$, satisfies the LDP in $\mathbb{R}$ with speed N and the GRF*

$$I^*(x)=\begin{cases}\beta\int_2^x\sqrt{(z/2)^2-1}\,dz, & x\ge 2\,,\\ +\infty, & \textit{otherwise}\,.\end{cases} \tag{102}$$

The next estimate is key to the proof of Theorem 3.5.

Lemma 3.6. *For every M large enough and all N,*

$$P^N_\beta\left(\max_{i=1}^N|\ell_i|\ge M\right)\le e^{-\beta NM^2/9}\,.$$

Observe that for any $|x| \geq M \geq 8$ and $\ell_i \in \mathbb{R}$,

$$|x - \ell_i| e^{-\frac{\ell_i^2}{8}} \leq (|x| + |\ell_i|) e^{-\frac{\ell_i^2}{8}} \leq 2|x| \leq e^{\frac{x^2}{8}} .$$

Therefore, integrating with respect to ℓ_1 yields, for $M \geq 8$,

$$\begin{aligned} P_\beta^N(|\ell_1| \geq M) &= \frac{Z_\beta^{N-1}}{Z_\beta^N} \int_{|x| \geq M} dx e^{-\frac{\beta x^2}{4} \frac{(N+1)}{2}} \int \prod_{i=2}^N \left(|x - \ell_i| e^{-\frac{\ell_i^2}{4} - \frac{x^2}{8}} \right)^\beta dP_\beta^{N-1}(\ell_j, j \geq 2) \\ &\leq e^{-\frac{\beta}{8} NM^2} \frac{Z_\beta^{N-1}}{Z_\beta^N} \int_{|x| \geq M} e^{-x^2/8} dx \int \prod_{i=2}^N (|x - \ell_i| e^{-\ell_i^2/4} e^{-x^2/8}) dP_\beta^{N-1}(\ell_j, j \geq 2) \\ &\leq e^{-\frac{\beta}{8} NM^2} \frac{Z_\beta^{N-1}}{Z_\beta^N} \int e^{-x^2/8} dx \end{aligned}$$

Further, following (101), we compute that

$$\lim_{N \to \infty} \frac{1}{N} \log \frac{Z_\beta^{N-1}}{Z_\beta^N} = -\frac{\beta}{4}. \tag{103}$$

It follows that for any $M \geq 8$, for N large enough,

$$P_\beta^N(\max_{i=1}^N |\ell_i| \geq M) \leq N P_\beta^N(|\ell_1| \geq M) \leq e^{-\frac{\beta}{9} NM^2} ,$$

and the lemma follows. ◇

Proof of Theorem 3.5. $I^*(x)$ is a good rate function since it is a continuous function (except at $x = 2$ where it is lower semi-continuous) and it goes to infinity at infinity. Moreover, with $I^*(x)$ continuous and strictly increasing on $[2, \infty[$ it suffices to show that for any $x < 2$,

$$\limsup_{N \to \infty} \frac{1}{N} \log P_\beta^N (\ell_N^* \leq x) = -\infty \tag{104}$$

whereas for any $x > 2$

$$\lim_{N \to \infty} \frac{1}{N} \log P_\beta^N (\ell_N^* \geq x) = -I^*(x) . \tag{105}$$

In fact, from these two estimates and since I^* increases on $[2, \infty[$, we find that for all $x < y$,

$$\lim_{N \to \infty} \frac{1}{N} \log P_\beta^N (\ell_N^* \in [x, y]) = - \inf_{z \in [x,y]} I^*(z),$$

the above right hand side being equal to $-\infty$ if $y \leq 2$, to zero if $x \leq 2 \leq y$ and to $I^*(x)$ if $x \geq 2$. By continuity of I^*, we also deduce that we have the same limits if we take (x, y) instead of $[x, y]$. Since $\mathcal{A} = \{[x, y], (x, y), x < y\}$ is a basis for the topology on $\mathbb{R}$, we conclude by Theorem **??**.

Starting with (104), fix $x < 2$ and $f \in \mathcal{C}_b(\mathbb{R})$ such that $f(y) = 0$ for all $y \le x$ whereas $\int f d\sigma > 0$. Note that $\{\ell_N^* \le x\} \subseteq \{\int f dL_{\mathbf{X}^N} = 0\}$, so (104) follows by applying the upper bound of the large deviation principle of Theorem 3.2 for the closed set $F = \{\mu : \int f d\mu = 0\}$, such that $\sigma \notin F$. Turning to the upper bound in (105), fix $M \ge x > 2$, noting that

$$P_\beta^N(\ell_N^* \ge x) = P_\beta^N(\max_{i=1}^N |\ell_i| > M) + P_\beta^N\left(\ell_N^* \ge x, \max_{i=1}^N |\ell_i| \le M\right) \tag{106}$$

By Lemma 3.6, the first term is exponentially negligible for all M large enough. To deal with the second term, let $P_N^{N-1}(\ell \in \cdot) = P_\beta^{N-1}((1-N^{-1})^{1/2}\ell \in \cdot)$, $L_{N-1} = (N-1)^{-1}\sum_{i=2}^N \mathrm{d}_{\ell_i}$ and

$$C_N := \frac{Z_\beta^{N-1}}{Z_\beta^N}(1-N^{-1})^{N(N-1)/4}.$$

Further, let $B(\sigma, \mathrm{d})$ denote an open ball in $\mathcal{P}(\mathbb{R})$ of radius $\mathrm{d} > 0$ and center σ, and $B_M(\sigma, \mathrm{d})$ its intersection with $\mathcal{P}([-M, M])$. Observe that for any $z \in [-M, M]$ and $\mu \in \mathcal{P}([-M, M])$,

$$\Phi(z, \mu) := \beta \int \log|z - y| d\mu(y) - \frac{\beta}{4}z^2 \le \beta \log(2M) .$$

