

Matrices Aléatoires

Léo Gayral

Ces notes sont basées sur le cours d'Édouard Maurel-Segala. On pourra se référer à l'ouvrage *An Introduction to Random Matrices* (Anderson, Guionnet, Zeitouni, 2009) pour une lecture complémentaire.

Table des matières

1 Étude macroscopique des matrices de Wigner par la méthode des moments	3
1.1 Définitions	3
1.2 Preuve du théorème	4
1.3 Variantes du théorème	7
2 Limite macroscopique pour le GUE par la méthode de la résolvante, la transformée de Stieljes	9
2.1 Définition	9
2.2 Transformée de Stieljes	10
2.3 Étude en espérance	11
2.4 Concentration	12
3 Étude microscopique du GUE	15
3.1 Étude en dimension N fixée	15
3.2 Mesure de Haar	16
3.3 Loi du spectre	17
3.4 Point de vue de Dyson	19
3.5 Polynômes d'Hermite	20
3.6 Comportement asymptotique	21

4	Détection de communautés, concentration de la norme d'une matrice sous-gaussienne	25
4.1	Détection de communautés : Stochastic Block Model	25
4.2	Contrôle de la norme d'une matrice à entrées sous-gaussiennes	28

Remarque 1 :

Par la suite, on s'intéressera tout particulièrement à des matrices hermitiennes dans $H_N(\mathbb{C})$, symétriques dans $S_N(\mathbb{R})$, et les matrices unitaires $U_N(\mathbb{C})$.

La question qui va nous intéresser tout du long sera le comportement du spectre, des valeurs et vecteurs propres, lorsque $N \rightarrow \infty$. On verra que les modèles complexes sont souvent plus faciles à manipuler que les réels.

Historiquement, ce sujet a émergé dans les années 1920, avec les matrices de Wishart, du côté des statistiques. On s'intéresse à N variables (Y_i) iid centrées, à valeurs dans \mathbb{R}^p . On cherche à estimer $C \in M_p(\mathbb{R})$ la matrice de covariance de la loi. Si on note $Y = (Y_1, \dots, Y_N) \in M_{p,N}(\mathbb{R})$, alors $X = \frac{1}{N}YY^T$ est un estimateur de C , avec convergence p.s. par la loi des grands nombres.

Ce sujet s'est plus tard popularisé chez les physiciens, avec les matrices de Wigner. Soit ainsi $X^N = \frac{1}{\sqrt{N}}(X_{i,j})_{1 \leq i,j \leq N} \in S_n(\mathbb{R})$, qui représente l'opérateur d'énergie d'un gros atome. Le spectre de X est réel, et on peut l'écrire $\lambda_1^N \geq \dots \geq \lambda_N^N$. Supposons que les $X_{i,j}$ sont iid, centrés réduits. Alors on a le théorème de Wigner, la convergence de $\frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j^N}$ vers la probabilité semi-circulaire σ , de densité $\frac{1}{2\pi} \sqrt{4-x^2}$ sur $[-2, 2]$. On montrera ce résultat avec des outils combinatoires, en mettant en lien les moments des lois empiriques avec des énumérations d'arbres plans.

Dans les années 1970, ce genre de raisonnements a créé d'autres types de matrices aléatoires en lien avec d'autres objets combinatoires.

Dans les années 1990, ces matrices sont apparues dans le cadre de la théorie non commutative des probabilités, afin d'étudier de grands groupes de façon probabiliste.

Récemment, les matrices aléatoires ont naturellement trouvé leur place au sein des statistiques en grande dimension. Considérons G un graphe à $2N$ sommets, répartis en deux communautés C_1 et C_2 à N individus chacune. On relie deux individus de la même communauté avec probabilité q , et deux individus de communautés différentes avec probabilité p . Implicitement, $p < q$. À partir de la matrice d'adjacence du graphe obtenu, on peut alors chercher à identifier à quelle communauté appartient un individu, à partitionner les sommets du graphe.

1 Étude macroscopique des matrices de Wigner par la méthode des moments

1.1 Définitions

La méthode des moments est robuste, s'adapte dans d'autres contextes, et est proche de l'approche originale de Wigner. Il existe d'autres preuves du résultat, en passant par l'étude d'une résolvante, ou bien via des polynômes orthogonaux.

Le terme *macroscopique* caractérise la nature globale de l'approche, qui considère le spectre entier et pas simplement λ_1 ou le rayon spectral. On peut aussi l'opposer à l'étude *microscopique*, qui caractérise souvent l'étude de l'écart entre des valeurs propres consécutives.

Définition 2 (Matrice de Wigner) :

On dit que $(W^N)_{N \in \mathbb{N}}$ est une famille de matrices de Wigner si $W^N \in S_N(\mathbb{R})$ est symétrique réelle, à coefficients centrés réduits, indépendants deux à deux. On suppose aussi que les moments des coefficients sont uniformément bornés :

$$\forall k \in \mathbb{N}, \sup_{N \in \mathbb{N}} \sup_{1 \leq i, j \leq N} \mathbb{E} \left[|W_{i,j}^N|^k \right] < \infty.$$

Théorème 3 :

Soient W une famille de matrices de Wigner, et $X^N = \frac{1}{\sqrt{N}} W^N$ la matrice renormalisée. Notons $\lambda_1^N \geq \dots \geq \lambda_N^N$ le spectre de X^N .

Pour toute fonction $f \in \mathcal{C}_b^0(\mathbb{R})$, on a $\mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N f(\lambda_j^N) \right] \xrightarrow[N \rightarrow \infty]{} \sigma(f)$, où σ est la loi semi-circulaire. On a de plus la mesure empirique $\frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j^N}$ qui converge en loi vers σ p.s. On montrera cette deuxième partie dans le cas où les (W^N) sont deux à deux indépendants.

Démonstration. Voir infra. □

Remarque 4 (Justification de la normalisation) :

Si $f(x) = x^2$, alors $\mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N \lambda_j^N(X)^2 \right] = \frac{1}{N^2} \mathbb{E} \left[\sum_{j=1}^N \lambda_j^N(W)^2 \right] = \frac{1}{N^2} \mathbb{E} \left[\text{Tr} \left((W^N)^2 \right) \right] = 1$. Le facteur de normalisation pour passer de W à X sert naturellement à stabiliser cette espérance, à empêcher qu'elle parte vers 0 ou ∞ .

1.2 Preuve du théorème

Commençons par étudier des propriétés combinatoires du système.

Lemme 5 :

Soit $G = (S, A)$ connexe non vide. Alors $|S| \leq |A| + 1$, avec égalité ssi G est un arbre.

Démonstration. On procède par récurrence sur $|A|$. Naturellement, si $|A| = 0$, G est forcément un sommet isolé, un arbre. Sinon, on considère une arête $a \in A$ quelconque qu'on retire de G .

Si on obtient G' connexe, alors on a l'inégalité voulue sur G' donc sur G , par hypothèse de récurrence ($S' = S$ et $A' = A \setminus \{a\}$). Cette inégalité est stricte dans le cas de G , mais ajouter a à G' garantit l'existence d'un cycle dans G , qui n'est pas un arbre.

Sinon, on obtient deux sous-graphes connexes G_1 et G_2 , et à nouveau par hypothèse de récurrence, on propage le résultat dans ce cas. \square

Définition 6 (Arbre plan) :

Un arbre plan est un arbre plongé dans \mathbb{R}^2 , enraciné en 0, vu à homéomorphisme du demi-plan (qui préserve l'orientation) près. On note C_p le nombre de tels arbres à p arêtes.

Alternativement, on peut voir un tel arbre comme un arbre où les fils de chaque sommet sont totalement ordonnés. Ainsi, $(\varepsilon, 1, 2, 21)$ et $(\varepsilon, 1, 12, 2)$ ne correspondent pas au même arbre, car il faut une symétrie pour passer de l'un à l'autre, même si le graphe sous-jacent est le même.

Proposition 7 (Nombres de Catalan) :

La suite (C_p) vérifie $C_0 = 1$ et $C_{p+1} = \sum_{k=0}^p C_k C_{p-k}$.

Si on considère $C(z) = \sum_{p \in \mathbb{N}} C_p z^p$ la série génératrice, c'est l'unique série formelle solution de $zC^2 - C + 1 = 0$ et $C(0) = 1$. En résolvant l'équation différentielle, on a $C(z) = \frac{1 - \sqrt{1-4z}}{2z}$. On en déduit donc que $C_p = \frac{1}{p+1} \binom{2p}{p}$.

Démonstration. Considérons l'arête a reliée à la racine la plus à gauche d'un arbre T à $p+1$ arêtes.

Si on la retire, on obtient deux arbres plans à k et $p-k$ arêtes. Inversement, tout tel couple d'arbres plans donne un arbre T à $p+1$ arêtes. On a donc une bijection naturelle, qui donne $C_{p+1} = \sum_{j=0}^p C_j C_{p-j}$. En initialisant $C_0 = 1$, cette relation définit entièrement la suite.

On vérifie aisément que l'équation donnée sur la série formelle C correspond à cette relation, et que la fonction analytique $z \mapsto \frac{1 - \sqrt{1-4z}}{2z}$ est solution également. \square

Lemme 8 :

Sous la loi semi-circulaire, on a $\int x^k d\sigma(x) = 0$ si k impair, et $C_{\frac{k}{2}}$ si k pair.

Démonstration. Par symétrie de la loi, le cas impair est évident, et naturellement, avec $x^0 = 1$, on a $\sigma(1) = 1 = C_0$. En outre, par intégration par parties :

$$\begin{aligned} \int_{-2}^2 x^{2(p+1)} d\sigma(x) &= \int_{-2}^2 x^{2p+1} \times \frac{x\sqrt{4-x^2}}{2\pi} dx \\ &= \frac{2p+1}{2\pi} \int_{-2}^2 x^{2p} \frac{4-x^2}{3} \sqrt{4-x^2} dx \\ &= \frac{2p+1}{3} \left(4 \int_{-2}^2 x^{2p} d\sigma(x) - \int_{-2}^2 x^{2(p+1)} d\sigma(x) \right). \end{aligned}$$

Il suffit alors de vérifier qu'on a la relation de récurrence $\frac{C_{p+1}}{C_p} = \frac{2(2p+1)}{p+2}$. □

On a désormais tous les ingrédients nécessaires pour montrer le théorème.

