

Matrices aléatoires et applications au traitement statistique du signal

Synthèse des résultats en français

Pascal Vallet

17 janvier 2012

1 Introduction

Matrices aléatoires et applications

Depuis le début du XX^{ème} siècle et l'apparition des statistiques multivariées, beaucoup de travaux ont été consacrés à la mise au point de nouveaux outils statistiques pour traiter de grandes quantités de données. Traditionnellement, ces outils ont été développés dans le cas où l'on dispose d'un grand nombre N d'observations statistiques (ou échantillons) de dimension raisonnable M , et se basent relativement souvent sur l'utilisation de la matrice de covariance empirique des observations. Si les N observations statistiques sont représentées par les vecteurs M -dimensionnels $\mathbf{y}_1, \dots, \mathbf{y}_N$, alors la matrice de covariance empirique est donnée par

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*.$$

Le comportement des matrices de covariance empirique a donné également lieu à de nombreux travaux, notamment sur la caractérisation de la loi jointe des valeurs propres et vecteurs propres. Le comportement asymptotique de ces matrices a également été étudié, dans le cas où le nombre d'observations disponibles tend vers l'infini, tandis que la dimension des observations est fixe (voir par exemple les références classiques Anderson [2] et Muirhead [24]).

Aujourd'hui, les progrès des différentes technologies ont entraîné une nette augmentation de la dimension des données à traiter (M grand), sous des contraintes de temps toujours plus fortes (limitations sur N). La plupart des outils classiques (estimateurs, tests) ont été étudiés quand le nombre d'échantillons N est bien plus grand que la dimension des observations M . Or, dans beaucoup de domaines, il n'est pas toujours possible de disposer d'un nombre trop grand d'observations, notamment quand les modèles ne sont stationnaires que sur une courte période de temps. Citons notamment comme exemple le domaine des communications multi-antennes où il est crucial de pouvoir estimer les canaux de transmission en un minimum de temps (et donc un minimum d'échantillons), car leurs propriétés statistiques évoluent très vite, et tout en ayant la possibilité d'avoir un grand nombre d'antennes (dimension des observations) qui permet d'augmenter le débit de transmission. Ces contraintes sont également présentes en finance par exemple, où l'on dispose de grands portefeuilles d'actions, mais de peu d'échantillons, car là encore les modèles évoluent très vite. Il apparaît dès lors que les outils statistiques classiques affichent des comportements différents, notamment ceux basés sur l'estimée empirique des matrices de covariance.

Ainsi, il est devenu nécessaire de disposer de nouveaux outils performants dans le cas où la dimension des observations est grande, et du même ordre de grandeur que le nombre d'échantillons disponibles. Ceci passe par une meilleure compréhension du comportement asymptotique de quantités fondamentales telles que les valeurs propres et vecteurs propres des matrices de covariance empirique, quand la dimension et le nombre des observations tendent vers l'infini au même rythme. Ces problèmes trouvent leur réponse grâce à la théorie des grandes matrices aléatoires. Le régime asymptotique considéré est donc celui où $M, N \rightarrow \infty$ de telle sorte que $M/N \rightarrow c > 0$.

Les premières applications des matrices aléatoires aux statistiques remontent à la fin des années 1980, avec les travaux de Girko (voir par exemple [16]), fondateur de la théorie de la "G-estimation" (G pour "Generalized"). Dans le cas de modèles dits "à covariance", les observations $\mathbf{y}_1, \dots, \mathbf{y}_N \in \mathbb{C}^M$ s'écrivent

$$\mathbf{y}_n = \mathbf{R}^{1/2} \mathbf{x}_n,$$

où $\mathbf{x}_1, \dots, \mathbf{x}_n$ sont des vecteurs indépendants et identiquement distribués (i.i.d.), dont les entrées sont i.i.d. centrées et de variance 1, et \mathbf{R} est la matrice de covariance $M \times M$ des observations. L'idée de la "G-estimation" repose sur le fait que la mesure spectrale empirique des valeurs propres de $\hat{\mathbf{R}}$, dont la fonction de répartition est donnée par

$$\hat{F}(\lambda) = \frac{1}{M} \text{card} \{k : \hat{\lambda}_k \leq \lambda\},$$

avec $\hat{\lambda}_1, \dots, \hat{\lambda}_M$ les valeurs propres de $\hat{\mathbf{R}}$, est proche, quand $M, N \rightarrow \infty$ et $M/N \rightarrow c > 0$, d'une distribution déterministe. Plus précisément (Silverstein [28], Girko [16]), on a presque sûrement

$$\hat{F}(\lambda) - F(\lambda) \rightarrow 0, \tag{1}$$

où F est la fonction de répartition d'une mesure de probabilité déterministe, fonction des valeurs propres de \mathbf{R} et du ratio M/N . Girko a développé un certain nombre d'estimateurs pour des fonctionnelles de la matrice de covariance \mathbf{R} , en utilisant la relation existante entre F et ces fonctionnelles de \mathbf{R} , puis en utilisant le fait que F est estimable de façon consistante par \hat{F} . Ces mêmes idées ont été reprises par El Karoui [14] en 2008, dans le cadre de l'estimation des valeurs propres de la matrice \mathbf{R} . Plus récemment en 2009, Mestre [36] a également proposé un nouveau point de vue sur l'estimation des valeurs propres et vecteurs propres de \mathbf{R} , basé sur des représentations en intégrales de contour facilement estimables.

L'application des matrices aléatoires au domaine des télécommunications est relativement récente, malgré l'importance des problèmes. Les premiers travaux concernent la théorie de l'information (Telatar [32] en 1999), où la capacité de certains canaux de communications MIMO (Multiple Inputs Multiple Outputs) est étudiée. Il est ainsi montré que cette capacité peut être interprétée en terme de fonctionnelle des valeurs propres de matrices aléatoires. En effet, la capacité ergodique par antenne d'un canal MIMO gaussien à M antennes de réception et N antennes d'émission est donnée par

$$I(\rho) = \mathbb{E} \left[\frac{1}{M} \log \det \left(I + \rho \frac{\mathbf{H}\mathbf{H}^*}{N} \right) \right],$$

où \mathbf{H} est la matrice $M \times N$ du canal dont les entrées sont i.i.d. gaussiennes complexes standard, ρ est le rapport signal à bruit. Cette capacité s'écrit comme

$$I(\rho) = \mathbb{E} \left[\int_{\mathbb{R}} \log(1 + \rho \lambda) d\hat{F}(\lambda) \right]$$

où $\hat{F}(\lambda)$ est la distribution empirique des valeurs propres de $\mathbf{H}\mathbf{H}^*$. En utilisant (1), il est ainsi possible de montrer que $I(\rho)$ converge vers une quantité constante, dont on connaît l'expression analytique. Ces idées basées sur les fonctionnelles de valeurs propres de matrices aléatoires ont été également reprises dans le contexte des grands systèmes CDMA (Code Division Multiple Access - voir Tse & Hanly [33], Verdu & Shamai [34]), où les matrices aléatoires en jeu ne sont plus gaussiennes. Plus récemment, ces idées ont également été utilisées dans les problèmes d'optimisation de la capacité, où le but est de précoder les signaux transmis de manière à maximiser la capacité de transmission (voir par exemple Chuah et al. [10], Moustakas et al. [23], Dumont et al. [13]).

Dans le contexte du traitement statistique du signal, qui est l'objet principal de cette thèse, les travaux sont peu nombreux. Une première contribution de Silverstein et Combettes [30] pose le problème de la détection de signal dans les grands réseaux de capteurs, quand le nombre d'antennes et le nombre d'échantillons disponibles sont grands et du même ordre de grandeur. Il est ainsi montré expérimentalement que les valeurs propres de la matrice de covariance empirique se séparent en deux groupes, le groupe des plus grandes valeurs propres étant directement relié au nombre de sources émettrices. Ces résultats ont été formalisés plus récemment avec l'introduction des "spiked models", qui sont des matrices aléatoires perturbées multiplicativement

par des matrices de petit rang (voir Baik et al. [5], Baik & Silverstein [6]). Des algorithmes de détection basés sur ce phénomène de séparation des valeurs propres de la covariance empirique ont ainsi été développés, notamment par Nadakuditi & Edelman [25], Kritchman & Nadler [21]. Tout récemment, Bianchi et al. [8] ont proposé une étude des performances du test du maximum de vraisemblance généralisé pour la détection d'une seule source, en développant des résultats de grandes déviations sur la plus grande valeur propre de la matrice de covariance empirique. Parallèlement à la détection de sources, des travaux ont été conduits dans le domaine de la localisation de sources. Ainsi, Mestre et Lagunas [38] ont proposé une méthode d'estimation des directions d'arrivée, basés sur l'algorithme MUSIC, qui présente de bonnes performances pour un nombre d'antennes et d'échantillons du même ordre de grandeur. Les travaux de Mestre-Lagunas constituent le point de départ du travail de cette thèse et sont résumés dans la section suivante.

Quelques résultats classiques et applications aux grands réseaux de capteurs

La théorie des grandes matrices aléatoires est bien antérieure aux problèmes statistiques évoqués précédemment, et trouve ses fondations en physique nucléaire dans les travaux de Wigner [35] de 1958. Wigner a notamment étudié le spectre de certaines matrices aléatoires hermitiennes et montré que la distribution empirique des valeurs propres converge vers la loi du demi-cercle.

Il faut attendre 10 ans plus tard pour voir apparaître les premiers résultats asymptotiques concernant les matrices de covariance empiriques, grâce aux travaux de Marcenko & Pastur [22]. Considérons $M, N \in \mathbb{N}^*$ tels que $M = M(N)$ est une fonction de N , et $c_N = M/N \rightarrow c \in]0, 1[$ quand $N \rightarrow \infty$, ainsi qu'une matrice

$$\mathbf{\Sigma}_N = \frac{\sigma}{\sqrt{N}} \mathbf{X}_N, \quad (2)$$

avec $\sigma > 0$ et \mathbf{X}_N une matrice de taille $M \times N$ dont les éléments sont des variables aléatoires i.i.d. centrés et de variance 1. Comme précédemment, nous définissons la mesure spectrale de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ (où distribution empirique de ses valeurs propres), par la mesure de probabilité aléatoire

$$\hat{\mu}_N = \frac{1}{M} \sum_{k=1}^M \delta_{\hat{\lambda}_{k,N}},$$

avec $\hat{\lambda}_{1,N} \leq \dots \leq \hat{\lambda}_{M,N}$ les valeurs propres de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ et δ_x la mesure de Dirac en x . Marcenko & Pastur ont notamment montré qu'il existe une suite de mesures de probabilité déterministes (ν_N) , telles que

$$\hat{\mu}_N - \nu_N \xrightarrow[N \rightarrow \infty]{e} 0, \quad (3)$$

presque sûrement, où $\xrightarrow[N \rightarrow \infty]{e}$ représente la convergence étroite. La loi ν_N est de plus donnée par

$$d\nu_N(\lambda) = \frac{\sqrt{(\lambda - \lambda_N^-)(\lambda_N^+ - \lambda)}}{2\sigma^2 c_N \pi \lambda} \mathbb{1}_{[\lambda_N^-, \lambda_N^+]}(\lambda) d\lambda, \quad (4)$$

où $\lambda_N^- = \sigma^2(1 - \sqrt{c_N})^2$ et $\lambda_N^+ = \sigma^2(1 + \sqrt{c_N})^2$. Quand $N \rightarrow \infty$, il est clair que ν_N tend vers une loi limite ν_∞ , appelée mesure spectrale limite de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$. Le support de ν_∞ est donné par l'intervalle $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$, communément appelé "bulk" dans la littérature. La convergence (3) signifie que l'histogramme des valeurs propres de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ se concentre autour de la distribution ν_∞ , pour N grand, comme montré dans l'exemple numérique donné en figure 1.

L'approche originale de la preuve de Marcenko & Pastur repose sur l'utilisation de la transformée de Stieltjes $\hat{m}_N(z)$ de $\hat{\mu}_N$ définie par l'intégrale

$$\hat{m}_N(z) = \int_{\mathbb{R}} \frac{d\hat{\mu}_N(\lambda)}{\lambda - z}, \text{ pour tout } z \in \mathbb{C} \setminus \mathbb{R},$$

et consiste à prouver que $\hat{m}_N(z) - \kappa_N(z) \rightarrow_N 0$ p.s., où $\kappa_N(z)$ est la transformée de Stieltjes de ν_N . Cette méthode est relativement puissante dans le sens où elle fait intervenir la résolvante $(\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - z\mathbf{I}_M)^{-1}$ par l'égalité

$$\hat{m}_N(z) = \frac{1}{M} \text{Tr} (\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - z\mathbf{I}_M)^{-1},$$

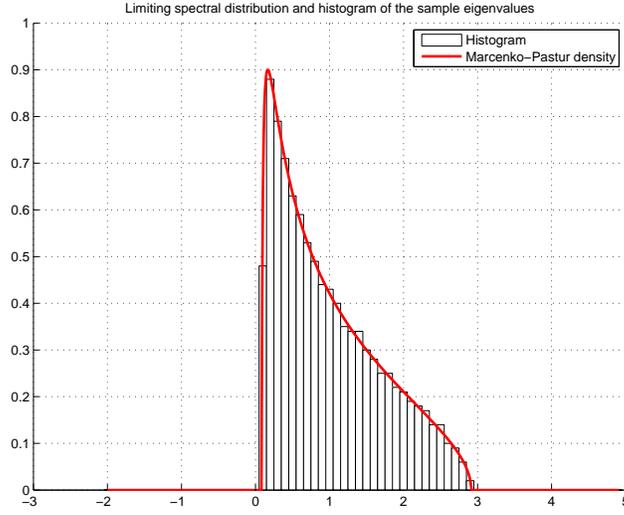


FIGURE 1 – Densité de la loi ν_∞ pour $\sigma = 1$, $c = 0.5$, et histogramme des valeurs propres de $\Sigma_N \Sigma_N^*$ dans le cas où $N = 2000$, $c_N = 0.5$ et $\sigma = 1$, pour une réalisation de la matrice aléatoire \mathbf{X}_N .

grâce à laquelle des calculs peuvent être réalisés aisément. De nombreux résultats ont été obtenus par la suite sur le comportement des valeurs propres de $\Sigma_N \Sigma_N^*$, prises individuellement. Il a ainsi été montré en particulier que (voir Bai & Silverstein [3])

$$\hat{\lambda}_{1,N} \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \sigma^2(1 - \sqrt{c})^2 \text{ et } \hat{\lambda}_{M,N} \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \sigma^2(1 + \sqrt{c})^2.$$

Il s'en suit que les valeurs propres de $\Sigma_N \Sigma_N^*$ sont ainsi toutes "absorbées", à partir d'un certain rang N , dans un voisinage du "bulk" $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$.

Pour comprendre les applications potentielles de ce résultat au traitement du signal, considérons un ensemble de K sources émettant sur un réseau de M antennes, avec $M > K$. A l'instant discret n , le signal reçu s'écrit

$$\mathbf{y}_n = \mathbf{A}\mathbf{s}_n + \mathbf{v}_n,$$

où

- \mathbf{A} est une matrice déterministe complexe de taille $M \times K$ contenant les coefficients de transmission entre les K sources et le réseau d'antennes,
- $\mathbf{s}_n = [s_{1,n}, \dots, s_{n,K}]^T$ le vecteur des signaux transmis (non observable), supposé gaussien centré de covariance $\mathbf{R}_S = \mathbb{E}[\mathbf{s}_n \mathbf{s}_n^*]$. Les signaux sont également supposés i.i.d. dans le domaine temporel.
- \mathbf{v}_n un vecteur gaussien complexe (moyenne nulle, covariance $\sigma^2 \mathbf{I}_M$), représentant le bruit de transmission indépendant du signal source.

En collectant N échantillons du précédent signal dans la matrice $\Sigma_N = \frac{1}{\sqrt{N}}[\mathbf{y}_1, \dots, \mathbf{y}_N]$, on obtient

$$\Sigma_N = \frac{\mathbf{A}\mathbf{S}_N}{\sqrt{N}} + \mathbf{W}_N, \quad (5)$$

où $\mathbf{S}_N = [\mathbf{s}_1, \dots, \mathbf{s}_N]$ et $\mathbf{W}_N = \frac{1}{\sqrt{N}}[\mathbf{v}_1, \dots, \mathbf{v}_N]$. Dans le cas où aucun signal n'émet sur le réseau d'antennes, $\Sigma_N = \mathbf{W}_N$ présente la même distribution que (2), et donc toutes les valeurs propres de la matrice de covariance empirique des observations sont concentrées dans un voisinage de $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$, pour un nombre d'antennes et d'échantillons suffisamment grands. En présence de signal, le modèle (5) est par contre équivalent à

$$\Sigma_N = \mathbf{\Gamma}_N^{1/2} \mathbf{X}_N \quad (6)$$

où $\mathbf{\Gamma}_N = \mathbf{A}\mathbf{R}_S\mathbf{A}^* + \sigma^2\mathbf{I}_M$ et \mathbf{X}_N une matrice $M \times N$ dont les entrées sont complexes gaussiennes standard et indépendantes. Sous l'hypothèse que \mathbf{A} et \mathbf{R}_S sont de rang plein, on a $\text{rang}(\mathbf{A}\mathbf{R}_S\mathbf{A}^*) = K$, et les $M - K$ plus petites

valeurs propres de Γ_N sont toutes égales à σ^2 sauf les K plus grandes qui sont strictement supérieures. En posant $\sigma = 1$ sans perte de généralité, Γ_N est donc une perturbation de rang K de la matrice identité \mathbf{I}_M , et (6) apparaît donc comme une perturbation multiplicative de (2). Il convient de quantifier la perturbation engendrée sur les valeurs propres de $\Sigma_N \Sigma_N^*$, dans un premier temps lorsque K est petit comparativement à M, N . Le comportement des valeurs propres dans un tel cadre est donné par les résultats sur les "spiked models" (voir Baik et al. [5], Baik & Silverstein [6]).

Considérons plus précisément le modèle (6), où Γ_N est une matrice déterministe définie positive de valeurs propres $\lambda_{1,N} \leq \dots \leq \lambda_{M,N}$ et \mathbf{X}_N une matrice dont les entrées sont i.i.d. centrées de variance 1. Supposons comme précédemment que M est une fonction de N et $c_N = M/N \rightarrow c \in]0, 1]$, posons K un entier indépendant de N . On suppose de plus que les $M - K$ plus petites valeurs propres de Γ_N sont égales à 1, i.e. $\lambda_{1,N} = \dots = \lambda_{M-K,N} = 1$, et que les K plus grandes valeurs propres $\lambda_{M-K+1,N}, \dots, \lambda_{M,N}$ convergent vers des limites $1 < \gamma_1 \leq \dots \leq \gamma_K$ quand $N \rightarrow \infty$. Comme Γ_N est une perturbation de l'identité de petit rang K comparé aux dimensions M, N , il est attendu que le comportement global des valeurs propres de $\Sigma_N \Sigma_N^*$ soit proche du cas (2), i.e. que leur distribution empirique soit proche de la loi de Marcenko-Pastur. En effet, si $\hat{\mu}_N$ désigne la mesure spectrale empirique de $\Sigma_N \Sigma_N^*$ dans le cadre du modèle (6), alors on a toujours $\hat{\mu}_N - \nu_N \xrightarrow[N \rightarrow \infty]{e} 0$, où ν_N est donnée par (4) (Silverstein [28]). Néanmoins, quand $N \rightarrow \infty$, toutes les valeurs propres de $\Sigma_N \Sigma_N^*$ ne se concentrent pas autour du bulk $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})]$. En effet, il est montré (voir Baik & Silverstein [6]) que les K plus grandes valeurs propres $\hat{\lambda}_{M-K+1,N}, \dots, \hat{\lambda}_{M,N}$ de $\Sigma_N \Sigma_N^*$ peuvent converger vers des limites en dehors du bulk si $\gamma_1, \dots, \gamma_K$, les valeurs propres limites correspondantes de Γ_N , sont supérieures à un certain seuil. Plus précisément, on a pour $k = 1, \dots, K$,

$$\hat{\lambda}_{M-K+k,N} \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \begin{cases} \gamma_k + \frac{c\gamma_k}{\gamma_k - 1} & \text{si } \gamma_k > 1 + \sqrt{c} \\ (1 + \sqrt{c})^2 & \text{sinon.} \end{cases} \quad (7)$$

Notons que $\gamma_k > 1 + \sqrt{c}$ implique $\gamma_k + \frac{c\gamma_k}{\gamma_k - 1} > (1 + \sqrt{c})^2$. Les résultats énoncés en (7) sont illustrés en figure 2.

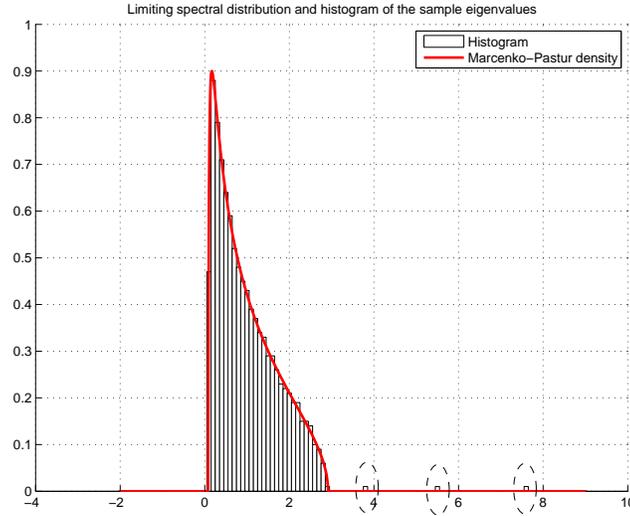


FIGURE 2 – Densité de la loi de Marcenko-Pastur ν_∞ pour $\sigma = 1$, $c = 0.5$, et histogramme des valeurs propres de $\Sigma_N \Sigma_N^*$ dans le cadre du modèle (6). Les paramètres sont $K = 3$, $N = 2000$, $c_N = 0.5$ et $\sigma = 1$, pour une réalisation de la matrice aléatoire \mathbf{X}_N . Les 3 plus grandes valeurs propres de Γ_N sont 3, 5, 7, et les valeurs propres correspondantes de $\Sigma_N \Sigma_N^*$ sont entourées en pointillés.

