THÈSE

soutenue le 28 novembre 2011 pour obtenir

le grade de **Docteur en Sciences** de l'Universté Paris-Est

Spécialité: Traitement du Signal

par

Pascal VALLET

Matrices aléatoires et applications au traitement statistique du signal

Random matrix theory and applications to statistical signal processing

Composition de la commision d'examen:

Président:	Mohammed NAJIM	Professeur, Université Bordeaux 1
Rapporteurs:	Mérouane DEBBAH	Professeur, Supélec
	Jian-Feng YAO	Professeur, University of Hong-Kong
Examinateurs:	Xavier MESTRE	Chercheur, CTTC
	Philippe FORSTER	Professeur, Université Paris-Ouest/Nanterre
Directeur de Thèse:	Philippe LOUBATON	Professeur, Université Paris-Est/Marne-la-Vallée

Contents

Remerciements

Mes premiers remerciements iront sans conteste à Philippe Loubaton. Je souhaite à tous les futurs doctorants un directeur de thèse avec autant de qualités que Philippe. Avec quatre étudiants en parallèle, des cours et des responsabilités administratives, Philippe a toujours su se montrer disponible, de bonne humeur, et ses conseils avisés ont évité qu'un certain nombre de talks ou de papiers ne tournent au vinaigre. Il a également constamment fait preuve d'une grande patience, notamment durant la période où je comblais de nombreuses lacunes accumulées au fil des ans. Sur le plan scientifique, le travail avec Philippe a été un vrai plaisir. Sa grande culture m'a permis de découvrir un nombre conséquent de nouveaux sujets, et j'ai aussi pu découvrir une méthode et une rigueur de travail, que j'espère pouvoir appliquer dans ma future carrière. Philippe, pour toutes ces choses, un grand merci à toi.

I also want especially to thank Xavier Mestre. My Ph.D topic is based on the beautiful results he obtained some years ago. I also had the pleasure to work with him, and I will be always impressed by his ingeniousness for solving intricky problems with simple ideas. I also thank him for welcoming me at CTTC during my post-doc, and for spending so much time to answer my e-mails and fill numerous administrative papers. Xavier, I am really happy you accepted to be in my Ph.D committee, muchas gracias por todo!

Ce fut également un réel plaisir de travailler avec Walid Hachem et Jamal Najim. Je les remercie profondément pour leur gentillesse et pour m'avoir intégré si vite. J'espère sincèrement pouvoir continuer à travailler avec eux dans l'avenir.

Mérouane Debbah et Jian-Feng Yao ont toute ma gratitude pour avoir accepté d'être rapporteurs de cette thèse, et je les remercie pour leurs commentaires positifs. J'ai également pu rencontrer Mérouane à différentes réunions sur les matrices aléatoires, mais également lors de tribulations à Séoul en 2009 dont je garderai le souvenir de franches rigolades (parfois méchamment épicées!). Merci également à Mohammed Najim et Philippe Forster d'avoir accepté de faire partie de mon jury de thèse.

Cette thèse a été financée par le CNRS que je remercie pour la confiance accordée. Je n'aurai probablement jamais eu l'occasion de faire cette thèse si je n'avais pas croisé la route de Jean-François Bercher et Antoine Chevreuil. Jean-François a éveillé mon intérêt pour le traitement de signal grâce à la grande qualité des cours qu'il a dispensé à l'ESIEE. Je remercie Antoine (et également J-F) pour le stage de 2007, qui a déclenché la suite des événements.

Durant ces trois années, j'ai travaillé au sein l'équipe signal du laboratoire IGM de Marne-la-Vallée. Je remercie tous les collègues pour la bonne ambiance qu'ils ont su apporter chaque jour: Florian, colocataire thésard du 4B048, avec qui j'ai eu de franches rigolades mais également des discussions scientifiques enrichissantes, Abdellatif, avec qui ce fut un plaisir de discuter, les thésards (ex ou actuels) Cédric, Lotfi, Nelly, Elena, Mounir, Anna, Mireille, et également Caroline pour sa bonne humeur de tous les jours, Emilie et Jean-Christophe. Line Fonfrede et Patrice Herault méritent amplement des remerciements, pour leur aide et leurs grandes compétences dans les domaines administratif et informatique, respectivement.

Mes derniers remerciements seront consacrés à mes proches. Je remercie mes parents pour leur soutien et toute leur aide, même si ça n'a pas été facile pour eux de me voir prendre une direction autre que celle de l'industrie il y a trois ans, mes beaux-parents, pour leur constante bonne humeur, leur aide et tous leurs bons petits plats, ainsi que tous mes amis, pour leur encouragement et pour avoir pris régulièrement des nouvelles. Finalement, mes plus grands remerciements iront à Ngoc-Tu, ma femme, qui pendant trois ans a fait preuve du plus grand altruisme, en supportant les périodes difficiles et en faisant l'impasse sur de nombreux projets, ainsi qu'en apportant une grande stabilité dans ma vie personnelle. Em yêu, cam on cho nhũng gì em dã làm cho anh.

Introduction

Matrices aléatoires et applications

Depuis le début du XXème 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 souvent sur l'utilisation de la matrice de covariance empirique des observations. Si les N observations statistiques sont représentées par les vecteurs $M-$dimensionnels $\mathbf{y}_1, \ldots, \mathbf{y}_N$, 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 [33]).

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 N. 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 $M, N \to \infty$ de telle sorte que $M/N \to c > 0$.

Les premières applications des matrices aléatoires aux statistiques remontent à la fin des années 1980, avec les travaux de Girko (voir par exemple [21]), fondateur de la théorie de la "G-estimation" (G pour "Generalized"). Dans le cas de modèles dits "à covariance", les observations $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{C}^M$ s'écrivent

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

où $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 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

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

avec $\hat{\lambda}_1, \dots, \hat{\lambda}_M$ les valeurs propres de $\hat{\mathbf{R}}$, est proche, quand $M, N \to \infty$ et $M/N \to c > 0$, d'une distribution déterministe. Plus précisément (Silverstein [39], Girko [21]), on a presque sûrement

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

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 [18] en 2008, dans le cadre de l'estimation des valeurs propres de la matrice $\mathbf{R}$. Plus récemment en 2009, Mestre [49] 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 [43] 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

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

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

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

où $\hat{F}(\lambda)$ est la distribution empirique des valeurs propres de $\mathbf{HH}^*$. 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 [45], Verdu & Shamai [47]), 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. [13], Moustakas et al. [32], Dumont et al. [17]).

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 [41] 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 experimentalement 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 perturbés multiplicativement par des matrices de petit rang (voir Baik et al. [6], Baik & Silverstein [7]). 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 [34], Kritchman & Nadler [30]. Tout récemment, Bianchi et al. [10] 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 conduit dans le domaine de la localisation de sources. Ainsi, Mestre et Lagunas [51] 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 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 [48] 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 [31]. Considérons $M, N \in \mathbb{N}^*$ tels que $M = M(N)$ est une fonction de N, et $c_N = M/N \to c \in]0,1]$ quand $N \to \infty$, ainsi qu'une matrice

$$\boldsymbol{\Sigma}_N = \frac{\sigma}{\sqrt{N}} \mathbf{X}_N, \tag{2}$$

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 $\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*$ (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 $\hat{\lambda}_{1,N} \leq \ldots \leq \hat{\lambda}_{M,N}$ les valeurs propres de $\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*$ 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

$$\hat{\mu}_N - \nu_N \xrightarrow[N\to\infty]{\mathrm{e}} 0, \tag{3}$$

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

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

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

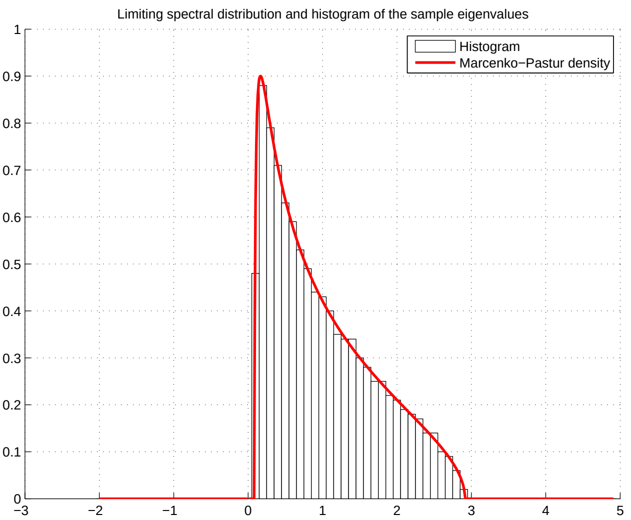


Figure 1: Densité de la loi ν_∞ pour $\sigma = 1$, $c = 0.5$, et histogramme des valeurs propres de $\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*$ dans le cas où $N = 2000$, $c_N = 0.5$ et $\sigma = 1$, pour une réalisation de la matrice aléatoire $\mathbf{X}_N$.

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

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

et consiste à prouver que $\hat{m}_N(z) - \kappa_N(z) \to_N 0$ 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 $\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^* - z\mathbf{I}_M\right)^{-1}$ par l'égalité

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

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 $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$, prises individuellement. Il a ainsi été montré en particulier que (voir Bai & Silverstein [4])

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

Les valeurs propres de $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$ sont ainsi toutes "absorbées" dans un voisinage du "bulk" $\left[\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2\right]$ à partir d'un certain rang N.

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,
- $\mathbf{s}_n = [s_{1,n}, \ldots, s_{n,K}]^T$ le vecteur des signaux transmis (non observable), supposé gaussien centré de covariance $\mathbf{R}_{\mathrm{S}} = \mathbb{E}[\mathbf{s}_n\mathbf{s}_n^*]$. Les signaux sont également supposés i.i.d. dans le domaine temporel.
- $\mathbf{v}_n$ un vecteur gaussien complexe (moyenne nulle, covariance $\sigma^2\mathbf{I}_M$), réprésentant le bruit de transmission indépendant du signal source.

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

$$\boldsymbol{\Sigma}_N = \frac{\mathbf{A}\mathbf{S}_N}{\sqrt{N}} + \mathbf{W}_N, \tag{5}$$

où $\mathbf{S}_N = [\mathbf{s}_1, \ldots, \mathbf{s}_N]$ et $\mathbf{W}_N = \frac{1}{\sqrt{N}}[\mathbf{v}_1, \ldots, \mathbf{v}_N]$. Dans le cas où aucun signal n'émet sur le réseau d'antennes, $\boldsymbol{\Sigma}_N = \mathbf{W}_N$ 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 $\left[\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2\right]$, pour un nombre d'antennes et d'échantillons suffisamment grands. En présence de signal, le modèle (5) est par contre équivalent à

$$\boldsymbol{\Sigma}_N = \boldsymbol{\Gamma}_N^{1/2}\mathbf{X}_N \tag{6}$$

où $\boldsymbol{\Gamma}_N = \mathbf{A}\mathbf{R}_{\mathrm{S}}\mathbf{A}^* + \sigma^2\mathbf{I}_M$ et $\mathbf{X}_N$ une matrice $M \times N$ dont les entrées sont complexes gaussiennes standards et indépendantes. Sous l'hypothèse que $\mathbf{A}$ et $\mathbf{R}_{\mathrm{S}}$ sont de rang plein, on a $\mathrm{rang}(\mathbf{A}\mathbf{R}_{\mathrm{S}}\mathbf{A}^*) = K$, et les $M-K$ plus petites valeurs propres de $\boldsymbol{\Gamma}_N$ sont toutes égales à σ^2 sauf les K plus grandes qui sont strictement supérieures. En posant $\sigma = 1$ sans perte de généralité, $\boldsymbol{\Gamma}_N$ 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 $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$, 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. [6], Baik & Silverstein [7]).

Considérons plus précisément le modèle (6), où $\boldsymbol{\Gamma}_N$ est une matrice déterministe definie positive de valeurs propres $\lambda_{1,N} \leq \ldots \leq \lambda_{M,N}$ 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 $c_N = M/N \to c \in]0,1]$, posons K un entier indépendant de N. On suppose de plus que les $M-K$ plus petites valeurs propres de $\boldsymbol{\Gamma}_N$ sont égales à 1, i.e. $\lambda_{1,N} = \ldots = \lambda_{M-K,N} = 1$, et que les K plus grandes valeurs propres $\lambda_{M-K+1,N}, \ldots, \lambda_{M,N}$ convergent vers des limites $1 < \gamma_1 \leq \ldots \leq \gamma_K$ quand $N \to \infty$. Comme $\boldsymbol{\Gamma}_N$ 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 $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$ 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 $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$ dans le cadre du modèle (6), alors on a toujours $\hat{\mu}_N - \nu_N \xrightarrow[N\to\infty]{\text{e}} 0$, où ν_N est donnée par (4) (Silverstein [39]). Neanmoins, quand

$N \to \infty$, toutes les valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ ne se concentrent pas autour du bulk $\left[\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})\right]$. En effet, il est montré (voir Baik & Silverstein [7]) que les K plus grandes valeurs propres $\hat{\lambda}_{M-K+1,N}, \ldots, \hat{\lambda}_{M,N}$ de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ peuvent converger vers des limites en dehors du bulk si $\gamma_1, \ldots, \gamma_K$, les valeurs propres limites correspondantes de $\mathbf{\Gamma}_N$, sont supérieures à un certain seuil. Plus précisément, on a pour $k = 1, \ldots, K$,

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

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

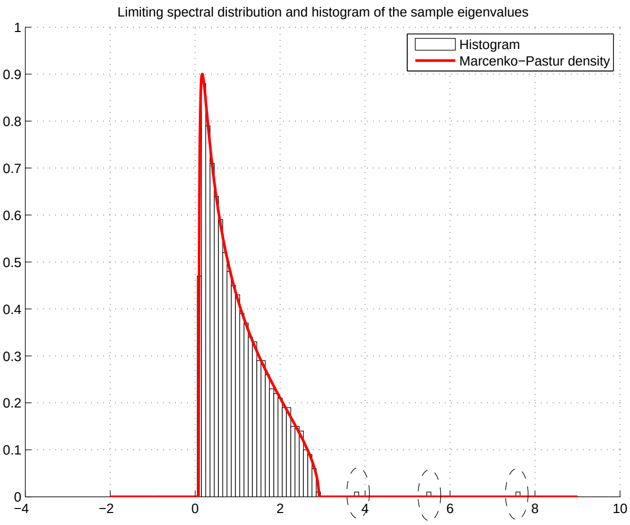


Figure 2: Densité de la loi de Marcenko-Pastur ν_∞ pour $\sigma = 1$, $c = 0.5$, et histogramme des valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ dans le cadre du modèle (6). Les paramètres sont $K = 3$, $N = 2000$, $c_N = 0.5$ et $\sigma = 1$, pour une réalisation de la matrice aléatoire $\mathbf{X}_N$. Les 3 plus grandes valeurs propres de $\mathbf{\Gamma}_N$ sont $3, 5, 7$, et les valeurs propres correspondantes de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ 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 $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ 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 [34], Kritchman & Nadler [30], Nadler [35], Bianchi et al. [10]).

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 $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. On considère donc le modèle (6) où $\mathbf{\Gamma}_N$ est une matrice définie positive telle que $\sup_N \|\mathbf{\Gamma}_N\| < \infty$. Dans ce contexte, il est montré (Silverstein [39]) que la mesure spectrale empirique $\hat{\mu}_N$ vérifie

$$\hat{\mu}_N - \mu_N \xrightarrow[N\to\infty]{\text{e}} 0, \tag{8}$$

presque sûrement, où μ_N est une mesure de probabilité déterministe et absolument continue de support compact (Silverstein & Choi [40]). De manière équivalente,

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

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

$$m_N(z) = \frac{1}{M}\mathrm{Tr}\left(\mathbf{\Gamma}_N(1 - c_N - c_N z m_N(z)) - z\mathbf{I}_M\right)^{-1}, \text{ pour tout } z \in \mathbb{C}\backslash\mathrm{supp}(\mu_N), \tag{10}$$

avec $\mathrm{supp}(\mu_N)$ 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 $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ 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 $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. Néanmoins, il est tout de même possible de mettre en évidence un phénomène de séparation des valeurs propres. En effet, définissons

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

Alors Bai & Silverstein [5] ont montré que, pour un intervalle $[a,b]$ tel que $[a-\epsilon, b+\epsilon] \cap \mathrm{supp}(\mu_N) = \emptyset$, pour un certain $\epsilon > 0$ et à partir d'un certain rang N, on a $w_N(a) < w_N(b)$ et

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

avec probabilité 1, pour N suffisamment grand. Autrement dit, si par exemple les valeurs propres de $\mathbf{\Gamma}_N$ sont scindées en deux groupes suffisamment séparés, de telle manière que

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

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

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

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

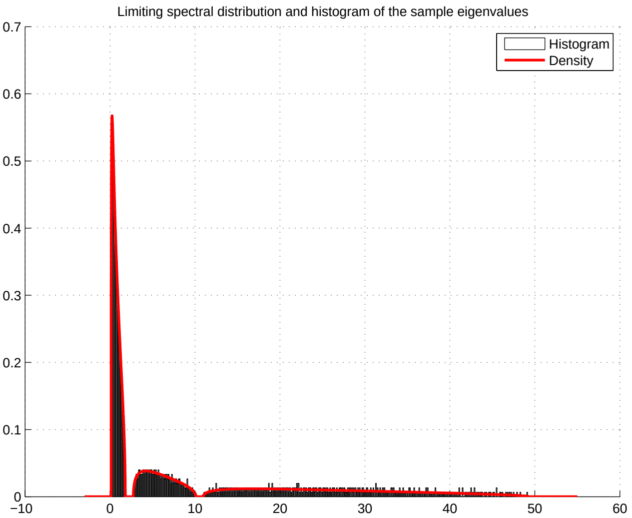


Figure 3: Densité de la loi μ_N pour $\sigma = 1$, $c_N = 0.5$, et histogramme des valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ dans le cadre du modèle (6). pour $N = 3000$, une réalisation de la matrice aléatoire $\mathbf{X}_N$. Les valeurs propres de $\mathbf{\Gamma}_N$ sont 1 (multiplicité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

Ces résultats généraux sur les modèles de matrices aléatoires à covariance ont été exploités par Mestre [51] 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 $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)]$, où $\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)$ 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 $\mathbf{R}_\mathrm{S}$ de rang plein, il est facile de voir que l'espace propre associé à la valeur propre σ^2 de $\mathbf{\Gamma}_N = \mathbf{A}\mathbf{R}_\mathrm{S}\mathbf{A}^* + \sigma^2\mathbf{I}_M$ est engendré par les vecteurs $\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)$. En notant $\mathbf{u}_{1,N}, \ldots, \mathbf{u}_{M,N}$ les vecteurs propres de $\mathbf{\Gamma}_N$ associés respectivement aux valeurs propres $\lambda_{1,N}, \ldots, \lambda_{M,N}$, ceci implique que les angles $\theta_1, \ldots, \theta_K$ sont des zéros de la fonction de localisation [1]

