

Introduction à la théorie des grandes matrices aléatoires

Jamal Najim et Romain Couillet

14 mars 2016

Résumé

Ces notes sont une introduction à la théorie des grandes matrices aléatoires, à destination d'un public non spécialiste. On y énonce et démontre en détail le théorème de Marčenko-Pastur qui décrit le spectre asymptotique d'une grande matrice de covariance. On introduit la transformée de Stieltjes et les techniques de calcul associées, ainsi que quelques techniques de calcul gaussien qui nous permettent d'aborder rapidement le théorème de Marčenko-Pastur isotrope. On aborde également les grandes matrices de covariance plus générales et les modèles à petites perturbations. Enfin, on évoque des applications de la théorie aux communications numériques ainsi qu'à l'apprentissage automatisé.

Table des matières

Introduction	2
1 Transformée de Stieltjes	9
1.1 Définition et premières propriétés	9
1.2 Convergence	13
1.3 Identifier une transformée de Stieltjes	16
1.4 Distribution de Marčenko-Pastur	17
2 Identités matricielles et inégalités	20
2.1 Notations	20
2.2 Identités matricielles	21
2.2.1 Inversion par blocs	21
2.2.2 Inverse d'une perturbation de matrice	22
2.2.3 Expression d'un élément diagonal de la résolvante	23
2.2.4 Inégalité de perturbation de rang 1	24
2.3 Inégalité d'Efron-Stein et estimations de variances	25
2.3.1 Estimation de la variance d'une forme quadratique	26
2.3.2 Estimation de la variance de la transformée de Stieltjes $g_n(z)$	27

3	Théorème de Marčenko-Pastur	31
3.1	Preuve du théorème de Marčenko-Pastur	31
3.1.1	Notations et identités importantes	31
3.1.2	Stratégie de la preuve	32
3.1.3	Equation satisfaite par $\mathbb{E} g_n(z)$	33
3.1.4	Stabilité de l'équation satisfaite par g_{MP}	34
3.1.5	Fin de la preuve	36
3.2	Calcul gaussien et théorème de Marčenko-Pastur isotrope	36
3.2.1	Formule d'intégration par parties et inégalité de Poincaré	36
3.2.2	Preuve gaussienne du théorème de Marčenko-Pastur . . .	39
3.2.3	Version isotrope du théorème de Marčenko-Pastur	40
4	Le modèle de grandes matrices de covariance	42
4.1	Exemples importants	44
4.2	Éléments de preuve	45
5	Etude de la mesure limite pour le modèle de grandes matrices de covariance	47
5.1	Compléments à la Proposition 2	48
5.2	Application au modèle de covariance	49
6	Inférence statistique pour le modèle de grandes matrices de covariance	57
6.1	Quelques rappels d'analyse complexe	57
6.2	Inférence statistique : Spécification des contours	58
6.3	Inférence statistique : Calcul des estimateurs	61
7	Petites perturbations	64
7.1	Etude de la plus grande valeur propre	65
7.2	Éléments de preuve	66
8	Grandes matrices aléatoires et communications numériques	70
8.1	le Canal MIMO	70
8.2	Le modèle de Rice	72
9	Grandes matrices aléatoires et apprentissage	75
9.1	Hypothèses pour un clustering non trivial	77
9.2	Etude de la matrice K	78
9.3	Etude de la matrice $L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}$	79
9.4	Application au clustering de la base MNIST	81

Introduction

La théorie des grandes matrices aléatoires prend ses racines dans les travaux de Wishart [34] qui a initialement considéré des matrices à entrées aléatoires gaussiennes. Wigner [32, 33] et plus tard Marčenko et Pastur [23] ont donné une

impulsion décisive au domaine en décrivant le spectre de certaines matrices hermitiennes à entrées aléatoires lorsque la dimension de ces matrices tend vers $+\infty$. Il s'est avéré que dans ce régime, une simplification majeure s'opère pour le comportement du spectre, à l'instar de la loi des grands nombres par exemple. Alors que le spectre d'une grande matrice aléatoire est un objet lui-même aléatoire et a priori compliqué, son comportement se stabilise lorsque la dimension des matrices sous-jacentes tend vers l'infini, pour finalement tendre vers un spectre complètement déterministe.

Du côté des applications en sciences de l'ingénieur, Telatar a dès 1995 (article paru en 1998) entrevu le potentiel applicatif de cette théorie aux communications numériques sans fil [30]. Ce point de vue s'est avéré extraordinairement fécond et a motivé un grand nombre de travaux (cf. par exemple [12] pour un panorama indicatif) en communications numériques mais aussi en traitement statistique du signal.

D'un point de vue mathématique, ces dix dernières années ont également été l'occasion d'une activité considérable, comme en témoigne la parution de monographies majeures dans le domaine [1, 25, 3, 29].

L'objectif de cet article est de présenter quelques éléments clé de cette théorie, avec suffisamment de détails pour permettre au lecteur intéressé d'aborder des lectures plus avancées, en particulier dans le domaine de l'application de la théorie des grandes matrices aléatoires aux sciences de l'ingénieur. On se concentrera principalement sur deux familles de techniques : les techniques fondées sur la transformée de Stieltjes et celles fondées sur le calcul gaussien. D'autres techniques existent, fondées sur la combinatoire (identification des moments des lois limites) ou sur la théorie des probabilités libres (cf. [1, 24, 29]), mais ne seront pas abordées ici.

Matrice de Wigner. Le théorème de Wigner concerne des matrices (W_N) , de dimension $N \times N$, dont toutes les entrées W_{ij} sur et au dessus de la diagonale sont réelles (le théorème s'étend aux entrées complexes), indépendantes et identiquement distribuées (i.i.d.) et de carré sommables. On complète les entrées sous la diagonale de sorte que la matrice W_N soit symétrique : $W_N^T = W_N$ où A^T est la matrice transposée de A . Dans le cas d'entrées complexes, on imposera la condition de symétrie $W_N^* = W_N$, où A^* représente la transconjuguée de la matrice A . On appellera matrice de Wigner une telle matrice. Si les entrées sont centrées de variance 1, et si on renormalise la matrice par $1/\sqrt{N}$, alors le spectre de W_N s'organise autour de la loi du demi-cercle $N \rightarrow \infty$, au sens où l'histogramme de ses valeurs propres va suivre la loi du demi-cercle ("*semi-circle law*"), de densité

$$f_{sc}(x) = \frac{1}{2\pi} \sqrt{4 - x^2} 1_{|x| < 2} ,$$

comme l'indique la figure 1. On notera \mathbb{P}_{sc} la probabilité associée à la densité f_{sc} . Notons que l'histogramme des valeurs propres a été construit à partir d'une seule réalisation de la matrice W_N . Pour une preuve combinatoire du théorème de Wigner, on pourra par exemple se référer aux notes de C. Bordenave [9] ainsi qu'au chapitre 21 du livre de D. Chafaï et F. Malrieu [11].

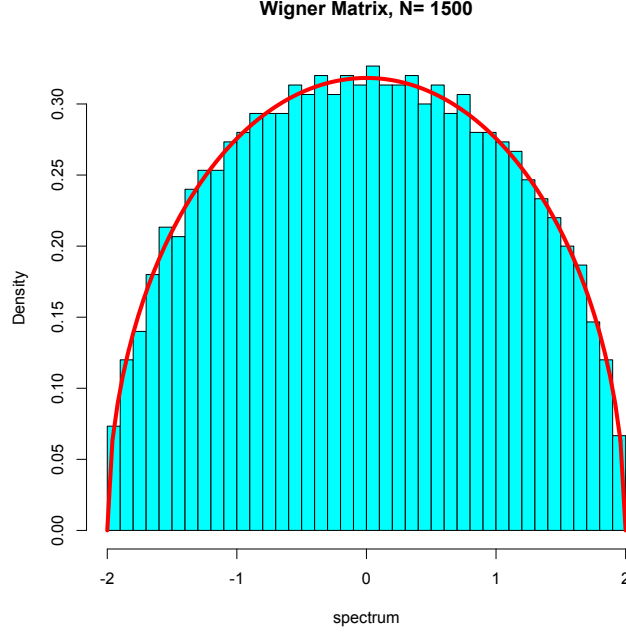


FIGURE 1 – Histogramme des valeurs propres d’une matrice de Wigner renormalisée de dimensions 1500×1500 . En rouge, la densité de la loi du demi-cercle.

Grandes matrices de covariance. Le théorème de Marčenko-Pastur concerne les grandes matrices de covariance. Soit X_N une matrice $N \times n$ dont toutes les entrées sont réelles ou complexes i.i.d., centrées, de variance unité et de carré sommables ; on considère la matrice $\frac{1}{n}X_N X_N^T$ si les entrées sont réelles et $\frac{1}{n}X_N X_N^*$ si les entrées sont complexes. Alors, quand les deux dimensions vont croître au même rythme, c’est à dire $N, n \rightarrow \infty$ et

$$\frac{N}{n} \rightarrow c \in]0, \infty[,$$

le spectre de la matrice $\frac{1}{n}X_N X_N^T$ (resp. de la matrice $\frac{1}{n}X_N X_N^*$ dans le cas d’entrées complexes) va s’organiser selon la distribution de Marčenko-Pastur dont la densité est donnée par

$$f_{MP}(x) = \frac{\sqrt{(\lambda^+ - x)(x - \lambda^-)}}{2\pi c x} 1_{[\lambda^-, \lambda^+]}(x) \quad \text{où} \quad \lambda^\pm = (1 \pm \sqrt{c})^2$$

(dans le cas où $c \leq 1$). On notera \mathbb{P}_{MP} la probabilité associée à f_{MP} . La figure 2 illustre ce résultat.

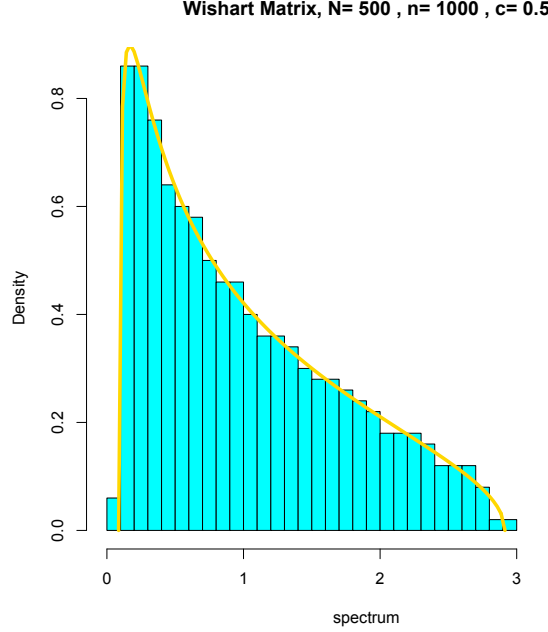


FIGURE 2 – Histogramme des valeurs propres d’une grande matrice de covariance de dimensions 500×1000 . En jaune, la densité de la loi du demi-cercle.

Grandes matrices de covariance (suite). Considérons une matrice X_N de dimensions $N \times n$ à entrées i.i.d., centrées, réduites et de carré sommables. On notera $X_N = [\vec{x}_1, \dots, \vec{x}_n]$ où \vec{x}_i représente le i ème vecteur-colonne de la matrice X_N . Soit R_N une matrice de covariance (hermitienne, semi-définie positive) $N \times N$, déterministe. On considère la matrice

$$R_N^{1/2} X_N = \left[R_N^{1/2} \vec{x}_1, \dots, R_N^{1/2} \vec{x}_n \right] .$$

Cette matrice modélise un échantillon de n observations $\vec{y}_i = R_N^{1/2} \vec{x}_i$, chacune de dimension N et de matrice de covariance $R_N = \mathbb{E} \vec{y}_i \vec{y}_i^*$. A ce titre, c’est un modèle très utile dans les applications en traitement statistique du signal ou autre. Posons

$$\Sigma_N = \frac{1}{\sqrt{n}} R_N^{1/2} X_N$$

et considérons la matrice de covariance empirique associée

$$\Sigma_N \Sigma_N^* = \frac{1}{n} R_N^{1/2} X_N X_N^* R_N^{1/2} = \frac{1}{n} \sum_{i=1}^n \vec{y}_i \vec{y}_i^* .$$

Cette matrice peut apporter des éléments de compréhension de la "vraie" matrice de covariance R_N (souvent appelée **matrice de covariance de la population**). En effet, dans le régime statistique conventionnel où N est fixé et $n \rightarrow \infty$ (données de dimension fixée N , taille de l'échantillon très grande $n \rightarrow \infty$), la convergence suivante est immédiate :

$$\Sigma_N \Sigma_N^* = \frac{1}{n} \sum_{i=1}^n \vec{y}_i \vec{y}_i^* \xrightarrow[N \text{ fixé}, n \rightarrow \infty]{p.s.} R_N . \quad (1)$$

Autrement dit, la matrice de covariance empirique est dans ce cas une bonne approximation de la matrice de covariance de la population.

La situation est sensiblement différente dans le régime asymptotique des grandes matrices $N, n \rightarrow \infty$ et $\frac{N}{n} \rightarrow c$. Dans ce cas, le ratio c représente le rapport entre la dimension N des données et la taille n de l'échantillon.

On peut montrer alors que les valeurs propres de la matrice $\Sigma_N \Sigma_N^*$ s'organisent autour d'une distribution limite dont on peut tracer la densité (on dira alors - cf. paragraphe suivant - que la mesure spectrale de $\Sigma_N \Sigma_N^*$ converge vers une distribution limite).

La figure 3 représente en rouge la distribution limite de la mesure spectrale de la matrice $\Sigma_N \Sigma_N^*$ lorsque la matrice R_N admet 3 valeurs propres distinctes, 1, 3 et 7 (représentées par 3 traits verticaux bleus), chacune de multiplicité identique ($\simeq \frac{N}{3}$).

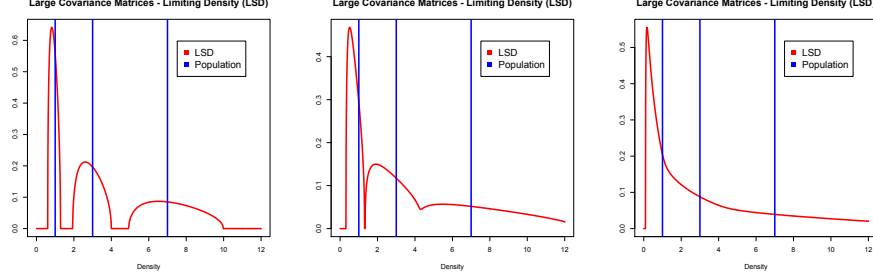


FIGURE 3 – En rouge, densité de la mesure limite dans les cas où $c = 0, 1$; $c = 0, 3$ et $c = 0, 6$. En bleu, les 3 valeurs propres de la matrice R_N de covariance de la population, chacune de multiplicité égale.

Il apparaît alors que lorsque le ratio c est petit ($c = 0, 1$, ce qui signifie que l'échantillon est 10 fois plus grand que la dimension des données), on peut, grâce à la mesure limite (en rouge) "pressentir" la matrice de covariance de la population (dont les valeurs propres sont en bleu) ; ce cas de figure est à rapprocher de la convergence (1). En revanche, quand ce ratio est plus important ($c = 0, 6$), il est rigoureusement impossible, à l'oeil, de deviner les valeurs propres de la matrice de covariance de la population. Autrement dit, l'information apportée par la matrice de covariance empirique sur la matrice de covariance de la population est beaucoup moins directe que dans le régime conventionnel !

Petites perturbations. Un cas particulier et intéressant du modèle de grandes matrices de covariance précédent $\Sigma_N \Sigma_N^*$ est celui où la matrice R_N est une petite perturbation de l'unité :

$$R_N = I_N + \sum_{\ell=1}^k \theta_\ell \vec{u}_\ell \vec{u}_\ell^*,$$

où les (θ_i) sont des réels fixés et déterministes, et les (\vec{u}_i) (qui dépendent de n) sont des vecteurs déterministes et orthonormés. Chaque matrice $\vec{u}_i \vec{u}_i^*$ est une matrice de rang 1 (matrice de perturbation) et on suppose que le nombre de perturbations k est fini et indépendant de N, n . La question suivante est alors naturelle : comment le spectre de la matrice $\frac{1}{n} X_N X_N^*$ est-il modifié par la présence de la perturbation $\sum_{\ell=1}^k \theta_\ell \vec{u}_\ell \vec{u}_\ell^*$? Des simulations apportent quelques éléments de réponse.

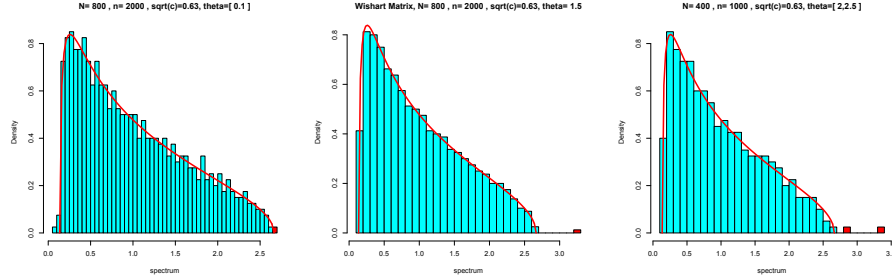


FIGURE 4 – En rouge, la densité de Marčenko-Pastur, en bleu l'histogramme des valeurs propres empiriques. Le(s) petit(s) bâton(s) rouge(s) représente(nt) la(les) plus grande(s) valeur(s) propre(s) empirique(s). Les deux premières simulations correspondent à une perturbation simple où $\theta = 0, 1$ et $\theta = 1, 5$; la dernière à une perturbation double où $\theta_1 = 2$ et $\theta_2 = 2, 5$.

Les simulations illustrent le fait que le comportement de toutes les valeurs propres, sauf des plus grandes, est bien prédit par la distribution de Marčenko-Pastur. L'intensité de la perturbation semble également jouer un rôle dans la capacité des plus grandes valeurs propres à se séparer de la masse des autres.

Mesure spectrale. La manière standard de formaliser la convergence du spectre d'une grande matrice de covariance est de considérer la mesure spectrale associée. Appelons $\lambda_1, \dots, \lambda_N$ les N valeurs propres (comptées avec leur multiplicité) d'une matrice symétrique ou hermitienne Z_N de dimension $N \times N$. Le théorème spectral nous assure que ces valeurs propres sont réelles. Appelons δ_x la mesure de Dirac au point x , définie par

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{si } x \notin A. \end{cases}$$

Une propriété importante de la mesure de Dirac est que

$$\int f(u) \delta_x(du) = f(x) .$$

On appellera **mesure spectrale de Z_N** (ou mesure empirique des valeurs propres associées à Z_N) la probabilité définie par :

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

Ainsi, $L_N(A)$ correspond à la proportion des valeurs propres contenues dans A :

$$L_N(A) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}(A) = \frac{\#\{i, \lambda_i \in A\}}{N} .$$

On dira que L_N **converge étroitement** vers une mesure de probabilité μ (notation $L_N \xrightarrow{etr} \mu$) si quelque soit la fonction f continue et bornée sur \mathbb{R} , on a

$$\int f(u) L_N(du) = \frac{1}{N} \sum_{i=1}^N f(\lambda_i) \xrightarrow{n \rightarrow \infty} \int f(u) \mu(du) .$$

Si de plus la matrice Z_N est aléatoire, ses valeurs propres le sont également et on dira que presque sûrement L_N converge étroitement vers μ si pour presque toute réalisation de la famille de matrices Z_N , on a convergence étroite de L_N (en tant qu'objet aléatoire) vers μ (qui est une probabilité certes, mais ne dépendant pas du hasard).

Ainsi, le théorème de Wigner s'exprimera de la manière suivante : soit L_N la mesure spectrale de $\frac{W_N}{\sqrt{N}}$, où W_N est une matrice de Wigner, alors presque sûrement

$$L_N \xrightarrow[N \rightarrow \infty]{etr} \mathbb{P}_{sc} .$$

Organisation de l'article. La preuve du théorème de Marčenko-Pastur structurera la première partie (rigoureuse) de l'article (sections 1 à 3) ; elle sera pour nous l'occasion d'introduire des concepts d'usage constant dans cette théorie tels que la transformée de Stieltjes (section 1), certaines identités matricielles et inégalités (section 2) et des techniques de calcul gaussien (section 3.2). Dans la deuxième partie de l'article, moins formelle (sections 4 & 5), nous décrirons la convergence du spectre de grandes matrices de covariance plus générales. Nous nous appuierons sur les résultats et ordres de grandeur développés dans la première partie. Enfin, nous nous intéresserons à l'impact de perturbations de rang fini sur le spectre de grandes matrices aléatoires. En section 8, nous détaillerons une application de la théorie aux communications numériques sans fil et plus particulièrement aux communications multi-antennes.

1 Transformée de Stieltjes

La transformée de Stieltjes d'une mesure s'avère être une transformation particulièrement adaptée pour l'étude asymptotique des mesures spectrales. Pour de telles mesures, la transformée de Stieltjes est la trace normalisée de la résolvante associée à la matrice sous-jacente.

1.1 Définition et premières propriétés

Soit μ une mesure positive sur \mathbb{R} muni de la tribu des boréliens, de masse totale finie, i.e. $\mu(\mathbb{R}) < \infty$. On note

$$\mathbb{C}^+ = \{z \in \mathbb{C}, \operatorname{Im}(z) > 0\} .$$

La **transformée de Stieltjes** de μ est définie par

$$g_\mu(z) = \int_{\mathbb{R}} \frac{\mu(d\lambda)}{\lambda - z}, \quad z \in \mathbb{C}^+ .$$

On remarque que la définition s'étend à \mathbb{C}^- , ainsi qu'à $\mathbb{R} \setminus \mathcal{S}$, où \mathcal{S} représente le support de μ .

Exemple 1. Soient $x, x_i \in \mathbb{R}$

1. Si $\mu(d\lambda) = \delta_x(d\lambda)$, alors $g_\mu(z) = \int \frac{\mu(d\lambda)}{\lambda - z} = \frac{1}{x - z}$.
2. Si $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$, alors $g_\mu(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{x_i - z}$.
3. Soit Z une matrice hermitienne $N \times N$ de décomposition spectrale

$$Z = U\Lambda U^* \quad \text{où} \quad UU^* = U^*U = I_N \quad \text{et} \quad \Lambda = \operatorname{diag}(\lambda_i) ,$$

les λ_i étant les valeurs propres de Z . Soit L_N la **mesure spectrale de Z** , aussi appelée **mesure empirique des valeurs propres de Z** :

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

Alors

$$g(z) = \frac{1}{N} \operatorname{Tr} (Z - zI_N)^{-1}$$

est la transformée de Stieltjes de la mesure L_N . En effet,

$$\begin{aligned} \frac{1}{N} \operatorname{Tr} (Z - zI_N)^{-1} &= \frac{1}{N} \operatorname{Tr} (U\Lambda U^* - zUU^*)^{-1} = \frac{1}{N} \operatorname{Tr} \{U(\Lambda - zI_N)^{-1}U^*\} \\ &= \frac{1}{N} \operatorname{Tr} (\Lambda - zI_N)^{-1} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z} = \int \frac{L_N(d\lambda)}{\lambda - z} = g(z) . \end{aligned}$$

4. Même contexte que précédemment. Soit a un vecteur $N \times 1$; on note u_i le vecteur propre associé à la valeur propre λ_i . Alors

$$g(z) = a^*(Z - zI_N)^{-1}a = \sum_{i=1}^N \frac{|\langle a, u_i \rangle|^2}{\lambda_i - z} \quad \text{où} \quad \langle a, u_i \rangle = a^* u_i ,$$

est la transformée de Stieltjes de la mesure

$$\sum_{i=1}^N |\langle a, u_i \rangle|^2 \delta_{\lambda_i} .$$

de masse totale $\sum_{i=1}^N |\langle a, u_i \rangle|^2 = \|a\|^2$.

Proposition 2. Soit g_μ la transformée de Stieltjes de la mesure μ , de masse totale $\mu(\mathbb{R})$ finie et de support \mathcal{S} , alors :

1. g_μ est analytique sur \mathbb{C}^+ ,
2. $g_\mu \in \mathbb{C}^+$ pour $z \in \mathbb{C}^+$,
3. si $\mathcal{S} \subset \mathbb{R}^+$, alors $\text{Im}(zg_\mu(z)) \geq 0$
4. $\lim_{y \rightarrow +\infty} iy g_\mu(iy) = -\mu(\mathbb{R})$
5. pour tout $z \in \mathbb{C}^+$,

$$|g_\mu(z)| \leq \frac{\mu(\mathbb{R})}{\text{Im}(z)} .$$

6. quelque soit $x \in \mathbb{R}$,

$$\mu\{x\} = \lim_{y \searrow 0} y \text{Im} g_\mu(x + iy) .$$

7. quelque soit $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée, on a

$$\int f d\mu = \lim_{y \searrow 0} \frac{1}{\pi} \text{Im} \int f(x) g_\mu(x + iy) dx$$

8. Si $a, b \in \mathbb{R}$ sont des points de continuité de μ , i.e. $\mu(\{a\}) = \mu(\{b\}) = 0$, alors

$$\mu(a, b) = \lim_{y \searrow 0} \frac{1}{\pi} \text{Im} \int_a^b g_\mu(x + iy) dx$$

On remarquera que lors de la reconstruction d'une mesure à partir de sa transformée de Stieltjes (cf. les points 6, 7 et 8 ci-dessus), c'est le comportement de celle-ci près de l'axe réel qui importe.

Démonstration. 1. Immédiat.

2. Un calcul direct donne

$$\text{Im} g_\mu(x + iy) = y \int \frac{\mu(d\lambda)}{(\lambda - x)^2 + y^2} > 0 \quad \text{si} \quad y > 0 .$$

3. Posons $z = x + iy$; un calcul de même nature que le précédent donne

$$\operatorname{Im}(zg_\mu(z)) = y \int_{\mathbb{R}^+} \frac{\lambda \mu(d\lambda)}{(\lambda - x)^2 + y^2} \geq 0 \quad \text{si } y > 0 \text{ et } \mathcal{S} \subset \mathbb{R}^+.$$

4. On a

$$iyg_\mu(iy) = \int \frac{iy}{\lambda - iy} \mu(d\lambda) = \int \frac{i\lambda y - y^2}{\lambda^2 + y^2} \mu(d\lambda).$$

On conclut par le théorème de convergence dominée :

$$\frac{|\lambda y|}{\lambda^2 + y^2} \leq \frac{1}{2}, \quad \frac{|\lambda y|}{\lambda^2 + y^2} \xrightarrow{y \rightarrow +\infty} 0 \quad \text{et} \quad \frac{|-y^2|}{\lambda^2 + y^2} \leq 1, \quad \frac{-y^2}{\lambda^2 + y^2} \xrightarrow{y \rightarrow +\infty} -1.$$

5. Si $z = x + iy$, alors $|\lambda - z| \geq y$ et

$$|g_\mu(z)| \leq \frac{\mu(\mathbb{R}^+)}{y}.$$

6. Un calcul immédiat donne

$$y \operatorname{Im} g_\mu(x + iy) = \int \frac{y^2 \mu(d\lambda)}{(\lambda - x)^2 + y^2}.$$

En remarquant que, à x fixé, $\frac{y^2}{(\lambda - x)^2 + y^2} \xrightarrow{y \searrow 0} 1_{\{x\}}(\lambda)$, une application directe du théorème de convergence dominée donne

$$\lim_{y \searrow 0} y \operatorname{Im} g_\mu(x + iy) = \int 1_{\{x\}}(\lambda) \mu(d\lambda) = \mu\{x\}.$$

Avant d'aborder les 2 derniers points (que l'on démontrera sans perte de généralité pour des mesures de probabilité), on fait un petit rappel sur les lois de Cauchy et la convolution : on dit que la variable aléatoire réelle Y suit une loi de Cauchy si sa densité $f_Y(y)$ est donnée par

$$f_Y(x) = \frac{1}{\pi(1 + x^2)}.$$

On démontre alors sans peine que yY (où $y > 0$) admet la densité

$$f_{yY}(x) = \frac{y}{\pi(y^2 + x^2)}.$$

Enfin, on rappelle que si X est une variable aléatoire réelle de loi μ , Y est indépendante de X et admet la densité f_Y , alors $X + Y$ admet la densité

$$f_{X+Y}(u) = \int f_Y(u - x) \mu(dx).$$

Forts de ces rappels, on utilisera la propriété suivante :

$$\frac{1}{\pi} \operatorname{Im} \int f(x) g_\mu(x + iy) dx = \mathbb{E} f(X + yY), \quad (2)$$

où X est une V.A. de loi μ , Y suit une loi de Cauchy et X et Y sont indépendantes. En effet, d'après les rappels, $X + yY$ a pour densité

$$\frac{1}{\pi} \int \frac{y}{(\lambda - x)^2 + y^2} \mu(d\lambda)$$

donc

$$\mathbb{E} f(X + yY) = \int f(x) \frac{1}{\pi} \int \frac{y \mu(d\lambda)}{(\lambda - x)^2 + y^2} dx .$$

D'autre part, un calcul direct de $\text{Im } g_\mu(x + iy)$ donne

$$\frac{1}{\pi} \int f(x) \text{Im } g_\mu(x + iy) dx = \int f(x) \frac{1}{\pi} \int \frac{y \mu(d\lambda)}{(\lambda - x)^2 + y^2} dx ,$$

ce qui établit (2). Revenons maintenant aux preuves des points 7 et 8.

7. Soit f une fonction continue bornée et $\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$, alors

$$f(X + yY) \xrightarrow{y \searrow 0} f(X) \quad \text{et} \quad |f(X + yY)| \leq \|f\|_\infty .$$

Par convergence dominée, il vient que

$$\lim_{y \searrow 0} \mathbb{E} f(X + yY) \stackrel{(a)}{=} \lim_{y \searrow 0} \frac{1}{\pi} \int f(x) \text{Im } g_\mu(x + iy) dx = \mathbb{E} f(X) = \int f(x) \mu(dx) ;$$

ici, (a) découle de l'identité (2).

8. On procède par approximation et on introduit deux fonctions continues bornées φ_n^- et φ_n^+ telles que

$$\varphi_n^- \leq 1_{[a,b]} \leq \varphi_n^+ , \quad \lim_{n \rightarrow \infty} \varphi_n^- = 1_{]a,b[} , \quad \lim_{n \rightarrow \infty} \varphi_n^+ = 1_{[a,b]} .$$

On pourra par exemple prendre pour φ_n^- et φ_n^+ les lignes brisées définies par

$$\begin{aligned} \varphi_n^-(x) &= \begin{cases} 0 & \text{si } x \in (-\infty, a) \cup (b, \infty) \\ 1 & \text{si } x \in (a + \frac{1}{n}, b - \frac{1}{n}) \end{cases} , \\ \varphi_n^+(x) &= \begin{cases} 0 & \text{si } x \in (-\infty, a - \frac{1}{n}) \cup (b + \frac{1}{n}, \infty) \\ 1 & \text{si } x \in (a, b) \end{cases} . \end{aligned}$$

et rectilignes ailleurs. Soit X une V.A. de loi μ , on a donc

$$\begin{aligned} \mathbb{E} \varphi_n^-(X) &= \lim_{y \searrow 0} \frac{1}{\pi} \int \varphi_n^-(x) \text{Im } g_\mu(x + iy) dx \\ &\leq \liminf_{y \searrow 0} \frac{1}{\pi} \int_a^b \text{Im } g_\mu(x + iy) dx \leq \limsup_{y \searrow 0} \frac{1}{\pi} \int_a^b \text{Im } g_\mu(x + iy) dx \\ &\leq \lim_{y \searrow 0} \frac{1}{\pi} \int \varphi_n^+(x) \text{Im } g_\mu(x + iy) dx = \mathbb{E} \varphi_n^+(X) . \end{aligned}$$

Par convergence dominée, en faisant maintenant tendre n vers l'infini, il vient $\mathbb{E} \varphi_n^-(X) \rightarrow \mu(]a, b[)$ et $\mathbb{E} \varphi_n^+(X) \rightarrow \mu([a, b])$. Soit finalement

$$\begin{aligned} \mu(]a, b[) &\leq \liminf_{y \searrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im} g_\mu(x + iy) dx \\ &\leq \limsup_{y \searrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im} g_\mu(x + iy) dx \leq \mu([a, b]) \end{aligned}$$

d'où l'égalité si a et b sont des points de continuité de μ .