Les implications dans le cadre applicatif du réseau d'antennes (5) sont immédiates. Si les K valeurs propres non nulles de $\mathbf{A} \mathbf{R}_S \mathbf{A}^*$ associées aux K sources émettrices sont suffisamment grandes, alors pour N, M suffisamment grands, les K plus grandes valeurs propres de $\Sigma_N \Sigma_N^*$ seront séparées des $M - K$ plus petites. Ces phénomènes spécifiques aux "spiked models" ont notamment été exploités dans des problèmes de détection

des sources (voir les travaux de Nadakuditi & Edelman [25], Kritchman & Nadler [21], Nadler [26], Bianchi et al. [8]).

Dans le cas où le nombre de sources n'est plus négligeable comparativement aux nombres d'antennes et d'échantillons, il est nécessaire de considérer un cadre plus général pour étudier le comportement des valeurs propres de $\Sigma_N \Sigma_N^*$. On considère donc le modèle (6) où Γ_N est une matrice définie positive telle que $\sup_N \|\Gamma_N\| < \infty$. Dans ce contexte, il est montré (Silverstein [28]) que la mesure spectrale empirique $\hat{\mu}_N$ vérifie

$$\hat{\mu}_N - \mu_N \xrightarrow[N \rightarrow \infty]{e} 0, \quad (8)$$

presque sûrement, où μ_N est une mesure de probabilité déterministe et absolument continue de support compact (Silverstein & Choi [29]). De manière équivalente,

$$\hat{m}_N(z) - m_N(z) \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0, \text{ pour tout } z \in \mathbb{C}^+, \quad (9)$$

où $\hat{m}_N(z) = \frac{1}{M} \text{Tr} (\Sigma_N \Sigma_N^* - z \mathbf{I}_M)^{-1}$ est la transformée de Stieltjes de $\hat{\mu}_N$ et où $m_N(z) = \int_{\mathbb{R}^+} \frac{d\mu_N(\lambda)}{\lambda - z}$, la transformée de Stieltjes de μ_N , vérifie l'équation

$$m_N(z) = \frac{1}{M} \text{Tr} (\Gamma_N (1 - c_N - c_N z m_N(z)) - z \mathbf{I}_M)^{-1}, \text{ pour tout } z \in \mathbb{C} \setminus \text{supp}(\mu_N), \quad (10)$$

avec $\text{supp}(\mu_N)$ le support de μ_N . La différence essentielle par rapport au modèle (2) réside à présent dans le fait que les différentes quantités décrivant le comportement global asymptotique des valeurs propres de $\Sigma_N \Sigma_N^*$ ne sont plus connues explicitement, mais uniquement par l'intermédiaire d'équations. Ainsi, il ne sera plus possible de déduire des résultats précis concernant le comportement des plus grandes valeurs propres de $\Sigma_N \Sigma_N^*$. Néanmoins, il est tout de même possible de mettre en évidence un phénomène de séparation des valeurs propres. En effet, définissons

$$w_N(z) = \frac{z}{1 - c_N - c_N z m_N(z)}. \quad (11)$$

Alors Bai & Silverstein [4] ont montré que, pour un intervalle $[a, b]$ tel que $[a - \epsilon, b + \epsilon] \cap \text{supp}(\mu_N) = \emptyset$, pour un certain $\epsilon > 0$ et à partir d'un certain rang N , on a $w_N(a) < w_N(b)$ et

$$\text{card}\{k : \hat{\lambda}_{k,N} < a\} = \text{card}\{k : \lambda_{k,N} < w_N(a)\} \text{ et } \text{card}\{k : \hat{\lambda}_{k,N} > b\} = \text{card}\{k : \lambda_{k,N} > w_N(b)\}, \quad (12)$$

avec probabilité 1, pour N suffisamment grand. Autrement dit, si par exemple les valeurs propres de Γ_N sont scindées en deux groupes suffisamment séparés, de telle manière que

$$\lambda_{1,N} \leq \dots \leq \lambda_{M-K,N} < w_N(a) < w_N(b) < \lambda_{M-K+1,N} \leq \dots \leq \lambda_{M,N},$$

où K peut dépendre de N , alors presque sûrement, pour N grand,

$$\hat{\lambda}_{1,N} \leq \dots \leq \hat{\lambda}_{M-K,N} < a < b < \hat{\lambda}_{M-K+1,N} \leq \dots \leq \hat{\lambda}_{M,N}.$$

Ce phénomène de séparation est illustré en figure 3, où l'on constate la séparation des valeurs propres de $\Sigma_N \Sigma_N^*$ en 3 groupes.

Ces résultats généraux sur les modèles de matrices aléatoires à covariance ont été exploités par Mestre [38] dans le cadre de la localisation de sources. Dans ce contexte, il est courant de supposer que la matrice \mathbf{A} du modèle (5) s'écrit $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$, où $\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)$ représentent les vecteurs directionnels linéairement indépendants associés aux K sources émettrices, et θ_k représente l'angle d'arrivée de la k -ième source. En supposant la matrice de covariance des signaux \mathbf{R}_S de rang plein, il est facile de voir que l'espace propre associé à la valeur propre σ^2 de $\Gamma_N = \mathbf{A} \mathbf{R}_S \mathbf{A}^* + \sigma^2 \mathbf{I}_M$ est engendré par les vecteurs $\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)$. En notant $\mathbf{u}_{1,N}, \dots, \mathbf{u}_{M,N}$ les vecteurs propres de Γ_N associés respectivement aux valeurs propres $\lambda_{1,N}, \dots, \lambda_{M,N}$, ceci implique que les angles $\theta_1, \dots, \theta_K$ sont des zéros de la fonction de localisation ¹

$$\eta_N(\theta) \triangleq \sum_{k=1}^{M-K} \mathbf{a}(\theta)^* \mathbf{u}_{k,N} \mathbf{u}_{k,N}^* \mathbf{a}(\theta), \quad (13)$$

1. Ils sont les uniques zéros sous certaines conditions portant sur la fonction $\theta \mapsto \mathbf{a}(\theta)$, par exemple pour le modèle classique $\mathbf{a}(\theta) = [1, e^{i\pi \sin(\theta)}, \dots, e^{i(M-1)\pi \sin(\theta)}]$ correspondant à un réseau d'antennes uniforme et linéaire, dont les capteurs sont espacés de la moitié de la longueur d'onde.

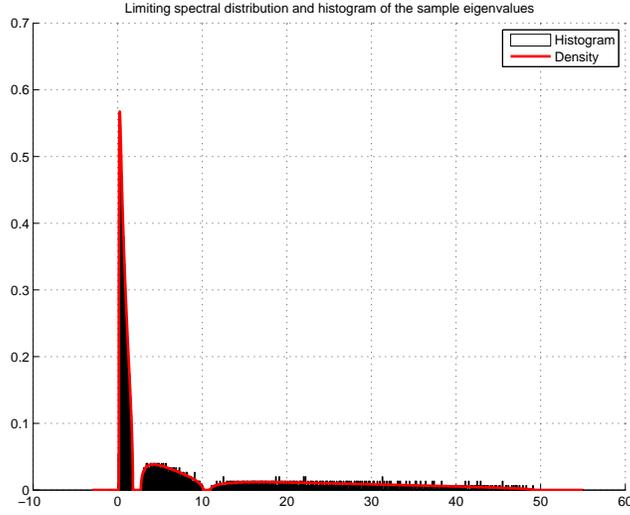


FIGURE 3 – Densité de la loi μ_N pour $\sigma = 1$, $c_N = 0.5$, et histogramme des valeurs propres de $\Sigma_N \Sigma_N^*$ dans le cadre du modèle (6). pour $N = 3000$, une réalisation de la matrice aléatoire \mathbf{X}_N . Les valeurs propres de Γ_N sont 1 (multiplicité 750), 7 (multiplicité 300) et 25 (multiplicité 450).

où $\sum_{k=1}^{M-K} \mathbf{u}_{k,N} \mathbf{u}_{k,N}^*$ est le projecteur sur le sous-espace propre de Γ_N associé à la valeur propre σ^2 . Quand M est indépendant de N , la forme quadratique $\eta_N(\theta)$ peut être estimée de façon consistante par

$$\hat{\eta}_N(\theta) \triangleq \sum_{k=1}^{M-K} \mathbf{a}(\theta)^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{a}(\theta),$$

quand $N \rightarrow \infty$, où $\sum_{k=1}^{M-K} \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^*$ est le projecteur sur le sous-espace propre associé aux $M - K$ plus petites valeurs propres de $\Sigma_N \Sigma_N^*$ (dont les vecteurs propres sont notés $\hat{\mathbf{u}}_{1,N}, \dots, \hat{\mathbf{u}}_{M,N}$). Les angles $\theta_1, \dots, \theta_K$ peuvent alors être estimés en considérant les K plus petits minima de $\theta \mapsto \hat{\eta}_N(\theta)$, c'est l'algorithme d'estimation MUSIC ("MUltiple Signal Classification"), mis au point par Schmidt [27] dans les années 1980. Cet estimateur a fait l'objet d'une littérature foisonnante en traitement statistique du signal, et sa consistance et normalité asymptotique ont en particulier été étudiés (voir Stoica & Nehorai [31]) dans le régime asymptotique où M est fixe et N tend vers l'infini. Dans le cas où M et N tendent vers l'infini au même rythme, $\hat{\eta}_N(\theta)$ n'est plus consistant, car $\left\| \Sigma_N \Sigma_N^* - \left(\frac{\mathbf{A} \Sigma_N \Sigma_N^* \mathbf{A}^*}{N} + \sigma^2 \mathbf{I}_M \right) \right\|$ ne tend plus vers 0. Récemment, Mestre [37] a proposé une approche séduisante pour étudier le cas où $M, N \rightarrow \infty$ et $M/N \rightarrow c \in]0, 1[$, en exploitant les principes de la G-estimation initiée par Girko et les résultats de localisation des valeurs propres (12) de Bai & Silverstein pour les modèles à covariance. L'idée part du principe que la fonction de localisation $\eta_N(\theta)$ peut être exprimée par l'intégrale de contour

$$\eta_N(\theta) = \frac{1}{2\pi i} \oint_{\mathcal{C}} \mathbf{a}(\theta)^* (\Gamma_N - w \mathbf{I}_M)^{-1} \mathbf{a}(\theta) dw, \quad (14)$$

où \mathcal{C} est un contour orienté dans le sens antitrigonométrique, entourant uniquement la valeur propre σ^2 de Γ_N . De plus, sous certaines hypothèses ne dépendant uniquement que du comportement du support de μ_N , il existe un rectangle $\mathcal{R} = \{u + iv : u \in [t^-, t^+], v \in [-\delta, \delta]\}$, avec $\delta, t^-, t^+ > 0$ indépendants de N , pour lesquels $w_N(\partial \mathcal{R})$ est un contour admissible pour \mathcal{C} , i.e. $w_N(\partial \mathcal{R})$ est un contour entourant uniquement la valeur σ^2 de Γ_N et laissant les autres valeurs propres à l'extérieur. Dans ces conditions, le changement de variable $w = w_N(z)$ dans (14) permet d'exprimer $\eta_N(\theta)$ sous la forme

$$\eta_N(\theta) = \frac{1}{2\pi i} \oint_{\partial \mathcal{R}} \mathbf{a}(\theta)^* \mathbf{T}_N(z) \mathbf{a}(\theta) (1 - c_N - c_N z m_N(z)) w'_N(z) dz. \quad (15)$$

où $\mathbf{T}_N(z) = (\Gamma_N (1 - c_N - c_N z m_N(z)) - z \mathbf{I}_M)^{-1}$ est la matrice intervenant dans la trace (10), et $w'_N(z)$ la dérivée de $w_N(z)$ définie en (11). En appliquant en particulier (9), l'intégrande

$$g_N(z) = \mathbf{a}(\theta)^* \mathbf{T}_N(z) \mathbf{a}(\theta) (1 - c_N - c_N z m_N(z)) w'_N(z)$$

peut être approchée par²

$$\hat{g}_N(z) = \mathbf{a}(\theta)^* \mathbf{Q}_N(z) \mathbf{a}(\theta) (1 - c_N - c_N z \hat{m}_N(z)) \hat{w}'_N(z),$$

avec $\mathbf{Q}_N(z) = (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* - z \mathbf{I}_M)^{-1}$ et $\hat{w}'_N(z)$ la dérivée de $\hat{w}_N(z) = \frac{z}{1 - c_N - c_N z \hat{m}_N(z)}$. $\hat{g}_N(z)$ est une fraction rationnelle dont les pôles sont les valeurs propres $\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{M,N}$ de $\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*$ et les zéros $\hat{\omega}_{1,N} < \dots < \hat{\omega}_{M,N}$ de la fonction $z \mapsto 1 - c_N - c_N z \hat{m}_N(z)$, qui sont tous réels et au nombre de M . Il est important de mentionner que les hypothèses formulées sur $\text{supp}(\mu_N)$ impliquent, du fait des propriétés de séparation exacte des valeurs propres de $\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*$, que p.s. pour N suffisamment grand,

$$\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{M-K,N}, \hat{\omega}_{1,N}, \dots, \hat{\omega}_{M-K,N} \in [t^-, t^+] \text{ et } \hat{\lambda}_{M-K+1,N}, \dots, \hat{\lambda}_{M,N}, \hat{\omega}_{M-K+1,N}, \dots, \hat{\omega}_{M,N} > t^+ + \epsilon,$$

pour un certain $\epsilon > 0$. Dans ces conditions, les pôles de $\hat{g}_N(z)$ sont suffisamment éloignés de $\partial \mathcal{R}$, et l'on peut montrer que

$$\left| \int_{\partial \mathcal{R}} (\hat{g}_N(z) - g_N(z)) dz \right| \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0. \quad (16)$$

En notant

$$\hat{\eta}_{\text{rmt},N}(\theta) = \frac{1}{2\pi i} \oint_{\partial \mathcal{R}_1} \hat{g}_N(z) dz, \quad (17)$$

on a donc $\hat{\eta}_{\text{rmt},N}(\theta) - \eta_N(\theta) \rightarrow 0$ p.s., et donc $\hat{\eta}_{\text{rmt},N}(\theta)$ est un estimateur consistant de la fonction de localisation $\eta_{\text{rmt},N}(\theta)$ dans le régime asymptotique $M, N \rightarrow \infty$, $c_N = M/N \rightarrow c \in]0, 1]$. Cette intégrale n'est pas calculable numériquement en pratique, car la connaissance du rectangle \mathcal{R} n'est pas accessible. Néanmoins, il est possible de calculer explicitement cette intégrale car la position des pôles de $\hat{g}_N(z)$ par rapport au rectangle \mathcal{R} est connue. Un calcul classique de résidus donne alors la formule explicite suivante, facilement implémentable,

$$\hat{\eta}_{\text{rmt},N}(\theta) = \sum_{k=1}^M \hat{\xi}_{k,N} \hat{\mathbf{a}}(\theta)^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \hat{\mathbf{a}}(\theta),$$

avec

$$\hat{\xi}_{k,N} = \begin{cases} 1 + \sum_{m=M-K+1}^M \left(\frac{\hat{\lambda}_{m,N}}{\hat{\lambda}_{k,N} - \hat{\lambda}_{m,N}} - \frac{\hat{\omega}_{m,N}}{\hat{\lambda}_{k,N} - \hat{\omega}_{m,N}} \right) & \text{pour } k \leq M-K, \\ - \sum_{m=1}^{M-K} \left(\frac{\hat{\lambda}_{m,N}}{\hat{\lambda}_{k,N} - \hat{\lambda}_{m,N}} - \frac{\hat{\omega}_{m,N}}{\hat{\lambda}_{k,N} - \hat{\omega}_{m,N}} \right) & \text{pour } k > M-K. \end{cases}$$

Bien que les travaux de Mestre aient une portée relativement générale³, l'application au contexte de localisation de sources n'est valide que pour des signaux gaussiens i.i.d. temporellement (la matrice \mathbf{S}_N a des colonnes i.i.d), ce qui élimine un bon nombre d'exemples pratiques, notamment le cas où les signaux sont corrélés temporellement (e.g. processus AR) ou dont la loi varie au cours du temps. Par ailleurs, il est important de noter que les travaux de Mestre établissent la consistance de l'estimateur de la fonction de localisation, mais la consistance de l'estimateur angulaire (les minima de la fonction de localisation) n'est pas traitée, et nécessite en particulier de montrer une consistance de type uniforme,

$$\sup_{\theta \in [-\pi, \pi]} |\hat{\eta}_{\text{rmt},N}(\theta) - \eta_N(\theta)| \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0, \quad (18)$$

un résultat qui n'est pas une conséquence directe de la convergence $\hat{\eta}_{\text{rmt},N}(\theta) - \eta_N(\theta) \rightarrow_N 0$.

L'objectif de cette thèse consiste donc à étendre l'approche de Mestre au cas où les signaux sont vus comme déterministes, ce qui permet de considérer un plus large panel d'hypothèses en pratique. Dans ce contexte, il est nécessaire de considérer le cas des matrices aléatoires dites "information plus bruit", et (5) peut être maintenant considéré comme le modèle plus général $\boldsymbol{\Sigma}_N = \mathbf{B}_N + \mathbf{W}_N$, où \mathbf{B}_N est une matrice déterministe (dans le cas de la localisation de source $\mathbf{B}_N = \frac{\mathbf{A} \mathbf{S}_N}{\sqrt{N}}$ est de rang K) et \mathbf{W}_N est une matrice aléatoire dont les entrées sont i.i.d gaussiennes complexes centrées de variance σ^2 . Les principales contributions du présent travail de recherche concernant le modèle information plus bruit sont données dans la section suivante.

2. On a également utilisé le fait que $|\mathbf{a}(\theta)^* (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{a}(\theta)| \rightarrow_N 0$ p.s. pour tout $z \in \mathbb{C}^+$.

3. Les mêmes idées peuvent être employées pour trouver des estimateurs consistants des valeurs propres et des formes quadratiques de projecteur sur les sous-espaces propres de Γ_N , pour le modèle (6), dans un cadre non nécessairement gaussien (voir notamment [36]).

Contributions de la thèse

Les contributions de la thèse s'articulent en deux axes principaux. Le premier axe concerne l'étude du spectre des matrices aléatoires de type information plus bruit, et notamment la localisation des valeurs propres en grande dimension (voir par exemple (12) dans le cas des modèles à covariance). Le second axe de recherche concerne l'étude d'un algorithme de localisation amélioré, en adaptant la procédure de Mestre évoquée dans la section précédente.

