

# Introduction à la théorie des grandes matrices aléatoires

Jamal Najim et Romain Couillet

version du 23 mars 2017



# Table des matières

<b>To Do List</b>	<b>7</b>
<b>Introduction</b>	<b>9</b>
<b>1 Transformée de Stieltjes</b>	<b>17</b>
1.1 Définition et premières propriétés . . . . .	17
1.2 Convergence . . . . .	21
1.3 Identifier une transformée de Stieltjes . . . . .	24
1.4 La formule d’Helffer-Sjöstrand . . . . .	25
1.5 Distribution de Marčenko-Pastur . . . . .	28
1.6 Exercices . . . . .	32
<b>2 Compléments d’algèbre linéaire</b>	<b>35</b>
2.1 Décomposition spectrale et résolvante . . . . .	35
2.2 Norme spectrale . . . . .	36
2.3 Identités matricielles . . . . .	38
2.4 Inégalités matricielles . . . . .	41
2.5 Calcul fonctionnel pour les matrices hermitiennes . . . . .	42
2.6 Exercices . . . . .	44
<b>3 Estimation de variances</b>	<b>45</b>
3.1 Inégalité d’Efron-Stein . . . . .	45
3.2 Estimation de la variance d’une forme quadratique . . . . .	47
3.3 Estimation de la variance de la transformée de Stieltjes $g_n(z)$ . . . . .	50
3.4 Variations autour de l’inégalité de Poincaré gaussienne . . . . .	53
3.5 Exercices . . . . .	54
<b>4 Théorème de Marčenko-Pastur</b>	<b>57</b>
4.1 Preuve du théorème de Marčenko-Pastur . . . . .	58
4.1.1 Notations et identités importantes . . . . .	58
4.1.2 Stratégie de la preuve . . . . .	58
4.1.3 Equation satisfaite par $\mathbb{E} g_n(z)$ . . . . .	59
4.1.4 Stabilité de l’équation satisfaite par $g_{\text{MP}}$ . . . . .	61
4.1.5 Fin de la preuve . . . . .	62

4.2	Calcul gaussien et applications . . . . .	62
4.2.1	Formule d'intégration par parties et inégalité de Poincaré . . . . .	63
4.2.2	Preuve gaussienne du théorème de Marčenko-Pastur . . . . .	65
4.2.3	Version isotrope du théorème de Marčenko-Pastur . . . . .	66
4.2.4	Confinement des valeurs propres . . . . .	69
4.3	Exercices . . . . .	71
<b>5</b>	<b>Le modèle de grandes matrices de covariance</b>	<b>73</b>
5.1	La matrice de covariance empirique . . . . .	73
5.2	Comportement de la matrice de covariance empirique en grande dimension . . . . .	74
5.2.1	Méthode heuristique de preuve pour le théorème 58 . . . . .	76
5.2.2	Exemples . . . . .	78
5.3	Démonstration du théorème 58 . . . . .	79
5.3.1	Démonstration dans le cas où $R_N$ est diagonale . . . . .	79
5.3.2	Démonstration dans le cas d'entrées gaussiennes . . . . .	84
5.3.3	Le cas général . . . . .	85
5.4	Etude de la mesure limite pour le modèle de grandes matrices de covariance . . . . .	89
5.4.1	Compléments à la Proposition 2 . . . . .	91
5.4.2	Application au modèle de covariance . . . . .	91
5.5	Inférence statistique pour le modèle de grandes matrices de covariance . . . . .	99
5.5.1	Quelques rappels d'analyse complexe . . . . .	99
5.5.2	Inférence statistique : Spécification des contours . . . . .	100
5.5.3	Inférence statistique : Calcul des estimateurs . . . . .	103
<b>6</b>	<b>Confinement des valeurs propres</b>	<b>107</b>
6.1	La formule d'Helffer-Sjöstrand . . . . .	107
6.2	Confinement des valeurs propres . . . . .	111
<b>7</b>	<b>Petites perturbations</b>	<b>115</b>
7.1	Etude de la plus grande valeur propre . . . . .	115
7.2	Éléments de preuve . . . . .	117
7.3	Grandes matrices aléatoires et apprentissage . . . . .	120
7.3.1	Hypothèses pour un clustering non trivial . . . . .	122
7.3.2	Etude de la matrice $K$ . . . . .	123
7.3.3	Etude de la matrice $L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ . . . . .	125
7.3.4	Application au clustering de la base MNIST . . . . .	126
<b>A</b>	<b>Rappels d'analyse complexe</b>	<b>129</b>
<b>B</b>	<b>Inégalités classiques</b>	<b>131</b>
B.1	Démonstration de l'inégalité d'Efron-Stein . . . . .	131
B.2	Inégalité de Poincaré gaussienne . . . . .	131
B.2.1	Démonstration du théorème 40 . . . . .	131

<i>TABLE DES MATIÈRES</i>	5
B.2.2 Démonstration du corollaire 41 . . . . .	132
<b>C Compléments</b>	<b>135</b>
C.1 Identification d'une transformée de Stieltjes . . . . .	135



# To Do list

- Stabilité de l'équation de Marchenko-Pastur peut être intégrée au 1.4
- Chapitre 2 : rajouter l'exemple  $\Im(\text{trace}Q^2)/(1 + \frac{1}{n}\text{trace}Q)$ .
- Reprendre chapitre 3 et l'intituler estimation de variances. + calculer moment d'ordre 4 d'une forme quadratique  $y^*Ay$  + moment d'ordre 4 d'une forme quadratique  $a^*Qa$  en exo
- Rajouter des exos concernant le modèle de Wigner
- au chapitre 5 rajouter l'étude du spectre limite

Chapitres à venir

- Petites perturbations
- Fluctuations globales (gaussienne), locales (TW, Pearcey, Bessel)
- Modèles non hermitiens



# Introduction

La théorie des grandes matrices aléatoires prend ses racines dans les travaux de Wishart [29] qui a initialement considéré des matrices à entrées aléatoires gaussiennes. Wigner [27, 28] et plus tard Marčenko et Pastur [18] 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 importante 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 [25]. Ce point de vue s'est avéré extrêmement 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 d'importantes monographies dans le domaine [1, 20, 3, 24].

L'objectif de ces notes de cours 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, 19, 24]), 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

$$f_{sc}(x) = \frac{1}{2\pi} \sqrt{(4-x^2)_+}, \quad \text{où } a_+ = \max(a, 0).$$

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].

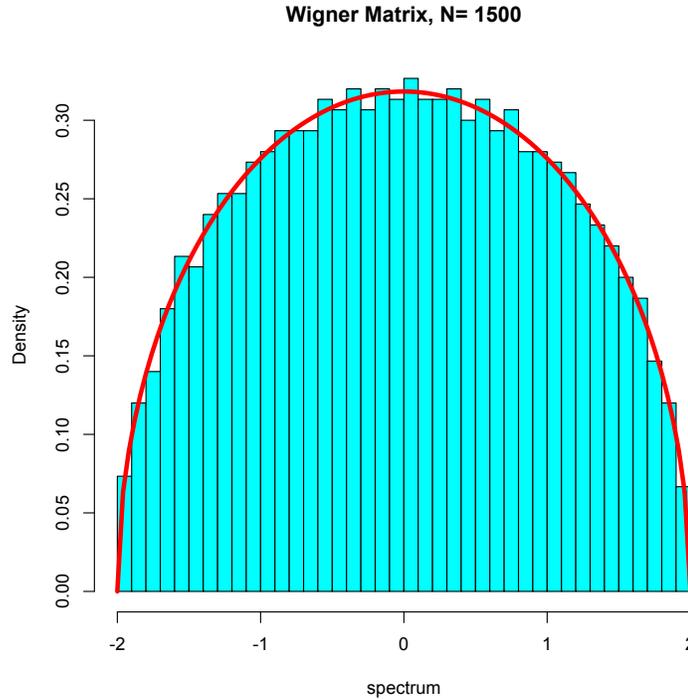


FIGURE 1 – Histogramme des valeurs propres d'une matrice de Wigner renormalisée de dimensions  $1500 \times 1500$ . En rouge, la densité de la loi du demi-cercle.

**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 croissent 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

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

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

**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  $X_N = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  où  $\mathbf{x}_i$  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

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

Cette matrice modélise un échantillon de  $n$  observations  $\mathbf{y}_i = R_N^{1/2} \mathbf{x}_i$ , chacune de dimension  $N$  et de matrice de covariance  $R_N = \mathbb{E} \mathbf{y}_i \mathbf{y}_i^*$ . 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

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

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 :

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

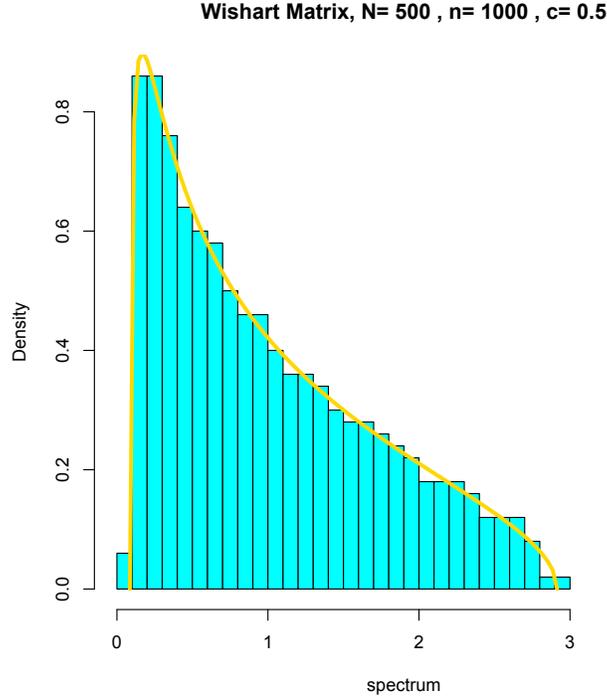


FIGURE 2 – Histogramme des valeurs propres d’une grande matrice de covariance de dimensions  $500 \times 1000$ . En jaune, la densité de la loi du demi-cercle.

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 ( $\simeq \frac{N}{3}$ ).

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 popula-

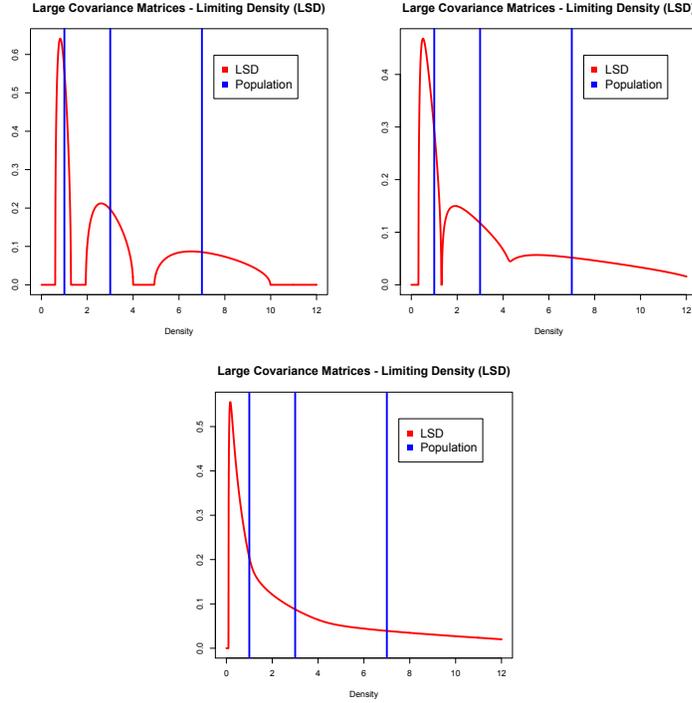


FIGURE 3 – En rouge, densité de la mesure limite dans les cas où  $c = 0, 1$  ;  $c = 0, 3$  et  $c = 0, 6$ . En bleu, les 3 valeurs propres de la matrice  $R_N$  de covariance de la population, chacune de multiplicité égale.

tion (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 :

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

où les  $(\theta_\ell)$  sont des réels fixés et déterministes, et les  $(\mathbf{u}_\ell)$  (qui dépendent de  $n$ ) sont des vecteurs déterministes et orthonormés. Chaque matrice  $\mathbf{u}_\ell \mathbf{u}_\ell^*$  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  $\sum_{\ell=1}^k \theta_\ell \mathbf{u}_\ell \mathbf{u}_\ell^*$  ? Des simulations apportent quelques éléments de réponse.

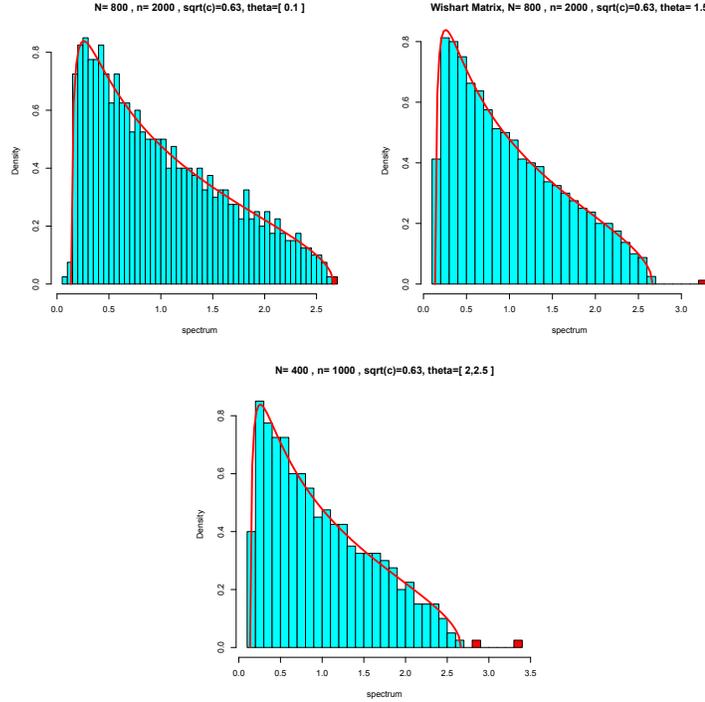


FIGURE 4 – En rouge, la densité de Marčenko-Pastur, en bleu l’histogramme des valeurs propres empiriques. Le(s) petit(s) bâton(s) rouge(s) représente(nt) la(les) plus grande(s) valeur(s) propre(s) empirique(s). Les deux premières simulations correspondent à une perturbation simple où  $\theta = 0,1$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

**Matrices non hermitiennes.** On s’intéresse maintenant au cas d’une matrice carrée  $X_n$  de dimension  $n \times n$  dont les entrées sont i.i.d, centrées, chacune de variance  $\frac{1}{\sqrt{n}}$ . Dans ce cas, les valeurs propres ne sont plus réelles mais complexes. En les représentant simplement dans le plan complexe, on peut constater

qu'elles ont tendance à se répartir uniformément dans le disque unité (cf. Figure 5).

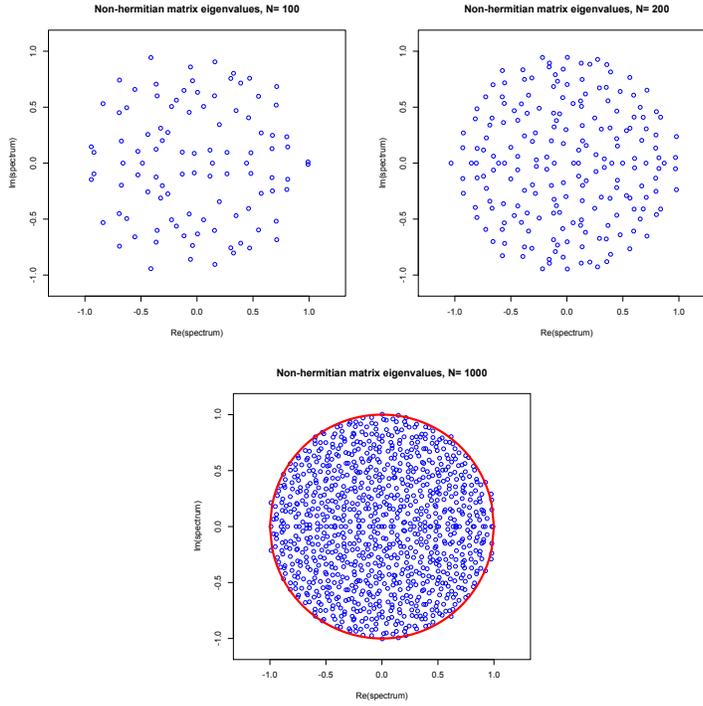


FIGURE 5 – On représente les valeurs propres d'une matrice  $X_N$  non hermitienne dans le plan complexe. Le graphique de gauche correspond à une matrice 100x100, celui du centre à une matrice 200x200, celui de droite à une matrice 1000x1000. Le cercle de rayon 1 est représenté en rouge.

**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  $A_N$  de dimension  $N \times N$ . Le théorème spectral nous assure que ces valeurs propres sont réelles. Soit  $\delta_x$  la mesure de Dirac au point  $x$ , définie par

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

Une propriété importante de la mesure de Dirac est que

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

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

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

Ainsi,  $L_N([a, b])$  correspond à la proportion des valeurs propres contenues dans  $[a, b]$  :

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

On dira que  $L_N$  **converge étroitement** vers une mesure de probabilité  $\mu$  (notation  $L_N \xrightarrow{\text{etr}} \mu$ ) 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  $A_N$  est aléatoire, ses valeurs propres le sont également et on dira que presque sûrement  $L_N$  converge étroitement vers  $\mu$  si pour presque toute réalisation de la famille de matrices  $A_N$ , on a convergence étroite de  $L_N$  (en tant qu'objet aléatoire) vers  $\mu$  (qui est une probabilité, mais qui ne dépend 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

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

Dans le cas où la matrice  $A_N$  n'est pas hermitienne, ses valeurs propres sont alors complexes et la mesure spectrale sera cette fois définie sur  $\mathbb{C}$  ; sa convergence éventuelle s'exprimera à l'aide de fonctions-test sur  $\mathbb{C}$ .

# Chapitre 1

## Transformée de Stieltjes

La transformée de Stieltjes d'une mesure est 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  $\mu$  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  $\mu$  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  $\mu$ .

**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  $A$  une matrice hermitienne  $N \times N$  de décomposition spectrale

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

les  $\lambda_i$  étant les valeurs propres de  $A$ . Soit  $L_N$  la **mesure spectrale de  $A$** , aussi appelée **mesure empirique des valeurs propres de  $A$**  :

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

Alors

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

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

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

4. Même contexte que précédemment. Soit  $\mathbf{a}$  un vecteur  $N \times 1$ ; on note  $\mathbf{u}_i$  le vecteur propre associé à la valeur propre  $\lambda_i$ . Alors la décomposition de  $A$  s'écrit  $A = \sum_{i=1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^*$  et

$$g(z) = \mathbf{a}^* (A - zI_N)^{-1} \mathbf{a} = \sum_{i=1}^N \frac{\mathbf{a}^* \mathbf{u}_i \mathbf{u}_i^* \mathbf{a}}{\lambda_i - z} = \sum_{i=1}^N \frac{|\mathbf{a}^* \mathbf{u}_i|^2}{\lambda_i - z},$$

est la transformée de Stieltjes de la mesure  $\sum_{i=1}^N |\mathbf{a}^* \mathbf{u}_i|^2 \delta_{\lambda_i}$  de masse totale  $\sum_{i=1}^N |\mathbf{a}^* \mathbf{u}_i|^2 = \|\mathbf{a}\|^2$ .

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

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

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

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

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

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

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

9. Si  $a, b \in \mathbb{R}$  sont des points de continuité de  $\mu$ , 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

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

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.

5. 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 :

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

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

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

7. 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  $\mu$ ,  $Y$  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 :

$$\frac{1}{\pi} \operatorname{Im} \int f(x)g_\mu(x + iy)dx = \mathbb{E}f(X + yY) , \quad (1.1)$$

où  $X$  est une V.A. de loi  $\mu$ ,  $Y$  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  $\operatorname{Im} g_\mu(x + iy)$  donne

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

ce qui établit (1.1). Revenons maintenant aux preuves des points 7 et 8.

8. 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

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

ici, (a) découle de l'identité (1.1).

9. On procède par approximation et on introduit deux fonctions continues bornées  $\varphi_n^-$  et  $\varphi_n^+$  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  $\varphi_n^-$  et  $\varphi_n^+$  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  $\mu$ , on a donc

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

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  $\mu$ .

□

## 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  $\mu_n$  de mesures de probabilités sur  $\mathbb{R}$ , on dit qu'elle **converge vaguement** vers une mesure  $\nu$  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  $\xrightarrow{\nu}$ .

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], [17, Th. 5.19]) : de toute suite de probabilités  $\mu_n$  sur  $\mathbb{R}$ , on peut extraire une sous-suite qui converge vaguement.

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

**Exemple 3.** 1.  $\delta_n \xrightarrow{n \rightarrow \infty} 0$ .

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

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

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

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

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

pour toute fonction  $f$  continue qui tend vers zéro en  $\pm\infty$ .

En effet, fixons  $\varepsilon > 0$  et  $K > 0$  tel que  $|f(x)| \leq \varepsilon$  si  $|x| > K$ , et soit  $\chi_\varepsilon : \mathbb{R} \rightarrow [0, 1]$  une fonction continue valant 1 sur  $[-K, K]$  et zéro pour  $|x| > K + 1$ , alors

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \left| \int f d\mathbb{P}_n - \int f d\nu \right| \\ &= \limsup_{n \rightarrow \infty} \left| \int f\chi_\varepsilon d\mathbb{P}_n - \int f\chi_\varepsilon d\nu + \int f(1 - \chi_\varepsilon) d\mathbb{P}_n - \int f(1 - \chi_\varepsilon) d\nu \right| \\ &\leq 2\varepsilon + \limsup_{n \rightarrow \infty} \left| \int f\chi_\varepsilon d\mathbb{P}_n - \int f\chi_\varepsilon d\nu \right|. \end{aligned}$$

La fonction  $f\chi_\varepsilon$  étant continue à support compact, la convergence étroite de  $\mathbb{P}_n$  vers  $\nu$  nous assure que cette dernière lim sup est nulle.

Cette remarque sera utile dans la preuve du théorème 8.

Soit  $\mu_n$  une suite de mesures de probabilités. On rappelle que  $\mu_n$  **converge étroitement** vers  $\mu$  si pour toute fonction continue bornée,

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

On notera  $\xrightarrow{\text{etr}}$ . Dans ce cas,  $\mu$  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.

Etant donnée une suite  $\mu_n$  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  $\varepsilon$  près, toutes les probabilités  $\mu_n$  ont leur support essentiellement dans  $[-K, K]$ .

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

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

On trouvera la démonstration de ce résultat dans [13, Th. 9.3.3] et [17, Prop. 5.21]; on vérifiera aisément que les suite de probabilités de l'exemple 3 ne sont pas tendues.

**Exemple 6.** Soit  $(\mu_n)$  une famille de probabilités sur  $\mathbb{R}$  vérifiant

$$\sup_n \int |\lambda| \mu_n(d\lambda) < \infty .$$

Alors  $(\mu_n)$  est tendue. En effet,

$$\sup_n \mu_n([-K, K]^c) \leq \frac{\sup_n \int |\lambda| \mu_n(d\lambda)}{K} \xrightarrow{K \rightarrow \infty} 0 .$$

**Proposition 7.** Étant donnée une suite de probabilités  $(\mu_n)$ , les deux conditions suivantes sont équivalentes :

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

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

On en déduit qu'à partir d'un certain rang,  $\mu_n([-(K+1), K+1]) \geq 1 - 2\varepsilon$ . 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  $(\mu_n)$  est tendue. Le théorème 5 permet alors de conclure.  $\square$

**Théorème 8.** Soit  $(\mu_n)$  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  $\nu$  de masse totale inférieure ou égale à 1 telle que

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

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

- (b) si de plus

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

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

**Remarque 9.** *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 8.* 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  $(\mu_n)$ . D'après le principe de sélection de Helly, il existe une sous-suite  $(\mu'_n)$  qui converge vaguement vers une sous-probabilité  $\nu$ . 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  $(\mu''_n)$  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 coïncident<sup>1</sup> aussi. Par suite  $\nu$  et  $\tilde{\nu}$  coïncident. On en déduit alors que de toute sous-suite de  $(\mu_n)$ , on peut extraire une sous-suite qui converge vaguement vers  $\nu$ , et finalement que la suite  $(\mu_n)$  converge vaguement vers  $\nu$ .

Si de plus  $iyg(iy)$  converge vers  $-1$ , alors  $\nu$  est une probabilité d'après le point (4) de la proposition 2. La proposition 7 permet alors de conclure. □

### 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.

1. On rappelle que si deux fonctions analytiques sur  $\mathbb{C}^+$  coïncident sur  $\mathcal{D}$ , alors elles coïncident sur  $\mathbb{C}^+$ , voir par exemple [21, Corollaire du théorème 10.18].

**Théorème 10** (Herglotz). *Si la fonction  $g : \mathbb{C}^+ \rightarrow \mathbb{C}$  satisfait les conditions suivantes :*

1.  $\text{Im}(g) \geq 0$ ,
2.  $g$  est analytique sur  $\mathbb{C}^+$ ,
3. il existe  $M > 0$  telle que

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

Alors il existe une unique mesure positive  $\mu$  sur  $\mathbb{R}$ , de masse totale  $\mu(\mathbb{R}) \leq M$ , telle que

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

De plus,  $\lim_{y \rightarrow \infty} -iyg(iy) = \mu(\mathbb{R})$ .

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

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

La démonstration du théorème est en annexe C.1.

**Exemple 11.** Soit  $g(z)$  la transformée de Stieltjes associée à une mesure positive, finie, de support inclus dans  $\mathbb{R}^+$ , alors

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

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

$$\text{Im}(h(z)) = \frac{\text{Im}(z) + \text{Im}(zg(z))}{|z(1+g(z))|^2} > 0 \quad \text{et} \quad \text{Im}(zh(z)) = \frac{\text{Im}(g(z))}{|1+g(z)|^2} > 0.$$

D'autre part,

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

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

## 1.4 La formule d'Helffer-Sjöstrand

Soit  $\mu$  une probabilité sur  $\mathbb{R}$  et  $g_\mu$  sa transformée de Stieltjes. Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction réelle à support compact, de classe  $C^{k+1}$ . On introduit l'extension [quasi-analytique? terminologie à vérifier] suivante de  $f$  sur  $\mathbb{C}^+$  définie par

$$\Phi_k(f)(x, y) = \sum_{\ell=0}^k \frac{(iy)^\ell}{\ell!} f^{(\ell)}(x) \chi(y), \quad x + iy \in \mathbb{C}^+, \quad (1.2)$$

où  $\chi : \mathbb{R} \rightarrow [0, 1]$  est une fonction paire  $C^\infty$  telle que

$$\chi(y) = 1 \quad \text{pour } 0 \leq y \leq \eta ; \quad \chi(y) = 0 \quad \text{pour } y \geq 1 .$$

L'extension (6.1) s'étend naturellement, si besoin est, au cas où  $y < 0$ .

On notera indifféremment  $\Phi_k(f)(x, y)$  ou  $\Phi_k(f)(z)$  avec  $z = x + iy$ . Ainsi,

$$\bar{\partial}\Phi_k(f) = \frac{\partial\Phi_k(f)}{\partial\bar{z}} = \frac{1}{2} \{ \partial_x \Phi_k(f) + i \partial_y \Phi_k(f) \} ,$$

où  $\partial_x = \frac{\partial}{\partial x}$  et  $\partial_y = \frac{\partial}{\partial y}$ .