□

1.2 Convergence

On va montrer dans ce paragraphe comment caractériser la convergence étroite de mesures de probabilités grâce à la convergence ponctuelle de leurs transformées de Stieltjes.

On appelle C_K l'ensemble des fonctions continues sur \mathbb{R} à support compact. Étant donnée une famille μ_n de mesures de probabilités sur \mathbb{R} , on dit qu'elle **converge vaguement** vers une mesure ν si pour toute fonction $f \in C_K$, on a

$$\int f(x) \mu_n(dx) \xrightarrow{n \rightarrow \infty} \int f(x) \nu(dx) .$$

On indiquera la convergence vague par \xrightarrow{v} .

La convergence vague est très flexible car on dispose du critère de compacité suivant, appelé **principe de sélection de Helly** (cf. [8, Appendix II, p.226], [21, Th. 5.19]) : de toute suite de probabilités μ_n sur \mathbb{R} , on peut extraire une sous-suite qui converge vaguement.

Malheureusement, cette convergence peut occasionner une **perte de masse** : la limite ν n'est plus forcément une probabilité (on peut d'ailleurs montrer que nécessairement, $\nu(\mathbb{R}) \leq 1$).

Exemple 3. 1. $\delta_n \xrightarrow[n \rightarrow \infty]{v} 0$.

2. Soit \mathbb{P} une probabilité, $\alpha \in (0, 1)$ et posons $\mu_n = (1 - \alpha)\mathbb{P} + \alpha\delta_n$, alors

$$\mu_n \xrightarrow[n \rightarrow \infty]{v} \nu \triangleq (1 - \alpha)\mathbb{P}$$

et $\nu(\mathbb{R}) = 1 - \alpha < 1$.

Remarque 4. Une propriété intéressante de la convergence vague est la suivante :

$$\mathbb{P}_n \xrightarrow[n \rightarrow \infty]{v} \mu \quad \Rightarrow \quad \int f d\mathbb{P}_n \xrightarrow[n \rightarrow \infty]{} \int f d\nu$$

pour toute fonction f continue qui tend vers zéro en $\pm\infty$. Cette remarque sera utile dans la preuve du théorème 7.

Soit μ_n une suite de mesures de probabilités. On rappelle que μ_n **converge étroitement** vers μ si pour toute fonction continue bornée,

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu .$$

On notera \xrightarrow{etr} . Dans ce cas, μ est nécessairement une probabilité (prendre $f = 1_{\mathbb{R}}$ pour montrer que $\mu(\mathbb{R}) = 1$) ; en revanche, on n'a plus de critère automatique de compacité.

Étant donnée une suite μ_n de probabilités, on dit qu'elle est **tendue** si pour tout $\varepsilon > 0$, il existe un compact $K \subset \mathbb{R}$ tel que

$$\forall n \geq 1, \quad \mu_n([-K, K]^c) \leq \varepsilon .$$

La condition de tension équivaut par passage au complémentaire à la condition suivante

$$\forall n \geq 1, \quad \mu_n([-K, K]) \geq 1 - \varepsilon$$

que l'on peut interpréter comme le fait qu'à ε près, toutes les probabilités μ_n ont leur support essentiellement dans $[-K, K]$.

Théorème 5. *Étant donnée une suite de probabilités (μ_n) , les deux conditions suivantes sont équivalentes :*

1. *De toute sous-suite de (μ_n) , on peut extraire une sous-suite qui converge étroitement,*
2. *La suite (μ_n) est tendue.*

On trouvera la démonstration de ce résultat dans [13, Th. 9.3.3] et [21, Prop. 5.21] ; on vérifiera aisément que les suite de probabilités de l'exemple 3 ne sont pas tendues.

Proposition 6. *Étant donnée une suite de probabilités (μ_n) , les deux conditions suivantes sont équivalentes :*

1. $\mu_n \xrightarrow[n \rightarrow \infty]{v} \mu$ et μ est une probabilité.
2. $\mu_n \xrightarrow[n \rightarrow \infty]{etr} \mu$.

Démonstration. Le fait que 2. \Rightarrow 1. est immédiat ; montrons la réciproque. Soit $K > 0$ tel que $\mu([-K, K]) \geq 1 - \varepsilon$. Considérons une fonction continue f , comprise entre 0 et 1, à support inclus dans $[-(K+1), K+1]$ et valant 1 sur le compact $[-K, K]$. Alors

$$\mu_n([-K, K]) \geq \int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu \geq 1 - \varepsilon .$$

On en déduit qu'à partir d'un certain rang, $\mu_n([-K, K]) \geq 1 - 2\varepsilon$. En agrandissant suffisamment le compact $[-(K+1), K+1]$ pour que l'inégalité précédente reste valable pour les indices plus petits, on démontre ainsi que (μ_n) est tendue. Le théorème 5 permet alors de conclure. \square

Théorème 7. Soit (μ_n) une suite de probabilités sur \mathbb{R} .

1. Si $\mu_n \xrightarrow[n \rightarrow \infty]{etr} \mu$, alors $g_{\mu_n}(z) \xrightarrow[n \rightarrow \infty]{} g_\mu(z)$, pour tout $z \in \mathbb{C}^+$.
2. (a) si $g_{\mu_n}(z) \xrightarrow[n \rightarrow \infty]{} g(z)$ pour $z \in \mathcal{D}$, \mathcal{D} étant un sous-ensemble de \mathbb{C}^+ contenant un point d'accumulation, alors il existe une mesure ν de masse totale inférieure ou égale à 1 telle que

$$g(z) = \int \frac{\nu(d\lambda)}{\lambda - z} \quad \text{pour } z \in \mathcal{D}$$

et $\mu_n \xrightarrow[n \rightarrow \infty]{v} \nu$.

(b) si de plus

$$\lim_{y \rightarrow +\infty} iy g(iy) = -1 ,$$

alors ν est une probabilité et $\mu_n \xrightarrow[n \rightarrow \infty]{etr} \nu$.

Remarque 8. Le théorème précédent sera d'un usage constant : bien souvent, il est plus facile de démontrer la convergence de la transformée de Stieltjes de la mesure spectrale L_N associée à une matrice aléatoire Z_N que la convergence directe de la mesure spectrale.

Preuve du théorème 7. Montrons le premier point. À $z = x + iy \in \mathbb{C}^+$ fixé, l'application

$$\lambda \mapsto \frac{1}{\lambda - z} = \frac{\lambda - x}{(\lambda - x)^2 + y^2} + i \frac{y}{(\lambda - x)^2 + y^2}$$

est continue bornée donc par définition de la convergence étroite, on a convergence de $g_{\mu_n}(z)$ vers $g_\mu(z)$ pour $z \in \mathbb{C}^+$.

Montrons le point 2-(a). Considérons la suite de probabilités (μ_n) . D'après le principe de sélection de Helly, il existe une sous-suite (μ'_n) qui converge vaguement vers une sous-probabilité ν . Comme l'application $\lambda \mapsto \frac{1}{\lambda - z}$ est continue et tend vers zéro en $\pm\infty$, cela entraîne (cf. la remarque 4) en particulier que

$$\int \frac{\mu'_n(d\lambda)}{\lambda - z} \xrightarrow[n \rightarrow \infty]{} \int \frac{\nu(d\lambda)}{\lambda - z}$$

quelque soit $z \in \mathbb{C}^+$. Du fait de l'hypothèse, on obtient l'identification suivante :

$$g(z) = \int \frac{\nu(d\lambda)}{\lambda - z} \quad \text{pour } z \in \mathcal{D} .$$

Le même raisonnement entraîne que si une autre sous-suite (μ''_n) converge vers $\tilde{\nu}$, alors la transformée de Stieltjes de $\tilde{\nu}$ est encore égale à g sur \mathcal{D} . Par suite, les fonctions

$$\int \frac{\nu(d\lambda)}{\lambda - z} \quad \text{et} \quad \int \frac{\tilde{\nu}(d\lambda)}{\lambda - z}$$

coïncident sur \mathcal{D} et, étant analytiques sur \mathbb{C}^+ , y coïncident¹ aussi. Par suite ν et $\tilde{\nu}$ coïncident. On en déduit alors que de toute sous-suite de (μ_n) , on peut extraire une sous-suite qui converge vaguement vers ν , et finalement que la suite (μ_n) converge vaguement vers ν .

Si de plus $iyg(iy)$ converge vers -1 , alors ν est une probabilité d'après le point (4) de la proposition 2. La proposition 6 permet alors de conclure. \square

1.3 Identifier une transformée de Stieltjes

Etant donnée une fonction $g : \mathbb{C}^+ \rightarrow \mathbb{C}^+$, il est important de disposer de critères permettant d'affirmer qu'une telle fonction est la transformée de Stieltjes d'une mesure, voire d'une probabilité.

Théorème 9. *Si la fonction $g : \mathbb{C}^+ \rightarrow \mathbb{C}^+$ satisfait les conditions suivantes :*

1. *g est analytique sur \mathbb{C}^+ ,*
2. *il existe $M > 0$ telle que*

$$|g(z)| \leq \frac{M}{\operatorname{Im}(z)}, \quad z \in \mathbb{C}^+.$$

3. *$\lim_{y \rightarrow \infty} -iyg(iy) = \kappa$.*

Alors il existe une unique mesure positive μ sur \mathbb{R} , de masse totale κ , i.e. $\mu(\mathbb{R}) = \kappa$, telle que

$$g(z) = \int_{\mathbb{R}} \frac{\mu(d\lambda)}{\lambda - z}.$$

4. *Si de plus $\operatorname{Im}(zg(z)) \geq 0$, alors $\mu[-\infty, 0] = 0$, i.e.*

$$g(z) = \int_{\mathbb{R}^+} \frac{\mu(d\lambda)}{\lambda - z}.$$

Pour une preuve de ce résultat, on pourra se référer à [31, Theorem B.3] et [22, Appendix].

Exemple 10. *Soit $g(z)$ la transformée de Stieltjes associée à une mesure de support inclus dans \mathbb{R}^+ , alors*

$$h(z) = -\frac{1}{z(1+g(z))}$$

est la transformée de Stieltjes d'une probabilité. En effet, h est analytique sur \mathbb{C}^+ . Par hypothèse sur g , $\operatorname{Im}(zg(z)) \geq 0$, ce qui implique que $h \in \mathbb{C}^+$, et

$$|h(z)| \leq \frac{1}{|z(1+g(z))|} \leq \frac{1}{\operatorname{Im}[z(1+g(z))]} \leq \frac{1}{\operatorname{Im}(z)}, \quad z \in \mathbb{C}^+.$$

Enfin, on montre facilement que $-iyh(iy) \xrightarrow{y \rightarrow +\infty} 1$. Le théorème précédent permet alors de conclure.

1. On rappelle que si deux fonctions analytiques sur \mathbb{C}^+ coïncident sur \mathcal{D} , alors elles coïncident sur \mathbb{C}^+ , voir par exemple [26, Corollaire du théorème 10.18].

1.4 Distribution de Marčenko-Pastur

On a mentionné le fait que l'une des approches possibles pour démontrer le théorème de Mačenko-Pastur est de considérer la convergence de la transformée de Stieltjes associée à la mesure spectrale. Il est donc important de bien savoir décrire la transformée de Stieltjes de la limite escomptée. C'est l'objet de ce paragraphe où on étudie la transformée de Stieltjes $g_{\check{M}P}$ associée à la distribution de Marčenko-Pastur.

On note \sqrt{z} la branche de la fonction racine définie pour $z = \rho e^{i\theta}$, $\theta \in (0, 2\pi)$ par

$$\sqrt{z} = \sqrt{\rho} e^{i\frac{\theta}{2}}, \quad \theta \in (0, 2\pi). \quad (3)$$

On remarque qu'avec cette définition, \sqrt{z} coïncide avec la branche principale de la fonction racine lorsque $\theta \in (0, \pi)$ et avec l'autre branche de la fonction racine lorsque $\theta \in (\pi, 2\pi)$. En particulier, soit $x \in \mathbb{R}$ et $x > 0$. Si $z = x e^{i\theta}$ alors

$$\sqrt{z} \xrightarrow{\theta \searrow 0} \sqrt{x} \quad \text{et} \quad \sqrt{z} \xrightarrow{\theta \nearrow 2\pi} -\sqrt{x}.$$

La fonction $z \mapsto \sqrt{z}$ ainsi définie est analytique lorsque $\theta \in (0, 2\pi)$.

Etant donné $c > 0$, $\sigma^2 > 0$, on note

$$\lambda^- = \sigma^2(1 - \sqrt{c})^2 \quad \text{et} \quad \lambda^+ = \sigma^2(1 + \sqrt{c})^2.$$

Enfin, on note $x_+ = \max(x, 0)$.

Théorème 11. Soit $z \in \mathbb{C}^+$, $c > 0$, $\sigma^2 > 0$ et le trinôme du second degré

$$zc\sigma^2 X^2 + [z + \sigma^2(c-1)]X + 1 = 0. \quad (4)$$

On considère la solution suivante du trinôme

$$g_{\check{M}P}(z) = \frac{-[z + \sigma^2(c-1)] + \sqrt{(z - \lambda^+)(z - \lambda^-)}}{2z\sigma^2 c}, \quad z \in \mathbb{C}^+,$$

où $z \mapsto \sqrt{z}$ est définie par (3). Alors

1. l'application $z \mapsto g_{\check{M}P}(z)$ est l'unique transformée de Stieltjes solution du trinôme pour tout $z \in \mathbb{C}^+$; c'est la transformée de Stieltjes d'une mesure de probabilité.
2. L'application $z \mapsto g_{\check{M}P}(z)$ est la transformée de Stieltjes de la distribution de Marčenko-Pastur, notée $\mathbb{P}_{\check{M}P}$, et définie par

$$\mathbb{P}_{\check{M}P}(d\lambda) = \left(1 - \frac{1}{c}\right)_+ \delta_0(d\lambda) + \frac{\sqrt{(\lambda^+ - \lambda)(\lambda - \lambda^-)}}{2\pi c \sigma^2 \lambda} 1_{[\lambda^-, \lambda^+]}(\lambda) d\lambda.$$

3. Si $z \in \mathbb{C}^+$, $x \in \mathbb{R}$, $x > \lambda^+$, alors

$$g_{\check{M}P}(z) \xrightarrow{z \rightarrow x} g_{\check{M}P}(x) \triangleq \frac{-[x + \sigma^2(c-1)] + \sqrt{(x - \lambda^+)(x - \lambda^-)}}{2x\sigma^2 c}$$

et $g_{\check{M}P}(x)$ est solution du trinôme (4) où z a été remplacé par x .

4. Si $z \in \mathbb{C}^+$, $x \in \mathbb{R} \setminus \{0\}$, $x < \lambda^-$, alors

$$g_{\tilde{M}P}(z) \xrightarrow{z \rightarrow x} g_{\tilde{M}P}(x) \triangleq \frac{-[x + \sigma^2(c-1)] - \sqrt{(x - \lambda^+)(x - \lambda^-)}}{2x\sigma^2c}$$

et $g_{\tilde{M}P}(x)$ est solution du trinôme (4) où z a été remplacé par x .

Remarque 12. Dans la suite, on se réfèrera souvent à l'équation (4) comme étant l'équation canonique de Marčenko-Pastur.

Démonstration. On démontre le premier point.

D'abord l'unicité. Soient g et \tilde{g} deux transformées de Stieltjes solutions de (4). En soustrayant les deux égalités correspondantes, il vient

$$(g - \tilde{g})(zc\sigma^2(g + \tilde{g}) + z + \sigma^2(c-1)) = 0.$$

Les fonctions g et \tilde{g} étant des transformées de Stieltjes, la quantité $z(g + \tilde{g})$ est bornée sur \mathbb{C}^+ . En particulier, pour $|z|$ assez grand (disons $|z| > A$), la quantité $zc\sigma^2(g + \tilde{g}) + z + \sigma^2(c-1)$ est toujours différente de zéro. Cela entraîne $g = \tilde{g}$ pour $|z| > A$. Les fonctions g et \tilde{g} étant analytiques sur \mathbb{C}^+ , elles y coïncident donc.

Montrons maintenant que $g_{\tilde{M}P}$, solution de (4), est une transformée de Stieltjes. Le calcul du discriminant du trinôme donne

$$\begin{aligned} [z + \sigma^2(c-1)]^2 - 4zc\sigma^2 &= z^2 - 2\sigma^2(1+c)z + \sigma^4(c-1)^2 \\ &= [z - \sigma^2(1+c)]^2 - 4c\sigma^4 \\ &= [z - \sigma^2(1+2\sqrt{c}+c)] [z - \sigma^2(1-2\sqrt{c}+c)] \\ &= (z - \lambda^+) (z - \lambda^-). \end{aligned}$$

Si $z \in \mathbb{C}^+$, alors les arguments des nombres complexes $z - \lambda^+$ et $z - \lambda^-$ sont compris entre 0 et π ; l'argument du discriminant $(z - \lambda^+) (z - \lambda^-)$ est alors compris entre 0 et 2π et par le choix (3) de la branche de la racine carrée, celui de sa racine est compris entre 0 et π . En particulier,

$$\operatorname{Im} \sqrt{(z - \lambda^+) (z - \lambda^-)} \geq 0.$$

On remarque en particulier que du fait de l'analyticit  de la branche choisie de la racine, le discriminant est analytique sur \mathbb{C}^+ . Considérons la solution

$$g_{\tilde{M}P}(z) = \frac{-[z + \sigma^2(c-1)] + \sqrt{(z - \lambda^+) (z - \lambda^-)}}{2c\sigma^2z}.$$

Celle-ci est aussi analytique sur \mathbb{C}^+ et s'écrit

$$\begin{aligned} g_{\tilde{M}P}(z) &= \frac{(z - \lambda^+) (z - \lambda^-) - [z + \sigma^2(c-1)]^2}{2c\sigma^2z \left(\sqrt{(z - \lambda^+) (z - \lambda^-)} + [z + \sigma^2(c-1)] \right)} \\ &= \frac{-4c\sigma^2z}{2c\sigma^2z \left(\sqrt{(z - \lambda^+) (z - \lambda^-)} + [z + \sigma^2(c-1)] \right)} \\ &= -\frac{2}{\sqrt{(z - \lambda^+) (z - \lambda^-)} + [z + \sigma^2(c-1)]}. \end{aligned} \tag{5}$$

Soit $D(z)$ le dénominateur ci-dessus ; alors si $z \in \mathbb{C}^+$

$$\operatorname{Im} D(z) = \underbrace{\operatorname{Im} \sqrt{(z - \lambda^+)(z - \lambda^-)}}_{\geq 0} + \underbrace{\operatorname{Im}(z)}_{> 0} > 0.$$

Par suite, $|g_{\check{\text{MP}}}(z)| \leq \frac{2}{\operatorname{Im}(z)}$ et

$$g_{\check{\text{MP}}}(z) = \frac{-2[\operatorname{Re} D(z) - i \operatorname{Im} D(z)]}{\operatorname{Re}^2 D(z) + \operatorname{Im}^2 D(z)} \in \mathbb{C}^+.$$

Ainsi $g_{\check{\text{MP}}}$ est la transformée de Stieltjes d'une mesure positive sur \mathbb{R} . Toujours grâce à la représentation (5), on montre facilement que

$$\lim_{\mathbb{R} \ni y \rightarrow \infty} i y g_{\check{\text{MP}}}(i y) = -1,$$

ce qui nous assure que la mesure dont $g_{\check{\text{MP}}}$ est la transformée de Stieltjes est une probabilité, que l'on notera $\mathbb{P}_{\check{\text{MP}}}$.

On montre maintenant les points 2,3 et 4 du théorème.

Soit $x \in]\lambda^+, \infty[$ et $z \in \mathbb{C}^+$, $z \rightarrow x$. Alors

$$|(z - \lambda^+)(z - \lambda^-)| \rightarrow (x - \lambda^+)(x - \lambda^-)$$

et $\operatorname{Arg}(z - \lambda^+)(z - \lambda^-) \searrow 0$. Ainsi,

$$\sqrt{(z - \lambda^+)(z - \lambda^-)} \xrightarrow{z \rightarrow x} \sqrt{(x - \lambda^+)(x - \lambda^-)}$$

et

$$g_{\check{\text{MP}}}(z) \xrightarrow{z \rightarrow x} g_{\check{\text{MP}}}(x) = \frac{-[x + \sigma^2(c - 1)] + \sqrt{(x - \lambda^+)(x - \lambda^-)}}{2x\sigma^2c}.$$

Comme il est d'autre part immédiat de montrer que $g_{\check{\text{MP}}}(x)$ est solution du trinôme (4) quand x remplace z , cela entraîne le point 3 du théorème.

Soit $x \in]-\infty, \lambda^-[$ et $z \in \mathbb{C}^+$ tel que $z \rightarrow x$. Un raisonnement similaire entraîne que $\operatorname{Arg}(z - \lambda^+)(z - \lambda^-) \nearrow 2\pi$. Ainsi,

$$\sqrt{(z - \lambda^+)(z - \lambda^-)} \xrightarrow{z \rightarrow x} -\sqrt{(x - \lambda^+)(x - \lambda^-)}.$$

Le point 4 en découle facilement.

On a montré en particulier que la limite de $g_{\check{\text{MP}}}(z)$ lorsque $z \rightarrow x$ et $x \in \mathbb{R}^* \setminus [\lambda^-, \lambda^+]$ est réelle, ce qui entraîne

$$\frac{1}{\pi} \operatorname{Im} g_{\check{\text{MP}}}(z) \xrightarrow{z \rightarrow x} 0 \tag{6}$$

Calculons maintenant $\mathbb{P}_{\check{\text{MP}}}(\{0\})$. Par propriété des transformées de Stieltjes, on a

$$\mathbb{P}_{\check{\text{MP}}}(\{0\}) = \lim_{y \searrow 0} y \operatorname{Im}(g_{\check{\text{MP}}}(i y)).$$

Notons que $\sqrt{(iy - \lambda^+)(iy - \lambda^-)}$ converge vers $-\sqrt{\lambda^+ \lambda^-}$ quand $y \rightarrow 0$. Il vient

$$g_{\check{\text{MP}}}(iy) = \frac{-(iy + \sigma^2(c-1)) - \sqrt{\lambda^+ \lambda^-} + o(y)}{2c\sigma^2 iy}$$

et

$$y \operatorname{Im}(g_{\check{\text{MP}}}(iy)) \xrightarrow{y \searrow 0} \frac{c-1}{2c} + \frac{\sqrt{\lambda^+ \lambda^-}}{2c\sigma^2} = \frac{1}{2} \left\{ \left(1 - \frac{1}{c}\right) + \left|1 - \frac{1}{c}\right| \right\} = \left(1 - \frac{1}{c}\right)_+.$$

Déterminons maintenant la densité de $\mathbb{P}_{\check{\text{MP}}}$ sur le segment $[\lambda^-, \lambda^+]$.

Soit $x \in [\lambda^-, \lambda^+]$ et $z \in \mathbb{C}^+$, $z \rightarrow x$. Alors

$$|(z - \lambda^+)(z - \lambda^-)| \xrightarrow{z \rightarrow x} (\lambda^+ - x)(x - \lambda^-) \quad \text{et} \quad \operatorname{Arg}(z - \lambda^+)(z - \lambda^-) \xrightarrow{z \rightarrow x} \pi.$$

Par suite,

$$\sqrt{(z - \lambda^+)(z - \lambda^-)} \xrightarrow{z \rightarrow x} i\sqrt{(\lambda^+ - x)(x - \lambda^-)}$$

et

$$\frac{1}{\pi} \operatorname{Im} g_{\check{\text{MP}}}(z) \xrightarrow{z \rightarrow x} \frac{\sqrt{(\lambda^+ - x)(x - \lambda^-)}}{2c\pi\sigma^2 x}. \quad (7)$$

On peut d'autre part facilement borner la transformée de Stieltjes $g_{\check{\text{MP}}}$ sur tout ensemble de la forme $[a, b] \times [0, \varepsilon]$ où $[a, b]$ ne contient pas zéro. En utilisant les convergences (6) et (7) et le théorème de convergence dominée, il vient

$$\int f(x) \mathbb{P}_{\check{\text{MP}}}(dx) = \int_{\lambda^-}^{\lambda^+} f(x) \frac{\sqrt{(\lambda^+ - x)(x - \lambda^-)}}{2c\pi\sigma^2 x} dx$$

pour toute fonction f à support inclus dans $[a, b]$ ne contenant pas zéro. \square

2 Identités matricielles et inégalités

2.1 Notations

On notera I la matrice identité (en indiquant parfois sa dimension $I = I_N$). Étant donné un vecteur $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{C}^N$, on notera $\|\mathbf{a}\|$ sa norme euclidienne :

$$\|\mathbf{a}\| = \left(\sum_{i=1}^N |a_i|^2 \right)^{1/2}.$$

Pour une matrice B de dimensions $N \times n$, on notera $\|B\|$ sa norme spectrale :

$$\|B\| = \max \left\{ \sqrt{\lambda}, \lambda \text{ valeur propre de } BB^* \right\}.$$

On rappelle qu'alors

$$\|B\| = \sup_{\|\mathbf{u}\| \leq 1, \|\mathbf{v}\| \leq 1} |\mathbf{u}^* B \mathbf{v}|, \quad (8)$$

où $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$.

Soient $u_n(z)$ et $v_n(z)$ des suites de fonctions définies sur \mathbb{C}^+ . On notera $u_n = \mathcal{O}_z(v_n)$ pour signifier qu'il existe $K > 0$, $p, q \in \mathbb{N}$ tels que

$$|u_n(z)| \leq K \frac{|z|^p}{\operatorname{Im}^q(z)} |v_n(z)|$$

pour n assez grand.

Étant donnée une matrice X_N de dimensions $N \times n$, à entrées centrées et i.i.d., on pose $Y_N = \frac{1}{\sqrt{n}} X_N$ et on note \mathbf{y}_i la i -ième colonne de Y_N , soit

$$Y_N = \frac{X_N}{\sqrt{n}} = [\mathbf{y}_1, \dots, \mathbf{y}_n] .$$

On introduit les résolvantes suivantes

$$\begin{aligned} Q(z) &= \left(\frac{1}{n} X_N X_N^* - zI \right)^{-1} = \left(\sum_{k=1}^n \mathbf{y}_k \mathbf{y}_k^* - zI \right)^{-1} , \\ Q_i(z) &= \left(\sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^* - zI \right)^{-1} , \end{aligned}$$

ainsi que la transformée de Stieltjes de la mesure spectrale de $Y_N Y_N^*$:

$$g_n(z) = \frac{1}{N} \operatorname{Tr} Q(z) .$$

On note $\mathbb{E}_{\{i\}}$ l'espérance conditionnelle² par rapport aux vecteurs $\{\mathbf{y}_k; k \neq i\}$ et $\operatorname{var}_{\{i\}}$, la variance conditionnelle correspondante.

2.2 Identités matricielles

On rappelle certaines identités matricielles et on les spécialise dans le contexte des grandes matrices aléatoires. Une référence classique sur le sujet est le livre de Horn et Johnson [19].

2.2.1 Inversion par blocs

On considère la matrice carrée de taille $n \times n$ partitionnée selon les blocs :

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} ,$$

les blocs diagonaux A et D étant carrés.

2. L'utilisation de l'espérance conditionnelle $\mathbb{E}_{\{i\}}$ revient à ne calculer l'espérance $\mathbb{E}_{\{i\}} f(\mathbf{y}_1, \dots, \mathbf{y}_n)$ que par rapport au vecteur \mathbf{y}_i et à considérer les autres vecteurs \mathbf{y}_k comme des constantes.

Proposition 13. *En supposant que tous les inverses considérés existent, l'inverse de la matrice M est donné par la formule*

$$M^{-1} = \begin{bmatrix} [A - BD^{-1}C]^{-1} & -A^{-1}B[D - CA^{-1}B]^{-1} \\ -[D - CA^{-1}B]^{-1}CA^{-1} & [D - CA^{-1}B]^{-1} \end{bmatrix} .$$

Démonstration. On écrit M^{-1} sous forme de matrice partitionnée

$$M^{-1} = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$$

et on exploite la condition :

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} .$$

Il vient :

$$AU_{11} + BU_{21} = I \quad (9)$$

$$AU_{12} + BU_{22} = 0 \quad (10)$$

$$CU_{11} + DU_{21} = 0 \quad (11)$$

$$CU_{12} + DU_{22} = I \quad (12)$$

En combinant (10) et (12), il vient

$$(D - CA^{-1}B)U_{22} = I \quad \Rightarrow \quad \begin{cases} U_{22} &= (D - CA^{-1}B)^{-1} \\ U_{12} &= -A^{-1}B(D - CA^{-1}B)^{-1} \end{cases} .$$

En procédant de manière similaire avec les équations (9) et (11), on en déduit les expressions de U_{11} et U_{21} . □

2.2.2 Inverse d'une perturbation de matrice

Etant donnée la matrice

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} ,$$

on appelle complément de Schur de A dans M la quantité :

$$S_A = D - CA^{-1}B .$$

L'expression du bloc U_{11} de la matrice inverse M^{-1} en fonction de S_A s'obtient à l'aide de l'équation (9) :

$$AU_{11} + BU_{21} = I \quad \Rightarrow \quad U_{11} = A^{-1} - A^{-1}BU_{21} .$$

En égalisant les deux expressions obtenues pour U_{11} , on obtient l'expression suivante :

$$\begin{aligned}(A - BD^{-1}C)^{-1} &= A^{-1} + A^{-1}B[D - CA^{-1}B]^{-1}CA^{-1} , \\ &= A^{-1} + A^{-1}B S_A CA^{-1} .\end{aligned}$$

En changeant D en $-D$ dans la formule précédente, on obtient l'identité suivante, parfois appelée **formule de Sherman-Morrison**.

Proposition 14. *Soit A une matrice $n \times n$, B , C et D des matrices $n \times k$, $k \times n$ et $k \times k$. En supposant que les inverses considérés existent, on a*

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B[D + CA^{-1}B]^{-1}CA^{-1} . \quad (13)$$

Cette identité est particulièrement utile pour exprimer l'inverse de $A + BD^{-1}C$, lorsque $BD^{-1}C$ est une perturbation de petit rang³, à l'aide de l'inverse A^{-1} de A .