Proposition 9 :

Pour tout $k \in \mathbb{N}$, on a $\mathbb{E} \left[\int x^k d \left(\frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j^N} \right) \right] \rightarrow \int x^k d\sigma(x)$.

Démonstration. Naturellement, l'intégrale de x^k contre la mesure empirique est en fait égale à la trace de la matrice $(X^N)^k$, avec un facteur $\frac{1}{N}$. On s'est donc ramené à :

$$\frac{1}{N} \mathbb{E} \left[\text{Tr} \left((X^N)^k \right) \right] = \frac{1}{N} \sum_{1 \leq i_1, \dots, i_k \leq N} \mathbb{E} [X_{i_1, i_2}^N \times \dots \times X_{i_k, i_1}^N] =: \frac{1}{N} \sum_{\vec{i} \in \llbracket 1, N \rrbracket^k} \mathbb{E} [T_{\vec{i}}].$$

On aimerait ainsi comprendre la contribution de $\mathbb{E} [T_{\vec{i}}]$ en fonction du vecteur \vec{i} .

Pour ce faire, on va associer à \vec{i} un graphe orienté, pour lequel \vec{i} correspond à un cycle qui visite toutes les arêtes : $G_{\vec{i}} = (S_{\vec{i}}, A_{\vec{i}})$, où $S_{\vec{i}} = \{i_j, 1 \leq j \leq k\} \subset \llbracket 1, N \rrbracket$, et $A_{\vec{i}} = \{(i_j, i_{j+1}), 1 \leq j \leq k\}$. Ce graphe induit naturellement un graphe non-orienté, avec les arêtes $A_{\vec{i}}$. On a donc l'égalité $T_{\vec{i}} = \prod_{u \in A_{\vec{i}}} X_u^{N\rho(u)}$, où $\rho(u)$ est le nombre de passages sur l'arête u par le cycle décrit par \vec{i} .

Par indépendance entre les termes du produit, si un des $\rho(u)$ est égal à 1, on a $\mathbb{E} [T_{\vec{i}}] = 0$. On se ramène donc à l'étude du cas où chaque arête est visitée *au moins* deux fois. Comme \vec{i} visite k arêtes en tout, le graphe non-orienté a donc *au plus* $\frac{k}{2}$ arêtes.

Par construction, le cycle \vec{i} visite tous les sommets, donc $G_{\vec{i}}$ est connexe. Par le lemme initial, $|S_{\vec{i}}| \leq \frac{k}{2} + 1$. Considérons une partie de la somme, correspondant au cas où l'inégalité précédente est stricte. Par inégalité de Hölder, on a la majoration :

$$\frac{1}{N} \sum_{\vec{i} \text{ tels que } \dots} \mathbb{E} [T_{\vec{i}}] \leq \sum_{l < \frac{k}{2} + 1} \sum_{|S_{\vec{i}}| = l} \mathbb{E} [T_{\vec{i}}] \leq \frac{1}{N} \sum_{l < \frac{k}{2} + 1} \binom{N}{l} l^k \frac{A_k}{\sqrt{N}^k} = o \left(\frac{1}{\sqrt{N}} \right).$$

Les seuls termes à prendre en compte correspondent donc au cas d'égalité. Dans ce cas, chaque arête est parcourue *exactement* deux fois, et le graphe est un arbre. Dans ce cas, $\mathbb{E} [T_{\vec{i}}] = \frac{1}{\sqrt{N}^k}$. Reste donc à compter le nombre de \vec{i} qui donnent un tel arbre. Quitte à voir le chemin \vec{i} comme un parcours en profondeur de l'arbre, au choix d'étiquette près. On peut donc mettre en bijection ces chemins avec les arbres plans enracinés étiquetés. On obtient donc :

$$\frac{1}{N} \mathbb{E} \left[\text{Tr} \left((X^N)^k \right) \right] = \frac{1}{N} \times \frac{C_{\frac{k}{2}}}{\sqrt{N}^k} \times \prod_{j=1}^{\frac{k}{2}+1} (N - j + 1) + o \left(\frac{1}{\sqrt{N}} \right) = \sigma(x^k) + o(1),$$

avec l'abus de notation $C_{\frac{k}{2}} = 0$ lorsque k impair. □

Corollaire 10 :

Par linéarité, on a cette convergence en moyenne pour tout polynôme $P \in \mathbb{R}[X]$. On étend alors le résultat aux fonctions continues bornées.

Démonstration. Par le théorème de densité de Weierstrass, par densité des polynômes pour $\|\cdot\|_\infty$ sur $\mathcal{C}^0([-3, 3])$, on approche f par P à ε près. On majore alors :

$$\begin{aligned} \left| \mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N f(\lambda_j^N) - \sigma(f) \right] \right| &\leq 2\varepsilon + \left| \mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N P(\lambda_j^N) \right] - \sigma(P) \right| \\ &+ \left| \mathbb{E} \left[\int_{|x| \geq 3} (f - P) d \left(\frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j^N} - \sigma \right) \right] \right|. \end{aligned}$$

Le terme central tend vers 0 par ce qui précède.

Pour le terme de droite, comme σ est à support dans $[-2, 2]$, on peut oublier cette mesure. Comme $f - P = O(x^{2k})$ pour un certain $k \in \mathbb{N}$, on se retrouve simplement à vouloir majorer $\mathbb{E} \left[\int_{|x| \geq 3} x^{2k} d \left(\frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j^N} \right) \right]$, dont on peut vérifier qu'il converge vers 0. \square

On obtient ainsi la convergence en espérance annoncée. Pour ce qui concerne la convergence p.s. on utilise le lemme ci-après, pour conclure ensuite par une méthode des moments.

Lemme 11 :

Pour tout $k \in \mathbb{N}$, on a $\text{Var} \left(\frac{1}{N} \text{Tr} \left((X^N)^k \right) \right) = O\left(\frac{1}{N^2}\right)$.

Démonstration. Admis. \square

1.3 Variantes du théorème

Le résultat précédent n'utilisait pas la même loi μ sur chaque coefficient $W_{i,j}$, mais requiert un contrôle sur tous les moments. Il existe d'autres variantes, comme on le voit ci-dessous.

Théorème 12 :

Soient μ une loi sur \mathbb{C} et ν une loi sur \mathbb{R} . Soient les $W^N \in H_n(\mathbb{C})$ hermitiennes à coefficients indépendants, avec $H_{i,j} \sim \mu$ si $i \neq j$ et $H_{i,i} \sim \nu$.

Supposons également que $\nu(x^2) = \mu(|x|^2) = 1$, que $\nu(x) = \mu(x) = 0$.

Dans ce cas spectre réel pour $\frac{1}{\sqrt{N}} W^N$, et $\frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j^N} \xrightarrow{\text{p.s.}} \sigma$.

On peut aussi pousser ces méthodes pour étudier λ_1^N , le rayon spectral.

Théorème 13 :

Si de plus $\mu(|x|^{4+\varepsilon}) < \infty$, alors $\lambda_1^N \xrightarrow{\text{p.s.}} 2$.

Si on s'affranchit des propriétés de symétrie, on peut obtenir d'autres résultats sur le spectre, qui n'est alors plus réel presque-sûrement.

Théorème 14 :

Soit μ une loi sur \mathbb{C} , centrée réduite. On considère une matrice G^N à coefficients iid, de loi μ . Alors la mesure empirique sur le spectre de G^N converge presque-sûrement vers la mesure uniforme sur le disque unité, pour $N \rightarrow \infty$.

2 Limite macroscopique pour le GUE par la méthode de la résolvante, la transformée de Stieljes

2.1 Définition

Définition 15 (Modèle unitaire gaussien) :

Dans ce modèle, $X^N \in H_N(\mathbb{C})$ est une matrice à coefficients indépendants, tels que :

- $X_{i,i} \sim \mathcal{N}\left(0, \frac{1}{N}\right)$,
- pour $i < j$, $\text{Re}(X_{i,j})$ et $\text{Im}(X_{i,j})$ sont indépendantes, de loi $\mathcal{N}\left(0, \frac{1}{2N}\right)$.

La normalisation est choisie telle que $\mathbb{E}\left[|X_{i,j}^N|^2\right] = 1$ indépendamment du choix de i et j .

Remarque 16 :

Notons que sans valeur absolue, $\mathbb{E}\left[(X_{i,j}^N)^2\right] = 0$. On a donc la covariance $\mathbb{E}[X_{i,j}^N X_{k,l}^N] = \frac{\delta_{i,l} \delta_{k,l}}{N}$.

En voyant X^N comme un vecteur gaussien, ces covariances permettent de caractériser sa loi.

Lemme 17 :

La loi GUE est invariante par conjugaison $X \mapsto U^* X U$ par une matrice unitaire $U \in H_N(\mathbb{C})$.

Démonstration. On peut voir la matrice aléatoire $U^* X U$ comme un vecteur gaussien centré. On peut expliciter sa covariance par un calcul direct, ce qui termine la preuve. \square

Remarque 18 :

On pourrait également s'intéresser au modèle orthogonal gaussien (GOE), réel, mais cet autre modèle possède en fait moins de propriétés de symétrie et est plus difficile à étudier.

Remarque 19 :

On peut munir $H_N(\mathbb{C})$ de la mesure de Lebesgue λ , en le voyant comme un espace vectoriel réel de dimension $N + 2 \times \frac{N(N-1)}{2} = N^2$, muni de la base $(\Delta_{k,k})_{1 \leq k \leq N}$ pour la diagonale réelle, des parties réelles $\left(\frac{\Delta_{k,l} - \Delta_{l,k}}{2}\right)_{1 \leq k < l \leq N}$, et des parties imaginaires analogues.