Thus, for the second term in (106),

$$\begin{aligned} P_\beta^N\left(\ell_N^* \ge x, \max_{i=1}^N |\ell_i| \le M\right) &\le N C_N \int_x^M d\ell_1 \int_{[-M,M]^{N-1}} e^{(N-1)\Phi(\ell_1, L_{N-1})} dP_N^{N-1}(\ell_j, j \ge 2) \\ &\le N C_N \left(\int_x^M e^{(N-1)\sup_{\mu \in B_M(\sigma,\mathrm{d})}\Phi(z,\mu)} dz + (2M)^N P_N^{N-1}(L_{N-1} \notin B(\sigma, \mathrm{d}))\right) \end{aligned} \tag{107}$$

For any h of Lipschitz norm at most 1 and $N \ge 2$,

$$|(N-1)^{-1}\sum_{i=2}^N (h((1-N^{-1})^{1/2}\ell_i) - h(\ell_i))| \le 3N^{-1}\max_{i=2}^N |\ell_i| .$$

Thus, by Lemma 3.6, the spectral measures L_{N-1} under σ^{N-1} are exponentially equivalent in $\mathcal{P}(\mathbb{R})$ to the spectral measures L_{N-1} under P_N^{N-1}, so Theorem 3.2 applies also for the latter (c.f Definition **??** and Lemma **??**). In particular, the second term in (107) is exponentially negligible as $N \to \infty$ for any $\mathrm{d} > 0$ and $M < \infty$ (since it behaves like $e^{-c(\mathrm{d})N^2}$). Therefore,

$$\begin{aligned} \limsup_{N\to\infty} \frac{1}{N} \log P_\beta^N\left(\ell_N^* \ge x, \max_{i=1}^N |\ell_i| \le M\right) &\le \limsup_{N\to\infty} \frac{1}{N}\log C_N \\ &+ \lim_{\mathrm{d}\downarrow 0} \sup_{\substack{z\in[x,M] \\ \mu \in B_M(\sigma,\mathrm{d})}} \Phi(z, \mu) \end{aligned} \tag{108}$$

Note that $\Phi(z,\mu) = \inf_{\eta>0} \Phi_\eta(z,\mu)$ with $\Phi_\eta(z,\mu) := \beta \int \log(|z-y| \vee \eta) d\mu(y) - \frac{\beta}{4} z^2$ continuous on $[-M,M] \times \mathcal{P}([-M,M])$. Thus, $(z,\mu) \mapsto \Phi(z,\mu)$ is upper semi-continuous, which implies

$$\lim_{\mathrm{d}\downarrow 0} \sup_{\substack{z\in[x,M] \\ \mu\in B_M(\sigma,\mathrm{d})}} \Phi(z,\mu) = \sup_{z\in[x,M]} \Phi(z,\sigma) \tag{109}$$

With σ supported on $[-2,2]$, $D(z) := \frac{d}{dz}\Phi(z,\sigma)$ exists for $z \geq 2$. Moreover, $D(z) = -\beta\sqrt{(z/2)^2 - 1} \leq 0$. It is shown in [8, Lemma 2.7] that $\Phi(2,\sigma) = -\beta/2$. Hence, for $x > 2$,

$$\sup_{z\geq x} \Phi(z,\sigma) = \Phi(x,\sigma) = -\frac{1}{2} - I^*(x)\,. \tag{110}$$

By (103), we deduce that

$$\lim_{N\to\infty} N^{-1} \log C_N = \frac{\beta}{2}.$$

Combining this with (108)–(110) completes the proof of the upper bound for (105). To prove the complementary lower bound, fix $y > x > r > 2$ and $\mathrm{d} > 0$, noting that for all N,

$$\begin{aligned}
P_\beta^N\left(\ell_N^* \geq x\right) &\geq P_\beta^N\left(\ell_1 \in [x,y], \max_{i=2}^N |\ell_i| \leq r\right) \\
&= C_N \int_x^y e^{-\ell_1^2/4} d\ell_1 \int_{[-r,r]^{N-1}} e^{(N-1)\Phi(\ell_1, L_{N-1})} dP_N^{N-1}(\ell_j, j \geq 2) \\
&\geq kC_N \exp\left((N-1) \inf_{\substack{z\in[x,y] \\ \mu\in B_r(\sigma,\mathrm{d})}} \Phi(z,\mu)\right) P_N^{N-1}\left(L_{N-1} \in B_r(\sigma,\mathrm{d})\right)
\end{aligned} \tag{111}$$

with $k = k(x,y) > 0$. Recall that the large deviation principle with speed N^2 and good rate function $I(\cdot)$ applies for the measures L_{N-1} under P_N^{N-1}. It follows by this LDP's upper bound that $P_N^{N-1}(L_{N-1} \notin B(\sigma,\mathrm{d})) \to 0$, whereas by the symmetry of $P_\beta^N(\cdot)$ and the upper bound of (105),

$$P_N^{N-1}\left(L_{N-1} \notin \mathcal{P}([-r,r])\right) \leq 2P_\beta^{N-1}(\ell_N^* \geq r) \to 0$$

as $N \to \infty$. Consequently,

$$\liminf_{N\to\infty} \frac{1}{N} \log P_\beta^N\left(\ell_N^* \geq x\right) \geq \frac{1}{2} + \beta \inf_{\substack{z\in[x,y] \\ \mu\in B_r(\sigma,\mathrm{d})}} \Phi(z,\mu)$$