Spectre des matrices aléatoires gaussiennes information plus bruit

Le premier axe de de cette thèse s'articule donc autour du modèle "information plus bruit"

$$\Sigma_N = \mathbf{B}_N + \mathbf{W}_N, \quad (19)$$

où \mathbf{B}_N est une matrice déterministe $M \times N$ de rang $K < M$ telle que $\sup_N \|\mathbf{B}_N\| < \infty$ et \mathbf{W}_N une matrice dont les entrées sont gaussiennes complexes i.i.d centrées de variance $\frac{\sigma^2}{N}$. Rappelons que l'on considère le régime asymptotique où M, K sont des fonctions de N , et $c_N = M/N \rightarrow c \in]0, 1[$ (le cas $c = 1$ sera éludé pour alléger le manuscrit). Ce modèle de matrice aléatoire a fait l'objet de peu de recherches, et les principales contributions sont dues à Dozier & Silverstein [12] [11] et Girko [16]. En particulier, ces travaux étudient la convergence de la distribution empirique $\hat{\mu}_N$ des valeurs propres de $\Sigma_N \Sigma_N^*$, de la même manière que pour (8). Ainsi on a $\hat{\mu}_N - \mu_N \xrightarrow[N \rightarrow \infty]{e} 0$ p.s., où μ_N est toujours une mesure de probabilité déterministe absolument continue, mais dont la transformée de Stieltjes $m_N(z)$ vérifie cette fois l'équation

$$m_N(z) = \frac{1}{M} \text{Tr } \mathbf{T}_N(z), \quad (20)$$

pour tout $z \in \mathbb{C} \setminus \mathbb{R}^+$, avec $\mathbf{T}_N(z)$ la matrice

$$\mathbf{T}_N(z) = \left(-z(1 + \sigma^2 c_N m_N(z)) \mathbf{I}_M + \sigma^2 (1 - c_N) \mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^*}{1 + \sigma^2 c_N m_N(z)} \right)^{-1}. \quad (21)$$

Une analyse de $\text{supp}(\mu_N)$, le support de μ_N , est donnée dans Dozier & Silverstein [11], mais la procédure de détermination de $\text{supp}(\mu_N)$ proposée n'est pas la meilleure. Inspirés par les travaux de Mestre, nous montrons que

$$\text{supp}(\mu_N) = \bigcup_{q=1}^Q [x_{q,N}^-, x_{q,N}^+], \quad (22)$$

où $x_{1,N}^- < x_{1,N}^+ < \dots < x_{Q,N}^- < x_{Q,N}^+$ sont les $2Q$ extrema positifs de la fonction

$$\phi_N(w) = w(1 - \sigma^2 c_N f_N(w))^2 + \sigma^2 (1 - c_N) (1 - \sigma^2 c_N f_N(w)),$$

où $f_N(w) = \frac{1}{M} \text{Tr} (\mathbf{B}_N \mathbf{B}_N^* - w \mathbf{I}_M)^{-1}$. L'obtention de (22) passe en particulier par l'analyse des propriétés de la fonction $w_N(z)$, l'analogue de (11), qui est donnée dans le contexte information plus bruit par

$$w_N(z) = z(1 + \sigma^2 c_N m_N(z))^2 - \sigma^2 (1 - c_N) (1 + \sigma^2 c_N m_N(z)). \quad (23)$$

Cette caractérisation du support est une étape indispensable à l'étude de résultats sur la localisation presque sûre des valeurs propres de $\Sigma_N \Sigma_N^*$, pour N grand, résultats qui n'ont pas encore établis dans des travaux antérieurs. Nous montrons dans un premier temps qu'aucune valeur propre de $\Sigma_N \Sigma_N^*$ n'appartient à $\text{supp}(\mu_N)$ p.s. à partir d'un certain rang N . Plus précisément, si $[a, b]$ est un intervalle et qu'il existe $\epsilon > 0$ tels que $[a - \epsilon, b + \epsilon] \cap \text{supp}(\mu_N) = \emptyset$ à partir d'un certain rang N , alors

$$\text{card} \{k : \hat{\lambda}_{k,N} \in [a, b]\} = 0, \quad (24)$$

p.s. pour N grand. Ce résultat est l'analogue d'un résultat démontré par Bai & Silverstein [3] dans le cadre des modèles à covariance, non nécessairement gaussiens. Pour montrer (24), nous utilisons une méthode plus

simple que [3], basée sur l'exploitation du caractère gaussien du modèle, et qui a été proposée par Haagerup & Thorbjornsen [17] dans le cas des matrices de Wigner (matrices hermitiennes à entrées gaussiennes). Nous déduisons également un ordre de décroissance pour la probabilité qu'une valeur propre $\hat{\lambda}_{k,N}$ de $\Sigma_N \Sigma_N^*$ sorte du support : si \mathcal{K} est un compact contenant \mathcal{S}_N à partir d'un certain rang N et \mathcal{V} un voisinage compact de \mathcal{K} , avec \mathcal{K}, \mathcal{V} indépendants de N , alors

$$\mathbb{P}(\exists k : \hat{\lambda}_{k,N} \notin \mathcal{V}) = \mathcal{O}\left(\frac{1}{N^l}\right), \text{ pour tout } l \in \mathbb{N}. \quad (25)$$

La méthode de [17] peut également être utilisée pour montrer une propriété de séparation des valeurs propres de $\Sigma_N \Sigma_N^*$ analogue à (12), où la fonction w_N est donnée désormais par (23). Ces résultats sont notamment exploités dans le cadre des "spiked models" gaussiens non centrés, pour établir la convergence des plus grandes valeurs propres de $\Sigma_N \Sigma_N^*$. En effet, sous l'hypothèse K indépendant de N et en supposant que les K valeurs propres non nulles de $\mathbf{B}_N \mathbf{B}_N^*$, i.e. $\lambda_{M-K+1,N}, \dots, \lambda_{M,N}$, convergent respectivement vers les limites $0 < \gamma_1 < \dots < \gamma_K$, nous montrons que pour $k = 1, \dots, K$,

$$\hat{\lambda}_{k,N} \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \begin{cases} (\gamma_k + \sigma^2 c)(\gamma_k + \sigma^2) & \text{si } \gamma > \sigma^2 \sqrt{c}, \\ \sigma^2 (1 + \sqrt{c})^2 & \text{sinon.} \end{cases} \quad (26)$$

Ainsi (26) montre que l'on peut exploiter les valeurs propres de $\Sigma_N \Sigma_N^*$ pour la détection de sources. En effet, dans le cadre du modèle à traitement d'antennes, où $\mathbf{B}_N = \frac{\mathbf{A} \Sigma_N}{\sqrt{N}}$, les K plus grandes valeurs propres de $\Sigma_N \Sigma_N^*$ se détacheront des autres si $\gamma_1 = \lim_N \lambda_{1,N} > \sigma^2 \sqrt{c}$. La quantité $\sigma^2 \sqrt{c}$ constitue en ce sens un seuil de détectabilité. Ces résultats ont été démontrés pour la première fois par Benaych & Nadakuditi [7], dans un cadre plus général que le cas gaussien, mais la méthodologie utilisée est relativement différente.

Application des modèles gaussiens information plus bruit à la localisation de source

Le second axe de cette thèse consiste donc à étendre au cas de signaux déterministes les résultats d'estimation sous-espace obtenus par Mestre pour les signaux gaussiens i.i.d (cadre des modèles à covariance). A l'aide des résultats développés préalablement sur le support de la distribution asymptotique μ_N , nous montrons ainsi que la fonction de localisation $\eta_N(\theta) = \sum_{k=1}^{M-K} \mathbf{a}(\theta)^* \mathbf{u}_{k,N} \mathbf{u}_{k,N}^* \mathbf{a}(\theta)$, où $\mathbf{u}_{1,N}, \dots, \mathbf{u}_{M,N}$ désignent désormais les vecteurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ associés aux valeurs propres $\lambda_{1,N} \leq \dots \leq \lambda_{M,N}$, peut s'écrire comme

$$\eta_N(\theta) = \frac{1}{2\pi i} \oint_{\partial \mathcal{R}} \mathbf{a}(\theta)^* \mathbf{T}_N(z) \mathbf{a}(\theta) \frac{w'_N(z)}{1 + \sigma^2 c_N m_N(z)} dz,$$

sous certaines hypothèses de séparation des sous-espaces bruit et signal (i.e. entre le sous-espace associé aux $M - K$ valeurs propres nulles de $\mathbf{B}_N \mathbf{B}_N^*$ et son complément orthogonal). L'étape suivante consiste à estimer de façon consistante le terme sous l'intégrale, par des quantités ne dépendant que de la matrice Σ_N , ce qui implique notamment de montrer que $\mathbf{a}(\theta)^* \mathbf{Q}_N(z) \mathbf{a}(\theta)$ est un estimateur consistant de la forme quadratique $\mathbf{a}(\theta)^* \mathbf{T}_N(z) \mathbf{a}(\theta)$ quand $N \rightarrow \infty$, où $\mathbf{Q}_N(z) = (\Sigma_N \Sigma_N^* - z \mathbf{I}_M)^{-1}$ est la résolvante de $\Sigma_N \Sigma_N^*$. Nous obtenons ainsi

$$\hat{\eta}_{\text{new},N}(\theta) = \frac{1}{2\pi i} \oint_{\partial \mathcal{R}} \mathbf{a}(\theta)^* \mathbf{Q}_N(z) \mathbf{a}(\theta) \frac{\hat{w}'_N(z)}{1 + \sigma^2 c_N \hat{m}_N(z)} dz, \quad (27)$$

où $\hat{w}'_N(z)$ est la dérivée de $\hat{w}_N(z) = z(1 + \sigma^2 c_N \hat{m}_N(z))^2 - \sigma^2 (1 - c_N)(1 + \sigma^2 c_N \hat{m}_N(z))$. Les résultats sur la localisation des valeurs propres de $\Sigma_N \Sigma_N^*$, et plus particulièrement les résultats de séparation, permettent ensuite d'évaluer l'intégrale estimée (27) à l'aide du théorème des résidus, pour obtenir une formule ne dépendant plus que de σ, c_N et des valeurs propres et vecteurs propres de $\Sigma_N \Sigma_N^*$. Ceci nécessite également un résultat de séparation pour les zéros du dénominateur $z \mapsto 1 + \sigma^2 c_N \hat{m}_N(z)$. Ayant développé un estimateur consistant de la fonction de localisation $\eta_N(\theta)$, nous nous intéressons ensuite à la consistance de l'estimateur des angles d'arrivées, définis comme étant les K minima les plus significatifs de cette fonction. Pour étudier cette consistance, nous proposons une approche basée sur un calcul de moments de l'estimateur, après régularisation par une fonction dont le but est de séparer et confiner les pôles de l'intégrande (27). Nous utilisons notamment (25) et prouvons un résultat similaire pour les M zéros de $z \mapsto 1 + \sigma^2 c_N \hat{m}_N(z)$, le dénominateur dans (27).

Organisation du document

La section 2 présente les bases du modèle information plus bruit, en particulier le résultat de la convergence de la distribution empirique des valeurs propres est donné, ainsi qu'une étude complète du support de la distribution asymptotique. En complément, nous examinons le cas où le rang de \mathbf{B}_N dans (19) est indépendant de N , référencé dans la littérature sous le nom de "spiked models".

La section 4 est consacrée à l'analyse de la localisation des valeurs propres de $\Sigma_N \Sigma_N^*$. En particulier, les propriétés de séparation des valeurs propres sont établies, et utilisées dans le cadre des "spiked models" pour étudier la limite des plus grandes valeurs propres.

La section 5 étudie l'équivalent de l'estimateur sous-espace de Mestre dans le cas où les signaux sources sont supposés déterministes. Les résultats des chapitres précédents sont utilisés pour montrer la consistance de l'estimateur de la fonction de localisation obtenu, ainsi que la consistance de l'estimateur des angles d'arrivée.

Notations

- Les entiers naturels sont notés \mathbb{N} et les corps des nombres réels et complexes respectivement \mathbb{R}, \mathbb{C} . \mathbb{R}^+ et \mathbb{R}^- représentent respectivement les réels positifs et négatifs, et nous notons $\mathbb{R}^* \equiv \mathbb{R} \setminus \{0\}$, $\mathbb{R}_*^+ \equiv \mathbb{R}^+ \setminus \{0\}$, $\mathbb{R}_*^- \equiv \mathbb{R}^- \setminus \{0\}$. Nous définissons également $\mathbb{C}^+ \equiv \{z \in \mathbb{C} : \text{Im}(z) > 0\}$, $\mathbb{C}^- \equiv \{z \in \mathbb{C} : \text{Im}(z) < 0\}$, et $\mathbb{C}^* \equiv \mathbb{C} \setminus \{0\}$.
- Si $\mathcal{E} \subset \mathbb{R}$, \mathcal{E}^c , $\text{Int}(\mathcal{E})$ et $\partial\mathcal{E}$ représentent respectivement le complémentaire, l'intérieur et la frontière de \mathcal{E} . Si \mathcal{E} a un nombre fini d'éléments, son cardinal est noté $|\mathcal{E}|$.
- Pour un nombre complexe z , nous notons $\text{Re}(z)$ et $\text{Im}(z)$ ses parties réelle et imaginaire, et z^* ou \bar{z} son conjugué. L'unité imaginaire est notée i . Les vecteurs et matrices seront notés respectivement avec des minuscules en gras et des capitales en gras, par exemple $\mathbf{a} \in \mathbb{C}^n$, $\mathbf{A} \in \mathbb{C}^{m \times n}$. Les transposée, conjuguée, et transposée-conjuguée de $\mathbf{A} \in \mathcal{M}_n(\mathbb{C})$ sont notées respectivement $\mathbf{A}^T, \bar{\mathbf{A}}$ et \mathbf{A}^* , sa trace et sa norme spectrale $\text{Tr}(\mathbf{A})$ et $\|\mathbf{A}\|$. Les vecteurs de la base canonique de \mathbb{R}^n sont notés $\mathbf{e}_1, \dots, \mathbf{e}_n$ avec \mathbf{e}_j ayant toutes ses entrées égales à 0 sauf la j -ième qui vaut 1.
- Pour un ensemble $\mathcal{E} \subset \mathbb{C}$, $\mathcal{C}(\mathbb{R}^n, \mathcal{E})$, $\mathcal{C}^1(\mathbb{R}^n, \mathcal{E})$ et $\mathcal{C}^\infty(\mathbb{R}^n, \mathcal{E})$ correspondent aux ensembles de fonctions définis sur \mathbb{R}^n à valeurs dans \mathcal{E} , et respectivement continues, continuellement différentiables et infiniment différentiables. Si les fonctions sont bornées, nous ajouterons l'indice "b", et si leur support est compact, nous ajouterons l'indice "c". Le support d'une fonction φ est noté $\text{supp}(\varphi)$.
- Une variable aléatoire complexe $Z = X + iY$ suit une loi gaussienne complexe circulaire $\mathcal{N}_{\mathbb{C}}(\alpha + i\beta, \sigma^2)$ si X et Y sont des variables gaussiennes réelles indépendantes $\mathcal{N}_{\mathbb{R}}(\alpha, \frac{\sigma^2}{2})$ et $\mathcal{N}_{\mathbb{R}}(\beta, \frac{\sigma^2}{2})$ respectivement. L'espérance de Z est notée $\mathbb{E}[Z] = \alpha + i\beta$ et sa variance $\mathbb{V}[Z] = \mathbb{E}|Z - \mathbb{E}[Z]|^2 = \sigma^2$. Si $(X_n), (Y_n)$ sont deux suites de variables aléatoires complexes, $X_n - Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0$ signifie que $X_n - Y_n$ tend presque-sûrement vers 0. Si (μ_n) et (ν_n) sont deux suites de mesures positives sur \mathbb{R} , nous notons $\mu_n - \nu_n \xrightarrow[n \rightarrow \infty]{e} 0$ la convergence étroite de $\mu_n - \nu_n$ to 0, i.e. $\int_{\mathbb{R}} \varphi d\mu_n - \int_{\mathbb{R}} \varphi d\nu_n \rightarrow_N 0$ pour toute $\varphi \in \mathcal{C}_b(\mathbb{R}, \mathbb{R})$.

2 Mesure spectrale asymptotique des matrices i.i.d. gaussiennes non centrées

2.1 La transformée de Stieltjes

Cette section introduit la transformée de Stieltjes, qui est constituée l'outil central dans l'étude des valeurs propres des grandes matrices aléatoires. Les résultats de cette section peuvent être trouvés dans Akhiezer [1] et dans Krein & Nudelman [20].

Soit μ une mesure finie de support $\text{supp}(\mu) \subset \mathbb{R}$. Sa transformée de Stieltjes Ψ_μ est la fonction définie par

$$\Psi_\mu(z) \triangleq \int_{\mathbb{R}} \frac{d\mu(\lambda)}{\lambda - z} \quad \forall z \in \mathbb{C} \setminus \text{supp}(\mu).$$

La transformée de Stieltjes est évidemment holomorphe et vérifie la propriété élémentaire $\Psi_\mu(z)^* = \Psi_\mu(z^*)$. Il est possible de caractériser les valeurs prises par la mesure μ à partir de sa transformée de Stieltjes. Ainsi, pour

$\varphi \in \mathcal{C}_c(\mathbb{R}, \mathbb{R})$ (l'espace des fonctions réelles continues à support compact dans \mathbb{R}), nous avons

$$\int_{\mathbb{R}} \varphi(\lambda) d\mu(\lambda) = \frac{1}{\pi} \lim_{y \downarrow 0} \text{Im} \left(\int_{\mathbb{R}} \varphi(x) \Psi_{\mu}(x + iy) dx \right). \quad (28)$$

En particulier, Si a, b sont des points de continuité de μ ,

$$\mu([a, b]) = \lim_{y \downarrow 0} \int_a^b \text{Im}(\Psi_{\mu}(x + iy)) dx. \quad (29)$$

Il existent des conditions suffisantes pour qu'une fonction soit une transformée de Stieltjes. En effet, soit $\Psi(z)$ une fonction holomorphe sur \mathbb{C}^+ . Si $z \in \mathbb{C}^+$ entraîne $\Psi(z) \in \mathbb{C}^+$ et si $\limsup_{y \rightarrow \infty} |iy\Psi(iy)| < \infty$, alors $\Psi(z)$ est la transformée de Stieltjes d'une mesure finie sur \mathbb{R} .

A l'instar de la transformée de Fourier, la transformée de Stieltjes est particulièrement adaptée pour montrer la convergence en loi. En effet, si (μ_n) , μ sont des mesures de probabilité sur \mathbb{R} et si (Ψ_{μ_n}) , Ψ_{μ} sont les transformées de Stieltjes associées, alors les deux assertions suivantes sont équivalentes

1. $\Psi_{\mu_n}(z) \xrightarrow[n \rightarrow \infty]{} \Psi_{\mu}(z)$ pour tout $z \in \mathbb{C}^+$,
2. $\mu_n \xrightarrow[n \rightarrow \infty]{w} \mu$.

2.2 Matrices gaussiennes i.i.d. non centrées et convergence de la mesure spectrale empirique

Soient $M, N \in \mathbb{N}^*$ tels que $M = M(N)$ est une fonction de N , $M < N$ et $c_N = M/N \rightarrow c \in]0, 1[$ quand $N \rightarrow \infty$. Une matrice gaussienne i.i.d. non centrée est définie comme la matrice aléatoire de taille $M \times N$

$$\Sigma_N = \mathbf{B}_N + \mathbf{W}_N,$$

où

- \mathbf{B}_N est une matrice déterministe telle que $\sup_N \|\mathbf{B}_N\| \leq B_{\max} < \infty$,
- Les entrées $W_{i,j}$ de \mathbf{W}_N sont i.i.d $W_{i,j} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ (loi gaussienne complexe circulaire centrée de variance σ^2).

Les décompositions en valeurs et vecteurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ et $\Sigma_N \Sigma_N^*$ sont notées respectivement par

$$\mathbf{B}_N \mathbf{B}_N^* = \mathbf{U}_N \Lambda_N \mathbf{U}_N^* \quad \text{et} \quad \Sigma_N \Sigma_N^* = \hat{\mathbf{U}}_N \hat{\Lambda}_N \hat{\mathbf{U}}_N^*, \quad (30)$$

avec $\mathbf{U}_N, \hat{\mathbf{U}}_N$ des matrices unitaires des vecteurs propres et $\Lambda_N = \text{Diag}(\lambda_{1,N}, \dots, \lambda_{M,N})$ les matrices des valeurs propres. $\hat{\Lambda}_N = \text{Diag}(\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{M,N})$. Les valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ et $\Sigma_N \Sigma_N^*$ sont ordonnées de telle manière que $0 \leq \lambda_{1,N} \leq \dots \leq \lambda_{M,N}$ et $0 \leq \hat{\lambda}_{1,N} \leq \dots \leq \hat{\lambda}_{M,N}$. Le comportement global des valeurs propres de $\Sigma_N \Sigma_N^*$ est étudié à partir de sa mesure spectrale empirique, définie comme la mesure de probabilité aléatoire

$$\hat{\mu}_N \triangleq \frac{1}{M} \sum_{k=1}^M \delta_{\hat{\lambda}_{k,N}},$$

où δ_x est la mesure de Dirac prise au point x . La transformée de Stieltjes $\hat{m}_N(z)$ de $\hat{\mu}_N$ s'exprime simplement en fonction de la résolvante de $\Sigma_N \Sigma_N^*$,

$$\mathbf{Q}_N(z) = (\Sigma_N \Sigma_N^* - z \mathbf{I}_M)^{-1},$$

définie pour tout $z \in \mathbb{C} \setminus \mathbb{R}^+$. En effet, nous avons

$$\hat{m}_N(z) \triangleq \int_{\mathbb{R}^+} \frac{d\hat{\mu}_N(\lambda)}{\lambda - z} = \frac{1}{M} \text{Tr} \mathbf{Q}_N(z).$$

Comme mentionné en section 2.1, la convergence de $\hat{\mu}_N$ peut être prouvée en étudiant la convergence ponctuelle de $\frac{1}{M} \text{Tr} \mathbf{Q}_N(z)$ quand $N \rightarrow \infty$, ce qui a notamment fait l'objet de travaux par Girko [15] puis plus récemment par Dozier & Silverstein [12] dans un cadre plus général.

Théorème 1. *Il existe une mesure de probabilité déterministe μ_N , telle que $\text{supp}(\mu_N) \subset \mathbb{R}^+$, et vérifiant $\hat{\mu}_N - \mu_N \xrightarrow{e} 0$ as $N \rightarrow \infty$ p.s. De manière équivalente, la transformée de Stieltjes m_N de μ_N vérifie $\hat{m}_N(z) - m_N(z) \rightarrow 0$ p.s. $\forall z \in \mathbb{C} \setminus \mathbb{R}^+$. De plus, $\forall z \in \mathbb{C} \setminus \mathbb{R}^+$, $m_N(z)$ est l'unique solution de l'équation*

$$m = \frac{1}{M} \text{Tr} \left(-z(1 + \sigma^2 c_N m) \mathbf{I}_M + \sigma^2 (1 - c_N) \mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^*}{1 + \sigma^2 c_N m} \right)^{-1} \quad (31)$$

telle que $\text{Im}(zm_N(z)) > 0$ pour $z \in \mathbb{C}^+$.

Ce résultat reste également valide dans le cas non gaussien (les éléments de la matrice \mathbf{W}_N ne sont plus gaussiens). La mesure de probabilité μ_N sera désignée par la suite comme la mesure spectrale asymptotique (ou distribution asymptotique des valeurs propres) de la matrice $\Sigma_N \Sigma_N^*$. Par souci de concision, nous notons

$$m_N(z) = \frac{1}{M} \text{Tr} \mathbf{T}_N(z), \quad (32)$$

avec la matrice $\mathbf{T}_N(z)$ définie par

$$\mathbf{T}_N(z) = \left(-z(1 + \sigma^2 c_N m_N(z)) \mathbf{I}_M + \sigma^2 (1 - c_N) \mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^*}{1 + \sigma^2 c_N m_N(z)} \right)^{-1}, \quad (33)$$

pour tout $z \in \mathbb{C} \setminus \mathbb{R}^+$. Du théorème 1, nous savons que $\frac{1}{M} \text{Tr} \mathbf{Q}_N(z)$ est une bonne approximation de $\frac{1}{M} \text{Tr} \mathbf{T}_N(z)$. On peut également se demander si les entrées de $\mathbf{Q}_N(z)$ sont également de bonnes approximations des entrées de la matrice déterministe $\mathbf{T}_N(z)$. La réponse, positive, est donnée dans le résultat suivant (montré dans le cadre de la thèse et publié dans [19], dans le cadre plus général où \mathbf{W}_N n'est pas nécessairement gaussienne).