**Propriétés 12.** — L'évaluation de  $\Phi_k(f)$  sur l'axe réel donne

$$\Phi_k(f)(x, 0) = f(x) .$$

- Le support de  $\Phi_k(f)$  est inclus dans  $\text{supp } f \times [0, 1]$ .
- Un calcul élémentaire lié au fait que  $\chi(y)$  est constante lorsque  $x + iy$  est proche de l'axe réel ( $y \leq \eta$ ) montre que

$$\bar{\partial}\Phi_k(f)(x, y) = \frac{(iy)^k}{k!} f^{(k+1)}(x) , \quad y \leq \eta . \quad (1.3)$$

**Proposition 13.** Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  de classe  $C^{k+1}$ , avec  $k \geq 1$ , à support compact et  $\Phi_k(f)$  son extension donnée par (6.1), alors

$$\int f(\lambda) \mu(d\lambda) = \frac{2}{\pi} \text{Re} \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial}\Phi_k(f)(x, y) g_\mu(x + iy) dx dy .$$

**Remarque 14.** Pour une fonction  $f \in C_c^{k+1}(\mathbb{R})$ , l'identité d'Helffer-Sjöstrand est valable pour toute extension  $\Phi_\ell(f)$  de  $f$  d'indice  $\ell$  compris entre 1 et  $k$ . Plus l'extension  $\Phi_\ell(f)$  a un indice  $\ell$  élevé, meilleure sera la convergence vers zéro de  $\bar{\partial}\Phi_\ell(f)$  lorsque  $z$  s'approche de la droite réelle, en vertu de la propriété

$$\bar{\partial}\Phi_k(f)(x, y) = \frac{(iy)^k}{k!} f^{(k+1)}(x) , \quad y \leq \eta .$$

Ceci est particulièrement utile lorsqu'on veut comparer

$$\int f d\mu - \int f d\nu$$

à partir d'une estimée sur la différence de leurs transformées de Stieltjes du type

$$g_\mu(z) - g_\nu(z) \propto \frac{1}{\text{Im}^k(z)} .$$

La régularité de  $f$  permettra alors de compenser la divergence de  $g_\mu(z) - g_\nu(z)$  près de l'axe réel :

$$\int f d\mu - \int f d\nu = \frac{2}{\pi} \text{Re} \int_{\mathbb{R} \times ]0, \infty[} \underbrace{\bar{\partial}\Phi_k(f)(z)}_{\propto y^k f^{(k+1)}(x)} \underbrace{(g_\mu - g_\nu)(z)}_{\propto \frac{1}{y^k}} dx dy .$$

[formaliser la remarque précédente sous la forme d'une proposition]

**Remarque 15.** *L'application de l'identité d'Helffer-Sjöstrand à la mesure empirique des valeurs propres d'une matrice hermitienne donne*

$$\frac{1}{N} \sum_{i=1}^N f(\lambda_i) = \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) g_n(x + iy) dx dy .$$

[expliquer comment on peut glisser une espérance à l'identité précédente]

*Démonstration de la proposition 83.* On montre facilement, en utilisant (6.2), que

$$(\lambda, x, y) \mapsto \bar{\partial} \Phi_k(f)(x, y) \frac{1}{\lambda - x - iy}$$

est bornée sur  $\mathbb{R}^2 \times \mathbb{R}^+$ . Le théorème de Fubini entraîne alors que

$$\begin{aligned} & \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) g_\mu(x + iy) dx dy \\ &= \int_{\mathbb{R}} \mu(d\lambda) \left\{ \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) \frac{1}{\lambda - x - iy} dx dy \right\} . \end{aligned}$$

Comme  $\frac{1}{\lambda - x - iy}$  est la transformée de Stieltjes de la mesure de Dirac en  $\lambda$ , il suffit d'établir la formule d'Helffer-Sjöstrand pour une telle mesure, soit :

$$f(\lambda) = \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) \frac{1}{\lambda - x - iy} dx dy \quad (1.4)$$

et d'utiliser l'argument du théorème de Fubini ci-dessus pour conclure. Montrons la formule pour la mesure de Dirac en zéro :

$$f(0) = -\frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \frac{\bar{\partial} \Phi_k(f)(x, y)}{x + iy} dx dy \quad (1.5)$$

la formule (6.3) suivra alors par simple changement de variable.

Dans la suite, on note simplement  $\Phi$  au lieu de  $\Phi_k(f)$ . On a :

$$\frac{\bar{\partial} \Phi(x, y)}{x + iy} = \frac{1}{2} \frac{\partial_x \Phi + i \partial_y \Phi}{x + iy} = \frac{1}{2} \frac{x \partial_x \Phi + y \partial_y \Phi + i(x \partial_y \Phi - y \partial_x \Phi)}{x^2 + y^2} .$$

On passe aux coordonnées polaires :

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases} , \quad dx dy = r dr d\theta , \quad \begin{cases} x \in \mathbb{R} \\ y \in ]0, \infty[ \end{cases} \Leftrightarrow \begin{cases} r \in ]0, \infty[ \\ \theta \in ]0, \pi[ \end{cases} .$$

Soit  $\Phi = \Phi(r \cos \theta, r \sin \theta)$ , alors

$$\partial_r \Phi = \cos \theta \partial_x \Phi + \sin \theta \partial_y \Phi \quad \Rightarrow \quad r \partial_r \Phi = x \partial_x \Phi + y \partial_y \Phi .$$

De même,

$$\partial_\theta \Phi = -r \sin \theta \partial_x \Phi + r \cos \theta \partial_y \Phi \quad \Rightarrow \quad \partial_\theta \Phi = -y \partial_x \Phi + x \partial_y \Phi ,$$

et

$$\frac{\bar{\partial} \Phi(x, y)}{x + iy} = \frac{1}{2} \left\{ \frac{\partial_r \Phi}{r} + i \frac{\partial_\theta \Phi}{r^2} \right\} ,$$

soit

$$\begin{aligned} \int_{\mathbb{R} \times ]0, \infty[} \frac{\bar{\partial} \Phi(f)(x, y)}{x + iy} dx dy &= \frac{1}{2} \int_{]0, \infty[ \times ]0, \pi[} \left\{ \frac{\partial_r \Phi}{r} + i \frac{\partial_\theta \Phi}{r^2} \right\} r dr d\theta \\ &= \frac{1}{2} \int_{]0, \infty[ \times ]0, \pi[} \partial_r \Phi dr d\theta + \frac{i}{2} \int_{]0, \infty[ \times ]0, \pi[} \frac{\partial_\theta \Phi}{r} dr d\theta \\ &= \frac{1}{2} \int_{]0, \pi[} d\theta [\Phi(r \cos \theta, r \sin \theta)]_{r=0}^{r=\infty} + \frac{i}{2} \int_{]0, \infty[} \frac{dr}{r} [\Phi(r \cos \theta, r \sin \theta)]_{\theta=0}^{\theta=\pi} \\ &= -\frac{\pi}{2} f(0) + \frac{i}{2} \int_0^\infty \frac{f(-r) - f(r)}{r} dr \end{aligned}$$

On obtient finalement bien le resultat escompté :

$$-\frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \frac{\bar{\partial} \Phi_k(f)(x, y)}{x + iy} dx dy = f(0) .$$

□

## 1.5 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  $g_{\text{MP}}$  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

$$\sqrt{z} = \sqrt{\rho} e^{i\frac{\theta}{2}} , \quad \theta \in (0, 2\pi) . \quad (1.6)$$

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

$$zc\sigma^2 X^2 + [z + \sigma^2(c-1)]X + 1 = 0. \quad (1.7)$$

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 (1.6). 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 (1.7) où  $z$  a été remplacé par  $x$ .

4. Si  $z \in \mathbb{C}^+$ ,  $x \in \mathbb{R} \setminus \{0\}$ ,  $x < \lambda^-$ , 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 (1.7) où  $z$  a été remplacé par  $x$ .

**Remarque 17.** Dans la suite, on se référera souvent à l'équation (1.7) 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 (1.7). 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  $g_{\tilde{\text{MP}}}$ , solution de (1.7), 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 strictement compris entre 0 et  $\pi$ ; l'argument du discriminant  $(z - \lambda^+) (z - \lambda^-)$  est alors strictement compris entre 0 et  $2\pi$  et par le choix (1.6) de la branche de la racine carrée, celui de sa racine est strictement compris entre 0 et  $\pi$ . En particulier,

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

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

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

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

$$\begin{aligned} g_{\tilde{\text{MP}}}(z) &= \frac{(z - \lambda^+) (z - \lambda^-) - [z + \sigma^2(c-1)]^2}{2c\sigma^2 z \left( \sqrt{(z - \lambda^+) (z - \lambda^-)} + [z + \sigma^2(c-1)] \right)} \\ &= \frac{-4c\sigma^2 z}{2c\sigma^2 z \left( \sqrt{(z - \lambda^+) (z - \lambda^-)} + [z + \sigma^2(c-1)] \right)} \\ &= -\frac{2}{\sqrt{(z - \lambda^+) (z - \lambda^-)} + [z + \sigma^2(c-1)]} . \end{aligned} \tag{1.8}$$

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

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

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

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

Ainsi  $g_{\tilde{\text{MP}}}$  est la transform e de Stieltjes d'une mesure positive sur  $\mathbb{R}$ . Toujours gr ce   la repr sentation (1.8), on montre facilement que

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

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  $\text{Arg}(z - \lambda^+)(z - \lambda^-) \searrow 0$ . 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 (1.7) quand  $x$  remplace  $z$ , cela entraîne le point 3 du théorème.

Soit  $x \in ]-\infty, \lambda^-[$ ,  $x \neq 0$  et  $z \in \mathbb{C}^+$  tel que  $z \rightarrow x$ . Un raisonnement similaire entraîne que  $\text{Arg}(z - \lambda^+)(z - \lambda^-) \nearrow 2\pi$ . 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

$$\frac{1}{\pi} \text{Im} g_{\check{\text{MP}}}(z) \xrightarrow{z \rightarrow x} 0. \quad (1.9)$$

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

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

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

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

et

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

soit

$$\mathbb{P}_{\check{\text{MP}}}(\{0\}) = \left(1 - \frac{1}{c}\right)_+. \quad (1.10)$$

Déterminons maintenant la densité de  $\mathbb{P}_{\check{\text{MP}}}$  sur le segment  $]\lambda^-, \lambda^+]$  (on exclut  $\lambda^-$  qui peut valoir zéro). Soit  $x \in ]\lambda^-, \lambda^+]$  et  $z \in \mathbb{C}^+$ ,  $z \rightarrow x$ . Alors

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

Par suite,

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

et

$$\frac{1}{\pi} \operatorname{Im} g_{\tilde{\text{MP}}}(z) \xrightarrow{z \rightarrow x} \frac{\sqrt{(\lambda^+ - x)(x - \lambda^-)}}{2c\pi\sigma^2 x}. \quad (1.11)$$

On peut d'autre part facilement borner la transformée de Stieltjes  $g_{\tilde{\text{MP}}}$  sur tout ensemble de la forme  $[a, b] \times [0, \varepsilon]$  où

$$[a, b] \subset ]-\infty, 0[ \cup ]0, \lambda^-[ \cup ]\lambda^-, \infty[. \quad (1.12)$$

En utilisant les convergences (1.9) et (1.11) et le théorème de convergence dominée, il vient

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

pour toute fonction  $f$  à support compact inclus dans  $[a, b]$  qui vérifie (1.12). En combinant ce résultat avec (1.10), on peut maintenant complètement décrire la fonction de répartition de  $\mathbb{P}_{\tilde{\text{MP}}}$ . □

## 1.6 Exercices

**Exercice 1** Soit  $(x_i, i \geq 1)$  une suite de vecteurs centrés i.i.d. à valeurs  $\mathbb{R}^N$  de matrice de covariance  $\mathbb{E}x_1 x_1^T = R_N$ . On note  $X_n = [x_1 \cdots x_n]$  la matrice de dimensions  $N \times n$ .

1. Montrer que presque sûrement

$$\frac{1}{n} X_n X_n^T = \frac{1}{n} \sum_{i=1}^n x_i x_i^T \xrightarrow{n \rightarrow \infty, N \text{ fixé}} R_N.$$

2. En déduire que presque sûrement  $L_N \rightarrow L^R$  où  $L_N$  est la mesure spectrale de  $\frac{1}{n} X_n X_n^*$  et  $L^R$ , la mesure spectrale de  $R_N$ . (*Indication : on pourra montrer que  $A \mapsto \frac{1}{N} \operatorname{Tr}(A - zI)^{-1}$  est continue.*)
3. Exprimer  $L^R$  en fonction des valeurs propres de  $R_N$ .
4. Quelle est la mesure spectrale de  $R_N = \sigma^2 I_N$ ? En déduire la limite de  $L_N$  dans ce cas.

### Exercice 2

1. Soit  $f$  la transformée de Stieltjes d'une mesure  $\mu$  sur  $\mathbb{R}$ . Montrer alors que la fonction  $g(z) = -(z + f(z))^{-1}$  est la transformée de Stieltjes d'une mesure de probabilité  $\nu$ . En déduire que  $|z + f(z)|^{-1} \leq \Im(z)^{-1}$  pour  $z \in \mathbb{C}^+$ .

2. Soit  $f$  la transformée de Stieltjes d'une mesure  $\mu$  sur  $\mathbb{R}^+$  ( $\mu(-\infty, 0) = 0$ ).  
Montrer alors que la fonction  $g$  :

$$g(z) = -\frac{1}{z(1 + \alpha f(z))}, \quad \alpha > 0,$$

est la transformée de Stieltjes d'une mesure de probabilité  $\nu$ . En déduire que :

$$\left| \frac{1}{1 + \alpha f(z)} \right| \leq \frac{|z|}{\Im(z)} \quad \forall z \in \mathbb{C}^+.$$

Le support de  $\nu$  est-il inclus dans  $\mathbb{R}^+$  ?

**Exercice 3** (TS DE LA SYMÉTRISÉE DES VALEURS SINGULIÈRES) Soit  $\mu$  une mesure positive sur  $\mathbb{R}^+$  de masse totale finie  $\mu(\mathbb{R}^+) = \kappa$ . On définit la mesure  $\nu$  sur  $\mathbb{R}^+$  par

$$\int_{\mathbb{R}^+} f(x) \nu(dx) = \int_{\mathbb{R}^+} f(\sqrt{\lambda}) \mu(d\lambda)$$

pour toute fonction  $f$  continue bornée.

Soit  $\tilde{\nu}$  la symétrisée de  $\nu$  définie sur  $\mathbb{R}$  par

$$\int_{\mathbb{R}} f(y) \tilde{\nu}(dy) = \frac{1}{2} \left\{ \int_{\mathbb{R}^+} f(x) \nu(dx) + \int_{\mathbb{R}^+} f(-x) \nu(dx) \right\}$$

pour toute fonction  $f$  continue bornée.

- Calculer, dans le cas où  $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ , les mesures  $\nu$  et  $\tilde{\nu}$ .
- Si  $m_\mu$  et  $m_{\tilde{\nu}}$  sont les transformées de Stieltjes associées aux mesures  $\mu$  et  $\tilde{\nu}$ , montrer que

$$m_{\tilde{\nu}}(z) = z m_\mu(z^2), \quad z \in \mathbb{C}^+.$$

Soit  $A$  une matrice  $n \times n$  et  $\mu$  la mesure spectrale de  $AA^*$ .

- Exprimer la transformée de Stieltjes de la mesure spectrale  $\tilde{\nu}$  associée à la matrice

$$\Gamma = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}$$

en fonction de la transformée de Stieltjes de  $\mu$  (*indication : on pourra utiliser les formules d'inversion par blocs de la proposition 26*).

**Exercice 4** (TRANSFORMÉE DE STIELTJES DE LA LOI DU DEMI-CERCLE) Dans la suite de l'exercice, on considèrera la branche suivante de la racine carrée, définie sur  $\mathbb{C} \setminus \mathbb{R}^+$  par

$$\sqrt{z} = \sqrt{r} e^{i\frac{\theta}{2}} \quad \text{si } z = r e^{i\theta}.$$

On rappelle alors que  $z \mapsto \sqrt{z}$  définit une fonction analytique sur  $\mathbb{C} \setminus \mathbb{R}^+$ .

- En analysant l'argument de  $(z-2)(z+2)$  pour  $z \in \mathbb{C}^+$ , montrer que  $z \mapsto \sqrt{z^2-4}$  est une application analytique de  $\mathbb{C}^+$  dans  $\mathbb{C}^+$ .
- Pour  $x$  réel, on pose  $x_+ = \max(x, 0)$ . Soit  $z = x + iy \in \mathbb{C}^+$ , montrer que

$$\lim_{z \in \mathbb{C}^+, \Im(z) \downarrow 0} \Im \sqrt{z^2-4} = \sqrt{(4-x^2)_+}.$$

On considère  $m(z)$  solution de l'équation  $X^2 + zX + 1 = 0$  définie par

$$m(z) = \frac{-z + \sqrt{z^2 - 4}}{2} .$$

3. Montrer que  $m(z)$  est la transformée de Stieltjes d'une mesure positive  $\mu$  dont on précisera la masse totale.
4. Montrer que pour  $z = x + iy \in \mathbb{C}^+$ ,

$$\lim_{y \downarrow 0} \frac{1}{\pi} \Im m(x + iy) = \frac{1}{2\pi} \sqrt{(4 - x^2)_+} .$$

5. En déduire que  $\mu(dx) = \frac{1}{2\pi} \sqrt{(4 - x^2)_+} dx$ .

## Chapitre 2

# Compléments d'algèbre linéaire

Le théorème spectral est supposé acquis.

### 2.1 Décomposition spectrale et résolvante

On notera  $I$  la matrice identité (en indiquant parfois sa dimension  $I = I_N$ ). Soient  $\mathbf{a} = (a_1, \dots, a_N)^T$  et  $\mathbf{b} = (b_1, \dots, b_N)^T$  deux vecteurs de dimension  $N$ , alors  $\mathbf{a}\mathbf{b}^* = (a_i \overline{b_j})$  est une matrice  $n \times n$  de rang 1 et  $\mathbf{a}^*\mathbf{b} = \sum_{i=1}^N \overline{a_i} b_i$  est le produit scalaire usuel de  $\mathbf{a}$  et  $\mathbf{b}$ .

Etant donnée une matrice hermitienne  $A$  de dimensions  $n \times n$ , on considère sa décomposition spectrale

$$A = U\Lambda U^*$$

où  $U$  est unitaire ( $UU^* = U^*U = I$ ) et  $\Lambda = \text{diag}(\lambda_i)$  est diagonale, d'éléments diagonaux les valeurs propres de  $A$ . Alors  $Q(z) = (A - zI)^{-1}$  [bien définie!] admet la décomposition spectrale

$$Q(z) = U(\Lambda - zI)^{-1}U^* .$$

On remarque que  $Q(z)$  n'est pas hermitienne car ses éléments diagonaux ne sont pas réels; on note que  $Q^*(z) = Q(\bar{z})$ . On note  $(\mathbf{u}_i)$  les vecteurs colonne de la matrice unitaire  $U$ , alors  $\mathbf{u}_i$  est le vecteur propre associé à  $\lambda_i$  et  $A$  s'écrit

$$A = \sum_{i=1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^*$$

(on vérifiera bien que  $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ ).

**Exemple 18** (Identités utiles). *La décomposition précédente entraîne*

$$Q(z) = \sum_{i=1}^N \frac{1}{\lambda_i - z} \mathbf{u}_i \mathbf{u}_i^*, \quad Q^*(z) = \sum_{i=1}^N \frac{1}{\lambda_i - \bar{z}} \mathbf{u}_i \mathbf{u}_i^*, \quad Q(z)Q^*(z) = \sum_{i=1}^N \frac{1}{|\lambda_i - z|^2} \mathbf{u}_i \mathbf{u}_i^* .$$

Soit  $\mathbf{x}$  un vecteur  $N \times 1$ , alors

$$\begin{aligned} \mathbf{x}^* Q(z) \mathbf{x} &= \sum_{i=1}^N \frac{\mathbf{x}^* \mathbf{u}_i \mathbf{u}_i^* \mathbf{x}}{\lambda_i - z} = \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{\lambda_i - z}, \\ \|Q(z) \mathbf{x}\|^2 &= \mathbf{x}^* Q(z)^* Q(z) \mathbf{x} = \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{|\lambda_i - z|^2} \end{aligned}$$

**Proposition 19** (Identité de la résolvante). *Soient  $A$  et  $B$  deux matrices  $N \times N$  inversibles, alors*

$$A^{-1} - B^{-1} = -A^{-1}(A - B)B^{-1} .$$

La démonstration est immédiate.

[rajouter quelques formules de différentiation de la résolvante par rapport aux entrées de la matrice  $A$ ]

## 2.2 Norme spectrale

Étant donné un vecteur  $\mathbf{a}$  de dimensions  $N \times 1$ , on note  $\|\mathbf{a}\|$  sa norme euclidienne :

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

Pour une matrice  $A$  de dimensions  $N \times N$ , on note  $\|A\|$  sa norme spectrale :

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

Étant donnée que les notations entre norme euclidienne et norme spectrale sont identiques, l'ambiguïté est levée en analysant l'objet dont on prend la norme : si c'est un vecteur, alors il s'agit de la norme euclidienne alors que si c'est une matrice, c'est la norme spectrale.

**Proposition 20.** *Étant donnée une matrice  $A$  de dimensions  $N \times N$ , alors*

1.  $\|A^*\| = \|A\|$ ,
2. Si  $A$  est hermitienne alors  $\|A\| = \max\{|\lambda|, \lambda \text{ valeur propre de } A\}$ ; en particulier,

$$\|AA^*\| = \max\{\lambda, \lambda \text{ valeur propre de } AA^*\} .$$

3. On a  $\|A\| = \max\{\|A\mathbf{u}\|, \|\mathbf{u}\| \leq 1\}$  et en particulier

$$\|A\mathbf{x}\| \leq \|A\| \times \|\mathbf{x}\|$$

pour  $\mathbf{x} \in \mathbb{C}^N$ .

4. On a  $\|A\| = \max\{|\mathbf{u}^* A \mathbf{v}|, \|\mathbf{u}\| \leq 1, \|\mathbf{v}\| \leq 1\}$  et en particulier

$$|\mathbf{x}^* A \mathbf{y}| \leq \|A\| \times \|\mathbf{x}\| \times \|\mathbf{y}\|$$

pour  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$ .

5. Soient  $A$  et  $B$  deux matrices de dimension  $N \times N$ , alors

$$\|AB\| \leq \|A\| \times \|B\| .$$

6. Soit  $A$  une matrice hermitienne, semi-définie positive et  $B$  quelconque, de même dimensions, alors

$$|\text{Tr}(AB)| \leq \|B\| \text{Tr}(A) .$$

En particulier, si  $B$  est une matrice  $N \times N$  alors  $|\text{Tr}(B)| \leq N\|B\|$ .

*Démonstration.* 1. Si  $\lambda$  est une valeur propre non nulle de  $AA^*$ , alors c'est aussi une valeur propre de  $A^*A$ . En effet, soit  $\mathbf{u} \neq 0$  le vecteur propre associé alors

$$AA^*\mathbf{u} = \lambda\mathbf{u} \quad \Rightarrow \quad A^*A(A^*\mathbf{u}) = \lambda(A^*\mathbf{u})$$

et nécessairement  $A^*\mathbf{u} \neq 0$  (sans quoi  $\lambda$  serait nulle). Par suite,

$$\begin{aligned} \|A^*\| &= \max\{\sqrt{\lambda}, \lambda \text{ valeur propre de } A^*A\} , \\ &= \max\{\sqrt{\lambda}, \lambda \text{ valeur propre de } AA^*\} = \|A\| . \end{aligned}$$

2. Le théorème spectral nous donne

$$A = U\Lambda U^* \quad \Rightarrow \quad AA^* = U\Lambda^2 U^* .$$

On en déduit donc que si  $\lambda \in \text{sp}(A)$  alors  $\lambda^2 \in \text{sp}(AA^*)$  et si  $\lambda^2 \in \text{sp}(AA^*)$ , alors  $\lambda$  ou  $-\lambda \in \text{sp}(A)$ , d'où le résultat. En particulier, on a  $\|AA^*\| = \max\{\lambda, \lambda \text{ valeur propre de } AA^*\}$ .

3. On a vu que  $\|A\| = \max\{\sqrt{\lambda}, \lambda \text{ valeur propre de } A^*A\}$ . Appelons  $\lambda_{\max}$  la plus grande valeur propre de  $A^*A$ . Alors

$$\lambda_{\max} = \sup_{\|\mathbf{x}\| \leq 1} |\mathbf{x}A^*A\mathbf{x}| = \sup_{\|\mathbf{x}\| \leq 1} \|A\mathbf{x}\|^2$$

(on montre l'inégalité  $\leq$  en particulierisant  $\mathbf{x}$  et en le prenant égal au vecteur propre unitaire associé à  $\lambda_{\max}$ , et l'inégalité  $\geq$  à l'aide du théorème spectral). L'égalité demandée en découle vu que  $\|A\| = \sqrt{\lambda_{\max}}$ .

4. On a  $\|A\mathbf{x}\| \leq \sup_{\|\mathbf{y}\| \leq 1} |\mathbf{y}^* A \mathbf{x}|$ . En effet, il suffit de prendre  $\mathbf{y} = \frac{A\mathbf{x}}{\|A\mathbf{x}\|}$ . D'autre part,

$$|\mathbf{y}^* A \mathbf{x}|^2 = \mathbf{x}^* A^* \mathbf{y} \mathbf{y}^* A \mathbf{x} = \|A\mathbf{x}\|^2 \frac{\mathbf{x}^* A^*}{\|A\mathbf{x}\|} \mathbf{y} \mathbf{y}^* \frac{A\mathbf{x}}{\|A\mathbf{x}\|} \leq \|A\mathbf{x}\|^2 \|\mathbf{y} \mathbf{y}^*\| .$$

Or la matrice  $\mathbf{y} \mathbf{y}^*$  est de rang 1 dont la seule valeur propre non nulle est  $\|\mathbf{y}\|^2$ . Par suite,

$$\sup\{|\mathbf{y}^* A \mathbf{x}|, \|\mathbf{x}\| \leq 1, \|\mathbf{y}\| \leq 1\} \leq \sup\{\|A\mathbf{x}\| \times \|\mathbf{y}\|, \|\mathbf{x}\| \leq 1, \|\mathbf{y}\| \leq 1\} = \|A\|$$

5. On a

$$\begin{aligned} \|AB\| &= \max\{|\mathbf{y}^* AB \mathbf{x}|, \|\mathbf{x}\| \leq 1, \|\mathbf{y}\| \leq 1\} , \\ &\leq \max\{\|A^* \mathbf{y}\| \times \|B \mathbf{x}\|, \|\mathbf{x}\| \leq 1, \|\mathbf{y}\| \leq 1\} , \\ &= \max\{\|B \mathbf{x}\|, \|\mathbf{x}\| \leq 1\} \times \max\{\|A^* \mathbf{y}\|, \|\mathbf{y}\| \leq 1\} \\ &= \|A\| \times \|B\| . \end{aligned}$$

6. Le théorème spectral donne  $A = U\Lambda U^*$  avec  $\Lambda$  matrice diagonale à entrées positives, soit  $A = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^*$ . On a donc

$$|\text{trace}(AB)| = \left| \sum_{i=1}^n \lambda_i \mathbf{u}_i^* B \mathbf{u}_i \right| \leq \sum_i \lambda_i \|B\| = \text{trace}(A) \|B\| .$$

En particulierisant  $A = I_N$ , on obtient la deuxième inégalité. □

**Exemple 21.** Soit  $A$  une matrice hermitienne et  $Q(z) = (q_{ij}(z))$  sa résolvante, alors

$$\|Q(z)\| \leq \frac{1}{\text{Im}(z)} , \quad |q_{ij}(z)| \leq \frac{1}{\text{Im}(z)} , \quad z \in \mathbb{C}^+ .$$

En effet, les valeurs propres de  $Q(z)Q^*(z)$  sont  $|\lambda_i - z|^{-2}$ , toutes majorées par  $\text{Im}(z)^{-2}$ . Cela entraîne la majoration de la norme spectrale de la résolvante. Le point 4 de la proposition précédente permet de conclure en considérant les vecteurs  $(\mathbf{e}_i)$  de la base canonique car  $\mathbf{e}_i^* Q(z) \mathbf{e}_j = q_{ij}(z)$ .

## 2.3 Identités matricielles

**Proposition 22** (Identité de Woodbury). Soit  $A$  une matrice  $N \times N$ ,  $U$  une matrice  $N \times k$ ,  $B$  une matrice  $k \times k$ ,  $V$  une matrice  $k \times N$ . On suppose que tous les inverses matriciels considérés existent, alors

$$(A + UB V)^{-1} = A^{-1} - A^{-1} U (B^{-1} + V A^{-1} U)^{-1} V A^{-1} .$$

**Remarque 23.** Comme  $B$  est une matrice  $k \times k$ , la matrice  $UBV$  est une matrice de rang (au plus)  $k$  et  $A + UBV$  est une perturbation de rang  $k$  de la matrice  $A$ . De même, la matrice  $(B^{-1} + VA^{-1}U)^{-1}$  est de rang  $k$ . L'identité de Woodbury exprime le fait que l'inverse d'une perturbation de rang  $k$  de  $A$  est une perturbation de rang  $k$  de la matrice  $A^{-1}$ .