Observe that $(z,\mu) \mapsto \Phi(z,\mu)$ is continuous on $[x,y] \times \mathcal{P}([-r,r])$, for $y > x > r > 2$. Hence, considering $\mathrm{d} \downarrow 0$ followed by $y \downarrow x$ results with the required lower bound

$$\liminf_{N\to\infty} \frac{1}{N} \log P_\beta^N(\ell_N^* \geq x) \geq \frac{\beta}{2} + \beta\Phi(x,\sigma)\,.$$

Exercise 3.7 (suggested by B. Collins). *Generalize the proof to obtain the large deviation principle for the joint law of the k^{th} largest eigenvalues (k finite) with good rate function given by*

$$I^*(x_1, \cdots, x_k) = \sum_{l=1}^{k} I^*(x_k)$$

if $x_1 \geq x_2 \cdots \geq x_k \geq 2$ and $+\infty$ otherwise.

Bibliographical notes This proof is taken from [6].

3.5 LDP for the largest eigenvalue with outliers and complexity of spiked tensor models

Let us now consider the model of GOE matrices with noise

$$\mathbf{Y}_N = \mathbf{X}_N + \gamma u u^T$$

for a unit vector u and a GOE matrix $\mathbf{X}_N$. Our goal is to prove large deviations for the law of the largest eigenvalue of this spiked model, see [18].

The distribution of the eigenvalues of $\mathbf{Y}_N$ is given by

$$d\mathbb{P}_N^\gamma = e^{-\frac{N}{4}\,\mathrm{Tr}(\mathbf{Y}_N - \gamma u u^T)^2} d\mathbf{Y}_N / Z_N$$

Diagonalizing $\mathbf{Y}_N = O\mathrm{diag}(\lambda_1, \ldots, \lambda_N) O^T$ with O Haar on the orthogonal group, we find

$$\begin{aligned} dP_N^\gamma(\lambda_1, \ldots \lambda_N) &= \int e^{-\frac{N}{4}\,\mathrm{Tr}((\mathrm{diag}(\lambda) - \gamma O u (O u)^T)^2} dO \Delta(\lambda) d\lambda / Z \\ &= e^{-\frac{N}{4}\sum \lambda_i^2} I_N(\lambda, \gamma) \Delta(\lambda) d\lambda \qquad (112) \end{aligned}$$

where I_N is the spherical integral

$$I_N(\mathrm{diag}(\lambda), \gamma) = \int e^{\frac{N}{2}\langle O u, \mathrm{diag}(\lambda) O u\rangle} dO\,.$$

The asymptotics of

$$J_N(X, \theta) = \frac{1}{N} \ln I_N(X, \theta)$$

were studied in [14] where it was proved that

Theorem 3.8. *[14, Theorem 6]*

If $(E_N)_{N \in \mathbb{N}}$ is a sequence of $N \times N$ real symmetric matrices when $\beta = 1$ and complex Hermitian matrices when $\beta = 2$ such that :

- *The sequence of empirical measures $\hat{\mu}_{E_N}^N$ weakly converges to a compactly supported measure μ,*

- *There are two reals $\lambda_{\min}(E), \lambda_{\max}(E)$ such that $\lim_{N\to\infty} \lambda_{\min}(E_N) = \lambda_{\min}(E)$ and $\lim_{N\to\infty} \lambda_{\max}(E_N) = \lambda_{\max}(E)$,*

and $\theta \geq 0$, then :

$$\lim_{N\to\infty} J_N(E_N, \theta) = J(\mu, \theta, \lambda_{\max}(E))\,.$$

The limit J is defined as follows. For a compactly supported probability measure we define its Stieltjes transform G_μ by

$$G_\mu(z) := \int_{\mathbb{R}} \frac{1}{z-t} d\mu(t)\,.$$

We assume hereafter that μ is supported on a compact $[a, b]$. Then G_μ is a bijection from $\mathbb{R}\backslash[a, b]$ to $]G_\mu(a), G_\mu(b)[\backslash\{0\}$ where $G_\mu(a), G_\mu(b)$ are taken as the (possibly infinite) limits of $G_\mu(t)$ when $t \to a^-$ and $t \to b^+$. We denote by K_μ its inverse and let $R_\mu(z) := K_\mu(z) - 1/z$ be its R-transform as defined by Voiculescu in [23] (defined on a neighborhood of the origin, but also on $]G_\mu(a), G_\mu(b)[$). In the sequel, for any compactly supported probability measure μ, we denote by $r(\mu)$ the right edge of the support of μ. In order to define the rate function, we now introduce, for any $\theta \geq 0$, and $\lambda \geq r(\mu)$,

$$J(\mu, \theta, \lambda) := \theta v(\theta, \mu, \lambda) - \frac{\beta}{2} \int \log\left(1 + \frac{2}{\beta}\theta v(\theta, \mu, \lambda) - \frac{2}{\beta}\theta y\right) d\mu(y), \qquad (113)$$

with

$$v(\theta, \mu, \lambda) := \begin{cases} R_\mu(\frac{2}{\beta}\theta), & \text{if } 0 \leq \frac{2\theta}{\beta} \leq H_{\max}(\mu, \lambda) := \lim_{z\downarrow\lambda} \int \frac{1}{z-y} d\mu(y), \\ \lambda - \frac{\beta}{2\theta}, & \text{if } \frac{2\theta}{\beta} > H_{\max}(\mu, \lambda). \end{cases}$$

An easy proof of Theorem 3.8 follows by first considering taht E_N has only finitely $p + m_1 + m_2$ distincts eigenvalues μ_i with multiplicities N_i and N_i/N goes to zero except for $i \in [1, p]$. We then have the following formula :

$$\langle e, E_N e\rangle = \sum_{i=-m_1+1}^{p+m_2} \mu_i \gamma_i^N$$

where we have denoted $\gamma_j^N = \sum_{i\in I_j} |u_i|^2$ with $I_i = [N_1 + ... + N_{i-1}, N_1 + ... + N_i]$.