Théorème 2. *Soit $(\mathbf{d}_{1,N})$ et $(\mathbf{d}_{2,N})$ deux suites de vecteurs déterministes tels que $\sup_N \|\mathbf{d}_{1,N}\|, \sup_N \|\mathbf{d}_{2,N}\| < \infty$. Alors, on a*

$$\mathbf{d}_{1,N}^* (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{d}_{2,N} \xrightarrow[N \rightarrow \infty]{p.s.} 0 \quad (34)$$

pour tout $z \in \mathbb{C} \setminus \mathbb{R}$.

Le résultat du théorème 2 sera notamment essentiel en section 5 pour les problèmes d'estimation sous-espace dans les grands réseaux de capteurs.

2.3 Propriétés de la mesure spectrale asymptotique

Dans cette section, nous étudions les propriétés de la mesure spectrale asymptotique μ_N , en particulier son support. La plupart des résultats présentés ci-dessous ont été montrés par Dozier & Silverstein [11]. Néanmoins, nous proposons une caractérisation plus explicite du support de la mesure μ_N , que nous noterons $\mathcal{S}_N = \text{supp}(\mu_N)$, et calculons explicitement la masse que μ_N associe à certaines parties de son support. Ces résultats auront un impact fondamental sur les résultats de localisation presque sûre des valeurs propres de $\Sigma_N \Sigma_N^*$ de la section 4, ainsi que pour l'estimation sous-espace de la section 5.

La caractérisation du support \mathcal{S}_N nécessite d'étudier quelques propriétés de la transformée de Stieltjes m_N , qui sont énoncées ci-dessous.

Propriété 1. *La fonction m_N peut être prolongée par continuité sur l'axe réel en considérant la limite de $m_N(z)$ quand $z \in \mathbb{C}^+$ converge vers $x \in \mathbb{R}$. La limite est notée $m_N(x)$. Définie ainsi sur \mathbb{R} , la fonction $x \mapsto m_N(x)$ est continue, et $m_N(x)$ vérifie l'équation (31) pour tout $x \in \mathbb{R} \setminus \mathcal{S}_N$. Finalement, μ_N est une mesure probabilité absolument continue dont la densité est donnée par*

$$\frac{d\mu_N(x)}{dx} = \frac{1}{\pi} \text{Im}(m_N(x)).$$

La propriété précédente nous montre donc que \mathcal{S}_N vérifie $\text{Int}(\mathcal{S}_N) = \{x \in \mathbb{R}^+ : \text{Im}(m_N(x)) > 0\}$, où $\text{Int}()$ représente l'intérieur. Pour étudier le support \mathcal{S}_N , Dozier & Silverstein ont caractérisé la fonction $x \mapsto 1 + \sigma^2 c_N m_N(x)$ parmi les solutions d'une certaine équation. Nous utiliserons ici la fonction

$$w_N(z) = z(1 + \sigma^2 c_N m_N(z))^2 - \sigma^2 (1 - c_N) (1 + \sigma^2 c_N m_N(z)), \quad (35)$$

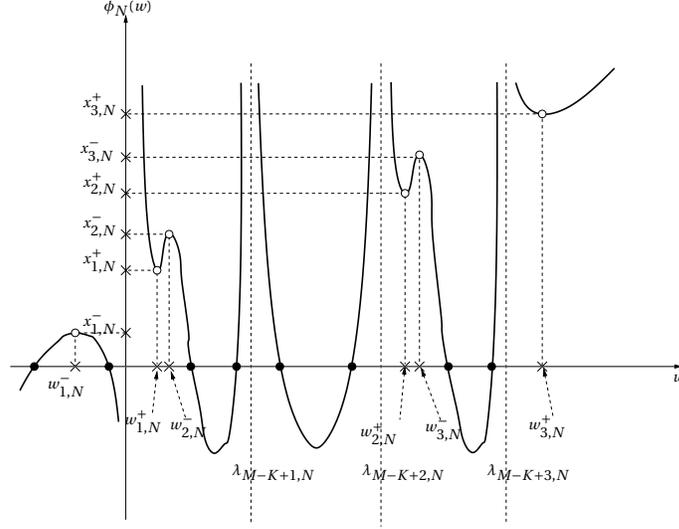


FIGURE 4 – Un exemple de graphe de la fonction $\phi_N(w)$, pour $K = 3$.

définie pour tout $z \in \mathbb{C}$, qui est plus adaptée pour l'analyse, et qui apparaîtra naturellement dans les problèmes d'estimation sous-espace de la section 5. Comme nous le verrons ci-dessous, la fonction w_N est étroitement liée à la fonction

$$\phi_N(w) = w(1 - \sigma^2 c_N f_N(w))^2 + \sigma^2(1 - c_N)(1 - \sigma^2 c_N f_N(w)), \quad (36)$$

où $f_N(w) = \frac{1}{M} \text{Tr}(\mathbf{B}_N \mathbf{B}_N^* - w \mathbf{I}_M)^{-1}$. Pour la suite, nous considérerons le cas particulier où \mathbf{B}_N est de rang déficient K . Les résultats peuvent bien sûr être étendus au cas \mathbf{B}_N de rang plein.

Hypothèse H-1: La matrice \mathbf{B}_N est de rang $K = K(N) < M$, et les valeurs propres non nulles de $\mathbf{B}_N \mathbf{B}_N^*$ ont multiplicité 1 pour tout N .

On a donc $0 = \lambda_{1,N} = \dots = \lambda_{M-K,N} < \lambda_{M-K+1,N} < \dots < \lambda_{M,N}$. Notons également que K peut aussi bien tendre vers l'infini avec N ou rester constant. Les principales propriétés de la fonction w_N sont énoncées ci-dessous.

Propriété 2. w_N vérifie $\text{Im}(w_N(z)) > 0$ si $\text{Im}(z) > 0$ et $w_N(x) \in \mathbb{R} \setminus \{\lambda_{1,N}, \dots, \lambda_{M,N}\}$ si $x \in \mathbb{R} \setminus \mathcal{S}_N$. De plus, pour $z \in \mathbb{C} \setminus \partial \mathcal{S}_N$, $w_N(z)$ est solution de l'équation $\phi_N(w) = z$. Finalement, $\text{Int}(\mathcal{S}_N) = \{x \in \mathbb{R} : \text{Im}(w_N(x)) > 0\}$.

La propriété 2 exprime le fait qu'il est suffisant d'étudier le comportement de $\text{Im}(w_N(x))$ pour caractériser le support \mathcal{S}_N . Le fait que $w_N(x)$ soit solution de l'équation $\phi_N(w) = x$, pour $x \in \mathbb{R} \setminus \partial \mathcal{S}_N$ constitue un point fondamental et cette observation peut être utilisée pour caractériser \mathcal{S}_N , en identifiant $w_N(x)$ parmi les solutions de l'équation $\phi_N(w) = x$. Une analyse de la fonction $\phi_N(w)$ permet d'obtenir la propriété suivante.

Propriété 3. La fonction ϕ_N admet exactement $2Q$ extrema locaux strictement positifs, avec $Q = Q(N)$ tel que $1 \leq Q \leq K + 1$, dont les antécédents, notés $w_{1,N}^- < 0 < w_{1,N}^+ < w_{2,N}^- < \dots < w_{Q,N}^- < w_{Q,N}^+$, appartiennent à l'ensemble $\{w \in \mathbb{R} : 1 - \sigma^2 c_N f_N(w) > 0\}$. ϕ_N est strictement croissante sur $(-\infty, w_{1,N}^-)$, $(w_{Q,N}^+, \infty)$, $(w_{q,N}^+, w_{q+1,N}^-)$ pour $q = 1, \dots, Q - 1$. Si ces extrema locaux sont notés $x_{q,N}^- \triangleq \phi_N(w_{q,N}^-)$ et $x_{q,N}^+ \triangleq \phi_N(w_{q,N}^+)$, alors

$$0 < x_{1,N}^- < x_{1,N}^+ < x_{2,N}^- < \dots < x_{Q,N}^- < x_{Q,N}^+, \quad (37)$$

De plus, chaque valeur propre $\lambda_{k,N}$ de $\mathbf{B}_N \mathbf{B}_N^*$ appartient à un intervalle $(w_{q,N}^-, w_{q,N}^+)$, $q = 1, \dots, Q$, et chaque intervalle $(w_{q,N}^-, w_{q,N}^+)$, $q = 1, \dots, Q$ contient de même au moins une valeur propre de $\mathbf{B}_N \mathbf{B}_N^*$,

Une représentation classique de la fonction ϕ_N est donnée en figure 4. Ces propriétés énoncées nous permettent de formuler la caractérisation du support de la mesure μ_N .

Théorème 3. *Sous l'hypothèse H-1, le support \mathcal{S}_N de μ_N est donné par*

$$\mathcal{S}_N = \bigcup_{q=1}^Q [x_{q,N}^-, x_{q,N}^+],$$

Démonstration. Comme mentionné précédemment, la preuve consiste à identifier $w_N(x)$ parmi les solutions de l'équation $\phi_N(w) = x$, pour $x \in \mathbb{R}$. Les grandes lignes sont données. Pour $x \in \mathbb{R} \setminus \left(\bigcup_{q=1}^Q [x_{q,N}^-, x_{q,N}^+] \right)$, on montre, en utilisant des propriétés élémentaires sur les variations de ϕ_N , que l'équation $\phi_N(w) = x$ n'admet que des solutions réelles. Donc $x \notin \text{Int}(\mathcal{S}_N)$ par la propriété 2, autrement $w_N(x)$ serait une des ces solutions et à partie imaginaire strictement positive. Donc $\text{Int}(\mathcal{S}_N) \subset \bigcup_{q=1}^Q [x_{q,N}^-, x_{q,N}^+]$. Si $x \in \bigcup_{q=1}^Q]x_{q,N}^-, x_{q,N}^+[$, on montre que l'équation $\phi_N(x) = x$ n'admet que des solutions réelles excepté deux solutions complexes conjuguées, et $w_N(x)$ correspond à la solution de partie imaginaire strictement positive. Donc $\bigcup_{q=1}^Q]x_{q,N}^-, x_{q,N}^+[\subset \text{Int}(\mathcal{S}_N)$. \square

Le théorème 3 montre donc que le support \mathcal{S}_N est la réunion disjointe d'intervalles compacts, qui seront référés par la suite sous le nom de "clusters". Chacun de ces clusters $[x_{q,N}^-, x_{q,N}^+]$ peut être calculé grâce à la fonction ϕ_N puisque les points frontières $x_{1,N}^- < x_{1,N}^+ < \dots < x_{q,N}^- < x_{q,N}^+$ sont des extrema locaux positifs de ϕ_N . De plus, chaque cluster est associé à un intervalle $[w_{q,N}^-, w_{q,N}^+]$, $q = 1 \dots Q$, au sens où $x_{q,N}^- = \phi_N(w_{q,N}^-)$ et $x_{q,N}^+ = \phi_N(w_{q,N}^+)$ (propriété 3). De plus, chaque valeur propre $\lambda_{k,N}$, $k = 1, \dots, M$, de $\mathbf{B}_N \mathbf{B}_N^*$ appartient toujours à un unique intervalle $[w_{q,N}^-, w_{q,N}^+]$ (propriété 3). Ceci motive la définition suivante

Définition 1. *Une valeur propre $\lambda_{k,N}$, $k = 1, \dots, M$ de $\mathbf{B}_N \mathbf{B}_N^*$ est dite associée au cluster $[x_{q,N}^-, x_{q,N}^+]$ si $\lambda_{k,N} \in [w_{q,N}^-, w_{q,N}^+]$.*

Notons que cette association n'est pas une bijection, au sens où plusieurs valeurs propres consécutives de $\mathbf{B}_N \mathbf{B}_N^*$ peuvent être associées au même cluster. Observons également que 0 est toujours associé au premier cluster $[x_{1,N}^-, x_{1,N}^+]$.

Remarque 1. *On peut également montrer que $w_N(x_{q,N}^-) = w_{q,N}^-$ et $w_N(x_{q,N}^+) = w_{q,N}^+$.*

Le résultat montre que la masse de chaque cluster par la mesure μ_N est directement reliée au nombre de valeurs propres associées. Définissons

$$\mathcal{I}_q = \{k \in \{1, \dots, K\} : \lambda_{M-K+k,N} \in (w_{q,N}^-, w_{q,N}^+)\}, \quad (38)$$

et $|\mathcal{I}_q| = \text{card}(\mathcal{I}_q) > 0$. $|\mathcal{I}_q|$ représente le nombre de valeurs propres non nulles de $\mathbf{B}_N \mathbf{B}_N^*$ associées au cluster $[x_{q,N}^-, x_{q,N}^+]$, et donc $|\mathcal{I}_q| > 0$ de part la propriété 3.

Propriété 4. *Sous l'hypothèse H-1,*

$$\mu_N \left([x_{1,N}^-, x_{1,N}^+] \right) = \frac{|\mathcal{I}_1|}{M} + \frac{M-K}{M},$$

et pour $q \geq 2$,

$$\mu_N \left([x_{q,N}^-, x_{q,N}^+] \right) = \frac{|\mathcal{I}_q|}{M}.$$

La propriété 4 exprime le fait que la masse d'un cluster par la mesure μ_N est exactement la proportion de valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ associées à ce cluster.

2.4 Le cas particulier des spiked models

Dans cette section, nous considérons le cas particulier où le rang de la matrice $\mathbf{B}_N \mathbf{B}_N^*$ est indépendant de N . Dans ce contexte, il est possible de calculer explicitement les points frontières du support \mathcal{S}_N , et ainsi obtenir une caractérisation précise de μ_N , en terme de son support et de concentration de sa masse. Nous remplaçons donc l'hypothèse H-1 par l'hypothèse suivante, plus restrictive.

Hypothèse H-2: Le rang $K > 0$ de $\mathbf{B}_N \mathbf{B}_N^*$ ne dépend pas de N et pour tout $k = 1, \dots, K$, les valeurs propres non nulles $(\lambda_{M-K+k, N})$ vérifient

$$\lambda_{M-K+k, N} = \gamma_k + \varepsilon_{k, N},$$

avec $\lim_{N \rightarrow +\infty} \varepsilon_{k, N} = 0$ et $\gamma_1 < \dots < \gamma_K$.

Avant d'étudier le devenir de \mathcal{S}_N dans ce contexte, discutons de quelques conséquences de l'hypothèse H-2. Du théorème 1, il est immédiat de voir que pour $z \in \mathbb{C}^+$, $\hat{m}_N(z) \rightarrow_N m(z)$ presque sûrement, avec

$$m(z) = \frac{1}{-z(1 + \sigma^2 c m(z)) + \sigma^2(1 - c)}. \quad (39)$$

$m(z)$ est la transformée de Stieltjes de la loi de Marcenko-Pastur μ , dont la densité vérifie

$$\frac{d\mu(x)}{dx} = \frac{\sqrt{(x - \sigma^2(1 - \sqrt{c}))^2 (\sigma^2(1 + \sqrt{c})^2 - x)}}{2\sigma^2 c \pi x} \mathbb{1}_{[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]}.$$

Etudions à présent le support \mathcal{S}_N de la mesure μ_N , sous l'hypothèse H-2.

Théorème 4. Sous l'hypothèse H-2, définissons $K_s = \text{card}\{k : \gamma_k > \sigma^2 \sqrt{c}\}$ et supposons que $\sigma^2 \sqrt{c} \notin \{\gamma_1, \dots, \gamma_K\}$, i.e

$$\gamma_1 < \dots < \gamma_{K-K_s} < \sigma^2 \sqrt{c} < \gamma_{K-K_s+1} < \dots < \gamma_K.$$

Pour N suffisamment grand, le support \mathcal{S}_N possède $Q = K_s + 1$ clusters, i.e $\mathcal{S}_N = \bigcup_{q=1}^{K_s+1} [x_{q, N}^-, x_{q, N}^+]$. Le premier cluster est associé à $\lambda_{1, N}, \dots, \lambda_{M-K_s, N}$ et est donnée par

$$x_{1, N}^- = \sigma^2(1 - \sqrt{c_N})^2 + \mathcal{O}\left(\frac{1}{N}\right) \quad \text{et} \quad x_{1, N}^+ = \sigma^2(1 + \sqrt{c_N})^2 + \mathcal{O}\left(\frac{1}{N}\right).$$

Pour $q = 2, \dots, K_s + 1$ et $k = q - 1$, le cluster $[x_{q, N}^-, x_{q, N}^+]$ est associé avec $\lambda_{M-K_s+k, N}$ et

$$x_{q, N}^- = \psi(\lambda_{M-K_s+k, N}, c_N) - \mathcal{O}^+\left(\frac{1}{\sqrt{N}}\right)$$

$$x_{q, N}^+ = \psi(\lambda_{M-K_s+k, N}, c_N) + \mathcal{O}^+\left(\frac{1}{\sqrt{N}}\right).$$

où $\psi(\lambda, c) = \frac{(\lambda + \sigma^2 c)(\lambda + c)}{\lambda}$ et $\mathcal{O}^+\left(\frac{1}{\sqrt{N}}\right)$ est un terme $\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$ positif.

Sous l'hypothèse des spiked models, il est intuitivement attendu que la mesure μ_N soit proche de la loi de Marcenko-Pastur μ , et particulièrement que \mathcal{S}_N soit proche de $\text{supp}(\mu) = [\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$. Le théorème 4 montre que le premier cluster $[x_{1, N}^-, x_{1, N}^+]$ est très proche du support de la loi Marcenko-Pastur et la présence de clusters supplémentaires peut être également vérifiée si les valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ sont suffisamment grandes. En effet, si K_s valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ convergent vers des limites différentes, et au-dessus du seuil $\sigma^2 \sqrt{c}$, alors K_s clusters additionnels seront présents dans le support de \mathcal{S}_N pour N suffisamment grand.

Le théorème 4 montre également que les $M - K_s$ plus petites valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ sont associées au premier cluster, où de manière équivalente que $\mu_N([x_{1, N}^-, x_{1, N}^+]) = \frac{M - K_s}{M}$, et $\mu_N([x_{k, N}^-, x_{k, N}^+]) = \frac{1}{M}$ pour $k = 2, \dots, K_s$ (propriété 4).

3 Illustrations

Dans cette section, nous donnons quelques illustrations des résultats des sections précédentes.

Les figures 5, 6 et 7 représentent le comportement typique de la fonction ϕ_N , pour une matrice \mathbf{B}_N de rang 3. Différentes situations sont représentées et commentées ci-dessous.

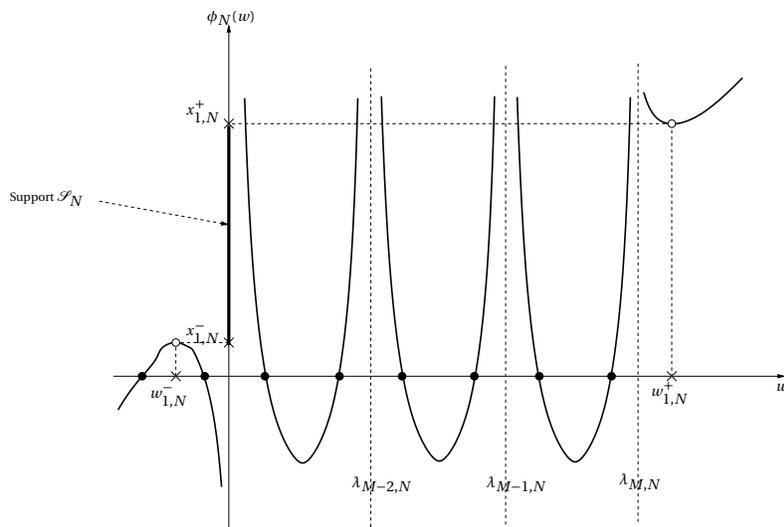


FIGURE 5 – $\phi_N(w)$: Cas 1

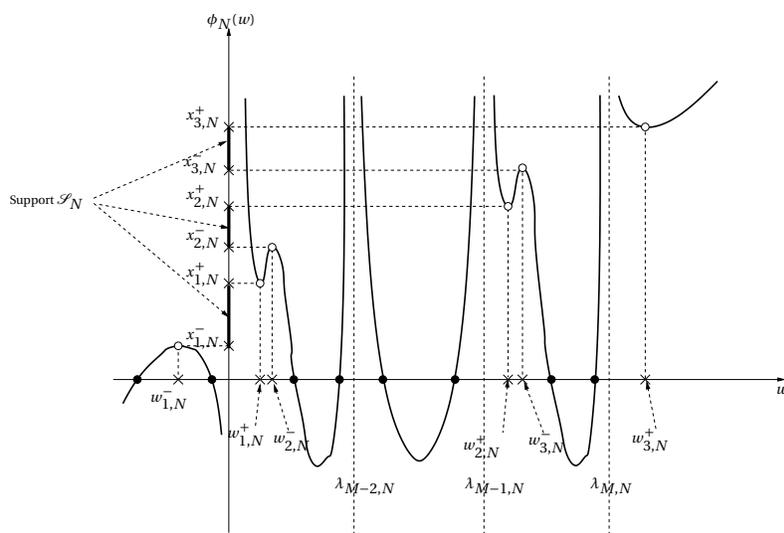


FIGURE 6 – $\phi_N(w)$: Cas 2

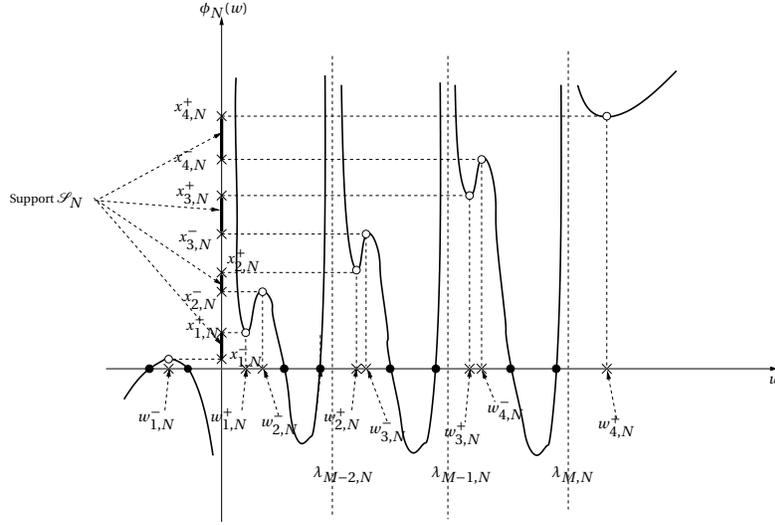


FIGURE 7 – $\phi_N(w)$: Cas 3

En figure 5, toutes les valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ sont associées à $[x_{1,N}^-, x_{1,N}^+]$. Il n'existe pas d'autres extrema locaux strictement positifs que $x_{1,N}^-$ et $x_{1,N}^+$, ce qui implique que \mathcal{S}_N n'a qu'un cluster, i.e $\mathcal{S}_N = [x_{1,N}^-, x_{1,N}^+]$.