*Démonstration.* On a

$$\begin{aligned}
& (A + UBV) (A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1}) \\
&= I + UBVA^{-1} - U(B^{-1} + VA^{-1}U)^{-1}VA^{-1} - UBVA^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1} \\
&= I + U [B - (B^{-1} + VA^{-1}U)^{-1} - BVA^{-1}U(B^{-1} + VA^{-1}U)^{-1}] VA^{-1} \\
&= I + UB [I - B^{-1}(B^{-1} + VA^{-1}U)^{-1} - VA^{-1}U(B^{-1} + VA^{-1}U)^{-1}] VA^{-1} \\
&= I + UB [I - (B^{-1} + VA^{-1}U)(B^{-1} + VA^{-1}U)^{-1}] VA^{-1} \\
&= I + UB [I - I] VA^{-1} \\
&= I.
\end{aligned}$$

□

La spécialisation de l'identité de Woodbury au cas d'une perturbation de rang 1 nous sera particulièrement utile :

**Corollaire 24** (Identité de Sherman-Morrison). *Soit  $A$  une matrice  $N \times N$  et  $\mathbf{u}, \mathbf{v}$  deux vecteurs de dimension  $N \times 1$ . On suppose que tous les inverses matriciels considérés existent, alors*

$$(A + \mathbf{u}\mathbf{v}^*)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^*A^{-1}}{1 + \mathbf{v}^*A^{-1}\mathbf{u}}.$$

De plus,

$$\begin{aligned}
\text{Tr} (A + \mathbf{u}\mathbf{v}^*)^{-1} &= \text{Tr} A^{-1} - \frac{\mathbf{v}^*A^{-2}\mathbf{u}}{1 + \mathbf{v}^*A^{-1}\mathbf{u}}, \\
(A + \mathbf{u}\mathbf{v}^*)^{-1}\mathbf{u}\mathbf{v}^* &= \frac{1}{1 + \mathbf{v}^*A^{-1}\mathbf{u}}A^{-1}\mathbf{u}\mathbf{v}^*.
\end{aligned}$$

**Exemple 25** (Perturbation de rang 1 pour la résolvante). *On considère les résolvantes*

$$Q(z) = \left( \sum_{k=1}^n \mathbf{y}_k \mathbf{y}_k^* - zI_N \right)^{-1} \quad \text{et} \quad Q_i(z) = \left( \sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^* - zI \right)^{-1},$$

où les  $(\mathbf{y}_i)$  sont des vecteurs aléatoires à valeurs  $\mathbb{C}^N$ , i.i.d. L'identité de Sherman-Morrison nous donne

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

D'un point de vue probabiliste, l'intérêt du membre de droite réside dans le fait que le vecteur  $\mathbf{y}_i$  est indépendant de la matrice  $Q_i(z)$ .

On s'intéresse maintenant à l'inversion par blocs d'une matrice carrée. On considère la matrice  $M$  de dimension  $N \times N$ , partitionnée selon les blocs :

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

les blocs diagonaux  $A$  et  $D$  étant carrés. Si ces blocs sont inversibles, on introduit les quantités :

$$\mathcal{S}_A = D - CA^{-1}B \quad \text{et} \quad \mathcal{S}_D = A - BD^{-1}C,$$

appelées compléments de Schur de  $A$  et  $D$  dans  $M$ . Les compléments de Schur  $\mathcal{S}_A$  et  $\mathcal{S}_D$  apparaissent dans la diagonalisation par blocs de la matrice  $M$  :

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & \mathcal{S}_A \end{bmatrix}, \quad (2.1)$$

et

$$\begin{bmatrix} I & -CD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -D^{-1}B & I \end{bmatrix} = \begin{bmatrix} \mathcal{S}_D & 0 \\ 0 & D \end{bmatrix}.$$

Ces identités permettent alors d'établir les formules d'inversion par bloc suivantes.

**Proposition 26** (Formules d'inversion par blocs). *En supposant que tous les inverses considérés existent, l'inverse de la matrice  $M$  est donné par la formule*

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B\mathcal{S}_A^{-1}CA^{-1} & -A^{-1}B\mathcal{S}_A^{-1} \\ -\mathcal{S}_A^{-1}CA^{-1} & \mathcal{S}_A^{-1} \end{bmatrix}$$

qui fait intervenir le complément de Schur  $\mathcal{S}_A$ . On obtient une formule similaire en utilisant  $\mathcal{S}_D$  :

$$M^{-1} = \begin{bmatrix} \mathcal{S}_D^{-1} & -\mathcal{S}_D^{-1}CD^{-1} \\ -D^{-1}B\mathcal{S}_D^{-1} & D^{-1} + D^{-1}B\mathcal{S}_D^{-1}CD^{-1} \end{bmatrix}.$$

*Démonstration.* On montre la première identité. L'équation (2.1) nous donne immédiatement

$$M^{-1} = \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \mathcal{S}_A^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}.$$

Il suffit alors d'effectuer le produit matriciel pour obtenir le résultat annoncé.  $\square$

La formule d'inversion par blocs, combinée à l'identité de Woodbury, va nous permettre d'obtenir une expression d'un élément diagonal  $q_{ii}(z)$  d'une résolvante  $Q(z)$ .

**Proposition 27** (Expression d'un élément diagonal de la résolvante). *Etant donnée une matrice  $\Sigma$  de dimensions  $N \times n$ , on dénote  $\xi_i$  la  $i$ ème ligne de  $\Sigma$  et  $\Sigma_i$  la matrice  $\Sigma$  privée de sa ligne  $\xi_i$ , de dimensions  $(N-1) \times n$ . Soit  $Q(z)$  la résolvante  $Q(z) = (\Sigma\Sigma^* - zI)^{-1} = (q_{ij}(z))$ . Alors*

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

*Démonstration.* Raisonnons sur  $q_{11}(z)$  et notons  $\Sigma_1$  la matrice  $\Sigma$  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 26, 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 Woodbury (proposition 22) en particulier

$$A = -zI, \quad U = \Sigma_1^*, \quad V = \Sigma_1, \quad B = I,$$

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

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

□

## 2.4 Inégalités matricielles

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.

La proposition suivante sera particulièrement utile dans la suite.

**Proposition 28** (Inégalité de perturbation de rang 1). *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 le corollaire 24 (identité de Sherman-Morrison), il vient

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

Posons  $Q(z) = (A - zI)^{-1}$ . On a alors

$$|\mathrm{Tr} B(A + \mathbf{x}\mathbf{x}^* - zI)^{-1} - \mathrm{Tr} B(A - zI)^{-1}| \leq \frac{|\mathbf{x}^*Q(z)BQ(z)\mathbf{x}|}{|\mathrm{Im}(1 + \mathbf{x}^*Q(z)\mathbf{x})|}. \quad (2.2)$$

La décomposition spectrale  $A = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^*$  entraîne (cf. l'exemple 18) :

$$\mathrm{Im}(1 + \mathbf{x}^*Q(z)\mathbf{x}) = \mathrm{Im}(z) \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{|\lambda_i - z|^2}. \quad (2.3)$$

car

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

D'autre part,

$$|\mathbf{x}^*Q(z)BQ(z)\mathbf{x}| \leq \|B\| \times \|Q(z)\mathbf{x}\| \times \|Q(\bar{z})\mathbf{x}\| = \|B\| \sum_{i=1}^N \frac{|\mathbf{x}^* \mathbf{u}_i|^2}{|\lambda_i - z|^2}.$$

Soit finalement,

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

□

## 2.5 Calcul fonctionnel pour les matrices hermitiennes

Etant donnée une matrice hermitienne  $A$  de dimensions  $n \times n$ , de décomposition spectrale  $U^* \Delta U$  où  $\Delta = \mathrm{diag}(\lambda_i)$  et une fonction  $f$  continue à support compact,

2.5. CALCUL FONCTIONNEL POUR LES MATRICES HERMITIENNES 43

on définit la matrice  $f(A)$  par

$$f(A) = U^* \begin{pmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{pmatrix} U .$$

**Remarque 29.** En considérant la mesure empirique des valeurs propres  $L_n = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$ , on obtient une expression alternative pour  $\int f d L_n$  :

$$\int_{\mathbb{R}} f(x) L_n(dx) = \frac{1}{N} \sum_{i=1}^N f(\lambda_i) = \frac{1}{N} \text{Tr } f(A) .$$

On remarque alors que, si la fonction  $f$  est suffisamment régulière ( $f \in C_c^{k+1}(\mathbb{R})$ ), la définition de  $f(A)$  est compatible avec la formule d'Helffer-Sjöstrand. En effet :

$$\begin{aligned} f(A) &= U^* \begin{pmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{pmatrix} U \\ &= \frac{2}{\pi} U^* \int_{\mathbb{C}^+} \begin{pmatrix} \text{Re} \left\{ \frac{\bar{\partial} \Phi_k(f)}{\lambda_1 - z} \right\} & & 0 \\ & \ddots & \\ 0 & & \text{Re} \left\{ \frac{\bar{\partial} \Phi_k(f)}{\lambda_n - z} \right\} \end{pmatrix} dx dy U \\ &= \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_k(f)(z) U^* (\Delta - zI)^{-1} U dx dy + \frac{1}{\pi} \int_{\mathbb{C}^+} \overline{\bar{\partial} \Phi_k(f)(z)} U^* (\Delta - \bar{z}I)^{-1} U dx dy \end{aligned}$$

Cela permet d'obtenir une expression alternative de la matrice  $f(A)$  à l'aide de l'extension  $\Phi_k(f)$  et de la résolvante  $(A - zI)^{-1}$  :

$$f(A) = \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_k(f)(z) (A - zI)^{-1} dx dy + \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_k(f)(\bar{z}) (A - \bar{z}I)^{-1} dx dy .$$

**Proposition 30.** Soit  $X$  une matrice  $N \times n$  à entrées complexes et  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction de classe  $C^3$  à support compact. On pose

$$A = \frac{\sigma^2}{n} X X^* .$$

Alors

$$\begin{aligned} \frac{\partial}{\partial \bar{X}_{k\ell}} \left\{ \frac{1}{N} \text{Tr } f(A) \right\} &= \frac{\sigma^2}{Nn} [f'(A)X]_{k\ell} , \\ \frac{\partial}{\partial X_{k\ell}} \left\{ \frac{1}{N} \text{Tr } f(A) \right\} &= \frac{\sigma^2}{Nn} [X^* f'(A)]_{\ell k} , \end{aligned}$$

où  $f'$  représente la dérivée de  $f$  par rapport à  $x$ .

*Démonstration.* Notons  $g_n(z) = \frac{1}{N} \text{Tr}(A - zI)^{-1}$ , alors

$$\frac{1}{N} \text{Tr} f(A) = \frac{2}{\pi} \text{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) g_n(z) dx dy .$$

Du fait que  $|g_n(z)| \leq y^{-1}$  et que  $\bar{\partial} \Phi_2(f)(z) = -y^2 f^{(3)}(x)/2$  près de l'axe réel, l'intégrande est uniformément majoré en  $\bar{X}_{k\ell}$  et le théorème de dérivation sous le signe intégral nous donne :

$$\begin{aligned} \frac{\partial}{\partial \bar{X}_{k\ell}} \left\{ \frac{1}{N} \text{Tr} f(A) \right\} &= \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) \frac{\partial g_n(z)}{\partial \bar{X}_{k\ell}} dx dy + \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(\bar{z}) \frac{\partial g_n(\bar{z})}{\partial \bar{X}_{k\ell}} dx dy , \\ &= -\frac{\sigma^2}{Nn} \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) [Q^2(z)X]_{k\ell} dx dy \\ &\quad - \frac{\sigma^2}{Nn} \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(\bar{z}) [Q^2(\bar{z})X]_{k\ell} dx dy \\ &= -\frac{\sigma^2}{Nn} \frac{2}{\pi} \sum_i X_{i\ell} \text{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) Q_{ki}^2(z) dx dy \end{aligned}$$

Remarquons que

$$Q^2(z) = Q^2(x + iy) = \frac{\partial}{\partial x} Q(x + iy) .$$

En effectuant une intégration par parties (selon la variable  $x$ ), il vient

$$\int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) Q_{ki}^2(z) dx dy = - \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f')(z) Q_{ki}(z) dx dy ,$$

soit finalement

$$\begin{aligned} \frac{\partial}{\partial \bar{X}_{k\ell}} \left\{ \frac{1}{N} \text{Tr} f(A) \right\} &= \frac{\sigma^2}{Nn} \frac{2}{\pi} \sum_i X_{i\ell} \text{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f')(z) Q_{ki}(z) dx dy , \\ &= \frac{\sigma^2}{Nn} \sum_i X_{i\ell} [f'(A)]_{ki} \\ &= \frac{\sigma^2}{Nn} [f'(A)X]_{k\ell} . \end{aligned}$$

Un calcul similaire nous permet d'obtenir la deuxième identité de la proposition.  $\square$

## 2.6 Exercices

**Exercice 1** Retrouver, à partir de la formule d'inversion par blocs, l'identité de Woodbury.

## Chapitre 3

# Estimation de variances

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.

### 3.1 Inégalité d'Efron-Stein

L'inégalité d'Efron-Stein est particulièrement efficace pour obtenir des bonnes estimations de variances dans un contexte de grandes matrices aléatoires. Elle ne nécessite pas de connaître l'espérance de la fonction dont on cherche à estimer la variance.

**Proposition 31.** *Soient  $y_1, \dots, y_n$   $n$  vecteurs aléatoires indépendants à valeurs  $\mathbb{R}^d$ , 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

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

où  $\mathbb{E}_{\{i\}}$  représente l'espérance conditionnelle par rapport aux variables  $y_k, k \neq i$ .

*Démonstration.* On introduit la filtration suivante  $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(y_1, \dots, y_i)$  et l'espérance conditionnelle associée  $\mathbb{E}_i = \mathbb{E}(\cdot | \mathcal{F}_i)$ . En remarquant que  $f = \mathbb{E}_n f$  et  $\mathbb{E} f = \mathbb{E}_0 f$ , on a

$$f - \mathbb{E} f = \sum_{i=1}^n (\mathbb{E}_i f - \mathbb{E}_{i-1} f) ,$$

soit

$$\mathbb{E}(f - \mathbb{E}f)^2 = \sum_{i=1}^n \mathbb{E}(\mathbb{E}_i f - \mathbb{E}_{i-1} f)^2 + 2 \sum_{i < j} \mathbb{E}(\mathbb{E}_i f - \mathbb{E}_{i-1} f)(\mathbb{E}_j f - \mathbb{E}_{j-1} f).$$

En remarquant que  $\mathbb{E}_i f$  et  $\mathbb{E}_{i-1} f$  sont  $\mathcal{F}_{j-1}$ -mesurables, on obtient

$$\begin{aligned} & \mathbb{E}(\mathbb{E}_i f - \mathbb{E}_{i-1} f)(\mathbb{E}_j f - \mathbb{E}_{j-1} f) \\ &= \mathbb{E} \mathbb{E}_{j-1}[(\mathbb{E}_i f - \mathbb{E}_{i-1} f)(\mathbb{E}_j f - \mathbb{E}_{j-1} f)] \\ &= \mathbb{E}[(\mathbb{E}_i f - \mathbb{E}_{i-1} f) \underbrace{\mathbb{E}_{j-1}(\mathbb{E}_j f - \mathbb{E}_{j-1} f)}_{=0}], \end{aligned}$$

soit finalement

$$\mathbb{E}(f - \mathbb{E}f)^2 = \sum_{i=1}^n \mathbb{E}(\mathbb{E}_i f - \mathbb{E}_{i-1} f)^2.$$

On exploite maintenant le fait que les vecteurs  $(y_i)$  sont indépendants, chacun de loi  $\mu_i$ . On a, pour  $i \leq n-1$ ,

$$\mathbb{E}_i f = \int f(y_1, \dots, y_n) \prod_{k=i+1}^n \mu_k(dy_k),$$

soit

$$\begin{aligned} \mathbb{E}_i f - \mathbb{E}_{i-1} f &= \int \prod_{k=i+1}^n \mu_k(dy_k) \left[ f(y_1, \dots, y_n) - \int f(y_1, \dots, y_n) \mu_i(dy_i) \right], \\ &= \mathbb{E}_i(f - \mathbb{E}_{\{i\}} f). \end{aligned}$$

En utilisant l'inégalité de Jensen par rapport à l'espérance conditionnelle  $\mathbb{E}_i$ , on obtient finalement

$$\begin{aligned} \mathbb{E}(\mathbb{E}_i f - \mathbb{E}_{i-1} f)^2 &= \mathbb{E}(\mathbb{E}_i(f - \mathbb{E}_{\{i\}} f))^2 \\ &\leq \mathbb{E} \mathbb{E}_i(f - \mathbb{E}_{\{i\}} f)^2 = \mathbb{E}(f - \mathbb{E}_{\{i\}} f)^2. \end{aligned}$$

On montre d'autre part facilement que

$$\mathbb{E}(f - \mathbb{E}_{\{i\}} f)^2 = \frac{1}{2} \mathbb{E}(f - f'_i)^2. \quad (3.1)$$

En effet, comme  $\mathbb{E}_{\{i\}} f = \mathbb{E}_{\{i\}} f'_i$  et que les variables  $f$  et  $f'_i$  sont indépendantes conditionnellement aux  $y_k$ , ( $k \neq i$ ), on a

$$\begin{aligned} \mathbb{E}(f - f'_i)^2 &= \mathbb{E} \mathbb{E}_i(f - \mathbb{E}_{\{i\}} f + \mathbb{E}_{\{i\}} f'_i - f'_i)^2 \\ &= \mathbb{E}(\text{var}_{\{i\}}(f) + \text{var}_{\{i\}}(f'_i)) = 2\mathbb{E} \text{var}_{\{i\}}(f) \\ &= 2\mathbb{E}(f - \mathbb{E}_{\{i\}} f)^2, \end{aligned}$$

ce qui établit (3.1) et achève la preuve.  $\square$

**Notations**

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 notation suivantes :

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

$$Q_i(z) = \left( \sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^* - zI \right)^{-1} ,$$

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

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

On note  $\mathbb{E}_{\{i\}}$  l'espérance conditionnelle<sup>1</sup> par rapport aux vecteurs  $\{\mathbf{y}_k; k \neq i\}$  et  $\text{var}_{\{i\}}$ , la variance conditionnelle correspondante.

## 3.2 Estimation de la variance d'une forme quadratique

**Proposition 32.** *Soit  $A$  une matrice  $n \times n$  déterministe, 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*

$$\begin{aligned} \text{var}(\mathbf{x}^* A \mathbf{x}) &\leq (3m_4 + 6m_2^2) \text{Tr}(AA^*) , \\ &\leq 9m_4 \text{Tr}(AA^*) , \end{aligned}$$

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

*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

$$f - f'_i = (|x_i|^2 - |x'_i|^2) A_{ii} + (\bar{x}_i - \bar{x}'_i) \sum_{\ell \neq i} x_\ell A_{i\ell} + (x_i - x'_i) \sum_{k \neq i} \bar{x}_k A_{ki} . \quad (3.2)$$

---

1. 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.

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

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

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

**Remarque 33.** 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

$$\text{var}(\mathbf{y}^* A \mathbf{y}) = \mathcal{O}\left(\frac{\|A\|^2}{n}\right).$$

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.

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

**Corollaire 34.** 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 3.1. 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 :

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

pour obtenir que

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

$\square$

### 3.2. ESTIMATION DE LA VARIANCE D'UNE FORME QUADRATIQUE 49

**Remarque 35.** 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

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

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

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

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  $\mathbb{E} |x_i|^2 = 1$ ).

**Proposition 36.** Soit  $A$  une matrice  $n \times n$  déterministe, et  $\mathbf{x}$  un vecteur  $n \times 1$  dont les composantes  $x_i$  sont des variables aléatoires i.i.d., centrées, de huitième moment fini, alors

$$\mathbb{E} |\mathbf{x}^* \mathbf{A} \mathbf{x} - \mathbb{E}(\mathbf{x}^* \mathbf{A} \mathbf{x})|^4 \leq K (m_8 \text{Tr}(AA^* AA^*) + m_4^2 [\text{Tr}(AA^*)]^2) ,$$

où  $m_8 = \mathbb{E} |x_1|^8$ ,  $m_4 = \mathbb{E} |x_1|^4$  et  $K$  est une constante indépendante de  $n$  et de la loi de  $\mathbf{x}$ .

*Démonstration.* On utilisera les notations suivantes :

$$f = \mathbf{x}^* \mathbf{A} \mathbf{x}, \quad \varphi = |f - \mathbb{E} f|^2, \quad M_4 = \mathbb{E} |f - \mathbb{E} f|^4 .$$

Dans la suite,  $K$  désignera une constante générique, indépendante de  $n$  et de la loi de  $\mathbf{x}$ , et dont la valeur peut changer d'une ligne à l'autre. On a

$$\text{var}(\varphi) = \mathbb{E} |f - \mathbb{E} f|^4 - \{\mathbb{E} |f - \mathbb{E} f|^2\}^2$$

soit

$$M_4 = \text{var}(\varphi) + \{\mathbb{E} |f - \mathbb{E} f|^2\}^2 \leq \text{var}(\varphi) + K m_4^2 (\text{Tr } AA^*)^2 \quad (3.3)$$

en vertu de la proposition 32. Evaluons  $\text{var}(\varphi)$  à l'aide de l'inégalité d'Efron-Stein. On a

$$\begin{aligned} \varphi - \varphi'_i &= (f - \mathbb{E} f)(\bar{f} - \mathbb{E} \bar{f}) - (f'_i - \mathbb{E} f'_i)(\bar{f}'_i - \mathbb{E} \bar{f}'_i) \\ &= (f - f'_i)(\bar{f} - \mathbb{E} \bar{f}) + (f'_i - \mathbb{E} f'_i)(\bar{f} - \bar{f}'_i) \end{aligned}$$

Soit

$$\mathbb{E} |\varphi - \varphi'_i|^2 \leq K \sqrt{\mathbb{E} |f - \mathbb{E} f|^4} \sqrt{\mathbb{E} |f - f'_i|^4} = K \sqrt{M_4} \sqrt{\mathbb{E} |f - f'_i|^4} . \quad (3.4)$$

Evaluons maintenant  $\mathbb{E} |f - f'_i|^4$ . En reprenant le calcul (3.2), il vient

$$\begin{aligned} \mathbb{E} |f - f'_i|^4 &\leq K m_8 |A_{ii}|^4 + m_4 \mathbb{E} \left| \sum_{\ell \neq i} x_\ell A_{i\ell} \right|^4 + m_4 \mathbb{E} \left| \sum_{k \neq i} \bar{x}_k A_{ki} \right|^4 \\ &\leq K m_8 |A_{ii}|^4 + K m_4^2 \sum_{\ell} |A_{i\ell}|^4 + K m_4 m_2^2 \sum_{k, \ell} |A_{i\ell}|^2 |A_{ik}|^2 \\ &\leq K m_8 |A_{ii}|^4 + K m_4^2 (AA^*)_{ii}^2 \leq K \left\{ \sqrt{m_8} |A_{ii}|^2 + m_4 (AA^*)_{ii} \right\}^2 \end{aligned}$$

En réinjectant cette inégalité dans (3.4) et en sommant sur  $i$ , il vient

$$\text{var}(\varphi) \leq K\sqrt{M_4} \left\{ \sqrt{m_8} \sum_i |A_{ii}|^2 + m_4 \text{Tr}(AA^*) \right\},$$

soit finalement, en réinjectant dans (3.3),

$$M_4 \leq K\sqrt{M_4} \left\{ \sqrt{m_8} \sum_i |A_{ii}|^2 + m_4 \text{Tr}(AA^*) \right\} + Km_4^2 (\text{Tr} AA^*)^2.$$

Nécessairement, l'équation

$$X^2 - KX \left\{ \sqrt{m_8} \sum_i |A_{ii}|^2 + m_4 \text{Tr}(AA^*) \right\} - Km_4^2 (\text{Tr} AA^*)^2 = 0$$

admet une racine positive  $X^*$  et le trinôme est négatif à l'intérieur des racines, soit  $M_4 \leq (X^*)^2$ . Le calcul de  $X^*$  donne l'estimation suivante :

$$(X^*)^2 \leq Km_8 \left( \sum_i |A_{ii}|^2 \right)^2 + Km_4^2 (\text{Tr} AA^*)^2.$$

Il suffit maintenant de remarquer que

$$\begin{aligned} \left( \sum_i |A_{ii}|^2 \right)^2 &\leq \sum_k \left( \sum_i |A_{ki}|^2 \right)^2 = \sum_k (AA^*)_{kk} (AA^*)_{kk} \\ &\leq \sum_{k\ell} (AA^*)_{k\ell} \overline{(AA^*)_{k\ell}} = \text{Tr}(AA^*)^2 \end{aligned}$$

□

### 3.3 Estimation de la variance de la transformée de Stieltjes $g_n(z)$

On montre l'estimation suivante :

**Proposition 37.** *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 précédemment. On suppose que  $Nn^{-1} \rightarrow c > 0$ , alors*

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

**Remarque 38.** *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  $\lambda_i$ . À titre de comparaison, si on considère  $N$  variables  $(U_i)$  i.i.d. de variance finie, alors*

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

### 3.3. ESTIMATION DE LA VARIANCE DE LA TRANSFORMÉE DE STIELTJES $G_N(Z)$ 51

*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 le corollaire 24, on obtient

$$f = \frac{1}{N} \text{Tr} Q_i - \frac{1}{N} \frac{\mathbf{y}_i^* Q_i^2 \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i}.$$

ainsi qu'une décomposition similaire pour  $f'_i$  où  $\mathbf{y}_i$  est remplacée par  $\mathbf{y}'_i$ . Soit

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

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

$$\begin{aligned} \mathbb{E} |f'_i - f|^2 &\leq \frac{2}{N^2} \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{B^\circ}{\mathbb{E}_{\{i\}} A} - \frac{BA^\circ}{A \mathbb{E}_{\{i\}} A} \right|^2 \\ &\leq \frac{4}{N^2} \left( \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{B^\circ}{\mathbb{E}_{\{i\}} A} \right|^2 + \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{BA^\circ}{A \mathbb{E}_{\{i\}} A} \right|^2 \right). \end{aligned} \quad (3.5)$$

Les estimées suivantes

$$\frac{1}{|\mathbb{E}_{\{i\}} A|^2} \leq \frac{|z|^2}{\text{Im}(z)^2}, \quad (3.6)$$

$$\frac{|B|}{|A|} \leq \frac{1}{\text{Im}(z)}, \quad (3.7)$$

$$\mathbb{E}_{\{i\}} |A^\circ|^2 \leq \frac{K}{n \text{Im}(z)^2}, \quad (3.8)$$

$$\mathbb{E}_{\{i\}} |B^\circ|^2 \leq \frac{K}{n \text{Im}(z)^4} \quad (3.9)$$

nous permettent de conclure. En effet, en les réinjectant dans (3.5), 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 (3.6)-(3.9). Montrons (3.6). 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 (3.7). On a

$$\frac{|B|}{|A|} \leq \frac{|B|}{\text{Im}(A)} = \frac{|\mathbf{y}_i^* Q_i^2 \mathbf{y}_i|}{\text{Im}(A)} \leq \frac{\mathbf{y}_i^* Q_i^* Q_i \mathbf{y}_i}{\text{Im}(A)}. \quad (3.10)$$

### 3.4. VARIATIONS AUTOUR DE L'INÉGALITÉ DE POINCARÉ GAUSSIENNE 53

Il reste à calculer  $\text{Im}(A)$ . On introduit la décomposition spectrale de la matrice  $\sum_{k \neq i} \mathbf{y}_k \mathbf{y}_k^*$  :

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

soit

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

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 (3.10) pour obtenir (3.7). Les estimées (3.8) et (3.9) s'obtiennent en appliquant la proposition 32 (voir aussi le corollaire 34).  $\square$

On complète maintenant le corollaire 34, en donnant une estimée de la vraie variance.