The vector γ^N follows a Dirichlet law of parameters $\frac{\beta}{2}(N_{1-m_1}, \ldots, N_p, \ldots N_{p+m_2})$, that is the distribution on $\Sigma = \{x \in [0, 1]^{m_1+m_2+p} : \sum_{i=1-m_1}^{m_2+p} x_i = 1\}$ given by

$$d\mathbb{P}_\alpha^N(\gamma) = \frac{1}{Z_\alpha^N} \prod_{i=1-m_1}^{p+m_2} \gamma_i^{\frac{\beta}{2}N_i} \prod_{j=1-m_1}^{m_2+p} d\gamma_j$$

We deduce the following large deviation principle

Theorem 3.9. *Assume that N_i/N converges towards α_i for all $i \in [1,p]$ and to zero otherwise. The law of $\gamma \in \Sigma$ satisfies a large deviation principle with scale N and good rate function I_α given for for $x \in \Sigma$ by*

$$I_\alpha(x_{1-m_1}, ..., x_{p+m_2}) = \frac{\beta}{2} \sum_{i=1}^{p} \alpha_i \log \frac{x_i}{\alpha_i} \,.$$

The proof is a direct consequence of Laplace's method. We deduce Proposition by Varadhan's lemma that

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\exp\left(\frac{\beta N}{2} \theta\langle e, E_N e\rangle\right)\right] = \frac{\beta}{2} \sup_{\substack{\gamma_i \geq 0 \\ \sum \gamma_i = 1}} \left\{ \theta \sum_{ß=-m_1+1}^{p+m_2} \lambda_i \gamma_i + \sum_{i=1}^{p} \alpha_i \ln \frac{\gamma_i}{\alpha_i} \right\}$$

Optimizing the above in γ shows that for $\theta \geq 0$

$$I(\gamma_{p+m_2}, \theta) = \sup_{\substack{\gamma_i \geq 0 \\ \sum \gamma_i = 1}} \left\{ \theta \sum_{ß=-m_1+1}^{p+m_2} \lambda_i \gamma_i + \sum_{i=1}^{p} \alpha_i \ln \frac{\gamma_i}{\alpha_i} \right\}$$

only depends on λ_{p+m_2}, θ and $\mu = \frac{1}{p}\sum_{i=1}^{p} \alpha_i \delta_{\lambda_i}$. It is given by

$$I(\gamma_{p+m_2}, \theta) = \theta \lambda_{m_2+p} + (v - \lambda_{p+m_2}) G_\mu(v) - \ln\theta - \int \ln|v-x| d\mu(x) - 1$$

where $v = \lambda_1$ if $G_\mu(\lambda_{p+m_2}) \leq \theta$ and $v = G_\mu^{-1}(\theta)$ if $G_\mu(\lambda_{p+m_2}) > \theta$. Here, G_μ denotes the Cauchy-Stieltjes transform $G_\mu(z) = \int (z-x)^{-1} d\mu(x)$.

We finally can deduce the case of continuous spectrum by approximation. It is then not too difficult to deduce a LDP for λ_{max} under P_N^γ

Theorem 3.10. *[18] The law of λ_{max} under P_N^γ follows a large deviation principle with speed N and good rate function*

$$I_\gamma(x) = I(x) - I(\gamma, x) - \inf_y \{I(y) - I(\gamma, y)\}$$

The proof follows from Varadhan's lemma after noticing that the empirical measure of the GOE converges towards the semi-circle law sub-Exponentially: for $\kappa > 0$ small enough

$$\mathbb{P}(d(\hat{\mu}_{\mathbf{X}_N}, \sigma) \geq N^{-\kappa}) \leq e^{-N^{1+\kappa}}$$

as well as the fact that the spherical integral is continuous (see [18] for details). This large deviation principle can be extended [9] to obtain the joint large deviation for the largest eigenvalue and $v_1(1) = \langle u, v_{max}\rangle$ if v_{max} is the eigenvector corresponding to the largest eigenvalue of $\mathbf{Y}_N$.

Theorem 3.11. *The joint law P_N of $(\lambda_1, |v_1(1)|^2)$ satisfies a large deviations principle in the scale N and good rate function I_β. In other words, for any closed set F of $\mathbb{R} \times [0,1]$*

$$\limsup_{N\to\infty} \frac{1}{N} \log P_N((\lambda_1, |v_1(1)|^2) \in F) \leq -\inf_F I_\beta \,.$$

and for any open set O of $\mathbb{R} \times [0,1]$

$$\liminf_{N\to\infty} \frac{1}{N} \log P_N((\lambda_1, |v_1(1)|^2) \in O) \geq -\inf_O I_\beta \,.$$

Moreover, I_β is a good rate function in the sense that it is non-negative and with compact level sets. I_β is infinite outside of $S = [2, +\infty) \times [0,1]$ and otherwise given by $I_\beta(x,u) = \beta(I(x,u) - \inf_S\{I\})$ where

$$I(x,u) = I(x) - \frac{1}{2}\theta x u - \frac{1}{2}\log|1-u| - \sup_{2\leq y\leq x} \{J(\sigma, \theta(1-u), y) - I(y)\} \,. \quad (114)$$

Here $\sigma(dx) = \sqrt{4-x^2}dx/2\pi$ is the semi-circle distribution and we denote

$$I(x) = \frac{x^2}{4} - \int \log|x-y| d\sigma(y) \,.$$

Moreover, if $G_\sigma(x) = \int (x-y)^{-1} d\sigma(y)$ denote the Cauchy transform of σ, we have for $\eta \leq G_\sigma(x)$,

$$J(\sigma, \eta, x) = \frac{1}{4}\eta^2, \quad (115)$$

whereas if $\eta > G_\sigma(x)$,

$$J(\sigma, \eta, x) = \frac{1}{2}\left(\eta x - 1 + \log\frac{1}{\eta} - \int \log|x-y| d\sigma(y)\right). \quad (116)$$

The second largest eigenvalue converges almost surely to the maximizer, y, of the variational problem (114) defined above.