Exemple 15. *L'inverse d'une perturbation de rang 1 de la matrice A est un cas particulier d'usage constant : soient \mathbf{x} et \mathbf{y} des vecteurs de dimension $n \times 1$, alors*

$$(A + \mathbf{x}\mathbf{y}^*)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{x}\mathbf{y}^*A^{-1}}{1 + \mathbf{y}^*A^{-1}\mathbf{x}} . \quad (14)$$

2.2.3 Expression d'un élément diagonal de la résolvante

Soit Σ une matrice $N \times n$ dont on désigne les lignes par $(\xi_i, 1 \leq i \leq N)$. On appelle Σ_1 la matrice Σ privée de sa première ligne. En remarquant que $\Sigma\Sigma^* = (\xi_i\xi_j^*)$, on obtient

$$\Sigma\Sigma^* - zI_N = \begin{bmatrix} \xi_1\xi_1^* - z & \xi_1\xi_2^* & \cdots & \xi_1\xi_N^* \\ \xi_2\xi_1^* & & & \\ \vdots & & \Sigma_1\Sigma_1^* - zI_{N-1} & \\ \xi_N\xi_1^* & & & \end{bmatrix} .$$

En appliquant la proposition 13, on peut obtenir une expression du premier élément diagonal de la résolvante $Q(z) = (\Sigma\Sigma^* - zI)^{-1} = (q_{ij}(z))$:

$$q_{11}(z) = \frac{1}{\xi_1\xi_1^* - z - \xi_1\Sigma_1^*(\Sigma_1\Sigma_1^* - zI)^{-1}\Sigma_1\xi_1^*} .$$

Appliquons maintenant la formule de la proposition 14 en particulierisant

$$A = -zI, \quad B = \Sigma_1^*, \quad C = \Sigma_1, \quad D = I ,$$

3. Par petite perturbation, on se réfère au cas où D est une petite matrice carrée ($k \ll n$), de sorte que la perturbation $BD^{-1}C$ est de rang k petit.

il vient

$$(-zI + \Sigma_1^* \Sigma_1)^{-1} = -\frac{I}{z} - \left(-\frac{I}{z}\right) \Sigma_1^* \left(\Sigma_1 \left(-\frac{I}{z}\right) \Sigma_1^* + I\right)^{-1} \Sigma_1 \left(-\frac{I}{z}\right),$$

soit finalement

$$-z(\Sigma_1^* \Sigma_1 - zI)^{-1} = I - \Sigma_1^* (\Sigma_1 \Sigma_1^* - zI)^{-1} \Sigma_1,$$

et

$$q_{11}(z) = -\frac{1}{z(1 + \boldsymbol{\xi}_1^* (\Sigma_1^* \Sigma_1 - zI)^{-1} \boldsymbol{\xi}_1)}.$$

On a montré la proposition suivante :

Proposition 16. *Etant donnée une matrice Σ de dimensions $N \times n$, on dénote $\boldsymbol{\xi}_i$ la i ème ligne de Σ et Σ_i la matrice Σ privée de sa ligne $\boldsymbol{\xi}_i$, de dimensions $(N-1) \times n$. Soit $Q(z)$ la résolvante $Q(z) = (\Sigma \Sigma^* - zI)^{-1} = (q_{ij}(z))$. Alors*

$$q_{ii}(z) = -\frac{1}{z(1 + \boldsymbol{\xi}_i^* (\Sigma_i^* \Sigma_i - zI)^{-1} \boldsymbol{\xi}_i^*)}, \quad 1 \leq i \leq N.$$

2.2.4 Inégalité de perturbation de rang 1

La proposition suivante sera particulièrement utile dans la suite.

Proposition 17. *Soit A une matrice $N \times N$ hermitienne, B une matrice $N \times N$ quelconque et \mathbf{x} un vecteur (colonne) de \mathbb{C}^N , alors*

$$|\mathrm{Tr} B(A + \mathbf{x} \mathbf{x}^* - zI)^{-1} - \mathrm{Tr} B(A - zI)^{-1}| \leq \frac{\|B\|}{\mathrm{Im}(z)}.$$

En particulier, si $B = B_N$ est de norme spectrale uniformément bornée en N , alors

$$\frac{1}{N} \mathrm{Tr} B(A + \mathbf{x} \mathbf{x}^* - zI)^{-1} = \frac{1}{N} \mathrm{Tr} B(A - zI)^{-1} + \mathcal{O}_z\left(\frac{1}{N}\right)$$

Démonstration. En utilisant l'identité de Sherman-Morrisson (14), il vient

$$\begin{aligned} & \mathrm{Tr} B(A + \mathbf{x} \mathbf{x}^* - zI)^{-1} - \mathrm{Tr} B(A - zI)^{-1} \\ &= -\mathrm{Tr} \left\{ B(A - zI)^{-1} \frac{\mathbf{x} \mathbf{x}^*}{1 + \mathbf{x}^* (A - zI)^{-1} \mathbf{x}} (A - zI)^{-1} \right\} \\ &= -\frac{\mathbf{x}^* (A - zI)^{-1} B (A - zI)^{-1} \mathbf{x}}{1 + \mathbf{x}^* (A - zI)^{-1} \mathbf{x}}. \end{aligned}$$

Soit

$$\begin{aligned} & |\mathrm{Tr} B(A + \mathbf{x} \mathbf{x}^* - zI)^{-1} - \mathrm{Tr} B(A - zI)^{-1}| \\ & \leq \frac{|\mathbf{x}^* (A - zI)^{-1} B (A - zI)^{-1} \mathbf{x}|}{|1 + \mathbf{x}^* (A - zI)^{-1} \mathbf{x}|} \\ & \leq \frac{|\mathbf{x}^* (A - zI)^{-1} B (A - zI)^{-1} \mathbf{x}|}{|\mathrm{Im}(1 + \mathbf{x}^* (A - zI)^{-1} \mathbf{x})|}. \end{aligned} \tag{15}$$

A l'aide du théorème spectral, on décompose $A = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^*$ où les λ_i sont les valeurs propres de A et les \mathbf{u}_i les vecteurs propres associés. On en déduit que

$$\operatorname{Im} (1 + \mathbf{x}^* (A - zI)^{-1} \mathbf{x}) = \operatorname{Im}(z) \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{|\lambda_i - z|^2} . \quad (16)$$

En effet,

$$1 + \mathbf{x}^* (A - zI)^{-1} \mathbf{x} = 1 + \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{\lambda_i - z} = 1 + \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{|\lambda_i - z|^2} (\lambda_i - \bar{z}) ,$$

ce qui entraîne (16). D'autre part, on a

$$\|(A - zI)^{-1} \mathbf{x}\| = \|(A - \bar{z}I)^{-1} \mathbf{x}\| = \left\{ \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{|\lambda_i - z|^2} \right\}^{1/2} . \quad (17)$$

En effet,

$$\begin{aligned} \|(A - zI)^{-1} \mathbf{x}\| &= \left\{ \mathbf{x}^* (A - \bar{z}I)^{-1} (A - zI)^{-1} \mathbf{x} \right\}^{1/2} \\ &= \left\{ \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{(\lambda_i - \bar{z})(\lambda_i - z)} \right\}^{1/2} = \|(A - \bar{z}I)^{-1} \mathbf{x}\| . \end{aligned}$$

En réinjectant les équations (16) et (17) dans (15), il vient

$$\begin{aligned} &|\operatorname{Tr} B(A + \mathbf{x} \mathbf{x}^* - zI)^{-1} - \operatorname{Tr} B(A - zI)^{-1}| \\ &\leq \frac{1}{\operatorname{Im}(z)} \left| \left(\frac{\mathbf{x}^* (A - zI)^{-1}}{\|(A - \bar{z}I)^{-1} \mathbf{x}\|} \right) B \left(\frac{(A - zI)^{-1} \mathbf{x}}{\|(A - zI)^{-1} \mathbf{x}\|} \right) \right| , \\ &\stackrel{(a)}{\leq} \frac{\|B\|}{\operatorname{Im}(z)} , \end{aligned}$$

où (a) découle de (8). □

2.3 Inégalité d'Efron-Stein et estimations de variances

L'inégalité d'Efron-Stein, dont on trouvera une preuve dans l'excellent livre de Boucheron et al. [10, Section 3.1], est particulièrement efficace pour obtenir des bonnes estimations de variances dans un contexte de grandes matrices aléatoires.

Proposition 18. *Soient y_1, \dots, y_n n vecteurs aléatoires à valeurs \mathbb{R}^d , i.i.d. et y'_1, \dots, y'_n n vecteurs de même loi que les (y_i) et indépendants d'eux. Soit $f = f(y_1, \dots, y_n)$ de carré sommable et*

$$f'_i = f(y_1, \dots, y_{i-1}, y'_i, y_{i+1}, \dots, y_n) .$$

Alors

$$\operatorname{var} f(y_1, \dots, y_n) \leq \frac{1}{2} \sum_{i=1}^n \mathbb{E} |f - f'_i|^2 .$$

2.3.1 Estimation de la variance d'une forme quadratique

Proposition 19. Soit A une matrice $n \times n$ déterministe et hermitienne, et \mathbf{x} un vecteur $n \times 1$ dont les composantes x_i sont des variables aléatoires i.i.d., centrées, de quatrième moment fini, alors

$$\text{var}(\mathbf{x}^* A \mathbf{x}) \leq 3m_4 \sum_{i=1}^N A_{ii}^2 + 6m_2^2 \text{Tr}(AA^*) ,$$

où $m_2 = \mathbb{E}|x_1|^2$ et $m_4 = \mathbb{E}|x_1|^4$.

Remarque 20. Posons $\mathbf{y} = \frac{\mathbf{x}}{\sqrt{n}}$ et supposons que la norme spectrale de la suite de matrices $A = A_n$ est uniformément bornée. Alors la proposition précédente entraîne que

$$\text{var}(\mathbf{y}^* A \mathbf{y}) \leq \frac{K}{n} .$$

On suppose de plus que $\mathbb{E}|x_i|^2 = 1$. En remarquant que $\mathbb{E}(\mathbf{y}^* A \mathbf{y}) = \frac{1}{n} \text{Tr}(A)$, cela entraîne en particulier que $\mathbf{y}^* A \mathbf{y} - \frac{1}{n} \text{Tr}(A) \xrightarrow[N \rightarrow \infty]{\mathcal{P}} 0$, où $\xrightarrow{\mathcal{P}}$ représente la convergence en probabilité.

Démonstration. Appliquons l'inégalité d'Efron-Stein à $f(x_1, \dots, x_n) = \mathbf{x}^* A \mathbf{x}$:

$$f(x_1, \dots, x_n) = |x_i|^2 A_{ii} + \bar{x}_i \sum_{\ell \neq i} x_\ell A_{i\ell} + x_i \sum_{k \neq i} \bar{x}_k A_{ki} + \sum_{k, \ell \neq i} \bar{x}_k x_\ell A_{k\ell} ,$$

soit

$$f - f'_i = (|x_i|^2 - |x'_i|^2) A_{ii} + (\bar{x}_i - \bar{x}'_i) \sum_{\ell \neq i} x_\ell A_{i\ell} + (x_i - x'_i) \sum_{k \neq i} \bar{x}_k A_{ki} .$$

En utilisant successivement les inégalités élémentaires

$$(a + b + c)^2 \leq 3(a^2 + b^2 + c^2) \quad \text{et} \quad (a + b)^2 \leq 2(a^2 + b^2)$$

pour a, b et c positifs, on obtient

$$\begin{aligned} \mathbb{E}|f - f'_i|^2 &\leq 6m_4 A_{ii}^2 + 6m_2 \mathbb{E} \left| \sum_{\ell \neq i} x_\ell A_{i\ell} \right|^2 + 6m_2 \mathbb{E} \left| \sum_{k \neq i} \bar{x}_k A_{ki} \right|^2 \\ &= 6m_4 A_{ii}^2 + 12m_2^2 \sum_{\ell \neq i} |A_{i\ell}|^2 \\ &\leq 6m_4 A_{ii}^2 + 12m_2^2 \sum_{\ell} |A_{i\ell}|^2 . \end{aligned}$$

Il suffit maintenant de sommer sur i et diviser par 2 pour obtenir le résultat annoncé. \square

Le corollaire suivant sera d'un usage constant dans la suite :

Corollaire 21. Soit X_N une matrice $N \times n$ dont les entrées sont i.i.d. centrées, de quatrième moment fini, et \mathbf{y}_i , Q_i tels que définis en section 2.1. On suppose que $Nn^{-1} \rightarrow c > 0$, alors

$$\text{var}_{\{i\}} (\mathbf{y}_i^* Q_i \mathbf{y}_i) = \mathcal{O}_z \left(\frac{1}{n} \right) .$$

On remarque que le terme de droite est déterministe (bien que celui de gauche dépende a priori des entrées de Q_i).

Démonstration. Il suffit de particulariser le résultat précédent :

$$m_2 = \frac{\mathbb{E} |X_{11}|^2}{n}, \quad m_4 = \frac{\mathbb{E} |X_{11}|^4}{n^2}, \quad |[Q_i]_{kk}| \leq \frac{1}{\text{Im}(z)} \quad \text{et} \quad \frac{1}{N} \text{Tr} Q_i Q_i^* \leq \frac{1}{\text{Im}^2(z)}$$

pour obtenir que

$$\text{var}_{\{i\}} (\mathbf{y}_i^* Q_i \mathbf{y}_i) \leq \frac{K}{n \text{Im}^2(z)} .$$

□

Remarque 22. Parfois, l'estimation de la variance ne suffit pas et il est nécessaire de "monter en moments" et d'obtenir un contrôle de quantités du type

$$\mathbb{E} |\mathbf{x}^* A \mathbf{x} - \text{Tr} A|^p .$$

Le lemme de Bai et Silverstein [3, Lemma B.26] fournit de telles estimées :

$$\mathbb{E} |\mathbf{x}^* A \mathbf{x} - \text{Tr} A|^p \leq C_p \left([m_4 \text{Tr} (AA^*)]^{p/2} + m_{2p} \text{Tr} (AA^*)^{p/2} \right) ,$$

où m_{2p} représente le moment d'ordre $2p$ de chaque composante x_i du vecteur \mathbf{x} (on exprime ici le résultat dans le cas où $\mathbb{E} |x_i|^2 = 1$).

2.3.2 Estimation de la variance de la transformée de Stieltjes $g_n(z)$

On montre l'estimation suivante :

Proposition 23. Soit X_N une matrice $N \times n$ dont les entrées sont i.i.d. centrées, de quatrième moment fini, et $g_n(z)$ telle que définie en section 2.1. On suppose que $Nn^{-1} \rightarrow c > 0$, alors

$$\text{var} g_n(z) = \mathcal{O}_z \left(\frac{1}{n^2} \right) .$$

Remarque 24. On note que la variance de g_n tend rapidement vers zéro, à la vitesse n^{-2} ; cela traduit la grande interdépendance des valeurs propres λ_i . À titre de comparaison, si on considère N variables (U_i) i.i.d. de variance finie, alors

$$\text{var} \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{U_i - z} \right) = \mathcal{O}_z \left(\frac{1}{N} \right) .$$

Démonstration. Soit \mathbf{y}_i la $i^{\text{ème}}$ colonne de $\frac{X_N}{\sqrt{n}}$. On pose

$$f(\mathbf{y}_1, \dots, \mathbf{y}_n) = \frac{1}{N} \text{Tr} \left(\frac{1}{n} X_N X_N^* - zI \right)^{-1} = \frac{1}{N} \text{Tr} \left(\sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^* - zI \right)^{-1}.$$

L'inégalité d'Efron-Stein donne $\text{var } f \leq 2^{-1} \sum_{i=1}^n \mathbb{E} |f - f'_i|^2$; il nous faut estimer la différence $f - f'_i$. En remarquant que

$$Q(z) = \left(\frac{1}{n} X_N X_N^* - zI \right)^{-1} = \left(\sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^* + \mathbf{y}_i \mathbf{y}_i^* - zI \right)^{-1}$$

et en utilisant la formule de perturbation de rang 1 pour la résolvante (cf. exemple 15), on obtient

$$f = \frac{1}{N} \text{Tr} \left\{ Q_i - \frac{Q_i \mathbf{y}_i \mathbf{y}_i^* Q_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right\}.$$

De même, avec \mathbf{y}'_i de même loi que \mathbf{y}_i mais indépendante de tous les $(\mathbf{y}_k; 1 \leq k \leq n)$, il vient

$$f'_i = \frac{1}{N} \text{Tr} \left(\sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^* + \mathbf{y}'_i (\mathbf{y}'_i)^* - zI \right)^{-1} = \frac{1}{N} \text{Tr} \left\{ Q_i - \frac{Q_i \mathbf{y}'_i (\mathbf{y}'_i)^* Q_i}{1 + (\mathbf{y}'_i)^* Q_i \mathbf{y}'_i} \right\},$$

soit

$$\begin{aligned} f - f'_i &= \frac{1}{N} \text{Tr} \left\{ \frac{Q_i \mathbf{y}'_i (\mathbf{y}'_i)^* Q_i}{1 + (\mathbf{y}'_i)^* Q_i \mathbf{y}'_i} - \frac{Q_i \mathbf{y}_i \mathbf{y}_i^* Q_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right\} \\ &= \frac{1}{N} \left\{ \frac{(\mathbf{y}'_i)^* Q_i^2 \mathbf{y}'_i}{1 + (\mathbf{y}'_i)^* Q_i \mathbf{y}'_i} - \frac{\mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right\} \end{aligned}$$

d'où l'on déduit que

$$\mathbb{E} |f'_i - f|^2 = \frac{1}{N^2} \mathbb{E} \left| \frac{(\mathbf{y}'_i)^* Q_i^2 \mathbf{y}'_i}{1 + (\mathbf{y}'_i)^* Q_i \mathbf{y}'_i} - \frac{\mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right|^2.$$

On a :

$$\begin{aligned} \mathbb{E} |f'_i - f|^2 &= \frac{1}{N^2} \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{(\mathbf{y}'_i)^* Q_i^2 \mathbf{y}'_i}{1 + (\mathbf{y}'_i)^* Q_i \mathbf{y}'_i} - \mathbb{E}_{\{i\}} \frac{(\mathbf{y}'_i)^* Q_i^2 \mathbf{y}'_i}{1 + (\mathbf{y}'_i)^* Q_i \mathbf{y}'_i} \right. \\ &\quad \left. + \mathbb{E}_{\{i\}} \frac{\mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} - \frac{\mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right|^2. \end{aligned}$$

On remarquera que du fait que les vecteurs \mathbf{y}_i et \mathbf{y}'_i ont même loi, les deux espérances conditionnelles introduites sont identiques. Comme la variance de la somme de variables indépendantes est égale à la somme des variances, et que

la variance d'une variable aléatoire X de carré sommable minimise son erreur quadratique :

$$\text{var}(X) = \inf_{a \in \mathbb{C}} \mathbb{E} |X - a|^2 ,$$

on obtient in fine

$$\begin{aligned} \mathbb{E} |f'_i - f|^2 &= \frac{2}{N^2} \mathbb{E} \text{var}_{\{i\}} \left\{ \frac{\mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right\} , \\ &\leq \frac{2}{N^2} \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{\mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} - \frac{\mathbb{E}_{\{i\}} \mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbb{E}_{\{i\}} \mathbf{y}_i^* Q_i \mathbf{y}_i} \right|^2 . \end{aligned}$$

L'idée est de faire apparaître les quantités recentrées

$$\mathbf{y}_i^* Q_i^2 \mathbf{y}_i - \mathbb{E}_{\{i\}} \mathbf{y}_i^* Q_i^2 \mathbf{y}_i \quad \text{et} \quad \mathbf{y}_i^* Q_i \mathbf{y}_i - \mathbb{E}_{\{i\}} \mathbf{y}_i^* Q_i \mathbf{y}_i$$

dont on maîtrisera mieux les variances. On introduit les notations suivantes pour alléger l'exposé :

$$A = 1 + \mathbf{y}_i^* Q_i \mathbf{y}_i , \quad B = \mathbf{y}_i^* Q_i^2 \mathbf{y}_i , \quad A^\circ = A - \mathbb{E}_{\{i\}} A \quad \text{et} \quad B^\circ = B - \mathbb{E}_{\{i\}} B .$$

Il vient

$$\frac{B}{A} - \frac{\mathbb{E}_{\{i\}} B}{\mathbb{E}_{\{i\}} A} = \frac{B \mathbb{E}_{\{i\}} A - A \mathbb{E}_{\{i\}} B + AB - AB}{A \mathbb{E}_{\{i\}} A} = \frac{-BA^\circ + AB^\circ}{A \mathbb{E}_{\{i\}} A} = \frac{B^\circ}{\mathbb{E}_{\{i\}} A} - \frac{BA^\circ}{A \mathbb{E}_{\{i\}} A}$$

et

$$\begin{aligned} \mathbb{E} |f'_i - f|^2 &\leq \frac{2}{N^2} \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{B^\circ}{\mathbb{E}_{\{i\}} A} - \frac{BA^\circ}{A \mathbb{E}_{\{i\}} A} \right|^2 \\ &\leq \frac{4}{N^2} \left(\mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{B^\circ}{\mathbb{E}_{\{i\}} A} \right|^2 + \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{BA^\circ}{A \mathbb{E}_{\{i\}} A} \right|^2 \right) . \quad (18) \end{aligned}$$

Les estimées suivantes

$$\frac{1}{|\mathbb{E}_{\{i\}} A|^2} \leq \frac{|z|^2}{\text{Im}(z)^2} , \quad (19)$$

$$\frac{|B|}{|A|} \leq \frac{1}{\text{Im}(z)} , \quad (20)$$

$$\mathbb{E}_{\{i\}} |A^\circ|^2 \leq \frac{K}{n \text{Im}(z)^2} , \quad (21)$$

$$\mathbb{E}_{\{i\}} |B^\circ|^2 \leq \frac{K}{n \text{Im}(z)^4} \quad (22)$$

nous permettent de conclure. En effet, en les réinjectant dans (18), il vient :

$$\mathbb{E} |f'_i - f|^2 \leq \frac{K}{n N^2 \text{Im}(z)^6} \implies \text{var } g_n(z) = \mathcal{O}_z \left(\frac{1}{N^2} \right) .$$

Pour parachever la preuve, il nous reste à établir les inégalités (19)-(22). Montrons (19). En remarquant que $\mathbb{E}_{\{i\}} A = 1 + \frac{\sigma^2}{n} \text{Tr } Q_i$, on note que $(-z\mathbb{E}_{\{i\}} A)^{-1}$ est la transformée de Stieltjes d'une probabilité et qu'à ce titre

$$\frac{1}{|-z\mathbb{E}_{\{i\}} A|} \leq \frac{1}{\text{Im}(z)} \implies \frac{1}{|\mathbb{E}_{\{i\}} A|^2} \leq \frac{|z|^2}{\text{Im}(z)^2} .$$

Montrons (20). On a

$$\frac{|B|}{|A|} \leq \frac{|B|}{\text{Im}(A)} = \frac{|\mathbf{y}_i^* Q_i^2 \mathbf{y}_i|}{\text{Im}(A)} \leq \frac{\mathbf{y}_i^* Q_i^* Q_i \mathbf{y}_i}{\text{Im}(A)} . \quad (23)$$

Il reste à calculer $\text{Im}(A)$. On procède comme lors de la preuve de la proposition 17 à l'aide du théorème spectral. On note $(\lambda_k^{\{i\}})$ les valeurs propres de $\sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^*$ et on pose $z = x + iy$. Le théorème spectral donne

$$\sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^* = U^{\{i\}} \text{diag} \left(\lambda_k^{\{i\}}; 1 \leq k \leq N \right) U^{\{i\}*} , \quad U^{\{i\}} U^{\{i\}*} = U^{\{i\}*} U^{\{i\}} = I_N ,$$

soit

$$\begin{aligned} A &= 1 + \mathbf{y}_i^* Q_i \mathbf{y}_i = 1 + \sum_{k=1}^N \mathbf{y}_i^* U^{\{i\}} \text{diag} \left(\frac{1}{\lambda_k^{\{i\}} - z} \right) U^{\{i\}*} \mathbf{y}_i \\ &= 1 + \mathbf{y}_i^* Q_i \mathbf{y}_i = 1 + \sum_{k=1}^N \left[\mathbf{y}_i^* U^{\{i\}} \right]_k \left(\frac{\lambda_k^{\{i\}} - x + iy}{|\lambda_k^{\{i\}} - z|^2} \right) \left[U^{\{i\}*} \mathbf{y}_i \right]_k , \end{aligned}$$

d'où l'on déduit que

$$\text{Im}(A) = y \sum_{k=1}^N \left[\mathbf{y}_i^* U^{\{i\}} \right]_k \left(\frac{1}{|\lambda_k^{\{i\}} - z|^2} \right) \left[U^{\{i\}*} \mathbf{y}_i \right]_k = y \mathbf{y}_i^* Q_i Q_i^* \mathbf{y}_i .$$

Il suffit de réinjecter cette identité dans (23) pour obtenir (20). Les estimées (21) et (22) s'obtiennent en appliquant la proposition 19 (voir aussi le corollaire 21). \square

On complète maintenant le corollaire 21, en donnant une estimée de la vraie variance.

Corollaire 25. *Soit X_N une matrice $N \times n$ dont les entrées sont i.i.d. centrées, de quatrième moment fini, et \mathbf{y}_i , Q_i tels que définis en section 2.1. Alors*

$$\text{var} (\mathbf{y}_i^* Q_i \mathbf{y}_i) = \mathcal{O}_z \left(\frac{1}{n} \right) .$$

Démonstration. On utilise la formule de la variance conditionnelle

$$\text{var} (\mathbf{y}_i^* Q_i \mathbf{y}_i) = \mathbb{E} \text{var}_{\{i\}} (\mathbf{y}_i^* Q_i \mathbf{y}_i) + \text{var} (\mathbb{E}_{\{i\}} \mathbf{y}_i^* Q_i \mathbf{y}_i) = \mathcal{O}_z \left(\frac{1}{n} \right) + \mathcal{O}_z \left(\frac{1}{n^2} \right) ,$$

en utilisant respectivement le corollaire 21 et la proposition 23. \square

3 Théorème de Marčenko-Pastur

Théorème 26. Soit X_N une matrice $N \times n$ dont les entrées X_{ij} sont des variables aléatoires i.i.d. vérifiant

$$\mathbb{E}(X_{ij}) = 0, \quad \mathbb{E}|X_{ij}|^2 = \sigma^2, \quad \mathbb{E}|X_{ij}|^4 < \infty.$$

On suppose que $N = N(n)$ et que

$$\lim_{n \rightarrow \infty} \frac{N}{n} = c \in (0, \infty),$$

condition notée $N, n \rightarrow \infty$ dans la suite. Alors la mesure spectrale L_n de la matrice $Z_n = \frac{1}{n} X_N X_N^*$ vérifie : presque sûrement,

$$L_n \xrightarrow[N, n \rightarrow \infty]{\text{etr}} \mathbb{P}_{MP}.$$

Remarque 27. Lorsque $c > 1$, on note la présence d'une masse de Dirac au point 0, pondérée par le facteur $(1 - \frac{1}{c})$, dans la distribution de Marčenko-Pastur. Il y a une interprétation matricielle très simple à cela. En effet, si $c > 1$, alors $N > n$ pour N, n assez grands. Cela signifie que le rang de la matrice X_N est au plus n et que la matrice $\frac{1}{n} X_N X_N^*$ admet la valeur propre zéro avec une multiplicité d'au moins $N - n$. La mesure spectrale L_N associée à $\frac{1}{n} X_N X_N^*$ va donc s'écrire

$$L_N = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i} = \frac{N-n}{N} \delta_0 + \frac{1}{N} \sum_{\text{autres } \lambda_i} \delta_{\lambda_i}.$$

Le premier terme converge précisément vers $(1 - \frac{1}{c}) \delta_0$, qui est la masse de Dirac pondérée apparaissant dans la distribution de Marčenko-Pastur.

Remarque 28. Dans le théorème 26, l'hypothèse optimale concernant les moments est d'avoir un deuxième moment $\mathbb{E}|X_{ij}|^2 = \sigma^2$ fini seulement.

3.1 Preuve du théorème de Marčenko-Pastur

3.1.1 Notations et identités importantes

On utilisera fréquemment les identités suivantes, variations autour de perturbations de rang 1 de la résolvante.

$$Q = Q_i - \frac{Q_i \mathbf{y}_i^* \mathbf{y}_i Q_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i}, \quad Q \mathbf{y}_i^* \mathbf{y}_i = \frac{Q_i \mathbf{y}_i^* \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i}. \quad (24)$$

On pourra facilement montrer que les fonctions suivantes

$$\begin{aligned} \alpha_1(z) &= -\frac{1}{z(1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1)}, & \alpha_2(z) &= -\frac{1}{z(1 + c_n \frac{\sigma^2}{N} \text{Tr } Q_1)}, \\ \alpha_3(z) &= -\frac{1}{z(1 + c_n \frac{\sigma^2}{N} \text{Tr } Q)} \end{aligned}$$

($z \in \mathbb{C}^+$) sont des transformées de Stieltjes de mesures de probabilité (cf. exemple 10). Cela implique alors les majorations suivantes :

$$\forall i \in \{1, 2, 3\}, \quad \forall z \in \mathbb{C}^+, \quad |\alpha_i(z)| \leq \frac{1}{\text{Im}(z)}. \quad (25)$$

3.1.2 Stratégie de la preuve

On veut montrer que

$$\forall z \in \mathbb{C}^+, \quad g_n(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} g_{\text{MP}}(z). \quad (26)$$

On en déduira alors que si \mathcal{D} est un ensemble dénombrable contenant un point d'accumulation, alors⁴ presque sûrement,

$$\forall z \in \mathcal{D}, \quad g_n(z) \xrightarrow[N, n \rightarrow \infty]{} g_{\text{MP}}(z). \quad (27)$$

En effet, supposons que (26) soit vraie. Posons $\mathcal{D} = \{z_k\}_{k \in \mathbb{N}}$; alors pour tout z_k il existe un ensemble Ω_k de probabilité 1 tel que la convergence

$$g_n(z_k) \xrightarrow[N, n \rightarrow \infty]{} g_{\text{MP}}(z_k)$$

ait lieu pour tout $\omega \in \Omega_k$. Il suffira alors de considérer $\tilde{\Omega} = \cap_{k \in \mathbb{N}} \Omega_k$, également de probabilité 1, pour obtenir (27). Cela nous assurera que presque sûrement, $L_n \xrightarrow[N, n \rightarrow \infty]{} \mathbb{P}_{\text{MP}}$ en vertu du théorème 7.

Pour établir (26), on va décomposer

$$g_n(z) - g_{\text{MP}}(z) = g_n(z) - \mathbb{E} g_n(z) + \mathbb{E} g_n(z) - g_{\text{MP}}(z)$$

puis

1. On utilisera l'estimation de la variance obtenue par l'inégalité d'Efron-Stein

$$\text{var}(g_n(z)) = \mathcal{O}\left(\frac{1}{n^2}\right)$$

pour conclure que presque sûrement $g_n(z) - \mathbb{E} g_n(z)$ tend vers 0.

2. On montrera ensuite que $\mathbb{E} g_n(z)$ satisfait approximativement l'équation satisfaite par g_{MP} :

$$zc\sigma^2 g_{\text{MP}}^2(z) + (z + \sigma^2(c-1)) g_{\text{MP}}(z) + 1 = 0 \quad (28)$$

3. Enfin, on montrera la stabilité de l'équation (28) ("équation proche \implies solutions proche").

4. On remarque qu'on inverse la propriété presque sûre avec le quantificateur $\forall z \in \{\cdot\}$ au prix d'une restriction de l'ensemble \mathbb{C}^+ à \mathcal{D} .

