Report on the spectrum of heavy-tailed random matrices

Ernest van Wijland

May 2021

Contents

1 Abstract

Let X_N an N by N random symmetric matrix, such that its coefficients $(X_{ij})_{i \leq j}$ are real, independent and equidistributed. We will study the convergence of the distribution of its eigenvalues, under different constraints on the law of the coefficients, P. If P has finite moments, Wigner [1] has shown that the spectral measure will converge towards the semicircular distribution. In a more recent paper, Ben Arous and Guionnet [4] have proved that when P is in the domain of attraction of an α-stable law, after renormalization, the corresponding spectral distribution converges towards a measure μ_α which only depends on the parameter α. We will also go over the localization of the eigenvectors for these different laws, as studied by Bordenave and Guionnet in [5].

If A is an N by N matrix, and $(\lambda_1, \ldots, \lambda_N)$ are its eigenvalues, then we note μ_A its spectral measure, *id est* $\mu_A := \frac{1}{N} \sum\limits_{i=1}^{N} \delta_{\lambda_i}$.

2 Finite second moment distributions

Let μ_{sc} the semicircular distribution, defined by $\mu_{sc}(dx) := \frac{1}{2\pi}\sqrt{4 - x^2} 1_{|x| \leq 2} dx$.

Then, if P has finite moments, the spectral measure will converge towards the semicircular distribution. This is illustrated by this simulation:

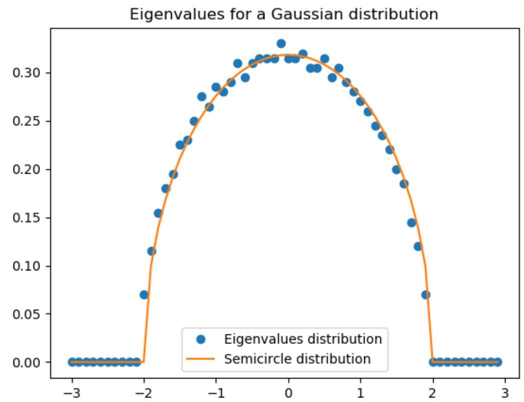

Figure 1: Eigenvalues distribution for $N = 2000$ and P the Gaussian law, step size $= 0.1$

Theorem 1 (Wigner's theorem)**.** *Assume that P has a unit second moment, and that all its higher moments are finite. Then, if we let $Y_N = X_N/\sqrt{N}$, μ_{Y_N} converges weakly in probability to μ_{sc}.*

Proof. We will use the notation $\langle \mu, \phi \rangle := \int_{\mathbb{R}} \phi(x) d\mu(x)$. The method used by Wigner, and explained in [2], relies on the following steps:

Lemma 1.1. *Let $\bar{\mu}_{Y_N}$ the average spectral measure. For any $k \in \mathbb{N}^*$,*

$$\langle \bar{\mu}_{Y_N}, x^k \rangle \to \langle \mu_{sc}, x^k \rangle$$

This means that the average measure has the same moments as the semicircular distribution. The following lemma allows us to work with the average measure:

Lemma 1.2. *For any $\varepsilon > 0$ and $k \in \mathbb{N}^*$,*

$$P(|\langle \mu_{Y_N}, x^k \rangle - \langle \bar{\mu}_{Y_N}, x^k \rangle| > \varepsilon) \to 0$$

The odd moments are equal to zero by symmetry of μ_{sc}. To compute the even moments, we can do the variable change $x = 2\sin\theta$, and find a recurrence relation between consecutive even moments. This gives us:

Lemma 1.3. *For any $k \in \mathbb{N}^*$, $\langle \mu_{sc}, x^{2k} \rangle = C_k = \frac{1}{k+1}\binom{2k}{k}$, the k-th Catalan number.*

This is enough to prove the theorem:

We are trying to prove that for any continuous bounded function $f : \mathbb{R} \to \mathbb{R}$, $\langle \mu_{Y_N}, f \rangle \to \langle \mu_{sc}, f \rangle$ in probability. We will use Weierstrass approximation theorem and replace f with a polynomial, which will allow use to rewrite the above integrals as linear combinations of the moments.

Weierstrass only holds for compactly supported functions, thus we show that we can assume that f is compactly supported by $[-5, 5]$:

Markov's inequality gives:

$$P(\langle \mu_{Y_N}, |x|^k 1_{|x|>5} \rangle > \varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}\langle \mu_{Y_N}, |x|^k 1_{|x|>5} \rangle \leq \frac{\langle \bar{\mu}_{Y_N}, x^{2k} \rangle}{\varepsilon 5^k}$$

where the last equality comes from putting $|x|^k/5^k > 1$ inside the integral.