Studying the minimizers of the rate function, we have more explicit results on the behavior of the second largest eigenvalue and the component of the associated eigenvector along $w = e_1$ associated to a given large deviation of the largest eigenvalue. We in particular see that when $\theta \geq 1$, conditionally on the largest eigenvalue of Y being close to x, $|\langle v_1, e_1\rangle|^2$ converges towards $u_{\theta,x}$ below, whereas if we additionally condition by $|\langle v_1, e_1\rangle|^2$ being close to u, the second largest eigenvalue of Y pops out of the semi-circle iff $u < 1 - 1/\theta$, and is equal to $\min\{\theta(1-u) + \frac{1}{\theta(1-u)}, x\}$.

Proposition 3.12. *Conditionally on the largest eigenvalue of Y being close to x, $|\langle v_1, e_1\rangle|^2$ is close to the minimizer of $v \to I_\beta(x,v)$. Moreover, conditionally on the largest eigenvalue of Y being close to x and $|\langle v_1, e_1\rangle|^2$ being close to u, the second largest eigenvalue of Y converges towards the maximizer of $K_{\theta(1-u)}(y) = J(\sigma, \theta(1-u), y) - I(y)$. Moreover, we have the following characterization of these optimizers :*

- *For $\theta \geq 1$ and $x > \theta + 1/\theta$: The minimum of $I_\beta(x,.)$, for a given x, is reached at $u_{\theta,x} = 1 - \frac{\theta x - \sqrt{(\theta x)^2 - 4\theta^2}}{2\theta^2}$. The maximizer of $K_{\theta(1-u)}$ pops out of the semicircle for $u < 1 - 1/\theta$, and is equal to $y(u) = \theta(1-u) + \frac{1}{\theta(1-u)}$.*

- *For $\theta \geq 1$ and $2 \leq x < \theta + 1/\theta$: The minimum of $I_\beta(x,.)$ is reached at $u_{\theta,x}$. The maximizer of $K_{\theta(1-u)}$ pops out of the semicircle for $u < 1 - 1/\theta$, and is equal to $\inf(y(u), x)$, i.e. it increases when u decreases until reaching the value x.*

- *For $\theta < 1$ and $x \geq 2$: The minimum of the large deviation function $I_\beta(x,.)$ is taken at $u = 0$, if $u_{\theta,x}$ is not positive, or at $u_{\theta,x}$ otherwise. The latter case corresponds to large enough values of x ($x \geq \theta + 1/\theta$). The maximizer of $K_{\theta(1-u)}$ sticks to two.*

The proof of Theorem 3.11 is based on the remark that again we can compute the joint law of $(v_1(1)^2, \lambda_1)$. Indeed, the law of $\mathbf{Y}_N = \mathbf{X}_N + \gamma u u^T$ is given by

$$d\mathbb{P}_N(Y) = \frac{1}{Z_N} \exp\left\{-\frac{N}{4} \operatorname{Tr}(Y - \theta w w^T)^2\right\} dY = \frac{1}{\tilde{Z}_N} \exp\left\{-\frac{N}{4} \operatorname{Tr} Y^2 + \frac{N}{2}\theta\langle w, Yw\rangle\right\} dY\,.$$

Therefore, since $\langle w, Yw\rangle = \sum \lambda_i |v_i(1)|^2$ when $w = e_1$, the joint law of $(\lambda_1, |v_1(1)|^2)$ is given by

$$dP_N(x,u) = \frac{1}{\bar{Z}_N} e^{-\frac{N}{4}x^2 + \frac{N}{2}\theta x u} \int \prod_{i=2}^{N} |x - \lambda_i| I_N(\lambda, \theta, u) dP^x_{N-1}(\lambda) dx dB_1(u) \tag{117}$$

where we denoted

$$I_N(\lambda, \theta, u) = \mathbb{E}[e^{\frac{N}{2}\theta \sum_{i=2}^N \lambda_i |v_i(1)|^2} || v_1(1)|^2](u) \tag{118}$$

and $\mathbb{E}[.||v_1(1)|^2](u)$ is the expectation on $\{|v_i(1)|, i \geq 2\}$ conditionally to $\{|v_1(1)|^2 = u\}$. $dB_1(u)$ is the distribution of $|v_1(1)|^2$. Moreover, P^x_{N-1} is the positive measure given by

$$dP^x_{N-1}(\lambda) = \frac{1}{Z^\infty_{N-1}} \prod_{2 \leq i < j \leq N} |\lambda_i - \lambda_j| e^{-\frac{N}{4}\sum_{i=2}^N \lambda_i^2} \prod_{i=2}^N 1_{\lambda_i \leq x} d\lambda_i\,,$$

where Z^∞_{N-1} is independent of x and such that P^∞_N is a probability measure :

$$Z^\infty_{N-1} = \int \prod_{2 \leq i < j \leq N} |\lambda_i - \lambda_j| e^{-\frac{N}{4}\sum_{i=2}^N \lambda_i^2} \prod_{i=2}^N d\lambda_i\,.$$