En figure 6, la valeur propre 0 est associée au cluster $[x_{1,N}^-, x_{1,N}^+]$, $\lambda_{M-2,N}$, $\lambda_{M-1,N}$ à $[x_{2,N}^-, x_{2,N}^+]$, et $\lambda_{M,N}$ est associée à $[x_{3,N}^-, x_{3,N}^+]$. Le support \mathcal{S}_N est donc ici composé de 3 clusters.

En figure 7, chaque valeur propre est associée à un cluster différent.

Les conditions pour lesquelles le support \mathcal{S}_N se sépare en plusieurs clusters dépendent de manière non triviale de σ , des valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ et de leurs espacements entre elles. Néanmoins, sous l'hypothèse des spiked models **H-2** (i.e. K indépendant de N et convergence des valeurs propres), nous avons obtenu des conditions explicites pour la séparation des valeurs propres : une valeur propre de $\mathbf{B}_N \mathbf{B}_N^*$ est séparée des autres si sa limite quand $N \rightarrow \infty$ est supérieure à $\sigma^2 \sqrt{c}$ (théorème 4). Les valeurs propres non séparées sont quant à elles associées au cluster $[x_{1,N}^-, x_{1,N}^+]$. Ainsi, dans le cas des spiked models, le comportement du support \mathcal{S}_N est complètement caractérisé.

En figure 8, la densité associée à μ_N est tracée dans le contexte où $N = 20$, $M = 10$, $\sigma = 1$ et $\mathbf{B}_N \mathbf{B}_N^*$ est diagonale avec 3 valeurs propres différentes : 0 de multiplicité 5, 5 de multiplicité 2 et 10 de multiplicité 3. La densité est tracée en calculant des valeurs de la fonction $x \rightarrow \pi^{-1} \text{Im}(m_N(x + iy))$ pour $y \ll 1$ (propriété 1). Notons que $m_N(z)$ pour $z \in \mathbb{C}^+$ est évaluée itérativement en utilisant l'équation point fixe (31), et cette procédure converge toujours, comme montré dans Hachem et al. [18].

En figure 9, la densité de μ_N est tracée dans le cas où $c_N = 0.5$, $\sigma = 1$ et où les valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ sont 0 (multiplicité $M - 2$), 5, 10, pour $N = 20, 100, 200, 2000$. Avec ces paramètres, nous avons donc $5 > \sigma^2 \sqrt{c_N}$ et les valeurs propres 5 et 10 sont donc associées à des clusters différents. La largeur de ces clusters tend vers 0 au rythme $\mathcal{O}(N^{-1/2})$ (théorème 4).

En figure 10, la fonction $w_N(x)$ est tracée pour $x \in \mathbb{R}$ et les mêmes paramètres que la figure 8. Plus précisément, la figure 10 représente $\text{Im}(w_N(x))$ en fonction de $\text{Re}(w_N(x))$. Comme mentionné dans la propriété 2, l'ensemble $w_N(\mathbb{R})$ ne rencontre aucune valeur propre, $w_N(x)$ est de partie imaginaire positive pour $x \in \mathcal{S}_N$ et est réel pour $x \in \mathbb{R} \setminus \mathcal{S}_N$. Si nous définissons $\mathcal{C}_{q,N} = \{w_N(x) : x \in [x_{q,N}^-, x_{q,N}^+]\} \cup \{w_N(x)^* : x \in [x_{q,N}^-, x_{q,N}^+]\}$, alors $\mathcal{C}_{q,N}$ entoure les valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ associées à $[x_{q,N}^-, x_{q,N}^+]$. Dans le présent exemple, la valeur propre 0 est associée à $[x_{1,N}^-, x_{1,N}^+]$, 5 est associée à $[x_{2,N}^-, x_{2,N}^+]$ et 10 à $[x_{3,N}^-, x_{3,N}^+]$. Ainsi, en figure 10, $\mathcal{C}_{1,N}$ entoure la valeur propre 0, et la valeur propre 5 est quant à elle entourée par $\mathcal{C}_{2,N}$, et 10 par $\mathcal{C}_{3,N}$. Cette propriété sera d'une importance fondamentale dans la section sur l'estimation sous-espace, dans laquelle nous utiliserons notamment une méthode basée sur des intégrales de contour.

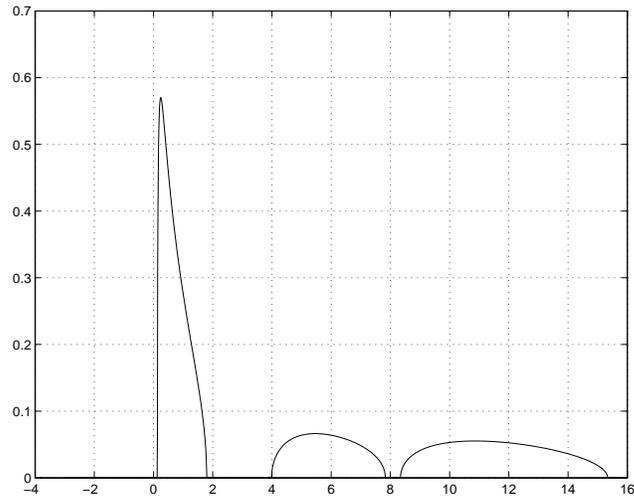


FIGURE 8 – Densité de μ_N

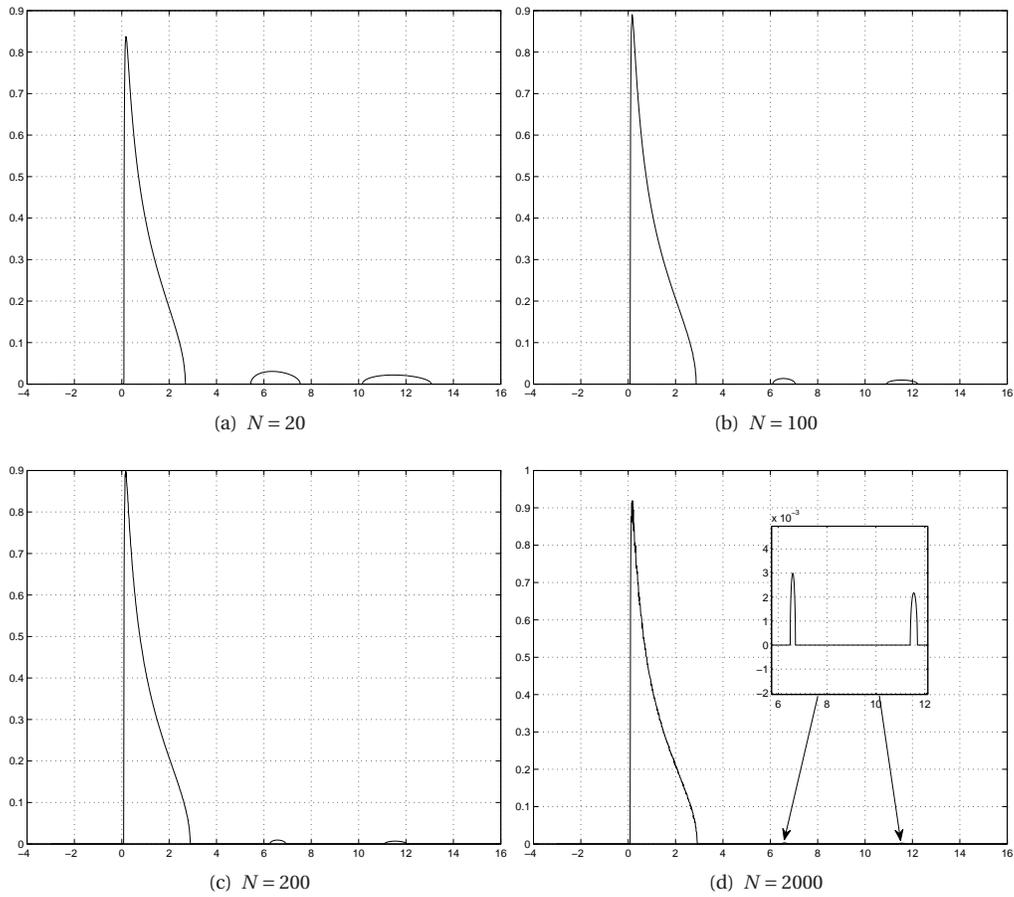


FIGURE 9 – Illustration de l'hypothèse des spiked models sur la densité de μ_N

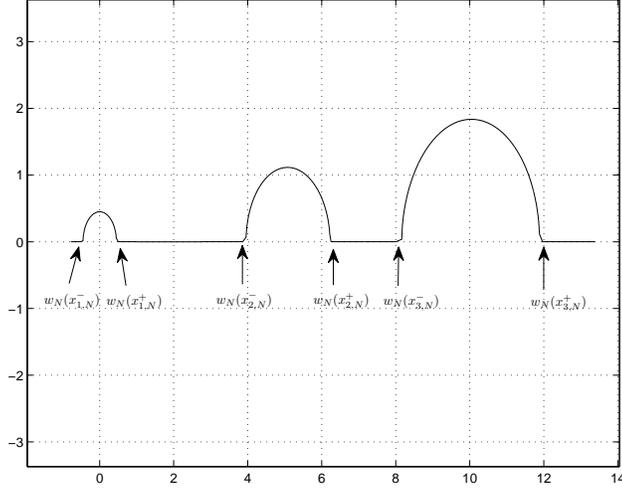


FIGURE 10 – $\text{Im}(w_N(x))$ versus $\text{Re}(w_N(x))$ pour $x \in \mathbb{R}$

4 Localisation du spectre des matrices i.i.d. gaussiennes non centrées

Cette courte section traite de la localisation presque-sûre des valeurs propres de $\Sigma_N \Sigma_N^*$ pour N grand. Nous établissons en premier lieu l'absence de valeurs propres de $\Sigma_N \Sigma_N^*$ en dehors du support \mathcal{S}_N de la mesure spectrale asymptotique μ_N . En complément, nous montrons que la probabilité qu'une valeur propre s'échappe d'un voisinage contenant \mathcal{S}_N pour N grand décroît au rythme $\frac{1}{N^p}$ pour tout $p \in \mathbb{N}$. Nous étudions un phénomène de séparation des valeurs propres, et plus précisément le fait que pour N suffisamment grand, les valeurs propres de $\Sigma_N \Sigma_N^*$ se séparent en plusieurs groupes liés aux différents clusters de \mathcal{S}_N .

4.1 Absence de valeurs propres en dehors du support

La méthode utilisée pour montrer l'absence presque sûre de valeurs propres en dehors du support est tirée de Haagerup & Thorbjornsen [17] et a également été utilisée par Capitaine et al. [9].

Théorème 5. *Soit $a, b \in \mathbb{R}$, et $\epsilon > 0$ tels que $(a - \epsilon, b + \epsilon) \cap \mathcal{S}_N = \emptyset$ pour N suffisamment grand. Alors, avec probabilité un,*

$$\text{card} \{k : \hat{\lambda}_{k,N} \in [a, b]\} = 0,$$

pour N suffisamment grand.

Le résultat du théorème 5 peut être amélioré en donnant un ordre de décroissance de la probabilité qu'une valeur propre de $\Sigma_N \Sigma_N^*$ s'échappe du support. Pour un compact $\mathcal{K} \subset \mathbb{R}$, nous noterons \mathcal{K}_ϵ l' ϵ -voisinage fermé de \mathcal{K} , i.e le compact

$$\mathcal{K}_\epsilon = \{x \in \mathbb{R} : \exists y \in \mathcal{K} \text{ s.t. } |x - y| \leq \epsilon\}.$$

Théorème 6. *Soit $\epsilon > 0$ un compact \mathcal{K} contenant \mathcal{S}_N pour N grand, et \mathcal{K}_ϵ l' ϵ -voisinage fermé de \mathcal{K} . Alors*

$$\mathbb{P}(\exists k : \hat{\lambda}_{k,N} \in \mathcal{K}_\epsilon^c) = \mathcal{O}\left(\frac{1}{N^l}\right),$$

pour tout $l \in \mathbb{N}$.

4.2 Séparation des valeurs propres

Dans cette section, nous étudions la séparation des valeurs propres de $\Sigma_N \Sigma_N^*$ en plusieurs groupes, liés aux valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$. Le résultat principal est le suivant.

Théorème 7. Soit $a, b \in \mathbb{R}$ et $\epsilon > 0$ tels que $(a - \epsilon, b + \epsilon) \cap \mathcal{S}_N = \emptyset$ pour N grand. Sous l'hypothèse **H-1**, avec probabilité un,

$$\text{card}\{k : \hat{\lambda}_{k,N} < a\} = \text{card}\{k : \lambda_{k,N} < w_N(a)\}, \quad (40)$$

$$\text{card}\{k : \hat{\lambda}_{k,N} > b\} = \text{card}\{k : \lambda_{k,N} > w_N(b)\}, \quad (41)$$

pour N suffisamment grand.

4.3 Le cas particulier des spiked models

Dans cette section, nous utilisons les résultats précédents sur la localisation des valeurs propres, pour calculer les limites explicites des plus grandes de ces valeurs propres dans le cas particulier des spiked models (voir section 2.4). Nous rappelons que sous l'hypothèse **H-2**, la quantité K_s a été définie Comme le nombre de valeurs propres limites de $\mathbf{B}_N \mathbf{B}_N^*$ strictement supérieures à $\sigma^2 \sqrt{c}$, et que la fonction ψ a été définie par $\psi(\lambda, c) = \frac{(\lambda + \sigma^2 c)(\lambda + c)}{\lambda}$. Le résultat principal est le suivant.

Théorème 8. Sous l'hypothèse **H-2**,

$$\hat{\lambda}_{M-K_s+k,N} \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \begin{cases} \sigma^2(1 + \sqrt{c})^2 & \text{pour } k = 0, \\ \psi(\gamma_k, c) & \text{pour } k = 1, \dots, K_s. \end{cases}$$

4.4 Discussions et exemples numériques

Dans cette section, nous discutons des résultats de localisation et séparation des valeurs propres de $\Sigma_N \Sigma_N^*$ en grandes dimensions (théorèmes 5, 6 et 7), ainsi que l'application aux spiked models (théorème 8), et donnons quelques illustrations numériques des phénomènes décrits.

Pour mieux comprendre le théorème 5, supposons par exemple que pour N grand, le nombre de clusters Q de \mathcal{S}_N est indépendant de N , et que pour chaque $q = 1, \dots, Q$, les suites de points frontières $(x_{q,N}^-)$ et $(x_{q,N}^+)$ convergent vers des limites x_q^- et x_q^+ , vérifiant $x_1^- < x_1^+ < x_2^- < x_2^+ < \dots < x_Q^- < x_Q^+$. Dans ce contexte, le théorème 5 entraîne que presque sûrement, pour tout $\epsilon > 0$, chaque valeur propre appartient à un des intervalles $[x_q^- - \epsilon, x_q^+ + \epsilon]$ pour N grand.

Pour interpréter le théorème 7, nous conservons les mêmes hypothèses. Dans ce cas, le résultat du théorème 7 signifie que presque sûrement pour N suffisamment grand, le nombre de valeurs propres de $\Sigma_N \Sigma_N^*$ appartenant à l'intervalle $[x_q^- - \epsilon, x_q^+ + \epsilon]$ coïncide avec le nombre de valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ associées au cluster $[x_{q,N}^-, x_{q,N}^+]$ pour N grand.

Ces deux faits sont illustrés en figure 11 où la densité de μ_N est tracée, avec les valeurs propres de $\Sigma_N \Sigma_N^*$. Les paramètres sont $N = 20$, $M = 10$, $\sigma = 1$ et la matrice $\mathbf{B}_N \mathbf{B}_N^*$ est diagonale avec comme valeurs propres 0 (de multiplicité 5), 5 (de multiplicité 2) et 10 (de multiplicité 3). Sur la figure 11, nous voyons clairement qu'aucune valeur propre de $\Sigma_N \Sigma_N^*$ n'est située en dehors du support de μ_N . Cette propriété est vérifiée en pratique pour de petites valeurs de N , ce qui confirme que l'échappement des valeurs propres en dehors du support est un événement dont la probabilité décroît très vite avec N (théorème 6). De plus, nous observons clairement le phénomène de séparation entre valeurs propres. En effet, les 3 clusters contiennent respectivement 5, 2 et 3 valeurs propres de $\Sigma_N \Sigma_N^*$. Notons de par les hypothèses fixées précédemment (**H-1**, que les valeurs propres non nulles de $\mathbf{B}_N \mathbf{B}_N^*$ sont supposées de multiplicité 1, mais le résultat du théorème 7 est bien sûr valable dans le cas général, en comptant chaque valeur $\mathbf{B}_N \mathbf{B}_N^*$ avec leur multiplicité respective.

En figure 12, nous avons représenté l'évolution de la densité de μ_N sous l'hypothèse des spiked models, quand $N = 20, 100, 200, 2000$. Les paramètres de l'exemple numérique en figure 11 ont été conservés, sauf les valeurs propres 5 et 10 qui sont à présent de multiplicité 1 (et donc la valeur propre 0 est de multiplicité $M - 2$). La densité de μ_N est donnée pour différentes valeurs de N , et les valeurs propres de $\Sigma_N \Sigma_N^*$ sont également représentées. Cet exemple est en accord avec les résultats des théorèmes 4 et 8.

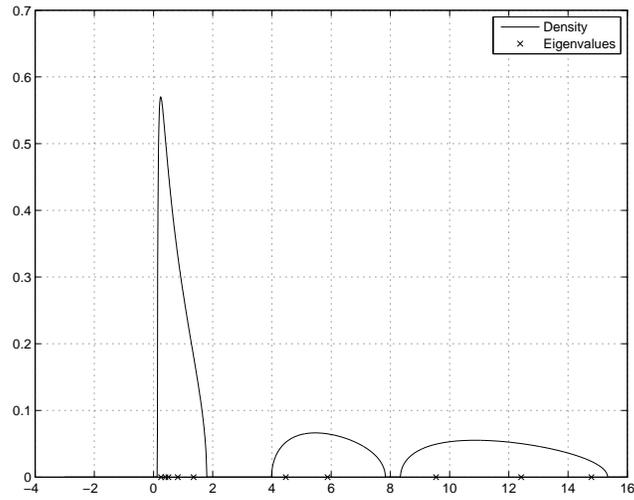
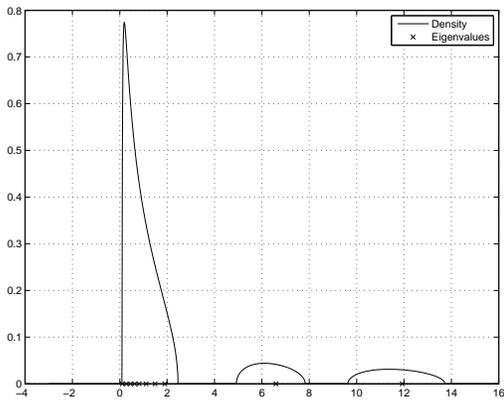
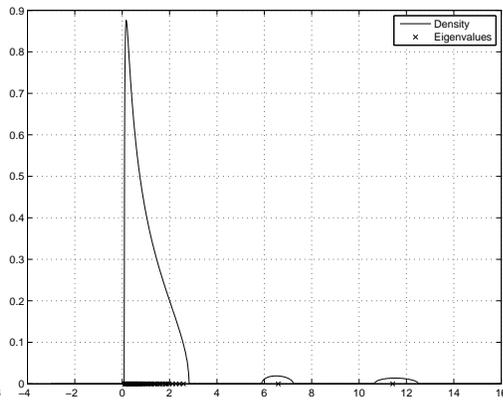


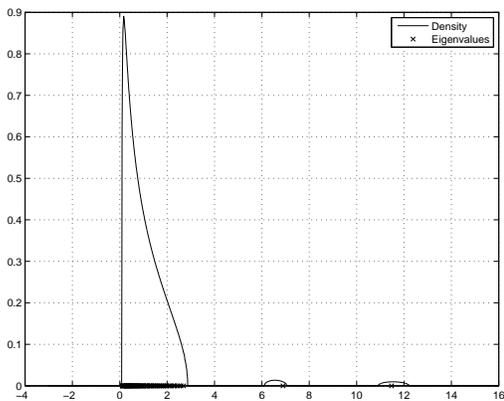
FIGURE 11 – Densité de μ_N et positions des valeurs propres de $\Sigma_N \Sigma_N^*$



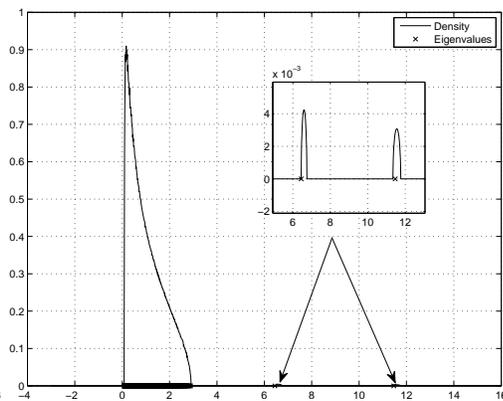
(a) $N = 20$



(b) $N = 100$



(c) $N = 200$



(d) $N = 2000$

FIGURE 12 – Effets de l'hypothèse des spiked models sur la densité de μ_N et le spectre de $\Sigma_N \Sigma_N^*$

5 Estimation sous-espace et localisation de sources dans les grands réseaux de capteurs

Ce chapitre est dédié à l'application des résultats des sections précédentes au problème de la localisation de sources par méthode sous-espace en traitement d'antennes.