On peut alors considérer la densité de X^N par cette mesure :

$$\begin{aligned} d\mu(X^N) &= C_N \prod_{k=1}^N e^{-NX_{i,i}^2/2} \prod_{k < l} e^{-N\operatorname{Re}(X_{k,l})^2 - N\operatorname{Im}(X_{k,l})^2} d\lambda(X^N) \\ &= C_N \prod_{1 \leq k, l \leq N} e^{-N|X_{k,l}|^2/2} d\lambda(X^N) \\ &= C_N e^{-N\operatorname{Tr}(XX^*)/2} d\lambda(X^N). \end{aligned}$$

Théorème 20 :

Pour un GUE, la mesure empirique $\hat{\mu}_N$ du spectre de X^N converge p.s. en loi vers la loi semi-circulaire.

2.2 Transformée de Stieljes

Définition 21 (Transformée de Stieljes) :

Pour une mesure réelle finie μ , on définit sa transformée S_μ sur $\mathbb{C} \setminus \mathbb{R}$ via $S_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-t} d\mu(t)$.

Lemme 22 :

L'application S_μ est holomorphe sur $\mathbb{C} \setminus \operatorname{Supp}(\mu)$, et $|S_\mu(z)| \leq \frac{\mu(\mathbb{R})}{|\operatorname{Im}(z)|}$.

Démonstration. On utilise le théorème d'holomorphic sous l'intégrale. □

Proposition 23 :

Si μ est à support compact, inclus dans $[-R, R]$, alors lorsque $|z| > R$, on peut décomposer $S_\mu(z) = \frac{1}{z} \sum_{k \geq 0} \frac{1}{z^k} \int t^k d\mu(t)$.

Démonstration. On utilise le fait que $\frac{1}{z-t} = \frac{1}{z} \times \frac{1}{1-\frac{t}{z}} = \frac{1}{z} \times \sum_{k \in \mathbb{N}} \left(\frac{t}{z}\right)^k$, puis on échange la somme et l'intégrale. □

Proposition 24 :

On peut reconstruire la mesure μ à partir de S_μ . Plus précisément, sur un intervalle borné I tel que $\mu(\partial I) = 0$, on a $\mu(I) = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0^+} \int_I \operatorname{Im}(S_\mu)(x + i\varepsilon) dx$. On en déduit que cette transformation est injective.

Dans ce cas, la convergence faible est équivalente à la convergence simple de la transformée sur $\mathbb{C} \setminus \mathbb{R}$.

Démonstration. Supposons que $\mu(\mathbb{R}) = 1$. On a :

$$\begin{aligned}
\operatorname{Im}\left(\int_I S_\mu(x + i\varepsilon) dx\right) &= \operatorname{Im}\left(\int_I \int_{\mathbb{R}} \frac{1}{x-t+i\varepsilon} d\mu(t) dx\right) \\
&= \int_I \int_{\mathbb{R}} \operatorname{Im}\left(\frac{x-t-i\varepsilon}{(x-t)^2+\varepsilon^2}\right) d\mu(t) dx \\
&= -\int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_I(x) \frac{\varepsilon}{(x-t)^2+\varepsilon^2} d\mu(t) dx \\
&= -\pi \int_{\mathbb{R}^2} \mathbf{1}_I(t + \varepsilon y) d\mu(t) \frac{1}{\pi(1+y^2)} dy.
\end{aligned}$$

Le terme qui apparaît ainsi à droite est une variable Y loi de Cauchy. On a donc ainsi obtenu $-\pi\mathbb{P}(X + \varepsilon Y \in I)$, qui converge vers $-\pi\mathbb{P}(X \in I)$ sous les hypothèses de l'énoncé. On adapte alors le résultat à toute mesure $\mu(\mathbb{R}) < \infty$.

Si on connaît S_μ , on peut obtenir $\mu * \mathcal{C}_\varepsilon$, donc en passant aux fonctions caractéristiques, on caractérise $\Phi_{\mu * \mathcal{C}_\varepsilon} = \Phi_\mu \Phi_{\mathcal{C}_\varepsilon}$, et cette fonction ne s'annule pas, on caractérise entièrement Φ_μ et donc μ .

Pour la convergence faible, on peut en particulier constater la convergence contre la fonction continue $f : t \mapsto \frac{1}{z-t}$, d'où la convergence simple des transformées. Réciproquement, si on a la convergence simple des transformées vers S_μ , toute valeur d'adhérence de la suite est forcément égale à μ par le sens direct, et par Banach-Alaoglu, on a une telle valeur d'adhérence. \square

Remarque 25 :

La loi semi-circulaire σ est à support dans $[-2, 2]$. Par calcul explicite, si $|z| > 2$, on a $S_\sigma(z) = \frac{1}{z} \sum_{k \geq 0} \frac{1}{z^k} \int t^k d\sigma(t) = \frac{1}{z} \sum_{k=0}^{\infty} \frac{C_k}{z^{2k}} = \frac{1}{z} C\left(\frac{1}{z^2}\right)$, avec (C_k) les nombres de Catalan et C la série génératrice associée.

On vérifie que cette transformée est caractérisée par $S^2 - zS + 1 = 0$ sur $\mathbb{C} \setminus \mathbb{R}$, et $S \xrightarrow{|z| \rightarrow \infty} 0$.

On va donc par la suite montrer que $S_{\hat{\mu}_N}^2 - zS_{\hat{\mu}_N} - 1 \rightarrow 0$, ce qui permettra de conclure la preuve du théorème.

2.3 Étude en espérance

Lemme 26 :

Soient $Z \sim \mathcal{N}(0, \sigma^2)$, et $f \in \mathcal{C}^1(\mathbb{R})$ telle que f' est à croissance sous-polynômiale. Alors $\mathbb{E}[Zf(Z)] = \sigma^2 \mathbb{E}[f'(Z)]$.

Démonstration. Il suffit de faire une IPP dans l'écriture explicite de l'intégrale à densité. \square

On aimerait désormais adapter cette propriété dans le cas où X est une matrice.

Définition 27 :

Soit $X = U^* \text{diag}(\lambda_i) U$ hermitienne. On étend $f : \mathbb{R} \rightarrow \mathbb{C}$ aux matrices hermitiennes via $f(X) = U^* \text{diag}(f(\lambda_i)) U$. Si f est un polynôme, cette définition coïncide naturellement avec les polynômes de matrices usuels.

Proposition 28 :

On a $\mathbb{E}[X_{i,j}^N \times X^N] = \frac{\Delta_{j,i}}{N}$. On propage alors ce résultat aux polynômes, ce qui nous permet finalement d'obtenir $\mathbb{E}[X_{i,j}^N \times \frac{1}{z-X^N}] = \frac{1}{N} \mathbb{E}[\frac{1}{z-X^N} \Delta_{j,i} \frac{1}{z-X^N}]$.

Corollaire 29 :

On a :

$$\begin{aligned} \mathbb{E}[S_{\hat{\mu}_N}(z)] &= \frac{1}{N} \mathbb{E} \left[\sum_{j=1}^N \frac{1}{z-\lambda_j^N} \right] \\ &= \frac{1}{z} + \frac{1}{Nz} \mathbb{E} \left[\text{Tr} \left(X^N \frac{1}{z-X^N} \right) \right] \\ &= \frac{1}{z} + \frac{1}{z} \mathbb{E} \left[\left(\frac{1}{N} \text{Tr} \left(\frac{1}{z-X^N} \right) \right)^2 \right] \\ &= \frac{1}{z} \left(1 + \mathbb{E} \left[S_{\hat{\mu}_N}^2(z) \right] \right). \end{aligned}$$

Autrement dit, on a montré que $\mathbb{E}[S_{\hat{\mu}_N}(z)^2 - zS_{\hat{\mu}_N}(z) + 1] = 0$.

2.4 Concentration

Supposons pour l'instant que $\text{Var}(S_{\hat{\mu}_N}(z)) \leq \frac{C_z}{N^2}$. On peut donc remplacer l'espérance du carré par le carré de l'espérance sans changer la limite.

En outre $\mathbb{E}[S_{\hat{\mu}_N}(z)]$ est localement borné par $\frac{1}{\text{Im}(z)}$, holomorphe, on peut en extraire des sous-suites convergentes par le théorème de Montel. Toute telle limite holomorphe S vérifie $S^2 - zS + 1 = 0$, et $|S(z)| \leq \frac{1}{|\text{Im}(z)|}$, donc c'est nécessairement la solution $S = S_\sigma$ du système.

La suite $\mathbb{E}[S_{\hat{\mu}_N}]$ a un unique point d'accumulation S_σ . En outre, avec probabilité $1 - O(\frac{1}{N^2})$, par inégalité de Chebychev, $S_{\hat{\mu}_N}(z)$ est à distance au plus t de sa moyenne, donc par Borel-Cantelli, $\overline{\text{lim}} |S_{\hat{\mu}_N}(z) - \mathbb{E}[S_{\hat{\mu}_N}(z)]| \leq t$ avec probabilité 1. Ceci est vrai pour tout t , d'où $S_{\hat{\mu}_N}(z) \rightarrow S_\sigma(z)$ p.s. Le résultat est vrai à z fixé, donc conjointement vrai pour une suite dénombrable dense, et finalement pour tout z .

On doit donc montrer cette propriété de concentration.

Proposition 30 (Talagrand) :

Soient X_1, \dots, X_n des gaussiennes indépendantes, dont les variances sont majorées par σ^2 , et une application Λ -lipschitzienne $F : \mathbb{R}^n \rightarrow \mathbb{R}$.

Dans ce cas, $\mathbb{P}(|F(X) - \mathbb{E}[F(X)]| > t) \leq 2 \exp\left(-c \frac{t^2}{\sigma^2 \Lambda^2}\right)$.