Our main goal is to estimate the density of P_N when N is large and to apply Laplace's method. We infer from concentration inequalities that with $\hat{\mu}^{N-1} = \frac{1}{N-1}\sum_{i=2}^N \delta_{\lambda_i}$

$$\begin{aligned} \sup_{x>2} \limsup_{N\to\infty} \frac{1}{N} \log P^x_{N-1}\left(d(\hat{\mu}^{N-1}, \sigma) > N^{-\kappa'}\right) &\leq \limsup_{N\to\infty} \frac{1}{N} \log P^\infty_{N-1}\left(d(\hat{\mu}^{N-1}, \sigma) > N^{-\kappa'}\right) \\ &= -\infty\,. \end{aligned} \tag{119}$$

Here, d denotes the Dudley distance and κ' belongs to $(0, \frac{1}{10})$ (see [13, Lemma 6.1] for details). We deduce that for $x > 2$ (the singularity of the log can be overcome as in [?]), with probability greater than $1 - e^{-MN}$ for all $M > 0$,

$$\prod_{i=1}^{N-1} |x - \lambda_i| = e^{N \int \log |t-x| d\sigma(t) + o(N)}, \tag{120}$$

i.e. we can replace the empirical distribution $\hat{\mu}^{N-1}$ by its limit. To estimate the other terms, we first observe that since $(v_i(1))_{1\le i\le N}$ is uniformly distributed on the sphere with radius one, we can represent $(v_i(1))_{1\le i\le N}$ as

$$v_i(1) = \frac{g_i}{(\sum_{i=2}^N |g_i|^2 + |g_1|^2)^{1/2}} \tag{121}$$

with independent standard Gaussian variables $g_i, 1 \le i \le N$, which are real when $\beta = 1$ and complex when $\beta = 2$. As a consequence, a simple change of variables shows that the distribution B_1 of $|v_1(1)|^2$ is the Beta-distribution

$$dB_1(u) = C_N u^{\beta/2-1}(1-u)^{(N-1)\beta/2-1} du \tag{122}$$

To estimate $I_N(\lambda, \theta, u)$, we first determine the distribution of $(|v_N(1)|^2, \dots, |v_2(1)|^2)$ conditionally to $|v_1(1)|^2$, again by using (121). In fact, if we fix $|v_1(1)|$ and denote $g^{N-1} = (g_2, \dots, g_N)$, we have

$$\|g\|_2^2 = \|g^{N-1}\|_2^2 + |v_1(1)|^2 \|g\|_2^2$$

so that $w_i = g_i/\|g^{N-1}\|_2, 2 \le i \le N$, follows the uniform law on the sphere $\mathbb{S}^{N-2}$ with radius one and

$$v_i(1) = w_i \frac{\|g^{N-1}\|_2}{\|g\|_2} = \sqrt{1 - |v_1(1)|^2} w_i$$

Observe that $(w_i)_{2\le i\le N}$ is independent of $|v_1(1)|^2$. Hence, we conclude that

$$I_N(\lambda, \theta, u) = \mathbb{E}_{w_i, i\ge 2}[e^{\frac{N}{2}\beta\theta(1-u)\sum_{i=2}^N \lambda_i |w_i|^2}]. \tag{123}$$

We next estimate this quantity as in Theorem 3.8. Hence, all terms can be estimated and Varadhan's lemma holds.

3.6 Universality and non universality of LDP's

So far we have considered large deviations for Gaussian ensembles. Let us discuss what we know for more general entries in the set up of Wigner matrices. Bordenave-Caputo [10] and Augeri [3] showed that if the entries have tails heavvier than Gaussian, namely such that

$$\mathbb{P}(|X_{ij}| \ge t) = e^{-aN^{\alpha/2} t^\alpha}$$

for N and t large, then the law of the empirical measure satisfies a large deviation but in the scale $N^{1+\alpha/2}$ whereas the largest eigenvalue satisfies a large deviation principle in the scale $N^{\alpha/2}$. The point is that the best strategy to creat such deviatioons is to change of order N, respectively a finite number, of entries so that they are of order one. Large deviations for sub-Gaussians entries are still unknow in the case of the empirical measure but recent progresses allowed to understand better what happens for the largest eigenvalue. In [13], it was shown that such large deviatioons are the same if the entries have “sharp”sub Gaussian bounds that is that for all t

$$\mathbb{E}[e^{tX_{ij}}] \le e^{\frac{t^2}{2N}}$$

Such a bound is verified for Rademacher variables and uniformly distributed variables. Then

Theorem 3.13. *The largest eigenvalue satisfies a large deviation principle with good rate function*

$$I(x) = I_{GOE}(x) = \int_2^x \sqrt{y^2-4}dy$$

In the case where for all t

$$\mathbb{E}[e^{tX_{ij}}] \le e^{\frac{At^2}{2N}}$$

with some $A > 1$ we proved with F. Augeri that this universality is not true anymore:

Theorem 3.14. *Under additional technical assumptions*

- *There exists B finite such that for any $x > B$*

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \frac{1}{N} \ln \mathbb{P}(|\lambda_1 - x| \le \delta) = -I_A(x)$$

  $I_A \simeq x^2/2A$ when $x \to \infty$. The same holds with the limsup replaced by a liminf.