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

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

où $\sum_{k=1}^{M-K} \mathbf{u}_{k,N}\mathbf{u}_{k,N}^*$ est le projecteur sur le sous-espace propre de $\mathbf{\Gamma}_N$ 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 $N \to \infty$, 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 la matrice de covariance empirique $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ (dont les vecteurs propres sont notés $\hat{\mathbf{u}}_{1,N},\ldots,\hat{\mathbf{u}}_{M,N}$). Les angles $\theta_1,\ldots,\theta_K$ 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 [38] 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 [42]) 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 $\left\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^* - \left(\frac{\mathbf{A}\mathbf{S}_N\mathbf{S}_N^*\mathbf{A}^*}{N} + \sigma^2\mathbf{I}_M\right)\right\|$ ne tend plus vers 0. Récemment, Mestre [50] a proposé une approche séduisante pour étudier le cas où $M,N \to \infty$ et $M/N \to c \in]0,1[$, 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

$$\eta_N(\theta) = \frac{1}{2\pi \mathrm{i}} \oint_{\mathscr{C}} \mathbf{a}(\theta)^* \left(\mathbf{\Gamma}_N - w\mathbf{I}_M\right)^{-1} \mathbf{a}(\theta)\mathrm{d}w, \tag{14}$$

où $\mathscr{C}$ est un contour orienté dans le sens antitrigonométrique, entourant uniquement la valeur propre σ^2 de $\mathbf{\Gamma}_N$. De plus, sous certaines hypothèses ne dépendant uniquement que du comportement du support de μ_N, il existe un rectangle $\mathscr{R} = \{u + \mathrm{i}v : u \in [t^-,t^+], v \in [-\delta,\delta]\}$, avec $\delta, t^-, t^+ > 0$ indépendants de N, pour lesquels $w_N(\partial\mathscr{R})$ est un contour admissible pour $\mathscr{C}$, i.e. $w_N(\partial\mathscr{R})$ est un contour entourant uniquement la valeur σ^2 de $\mathbf{\Gamma}_N$ 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

$$\eta_N(\theta) = \frac{1}{2\pi \mathrm{i}} \oint_{\partial\mathscr{R}} \mathbf{a}(\theta)^* \mathbf{T}_N(z)\mathbf{a}(\theta)\left(1 - c_N - c_N z m_N(z)\right) w_N'(z)\mathrm{d}z. \tag{15}$$

où $\mathbf{T}_N(z) = \left(\mathbf{\Gamma}_N\left(1-c_N-c_N z m_N(z)\right) - z\mathbf{I}_M\right)^{-1}$ 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

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

peut être approchée par [2]

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

avec $\mathbf{Q}_N(z) = \left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^* - z\mathbf{I}_M\right)^{-1}$ 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 $\hat{\lambda}_{1,N},\ldots,\hat{\lambda}_{M,N}$ de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ et les zéros $\hat{\omega}_{1,N} < \ldots < \hat{\omega}_{M,N}$ 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 $\mathrm{supp}(\mu_N)$ impliquent, du fait des propriétés de séparation exacte des valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$, que p.s. pour N suffisamment grand,

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

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

$$\left|\int_{\partial\mathscr{R}} \left(\hat{g}_N(z) - g_N(z)\right)\mathrm{d}z\right| \xrightarrow[N\to\infty]{\text{p.s.}} 0. \tag{16}$$

En notant

$$\hat{\eta}_{\mathrm{rmt},N}(\theta) = \frac{1}{2\pi \mathrm{i}} \oint_{\partial\mathscr{R}_1} \hat{g}_N(z)\mathrm{d}z, \tag{17}$$

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

on a donc $\hat{\eta}_{\mathrm{rmt},N}(\theta) - \eta_N(\theta) \to 0$ p.s., et donc $\hat{\eta}_{\mathrm{rmt},N}(\theta)$ est un estimateur consistant de la fonction de localisation $\hat{\eta}_{\mathrm{rmt},N}(\theta)$ dans le régime asymptotique $M, N \to \infty$, $c_N = M/N \to c \in]0,1]$. Cette intégrale n'est par calculable numériquement en pratique, car la connaissance du rectangle $\mathscr{R}$ 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 $\mathscr{R}$ est connue. Un calcul classique de résidus donne alors la formule explicite suivante, facilement implémentable,

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

avec

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

Bien que les travaux de Mestre aient un portée relativement générale [3], 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,

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

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

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 $\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{W}_N$, où $\mathbf{B}_N$ est une matrice déterministe (dans le cas de la localisation de source $\mathbf{B}_N = \frac{\mathbf{AS}_N}{\sqrt{N}}$ 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.

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"

$$\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{W}_N, \tag{19}$$

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 $c_N = M/N \to c \in]0,1[$ (le cas $c = 1$ sera éludé pour alléger le manuscript). Ce modèle de matrice aléatoire a fait l'objet de peu de recherches, et les principales contributions sont dues à Dozier & Silverstein [16] [15] et Girko [21]. En particulier, ces travaux étudient la convergence de la distribution empirique $\hat{\mu}_N$ des valeurs propres de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$, de la même manière que pour (8). Ainsi on a $\hat{\mu}_N -$

[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 $\mathbf{\Gamma}_N$, pour le modèle (6), dans un cadre non nécessairement gaussien (voir notamment [49]).

$\mu_N \xrightarrow[N\to\infty]{\mathrm{e}} 0$ 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

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

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

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

Une analyse de supp(μ_N), le support de μ_N, est donnée dans Dozier & Silverstein [15], mais la procédure de détermination de supp(μ_N) proposée n'est pas la meilleure. Inspirés par les travaux de Mestre, nous montrons que

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

où $x_{1,N}^- < x_{1,N}^+ < \ldots < x_{Q,N}^- < x_{Q,N}^+$ 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 $f_N(w) = \frac{1}{M}\mathrm{Tr}\left(\mathbf{B}_N\mathbf{B}_N^* - w\mathbf{I}_M\right)^{-1}$. 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

$$w_N(z) = z\left(1+\sigma^2 c_N m_N(z)\right)^2 - \sigma^2(1-c_N)\left(1+\sigma^2 c_N m_N(z)\right). \tag{23}$$

Cette caractérisation du support est une étape indispensable à l'étude de résultats sur la localisation presque sûre des valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$, 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 $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ n'appartient à supp(μ_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 $[a-\epsilon, b+\epsilon]\cap\mathrm{supp}(\mu_N) = \emptyset$ à partir d'un certain rang N, alors

$$\mathrm{card}\left\{k : \hat\lambda_{k,N} \in [a,b]\right\} = 0, \tag{24}$$

p.s. pour N grand. Ce résultat est l'analogue d'un résultat démontré par Bai & Silverstein [4] dans le cadre des modèles à covariance, non nécessairement gaussiens. Pour montrer (24), nous utilisons une méthode plus simple que [4], basée sur l'exploitation du caractère gaussien du modèle, et qui a été proposée par Haagerup & Thorbjornsen [22] 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 $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ sorte du support: si $\mathscr{K}$ est un compact contenant $\mathscr{S}_N$ à partir d'un certain rang N et $\mathscr{V}$ un voisinage compact de $\mathscr{K}$, avec $\mathscr{K},\mathscr{V}$ indépendants de N, alors

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

La méthode de [22] peut également être utilisée pour montrer une propriété de séparation des valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ 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 $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. 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. $\lambda_{M-K+1,N},\ldots,\lambda_{M,N}$, convergent respectivement vers les limites $0 < \gamma_1 < \ldots < \gamma_K$, nous montrons que pour $k = 1,\ldots,K$,

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

Ainsi (26) montre que l'on peut exploiter les valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ pour la détection de sources. En effet, dans le cadre du modèle à traitement d'antennes, où $\mathbf{B}_N = \frac{\mathbf{A}\mathbf{S}_N}{\sqrt{N}}$, les K plus grandes valeurs propres de $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ 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 [9], 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 $\mathbf{u}_{1,N}, \ldots, \mathbf{u}_{M,N}$ désignent désormais les vecteurs propres de $\mathbf{B}_N \mathbf{B}_N^*$ associés aux valeurs propres $\lambda_{1,N} \leq \ldots \leq \lambda_{M,N}$, peut s'écrire comme

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

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 $\mathbf{\Sigma}_N$, 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 $N \to \infty$, où $\mathbf{Q}_N(z) = \left(\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - z \mathbf{I}_M\right)^{-1}$ est la résolvante de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$. Nous obtenons ainsi

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

où $\hat{w}_N'(z)$ est la dérivée de $\hat{w}_N(z) = z\left(1 + \sigma^2 c_N \hat{m}_N(z)\right)^2 - \sigma^2 (1 - c_N)\left(1 + \sigma^2 c_N \hat{m}_N(z)\right)$. Les résultats sur la localisation des valeurs propres de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$, 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 $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$. Ceci nécessite également un résultat de séparation pour les zeros 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 zeros de $z \mapsto 1 + \sigma^2 c_N \hat{m}_N(z)$, le dénominateur dans (27).

Organisation du manuscript

Le chapitre 1 est consacré au résumé des outils fondamentaux qui seront utilisés tout au long de ce travail. En particulier, les principales propriétés concernant la transformée de Stieltjes, l'outil central dans l'étude des valeurs propres, sont données.
Le chapitre 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".
Le chapitre 3 est consacré à l'analyse de la localisation des valeurs propres de $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$. 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.
Le dernier chapitre 4 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.

Publications associées au manuscript

Articles de journaux

1. P.Vallet, P.Loubaton, et X.Mestre, Improved subspace estimation for multivariate observations of high dimension: the deterministic signal case, *A paraître dans IEEE Transactions on Information Theory*, 2011, arXiv: 1002.3234.

2. P.Loubaton et P.Vallet, Almost sure localization of the eigenvalues in a Gaussian information plus noise model. Application to the spiked models., *Electronic Journal of Probability*, Vol. 16, p. 1934-1959, 2011, arXiv: 1009.5807.

3. W.Hachem, P.Loubaton, J.Najim, et P.Vallet, On bilinear forms based on the resolvent of large random matrices, *A paraître dans Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 2011, arXiv:1004.3848.

4. W.Hachem, P.Loubaton, X.Mestre, J.Najim et P.Vallet, Large information plus noise random matrix models and consistent subspace estimation in large sensor networks, *A paraître dans Random Matrix: Theory and Applications*, 2011, arXiv:1106.5119.

Articles de conférences

1. P.Vallet, P.Loubaton, X.Mestre, Improved subspace DoA estimation methods with large arrays: the deterministic signals case, *IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei (Taiwan)*, 2137–2140, 2009.

2. P.Vallet, W.Hachem, P.Loubaton, X.Mestre, J.Najim, On the consistency of the G-MUSIC DoA Estimator, *IEEE International Workshop on Statistical Signal Processing (SSP), Nice (France)*, 2011.

3. P.Vallet, W.Hachem, P.Loubaton, X.Mestre, J.Najim, An improved MUSIC algorithm based on low-rank perturbation of large random matrices, *IEEE International Workshop on Statistical Signal Processing (SSP), Nice (France)*, 2011.

Chapter 1

Notations and basic tools

In this chapter, we introduce in the first section the notations which will be used throughout the manuscript. The remaining sections are devoted to introduce fundamental tools and basic results.

1.1 Notations

We give here a list of the main notations.

- The set of non-negative integers is $\mathbb{N}$ and the fields of real and complex numbers are denoted respectively $\mathbb{R}, \mathbb{C}$. $\mathbb{R}^+$ and $\mathbb{R}^-$ represents respectively the set non-negative numbers and non-positive numbers, and we denote $\mathbb{R}^* \equiv \mathbb{R}\backslash\{0\}$, $\mathbb{R}^+_* \equiv \mathbb{R}^+\backslash\{0\}$, $\mathbb{R}^-_* \equiv \mathbb{R}^-\backslash\{0\}$. We also define $\mathbb{C}^+ \equiv \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ and $\mathbb{C}^- \equiv \{z \in \mathbb{C} : \mathrm{Im}(z) < 0\}$, and write $\mathbb{C}^* \equiv \mathbb{C}\backslash\{0\}$. The set of $m \times n$ matrices with entries in the field $\mathbb{K}$ is written $\mathbb{K}^{m\times n}$ or $\mathscr{M}_{m\times n}(\mathbb{K})$. The set of $n \times n$ square matrices is simply denoted by $\mathscr{M}_n(\mathbb{K})$ and $\mathscr{M}_n^{\mathrm{s.a}}(\mathbb{K})$ is the set of $n \times n$ Hermitian matrices. If $\mathscr{E} \subset \mathbb{R}$, $\mathscr{E}^c$, $\mathrm{Int}(\mathscr{E})$ and $\partial\mathscr{E}$ represent respectively the complementary, the interior and the boundary of $\mathscr{E}$, If $\mathscr{E}$ has a finite number of elements, its cardinal is denoted by $|\mathscr{E}|$.

- For a complex number z, we denote by $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ its real and imaginary part, as well as z^* or $\overline{z}$ its conjuguate. The imaginary unit is denoted i. The vectors and matrices will be respectively written with bold lower-case and bold upper-case letters, e.g $\mathbf{a} \in \mathbb{C}^n$, $\mathbf{A} \in \mathscr{M}_n(\mathbb{C})$. The transpose, conjuguate, and conjuguate transpose of $\mathbf{A} \in \mathscr{M}_n(\mathbb{C})$ are respectively denoted by $\mathbf{A}^T, \overline{\mathbf{A}}$ and $\mathbf{A}^*$, its trace and spectral norm by $\mathrm{Tr}\,(\mathbf{A})$ and $\|\mathbf{A}\|$. If $\mathbf{A}, \mathbf{B} \in \mathscr{M}_n(\mathbb{C})$, $\mathbf{A} \geq \mathbf{B}$ stands for $\mathbf{A} - \mathbf{B}$ non-negative definite. The canonical basis vectors of $\mathbb{R}^n$ are denoted $\mathbf{e}_1, \ldots, \mathbf{e}_n$ with $\mathbf{e}_j$ having all entries equal to 0 except the j-th which is 1.

- For a set $\mathscr{E} \subset \mathbb{C}$, $\mathscr{C}(\mathbb{R}^n, \mathscr{E})$, $\mathscr{C}^1(\mathbb{R}^n, \mathscr{E})$ and $\mathscr{C}^\infty(\mathbb{R}^n, \mathscr{E})$ correspond to the set of functions on $\mathbb{R}^n$ taking values in $\mathscr{E}$, which are respectively continuous, continuously differentiable and infinitely differentiable. If we consider bounded functions, we add the index "b" to the previous sets, and if the functions are compactly supported, we add the index "c". The support of a function φ is denoted $\mathrm{supp}(\phi)$.

- A function $\tilde{\gamma} : \mathbb{R}^2 \mapsto \mathbb{C}$ can be written as $\tilde{\gamma}\left(x, y\right) = \gamma(z, \overline{z})$ with $x = \mathrm{Re}(z)$ and $y = \mathrm{Im}(z)$. If $\tilde{\gamma} \in \mathscr{C}^1(\mathbb{R}^2, \mathbb{C})$, we define the classical differential operators

$$\frac{\partial\gamma(z,\overline{z})}{\partial z} = \frac{1}{2}\left(\frac{\partial\tilde{\gamma}(x,y)}{\partial x} - \mathrm{i}\frac{\partial\tilde{\gamma}(x,y)}{\partial y}\right) \quad \text{and} \quad \frac{\partial\gamma(z,\overline{z})}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial\tilde{\gamma}(x,y)}{\partial x} + \mathrm{i}\frac{\partial\tilde{\gamma}(x,y)}{\partial y}\right).$$

  In this context, we directly write $\gamma \in \mathscr{C}^1(\mathbb{R}^2, \mathbb{C})$ instead of $\tilde{\gamma} \in \mathscr{C}^1(\mathbb{R}^2, \mathbb{C})$.

- A complex valued random variable $Z = X + \mathrm{i}Y$ follows the distribution $\mathscr{CN}\left(\alpha + i\beta, \sigma^2\right)$ if X and Y are independent real Gaussian random variables $\mathscr{N}\left(\alpha, \frac{\sigma^2}{2}\right)$ and $\mathscr{N}\left(\beta, \frac{\sigma^2}{2}\right)$ respectively. The expectation of Z is denoted $\mathbb{E}[Z] = \alpha + \mathrm{i}\beta$ and its variance as $\mathrm{Var}[Z] = \mathbb{E}|Z - \mathbb{E}[Z]|^2 = \sigma^2$. If (X_n), (Y_n) are two sequences of complex random variables, $X_n - Y_n \xrightarrow[n\to\infty]{\text{a.s}} 0$ stands for the almost-sure convergence to 0 of $X_n - Y_n$. If (μ_n) and (ν_n) are two sequences of positive measures on $\mathbb{R}$, we denote by $\mu_n - \nu_n \xrightarrow[n\to\infty]{w} 0$ the weak convergence of $\mu_n - \nu_n$ to 0, i.e. the property $\int_{\mathbb{R}} \varphi \mathrm{d}\mu_n - \int_{\mathbb{R}} \varphi \mathrm{d}\nu_n \to_N 0$ for all $\varphi \in \mathscr{C}_b(\mathbb{R}, \mathbb{R})$.

1.2 The Stieltjes transform and its properties

The results of this section can be found in Akhiezer [1] and in Krein & Nudelman [29].

Let μ be a finite measure with support $\operatorname{supp}(\mu) \subset \mathbb{R}$. Its Stieltjes transform Ψ_μ is the function

$$\Psi_\mu(z) \triangleq \int_{\mathbb{R}} \frac{\mathrm{d}\mu(\lambda)}{\lambda - z} \quad \forall z \in \mathbb{C} \backslash \operatorname{supp}(\mu),$$

which satisfies the following straightforward property.

Property 1.2.1. *Ψ_μ is holomorphic on $\mathbb{C}\backslash\operatorname{supp}(\mu)$ and such that $\Psi_\mu(z)^* = \Psi_\mu(z^*)$. Moreover,*

- *$z \in \mathbb{C}^+$ implies $\Psi_\mu(z) \in \mathbb{C}^+$,*
- *$\operatorname{supp}(\mu) \subset \mathbb{R}^+$ and $z \in \mathbb{C}^+$ imply $z\Psi_\mu(z) \in \mathbb{C}^+$.*

Finally, Ψ_μ satisfies the bound $\left|\Psi_\mu(z)\right| \leq \frac{\mu(\mathbb{R})}{\operatorname{dist}(z,\operatorname{supp}(\mu))}$ and $\left|\Psi_\mu(z)\right| \leq \frac{\mu(\mathbb{R})}{|\operatorname{Im}(z)|}$.