Lemma 1 and $C_k \leq 4^k$ give:

$$\limsup P(\langle \mu_{Y_N}, |x|^k 1_{|x|>5} \rangle > \varepsilon) \leq \frac{\langle \mu_{sc}, x^{2k} \rangle}{\varepsilon 5^k} \leq \frac{4^k}{\varepsilon 5^k}$$

The left side of the inequality is either zero or increasing with k while the right side is strictly decreasing with k, therefore, we have

$$\limsup P(\langle \mu_{Y_N}, |x|^k 1_{|x|>5} \rangle > \varepsilon) = 0 \quad (*)$$

Let f continuous and bounded, as shown, we can assume that it is compactly supported by $[-5, 5]$. Let $\delta > 0$. Let p_δ a polynomial such that $\forall x \in [-5, 5]$, $|p_\delta(x) - f(x)| \leq \delta/4$.

We have:

$$\begin{aligned}
|\langle \mu_{Y_N}, f\rangle - \langle \mu_{sc}, f\rangle| &\leq |\langle \mu_{Y_N}, f - p_\delta\rangle - \langle \mu_{sc}, f - p_\delta\rangle| \\
&+ |\langle \mu_{Y_N}, p_\delta\rangle - \langle \mu_{sc}, p_\delta\rangle| \\
&\leq |\langle \mu_{Y_N}, (f - p_\delta)1_{|x|\leq 5}\rangle| \\
&+ |\langle \mu_{sc}, (f - p_\delta)1_{|x|\leq 5}\rangle| \\
&+ |\langle \mu_{Y_N}, p_\delta 1_{|x|>5}\rangle| \\
&+ |\langle \mu_{Y_N}, p_\delta\rangle - \langle \mu_{sc}, p_\delta\rangle|
\end{aligned}$$

The first two terms are bounded by $\delta/4$ by definition of p_δ. Therefore:

$$\begin{aligned}
P(|\langle \mu_{Y_N}, f\rangle - \langle \mu_{sc}, f\rangle| > \delta) &\leq P(|\langle \mu_{Y_N}, p_\delta 1_{|x|>5}\rangle| > \delta/2) \\
&+ P(|\langle \bar{\mu}_{Y_N}, p_\delta\rangle - \langle \mu_{sc}, p_\delta\rangle| > \delta/2) \\
&+ P(|\langle \mu_{Y_N}, p_\delta\rangle - \langle \bar{\mu}_{Y_N}, p_\delta\rangle| > \delta/2)
\end{aligned}$$

In virtue of $(*)$ the first term converges to zero, lemma 1 guarantees that the second term stagnates at zero for N large enough, and lemma 2 guarantees that the third term goes towards zero as $N \to \infty$.

Hence, for any $\delta > 0$ and f continuous and bounded,

$$P(|\langle \mu_{Y_N}, f\rangle - \langle \mu_{sc}, f\rangle| > \delta) \to 0$$

therefore, $\mu_{Y_N} \to \mu_{sc}$ weakly in probability. □

There exists a stronger version of the theorem that states that $\mu_{Y_N} \to \mu_{sc}$ weakly almost surely, whose proof we will not discuss in this report.

3 Heavy-tailed distributions

3.1 Introduction

We now consider the case of heavy-tailed coefficients, which means that their second moment is infinite. Let $\alpha \in]0, 2[$, we will assume that the entries are in the domain of attraction of an α-stable law, *id est*: there exists a slowly varying function (for example, a constant function), L such that

$P(|x_{ij}| \geq u) = \frac{L(u)}{u^\alpha}$.

We normalize the matrix by defining $a_N := \inf(u, P(|x_{ij}| \geq u) \leq 1/N)$, and letting $A_N = a_N^{-1} X_N$, and $(\lambda_1, \ldots, \lambda_N)$ its eigenvalues.

We first want to prove the following theorem, that gives us a limit ditribution for μ_{A_N}:

Theorem 2. *1. There exists a probability measure μ_α such that the mean spectral measure $\mathbb{E}(\mu_{A_N})$ converges weakly to μ_α.*
2. μ_{A_N} converges weakly in probability to μ_α.