- *If $A < 2$ and $x \to \frac{1}{x^2} \log \mathbb{E}[e^{xX_{ij}}]$ is increasing, the same results holds with $I = I_{GOE}$ for*

$$x \in \big[2, \frac{1}{\sqrt{A-1}} + \sqrt{A-1}\big],\ I(x) = I_{GOE}(x).$$

The proof of these results uses again spherical integrals. Let us give a sketch of the proof of Theorem 3.13. It uses a tilt of the measure by spherical integrals:

$$\begin{aligned}
\mathbb{P}(|\lambda_{max} - x| < \delta) &= \mathbb{E}[\frac{I_N(X,\theta)}{I_N(X,theta)} 1_{|\lambda_{max}-x|<\delta}] \\
&\simeq e^{-N(I(\sigma,\theta,x)+o(\delta))}\mathbb{E}[I_N(X,\theta)1_{|\lambda_{max}-x|<\delta}] \\
&\le e^{-N(I(\sigma,\theta,x)+o(\delta))}\mathbb{E}[I_N(X,\theta)]
\end{aligned}$$

where in the second line we used the continuity of spherical integrals and their asymptotics. We next compute the expectation of spherical integrals:

$$\begin{aligned}
\mathbb{E}[I_N(X,\theta)] &= \mathbb{E}_e[\prod_{i\le j}\mathbb{E}[e^{\sqrt{N}2^{-1_{i=j}}\theta e_i e_j x_{ij}}] \\
&\le \mathbb{E}[\prod_{i\le j} e^{\frac{1}{2}\theta^2 2^{-1_{i=j}} N e_i^2 e_j^2}] \\
&= e^{\frac{1}{2}N\theta^2}
\end{aligned}$$

where we used that the covariance is 2 on the diagonal (and one outside)

For the lower bound we notice that

$$\begin{aligned}
\mathbb{E}[I_N(X,\theta)] &= \mathbb{E}_e[\prod_{i\le j}\mathbb{E}[e^{\sqrt{N}2^{-1_{i=j}}\theta e_i e_j x_{ij}}]] \\
&\ge \mathbb{E}_e[\prod_{i\le j} 1_{|\sqrt{N}e_i e_j|\le N^{-1/4}}\mathbb{E}[e^{\sqrt{N}2^{-1_{i=j}}\theta e_i e_j x_{ij}}]] \\
&\ge \mathbb{E}[\prod_{i\le j} prod_{i\le j} 1_{|\sqrt{N}e_i e_j|\le N^{-1/4}} e^{\frac{1}{2}\theta^2 2^{-1_{i=j}} N e_i^2 e_j^2}] \\
&= e^{\frac{1}{2}N\theta^2}\mathbb{P}(\max_{i\le j}|\sqrt{N}e_i e_j|\le N^{-1/4})
\end{aligned}$$

We conclude the proof by noticing that $\mathbb{P}(\max_{i\le j}|\sqrt{N}e_i e_j|\le N^{-1/4})$ goes to one. Hence

$$\lim_{N\to\infty}\frac{1}{N}\ln\mathbb{E}_e[I_N(X,\theta)] = \frac{1}{2}\theta^2$$

and we have the upper bound

$$\limsup_{\delta\downarrow 0}\limsup_{N\to\infty}\frac{1}{N}\log N\mathbb{P}(|\lambda_{max}-x|<\delta)\le\frac{1}{2}\theta^2 - I(\sigma,\theta,x)$$

Optimizing over θ yields

$$\limsup_{\delta\downarrow 0}\limsup_{N\to\infty}\frac{1}{N}\log N\mathbb{P}(|\lambda_{max}-x|<\delta)\le -\sup_{\theta\ge 0}\{I(\sigma,\theta,x)-\frac{1}{2}\theta^2\}$$

and the above RHS can be seen to be equal to I_{GOE}. To prove the lower bound we remark that

$$\begin{aligned}
\mathbb{P}(|\lambda_{max}-x|<\delta) &= \mathbb{E}[\frac{I_N(X,\theta)}{I_N(X,\theta)}1_{|\lambda_{max}-x|<\delta}] \\
&\simeq e^{-N(I(\sigma,\theta,x)+o(\delta))}\mathbb{E}[I_N(X,\theta)1_{|\lambda_{max}-x|<\delta}] \\
&\simeq e^{-N(I(\sigma,\theta,x)+N\theta^2+o(\delta))}\frac{\mathbb{E}[I_N(X,\theta)1_{|\lambda_{max}-x|<\delta}]}{\mathbb{E}[I_N(X,\theta)]}
\end{aligned}$$

and therefore it is enough to show that there exists θ such that

$$\liminf_{N\to\infty} \frac{\mathbb{E}[I_N(X,\theta)1_{|\lambda_{max}-x|<\delta}]}{\mathbb{E}[I_N(X,\theta)]} = 1\,.$$

Let us notice that

$$\frac{\mathbb{E}[I_N(X,\theta)1_{|\lambda_{max}-x|<\delta}]}{\mathbb{E}[I_N(X,\theta)]} = \frac{1}{\mathbb{E}[I_N(X,\theta)]}\mathbb{E}_e[\mathbb{E}_X[e^{\frac{N}{2}\theta\langle e,Xe\rangle}]\frac{\mathbb{E}_X[e^{\frac{N}{2}\theta\langle e,Xe\rangle}1_{|\lambda_{max}-x|<\delta}]}{\mathbb{E}_X[e^{\frac{N}{2}\theta\langle e,Xe\rangle}]}]$$