The following property shows how to recover the measure μ from the Stieltjes transform Ψ_μ.

Property 1.2.2. *The mass $\mu(\mathbb{R})$ can be recovered through the formula*

$$\mu(\mathbb{R}) = \lim_{y \to \infty} -iy\Psi_\mu(iy). \tag{1.1}$$

Moreover, for all $\varphi \in \mathscr{C}_c(\mathbb{R}, \mathbb{R})$,

$$\int_{\mathbb{R}} \varphi(\lambda)\mathrm{d}\mu(\lambda) = \frac{1}{\pi} \lim_{y \downarrow 0} \operatorname{Im}\left(\int_{\mathbb{R}} \varphi(x)\Psi_\mu(x+iy)\mathrm{d}x\right). \tag{1.2}$$

Particularly, if a, b are continuity points of μ,

$$\mu([a,b]) = \lim_{y \downarrow 0} \int_a^b \operatorname{Im}\left(\Psi_\mu(x+iy)\right)\mathrm{d}x. \tag{1.3}$$

We now give sufficient conditions for a function to be a Stieltjes transform. The second set of conditions can be found in Tillmann [44], and extends the Stieltjes transform for signed measures.

Property 1.2.3.
- *If*
  1. *$\Psi(z)$ is holomorphic on $\mathbb{C}^+$*
  2. *$z \in \mathbb{C}^+$ implies $\Psi(z) \in \mathbb{C}^+$*
  3. *$\limsup_{y\to\infty} \left|iy\Psi(iy)\right| < \infty$*

  then $\Psi(z)$ is the Stieltjes transform of a positive finite measure on $\mathbb{R}$.
- *If*
  1. *$\Psi(z)$ is holomorphic on $\mathbb{C}\backslash\mathscr{K}$ with $\mathscr{K}$ a compact set,*
  2. *$\Psi(z) \to 0$ as $|z| \to \infty$,*
  3. *there exists a constant $C > 0$ and $n \in \mathbb{N}$ such that for all $z \in \mathbb{C}\backslash\mathbb{R}$, $|\Psi(z)| \leq \max\left\{\operatorname{dist}(z,\mathscr{K})^{-n}, 1\right\}$,*

  then $\Psi(z)$ is the Stieltjes transform of a finite signed measure on $\mathbb{R}$ whose support is the set of singularities of Ψ contained in $\mathscr{K}$.

We now give the fundamental result, which states the equivalence between pointwise convergence of Stieltjes transform and weak convergence of probability measures.

Theorem 1.2.1. *Let (μ_n), μ be probability measures on $\mathbb{R}$ and (Ψ_{μ_n}), Ψ_μ the associated Stieltjes transform. Then the following two statements are equivalent.*

1. *$\Psi_{\mu_n}(z) \xrightarrow[n\to\infty]{} \Psi_\mu(z)$ for all $z \in \mathbb{C}^+$,*
2. *$\mu_n \xrightarrow[n\to\infty]{w} \mu$.*

Remark 1.2.1. *If (μ_n) and (ν_n) are sequences of probability measures, it also holds that $\mu_n - \nu_n \xrightarrow[n\to\infty]{w} 0$ iff $\Psi_{\mu_n}(z) - \Psi_{\nu_n}(z) \to_n 0$ for all $z \in \mathbb{C}^+$. This will be useful in the next chapters because we will often deal with sequences of Stieltjes transforms which do not necessarily converge, but are close to other Stieltjes transform sequences.*

1.3 Standard results of matrix analysis

1.3.1 Various inequalities

The following two properties give various inequalities for trace and quadratic forms of matrices. They are straightforward applications of singular value decomposition, Cauchy-Schwarz and Jensen's inequalities.

Property 1.3.1. *Let* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n(\mathbb{C})$. *Then it holds*

- $\mathbf{A} \geq 0$ *implies* $|\mathrm{Tr}\,(\mathbf{AB})| \leq \|\mathbf{B}\| \mathrm{Tr}\,(\mathbf{A})$,
- $|\mathrm{Tr}\,(\mathbf{AB})| \leq \sqrt{\mathrm{Tr}\,(\mathbf{AA}^*)}\sqrt{\mathrm{Tr}\,(\mathbf{BB}^*)}$.

Property 1.3.2. *Let* $\mathbf{A} \in \mathcal{M}_n(\mathbb{C})$. *Then for all* $\mathbf{x} \in \mathbb{C}^n$,

- $|\mathbf{x}^* \mathbf{A} \mathbf{x}| \leq \|\mathbf{A}\| \|\mathbf{x}\|^2$,
- $|\mathbf{x}^* \mathbf{A} \mathbf{x}| \leq \sqrt{\mathbf{x}^* (\mathbf{AA}^*)^{1/2} \mathbf{x}} \sqrt{\mathbf{x}^* (\mathbf{A}^* \mathbf{A})^{1/2} \mathbf{x}}$.

Property 1.3.3. *If* $\mathbf{A} \in \mathcal{M}_n(\mathbb{C})$ *is Hermitian and* $\mathbf{x} \in \mathbb{C}^n$ *with* $\|\mathbf{x}\| = 1$, *then*

$$\left(\mathbf{x}^* \mathbf{A} \mathbf{x}\right)^2 \leq \mathbf{x}^* \mathbf{A}^2 \mathbf{x}.$$

1.3.2 Resolvent formulas

In this subsection, we give some properties of the resolvent of Gram matrices, a fundamental tool in the analysis of spectral distribution of random matrices. Let $\mathbf{\Sigma} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ a $m \times n$ complex matrix. The resolvent of $\mathbf{\Sigma\Sigma}^*$ is defined as

$$\mathbf{Q}(z) = \left(\mathbf{\Sigma\Sigma}^* - z\mathbf{I}_m\right)^{-1} \quad \text{for all} \quad z \in \mathbb{C} \backslash \mathbb{R}^+.$$

Remark 1.3.1. *If* $\hat{F}(\lambda) = \frac{1}{m}\mathrm{card}\{$*eigenvalues of* $\mathbf{\Sigma\Sigma}^* \leq \lambda\}$, *commonly referred as the spectral distribution of matrix* $\mathbf{\Sigma\Sigma}^*$, *then it is easily seen that the Stieltjes transform of* $\hat{F}(\lambda)$ *is* $\frac{1}{m}\mathrm{Tr}\,\mathbf{Q}(z)$. *Thus Stieltjes transforms of spectral distributions are closely related with resolvents, and this link provides an approach to the study of asymptotic spectral distribution of random matrices, as we will see in the next chapter.*

Note that $\mathbf{Q}(z)$ satisfies the bounds $\|\mathbf{Q}(z)\| \leq \left|\mathrm{Im}(z)^{-1}\right|$ and $\|\mathbf{Q}(z)\| \leq \mathrm{dist}\left(z, \mathrm{spec}(\mathbf{\Sigma\Sigma}^*)\right)^{-1}$, where $\mathrm{spec}(\mathbf{\Sigma\Sigma}^*)$ is the spectrum of $\mathbf{\Sigma\Sigma}^*$.

We now set $\mathbf{\Sigma} = \mathbf{B} + \mathbf{W}$ and denote as above $\mathbf{Q}(z) = (\mathbf{\Sigma\Sigma}^* - z\mathbf{I}_n)^{-1}$. It will be useful for the next chapters to have differentiation formulas with respect to the elements $W_{k,l}$ of $\mathbf{W}$. Using the differential of the inverse $\partial \mathbf{A}^{-1} = -\mathbf{A}^{-1}(\partial \mathbf{A})\mathbf{A}^{-1}$, classical differential calculus leads to the following property.

Property 1.3.4. *The resolvent* $\mathbf{Q}(z)$ *is a continuously differentiable function of* $\mathbf{W}$ *and we have*

$$\frac{\partial [\mathbf{Q}(z)]_{i,j}}{\partial W_{k,l}} = -[\mathbf{Q}(z)]_{i,k} [\mathbf{\Sigma}^* \mathbf{Q}(z)]_{l,j} \quad \textit{and} \quad \frac{\partial [\mathbf{Q}(z)]_{i,j}}{\partial W^*_{k,l}} = -[\mathbf{Q}(z)]_{k,j} [\mathbf{Q}(z)\mathbf{\Sigma}]_{i,l}.$$

1.3.3 Eigenvalues perturbation

The following inequality is particularly useful to perform additive perturbation of eigenvalues.

Theorem 1.3.1 (Ky Fan [19])**.** *Let* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n(\mathbb{C})$ *and denote* $\sigma_k(\mathbf{A})$, $\sigma_k(\mathbf{B})$ *and* $\sigma_k(\mathbf{A}+\mathbf{B})$ *the respective singular values (in increasing order) of* $\mathbf{A}, \mathbf{B}$ *and* $\mathbf{A}+\mathbf{B}$. *Then,*

$$\sigma_{n-i-j}(\mathbf{A}+\mathbf{B}) \leq \sigma_{n-i}(\mathbf{A}) + \sigma_{n-j}(\mathbf{B}).$$

for proper indexes i, j.

Next, we state the well-known result on the continuity of the eigenvalues, as functions of the coefficients of the matrix.

Theorem 1.3.2. *Let* $f(z) \in \mathbb{C}^n[z]$ *given by* $f(z) = z^n + \alpha_{n-1}z^{n-1} + \ldots + \alpha_0$. *Denote* $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{n-1}]$. *Then its n roots,* $z_1(\boldsymbol{\alpha}), \ldots, z_n(\boldsymbol{\alpha})$, *viewed as functions of the coefficients, are continuous.*

Next, we make use of this result to express the derivative of regular functions of the eigenvalues with respect to the entries of the matrix. We will consider the special case of additive model. For a real function φ defined on $\mathbb{R}$ and an Hermitian matrix $\mathbf{A}$, we define

$$\varphi(\mathbf{A}) = \mathbf{U}\mathrm{Diag}\left(\phi(\lambda_1),\dots,\phi(\lambda_m)\right)\mathbf{U}^*,$$

with $\lambda_1,\dots,\lambda_m$ and $\mathbf{U}$ the eigenvalues and the eigenvectors matrix of $\mathbf{A}$ respectively.

Lemma 1.3.1 (Haagerup & Thorbjornsen [22]). *Let $\mathscr{I} \subset \mathbb{R}$ an open interval and $t \mapsto \mathbf{A}(t) \in \mathscr{M}_n(\mathbb{C})$ a $\mathscr{C}^1$ function defined on $\mathscr{I}$. If $\varphi \in \mathscr{C}^1(\mathbb{R},\mathbb{R})$, then $t \mapsto \mathrm{Tr}\,\varphi(\mathbf{A}(t))$ is $\mathscr{C}^1$ on $\mathscr{I}$ and*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\mathrm{Tr}\,\varphi(\mathbf{A}(t))\right\} = \mathrm{Tr}\left(\varphi'(\mathbf{A}(t))\frac{\mathrm{d}\mathbf{A}(t)}{\mathrm{d}t}\right).$$

In particular, if $\mathbf{B}$ a fixed matrix, then the function

$$\mathbf{W} \to \mathrm{Tr}\,\varphi((\mathbf{B}+\mathbf{W})(\mathbf{B}+\mathbf{W})^*)$$

is continuously differentiable, and

$$\frac{\partial \mathrm{Tr}\,\varphi((\mathbf{B}+\mathbf{W})(\mathbf{B}+\mathbf{W})^*)}{\partial W_{i,j}} = \mathbf{e}_j^*(\mathbf{B}+\mathbf{W})^*\varphi'\left((\mathbf{B}+\mathbf{W})(\mathbf{B}+\mathbf{W})^*\right)\mathbf{e}_i.$$

As a consequence of lemma 1.3.1, we can express the derivative of the eigenvalues, up to an additional condition.

Lemma 1.3.2. *Let $\mathbf{B}$ a fixed matrix and let $\mathbf{W} \mapsto \lambda_k(\mathbf{W})$ the k-th eigenvalue of $(\mathbf{B}+\mathbf{W})(\mathbf{B}+\mathbf{W})^*$, viewed as a function of the matrix $\mathbf{W}$. Let $\mathbf{W}^{(0)}$ such that $\lambda_k(\mathbf{W}^{(0)})$ has multiplicity one. Then the function λ_k is continuously differentiable at $\mathbf{W}^{(0)}$ and*

$$\left.\frac{\partial \lambda_k}{\partial W_{i,j}}\right|_{\mathbf{W}^{(0)}} = \mathbf{e}_j^*(\mathbf{B}+\mathbf{W}^{(0)})^*\mathbf{\Pi}_k^{(0)}\mathbf{e}_i,$$

with $\mathbf{\Pi}_k^{(0)}$ the projector onto the eigenspace associated with the k-th eigenvalue of $(\mathbf{B}+\mathbf{W}^{(0)})(\mathbf{B}+\mathbf{W}^{(0)})^*$.

Proof. From theorem 1.3.2, one can find a bounded neighborhood $\mathscr{W}_0$ of $\mathbf{W}^{(0)}$ and a bounded open interval (a,b) such that $\lambda_k(\mathbf{W}) \in (a,b)$ and $\lambda_l(\mathbf{W}) \notin [a-\epsilon, b+\epsilon]$ for $l \neq k$ and some $\epsilon > 0$ small enough. The proof follows from theorem 1.3.1 by taking a function $\varphi \in \mathscr{C}_c^1(\mathbb{R},\mathbb{R})$ such that

$$\varphi(\lambda) = \begin{cases} \lambda & \text{for all } \lambda \in [a,b] \\ 0 & \text{for all } \lambda \notin [a-\epsilon, b+\epsilon]. \end{cases}$$

□

1.4 Standard probability results

1.4.1 Tools for Gaussian variables

This section is devoted to the introduction of two fundamental tools, a correlation identity and a Poincaré inequality for Gaussian random variables, which will be of constant use for the computations in the next chapters. Let $\tilde{f}: \mathbb{R}^{2n} \mapsto \mathbb{C}$ and its associated function $f(\mathbf{z},\overline{\mathbf{z}}) = \tilde{f}(\mathbf{x},\mathbf{y})$, with $\mathbf{z} = \mathbf{x} + \mathbf{i}\mathbf{y}$. If $f \in \mathscr{C}^1(\mathbb{R}^{2n},\mathbb{C})$, we define the following differentiation operators

$$\nabla_{\mathbf{z}} f = \left[\frac{\partial f(\mathbf{z},\overline{\mathbf{z}})}{\partial z_1},\dots,\frac{\partial f(\mathbf{z},\overline{\mathbf{z}})}{\partial z_n}\right]^T \quad \text{and} \quad \nabla_{\overline{\mathbf{z}}} f = \left[\frac{\partial f(\mathbf{z},\overline{\mathbf{z}})}{\partial \overline{z_1}},\dots,\frac{\partial f(\mathbf{z},\overline{\mathbf{z}})}{\partial \overline{z_n}}\right]^T.$$

The following result is a well-known correlation identity.

Theorem 1.4.1. *Let $f: \mathbb{C}^n \mapsto \mathbb{C}$ and assume $f \in \mathscr{C}^1(\mathbb{R}^{2n},\mathbb{C})$ such that f and its partial derivative are polynomially bounded. Let $\mathbf{x},\mathbf{y}$ be n-dimensional i.i.d. real Gaussian vector with zero mean and covariance $\frac{\mathbf{\Gamma}}{2}$, and $\mathbf{z} = \mathbf{x} + i\mathbf{y}$. Then*

$$\mathbb{E}\left[z_k f(\mathbf{z})\right] = \mathbf{e}_k^*\mathbf{\Gamma}\mathbb{E}\left[\nabla_{\overline{\mathbf{z}}} f\right] \quad \textit{and} \quad \mathbb{E}\left[\overline{z_k} f(\mathbf{z})\right] = \mathbf{e}_k^*\mathbf{\Gamma}\mathbb{E}\left[\nabla_{\mathbf{z}} f\right].$$

Proof. The result is well-known and has been used e.g. by Novikov [36]. To give some insights, we prove the result for a real Gaussian variable (the complex multivariate case can be obtained with the same technic). Let $X \sim \mathcal{N}(0,\sigma^2)$. Using an integration by part, we get

$$\mathbb{E}\left[Xf(X)\mathbb{1}_{|X|\leq a}\right] = \frac{\sigma}{\sqrt{2\pi}}\mathrm{e}^{-a^2/2\sigma^2}\left(f(-a)-f(a)\right)+\sigma^2\mathbb{E}\left[f'(X)\mathbb{1}_{|X|\leq a}\right]$$

and using the polynomial bound on f, f', dominated convergence theorem yields the desired result by taking the limit $a \to \infty$. □

The next result is known in the literature as the Poincaré inequality.