Démonstration. Par densité, on se ramène à une fonction \mathcal{C}^1 . Quitte à normaliser, $\sigma = \Lambda = 1$.

Quitte à augmenter la variance, on suppose que toutes les gaussiennes ont une variance 1. Quitte à translater F , on suppose que l'espérance est nulle. On veut alors majorer $\mathbb{P}(F(X) > t)$. Le facteur 2 sortira en considérant cette inégalité pour F et $-F$.

Soit \tilde{X} une copie indépendante de X . Par Jensen, on a $1 = e^{-u\mathbb{E}[F(X)]} \leq \mathbb{E}[e^{-uF(X)}]$, donc :

$$\mathbb{E}[e^{uF(X)}] \leq \mathbb{E}\left[e^{u(F(X)-F(\tilde{X}))}\right].$$

On interpole $X(\theta) = \sin(\theta)X + \cos(\theta)\tilde{X}$, qui correspond à \tilde{X} en 0 et à X en $\frac{\pi}{2}$. On étudie donc $\exp(u(F(X(\frac{\pi}{2}) - F(X(0)))) = \frac{u\pi}{2} \int_0^{\frac{\pi}{2}} \partial_\theta F(X(\theta)) \frac{d\theta}{\pi/2}$. On utilise à nouveau Jensen, ce qui donne donc :

$$\mathbb{E}[e^{uF(X)}] \leq \int_0^{\frac{\pi}{2}} \mathbb{E}\left[\exp\left(u\frac{\pi}{2}\langle \nabla_{X(\theta)} F, \partial_\theta X(\theta) \rangle\right)\right] \frac{2}{\pi} d\theta.$$

Notons que $\partial_\theta X_i(\theta) = \cos(\theta)X_i - \sin(\theta)\tilde{X}_i$ est une gaussienne de variance $\cos^2 + \sin^2 = 1$. On vérifie alors que $(X(\theta), \partial_\theta X(\theta))$ est un vecteur gaussien, et que les deux termes sont indépendants. On peut alors majorer l'espérance dans l'intégrale précédente par :

$$\mathbb{E}\left[\exp\left(u\frac{\pi}{2}\langle \nabla_{X(\theta)} F, \partial_\theta X(\theta) \rangle\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\exp\left(u\frac{\pi}{2}\langle \nabla_{X(\theta)} F, \partial_\theta X(\theta) \rangle\right) \middle| X(\theta)\right]\right] \leq e^{u^2\pi^2/8}.$$

On utilise alors l'inégalité de Chebychev exponentielle, puis on optimise le paramètre u pour obtenir la borne souhaitée, avec la constante $c = \frac{2}{\pi^2}$. \square

Remarque 31 :

Avec $F = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$ une application 1-lipschitz, on vérifie l'optimalité de l'inégalité à n fixé.

Il nous reste à justifier que l'application qui nous intéresse est effectivement Lipschitz.

Lemme 32 (Hoffman Wielandt) :

Soient λ^X et λ^Y les spectres ordonnés de X et Y hermitiennes. On a $\|\lambda^X - \lambda^Y\|_2 \leq \|X - Y\|_2$, où la norme de droite correspond à la norme euclidienne en dimension N^2 , qu'on peut réécrire $\sqrt{\text{Tr}((X - Y)^2)}$.

Démonstration. On peut également réécrire le terme de gauche comme $\sqrt{\text{Tr}((\Lambda^X - \Lambda^Y)^2)}$, avec des matrices diagonales. Quitte à développer les carrés, par linéarité, on veut de façon équivalente montrer que $\text{Tr}(XY) \leq \sum_{i=1}^N \lambda_i^X \lambda_i^Y$.

Si X est déjà diagonale, alors $\text{Tr}(XY) = \sum_{i,j} \lambda_i^X \lambda_j^Y |V_{i,j}|^2$. Si V est l'identité on a l'inégalité souhaitée. Dans le cas général, on a en particulier une matrice bistochastique, combinaison convexe de permutations. Il suffit de trouver la permutation de coefficients qui maximise cette somme. Si la permutation n'est pas l'identité (dans le cas où les valeurs propres sont strictement ordonnés), quitte à intervertir deux coefficients dont l'ordre est interverti, on augmente le résultat,

d'où la majoration.

Il en va de même si X n'est pas diagonale, en conjuguant X et Y conjointement. \square

Lemme 33 (Klein) :

Si $f : \mathbb{R} \rightarrow \mathbb{C}$ est Λ -Lipschitz, alors la fonction $X \mapsto \frac{1}{N} \text{Tr}(f(X))$ définie sur les matrices hermitiennes est $\frac{\Lambda}{\sqrt{N}}$ -Lipschitz.

Démonstration. On applique directement le résultat précédent. \square

Comme $S_{\hat{\mu}_N}(z) = \frac{1}{N} \text{Tr}\left(\frac{1}{z - X^N}\right)$, du point de vue matriciel on a une fonction $\frac{\Lambda_z}{\sqrt{N}}$ -Lipschitz, avec $\Lambda_z = \frac{1}{|\text{Im}(z)|^2}$ le facteur Lipschitz de l'application $x \mapsto \frac{1}{z-x}$. On peut donc utiliser la propriété de concentration de Talagrand, puis enfin majorer la variance par l'intégrale de la probabilité de s'écarter de la moyenne pour obtenir une majoration sur la variance comme annoncé.

3 Étude microscopique du GUE

3.1 Étude en dimension N fixée

Définition 34 :

On considère ici une autre variante du GUE non normalisée : $X_{i,i}^N \sim \mathcal{N}(0, 1)$ sur la diagonale, et $X_{i,j}^N \sim \mathcal{N}(0, \frac{1}{2}) + \sqrt{-1}\mathcal{N}(0, \frac{1}{2})$ de façon symétrique.

Ce faisant, les valeurs propres se localisent dans l'intervalle $[\pm 2\sqrt{N}]$.

Théorème 35 :

Notons $\Delta(\lambda) = \prod_{i < j} (\lambda_j - \lambda_i)$ le déterminant de Vandermonde de $\lambda \in \mathbb{R}^N$. On considère alors la variable λ à densité $\Delta(\lambda)^2 e^{-\|\lambda\|_2^2/2} d\lambda$. Soit de plus $U \sim \text{Haar}(U_N(\mathbb{C}))$, une matrice unitaire uniforme, indépendante de λ . Alors $U \text{diag}(\lambda) U^*$ a la loi du GUE non normalisé.

Démonstration. Voir infra. □

On va par la suite chercher à prouver ce résultat.

Remarque 36 :

On n'a pas unicité de la décomposition (U, λ) de X . En effet, on peut permuter les coordonnées de λ , ou remplacer U par $U \text{diag}(\omega)$ avec $\omega \in \mathbb{U}^N$.

Remarque 37 :

D'après ce théorème, les valeurs propres se repoussent, car la densité s'annule lorsque $\lambda_i = \lambda_j$.

Théorème 38 :

Soit $V : \mathbb{R} \rightarrow \mathbb{R}$ tel que $\lim_{x \rightarrow \pm\infty} \frac{V(x)}{2 \ln(1+|x|)} > 1$.

Si X^N est une matrice à densité $e^{-\text{Tr}(V(X^N))} dX^N$ renormalisée par Z_V^N , alors on a la même propriété de décomposition, avec un spectre λ à densité $\Delta(\lambda)^2 e^{-\sum_{i=1}^N V(\lambda_i)} d\lambda$.

Théorème 39 :

Pour une matrice X cette fois-ci symétrique réelle, on a la décomposition $O \text{diag}(\lambda) O^T$, avec $O \sim \text{Haar}(O_N(\mathbb{R}))$, et λ à densité $|\Delta(\lambda)| e^{-\sum V(\lambda_i)} d\lambda$ normalisée.

3.2 Mesure de Haar

Définition 40 (Mesure de Haar) :

Soit G un groupe topologique. Une mesure de Haar sur G est une mesure μ invariante sous l'action du groupe, telle que $\mu(gA) = \mu(A)$ pour tout borélien A et tout élément $g \in G$.

Remarque 41 :

C'est en particulier le cas de la mesure de Lebesgue sur \mathbb{R}^d , ou \mathbb{T}^d . C'est aussi le cas du groupe multiplicatif \mathbb{R}^{+*} pour la mesure à densité $\frac{1}{x} dx$.

Sur un groupe fini, la mesure de comptage et la mesure uniforme conviennent.

Théorème 42 :

Si G est localement compact, alors il existe une unique mesure de Haar, à une constante multiplicative près. Si de plus G est compact, il existe une mesure de Haar finie.

Ce théorème d'existence est abstrait, on va ici remonter certaines propriétés dans le cas qui nous intéresse.

Proposition 43 :

Si G est compact, la probabilité de Haar est unique, et invariante par translation à droite.

Démonstration. Soit μ une probabilité de Haar. Remarquons que :

$$\mu(A) = \int \mu(gA) d\mu(g) = \int \int \mathbf{1}_{z \in gA} d\mu(z) d\mu(g) = \int \mathbf{1}_{g \in zA^{-1}} d\mu(g) d\mu(z) = \mu(A^{-1}).$$

En conséquence, $\mu(Ag) = \mu(g^{-1}A^{-1}) = \mu(A^{-1}) = \mu(A)$. Ceci conclut la translation à droite.

Pour l'unicité, considérons une autre mesure de Haar ν . Alors $\mu(A) = \int \mu(gA) d\nu(g)$. En développant les calculs comme précédemment, $\mu(A) = \nu(A^{-1})$, ce qui conclut la preuve. \square

Remarque 44 (Construction d'une probabilité de Haar sur $U_N(\mathbb{C})$) :

Partons de la matrice X^N à coefficients $X_{i,j}^N \sim \mathcal{N}(0, \frac{1}{2}) + \sqrt{-1}\mathcal{N}(0, \frac{1}{2})$.