We shall conclude by proving that

$$\liminf_{N\to\infty} \inf_{\max_{i\le j}|\sqrt{N}e_ie_j|\le N^{-1/4}} \frac{\mathbb{E}_X[e^{\frac{N}{2}\theta\langle e,Xe\rangle}1_{|\lambda_{max}-x|<\delta}]}{\mathbb{E}_X[e^{\frac{N}{2}\theta\langle e,Xe\rangle}]} = 1$$

provided $x = \theta + \frac{1}{\theta}$. In fact, under the tilted measure

$$\mathbb{P}^{\theta,e}(dX) = \frac{1}{E_X[e^{\frac{N}{2}\theta\langle e,Xe\rangle}]} e^{\frac{N}{2}\theta\langle e,Xe\rangle}\prod d\mu(x_{ij})$$

X still has independent entries above the diagonal. When $\max_{i\le j}|\sqrt{N}e_ie_j| \le N^{-1/4}$ we can easily compute the mean m_{ij}^θ and the variance σ_{ij}^θ of the entries under this tilted law by simply expanding the density: we find that

$$m_{ij}^\theta = \sqrt{N}2^{-1_{i=j}}\theta e_ie_j(1+o(1)), \quad \sigma_{ij}^\theta = 1+o(1)$$

Hence under $\mathbb{P}^{\theta,e}$, $\mathbf{X}_N$ has approximately the law of $\mathbf{X}_N^\theta = \mathbf{Y}_N + \theta ee^T$ where $\mathbf{Y}_N$ has the distribution of a standard Gaussian variable. We have seen that the largest eigenvalue of $\mathbf{X}_N^\theta$ goes towards $\theta + \frac{1}{\theta}$ if $\theta \ge 1$, which concludes the proof.

3.7 Generalization

- This approach allows to get the LDP also for $D + \mathbf{X}_N$ provided that D has no outliers (B. Mc Kenna), to matrices with a more elaborate variance profile (J. Husson), to the largest eigenvalue of $A + UBU^*$ (Guionnet-Maida), it generalizes as well to the pth largest eigenvalues (JW Husson-Guionnet).
- The LDP for the largest eigenvalue of $\mathbf{X}_N + \gamma ee^T$ for non Gaussian entries is under investigation.
- The LDP for the spectral measure for Wigner matrices with sub-Gaussian tails is open. A partial answer was obtained for $A + UBU^*$ with U Haar (Belinschi, Guionnet, Huang).

References

[1] G. W. Anderson and O. Zeitouni, *A CLT for a band matrix model*, Probab. Theory Rel. Fields **134** (2005), 283–338.

[2] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR 2760897

[3] Fanny Augeri, *Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails*, Electron. J. Probab. **21** (2016), Paper No. 32, 49. MR 3492936

[4] Z. D. Bai and Jack W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab. **26** (1998), no. 1, 316–345. MR 1617051

[5] J. Baik, G. Ben Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Annals Probab. **33** (2005), 1643–1697.

[6] G. Ben Arous, A. Dembo, and A. Guionnet, *Aging of spherical spin glasses*, Probab. Theory Related Fields **120** (2001), no. 1, 1–67. MR 1856194

[7] G. Ben Arous and A. Guionnet, *Large deviations for Wigner's law and Voiculescu's non-commutative entropy*, Probab. Theory Rel. **108** (1997), 517–542.

[8] G. Ben Arous and A. Guionnet, *Large deviations for Wigner's law and Voiculescu's non-commutative entropy*, Probab. Theory Related Fields **108** (1997), no. 4, 517–542. MR 1465640

[9] Giulio Biroli and Alice Guionnet, *Large deviations of spiked random matrices*, To appear in ECP (2020).

[10] Charles Bordenave and Pietro Caputo, *A large deviation principle for Wigner matrices without Gaussian tails*, Ann. Probab. **42** (2014), no. 6, 2454–2496. MR 3265172

[11] R. N. Rao. F. Benaych-Georges, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Advances in Math. (2009).

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

[13] A. Guionnet and J. Husson, *Large deviations for the largest eigenvalue of Rademacher matrices*, To appear in Annals of Probability **arXiv:1810.0118** (2019).

[14] A. Guionnet and M. Maï da, *A Fourier view on the R-transform and related asymptotics of spherical integrals*, J. Funct. Anal. **222** (2005), no. 2, 435–490. MR 2132396

[15] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Commun. Prob. **5** (2000), 119–136 (electronic).

[16] Alice Guionnet, *Large random matrices: lectures on macroscopic asymptotics*, Lecture Notes in Mathematics, vol. 1957, Springer-Verlag, Berlin, 2009, Lectures from the 36th Probability Summer School held in Saint-Flour, 2006. MR 2498298

[17] Cosme Louart, Zhenyu Liao, and Romain Couillet, *A random matrix approach to neural networks*, Ann. Appl. Probab. **28** (2018), no. 2, 1190–1248. MR 3784498

[18] Mylène Maï da, *Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles*, Electron. J. Probab. **12** (2007), 1131–1150. MR 2336602

[19] M.L. Mehta, *Random matrices*, Academic Press, 1991.

[20] Catherine Donati-Martin Mireille Capitaine and Delphine Fral, *The largest eigenvalues of finite rank deformation of large wigner matrices: Convergence and nonuniversality of the fluctuations*, Annals of Probability, 1 – 47.

[21] J. Silverstein and Z. D. Bai, *On the empirical distribution of eigenvalues of large dimensional random matrices*, J. Multivariate Anal. **54** (1995), 175–192.

[22] Ya. G. Sinaĭ and A. B. Soshnikov, *A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices*, Funktsional. Anal. i Prilozhen. **32** (1998), no. 2, 56–79, 96. MR 1647832

[23] Dan Voiculescu, *The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms*, Invent. Math. **132** (1998), no. 1, 189–227. MR 99d:46087

[24] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals Math. **62** (1955), 548–564.