Theorem 1.4.2. *Let $f : \mathbb{C}^n \mapsto \mathbb{C}$ and assume $f \in \mathscr{C}^1(\mathbb{R}^{2n},\mathbb{C})$ such that f and its partial derivative are polynomially bounded. Let $\mathbf{x}, \mathbf{y}$ be n-dimensional i.i.d. real Gaussian vector with zero mean and covariance $\frac{\mathbf{\Gamma}}{2}$, and $\mathbf{z} = \mathbf{x} + i\mathbf{y}$. Then*

$$\mathrm{Var}\left[f(\mathbf{z})\right] = \mathrm{Var}\left[f(\mathbf{z})\right] \leq \mathbb{E}\left[(\nabla_{\mathbf{z}} f)^* \,\mathbf{\Gamma}\left(\nabla_{\mathbf{z}} f\right)\right] + \mathbb{E}\left[(\nabla_{\overline{\mathbf{z}}} f)^* \,\mathbf{\Gamma}\left(\nabla_{\overline{\mathbf{z}}} f\right)\right].$$

Proof. A complete proof can be found in Chen [12]. We give here a proof from Pastur [37] in the simpler real scalar case and when f is real and twice continuously differentiable, bounded together with its first and second derivatives. Let $X \sim \mathcal{N}(0,\sigma^2)$ and X_1, X_2 i.i.d copies of X. Then, we can write

$$\mathrm{Var}\left[f(X)\right] = \mathbb{E}\left[f(X)^2\right] - \left(\mathbb{E}\left[f(X)\right]\right)^2 = \mathbb{E}\left[g(\mathbf{y}) - g(\mathbf{z})\right],$$

with $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $g(a_1, a_2) = f(a_1)f(a_2)$ and $\mathbf{y} = [X, X]^T$, $\mathbf{z} = [X_1, X_2]^T$. Clearly,

$$\mathbb{E}\left[g(\mathbf{y}) - g(\mathbf{z})\right] = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left\{\mathbb{E}\left[g(\mathbf{y}\sqrt{t}+\mathbf{z}\sqrt{1-t})\right]\right\}\mathrm{d}t.$$

From the previous definitions,

$$\mathbb{E}\left[g(\mathbf{y}\sqrt{t}+\mathbf{z}\sqrt{1-t})\right] = \mathbb{E}\left[f(X\sqrt{t}+X_1\sqrt{1-t})f(X\sqrt{t}+X_2\sqrt{1-t})\right],$$

and it is easy to obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}&\left\{\mathbb{E}\left[g(\mathbf{y}\sqrt{t}+\mathbf{z}\sqrt{1-t})\right]\right\} = \\ &\frac{1}{2}\mathbb{E}\left[\left(\frac{X}{\sqrt{t}}-\frac{X_1}{\sqrt{1-t}}\right)f'(X\sqrt{t}+X_1\sqrt{1-t})f(X\sqrt{t}+X_2\sqrt{1-t})\right] \\ &+\frac{1}{2}\mathbb{E}\left[\left(\frac{X}{\sqrt{t}}-\frac{X_2}{\sqrt{1-t}}\right)f'(X\sqrt{t}+X_2\sqrt{1-t})f(X\sqrt{t}+X_1\sqrt{1-t})\right].\end{aligned}$$

Applying theorem 1.4.1 to the righthandside of the previous expression, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\mathbb{E}\left[g(\mathbf{y}\sqrt{t}+\mathbf{z}\sqrt{1-t})\right]\right\} = \sigma^2\mathbb{E}\left[f'(X\sqrt{t}+X_1\sqrt{1-t})f'(X\sqrt{t}+X_2\sqrt{1-t})\right].$$

Using Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left\{\mathbb{E}\left[g(\mathbf{y}\sqrt{t}+\mathbf{z}\sqrt{1-t})\right]\right\} &\leq \sigma^2\sqrt{\mathbb{E}\left[f'(X\sqrt{t}+X_1\sqrt{1-t})^2\right]}\sqrt{\mathbb{E}\left[f'(X\sqrt{t}+X_2\sqrt{1-t})^2\right]} \\ &= \sigma^2\mathbb{E}\left[f'(X\sqrt{t}+X_1\sqrt{1-t})^2\right],\end{aligned} \tag{1.4}$$

since X_1 and X_2 have the same law. Moreover, it is obvious that $X\sqrt{t}+X_1\sqrt{1-t} \sim \mathcal{N}(0,\sigma^2)$, and consequently the righthandside of (1.4) does not depend on t, which concludes the proof. □

1.4.2 Results for Wishart matrices

Let $M, N \in \mathbb{N}^*$ with $M \leq N$ and $\mathbf{W} \in \mathscr{M}_{M,N}(\mathbb{C})$ having i.i.d $\mathscr{CN}\left(0, \frac{1}{N}\right)$ entries. It is well-known (see e.g. James [27]) that the distribution of the M eigenvalues of $\mathbf{W}\mathbf{W}^*$ is absolutely continuous, which implies that they are all different with probability one.

The following concentration result is also well-known.

Theorem 1.4.3 (Davidson & Szarek [14, Th. II.13])**.** *Let* $c = \frac{M}{N}$ *and* $X = \|\mathbf{W}\|$*. Then it holds that* $\mathbb{E}[X] \leq 1 + \sqrt{c}$ *and for all* $t > 0$*,*

$$\mathbb{P}\left(X > 1 + \sqrt{c} + t\right) \leq \mathrm{e}^{-N\frac{t^2}{2}}.$$

Using theorem 1.4.3, we can easily obtain a bound independent of M, N for X^p $(p \geq 1)$ by writing,

$$\mathbb{E}[X^p] = \int_0^{+\infty} \mathbb{P}(X \geq t)\, p t^{p-1} \mathrm{d}t \leq p\left(1 + \sqrt{c}\right)^p + \int_0^{+\infty} \mathbb{P}\left(X \geq t + 1 + \sqrt{c}\right) p(t + 1 + \sqrt{c})^{p-1} \mathrm{d}t,$$

and thus $\mathbb{E}[X^p] \leq K < \infty$, with K a constant independent of M, N, for all $p \in \mathbb{N}$.

Finally, from [27], for $\mathbf{B} \in \mathscr{M}_{M,N}(\mathbb{C})$ deterministic, the eigenvalues of the matrix $(\mathbf{B} + \mathbf{W})(\mathbf{B} + \mathbf{W})^*$ also have an absolutely continuous distribution, and thus have also multiplicity one almost surely.

Chapter 2

Asymptotic spectral distribution of complex Gaussian information plus noise models

The present chapter is devoted to the introduction of the model of random matrices which will be used throughout the manuscript, namely the information plus noise model. A $M \times N$ information plus noise matrix $\mathbf{\Sigma}$ of this model writes

$$\mathbf{\Sigma} = \mathbf{B} + \mathbf{W}, \tag{2.1}$$

where matrix $\mathbf{B}$ is a complex deterministic matrix (the "information" part) and $\mathbf{W}$ a complex random matrix (the "noise" part) with i.i.d zero mean and variance σ^2/N entries. We will review some of the basic results concerning the information plus noise model. Roughly speaking, the most important one, states that the empirical eigenvalues distribution of the Gram matrix $\mathbf{\Sigma}\mathbf{\Sigma}^*$ (referred as spectral distribution of $\mathbf{\Sigma}\mathbf{\Sigma}^*$), i.e

$$\hat{F}(\lambda) = \frac{1}{M}\text{card}\{\text{eigenvalues of } \mathbf{\Sigma}\mathbf{\Sigma}^* \leq \lambda\}, \tag{2.2}$$

which is a random function, is almost surely close to a deterministic probability distribution $F(\lambda)$ (referred to as "deterministic equivalent spectral distribution"), when $M, N \to \infty$ such that the ratio $M/N \to c > 0$. We will considered the case where $c < 1$, to enlight the exposition. This asymptotic regime, which will be considered in all the manuscript, suggests to adapt the notation, and we will consider M as a function of N, and all matrices and related quantities will be denoted with index N. In this context, the result writes $\hat{F}_N(\lambda) - F_N(\lambda) \to_N 0$ a.s. The proof of this result relies on the use of the Stieltjes transform, and especially on theorem 1.2.1 in chapter 1, i.e it is equivalent to prove the convergence $\hat{m}_N(z) - m_N(z) \to_N 0$, where $\hat{m}_N$ and m_N are the Stieljes transform of $\hat{F}_N$ and F_N given by $\hat{m}_N(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\hat{F}_N(\lambda)}{\lambda - z}$ and $m_N(z) = \int_{\mathbb{R}} \frac{\mathrm{d}F_N(\lambda)}{\lambda - z}$.

In this chapter, we will review the main results concerning the convergence of $\hat{m}_N(z) - m_N(z)$ to zero, and provide further results. We will refine a result of Dumont et al. [17] and prove that $N^2|\mathbb{E}[\hat{m}_N(z)] - m_N(z)|$ is bounded by a certain product of polynomials in $|\text{Im}(z)|$ and $|z|$, independent of N. independent of N. This result will be the corner stone of the technics used to prove results about the localization of the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$, developped in the next chapter 3. We will also characterize the asymptotic behaviour of certain bilinear forms related with $\hat{m}_N(z)$, a result which will be useful in chapter 4 for the subspace estimation.

We also give several properties of the deterministic probability distribution F_N. In particular Dozier & Silverstein [15] proved that F_N is absolutely continuous with density $\pi^{-1}\text{Im}(m_N(x))$ where $m_N(x)$ is the continuity extension of m_N on the real axis. [15] also provides a way to numerically compute the support $\mathcal{S}_N$ of F_N. We will refine here this analysis and prove formally that $\mathcal{S}_N$ is the union of a finite number of compact disjoint intervals. Moreover, it appears, in the same way as [15], that the boundary points of $\mathcal{S}_N$ can be computed by looking at the local extrema of a certain function depending on the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$, σ^2 and the ratio $c_N = M/N$. Finally, once the support $\mathcal{S}_N$ of the limiting spectral distribution is identified, it is possible, using a contour integration technique developed in Mestre [49], to obtain the mass, by F_N, of any connected component composing $\mathcal{S}_N$.

This chapter is organized as follows. In section 2.1, we introduce the information plus noise model and the main related assumptions. The main result concerning the convergence of the empirical spectral distribution is also given, as well as the refinement of the bound of $N^2(\mathbb{E}[\hat{m}_N(z)] - m_N(z))$ mentionned above. We also prove a similar property for bilinear forms of the resolvent $(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^* - z\mathbf{I}_M)^{-1}$, a central tool in the study of spectrum of random matrices (see remark 1.3.1 in the previous chapter), and which will be also useful in the applicative part of this thesis, concerning the subspace estimation. Sections 2.3, 2.4 and 2.5 provide several properties of the deterministic equivalent spectral distribution, and particularly a complete study of its support. Finally, we postpone in section 2.7 several long proofs.

2.1 The complex Gaussian information plus noise model and notations

In this section, we present the information plus noise model of random matrices, which will be used throughout this thesis.

We consider $M, N \in \mathbb{N}^*$ such that $M = M(N)$, $M < N$ and $c_N = M/N \to c \in (0,1)$ as $N \to \infty$. A Gaussian information plus noise matrix is a $M \times N$ random matrix defined by

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

where

- matrix $\mathbf{B}_N$ is deterministic such that $\sup_N \|\mathbf{B}_N\| \leq B_{\max} < \infty$,
- the entries $W_{i,j,N}$ of $\mathbf{W}_N$ are i.i.d and satisfy $W_{i,j,N} \sim \mathcal{CN}\left(0, \sigma^2\right)$.

Remark 2.1.1. *Note that most of the results derived in this manuscript hold in the general case where $c_N, c \leq 1$, but we choose to consider $c_N, c < 1$ to enlight the computations. Some results can be also extended to the non-Gaussian case.*

The spectral decomposition of $\mathbf{B}_N\mathbf{B}_N^*$ and $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ will be denoted respectively by

$$\mathbf{B}_N\mathbf{B}_N^* = \mathbf{U}_N\mathbf{\Lambda}_N\mathbf{U}_N^* \quad \text{and} \quad \mathbf{\Sigma}_N\mathbf{\Sigma}_N^* = \hat{\mathbf{U}}_N\hat{\mathbf{\Lambda}}_N\hat{\mathbf{U}}_N^*, \tag{2.3}$$

with $\mathbf{U}_N, \hat{\mathbf{U}}_N$ unitary matrices and $\mathbf{\Lambda}_N = \mathrm{Diag}\left(\lambda_{1,N}, \ldots, \lambda_{M,N}\right)$, $\hat{\mathbf{\Lambda}}_N = \mathrm{Diag}\left(\hat{\lambda}_{1,N}, \ldots, \hat{\lambda}_{M,N}\right)$. The eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ and $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ will be ordered such that $0 \leq \lambda_{1,N} \leq \ldots \leq \lambda_{M,N}$ and $0 \leq \hat{\lambda}_{1,N} \leq \ldots \leq \hat{\lambda}_{M,N}$.

2.2 Resolvent of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ and convergence of the empirical spectral measure

In this section, we review the main results related to the convergence of the empirical spectral measure of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ defined by

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

with δ_x the Dirac measure at point x.

First, we define the resolvent of matrix $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$, by

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

for all $z \in \mathbb{C}\backslash\mathbb{R}^+$. Its normalized trace $\frac{1}{M}\mathrm{Tr}\,\mathbf{Q}_N(z)$ can be expressed in terms of the Stieltjes transform of $\hat{\mu}_N$, i.e

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

The weak convergence of $\hat{\mu}_N$ can be studied by characterizing the convergence of $\frac{1}{M}\mathrm{Tr}\,\mathbf{Q}_N(z)$ as $N \to \infty$ (theorem 1.2.1 in chapter 1). The main result of this section is the following.

Theorem 2.2.1. *It exists a deterministic probability measure μ_N, satisfying $\mathrm{supp}(\mu_N) \subset \mathbb{R}^+$, and such that $\hat{\mu}_N - \mu_N \xrightarrow{w} 0$ as $N \to \infty$ with probability one. Equivalently, the Stieltjes transform m_N of μ_N satisfies $\hat{m}_N(z) - m_N(z) \to 0$ almost surely $\forall z \in \mathbb{C}\backslash\mathbb{R}^+$. Moreover, $\forall z \in \mathbb{C}\backslash\mathbb{R}^+$, $m_N(z)$ is the unique solution of the equation*

$$m = \frac{1}{M}\mathrm{Tr}\left(-z(1+\sigma^2 c_N m)\mathbf{I}_M + \sigma^2(1-c_N)\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma^2 c_N m}\right)^{-1} \tag{2.4}$$

satisfying $\mathrm{Im}(zm_N(z)) > 0$ for $z \in \mathbb{C}^+$.

This result has been first proved by Girko [20] and later by Dozier-Silverstein [16]. Note that this result is also true in the non Gaussian case. The measure μ_N is referred from now on as the asymptotic spectral measure of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. The Stieltjes transform m_N of μ_N is defined by

$$m_N(z) = \frac{1}{M}\mathrm{Tr}\,\mathbf{T}_N(z), \tag{2.5}$$

with

$$\mathbf{T}_N(z) = \left(-z(1+\sigma^2 c_N m_N(z))\mathbf{I}_M + \sigma^2(1-c_N)\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma^2 c_N m_N(z)}\right)^{-1}, \tag{2.6}$$

for all $z \in \mathbb{C}\backslash\mathbb{R}^+$.

Remark 2.2.1. *If the spectral distribution $F_N(x) \triangleq \frac{1}{M}\mathrm{card}\left\{k : \lambda_{k,N} \le x\right\}$ of matrix $\mathbf{B}_N\mathbf{B}_N^*$ converges to the distribution $F(x)$ as $N\to\infty$, then $\mu_N \xrightarrow{w} \mu$, with μ a probability measure, whose Stieltjes transform $m(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu(\lambda)}{\lambda - z}$ satisfies*

$$m(z) = \int_{\mathbb{R}} \frac{\mathrm{d}F(\lambda)}{\frac{\lambda}{1+\sigma^2 c m(z)} - z(1+\sigma^2 c m(z)) + \sigma^2(1-c)}.$$

It is also possible to obtain further results concerning the convergence of $\mathbb{E}[\hat{m}_N(z)]$. The following result (proved in appendix 2.7.2) is a consequence of the complex Gaussian assumption, and does not hold in the general case.

Theorem 2.2.2. *For all $z \in \mathbb{C}\backslash\mathbb{R}$,*

$$\left|\mathbb{E}\left[\hat{m}_N(z)\right] - m_N(z)\right| \le \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$

for all large N, with $\mathrm{P}_1, \mathrm{P}_2$ two polynomials with positive coefficients independent of N, z.

From theorem 2.2.1, we know that $\frac{1}{M}\mathrm{Tr}\,\mathbf{Q}_N(z)$ is a good approximation of $\frac{1}{M}\mathrm{Tr}\,\mathbf{T}_N(z)$. A natural question is to know if the entries of $\mathbf{Q}_N(z)$ also approximate the entries of matrix $\mathbf{T}_N(z)$. The answer is given by the following result (proved in appendix 2.7.3).

Theorem 2.2.3. *Let $(\mathbf{d}_{1,N})$ and $(\mathbf{d}_{2,N})$ be two sequences of deterministic vectors such that $\sup_N \|\mathbf{d}_{1,N}\|, \sup_N \|\mathbf{d}_{2,N}\| < \infty$. Then, $\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z) - \mathbf{T}_N(z)\right)\mathbf{d}_{2,N} \to_N 0$ almost surely for all $z \in \mathbb{C}\backslash\mathbb{R}$. Moreover,*

$$\left|\mathbb{E}\left[\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z) - \mathbf{T}_N(z)\right)\mathbf{d}_{2,N}\right]\right| \le \frac{1}{N^{3/2}}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$

for all large N, with $\mathrm{P}_1, \mathrm{P}_2$ two polynomials with positive coefficients independent of N, z.

Note that the rate of convergence for the entries of $\mathbb{E}[\mathbf{Q}_N(z)]$ is slower than the rate for the trace $\mathbb{E}[\frac{1}{M}\mathrm{Tr}\,\mathbf{Q}_N(z)]$ in theorem 2.2.2. The result of theorem 2.2.3 has also been extended to the non-Gaussian case [24], but at a slower rate of $N^{-1/2}$. The almost-sure convergence $\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z) - \mathbf{T}_N(z)\right)\mathbf{d}_{2,N} \to_N 0$ will be essential in chapter 4 for the subspace estimation problem.

2.3 Function m_N and density of μ_N

In this section, we review some properties of the Stieltjes transform m_N of the deterministic probability measure μ_N defined in theorem 2.2.1. Most of the results have been shown in Dozier & Silverstein [15]. We define $\mathcal{S}_N = \mathrm{supp}(\mu_N)$, the support of μ_N.

First, we give the following fundamental bounds.

Lemma 2.3.1. *It holds that*

$$|m_N(z)| \le \frac{1}{\mathrm{dist}(z, \mathcal{S}_N)}, \tag{2.7}$$

and moreover,

$$\|\mathbf{T}_N(z)\| \le \frac{1}{\mathrm{dist}(z, \mathcal{S}_N)}. \tag{2.8}$$

Proof. (2.7) is a consequence of property 1.2.1, and (2.8) is proved in appendix 2.7.1. □

The main properties of m_N are gathered in the following result.

Theorem 2.3.1. 0 *does not belong to* $\mathscr{S}_N$ *and the function* m_N *satisfies the following properties.*

1. *The limit of* $m_N(z)$, *as* $z \in \mathbb{C}^+$ *converges to* $x \in \mathbb{R}$, *exists and is still denoted by* $m_N(x)$.
2. *The function* $x \to m_N(x)$ *thus defined is continuous on* $\mathbb{R}$ *and analytic on* $\mathbb{R}\backslash\partial\mathscr{S}_N$.
3. $m_N(x)$ *is solution of* (2.4) *for* $x \in \mathbb{R}\backslash\partial\mathscr{S}_N$.
4. μ_N *is absolutely continuous with density given by* $f_{\mu_N}(x) = \pi^{-1}\mathrm{Im}(m_N(x))$.

Proof. Items 1, 2 and 3 have been proved in [15], but only for $x \neq 0$. We show in appendix 2.7.4 that $0 \notin \mathscr{S}_N$ if $c_N < 1$, a property not established in [15]. This immediately implies that $m_N(0)$ is well-defined and satisfies (2.4), because the Stieltjes transform m_N is holomorphic in the neighborhood of 0, and item 3 also follows. Item 4 was also proved in [15]. □

Theorem 2.3.1 essentially states that the definition of $m_N(z)$ can be extended to the real axis by continuity on the upper half plane.