Then, we will show the following specifications on the shape of μ_α:

Theorem 3. *1. μ_α is symmetric.*
2. μ_α has unbounded support.
3. μ_α has heavy tails, which means that there exists a constant $L_\alpha > 0$ such that, when $|x| \to \infty$, noting ρ_α its density:

$$\rho_\alpha(x) \sim \frac{L_\alpha}{|x|^{\alpha+1}}$$

3.2 Truncating of the entries

To make it easier to work with our random matrices, we will truncate them. In this section, we introduce the adequate truncations, and prove that it is valid to approximate A_N's spectral measure by that of its truncated counterparts.

For $B > 0$ we define X_N^B the matrix with coefficients $x_{ij} 1_{|x_{ij}| \leq B a_N}$, for $\kappa > 0$, we define X_N^κ the matrix with coefficients $x_{ij} 1_{|x_{ij}| \leq N^\kappa a_N}$, and the matrices $A_N^B := a_N^{-1} X_N^B$ and $A_N^\kappa := a_N^{-1} X_N^\kappa$.

We define d the Dudley distance and d_1 a variant:

$$d(\mu, \nu) := \sup_{||f||_{\mathcal{L}} \leq 1} |\int f d\nu - \int f d\mu|$$

$$d_1(\mu, \nu) := \sup_{||f||_{\mathcal{L}} \leq 1, f \uparrow} |\int f d\nu - \int f d\mu|$$

where the sup is taken over the set of non-decreasing Lipschitz function with norm less than 1. These distances are compatible with the weak topology on $\mathcal{P}(\mathbb{R})$.

And we prove that for this distance, truncating doesn't change the spectral measure a lot:

Theorem 4. *1. For any $\varepsilon > 0$, there exists $B(\varepsilon) < \infty$ and $\delta(\varepsilon, B) > 0$ when $B > B(\varepsilon)$ such that, for N large enough:*

$$P(d_1(\mu_{A_N}, \mu_{A_N^B}) > \varepsilon) \leq e^{-\delta(\varepsilon,B)N}$$

2. For $\kappa > 0$ and $a \in]1-\alpha\kappa, 1[$, there exists a finite constant $C(\alpha, \kappa, a)$ such that for any $N \in \mathbb{N}$,

$$P(d_1(\mu_{A_N}, \mu_{A_N^\kappa}) > N^{a-1}) \leq e^{-CN^a \log N}$$

3.3 Tightness

This section proves that the mean spectral distributions of A_N, A_N^B and A_N^κ are tight.

Theorem 5. *The sequences $(\mathbb{E}(\mu_{A_N}))_N$, $(\mathbb{E}(\mu_{A_N^B}))_N$, $(\mathbb{E}(\mu_{A_N^\kappa}))_N$ are tight for the weak topology on $\mathcal{P}(\mathbb{R})$.*

3.4 Limiting equation

To prove the vague convergence of $(\mathbb{E}(\mu_{A_N}))_N$, we study the asymptotic behavior of the following probability measure on $\mathbb{C}$, for $z \in \mathbb{C}$ and $f \in \mathcal{C}_b(\mathbb{C})$:

$$L_N^z(f) := \mathbb{E}\left(\frac{1}{N}\sum_{k=1}^{N} f(((zI_N - A_N)^{-1})_{kk})\right)$$

similarly:

$$L_N^{z,\kappa}(f) := \mathbb{E}\left(\frac{1}{N}\sum_{k=1}^{N} f(((zI_N - A_N^\kappa)^{-1})_{kk})\right)$$

For $f : x \mapsto x$, $L_N^z(f) = \mathbb{E}\left(\frac{1}{N}\mathrm{tr}(zI_N - A_N)^{-1}\right)$ is the Stieltjes transform of $\mathbb{E}(\mu_{A_N})$. Therefore, the weak convergence of L_N^z for all z in a set with accumulation points would be enough to prove the vague convergence of $\mathbb{E}(\mu_{A_N})$, because it is a consequence of the convergence of its Stieltjes transform, which is an analytic function.