3.1.3 Equation satisfaite par $\mathbb{E} g_n(z)$

Proposition 29. *On pose $c_n = \frac{N}{n}$, alors*

$$z\sigma^2 c_n (\mathbb{E} g_n)^2 + (z + \sigma^2(c_n - 1))\mathbb{E} g_n + 1 = \mathcal{O}_z \left(\frac{1}{\sqrt{n}} \right)$$

Démonstration. On a

$$I = Q(Y_n Y_n^* - zI) = QY_n Y_n^* - zQ.$$

Soit

$$zQ = -I + QY_n Y_n^* = -I + Q \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^* = -I + \sum_{i=1}^n \frac{Q_i \mathbf{y}_i \mathbf{y}_i^*}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i}$$

où on a utilisé la seconde identité de (24). En prenant la trace puis l'espérance et en normalisant, il vient :

$$\begin{aligned} z \mathbb{E} g_n(z) &= -1 + \frac{1}{N} \sum_{i=1}^n \mathbb{E} \left\{ \frac{\mathbf{y}_i^* Q_i \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right\} \\ &= -1 + \frac{1}{N} \sum_{i=1}^n \mathbb{E} \left(1 - \frac{1}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i} \right) = -1 + \frac{1}{c_n} \left(1 - \mathbb{E} \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right) \end{aligned}$$

où on a utilisé le fait que la distribution de $(1 + \mathbf{y}_i^* Q_i \mathbf{y}_i)^{-1}$ ne dépend pas de i . Soit finalement

$$c_n z \mathbb{E} g_n(z) = -c_n + 1 - \mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right\} \quad (29)$$

On voudrait montrer

$$\mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right\} \simeq \frac{1}{1 + \mathbb{E} \mathbf{y}_1^* Q_1 \mathbf{y}_1} = \frac{1}{1 + \frac{\sigma^2}{n} \mathbb{E} \text{Tr} Q_1} \simeq \frac{1}{1 + \sigma^2 c_n \mathbb{E} g_n}.$$

En effet, en reportant cette approximation dans (29), on obtiendrait alors :

$$\sigma^2 c_n z (\mathbb{E} g_n)^2 + [z + \sigma^2(c_n - 1)] \mathbb{E} g_n + 1 \simeq 0,$$

qui est l'équation canonique de Marčenko-Pastur, satisfaite par g_{MP} . Évaluons la différence

$$\begin{aligned} &\mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right\} - \frac{1}{1 + \sigma^2 c_n \mathbb{E} g_n} \\ &= \mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right\} - \mathbb{E} \left\{ \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr} Q_1} \right\} \\ &\quad + \mathbb{E} \left\{ \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr} Q_1} \right\} - \mathbb{E} \left\{ \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr} Q} \right\} \\ &\quad + \mathbb{E} \left\{ \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr} Q} \right\} - \frac{1}{1 + \sigma^2 c_n \mathbb{E} g_n} \triangleq T_1 + T_2 + T_3. \end{aligned} \quad (30)$$

Commençons par le terme T_1 .

$$\begin{aligned}
|T_1| &= \left| \mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} - \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr } Q_1} \right\} \right| \\
&= \left| \mathbb{E} \mathbb{E}_{\{1\}} \left\{ \frac{\frac{\sigma^2}{n} \text{Tr } Q_1 - \mathbf{y}_1^* Q_1 \mathbf{y}_1}{(1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1)(1 + \sigma^2 n \text{Tr } Q_1)} \right\} \right| \\
&\stackrel{(a)}{\leq} \frac{|z|^2}{\text{Im}^2(z)} \mathbb{E} \left| \frac{\sigma^2}{n} \text{Tr } Q_1 - \mathbf{y}_1^* Q_1 \mathbf{y}_1 \right| \leq \frac{|z|^2}{\text{Im}^2(z)} \mathbb{E}^{1/2} \{ \text{var}_{\{1\}} (\mathbf{y}_1^* Q_1 \mathbf{y}_1) \} \\
&\stackrel{(b)}{\leq} \frac{|z|^2}{\text{Im}(z)} \mathcal{O}_z \left(\frac{1}{\sqrt{n}} \right) = \mathcal{O}_z \left(\frac{1}{\sqrt{n}} \right)
\end{aligned}$$

où (a) découle des inégalités (25) et (b), du corollaire 21. Le terme T_2 maintenant. On a

$$\begin{aligned}
|T_2| &= \left| \mathbb{E} \left\{ \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr } Q_1} - \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr } Q} \right\} \right| = \left| \mathbb{E} \frac{\frac{1}{n} \text{Tr } (Q_1 - Q)}{(1 + \frac{\sigma^2}{n} \text{Tr } Q_1)(1 + \frac{\sigma^2}{n} \text{Tr } Q)} \right| \\
&\stackrel{(a)}{\leq} \frac{|z|^2}{n \text{Im}^3(z)} = \mathcal{O}_z \left(\frac{1}{n} \right).
\end{aligned}$$

où (a) découle des inégalités (25) et de la proposition 17. Enfin, le terme T_3 . On a

$$\begin{aligned}
|T_3| &= \left| \mathbb{E} \left\{ \frac{1}{1 + \frac{\sigma^2}{n} \text{Tr } Q} - \frac{1}{1 + \frac{\sigma^2}{n} \mathbb{E} \text{Tr } Q} \right\} \right| = \left| \mathbb{E} \left\{ \frac{\frac{\sigma^2}{n} \text{Tr } Q - \frac{\sigma^2}{n} \mathbb{E} \text{Tr } Q}{(1 + \frac{\sigma^2}{n} \text{Tr } Q)(1 + \frac{\sigma^2}{n} \text{Tr } Q)} \right\} \right| \\
&\leq \frac{\sigma^2 c_n |z|^2}{\text{Im}^2(z)} (\text{var } g_n(z))^{1/2} = \mathcal{O}_z \left(\frac{1}{n} \right).
\end{aligned}$$

En rassemblant les estimées obtenues pour T_1 , T_2 et T_3 , on en déduit une estimée pour (30) qui réinjectée dans (29) donne finalement

$$c_n z \mathbb{E} g_n(z) = -c_n + 1 - \frac{1}{1 + c_n \sigma^2 \mathbb{E} g_n(z)} + \mathcal{O}_z \left(\frac{1}{\sqrt{n}} \right).$$

Soit

$$\begin{aligned}
z \sigma^2 c_n (\mathbb{E} g_n)^2 + (z + \sigma^2 (c_n - 1)) \mathbb{E} g_n + 1 &= (1 + c_n \sigma^2 \mathbb{E} g_n(z)) \mathcal{O}_z \left(\frac{1}{\sqrt{n}} \right) \\
&= \mathcal{O}_z \left(\frac{1}{\sqrt{n}} \right).
\end{aligned}$$

La proposition 29 est alors démontrée. \square

3.1.4 Stabilité de l'équation satisfaite par g_{MP}

On rappelle que $z \mapsto g_{\text{MP}}(z)$ est la transformée de Stieltjes de la distribution de Marčenko-Pastur, solution de l'équation

$$z c \sigma^2 X^2 + [z + \sigma^2 (c - 1)] X + 1 = 0, \quad z \in \mathbb{C}^+.$$

Lemme 30. Soit $z \in \mathbb{C}^+$ et $X = X(z)$ et $X_\delta = X_\delta(z)$ deux transformées de Stieltjes solutions des équations

$$\begin{aligned} zc\sigma^2 X^2 + [z + \sigma^2(c-1)]X + 1 &= 0, \\ zc_\delta\sigma^2 X_\delta^2 + [z + \sigma^2(c_\delta-1)]X_\delta + 1 &= \delta, \end{aligned}$$

avec $c, c_\delta > 0$. Alors

$$|X - X_\delta| = \mathcal{O}_z(|\delta|) + \mathcal{O}_z(|c - c_\delta|) .$$

Démonstration. Par unicité de la solution dans \mathbb{C}^+ de l'équation de Marčenko-Pastur, on a $X(z) = g_{\text{MP}}(z)$, en particulier, si $c > 1$

$$X(z) = \frac{1}{c}h(z) - \left(1 - \frac{1}{c}\right)\frac{1}{z} \quad (31)$$

où $h(z)$ est la transformée de Stieltjes d'une mesure de probabilité. Cela nous sera utile pour la suite.

On réécrit l'équation approchée.

$$zc\sigma^2 X_\delta^2 + [z + \sigma^2(c-1)]X_\delta + 1 = \delta - z(c_\delta - c)\sigma^2 X_\delta^2 - \sigma^2(c_\delta - c)X_\delta .$$

Du fait que X_δ est une transformée de Stieltjes, $|X_\delta| \leq \text{Im}^{-1}(z)$. Cela entraîne en particulier que

$$z(c_\delta - c)\sigma^2 X_\delta^2 = \mathcal{O}_z(|c_\delta - c|) , \quad \sigma^2(c_\delta - c)X_\delta = \mathcal{O}_z(|c_\delta - c|) .$$

L'équation approchée se réécrit donc

$$zc\sigma^2 X_\delta^2 + [z + \sigma^2(c-1)]X_\delta + 1 = \delta' \quad \text{avec} \quad \delta' = \delta + \mathcal{O}_z(|c_\delta - c|) .$$

En soustrayant les 2 équations, il vient :

$$\begin{aligned} zc\sigma^2 (X^2 - X_\delta^2) + [z + \sigma^2(c-1)](X - X_\delta) &= -\delta' , \\ \Leftrightarrow (X - X_\delta) [zc\sigma^2 (X + X_\delta) + z + \sigma^2(c-1)] &= -\delta' , \\ \Leftrightarrow X - X_\delta &= -\frac{\delta'}{z \left(1 + c\sigma^2 (X + X_\delta) + \frac{\sigma^2(c-1)}{z}\right)} , \\ \Leftrightarrow X - X_\delta &= -\frac{\delta'}{z \left(1 + c\sigma^2 X_\delta + c\sigma^2 \left(X + \left(1 - \frac{1}{c}\right)\frac{1}{z}\right)\right)} . \end{aligned}$$

Si $c \leq 1$ alors $c\sigma^2 X_\delta$, $c\sigma^2 X$ et $-c\sigma^2 \left(1 - \frac{1}{c}\right)\frac{1}{z}$ sont des transformées de Stieltjes de mesures, par conséquent

$$-\frac{1}{z \left(1 + c\sigma^2 X_\delta + c\sigma^2 \left(X + \left(1 - \frac{1}{c}\right)\frac{1}{z}\right)\right)}$$

est la transformée de Stieltjes d'une mesure de probabilité. En particulier

$$\left| -\frac{1}{z(1 + c\sigma^2 X_\delta + c\sigma^2 (X - (1 - \frac{1}{c}) \frac{1}{z}))} \right| \leq \frac{1}{\text{Im}(z)} \quad \Rightarrow \quad |X - X_\delta| \leq \frac{|\delta'|}{\text{Im}(z)} .$$

Si $c > 1$, le même argument s'applique en remplaçant $X + (1 - \frac{1}{c}) \frac{1}{z}$ par $c^{-1}h$ (cf. (31)). Il suffit maintenant de remplacer δ' par sa valeur $\delta + \mathcal{O}_z(|c - c_\delta|)$ pour conclure. \square

3.1.5 Fin de la preuve

En combinant la proposition 29 et le lemme 30, on obtient

$$\mathbb{E} g_n(z) - g_{\text{MP}}(z) = \mathcal{O}_z \left(\frac{1}{\sqrt{n}} \right) , \quad z \in \mathbb{C}^+ .$$

Comme d'autre part $\text{var } g_n(z) = \mathcal{O}_z(n^{-2})$, on obtient la majoration suivante :

$$\mathbb{P}(|g_n(z) - \mathbb{E} g_n(z)| > t) \leq \frac{\text{var } g_n(z)}{t^2} = \mathcal{O}_z \left(\frac{1}{n^2} \right) .$$

Le lemme de Borel-Cantelli permet de conclure : pour $z \in \mathbb{C}^+$,

$$g_n(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} g_{\text{MP}}(z) .$$

Le théorème 7 permet alors d'obtenir la convergence de la mesure spectrale L_n vers la probabilité de Marčenko-Pastur.

3.2 Calcul gaussien et théorème de Marčenko-Pastur isotrope

Quand les variables aléatoires sont gaussiennes et en particulier gaussiennes complexes, on dispose de techniques spécifiques, souvent assez rapides. On va les présenter dans cette section et démontrer une version isotrope du théorème de Marčenko-Pastur.

On dit qu'une variable aléatoire X suit une loi gaussienne complexe standard, notée $\mathcal{N}_{\mathbb{C}}(0, 1)$, si $X = \frac{U + iV}{\sqrt{2}}$ avec (U, V) un vecteur gaussien réel bidimensionnel, centré et de matrice de covariance l'identité. Pour une telle variable, on vérifie que

$$\mathbb{E} X = 0 , \quad \mathbb{E} (X^2) = 0 , \quad \mathbb{E} |X|^2 = 1 .$$

3.2.1 Formule d'intégration par parties et inégalité de Poincaré

On rappelle que si $z = x + iy$, alors

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) \quad \text{et} \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) .$$

Proposition 31. Soit $F : \mathbb{C}^{2n} \rightarrow \mathbb{C}$ une fonction différentiable dont la croissance ainsi que celle de ses dérivées partielles est au plus polynomiale. Soient X_1, \dots, X_n des v.a. i.i.d. de loi $\mathcal{N}_{\mathbb{C}}(0, 1)$, alors

$$\mathbb{E} X_i F(X_1, \dots, X_n, \bar{X}_1, \dots, \bar{X}_n) = \mathbb{E} \frac{\partial F}{\partial \bar{X}_i}.$$

Cette identité est souvent appelée **formule d'intégration par parties**. C'est une généralisation du cas réel unidimensionnel : si $Z \sim \mathcal{N}(0, 1)$,

$$\begin{aligned} \mathbb{E} Z F(Z) &= \int_{\mathbb{R}} u F(u) e^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}} \\ &= \left[-F(u) e^{-\frac{u^2}{2}} \right]_{u=-\infty}^{u=+\infty} + \int_{-\infty}^{\infty} F'(u) e^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}} = \mathbb{E} F'(Z). \end{aligned}$$

L'inégalité suivante est l'homologue complexe de l'**inégalité de Poincaré** pour un vecteur gaussien réel standard.

$$\text{var } F(X_1, \dots, X_n, \bar{X}_1, \dots, \bar{X}_n) \leq \sum_{i=1}^n \mathbb{E} \left| \frac{\partial F}{\partial X_i} \right|^2 + \sum_{i=1}^n \mathbb{E} \left| \frac{\partial F}{\partial \bar{X}_i} \right|^2,$$

voir par exemple [10, Théorème 3.20]. Soit X_N une matrice $N \times n$ à entrées i.i.d $\mathcal{N}_{\mathbb{C}}(0, 1)$, on note

$$Q_N(z) = \left(\frac{\sigma^2}{n} X_N X_N^* - z I_N \right)^{-1}.$$

On écrira souvent simplement Q à la place de Q_N . Les formules de différentiation suivantes seront d'un usage constant dans la suite.

$$\frac{\partial Q_{ij}}{\partial \bar{X}_{k\ell}} = -\frac{\sigma^2}{n} (Q X_N)_{i\ell} Q_{kj}, \quad (32)$$

$$\frac{\partial Q_{ij}}{\partial X_{k\ell}} = -\frac{\sigma^2}{n} Q_{ik} (X_N^* Q)_{\ell j}. \quad (33)$$

On notera le motif croisé en γ au niveau des indices : $i \rightarrow \ell \rightarrow k \rightarrow j$ quand on différencie par rapport au conjugué d'une entrée, et le motif en U : $i \rightarrow k \rightarrow \ell \rightarrow j$ quand on différencie par rapport à une entrée.

Exemple 32 (Dérivées partielles pour la transformée de Stieltjes). Soit $g_n(z) = \frac{1}{N} \text{Tr } Q(z)$, alors

$$\frac{\partial g_n(z)}{\partial \bar{X}_{k\ell}} = \frac{1}{N} \sum_{i=1:N} \frac{\partial Q_{ii}(z)}{\partial \bar{X}_{k\ell}} = -\frac{\sigma^2}{Nn} \sum_{i=1:N} (Q X_N)_{i\ell} Q_{ki} = -\frac{\sigma^2}{Nn} (Q^2 X_N)_{k\ell}.$$

De même,

$$\frac{\partial g_n(z)}{\partial X_{k\ell}} = \frac{1}{N} \sum_{i=1:N} \frac{\partial Q_{ii}(z)}{\partial X_{k\ell}} = -\frac{\sigma^2}{Nn} \sum_{i=1:N} Q_{ik} (X_N^* Q)_{\ell i} = -\frac{\sigma^2}{Nn} (X_N^* Q^2)_{\ell k}.$$

Exemple 33 (Dérivées partielles pour une forme quadratique de la résolvante). Soit \mathbf{u}, \mathbf{v} deux vecteurs déterministes $N \times 1$ et $h_n(z) = \mathbf{u}^* Q(z) \mathbf{v}$. On a

$$\frac{\partial h_n(z)}{\partial X_{k\ell}} = \sum_{i,j=1:N} \bar{u}_i v_j \frac{\partial Q_{ij}}{\partial X_{k\ell}} = -\frac{\sigma^2}{n} \sum_{i,j=1:N} \bar{u}_i v_j (QX)_{i\ell} Q_{kj} = -\frac{\sigma^2}{n} \mathbf{u}^* (QX)_{\cdot\ell} Q_{k\cdot} \mathbf{v}$$

où $(QX)_{\cdot\ell}$ désigne le $\ell^{\text{ième}}$ vecteur colonne et $Q_{k\cdot}$ le $k^{\text{ième}}$ vecteur ligne. De même

$$\frac{\partial h_n(z)}{\partial X_{k\ell}} = -\frac{\sigma^2}{n} \sum_{i,j=1:N} \bar{u}_i v_j Q_{ik} (X^* Q)_{\ell j} = -\frac{\sigma^2}{n} \mathbf{u}^* Q_{\cdot k} (X^* Q)_{\ell} \mathbf{v}$$

Exemple 34 (Estimation de la variance de g_n par techniques gaussiennes). En appliquant l'inégalité de Poincaré, il vient :

$$\text{var } g_n(z) \leq \sum_{k,\ell} \mathbb{E} \left| \frac{\partial g_n(z)}{\partial X_{k\ell}} \right|^2 + \sum_{k,\ell} \mathbb{E} \left| \frac{\partial g_n(z)}{\partial X_{k\ell}} \right|^2.$$

Estimons le premier terme du membre de droite.

$$\begin{aligned} \sum_{k,\ell} \mathbb{E} \left| \frac{\partial g_n(z)}{\partial X_{k\ell}} \right|^2 &\stackrel{(a)}{=} \frac{\sigma^4}{N^2 n^2} \mathbb{E} \text{Tr} (Q^2(z) X_N X_N^* Q^2(\bar{z})) \\ &= \frac{\sigma^2}{N^2 n} \mathbb{E} \text{Tr} \left(Q^2(z) \left(\frac{\sigma^2}{n} X_N X_N^* - zI + zI \right) Q^2(\bar{z}) \right) \\ &= \frac{\sigma^2}{N^2 n} \mathbb{E} \text{Tr} (Q(z) Q^2(\bar{z})) + \frac{z\sigma^2}{N^2 n} \mathbb{E} \text{Tr} (Q^2(z) Q^2(\bar{z})) \\ &\stackrel{(b)}{\leq} \frac{\sigma^2}{Nn} \frac{1}{\text{Im}^3(z)} + \frac{\sigma^2}{Nn} \frac{|z|}{\text{Im}^4(z)}. \end{aligned}$$

On a utilisé en (a) le fait que $Q^*(z) = Q(\bar{z})$ et en (b) le fait que la norme spectrale de Q est toujours majorée par $\text{Im}(z)^{-1}$. En estimant l'autre terme, on en déduit

$$\text{var } g_n(z) = \mathcal{O}_z \left(\frac{1}{n^2} \right),$$

ce qui est comparable à l'estimée obtenue en utilisant l'inégalité d'Efron-Stein (cette dernière étant néanmoins plus générale puisque s'appliquant à des entrées pas forcément gaussiennes).

Exemple 35 (Estimation de la variance de h_n par des techniques gaussiennes). On procède comme précédemment en appliquant l'inégalité de Poincaré :

$$\text{var } h_n(z) \leq \sum_{k,\ell} \mathbb{E} \left| \frac{\partial h_n(z)}{\partial X_{k\ell}} \right|^2 + \sum_{k,\ell} \mathbb{E} \left| \frac{\partial h_n(z)}{\partial X_{k\ell}} \right|^2.$$

Estimons le premier terme.

$$\begin{aligned} \sum_{k,\ell} \mathbb{E} \left| \frac{\partial h_n(z)}{\partial \overline{X_{k\ell}}} \right|^2 &= \frac{\sigma^4}{n^2} \mathbf{v}^* Q X X^* Q^* \mathbf{v} \times \mathbf{u}^* Q Q^* \mathbf{u} \\ &= \frac{\sigma^2}{n} \mathbf{v}^* Q \left(\frac{\sigma^2}{n} X X^* - zI + zI \right) Q^* \mathbf{v} \times \mathbf{u}^* Q Q^* \mathbf{u} = \mathcal{O}_z \left(\frac{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2}{n} \right). \end{aligned}$$

On en déduit, en estimant le terme restant de la même manière, que

$$\text{var } h_n(z) = \mathcal{O}_z \left(\frac{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2}{n} \right).$$

3.2.2 Preuve gaussienne du théorème de Marčenko-Pastur

L'estimée pour la variance de g_n associée au lemme de Borel-Cantelli nous assure que $g_n(z) - \mathbb{E} g_n(z) \rightarrow 0$ converge presque sûrement vers zéro. Reste à montrer que $\mathbb{E} g_n(z) - g_{\text{MP}}(z)$ tend vers zéro également.

L'identité $Q^{-1}Q = I$ donne pour l'entrée (i, j)

$$\left[\left(\frac{\sigma^2}{n} X X^* - zI \right) Q \right]_{ij} = \delta_{ij}$$

où δ_{ij} est le symbole de Kroneker, valant 1 si $i = j$, zéro sinon. En prenant l'espérance, il vient :

$$\frac{\sigma^2}{n} \sum_{k=1:n, \ell=1:N} \mathbb{E} X_{ik} [X^*]_{k\ell} Q_{\ell j} - z \mathbb{E} Q_{ij} = \delta_{ij}.$$

Dans la suite des calculs suivants, on va successivement utiliser l'identité d'intégration par parties, puis la formule de différentiation d'une entrée de la résolvante. On note que $[X^*]_{k\ell} = \overline{X_{\ell k}}$.

$$\begin{aligned} \frac{\sigma^2}{n} \sum_{k=1:n, \ell=1:N} \mathbb{E} \frac{\partial}{\partial \overline{X_{ik}}} \{ \overline{X_{\ell k}} Q_{\ell j} \} - z \mathbb{E} Q_{ij} &= \delta_{ij}, \\ \iff \frac{\sigma^2}{n} \sum_{k=1:n, \ell=1:N} \mathbb{E} \frac{\partial \overline{X_{\ell k}}}{\partial \overline{X_{ik}}} Q_{\ell j} + \frac{\sigma^2}{n} \sum_{k=1:n, \ell=1:N} \mathbb{E} \overline{X_{\ell k}} \frac{\partial Q_{\ell j}}{\partial \overline{X_{ik}}} - z \mathbb{E} Q_{ij} &= \delta_{ij}, \\ \iff \sigma^2 \mathbb{E} Q_{ij} - \frac{\sigma^2}{n} \sum_{k=1:n, \ell=1:N} \mathbb{E} \overline{X_{\ell k}} \frac{\sigma^2}{n} (Q X)_{\ell k} Q_{ij} - z \mathbb{E} Q_{ij} &= \delta_{ij}, \\ \iff \sigma^2 \mathbb{E} Q_{ij} - \frac{\sigma^2}{n} \mathbb{E} \text{Tr} \left(Q \frac{\sigma^2}{n} X X^* \right) Q_{ij} - z \mathbb{E} Q_{ij} &= \delta_{ij} \\ \iff \sigma^2 \mathbb{E} Q_{ij} - \frac{\sigma^2}{n} \mathbb{E} \text{Tr} \left(Q \left(\frac{\sigma^2}{n} X X^* - zI_N + zI_N \right) \right) Q_{ij} - z \mathbb{E} Q_{ij} &= \delta_{ij} \\ \iff \sigma^2 \mathbb{E} Q_{ij} - z \sigma^2 \frac{N}{n} \mathbb{E} Q_{ij} - \frac{\sigma^2}{n} \mathbb{E} \text{Tr} (Q) Q_{ij} - z \mathbb{E} Q_{ij} &= \delta_{ij} \end{aligned}$$

Soit finalement, en posant $c_n = \frac{N}{n}$:

$$\mathbb{E} Q_{ij} (\sigma^2(1 - c_n) - z) - z \frac{\sigma^2}{n} \mathbb{E} Q_{ij} \text{Tr} Q = \delta_{ij} \quad (34)$$

On spécifie maintenant $i = j$, puis on somme sur i en normalisant. Il vient

$$\mathbb{E} g_n(z) (\sigma^2(1 - c_n) - z) - z \sigma^2 c_n \mathbb{E} g_n^2 = 1 .$$

L'estimation pour la variance de g_n nous permet de découpler et d'obtenir

$$\mathbb{E} g_n^2(z) = (\mathbb{E} g_n(z))^2 + \mathcal{O}_z \left(\frac{1}{n^2} \right)$$

soit finalement

$$z \sigma^2 c_n (\mathbb{E} g_n(z))^2 + [z + \sigma^2(c_n - 1)] \mathbb{E} g_n(z) + 1 = \mathcal{O}_z \left(\frac{1}{n^2} \right) .$$

Il suffit maintenant d'invoquer la stabilité de l'équation canonique de Marčenko-Pastur pour conclure.

3.2.3 Version isotrope du théorème de Marčenko-Pastur

Théorème 36. *soit X_N une matrice $N \times n$ à entrées i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$ et $(\mathbf{u}_N)_{n \geq 1}$ et $(\mathbf{v}_N)_{n \geq 1}$ deux familles de vecteurs $N \times 1$ déterministes et de norme euclidienne uniformément bornée en N :*

$$\sup_N (\|\mathbf{u}_N\|, \|\mathbf{v}_N\|) \leq K < \infty .$$

Alors pour tout $z \in \mathbb{C}^+$

$$\mathbf{u}_N^* Q_N(z) \mathbf{v}_N - \langle \mathbf{u}_N, \mathbf{v}_N \rangle g_{\check{M}P}(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} 0 ,$$

où $\langle \mathbf{u}_n, \mathbf{v}_n \rangle = \mathbf{u}_N^* \mathbf{v}_N$ représente le produit scalaire entre les vecteurs \mathbf{u}_N et \mathbf{v}_N .

Remarque 37. 1. *Quelque soit la suite de vecteurs (\mathbf{u}_N) de norme 1, on a $\mathbf{u}_N^* Q_N(z) \mathbf{u}_N \rightarrow g_{\check{M}P}(z)$. Ainsi, la limite ne dépend pas de la direction du vecteur \mathbf{u}_N , ce qui explique l'utilisation du terme isotrope.*

2. *En choisissant pour vecteurs \mathbf{u}_N et \mathbf{v}_N les vecteurs canoniques \mathbf{e}_i et \mathbf{e}_j , on obtient les convergences presque sûres suivantes :*

$$Q_{ii}(z) \xrightarrow[N, n \rightarrow \infty]{} g_{\check{M}P}(z) , \quad Q_{ij}(z) \xrightarrow[N, n \rightarrow \infty]{} 0 .$$

Asymptotiquement, la résolvante ressemble ainsi à la matrice $g_{\check{M}P}(z) I_N$.

3. *Les conclusions du théorème 36 demeurent vraies si les entrées de la matrice X_N sont i.i.d., centrées de variance unité avec un moment fini d'ordre 4 ($\mathbb{E} |X_{ij}|^4 < \infty$, voir par exemple [18]).*

Démonstration. La preuve comporte 2 étapes :

- (a) montrer que $\mathbf{u}_N^* Q_N(z) \mathbf{v}_N$ se concentre suffisamment vite autour de son espérance (on notera que le calcul de l'exemple 35 montre que l'estimation de sa variance, en $\mathcal{O}_z(n^{-1})$, ne suffit pas dans ce cas) ;
- (b) établir la convergence de $\mathbb{E} \mathbf{u}_N^* Q_N(z) \mathbf{v}_N$ vers $g_{\text{MP}}(z) \langle \mathbf{u}_N, \mathbf{v}_N \rangle$. La preuve de ce dernier point repose sur la preuve gaussienne du théorème de Marčenko-Pastur.

Commençons par démontrer le point (a). On va établir l'estimée suivante :

$$\mathbb{E} |\mathbf{u}^* Q(z) \mathbf{v} - \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v}|^4 = \mathcal{O}_z \left(\frac{1}{n^2} \right). \quad (35)$$

On a

$$\begin{aligned} & \text{var} \left(|\mathbf{u}^* Q(z) \mathbf{v} - \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v}|^2 \right) \\ &= \mathbb{E} |\mathbf{u}^* Q(z) \mathbf{v} - \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v}|^4 - \left(\mathbb{E} |\mathbf{u}^* Q(z) \mathbf{v} - \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v}|^2 \right)^2 \\ &= \mathbb{E} |\mathbf{u}^* Q(z) \mathbf{v} - \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v}|^4 + \mathcal{O}_z \left(\frac{1}{n^2} \right). \end{aligned}$$

où l'estimée de la dernière ligne provient de l'étude faite à l'exemple 35. Il nous suffit donc d'estimer la variance initiale, ce que l'on va faire en utilisant l'inégalité de Poincaré.

Posons $h = \mathbf{u}^* Q(z) \mathbf{v}$ et $\varphi = |\mathbf{u}^* Q(z) \mathbf{v} - \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v}|^2$. on a

$$\varphi = h\bar{h} - h\mathbb{E}\bar{h} - \bar{h}\mathbb{E}h + |\mathbb{E}h|^2$$

Soit

$$\begin{aligned} \frac{\partial \phi}{\partial \bar{X}_{k\ell}} &= \frac{\partial h}{\partial \bar{X}_{k\ell}} \bar{h} + h \frac{\partial \bar{h}}{\partial \bar{X}_{k\ell}} - \frac{\partial h}{\partial \bar{X}_{k\ell}} \mathbb{E} \bar{h} - \mathbb{E} h \frac{\partial \bar{h}}{\partial \bar{X}_{k\ell}} \\ &= \frac{\partial h}{\partial \bar{X}_{k\ell}} (\bar{h} - \mathbb{E} \bar{h}) + (h - \mathbb{E} h) \frac{\partial \bar{h}}{\partial \bar{X}_{k\ell}} \end{aligned}$$

et

$$\mathbb{E} \left| \frac{\partial \phi}{\partial \bar{X}_{k\ell}} \right|^2 \leq 2 \text{var}(h) \mathbb{E} \left| \frac{\partial h}{\partial \bar{X}_{k\ell}} \right|^2 \implies \sum_{k,\ell} \mathbb{E} \left| \frac{\partial \phi}{\partial \bar{X}_{k\ell}} \right|^2 \leq 2 \text{var}(h) \sum_{k,\ell} \mathbb{E} \left| \frac{\partial h}{\partial \bar{X}_{k\ell}} \right|^2.$$

Les deux termes du membre de droite ont été estimés à l'exemple 35, ce qui nous donne l'estimée :

$$\sum_{k,\ell} \mathbb{E} \left| \frac{\partial \phi}{\partial \bar{X}_{k\ell}} \right|^2 = \mathcal{O}_z \left(\frac{1}{n^2} \right).$$

Il suffit maintenant d'appliquer l'inégalité de Poincaré (le terme supplémentaire faisant intervenir $\partial/\partial X_{k\ell}$ se calculant de la même manière) pour obtenir l'estimée souhaitée : $\text{var} \phi = \mathcal{O}_z(n^{-2})$. Cela établit du coup (35) et le lemme de

Borel-Cantelli nous permet d'obtenir la convergence presque sûre souhaitée : pour tout $z \in \mathbb{C}^+$,

$$\mathbf{u}^* Q(z) \mathbf{v} - \mathbf{u}^* \mathbb{E} Q(z) \mathbf{v} \xrightarrow[N, n \rightarrow \infty]{p.s.} 0 .$$

On démontre maintenant le point (b). En multipliant l'équation (34) de part et d'autre par les vecteurs \mathbf{u} et \mathbf{v} , il vient :

$$\mathbb{E} \mathbf{u}^* Q(z) \mathbf{v} (\sigma^2(1 - c_n) - z) - z \frac{\sigma^2}{n} \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v} \text{Tr} Q(z) = \langle \mathbf{u}, \mathbf{v} \rangle .$$

Les estimées des variances obtenues aux exemples 34 et 35 permettent de découpler le terme $\mathbb{E} \mathbf{u}^* Q(z) \mathbf{v} \text{Tr} Q(z)$ et d'obtenir :

$$\frac{1}{n} \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v} \text{Tr} Q(z) = \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v} \times \frac{1}{n} \mathbb{E} \text{Tr} Q(z) + \mathcal{O}_z \left(\frac{1}{n^{3/2}} \right) .$$

Soit finalement

$$\mathbb{E} \mathbf{u}^* Q(z) \mathbf{v} \{ (\sigma^2(1 - c_n) - z) - z \sigma^2 c_n \mathbb{E} g_n(z) \} = \langle \mathbf{u}, \mathbf{v} \rangle + \mathcal{O}_z \left(\frac{1}{n^{3/2}} \right) .$$

Il suffit maintenant de diviser sereinement, en remarquant que

$$\left| \frac{1}{-z(1 + \sigma^2 c_n \mathbb{E} g_n(z) - \sigma^2(1 - c_n)/z)} \right| \leq \frac{1}{\text{Im}(z)} ,$$

en tant que transformée de Stieltjes, ce qui permet d'absorber le dénominateur dans le \mathcal{O}_z :

$$\mathbb{E} \mathbf{u}^* Q(z) \mathbf{v} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{-z(1 + \sigma^2 c_n \mathbb{E} g_n(z)) + \sigma^2(1 - c_n)} + \mathcal{O}_z \left(\frac{1}{n^{3/2}} \right) .$$

On conclut aisément. \square

4 Le modèle de grandes matrices de covariance

Soit $X_N = (X_{ij})$ une matrice $N \times n$, à entrées i.i.d. et telle que

$$\mathbb{E} X_{ij} = 0, \quad \mathbb{E} |X_{ij}|^2 = 1, \quad \mathbb{E} |X_{ij}|^4 < \infty .$$

Soit R_N une matrice $N \times N$ déterministe, hermitienne, semi-définie positive et de norme spectrale uniformément bornée, i.e.

$$\mathbf{R} \triangleq \sup_N \|R_N\|_{\text{sp}} < \infty .$$

On pose $\Sigma_n = \frac{1}{\sqrt{n}} R_N^{1/2} X_N$.