Remark 2.3.1. *We note that as* m_N *is a Stieltjes transform, it also satisfies* $m_N(z^*) = m_N(z)^*$. *Therefore, it holds that*

$$\lim_{\substack{z\in\mathbb{C}^- \\ z\to x}} m_N(z) = m_N(x)^*,$$

for $x \in \mathbb{R}$ *if* $c_N < 1$.

Remark 2.3.2. *From the last item of theorem 2.3.1, we have* $\mathrm{Int}(\mathscr{S}_N) = \left\{x \in \mathbb{R}^+ : \mathrm{Im}(m_N(x)) > 0\right\}$, *and* $m_N(x) \in \mathbb{R}$ *for* $x \in \mathbb{R}\backslash\mathscr{S}_N$.

We now end this section with the following fundamental bound, derived in appendix 2.7.5.

Property 2.3.1. *The following inequality* $\mathrm{Re}(1+\sigma^2 c_N m_N(z)) \geq 1/2$ *holds for each* $z \in \mathbb{C}$.

2.4 Functions w_N, ϕ_N and support of μ_N

This subsection is devoted to characterize the support $\mathscr{S}_N$ of μ_N. A study of the support is already provided in Dozier & Silverstein [15], and we will refine here the analysis.

The study of [15] is essentially based on characterizing $1+\sigma^2 c_N m_N(x)$ for $x \in \mathscr{S}_N$ among the solutions of a certain equation. We will use here the function

$$w_N(z) = z\left(1+\sigma^2 c_N m_N(z)\right)^2 - \sigma^2(1-c_N)\left(1+\sigma^2 c_N m_N(z)\right), \tag{2.9}$$

defined for all $z \in \mathbb{C}$, which is believed to simplify the analysis, and will be crucial for the subspace estimation part [1]. As we will see, w_N is closely related to the function

$$\phi_N(w) = w(1-\sigma^2 c_N f_N(w))^2 + \sigma^2(1-c_N)(1-\sigma^2 c_N f_N(w)), \tag{2.10}$$

with $f_N(w) = \frac{1}{M}\mathrm{Tr}\,\left(\mathbf{B}_N\mathbf{B}_N^* - w\mathbf{I}_M\right)^{-1}$. We will make the following additional assumption on the rank of $\mathbf{B}_N$.

Assumption A-1: *Matrix* $\mathbf{B}_N$ *has rank* $K = K(N) < M$, *and the eigenvalues of* $\mathbf{B}_N\mathbf{B}_N^*$ *have multiplicity one for all* N.

Thus $0 = \lambda_{1,N} = \ldots = \lambda_{M-K,N} < \lambda_{M-K+1,N} < \ldots < \lambda_{M,N}$. Note that K may scale-up with N or stay constant.

Remark 2.4.1. *Assumption* **A-1** *is not essential to study the properties of function* ϕ_N, *but this assumption will be necessary for the chapter 4. In fact,* ϕ_N *will have different behaviours depending on whether* $\mathbf{B}_N$ *is full rank or not,* $c_N < 1$ *or* $c_N = 1$. *These different cases will impact the support* $\mathscr{S}_N$, *in particular the property if* 0 *belongs or not to* $\mathscr{S}_N$.

[1] The characterization of the support given in this section (in theorem 2.4.1), can also be directly deduced by using the results of [15]. We choose here to develop and use the properties of the function w_N, which will often appear in the next chapters.

We now turn to the function w_N whose main properties are gathered in the following result, proved in appendix 2.7.6.

Property 2.4.1. *The following assertions hold:*

1. *The function $z \mapsto w_N(z)$ is continuous on $\mathbb{C}^+ \cup \mathbb{R}$ and the function $x \mapsto w_N(x)$ is continuously differentiable on $\mathbb{R}\backslash\partial\mathcal{S}_N$.*
2. $\mathrm{Im}(w_N(z)) > 0$ *if* $\mathrm{Im}(z) > 0$ *and* $w_N(x) \in \mathbb{R}\backslash\{\lambda_{1,N},\dots,\lambda_{M,N}\}$ *for* $x \in \mathbb{R}\backslash\mathcal{S}_N$.
3. *For $z \in \mathbb{C}\backslash\partial\mathcal{S}_N$, it holds that*
$$1+\sigma^2 c_N m_N(z) = (1-\sigma^2 c_N f_N(w_N(z)))^{-1} \tag{2.11}$$
and $w_N(z)$ is solution to the equation $\phi_N(w) = z$.
4. *For $x \in \mathbb{R}\backslash\mathcal{S}_N$, it holds that $w'_N(x), \phi'_N(w_N(x)), 1-\sigma^2 c_N f_N(w_N(x)) > 0$.*
5. $\mathrm{Int}(\mathcal{S}_N) = \{x \in \mathbb{R} : \mathrm{Im}(w_N(x)) > 0\}$.

Property 2.4.1 is basically stating the fact that it is sufficient to study the behavior of $\mathrm{Im}(w_N(x))$ in order to characterize the support $\mathcal{S}_N$. Item 3 is the most important one since it shows that $w_N(x)$ is solution of the equation $\phi_N(w) = x$, for $x \in \mathbb{R}\backslash\partial\mathcal{S}_N$. This observation will be the basis of the method used to characterize the support $\mathcal{S}_N$, which consists in identifying $w_N(x)$ out of the solutions of $\phi_N(w) = x$.

Remark 2.4.2. *By taking derivatives with respect to z on both sides of the equation $\phi_N(w_N(z)) = z$, we see that*
$$w'_N(z)\phi'_N(w_N(z)) = 1$$
holds for $z \in \mathbb{C}\backslash\partial\mathcal{S}_N$ (if $z \in \mathbb{R}\backslash\partial\mathcal{S}_N$, the derivative is taken in the real sense). From item 2,
$$\phi'_N(w_N(x)) = \lim_{y\downarrow 0}\phi'_N(w_N(x+iy))$$
is well defined, as well as $w'_N(x)$ (item 1). Thus, we deduce $w'_N(x) = \lim_{y\downarrow 0} w'_N(x+iy)$.

Remark 2.4.3. *Under Assumption* **A-1**, *the equation in $\phi_N(w) = x$ for $x \in \mathbb{R}$ is in fact equivalent to a polynomial equation of degree $2(K+1)$. This can be readily seen by using the expression of $f_N(w)$, so that we can express $\phi_N(w)$ as sums of quotients of polynomials in w, i.e.*
$$\begin{aligned}\phi_N(w) = w&\left(1+\sigma^2\frac{M-K}{M}\frac{c_N}{w} - \sigma^2\frac{c_N}{M}\sum_{m=M-K+1}^{M}\frac{1}{\lambda_{m,N}-w}\right)^2 \\ &+\sigma^2(1-c_N)\left(1+\sigma^2\frac{M-K}{M}\frac{c_N}{w} - \sigma^2\frac{c_N}{M}\sum_{m=M-K+1}^{M}\frac{1}{\lambda_{m,N}-w}\right).\end{aligned}$$
Hence, multiplying both sides of equation $\phi_N(w) = x$ by $w\prod_{m=M-K+1}^{M}\left(\lambda_{m,N}-w\right)^2$, we end up with a polynomial equation of degree $2(K+1)$.

The following result gather the main properties of function ϕ_N.

Property 2.4.2. 1. *Function ϕ_N satisfies the following limits,*
$$\lim_{\substack{w\in\mathbb{R}\\ w\downarrow 0}}\phi_N(w) = +\infty, \quad \lim_{\substack{w\in\mathbb{R}\\ w\uparrow 0}}\phi_N(w) = -\infty, \quad \text{and} \quad \lim_{\substack{w\in\mathbb{R}\\ w\to\lambda_{M-K+k,N}}}\phi_N(w) = +\infty,$$

2. *Function ϕ_N admits $2K+2$ real different zeros denoted $z^-_{0,N} < z^+_{0,N} < \ldots < z^-_{K,N} < z^+_{K,N}$, with $z^+_{0,N}, z^-_{1,N}, \ldots, z^-_{K,N}$ the $K+1$ zeros of function $w \mapsto 1-\sigma^2 c_N f_N(w)$. Moreover,*
   - $z^-_{0,N}, z^+_{0,N} < 0$ *and* $\phi'_N(z^-_{0,N}) > 0$, $\phi'_N(z^+_{0,N}) < 0$,
   - $z^-_{1,N}, z^+_{1,N} \in (0, \lambda_{M-K+1,N})$ *and* $\phi'_N(z^-_{1,N}) < 0$, $\phi'_N(z^+_{1,N}) > 0$,
   - *for* $k = 2,\ldots,K$, $z^-_{k,N}, z^+_{k,N} \in (\lambda_{M-K+k-1,N}, \lambda_{M-K+k,N})$ *and* $\phi'_N(z^-_{k,N}) < 0$, $\phi'_N(z^+_{k,N}) > 0$.

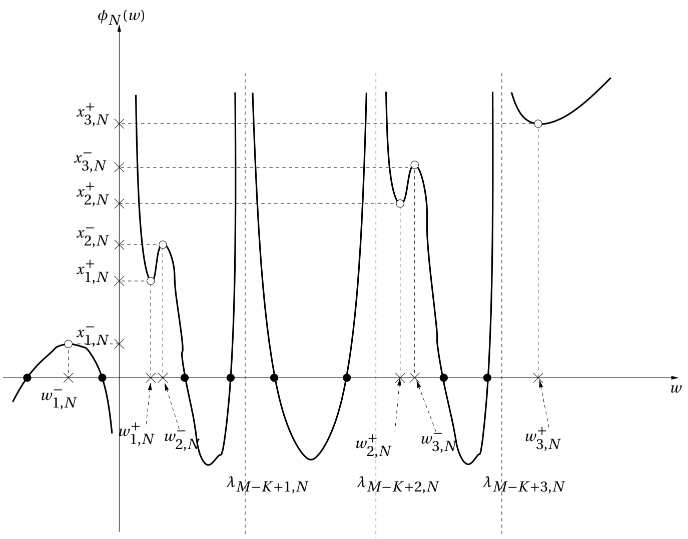


Figure 2.1: Typical representation of $\phi_N(w)$: in this case, $K = 3$ and black dots represent the zeros $z^-_{k,N}, z^+_{k,N}$.

3. *Function ϕ_N admits $2Q$ positive local extrema, with $Q = Q(N)$, $1 \le Q \le K+1$, whose preimages, denoted $w^-_{1,N} < 0 < w^+_{1,N} < w^-_{2,N} < \ldots < w^-_{Q,N} < w^+_{Q,N}$, belong to the set $\{w \in \mathbb{R} : 1 - \sigma^2 c_N f_N(w) > 0\}$. Moreover, we also have*

   - *the local extrema $x^-_{q,N} \triangleq \phi_N(w^-_{q,N})$ and $x^+_{q,N} \triangleq \phi_N(w^+_{q,N})$ satisfy*

$$0 < x^-_{1,N} < x^+_{1,N} < x^-_{2,N} < \ldots < x^-_{Q,N} < x^+_{Q,N}, \tag{2.12}$$

   - *each eigenvalue $\lambda_{k,N}$ of $\mathbf{B}_N\mathbf{B}_N^*$ belongs to an interval $(w^-_{q,N}, w^+_{q,N})$, $q = 1,\ldots,Q$,*
   - *each interval $(w^-_{q,N}, w^+_{q,N})$, $q = 1,\ldots,Q$ contains an eigenvalue of $\mathbf{B}_N\mathbf{B}_N^*$,*
   - *ϕ_N is increasing on $(-\infty, w^-_{1,N})$, $(w^+_{Q,N}, \infty)$, $(w^+_{q,N}, w^-_{q+1,N})$ for $q = 1,\ldots,Q-1$.*

4. *For $x \in \mathbb{R}\backslash\bigcup_{q=1}^{Q}[x^-_{q,N}, x^+_{q,N}]$, the equation $\phi_N(w) = x$ admits a unique solution satisfying $1 - \sigma^2 c_N f_N(w) > 0$ and $\phi'_N(w) > 0$.*

Proof. Property 2.4.2 follows from an elementary analysis of the function ϕ_N, except the ordering (2.12) (proved in appendix 2.7.7), which is non-trivial. □

A typical representation of function ϕ_N is given in figure 2.1. We are now in position to completely characterize the support $\mathcal{S}_N$.

Theorem 2.4.1. *Under Assumption* **A-1***, the support $\mathcal{S}_N$ of μ_N is given by*

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

Proof. 1. $\mathrm{Int}(\mathcal{S}_N) \subset \bigcup_{q=1}^{Q}[x^-_{q,N}, x^+_{q,N}]$: Let $x \in \mathbb{R}^+\backslash\bigcup_{q=1}^{Q}[x^-_{q,N}, x^+_{q,N}]$. From remark 2.4.3, the equation $\phi_N(w) = x$ has $2K+2$ solutions. Therefore, to prove that $x \notin \mathrm{Int}(\mathcal{S}_N)$, we have to verify that $w_N(x)$ can not be a non-real solution of $\phi_N(w) = x$, from property 2.4.1 items 3 and 5, and thus it is sufficient to prove that every solution of $\phi_N(w) = x$ is real. From property 2.4.2 items 1 and 2, we deduce that the equation $\phi_N(w) = x$ has at least $2K$ real solutions (see figure 2.1 for an example):

   - one solution in $(0, z^-_{1,N})$ and one in $(z^+_{1,N}, \lambda_{M-K+1,N})$,
   - one solution in $(\lambda_{M-K+k,N}, z^-_{k+1,N})$ and one in $(z^+_{k+1,N}, \lambda_{M-K+k+1,N})$ for $k = 1,\ldots,K-1$.

Assume moreover $x \in (0, x^-_{1,N})$, then using property 2.4.2, we easily deduce the existence of two real solutions located in $(z^-_{0,N}, z^+_{0,N})$, and therefore the equation $\phi_N(w) = x$ has $2K+2$ real solutions. If $x \in (x^+_{Q,N}, \infty)$, it is also clear from property 2.4.2 that there exists two real solutions in $(\lambda_{M,N}, \infty)$. Finally set $x \in (x^+_{q,N}, x^-_{q+1,N})$ for some $q \in \{1, \dots, Q-1\}$. By property 2.4.2 item 3, $x^+_{q,N}$ and $x^-_{q+1,N}$ are two positive local extrema with preimages $w^+_{q,N}$ and $w^-_{q+1,N}$ located in one of the intervals $(0, \lambda_{M-K+1,N})$, $(\lambda_{M-K+k,N}, \lambda_{M-K+k+1,N})$ for $k = 1, \dots, K-1$ and which satisfy $w^+_{q,N}, w^-_{q+1,N} \in \left\{w : 1 - \sigma^2 c_N f_N(w) > 0\right\}$. More precisely, this means that $w^+_{q,N}$, $w^-_{q+1,N}$ are located in one of the intervals $(0, z^-_{1,N})$, $(\lambda_{M-K+k,N}, z^-_{k+1,N})$ for $k = 1, \dots, K-1$. Since ϕ_N is increasing on $(w^+_{q,N}, w^-_{q+1,N})$ the equation $\phi_N(w) = x$ will admit two more solutions in one of the intervals $(0, z^-_{1,N})$, $(\lambda_{M-K+k,N}, z^-_{k+1,N})$ for $k = 1, \dots, K-1$ (see further figure 2.1). This proves once again that the equation $\phi_N(w) = x$ has $2K+2$ real solutions for $x \in (x^+_{q,N}, x^-_{q+1,N})$. Therefore all the solutions of equation $\phi_N(w) = x$ are real for $x \in \mathbb{R}^+ \setminus \bigcup_{q=1}^{Q} [x^-_{q,N}, x^+_{q,N}]$.

2. $\bigcup_{q=1}^{Q} (x^-_{q,N}, x^+_{q,N}) \subset \mathrm{Int}(\mathscr{S}_N)$: Let $x \in \bigcup_{q=1}^{Q} (x^-_{q,N}, x^+_{q,N})$. From the previous arguments, we deduce that equation $\phi_N(w) = x$ has at least $2K$ real solutions:

   - one solution in $(0, z^-_{1,N})$ and one in $(z^+_{1,N}, \lambda_{M-K+1,N})$,
   - one solution in $(\lambda_{M-K+k,N}, z^-_{k+1,N})$ and one in $(z^+_{k+1,N}, \lambda_{M-K+k+1,N})$ for $k = 1, \dots, K-1$,

   and we can check that none of these solutions satisfy both the conditions $1 - \sigma^2 c_N f_N(w) > 0$ and $\phi'_N(w) > 0$. Since there are $2K+2$ solutions to the equations, it remains two more solutions. If the two remaining solutions of the equation are real, then they are necessarily located in one of the intervals $(z^+_{k,N}, \lambda_{M-K+k,N})$ for $k = 1, \dots, K$ (see figure 2.1) and therefore do not verify the inequality $1 - \sigma^2 c_N f_N(w) > 0$. Therefore, $x \notin \mathbb{R} \setminus \mathscr{S}_N$, otherwise $w_N(x)$ would be one of the real solutions and would satisfies the inequality $1 - \sigma^2 c_N f_N(w) > 0$ (property 2.4.1 items 3 and 4). On the other hand, if the two remaining solutions are complex conjuguate, then again $x \notin \mathbb{R} \setminus \mathscr{S}_N$ (otherwise $w_N(x)$ would be one of the real solutions). Therefore, we have

$$\mathbb{R} \setminus \mathscr{S}_N \subset \mathbb{R} \setminus \bigcup_{q=1}^{Q}]x^-_{q,N}, x^+_{q,N}[,$$

   which in turns implies $\bigcup_{q=1}^{Q} (x^-_{q,N}, x^+_{q,N}) \subset \mathrm{Int}(\mathscr{S}_N)$.

□

Remark 2.4.4. *If $x \in \mathbb{R} \setminus \mathscr{S}_N$, it is easy to see from property 2.4.2 and the arguments of the proof of theorem 2.4.1 (see also figure 2.1), that $w_N(x)$ is the unique solution of $\phi_N(w) = x$ satisfying $1 - \sigma^2 c_N f_N(w) > 0$ and $\phi'_N(w) > 0$. Moreover, w_N maps $(x^+_{q,N}, x^-_{q+1,N})$ to $(w^+_{q,N}, w^-_{q+1,N})$, and property 2.4.1 implies that $w_N(x)$ is increasing on $(x^+_{q,N}, x^-_{q+1,N})$. Therefore, since w_N is continuous on $\mathbb{R}$, we get*

$$w_N(x^-_{q,N}) = w^-_{q,N} \text{ and } w_N(x^+_{q,N}) = w^+_{q,N}. \tag{2.13}$$

Moreover, from remark 2.4.2, we have the equality $\phi'_N(w_N(x))\, w'_N(x) = 1$ for $x \in \mathbb{R} \setminus \partial\mathscr{S}_N$. But

$$\phi'_N(w_N(x)) \xrightarrow[x \to x^-_{q,N}]{} \phi'_N(w^-_{q,N}) = 0$$

(and similarly for $x^+_{q,N}$), which shows that the derivative $w'_N(x)$ is unbounded when x approaches the boundary of the support $\partial\mathscr{S}_N$.