En toute généralité, le problème consiste à inférer sur K paramètres à partir d'un ensemble de N observations multivariées de dimension M , qui sont composées d'une partie signal et d'une partie bruit. Habituellement, la partie signal a une matrice de corrélation de petit rang, tandis que la partie bruit a une matrice de corrélation de rang plein. Pour extraire les K paramètres d'intérêt des observations, une procédure classique est d'utiliser le fait que l'espace image de la matrice de corrélation des observations se sépare en deux sous-espaces orthogonaux : un sous-espace signal de dimension K (correspondant au sous-espace propre de la matrice de corrélation signal), et un sous-espace bruit de dimension $M - K$. En général, les estimateurs résultants sont relativement peu coûteux en terme de calculs que d'autres estimateurs comme ceux basés sur le maximum de vraisemblance, qui ont généralement de meilleures performances mais qui impliquent une minimisation dans un espace de paramètres multidimensionnel.

Pour formuler un estimateur sous-espace générique, nous devons en premier lieu inférer sur les vecteurs propres de la matrice de corrélation des observations, qui n'est pas disponible. En conséquence, les méthodes classiques d'estimation sous-espace utilisent la matrice de corrélation empirique des observations, et approximativement les vecteurs propres de la vraie matrice de corrélation par les vecteurs propres de l'estimée empirique. Cette procédure est clairement optimale quand le nombre d'échantillons N tend vers l'infini tandis que la dimension M des observations reste constante. En effet, sous certaines conditions d'ergodicité, quand $N \rightarrow \infty$ avec M constant, la matrice de corrélation empirique des observations converge presque sûrement vers la vraie matrice de corrélation, et en conséquence, quand $N \gg M$, les vecteurs propres de l'estimée empirique sont de bonnes approximations des vrais vecteurs propres. Cependant, en pratique, le nombre d'observations N et la dimension des observation M sont du même ordre de grandeur, ce qui conduit à des différences non négligeables entre les vecteurs propres de la matrice de corrélation empirique des observations et ceux de la vraie matrice de corrélation.

Le fait que les vecteurs propres empiriques ne sont pas les meilleurs estimateurs des vrais vecteurs propres est connu depuis des décennies, bien que l'étude d'alternatives aux estimateurs classiques n'ait été conduite exclusivement que dans le régime où $N \gg M$. Cependant, Mestre [38] a récemment montré que le régime où N et M sont du même ordre de grandeur peut être étudié en considérant le régime asymptotique où M, N convergent vers l'infini au même rythme, i.e. $M, N \rightarrow +\infty$, de telle manière que le ratio $c_N = \frac{M}{N}$ converge vers une constante strictement positive. En utilisant la théorie des matrices aléatoires, Mestre [38] a montré que les estimateurs traditionnels sous-espaces sont asymptotiquement biaisés dans ce nouveau régime. Dans ce contexte, la théorie des matrices aléatoires s'avère utile pour quantifier la différence entre les vecteurs propres de la matrice de corrélation des observations et leurs estimées empiriques dans un scénario où M, N sont du même ordre de grandeur, et pour construire des estimateurs alternatifs des vecteurs propres qui convergent, non seulement dans le régime où $N \rightarrow +\infty$ et M constant, mais également dans le régime où $M, N \rightarrow +\infty$ même rythme. De telles considérations ont notamment été traitées dans [37] et [36], qui considèrent respectivement l'étude des vecteurs propres empiriques quand $M, N \rightarrow +\infty$ au même rythme, et proposent des estimateurs consistants de ces quantités dans ce nouveau régime asymptotique.

Cependant, [37] et [36] ne s'appliquent pas au modèle information plus bruit considéré ici, à moins que les signaux source en jeu ne soient des vecteurs aléatoires gaussiens i.i.d. dans le domaine temporel. En pratique, il existe de nombreuses applications où les observations ne suivent pas cette structure, et sont plutôt modélisés comme une composante déterministe (correspondant à la partie signal), plus un bruit additif généralement gaussien. Ceci constitue le modèle de matrices aléatoires dit "information plus bruit", étudié dans les sections précédentes. Les études [38, 37, 36] ont considérées le modèle centré corrélé, développé dans Silverstein [28], Le but de cette section est donc de proposer un estimateur sous-espace amélioré pour le modèle information plus bruit, qui représentera la situation où les signaux sources sont modélisés comme des suites déterministes et non observables, en utilisant les résultats des sections précédentes.

5.1 Modèle statistique et estimation sous-espace traditionnelle

Dans cette section, nous introduisons le modèle statistique usuel associé au problème de l'estimation sous-espace en traitement d'antennes. Nous considérons la situation où K signaux sources déterministes à bande étroite $(s_k)_{k=1,\dots,K}$ sont reçus par un réseau d'antennes composés de M éléments, $K < M$. Le signal reçu à l'instant discret n , noté \mathbf{y}_n (et de dimension M) peut être modélisé par

$$\mathbf{y}_n = \mathbf{A}\mathbf{s}_n + \mathbf{v}_n,$$

où \mathbf{A} est une matrice complexe $M \times K$ dont les entrées représentent l'atténuation entre les K signaux sources et les M antennes de réception, \mathbf{s}_n est un vecteur complexe de dimension K contenant les signaux transmis par les K sources à l'instant n , et \mathbf{v}_n un bruit blanc additif gaussien complexe de moyenne nulle et de matrice de covariance $\mathbb{E}[\mathbf{v}_n \mathbf{v}_n^*] = \sigma^2 \mathbf{I}_M$. Nous supposons que la période d'observations s'étend de $n = 1$ à $n = N$, et que $c_N = \frac{M}{N} < 1$. Pour la suite, nous adopterons les conventions des sections précédentes, i.e $M = M(N) < N$ et $K = K(N) < M$ fonctions de N .

Nous notons $\mathbf{Y}_N = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ la matrice des observations (de taille $M \times N$) qui peut évidemment se réécrire

$$\mathbf{Y}_N \triangleq \mathbf{A}\mathbf{S}_N + \mathbf{V}_N, \quad (42)$$

avec $\mathbf{S}_N = [\mathbf{s}_1, \dots, \mathbf{s}_N]$ et $\mathbf{V}_N = [\mathbf{v}_1, \dots, \mathbf{v}_N]$. La matrice de covariance empirique des observations est donnée par $\frac{1}{N} \mathbf{Y}_N \mathbf{Y}_N^*$, et la matrice de covariance empirique des observations non bruitées est donnée quant à elle par $\frac{1}{N} \mathbf{A}\mathbf{S}_N \mathbf{S}_N^* \mathbf{A}^*$.

Pour simplifier les notations, nous posons $\mathbf{\Sigma}_N = N^{-1/2} \mathbf{Y}_N$, $\mathbf{B}_N = N^{-1/2} \mathbf{A}\mathbf{S}_N$ et $\mathbf{W}_N = N^{-1/2} \mathbf{V}_N$ de telle manière que le modèle (42) peut se réécrire

$$\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{W}_N, \quad (43)$$

Sous l'hypothèse où \mathbf{A} et \mathbf{S}_N sont de rang plein K , nous retrouvons les principales hypothèses faites sur le modèle information plus bruit en section 2 i.e

- \mathbf{B}_N est une matrice déterministe de rang K ,
- \mathbf{W}_N est une matrice aléatoire dont les entrées sont i.i.d. $\mathcal{N}_C(0, \sigma^2/N)$.

Remarque 2. Dans le contexte de la localisation de sources (estimation des directions d'arrivées des K sources), un modèle typique pour la matrice \mathbf{A} est donnée par $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$ où $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ sont les angles d'arrivées des K sources émettant sur le réseau d'antennes. La colonne $\mathbf{a}(\theta_k)$ est appelée dans ce contexte "vecteur directionnel" de la k -ième source et dépend de l'angle d'arrivée θ_k et la géométrie des antennes. Ce modèle sera utilisé en section 5.3 dans le contexte de l'algorithme MUSIC.

Nous supposons sans perte de généralité que les valeurs propres non nulles de $\mathbf{B}_N \mathbf{B}_N^*$ sont de multiplicité un. Nous notons, comme dans les sections précédentes, $\hat{\lambda}_{1,N} \leq \dots \leq \hat{\lambda}_{M,N}$ et $0 = \lambda_{1,N} = \dots = \lambda_{M-K,N} < \lambda_{M-K+1} < \dots < \lambda_{M,N}$ les valeurs propres respectives de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ et $\mathbf{B}_N \mathbf{B}_N^*$. Les vecteurs propres associés sont $\hat{\mathbf{u}}_{1,N}, \dots, \hat{\mathbf{u}}_{M,N}$ et $\mathbf{u}_{1,N}, \dots, \mathbf{u}_{M,N}$.

Dans la terminologie de l'estimation sous-espace, "sous-espace bruit" designera $\text{span}\{\mathbf{u}_{1,N}, \dots, \mathbf{u}_{M-K,N}\}$, i.e le sous-espace propre associé à la valeur propre 0 de $\mathbf{B}_N \mathbf{B}_N^*$ et "sous-espace signal" le sous-espace supplémentaire, i.e le sous-espace propre associé aux valeurs propres non nulles de $\mathbf{B}_N \mathbf{B}_N^*$. Le but de l'estimation sous-espace est donc d'inférer un de ces deux sous-espaces, en s'intéressant aux projecteurs. Ainsi, le but est d'estimer la matrice de projection sur le sous-espace bruit, i.e $\mathbf{\Pi}_N = \sum_{k=1}^{M-K} \mathbf{u}_{k,N} \mathbf{u}_{k,N}^*$. Ce problème implique notamment un grand nombre de paramètres à estimer, si M, N sont grands, et on préférera donc estimer dans ce cas des formes bilinéaires de ce projecteur, i.e des quantités comme $\mathbf{d}_{1,N}^* \mathbf{\Pi}_N \mathbf{d}_{2,N}$, où $\mathbf{d}_{1,N}$ et $\mathbf{d}_{2,N}$ sont deux vecteurs déterministes. Par une identité de polarisation classique, le problème se ramène à estimer n'importe quelle forme quadratique de $\mathbf{\Pi}_N$. Le problème d'estimation sous-espace que nous considérerons ici est donc de trouver un estimateur consistant de

$$\eta_N = \mathbf{d}_N^* \mathbf{\Pi}_N \mathbf{d}_N, \text{ quand } N \rightarrow \infty, \quad (44)$$

où (\mathbf{d}_N) est une suite de vecteurs déterministes tels que $\sup_N \|\mathbf{d}_N\| < \infty$. Traditionnellement, η_N est estimé par $\hat{\eta}_N = \mathbf{d}_N^* \hat{\mathbf{\Pi}}_N \mathbf{d}_N = \sum_{k=1}^{M-K} \mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N$, i.e en remplaçant les vecteurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ par leurs estimées

empiriques. Cet estimateur est pertinent dans le cas où M est indépendant de N (et donc $c_N \rightarrow_N 0$), car la loi des grands nombres donne dans ce cas

$$\|\Sigma_N \Sigma_N^* - (\mathbf{B}_N \mathbf{B}_N^* + \sigma^2 \mathbf{I}_M)\| \xrightarrow[N \rightarrow \infty]{p.s.} 0.$$

Cependant, ce résultat de convergence n'est pas vérifié en général, si $c_N \rightarrow_N c > 0$ (voir les résultats de la section 2, ou les résultats sur les spiked models section 4.3 pour un contre-exemple immédiat). En particulier, il peut être montré que $\eta_N - \hat{\eta}_N$ ne converge pas vers 0.

Le but de la section suivante est construire un estimateur consistant de η_N en utilisant les résultats concernant la convergence des formes bilinéaires de la résolvante de $\Sigma_N \Sigma_N^*$, fournis en section 2.

5.2 Estimation sous-espace généralisée

Dans cette section, nous faisons deux hypothèses supplémentaires, portant sur la séparation entre les valeurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ associées au sous-espace bruit et celles associées au sous-espace signal. Nous renvoyons le lecteur à la section 2, pour une étude de la fonction w_N et la caractérisation du support \mathcal{S}_N de la mesure μ_N .

Hypothèse H-3: *Pour N suffisamment grand, $\lambda_{M-K+1,N} > w_N(x_{2,N}^-)$, i.e les valeurs propres non nulles de $\mathbf{B}_N \mathbf{B}_N^*$ ne sont pas associées au premier cluster $[x_{1,N}^-, x_{1,N}^+]$ de \mathcal{S}_N .*

Hypothèse H-4: *Il existe $t_{i,N}^-, t_{i,N}^+ > 0$ ($i = 1, 2$) tels que*

$$0 < t_1^- < \liminf_N x_{1,N}^- < \limsup_N x_{1,N}^+ < t_1^+ < t_2^- < \liminf_N x_{2,N}^- < \limsup_N x_{2,N}^+ < t_2^+.$$

L'hypothèse H-4 est essentiellement technique et l'hypothèse H-3 est la plus importante, son but, qui sera plus clair par la suite, est de ne prendre en compte aucune contribution du sous-espace signal.

5.2.1 Résultats préliminaires

Le résultat suivant améliore certains résultats de convergence donnés en section 2, et notamment permet de travailler avec des convergences uniformes sur certains ensembles compacts.

Lemme 1. *Soit $\mathcal{K} \subset \mathbb{C} \setminus ([t_1^-, t_1^+] \cup [t_2^-, \infty))$ un compact de \mathbb{C} . Alors sous l'hypothèse H-4,*

$$\begin{aligned} \sup_{z \in \mathcal{K}} |\hat{m}_N(z) - m_N(z)| &\xrightarrow[N \rightarrow \infty]{p.s.} 0, \\ \sup_{z \in \mathcal{K}} |\hat{m}'_N(z) - m'_N(z)| &\xrightarrow[N \rightarrow \infty]{p.s.} 0, \end{aligned} \quad (45)$$

et pour $(\mathbf{d}_{1,N}), (\mathbf{d}_{2,N})$ comme au théorème 2

$$\sup_{z \in \mathcal{K}} |\mathbf{d}_{1,N}^* (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{d}_{2,N}| \xrightarrow[N \rightarrow \infty]{p.s.} 0. \quad (46)$$

Finalement, si $0 \notin \mathcal{K}$,

$$\sup_{z \in \mathcal{K}} \left| \frac{1}{1 + \sigma^2 c_N \hat{m}_N(z)} - \frac{1}{1 + \sigma^2 c_N m_N(z)} \right| \xrightarrow[N \rightarrow \infty]{p.s.} 0. \quad (47)$$

Nous introduisons à présent de nouvelles quantités. Soit la matrice de taille $M \times M$

$$\hat{\mathbf{\Omega}}_N = \hat{\mathbf{\Lambda}}_N + \frac{\sigma^2 c_N}{M} \mathbf{1} \mathbf{1}^T, \quad (48)$$

avec $\hat{\mathbf{\Lambda}}_N = \text{Diag}(\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{M,N})$ et $\mathbf{1} = [1, \dots, 1]^T$. Nous notons $\hat{\omega}_{1,N}, \dots, \hat{\omega}_{M,N}$ les valeurs propres de $\hat{\mathbf{\Omega}}_N$ (en ordre croissant). Avec cette définition, les solutions de l'équation $1 + \sigma^2 c_N \hat{m}_N(z) = 0$ sont incluses dans l'ensemble $\{\hat{\omega}_{1,N}, \dots, \hat{\omega}_{M,N}\}$. Ceci peut être vu en écrivant le polynôme caractéristique de $\hat{\mathbf{\Omega}}_N$, i.e

$$\det(\hat{\mathbf{\Omega}}_N - z \mathbf{I}_M) = \det(\hat{\mathbf{\Lambda}}_N - z \mathbf{I}_M) (1 + \sigma^2 c_N \hat{m}_N(z))$$

Plus précisément, si $\hat{\lambda}_{k,N}$ est de multiplicité i , i.e. $\hat{\lambda}_{k-1,N} < \hat{\lambda}_{k,N} = \dots = \hat{\lambda}_{k+i-1,N} < \hat{\lambda}_{k+i,N}$, alors

$$\hat{\omega}_{k-1,N} < \hat{\lambda}_{k,N} = \hat{\omega}_{k,N} = \dots = \hat{\lambda}_{k+i-1,N} = \hat{\omega}_{k+i-1,N} < \hat{\omega}_{k+i,N} < \hat{\lambda}_{k+i,N}.$$

Remarque 3. Les valeurs propres $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$ sont de multiplicité un presque sûrement (car elles admettent une densité jointe). Ceci implique donc que presque sûrement $\hat{\omega}_{1,N} < \dots < \hat{\omega}_{M,N}$, et ainsi l'équation

$$1 + \sigma^2 c_N \hat{m}_N(z) = 0$$

admet M solutions satisfaisant

$$\hat{\lambda}_{1,N} < \hat{\omega}_{1,N} < \dots < \hat{\lambda}_{M,N} < \hat{\omega}_{M,N}.$$

La remarque précédente implique que si les valeurs propres $\hat{\omega}_{1,N}, \dots, \hat{\omega}_{M,N}$ apparaissent dans un énoncé impliquant un événement de probabilité un, on pourra toujours supposer qu'elles sont de multiplicité un. Le principal résultat concerne le fait que $\hat{\omega}_{1,N}, \dots, \hat{\omega}_{M,N}$ vérifient une propriété de séparation, similaire au théorème 7. En appliquant les résultats du théorème 7, les hypothèses **H-3** et **H-4** impliquent immédiatement qu'avec probabilité un,

$$\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{M-K,N} \in [t_1^-, t_1^+] \quad \text{et} \quad \hat{\lambda}_{M-K+1,N}, \dots, \hat{\lambda}_{M,N} \in [t_2^-, t_2^+], \quad (49)$$

pour N suffisamment grand. Nous avons donc également le résultat suivant.

Corollaire 1. Sous les hypothèses **H-1**, **H-3** et **H-4**, avec probabilité un pour N grand, nous avons

$$\hat{\omega}_{1,N}, \dots, \hat{\omega}_{M-K,N} \in [t_1^-, t_1^+] \quad \text{et} \quad \hat{\omega}_{M-K+1,N} \geq t_2^-. \quad (50)$$

5.2.2 Le cas général

Les résultats énoncés précédemment nous permettent à présent de construire un estimateur consistant de $\eta_N = \sum_{k=1}^{M-K} \mathbf{d}_N^* \mathbf{u}_{k,N} \mathbf{u}_{k,N}^* \mathbf{d}_N$ avec $(\mathbf{d}_N)_{N \geq 1}$ une suite de vecteurs déterministes tels que $\sup_N \|\mathbf{d}_N\| < \infty$. La construction de l'estimateur est donnée pour montrer notamment l'utilité des résultats développés précédemment.