**Corollaire 39.** *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 3.1. 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 34 et la proposition 37.  $\square$

## 3.4 Variations autour de l'inégalité de Poincaré gaussienne

**Théorème 40.** *Soit  $\mathbf{x}$  un vecteur gaussien standard  $\mathbf{x} \sim \mathcal{N}_n(0, I)$  et  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continûment différentiable, alors*

$$\text{var} f(\mathbf{x}) \leq \sum_{i=1}^n \mathbb{E} \left| \frac{\partial f(\mathbf{x})}{\partial x_i} \right|^2 .$$

La preuve du théorème est en annexe B.2.

[Variable complexe gaussienne et vecteur complexe gaussien à introduire qq part]

**Corollaire 41.** 1. Soit  $\mathbf{x} \sim \mathcal{N}_n(0, R)$  et  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continûment différentiable, alors

$$\text{var } f(\mathbf{x}) \leq \sum_{i,j} R_{ij} \mathbb{E} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j}.$$

En particulier, si les composantes du vecteur  $\mathbf{x}$  sont indépendantes, chacune de variance  $\text{var}(x_i) = \sigma_i^2$ , l'inégalité de Poincaré prend la forme suivante :

$$\text{var } f(\mathbf{x}) \leq \sum_i \sigma_i^2 \mathbb{E} \left| \frac{\partial f}{\partial x_i} \right|^2.$$

2. Soit  $\mathbf{x} \sim \mathcal{N}_n^{\mathbb{C}}(0, I)$  et  $\Phi : \mathbb{C}^n \rightarrow \mathbb{C}$  une fonction continûment différentiable alors

$$\text{var } \Phi(\mathbf{x}) \leq \sum_i \left\{ \mathbb{E} \left| \frac{\partial \Phi}{\partial x_i} \right|^2 + \mathbb{E} \left| \frac{\partial \Phi}{\partial \bar{x}_i} \right|^2 \right\}$$

### 3.5 Exercices

**Exercice 1** (VARIANCE D'UNE FORME QUADRATIQUE DE LA RÉSOVANTE) Soit  $X_n$  une matrice  $N \times n$  dont les entrées sont i.i.d. centrées, de quatrième moment fini. On suppose que

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

Soit  $Q(z) = (q_{ij}(z))_{ij}$  la résolvante de la matrice  $\frac{1}{n} X_n X_n^*$ ,

$$Q(z) = \left( \frac{1}{n} X_n X_n^* - z I_N \right)^{-1}$$

et  $\mathbf{a}$  un vecteur  $N \times 1$ .

1. Montrer que

$$\text{var}(\mathbf{a}^* Q(z) \mathbf{a}) = \mathcal{O}_z \left( \frac{\|\mathbf{a}\|^4}{n} \right).$$

*Indications* : on pourra s'inspirer de la preuve permettant d'estimer la variance de la trace normalisée de la résolvante. Plus précisément :

(a) En introduisant une fonction  $f$  appropriée, montrer que

$$\mathbb{E} |f - f'_i|^2 \leq 2 \mathbb{E} \mathbb{E}_{\{i\}} \left| \frac{B}{A} - \frac{\mathbb{E}_{\{i\}} B}{\mathbb{E}_{\{i\}} A} \right|^2$$

où  $A = 1 + \mathbf{y}_i^* Q_i \mathbf{y}_i$  et  $B = \mathbf{a}^* Q_i \mathbf{y}_i \mathbf{y}_i^* Q_i \mathbf{a}$ .

(b) Montrer que

$$|\mathbb{E}_{\{i\}} B| \leq K \frac{\|\mathbf{a}\|^2}{n\Im(z)^2} \quad \text{et} \quad \text{var}_{\{i\}}(B) \leq K \frac{\|\mathbf{a}\|^4}{n^2\Im(z)^4}.$$

(c) Conclure.

2. Quelle est l'ordre de grandeur de la variance d'un élément  $q_{ij}(z)$  de la résolvante ?



## Chapitre 4

# Théorème de Marčenko-Pastur

**Théorème 42.** 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,

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

**Remarque 43.** 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 44.** Dans le théorème 42, l'hypothèse optimale concernant les moments est d'avoir un deuxième moment  $\mathbb{E}|X_{ij}|^2 = \sigma^2$  fini seulement.

## 4.1 Preuve du théorème de Marčenko-Pastur

### 4.1.1 Notations et identités importantes

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

$$Q = Q_i - \frac{Q_i \mathbf{y}_i^* \mathbf{y}_i Q_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i}, \quad Q \mathbf{y}_i^* \mathbf{y}_i = \frac{Q_i \mathbf{y}_i^* \mathbf{y}_i}{1 + \mathbf{y}_i^* Q_i \mathbf{y}_i}. \quad (4.1)$$

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 11). Cela implique alors les majorations suivantes :

$$\forall i \in \{1, 2, 3\}, \quad \forall z \in \mathbb{C}^+, \quad |\alpha_i(z)| \leq \frac{1}{\text{Im}(z)}. \quad (4.2)$$

### 4.1.2 Stratégie de la preuve

On veut montrer que

$$\forall z \in \mathbb{C}^+, \quad g_n(z) \xrightarrow[N, n \rightarrow \infty]{p.s.} g_{\check{\text{M}}\text{P}}(z). \quad (4.3)$$

On en déduira alors que si  $\mathcal{D}$  est un ensemble dénombrable contenant un point d'accumulation, alors<sup>1</sup> presque sûrement,

$$\forall z \in \mathcal{D}, \quad g_n(z) \xrightarrow[N, n \rightarrow \infty]{} g_{\check{\text{M}}\text{P}}(z). \quad (4.4)$$

En effet, supposons que (4.3) soit vraie. Posons  $\mathcal{D} = \{z_k\}_{k \in \mathbb{N}}$ ; alors pour tout  $z_k$  il existe un ensemble  $\Omega_k$  de probabilité 1 tel que la convergence

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

ait lieu pour tout  $\omega \in \Omega_k$ . Il suffira alors de considérer  $\tilde{\Omega} = \bigcap_{k \in \mathbb{N}} \Omega_k$ , également de probabilité 1, pour obtenir (4.4). Cela nous assurera que presque sûrement,  $L_n \xrightarrow[N, n \rightarrow \infty]{} \mathbb{P}_{\check{\text{M}}\text{P}}$  en vertu du théorème 8.

Pour établir (4.3), on va décomposer

$$g_n(z) - g_{\check{\text{M}}\text{P}}(z) = g_n(z) - \mathbb{E} g_n(z) + \mathbb{E} g_n(z) - g_{\check{\text{M}}\text{P}}(z)$$

puis

---

1. 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  $\mathcal{D}$ .

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  $g_{\text{MP}}$  :

$$zc\sigma^2 g_{\text{MP}}^2(z) + (z + \sigma^2(c-1))g_{\text{MP}}(z) + 1 = 0 \quad (4.5)$$

3. Enfin, on montrera la stabilité de l'équation (4.5) ("équation proche  $\implies$  solution proche").

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

**Proposition 45.** 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 (4.1). 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

$$c_n z \mathbb{E} g_n(z) = -c_n + 1 - \mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right\} \quad (4.6)$$

Remarquons que  $\mathbb{E} \mathbf{y}_1^* Q_1 \mathbf{y}_1 = \frac{\sigma^2}{n} \mathbb{E} \text{Tr} Q_1$ . 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 ci-dessus, 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  $g_{\text{MP}}$ . Évaluons la différence

$$\begin{aligned} & \mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right\} - \frac{1}{1 + \sigma^2 c_n \mathbb{E} g_n} \\ &= \mathbb{E} \left\{ \frac{1}{1 + \mathbf{y}_1^* Q_1 \mathbf{y}_1} \right\} - \frac{1}{1 + \mathbb{E} (\mathbf{y}_1^* Q_1 \mathbf{y}_1)} \\ & \quad + \frac{1}{1 + \mathbb{E} (\mathbf{y}_1^* Q_1 \mathbf{y}_1)} - \frac{1}{1 + \sigma^2 c_n \mathbb{E} g_n(z)} \triangleq T_1 + T_2 . \end{aligned} \quad (4.7)$$

Commençons par le terme  $T_1$ .

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

où (a) découle des inégalités (4.2) et (b), du corollaire 39. Le terme  $T_2$  maintenant. On a

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

où (a) découle des inégalités (4.2) et de la proposition 28.

En rassemblant les estimées obtenues pour  $T_1$  et  $T_2$ , on en déduit une estimée pour (4.7) qui réinjectée dans (4.6) 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 45 est alors démontrée.  $\square$

#### 4.1.4 Stabilité de l'équation satisfaite par $g_{\check{\text{MP}}}$

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

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

**Lemme 46.** *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  $X(z) = g_{\check{\text{MP}}}(z)$ , en particulier, si  $c > 1$

$$X(z) = \frac{1}{c}h(z) - \left(1 - \frac{1}{c}\right)\frac{1}{z} \quad (4.8)$$

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  $X_\delta$  est une transformée de Stieltjes,  $|X_\delta| \leq K \operatorname{Im}^{-1}(z)$ . 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

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

Si  $c > 1$ , le même argument s'applique en remplaçant  $X + \left(1 - \frac{1}{c}\right) \frac{1}{z}$  par  $c^{-1}h$  (cf. (4.8)). Il suffit maintenant de remplacer  $\delta'$  par sa valeur  $\delta + \mathcal{O}_z(|c - c_\delta|)$  pour conclure.  $\square$

#### 4.1.5 Fin de la preuve

En combinant la proposition 45 et le lemme 46, on obtient

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

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 raisonnement expliqué en section 4.1.2 et le théorème 8 permettent alors d'obtenir la convergence de la mesure spectrale  $L_n$  vers la probabilité de Marčenko-Pastur.

## 4.2 Calcul gaussien et applications

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

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

### 4.2.1 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 47.** 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 en l'infini. 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 complexe 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.

$$\frac{\partial Q_{ij}}{\partial \bar{X}_{kl}} = -\frac{\sigma^2}{n} (Q X_N)_{i\ell} Q_{kj}, \quad (4.9)$$

$$\frac{\partial Q_{ij}}{\partial X_{kl}} = -\frac{\sigma^2}{n} Q_{ik} (X_N^* Q)_{\ell j}. \quad (4.10)$$

On notera le motif croisé en  $\gamma$  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 48** (Dérivées partielles pour la transformée de Stieltjes). Soit  $g_n(z) = \frac{1}{N} \text{Tr} Q(z)$ , alors

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

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 49** (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

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

**Exemple 50** (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.

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

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

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

Estimons le premier terme.

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

On en déduit, en prenant l'espérance et en estimant le terme restant de la même manière, que

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

## 4.2.2 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  $\delta_{ij}$  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  $[X^*]_{k\ell} = \overline{X_{\ell k}}$ .

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

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

$$\mathbb{E} Q_{ij} (\sigma^2(1 - c_n) - z) - z \frac{\sigma^2}{n} \mathbb{E} Q_{ij} \operatorname{Tr} Q = \delta_{ij} \quad (4.11)$$

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

$$z \sigma^2 c_n (\mathbb{E} g_n(z))^2 + [z + \sigma^2(c_n - 1)] \mathbb{E} g_n(z) + 1 = \mathcal{O}_z \left( \frac{1}{n^2} \right) . \quad (4.12)$$

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

### 4.2.3 Version isotrope du théorème de Marčenko-Pastur

**Théorème 52.** *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 53.** 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 52 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 [15]).

*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 51 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  $g_{\check{M}P}(z) \langle \mathbf{u}_N, \mathbf{v}_N \rangle$ . 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 :

$$\mathbb{E} |\mathbf{u}^* Q(z) \mathbf{v} - \mathbb{E} \mathbf{u}^* Q(z) \mathbf{v}|^4 = \mathcal{O}_z \left( \frac{1}{n^2} \right) . \quad (4.13)$$

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 51. 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

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

et

$$\mathbb{E} \left| \frac{\partial \phi}{\partial X_{k\ell}} \right|^2 \leq 2 \mathbb{E} \left| (h - \mathbb{E} h) \frac{\partial \bar{h}}{\partial X_{k\ell}} \right|^2 .$$

Soit

$$\sum_{k,\ell} \mathbb{E} \left| \frac{\partial \phi}{\partial X_{k\ell}} \right|^2 \leq 2 \mathbb{E} \left\{ |h - \mathbb{E} h|^2 \sum_{k,\ell} \left| \frac{\partial \bar{h}}{\partial X_{k\ell}} \right|^2 \right\}$$

La somme du membre de droite a été estimée à l'exemple 51 et peut être majorée par une quantité déterministe en  $\mathcal{O}_z(\|\mathbf{u}\|^2 \|\mathbf{v}\|^2 n^{-1})$ . Il reste alors la variance de  $h$  qui est du même ordre, soit finalement l'estimée :

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

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 (4.13) 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 (4.11) 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 50 et 51 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

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

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 en faisant appel à l'équation (4.12), ce qui nous donne

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

Il reste à invoquer la stabilité de l'équation de Marčenko-Pastur pour remplacer  $\mathbb{E} g_n(z)$  par  $g_{\text{MP}}(z)$ .  $\square$

#### 4.2.4 Confinement des valeurs propres

En combinant le lemme de stabilité pour l'équation de Marčenko-Pastur (lemme 46) et l'estimée (4.12) obtenue pour les entrées gaussiennes, on obtient l'estimée suivante :

$$\mathbb{E} g_n(z) - g_{\text{MP}}^{c_n}(z) = \mathcal{O}_z \left( \frac{1}{n^2} \right). \quad (4.14)$$

L'identité d'Helffer-Sjöstrand permet de transférer cette estimée à une fonction régulière à support compact.

**Proposition 54.** *Soit  $f$  une fonction  $C^\infty$  à support compact, alors*

$$\mathbb{E} \frac{1}{N} \text{Tr} f \left( \frac{\sigma^2}{n} X X^* \right) - \int f(x) \mathbb{P}_{\check{M}P}(dx) = \mathcal{O} \left( \frac{1}{n^2} \right).$$

*Démonstration.* D'après l'estimation (6.5), il existe  $p, q$  entiers tels que

$$|\mathbb{E} g_n(z) - g_{\check{M}P}(z)| \leq \frac{K}{n^2} \frac{|z|^p}{\text{Im}(z)^q}.$$

En utilisant l'identité d'Helffer-Sjöstrand avec  $\Phi_q(f)$ , il vient

$$\mathbb{E} \frac{1}{N} \text{Tr} f \left( \frac{\sigma^2}{n} X X^* \right) - \int f(x) \mathbb{P}_{\check{M}P}(dx) = \frac{2}{\pi} \text{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_q(f) \{ \mathbb{E} g_n(z) - g_{\check{M}P}(z) \} dx dy$$

Comme  $\bar{\partial} \Phi_q(f)(z) = \frac{(iy)^q}{q!} f^{(q+1)}(x)$  près de l'axe réel ( $y \leq \eta$ ), on obtient alors

$$\begin{aligned} & \left| \mathbb{E} \frac{1}{N} \text{Tr} f \left( \frac{\sigma^2}{n} X X^* \right) - \int f(x) \mathbb{P}_{\check{M}P}(dx) \right| \\ & \leq \frac{2}{\pi} \text{Re} \int_{\mathbb{C}^+} \left| \bar{\partial} \Phi_q(f) \frac{K}{n^2} \frac{|z|^p}{\text{Im}(z)^q} \right| dx dy \\ & \leq \frac{K'}{n^2}, \end{aligned}$$

ce qui achève la démonstration.  $\square$

On notera à l'occasion  $A = \frac{\sigma^2}{n} XX^*$ .

**Proposition 55.** *Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction  $C^\infty$  à support compact, alors*

$$\text{var} \left( \frac{1}{N} \text{Tr} f \left( \frac{\sigma^2}{n} XX^* \right) \right) \leq \frac{2\sigma^4}{N^2 n^2} \mathbb{E} \text{Tr} (f'(A) XX^* f'(A)) .$$

*Démonstration.* Il suffit d'appliquer l'inégalité de Poincaré et la proposition 87 :

$$\begin{aligned} \text{var} \left( \frac{1}{N} \text{Tr} f(A) \right) &\leq \sum_{k\ell} \mathbb{E} \left| \frac{\partial}{\partial X_{k\ell}} \left( \frac{1}{N} \text{Tr} f(A) \right) \right|^2 + \sum_{k\ell} \mathbb{E} \left| \frac{\partial}{\partial X_{k\ell}} \left( \frac{1}{N} \text{Tr} f(A) \right) \right|^2 \\ &\leq \frac{\sigma^4}{n^2 N^2} \sum_{k\ell} \mathbb{E} |[f'(A)X]_{k\ell}|^2 + \frac{\sigma^4}{n^2 N^2} \sum_{k\ell} \mathbb{E} |[X^* f'(A)]_{k\ell}|^2 \\ &\leq \frac{2\sigma^4}{N^2 n^2} \mathbb{E} \text{Tr} (f'(A) XX^* f'(A)) . \end{aligned}$$

□

**Théorème 56.** *Supposons que  $\frac{N}{n} \rightarrow c \in (0, 1]$ . Soit  $\mathcal{S}_n$  le spectre empirique des valeurs propres de  $A = \frac{\sigma^2}{n} XX^*$  et  $\mathcal{S}_\infty$  le support de la loi de Marčenko-Pastur de paramètre  $c$  :*

$$\mathcal{S}_\infty^c = [\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2] .$$

Alors pour tout  $\varepsilon > 0$ , presque sûrement

$$\mathcal{S}_n \subset \mathcal{S}_\infty + (-\varepsilon, \varepsilon)$$

pour  $n$  suffisamment grand.

Autrement dit, presque sûrement, les valeurs propres empirique sont confinées près du spectre limite.

*Démonstration.* Soit  $\varepsilon > 0$  fixé et  $\varphi_n$  une fonction  $C^\infty$ , à valeurs dans  $[0, 1]$ , valant 1 sur  $\mathcal{S}_\infty^{c_n}$  et zéro sur le complémentaire de  $\mathcal{S}_\infty^{c_n} + (-\frac{\varepsilon}{2}, \frac{\varepsilon}{2})$ . On pose

$$\psi_n = 1 - \varphi_n$$

et on va démontrer que

$$\text{Tr} \psi_n \left( \frac{\sigma^2}{n} XX^* \right) \xrightarrow[N, n \rightarrow \infty]{p.s.} 0 . \quad (4.15)$$

Cela entraînera que presque sûrement, pour  $n$  assez grand, les valeurs propres  $\lambda_i$  de  $A$  satisfont :

$$\lambda_i \in \mathcal{S}_\infty^{c_n} + \left( -\frac{\varepsilon}{2}, \frac{\varepsilon}{2} \right) ,$$

autrement, on aurait  $\text{Tr} \psi_n \left( \frac{\sigma^2}{n} XX^* \right) = \sum_i \psi_n(\lambda_i) \geq 1$ .

On commence par montrer que

$$\mathbb{E} \operatorname{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right) = O \left( \frac{1}{N} \right). \quad (4.16)$$

En utilisant la proposition 88 appliquée à la fonction à support compact  $\varphi_n$ , il vient

$$\begin{aligned} \mathbb{E} \operatorname{Tr} \varphi_n \left( \frac{\sigma^2}{n} X X^* \right) &= N \int \varphi_n(x) \mathbb{P}_{\text{MP}}^{c_n}(dx) + O \left( \frac{1}{N} \right) \\ &= N + O \left( \frac{1}{N} \right). \end{aligned}$$

Soit finalement

$$\mathbb{E} \operatorname{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right) = N - \mathbb{E} \operatorname{Tr} \varphi_n \left( \frac{\sigma^2}{n} X X^* \right) = O \left( \frac{1}{N} \right),$$

ce qui établit (6.7). Évaluons maintenant la variance de  $\operatorname{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right)$  à l'aide de la proposition 89. On a

$$\begin{aligned} \operatorname{var} \left( \operatorname{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right) \right) &= \operatorname{var} \left( \operatorname{Tr} \varphi_n \left( \frac{\sigma^2}{n} X X^* \right) \right) \\ &\leq \frac{2\sigma^4}{n^2} \mathbb{E} \operatorname{Tr} \varphi_n'(A) X X^* \varphi_n'(A) \\ &= \frac{2\sigma^2}{n} \mathbb{E} \operatorname{Tr} \varphi_n'(A) A \varphi_n'(A) \\ &= \frac{2\sigma^2}{n} \mathbb{E} \operatorname{Tr} \xi_n(A) \end{aligned}$$

avec  $\xi_n(x) = [\varphi_n'(x)]^2 x$ , fonction  $C^\infty$  à support compact et valant 0 sur  $\mathcal{S}_\infty^{c_n}$ . Il suffit à nouveau d'appliquer la proposition 88 pour conclure :

$$\frac{1}{N} \mathbb{E} \operatorname{Tr} \xi_n(A) = \int \xi_n(x) \mathbb{P}_{\text{MP}}^{c_n}(dx) + O \left( \frac{1}{n^2} \right) = O \left( \frac{1}{n^2} \right).$$

Soit finalement

$$\operatorname{var} \left( \operatorname{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right) \right) = O \left( \frac{1}{n^2} \right).$$

En appliquant maintenant le lemme de Borel-Cantelli et (6.7), on obtient (6.6), ce qui conclut la preuve du théorème.  $\square$

[Rajouter la convergence de la plus grande valeur propre]

## 4.3 Exercices

[Modèle de Wigner : preuve gaussienne et confinement des valeurs propres]

[Éventuellement, thm de MP isotrope pour des entrées non gaussiennes]



# Chapitre 5

## Le modèle de grandes matrices de covariance

### 5.1 La matrice de covariance empirique

Soit  $X_n = (X_{ij})$  une matrice  $N \times n$ , à entrées i.i.d. centrées et de variance unité et  $R_N$  une matrice  $N \times N$  déterministe, hermitienne et semi-définie positive. En notant  $(\mathbf{x}_i)$  les vecteurs colonne de la matrice  $X_n$ , la matrice  $Y_n$  de dimensions  $N \times n$

$$Y_n = [R_N^{1/2} \mathbf{x}_1, \dots, R_N^{1/2} \mathbf{x}_n]$$

représente un échantillon de  $n$  observations indépendantes  $(\mathbf{y}_1, \dots, \mathbf{y}_n)$ , centrées, à valeurs dans  $\mathbb{R}^N$  ou  $\mathbb{C}^N$  et chacune de matrice de covariance

$$\mathbb{E} \mathbf{y}_1 \mathbf{y}_1^* = R_N^{1/2} \mathbb{E} \mathbf{x}_1 \mathbf{x}_1^* R_N^{1/2} = R_N .$$

On appelle  $R_N$  la **matrice de covariance de la population**  $(\mathbf{y}_1, \dots, \mathbf{y}_n)$ .

La matrice

$$\hat{R}_N = \frac{1}{n} R_N^{1/2} X_n X_n^* R_N^{1/2} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^*$$

est la **matrice de covariance empirique**.

Une question importante en statistiques est d'estimer la matrice de covariance de la population  $R_N$  à partir de la matrice de covariance empirique

$$\hat{R}_N = \frac{1}{n} Y_n Y_n^* .$$

Dans le cas classique de données de petite dimension  $N$  et d'un échantillon de grande dimension  $n$ , la convergence de la mesure spectrale de  $\hat{R}_n$  vers celle de  $R_N$  est immédiate :

**Proposition 57.** Soit  $L_n$  la mesure spectrale de  $\hat{R}_n$  et  $L_N^R$  celle de la matrice  $R_N$ . Alors, presque sûrement

$$L_n \xrightarrow[N \text{ fixé, } n \rightarrow \infty]{\text{etr}} L_N^R .$$

*Démonstration.* La loi des grands nombres entraîne que

$$\hat{R}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^* \xrightarrow[N \text{ fixé, } n \rightarrow \infty]{p.s.} \mathbb{E} \mathbf{y}_1 \mathbf{y}_1^* = R_N ,$$

la convergence ayant lieu pour n'importe quelle norme de  $\mathbb{C}^{N \times N}$ . D'autre part, l'application de l'ensemble des matrices hermitiennes de dimension  $N$  dans lui-même

$$A \mapsto Q(A) = (A - zI_N)^{-1}$$

est continue comme simple conséquence de l'identité de la résolvante associée à la majoration de la norme spectrale d'une résolvante :

$$\|Q(B) - Q(A)\| = \|-Q(B)(B - A)Q(A)\| \leq \frac{1}{\text{Im}(z)^2} \|B - A\| .$$

Par suite,

$$Q(\hat{R}_N) \xrightarrow[N \text{ fixé, } n \rightarrow \infty]{p.s.} Q(R_N) ,$$

pour  $z \in \mathbb{C}^+$ , ce qui entraîne la convergence des transformées de Stieltjes associées et donc des mesures spectrales associées.  $\square$

## 5.2 Comportement de la matrice de covariance empirique en grande dimension

On suppose maintenant que les entrées  $(X_{ij})$  de la matrice  $X_n$  admettent en plus un quatrième moment fini :

$$\mathbb{E} |X_{ij}|^4 < \infty .$$

Soit  $(R_N)$  une famille de matrices hermitiennes, semi-définies positives de norme spectrale uniformément bornée, i.e.

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

On considère le régime asymptotique dans lequel  $N = N(n)$ ,  $n \rightarrow \infty$  et

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

On notera ce régime  $N \sim n$  et on utilisera le ratio  $c_n = \frac{N}{n}$  dans la suite.

**Théorème 58.** *On considère les hypothèses énoncées précédemment : quatrième moment fini pour les entrées  $(X_{ij})$ , norme spectrale uniformément bornée pour la famille  $(R_N)$  et  $N \sim n$ . Soit  $L_N$  la mesure spectrale de la matrice de covariance empirique  $\hat{R}_n$  et  $g_n$  sa transformée de Stieltjes. Alors*

(a) l'équation

$$t_n(z) = \frac{1}{N} \text{Tr} [-z (I_N + c_n R_N t_n(z)) + (1 - c_n) R_N]^{-1}, \quad z \in \mathbb{C}^+ \quad (5.1)$$

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  $g_n(z) - t_n(z) \xrightarrow[N \sim n]{p.s} 0$ .

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

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

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

**Remarque 59.** L'équation (5.1) peut se réécrire à l'aide de la mesure spectrale de  $R_N$ . Considérons la décomposition spectrale de la matrice  $R_N = U \Lambda_N U^*$ ; on notera  $(\rho_i)$  les valeurs propres de la matrice  $R_N$ , soit  $\Lambda_N = \text{diag}(\rho_i)$ . Alors

$$\begin{aligned} t_n(z) &= \frac{1}{N} \text{Tr} [-z (I_N + c_n R_N t_n(z)) + (1 - c_n) R_N]^{-1} \\ &= \frac{1}{N} \text{Tr} U [-z (I_N + c_n \Lambda_N t_n(z)) + (1 - c_n) \Lambda_N]^{-1} U^* \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + c_n \rho_i t_n(z)) + (1 - c_n) \rho_i} \\ &= \int \frac{L_N^R(du)}{-z(1 + u c_n t_n(z)) + (1 - c_n) u}. \end{aligned}$$

**Remarque 60** (Le cas Marčenko-Pastur). Considérons le cas où  $R_N = \sigma^2 I_N$ , alors l'équation précédente prend la forme

$$t_n(z) = \frac{1}{-z(1 + c_n \sigma^2 t_n(z)) + (1 - c_n) \sigma^2},$$

soit

$$z c_n \sigma^2 t_n^2(z) + (z + \sigma^2(c_n - 1)) t_n(z) + 1 = 0.$$

On reconnaît l'équation de Marčenko-Pastur et dans ce cas, la probabilité  $\mathcal{F}_n$  associée à  $t_n(z)$  est la distribution de Marčenko-Pastur associée au paramètre  $c_n$ .

**Remarque 61** (Le cas limite où  $N$  est fixé et  $n \rightarrow \infty$ ). Dans ce cas, l'hypothèse  $N \sim n$  n'est plus vérifiée et  $c_n$  tend vers zéro. En remplaçant brutalement  $c_n$  par zéro dans l'équation (5.1), on obtient l'équation

$$t_n(z) = \frac{1}{N} \text{Tr} (R_N - zI_N)^{-1} .$$

La transformée de Stieltjes  $t_n(z)$  est dans ce cas associée à la distribution spectrale de  $R_N$  et on retrouve (formellement) le résultat de la proposition 57.