The next result states that the support of μ_N does not escape to infinity.

Lemma 2.4.1. *Under Assumption* **A-1***,*

$$\sup_N x^+_{Q,N} < \infty.$$

Proof. Recall that $\mathrm{B}_{\max} = \sup_N \|\mathbf{B}_N\|$. Thus, for $w > \mathrm{B}_{\max}$

$$\sup_N |f_N(w)| \le \frac{1}{|\mathrm{B}_{\max} - w|},$$
$$\sup_N |f'_N(w)| \le \frac{1}{|\mathrm{B}_{\max} - w|^2},$$
$$\sup_N |w f'_N(w)| \le \frac{w}{|\mathrm{B}_{\max} - w|^2},$$

and since $\phi'_N(w) = (1-\sigma^2 c_N f_N(w))^2 - 2\sigma^2 c_N w f'_N(w)(1-\sigma^2 c_N f_N(w)) - \sigma^4 c_N(1-c_N) f'_N(w)$ converges towards 1 when $w \to +\infty$, we deduce that for $\epsilon > 0$, $\exists w_\epsilon > \mathrm{B}_{\max}$ such that $\forall w > w_\epsilon$, $\phi'_N(w) > \epsilon$ for all N. Since $\phi'_N(w^+_{Q,N}) = 0$, this implies that

$$\sup_N w^+_{Q,N} \leq w_\epsilon < +\infty.$$

Moreover, using $w^+_{Q,N} = w_N(x^+_{Q,N}) = x^+_{Q,N}(1+\sigma^2 c_N m_N(x^+_{Q,N}))^2 - \sigma^2(1-c_N)(1+\sigma^2 c_N m_N(x^+_{Q,N}))$, and property 2.3.1 in section 2.3, we get

$$x^+_{Q,N} \leq \frac{w_\epsilon}{(1+\sigma^2 c_N m_N(x^+_{Q,N}))^2} + \frac{\sigma^2(1-c_N)}{1+\sigma^2 c_N m_N(x^+_{Q,N})} < 4w_\epsilon + 2\sigma^2,$$

which concludes the proof. □

Theorem 2.4.1 states that the support $\mathscr{S}_N$ is a disjoint reunion of compact intervals, which will be referred to as "clusters". Each of these clusters $[x^-_{q,N}, x^+_{q,N}]$ can be computed from function ϕ_N since $x^-_{1,N} < x^+_{1,N} < \ldots < x^-_{Q,N} < x^+_{Q,N}$ coincide with the set of all positive extrema of function ϕ_N. Moreover, each cluster is associated to an interval of the type $[w^-_{q,N}, w^+_{q,N}]$, $q = 1\ldots Q$, in the sense that $x^-_{q,N} = \phi_N(w^-_{q,N})$ and $x^+_{q,N} = \phi_N(w^+_{q,N})$ (property 2.4.2 item 3). On the other hand, we also see that a specific eigenvalue $\lambda_{k,N}$, $k = 1,\ldots,M$, always belongs to one, and only one of the intervals $[w^-_{q,N}, w^+_{q,N}]$ (property 2.4.2 item 3). This motivates the following terminology.

> **We will say that eigenvalue $\lambda_{k,N}$, $k = 1,\ldots,M$ of $\mathbf{B}_N\mathbf{B}_N^*$ is "associated" with the cluster $[x^-_{q,N}, x^+_{q,N}]$ if $\lambda_{k,N} \in [w^-_{q,N}, w^+_{q,N}]$.**

Observe that this association is not a one-to-one correspondence, in the sense that multiple consecutive eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ may be associated with the same cluster. Notice moreover that 0 is always associated with the cluster $[x^-_{1,N}, x^+_{1,N}]$. We now show that the mass of any cluster by μ_N is directly related to the number of associated eigenvalues. Define

$$\mathscr{I}_q = \{k \in \{1,\ldots,K\} : \lambda_{M-K+k,N} \in (w^-_{q,N}, w^+_{q,N})\}, \tag{2.14}$$

and $|\mathscr{I}_q| = \mathrm{card}(\mathscr{I}_q) > 0$. Note that $|\mathscr{I}_q| > 0$ from property 2.4.2, and $|\mathscr{I}_q|$ is the number of non zero eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ associated with $[x^-_{q,N}, x^+_{q,N}]$. The following result in proved in appendix 2.7.8.

Property 2.4.3. *Under assumption* **A-1***, it holds that*

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

and for $q \geq 2$,

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

Property 2.4.3 basically states that the mass of a cluster by μ_N is exactly the proportion of eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ associated with this cluster.

2.5 The special case of spiked models

In this section, we consider the special case where the rank of $\mathbf{B}_N\mathbf{B}_N^*$ is independent of N. In this case, it is possible to compute explicitely the boundary points of the support $\mathscr{S}_N$, and thus to obtain the precise characterization of μ_N, in terms of its support and concentration of its mass. We replace Assumption A-1 by the following stronger assumption.

Assumption A-2: *The rank $K > 0$ of $\mathbf{B}_N\mathbf{B}_N^*$ does not depend on N and for all $k = 1,\ldots,K$, the positive sequence $(\lambda_{M-K+k,N})$ writes*

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

with $\lim_{N\to+\infty} \varepsilon_{k,N} = 0$ and $\gamma_1 < \ldots < \gamma_K$.

Before characterizing the support $\mathscr{S}_N$, we discuss some consequences of assumption A-2. From theorem 2.2.1 (section 2.1), it is straightforward to see that for $z \in \mathbb{C}^+$, $\hat{m}_N(z) \to_N m(z)$ almost surely, with

$$m(z) = \frac{1}{-z(1+\sigma^2 c m(z)) + \sigma^2(1-c)}. \tag{2.15}$$

$m(z)$ the Stieltjes transform of the Marcenko-Pastur μ, given by

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

We now study the support $\mathscr{S}_N$ of measure μ_N, under the assumption A-2. We recall that an eigenvalue $\lambda_{k,N}$ of $\mathbf{B}_N\mathbf{B}_N^*$ is associated with the cluster $[x_{q,N}^-, x_{q,N}^+]$ if $\lambda_{k,N} \in (w_{q,N}^-, w_{q,N}^+)$ (see section 2.4 for details). Recall also that eigenvalue 0 is always associated with $[x_{1,N}^-, x_{1,N}^+]$. The following theorem is proved in appendix 2.7.9.

Theorem 2.5.1. *Under Assumption A-2, define $K_s = \mathrm{card}\{k : \gamma_k > \sigma^2\sqrt{c}\}$ and assume that $\sigma^2\sqrt{c} \notin \{\gamma_1, \ldots, \gamma_K\}$, i.e*

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

Thus, for N large enough, the support $\mathscr{S}_N$ has $Q = K_s + 1$ clusters, i.e $\mathscr{S}_N = \bigcup_{q=1}^{K_s+1}[x_{q,N}^-, x_{q,N}^+]$. The first cluster is associated with $\lambda_{1,N}, \ldots, \lambda_{M-K_s,N}$ and is given by

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

For $q = 2, \ldots, K_s+1$ and $k = q-1$, the cluster $[x_{q,N}^-, x_{q,N}^+]$ is associated with $\lambda_{M-K_s+k,N}$ and

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

where $\psi(\lambda, c) = \frac{(\lambda+\sigma^2 c)(\lambda+c)}{\lambda}$ and $\mathscr{O}^+\left(\frac{1}{\sqrt{N}}\right)$ is a positive $\mathscr{O}\left(\frac{1}{\sqrt{N}}\right)$ term.

Under the spiked model assumption, measure μ_N is intuitively expected to be very close to the Marcenko-Pastur distribution μ, and particularly $\mathscr{S}_N$ should be close to $\mathrm{supp}(\mu) = [\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2]$. Theorem 2.5.1 shows the first cluster $[x_{1,N}^-, x_{1,N}^+]$ is very close to the support of the Marcenko-Pastur distribution and we have presence of additional clusters if the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ are large enough. Indeed, if K_s eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ converge to different limits, above the threshold $\sigma^2\sqrt{c}$, then there will be K_s additional clusters in the support of $\mathscr{S}_N$ for all large N.

Theorem 2.5.1 also states that the smallest $M - K_s$ eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ are associated with the first cluster, or equivalently that $\mu_N([x_{1,N}^-, x_{1,N}^+]) = \frac{M-K_s}{M}$, and that $\mu_N([x_{k,N}^-, x_{k,N}^+]) = \frac{1}{M}$ for $k = 2, \ldots, K_s$

2.6 Discussion and numerical examples

In this section, we discuss the results of theorem 2.4.1 of section 2.4, and give some numerical examples.

The analysis provided in the previous sections shows that the boundary points of $\mathscr{S}_N$ $x_{1,N}^- < x_{1,N}^+ < \ldots < x_{Q,N}^- < x_{Q,N}^+$ coincide with the set of all positive extrema of function ϕ_N, and moreover we have seen that the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ are associated to the clusters of $\mathscr{S}_N$, i.e. $\lambda_{k,N}$ is associated with $[x_{q,N}^-, x_{q,N}^+]$ if $\lambda_{k,N} \in (w_{q,N}^-, w_{q,N}^+)$, with $w_{q,N}^-, w_{q,N}^+$ the preimages of $x_{q,N}^-, x_{q,N}^+$ by ϕ_N.

In figures 2.2, 2.3 and 2.4, we have represented the typical behaviour of function ϕ_N, for a rank $K = 3$ matrix $\mathbf{B}_N$. Several situations are drawn, which are commented below.

In figure 2.2, we see that all the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ are associated to $[x_{1,N}^-, x_{1,N}^+]$. There exists no positive local extrema other than $x_{1,N}^-$ and $x_{1,N}^+$, and thus $\mathscr{S}_N$ has only one cluster, i.e $\mathscr{S}_N = [x_{1,N}^-, x_{1,N}^+]$.

In figure 2.3, eigenvalue 0 is associated to $[x_{1,N}^-, x_{1,N}^+]$, $\lambda_{M-2,N}, \lambda_{M-1,N}$ are associated with $[x_{2,N}^-, x_{2,N}^+]$, and $\lambda_{M,N}$ is associated with $[x_{3,N}^-, x_{3,N}^+]$. The support $\mathscr{S}_N$ is composed of 3 clusters.

In figure 2.4, each eigenvalue is associated to a different cluster.

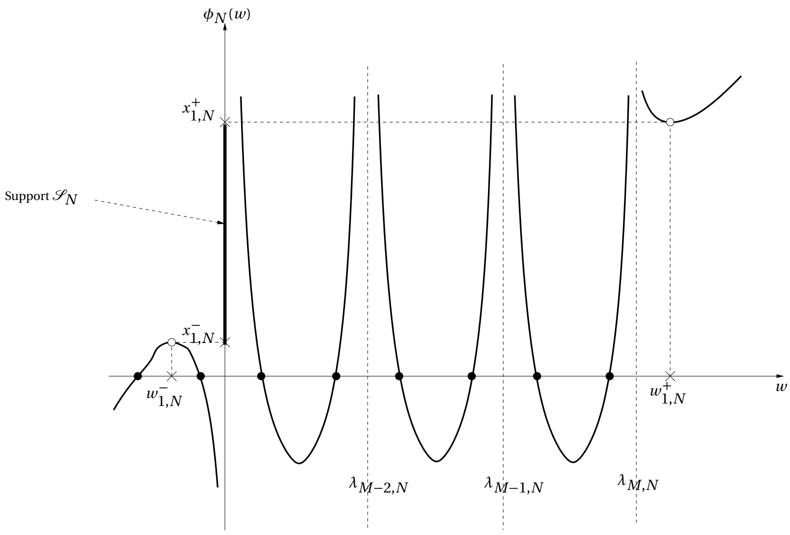


Figure 2.2: $\phi_N(w)$: Case 1

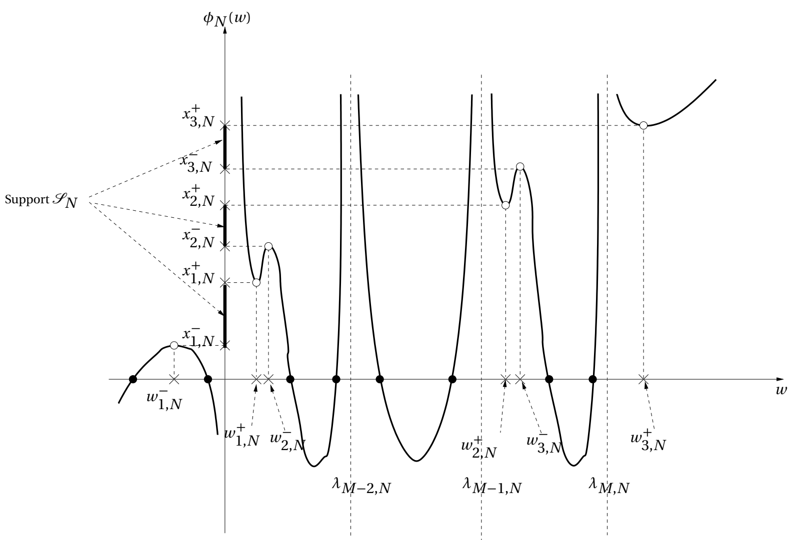


Figure 2.3: $\phi_N(w)$: Case 2

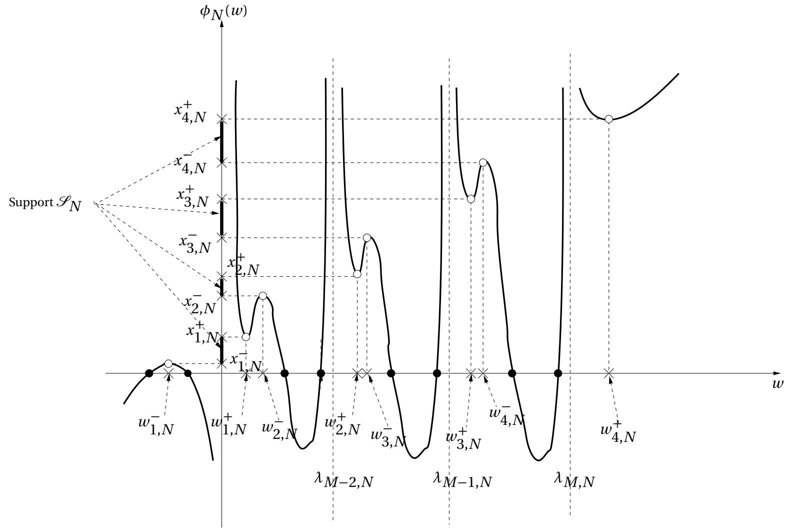


Figure 2.4: $\phi_N(w)$: Case 3

The conditions for the support $\mathscr{S}_N$ to split into several clusters depend in a non-trivial way of σ, the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ and the distance between them. Nevertheless, under the stronger Assumption **A-2** (i.e. K independent of N and convergence of the eigenvalues to different limits), we have obtained explicit conditions for the separation of the eigenvalues: an eigenvalue of $\mathbf{B}_N\mathbf{B}_N^*$ is separated from the others if its limit is greater than $\sigma^2\sqrt{c}$ (see theorem 2.5.1). The non-separated eigenvalues are those associated with $[x^-_{1,N}, x^+_{1,N}]$. Therefore in the spiked model case, the behaviour of the clusters of $\mathscr{S}_N$ is completely characterized.

In figure 2.5, we plot the density of μ_N in the case where $c_N = 0.5$, $\sigma = 1$ and the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ are 0 (multiplicity $M-2$), 5, 10, for $N = 20, 100, 200, 2000$. With this parameters, $5 > \sigma^2\sqrt{c_N}$ and thus there are two clusters associated with eigenvalue 5 and 10. Moreover, the width of each cluster tends to 0 at rate $\mathscr{O}(N^{-1/2})$ (theorem 2.5.1).

In figure 2.6, we have plotted the density of μ_N, in the context where $N = 20$, $M = 10$, $\sigma = 1$ and $\mathbf{B}_N\mathbf{B}_N^*$ is diagonal with three different eigenvalues, 0 with multiplicity 5, 5 with multiplicity 2 and 10 with multiplitticy 3. The density is plotted by computing $x \mapsto \pi^{-1}\mathrm{Im}(m_N(x+iy))$ for $y << 1$ (see theorem 2.3.1). Note that $m_N(z)$ for $z \in \mathbb{C}^+$ is evaluated iteratively using the fixed point equation (2.4), and this procedure always converges as proved in [23]. On figure 2.6, we also notice that the density behaves as a square root near the boundary points of the support, a result already proved in [15]. This result will be of importance, for proving that w'_N is integrable in a neighborhood of the boundary points (property 2.7.1 in appendix 2.7.8).