We can then prove the following lemma, which will prove useful in proving convergence for $L_N^{z,\kappa}$:

Lemma 5.1. *For $\kappa \in]0, \frac{1}{2-\alpha}$, let $\varepsilon = 1 - \kappa(2-\alpha) > 0$. There exists a finite constant c such that for $z \in \mathbb{C} \setminus \mathbb{R}$ and f a Lipschitz function on $\mathbb{C}$:*

$$P(|L_N^{z,\kappa}(f) - \mathbb{E}(L_N^{z,\kappa}(f))| \geq \delta) \leq \frac{c||f||_{\mathcal{L}}^2}{|Im(z)|^4\delta^2} N^{-\varepsilon}$$

3.5 Proof of the convergence theorem

Theorem 6. *For any $\kappa \in]0, \frac{1}{2(2-\alpha}[$, any $z \in \mathbb{C}_+$, the Stieltjes transform of $\mathbb{E}(\mu_{A_N^\kappa})$, $\mathbb{E}\left(\frac{1}{N} tr((zI_N - A_N^\kappa)^{-1})\right)$ converges to $G_\alpha(z)$ defined by:*

$$G_\alpha(z) := i\int_0^\infty e^{itz} e^{-c(\alpha)(it)^{\frac{\alpha}{2}} X_z} dt$$

where we admit that X_z exists and is the unique analytic solution of the equation:

$$X_z = iC(\alpha)\int_0^\infty (it)^{\frac{\alpha}{2}-1} e^{itz} e^{-c(\alpha)(it)^{\frac{\alpha}{2}} X_z} dt$$

with $C(\alpha) := \frac{e^{i\frac{\pi\alpha}{2}}}{\Gamma(\frac{\alpha}{2})}$ and $c(\alpha) := \Gamma(1 - \frac{\alpha}{2})$.

This last result allows us to conclude:

From theorem 5 we get that $(\mathbb{E}(\mu_{A_N^\kappa}))_N$ is tight for the weak topology. If we take a subsequence, we obtain that any limit point μ will have a Stieltjes transform equal to $G_\alpha(z)$, for all $z \in \mathbb{C}_+$. Hence, μ is unique, and verifies:

$$\int (z-x)^{-1} d\mu(x) = G_\alpha(z),\ \forall z \in \mathbb{C}_+$$

and this ensures the weak convergence of $(\mathbb{E}(\mu_{A_N^\kappa}))_N$.

Since $d_1(\mathbb{E}(\mu_{A_N^\kappa}), \mathbb{E}(\mu_{A_N})) \leq \mathbb{E}(d_1(\mu_{A_N^\kappa}, \mu_{A_N}))$ and theorem 4, $\mathbb{E}(\mu_{A_N})$ converges weakly towards μ.

Lemma 5.1 gives us that $\forall z \in \mathbb{C} \setminus \mathbb{R}$, $L_N^{z,\kappa}(x) = \int (z-x)^{-1} d\mu_{A_N^\kappa}(x)$ converges in probability towards $G_\alpha(z)$. The set $\{(z-x)^{-1}, z \in \mathbb{C} \setminus \mathbb{R}\}$ is dense in the set $\mathcal{C}_0$ of functions on $\mathbb{R}$ that converge to zero at infinity, $\int f d\mu_{A_N^\kappa}$ converges in probability towards $\int f d\mu$, for all $f \in \mathcal{C}_0$. This vague convergence can easily be strengthened into a weak convergence. Furthermore, theorem 4 allows us to replace $\mu_{A_N^\kappa}$ by μ_{A_N}. Therefore, μ_{A_N} converges weakly in probability towards μ, which ends the proof of theorem 2.

3.6 Shape of the limit distribution

We have the following result on the shape of the limit distribution μ_α:

Theorem 2. *1. μ_α is symmetric.*
2. μ_α has unbounded support.
3. μ_α has heavy tails, which means that there exists a constant $L_\alpha > 0$ such that, when $|x| \to \infty$, noting ρ_α its density:

$$\rho_\alpha(x) \sim \frac{L_\alpha}{|x|^{\alpha+1}}$$

This is illustrated by this simulation, where the law I used for the coefficients has density $K\frac{1}{|x|^{\alpha+1}}1_{|x|\geq 1}$ with $\alpha = 1.5$:

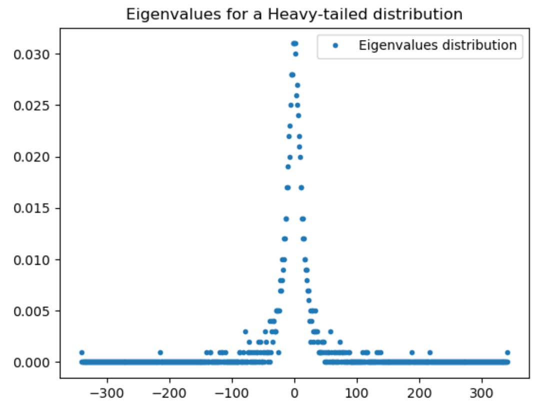

Figure 2: Eigenvalues distribution for $N = 1000$ and P a power law in the domain of attraction of an α-stable law, with $\alpha = 1.5$, step size = 1

4 Localization of the eigenvectors

Another aspect of the spectrum of great interest is the localization of the eigenvectors. Indeed, depending on the law of the coefficients, the repartition of the "mass" of the vectors can be more or less homogeneous.