Théorème 38. Soit L_N la mesure spectrale de la matrice $\Sigma_n \Sigma_n^*$, et g_n sa transformée de Stieltjes. Alors

(a) l'équation

$$t_n(z) = \frac{1}{N} \text{Tr} [-z (I_N + c_n R_N t_n(z)) + (1 - c_n) R_N]^{-1}, \quad z \in \mathbb{C}^+ \quad (36)$$

admet une unique solution $z \mapsto t_n(z)$ qui est la transformée de Stieltjes d'une mesure de probabilité, soit

$$t_n(z) = \int \frac{\mathcal{F}_n(d\lambda)}{\lambda - z}$$

où \mathcal{F}_n est une mesure de probabilité sur \mathbb{R}^+ .

(b) Pour tout $z \in \mathbb{C}^+$, presque sûrement $g_n(z) - t_n(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} 0$.

(c) Presque sûrement, pour toute fonction f continue bornée,

$$\int f(\lambda) L_N(d\lambda) - \int f(\lambda) \mathcal{F}_N(d\lambda) \xrightarrow[N, n \rightarrow \infty]{} 0.$$

On appelle t_n l'**équivalent déterministe** de la transformée de Stieltjes de g_n .

Remarque 39. L'équation (36) peut se réécrire à l'aide de la mesure spectrale de R_N . Soit L_N^R la mesure spectrale de R_N , i.e. $L_N^R = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(R_N)}$, alors (36) équivaut à :

$$t_n(z) = \int \frac{L_N^R(du)}{-z(1 + u c_n t_n(z)) + (1 - c_n)u}, \quad z \in \mathbb{C}^+.$$

Remarque 40. On n'a pas besoin de faire l'hypothèse que $\frac{N}{n} \rightarrow c \in (0, \infty)$, il suffit que

$$0 < \liminf_n \frac{N}{n} \leq \limsup_n \frac{N}{n} < \infty$$

pour que le théorème précédent soit valide.

Dans le cas où $\frac{N}{n} \rightarrow c$ et où $L_n^R \xrightarrow[N, n \rightarrow \infty]{etr} L_\infty^R$, le théorème précédent prend la forme suivante, avec de vraies limites cette fois, et pas des équivalents déterministes.

Théorème 41. Soit L_N la mesure spectrale de la matrice $\Sigma_n \Sigma_n^*$, et g_n sa transformée de Stieltjes. Soit L_N^R la mesure spectrale de la matrice R_N , si

$$L_N^R \xrightarrow[N, n \rightarrow \infty]{etr} L_\infty^R,$$

alors

(a) l'équation

$$t(z) = \int \frac{L_\infty^R(du)}{-z(1 + uct(z)) + (1 - c)u} , \quad z \in \mathbb{C}^+$$

admet une unique solution $z \mapsto t(z)$ qui est la transformée de Stieltjes d'une mesure de probabilité, soit

$$t(z) = \int \frac{\mathcal{F}(d\lambda)}{\lambda - z}$$

où \mathcal{F} est une probabilité sur \mathbb{R}^+ .

(b) Pour tout $z \in \mathbb{C}^+$, presque sûrement $g_n(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} t(z)$.

(c) Presque sûrement, pour toute fonction f continue bornée,

$$L_N \xrightarrow[N, n \rightarrow \infty]{etr} \mathcal{F} .$$

On pourra se référer à l'article historique [23], à l'article de référence [27] (voir aussi [3]) et aussi à l'article [17] pour les notions d'équivalents déterministes dans des modèles plus compliqués.

4.1 Exemples importants

Cas où R_N admet un nombre fini de valeurs propres

On suppose que la mesure spectrale de R_N prend la forme suivante

$$L_N^R = \frac{1}{N} \sum_{\ell=1}^K N_\ell \delta_{\lambda_\ell^R}$$

où K est fini et indépendant de N, n et où

$$\frac{N_\ell}{N} \xrightarrow[N, n \rightarrow \infty]{} \alpha_\ell > 0 .$$

Dans ce cas,

$$L_N^R \xrightarrow[N \rightarrow \infty]{etr} L_\infty^R = \sum_{\ell=1}^K \alpha_\ell \delta_{\lambda_\ell^R} ,$$

et la limite $t(z)$ de $g_n(z)$ satisfait l'équation :

$$t(z) = \sum_{\ell=1}^K \alpha_\ell \frac{1}{-z(1 + c\lambda_\ell^R t(z)) + (1 - c)\lambda_\ell^R} . \quad (37)$$

En réduisant au même dénominateur, on constate que $t(z)$ est solution d'un polynôme de degré $K + 1$. A l'exception du cas $K = 1$, où les solutions du polynôme sont explicites (équation canonique de Marčenko-Pastur), on ne peut guère espérer de solution explicite pour $t(z)$.

Remarque 42. On peut obtenir facilement (d'un point de vue numérique) la densité associée à la probabilité dont la transformée de Stieltjes est $t(z)$. La procédure suivante repose sur des résultats théoriques de Silverstein et Choi [28] :

1. Résoudre numériquement l'équation (37) pour $z = x \in \mathbb{R}^*$,
2. Choisir l'unique solution (quand elle existe) $t(x)$ telle que $\text{Im } t(x) > 0$,
3. La densité au point x est donnée par $f(x) = \frac{1}{\pi} \text{Im } t(x)$.

C'est ainsi qu'ont été obtenues les courbes de la figure 3.

Cas où R_N est une perturbation de l'identité

Dans cette section, on suppose que R_N est proportionnelle à une perturbation de l'identité, i.e.

$$R_N = \sigma^2 \left(I_N + \sum_{\ell=1}^K \theta_\ell \mathbf{u}_\ell \mathbf{u}_\ell^* \right)$$

où K est fixé et ne dépend pas de N, n , les θ_ℓ sont positifs et ne dépendent pas de N, n et les vecteurs (\mathbf{u}_ℓ) sont orthonormés.

On vérifie immédiatement que $R_N \mathbf{u}_\ell = \sigma^2(1 + \theta_\ell) \mathbf{u}_\ell$, par suite la mesure spectrale de R_N s'écrit

$$L_N^{\mathbf{R}} = \frac{1}{N} \sum_{\ell=1}^K \delta_{\sigma^2(1+\theta_\ell)} + \frac{N-K}{N} \delta_{\sigma^2}$$

et la convergence suivante est immédiate :

$$L_N^{\mathbf{R}} \xrightarrow[N \rightarrow \infty]{\text{etr}} \delta_{\sigma^2} \quad \Rightarrow \quad t(z) = \frac{1}{-z(1 + c\sigma^2 t(z)) + (1 - c)\sigma^2} .$$

On reconnaît l'équation canonique de Marčenko-Pastur ; on a donc convergence de la mesure spectrale vers la distribution de Marčenko-Pastur.

On verra en section 7 que même si le comportement asymptotique de la mesure spectrale n'est pas affecté par ce type de perturbations, il n'en va pas de même pour la plus grande valeur propre.

4.2 Éléments de preuve

On introduit ici une méthode heuristique permettant de "deviner" l'équation limite satisfaite par la transformée de Stieltjes de la mesure spectrale associée au modèle

$$\Sigma_N \Sigma_N^* = \frac{1}{n} R_N^{1/2} X_N X_N^* R_N^{1/2} \quad \text{avec} \quad \Sigma_N = \frac{1}{\sqrt{n}} R_N^{1/2} X_N$$

L'heuristique fonctionne dans le cas où la matrice R_N est diagonale. On suppose donc que $R_N = \text{diag}(\rho_1, \dots, \rho_N)$. On utilisera les notations suivantes

$$g_n(z) = \frac{1}{N} \text{Tr} (\Sigma_N \Sigma_N^* - z I_N)^{-1}, \quad \tilde{g}_n(z) = \frac{1}{n} \text{Tr} (\Sigma_N^* \Sigma_N - z I_n)^{-1}, \quad c_n = \frac{N}{n}.$$

Les ingrédients principaux sont les éléments suivants :

- ▷ Expression de l'élément diagonal de la résolvante (cf. proposition 16).
- ▷ Lien entre la résolvante g_n et la co-résolvante \tilde{g}_n :

$$\tilde{g}_n(z) = c_n g_n(z) + (1 - c_n) \left(-\frac{1}{z} \right).$$

- ▷ Perturbation de rang 1 de la résolvante : soit A une matrice hermitienne $n \times n$ alors

$$\frac{1}{n} \text{Tr} (A + \mathbf{u} \mathbf{u}^* - z I_n)^{-1} = \frac{1}{n} \text{Tr} (A - z I_n)^{-1} + \mathcal{O}_z \left(\frac{1}{n} \right).$$

- ▷ Approximation d'une forme quadratique $\frac{1}{n} x^* A x$ (où x est un vecteur à entrées i.i.d. centrées réduites) par la trace normalisée $\frac{1}{n} \text{Tr} A$. L'approximation est justifiée par l'estimation de la variance de la proposition 19, cf. aussi la remarque 20.

Dans la suite, Σ_i représentera la i ème ligne de la matrice Σ et $\Sigma_{(i)}$ représente la matrice Σ privée de sa i ème ligne. On commence par exprimer g_n comme la moyenne des éléments diagonaux de la résolvante.

$$\begin{aligned} g_n(z) &= \frac{1}{N} \sum_{i=1}^N q_{ii}(z) \\ &\stackrel{(a)}{=} -\frac{1}{N} \sum_{i=1}^N \frac{1}{z \left(1 + \Sigma_i \cdot \left(\Sigma_{(i)}^* \Sigma_{(i)} - z I_n \right)^{-1} \Sigma_i^* \right)} \\ &\stackrel{(b)}{\approx} \frac{1}{N} \sum_{i=1}^N \frac{1}{-z \left(1 + \frac{\rho_i}{n} \text{Tr} \left(\Sigma_{(i)}^* \Sigma_{(i)} - z I_n \right)^{-1} \right)} \\ &\stackrel{(c)}{\approx} \frac{1}{N} \sum_{i=1}^N \frac{1}{-z \left(1 + \frac{\rho_i}{n} \text{Tr} \left(\Sigma^T \Sigma - z I_n \right)^{-1} \right)} = \frac{1}{N} \sum_{i=1}^N \frac{1}{-z (1 + \rho_i \tilde{g}_n(z))} \\ &\stackrel{(d)}{\approx} \frac{1}{N} \sum_{i=1}^N \frac{1}{-z (1 + \rho_i c_n g_n(z)) + (1 - c_n) \rho_i} \end{aligned}$$

En (a), on utilise l'expression exacte, algébrique, d'un élément diagonal de la résolvante ; en (b), on approxime la forme quadratique obtenue au dénominateur

par la trace normalisée de la matrice ; en (c), on approxime en utilisant la formule de perturbation de rang un ; en (d), on utilise finalement le lien entre la corésolvante et la résolvante. L'équation approchée finalement obtenue

$$g_n(z) \approx \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + \rho_i c_n g_n(z)) + (1 - c_n) \rho_i}$$

est à rapprocher de l'équation dont t_n est solution :

$$t_n(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + \rho_i c_n t_n(z)) + (1 - c_n) \rho_i} .$$

5 Etude de la mesure limite pour le modèle de grandes matrices de covariance

Nous nous intéressons précisément ici à la structure du support de la distribution \mathcal{S} dans le cas de grandes matrices de covariance. Cette étude est fortement inspirée (à pas mal de simplifications de preuves près) par l'important article [28].

Au lieu de travailler ici avec la matrice $\Sigma_n \Sigma_n^*$, on s'intéressera de préférence à la matrice $\Sigma_n^* \Sigma_n$ de taille $n \times n$ (où, comme précédemment, $\Sigma_n = \frac{1}{\sqrt{n}} R_N^{1/2} X_N$ avec $X_N = (X_{ij})$ une matrice $N \times n$ à entrées i.i.d. et telle que $\mathbb{E} X_{ij} = 0$, $\mathbb{E} |X_{ij}|^2 = 1$ et $\mathbb{E} |X_{ij}|^4 < \infty$). Bien évidemment, avec les notations de la section précédente, on trouve que la mesure spectrale \tilde{L}_N de $\Sigma_n^* \Sigma_n$ vérifie

$$\tilde{L}_N = \frac{n - N}{n} \delta_0 + \frac{N}{n} L_N$$

et donc, dans le cas où $L_n^R \xrightarrow[N, n \rightarrow \infty]{etr} L_\infty^R$ et $\frac{N}{n} \rightarrow c \in (0, \infty)$,

$$\tilde{L}_N \xrightarrow[N, n \rightarrow \infty]{etr} \tilde{\mathcal{F}} = (1 - c) \delta_0 + c \mathcal{F}.$$

Lemme 43. *Dans les conditions du Théorème 41, l'équation*

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u \tilde{t}(z)} \right)^{-1}, \quad z \in \mathbb{C}^+$$

admet une unique solution $z \mapsto \tilde{t}(z)$ qui est la transformée de Stieltjes de $\tilde{\mathcal{F}}$.

Démonstration. En partant de l'équation

$$\frac{1}{\tilde{t}(z)} = -z + c \int \frac{u L_\infty^R(du)}{1 + u \tilde{t}(z)}$$

et en multipliant chaque côté par $\tilde{t}(z)$, on trouve

$$1 = -z\tilde{t}(z) + c \left(1 - \int \frac{L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)$$

de sorte qu'avec $czt(z) = 1 - c + z\tilde{t}(z)$, on trouve

$$t(z) = -\frac{1}{z} \int \frac{L_\infty^R(du)}{1 + u\tilde{t}(z)} = \int \frac{L_\infty^R(du)}{-z - zut(z)}$$

et finalement, à nouveau grâce à $czt(z) = 1 - c + z\tilde{t}(z)$,

$$t(z) = \int \frac{L_\infty^R(du)}{-z(1 + uct(z)) + (1 - c)u}.$$

□

L'équation

$$\tilde{t}(z) = \left(-z + c \int \frac{uL_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1}$$

est très intéressante, car elle permet immédiatement d'obtenir une formule d'inversion pour $\tilde{t}(z)$, à savoir

$$z = -\frac{1}{\tilde{t}(z)} + c \int \frac{uL_\infty^R(du)}{1 + u\tilde{t}(z)}.$$

Cette formule sera fondamentale dans ce qui suit.

L'objectif de cette section est de caractériser \mathcal{F} (ou de manière équivalente $\tilde{\mathcal{F}}$) de manière assez fine. Rappelons pour cela que la connaissance de l'équation régissant $t(z)$ nous permet de "tracer" empiriquement la forme de la loi \mathcal{F} du fait de la Proposition 2-8. Cependant, cette proposition n'assure pas que \mathcal{F} soit une mesure à densité et la tracé (qui se base sur une estimation de $t(x + iy)$ pour y faible) ne nous permet pas d'en définir le support. C'est le premier point que nous étudierons.

En fait, cette étude nous permettra, au delà de la caractérisation du support, de comprendre un phénomène crucial de *changement de variable* qui nous permettra de connecter plus immédiatement les lois L_∞^R et \mathcal{F} . Cette connexion sera fondamentale pour les questions d'inférences (à savoir étudier les propriétés de L_∞^R étant données celles de \mathcal{F}).

5.1 Compléments à la Proposition 2

Avant tout, nous devons ajouter quelques éléments techniques à la Proposition 2 nous permettant de déterminer la densité d'une mesure dont on connaît la transformée de Stieltjes.

Proposition 44. Soit g_μ la transformée de Stieltjes de la mesure μ de support réel et de masse finie. Supposons que $\lim_{y \searrow 0} \Im[g_\mu(x + iy)]$ existe dans un voisinage de x_0 . Alors μ a une densité en x_0 égale à

$$\frac{1}{\pi} \lim_{y \searrow 0} \Im[g_\mu(x_0 + iy)].$$

Démonstration. Dénotons $I(x) = \lim_{y \searrow 0} \Im[g_\mu(x + iy)]$. Par continuité de g_μ sur $\{z, \Im[z] > 0\}$, on peut choisir $\delta > 0$ tel que, pour tout t et y satisfaisant $|t - x| < \delta$ et $|y| < \delta$,

$$|I(x) - \Im[g_\mu(t + iy)]| < \varepsilon.$$

Prenons x_1, x_2 deux points de continuité de μ tels que $|x_i - x| < \delta$. On a alors par la Proposition 2-8 que, par exemple,

$$\left| \mu([x_1, x_2]) - \frac{1}{\pi} \int_{x_1}^{x_2} \Im[g_\mu(t + iy)] dt \right| < \varepsilon |x_2 - x_1|$$

de sorte que

$$\left| \frac{\mu([x_1, x_2])}{|x_2 - x_1|} - \frac{1}{\pi} \frac{1}{|x_2 - x_1|} \int_{x_1}^{x_2} \Im[g_\mu(t + iy)] dt \right| < \varepsilon.$$

Ainsi

$$\left| \frac{\mu([x_1, x_2])}{|x_2 - x_1|} - \frac{1}{\pi} I(x) \right| < \varepsilon + \left| \frac{1}{\pi} \frac{1}{|x_2 - x_1|} \int_{x_1}^{x_2} \Im[g_\mu(t + iy)] dt - \frac{1}{\pi} I(x) \right| < 2\varepsilon.$$

On peut alors prendre une séquence $\varepsilon_n \rightarrow 0$ et des $x_{1,n}, x_{2,n}$ points de continuités tels que

$$\left| \frac{\mu([x_{1,n}, x_{2,n}])}{|x_{2,n} - x_{1,n}|} - \frac{1}{\pi} I(x) \right| \rightarrow 0.$$

On généralise alors à toute séquence. Ceci montre que μ est dérivable en x avec dérivée $\frac{1}{\pi} I(x)$.

Il reste à démontrer que I est continue en x_0 . Pour cela, on utilise la continuité de $z \mapsto \Im[g_\mu(z)]$. Pour z avec $\Im[z] > 0$ et $|z - x_0| < \delta$, $|I(x_0) - \Im[g_\mu(z)]| < \varepsilon$. Pour x réel et $|x - x_0| < \delta$, il existe z avec $\Im[z] > 0$ tel que $\max(|z - x|, |z - x_0|) < \delta$ implique $|I(x) - \Im[g_\mu(z)]| < \varepsilon$. Et donc $|I(x) - I(x_0)| < 2\varepsilon$. \square

5.2 Application au modèle de covariance

Dans toute la suite, nous allons de manière fondamentale nous baser sur l'identité suivante. Pour tout $z_1, z_2 \in \mathbb{C} \setminus \mathbb{R}$,

$$(\tilde{t}(z_1) - \tilde{t}(z_2)) \left(1 - c \int \frac{\tilde{t}(z_1) \tilde{t}(z_2) u^2 L_\infty^R(du)}{(1 + u \tilde{t}(z_1))(1 + u \tilde{t}(z_2))} \right) = (z_1 - z_2) \tilde{t}(z_1) \tilde{t}(z_2). \quad (38)$$

Cette formule fait suite immédiatement à l'identité scalaire de la résolvante $a^{-1} - b^{-1} = a^{-1}b^{-1}(b - a)$,

$$\tilde{t}(z_1) - \tilde{t}(z_2) = \tilde{t}(z_1)\tilde{t}(z_2) \left(z_1 - z_2 + c \int \frac{(\tilde{t}(z_1) - \tilde{t}(z_2)) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))} \right).$$

En particulier, pour $z_1 = z \in \mathbb{C}^+$ et $z_2 = z^*$, on trouve

$$2i\Im[\tilde{t}(z)] \left(1 - c|\tilde{t}(z)|^2 \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(z)|^2} \right) = 2i\Im[z]|\tilde{t}(z)|^2$$

et donc, comme $\Im[z] > 0$ et $\Im[\tilde{t}(z)] > 0$ (en tant que transformée de Stieltjes d'une mesure réelle)

$$1 - c|\tilde{t}(z)|^2 \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(z)|^2} > 0 \quad (39)$$

pour tout $z \in \mathbb{C}^+$.

Rappelons ici que $\mathbb{C}^+ = \{z, \Im[z] > 0\}$. Avec les notations du modèle de covariance évoquées plus haut, nous avons tout d'abord le résultat suivant.

Théorème 45. *Pour tout $x \in \mathbb{R}^*$,*

$$\lim_{z \in \mathbb{C}^+ \rightarrow x} \tilde{t}(z) = \tilde{t}^\circ(z)$$

existe.

Cela implique donc par la Proposition 44 que $\tilde{\mathcal{F}}$ a une densité partout sur \mathbb{R}^* égale à $\frac{1}{\pi}\Im[\tilde{t}(z)]$.

Démonstration. La preuve se déroule en deux étapes. On montrera tout d'abord que $\tilde{t}(z)$ est bornée dans un voisinage $B \subset \mathbb{C}^+$ de x . On pourra alors extraire des sous-suites d'éléments de \mathbb{C}^+ convergeant vers x telles que \tilde{t} converge également sur ces sous-suites et on cherchera à montrer qu'alors \tilde{t} a la même limite sur ces sous-suites.

Étape 1. Prenons $B \subset \mathbb{C}^+$ un voisinage de $x \in \mathbb{R}^*$. Supposons que pour une suite $z_n \in B$ loin de zéro, $|\tilde{t}(z_n)| \rightarrow \infty$. Prenons également $z_0 \in \mathbb{C}^+$ fixe. Alors, par un calcul similaire à celui permettant d'obtenir (38), on trouve que

$$\begin{aligned} & (z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)) \left(1 - c \int \frac{\tilde{t}(z_n) \tilde{t}(z_0) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))} \right) \\ &= (z_n - z_0) c \int \frac{\tilde{t}(z_n) \tilde{t}(z_0) u L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}. \end{aligned} \quad (40)$$

Par Cauchy–Schwarz, nous avons les deux résultats :

$$\begin{aligned} \left| \int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))} \right| &\leq \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2}} \sqrt{\int \frac{|\tilde{t}(z_0)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_0)|^2}} \\ \left| \int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))} \right| &\leq \sqrt{\int \frac{|\tilde{t}(z_n)|^2 u^2 L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2}} \sqrt{\int \frac{|\tilde{t}(z_0)|^2 L_\infty^R(du)}{|1+u\tilde{t}(z_0)|^2}}. \end{aligned}$$

Et donc, d'après (39), dans la limite $n \rightarrow \infty$,

$$\begin{aligned} \limsup_n c \left| \int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))} \right| &< 1 \\ \limsup_n c \left| \int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1+u\tilde{t}(z_n))(1+u\tilde{t}(z_0))} \right| &< \infty. \end{aligned}$$

Attention : la première inégalité est stricte du fait que $c \int \frac{|\tilde{t}(z_0)|^2 u^\beta L_\infty^R(du)}{|1+u\tilde{t}(z)|^2} < 1$ alors qu'on a seulement $\limsup_n c \int \frac{|\tilde{t}(z_n)|^2 u^\beta L_\infty^R(du)}{|1+u\tilde{t}(z_n)|^2} \leq 1$.

Et donc, d'après ces relations replacées dans (40), on trouve que $\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| < \infty$. Mais d'après notre hypothèse de départ,

$$\limsup_n |z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)| \geq \limsup_n |z_n| \cdot |\tilde{t}(z_n)| - |z_0| \cdot |\tilde{t}(z_0)| \rightarrow \infty$$

(rappelons que $z_n \not\rightarrow 0$) ce qui est contradictoire.

Etape 2. Comme on sait que $\tilde{t}(z_n)$ est borné pour z_n dans un voisinage B de $x \in \mathbb{R}^*$, considérons deux suites $z_{1,n}$ et $z_{2,n}$ de limite commune x et telles que $\tilde{t}(z_{1,n}) \rightarrow \tilde{t}_1$ et $\tilde{t}(z_{2,n}) \rightarrow \tilde{t}_2$. D'après (38), on a alors

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left(1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))} \right) = (z_{1,n} - z_{2,n})\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}). \quad (41)$$

Dans cette expression, le membre de droite tend vers 0 lorsque $n \rightarrow \infty$. Pour prouver que $\tilde{t}_1 = \tilde{t}_2$, il s'agit alors de prouver que

$$\liminf_n \left| 1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))} \right| > 0.$$

L'inégalité de Cauchy–Schwarz n'est pas suffisante ici. L'approche que nous proposons consiste en l'étude de

$$\begin{aligned} &\Re \left[c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))} \right] \\ &= \frac{1}{4} \left[c \int \left| \frac{u\tilde{t}(z_{1,n})}{1+u\tilde{t}(z_{1,n})} + \frac{u\tilde{t}(z_{2,n})}{1+u\tilde{t}(z_{2,n})} \right|^2 L_\infty^R(du) - c \int \left| \frac{u\tilde{t}(z_{1,n})}{1+u\tilde{t}(z_{1,n})} - \frac{u\tilde{t}(z_{2,n})}{1+u\tilde{t}(z_{2,n})} \right|^2 L_\infty^R(du) \right] \end{aligned}$$

où on a utilisé ici l'identité de polarisation $\Re[\int f^*g] = \frac{1}{4}[\|f+g\|^2 - \|f-g\|^2]$. Pour le premier terme de droite, par $|a+b|^2 \leq 2(|a|^2 + |b|^2)$ et l'inégalité (39),

$$\limsup_n c \int \left| \frac{\tilde{t}(z_{1,n})}{1 + u\tilde{t}(z_{1,n})} + \frac{\tilde{t}(z_{2,n})}{1 + u\tilde{t}(z_{2,n})} \right|^2 L_\infty^R(du) \leq 4.$$

Ainsi,

$$\begin{aligned} 1 - \Re \left[c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))} \right] \\ \geq |\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})|^2 c \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(z_{1,n})|^2 |1 + u\tilde{t}(z_{2,n})|^2} \end{aligned}$$

où l'intégrale de droite, du fait de $\tilde{t}(z_{2,n}) < K$ pour une constante K donnée, est bornée inférieurement par une valeur non nulle (à moins que $L_\infty^R = \delta_0$, ce qu'on exclue évidemment). On utilise alors $|1 - w| = \sqrt{|1 - \Re[w]|^2 + |\Im[z]|^2} \geq |1 - \Re[w]|$ pour conclure que, lorsque $n \rightarrow \infty$,

$$\liminf_n \left| 1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_{1,n}))(1 + u\tilde{t}(z_{2,n}))} \right| \geq C|\tilde{t}_1 - \tilde{t}_2|$$

pour une certaine constante $C > 0$. Si $\tilde{t}_1 \neq \tilde{t}_2$, alors cette limite est non nulle et donc, en prenant (41) dans la limite $n \rightarrow \infty$, on trouve $\tilde{t}_1 = \tilde{t}_2$. Ce qui est contradictoire avec l'hypothèse $\tilde{t}_1 \neq \tilde{t}_2$ et implique donc que $\tilde{t}_1 = \tilde{t}_2$. \square

Avant de continuer, penchons nous plus précisément sur l'équation

$$\tilde{t}(z) = \left(-z + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)} \right)^{-1}$$

que, nous l'avons vu, peut se réécrire sous la forme

$$z = -\frac{1}{\tilde{t}(z)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(z)}.$$

Comme $\tilde{t}(z)$ est uniquement définie pour $z \in \mathbb{C}^+$ avec pour image $\tilde{t}(\mathbb{C}^+)$, on peut d'après ce qui précède définir explicitement son inverse $\tilde{t}(\mathbb{C}^+) \rightarrow \mathbb{C}^+$,

$$z(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}}$$

qu'il est tentant d'étendre sur \mathbb{C}^+ . De la même manière, souvenons nous que $\tilde{t}(z)$ doit s'étendre par continuité sur $\mathbb{R}^* \setminus \mathcal{S}$ avec \mathcal{S} le support de \mathcal{F} . On est donc à nouveau tenté de définir une nouvelle fonction, réelle cette fois-ci, définie par $\mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \text{supp}(L_\infty^R)\} \rightarrow \mathbb{R}$,

$$x(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}}.$$

Cette fonction semble naturellement être :

- ▷ sur $\tilde{t}(\mathbb{R}^* \setminus \mathcal{S})$, l'inverse de la restriction de (l'extension de) \tilde{t} à $\mathbb{R}^* \setminus \mathcal{S}$
- ▷ sur le reste de son ensemble de définition, une extension “sans réelle interprétation physique” de cet inverse.

En particulier, la motivation de ce qui va suivre est la suivante : en se rappelant que \tilde{t} , en tant que transformée de Stieltjes d'une mesure réelle, est nécessairement *croissante* sur $\mathbb{R} \setminus \mathcal{S}$ (rappelons que $\tilde{t}'(x) = \int (u-x)^2 \tilde{\mathcal{F}}(du) > 0$), on s'attend à ce que

- ▷ sur $\tilde{t}(\mathbb{R}^* \setminus \mathcal{S})$, $x(\tilde{t})$ soit une fonction croissante (en tant qu'inverse locale d'une fonction croissante)
- ▷ sur le reste de son ensemble de définition, $x(\tilde{t})$ ne puisse pas être croissante (sinon elle aurait une inverse locale satisfaisant la définition de $\tilde{t}(x)$ qui serait croissante).

Cette intuition, originellement due à Marcenko et Pastur dans leur article [23], n'a été rigoureusement prouvée que dans [28]. Nous donnons (une version parfois améliorée) de cette preuve ici.

Théorème 46. *Définissons $z(\cdot)$ et $x(\cdot)$ comme ci-dessus et rappelons que \mathcal{S} est le support de \mathcal{F} . Alors,*

1. *Si $x_0 \in \mathcal{S}$, $x_0 \neq 0$, est un point de densité positive, alors $\tilde{t}^\circ(x_0)$ est l'unique solution \tilde{t} de partie imaginaire positive de l'équation $z(\tilde{t}) = x_0$.*
2. *Si $x_0 \notin \mathcal{S}$, $x_0 \neq 0$, alors $\tilde{t}^\circ(x_0)$ appartient à l'ensemble $\mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \text{supp}(L_\infty^R)\}$ et est l'unique solution réelle de $x(\tilde{t}) = x_0$ telle que $x'(\tilde{t}) > 0$. Réciproquement, pour $\tilde{t}_0 \in \mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \text{supp}(L_\infty^R)\}$ tel que $x'(\tilde{t}_0) > 0$, $x(\tilde{t}_0) \notin \mathcal{S}$.*

Démonstration. Commençons par prouver le résultat 1. On a évidemment $0 \neq \tilde{t}^\circ(x_0) = \lim_{y \searrow 0} \tilde{t}(x_0 + iy)$ avec

$$x_0 + iy = -\frac{1}{\tilde{t}(x_0 + iy)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0 + iy)}.$$

Par ailleurs,

$$\left| \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0 + iy)} \right|^2 \leq \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(x_0 + iy)|^2} \leq \frac{1}{\Im[\tilde{t}(x_0 + iy)]^2}$$

puisque $|1 + u\tilde{t}(x_0 + iy)|^2 \geq t^2 \Im[\tilde{t}(x_0 + iy)]^2$. Du fait de la convergence de $\Im[\tilde{t}(x_0 + iy)]$ vers une valeur non nulle, le membre de droite est borné pour tout y petit. Cela nous permet d'appliquer le théorème de convergence dominée pour obtenir, dans la limite $y \searrow 0$,

$$x_0 = -\frac{1}{\tilde{t}^\circ(x_0)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0)}.$$

Cela nous donne l'existence de la solution.