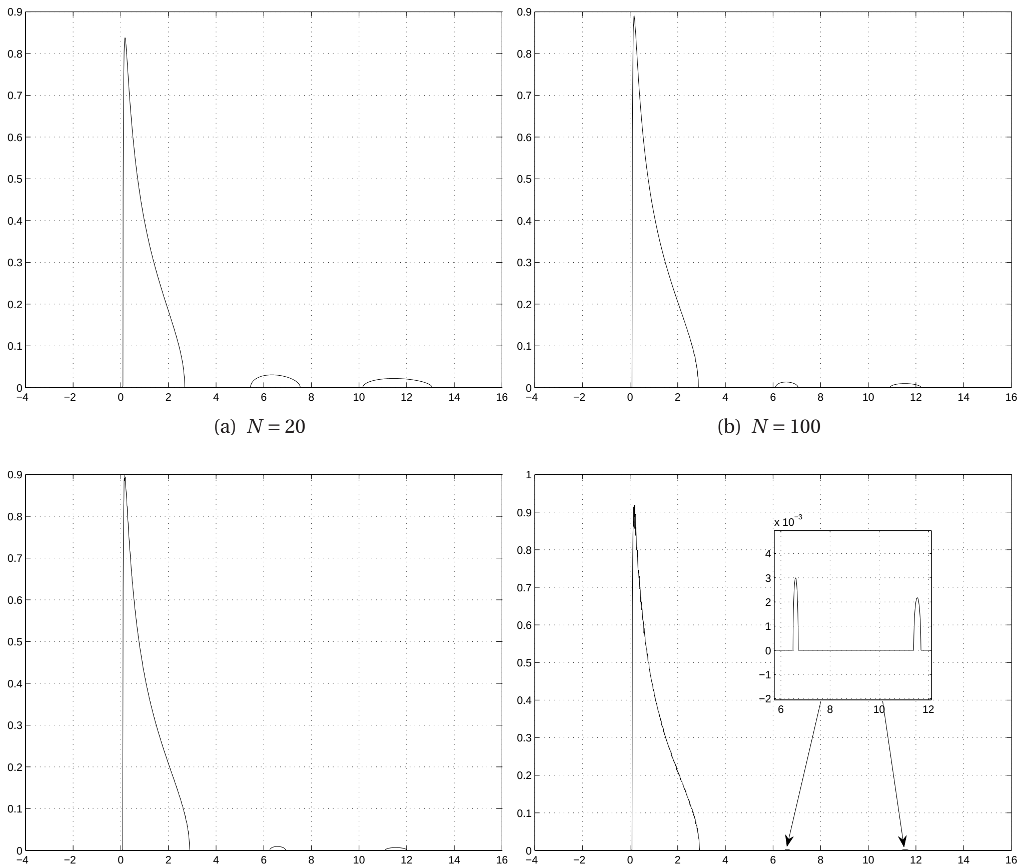
In figure 2.7, we have plotted the function $w_N(x)$ for $x \in \mathbb{R}$, in the same settings as for figure 2.6. More precisely, figure 2.7 represents $\mathrm{Im}(w_N(x))$ versus $\mathrm{Re}(w_N(x))$. As stated in property 2.4.1, we see that $w_N(\mathbb{R})$ does not meet any eigenvalue, $w_N(x)$ has positive imaginary part for $x \in \mathscr{S}_N$ and is real for $x \in \mathbb{R}\backslash\mathscr{S}_N$. If $\mathscr{C}_{q,N} = \{w_N(x) : x \in [x^-_{q,N}, x^+_{q,N}]\} \cup \{w_N(x)^* : x \in [x^-_{q,N}, x^+_{q,N}]\}$, this shows that $\mathscr{C}_{q,N}$ encloses the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ associated with $[x^-_{q,N}, x^+_{q,N}]$. In the present example, the eigenvalue 0 is associated with $[x^-_{1,N}, x^+_{1,N}]$, 5 is associated with $[x^-_{2,N}, x^+_{2,N}]$ and 10 with $[x^-_{3,N}, x^+_{3,N}]$. Therefore, in figure 2.7, $\mathscr{C}_{1,N}$ encloses eigenvalue 0, and eigenvalue 5 is in turn enclosed by $\mathscr{C}_{2,N}$ and 10 by $\mathscr{C}_{3,N}$. This property will be fundamental in the chapter on subspace estimation, to perform contour integration. Note that this property is also used in the proof of 2.4.3 in appendix 2.7.8.

2.7 Appendix

2.7.1 Bound on $\|\mathbf{T}_N(z)\|$

In this section, we prove the inequality $\|\mathbf{T}_N(z)\| \le \mathrm{dist}(z, \mathscr{S}_N)^{-1}$ for all $z \in \mathbb{C}\backslash\mathscr{S}_N$, or equivalently

$$\mathbf{T}_N(z)\mathbf{T}_N(z)^* \le \frac{\mathbf{I}_M}{\mathrm{dist}(z, \mathscr{S}_N)^2}. \tag{2.16}$$



(a) $N = 20$ (b) $N = 100$

(c) $N = 200$ (d) $N = 2000$

Figure 2.5: Illustration of the spiked model assumption on the density of μ_N

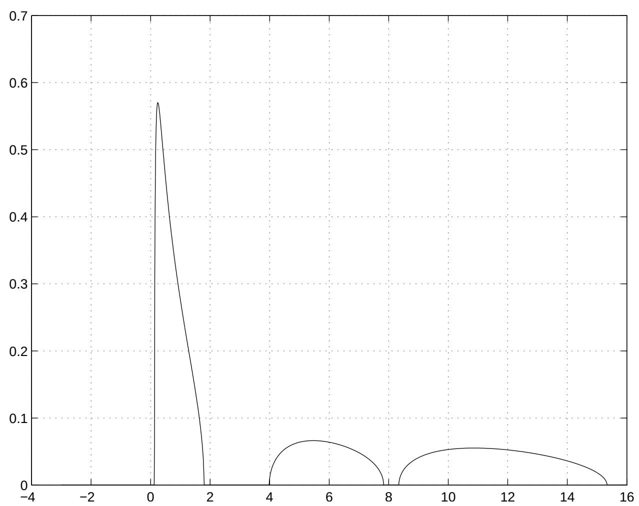


Figure 2.6: Density of μ_N

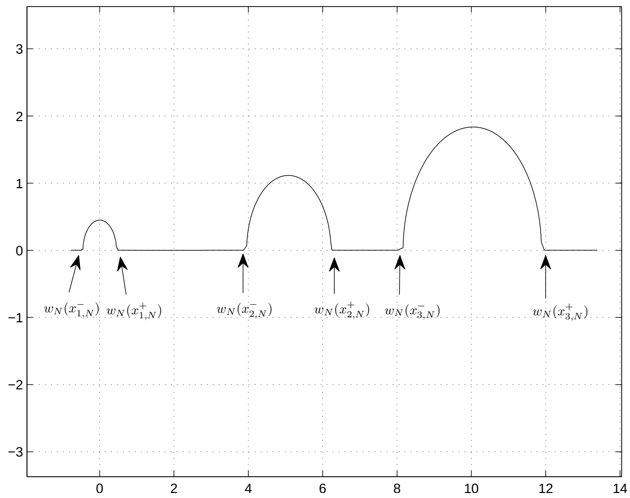


Figure 2.7: $\mathrm{Im}(w_N(x))$ versus $\mathrm{Re}(w_N(x))$ for $x \in \mathbb{R}$

with $\mathscr{S}_N = \mathrm{supp}(\mu_N)$. It is shown in Hachem et al [23, Th. 2.4 & Prop. 2.2]) that $\mathbf{T}_N$ coincides with the Stieltjes transform of a positive matrix-valued measure $\boldsymbol{\mu}_N$ with support $\mathscr{S}_N$ such that $\boldsymbol{\mu}_N(\mathscr{S}_N) = \mathbf{I}_M$, i.e.

$$\mathbf{T}_N(z) = \int_{\mathscr{S}_N} \frac{d\boldsymbol{\mu}_N(\lambda)}{\lambda - z}.$$

Since $m_N(z)$ is solution to the equation (2.4), it is clear that $\frac{1}{M}\mathrm{Tr}\, \boldsymbol{\mu}_N = \mu_N$. In order to establish (2.16), we follow the proof of in Hachem et al. [23, Prop.5.1]. We first remark that function $\tilde{m}_N(z)$ defined by

$$\tilde{m}_N(z) = c_N m_N(z) - \frac{1-c_N}{z}$$

is the Stieltjes transform of the probability measure $\tilde{\mu}_N = c_N\mu_N + (1-c_N)\delta_0$. The support of $\tilde{\mu}_N$ thus coincides with $\mathscr{S}_N \cup \{0\}$, and is included in $\mathbb{R}^+$. Therefore, it holds that $\frac{\mathrm{Im}(z\tilde{m}_N(z))}{\mathrm{Im}(z)} > 0$ if $z \in \mathbb{C}\backslash\mathbb{R}$ (see property 1.2.1). We remark that

$$\frac{\mathbf{T}_N(z) - \mathbf{T}_N(z)^*}{2i} = \mathrm{Im}(z) \int_{\mathscr{S}_N} \frac{d\boldsymbol{\mu}_N(\lambda)}{|\lambda - z|^2}.$$

By using the identity,

$$\mathbf{T}_N(z) - \mathbf{T}_N(z)^* = \mathbf{T}_N(z)\left(\mathbf{T}_N(z)^{-*} - \mathbf{T}_N(z)^{-1}\right)\mathbf{T}_N(z)^*,$$

we get after some algebra

$$\mathrm{Im}(z)\int_{\mathscr{S}_N} \frac{d\boldsymbol{\mu}_N(\lambda)}{|\lambda - z|^2} =$$

$$\mathrm{Im}(z)\mathbf{T}_N(z)\mathbf{T}_N(z)^* + \sigma^2 \mathrm{Im}(z\tilde{m}_N(z))\mathbf{T}_N(z)\mathbf{T}_N(z)^* + \frac{\sigma^2 c_N}{|1+\sigma^2 c_N m_N(z)|^2}\mathrm{Im}(m_N(z))\mathbf{T}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)^*,$$

for each $z \in \mathbb{C}\backslash\mathbb{R}$, or equivalently

$$\int_{\mathscr{S}_N} \frac{d\boldsymbol{\mu}_N(\lambda)}{|\lambda - z|^2} =$$

$$\mathbf{T}_N(z)\mathbf{T}_N(z)^* + \sigma^2 \frac{\mathrm{Im}(z\tilde{m}_N(z))}{\mathrm{Im}(z)}\mathbf{T}_N(z)\mathbf{T}_N(z)^* + \frac{\sigma^2 c_N}{|1+\sigma^2 c_N m_N(z)|^2}\frac{\mathrm{Im}(m_N(z))}{\mathrm{Im}(z)} \times \mathbf{T}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)^*.$$

Consequently, we obtain that

$$\mathbf{T}_N(z)\mathbf{T}_N(z)^* \leq \int_{\mathscr{S}_N} \frac{d\boldsymbol{\mu}_N(\lambda)}{|\lambda - z|^2}$$

for $z \in \mathbb{C}\backslash\mathbb{R}$, but also for $z \in \mathbb{C}\backslash\mathscr{S}_N$ because both members of the above inequality are continuous on $\mathbb{C}\backslash\mathscr{S}_N$. This immedialely leads to (2.16).

2.7.2 Proof of theorem 2.2.2

In this section, we prove that it exists two polynomials $\mathrm{P}_1, \mathrm{P}_2$ independent of N, with positive coefficients, such that

$$|\mathbb{E}[\hat{m}_N(z)] - m_N(z)| \leq \frac{1}{N^2} \mathrm{P}_1(|z|) \mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right). \tag{2.17}$$

for all $z \in \mathbb{C} \backslash \mathbb{R}$. Since $\hat{m}_N$ and m_N are Stieltjes transforms, $\hat{m}_N(z^*) = \hat{m}_N(z)^*$ and $m_N(z^*) = m_N(z)^*$, and thus we only need to prove the result for $z \in \mathbb{C}^+$. For the remainder of this section, $\mathrm{P}_1, \mathrm{P}_2$ will be used as generic polynomials whose values may change from one line to another.

Some auxiliary quantities

For ease of reading, we will use the notations

$$\alpha_N(z) = \mathbb{E}[\sigma c_N \hat{m}_N(z)] = \mathbb{E}\left[\frac{\sigma}{N} \mathrm{Tr}\, \mathbf{Q}_N(z)\right] \quad \text{and} \quad \delta_N(z) = \sigma c_N m_N(z).$$

We also define

$$\tilde{\alpha}_N(z) = \alpha_N(z) - \frac{\sigma(1-c_N)}{z} \quad \text{and} \quad \tilde{\delta}_N(z) = \delta_N(z) - \frac{\sigma(1-c_N)}{z}.$$

From section 1.2 in chapter 1, it is easy to check that $\alpha_N(z)$ and $\tilde{\alpha}_N(z)$ are Stieltjes transforms of respectively the finite measures $\sigma c_N \mathbb{E}[\hat{\mu}_N]$ and $\sigma c_N \mathbb{E}[\hat{\mu}_N] + \sigma(1-c_N)\delta_0$, carried by $\mathbb{R}^+$. Thus, for $z \in \mathbb{C}^+$, $\alpha_N(z)$, $z\alpha_N(z)$, $z\tilde{\alpha}_N(z)$ and $\tilde{\alpha}_N(z)$ belong to $\mathbb{C}^+$. Similarly, $\delta_N(z)$ and $\tilde{\delta}_N(z)$ are Stieltjes transforms of the finite measures $\sigma c_N \mu_N$ and $\sigma c_N \mu_N + \sigma(1-c_N)\delta_0$.

Moreover, $[-z(1+\sigma\alpha_N(z))]^{-1}$ and $[-z(1+\sigma\tilde{\alpha}_N(z))]^{-1}$ are also Stieltjes transform of probability measures carried by $\mathbb{R}^+$, which in particular imply $\left|\frac{1}{1+\sigma\alpha_N(z)}\right| \leq \frac{|z|}{\mathrm{Im}(z)}$. Similarly, we have the same bounds for $[-z(1+\sigma\delta_N(z))]^{-1}$ and $\left[-z(1+\sigma\tilde{\delta}_N(z))\right]^{-1}$.

We remark also that $\tilde{\alpha}_N(z) = \mathbb{E}\left[\frac{\sigma}{N} \mathrm{Tr}\, \tilde{\mathbf{Q}}_N(z)\right]$ with $\tilde{\mathbf{Q}}_N(z) = \left(\mathbf{\Sigma}_N^* \mathbf{\Sigma}_N - z\mathbf{I}_N\right)^{-1}$. Matrix $\mathbf{T}_N(z)$ defined in (2.6) writes

$$\mathbf{T}_N(z) = \left(-z(1+\sigma\tilde{\delta}_N(z))\mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^*}{1+\sigma\delta_N(z)}\right)^{-1},$$

and $\delta_N(z) = \frac{\sigma}{N} \mathrm{Tr}\, \mathbf{T}_N(z)$. We also define

$$\tilde{\mathbf{T}}_N(z) = \left(-z(1+\sigma\delta_N(z))\mathbf{I}_N + \frac{\mathbf{B}_N^* \mathbf{B}_N}{1+\sigma\tilde{\delta}_N(z)}\right)^{-1},$$

and remark, after simple calculations, that $\tilde{\delta}_N(z) = \frac{\sigma}{N} \mathrm{Tr}\, \tilde{\mathbf{T}}_N(z)$. We finally denote by $\mathbf{R}_N(z)$ and $\tilde{\mathbf{R}}_N(z)$ the matrices defined by

$$\mathbf{R}_N(z) = \left(-z(1+\sigma\tilde{\alpha}_N(z))\mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^*}{1+\sigma\alpha_N(z)}\right)^{-1} \quad \text{and} \quad \tilde{\mathbf{R}}_N(z) = \left(-z(1+\sigma\alpha_N(z))\mathbf{I}_N + \frac{\mathbf{B}_N^* \mathbf{B}_N}{1+\sigma\tilde{\alpha}_N(z)}\right)^{-1}. \tag{2.18}$$

The above remarks show that

$$\|\mathbf{R}_N(z)\| = \min_{k=1,\dots,M} \left|\frac{\lambda}{1+\sigma\alpha_N(z)} - z(1+\sigma\tilde{\alpha}_N(z))\right|^{-1} \leq \min_{k=1,\dots,M} \left|\mathrm{Im}\left(\frac{\lambda}{1+\sigma\alpha_N(z)} - z(1+\sigma\tilde{\alpha}_N(z))\right)\right|^{-1} \leq \frac{1}{\mathrm{Im}(z)}.$$

The same bound of course holds for $\tilde{\mathbf{R}}_N(z)$, as well as $\mathbf{T}_N(z)$, $\tilde{\mathbf{T}}_N(z)$. We finally remark that $\frac{1}{N} \mathrm{Tr}\, \mathbf{R}_N(z)$ is the Stieltjes transform of a probability measure carried by $\mathbb{R}^+$.

A fundamental system of equations

We now consider the difference $\mathbf{R}_N(z) - \mathbf{T}_N(z) = \mathbf{R}_N(z)\left(\mathbf{T}_N(z)^{-1} - \mathbf{R}_N(z)^{-1}\right)\mathbf{T}_N(z)$ and use the mere definition of $\mathbf{R}_N(z)$, $\mathbf{T}_N(z)$ to get

$$\frac{\sigma}{N} \mathrm{Tr}\, (\mathbf{R}_N(z) - \mathbf{T}_N(z)) = \left(\tilde{\alpha}_N(z) - \tilde{\delta}_N(z)\right) z v_N(z) + (\alpha_N(z) - \delta_N(z))\, u_N(z),$$

with $u_N(z) \triangleq \frac{\sigma^2}{N} \mathrm{Tr}\, \frac{\mathbf{R}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)}{(1+\sigma\alpha_N(z))(1+\sigma\delta_N(z))}$ and $v_N(z) \triangleq \frac{\sigma^2}{N} \mathrm{Tr}\, \mathbf{R}_N(z)\mathbf{T}_N(z)$. Similar computations give

$$\frac{\sigma}{N}\mathrm{Tr}\, \left(\tilde{\mathbf{R}}_N(z) - \tilde{\mathbf{T}}_N(z)\right) = (\alpha_N(z) - \delta_N(z))\, z\tilde{v}_N(z) + \left(\tilde{\alpha}_N(z) - \tilde{\delta}_N(z)\right)\tilde{u}_N(z),$$

with $\tilde{u}_N(z) \triangleq \frac{\sigma^2}{N} \mathrm{Tr}\, \frac{\tilde{\mathbf{R}}_N(z)\mathbf{B}_N^*\mathbf{B}_N\tilde{\mathbf{T}}_N(z)}{(1+\sigma\tilde{\alpha}_N(z))(1+\sigma\tilde{\delta}_N(z))}$ and $\tilde{v}_N(z) \triangleq \frac{\sigma^2}{N} \mathrm{Tr}\, \tilde{\mathbf{R}}_N(z)\tilde{\mathbf{T}}_N(z)$. Define

$$\epsilon_N(z) = \alpha_N(z) - \frac{\sigma}{N}\mathrm{Tr}\, \mathbf{R}_N(z). \tag{2.19}$$

Since

$$\frac{\sigma}{N}\mathrm{Tr}\, \mathbf{R}_N(z) = \frac{1+\sigma\alpha_N(z)}{1+\sigma\tilde{\alpha}_N(z)}\frac{\sigma}{N}\mathrm{Tr}\, \tilde{\mathbf{R}}_N(z) + \frac{\sigma(1-c_N)}{z(1+\sigma\tilde{\alpha}_N(z))},$$

we easily deduce that

$$\tilde{\epsilon}_N(z) \triangleq \tilde{\alpha}_N(z) - \frac{\sigma}{N}\mathrm{Tr}\, \tilde{\mathbf{R}}_N(z) = \epsilon_N(z)\frac{1+\sigma\tilde{\alpha}_N(z)}{1+\sigma\alpha_N(z)}. \tag{2.20}$$

Moreover, from matrix inversion lemma, $\frac{\mathbf{B}_N^*\mathbf{T}_N(z)}{1+\sigma\delta_N(z)} = \frac{\mathbf{T}_N(z)\mathbf{B}_N^*}{1+\sigma\tilde{\delta}_N(z)}$ and $\frac{\mathbf{B}_N^*\mathbf{R}_N(z)}{1+\sigma\alpha_N(z)} = \frac{\mathbf{R}_N(z)\mathbf{B}_N^*}{1+\sigma\tilde{\alpha}_N(z)}$, which implies $u_N(z) = \tilde{u}_N(z)$. Therefore, gathering the previous expressions, we obtain the following 2×2 linear system

$$\begin{bmatrix}\alpha_N(z)-\delta_N(z)\\ \tilde{\alpha}_N(z)-\tilde{\delta}_N(z)\end{bmatrix} = \begin{bmatrix} u_N(z) & zv_N(z)\\ z\tilde{v}_N(z) & u_N(z)\end{bmatrix}\begin{bmatrix}\alpha_N(z)-\delta_N(z)\\ \tilde{\alpha}_N(z)-\tilde{\delta}_N(z)\end{bmatrix} + \begin{bmatrix}\epsilon_N(z)\\ \tilde{\epsilon}_N(z)\end{bmatrix}. \tag{2.21}$$

The determinant associated with this system is defined by

$$\Delta_N(z) = (1-u_N(z))^2 - z^2 v_N(z)\tilde{v}_N(z).$$

Bound for $\epsilon_N(z)$

We first prove that

$$|\epsilon_N(z)| \le \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right), \tag{2.22}$$

Of course, the same type of bound holds also for $\tilde{\epsilon}_N(z)$ from (2.20) .