On peut décomposer X en N vecteurs colonnes (X_i^N) . Soit (U_i^N) l'orthonormalisation de X par Gram-Schmidt. La matrice X est presque-sûrement inversible, donc U est presque-sûrement définie.

Soit $V \in U_N(\mathbb{C})$. On vérifie que $VX^N \stackrel{d}{=} X^N$. En outre, comme V laisse la norme invariante, on vérifie qu'orthonormaliser VX^N équivaut à appliquer V à U . On en conclut alors que U suit bien la loi de Haar.

Remarque 45 :

Une construction alternative consiste à partir de U_1^N uniforme sur la sphère unité de C^N , puis U_2^N uniforme sur la sphère de dimension $N - 2$ orthogonale à U_1^N , et ainsi de suite.

Lemme 46 :

Soit X un GUE. Presque-sûrement, les valeurs propres de X sont distinctes.

Démonstration. Quitte à ordonner les valeurs propres de façon décroissante, la k -ième valeur propre vérifie $\lambda_k = \sup_{\dim(E)=k} \inf_{y \in E, \|y\|_2=1} \langle y, X^N y \rangle$, donc ces valeurs propres sont bien des variables aléatoires en tant que fonctions mesurables de X^N . Le polynôme caractéristique P , dont les racines sont les λ_i , est lui aussi une variable aléatoire à densité pour la mesure de Lebesgue, en tant que combinaison polynomiale des $X_{i,j}^N$.

La matrice X^N a une valeur propre multiple ssi $\Delta(\lambda) = 0$. Or $\Delta(\lambda)$, en tant que fonction symétrique des λ_i , peut être autrement exprimée comme une fonction de P , et donc comme une fonction polynomiale $R(X)$ non triviale. En conséquence, $\mathbb{P}(R(X) = 0) = 0$, d'où le résultat. \square

Remarque 47 :

Le résultat précédent est plus généralement vrai dès que X a une densité pour la mesure de Lebesgue.

La propriété est nettement plus complexe à étudier si on ne considère plus des variables à densité. Ainsi, pour une matrice symétrique réelle $X_{i,j}^N \stackrel{\text{iid}}{\sim} \mathcal{B}(\frac{1}{2})$, on a des encadrements mais toujours pas d'asymptotique exacte sur $\mathbb{P}(\det(X^N) = 0)$.

Proposition 48 :

Il existe une probabilité ν sur \mathbb{R}^N , et λ de loi ν indépendamment de $U \sim \text{Haar}(U_N(\mathbb{C}))$, tel que $U \text{diag}(\lambda) U^*$ a la loi du GUE non normalisé.

Démonstration. Par le lemme précédent, le spectre strictement ordonné $(\lambda_1 > \dots > \lambda_N)$ est bien défini. Notons Λ la matrice diagonale associée.

Soit (U_i) une famille orthonormale de vecteurs propres associée. On a $X^N = U \Lambda U^*$.

Soit de plus $V \sim \text{Haar}(U_N(\mathbb{C}))$ indépendant de X , et $W = VU$. On a sans soucis la loi $W \sim \text{Haar}(U_N(\mathbb{C}))$, indépendamment de Λ cette fois-ci. En outre, $W \Lambda W^* = V X V^* \stackrel{d}{=} X$, donc on a bien un GUE. \square

3.3 Loi du spectre

Proposition 49 :

La loi ν précédemment introduite est à densité $\Delta(\lambda)^2 e^{-\sum_{i=1}^N \lambda_i^2/2} d\lambda$.

Démonstration. Soit U issue de la décomposition $U\Lambda U^*$. Soit f une fonction réelle bornée du spectre de X^N , d'où $f(U^*XU) = f(X)$. Par indépendance entre U et Λ , on peut écrire :

$$\mathbb{E}[f(X^N)] = c_\varepsilon \mathbb{E}[f(X^N) \mathbf{1}_{U \in V_\varepsilon}].$$

avec $V_\varepsilon = \{e^{iH}, H \text{ hermitien à diagonale nulle, } \|H\| \leq \varepsilon\}$. Dans ce voisinage, on peut réécrire $X^N = e^{iH} \Lambda e^{-iH} =: \psi(H, \Lambda)$ de façon difféomorphe. On a donc :

$$D_{H,\Lambda} \psi(dH, d\Lambda) = e^{-iH} (d\Lambda + i\Lambda \times dH - dH \times \Lambda) e^{iH} (1 + o(\varepsilon)).$$

Notons $\varphi(dH, d\Lambda)$ l'application au centre, en oubliant les $e^{\pm iH}$.

Notons alors $C_U : X \mapsto U^*XU$ la conjugaison par U . On a $C_U C_V = C_{UV}$. Par compacité du groupe unitaire, la suite (U^k) a un point d'accumulation V , donc $|\det(C_{U^k})|$ a un point d'accumulation en $|\det(C_V)| \in \mathbb{R}^{+*}$, ce qui implique directement $|\det(C_U)| = 1$.

On a $\varphi(0, E_{i,i}) = E_{i,i}$. En outre, $\varphi(E_{i,j} + E_{j,i}, 0) = (\lambda_i - \lambda_j)(E_{i,j} + E_{j,i})$ et de même pour la partie imaginaire. On a donc entièrement déterminé le spectre de φ . Ainsi, par la remarque précédente sur le déterminant d'une conjugaison, $|\det(D\psi)| = |\det(\varphi)| = \prod_{i < j} |\lambda_i - \lambda_j|^2 (1 + o(\varepsilon))$.

En réinjectant ce déterminant dans l'espérance initiale, par formule du changement de variable, on obtient le résultat attendu en prenant $\varepsilon \rightarrow 0$. □

Remarque 50 (Réécriture de la densité) :

Que vaut la constante de normalisation C_N dans la densité de λ ?

Quitte à développer $\Delta(\lambda)$ en tant que déterminant de la matrice de Vandermonde, quitte à utiliser des polynômes unitaires orthogonaux (P_j) , on a $\Delta(\lambda) = \sum_{\sigma} \prod_{i=1}^N \varepsilon(\sigma) P_{\sigma(i)-1}(\lambda_i)$. Les polynômes d'Hermite orthogonaux pour $\langle f, g \rangle = \int f g e^{-x^2/2} dx$ conviennent.

Avec ces polynômes, en développant le calcul, on obtient :

$$\int \Delta(\lambda)^2 e^{-\sum \lambda_i^2/2} d\lambda = \sum_{\sigma} \prod_{i=1}^N \|h_{\sigma(i)-1}\|^2 = N! \prod_{j=0}^{N-1} \|h_j\|^2.$$

3.4 Point de vue de Dyson

Définition 51 (Mouvement brownien hermitien) :

Partons du constat que $\mathcal{N}(0, 1)$ est la valeur du mouvement brownien au temps $t = 1$.

On peut donc généraliser le GUE non normalisé en un processus $X^N(t)$, où $X_{i,i}^N(t) = B_{i,i}(t)$ et $X_{i,j}^N(t) = \frac{B_{i,j}(t) + \sqrt{-1}B'_{i,j}(t)}{\sqrt{2}}$. On appelle ce processus le mouvement brownien hermitien.

Remarque 52 :

On s'intéresse alors au processus des valeurs et vecteurs propres. Plus précisément, on cherche un processus $U^N(t)$ à valeurs dans $U_N(\mathbb{C})$, et un processus $\Lambda^N(t)$ diagonal réel, tels que :

$$X^N(t)U^N(t) = U^N(t)\Lambda^N(t).$$

Admettons que ces processus sont des semi-martingales. On décompose alors le processus $U^N = M^U + \int_0^t V^U(s) ds$ entre une martingale locale et un processus à variations bornées, et de même pour Λ^N . En particulier, $dU^N = dM^U(t) + V^U(t) dt$.

Or $U^N U^{N*} = \text{Id}$, donc :

$$0 = d(UU^*) = [dM^U U^* + U dM^{U*}] + [V^U U^* dt + U^* V^U dt + d\langle M^U, M^{U*} \rangle],$$

avec le terme de gauche correspondant à une martingale locale et celui de droite à un processus à variations bornées. On en déduit que $dM^U U^{N*}$ est anti-hermitienne.

Quitte à faire un changement de base unitaire via $U^N(t_0)$, on obtient un nouveau processus qui a la même loi. on peut ainsi supposer qu'en t_0 , on a $U^N(t_0) = \text{Id}$ et $X^N(t_0) = \Lambda^N(t_0)$. Comme à tout instant on a $XU = U\Lambda$, on a :

$$dXU + X dM^U = dM^U \Lambda + U dM^\Lambda,$$

pour la partie martingale locale, et :

$$XV^U dt + d\langle X, M^U \rangle = V^U dt + \Lambda + UV^\Lambda dt + d\langle M^U, M^\Lambda \rangle,$$

pour la partie à variations finies. Dans ce cas, à t_0 , la partie martingale se simplifie en :

$$dX^N(t_0) = [dM^U \Lambda - \Lambda dM^U] + dM^\Lambda,$$

avec le terme de gauche à diagonale nulle, et celui de droite diagonal réel. D'autre part, pour la partie VF, en calculant le crochet $\langle X, M^U \rangle$:

$$V^\Lambda dt = (\Lambda V^U - V^U \Lambda) dt + \sum_{k \neq i} \frac{1}{\lambda_k - \lambda_i} dt,$$

à nouveau avec un terme gauche à diagonale nulle. En mettant bout-à-bout ces résultats, on finit par obtenir un système d'EDS stochastique :

$$d\lambda_i(t) = dB_{i,i}(t) - \sum_{k \neq i} \frac{1}{\lambda_k(t) - \lambda_i(t)} dt,$$

qui fait ressortir l'effet répulsif des valeurs propres les unes sur les autres.

Ce nouveau processus des valeurs propres est aussi appelé le mouvement brownien de Dyson, ou autrement dit la loi de N mouvement browniens réels conditionnés pour ne pas se croiser.