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

**Théorème 62.** Sous les hypothèses du théorème précédent, soit  $L_N$  la mesure spectrale de la matrice  $\hat{R}_n$  et  $g_n$  sa transformée de Stieltjes. Soit  $L_N^R$  la mesure spectrale de la matrice  $R_N$ . Si

$$c_n \rightarrow c > 0 \quad \text{et} \quad L_n^R \xrightarrow[N, n \rightarrow \infty]{\text{etr}} L_\infty^R ,$$

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,

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

On pourra se référer à l'article historique [18], à l'article de référence [22] (voir aussi [3]) et aussi à l'article [14] pour les notions d'équivalents déterministes dans des modèles plus compliqués.

### 5.2.1 Méthode heuristique de preuve pour le théorème 58

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

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

Les ingrédients principaux sont les éléments suivants :

- ▷ Expression de l'élément diagonal de la résolvante (cf. proposition 27).
- ▷ 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

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

- ▷ 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 32, cf. aussi la remarque 33.

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

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

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 :

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

### 5.2.2 Exemples

#### Cas où $R_N$ admet un nombre fini de valeurs propres

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

$$L_N^R = \frac{1}{N} \sum_{\ell=1}^K n_\ell \delta_{\rho_\ell}$$

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,

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

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

$$t(z) = \sum_{\ell=1}^K \alpha_\ell \frac{1}{-z(1 + c \rho_\ell t(z)) + (1 - c) \rho_\ell} . \quad (5.2)$$

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 63.** *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 [23] :*

1. Résoudre numériquement l'équation (5.2) pour  $z = x \in \mathbb{R}^*$ ,
2. S'il n'y a que des racines réelles, alors la densité  $f(x)$  vaut zéro.
3. S'il y a des racines complexes, choisir l'unique solution  $t(x)$  vérifiant  $\text{Im} t(x) > 0$ . La densité au point  $x$  est alors 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  $\theta_\ell$  sont positifs et ne dépendent pas de  $N, n$  et les vecteurs  $(\mathbf{u}_\ell)$  sont orthonormés.

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

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

La convergence suivante est alors 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.

## 5.3 Démonstration du théorème 58

Nous allons dans un premier temps démontrer le théorème pour  $R_N$  diagonale en section 5.3.1, puis en déduire le théorème pour  $R_N$  quelconque mais dans le cas d'entrées  $(X_{ij})$  gaussiennes en section 5.3.2, et enfin dans le cas général en section ??.

### 5.3.1 Démonstration dans le cas où $R_N$ est diagonale

On pose  $R_N = \text{diag}(\rho_1, \dots, \rho_N)$ . On rappelle que  $g_n(z) = \frac{1}{N} \text{Tr} (\Sigma_n \Sigma_n^T - z I_N)^{-1}$  et on introduit

$$\tilde{g}_n(z) = \frac{1}{n} \text{Tr} (\Sigma_n^T \Sigma_n - z I_n)^{-1} .$$

On va procéder en 3 étapes :

1. établir les équations approchées satisfaites par  $\mathbb{E} g_n(z)$  et  $\mathbb{E} \tilde{g}_n(z)$
2. étudier les équations exactes satisfaites par  $t_n$  et son homologue  $\tilde{t}_n$  (que l'on introduira bientôt)
3. étudier la stabilité de l'équation exacte satisfaite par  $t_n$ .

Ces 3 étapes correspondent aux 3 propositions suivantes.

**Proposition 64.** *Sous les hypothèses du théorème 58 et dans le cas où  $R_N$  est diagonale, i.e.  $R_N = \text{diag}(\rho_i, 1 \leq i \leq N)$ ,*

$$\mathbb{E} g_n(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + c_n \rho_i \mathbb{E} g_n(z)) + (1 - c_n) \rho_i} + \mathcal{O}_z \left( \frac{1}{\sqrt{n}} \right), \quad (5.3)$$

$$\mathbb{E} \tilde{g}_n(z) = \left[ -z + \frac{1}{n} \sum_{i=1}^N \frac{\rho_i}{1 + \rho_i \mathbb{E} \tilde{g}_n(z)} \right]^{-1} + \mathcal{O}_z \left( \frac{1}{\sqrt{n}} \right), \quad (5.4)$$

$$\mathbb{E} \tilde{g}_n(z) = c_n \mathbb{E} g_n(z) + (1 - c_n) \left( -\frac{1}{z} \right). \quad (5.5)$$

*Démonstration.* L'équation (5.5) est trivialement satisfaite et nous assure en particulier que

$$z \mapsto \frac{1}{-z(1 + c_n \rho_i \mathbb{E} g_n(z)) + (1 - c_n) \rho_i} = \frac{1}{-z(1 + \rho_i \mathbb{E} \tilde{g}_n(z))}$$

est la transformée de Stieltjes d'une probabilité, ce qui nous permettra de majorer sa valeur absolue par  $1/\text{Im}(z)$  pour  $z \in \mathbb{C}^+$ .

Commençons par démontrer (5.3). On écrit

$$\begin{aligned} \mathbb{E} g_n(z) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + c_n \rho_i \mathbb{E} g_n(z)) + (1 - c_n) \rho_i} \\ &= \mathbb{E} g_n(z) - \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + c_n \rho_i \mathbb{E} \tilde{g}_n(z))} \\ &\triangleq T_1 + T_2 + T_3 \end{aligned}$$

où

$$\begin{aligned} T_1 &= \mathbb{E} g_n(z) - \frac{1}{N} \sum_{i=1}^N \mathbb{E} \frac{1}{-z \left( 1 + \frac{\rho_i}{n} \text{Tr} \left( \Sigma_{(i)}^T \Sigma_{(i)} - z I_n \right)^{-1} \right)} \\ T_2 &= \frac{1}{N} \sum_{i=1}^N \mathbb{E} \frac{1}{-z \left( 1 + \frac{\rho_i}{n} \text{Tr} \left( \Sigma_{(i)}^T \Sigma_{(i)} - z I_n \right)^{-1} \right)} \\ &\quad - \frac{1}{N} \sum_{i=1}^N \mathbb{E} \frac{1}{-z \left( 1 + \frac{\rho_i}{n} \text{Tr} \left( \Sigma^T \Sigma - z I_n \right)^{-1} \right)} \\ T_3 &= \frac{1}{N} \sum_{i=1}^N \mathbb{E} \frac{1}{-z \left( 1 + \frac{\rho_i}{n} \text{Tr} \left( \Sigma^T \Sigma - z I_n \right)^{-1} \right)} - \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + \rho_i \mathbb{E} \tilde{g}_n(z))} \end{aligned}$$

Le terme  $T_1$  se gère en exprimant  $g_n$  comme moyenne des éléments diagonaux de la résolvante, puis en utilisant l'expression exacte d'un élément diagonal de la résolvante :

$$g_n(z) = \frac{1}{N} \sum_{i=1}^N q_{ii}(z) = -\frac{1}{N} \sum_{i=1}^N \frac{1}{z \left( 1 + \Sigma_{i \cdot} \left( \Sigma_{(i)}^T \Sigma_{(i)} - zI_n \right)^{-1} \Sigma_{i \cdot}^T \right)}$$

où  $\Sigma_{i \cdot}$  représente la  $i$ ème ligne de la matrice  $\Sigma$  et  $\Sigma_{(i)}$  représente la matrice  $\Sigma$  privée de sa  $i$ ème ligne. On notera cette fois-ci  $\mathbb{E}_{\{i\}}$  l'espérance conditionnelle par rapport aux vecteurs-ligne de  $\Sigma_n$  autres que  $\Sigma_{i \cdot}$ , i.e.  $\mathbb{E}_{\{i\}} = \mathbb{E}(\cdot \mid \Sigma_{k \cdot}, k \neq i)$ ;  $\text{var}_{\{i\}}$  représentera la variance conditionnelle. Par suite,

$$\begin{aligned} |T_1| &\leq \frac{|z|}{N} \sum_{i=1}^N \mathbb{E} \left\{ \frac{|q_{ii}(z)|}{\left| -z \left( 1 + \frac{\rho_i}{n} \text{Tr} \left( \Sigma_{(i)}^T \Sigma_{(i)} - zI_n \right)^{-1} \right) \right|} \right. \\ &\quad \left. \times \left| \Sigma_{i \cdot} \left( \Sigma_{(i)}^T \Sigma_{(i)} - zI_n \right)^{-1} \Sigma_{i \cdot}^T - \frac{\rho_i}{n} \text{Tr} \left( \Sigma_{(i)}^T \Sigma_{(i)} - zI_n \right)^{-1} \right| \right\} \\ &\leq \frac{|z|}{N \text{Im}^2(z)} \sum_{i=1}^N \mathbb{E} \left[ \text{var}_{\{i\}} \left( \Sigma_{i \cdot} \left( \Sigma_{(i)}^T \Sigma_{(i)} - zI_n \right)^{-1} \Sigma_{i \cdot}^T \right) \right]^{1/2} \\ &= \mathcal{O}_z \left( \frac{1}{\sqrt{n}} \right). \end{aligned}$$

L'estimée finale provenant de l'estimée de la variance (ici conditionnelle mais ça ne change rien) d'une forme quadratique.

Des arguments similaires permettent d'obtenir les estimées suivantes

$$\begin{aligned} |T_2| &\leq \frac{\mathbf{R}|z|}{N \text{Im}^2(z)} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left| \text{Tr} \left( \Sigma^T \Sigma - zI_n \right)^{-1} - \text{Tr} \left( \Sigma_{(i)}^T \Sigma_{(i)} - zI_n \right)^{-1} \right| \\ &\leq \frac{\mathbf{R}|z|}{N \text{Im}^3(z)} = \mathcal{O}_z \left( \frac{1}{n} \right), \end{aligned}$$

l'estimée finale provenant de l'estimée de différence de traces de résolvantes de matrices perturbées.

Enfin, le terme  $T_3$  se gère de manière proche :

$$|T_3| \leq \frac{\mathbf{R}}{\text{Im}^2(z)} \mathbb{E} |\tilde{g}_n(z) - \mathbb{E} \tilde{g}_n(z)| \leq \frac{\mathbf{R}}{\text{Im}^2(z)} (\text{var} \tilde{g}_n(z))^{1/2} = \mathcal{O}_z \left( \frac{1}{n} \right).$$

Ceci achève la démonstration de (5.3), l'équation (5.4) se démontre de manière identique.  $\square$

**Proposition 65.** *Le système des trois équations suivantes, d'inconnues  $t_n$  et  $\tilde{t}_n$ , admet un unique couple de solutions  $(t_n, \tilde{t}_n)$ , chaque fonction  $t_n$  et  $\tilde{t}_n$  étant*

la transformée de Stieltjes d'une probabilité à support inclus dans  $\mathbb{R}^+$  :

$$t_n(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + c_n \rho_i t_n(z)) + (1 - c_n) \rho_i}, \quad (5.6)$$

$$\tilde{t}_n(z) = \left[ -z + \frac{1}{n} \sum_{i=1}^N \frac{\rho_i}{1 + \rho_i \tilde{t}_n(z)} \right]^{-1}, \quad (5.7)$$

$$\tilde{t}_n(z) = c_n t_n(z) + (1 - c_n) \left( -\frac{1}{z} \right). \quad (5.8)$$

*Démonstration.* On va dans un premier temps démontrer l'existence de telles solutions. On montrera ensuite l'unicité. Considérons le **méta-modèle** suivant :

$$R_N^M = \begin{pmatrix} R_N & 0 & \cdots \\ & \ddots & \\ \cdots & 0 & R_N \end{pmatrix}$$

où  $R_N^M$  est une matrice  $NM \times NM$ , diagonale, chaque valeur propre  $\rho_i$  de  $R_N$  ayant maintenant une multiplicité  $M$  en tant que valeur propre de  $R_N^M$ . On introduit les grandeurs associées à  $R_N^M$  :

- $X_n^M$  est une matrice  $NM \times nM$  à entrées i.i.d.,
- $\Sigma_n^M = \frac{1}{\sqrt{nM}} (R_N^M)^{1/2} X_n^M$ ,
- $g_n^M(z) = \frac{1}{NM} \text{Tr} \left( \Sigma_n^M (\Sigma_n^M)^T - z I_{NM} \right)^{-1}$ ,
- $\tilde{g}_n^M(z) = \frac{1}{nM} \text{Tr} \left( (\Sigma_n^M)^T \Sigma_n^M - z I_{nM} \right)^{-1}$ ,
- $c_n^M = \frac{NM}{nM} = \frac{N}{n} = c_n$ .

Fixons les dimensions  $N, n$  et faisons tendre le 3ème paramètre  $M$  vers l'infini. Les résultats de la proposition 64 demeurent valables et étant donné le caractère bloc diagonal de la matrice  $R_N^M$ , les équations s'écrivent :

$$\mathbb{E} g_n^M(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + c_n \rho_i \mathbb{E} g_n^M(z)) + (1 - c_n) \rho_i} + \mathcal{O}_z \left( \frac{1}{\sqrt{nM}} \right) \quad (5.9)$$

$$\mathbb{E} \tilde{g}_n^M(z) = \left[ -z + \frac{1}{n} \sum_{i=1}^N \frac{\rho_i}{1 + \rho_i \mathbb{E} \tilde{g}_n^M(z)} \right]^{-1} + \mathcal{O}_z \left( \frac{1}{\sqrt{nM}} \right), \quad (5.10)$$

$$\mathbb{E} \tilde{g}_n^M(z) = c_n \mathbb{E} g_n^M(z) + (1 - c_n) \left( -\frac{1}{z} \right). \quad (5.11)$$

En tant que transformées de Stieltjes, les suites  $(\mathbb{E} g_n^M)_{M \geq 1}$  et  $(\mathbb{E} \tilde{g}_n^M)_{M \geq 1}$  sont des familles normales sur  $\mathbb{C}^+$  dont on peut extraire deux sous-suites (que l'on

notera encore par abus de notation  $\mathbb{E} g_n^M$  et  $\mathbb{E} \tilde{g}_n^M$ ) qui convergent uniformément sur tout compact de  $\mathbb{C}^+$  vers des limites analytiques sur  $\mathbb{C}^+$   $t_n^*$  et  $\tilde{t}_n^*$ . La convergence étant en particulier ponctuelle sur  $\mathbb{C}^+$ , on en déduit (par propriété de convergence ponctuelle de transformées de Stieltjes) que  $t_n^*$  et  $\tilde{t}_n^*$  sont les transformées de Stieltjes de sous-probabilités (=mesures positives de masse totale inférieure ou égale à 1) de support inclus dans  $\mathbb{R}^+$ , cette dernière propriété découlant de

$$zt_n^*(z) = \lim_{M \rightarrow \infty} z \mathbb{E} g_n^M(z) \geq 0 \quad \text{pour } z \in \mathbb{C}^+$$

et du résultat analogue pour  $\tilde{t}_n^*$ . D'autre part, l'équation (5.11) entraîne immédiatement que

$$\tilde{t}_n^*(z) = c_n t_n^*(z) + (1 - c_n) \left( -\frac{1}{z} \right). \quad (5.12)$$

Réécrivons l'équation (5.9) sous la forme

$$\mathbb{E} g_n^M(z) = \frac{1}{N} \sum_{i=1}^N -\frac{1}{z(1 + \rho_i \mathbb{E} \tilde{g}_n^M(z))} + \mathcal{O}_z \left( \frac{1}{\sqrt{nM}} \right). \quad (5.13)$$

Si  $z \in \mathbb{C}^+$ , alors  $\tilde{t}_n^*(z) \in \mathbb{C}^+$  en tant que transformée de Stieltjes et  $\mathbb{E} \tilde{g}_n^M(z) \xrightarrow{M \rightarrow \infty} \tilde{t}_n^*(z)$ . Comme l'application de  $\mathbb{C}^+$  dans  $\mathbb{C}^+$  :

$$Z \mapsto \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + \rho_i Z)}$$

y est continue, le passage à la limite dans (5.13) entraîne que

$$t_n^*(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + \rho_i \tilde{t}_n^*(z))} = \frac{1}{N} \sum_{i=1}^N \frac{1}{-z(1 + c_n \rho_i \tilde{t}_n^*(z)) + (1 - c_n) \rho_i}$$

pour  $z \in \mathbb{C}^+$ . On démontre de même que  $\tilde{t}_n^*$  satisfait l'équation (5.7). On a ainsi démontré l'existence de solutions  $t_n^*$  et  $\tilde{t}_n^*$  qui satisfont les 3 équations de la proposition.

Reste à démontrer que  $t_n^*$  et  $\tilde{t}_n^*$  sont les transformées de Stieltjes de probabilités et qu'elles sont uniques.

La fonction  $\tilde{t}_n^*$  étant la transformée de Stieltjes d'une sous probabilité, on en déduit l'existence de  $\kappa \in (0, 1]$  tel que pour  $z \in \mathbb{C}^+$ ,

$$|\tilde{t}_n^*(z)| \leq \frac{\kappa}{\text{Im}(z)} \quad \Rightarrow \quad \lim_{y \rightarrow \infty} \tilde{t}_n^*(iy) = 0.$$

Passons maintenant à la limite en  $y \rightarrow \infty$  dans l'équation

$$iy t_n^*(iy) = \frac{1}{N} \sum_{i=1}^N -\frac{1}{1 + \rho_i \tilde{t}_n^*(iy)} \quad \Rightarrow \quad \lim_{y \rightarrow \infty} iy t_n^*(iy) = -1.$$

On en déduit que  $t_n^*$  est la transformée de Stieltjes d'une probabilité ; idem pour  $\tilde{t}_n^*$  par le même type d'arguments ou l'utilisation de (5.12).

Démontrons maintenant l'unicité et supposons qu'il existe deux couples  $(t, \tilde{t})$  et  $(\tau, \tilde{\tau})$  qui sont des transformées de Stieltjes et satisfaisant le système de la proposition. En particulier,  $t$  et  $\tau$  satisfont l'équation (5.6). En soustrayant et en majorant, il vient :

$$|t - \tau| \leq \frac{\mathbf{R}|z|c_n}{\text{Im}^2(z)} |t - \tau| .$$

En se plaçant dans un domaine de  $\mathbb{C}^+$  contenant un point d'accumulation où  $\frac{\mathbf{R}|z|}{\text{Im}^2(z)} < 1$ , on en déduit que  $t$  et  $\tau$  y coïncident mais étant analytiques sur  $\mathbb{C}^+$ , elles y sont donc égales. Cela entraîne immédiatement le fait que  $\tilde{t} = \tilde{\tau}$ .  $\square$

**Proposition 66.** *Pour tout  $z \in \mathbb{C}^+$ , on a :*

$$\mathbb{E} g_n(z) - t_n(z) \xrightarrow{N, n \rightarrow \infty} 0 \quad \text{et} \quad \mathbb{E} \tilde{g}_n(z) - \tilde{t}_n(z) \xrightarrow{N, n \rightarrow \infty} 0 .$$

*Démonstration.* On démontre la première convergence. La famille  $(\mathbb{E} g_n - t_n)$  étant normale sur  $\mathbb{C}^+$ , on en extrait une sous-suite convergente vers  $f_\infty$ , analytique sur  $\mathbb{C}^+$ . En soustrayant les équations (5.3) et (5.6), il vient :

$$|\mathbb{E} g_n(z) - t_n(z)| \leq \frac{\mathbf{R}|z|c_n}{\text{Im}^2(z)} |\mathbb{E} g_n(z) - t_n(z)| + \mathcal{O}_z \left( \frac{1}{\sqrt{n}} \right) .$$

On se place maintenant dans un domaine de  $\mathbb{C}^+$  contenant un point d'accumulation tel que  $\sup_n \frac{\mathbf{R}|z|c_n}{\text{Im}^2(z)} < 1$ . Sur un tel domaine  $\mathcal{D}$ , on a  $\mathbb{E} g_n(z) - t_n(z) = \mathcal{O}_z(1/\sqrt{n})$ . En particulier,  $\mathbb{E} g_n(z) - t_n(z)$  tend vers zéro pour tout  $z \in \mathcal{D}$ .

Cela implique en particulier que la limite  $f_\infty$  analytique et coïncidant avec 0 sur  $\mathcal{D}$  y est nulle sur  $\mathbb{C}^+$ . Autrement dit, de toute sous-suite, je peux extraire une sous-suite qui converge vers zéro. Nécessairement, la suite  $(\mathbb{E} g_n(z) - t_n(z))$  converge vers zéro pour tout  $z \in \mathbb{C}^+$ .  $\square$

### 5.3.2 Démonstration dans le cas d'entrées gaussiennes

Soit  $X_n$  une matrice  $N \times n$  à entrées i.i.d.  $\mathcal{N}(0, 1)$  et  $R_N$  une matrice  $N \times N$  réelle, déterministe, semi-définie positive, de décomposition spectrale

$$R_N = O_N^T D_N O_N , \quad O_N O_N^T = O_N^T O_N = I_N$$

**Proposition 67.** *On pose*

$$\Sigma_n = \frac{1}{\sqrt{n}} R_n^{1/2} X_n , \quad \tilde{X}_n = O_N X_n , \quad \Gamma_n = \frac{1}{\sqrt{n}} D_N^{1/2} \tilde{X}_N$$

*alors la matrice  $\tilde{X}_n$  est à entrées i.i.d.  $\mathcal{N}(0, 1)$  et les matrices  $\Sigma_n \Sigma_n^*$  et  $\Gamma_n \Gamma_n^*$  ont même mesure spectrale.*

La proposition 67 stipule que dans le cas gaussien, il suffit d'étudier la mesure spectrale pour  $R_N$  diagonale.

*Démonstration de la proposition 67.* Introduisons les vecteurs colonne  $(\mathbf{x}_i)$  et  $(\tilde{\mathbf{x}}_i)$  des matrices  $X_n$  et  $\tilde{X}_n$ . On a

$$\tilde{X}_n = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n] = [O_N \mathbf{x}_1, \dots, O_N \mathbf{x}_n]$$

De manière évidente, les vecteurs  $\tilde{\mathbf{x}}_i$  sont des vecteurs gaussiens indépendants de loi  $\mathcal{N}_N(0, O_N O_N^T) = \mathcal{N}_N(0, I_N)$ . Le caractère i.i.d. gaussien des entrées de la matrice  $\tilde{X}_n$  en découle. Montrons maintenant que les transformées de Stieltjes  $g_n$  et  $\tilde{g}_n$  des mesures spectrales de  $\Sigma_n \Sigma_n^*$  et  $\Gamma_n \Gamma_n^*$  coïncident.

$$\begin{aligned} (\Sigma_n \Sigma_n^T - z I_N)^{-1} &= \left( \frac{1}{n} R_N^{1/2} X_n X_n^T R_N^{1/2} - z I_N \right)^{-1} \\ &= \left( \frac{1}{n} O_N^T D_N^{1/2} O_N X_n X_n^T O_N^T D_N^{1/2} O_N - z I_N \right)^{-1} \\ &= O_N^T \left( \frac{1}{n} D_N^{1/2} \tilde{X}_n \tilde{X}_n^T D_N^{1/2} - z I_N \right)^{-1} O_N \\ &= O_N^T (\Gamma_n \Gamma_n^T - z I_N)^{-1} O_N \end{aligned}$$

En prenant les traces normalisées de part et d'autre, il vient

$$\begin{aligned} \frac{1}{N} \text{Tr} (\Sigma_n \Sigma_n^T - z I_N)^{-1} &= \frac{1}{N} \text{Tr} O_N^T (\Gamma_n \Gamma_n - z I_N)^{-1} O_N \\ &= \frac{1}{N} \text{Tr} (\Gamma_n \Gamma_n - z I_N)^{-1} \end{aligned}$$

soit  $g_n = \tilde{g}_n$  □

### 5.3.3 Le cas général

#### Interpolation avec le cas gaussien

**Proposition 68.** *Soit  $A$  une matrice  $N \times N$  déterministe de résolvante  $Q(z) = (A - zI)^{-1}$  et  $R$  une matrice hermitienne  $N \times N$ . Soit  $\mathbf{x} = (x_1, \dots, x_n)^T$  un vecteur aléatoire  $n \times 1$  réel ou complexe, à composantes i.i.d. vérifiant*

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

*Soit par ailleurs un vecteur gaussien  $\tilde{\mathbf{x}}$  à composantes i.i.d. gaussiennes complexes standard, i.e.  $\tilde{x}_i \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ . On suppose que  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  sont indépendants. Alors*

$$\begin{aligned} \frac{1}{N} \mathbb{E} \text{Tr} \left( A + \frac{1}{n} R^{1/2} \mathbf{x} \mathbf{x}^* R^{1/2} - z I \right)^{-1} \\ - \frac{1}{N} \mathbb{E} \text{Tr} \left( A + \frac{1}{n} R^{1/2} \tilde{\mathbf{x}} \tilde{\mathbf{x}}^* R^{1/2} - z I \right)^{-1} &= \mathcal{O}_z \left( \frac{\|R\|^2}{n^{3/2}} \right) \end{aligned}$$

pour

$$0 < \liminf \frac{N}{n} \leq \limsup \frac{N}{n} < \infty .$$

*Démonstration.* Appelons  $\Delta$  le membre de gauche de l'équation ci-dessus. En utilisant la formule de perturbation de rang 1, il vient

$$\begin{aligned} \Delta &= \frac{1}{N} \mathbb{E} \left( \frac{\frac{1}{n} \mathbf{x}^* R^{1/2} Q^2 R^{1/2} \mathbf{x}}{1 + \frac{1}{n} \mathbf{x}^* R^{1/2} Q R^{1/2} \mathbf{x}} \right) - \frac{1}{N} \mathbb{E} \left( \frac{\frac{1}{n} \tilde{\mathbf{x}}^* R^{1/2} Q^2 R^{1/2} \tilde{\mathbf{x}}}{1 + \frac{1}{n} \tilde{\mathbf{x}}^* R^{1/2} Q R^{1/2} \tilde{\mathbf{x}}} \right) \\ &\stackrel{(a)}{=} \frac{1}{N} \mathbb{E} \left( \frac{1}{n} \mathbf{x}^* R^{1/2} Q^2 R^{1/2} \mathbf{x} \right) \left( \frac{1}{1 + \frac{1}{n} \mathbf{x}^* R^{1/2} Q R^{1/2} \mathbf{x}} - \frac{1}{1 + \frac{1}{n} \text{Tr } R^{1/2} Q R^{1/2}} \right) \\ &\quad - \frac{1}{N} \mathbb{E} \left( \frac{1}{n} \tilde{\mathbf{x}}^* R^{1/2} Q^2 R^{1/2} \tilde{\mathbf{x}} \right) \left( \frac{1}{1 + \frac{1}{n} \tilde{\mathbf{x}}^* R^{1/2} Q R^{1/2} \tilde{\mathbf{x}}} - \frac{1}{1 + \frac{1}{n} \text{Tr } R^{1/2} Q R^{1/2}} \right) \\ &:= \Delta_{\mathbf{x}} + \Delta_{\tilde{\mathbf{x}}} . \end{aligned}$$

Notons que l'identité (a) provient du fait que les vecteurs  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  ayant les mêmes moments d'ordre 1 et 2, on a l'égalité

$$\mathbb{E} \left( \frac{1}{n} \mathbf{x}^* R^{1/2} Q^2 R^{1/2} \mathbf{x} \right) = \mathbb{E} \left( \frac{1}{n} \tilde{\mathbf{x}}^* R^{1/2} Q^2 R^{1/2} \tilde{\mathbf{x}} \right) = \frac{1}{n} \text{Tr} (R^{1/2} Q R^{1/2}) .$$

Avant de nous attaquer à la majoration du module de  $\Delta_{\mathbf{x}}$ , remarquons que :

$$\left| \frac{\frac{1}{n} \mathbf{x}^* R^{1/2} Q^2 R^{1/2} \mathbf{x}}{1 + \frac{1}{n} \mathbf{x}^* R^{1/2} Q R^{1/2} \mathbf{x}} \right| \leq \frac{|\frac{1}{n} \mathbf{x}^* R^{1/2} Q^2 R^{1/2} \mathbf{x}|}{\text{Im} \left( \frac{1}{n} \mathbf{x}^* R^{1/2} Q R^{1/2} \mathbf{x} \right)} \leq \frac{1}{\text{Im}(z)}$$

et que

$$\left| \frac{1}{1 + \frac{1}{n} \text{Tr } R^{1/2} Q R^{1/2}} \right| \leq \frac{|z|}{\text{Im}(z)} .$$

Par suite, en réduisant au même dénominateur, il vient :

$$|\Delta_{\mathbf{x}}| \leq \frac{|z|}{N \text{Im}^2(z)} \text{var}^{1/2} \left( \frac{1}{n} \mathbf{x}^* R^{1/2} Q^2 R^{1/2} \mathbf{x} \right) = \mathcal{O}_z \left( \frac{\|R\|^2}{n^{3/2}} \right) .$$

La même estimation prévaut pour  $\Delta_{\tilde{\mathbf{x}}}$  ; cela nous permet alors de conclure.  $\square$

**Remarque 69.** *Il y a beaucoup de similitudes entre la preuve ci-dessus et celle permettant l'estimation de la variance de la transformée de Stieltjes à l'aide de l'inégalité d'Efron-Stein. La principale différence est liée au fait qu'ici, on ne fait pas l'hypothèse que  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  ont même loi, seulement les mêmes moments d'ordre 1 et 2.*

On peut maintenant interpoler entre les cas gaussien et non gaussien. On rappelle que

$$\Sigma_n = \frac{1}{\sqrt{n}} R_n^{1/2} X_n = [\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n]$$

et on introduit

$$\Sigma_n^{\mathbb{C}} = \frac{1}{\sqrt{n}} R_n^{1/2} X_n^{\mathbb{C}} = [\xi_1^{\mathbb{C}}, \dots, \xi_n^{\mathbb{C}}]$$

où  $X_n^{\mathbb{C}}$  est une matrice  $N \times n$  à entrées i.i.d.  $\mathcal{N}_{\mathbb{C}}(0, 1)$ . On interpole entre les résolvantes des matrices de covariance associées en changeant à chaque fois un vecteur colonne ; on introduit.

$$Q_n(z) = (\Sigma_n \Sigma_n^* - zI)^{-1}, \quad Q_n^{\mathbb{C}}(z) = (\Sigma_n^{\mathbb{C}} (\Sigma_n^{\mathbb{C}})^* - zI)^{-1}$$

et les matrices

$$Q_n^{(i)}(z) = \left( \sum_{k=1}^i \xi_k \xi_k^* + \sum_{k=i+1}^n \xi_k^{\mathbb{C}} (\xi_k^{\mathbb{C}})^* - zI \right)^{-1}, \quad 1 \leq i \leq n-1$$

en posant  $Q_n^{(0)} = Q_n^{\mathbb{C}}$  et  $Q_n^{(n)} = Q_n$ .