Pour l'unicité, prenons $\tilde{t}_1 \neq \tilde{t}_2 \in \mathbb{C}^+$ tels que $z(\tilde{t}_1) = z(\tilde{t}_2) = x_0$ et considérons des boules $B_1 = B(\tilde{t}_1, \varepsilon)$ et $B_2 = B(\tilde{t}_2, \varepsilon)$ telles que $B(\tilde{t}_1, \varepsilon) \cap$

$B(\tilde{t}_2, \varepsilon) = \emptyset$. Alors $x_0 \in z(B_1) \cap z(B_2)$ qui est ouvert par le théorème de l'image ouverte ($z(\cdot)$ étant une fonction analytique aux voisinages concernés). Donc il existe $z' \in \mathbb{C}^+ \cap (z(B_1) \cap z(B_2))$ tel que $z' = z(\tilde{t}'_1) = z(\tilde{t}'_2)$ pour un certain couple $\tilde{t}'_1 \in B_1$ et $\tilde{t}'_2 \in B_2$. Mais cela est impossible car alors $\tilde{t}'_1 = \tilde{t}(z')$ et $\tilde{t}'_2 = \tilde{t}(z')$, ce qui rompt l'unicité de la définition de \tilde{t} sur \mathbb{C}^+ .

Passons maintenant au résultat 2. Supposons tout d'abord que $\tilde{t}^\circ(x_0) \neq 0$ (on montrera plus bas que c'est en effet toujours le cas). Comme $x_0 \notin \mathcal{S}$, \tilde{t} peut être défini en x_0 par $\tilde{t}(x_0) = \tilde{t}^\circ(x_0) = \int \frac{1}{u-x_0} \tilde{\mathcal{F}}(du)$ et, restreint à \mathbb{R} , on a évidemment $\tilde{t}'(x_0) > 0$. Ainsi \tilde{t} , maintenant vue comme une fonction complexe, est analytique dans un voisinage de $(x_0, \tilde{t}(x_0))$.

Remarquons alors que, pour tout \tilde{t} dans un voisinage de $\tilde{t}(x_0)$,

$$z(\tilde{t}) = -\frac{1}{\tilde{t}} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}} = -\frac{1}{\tilde{t}} + \frac{c}{\tilde{t}} \left(1 - \frac{1}{\tilde{t}} r \left(-\frac{1}{\tilde{t}} \right) \right)$$

où on définit $r(z) = \int (u-z)^{-1} L_\infty^R(du)$ la transformée de Stieltjes de la mesure L_∞^R . Comme $\tilde{t}(x_0) \neq 0$, dans la limite $\tilde{t} \in \mathbb{C}^+ \rightarrow \tilde{t}(x_0)$, on a alors

$$r \left(-\frac{1}{\tilde{t}} \right) \rightarrow x_0 + \frac{1}{\tilde{t}(x_0)} \in \mathbb{R}.$$

Comme cela est vrai aussi pour tout x dans un voisinage de x_0 , on a $r(-\frac{1}{\tilde{t}}) \rightarrow \ell \in \mathbb{R}$ dans la limite $\tilde{t} \rightarrow \tilde{t}'$ pour tout \tilde{t}' dans un ouvert contenant $\tilde{t}(x_0)$. Mais d'après la Proposition 44, cela implique que $-\frac{1}{\tilde{t}(x_0)} \notin \text{supp}(L_\infty^R)$ (puisque $\Im[r(-\frac{1}{\tilde{t}})] \rightarrow 0$ pour tout $\tilde{t} \rightarrow \tilde{t}'$ dans un voisinage de $\tilde{t}(x_0)$). Ainsi nous avons montré que $\tilde{t}(x_0) \notin \{\tilde{t}; -\frac{1}{\tilde{t}} \in \text{supp}(L_\infty^R)\}$.

Cela implique donc que $x(\tilde{t}(x_0))$ est bien défini et $\tilde{t}(x_0)$ est l'une des solutions de $x(\tilde{t}(x_0)) = -\frac{1}{\tilde{t}(x_0)} + c \int \frac{u L_\infty^R(du)}{1 + u\tilde{t}(x_0)}$ (par extension analytique possible de $z(\tilde{t})$ autour de $(\tilde{t}(x_0), x_0)$). Comme $\tilde{t}'(x_0) > 0$, nécessairement, par inversion locale, $x'(\tilde{t}(x_0)) > 0$.

Pour l'unicité, considérons $\tilde{t}_1 \neq \tilde{t}_2$ deux solutions réelles de $x(\tilde{t}_i) = x_0$ telles que $x'(\tilde{t}_i) > 0$. Alors,

$$(\tilde{t}_1 - \tilde{t}_2) \left(1 - c \int \frac{u^2 L_\infty^R(du)}{(1 + u\tilde{t}_1)(1 + u\tilde{t}_2)} \right) = 0.$$

Comme $\tilde{t}_1 \neq \tilde{t}_2$, nécessairement $c \int \frac{u^2 L_\infty^R(du)}{(1 + u\tilde{t}_1)(1 + u\tilde{t}_2)} = 1$. Mais par Cauchy-Schwarz et par (39) prise dans la limite des $z \rightarrow x_0$,

$$c \int \frac{u^2 L_\infty^R(du)}{(1 + u\tilde{t}_1)(1 + u\tilde{t}_2)} \leq \sqrt{c \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}_1|^2}} \sqrt{c \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}_2|^2}} \leq 1$$

avec la première inégalité devenant une égalité si et seulement si les fonctions $u\tilde{t}_1/(1 + u\tilde{t}_1)$ et $u\tilde{t}_2/(1 + u\tilde{t}_2)$ sont proportionnelles; ceci ne peut arriver que

si L_∞^R est un Dirac et on se trouve alors dans le cas de Marcenko–Pastur qui est complètement connu. Pour le cas général, on a donc inégalité stricte et donc contradiction avec l’hypothèse. D’où $\tilde{t}_1 = \tilde{t}_2$.

Supposons maintenant que $\tilde{t}^\circ(x_0) = 0$. Toujours par analyticité de \tilde{t} au voisinage de x_0 , écrivons

$$z_n \mathbf{i} \varepsilon_n = -1 + c \int \frac{u \mathbf{i} \varepsilon_n L_\infty^R(du)}{1 + u \mathbf{i} \varepsilon_n}$$

pour une certaine séquence $z_n \rightarrow x_0$ et $\mathbf{i} \varepsilon_n = \tilde{t}(z_n)$ avec $\varepsilon_n \searrow 0$. Dans la limite, on trouve, par convergence dominée, $0 = -1$, ce qui implique l’impossibilité d’avoir $\tilde{t}^\circ(x_0) = 0$.

Il nous reste à démontrer la relation réciproque. Si $\tilde{t}_0 \in \mathbb{R}^* \setminus \{\tilde{t}; -\frac{1}{\tilde{t}} \in \text{supp}(L_\infty^R)\}$, alors pour $z \in \mathbb{C}^+$ et \tilde{t} dans le voisinage de \tilde{t}_0 ,

$$(\tilde{t}(z) - \tilde{t}) \left(1 - c \int \frac{u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z))(1 + u\tilde{t})} \right) = (z - x(\tilde{t}))\tilde{t}(z)\tilde{t}.$$

En prenant $z \in \mathbb{C}^+ \rightarrow x(\tilde{t})$ et en utilisant Cauchy–Schwarz avec le fait que $x'(\tilde{t}) > 0$ au voisinage de \tilde{t}_0 (et donc $1 - c \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}|^2} > 0$), on a finalement $\tilde{t}(z) \rightarrow \tilde{t} \in \mathbb{R}$. Ainsi, $x(\tilde{t}_0)$ n’est pas dans le support de \mathcal{S} d’après la Proposition 44. \square

Une représentation graphique de la fonction $x(\tilde{t})$ est donnée en Figure 5. Le comportement de la fonction est très particulier. Par des arguments similaires à ceux évoqués dans les preuves, il peut précisément être démontré que :

- ▷ pour $\tilde{t}_1 \neq \tilde{t}_2$ et $x'(\tilde{t}_i) > 0$, $x(\tilde{t}_1) \neq x(\tilde{t}_2)$
- ▷ si $\tilde{t}_1 < \tilde{t}_2$ et $[\tilde{t}_1, \tilde{t}_2] \subset \mathbb{R}^* \subset \{\tilde{t}; -\frac{1}{\tilde{t}} \in \text{supp}(L_\infty^R)\}$, alors $x(\tilde{t}_1) < x(\tilde{t}_2)$.

Il n’est pas non plus toujours vrai que des asymptotes apparaissent comme en Figure 5 ; cela dépend de la forme spécifique de L_∞^R .

D’après la Figure 5, il est intéressant de noter l’existence ou non de dérivées nulles pour $x(\tilde{t})$ aux positions correspondant à l’entrée ou la sortie du support \mathcal{S} . Le signe de la dérivée prise au point d’inflexion nécessairement existant dans la représentation permet en particulier d’assurer l’existence ou non d’une partie croissante pour $x(\tilde{t})$ dans chaque intervalle de définition.

Nous concluons la section par la caractérisation, plus simple, de la masse en zéro.

Théorème 47. *Sous les hypothèses de travail de cette section, la mesure $\tilde{\mathcal{F}}$ a une masse en zéro donnée par*

$$\tilde{\mathcal{F}}(\{0\}) = \max(0, 1 - c(1 - L_\infty^R(\{0\}))).$$

Démonstration. On utilise ici la formule (voir Proposition 2-6)

$$\tilde{\mathcal{F}}(\{0\}) = \lim_{y \searrow 0} -\mathbf{i} y \tilde{t}(\mathbf{i} y) = 1 - c + \lim_{y \searrow 0} c \int \frac{L_\infty^R(du)}{1 + u\tilde{t}(\mathbf{i} y)}.$$

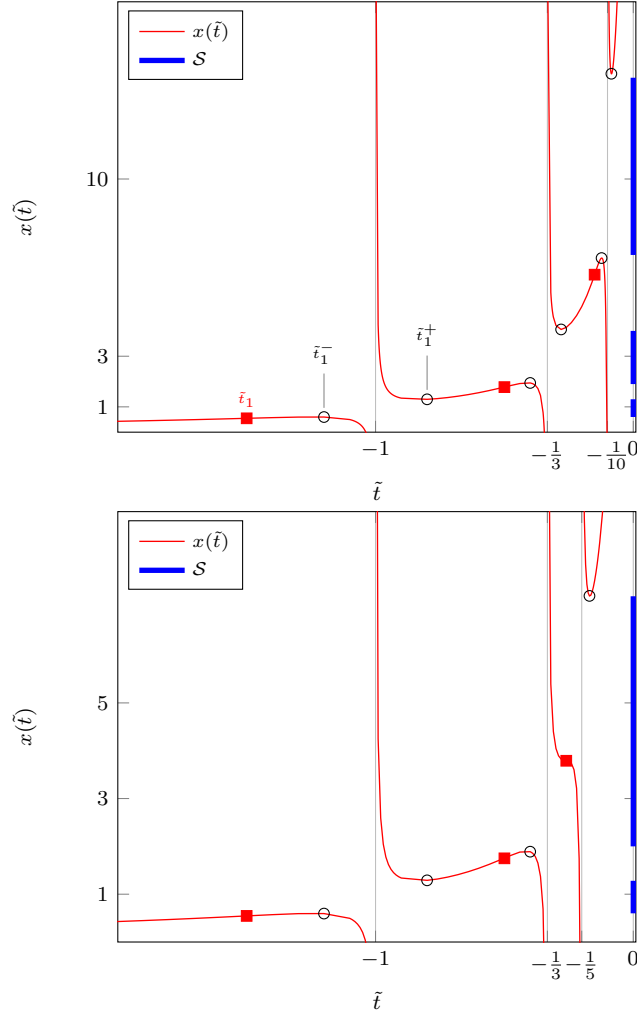


FIGURE 5 – $x(\tilde{t})$ pour \tilde{t} réel, R_N diagonal composé de trois masses en 1, 3 et 10 (figure supérieure) et 1, 3 et 5 (figure inférieure), $c = 1/10$ dans les deux cas. Extrema locaux marqués par des cercles, points d’inflexion par des carrés. Le support \mathcal{S} de \mathcal{F} peut être lu sur l’axe des ordonnées.

Grâce à l'identité (qui se démontre en dérivant le membre de gauche par rapport à u pour en trouver le maximum)

$$\frac{1}{|1 + u\tilde{t}(\mathbf{i}y)|^2} \leq 1 + \left(\frac{\Re[\tilde{t}(\mathbf{i}y)]}{\Im[\tilde{t}(\mathbf{i}y)]} \right)^2$$

on trouve que, si $\tilde{\mathcal{F}}(\{0\}) > 0$, alors $|y\Im[\tilde{t}(\mathbf{i}y)]| \rightarrow \tilde{\mathcal{F}}(\{0\})$ et $y\Re[\tilde{t}(\mathbf{i}y)] \rightarrow 0$, d'où par convergence dominée, $\tilde{\mathcal{F}}(\{0\}) = 1 - c(1 - L_\infty^R(\{0\}))$. \square

6 Inférence statistique pour le modèle de grandes matrices de covariance

Une des conséquence inattendue des résultats de la section précédente est qu'ils vont nous permettre d'effectuer de l'inférence statistique, en ce sens que, étant donné un estimateur pour la mesure \mathcal{F} , il va nous être possible d'évaluer certaines fonctionnelles de la mesure L_∞^R (et donc de sa version à dimension finie L_N^R). En particulier, une formule *a priori* seulement utile dans le cadre d'une preuve évoquée plus haut va se révéler fondamentale. A savoir la relation

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right)$$

avec

$$r(w) = \int \frac{L_\infty^R(du)}{u - w}.$$

Cette formule nous permet, en une seule écriture, de relier directement les transformées de Stieltjes $r(\cdot)$ et $\tilde{t}(\cdot)$ des informations cachée (le spectre de la matrice R_N) et visible (le spectre de $\Sigma_n^* \Sigma_n$) dans un contexte d'inférence.

6.1 Quelques rappels d'analyse complexe

Avant de poursuivre, rappelons rapidement quelques éléments d'analyse complexe.

Proposition 48 (Intégrale de Cauchy). *Soit \mathcal{C} un contour fermé orienté positivement dans \mathbb{C}^+ (et n'effectuant qu'une seule boucle). On a dans ce cas*

▷ *si z est un point de la surface décrite par \mathcal{C} , et f est holomorphe sur $U \supset \mathcal{C}$, alors*

$$\frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{f(w)}{w - z} dw = f(z).$$

▷ *si z n'est pas dans cette surface, alors sous les mêmes hypothèses*

$$\frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{f(w)}{w - z} dw = 0.$$

Proposition 49 (Calcul de résidus). *Soit \mathcal{C} un contour fermé orienté positivement dans \mathbb{C}^+ (et n'effectuant qu'une seule boucle) et f une fonction holomorphe sur un ouvert incluant \mathcal{C} . Soient a_1, \dots, a_L les points singuliers de f à l'intérieur du contour \mathcal{C} . Alors,*

$$\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) dz = \sum_{i=1}^L \text{Res}(f, a_i)$$

où $\text{Res}(f, a)$ est donné par

$$\text{Res}(f, a) = \frac{1}{(n-1)!} \lim_{z \rightarrow a} \frac{d^{n-1}}{dz^{n-1}} ((z-a)^n f(z))$$

avec $n \in \{1, 2, \dots\}$ le plus index tel que cette quantité est finie.

Proposition 50 (Théorème de convergence de Vitali). *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonction analytiques dans $D \subset \mathbb{C}$ telles que $|f_n(z)| \leq M$ pour tout $z \in D$. Supposons que $f_n(z_k) \rightarrow f(z_k)$ pour z_1, z_2, \dots un ensemble admettant un point d'accumulation dans D . Alors, f_n converge uniformément dans tout ensemble B strictement inclus dans D et f est analytique.*

6.2 Inférence statistique : Spécification des contours

Notre objectif dans cette section est d'évaluer une quantité du type

$$G(f) = \int f(t) L_{\infty}^R(dt)$$

pour une fonction f que nous voulons extensible en une fonction complexe analytique au voisinage de chaque composante connexe du support de L_{∞}^R . En particulier, nous pourrions demander à f d'être une quelconque fonction analytique sur une composante connexe de L_{∞}^R et nulle partout ailleurs. Encore plus spécifiquement, si $L_{\infty}^R = \sum_{\ell=1}^K \alpha_{\ell} \delta_{\lambda_{\ell}^R}$, nous pourrions souhaiter estimer les quantités λ_{ℓ}^R individuellement (auquel cas, si on connaît les α_{ℓ} , on prendra $f(t) = t$ au voisinage de λ_{ℓ}^R et zéro ailleurs).

Le cœur de l'approche consiste à observer que, grâce au théorème de Cauchy (Proposition 48),

$$G(f) = \int \left(\frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{f(w)}{w-t} dw \right) L_{\infty}^R(dt) = -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(w) r(w) dw$$

où $r(w) = \int (t-w)^{-1} L_{\infty}^R(dt)$, transformée de Stieltjes de L_{∞}^R .

Il suffit alors d'utiliser le lien entre $r(z)$ et $\tilde{t}(z)$ évoqué plus haut pour lier $G(f)$ à $\tilde{t}(z)$. En particulier, en réécrivant de manière plus élégante la relation

$$z = -\frac{1}{\tilde{t}(z)} + \frac{c}{\tilde{t}(z)} r\left(-\frac{1}{\tilde{t}(z)}\right)$$

nous avons

$$r\left(-\frac{1}{\tilde{t}}(z)\right) = -z\tilde{t}(z)t(z)$$

où nous rappelons que $-czt(z) = c - 1 - z\tilde{t}(z)$.

Evidemment, par analyticit  de \tilde{t} , t et r sur \mathbb{C}^+ et \mathbb{C}^- , si z parcourt un contour restreint   \mathbb{C}^+ ou \mathbb{C}^- , il est possible d'effectuer le changement de variable $w = -\frac{1}{\tilde{t}(z)}$, ce qu'on souhaite faire pour  crire $G(f)$ comme fonction de \tilde{t} et non plus de r . Cependant le contour \mathcal{C} traverse l'axe r el (afin de pouvoir encadrer les composante connexes du support de L_∞^R). Il s'agit alors de bien comprendre si le changement de variable $w = -\frac{1}{\tilde{t}(z)}$ supporte la travers e de l'axe r el.

En fait, pos e de telle mani re, la question n'a pas de r ponse claire. Cependant, en prenant le probl me diff remment, on parvient   prouver dans certains cas que le changement de variable est valable pour un contour \mathcal{C} qui convient (mais pas forc ment celui qu'on a construit au d part).

Nous avons vu pr c demment que pour $z \notin \mathcal{S}$, $-\frac{1}{\tilde{t}(z)} \notin \text{supp}(L_\infty^R)$ (Th or me 46-2). Ainsi il est possible de g n rer un contour $\mathcal{C}^\mathcal{F}$ qui entoure une ou plusieurs composantes connexes de \mathcal{S} de sorte que son image $-\frac{1}{\tilde{t}(\mathcal{C}^\mathcal{F})}$ ne traverse pas le support de L_∞^R . Les discussions de la section pr c dente, et en particulier la visualisation de la Figure 5, nous permet de comprendre ce changement de variable :

- ▷ si c est suffisamment faible ou alternativement, si les composantes connexes du support de L_∞^R sont suffisamment disjointes (cas de la Figure 5 sup rieure), alors, pour $\mathcal{C}^\mathcal{F}$ entourant la k -i me composante connexe de \mathcal{S} (disons, en partant de z ro et en excluant la masse en 0), l'image $-\frac{1}{\tilde{t}(\mathcal{C}^\mathcal{F})}$ entourera exactement la k -i me composante connexe de $\text{supp}(L_\infty^R)$;
- ▷ si au contraire on se trouve dans le cas de la Figure 5 inf rieure, alors l'image $-\frac{1}{\tilde{t}(\mathcal{C}^\mathcal{F})}$ pourra entourer plus d'une composante connexe de $\text{supp}(L_\infty^R)$. Dans ce cas, il sera impossible, par cette m thode, d'obtenir une estim e de $G(f)$ si f n'est non nulle que sur une des composante connexe de $\text{supp}(L_\infty^R)$ non dissociable par le changement de variable.

Cependant, de mani re g n rale, il sera tr s souvent (en fait pas forc ment toujours) possible d'estimer $G(f)$ si f est une fonction analytique du support de L_∞^R entier. Gr ce   l' tude effectu e dans la section pr c dente, en particulier l' tude du signe de la d riv e de la fonction $x(\tilde{t})$ aux points d'inflexions, il est possible d' tablir les zones dites de *s parabilit * de \mathcal{S} en autant de composantes connexes qu'en contient $\text{supp}(L_\infty^R)$ (toujours Dirac en z ro mis   part). Les Figures 6 et 7 donnent une repr sentation des zones de s parabilit  dans le cas o  L_∞^R est la somme de deux ou trois masses de Dirac.

La Figure 8 donne une id e visuelle de l'allure des $\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}^\mathcal{F})}$ pour $\mathcal{C}^\mathcal{F}$ entourant successivement chaque composante connexe de \mathcal{S} dans le cadre de la Figure 5 sup rieure.

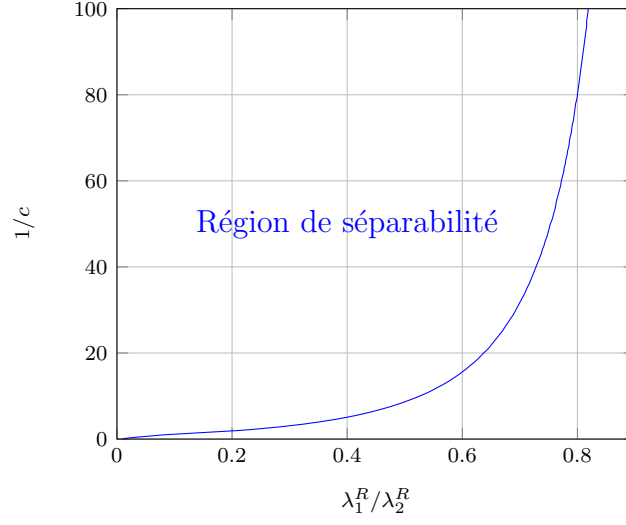


FIGURE 6 – Rapport limite c pour obtenir la séparabilité de $\lambda_1^R \leq \lambda_2^R$ pour $L_\infty^R = \frac{1}{2}(\delta_{\lambda_1^R} + \delta_{\lambda_2^R})$.

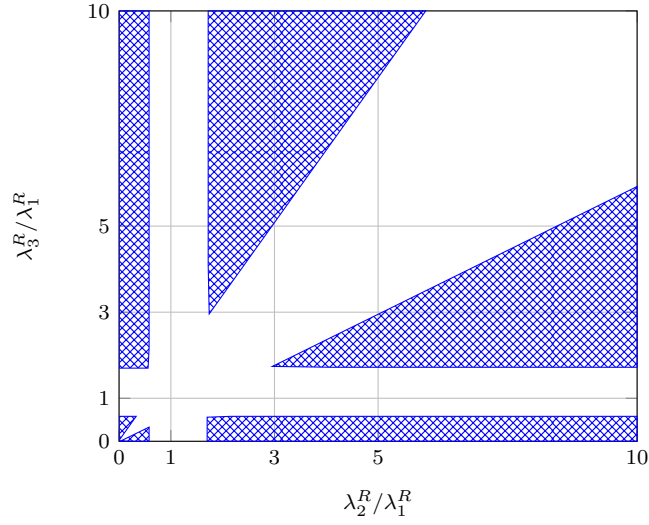


FIGURE 7 – Sous-ensemble des $\lambda_1^R \leq \lambda_2^R \leq \lambda_3^R$ (région hachurée) assurant la condition de séparabilité pour $L_\infty^R = \frac{1}{3}(\delta_{\lambda_1^R} + \delta_{\lambda_2^R} + \delta_{\lambda_3^R})$, $c = 1/10$.

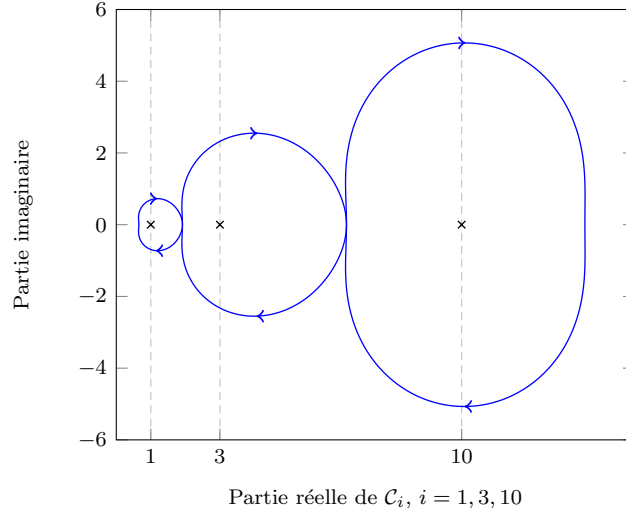


FIGURE 8 – Contours \mathcal{C}_1 , \mathcal{C}_3 et \mathcal{C}_{10} ante-images des $\mathcal{C}_i^{\mathcal{F}}$, contours circulaires entourant les composantes connexes de \mathcal{S} . Configuration de la Figure 5 supérieure.

6.3 Inférence statistique : Calcul des estimateurs

Ayant établi le domaine de validité du changement de variable, supposons que les conditions de séparabilité souhaitées sont remplies dans ce qui suit. Rappelons aussi que

$$G(f) = -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(w) r(w) dw$$

avec $r(w) = \int (u - w)^{-1} L_{\infty}^R(du)$ et ici $\mathcal{C} = -\frac{1}{\tilde{t}(\mathcal{C}^{\mathcal{F}})}$, où $\mathcal{C}^{\mathcal{F}}$ est un contour quelconque entourant une ou plusieurs composantes connexes de \mathcal{S} .

En opérant le changement de variable $w = -\frac{1}{\tilde{t}(z)}$, on trouve alors

$$G(f) = -\frac{1}{2\pi i} \oint_{\mathcal{C}^{\mathcal{F}}} f\left(-\frac{1}{\tilde{t}(z)}\right) r\left(-\frac{1}{\tilde{t}(z)}\right) \frac{\tilde{t}'(z)}{\tilde{t}(z)^2} dz.$$

Du fait de la relation

$$r\left(-\frac{1}{\tilde{t}(z)}\right) = -z\tilde{t}(z)t(z)$$

il suit que

$$G(f) = \frac{1}{2\pi i} \oint_{\mathcal{C}^{\mathcal{F}}} f\left(-\frac{1}{\tilde{t}(z)}\right) zt(z) \frac{\tilde{t}'(z)}{\tilde{t}(z)} dz.$$

Par ailleurs, dénotons, pour tout $z \in \mathbb{C} \setminus \mathcal{S}$ et pour tout n large,

$$\tilde{g}_n(z) = \frac{1}{n} \text{Tr} (\Sigma_n^* \Sigma_n - zI_n)^{-1}$$

(rappelons que nous avons défini dans les sections précédentes $g_n(z) = \frac{1}{N} \text{Tr} (\Sigma_n \Sigma_n^* - z I_N)^{-1}$). Il suit d'une application immédiate du Théorème 41-(b) que

$$\tilde{g}_n(z) - \tilde{t}(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} 0.$$

Il est donc tentant de remplacer $\tilde{t}(z)$ par $\tilde{g}_n(z)$ dans l'expression précédente de $G(f)$ pour obtenir un estimateur de $G(f)$. Cependant, pour s'assurer de la validité de cette opération, nous devons étendre Théorème 41-(b) à une version uniforme de la convergence sur le compacte $\mathcal{C}^{\mathcal{F}}$. Pour cela, nous ferons appel au théorème de Vitali (Proposition 50) et au résultat fondamental complémentaire suivant (que nous ne prouverons pas), adapté de [2] :

Théorème 51 (Pas de valeurs propres en dehors du support). *Dans les conditions du Théorème 41, supposons que $\lambda_1^R, \dots, \lambda_N^R$ les valeurs propres de R_N sont telles que*

$$\max_{1 \leq i \leq N} \{ \text{dist}(\lambda_i^R, \text{supp}(L_\infty^R)) \} \rightarrow 0.$$

Pour $\varepsilon > 0$, fixons également $A \supset \mathcal{S}$ l' ε -couverture de \mathcal{S} , support de \mathcal{F} . Alors, avec probabilité un, pour tout n large, il n'existe aucune valeur propre de $\Sigma_n \Sigma_n^$ dans A .*

Ce résultat, que nous reprendrons plus tard dans le cas de l'étude du modèle dit perturbé (voir Théorème 54) est bien plus fort que la convergence faible établie au Théorème 41 qui ne nous permettait jusqu'alors que d'assurer la convergence de la “masse globale” des valeurs propres de $\Sigma_n \Sigma_n^*$ vers la loi limite \mathcal{F} , mais pas d'établir des résultats pour les valeurs propres individuelles. Ici, le Théorème 51 garantit qu'il n'existe pas de valeur propre isolée pour n suffisamment large.

Plaçons nous donc sous les hypothèses du Théorème 51 (à savoir, nous demandons que R_N n'ait pas de valeur propre isolées). Il apparaît alors d'après le théorème que pour $z \in \mathcal{C}^{\mathcal{S}}$, sur un espace Ω_z de probabilité un, $\tilde{g}_n(z)$ est une fonction bornée pour tout n large, la borne et sa dépendance en n étant liée à la distance de $\mathcal{C}^{\mathcal{S}}$ à \mathcal{S} (et donc ne dépend pas de z). Afin de pouvoir appliquer le théorème de Vitali (Proposition 50) pour étendre la convergence à $\mathcal{C}^{\mathcal{S}}$ entier, on se place sur l'intersection $\Omega = \bigcap_{z_i} \Omega_{z_i}$ pour un ensemble discret de $\{z_i\}$ ayant un point d'accumulation dans $\mathcal{C}^{\mathcal{S}}$. Cet ensemble est de probabilité un. On applique alors Vitali pour obtenir

$$\sup_{z \in \mathcal{C}^{\mathcal{S}}} |\tilde{g}_n(z) - \tilde{t}(z)| \rightarrow 0$$

presque surement. Il vient alors, par l'analyticité des fonctions utilisées, que

$$G(f) - \frac{1}{2\pi i} \oint f \left(-\frac{1}{\tilde{g}_n(z)} \right) z g_n(z) \frac{\tilde{g}'_n(z)}{\tilde{g}_n(z)} dz \xrightarrow[N, n \rightarrow \infty]{p.s.} 0.$$

Ceci donne donc lieu à un estimateur consistant de $G(f)$. Mais on peut en général aller plus loin. Du fait des relations

$$\begin{aligned}\tilde{g}_n(z) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i - z} \\ g_n(z) &= \frac{n}{N} \tilde{g}_n(z) + \frac{N-n}{Nz} \\ \tilde{g}'_n(z) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{(\lambda_i - z)^2}\end{aligned}$$

avec $\lambda_1 \geq \dots \geq \lambda_n$ les valeurs propres de $\Sigma_n^* \Sigma_n$, les fonctions dans l'intégrande sont rationnelles. Si $f(w)$ est une fonction simple, alors on peut évaluer l'intégrale par calcul de résidu.