Remarque 53 :

On peut envisager une variante du modèle précédent, où $dY = dX - Y dt$. On peut alors vérifier que les valeurs propres vérifient des équations du type :

$$d\lambda_i(t) = dB_{i,i}(t) - \sum_{k \neq i} \frac{1}{\lambda_k(t) - \lambda_i(t)} dt - \lambda_i dt,$$

Ce cas de figure correspond en fait à la fonction $V(x) = \sum_{i < j} \ln(|x_i - x_j|) + \frac{\|x\|_2^2}{2}$.

3.5 Polynômes d'Hermite

Faisons une brève digression sur les polynômes d'Hermite, utilisés pour calculer la constante de normalisation C_N de la loi ν .

Définition 54 :

Soit $h_n = (-1)^n e^{x^2/2} \partial_x^n [e^{-x^2/2}]$. On peut voir h_n comme un polynôme.

Lemme 55 :

Le polynôme h_n est unitaire, de degré n , avec la même parité que n .

Si on se place dans $L^2(\mathbb{R})$ muni de la mesure $e^{-x^2/2} dx$, on a $\langle h_n, h_m \rangle_{L^2} = \delta_{m,n} \sqrt{2\pi n!}$.

On a l'équation $h_{n+1} = xh_n - nh_{n-1}$.

Lemme 56 (Formule de Christoffel-Darboux) :

On a pour tout n :

$$\sum_{j=0}^{n-1} \frac{h_j(x)h_j(y)}{j!} = \frac{1}{(n-1)!} \times \frac{h_n(x)h_{n-1}(y) - h_n(y)h_{n-1}(x)}{x-y}.$$

Définition 57 (Fonction d'Hermite) :

Les fonctions $\psi_n = \frac{h_n}{\sqrt{n!} \sqrt{2\pi}} e^{-\frac{x^2}{4}}$ forment une famille orthonormale de $L^2(\mathbb{R}, dx)$.

On peut adapter la formule de Christoffel-Darboux :

$$\sum_{k=0}^{n-1} \psi_k(x)\psi_k(y) = \sqrt{n} \times \frac{\psi_n(x)\psi_{n-1}(y) - \psi_n(y)\psi_{n-1}(x)}{x-y}.$$

3.6 Comportement asymptotique

Définition 58 :

Posons $Z^N = \{\lambda_1, \dots, \lambda_N\}$ une variable aléatoire à valeurs dans l'ensemble des parties de \mathbb{R} localement finies, qu'on note $\mathcal{P}_{loc}(\mathbb{R})$.

On munit cet espace de la tribu qui rend mesurable les applications $Z \mapsto |Z \cap A|$ pour tout borélien A .

On a vu que Z^N était typiquement compris dans $[\pm 2\sqrt{N}]$. On s'intéresse alors au comportement asymptotique de $\sqrt{N}Z^N$.

Définition 59 (Fonction de corrélation) :

Soit Z à valeurs dans $\mathcal{P}_{loc}(\mathbb{R})$. Sa fonction de corrélation est la famille (ρ_k) , où $\rho_k : \mathbb{R}^k \rightarrow \mathbb{R}^+$ est localement intégrable, telle que pour tous boréliens disjoints :

$$\mathbb{E}[|Z \cap B_1| \times \dots \times |Z \cap B_k|] = \int \mathbf{1}_{B_1}(x_1) \dots \mathbf{1}_{B_k}(x_k) \rho_k(x) dx.$$

Remarque 60 :

La fonction de corrélation du processus ponctuel de Poisson d'intensité λ est la fonction constante $\rho_k(x) = \lambda^k$.

On peut ainsi généraliser ce processus par une fonction de densité $\lambda : \mathbb{R} \rightarrow \mathbb{R}^+$ variable dans l'espace, telle que $\rho_k(x) = \prod_{j=1}^k \lambda(x_j)$, ce qui donne un processus de Poisson inhomogène.

Remarque 61 :

Tout processus ponctuel Z n'a pas de fonction de corrélation. En particulier, c'est le cas de $Z = \{0\}$ constant, pour lequel on devrait avoir le Dirac δ_0 qui n'admet pas de densité ρ_1 pour la mesure de Lebesgue.

Remarque 62 :

Il n'est a priori pas évident que (ρ_k) caractérise la loi de Z . Pour caractériser cette loi, il faut par exemple pouvoir calculer $\mathbb{E} \left[\prod_{j=1}^n |Z \cap A_j|^{k_j} \right]$ pour toutes puissances (k_j) , et tous boréliens (A_j) , ce qu'on peut faire par découpages, via des subdivisions arbitrairement fines.

Quelles sont les fonctions de corrélation de $Z^N = \{\lambda_1, \dots, \lambda_N\}$? Naturellement, lorsque $k > N$, au moins un des ensembles disjoints ne rencontre pas Z , on a $\rho_k = 0$.

Lemme 63 :

Par invariance de la loi des valeurs propres par permutation, on a :

$$\begin{aligned} \mathbb{E} \left[\prod_{j=1}^N |Z \cap B_j| \right] &= N! \mathbb{E} \left[\prod_{j=1}^N \mathbf{1}_{B_j}(\lambda_j) \right] \\ &= \int_{\prod_{j=1}^N B_j} N! f_N(x) dx, \end{aligned}$$

avec $f_N(\lambda) = C_N \Delta(\lambda)^2 \exp(-\|\lambda\|_2^2/2) d\lambda$, et $C_N^{-1} = N! \sqrt{2\pi}^N \prod_{j=0}^{N-1} j!$. Notons qu'on peut alors réécrire :

$$\begin{aligned} f_N(\lambda) &= \frac{1}{N!} \det \left(\left(\frac{h_{j-1}(\lambda_i)}{\sqrt{(j-1)!} \sqrt{2\pi}} e^{-\lambda_i^2/2} \right)_{1 \leq i, j \leq N} \right)^2 \\ &= \frac{1}{N!} \det \left((\psi_{j-1}(\lambda_i))_{1 \leq i, j \leq N} \right)^2 \\ &= \frac{1}{N!} \det \left(\left(\sum_{k=0}^{n-1} \psi_k(\lambda_i) \psi_k(\lambda_j) \right)_{1 \leq i, j \leq N} \right) \\ &=: \frac{1}{N!} \det \left((K^N(\lambda_i, \lambda_j))_{1 \leq i, j \leq N} \right), \end{aligned}$$

avec le noyau $K^N(x, y) = \sqrt{N} \frac{\psi_N(x) \psi_{N-1}(y) - \psi_N(y) \psi_{N-1}(x)}{x-y}$. Finalement, on obtient la fonction de corrélation $\rho_N(x) = \det \left((K^N(x_i, x_j))_{1 \leq i, j \leq N} \right)$.

De même, pour $k < N$:

$$\mathbb{E} \left[\prod_{j=1}^k |Z \cap B_j| \right] = \frac{N!}{(N-k)!} \int_{\prod_{j=1}^k B_j \times \mathbb{R}^{N-k}} f_N(\lambda) d\lambda.$$

On aimerait donc intégrer le déterminant de K^N contre les variables qui parcourent \mathbb{R} . Pour ce faire, quitte à développer le déterminant en tant que somme sur les permutations, on a besoin

de calculer $\int_{\mathbb{R}} K^N(x_k, x_{\sigma(k)}) dx_k$. Si $\sigma(k) = k$, alors $\int_{\mathbb{R}} K^N(x, x) dx = \sum_{j=0}^{N-1} \langle \psi_j, \psi_j \rangle = N$. Avec un calcul analogue dans le cas où $\sigma(k) \neq k$, on arrive à la conclusion que :

$$\int_{\mathbb{R}} \det\left(\left(K^N(x_i, x_j)\right)_{1 \leq i, j \leq k}\right) dx_k = (N - k + 1) \det\left(\left(K^N(x_i, x_j)\right)_{1 \leq i, j \leq k-1}\right).$$

En réinjectant cette expression dans le calcul précédent pour $k < N$, pour intégrer une par une les $N - k$ coefficients pris sur tout \mathbb{R} , on compense en fait le $\frac{1}{(N-k)!}$, ce qui finalement donne :

$$\rho_k(x) = \det\left(\left(K^N(x_i, x_j)\right)_{1 \leq i, j \leq k}\right).$$

On peut vérifier en outre que ce déterminant est nul lorsque $k > N$, puisque les fonctions $(\psi_k)_{k \leq N}$ ne peuvent qu'engendrer un espace vectoriel de dimension N . ce qui donne donc une expression pour ρ_k générale.

Définition 64 (Processus déterminantal) :

On dit que le processus ponctuel X est déterminantal lorsqu'il existe un *noyau* $K : \mathbb{R}^2 \rightarrow \mathbb{R}$ symétrique, localement intégrable, tel que pour tout $k \in \mathbb{N}$:

$$\rho_k(x) = \det\left(\left(K(x_i, x_j)\right)_{1 \leq i, j \leq k}\right).$$

Remarque 65 :

C'est le cas du processus de Poisson d'intensité λ , pour lequel $K(x, y) = \lambda \delta_{x, y}$.

Remarque 66 :

Pour un tel processus, on a une forme de répulsion :

$$\mathbb{P}(Z \cap [x \pm \varepsilon/2] \neq \emptyset, Z \cap [y \pm \varepsilon/2]) \approx \varepsilon^2 (K(x, x)K(y, y) - K(x, y)^2),$$

pour $\varepsilon \rightarrow 0$. Le terme $\varepsilon^2 K(x, x)K(y, y)$ correspondrait au cas indépendant. Le terme de droite, négatif, ne peut que diminuer cette valeur, et donc ajouter une forme de répulsion entre x et y .