**Proposition 70.** *Pour  $z \in \mathbb{C}^+$ , on a*

$$\frac{1}{n} \text{Tr} \mathbb{E} Q_n(z) - \frac{1}{n} \text{Tr} \mathbb{E} Q_n^{\mathbb{C}}(z) \xrightarrow[n \rightarrow \infty]{} 0$$

*Démonstration.* On a

$$\begin{aligned} \frac{1}{n} \text{Tr} \mathbb{E} Q_n(z) - \frac{1}{n} \text{Tr} \mathbb{E} Q_n^{\mathbb{C}}(z) &= \sum_{i=0}^{n-1} \left( \frac{1}{n} \text{Tr} \mathbb{E} Q_n^{(i)}(z) - \frac{1}{n} \text{Tr} \mathbb{E} Q_n^{(i+1)}(z) \right), \\ &= \mathcal{O}_z \left( \frac{1}{\sqrt{n}} \right), \end{aligned}$$

d'après la proposition 68. □

On peut maintenant démontrer le théorème 58

*Preuve du théorème 58.* Concernant le point (a), l'existence et l'unicité d'une solution à l'équation (5.1), ainsi que ses propriétés, ont été établies à la proposition 65.

Abordons maintenant le point (b). L'estimation de la variance pour la transformée de Stieltjes  $g_n(z)$ , en  $\mathcal{O}_z(n^{-2})$ , nous permet d'établir

$$g_n(z) - \mathbb{E} g_n(z) \xrightarrow[n \rightarrow \infty]{p.s.} 0.$$

L'interpolation avec le modèle à entrées gaussiennes (cf. proposition 70) entraîne que

$$\mathbb{E} g_n(z) - \mathbb{E} g_n^{\mathbb{C}}(z) \xrightarrow[n \rightarrow \infty]{} 0 \text{ où } g_n^{\mathbb{C}}(z) = \frac{1}{N} \text{Tr} Q_n^{\mathbb{C}}(z).$$

La proposition 67 nous permet de nous ramener alors à une matrice  $R_N$  de type diagonal pour laquelle la proposition 66 s'applique : pour  $z \in \mathbb{C}$ ,

$$\mathbb{E} g_n^{\mathbb{C}}(z) - t_n(z) \xrightarrow[n \rightarrow \infty]{} 0,$$

soit finalement, en collectant ces résultats intermédiaires,

$$g_n(z) - t_n(z) \xrightarrow[n \rightarrow \infty]{p.s.} 0 .$$

Démontrons maintenant le point (c). Montrons dans un premier temps que la famille de mesures spectrales  $(L_n)$  ainsi que la famille de probabilités  $(\mathcal{F}_n)$  sont tendues. On va s'appuyer pour cela sur le critère de tension utilisé dans l'exemple 6. On a

$$\int_{\mathbb{R}^+} \lambda L_n(d\lambda) = \frac{1}{N} \sum_{i=1}^N \lambda_i = \frac{1}{N} \text{Tr} R_N^{1/2} X_N X_N^* R_N^{1/2} \leq \frac{\|R_N\|}{N} \text{Tr} X_N X_N^* .$$

Donc presque sûrement,

$$\limsup_n \int_{\mathbb{R}^+} \lambda L_n(d\lambda) \leq \limsup_n \frac{\mathbf{R}}{Nn} \sum_{i=1:N, j=1:n} |X_{ij}|^2 < \infty .$$

Cela assure que presque sûrement,  $(L_n)$  est tendue.

Considérons maintenant la probabilité  $\mathcal{F}_n$  sur  $\mathbb{R}^+$  dont la transformée de Stieltjes est  $t_n$ . On montre facilement que

$$\int_{\mathbb{R}^+} \lambda \mathcal{F}_n(d\lambda) = \lim_{u \rightarrow +\infty} u [1 - u t_n(-u)] .$$

En effet,

$$u [1 - u t_n(-u)] = u \int \frac{\lambda}{\lambda + u} \mathcal{F}_n(d\lambda) \xrightarrow{u \rightarrow \infty} \int_{\mathbb{R}^+} \lambda \mathcal{F}_n(d\lambda) .$$

Remarquons aussi que  $\lim_{u \rightarrow \infty} u t_n(-u) = 1$ . Évaluons maintenant la limite de  $u [1 - u t_n(-u)]$  quand  $u \rightarrow \infty$  à l'aide de l'équation satisfaite<sup>1</sup> par  $t_n(z)$ . On a

$$\begin{aligned} u [1 - u t_n(-u)] &= u \left( 1 - \frac{1}{N} \sum_{i=1}^N \frac{u}{u(1 + c_N \rho_i t_n(-u)) + (1 - c_N) \rho_i} \right) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{u^2 c_N \rho_i t_n(-u) + u(1 - c_N) \rho_i}{u(1 + c_N \rho_i t_n(-u)) + (1 - c_N) \rho_i} \\ &\xrightarrow{u \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \rho_i = \frac{1}{N} \text{Tr} R_N \leq \sup_n \|R_N\| < \infty . \end{aligned}$$

Soit finalement

$$\sup_n \int \lambda \mathcal{F}_n(d\lambda) \leq \sup_n \|R_N\| < \infty ,$$

ce qui nous assure que la famille de probabilités  $(\mathcal{F}_n)$  est tendue.

<sup>1</sup>. On pourra justifier que l'équation satisfaite par  $t_n(z)$  reste satisfaite si  $z$  est un réel négatif.

Soit  $\mathcal{D}$  un sous ensemble de  $\mathbb{C}^+$  dénombrable et contenant un point d'accumulation tel que presque sûrement, pour tout  $z \in \mathcal{D}$ , on ait

$$g_n(z) - t_n(z) \xrightarrow[n \rightarrow \infty]{} 0$$

et  $(L_n)$  tendue. Supposons l'existence d'une fonction  $f$  continue bornée pour laquelle  $\int f(\lambda)L_n(d\lambda) - \int f(\lambda)\mathcal{F}_n(d\lambda)$  ne converge pas vers zéro. Alors il existe  $\varepsilon > 0$  et deux sous-suites  $(L'_n)$  et  $(\mathcal{F}'_n)$  telles que

$$\left| \int f(\lambda)L'_n(d\lambda) - \int f(\lambda)\mathcal{F}'_n(d\lambda) \right| > \varepsilon. \quad (5.14)$$

De la sous-suite  $(\mathcal{F}'_n)$  considérons une sous-suite extraite  $(\mathcal{F}''_n)$  qui converge étroitement vers  $\mathcal{F}^*$ , de transformée de Stieltjes  $t^*$ . Alors  $t''_n(z) \rightarrow t^*(z)$  pour tout  $z \in \mathcal{D}$  et en particulier  $g''_n(z) \rightarrow t^*(z)$  pour tout  $z \in \mathcal{D}$ . Cela entraîne alors que  $L''_n$  converge étroitement vers  $\mathcal{F}^*$ , ce qui contredit (5.14).  $\square$

## 5.4 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 [23].

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

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

**Lemme 71.** *Dans les conditions du Théorème 62, 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

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

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

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

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  $L_{\infty}^R$  et  $\mathcal{F}$ . Cette connexion sera fondamentale pour les questions d'inférences (à savoir étudier les propriétés de  $L_{\infty}^R$  étant données celles de  $\mathcal{F}$ ).

### 5.4.1 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 72.** *Soit  $g_\mu$  la transformée de Stieltjes de la mesure  $\mu$  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  $\mu$  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  $g_\mu$  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  $\mu$  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  $\mu$  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  $|I(x) - I(x_0)| < 2\varepsilon$ .  $\square$

### 5.4.2 Application au modèle de covariance

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

$$(\tilde{t}(z_1) - \tilde{t}(z_2)) \left( 1 - c \int \frac{\tilde{t}(z_1) \tilde{t}(z_2) u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_1))(1 + u\tilde{t}(z_2))} \right) = (z_1 - z_2) \tilde{t}(z_1) \tilde{t}(z_2). \quad (5.15)$$

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)

$$1 - c|\tilde{t}(z)|^2 \int \frac{u^2 L_\infty^R(du)}{|1 + u\tilde{t}(z)|^2} > 0 \quad (5.16)$$

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 73.** *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 72 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 (5.15), on trouve que

$$\begin{aligned} & (z_n \tilde{t}(z_n) - z_0 \tilde{t}(z_0)) \left( 1 - c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u^2 L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))} \right) \\ &= (z_n - z_0)c \int \frac{\tilde{t}(z_n)\tilde{t}(z_0)u L_\infty^R(du)}{(1 + u\tilde{t}(z_n))(1 + u\tilde{t}(z_0))}. \end{aligned} \quad (5.17)$$

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 (5.16), 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 (5.17), 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 (5.15), on a alors

$$(\tilde{t}(z_{1,n}) - \tilde{t}(z_{2,n})) \left( 1 - c \int \frac{\tilde{t}(z_{1,n})\tilde{t}(z_{2,n})u^2 L_\infty^R(du)}{(1+u\tilde{t}(z_{1,n}))(1+u\tilde{t}(z_{2,n}))} \right) = (z_{1,n} - z_{2,n})\tilde{t}(z_{1,n})\tilde{t}(z_{2,n}). \quad (5.18)$$

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  $\Re[f f^* g] = \frac{1}{4}[\|f + g\|^2 - \|f - g\|^2]$ . Pour le premier terme de droite, par  $|a + b|^2 \leq 2(|a|^2 + |b|^2)$  et l'inégalité (5.16),

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

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 (5.18) 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 [18], n'a été rigoureusement prouvée que dans [23]. Nous donnons (une version parfois améliorée) de cette preuve ici.

**Théorème 74.** *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

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

Par ailleurs,

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

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  $y \searrow 0$ ,

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

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  $\tilde{t}(x_0)$ ,

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

où on définit  $r(z) = \int (u-z)^{-1} L_\infty^R(du)$  la transformée de Stieltjes de la mesure  $L_\infty^R$ . 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 72, 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  $x(\tilde{t}(x_0)) = -\frac{1}{\tilde{t}(x_0)} + c \int \frac{uL_\infty^R(du)}{1+u\tilde{t}(x_0)}$  (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 (5.16) 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  $L_\infty^R$  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 analyticit  de  $\tilde{t}$  au voisinage de  $x_0$ ,  crivons

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

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

Il nous reste   d montrer la relation r ciproque. 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’apr s la Proposition 72.  $\square$

Une repr sentation graphique de la fonction  $x(\tilde{t})$  est donn e en Figure 5.1. Le comportement de la fonction est tr s particulier. Par des arguments similaires   ceux  voqu s dans les preuves, il peut pr cis ment  tre d montr  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 5.1 ; cela d pend de la forme sp cifique de  $L_\infty^R$ .

Dapr s la Figure 5.1, il est int ressant de noter l’existence ou non de d riv es nulles pour  $x(\tilde{t})$  aux positions correspondant   l’entr e ou la sortie du support  $\mathcal{S}$ . Le signe de la d riv e prise au point d’inflexion n cessairement existant dans la repr sentation permet en particulier d’assurer l’existence ou non d’une partie croissante pour  $x(\tilde{t})$  dans chaque intervalle de d finition.

Nous concluons la section par la caract risation, plus simple, de la masse en z ro.

**Th or me 75.** *Sous les hypoth ses de travail de cette section, la mesure  $\tilde{\mathcal{F}}$  a une masse en z ro donn e par*

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

*D monstration.* On utilise ici la formule (voir Proposition 2-6)

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

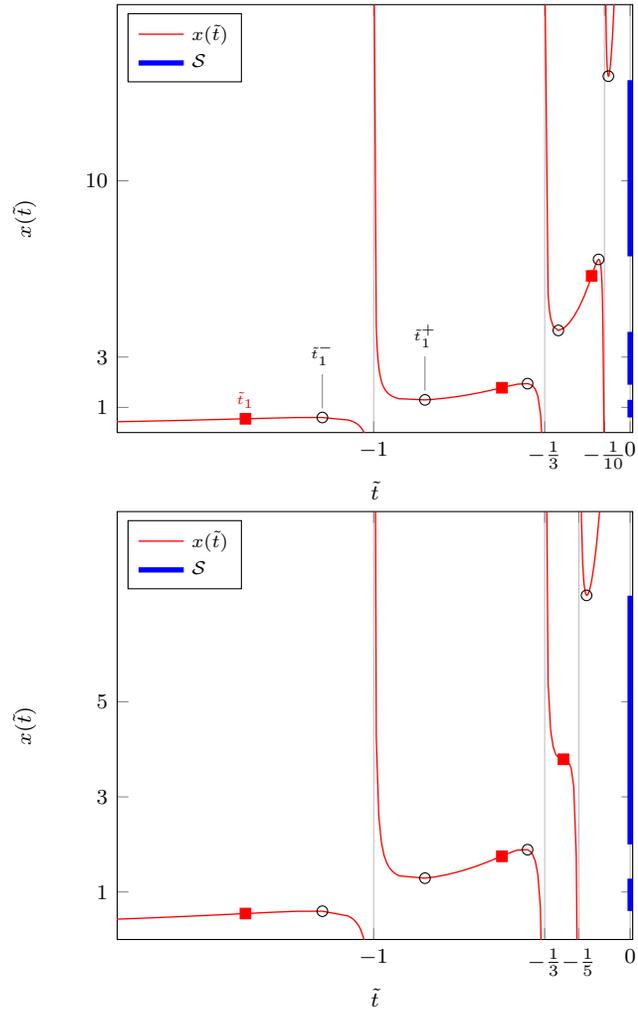


FIGURE 5.1 –  $x(\tilde{t})$  pour  $\tilde{t}$  réel,  $R_N$  diagonal composé de trois masses en 1, 3 et 10 (figure supérieure) et 1, 3 et 5 (figure inférieure),  $c = 1/10$  dans les deux cas. Extrema locaux marqués par des cercles, points d’inflexion par des carrés. Le support  $\mathcal{S}$  de  $\mathcal{F}$  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)

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

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

## 5.5 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  $L_\infty^R$  (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.

### 5.5.1 Quelques rappels d'analyse complexe

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

**Proposition 76** (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 77** (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 78** (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.*

### 5.5.2 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  $L_{\infty}^R$ . En particulier, nous pourrions demander à  $f$  d'être une quelconque fonction analytique sur une composante connexe de  $L_{\infty}^R$  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  $\lambda_{\ell}^R$  individuellement (auquel cas, si on connaît les  $\alpha_{\ell}$ , on prendra  $f(t) = t$  au voisinage de  $\lambda_{\ell}^R$  et zéro ailleurs).

Le cœur de l'approche consiste à observer que, grâce au théorème de Cauchy (Proposition 76),

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

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

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

En fait, posée de telle manière, la question n'a pas de réponse claire. Cependant, en prenant le problème différemment, on parvient à prouver dans certains cas que le changement de variable est valable pour un contour  $\mathcal{C}$  qui convient (mais pas forcément celui qu'on a construit au départ).

Nous avons vu précédemment que pour  $z \notin \mathcal{S}$ ,  $-\frac{1}{\tilde{t}(z)} \notin \text{supp}(L_\infty^R)$  (Théorème 74-2). Ainsi il est possible de générer 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  $L_\infty^R$ . Les discussions de la section précédente, et en particulier la visualisation de la Figure 5.1, nous permet de comprendre ce changement de variable :

- ▷ si  $c$  est suffisamment faible ou alternativement, si les composantes connexes du support de  $L_\infty^R$  sont suffisamment disjointes (cas de la Figure 5.1 supérieure), alors, pour  $\mathcal{C}^{\mathcal{F}}$  entourant la  $k$ -ième composante connexe de  $\mathcal{S}$  (disons, en partant de zéro et en excluant la masse en 0), l'image  $-\frac{1}{\tilde{t}(\mathcal{C}^{\mathcal{F}})}$  entourera exactement la  $k$ -ième composante connexe de  $\text{supp}(L_\infty^R)$ ;
- ▷ si au contraire on se trouve dans le cas de la Figure 5.1 inférieure, 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 méthode, d'obtenir une estimée 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 manière générale, il sera très souvent (en fait pas forcément toujours) possible d'estimer  $G(f)$  si  $f$  est une fonction analytique du support de  $L_\infty^R$  entier. Grâce à l'étude effectuée dans la section précédente, en particulier l'étude du signe de la dérivée de la fonction  $x(\tilde{t})$  aux points d'inflexions, il est possible d'établir les zones dites de *séparabilité* de  $\mathcal{S}$  en autant de composantes connexes qu'en contient  $\text{supp}(L_\infty^R)$  (toujours Dirac en zéro mis à part). Les Figures 5.2 et 5.3 donnent une représentation des zones de séparabilité dans le cas où  $L_\infty^R$  est la somme de deux ou trois masses de Dirac.

La Figure 5.4 donne une idée 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 5.1 supérieure.

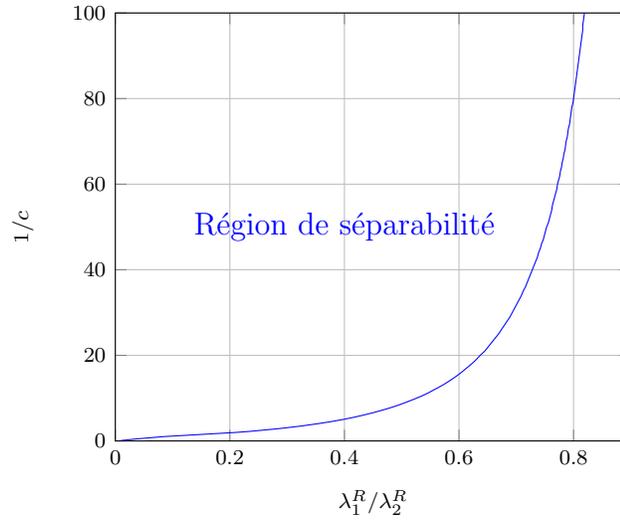


FIGURE 5.2 – Rapport limite  $c$  pour obtenir la séparabilité de  $\lambda_1^R \leq \lambda_2^R$  pour  $L_\infty^R = \frac{1}{2}(\delta_{\lambda_1^R} + \delta_{\lambda_2^R})$ .

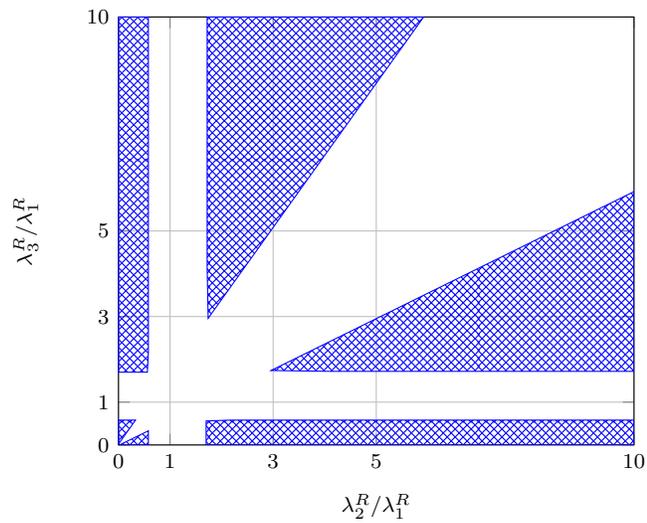


FIGURE 5.3 – Sous-ensemble des  $\lambda_1^R \leq \lambda_2^R \leq \lambda_3^R$  (région hachurée) assurant la condition de séparabilité pour  $L_\infty^R = \frac{1}{3}(\delta_{\lambda_1^R} + \delta_{\lambda_2^R} + \delta_{\lambda_3^R})$ ,  $c = 1/10$ .

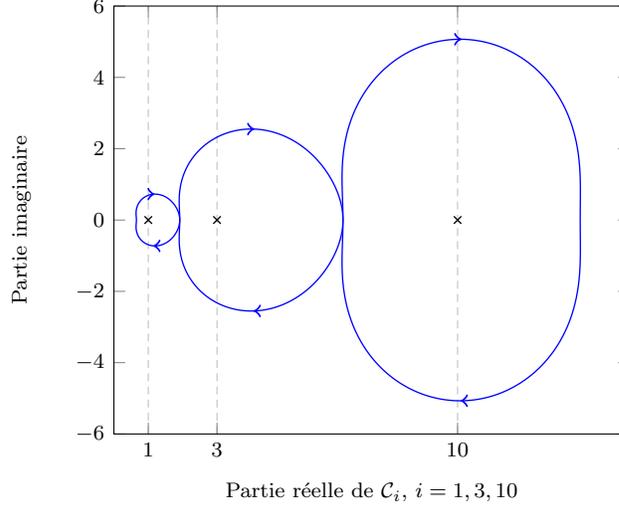


FIGURE 5.4 – Contours  $\mathcal{C}_1$ ,  $\mathcal{C}_3$  et  $\mathcal{C}_{10}$  ante-images des  $\mathcal{C}_i^{\mathcal{F}}$ , contours circulaires entourant les composantes connexes de  $\mathcal{S}$ . Configuration de la Figure 5.1 supérieure.

### 5.5.3 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 62-(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 62-(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 78) et au résultat fondamental complémentaire suivant (que nous ne prouverons pas), adapté de [2] :

**Théorème 79** (Pas de valeurs propres en dehors du support). *Dans les conditions du Théorème 62, 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' $\varepsilon$ -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 ??) est bien plus fort que la convergence faible établie au Théorème 62 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 79 garantit qu'il n'existe pas de valeur propre isolée pour  $n$  suffisamment large.

Plaçons nous donc sous les hypothèses du Théorème 79 (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  $\Omega_z$  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 78) 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 sûrement. 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 80** (Monôme d'ordre 2). Prenons  $f(w) = w^2$ . Alors, avec probabilité un,

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

Par intégration par partie, cela donne finalement

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

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  $\mathcal{C}^{\mathcal{F}}$  telles que  $\tilde{g}_n(z) = 0$ . Grâce au lemme suivant, ces valeurs sont toutes identifiées.

**Lemme 81.** 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  $\Lambda$ , 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  $\Lambda$ , 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 Sylverster. □

Dans notre cas, on applique le lemme pour  $\Lambda = \text{diag}(\lambda_i)$  et  $a = \sqrt{\lambda/n}$  avec  $\lambda$  le vecteur des  $\lambda_i$ , 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  $\lambda_i$  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  $\lambda_i$  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  $\lambda_i$  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 77,

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

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  $\lambda_i$  est inclus dans le contour  $\mathcal{C}^S$ .

## Chapitre 6

# Confinement des valeurs propres

### 6.1 La formule d'Helffer-Sjöstrand

Soit  $\mu$  une probabilité sur  $\mathbb{R}$  et  $g_\mu$  sa transformée de Stieltjes. Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction réelle à support compact, de classe  $C^{k+1}$ . On introduit l'extension suivante de  $f$  sur  $\mathbb{C}^+$  définie par

$$\Phi_k(f)(x, y) = \sum_{\ell=0}^k \frac{(iy)^\ell}{\ell!} f^{(\ell)}(x) \chi(y), \quad x + iy \in \mathbb{C}^+, \quad (6.1)$$

où  $\chi : \mathbb{R} \rightarrow [0, 1]$  est une fonction paire  $C^\infty$  telle que

$$\chi(y) = 1 \quad \text{pour } 0 \leq y \leq \eta; \quad \chi(y) = 0 \quad \text{pour } y \geq 1.$$

L'extension (6.1) s'étend naturellement, si besoin est, au cas où  $y < 0$ .

On notera indifféremment  $\Phi_k(f)(x, y)$  ou  $\Phi_k(f)(z)$  avec  $z = x + iy$ . Ainsi,

$$\bar{\partial} \Phi_k(f) = \frac{\partial \Phi_k(f)}{\partial \bar{z}} = \frac{1}{2} \{ \partial_x \Phi_k(f) + i \partial_y \Phi_k(f) \},$$

où  $\partial_x = \frac{\partial}{\partial x}$  et  $\partial_y = \frac{\partial}{\partial y}$ .