Exemple 52 (Monôme d'ordre 2). Prenons $f(w) = w^2$. Alors, avec probabilité un,

$$\begin{aligned}G(f) &= \frac{1}{2\pi i} \oint_{\mathcal{C}^F} z g_n(z) \frac{\tilde{g}'_n(z)}{\tilde{g}_n(z)^3} dz + o(1) \\ &= \frac{1}{2\pi i} \oint_{\mathcal{C}^F} \left(\frac{n}{N} \frac{z \tilde{g}'_n(z)}{\tilde{g}_n(z)^2} + \underbrace{\frac{n-N}{N} \frac{\tilde{g}'_n(z)}{\tilde{g}_n(z)^2}}_0 \right) dz + o(1) \\ &= \frac{1}{2\pi i} \oint_{\mathcal{C}^F} \frac{n}{N} \frac{z \tilde{g}'_n(z)}{\tilde{g}_n(z)^2} dz + o(1).\end{aligned}$$

Par intégration par partie, cela donne finalement

$$G(f) = -\frac{1}{2\pi i} \oint_{\mathcal{C}^F} \frac{n}{N} \frac{1}{\tilde{g}_n(z)} dz + o(1)$$

presque sûrement. En écrivant $\tilde{g}_n(z)$ sous forme d'une fonction rationnelle, on trouve alors que les pôles de l'intégrande sont les valeurs de z dans l'ensemble décrit par le contour \mathcal{C}^F telles que $\tilde{g}_n(z) = 0$. Grâce au lemme suivant, ces valeurs sont toutes identifiées.

Lemme 53. Soit $\Lambda \in \mathbb{R}^{n \times n}$ une matrice diagonale et $a \in \mathbb{R}^n$ un vecteur quelconque. Alors, les valeurs propres de $\Lambda - aa^*$ sont soit valeurs propres de Λ , soit racines de l'équation $1 = a^*(\Lambda - xI_n)^{-1}a$.

Démonstration. Il suffit de calculer, pour x différent d'une valeur propre de Λ , le déterminant

$$\begin{aligned}\det(\Lambda - aa^* - xI_n) &= \det(\Lambda - xI_n) \det(I_n - aa^*(\Lambda - xI_n)^{-1}) \\ &= \det(\Lambda - xI_n) (1 - a^*(\Lambda - xI_n)^{-1}a)\end{aligned}$$

par l'identité de Sylvester. □

Dans notre cas, on applique le lemme pour $\Lambda = \text{diag}(\lambda_i)$ et $a = \sqrt{\lambda/n}$ avec λ le vecteur des λ_i , de sorte que les valeurs propres de

$$\Lambda - \frac{1}{n} \sqrt{\lambda} \sqrt{\lambda}$$

sont racines de $1 = \frac{1}{n} \sum_{i=1}^n \lambda_i (\lambda_i - x)$, ou de manière équivalente de $0 = \frac{1}{n} \sum_{i=1}^n (\lambda_i - x) = \tilde{g}_n(x)$.

Par le lemme d'entrelacement de Weyl, il est clair que les λ_i sont entrelacés avec les solutions de $0 = g_n(x)$. Au contrôle près de la solution x_i correspondant à la valeur directement inférieure à la plus petite des valeurs propres λ_i d'une "composante connexe" du spectre de $\Sigma_n^* \Sigma_n$, on situe donc toutes ces solutions. Pour montrer que la dite valeur d'exception x_i reste proche du λ_i extrémal, on peut en fait effectuer notre calcul pour la fonction $f(w) = 1$ sur la composante connexe en question et se rendre compte que nécessairement elle doit être asymptotiquement incluse dans le contour, presque sûrement.

Il ne reste plus qu'à calculer le résidu associé à chacun de ces x_i . Précisément, on a d'après la Proposition 49,

$$\text{Res}(x_i) = \lim_{z \rightarrow x_i} \frac{n}{N} \frac{z - x_i}{\tilde{g}_n(z)}.$$

Le ratio de droite est une forme indéterminée dans la limite, de sorte qu'en appliquant la règle de l'Hopital (ou un développement limité au premier ordre), on trouve

$$\text{Res}(x_i) = \lim_{z \rightarrow x_i} \frac{n}{N} \frac{1}{\tilde{g}'_n(z)} = \frac{n}{N} \frac{1}{\tilde{g}'_n(x_i)}.$$

On en déduit finalement l'estimateur :

$$G(f) = \sum_{i \in S} \frac{n}{N} \frac{1}{\tilde{g}'_n(x_i)} + o(1)$$

presque sûrement, où l'ensemble S est l'ensemble des indices i tels que λ_i est inclus dans le contour \mathcal{C}^S .

7 Petites perturbations

Nous nous intéressons ici au modèle

$$\Sigma_N = \frac{1}{\sqrt{n}} R_N^{1/2} X_N$$

dans le cas où la matrice de population R_N est une petite perturbation de l'unité telle que définie en section 4.1

$$R_N = \sigma^2 \left(I_N + \sum_{\ell=1}^K \theta_\ell \mathbf{u}_\ell \mathbf{u}_\ell^* \right),$$

où K est un nombre fixé, les (θ_ℓ) sont des nombre positifs et les (\mathbf{u}_ℓ) des vecteurs orthonormés. On a vu précédemment que la mesure spectrale associée à la matrice $\Sigma_N \Sigma_N^*$ converge vers la loi de Marčenko-Pastur. On se pose maintenant la question de la convergence des plus grandes valeurs propres de la matrice, et comment les paramètres de la perturbation, (θ_ℓ) et (\mathbf{u}_ℓ) influent sur cette convergence.

L'étude de ce type de modèles a été motivée par I. Johnstone dans son article [20, Section 1.4] où il est fait mention de "*spiked models*".

On étudiera dans la suite une perturbation de rang 1 :

$$R_N = \sigma^2 (I_N + \theta \mathbf{u} \mathbf{u}^*) , \quad \|\mathbf{u}\| = 1 . \quad (42)$$

7.1 Etude de la plus grande valeur propre

Théorème 54 (Modèle non perturbé). *Considérons la matrice $\frac{\sigma}{\sqrt{n}} X_N$ de dimensions $N \times n$. C'est une matrice du type*

$$\frac{1}{\sqrt{n}} R_N^{1/2} X_N \quad \text{avec} \quad R_N = \sigma^2 I_N .$$

On suppose que $Nn^{-1} \rightarrow c > 0$. Alors, la plus grande valeur propre λ_{\max} de la matrice $\frac{\sigma^2}{n} X_N X_N^$ vérifie : presque sûrement,*

$$\lambda_{\max} \xrightarrow[N, n \rightarrow \infty]{} \sigma^2 (1 + \sqrt{c})^2 .$$

Autrement dit, la plus grande valeur propre converge presque sûrement vers le bord droit $\sigma^2(1 + \sqrt{c})^2$ du support de la distribution de Marčenko-Pastur.

Théorème 55 (Perturbation de rang 1). *Considérons la matrice $\Sigma_N = \frac{1}{\sqrt{n}} R_N^{1/2} X_N$ de dimensions $N \times n$, où R_N est donnée par (42), i.e.*

$$R_N = \sigma^2 (I_N + \theta \mathbf{u} \mathbf{u}^*) , \quad \|\mathbf{u}\| = 1 ,$$

et supposons que $Nn^{-1} \rightarrow c > 0$. Alors, de deux choses l'une :

1. *Si $\theta \leq \sqrt{c}$, la plus grande valeur propre λ_{\max} de la matrice $\frac{1}{n} \Sigma_N \Sigma_N^*$ vérifie*

$$\lambda_{\max} \xrightarrow[N, n \rightarrow \infty]{p.s.} \sigma^2 (1 + \sqrt{c})^2 .$$

2. *Si $\theta > \sqrt{c}$, alors la plus grande valeur propre λ_{\max} vérifie*

$$\lambda_{\max} \xrightarrow[N, n \rightarrow \infty]{p.s.} \sigma^2 (1 + \theta) \left(1 + \frac{c}{\theta} \right) .$$

Pour plus d'informations et la démonstration de ce type de résultats, on pourra se référer aux articles [4, 5, 6]. Pour une application statistique, cf. [7].

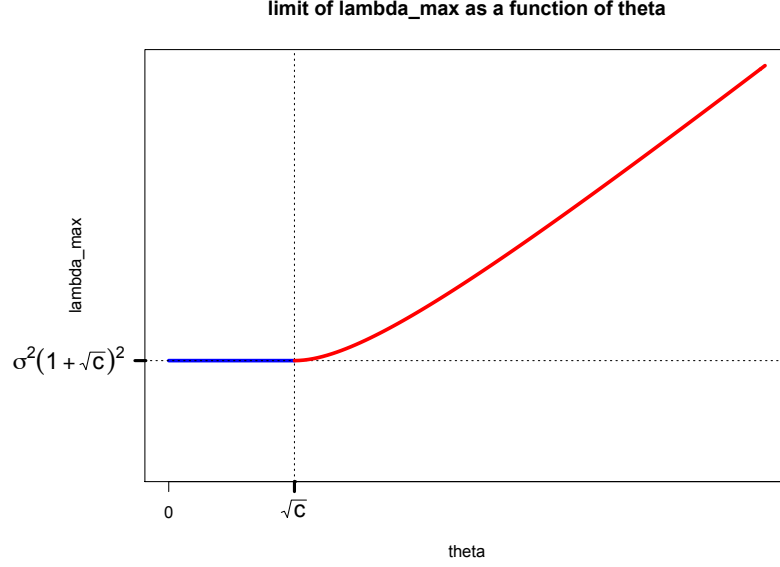


FIGURE 9 – Limite de la plus grande valeur propre λ_{\max} en tant que fonction de la perturbation θ

Remarque 56. On vérifie très facilement, par une étude de fonction élémentaire, que si $\theta > \sqrt{c}$, alors

$$\sigma^2(1 + \theta) \left(1 + \frac{c}{\theta}\right) > \sigma^2(1 + \sqrt{c})^2 ,$$

cf. la figure 7.1.

Remarque 57. Le résultat précédent illustre le fait que l'intensité de la perturbation θ a une influence sur le comportement de la plus grande valeur propre λ_{\max} : Celle-ci va converger vers le bord droit du support MP si la perturbation est suffisamment petite ($\theta < \sqrt{c}$) et va se détacher du support sinon, cf. figure 7.1.

Remarque 58. Les résultats précédents peuvent servir à tester la présence ou non d'un signal dans une série d'observations de grandes dimensions. Le lecteur intéressé pourra se reporter à l'article [7] où un tel test statistique est intégralement analysé.

7.2 Eléments de preuve

Dans cette section, on va développer de manière rapide et non rigoureuse des arguments permettant de déterminer le seuil \sqrt{c} au delà duquel la plus grande

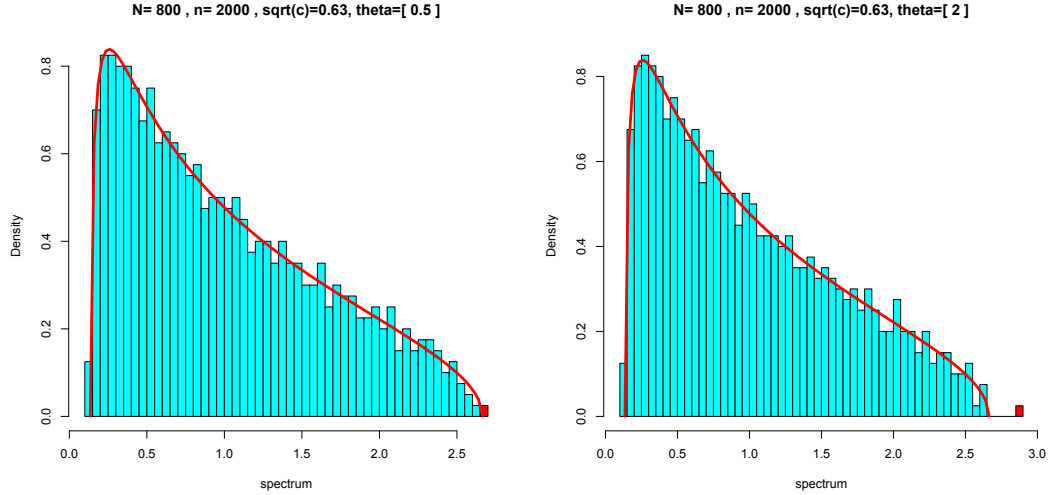


FIGURE 10 – Influence de la perturbation θ sur le comportement de la plus grande valeur propre. Dans les deux figures, $c = 0.63$; à gauche, $\theta = 0.5$, à droite $\theta = 2$. Le carré rouge correspond à la plus grande valeur propre; la ligne rouge correspond à la densité de Marčenko-Pastur associée à la valeur de c (ici, $\theta = 1$).

valeur propre se détache du support de la loi $\check{\text{MP}}$, ainsi que ceux permettant alors de trouver sa limite.

Stratégie de la preuve. On va procéder en 3 étapes.

1. On exprime d'abord une condition (condition du déterminant) pour laquelle

$$\lambda_{\max}(\Sigma_N \Sigma_N^*)$$

se détache du support $\check{\text{MP}}$.

2. En exploitant la théorie des grandes matrices aléatoires, on simplifie cette condition pour obtenir une condition asymptotique.
3. Enfin, on conclut, en obtenant la condition $\theta > \sqrt{c}$ pour laquelle la limite λ_{\max} se sépare du support $\check{\text{MP}}$, et en calculant cette limite.

Notations. On utilisera les notations suivantes, pour le modèle $\check{\text{MP}}$ non perturbé :

$$Z_N = \frac{\sigma^2}{n} X_N X_N^*, \quad Q_N(z) = (-zI_N + Z_N)^{-1};$$

pour le modèle perturbé

$$\tilde{Z}_N = \Sigma_N \Sigma_N^*.$$

On introduit la matrice $\Pi_N = I_N + \theta \mathbf{u} \mathbf{u}^*$, de sorte que

$$R_N = \sigma^2 \Pi_N \quad \text{et} \quad \tilde{Z}_N = \Pi_N^{\frac{1}{2}} Z_N \Pi_N^{\frac{1}{2}} .$$

On rappelle la formule d'inversion suivante (formule d'une perturbation de rang 1 de l'unité, cf. exemple 15) :

$$\Pi_N^{-1} = (I_N + \theta \mathbf{u} \mathbf{u}^*)^{-1} = I_N - \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^* .$$

Condition du déterminant. On veut trouver λ^θ valeur propre du modèle perturbé \tilde{Z}_N , telle que λ^θ ne soit pas valeur propre du modèle non perturbé Z_N , autrement dit, on veut que :

$$\det(-\lambda^\theta I_N + \tilde{Z}_N) = 0 \quad \text{et} \quad \det(-\lambda^\theta I_N + Z_N) \neq 0 .$$

En retravaillant cette condition, il vient :

$$\begin{aligned} \det(-\lambda^\theta I_N + \tilde{Z}_N) = 0 &\Leftrightarrow \det(-\lambda^\theta I_N + \Pi_N^{1/2} Z_N \Pi_N^{1/2}) = 0 \\ &\Leftrightarrow \det(-\lambda^\theta \Pi_N^{-1} + Z_N) = 0 \\ &\Leftrightarrow \det\left(-\lambda^\theta \left(I_N - \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^*\right) + Z_N\right) = 0 \\ &\Leftrightarrow \det\left(-\lambda^\theta I_N + Z_N + \lambda^\theta \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^*\right) = 0 \\ &\Leftrightarrow \det\left[(-\lambda^\theta I_N + Z_N) \left(I_N + \lambda^\theta \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^* Q_N(\lambda^\theta)\right)\right] = 0 \\ &\Leftrightarrow \det\left[I_N + \lambda^\theta \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^* Q_N(\lambda^\theta)\right] = 0 \end{aligned} \quad (43)$$

car $\det(-\lambda^\theta I_N + Z_N) \neq 0$. L'intérêt de cette expression réside dans le fait que les caractéristiques de la perturbation \mathbf{u} et θ sont clairement séparées du modèle non perturbé Q_N .

La matrice $\lambda^\theta \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^* Q_N(\lambda^\theta)$ étant de rang 1 admet la valeur propre 0 avec multiplicité $N - 1$. Pour que le déterminant précédent soit nul (cf. condition (43)), il faut que sa valeur propre non nulle soit égale à -1, ce qui entraîne que sa trace doit être égale à -1. Ainsi, la condition (43) se réexprime de la manière suivante :

$$\text{Tr} \left[\lambda^\theta \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^* Q_N(\lambda^\theta) \right] \Leftrightarrow \mathbf{u}^* Q_N(\lambda^\theta) \mathbf{u} = -\frac{1 + \theta}{\theta} .$$

Condition asymptotique. En utilisant maintenant le théorème de Marčenko-Pastur isotrope (théorème 36), il vient

$$\mathbf{u}^* Q_N(\lambda^\theta) \mathbf{u} \xrightarrow[N, n \rightarrow \infty]{p.s.} g_{\text{MP}}(\lambda^\theta) .$$

On obtient finalement la condition asymptotique suivante :

$$g_{\check{\text{MP}}}(\lambda^\theta) = -\frac{1+\theta}{\theta} .$$

Introduisons la fonction auxiliaire $\rho_{\check{\text{MP}}}(x) = 1 + xg_{\check{\text{MP}}}(x)$; la condition précédente devient :

$$\rho_{\check{\text{MP}}}(\lambda^\theta) = -\frac{1}{\theta} . \quad (44)$$

En utilisant le résultat du théorème 11-(3), on obtient une expression explicite pour la fonction $\rho_{\check{\text{MP}}}$, donnée par :

$$\rho_{\check{\text{MP}}}(x) = 1 + \frac{1}{2\sigma^2 c} \left\{ (\sigma^2(1-c) - x) + \sqrt{(x - \lambda^+)(x - \lambda^-)} \right\} ,$$

pour $x > \sigma^2(1 + \sqrt{c})^2$.

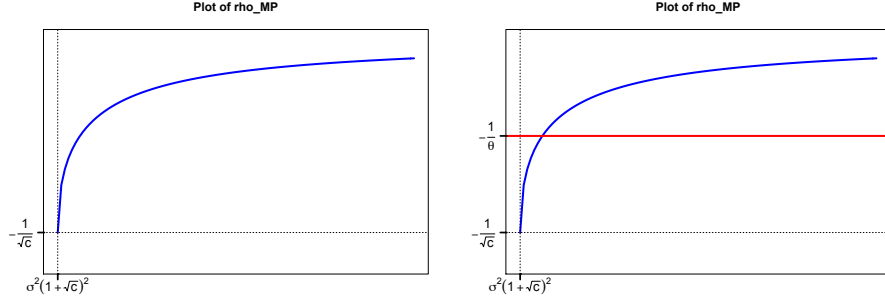


FIGURE 11 – Graphe de la fonction $\rho_{\check{\text{MP}}}$ sur $(\sigma^2(1 + \sqrt{c})^2, \infty)$

Condition de séparation de la plus grande valeur propre et sa limite.

La condition (44) est vérifiée dès lors que

$$\rho_{\check{\text{MP}}}(\lambda^\theta) = -\frac{1}{\theta} \Leftrightarrow -\frac{1}{\theta} > -\frac{1}{\sqrt{c}} \Leftrightarrow \theta > \sqrt{c} .$$

Calculons maintenant la limite λ^θ .

$$\rho_{\check{\text{MP}}}(\lambda^\theta) = -\frac{1}{\theta} \Leftrightarrow \lambda^\theta = \rho_{\check{\text{MP}}}^{-1}\left(-\frac{1}{\theta}\right) .$$

On cherche donc à inverser $\rho_{\check{\text{MP}}}$. En utilisant l'équation de $\check{\text{MP}}$ et la relation entre $g_{\check{\text{MP}}}$ et $\rho_{\check{\text{MP}}}$,

$$g_{\check{\text{MP}}}(x) = \frac{1}{\sigma^2(1-c) - x - x\sigma^2 c g_{\check{\text{MP}}}(x)} , \quad \rho_{\check{\text{MP}}}(x) = 1 + xg_{\check{\text{MP}}}(x) .$$

on obtient

$$x = \frac{\sigma^2}{\rho_{\text{MP}}(x)} (\rho_{\text{MP}}(x) - 1) (1 - c\rho_{\text{MP}}(x)) .$$

Remplaçons maintenant $x = \rho_{\text{MP}}^{-1}(-\frac{1}{\theta})$ dans les équations. On obtient

$$\lambda^\theta = \rho_{\text{MP}}^{-1}\left(-\frac{1}{\theta}\right) = \sigma^2(1 + \theta) \left(1 + \frac{c}{\theta}\right) ,$$

ce qui est le résultat souhaité.

8 Grandes matrices aléatoires et communications numériques

A la suite des travaux précurseurs de Telatar [30], Foschini [15], etc. la théorie des grandes matrices aléatoires a été appliquée de manière systématique à de nombreux problèmes de communications numériques sans fil. Les problèmes soulevés ont d'ailleurs motivé le développement de nombreux résultats mathématiques (cf. le livre [12] par exemple). On donne ici un aperçu de ces résultats dans le contexte des communications multi-antennes.

8.1 le Canal MIMO

Les années 90 ont vu l'émergence de nouvelles techniques de communication sans fil : au lieu de considérer une transmission sans fil point à point pour transmettre une information, l'intérêt s'est porté sur l'utilisation de plusieurs antennes en émission ainsi qu'en réception. Il en a résulté une augmentation substantielle de la capacité du canal, c'est à dire de la quantité d'information par unité de temps que l'on pouvait transmettre sans perte. L'étude de ce type de canaux, baptisés canaux MIMO (pour "*Multiple Input Multiple Output*") a engendré une activité très importante en sciences de l'ingénieur.

Dans un canal MIMO, le signal reçu \mathbf{y} s'écrit

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

où \mathbf{x} est un vecteur de dimension t représentant le vecteur transmis, \mathbf{H} est une matrice de dimensions $r \times t$ représentant le canal et \mathbf{n} est un vecteur de dimension r représentant le bruit blanc additif (de matrice de covariance $\sigma^2 I_r$).

Soyons plus précis :

- Il y a t antennes en émission et la $j^{\text{ième}}$ composante du vecteur \mathbf{x} représente le signal transmis par la $j^{\text{ième}}$ antenne.
- De même, il y a r antennes en réception et la $i^{\text{ième}}$ composante de \mathbf{y} représente le signal reçu par la $i^{\text{ième}}$ antenne.
- La $i^{\text{ième}}$ composante du vecteur \mathbf{n} représente le bruit reçu par la $i^{\text{ième}}$ antenne.
- Enfin, l'entrée H_{ij} de la matrice \mathbf{H} représente le gain complexe entre l'antenne de transmission j et l'antenne de réception i .

La matrice de canal \mathbf{H}

Dans la suite, on fera l'hypothèse que chaque entrée H_{ij} est aléatoire. Cette hypothèse est fondée sur le fait qu'entre une antenne en émission j et une antenne en réception i , il y a un grand nombre de chemins indépendants possibles pour le signal. Le gain H_{ij} apparaît alors comme une somme de variables aléatoires complexes et un argument du type "Théorème de la Limite Centrale" entraîne le fait que H_{ij} peut-être modélisé comme une variable aléatoire gaussienne complexe.

De ce point de vue, les caractéristiques physiques du canal se traduiront par des hypothèses probabilistes sur la loi de la matrice aléatoire \mathbf{H} . D'un point de vue du récepteur, celui-ci connaîtra la réalisation de la matrice \mathbf{H} ; en revanche, l'émetteur n'aura accès qu'à la distribution de \mathbf{H} .

Dans son article pionnier [30], Telatar fait l'hypothèse que les entrées de la matrice \mathbf{H} sont des variables aléatoires i.i.d. complexes gaussiennes centrées et la justifie ainsi :

"Equivalently, each entry of \mathbf{H} has uniform phase and Rayleigh magnitude. This choice models a Rayleigh fading environment with enough separation within the receiving antennas and the transmitting antennas such that the fades for each transmitting-receiving antenna pair are independent".

Il se peut néanmoins que cette hypothèse de séparation entre les antennes ne soit pas satisfaite et il faut alors tenir compte d'une certaine corrélation entre les gains.

Dans le modèle de Rice, populaire car à la fois réaliste et permettant une exploitation mathématique, la matrice de gain \mathbf{H} s'écrit :

$$\mathbf{H} = \mathbf{A} + \mathbf{R}^{1/2} \mathbf{W} \mathbf{T}^{1/2} .$$

Dans ce modèle, la matrice \mathbf{A} est déterministe et modélise le trajet moyen de la transmission (en anglais "*Line-of-sight component*"); la matrice \mathbf{T} est déterministe et modélise les corrélations entre les antennes à l'émission; la matrice \mathbf{R} est déterministe et modélise les corrélations entre antennes à la réception. Les matrices \mathbf{R} et \mathbf{T} sont semi-définies positives.

Performances du canal de transmission

On considèrera deux scénarios : canal à évanouissement rapide et lent.

Dans le cas du canal à évanouissement rapide ("*fast fading channel*"), à chaque vecteur \mathbf{x} transmis, de covariance \mathbf{Q} , correspond une nouvelle réalisation du canal \mathbf{H} . L'information mutuelle entre \mathbf{x} et \mathbf{y} est alors donnée par

$$\mathbb{E} \log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2 r} \mathbf{H} \mathbf{Q} \mathbf{H}^* \right) ,$$

où σ^2 est la variance du bruit \mathbf{n} . La distribution du canal étant connue à l'émetteur, on pourra (tenter d')optimiser la matrice de covariance \mathbf{Q} , par

précodage par exemple, pour obtenir la capacité ergodique :

$$\sup_{\substack{\mathbf{Q} \geq 0 \\ \frac{1}{t} \text{Tr}(\mathbf{Q}) \leq 1}} \mathbb{E} \log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2 r} \mathbf{H} \mathbf{Q} \mathbf{H}^* \right) ,$$

où la contrainte de trace s'interprète simplement comme une contrainte sur l'énergie totale en transmission. D'un point de vue de la théorie de l'information, la capacité ergodique représente le taux théorique maximum auquel on peut transmettre des données sans perte dans le scénario de canal à évanouissement rapide.

Dans le scénario de canal à évanouissement lent (en anglais "*slow fading channel*"), une même réalisation du canal \mathbf{H} est utilisée pour la transmission de plusieurs vecteurs \mathbf{x} en transmission. Il n'est alors plus légitime de moyenner sur les réalisations du canal \mathbf{H} (i.e. de prendre l'espérance \mathbb{E} sur \mathbf{H}). L'information mutuelle s'écrit alors

$$\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2 r} \mathbf{H} \mathbf{Q} \mathbf{H}^* \right)$$

et peut être médiocre si la réalisation du canal \mathbf{H} est atypique par exemple. Dans ce cas, la mesure de la performance est donnée par la probabilité de dépassement (*outage probability*) :

$$\mathbb{P}_{\text{out}}(R) = \inf_{\substack{\mathbf{Q} \geq 0 \\ \frac{1}{t} \text{Tr}(\mathbf{Q}) \leq 1}} \mathbb{P} \left(\log \det \left(\mathbf{I}_r + \frac{1}{\sigma^2 r} \mathbf{H} \mathbf{Q} \mathbf{H}^* \right) < R \right)$$

et un compromis doit être réalisé entre le taux de transmission souhaité R et la probabilité $\mathbb{P}_{\text{out}}(R)$ que ce taux ne soit pas atteint. Le calcul de la probabilité de dépassement est un problème difficile. L'étude des fluctuations de l'information mutuelle, quoique moins informative, peut être une alternative plus accessible.

Et les matrices aléatoires dans tout ça ?

Même lorsque le nombre d'antennes en émission et en réception est petit et comparable, il est difficile de calculer l'information mutuelle ou la capacité ergodique. La théorie des grandes matrices aléatoires permet de remplacer des quantités qui dépendent de la loi des variables sous-jacentes (à travers l'espérance \mathbb{E} sur \mathbf{H}) par leur équivalents déterministes. Le succès de l'utilisation de cette théorie dans ce contexte est sa forte capacité prédictive (équivalents déterministes très proches des "vraies quantités") même pour un petit nombre d'antennes.

On va décrire dans le paragraphe suivant les résultats développés dans le cadre du modèle de Rice.

8.2 Le modèle de Rice

On s'intéresse au modèle matriciel suivant :

$$\mathbf{H}_N = \mathbf{A}_N + \frac{1}{\sqrt{n}} \mathbf{D}_N^{1/2} \mathbf{X}_N \tilde{\mathbf{D}}_n^{1/2} ,$$

où

- ▷ la matrice A_N est déterministe, de dimensions $N \times n$,
- ▷ les matrices D_N et \tilde{D}_n sont déterministes, de dimensions respectives $N \times N$ et $n \times n$, diagonales, à éléments diagonaux positifs ou nuls,
- ▷ La matrice X_N est aléatoire, de dimensions $N \times n$, à entrées i.i.d., centrées et réduites.

On décrit ci-après le comportement asymptotique, dans le régime des grandes matrices aléatoires $N \propto n$, de la transformée de Stieltjes g_n de la mesure empirique des valeurs propres de $H_N H_N^*$, i.e.

$$g_n(z) = \frac{1}{N} \text{Tr} (H_N H_N^* - z I_N)^{-1} .$$

Théorème 59 (cf. [17]). *On suppose que les dimensions de la matrice X_N vérifient $N, n \rightarrow \infty$ de sorte que*

$$0 \leq \liminf_n \frac{N}{n} \leq \limsup_n \frac{N}{n} < \infty$$

et que les entrées (X_{ij}) de la matrice X_N admettent un 4^{ème} moment fini $\mathbb{E}|X_{ij}|^4 < \infty$. On suppose de plus que les normes spectrales des familles de matrices (A_N) , (D_N) et (\tilde{D}_n) sont uniformément bornées. Le système d'équations

$$\begin{cases} \delta_n(z) &= \frac{1}{n} \text{Tr} \left[D_N \left(-z(I_N + \tilde{\delta}_n(z) D_N) + A_N (I_n + \delta_n(z) \tilde{D}_n)^{-1} A_N^* \right)^{-1} \right] \\ \tilde{\delta}_n(z) &= \frac{1}{n} \text{Tr} \left[\tilde{D}_n \left(-z(I_n + \delta_n(z) \tilde{D}_n) + A_N^* (I_N + \tilde{\delta}_n(z) D_N)^{-1} A_N \right)^{-1} \right] \end{cases} \quad (45)$$

admet un unique couple de solutions $(\delta_n, \tilde{\delta}_n)$ chacune étant une transformée de Stieltjes d'une mesure⁵ (positive) à support inclus dans \mathbb{R}^+ . Soit $T_N(z)$ la matrice définie par

$$T_N(z) = \left(-z(I_N + \tilde{\delta}_n(z) D_N) + A_N (I_n + \delta_n(z) \tilde{D}_n)^{-1} A_N^* \right)^{-1} , \quad (46)$$

alors $t_n(z) = \frac{1}{N} \text{Tr} T_N(z)$ est la transformée de Stieltjes d'une mesure de probabilité \mathbb{P}_n et

$$g_n(z) - \frac{1}{N} \text{Tr} T_N(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} 0 .$$

De manière équivalente, presque sûrement, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée

$$\frac{1}{N} \sum_{i=1}^N f(\lambda_i) - \int_{\mathbb{R}} f(\lambda) \mathbb{P}_n(d\lambda) \xrightarrow[N, n \rightarrow \infty]{} 0 ,$$

où les (λ_i) représentent les valeurs propres de $H_N H_N^$.*

5. On montre facilement que δ_n (respectivement $\tilde{\delta}_n$) est la transformée de Stieltjes d'une mesure de masse totale égale à $\frac{1}{n} \text{Tr} D_n$ (respectivement égale à $\frac{1}{n} \text{Tr} \tilde{D}_n$).