Théorème 9. Définissons

$$\hat{\eta}_{\text{new},N} = \frac{1}{2\pi i} \oint_{\partial \mathcal{R}^-} \mathbf{d}_N^* \mathbf{Q}_N(z) \mathbf{d}_N \frac{\hat{w}'_N(z)}{1 + \sigma^2 c \hat{m}_N(z)} dz,$$

où \mathcal{R} est le rectangle $\mathcal{R} = \{u + iv : u \in [t_1^- - \delta, t_1^+ + \delta], v \in [-y, y]\}$, avec $\delta > 0$ tel que $t_1^- - \delta > 0$ et $t_1^+ + \delta < t_2^-$, et $y > 0$. Alors, sous les hypothèses **H-1**, **H-3** et **H-4**,

$$\hat{\eta}_{\text{new},N} - \eta_N \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0,$$

où $\hat{\eta}_{\text{new},N} = \sum_{k=1}^M \hat{\xi}_{k,N} \mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N$ p.s. pour N grand, avec, pour $k = 1, \dots, M-K$,

$$\hat{\xi}_{k,N} = 1 + \frac{\sigma^2 c_N}{M} \sum_{l=M-K+1}^M \frac{\hat{\lambda}_{k,N} + \hat{\lambda}_{l,N}}{(\hat{\lambda}_{k,N} - \hat{\lambda}_{l,N})^2} + \sigma^2 (1 - c_N) \sum_{l=M-K+1}^M \left(\frac{1}{\hat{\lambda}_{k,N} - \hat{\lambda}_{l,N}} - \frac{1}{\hat{\lambda}_{k,N} - \hat{\omega}_{l,N}} \right), \quad (51)$$

et pour $k = M-K+1, \dots, M$,

$$\hat{\xi}_{k,N} = -\frac{\sigma^2 c_N}{M} \sum_{l=1}^{M-K} \frac{\hat{\lambda}_{k,N} + \hat{\lambda}_{l,N}}{(\hat{\lambda}_{k,N} - \hat{\lambda}_{l,N})^2} - \sigma^2 (1 - c_N) \sum_{l=1}^{M-K} \left(\frac{1}{\hat{\lambda}_{k,N} - \hat{\lambda}_{l,N}} - \frac{1}{\hat{\lambda}_{k,N} - \hat{\omega}_{l,N}} \right). \quad (52)$$

Démonstration. De part les propriétés de la fonction w_N (propriété 2), l'ensemble $\mathcal{W}_{1,N} = \{w_N(z) : z \in \partial \mathcal{R}\}$ est un lacet \mathcal{C}^1 rencontrant l'axe réel aux points $w_N(t_1^- - \delta)$, $w_N(t_1^+ + \delta)$, entourant l'intervalle $]w_{1,N}^-, w_{1,N}^+[$ et

laissant $[w_N(t_2^-), +\infty[$ en dehors, pour N grand. L'hypothèse **H-3** impliquent que 0 est l'unique valeur propre de $\mathbf{B}_N \mathbf{B}_N^*$ entourée par $\mathcal{W}_{1,N}$ pour N grand. Le théorème des résidus entraîne que

$$\begin{aligned}\eta_N &= \sum_{k=1}^{M-K} \mathbf{d}_N^* \mathbf{u}_{k,N} \mathbf{u}_{k,N}^* \mathbf{d}_N \\ &= \frac{1}{2\pi i} \oint_{\mathcal{W}_{1,N}^-} \mathbf{d}_N^* (\mathbf{B}_N \mathbf{B}_N^* - \lambda \mathbf{I}_M)^{-1} \mathbf{d}_N d\lambda,\end{aligned}$$

où $\mathcal{W}_{1,N}^-$ signifie que $\mathcal{W}_{1,N}$ est orienté dans le sens anti-trigonométrique. Ceci mène directement à

$$\eta_N = \frac{1}{2\pi i} \oint_{\partial\mathcal{R}^-} \mathbf{d}_N^* (\mathbf{B}_N \mathbf{B}_N^* - w_N(z) \mathbf{I}_M)^{-1} \mathbf{d}_N w'_N(z) dz.$$

En utilisant l'égalité $(1 + \sigma^2 c m_N(z)) (\mathbf{B}_N \mathbf{B}_N^* - w_N(z) \mathbf{I}_M)^{-1} = \mathbf{T}_N(z)$ (voir l'équation 31), nous pouvons écrire

$$\eta_N = \frac{1}{2\pi i} \oint_{\partial\mathcal{R}^-} g_N(z) dz,$$

avec $g_N(z) = \mathbf{d}_N^* \mathbf{T}_N(z) \mathbf{d}_N \frac{w'_N(z)}{1 + \sigma^2 c m_N(z)}$. A présent, le point-clé de la preuve est basé sur l'observation que $g_N(z)$ peut être estimé de façon consistante (et uniforme) à partir des éléments de la matrice de covariance empirique $\Sigma_N \Sigma_N^*$. En effet, nous avons

$$\sup_{z \in \partial\mathcal{R}} |\hat{g}_N(z) - g_N(z)| \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0. \quad (53)$$

où

$$\hat{g}_N(z) = \mathbf{d}_N^* \mathbf{Q}_N(z) \mathbf{d}_N \frac{\hat{w}'_N(z)}{1 + \sigma^2 c \hat{m}_N(z)},$$

avec $\hat{w}'_N(z)$ la dérivée de $\hat{w}_N(z) = z(1 + \sigma^2 c_N \hat{m}_N(z))^2 - \sigma^2(1 - c_N)(1 + \sigma^2 c_N \hat{m}_N(z))$. La convergence (53) est une conséquence directe du lemme 1. Ceci implique que

$$\left| \frac{1}{2\pi i} \oint_{\partial\mathcal{R}^-} (\hat{g}_N(z) - g_N(z)) dz \right| \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0.$$

Pour conclure cette preuve, il reste à vérifier que $\hat{\eta}_{\text{new},N}$ défini au théorème 9 vérifie

$$\hat{\eta}_{\text{new},N} = \frac{1}{2\pi i} \oint_{\partial\mathcal{R}^-} \hat{g}_N(z) dz. \quad (54)$$

\hat{g}_N est une fraction rationnelle et ses pôles sont

- $\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{M,N}$ les valeurs propres de $\Sigma_N \Sigma_N^*$,
- $\hat{\omega}_{1,N}, \dots, \hat{\omega}_{M,N}$ les zéros de $z \mapsto 1 + \sigma^2 c_N \hat{m}_N(z)$.

En conséquence, grâce aux résultats de localisation des pôles (49) et (50), nous obtenons par le théorème des résidus

$$\frac{1}{2\pi i} \oint_{\partial\mathcal{R}^-} \hat{g}_N(z) dz = \sum_{k=1}^{M-K} (\text{Ind}_{\partial\mathcal{R}^-}(\hat{\lambda}_{k,N}) \text{Res}(\hat{g}_N, \hat{\lambda}_{k,N}) + \text{Ind}_{\partial\mathcal{R}^-}(\hat{\omega}_{k,N}) \text{Res}(\hat{g}_N, \hat{\omega}_{k,N})),$$

avec $\text{Res}(\hat{g}_N, \lambda)$ le résidu de la fonction \hat{g}_N au point λ et $\text{Ind}_{\partial\mathcal{R}^-}(\lambda)$ l'indice de $\partial\mathcal{R}^-$ autour de λ . Un calcul de résidus long mais direct permet d'obtenir la formule du théorème. \square

Le nouvel estimateur consistant introduit au théorème 9 peut être vu comme une extension du travail de Mestre [38], qui suppose les signaux utiles comme des suites i.i.d. gaussiennes mutuellement indépendantes.

Nous remarquons que le nouvel estimateur consistant $\hat{\eta}_{\text{new},N}$ est une combinaison linéaire des termes $\mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N$ pour $k = 1, \dots, M$ et en contraste avec l'estimateur traditionnel $\hat{\eta}_{\text{trad},N} = \sum_{k=1}^{M-K} \mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N$, ce nouvel estimateur inclue des contributions à la fois du sous-espace bruit et du sous-espace signal de la matrice de covariance empirique.

Remarque 4. Il convient de remarquer que lorsque le nombre d'échantillons est bien plus grand que le nombre d'antennes ($N \gg M$ ou de manière équivalente $c_N \rightarrow_N 0$), le nouvel estimateur converge vers l'estimateur traditionnel $\hat{\eta}_N$. Ceci peut être vu en prenant la limite $c_N \rightarrow_N 0$ des coefficients (51) et (52), et en remarquant que $\hat{\omega}_{l,N} \rightarrow_N \hat{\lambda}_{l,N}$ quand $c_N \rightarrow_N 0$. Ainsi, lorsque $c_N \rightarrow_N 0$, nous avons $\hat{\xi}_{k,N} \rightarrow_N 1$ pour $k = 1, \dots, M-K$, et $\hat{\xi}_{k,N} \rightarrow_N 0$ pour $k = M-K+1, \dots, M$, et donc $\hat{\eta}_{\text{new},N} - \hat{\eta}_{\text{trad},N} \rightarrow_N 0$. Ceci montre que l'estimateur proposé est en fait une généralisation de l'estimateur traditionnel.

5.2.3 Le cas particulier des spiked models

Dans cette section, nous considérons le cas particulier des spiked models, précédemment introduit en section 2.4 et dont les principales implications sur les valeurs propres ont été données en section 4.3. En utilisant l'hypothèse **H-2**, il est possible de construire, à partir du théorème 9, un estimateur simplifié.

Théorème 10. Soit

$$\hat{\eta}_{\text{spike},N} = \mathbf{d}_N^* \hat{\mathbf{\Pi}}_N \mathbf{d}_N + \sum_{k=M-K+1}^M \mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N \left(1 - \frac{\Gamma'(\hat{\lambda}_{k,N})}{\Gamma(\hat{\lambda}_{k,N}) m(\hat{\lambda}_{k,N})} \right), \quad (55)$$

où $\Gamma(x) = xm(x)\tilde{m}(x)$, et $m(x)$ est la transformée de Stieltjes de la loi de Marcenko-Pastur, définie en (39) et $\tilde{m}(x) = cm(x) - \frac{1-c}{x}$. Alors sous l'hypothèse **H-2**, si $\lim_N \lambda_{M-K+1,N} = \gamma_1 > \sigma^2 \sqrt{c}$, nous avons

$$\hat{\eta}_{\text{spike},N} - \eta_N \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0.$$

La preuve de ce résultat (omise ici) repose notamment en partie sur le lemme suivant, qui montre que les valeurs propres $(\hat{\omega}_{M-K+k,N})_{k=0,\dots,K}$ ont le même comportement que les valeurs propres $(\hat{\lambda}_{M-K+k,N})_{k=0,\dots,K}$.

Lemme 2. Sous l'hypothèse **H-2**, si $\lim_N \lambda_{M-K+1,N} = \gamma_1 > \sigma^2 \sqrt{c}$, nous avons

$$\hat{\omega}_{M-K+k,N} \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \psi(\gamma_k, c),$$

pour $k = 1, \dots, K$ et $\hat{\omega}_{M-K,N} \xrightarrow[N \rightarrow \infty]{\text{p.s.}} \sigma^2(1 + \sqrt{c})^2$.

5.3 Estimation des angles d'arrivées

En section 5.1, nous avons considéré la matrice des observations de taille $M \times N$

$$\mathbf{Y}_N = \mathbf{A} \mathbf{S}_N + \mathbf{V}_N,$$

où \mathbf{S}_N représente les signaux déterministes inconnus, \mathbf{V}_N le bruit blanc gaussien et \mathbf{A} la matrice $M \times K$ des vecteurs directionnels, contenant les coefficients de transmission entre les K sources et les M antennes de réception. Comme mentionné dans la remarque 2, un modèle usuelle pour la matrice \mathbf{A} est donné par

$$\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)],$$

où $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]$ contient les angles d'arrivée des K sources.

Le problème classique de localisation de sources en traitement d'antennes consiste à estimer le vecteur $\boldsymbol{\theta}$ à partir des N échantillons collectés dans la matrice \mathbf{Y}_N . L'estimateur dit "sous-espace" de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^T$ repose sur l'observation que si les matrices $\mathbf{A}(\boldsymbol{\theta})$ et \mathbf{S}_N sont toutes les deux de rang plein K , alors les angles $\theta_1, \dots, \theta_K$ sont solutions⁴ de l'équation $\mathbf{a}(\theta)^* \mathbf{\Pi}_N \mathbf{a}(\theta) = 0$, où nous rappelons que $\mathbf{\Pi}_N$ est le projecteur orthogonal sur le noyau de la matrice $\mathbf{A}(\boldsymbol{\theta}) \mathbf{S}_N \mathbf{S}_N^* \mathbf{A}(\boldsymbol{\theta})^*$. Dans ce contexte, l'algorithme MUSIC (MUltiple Signal Classification), proposé par Schmidt [27], consiste à estimer pour chaque θ la quantité $\eta_N(\theta) = \mathbf{a}(\theta)^* \mathbf{\Pi}_N \mathbf{a}(\theta)$ of $\mathbf{\Pi}_N$ par un certain $\hat{\eta}_N(\theta)$, et d'estimer les K angles d'arrivée comme les antécédents des K plus petits minima de la fonction $\theta \rightarrow \hat{\eta}_N(\theta)$.

4. Ceci vient du fait que $\text{span}\{\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)\} = \text{span}\{\mathbf{u}_{M-K+1,N}, \dots, \mathbf{u}_{M,N}\}$, ou de manière équivalente que le supplémentaire de l'image de \mathbf{A} coïncide avec le sous-espace bruit. Notons que les K angles sont les solutions uniques sous certaines conditions sur la fonction $\theta \rightarrow \mathbf{a}(\theta)$.

Cette approche a été développée pour le régime $N \rightarrow +\infty$ et M constant. Comme explicité en section 5.1, $\eta_N(\theta)$ peut être estimé de façon consistante dans ce contexte, pour chaque θ , par $\hat{\eta}_N(\theta) = \mathbf{a}(\theta)^* \hat{\mathbf{\Pi}}_N \mathbf{a}(\theta)$ avec $\hat{\mathbf{\Pi}}_N$ la matrice de projection orthogonale sur le sous-espace propre associé aux $M - K$ plus petites valeurs propres de la matrice $\frac{1}{N} \mathbf{Y}_N \mathbf{Y}_N^*$. On a alors

$$\sup_{\theta \in [-\pi, \pi]} |\hat{\eta}_N(\theta) - \eta_N(\theta)| \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0, \quad (56)$$

et (56) implique directement que les estimateurs correspondants des angles d'arrivée sont consistants. Evidemment, comme mentionné en section 5.1, cette convergence n'est plus vraie sous les hypothèses données en section 2.2.

Nous montrons ici que les estimées des angles d'arrivées obtenus depuis l'estimateur sous-espace amélioré du théorème 9 sont consistants quand $N \rightarrow \infty$. Nous considérons un réseau linéaire et uniforme d'antennes, dont les éléments sont espacés de la moitié de la longueur d'onde. Les vecteurs directionnels $\mathbf{a}(\theta)$ sont donc donnés par

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \left[1, e^{i\pi \sin(\theta)}, \dots, e^{i(M-1)\pi \sin(\theta)} \right]^T, \quad (57)$$

pour $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Cependant, nous considérerons le modèle

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \left[1, e^{i\theta}, \dots, e^{i(M-1)\theta} \right]^T, \quad (58)$$

pour $\theta \in [-\pi, \pi]$, équivalent à (57), mais plus simple à manipuler.

La consistance des angles d'arrivées nécessite de disposer d'une consistance uniforme comme en (56).

Proposition 1. *Sous les hypothèses H-1, H-3 et H-4, et en supposant que les vecteurs directionnels suivent le modèle (58)*

$$\sup_{\theta \in [-\pi, \pi]} |\hat{\eta}_{\text{new}, N}(\theta) - \eta_N(\theta)| \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0.$$

Remarque 5. *La proposition 1 reste vraie pour des fonctions plus générale que (58). En particulier, la preuve de la proposition 1 peut être facilement adaptée pour toute fonction $\mathbf{a}(\theta) : [-\pi, \pi] \rightarrow \mathbb{C}^M$ telle que*

$$\|\mathbf{a}(\theta) - \mathbf{a}(\theta')\| \leq CN^r |\theta - \theta'|^s \text{ et } \sup_N \sup_{\theta \in [-\pi, \pi]} \|\mathbf{a}(\theta)\|. \quad (59)$$

avec $s > 0$, $r \geq 0$ et $C > 0$ une constante positive.

La consistance uniforme de la proposition 1 peut être transférée aux estimées angulaires. Pour définir les estimateurs de $\theta_1, \dots, \theta_K$ de manière rigoureuse, nous devons supposer que le nombre de sources K est indépendant de N , et nous considérons ainsi K intervalles disjoints $\mathcal{I}_1, \dots, \mathcal{I}_K$, tels que $\theta_k \in \mathcal{I}_k$, et définissons pour chaque k l'estimateur $\hat{\theta}_{k, N}$ de θ_k par $\hat{\theta}_{k, N} = \operatorname{argmin}_{\theta \in \mathcal{I}_k} |\hat{\eta}_{\text{new}, N}(\theta)|$. Il est ainsi possible de montrer le résultat suivant.

Théorème 11. *Sous les hypothèses H-3 et H-4, et si $K > 0$ est indépendant de N , alors,*

$$N(\hat{\theta}_{k, N} - \theta_k) \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0,$$

pour $k = 1, \dots, K$

5.3.1 Discussions et illustrations numériques

Nous considérons un réseau d'antennes linéaire et uniforme dont les éléments sont espacés de la moitié de la longueur d'onde, et dont le modèle de vecteurs directionnels $\mathbf{a}(\theta)$ est donné par (57). Dans les exemples numériques qui suivent, les signaux sources sont des réalisations de processus AR(1) mutuellement indépendants, de variance 1 et de corrélation 0.9. Le RSB est défini comme $10 \log(\sigma^{-2})$. Le bruit additif varie de tirages.

Comparaison avec l'estimateur traditionnel et l'estimateur de Mestre Nous comparons en premier lieu les résultats fournis par l'estimateur sous-espace traditionnel, le nouvel estimateur $\hat{\eta}_{\text{new},N}(\theta)$ défini par $\hat{\eta}_{\text{new},N}$ pour $\mathbf{d}_N = \mathbf{a}(\theta)$ (référéncé dans les figures comme "conditional estimator"), et l'estimateur amélioré de Mestre [38] construit sous l'hypothèse que les sources sont des processus i.i.d. gaussiens. (référéncé sur les figures comme "unconditional estimator"), et dénoté $\hat{\eta}_{\text{unc},N}(\theta)$. Les estimées angulaires correspondantes (définies comme les antécédents des K plus petits minima locaux du module de la fonction de localisation estimée) sont notés respectivement "Trad-MUSIC", "Conditional G-MUSIC" et "Unconditional G-MUSIC" dans cette section. La "condition de séparation" signifie que la valeur propre 0 de la matrice $\mathbf{B}_N \mathbf{B}_N^*$ est séparée des clusters correspondants aux valeurs propres non nulles, i.e. pour chaque σ^2 , M et N , $0 < w_{1,N}^+ < w_{2,N}^- < \lambda_{M-K+1,N}$.

Nous mentionnons que l'estimateur de Mestre [38] n'est pas supposé être consistant dans le contexte des exemples suivats, car les signaux sources ne sont pas i.i.d temporellement. Cependant, nous verrons que les performances des estimateurs améliorés $\hat{\eta}_{\text{new},N}(\theta)$ et $\hat{\eta}_{\text{unc},N}(\theta)$ sont très proches, une propriété qui nécessite de plus amples recherches (voir les remarques ci-dessous).

- Dans l'exemple 1, nous considérons 2 sources très proches, à $\theta_1 = 16^\circ$ et $\theta_2 = 18^\circ$. Le nombre d'antennes est $M = 20$ et le nombre d'échantillons $N = 40$. La condition de séparation est vérifiée si le RSB est supérieur à 10 dB. Dans le but d'évaluer les performances des estimées de la fonction de localisation, pour chaque estimateur amélioré (conditional et unconditional), nous traçons en fonction de θ en figure 13 le ratio de l'EQM (erreur quadratique moyenne) de l'estimateur traditionnel de $\mathbf{a}(\theta)^H \mathbf{\Pi} \mathbf{a}(\theta)$, i.e. $\hat{\eta}_{\text{trad},N}(\theta) = \mathbf{a}(\theta)^H \hat{\mathbf{\Pi}} \mathbf{a}(\theta)$ sur l'EQM de l'estimateur amélioré $\hat{\eta}_{\text{new},N}(\theta)$. Le RSB est égal à 16 dB. La figure 13 montre que les 2 estimateurs améliorés ont les mêmes performances, et qu'ils améliorent considérablement l'estimateur traditionnel autour des 2 angles. Cependant, les 3 estimateurs ont les mêmes performances si θ est éloigné de $\theta_1 = 16^\circ$ et $\theta_2 = 18^\circ$. Pour évaluer plus précisément les améliorations

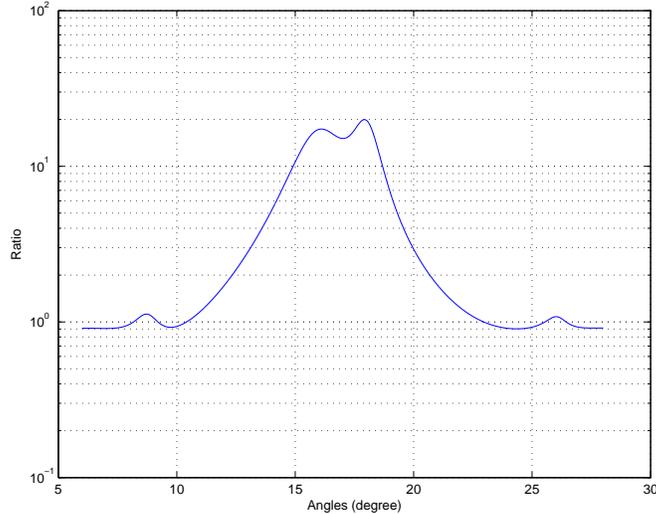


FIGURE 13 – Ratio (en dB) de l'EQM de $\hat{\eta}_{\text{trad},N}(\theta)$ sur l'EQM de $\hat{\eta}_{\text{new},N}(\theta)$ en fonction de θ .

apportées par les estimateurs "conditional" et "unconditional" autour de θ_1 et θ_2 , nous traçons les quantités $\frac{1}{2} \sum_{k=1}^2 \mathbb{E} |\hat{\eta}_{\text{trad},N}(\theta_k) - \eta_N(\theta_k)|^2$, $\frac{1}{2} \sum_{k=1}^2 \mathbb{E} |\hat{\eta}_{\text{new},N}(\theta_k) - \eta_N(\theta_k)|^2$ et $\frac{1}{2} \sum_{k=1}^2 \mathbb{E} |\hat{\eta}_{\text{unc},N}(\theta_k) - \eta_N(\theta_k)|^2$ en fonction du RSB (notons que $\eta_N(\theta_k) = 0$), en figure 14. En figure 15, nous traçons pour chaque méthode la moyenne des EQM des 2 angles estimées, i.e. $0.5(\mathbb{E} |\hat{\theta}_1 - \theta_1|^2 + \mathbb{E} |\hat{\theta}_2 - \theta_2|^2)$, où $\hat{\theta}_{1,2}$ représentent les angles estimés. La moyenne des deux bornes de Cramer-Rao est aussi tracée. Les performances des estimateurs améliorés sont encore une fois similaires, et montrent une amélioration de 4 dB par rapport à l'estimateur traditionnel dans la région de RSB 15dB-25dB. Nous traçons à présent la probabilité d'outlier, définie ici comme la probabilité qu'un des 2 angles estimés soit séparé du vrai angle de plus de la moitié de la distance entre les 2 vrais angles θ_1 et θ_2 . En figure 16, nous comparons la probabilité d'outlier des 3 méthodes en fonction du RSB. Pour une probabilité d'outlier cible de 0.5, les deux estimateurs améliorés fournissent une amélioration de 8 dB par rapport à l'estimateur traditionnel. Pour conclure l'exemple 1, nous évaluons l'influence de M et N sur les performances. Nous faisons varier N de 20 à 200 tout en

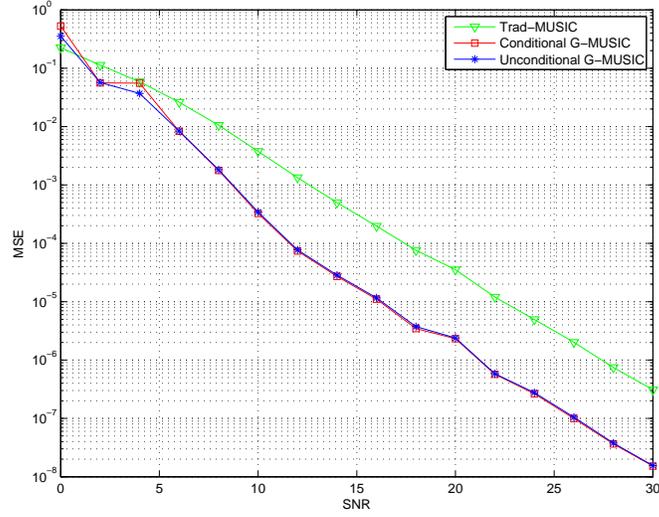


FIGURE 14 – Moyennes des EQM des estimées de $\eta_N(\theta_1) = \mathbf{a}(\theta_1)^H \mathbf{\Pi} \mathbf{a}(\theta_1)$ et $\eta_N(\theta_2) = \mathbf{a}(\theta_2)^H \mathbf{\Pi} \mathbf{a}(\theta_2)$.

maintenant le ratio c_N constant à 0.5, et le RSB à 15 dB. En figure 17 nous avons tracé la moyenne des EQM des estimées de $\eta_N(\theta_k) = \mathbf{a}(\theta_k)^H \mathbf{\Pi} \mathbf{a}(\theta_k)$ pour $i = 1, 2$. La condition de séparation a lieu pour $N \geq 32$. La figure 17 illustre clairement la non-consistance de l'estimateur traditionnel.