**Propriétés 82.** — L'évaluation de  $\Phi_k(f)$  sur l'axe réel donne

$$\Phi_k(f)(x, 0) = f(x).$$

- Le support de  $\Phi_k(f)$  est inclus dans  $\text{supp } f \times [0, 1]$ .
- Un calcul élémentaire lié au fait que  $\chi(y)$  est constante lorsque  $x + iy$  est proche de l'axe réel ( $y \leq \eta$ ) montre que

$$\bar{\partial} \Phi_k(f)(x, y) = \frac{(iy)^k}{k!} f^{(k+1)}(x), \quad y \leq \eta. \quad (6.2)$$

**Proposition 83.** Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  de classe  $C^{k+1}$ , avec  $k \geq 1$ , à support compact et  $\Phi_k(f)$  son extension donnée par (6.1), alors

$$\int f(\lambda) \mu(d\lambda) = \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) g_\mu(x + iy) dx dy .$$

**Remarque 84.** Pour une fonction  $f \in C_c^{k+1}(\mathbb{R})$ , l'identité d'Helffer-Sjöstrand est valable pour toute extension  $\Phi_\ell(f)$  de  $f$  d'indice  $\ell$  compris entre 1 et  $k$ . Plus l'extension  $\Phi_\ell(f)$  a un indice  $\ell$  élevé, meilleure sera la convergence vers zéro de  $\bar{\partial} \Phi_\ell(f)$  lorsque  $z$  s'approche de la droite réelle, en vertu de la propriété

$$\bar{\partial} \Phi_k(f)(x, y) = \frac{(iy)^k}{k!} f^{(k+1)}(x) , \quad y \leq \eta .$$

Ceci est particulièrement utile lorsqu'on veut comparer

$$\int f d\mu - \int f d\nu$$

à partir d'une estimée sur la différence de leurs transformées de Stieltjes du type

$$g_\mu(z) - g_\nu(z) \propto \frac{1}{\operatorname{Im}^k(z)} .$$

La régularité de  $f$  permettra alors de compenser la divergence de  $g_\mu(z) - g_\nu(z)$  près de l'axe réel :

$$\int f d\mu - \int f d\nu = \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \underbrace{\bar{\partial} \Phi_k(f)(z)}_{\propto y^k f^{(k+1)}(x)} \underbrace{(g_\mu - g_\nu)(z)}_{\propto \frac{1}{y^k}} dx dy .$$

**Remarque 85.** L'application de l'identité d'Helffer-Sjöstrand à la mesure empirique des valeurs propres d'une matrice hermitienne donne

$$\frac{1}{N} \sum_{i=1}^N f(\lambda_i) = \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) g_n(x + iy) dx dy .$$

*Démonstration de la proposition 83.* On montre facilement, en utilisant (6.2), que

$$(\lambda, x, y) \mapsto \bar{\partial} \Phi_k(f)(x, y) \frac{1}{\lambda - x - iy}$$

est bornée sur  $\mathbb{R}^2 \times \mathbb{R}^+$ . Le théorème de Fubini entraîne alors que

$$\begin{aligned} & \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) g_\mu(x + iy) dx dy \\ &= \int_{\mathbb{R}} \mu(d\lambda) \left\{ \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) \frac{1}{\lambda - x - iy} dx dy \right\} . \end{aligned}$$

Comme  $\frac{1}{\lambda-x-iy}$  est la transformée de Stieltjes de la mesure de Dirac en  $\lambda$ , il suffit d'établir la formule d'Helffer-Sjöstrand pour une telle mesure, soit :

$$f(\lambda) = \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \bar{\partial} \Phi_k(f)(x, y) \frac{1}{\lambda - x - iy} dx dy \quad (6.3)$$

et d'utiliser l'argument du théorème de Fubini ci-dessus pour conclure. Montrons la formule pour la mesure de Dirac en zéro :

$$f(0) = -\frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \frac{\bar{\partial} \Phi_k(f)(x, y)}{x + iy} dx dy \quad (6.4)$$

la formule (6.3) suivra alors par simple changement de variable.

Dans la suite, on note simplement  $\Phi$  au lieu de  $\Phi_k(f)$ . On a :

$$\frac{\bar{\partial} \Phi(x, y)}{x + iy} = \frac{1}{2} \frac{\partial_x \Phi + i \partial_y \Phi}{x + iy} = \frac{1}{2} \frac{x \partial_x \Phi + y \partial_y \Phi + i(x \partial_y \Phi - y \partial_x \Phi)}{x^2 + y^2}.$$

On passe aux coordonnées polaires :

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases}, \quad dx dy = r dr d\theta, \quad \begin{cases} x \in \mathbb{R} \\ y \in ]0, \infty[ \end{cases} \Leftrightarrow \begin{cases} r \in ]0, \infty[ \\ \theta \in ]0, \pi[ \end{cases}.$$

Soit  $\Phi = \Phi(r \cos \theta, r \sin \theta)$ , alors

$$\partial_r \Phi = \cos \theta \partial_x \Phi + \sin \theta \partial_y \Phi \quad \Rightarrow \quad r \partial_r \Phi = x \partial_x \Phi + y \partial_y \Phi.$$

De même,

$$\partial_\theta \Phi = -r \sin \theta \partial_x \Phi + r \cos \theta \partial_y \Phi \quad \Rightarrow \quad \partial_\theta \Phi = -y \partial_x \Phi + x \partial_y \Phi,$$

et

$$\frac{\bar{\partial} \Phi(x, y)}{x + iy} = \frac{1}{2} \left\{ \frac{\partial_r \Phi}{r} + i \frac{\partial_\theta \Phi}{r^2} \right\},$$

soit

$$\begin{aligned} \int_{\mathbb{R} \times ]0, \infty[} \frac{\bar{\partial} \Phi(f)(x, y)}{x + iy} dx dy &= \frac{1}{2} \int_{]0, \infty[ \times ]0, \pi[} \left\{ \frac{\partial_r \Phi}{r} + i \frac{\partial_\theta \Phi}{r^2} \right\} r dr d\theta \\ &= \frac{1}{2} \int_{]0, \infty[ \times ]0, \pi[} \partial_r \Phi dr d\theta + \frac{i}{2} \int_{]0, \infty[ \times ]0, \pi[} \frac{\partial_\theta \Phi}{r} dr d\theta \\ &= \frac{1}{2} \int_{]0, \pi[} d\theta [\Phi(r \cos \theta, r \sin \theta)]_{r=0}^{r=\infty} + \frac{i}{2} \int_{]0, \infty[} \frac{dr}{r} [\Phi(r \cos \theta, r \sin \theta)]_{\theta=0}^{\theta=\pi} \\ &= -\frac{\pi}{2} f(0) + \frac{i}{2} \int_0^\infty \frac{f(-r) - f(r)}{r} dr \end{aligned}$$

On obtient finalement bien le resultat escompté :

$$-\frac{2}{\pi} \operatorname{Re} \int_{\mathbb{R} \times ]0, \infty[} \frac{\bar{\partial} \Phi_k(f)(x, y)}{x + iy} dx dy = f(0).$$

□

### Calcul fonctionnel pour les matrices hermitiennes

Etant donnée une matrice hermitienne  $A$  de dimensions  $n \times n$ , de décomposition spectrale  $U^* \Delta U$  où  $\Delta = \text{diag}(\lambda_i)$  et une fonction  $f$  continue à support compact, on définit la matrice  $f(A)$  par

$$f(A) = U^* \begin{pmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{pmatrix} U .$$

**Remarque 86.** En considérant la mesure empirique des valeurs propres  $L_n = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$ , on obtient une expression alternative pour  $\int f dL_n$  :

$$\int_{\mathbb{R}} f(x) L_n(dx) = \frac{1}{N} \sum_{i=1}^N f(\lambda_i) = \frac{1}{N} \text{Tr} f(A) .$$

On remarque alors que, si la fonction  $f$  est suffisamment régulière ( $f \in C_c^{k+1}(\mathbb{R})$ ), la définition de  $f(A)$  est compatible avec la formule d'Helffer-Sjöstrand. En effet :

$$\begin{aligned} f(A) &= U^* \begin{pmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{pmatrix} U \\ &= \frac{2}{\pi} U^* \int_{\mathbb{C}^+} \begin{pmatrix} \text{Re} \left\{ \frac{\bar{\partial} \Phi_k(f)}{\lambda_1 - z} \right\} & & 0 \\ & \ddots & \\ 0 & & \text{Re} \left\{ \frac{\bar{\partial} \Phi_k(f)}{\lambda_n - z} \right\} \end{pmatrix} dx dy U \\ &= \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_k(f)(z) U^* (\Delta - zI)^{-1} U dx dy + \frac{1}{\pi} \int_{\mathbb{C}^+} \overline{\bar{\partial} \Phi_k(f)(z)} U^* (\Delta - \bar{z}I)^{-1} U dx dy \end{aligned}$$

Cela permet d'obtenir une expression alternative de la matrice  $f(A)$  à l'aide de l'extension  $\Phi_k(f)$  et de la résolvante  $(A - zI)^{-1}$  :

$$f(A) = \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_k(f)(z) (A - zI)^{-1} dx dy + \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_k(f)(\bar{z}) (A - \bar{z}I)^{-1} dx dy .$$

**Proposition 87.** Soit  $X$  une matrice  $N \times n$  à entrées complexes et  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction de classe  $C^3$  à support compact. On pose

$$A = \frac{\sigma^2}{n} X X^* .$$

Alors

$$\begin{aligned} \frac{\partial}{\partial \bar{X}_{k\ell}} \left\{ \frac{1}{N} \text{Tr} f(A) \right\} &= \frac{\sigma^2}{Nn} [f'(A)X]_{k\ell} , \\ \frac{\partial}{\partial X_{k\ell}} \left\{ \frac{1}{N} \text{Tr} f(A) \right\} &= \frac{\sigma^2}{Nn} [X^* f'(A)]_{\ell k} , \end{aligned}$$

où  $f'$  représente la dérivée de  $f$  par rapport à  $x$ .

*Démonstration.* Notons  $g_n(z) = \frac{1}{N} \text{Tr}(A - zI)^{-1}$ , alors

$$\frac{1}{N} \text{Tr} f(A) = \frac{2}{\pi} \text{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) g_n(z) dx dy .$$

Du fait que  $|g_n(z)| \leq y^{-1}$  et que  $\bar{\partial} \Phi_2(f)(z) = -y^2 f^{(3)}(x)/2$  près de l'axe réel, l'intégrande est uniformément majoré en  $\bar{X}_{k\ell}$  et le théorème de dérivation sous le signe intégral nous donne :

$$\begin{aligned} \frac{\partial}{\partial \bar{X}_{k\ell}} \left\{ \frac{1}{N} \text{Tr} f(A) \right\} &= \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) \frac{\partial g_n(z)}{\partial \bar{X}_{k\ell}} dx dy + \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(\bar{z}) \frac{\partial g_n(\bar{z})}{\partial \bar{X}_{k\ell}} dx dy , \\ &= -\frac{\sigma^2}{Nn} \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) [Q^2(z)X]_{k\ell} dx dy \\ &\quad - \frac{\sigma^2}{Nn} \frac{1}{\pi} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(\bar{z}) [Q^2(\bar{z})X]_{k\ell} dx dy \\ &= -\frac{\sigma^2}{Nn} \frac{2}{\pi} \sum_i X_{i\ell} \text{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) Q_{ki}^2(z) dx dy \end{aligned}$$

Remarquons que

$$Q^2(z) = Q^2(x + iy) = \frac{\partial}{\partial x} Q(x + iy) .$$

En effectuant une intégration par parties (selon la variable  $x$ ), il vient

$$\int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f)(z) Q_{ki}^2(z) dx dy = - \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f')(z) Q_{ki}(z) dx dy ,$$

soit finalement

$$\begin{aligned} \frac{\partial}{\partial \bar{X}_{k\ell}} \left\{ \frac{1}{N} \text{Tr} f(A) \right\} &= \frac{\sigma^2}{Nn} \frac{2}{\pi} \sum_i X_{i\ell} \text{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_2(f')(z) Q_{ki}(z) dx dy , \\ &= \frac{\sigma^2}{Nn} \sum_i X_{i\ell} [f'(A)]_{ki} \\ &= \frac{\sigma^2}{Nn} [f'(A)X]_{k\ell} . \end{aligned}$$

Un calcul similaire nous permet d'obtenir la deuxième identité de la proposition.  $\square$

## 6.2 Confinement des valeurs propres

En combinant le lemme de stabilité pour l'équation de Marčenko-Pastur (lemme 46) et l'estimée (4.12) obtenue pour les entrées gaussiennes, on obtient l'estimée suivante :

$$\mathbb{E} g_n(z) - g_{\text{MP}}^{c_n}(z) = \mathcal{O}_z \left( \frac{1}{n^2} \right) . \quad (6.5)$$

L'identité d'Helffer-Sjöstrand permet de transférer cette estimée à une fonction régulière à support compact.

**Proposition 88.** *Soit  $f$  une fonction  $C^\infty$  à support compact, alors*

$$\mathbb{E} \frac{1}{N} \operatorname{Tr} f \left( \frac{\sigma^2}{n} X X^* \right) - \int f(x) \mathbb{P}_{\check{M}P}(dx) = O \left( \frac{1}{n^2} \right) .$$

*Démonstration.* D'après l'estimation (6.5), il existe  $p, q$  entiers tels que

$$|\mathbb{E} g_n(z) - g_{\check{M}P}(z)| \leq \frac{K}{n^2} \frac{|z|^p}{\operatorname{Im}(z)^q} .$$

En utilisant l'identité d'Helffer-Sjöstrand avec  $\Phi_q(f)$ , il vient

$$\mathbb{E} \frac{1}{N} \operatorname{Tr} f \left( \frac{\sigma^2}{n} X X^* \right) - \int f(x) \mathbb{P}_{\check{M}P}(dx) = \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \bar{\partial} \Phi_q(f) \{ \mathbb{E} g_n(z) - g_{\check{M}P}(z) \} dx dy$$

Comme  $\bar{\partial} \Phi_q(f)(z) = \frac{(iy)^q}{q!} f^{(q+1)}(x)$  près de l'axe réel ( $y \leq \eta$ ), on obtient alors

$$\begin{aligned} & \left| \mathbb{E} \frac{1}{N} \operatorname{Tr} f \left( \frac{\sigma^2}{n} X X^* \right) - \int f(x) \mathbb{P}_{\check{M}P}(dx) \right| \\ & \leq \frac{2}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \left| \bar{\partial} \Phi_q(f) \frac{K}{n^2} \frac{|z|^p}{\operatorname{Im}(z)^q} \right| dx dy \\ & \leq \frac{K'}{n^2} , \end{aligned}$$

ce qui achève la démonstration.  $\square$

On notera à l'occasion  $A = \frac{\sigma^2}{n} X X^*$ .

**Proposition 89.** *Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction  $C^\infty$  à support compact, alors*

$$\operatorname{var} \left( \frac{1}{N} \operatorname{Tr} f \left( \frac{\sigma^2}{n} X X^* \right) \right) \leq \frac{2\sigma^4}{N^2 n^2} \mathbb{E} \operatorname{Tr} (f'(A) X X^* f'(A)) .$$

*Démonstration.* Il suffit d'appliquer l'inégalité de Poincaré et la proposition 87 :

$$\begin{aligned} \operatorname{var} \left( \frac{1}{N} \operatorname{Tr} f(A) \right) & \leq \sum_{k\ell} \mathbb{E} \left| \frac{\partial}{\partial \bar{X}_{k\ell}} \left( \frac{1}{N} \operatorname{Tr} f(A) \right) \right|^2 + \sum_{k\ell} \mathbb{E} \left| \frac{\partial}{\partial X_{k\ell}} \left( \frac{1}{N} \operatorname{Tr} f(A) \right) \right|^2 \\ & \leq \frac{\sigma^4}{n^2 N^2} \sum_{k\ell} \mathbb{E} |[f'(A)X]_{k\ell}|^2 + \frac{\sigma^4}{n^2 N^2} \sum_{k\ell} \mathbb{E} |[X^* f'(A)]_{k\ell}|^2 \\ & \leq \frac{2\sigma^4}{N^2 n^2} \mathbb{E} \operatorname{Tr} (f'(A) X X^* f'(A)) . \end{aligned}$$

$\square$

**Théorème 90.** *Supposons que  $\frac{N}{n} \rightarrow c \in (0, 1]$ . Soit  $\mathcal{S}_n$  le spectre empirique des valeurs propres de  $A = \frac{\sigma^2}{n} XX^*$  et  $\mathcal{S}_\infty$  le support de la loi de Marčenko-Pastur de paramètre  $c$  :*

$$\mathcal{S}_\infty^c = [\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2] .$$

Alors pour tout  $\varepsilon > 0$ , presque sûrement

$$\mathcal{S}_n \subset \mathcal{S}_\infty + (-\varepsilon, \varepsilon)$$

pour  $n$  suffisamment grand.

Autrement dit, presque sûrement, les valeurs propres empirique sont confinées près du spectre limite.

*Démonstration.* Soit  $\varepsilon > 0$  fixé et  $\varphi_n$  une fonction  $C^\infty$ , à valeurs dans  $[0, 1]$ , valant 1 sur  $\mathcal{S}_\infty^{c_n}$  et zéro sur le complémentaire de  $\mathcal{S}_\infty^{c_n} + (-\frac{\varepsilon}{2}, \frac{\varepsilon}{2})$ . On pose

$$\psi_n = 1 - \varphi_n$$

et on va démontrer que

$$\mathrm{Tr} \psi_n \left( \frac{\sigma^2}{n} XX^* \right) \xrightarrow[N, n \rightarrow \infty]{p.s.} 0 . \quad (6.6)$$

Cela entraînera que presque sûrement, pour  $n$  assez grand, les valeurs propres  $\lambda_i$  de  $A$  satisfont :

$$\lambda_i \in \mathcal{S}_\infty^{c_n} + \left( -\frac{\varepsilon}{2}, \frac{\varepsilon}{2} \right) ,$$

autrement, on aurait  $\mathrm{Tr} \psi_n \left( \frac{\sigma^2}{n} XX^* \right) = \sum_i \psi_n(\lambda_i) \geq 1$ .

On commence par montrer que

$$\mathbb{E} \mathrm{Tr} \psi_n \left( \frac{\sigma^2}{n} XX^* \right) = O \left( \frac{1}{N} \right) . \quad (6.7)$$

En utilisant la proposition 88 appliquée à la fonction à support compact  $\varphi_n$ , il vient

$$\begin{aligned} \mathbb{E} \mathrm{Tr} \varphi_n \left( \frac{\sigma^2}{n} XX^* \right) &= N \int \varphi_n(x) \mathbb{P}_{\mathrm{MP}}^{c_n}(dx) + O \left( \frac{1}{N} \right) \\ &= N + O \left( \frac{1}{N} \right) . \end{aligned}$$

Soit finalement

$$\mathbb{E} \mathrm{Tr} \psi_n \left( \frac{\sigma^2}{n} XX^* \right) = N - \mathbb{E} \mathrm{Tr} \varphi_n \left( \frac{\sigma^2}{n} XX^* \right) = O \left( \frac{1}{N} \right) ,$$

ce qui établit (6.7). Évaluons maintenant la variance de  $\text{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right)$  à l'aide de la proposition 89. On a

$$\begin{aligned} \text{var} \left( \text{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right) \right) &= \text{var} \left( \text{Tr} \varphi_n \left( \frac{\sigma^2}{n} X X^* \right) \right) \\ &\leq \frac{2\sigma^4}{n^2} \mathbb{E} \text{Tr} \varphi'_n(A) X X^* \varphi'_n(A) \\ &= \frac{2\sigma^2}{n} \mathbb{E} \text{Tr} \varphi'_n(A) A \varphi'_n(A) \\ &= \frac{2\sigma^2}{n} \mathbb{E} \text{Tr} \xi_n(A) \end{aligned}$$

avec  $\xi_n(x) = [\varphi'_n(x)]^2 x$ , fonction  $C^\infty$  à support compact et valant 0 sur  $\mathcal{S}_\infty^{c_n}$ . Il suffit à nouveau d'appliquer la proposition 88 pour conclure :

$$\frac{1}{N} \mathbb{E} \text{Tr} \xi_n(A) = \int \xi_n(x) \mathbb{P}_{\text{MP}}^{c_n}(dx) + O\left(\frac{1}{n^2}\right) = O\left(\frac{1}{n^2}\right).$$

Soit finalement

$$\text{var} \left( \text{Tr} \psi_n \left( \frac{\sigma^2}{n} X X^* \right) \right) = O\left(\frac{1}{n^2}\right).$$

En appliquant maintenant le lemme de Borel-Cantelli et (6.7), on obtient (6.6), ce qui conclut la preuve du théorème.  $\square$

# Chapitre 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 5.2.2

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

où  $K$  est un nombre fixé, les  $(\theta_\ell)$  sont des nombre positifs et les  $(\mathbf{u}_\ell)$  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,  $(\theta_\ell)$  et  $(\mathbf{u}_\ell)$  influent sur cette convergence.

L'étude de ce type de modèles a été motivée par I. Johnstone dans son article [16, Section 1.4] où il est fait mention de "spiked models".

On étudiera dans la suite une perturbation de rang 1 :

$$R_N = \sigma^2 (I_N + \theta \mathbf{u} \mathbf{u}^*), \quad \|\mathbf{u}\| = 1. \quad (7.1)$$

### 7.1 Etude de la plus grande valeur propre

**Théorème 91** (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  $\lambda_{\max}$  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 92** (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 (7.1), 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  $\lambda_{\max}$  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  $\lambda_{\max}$  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].

**Remarque 93.** *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 94.** *Le résultat précédent illustre le fait que l'intensité de la perturbation  $\theta$  a une influence sur le comportement de la plus grande valeur propre  $\lambda_{\max}$  : 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 95.** *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.*

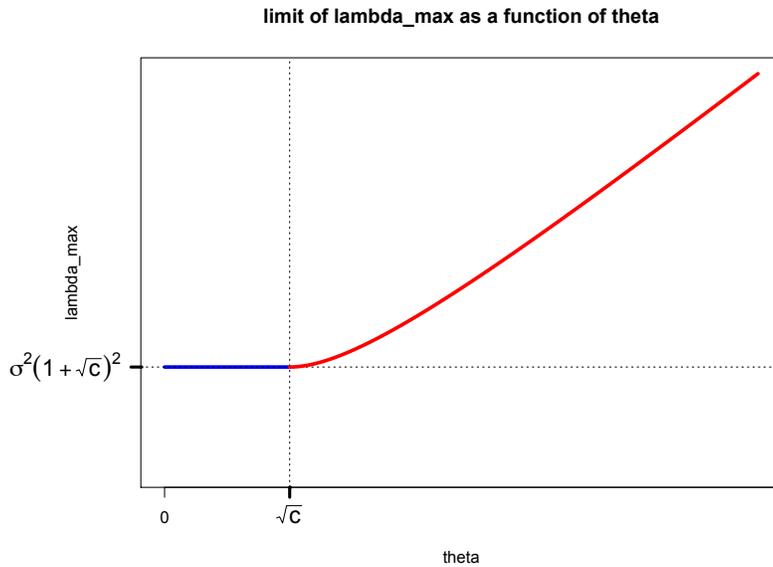


FIGURE 7.1 – Limite de la plus grande valeur propre  $\lambda_{\max}$  en tant que fonction de la perturbation  $\theta$

## 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 valeur propre se détache du support de la loi 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 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  $\lambda_{\max}$  se sépare du support MP, et en calculant cette limite.

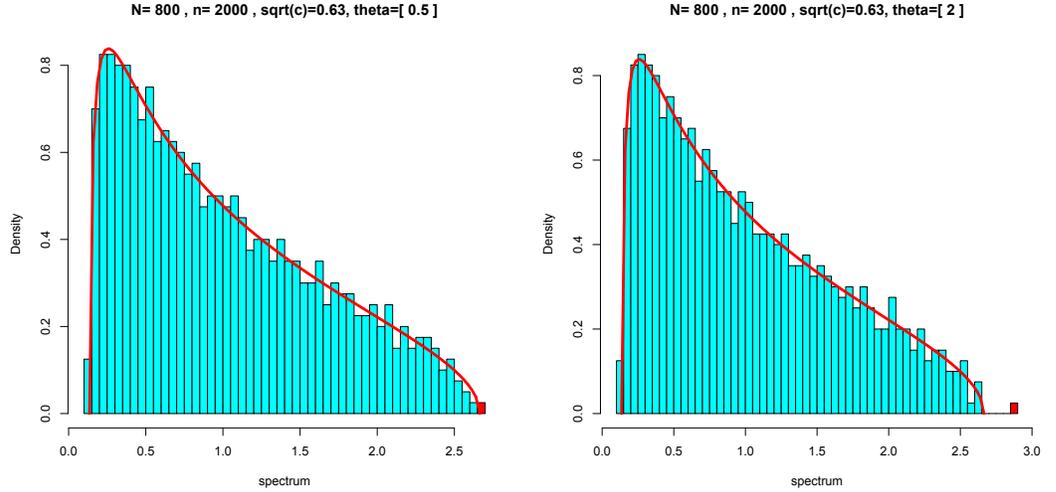


FIGURE 7.2 – Influence de la perturbation  $\theta$  sur le comportement de la plus grande valeur propre. Dans les deux figures,  $c = 0.63$ ; à gauche,  $\theta = 0.5$ , à droite  $\theta = 2$ . Le carré rouge correspond à la plus grande valeur propre; la ligne rouge correspond à la densité de Marčenko-Pastur associée à la valeur de  $c$  (ici,  $\theta = 1$ ).

**Notations.** On utilisera les notations suivantes, pour le modèle  $\check{M}P$  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) :

$$\Pi_N^{-1} = (I_N + \theta \mathbf{u} \mathbf{u}^*)^{-1} = I_N - \frac{\theta}{1 + \theta} \mathbf{u} \mathbf{u}^*.$$

**Condition du déterminant.** On veut trouver  $\lambda^\theta$  valeur propre du modèle perturbé  $\tilde{Z}_N$ , telle que  $\lambda^\theta$  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 :

$$\begin{aligned}
\det\left(-\lambda^\theta I_N + \tilde{Z}_N\right) = 0 &\Leftrightarrow \det\left(-\lambda^\theta I_N + \Pi_N^{1/2} Z_N \Pi_N^{1/2}\right) = 0 \\
&\Leftrightarrow \det\left(-\lambda^\theta \Pi^{-1} + Z_N\right) = 0 \\
&\Leftrightarrow \det\left(-\lambda^\theta \left(I_N - \frac{\theta}{1+\theta} \mathbf{u}\mathbf{u}^*\right) + Z_N\right) = 0 \\
&\Leftrightarrow \det\left(-\lambda^\theta I_N + Z_N + \lambda^\theta \frac{\theta}{1+\theta} \mathbf{u}\mathbf{u}^*\right) = 0 \\
&\Leftrightarrow \det\left[(-\lambda^\theta I_N + Z_N) \left(I_N + \lambda^\theta \frac{\theta}{1+\theta} \mathbf{u}\mathbf{u}^* Q_N(\lambda^\theta)\right)\right] = 0 \\
&\Leftrightarrow \det\left[I_N + \lambda^\theta \frac{\theta}{1+\theta} \mathbf{u}\mathbf{u}^* Q_N(\lambda^\theta)\right] = 0 \tag{7.2}
\end{aligned}$$

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  $\theta$  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 (7.2)), 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 (7.2) se réexprime de la manière suivante :

$$\text{Tr}\left[\lambda^\theta \frac{\theta}{1+\theta} \mathbf{u}\mathbf{u}^* Q_N(\lambda^\theta)\right] = -1 \Leftrightarrow \mathbf{u}^* Q_N(\lambda^\theta) \mathbf{u} = -\frac{1+\theta}{\theta}.$$

**Condition asymptotique.** En utilisant maintenant le théorème de Marčenko-Pastur isotrope (théorème 52), 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 :

$$g_{\text{MP}}(\lambda^\theta) = -\frac{1+\theta}{\theta}.$$

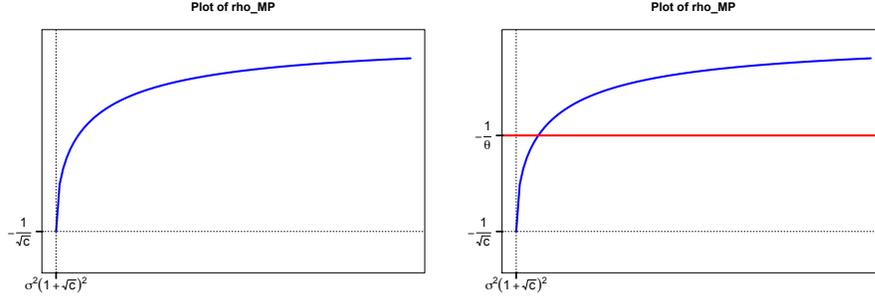
Introduisons la fonction auxiliaire  $\rho_{\text{MP}}(x) = 1 + x g_{\text{MP}}(x)$ ; la condition précédente devient :

$$\rho_{\text{MP}}(\lambda^\theta) = -\frac{1}{\theta}. \tag{7.3}$$

En utilisant le résultat du théorème 16-(3), on obtient une expression explicite pour la fonction  $\rho_{\text{MP}}$ , donnée par :

$$\rho_{\text{MP}}(x) = 1 + \frac{1}{2\sigma^2 c} \left\{ (\sigma^2(1-c) - x) + \sqrt{(x - \lambda^+)(x - \lambda^-)} \right\},$$

pour  $x > \sigma^2(1 + \sqrt{c})^2$ .