Lemme 67 :

Pour $\sqrt{N}Z^N$, après un changement de variable, on obtient la fonction de corrélation :

$$\tilde{\rho}_k^N(x) = \det\left(\left(\frac{1}{\sqrt{N}} K^N\left(\frac{x_i}{\sqrt{N}}, \frac{x_j}{\sqrt{N}}\right)\right)_{1 \leq i, j \leq k}\right).$$

Lemme 68 :

On a la convergence de $(x, y) \mapsto \frac{1}{\sqrt{N}} K^N \left(\frac{x}{\sqrt{N}}, \frac{y}{\sqrt{N}} \right)$ vers $(x, y) \mapsto \text{sinc}(\pi(y-x)) =: K_{\sin}(x, y)$, uniformément sur tout compact.

Corollaire 69 :

On a la convergence uniforme sur tout compact de $\tilde{\rho}_k^N$ vers $\det \left((K_{\sin}(x_i, x_j))_{1 \leq i, j \leq k} \right)$, pour tout $k \in \mathbb{N}^*$ fixé.

Théorème 70 :

Pour tout $\alpha \in]-2, 2[$, On a la convergence en loi de $\sqrt{1 - \frac{\alpha^2}{2}} \times \sqrt{N} \left(Z^N - \alpha \sqrt{N} \right)$ vers un processus ponctuel déterminantal, de noyau K_{\sin} .

Remarque 71 :

Pour le processus limite, en prenant deux λ_i et λ_j uniformément dans un grand compact, la proportion des couples dont l'écart $\lambda_i - \lambda_j$ est compris entre a et b tend vers $\int_a^b 1 - K_{\sin}(0, x)^2 dx$.

Cette quantité est remarquable, car elle correspond aux écarts remarqués entre les niveaux d'énergie dans un gros d'atome, ou bien entre les parties imaginaires des zéros de la fonction de Riemann.

4 Détection de communautés, concentration de la norme d'une matrice sous-gaussienne

4.1 Détection de communautés : Stochastic Block Model

Définition 72 :

Soient $0 < q < p < 1$. On a deux communautés $\llbracket 1, n \rrbracket = C_1 \sqcap C_2$. On suppose ici que $C_1 = \llbracket 1, \lfloor \frac{n}{2} \rfloor \rrbracket$. Par la suite, on supposera implicitement n pair, pour avoir $|C_1| = |C_2| = \frac{n}{2}$ et alléger les notations.

On définit alors le graphe aléatoire $G(n, p, q)$ sur les sommets $\llbracket 1, n \rrbracket$, où on relie i et j avec probabilité p s'ils sont dans la même communauté, et avec probabilité q sinon, indépendamment du reste.

Remarque 73 :

Sans perte de généralité, on peut imposer le fait que i est relié à lui-même avec probabilité p , pour que la diagonale de la matrice ait aussi de l'aléa.

Remarque 74 :

On peut généraliser ce modèle à plus de deux classes, dont les tailles peuvent être aléatoires.

Remarque 75 :

On travaille à p et q fixés, ce qui correspond à une phase dense lorsque $n \rightarrow \infty$.

Remarque 76 :

Par définition du modèle, on connaît déjà C_1 . On va en fait chercher une méthode utilisant les propriétés spectrales de la matrice d'adjacence, qui donnera la même communauté au final, indépendamment de toute permutation sur les sommets.

Lemme 77 :

La matrice d'adjacence X_n du graphe aléatoire $G(n, p, q)$ est sous la forme :

$$\begin{pmatrix} \mathcal{B}(p) & \mathcal{B}(q) \\ \mathcal{B}(q) & \mathcal{B}(p) \end{pmatrix},$$

avec des blocs dont tous les coefficients sont des variables de Bernoulli indépendantes. En passant

à l'espérance on a donc naturellement :

$$\mathbb{E}[X_n] = \begin{pmatrix} p & q \\ q & p \end{pmatrix}.$$

Naturellement, le vecteur u_1 constant égal à 1 est vecteur propre pour la valeur propre $\frac{p+q}{2}n$. En outre, $u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \mathbb{1}_{C_1} - \mathbb{1}_{C_2}$ est un $\frac{p-q}{2}n$ -vecteur propre, qui permet de distinguer les deux communautés. La matrice étant de rang 2, les valeurs propres suivantes sont nulles.

C'est donc u_2 qu'on aimerait estimer, en pratique. Le problème est qu'on n'a accès qu'à X_n , pas à son espérance. On aimerait donc avoir un contrôle sur l'écart entre les deux.

Définition 78 (Estimateur de C_1) :

On considère $(\lambda_i^{X_n})_{1 \leq i \leq n}$ le spectre de X_n , classé dans l'ordre décroissant. La valeur propre $u_2^{X_n}$ associée à $\lambda_2^{X_n}$ nous permet de définir :

$$\widehat{C}_1 = \{i \in \llbracket 1, n \rrbracket, u_2^{X_n}(i) > 0\}.$$

Remarque 79 :

Les coefficients de X_n sont de l'ordre de 1 indépendamment de n . Pour retrouver l'échelle d'une matrice de Wigner, il faudrait diviser par \sqrt{n} . On s'attend à ce que $\left\| \frac{X_n - \mathbb{E}[X_n]}{\sqrt{n}} \right\| = O(1)$.

Admettons pour l'instant le théorème de concentration suivant :

Théorème 80 :

Il existe des constantes $C > 0$ telle que, pour tout $n \in \mathbb{N}$:

$$\mathbb{P}(\|X_n - \mathbb{E}[X_n]\| > C\sqrt{n}) \leq 4e^{-n}.$$

Lemme 81 (Perturbation des valeurs propres) :

Soient A et B symétriques, de spectres $(\lambda_i^A)_{i \leq n}$ et $(\lambda_i^B)_{i \leq n}$ décroissants. Alors on a l'inégalité $|\lambda_i^A - \lambda_i^B| \leq \|A - B\|$.

Démonstration. On utilise théorème de Courant-Fischer :

$$\lambda_i^A = \sup_{\dim(V)=i} \inf_{x \in V, \|x\|_2=1} \langle x, Ax \rangle \leq \sup_{\dim(V)=i} \inf_{x \in V, \|x\|_2=1} \langle x, Bx \rangle + \langle x, (A - B)x \rangle \leq \lambda_i^B + \|A - B\|,$$

et de même dans l'autre sens, d'où le résultat attendu. \square

Lemme 82 (Davis-Kahan, perturbation de vecteurs propres) :

Soient A et B symétriques. On considère e (resp. f) une valeur propre de λ pour A (resp. μ pour B) de norme 1.

$$\text{Si } |\lambda - \mu| \leq \delta, \text{ alors } |\langle f, e \rangle| \leq \frac{\|A-B\|}{\delta}.$$

Démonstration. On a $\langle f, Be \rangle = \mu \langle f, e \rangle$. En outre, par symétrie de A :

$$\langle f, Be \rangle = \langle f, (B - A)e \rangle + \langle f, Ae \rangle = \langle f, (B - A)e \rangle + \lambda \langle f, e \rangle,$$

d'où le résultat souhaité. □

Corollaire 83 :

Soit (μ_j) le spectre de B , et (f_j) une base de vecteurs propres associée. Si on a $|\mu_i - \lambda| \leq \delta$, et $|\mu_j - \lambda| > \delta$ pour $j \neq i$, alors $|\sin(e, f_i)| \leq \frac{\|B-A\|}{\delta}$, et $\|f_i \pm e\|_2 \leq \sqrt{2} \frac{\|B-A\|}{\delta}$, où $f \pm e$ désigne ici spécifiquement le choix de signe qui minimise la norme.

Démonstration. On a $\langle f_j, e \rangle^2 \leq \frac{1}{\delta^2} \langle f_j, (A - B)e \rangle^2$. En conséquence :

$$1 = \|e\|^2 = \sum_{j=1}^n \langle f_j, e \rangle^2 \leq \langle f_i, e \rangle^2 + \frac{1}{\delta^2} \sum_{j \neq i} \langle f_j, (A - B)e \rangle^2 \leq \langle f_i, e \rangle^2 + \frac{1}{\delta^2} \|(A - B)e\|_2^2,$$

et donc :

$$|\sin(e, f_i)| = \sqrt{1 - \cos(e, f_i)^2} = \sqrt{1 - \langle e, f_i \rangle^2} \leq \frac{\|A - B\|}{\delta}.$$

On obtient alors le résultat voulu sur la norme via :

$$\|f_i \pm e\|^2 = 2(1 - |\langle f_i, e \rangle|) = 2 \left(1 - \sqrt{1 - \sin(f_i, e)^2} \right) \leq 2 \times |\sin(f_i, e)|,$$

d'où le résultat en prenant la racine. □

Proposition 84 :

L'ensemble aléatoire \widehat{C}_1 estime C_1 avec grande probabilité. Plus précisément, il existe une constante \widetilde{C} indépendante de n , telle que la probabilité de faire plus de \widetilde{C} erreurs est majorée par $4e^{-n}$.

Démonstration. On se place sur l'évènement $\|X_n - \mathbb{E}[X_n]\| \leq C\sqrt{n}$. Par perturbation des valeurs propres, on a $|\lambda_2^{X_n} - n \frac{p-q}{2}| \leq C\sqrt{n}$, et $\lambda_i^{X_n} \leq C\sqrt{n}$ pour $i \geq 3$. Soit $v_2 = \frac{1}{\sqrt{n}} u_2$ le vecteur propre de $\mathbb{E}[X_n]$ normalisé.

Soit $\alpha = \min\left(\frac{p-q}{2}, q\right)$. On a donc les autres valeurs propres à distance au moins $\delta = \alpha n$.

Soit $v_2^{X_n}$ unitaire associé à la valeur propre $\lambda_2^{X_n}$ de X_n . Par le lemme précédent, on a $\|v_2^{X_n} - v_2\|_2^2 \leq 2 \frac{\|X_n - \mathbb{E}[X_n]\|_2^2}{\delta^2} = \frac{2C^2}{n\alpha^2}$. En conséquence, $\|\sqrt{n}v_2^{X_n} - u_2\|_2^2 \leq 2 \frac{C^2}{\alpha^2}$.