For finite second-moment distributions, it is a well-known result that the eigenvectors are delocalized, meaning the mass is homogeneously distributed among the eigenvector's entries:

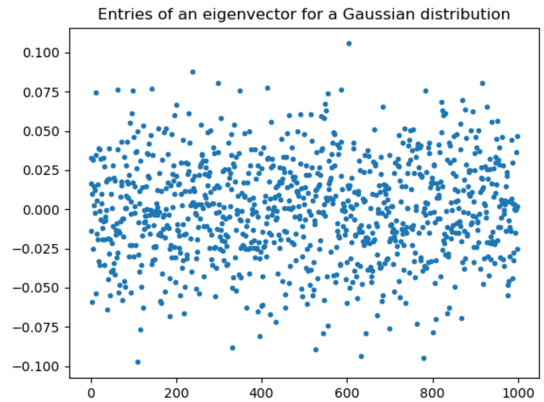

Figure 3: Entries of an eigenvector for a Gaussian distribution with $N = 1000$

For heavy-tails distribution, article [5] by Bordenave and Guionnet study this property, which we can summarize as:

- for $\alpha \in]1, 2[$, all but $o(N)$ of the eigenvectors are delocalized:

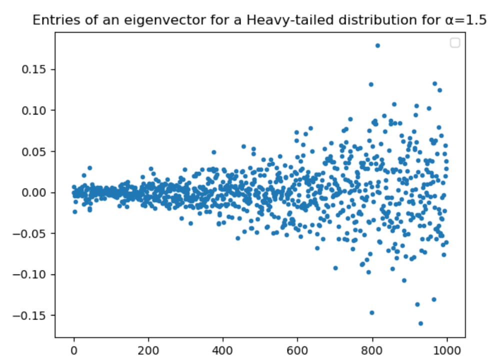

Figure 4: Entries of an eigenvector for a Heavy-tailed distribution with $N = 1000$ and $\alpha = 1.5$

- for $\alpha \in]0, 1[$, the localization depends on the eigenvalue we are looking at, if it is above a certain threshold, the eigenvector will be localized:

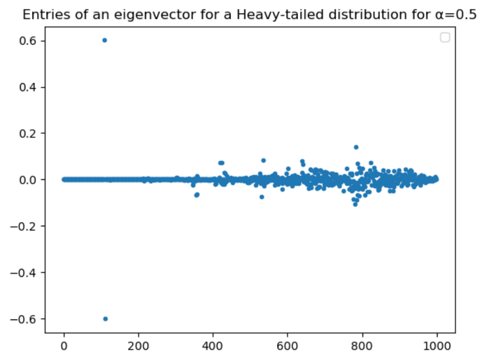

Figure 5: Entries of an eigenvector for a Heavy-tailed distribution with $N = 1000$ and $\alpha = 0.5$

A more recent article, from last month, [6], links those results to the localization of the eigenvectors of the adjacency matrix of Erdös-Renyi graphs. The same delocalization transition phenomenon appears in the case of $G(N, d/N)$ graphs, with d of order $\log N$.

5 Conclusion

Working on this report, I learned fundamental proof methods and approaches on random matrix theory results. It allowed me to go beyond my classic Integration and Probability course, and to enjoy a more in-depth discovery of this fruitful area, with countless applications, ranging from Physics to Graph Theory. I also got to find out how real research papers were structured, and the extent to which they were built on top of each other.

With this report, I hope to provide a detailed synopsis of Wigner's classic result, and Ben Arous and Guionnet's more specialized characterisation of the limit spectral distribution of heavy-tailed random matrices. I tried to provide visual illustrations, and clear guidelines for the constructions of the proofs.

6 References

[**1**] WIGNER, E. P. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2) 67* (1958), 325–327.

[**2**] FEIER, A. R. Methods of proof in random matrix theory.

[**3**] FELLER, W. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[**4**] BEN AROUS, G., GUIONNET, A. The Spectrum of Heavy Tailed Random Matrices.

[**5**] BORDENAVE, C., GUIONNET, A. Localization and delocalization of eigenvectors for heavy-tailed random matrices.

[**6**] ALT, J., DUCATEZ, R., KNOWLES, A. Delocalization transition for critical Erdős–Rényi graphs.