- Pour l'exemple 2, nous supposons à présent que le nombre de sources K est du même ordre de grandeur que M et N , i.e. $K = 10, M = 20, N = 40$. Les 10 angles $(\theta_k)_{k=1, \dots, 10}$ sont $\theta_k = -40^\circ + (k-1)10^\circ$ pour $k = 1, \dots, 10$. La condition de séparation a lieu si le RSB est supérieur à 15 dB. Nous traçons en fonction de θ en figure 18 le ratio de l'EQM de l'estimateur traditionnel de la fonction de localisation sur l'EQM des estimateurs améliorés. Le RSB est de 16 dB. La figure 18 montre l'amélioration apportés par le nouvel estimateur et l'estimateur de Mestre, en particulier autour des vrais angles $(\theta_k)_{k=1, \dots, 10}$. La figure 19 représente la moyenne des EQM des différents estimateurs de $\eta_N(\theta_k) = \mathbf{a}(\theta_k)^H \mathbf{\Pi} \mathbf{a}(\theta_k)$ pour $k = 1, \dots, 10$ en fonction du RSB, et montre la supériorité des deux estimateurs améliorés quand la condition de séparation est vérifiée.

Les précédents tracés montrent clairement la supériorité du nouvel estimateur sur l'estimateur traditionnel, mais sa différence avec l'estimateur de Mestre est négligeable. C'est un fait relativement surprenant. En effet, l'estimateur de Mestre a été construit [38] sous l'hypothèse que la matrice \mathbf{S}_N a des entrées i.i.d gaussiennes de variance 1. Dans ce contexte, les entrées de $(\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - z \mathbf{I}_M)^{-1}$ ont le même comportement que les entrées de la matrice $\mathbf{T}_{\text{iid}, N}(z)$ donnée par l'équation

$$m_{\text{iid}, N}(z) = \frac{1}{M} \text{Tr} \mathbf{T}_{\text{iid}, N}(z),$$

$$\mathbf{T}_{\text{iid}, N}(z) = [(\mathbf{A} \mathbf{A}^* + \sigma^2 \mathbf{I}_M) (1 - c_N - c_N z m_{\text{iid}, N}(z)) - z \mathbf{I}_M]^{-1}.$$

on peut vérifier aisément que les entrées de $\mathbf{T}_N(z)$ données par (33), qui dépendent de \mathbf{S}_N , ont le même comportement asymptotique que les entrées de $\mathbf{T}_{\text{iid}, N}(z)$ quand \mathbf{S}_N est la réalisation d'une matrice i.i.d. Dans ce cas, le nouvel estimateur et l'estimateur de Mestre ont bien entendu le même comportement. Cependant, si \mathbf{S}_N n'est pas une matrice à entrées i.i.d, alors les entrées de $(\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - z \mathbf{I}_M)^{-1}$ ne se comportent plus comme les entrées de $\mathbf{T}_{\text{iid}, N}(z)$, et l'estimateur de Mestre n'est plus consistant en principe. Les exemples précédents tendent à montrer que ce n'est pas le cas. La compréhension de ce phénomène est un sujet pour de futures recherches.

Comparaison avec l'estimateur (55) Nous présentons à présent quelques comparaisons numériques entre les estimateurs sous-espaces $\hat{\eta}_{\text{new}, N}(\theta)$ et $\hat{\eta}_{\text{spike}, N}(\theta)$ défini par (55) pour $\mathbf{d}_N = \mathbf{a}(\theta)$.

- Nous considérons une fois de plus les paramètres de l'exemple 1 précédent, i.e. 2 sources à $\theta_1 = 16^\circ$ et $\theta_2 = 18^\circ$. Le nombre d'antennes est $M = 20$ et le nombre d'échantillons est $N = 40$. En figure 20, nous

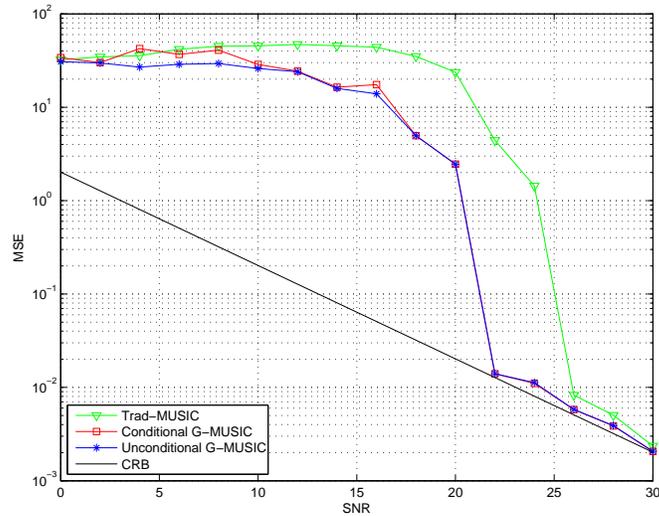


FIGURE 15 – Moyenne des EQM des estimées angulaires en fonction du RSB

évaluons par méthode de Monte-Carlo la quantité $0.5(\mathbb{E}|\hat{\theta}_1 - \theta_1|^2 + \mathbb{E}|\hat{\theta}_2 - \theta_2|^2)$, qui est la moyenne des EQM des 2 angles estimés, en fonction du RSB. La performance des 2 estimateurs est très proche. La moyennes des bornes de Cramer-Rao est également tracée. En figure 21, nous calculons par méthode de Monte-Carlo $\frac{1}{2} \sum_{k=1}^2 \mathbb{E}|\hat{\eta}(\theta_k) - \eta(\theta_k)|^2$, i.e la moyenne des EQM de la fonction de localisation estimée, évaluée aux vrais angles. Pour un RSB supérieur à 10 dB, la performance $\hat{\eta}_{\text{spike},N}$ est proche une fois de plus de $\hat{\eta}_{\text{new},N}$.

- Dans l'exemple 3, nous considérons $K = 5$ sources à -20° , -10° , 0° , 10° et 20° , et M, N sont toujours égaux à 20 et 40. La condition de séparation est vérifiée pour toute valeur du RSB comprise entre 10 dB et 30 dB. En figure 22, nous traçons le même graphe qu'en figure 21. Nous remarquons que l'estimateur $\hat{\eta}_{\text{spike},N}$ n'est plus proche de l'estimateur $\hat{\eta}_{\text{new},N}$ notamment car le ratio K/N n'est plus négligeable. Cependant, l'estimateur spike reste meilleur que l'estimateur traditionnel.

Références

- [1] NI Akhiezer. *The Classical Moment Problem*. Oliver & Boyd, 1965.
- [2] T.W. Anderson. *An introduction to multivariate statistical analysis*. John Wiley & Sons, 1958.
- [3] Z.D. Bai and J.W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of Probability*, 26(1) :316–345, 1998.
- [4] Z.D. Bai and J.W. Silverstein. Exact separation of eigenvalues of large dimensional sample covariance matrices. *Annals of Probability*, 27(3) :1536–1555, 1999.
- [5] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, 33(5) :1643–1697, 2005.
- [6] J. Baik and J.W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6) :1382–1408, 2006.
- [7] F. Benaych-Georges and R.R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 2011.
- [8] P. Bianchi, M. Debbah, M. Maida, and J. Najim. Performance of statistical tests for single-source detection using random matrix theory. *IEEE Transactions on Information Theory*, 57(4) :2400–2419, 2011.
- [9] M. Capitaine and C. Donati-Martin. Strong asymptotic freeness of Wigner and Wishart matrices. *Indiana Univ. Math. Journal*, 56 :295–309, 2007.
- [10] C.N. Chuah, D.N.C. Tse, J.M. Kahn, and R.A. Valenzuela. Capacity scaling in mimo wireless systems under correlated fading. *IEEE Transactions on Information Theory*, 48(3) :637–650, 2002.

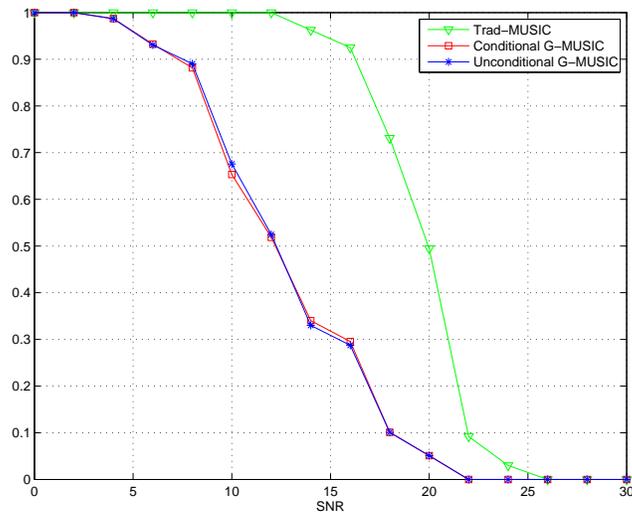


FIGURE 16 – Probabilité d'outlier en fonction du RSB

- [11] R.B. Dozier and J.W. Silverstein. Analysis of the limiting spectral distribution of large dimensional information plus noise type matrices. *Journal of Multivariate Analysis*, 98(6) :1099–1122, 2007.
- [12] R.B. Dozier and J.W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information plus noise type matrices. *Journal of Multivariate Analysis*, 98(4) :678–694, 2007.
- [13] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the Capacity Achieving Covariance Matrix for Rician MIMO Channels : An Asymptotic Approach. *IEEE Transactions on Information Theory*, 56(3) :1048–1069, 2010.
- [14] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, 36(6) :2757–2790, 2008.
- [15] V. L. Girko. *Theory of stochastic canonical equations. Vol. I*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.
- [16] V.L. Girko. *An introduction to statistical analysis of random arrays*. Vsp, 1998.
- [17] U. Haagerup and S. Thorbjornsen. A new application of random matrices : $\text{Ext}(C_{red}^*(F_2))$ is not a group. *Annals of Mathematics*, 162(2) :711, 2005.
- [18] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Annals of Applied Probability*, 17(3) :875–930, 2007.
- [19] W. Hachem, P. Loubaton, J. Najim, and P. Vallet. On bilinear forms based on the resolvent of large random matrices. *To appear in Annales de l'Institut Henri Poincaré, Probabilités et statistiques*, 2010. arXiv :1004.3848.
- [20] M. Krein and A. Nudelman. The markov moment problem and extremal problems. translations of mathematical monographs 50. *American Mathematical Society. Providence*, 1977.
- [21] S. Kritchman and B. Nadler. Non-parametric detection of the number of signals : Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, 57(10) :3930–3941, 2009.
- [22] V.A. Marcenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1 :457, 1967.
- [23] A.L. Moustakas, S.H. Simon, and A.M. Sengupta. Mimo capacity through correlated channels in the presence of correlated interferers and noise : A (not so) large n analysis. *IEEE Transactions on Information Theory*, 49(10) :2545–2561, 2003.
- [24] R.J. Muirhead. *Aspects of multivariate statistical theory*. Wiley Online Library, 1982.
- [25] R.R. Nadakuditi and A. Edelman. Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Transactions on Signal Processing*, 56(7) :2625–2638, 2008.

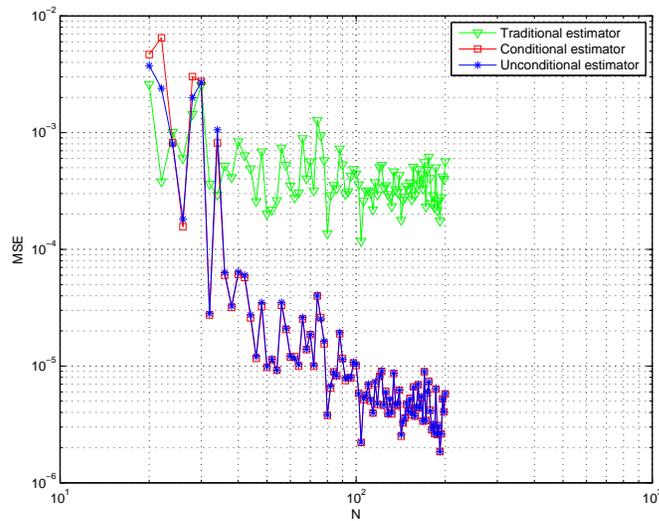


FIGURE 17 – EQM des estimateurs de la fonction de localisation en fonction de N

- [26] B. Nadler. Nonparametric Detection of Signals by Information Theoretic Criteria : performance analysis and an improved estimator. *IEEE Transactions on Signal Processing*, 58(5) :2746–2756, 2010.
- [27] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3) :276–280, 1986.
- [28] J.W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2) :331–339, 1995.
- [29] J.W. Silverstein and S. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, 52(2) :175–192, 1995.
- [30] J.W. Silverstein and P.L. Combettes. Signal detection via spectral theory of large dimensional random matrices. *IEEE Transactions on Signal Processing*, 40(8) :2100–2105, 1992.
- [31] P. Stoica and A. Nehorai. Music, maximum likelihood, and cramer-rao bound. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(5) :720–741, 1989.
- [32] E. Telatar. Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6) :585–595, 1999.
- [33] D.N.C. Tse and S.V. Hanly. Linear multiuser receivers : Effective interference, effective bandwidth and user capacity. *IEEE Transactions on Information Theory*, 45(2) :641–657, 1999.
- [34] S. Verdú and S. Shamai. Spectral efficiency of cdma with random spreading. *IEEE Transactions on Information Theory*, 45(2) :622–640, 1999.
- [35] E.P. Wigner. On the distribution of the roots of certain symmetric matrices. *The Annals of Mathematics*, 67(2) :325–327, 1958.
- [36] X.Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11) :5113–5129, 2008.
- [37] X.Mestre. On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. *IEEE Transactions on Signal Processing*, 56(11) :5353–5368, 2008.
- [38] X.Mestre and M.A.Lagunas. Modified subspace algorithms for DoA estimation with large arrays. *IEEE Transactions on Signal Processing*, 56(2) :598, 2008.

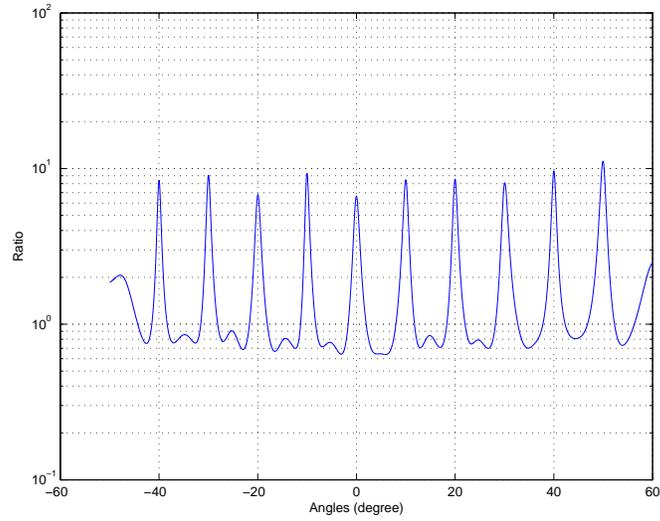


FIGURE 18 – Ratio (en dB) de l'EQM de l'estimateur traditionnel de la fonction de localisation sur l'EQM des estimateurs améliorés, en fonction de θ

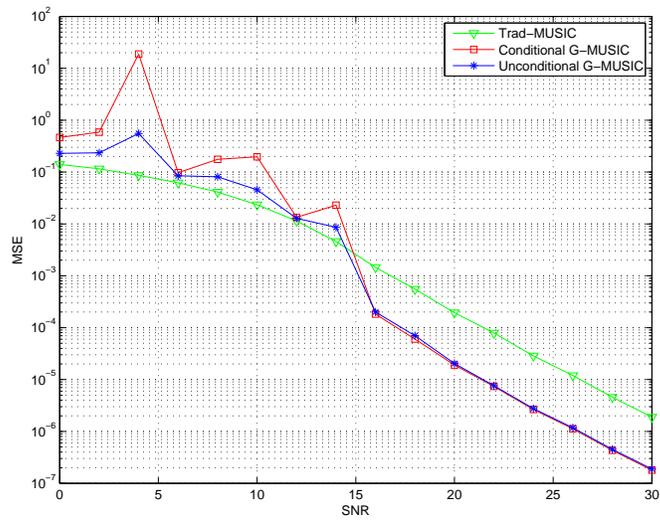


FIGURE 19 – Moyenne des EQM des estimateurs de $\mathbf{a}(\theta_k)^H \mathbf{\Pi} \mathbf{a}(\theta_k)$ pour $k = 1, \dots, 10$, en fonction du RSB.

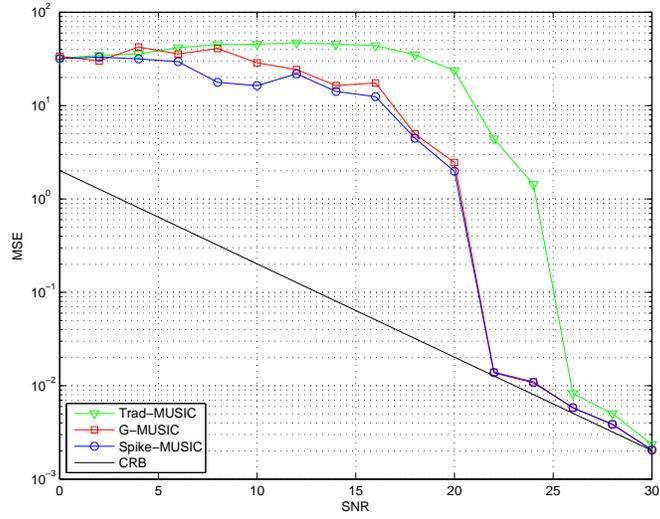


FIGURE 20 – Moyenne des EQM des 2 estimateurs angulaires

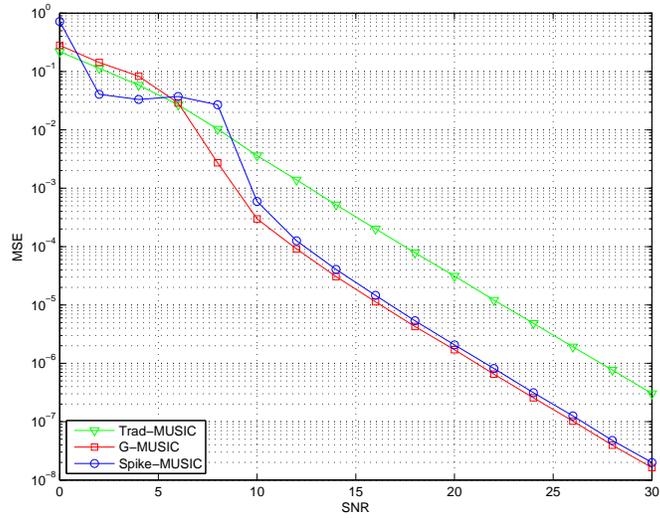


FIGURE 21 – Moyenne des MSE (fonction de localisation), $K = 2$.

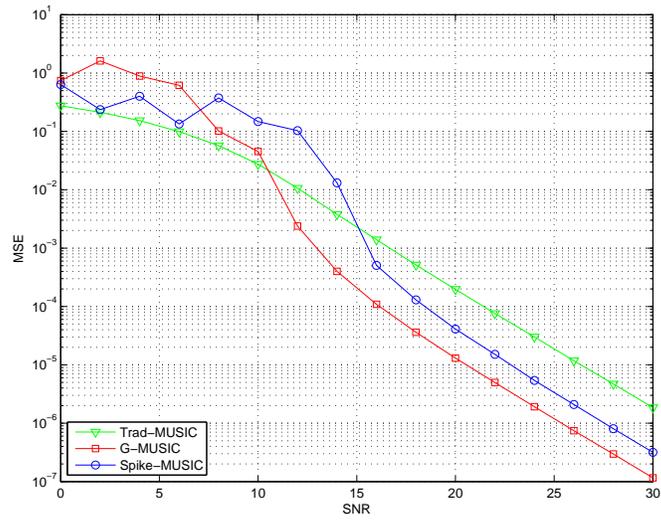


FIGURE 22 – Moyenne des EQM (fonction de localisation), $K = 5$.