FIGURE 7.3 – 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 (7.3) 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  $\lambda^\theta$ .

$$\rho_{\check{\text{MP}}}(\lambda^\theta) = -\frac{1}{\theta} \Leftrightarrow \lambda^\theta = \rho_{\check{\text{MP}}}^{-1}\left(-\frac{1}{\theta}\right).$$

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  $\rho_{\check{\text{MP}}}$ ,

$$g_{\check{\text{MP}}}(x) = \frac{1}{\sigma^2(1-c) - x - x\sigma^2cg_{\check{\text{MP}}}(x)}, \quad \rho_{\check{\text{MP}}}(x) = 1 + xg_{\check{\text{MP}}}(x).$$

on obtient

$$x = \frac{\sigma^2}{\rho_{\check{\text{MP}}}(x)} (\rho_{\check{\text{MP}}}(x) - 1) (1 - c\rho_{\check{\text{MP}}}(x)).$$

Remplaçons maintenant  $x = \rho_{\check{\text{MP}}}^{-1}\left(-\frac{1}{\theta}\right)$  dans les équations. On obtient

$$\lambda^\theta = \rho_{\check{\text{MP}}}^{-1}\left(-\frac{1}{\theta}\right) = \sigma^2(1 + \theta) \left(1 + \frac{c}{\theta}\right),$$

ce qui est le résultat souhaité.

**7.3 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 :

$$(\text{RatioCut}) \operatorname{argmin}_{\mathcal{C}_1 \cup \dots \cup \mathcal{C}_k = \{x_1, \dots, x_n\}} \sum_{a=1}^k \sum_{\substack{i \in \mathcal{C}_a \\ j \notin \mathcal{C}_a}} \frac{\kappa(x_i, x_j)}{|\mathcal{C}_a|}.$$

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

$$(\text{RatioCut}) \operatorname{argmin}_{M \in \mathcal{M}} \operatorname{Tr} M^*(D - K)M$$

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

$$(\text{RelaxedRatioCut}) \operatorname{argmin}_{\substack{M \in \mathbb{R}^{n \times k} \\ M^*M = I_k}} \operatorname{Tr} M^*(D - K)M$$

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*  $\kappa(x, y)$ .

### 7.3.1 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  $a = 1, \dots, k$ ,

$$\limsup_n \max_{1 \leq a \leq k} \|C_a^\circ\| < \infty \text{ et } \limsup_n \max_{1 \leq a \leq k} \frac{1}{\sqrt{p}} \text{Tr } C_a^\circ < \infty.$$

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.

### 7.3.2 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  $x_j \in \mathcal{C}_b$ ,

$$\begin{aligned} \frac{1}{p} \|x_j - x_i\|^2 &= \|w_j - w_i\|^2 + \frac{1}{p} \|\mu_b - \mu_a\|^2 + \frac{2}{\sqrt{p}} (\mu_b - \mu_a)^* (w_j - w_i) \\ &= \tau + \frac{1}{p} \text{Tr } C_a^o + \frac{1}{p} \text{Tr } C_b^o + \psi_j + \psi_i - 2w_i^* w_j \\ &\quad + \frac{\|\mu_b - \mu_a\|^2}{p} + \frac{2}{\sqrt{p}} (\mu_b - \mu_a)^* (w_j - w_i) \end{aligned}$$

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  $\psi_i$ ) 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  $\tau$ , 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,  $f'(\tau)$ , multipliera
  - la matrice  $\psi 1_n^* + 1_n \psi^*$  où  $\psi \in \mathbb{R}^n$  est le vecteur des  $\psi_i$ . 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  $\mu_a$ ,  $\text{Tr } C_a$  et  $\text{Tr } C_a C_b$  le long des vecteurs canoniques  $j_a = [0, \dots, 0, 1_{|C_a|}, 0, \dots, 0]^*$ .

### 7.3.3 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  $D = \text{diag}(K1_n)$  étant donné le résultat précédent (et donc l'approximation de  $D^\alpha$  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 96.** *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  $C_a$ ,  $P = I_n - \frac{1}{n} 1_n 1_n^*$ ,  $v_i = PW^* \mu_a^\circ$ ,

$\psi = \{\|w_i\|^2 - \mathbb{E}[\|w_i\|^2]\}_{i=1}^n$ , et

$$B = \begin{bmatrix} B_{11} & I_k - \mathbf{1}_k \{c_b\}_{b=1}^k & \left( \frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \right) t \\ I_k - \{c_a\}_{a=1}^k \mathbf{1}_k^* & 0_{k \times k} & 0_{k \times 1} \\ \left( \frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \right) t^* & 0_{1 \times k} & \frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \end{bmatrix} \in \mathbb{R}^{(2k+4) \times (2k+4)}$$

$$B_{11} = M^* M + \left( \frac{5f'(\tau)}{8f(\tau)} - \frac{f''(\tau)}{2f'(\tau)} \right) tt^* - \frac{f''(\tau)}{f'(\tau)} T + c \frac{f(0) - f(\tau) + \tau f'(\tau)}{2f'(\tau)} \mathbf{1}_k \mathbf{1}_k^*.$$

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.

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.

### 7.3.4 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



FIGURE 7.4 – Echantillons de la base MNIST.

(images de dimension  $28 \times 28$ ) des chiffres 0, 1 et 2, voir Figure 7.4. 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 7.5 et 7.6.

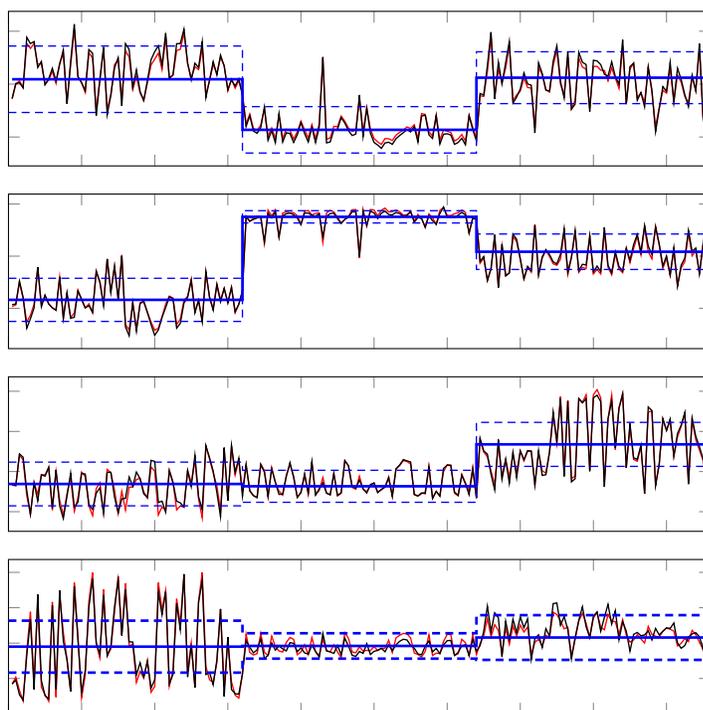


FIGURE 7.5 – Quatre vecteurs propres dominants de  $L$  (rouge), de son approximation asymptotique  $\hat{L}$  (noir) versus les moyennes et écart-types théoriques ( $1\sigma$  et  $2\sigma$ ) (bleu); base MNIST.

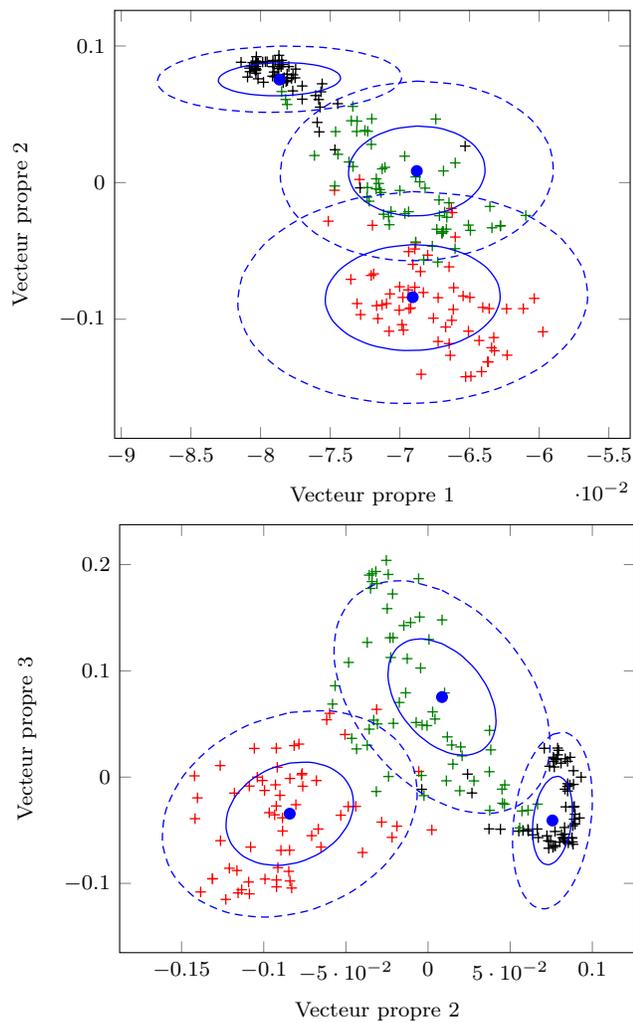


FIGURE 7.6 – Représentation 2D des vecteurs propres 1 et 2 (figure supérieure), et 2 et 3 (figure inférieure) de  $L$ , pour la base MNIST. Moyenne et écarts-types théoriques en bleu.

**Annexe A**

**Rappels d'analyse complexe**



# Annexe B

## Inégalités classiques

### B.1 Démonstration de l'inégalité d'Efron-Stein

### B.2 Inégalité de Poincaré gaussienne

#### B.2.1 Démonstration du théorème 40

**Réduction au cas d'une fonction  $C^2$  à support compact.** On note  $\nabla f$  le gradient de la fonction  $f$ . Dans le cas où  $\mathbb{E}|\nabla f(\mathbf{x})|^2 = \infty$ , l'inégalité est triviale ; dans le cas contraire, il suffit, par approximation, de démontrer l'inégalité pour une fonction  $f \in C^2(\mathbb{R}^n)$  à support compact. C'est ce qu'on suppose dorénavant.

**Réduction au cas unidimensionnel.** Par application de l'inégalité d'Efron-Stein, on a

$$\text{var } f(\mathbf{x}) \leq \frac{1}{2} \sum_i \mathbb{E} |f - f'_i|^2 .$$

En remarquant que

$$\mathbb{E} |f - f'_i|^2 = \mathbb{E} |f - \mathbb{E}_{\{i\}} f - (f'_i - \mathbb{E}_{\{i\}} f'_i)|^2 = \mathbb{E} \text{var}_{\{i\}} (f - f'_i) = 2\mathbb{E} \text{var}_{\{i\}} (f) ,$$

il suffit de démontrer

$$\text{var } f(\mathbf{x}) \leq \sum_i \mathbb{E} \text{var}_{\{i\}} (f) .$$

Nous allons montrer que  $\text{var}_{\{i\}} (f) \leq \mathbb{E}_i \left| \frac{\partial f}{\partial x_i} \right|^2$ . Comme on conditionne par rapport aux variables  $x_k, (k \neq i)$ , cela revient au même de démontrer que

$$\text{var } f(X) \leq \mathbb{E} |f'(X)|^2$$

pour  $X \sim \mathcal{N}(0, 1)$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction  $C^2$  à support compact, et  $\mathbb{E} |f'(X)|^2 < \infty$ .

**Démonstration du cas unidimensionnel.** On rappelle l'identité suivante,

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

valable pour  $X \sim \mathcal{N}(0, 1)$  sous l'hypothèse  $\mathbb{E} |f'(X)|^2 < \infty$  et qui s'obtient facilement par une intégration par parties.

Soit  $Y \sim \mathcal{N}(0, 1)$  une variable aléatoire indépendante de  $X$ , on note

$$X_t = \sqrt{t}X + \sqrt{1-t}Y .$$

On remarque que  $X_t \sim \mathcal{N}(0, 1)$ . alors

$$\begin{aligned} \text{var } f(X) &= \mathbb{E} f(X)f(X) - \mathbb{E} f(X)f(Y) = \mathbb{E} f(X) \int_0^1 \frac{d}{dt} f(X_t) dt \\ &= \int_0^1 \mathbb{E} f(X) \left\{ \frac{X}{2\sqrt{t}} - \frac{Y}{2\sqrt{1-t}} \right\} f'(X_t) dt \\ &= \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E} \frac{\partial}{\partial X} [f(X)f'(X_t)] dt - \int_0^1 \frac{1}{2\sqrt{1-t}} \mathbb{E} f(X) \frac{\partial}{\partial Y} [f'(X_t)] dt \\ &= \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E} f'(X)f'(X_t) dt + \int_0^1 \frac{1}{2} \mathbb{E} f(X)f''(X_t) dt - \int_0^1 \frac{1}{2} \mathbb{E} f(X)f''(X_t) dt \\ &\leq \int_0^1 \frac{1}{2\sqrt{t}} (\mathbb{E} |f'(X)|^2)^{1/2} (\mathbb{E} |f'(X_t)|^2)^{1/2} dt \end{aligned}$$

Comme  $X_t$  est une gaussienne centrée réduite, l'espérance  $\mathbb{E} |f'(X_t)|^2$  est égale à  $\mathbb{E} |f'(X)|^2$  et ne dépend pas de  $t$ . Soit finalement,

$$\text{var } f(X) \leq \mathbb{E} |f'(X)|^2 \int_0^1 \frac{dt}{2\sqrt{t}} = \mathbb{E} |f'(X)|^2 .$$

L'inégalité est établie dans ce cas. Le cas multidimensionnel s'en déduit.

## B.2.2 Démonstration du corollaire 41

Pour montrer le premier point, on représente le vecteur  $\mathbf{x}$  sous la forme  $\mathbf{x} = R^{1/2}\mathbf{y}$  avec  $\mathbf{y} \sim \mathcal{N}_n(0, 1)$ . Alors  $f(\mathbf{x}) = f(R^{1/2}\mathbf{y}) := g(\mathbf{y})$ . La composée de différentielles donne

$$g(\mathbf{y} + \mathbf{h}) - g(\mathbf{y}) = f(R^{1/2}(\mathbf{y} + \mathbf{h})) - f(R^{1/2}\mathbf{y}) = \nabla f \cdot R^{1/2}\mathbf{h} + o(R^{1/2}\mathbf{h}) ,$$

soit

$$\frac{\partial g(\mathbf{y})}{\partial y_k} = \sum_j R_{jk}^{1/2} \frac{\partial f(\mathbf{x})}{\partial x_j} ,$$

et

$$\mathbb{E} \left| \frac{\partial g(\mathbf{y})}{\partial y_k} \right|^2 = \sum_{ij} R_{ik}^{1/2} R_{jk}^{1/2} \mathbb{E} \frac{\partial f(\mathbf{x})}{\partial x_i} \frac{\partial f(\mathbf{x})}{\partial x_j} .$$

Il suffit alors de sommer sur  $k$  et de remarquer que  $R_{ij} = \sum_k R_{ik}^{1/2} R_{jk}^{1/2}$  (car  $R$  est symétrique) pour conclure.

Montrons le second point. Soit  $\Phi$  fonction complexe de parties réelle  $\Phi^{\text{Re}}$  et imaginaire  $\Phi^{\text{Im}}$ , alors

$$\text{var } \Phi(\mathbf{x}) = \mathbb{E} |\Phi(\mathbf{x}) - \mathbb{E} \Phi(\mathbf{x})|^2 = \text{var } \Phi^{\text{Re}}(\mathbf{x}) + \text{var } \Phi^{\text{Im}}(\mathbf{x}).$$

La fonction réelle  $\Phi^{\text{Re}}(\mathbf{x}) = \Phi^{\text{Re}}(\text{Re}(\mathbf{x}), \text{Im}(\mathbf{x}))$  est une fonction du vecteur gaussien réel  $(\text{Re}(\mathbf{x}), \text{Im}(\mathbf{x}))^t \sim \mathcal{N}_{2n}(0, I/2)$ . En lui appliquant la première partie de la proposition, il vient

$$\text{var}(\Phi^{\text{Re}}(\mathbf{x})) \leq \frac{1}{2} \sum_i \left\{ \mathbb{E} \left| \frac{\partial \Phi^{\text{Re}}}{\partial \text{Re}(x_i)} \right|^2 + \mathbb{E} \left| \frac{\partial \Phi^{\text{Re}}}{\partial \text{Im}(x_i)} \right|^2 \right\}.$$

Un calcul similaire pour  $\Phi^{\text{Im}}$  entraîne finalement que

$$\begin{aligned} & \text{var}(\Phi(\mathbf{x})) \\ & \leq \frac{1}{2} \sum_i \left\{ \mathbb{E} \left| \frac{\partial \Phi^{\text{Re}}}{\partial \text{Re}(x_i)} \right|^2 + \mathbb{E} \left| \frac{\partial \Phi^{\text{Re}}}{\partial \text{Im}(x_i)} \right|^2 + \mathbb{E} \left| \frac{\partial \Phi^{\text{Im}}}{\partial \text{Re}(x_i)} \right|^2 + \mathbb{E} \left| \frac{\partial \Phi^{\text{Im}}}{\partial \text{Im}(x_i)} \right|^2 \right\} \\ & = \frac{1}{2} \sum_i \left\{ \mathbb{E} \left| \frac{\partial \Phi}{\partial \text{Re}(x_i)} \right|^2 + \mathbb{E} \left| \frac{\partial \Phi}{\partial \text{Im}(x_i)} \right|^2 \right\}. \end{aligned} \quad (\text{B.1})$$

Calculons maintenant les dérivées partielles de  $\Phi$  par rapport à  $x_i$  et  $\bar{x}_i$ . Par définition,

$$\begin{aligned} \left| \frac{\partial \Phi}{\partial x_i} \right|^2 &= \left| \frac{1}{2} \left( \frac{\partial \Phi}{\partial \text{Re}(x_i)} - i \frac{\partial \Phi}{\partial \text{Im}(x_i)} \right) \right|^2 \\ &= \frac{1}{4} \left( \frac{\partial \Phi}{\partial \text{Re}(x_i)} - i \frac{\partial \Phi}{\partial \text{Im}(x_i)} \right) \overline{\left( \frac{\partial \Phi}{\partial \text{Re}(x_i)} - i \frac{\partial \Phi}{\partial \text{Im}(x_i)} \right)} \\ &= \frac{1}{4} \left| \frac{\partial \Phi}{\partial \text{Re}(x_i)} \right|^2 + \frac{1}{4} \left| \frac{\partial \Phi}{\partial \text{Im}(x_i)} \right|^2 + \frac{1}{2} \text{Im} \left( \frac{\partial \Phi}{\partial \text{Re}(x_i)} \overline{\frac{\partial \Phi}{\partial \text{Im}(x_i)}} \right) \end{aligned}$$

Un calcul similaire donne

$$\left| \frac{\partial \Phi}{\partial \bar{x}_i} \right|^2 = \frac{1}{4} \left| \frac{\partial \Phi}{\partial \text{Re}(x_i)} \right|^2 + \frac{1}{4} \left| \frac{\partial \Phi}{\partial \text{Im}(x_i)} \right|^2 - \frac{1}{2} \text{Im} \left( \frac{\partial \Phi}{\partial \text{Re}(x_i)} \overline{\frac{\partial \Phi}{\partial \text{Im}(x_i)}} \right)$$

Soit finalement

$$\left| \frac{\partial \Phi}{\partial x_i} \right|^2 + \left| \frac{\partial \Phi}{\partial \bar{x}_i} \right|^2 = \frac{1}{2} \left\{ \left| \frac{\partial \Phi}{\partial \text{Re}(x_i)} \right|^2 + \left| \frac{\partial \Phi}{\partial \text{Im}(x_i)} \right|^2 \right\}$$

Il suffit maintenant de reporter cette identité dans (B.1) pour conclure.



# Annexe C

## Compléments

### C.1 Identification d'une transformée de Stieltjes

On démontre dans cette section le théorème 10 ; preuve d'après [26, Theorem B.3].

Etant donnés  $\varepsilon, r > 0$  tels que  $0 < \varepsilon < r$ , on introduit les ensembles suivants :

$$\begin{aligned}\Gamma_{\varepsilon,r} &= \{z \in \mathbb{C}^+, \operatorname{Im}(z) = \varepsilon, |z| \leq r\}, \\ \Gamma'_{\varepsilon,r} &= \{z \in \mathbb{C}^+, \operatorname{Im}(z) \geq \varepsilon, |z| = r\}.\end{aligned}$$

On définit le contour  $\Gamma$  comme l'union des deux contours précédents. [\[figure à inclure\]](#)

Soit  $z = x + iy \in \mathbb{C}^+$  tel que  $\operatorname{Im}(z) > \varepsilon$  et  $|z| < r$  et tel que  $\bar{z} + 2i\varepsilon$  soit à l'extérieur du contour  $\Gamma_r$ . Alors la formule de Cauchy entraîne que :

$$\begin{aligned}g(z) &= \frac{1}{2i\pi} \int_{\Gamma} \frac{g(u)}{u-z} du, \\ &= \frac{1}{2i\pi} \int_{\Gamma} \left\{ \frac{1}{u-z} - \frac{1}{u-(\bar{z}+2i\varepsilon)} \right\} g(u) du, \\ &= \frac{1}{\pi} \int_{\Gamma} \frac{y-\varepsilon}{(u-z)(u-\bar{z}-2i\varepsilon)} g(u) du.\end{aligned}$$

On souhaite montrer que

$$\lim_{r \rightarrow \infty} \frac{1}{\pi} \int_{\Gamma'_{\varepsilon,r}} \frac{y-\varepsilon}{(u-z)(u-\bar{z}-2i\varepsilon)} g(u) du = 0. \quad (\text{C.1})$$

Si  $u \in \Gamma'_{\varepsilon,r}$ , alors  $u = re^{i\theta}$  pour  $\theta \in (0, \pi)$ ,

$$|g(u)| \leq \frac{M}{\varepsilon}$$

et

$$\begin{aligned} |(u-z)(u-\bar{z}-2i\varepsilon)| &\geq |(u-z)(u-\bar{z})| - 2\varepsilon r - 2\varepsilon|z|, \\ &\geq \frac{r^2}{2} - |z|^2 - 2\varepsilon r - 2\varepsilon|z|. \end{aligned}$$

Soit finalement

$$\begin{aligned} \left| \frac{1}{\pi} \int_{\Gamma'_{\varepsilon,r}} \frac{y-\varepsilon}{(u-z)(u-\bar{z}-2i\varepsilon)} g(u) du \right| &\leq \frac{M(y-\varepsilon)}{\pi\varepsilon} \int_0^\pi \frac{r}{\frac{r^2}{2} - |z|^2 - 2\varepsilon r - 2\varepsilon|z|} d\theta \\ &\leq \frac{rM(y-\varepsilon)}{\varepsilon(\frac{r^2}{2} - |z|^2 - 2\varepsilon r - 2\varepsilon|z|)} \xrightarrow{r \rightarrow \infty} 0, \end{aligned}$$

ce qui établit (C.1). En notant  $\Gamma_{\varepsilon,\infty} = \{z \in \mathbb{C}^+, \text{Im}(z) = \varepsilon\}$ , on obtient finalement la représentation suivante de  $g(z)$  :

$$\begin{aligned} g(z) &= \frac{1}{\pi} \int_{\Gamma_{\varepsilon,\infty}} \frac{y-\varepsilon}{(u-z)(u-\bar{z}-2i\varepsilon)} g(u) du, \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y-\varepsilon}{(t+i\varepsilon-z)(t-i\varepsilon-\bar{z})} g(t+i\varepsilon) dt, \quad (t+i\varepsilon = u). \end{aligned}$$

Introduisons maintenant  $v(z) = \text{Im}(g(z))$ , positive par hypothèse, alors

$$v(z) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y-\varepsilon}{(t-x)^2 + (\varepsilon-y)^2} v(t+i\varepsilon) dt.$$

On a  $|yv(x+iy)| \leq |yg(x+iy)| \leq M$ . On déduit de la représentation précédente que

$$\begin{aligned} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(y-\varepsilon)^2}{(t-x)^2 + (\varepsilon-y)^2} v(t+i\varepsilon) dt &= (y-\varepsilon)v(z) \\ &\leq yv(z) \leq M \end{aligned}$$

En faisant tendre  $y$  vers  $+\infty$  et en appliquant le lemme de Fatou, on en déduit que

$$\frac{1}{\pi} \int_{-\infty}^{\infty} v(t+i\varepsilon) dt \leq M.$$

En particulier, les mesures positives de densité  $t \mapsto \frac{1}{\pi} v(t+i\varepsilon)$  sont de masses totales finies uniformément bornées par  $M$ . Ce point nous sera utile par la suite.

Montrons que

$$v(z) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y}{(t-x)^2 + y^2} v(t+i\varepsilon) dt. \quad (\text{C.2})$$

On a

$$\begin{aligned} \left| \frac{y-\varepsilon}{(t-x)^2 + (y-\varepsilon)^2} - \frac{y}{(t-x)^2 + y^2} \right| &\leq \frac{\varepsilon}{(t-x)^2 + y^2/2} + \frac{y(y^2 - (y-\varepsilon)^2)}{((t-x)^2 + y^2/2)^2} \\ &\leq \frac{2\varepsilon}{y^2} + \frac{8\varepsilon}{y^2} \end{aligned}$$

Soit finalement

$$\begin{aligned} & \left| \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y}{(t-x)^2 + y^2} v(t + i\varepsilon) dt - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y}{(t-x)^2 + y^2} v(t + i\varepsilon) dt \right| \\ & \leq \frac{10\varepsilon}{\pi y^2} \int_{-\infty}^{\infty} v(t + i\varepsilon) dt \leq \frac{10M\varepsilon}{\pi y^2} \xrightarrow{\varepsilon \searrow 0} 0 . \end{aligned}$$

ce qui établit (C.2).

Les mesures de densité  $\frac{1}{\pi}v(t + i\varepsilon)$  étant de masse totale uniformément bornées, on peut en extraire par le théorème d'Helly-Bray, une sous-suite qui converge vaguement le long de  $\varepsilon'$  vers une mesure positive de masse totale inférieure ou égale à  $M$ , soit

$$\frac{1}{\pi}v(t + i\varepsilon') dt \xrightarrow{\varepsilon' \rightarrow 0} \mathcal{V}(dt) .$$

Comme de plus la fonction  $t \mapsto \frac{y}{(t-x)^2 + y^2}$  tend vers zéro lorsque  $|t| \rightarrow \infty$ , on obtient la convergence suivante :

$$\lim_{\varepsilon' \rightarrow 0} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y}{(t-x)^2 + y^2} v(t + i\varepsilon') dt = \int_{-\infty}^{\infty} \frac{y}{(t-x)^2 + y^2} \mathcal{V}(dt) .$$

Soit finalement

$$v(z) = \int_{-\infty}^{\infty} \frac{y}{(t-x)^2 + y^2} \mathcal{V}(dt) = \operatorname{Im} \left( \int_{-\infty}^{\infty} \frac{\mathcal{V}(dt)}{t-z} \right) .$$

Les deux fonctions analytiques sur  $\mathbb{C}^+$ ,  $g(z)$  et  $z \mapsto \int_{-\infty}^{\infty} \frac{\mathcal{V}(dt)}{t-z}$ , ont la même partie imaginaire. Les équations de Cauchy-Riemann entraînent alors que

$$g(z) = \int_{-\infty}^{\infty} \frac{\mathcal{V}(dt)}{t-z} + C$$

où  $C$  est une constante à déterminer. Comme d'autre part  $|\operatorname{Im}(z)g(z)|$  et  $|\operatorname{Im}(z) \int_{-\infty}^{\infty} \frac{\mathcal{V}(dt)}{t-z}|$  sont majorées par  $M$  sur  $\mathbb{C}^+$ , on aura  $|\operatorname{Im}(z)C| \leq 2M$  pour tout  $z \in \mathbb{C}^+$ , ce qui n'est possible que pour  $C = 0$ . Finalement,

$$g(z) = \int_{-\infty}^{\infty} \frac{\mathcal{V}(dt)}{t-z} ,$$

ce qui achève la démonstration du théorème.



# Bibliographie

- [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.
- [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] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.*, 17(3) :875–930, 2007.
- [15] 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.
- [16] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2) :295–327, 2001.
- [17] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
- [18] 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.
- [19] 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.
- [20] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.
- [21] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [22] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2) :331–339, 1995.
- [23] 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.
- [24] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.
- [25] E. Telatar. Capacity of multi-antenna Gaussian channels. *European transactions on telecommunications*, 10(6) :585–595, 1998.
- [26] 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 Szücs.
- [27] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62 :548–564, 1955.
- [28] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67 :325–327, 1958.
- [29] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A((1-2)) :32–52, 1928.