On s'intéresse maintenant au comportement asymptotique de l'information mutuelle associée au canal de matrice de gain H_n :

$$\mathcal{I}_n(\rho) = \frac{1}{N} \log \det (H_N H_N^* + \rho I_N) = \frac{1}{N} \sum_{i=1}^N \log(\lambda_i + \rho) , \quad \rho > 0 .$$

L'information mutuelle se présente comme une statistique linéaire des valeurs propres de $H_N H_N^*$ (on appelle statistique linéaire des valeurs propres une fonction des valeurs propres de la forme $\sum_i \varphi(\lambda_i)$). Elle est aussi liée de manière plus profonde à la transformée de Stieltjes g_n car

$$\mathcal{I}_n(\rho) = \log \rho + \int_{\rho}^{\infty} \left(\frac{1}{u} - g_n(-u) \right) du . \quad (47)$$

En effet,

$$\frac{1}{u} - g_n(-u) = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{u} - \frac{1}{u + \lambda_i} \right)$$

et

$$\begin{aligned} \int_{\rho}^K \left(\frac{1}{u} - \frac{1}{u + \lambda_i} \right) du &= \left[\log(u) \right]_{\rho}^K - \left[\log(u + \lambda_i) \right]_{\rho}^K \\ &= -\log(\rho) + \log(\rho + \lambda_i) + \log \left(\frac{K}{K + \lambda_i} \right) \xrightarrow{K \rightarrow \infty} -\log(\rho) + \log(\rho + \lambda_i) , \end{aligned}$$

ce qui démontre la formule (47). Ce lien permet de décrire précisément et de manière explicite le comportement asymptotique de $\mathbb{E} \mathcal{I}_n(\rho)$. On se place à nouveau sous les hypothèses du théorème 59.

Théorème 60 (cf. [17]). *Soient $(\delta_n, \tilde{\delta}_n)$ les solutions du système (45) évaluées en $-\rho$. On pose*

$$\begin{aligned} V_n(\rho) &= \frac{1}{N} \log \det \left[\rho \left(I_N + \tilde{\delta}_n D_N \right) + A_N (I_n + \delta_n \tilde{D}_n)^{-1} A_N^* \right] \\ &\quad + \frac{1}{N} \log \det \left(I_n + \delta_n \tilde{D}_n \right) - \rho \frac{n}{N} \delta_n \tilde{\delta}_n . \end{aligned}$$

Alors

$$\mathbb{E} \mathcal{I}_n(\rho) - V_n(\rho) \xrightarrow{N, n \rightarrow \infty} 0 .$$

On constate que l'équivalent déterministe $V_n(\rho)$ de l'information mutuelle $\mathcal{I}_n(\rho)$ dépend des deux quantités δ_n et $\tilde{\delta}_n$, solutions du système (45). Une étude plus poussée (cf. [14]) permet d'obtenir un équivalent déterministe de la capacité ergodique.

Dans le cas d'un canal à évanouissement lent, il peut être intéressant de décrire les fluctuations de l'information mutuelle. C'est l'objet du théorème suivant. On se place à nouveau sous les hypothèses du théorème 59. On rappelle

que $T = T_N$ est donnée par la formule (46). On définit la matrice associée $\tilde{T} = \tilde{T}_n$ par :

$$\tilde{T}_n = \left(-z(I_n + \delta_n(z)\tilde{D}_n) + A_N^*(I_N + \tilde{\delta}_n(z)D_N)^{-1}A_N \right)^{-1}$$

Théorème 61 (cf. [16]). *On fait maintenant l'hypothèse que les entrées de la matrice X_n sont des gaussiennes complexes standard i.i.d., i.e. chaque entrée X_{ij} s'écrit $X_{ij} = \frac{U+iV}{2}$ où U et V sont des gaussiennes réelles indépendantes, centrées et réduites. On suppose de plus que*

$$\inf_n \frac{1}{n} \text{Tr } D_N > 0 \quad \text{et} \quad \inf_n \frac{1}{n} \text{Tr } \tilde{D}_n > 0 .$$

On pose

$$\gamma = \frac{1}{n} \text{Tr } D T D T \quad \text{et} \quad \tilde{\gamma} = \frac{1}{n} \text{Tr } \tilde{D} \tilde{T} \tilde{D} \tilde{T} .$$

Soit

$$\Theta_n = -\log \left(\left(1 - \frac{1}{n} \text{Tr } D^{1/2} T A (I + \delta \tilde{D})^{-1} \tilde{D} (I + \delta \tilde{D})^{-1} A^* T D^{1/2} \right)^2 - \rho^2 \gamma \tilde{\gamma} \right) .$$

Alors

$$\frac{N}{\sqrt{\Theta_n}} (\mathcal{I}_n(\rho) - V_n(\rho)) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1) ,$$

où $\xrightarrow{\mathcal{D}}$ représente la convergence en distribution.

9 Grandes matrices aléatoires et apprentissage

On considère dans cette section le problème de la *classification non supervisée* d'une séquence de données vectorielles $x_1, \dots, x_n \in \mathbb{R}^p$ en k classes d'affinité $\mathcal{C}_1, \dots, \mathcal{C}_k$. L'idée au centre de cette étude est de comprendre les performances de l'algorithme standard de Ng–Weiss–Jordan dit de clustering spectral à noyau.

Effectuons pour cela un petit rappel. Si on note $\kappa(x, y) \geq 0$ un noyau permettant de tenir compte de l'affinité des vecteurs $x, y \in \mathbb{R}^p$, tel que $\kappa(x, y)$ est large pour “ x proche de y ” et $\kappa(x, y)$ proche de zéro dans le cas contraire, la minimisation du “RatioCut” consiste en le problème d'optimisation discret suivant :

$$(\text{RatioCut}) \quad \text{argmin}_{\mathcal{C}_1 \cup \dots \cup \mathcal{C}_k = \{x_1, \dots, x_n\}} \sum_{a=1}^k \sum_{\substack{i \in \mathcal{C}_a \\ j \notin \mathcal{C}_a}} \frac{\kappa(x_i, x_j)}{|\mathcal{C}_a|} .$$

Cette métrique tend à minimiser la somme des coûts (faibles) d'affinités inter-classes tout en imposant une taille non négligeable de chaque classe par la division par $|\mathcal{C}_a|$.

En définissant $M \in \mathbb{R}^{n \times k}$ la matrice telle que $[M]_{ia} = |\mathcal{C}_a|^{-\frac{1}{2}} \delta_{x_i \in \mathcal{C}_a}$, ce problème peut se réécrire de manière équivalente

$$(\text{RatioCut}) \quad \operatorname{argmin}_{M \in \mathcal{M}} \operatorname{Tr} M^*(D - K)M$$

avec \mathcal{M} l'ensemble de telles matrices M et D la matrice diagonale telle que $D_{ii} = \sum_{j=1}^n K_{ij}$. Ce problème n'est pas plus simple à résoudre que le précédent mais, en remarquant que $M^*M = I_k$, il est tentant de relaxer le caractère discret du problème pour obtenir

$$(\text{RelaxedRatioCut}) \quad \operatorname{argmin}_{\substack{M \in \mathbb{R}^{n \times k} \\ M^*M = I_k}} \operatorname{Tr} M^*(D - K)M$$

qui est un simple problème aux valeurs propres (minimales). La solution de ce problème est simplement donnée par $M = [m_1, \dots, m_k]$ où m_i est le vecteur propre associé à la i -ème plus petite valeur propre de $D - K$. En cas de multiplicité, toute famille libre de vecteurs propres convient.

En supposant que la version relaxée du problème donne une bonne approximation de la solution du problème originel, on s'attend donc à pouvoir lire directement dans les vecteurs m_a les indices des x_i correspondants à la classe \mathcal{C}_a . Cela a donné lieu à plusieurs algorithmes qui partent de cette idée, mais l'améliorent en changeant en particulier la forme de la matrice $D - K$ (que l'on peut voir comme la Laplacienne d'un graphe de matrice d'adjacence K). L'algorithme de Ng–Jordan–Weiss que nous considérerons ici se base sur une version normalisée de la Laplacienne, à savoir

$$D^{-\frac{1}{2}}(D - K)D^{-\frac{1}{2}} = I_n - D^{-\frac{1}{2}}KD^{-\frac{1}{2}}.$$

Evidemment, comme les vecteurs propres associés aux plus petites valeurs propres de $I_n - D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ sont les mêmes que les vecteurs propres associés aux plus grandes valeurs propres de $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$, on étudiera en fait cette matrice et (pour des raisons que nous verrons plus tard), plus précisément la matrice

$$L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}.$$

L'objectif de l'étude est de comprendre les performances et les limitations de l'algorithme de Ng–Jordan–Weiss pour les classes définies comme suit :

$$x \in \mathcal{C}_a \Leftrightarrow x \sim \mathcal{N}(\mu_a, C_a)$$

pour une famille de $\mu_a \in \mathbb{R}^p$ et $C_a \in \mathbb{R}^{p \times p}$, $a = 1, \dots, k$.

Il s'agira donc de comprendre, dans la limite $n, p \rightarrow \infty$, comment se comporte la matrice L et d'identifier la position et les valeurs prises à la fois par ses valeurs et vecteurs propres. Ce problème est en général très difficile, et n'a jamais vraiment été résolu dans le cas pratique où p et n sont de dimensions similaires. Du point de vue des matrices aléatoires, la difficulté majeure est d'arriver à *passer outre la non linéarité du noyau* $\kappa(x, y)$.

9.1 Hypothèses pour un clustering non trivial

Afin de résoudre le problème de non linéarité de $\kappa(x, y)$, nous allons nous placer dans des conditions “limites” qui rendent le problème de clustering non trivial. A savoir, nous allons demander que les classes $\mathcal{C}_1, \dots, \mathcal{C}_k$ ne soient que peu distinguables (sinon, il n’y a pas vraiment de problème). Cette hypothèse va nous placer dans une situation permettant de *linéariser le noyau* sans erreur d’approximation, dans la limite des grands n, p .

Précisément, nous allons tout d’abord travailler avec le noyau suivant :

$$\kappa(x, y) = f\left(\frac{1}{p}\|x - y\|^2\right)$$

pour une fonction f suffisamment lisse (on aurait pu également prendre un noyau basé sur le produit scalaire). Plus important, nous allons poser les hypothèses suivantes, lorsque $n \rightarrow \infty$,

- ▷ $p/n \rightarrow c \in (0, \infty)$
- ▷ $|\mathcal{C}_a|/n \rightarrow c_a \in (0, 1)$ pour tout $a = 1, \dots, k$
- ▷ en définissant $\mu^\circ = \sum_{a=1}^k c_a \mu_a$ et $\mu_a^\circ = \mu_a - \mu^\circ$ pour tout $a = 1, \dots, k$,

$$\limsup_n \max_{1 \leq a \leq k} \|\mu_a^\circ\| < \infty$$

- ▷ en définissant $C^\circ = \sum_{a=1}^k c_a C_a$ et $C_a^\circ = C_a - C^\circ$ pour tout $a = 1, \dots, k$,

$$\limsup_n \max_{1 \leq a \leq k} \|C_a\| < \infty \text{ et } \limsup_n \max_{1 \leq a \leq k} \frac{1}{\sqrt{p}} \text{Tr } C_a^\circ < \infty.$$

Sous ces conditions, on assure en fait qu’il n’est pas possible asymptotiquement de garantir une identification parfaite d’au moins une classe. Si c’était le cas, alors on pourrait isoler cette classe et procéder au clustering des autres classes.

Une conséquence, heureuse, de ces hypothèses est le fait qu’en écrivant

$$\tau = \frac{2}{p} \text{Tr } C^\circ$$

il vient facilement qu’avec probabilité un,

$$\max_{1 \leq i \neq j \leq n} \left| \frac{1}{p} \|x_i - x_j\|^2 - \tau \right| \rightarrow 0.$$

Ce résultat est fondamental dans ce qui suit, car il nous permet d’assurer que

$$\max_{1 \leq i \neq j \leq n} |K_{ij} - f(\tau)| \rightarrow 0$$

ce qui donne a priori l’impression que K est assimilable (à une diagonale près) à la matrice $f(\tau)1_n 1_n^*$ de rang 1. Cependant, cette convergence uniforme sur les entrées n’implique pas du tout une convergence en norme spectrale. L’étude fine de la matrice K se révèle apporter bien plus d’informations qu’il n’y paraît à ce niveau.

9.2 Etude de la matrice K

Dans tout ce qui suit, nous supposons les x_i ordonnés par classe pour des raisons de lisibilité. Cela n'impacte cependant pas l'étude.

L'idée de l'étude est d'effectuer un développement limité des entrées non diagonales de la matrice K autour de la valeur $f(\tau)$. Il s'agira alors d'identifier, parmi les termes de l'expansion, ceux qui contribueront à l'approximation finale de K par une matrice asymptotiquement équivalente \hat{K} portant l'information sur les classes $\mathcal{C}_1, \dots, \mathcal{C}_k$. Cette étude des termes utiles est particulièrement délicate à mener car assez contre-intuitive de prime abord.

Commençons par écrire, pour $x_i \in \mathcal{C}_a$,

$$x_i = \mu_a + \sqrt{p}w_i$$

avec $w_i \sim \mathcal{N}(0, \frac{1}{p}C_a)$. On a alors, pour $x_i \in \mathcal{C}_a$ et $x_j \in \mathcal{C}_b$,

$$\begin{aligned} \frac{1}{p}\|x_j - x_i\|^2 &= \|w_j - w_i\|^2 + \frac{1}{p}\|\mu_b - \mu_a\|^2 + \frac{2}{\sqrt{p}}(\mu_b - \mu_a)^*(w_j - w_i) \\ &= \tau + \frac{1}{p}\text{Tr } C_a^\circ + \frac{1}{p}\text{Tr } C_b^\circ + \psi_j + \psi_i - 2w_i^*w_j \\ &\quad + \frac{\|\mu_b - \mu_a\|^2}{p} + \frac{2}{\sqrt{p}}(\mu_b - \mu_a)^*(w_j - w_i) \end{aligned}$$

où, pour $x_i \in \mathcal{C}_a$,

$$\psi_i = \|w_i\|^2 - \frac{1}{p}\text{Tr } C_a.$$

L'intérêt de cette décomposition est de faire apparaître des quantités aléatoires de moyenne nulle (les w_i et les ψ_i) et de recentrer les quantités déterministes $\frac{1}{p}\text{Tr } C_a$ par rapport à leur moyenne $\frac{1}{2}\tau$. En particulier, remarquons que, mis à part τ , les termes dans le développement sont bien d'amplitudes $o(1)$. Plus précisément, il n'est pas dur de voir que $\psi_i = O(n^{-\frac{1}{2}})$, $w_i^*w_j = O(n^{-\frac{1}{2}})$ alors que $\frac{1}{\sqrt{p}}(\mu_b - \mu_a)^*(w_j - w_i) = O(n^{-1})$.

Nous devons maintenant effectuer l'expansion de Taylor de $f(\frac{1}{p}\|x_i - x_j\|^2)$ autour de $f(\tau)$ pour chaque couple (i, j) et rassembler le résultat sous forme matricielle. Il nous faut alors identifier dans cette expansion quels termes contribuent significativement *en norme spectrale*. Cela amène des calculs assez fastidieux, dont nous ne retiendrons ici que les ingrédients fondamentaux :

- ▷ hormis sur la diagonale que nous excluons aisément, le terme $f(\tau)$ apparaît pour tout (i, j) comme le coefficient dominant. Ainsi la matrice $f(\tau)1_n 1_n^*$, de rang 1 et de norme spectrale $O(n)$, constitue le terme dominant de K . Elle n'apporte aucune information sur les classes et ne nous sera guère utile.
- ▷ le terme suivant du développement, $f'(\tau)$, multiplierà
 - la matrice $\psi 1_n^* + 1_n \psi^*$ où $\psi \in \mathbb{R}^n$ est le vecteur des ψ_i . Cette matrice est de rang 2 et de norme spectrale $O(\sqrt{n})$;

- la matrice $\{\frac{1}{p}\text{Tr } C_a^\circ 1_{|C_a|}\}_{a=1}^k 1_n^* + 1_n \{\frac{1}{p}\text{Tr } C_b^\circ 1_{|C_b|}\}_{b=1}^k$, où les a sont arrangés en colonne et les b en ligne. Cette matrice de rang 2 est de norme $O(\sqrt{n})$;
- la matrice $\{\frac{1}{p}\|\mu_a - \mu_b\|^2 1_{|C_a|} 1_{|C_b|}^*\}_{a,b=1}^k$ de rang maximal k et de norme $O(1)$.
- la matrice W^*W , avec $W = [w_1, \dots, w_n]$, qui est une version un peu améliorée du modèle de matrice de covariance vu précédemment en cours; cette matrice a une norme spectrale d'ordre $O(1)$;
- les autres termes sont petits en comparaison.
- ▷ le terme suivant du développement, $\frac{1}{2}f''(\tau)$, multipliera notamment
 - la matrice $(\psi \circ \psi) 1_n^* + 1_n (\psi \circ \psi)^*$ où $\psi \in \mathbb{R}^n$, de rang 2 et de norme spectrale $O(1)$;
 - la matrice $\{(\frac{1}{p}\text{Tr } C_a^\circ)^2 1_{|C_a|}\}_{a=1}^k 1_n^* + 1_n \{(\frac{1}{p}\text{Tr } C_b^\circ)^2 1_{|C_b|}\}_{b=1}^k$ de rang 2 et de norme $O(1)$;
 - les produits croisés $2\{\frac{1}{p}\text{Tr } C_a^\circ 1_{|C_a|}\}_{a=1}^k \psi^* + 2\psi \{\frac{1}{p}\text{Tr } C_b^\circ 1_{|C_b|}\}_{b=1}^k$, également de rang 2 et de norme $O(1)$;
 - la matrice $\psi \psi^*$ de rang 1 et de norme $O(1)$;
 - la matrice $2(1_n \psi^* \text{diag}(\frac{1}{p}\text{Tr } C_a^\circ 1_{|C_a|}) + \text{diag}(\frac{1}{p}\text{Tr } C_a^\circ 1_{|C_a|}) \psi 1_n^*)$ de rang 2 et de norme $O(1)$;
 - la matrice $\{\frac{1}{p}\text{Tr } C_a^\circ \frac{1}{p}\text{Tr } C_b^\circ 1_{|C_a|} 1_{|C_b|}^*\}_{a,b=1}^k$ de rang 1 et de norme $O(1)$;
 - la matrice $4\{\frac{1}{p^2}\text{Tr } (C_a C_b) 1_{|C_a|} 1_{|C_b|}^*\}_{a,b=1}^k$ de rang maximal k et de norme $O(1)$; cette matrice provient de la décomposition de $\{(w_i^* w_j)^2\}_{i,j}$ en une matrice déterministe (la moyenne statistique) et une matrice aléatoire de norme $O(n^{-\frac{1}{2}})$.
- ▷ les termes suivants du développement multiplieront seulement des termes de norme spectrale $O(n^{-\frac{1}{2}})$.

On s'aperçoit ici que la matrice K a un comportement spectral assez complexe, avec

- ▷ une valeur propre isolée d'ordre $O(n)$ (en fait sensiblement égale à n) de vecteur propre associé très proche de 1_n
- ▷ des contributions de rang faible d'ordre $O(\sqrt{n})$ (qui peuvent ou non générer des valeurs propres isolées de cet ordre, ce qui est en effet confirmé par simulations)
- ▷ une contribution de rang élevé de norme spectrale $O(1)$ qui est la somme de W^*W (qui se comprend comme un “bulk” de bruit) et de matrices de petits rangs portant l'information sur les μ_a , $\text{Tr } C_a$ et $\text{Tr } C_a C_b$ le long des vecteurs canoniques $j_a = [0, \dots, 0, 1_{|C_a|}, 0, \dots, 0]^*$.

9.3 Etude de la matrice $L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}$

La suite du calcul consiste en l'expansion de Taylor de la matrice $D = \text{diag}(K 1_n)$ étant donné le résultat précédent (et donc l'approximation de D^α pour tout $\alpha \in \mathbb{R}$), puis le calcul de $nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}$, où le facteur n permet de normaliser les matrices $D^{-\frac{1}{2}}$. Ce calcul donne lieu à une forme similaire à celle

obtenue pour K que nous ne développerons pas plus avant ici.

Il est très facile de se rendre compte que $nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ a pour vecteur propre $D^{\frac{1}{2}}1_n$ avec valeur propre associée n (et donc liée à la valeur propre $O(n)$ de K vue précédemment). Ce vecteur propre pourra être étudié de manière autonome, et il suffit alors d'étudier la matrice

$$L' = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}} - n \frac{D^{\frac{1}{2}}1_n 1_n^* D^{\frac{1}{2}}}{1_n^* D 1_n}$$

qui n'est autre que la projection orthogonale de L' orthogonalement à l'espace engendré par $D^{\frac{1}{2}}1_n$.

Bien que cela soit loin d'être immédiat, le fait que $D^{\frac{1}{2}}1_n$ soit sensiblement colinéaire au vecteur 1_n va avoir pour effet d'éliminer dans l'expansion de Taylor de $nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ pratiquement tous les termes de la forme $1_n v^*$ et $v 1_n^*$. En particulier, l'approximation finale de L' ne contiendra que des termes de norme spectrale $O(1)$. Tout calcul fait, on a le résultat final suivant

Théorème 62. *Sous les hypothèses de croissance du modèle discutées précédemment,*

$$L' = -2 \frac{f'(\tau)}{f(\tau)} [PW^*WP + UBU^*] + \frac{f(0) - f(\tau) + \tau f'(\tau)}{f(\tau)} I_n + O_{\|\cdot\|}(n^{-\frac{1}{2}})$$

où

$$U = \left[\frac{j_1}{\sqrt{p}}, \dots, \frac{j_k}{\sqrt{p}}, v_1, \dots, v_k, \psi \right]$$

avec $j_a \in \mathbb{R}^n$ le vecteur canonique de la classe \mathcal{C}_a , $P = I_n - \frac{1}{n} 1_n 1_n^*$, $v_i = PW^* \mu_a^\circ$, $\psi = \{\|w_i\|^2 - \mathbb{E}[\|w_i\|^2]\}_{i=1}^n$, et

$$B = \begin{bmatrix} B_{11} & I_k - 1_k \{c_b\}_{b=1}^k & \left(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \right) t \\ I_k - \{c_a\}_{a=1}^k 1_k^* & 0_{k \times k} & 0_{k \times 1} \\ \left(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \right) t^* & 0_{1 \times k} & \frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \end{bmatrix} \in \mathbb{R}^{(2k+4) \times (2k+4)}$$

$$B_{11} = M^* M + \left(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \right) tt^* - \frac{f''(\tau)}{f'(\tau)} T + c \frac{f(0) - f(\tau) + \tau f'(\tau)}{2f'(\tau)} 1_k 1_k^*.$$

où $M = [\mu_1^\circ, \dots, \mu_k^\circ]$, $t = \{\frac{1}{\sqrt{p}} \text{Tr } C_a^\circ\}_{a=1}^k$ et $T = \{\frac{1}{p} \text{Tr } C_a^\circ C_b^\circ\}$.

Par ailleurs, le cas $f'(\tau) = 0$ s'obtient par la limite continue $f'(\tau) \rightarrow 0$ du résultat précédent.

Le théorème fait apparaître un modèle de type “perturbation de petit rang” comme étudié précédemment dans le cours, à ceci près que :

- ▷ la matrice de bruit PW^*WP fait intervenir un modèle un peu plus compliqué pour W que vu en cours ;
- ▷ la perturbation UBU^* contient une partie aléatoire.



FIGURE 12 – Echantillons de la base MNIST.

Ces problèmes se gèrent plutôt bien et les conséquences d’une étude plus fine du modèle sont les suivantes :

- ▷ dans le cas où les valeurs propres associées à la matrice UBU^* sont suffisamment larges, un certain nombre de valeurs propres de L' s’échappent du support limite de la mesure spectrale de PW^*WP
- ▷ les valeurs propres isolées sont dimensionnées par les valeurs propres de $f'(\tau)M$, $(\frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)})tt^*$ et $\frac{f''(\tau)}{f'(\tau)}T$;
- ▷ pour la plupart des valeurs propres isolées (mais pas forcément toutes !), les vecteurs propres de UBU^* s’alignent à des combinaisons linéaires des vecteurs j_1, \dots, j_k d’autant plus que les valeurs propres en question sont larges

En remarquant que

$$\hat{\tau} = \frac{2}{n} \sum_{i=1}^n \|x_i\|^2 \rightarrow \tau$$

presque sûrement, il est possible d’exploiter la 2e remarque ci-dessus pour adapter le choix du noyau à une tâche de clustering particulière (basée sur les covariances ou sur les moyennes).

Une étude complète de la position des valeurs propres et des projections des vecteurs propres associés sur la base des j_a peut être effectuée. Les calculs sont assez longs mais semblables aux approches évoquées dans les sections précédentes. Nous ne fournirons ici que des illustrations comparatives des performances théoriques dans des contextes pratiques.

9.4 Application au clustering de la base MNIST

Nous nous intéressons dans cette section au clustering d’un certain nombre de vecteurs de la base MNIST. Nous sélectionnerons ici des images vectorisées (images de dimension 28×28) des chiffres 0, 1 et 2, voir Figure 12. Nous prenons $n = 192$ échantillons composés de 64 zéros, 64 uns et 64 deux.

En faisant l’hypothèse (plus qu’approximative !) que les vecteurs-images de la base MNIST sont gaussiens avec pour moyennes et matrices de covariance les valeurs empiriquement obtenues de la base totale des 60 000 vecteurs-images, nous obtenons alors la comparaison théorie-pratique donnée par les Figures 13 et 14.

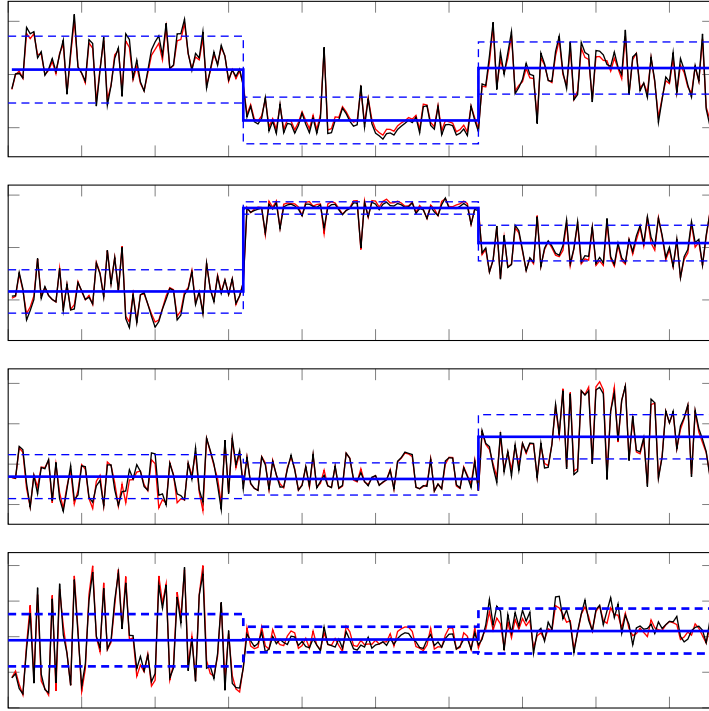


FIGURE 13 – Quatre vecteurs propres dominants de L (rouge), de son approximation asymptotique \hat{L} (noir) versus les moyennes et écart-types théoriques (1σ et 2σ) (bleu) ; base MNIST.

Références

- [1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
- [2] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1) :316–345, 1998.
- [3] Z. D. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.
- [4] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5) :1643–1697, 2005.
- [5] J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6) :1382–1408, 2006.

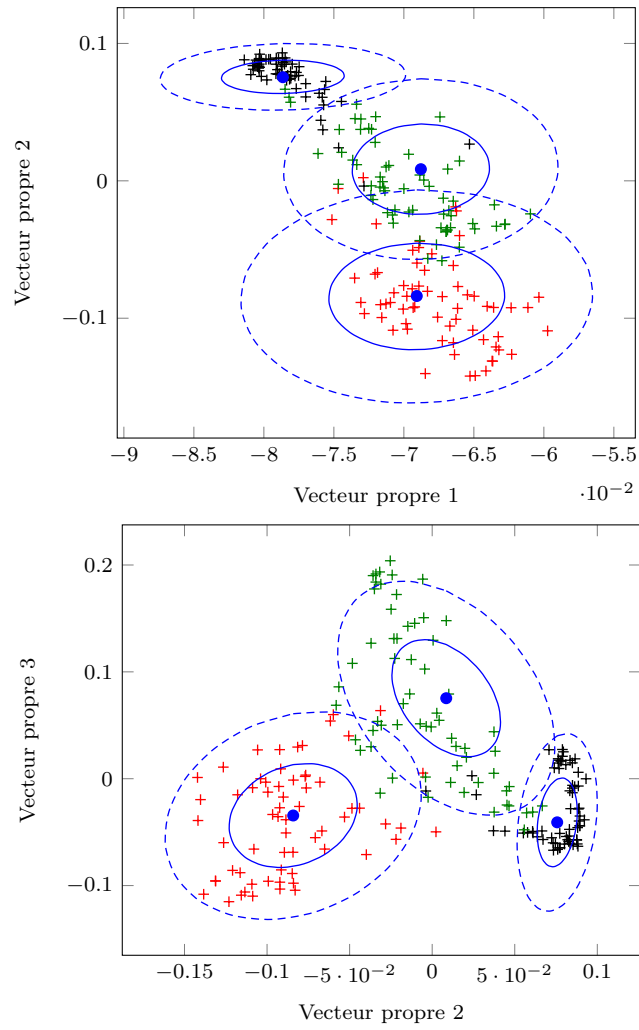


FIGURE 14 – Représentation 2D des vecteurs propres 1 et 2 (figure supérieure), et 2 et 3 (figure inférieure) de L , pour la base MNIST. Moyenne et écarts-types théoriques en bleu.

- [6] F. Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, 227(1) :494–521, 2011.
- [7] P. Bianchi, M. Debbah, M. Maida, and J. Najim. Performance of statistical tests for single-source detection using random matrix theory. *Information Theory, IEEE Transactions on*, 57(4) :2400–2419, april 2011.
- [8] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
- [9] C. Bordenave. *A short course on random matrices (preliminary draft)*. available on <http://www.math.univ-toulouse.fr/~bordenave/coursRMT.pdf>, 2013.
- [10] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities : A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.
- [11] D. Chafaï and F. Malrieu. Recueil de modèles aléatoires. disponible sur la page web de D. Chafaï, 2015.
- [12] R. Couillet and M. Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, 2011.
- [13] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [14] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for rician mimo channels : An asymptotic approach. *Information Theory, IEEE Transactions on*, 56(3) :1048–1069, march 2010.
- [15] G.J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs technical journal*, 1 :41–59, 1996.
- [16] W. Hachem, M. Kharouf, J. Najim, and J. W. Silverstein. A CLT for information-theoretic statistics of non-centered Gram random matrices. *Random Matrices Theory Appl.*, 1(2) :1150010, 50, 2012.
- [17] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.*, 17(3) :875–930, 2007.
- [18] W. Hachem, P. Loubaton, J. Najim, and P. Vallet. On bilinear forms based on the resolvent of large random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(1) :36–63, 2013.
- [19] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.
- [20] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2) :295–327, 2001.
- [21] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

- [22] M. G. Kreĭn and A. A. Nudel'man. *The Markov moment problem and extremal problems*. American Mathematical Society, Providence, R.I., 1977. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.
- [23] V. A. Marcenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114) :507–536, 1967.
- [24] A. Nica and R. Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.
- [25] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.
- [26] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [27] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2) :331–339, 1995.
- [28] J. W. Silverstein and S. Choi. Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2) :295–309, 1995.
- [29] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.
- [30] E. Telatar. Capacity of multi-antenna Gaussian channels. *European transactions on telecommunications*, 10(6) :585–595, 1998.
- [31] J. Weidmann. *Linear operators in Hilbert spaces*, volume 68 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1980. Translated from the German by Joseph Szűcs.
- [32] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62 :548–564, 1955.
- [33] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67 :325–327, 1958.
- [34] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A((1-2)) :32–52, 1928.