Si l'algorithme fait K erreurs, alors on a K coordonnées où on a des signes opposés entre $\sqrt{nv_2^{X_n}}(i)$ et $u_2(i) = \pm 1$, d'où $\|\sqrt{nv_2^{X_n}} - u_2\|_2^2 \geq K$.

Autrement dit, avec grande probabilité, on ne fera pas plus de $2\frac{C^2}{\alpha^2}$ erreurs. \square

4.2 Contrôle de la norme d'une matrice à entrées sous-gaussiennes

Définition 85 :

Une variable réelle X est sous-gaussienne si elle vérifie une des propriétés équivalentes :

1. $\exists K_1 > 0, \mathbb{P}(|X| \geq t) \leq 2 \exp\left(-\frac{t^2}{K_1^2}\right)$,
2. $\exists K_2 > 0, \forall p \in \mathbb{N}^*, \|X\|_p \leq K_2 \sqrt{p}$,
3. $\exists K_3 > 0, \forall |\lambda| \leq \frac{1}{K_3}, \mathbb{E}\left[e^{\lambda^2 X^2}\right] \leq e^{\lambda^2 K_3^2}$,
4. $\exists K_4 > 0, \mathbb{E}\left[\exp\left(\frac{X^2}{K_4^2}\right)\right] \leq 2$.

De plus, on peut toujours passer de K_i à K_j au prix d'un facteur multiplicatif universel, qui ne dépend pas de X . Si X est une variable aléatoire centrée, on a une cinquième équivalence :

$$\exists K_5 > 0, \mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2 K_5^2}.$$

Démonstration. Montrons (1 \Rightarrow 2). On a $\|X\|_p^p = \int_0^\infty pt^{p-1} \mathbb{P}(|X| > t) dt \leq 2p \int_0^\infty t^{p-1} e^{-t^2/K_1^2} dt$. On n'a plus qu'à faire un changement de variable pour faire apparaître un $2pK_1^p$ en facteur d'une intégrale correspondant à $\Gamma\left(\frac{p}{2}\right)$. Une fois qu'on passe à la puissance $\frac{1}{p}$, par la formule de Stirling, on peut borner ce terme par une constante.

Pour (2 \Rightarrow 3), il suffit de développer $e^{\lambda^2 X^2}$ comme série entière puis d'invertir la somme avec l'espérance. Pour (3 \Rightarrow 4), le résultat est direct avec $\lambda = \frac{1}{K_3}$.

Pour (4 \Rightarrow 1), on a $\mathbb{P}(|X| \geq t) = \mathbb{P}\left(e^{-X^2} \geq e^{-t^2}\right)$, d'où le résultat par inégalité de Markov.

Pour (5 \Rightarrow 1), on a $\mathbb{P}(X \geq t) \leq \mathbb{E}[e^{\lambda(X-t)}] \leq e^{\lambda(\lambda-t)}$. En optimisant en λ , on obtient la majoration souhaitée.

Enfin, montrons (3 \Rightarrow 5). Naturellement, $\mathbb{E}[e^{\lambda X}] \leq \mathbb{E}\left[\lambda X + e^{\lambda^2 X^2}\right] = \mathbb{E}\left[e^{\lambda^2 X^2}\right]$. \square

Remarque 86 :

Si X est bornée, elle est en particulier sous-exponentielle.

Définition 87 :

Si X est sous-gaussienne, on pose :

$$\|X\|_{\psi_2} = \inf \left\{ t \geq 0, \mathbb{E}\left[\exp\left(\frac{X^2}{t^2}\right)\right] \leq 2 \right\}.$$

On obtient ainsi la valeur optimale pour $K_4(X)$.

Lemme 88 :

On a $\|X\|_{\psi_2} < \infty$ ssi X est sous-gaussienne. On a $\|\lambda X\| = |\lambda| \times \|X\|$. Si $\|X\| = 0$, alors par inégalité de Markov inversée on a bien $X = 0$.

Pour la sous-additivité, soient $s > \|X\|$ et $t > \|Y\|$:

$$\begin{aligned} \mathbb{E}\left[\exp\left(\frac{(X+Y)^2}{(s+t)^2}\right)\right] &\leq \mathbb{E}\left[\exp\left(\frac{1}{p}\left(\frac{X}{s}\right)^2 + \frac{1}{q}\left(\frac{Y}{t}\right)^2\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(\frac{X^2}{s^2}\right)\right]^{\frac{1}{p}} \mathbb{E}\left[\exp\left(\frac{Y^2}{t^2}\right)\right]^{\frac{1}{q}} \\ &\leq 2^{1/p} 2^{1/q} = 2, \end{aligned}$$

en utilisant la convexité puis l'inégalité Hölder avec $p = \frac{t+s}{s}$ et $q = \frac{t+s}{t}$. On en déduit bien $\|X + Y\| \leq t + s$, et donc la sous-additivité en prenant l'infimum.

Remarque 89 :

Soient $(X_i)_{i \leq n}$ des variables *indépendantes* centrées. Comme $K_5 \leq CK_4$, on a :

$$\mathbb{E}\left[e^{\lambda \sum_{i=1}^n X_i}\right] = \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}] \leq e^{\lambda^2 C^2 \left(\sum_{i=1}^n \|X_i\|_{\psi_2}\right)}.$$

En conséquence, $\sum_{i=1}^n X_i$ vérifie la propriété 5 avec $K_5^2 \leq C^2 \left(\sum_{i=1}^n \|X_i\|_{\psi_2}\right)$. Il existe donc une constante universelle D telle que :

$$\left\| \sum_{i=1}^n X_i \right\|_{\psi_2}^2 \leq D \sum_{i=1}^n \|X_i\|_{\psi_2}^2.$$

Finalement, lorsque $\sup_{i=1}^n \|X_i\|_{\psi_2}^2 \leq K$, on a $\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right\|_{\psi_2}^n \leq DK$.

Théorème 90 :

Soit X^n une matrice symétrique à coefficients indépendants, centrés, avec la majoration $\sup_{i,j} \|X_{i,j}^N\|_{\psi_2}^2 \leq K$. Alors il existe $C(K) > 0$ tel que $\mathbb{P}(\|X^n\| \geq C\sqrt{n}) \leq 4e^{-n}$.

Démonstration. Pour se débarrasser de la symétrie, qui crée des dépendances, on décompose $X^n = T^+ + T^-$ où $T_{i,j}^+ = 0$ quand $i > j$ et $T_{i,j}^- = 0$ quand $i \leq j$. Comme $\|X^n\| \leq \|T^+\| + \|T^-\|$, il suffit de montrer le résultat pour une matrice non symétrique, avec une majoration en $2e^{-n}$, pour en déduire le résultat attendu.

Considérons donc Y à coefficients centrés, indépendants, uniformément sous-gaussiens. Dans ce cas, on veut majorer $\mathbb{P}\left(\sup_{\|x\|_2=1} \langle x, Yx \rangle \geq C\sqrt{n}\right)$. Le problème est qu'on doit considérer un

supremum sur la sphère \mathbb{S}^{n-1} . On aimerait se ramener à un maximum sur un ensemble fini.

Soit $(x_i)_{i \leq p}$ un ε -recouvrement donné par le lemme ci-après. Comme tout $x \in \mathbb{S}^{n-1}$ est à distance au plus ε d'un x_i , on a donc $\langle x, Yx \rangle \leq \langle x_i, Yx_i \rangle + 2\varepsilon\|Y\| + \varepsilon^2\|Y\|^2$. Autrement dit, on a $\|Y\| \leq \frac{1}{1-2\varepsilon-\varepsilon^2} \sup_{i=1}^p \langle x_i, Yx_i \rangle$.

Si $\varepsilon = \frac{1}{4}$, on peut prendre $p \leq 9^n$, d'où $\mathbb{P}(\|Y^n\| \geq C\sqrt{n}) \leq 9^n \sup_{x \in \mathbb{S}^{n-1}} \mathbb{P}(\langle x, Yx \rangle \geq \tilde{C}\sqrt{n})$. En outre, à x fixé, $\langle x, Yx \rangle = \sum_{i,j} x_i x_j Y_{i,j}$ est la somme de n^2 variables centrées indépendantes, de normes majorées par $x_i^2 x_j^2 K$, d'où finalement $\|\langle x, Yx \rangle\|_{\psi_2}^2 \leq D\|x\|_2^4 K = DK$. Finalement :

$$\mathbb{P}(Y^n \geq \tilde{C}\sqrt{n}) \leq 9^n \times 2 \exp\left(-\frac{\tilde{C}^2 n}{BDK}\right).$$

Quitte enfin à ajuster la constante \tilde{C} , on peut finalement obtenir une majoration en $2e^{-n}$, ce qui conclut la preuve. \square

Lemme 91 :

Il existe une famille $(x_i)_{1 \leq i \leq p}$, avec $p \leq \left(1 + \frac{2}{\varepsilon}\right)^n$, telle que $\mathbb{S}^{n-1} \subset \bigcup_{i=1}^p B(x_i, \varepsilon)$.

Démonstration. On construit une suite de sommets à distance au moins ε les uns des autres. On obtient donc des boules de rayon $\frac{\varepsilon}{2}$ disjointes. Toutes ces boules sont incluses dans $B(0, 1 + \frac{\varepsilon}{2})$. En conséquence, si on a p points, on a $p \times \left(\frac{\varepsilon}{2}\right)^n \leq \left(1 + \frac{\varepsilon}{2}\right)^n$, d'où la borne sur p annoncée. En particulier, le processus itératif donne un recouvrement en un nombre fini d'étapes. \square

Remarque 92 :

Dans le cadre précédent, on avait $K = \max\left(\|\mathcal{B}(p)\|_{\psi_2}^2, \|\mathcal{B}(q)\|_{\psi_2}^2\right)$.