We first consider the following useful result.

Lemma 2.7.1. *For $z \in \mathbb{C}\backslash\mathbb{R}$, let $(\mathbf{M}_N(z))$ a sequence of deterministic matrix such that*

$$\|\mathbf{M}_N(z)\| \le \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

Then,

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{N}\mathrm{Tr}\, \mathbf{Q}_N(z)\mathbf{M}_N(z)\right] &\le \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),\\ \mathrm{Var}\left[\frac{1}{N}\mathrm{Tr}\, \mathbf{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{M}_N(z)\right] &\le \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).\end{aligned} \tag{2.23}$$

Proof. As the proofs of the two statements are similar, we just prove the first statement of the lemma. We first remark that

$$\frac{\partial[\mathbf{Q}_N(z)]_{p,q}}{\partial W_{i,j,N}} = -[\mathbf{Q}_N(z)]_{p,i}\left[\mathbf{\Sigma}_N^*\mathbf{Q}_N(z)\right]_{j,q} \quad \text{and} \quad \frac{\partial[\mathbf{Q}_N(z)]_{p,q}}{\partial W_{i,j,N}^*} = -[\mathbf{Q}_N(z)]_{i,q}\left[\mathbf{Q}_N(z)\mathbf{\Sigma}_N\right]_{p,j},$$

and the Poincaré inequality (theorem 1.4.2 in section chapter 1) gives

$$
\begin{aligned}
&\operatorname{Var}\left[\frac{1}{N}\operatorname{Tr}\mathbf{Q}_N(z)\mathbf{M}_N(z)\right] \\
&\leq \frac{\sigma^2}{N}\sum_{i,j}\left[\mathbb{E}\left|\frac{1}{N}\sum_{p,q}\frac{\partial[\mathbf{Q}_N(z)]_{p,q}}{\partial \mathbf{W}_{i,j,N}}[\mathbf{M}_N(z)]_{q,p}\right|^2+\mathbb{E}\left|\frac{1}{N}\sum_{p,q}\frac{\partial[\mathbf{Q}_N(z)]_{p,q}}{\partial \mathbf{W}^*_{i,j,N}}[\mathbf{M}_N(z)]_{q,p}\right|^2\right] \\
&\leq \frac{C}{N^3}\sum_{i,j}\left[\mathbb{E}\left|\left[\boldsymbol{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{M}_N(z)\mathbf{Q}_N(z)\right]_{j,i}\right|^2+\mathbb{E}\left|\left[\mathbf{Q}_N(z)\mathbf{M}_N(z)\mathbf{Q}_N(z)\boldsymbol{\Sigma}_N^*\right]_{i,j}\right|^2\right] \\
&\leq \frac{C}{N^3}\mathbb{E}\left[\operatorname{Tr}\mathbf{Q}_N(z)\mathbf{M}_N(z)\mathbf{Q}_N(z)\mathbf{Q}_N(z)^*\mathbf{M}_N(z)^*\mathbf{Q}_N(z)^*\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right] \\
&\quad+\frac{C}{N^3}\mathbb{E}\left[\operatorname{Tr}\mathbf{Q}_N(z)^*\mathbf{M}_N(z)^*\mathbf{Q}_N(z)^*\mathbf{Q}_N(z)\mathbf{M}_N(z)\mathbf{Q}_N(z)\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right].
\end{aligned}
$$

Using the resolvent identity $\mathbf{Q}_N(z)\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^* = \mathbf{I}_M + z\mathbf{Q}_N(z)$ together with the bounds $\|\mathbf{Q}_N(z)\| \leq \frac{1}{\operatorname{Im}(z)}$ and $\|\mathbf{M}_N(z)\| \leq \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{\operatorname{Im}(z)}\right)$ leads the desired result. □

To prove (2.22), we make use of a n equation involving $\mathbb{E}[\mathbf{Q}_N(z)]$. Define

$$\tilde{\tau}_N(z) = \frac{-\sigma}{z(1+\sigma\alpha_N(z))}\left(1-\frac{1}{N}\operatorname{Tr}\left(\frac{\mathbf{B}_N^*\mathbb{E}[\mathbf{Q}_N(z)]\mathbf{B}_N}{1+\sigma\alpha_N(z)}\right)\right). \tag{2.24}$$

It is shown in Dumont et al. [17] that for all $z \in \mathbb{R}_*^-$,

$$\mathbf{I}_M + \boldsymbol{\Delta}_N(z) = \mathbb{E}[\mathbf{Q}_N(z)]\left(-z(1+\sigma\tilde{\tau}_N(z))\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma\alpha_N(z)}\right), \tag{2.25}$$

where

$$\boldsymbol{\Delta}_N(z) = \boldsymbol{\Delta}_{1,N}(z) + \boldsymbol{\Delta}_{2,N}(z) + \boldsymbol{\Delta}_{3,N}(z), \tag{2.26}$$

with

$$\boldsymbol{\Delta}_{1,N}(z) = \frac{\sigma}{1+\sigma\alpha_N(z)}\mathbb{E}\left[\mathbf{Q}_N(z)\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\frac{\sigma}{N}\operatorname{Tr}\left(\mathbf{Q}_N(z)-\mathbb{E}[\mathbf{Q}_N(z)]\right)\right] \tag{2.27}$$

$$\boldsymbol{\Delta}_{2,N}(z) = \frac{\sigma^2}{1+\sigma\alpha_N(z)}\mathbb{E}\left[\left(\mathbf{Q}_N(z)-\mathbb{E}[\mathbf{Q}_N(z)]\right)\frac{\sigma}{N}\operatorname{Tr}\boldsymbol{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{B}_N\right] \tag{2.28}$$

$$\boldsymbol{\Delta}_{3,N}(z) = -\frac{\sigma^2}{(1+\sigma\alpha_N(z))^2}\mathbb{E}[\mathbf{Q}_N(z)]\,\mathbb{E}\left[\frac{\sigma}{N}\operatorname{Tr}\left(\mathbf{Q}_N(z)-\mathbb{E}[\mathbf{Q}_N(z)]\right)\frac{\sigma}{N}\operatorname{Tr}\boldsymbol{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{B}_N\right] \tag{2.29}$$

It is easy to see that the lefthandside and the righthandside of (2.25) are holomorphic matrix-valued functions on $\mathbb{C}\backslash\mathbb{R}^+$, and thus (2.25) holds not only on $\mathbb{R}_*^-$, but also on $\mathbb{C}\backslash\mathbb{R}^+$ by analytic continuation. As it will become apparent below, the entries of matrix $\boldsymbol{\Delta}_N(z)$ converge towards 0.

The general expression of $\tilde{\alpha}_N(z)-\tilde{\tau}_N(z)$ given in [17] is complicated. However, the simplicity of the model considered in this work (matrices $\mathbf{D}$ and $\tilde{\mathbf{D}}$ in [17] are reduced to $\sigma\mathbf{I}_M$ and $\sigma\mathbf{I}_N$) allows to derive a simpler expression. Indeed, taking the trace on both sides of (2.25) yields

$$\frac{\sigma}{N}\operatorname{Tr}\left(\frac{\mathbf{B}_N^*\mathbb{E}[\mathbf{Q}_N(z)]\mathbf{B}_N}{1+\sigma\alpha_N(z)}\right) = \sigma c_N + \frac{\sigma}{N}\operatorname{Tr}\boldsymbol{\Delta}_N(z) + z\left(1+\sigma\tilde{\tau}_N(z)\right)\alpha_N(z),$$

and from the definition of $\tilde{\tau}_N(z)$ in (2.24), we also have

$$\frac{\sigma}{N}\operatorname{Tr}\left(\frac{\mathbf{B}_N^*\mathbb{E}[\mathbf{Q}_N(z)]\mathbf{B}_N}{1+\sigma\alpha_N(z)}\right) = z\tilde{\tau}_N(z)(1+\sigma\alpha_N(z)) + \sigma c_N.$$

The two previous equalities imply that

$$\tilde{\tau}_N(z) = \tilde{\alpha}_N(z) + \frac{\sigma}{z}\frac{1}{N}\operatorname{Tr}\boldsymbol{\Delta}_N(z). \tag{2.30}$$

Inserting (2.30) in (2.25) and taking the trace, we finally obtain

$$\frac{1}{N}\mathrm{Tr}\ (\mathbb{E}[\mathbf{Q}_N(z)]-\mathbf{R}_N(z))=\frac{\sigma}{N}\mathrm{Tr}\,\mathbb{E}[\mathbf{Q}_N(z)]\,\mathbf{R}_N(z)\frac{\sigma}{N}\mathrm{Tr}\ \boldsymbol{\Delta}_N(z)+\frac{1}{N}\mathrm{Tr}\ \boldsymbol{\Delta}_N(z)\mathbf{R}_N(z). \tag{2.31}$$

Since $\left|\frac{\sigma}{N}\mathrm{Tr}\,\mathbb{E}[\mathbf{Q}_N(z)]\,\mathbf{R}_N(z)\right|\leq\sigma c_N|\mathrm{Im}(z)|^{-1}$, to prove (2.22), it is sufficient to check that for $i=1,2,3$,

$$\left|\frac{1}{N}\mathrm{Tr}\ \boldsymbol{\Delta}_{i,N}\mathbf{M}_N(z)\right|\leq\frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right), \tag{2.32}$$

with $\mathbf{M}_N(z)$ as in lemma 2.7.1. We just prove the result for $i=1$, the case $i=2,3$ being similar. Using the classical identity $\mathbf{Q}_N(z)\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*=\mathbf{I}_M+z\mathbf{Q}_N(z)$ yields

$$\frac{1}{N}\mathrm{Tr}\ \boldsymbol{\Delta}_{1,N}(z)\mathbf{M}_N(z)=\frac{\sigma^2 z}{1+\sigma\alpha_N(z)}\mathbb{E}\left[\left(\frac{1}{N}\mathrm{Tr}\ (\mathbf{Q}_N(z)-\mathbb{E}[\mathbf{Q}_N(z)])\right)^2\right].$$

Since $|1+\sigma\alpha_N(z)|^{-1}\leq|z||\mathrm{Im}(z)|^{-1}$, lemma 2.7.1 gives immediately (2.32). This concludes the proof of (2.22).

Convergence of $\alpha_N(z)-\delta_N(z)$

We now prove that

$$\alpha_N(z)-\delta_N(z)\xrightarrow[N\to\infty]{}0, \tag{2.33}$$

for all $z\in\mathbb{C}^+$. For this, we use the system of equations (2.21) and (2.22). From the bounds developed above, we have $\|\mathbf{R}_N(z)\|,\|\mathbf{T}_N(z)\|\leq\mathrm{Im}(z)^{-1}$ as well as $|1+\sigma^2c_N\alpha_N(z)|^{-1},|1+\sigma^2c_N\delta_N(z)|^{-1}\leq|z|\mathrm{Im}(z)^{-1}$ and therefore,

$$\begin{aligned}1-|u_N(z)|&\geq1-\frac{\sigma^2B_{\max}|z|^2}{\mathrm{Im}(z)^4},\\|z|^2|v_N(z)||\tilde{v}_N(z)|&\leq\frac{\sigma^4|z|^2}{\mathrm{Im}(z)^4},\end{aligned}$$

where we recall that $B_{\max}\geq\sup_N\|\mathbf{B}_N\|<\infty$. Therefore, we have $|\Delta_N(z)|\geq(1-|u_N(z)|)^2-|z|^2|v_N(z)||\tilde{v}_N(z)|>\frac{1}{2}$ for all z in the open set

$$\mathscr{E}=\left\{z\in\mathbb{C}^+:1-\frac{\sigma^2B_{\max}|z|^2}{\mathrm{Im}(z)^4}>\frac{1}{2}\right\}\cap\left\{z\in\mathbb{C}^+:\left(1-\frac{\sigma^2B_{\max}|z|^2}{\mathrm{Im}(z)^4}\right)^2-\frac{\sigma^4|z|^2}{\mathrm{Im}(z)^4}>\frac{1}{2}\right\}.$$

Therefore, by inverting the system (2.21), we obtain $g_N(z)=\alpha_N(z)-\delta_N(z)\to_N 0$ for each $z\in\mathscr{E}$, from the convergence (2.22). But since $\alpha_N(z)$ and $\delta_N(z)$ are Stieltjes transforms of finite measures with support included in $\mathbb{R}^+$, we deduce that g_N is holomorphic on $\mathbb{C}^+$ and $|g_N(z)|\leq\frac{2\sigma c_N}{\mathrm{Im}(z)}$, which means in particular that the sequence (g_N) is uniformly bounded on each compact subset of $\mathbb{C}^+$. Therefore (g_N) is a normal family by Montel's theorem. If $(g_{\varphi(N)})$ is a subsequence which converges uniformly on each compact of $\mathbb{C}^+$ to the holomorphic function g, then $g(z)=0$ for $z\in\mathscr{E}$, which implies g is identically zero on $\mathbb{C}^+$, by analytic continuation. Thus all converging subsequence of the normal family (g_N) converge to 0, therefore the whole sequence (g_N) converge to 0 uniformly on each compact of $\mathbb{C}^+$, which proves (2.33).

Bound for $\Delta_N(z)$

We now tackle the most demanding step of the proof, namely computing a polynomial bound for $|\Delta_N(z)|^{-1}$. First, from the definition of $\Delta_N(z)$, we clearly have the inequality

$$|\Delta_N(z)|\geq(1-|u_N(z)|)^2-|z|^2|v_N(z)||\tilde{v}_N(z)|.$$

Applying Cauchy-Schwarz inequality (via property 1.3.1 in chapter 1), we obtain

$$|\Delta_N(z)|\geq\left(1-\left|u_{1,N}(z)\right|^{1/2}\left|u_{2,N}(z)\right|^{1/2}\right)^2-|z|^2\left|v_{1,N}(z)\right|^{1/2}\left|v_{2,N}(z)\right|^{1/2}\left|\tilde{v}_{1,N}(z)\right|^{1/2}\left|\tilde{v}_{2,N}(z)\right|^{1/2}, \tag{2.34}$$

with $u_{1,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\frac{\mathbf{T}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)^*}{|1+\sigma\delta_N(z)|^2}$, $u_{2,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\frac{\mathbf{R}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{R}_N(z)^*}{|1+\sigma\alpha_N(z)|^2}$, and

$$v_{1,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{T}_N(z)\mathbf{T}_N(z)^*, \qquad v_{2,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{R}_N(z)\mathbf{R}_N(z)^*,$$
$$\tilde{v}_{1,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\tilde{\mathbf{T}}_N(z)\tilde{\mathbf{T}}_N(z)^*, \qquad \tilde{v}_{2,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\tilde{\mathbf{R}}_N(z)\tilde{\mathbf{R}}_N(z)^*.$$

For the remainder, we define

$$\Delta_{1,N}(z) \triangleq \left(1-u_{1,N}(z)\right)^2 - z^2 v_{1,N}(z)\tilde{v}_{1,N}(z) \quad \text{and} \quad \Delta_{2,N}(z) \triangleq \left(1-u_{2,N}(z)\right)^2 - z^2 v_{2,N}(z)\tilde{v}_{2,N}(z).$$

The following two lemmas are dedicated to study separately polynomial bounds for $|\Delta_{1,N}(z)|^{-1}$ and $|\Delta_{2,N}(z)|^{-1}$.

Lemma 2.7.2. *For $z \in \mathbb{C}^+$,*

$$\Delta_{1,N}(z)^{-1} \leq \mathrm{P}_1\left(|z|\right)\mathrm{P}_2\left(\mathrm{Im}(z)^{-1}\right).$$

Proof. Using the formula $\mathbf{T}_N(z) - \mathbf{T}_N(z)^* = \mathbf{T}_N(z)\left(\mathbf{T}_N^{-*} - \mathbf{T}_N(z)\right)\mathbf{T}_N(z)^*$, it is easy to see that

$$\mathrm{Im}\left(\delta_N(z)\right) = u_{1,N}(z)\mathrm{Im}\left(\delta_N(z)\right) + v_{1,N}(z)\mathrm{Im}\left(z\tilde{\delta}_N(z)\right) + \frac{\mathrm{Im}(z)}{\sigma}v_{1,N}(z).$$

with $u_{1,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\frac{\mathbf{T}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)^*}{\left|1+\sigma\tilde{\delta}_N(z)\right|^2}$ and $v_{1,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{T}_N(z)\mathbf{T}_N(z)^*$. Similarly, we obtain

$$\mathrm{Im}\left(z\tilde{\delta}_N(z)\right) = |z|^2\tilde{v}_{1,N}(z)\mathrm{Im}\left(\delta_N(z)\right) + \tilde{u}_{1,N}(z)\mathrm{Im}\left(z\tilde{\delta}_N(z)\right) + \frac{\mathrm{Im}(z)}{\sigma}\tilde{u}_{1,N}(z).$$

with $\tilde{u}_{1,N} \triangleq \frac{\sigma^2}{N}\frac{\tilde{\mathbf{T}}_N(z)\mathbf{B}_N^*\mathbf{B}_N\tilde{\mathbf{T}}_N(z)^*}{\left|1+\sigma\delta_N(z)\right|^2}$ and $\tilde{v}_{1,N}(z) \triangleq \frac{\sigma^2}{N}\mathrm{Tr}\,\tilde{\mathbf{T}}_N(z)\tilde{\mathbf{T}}_N(z)^*$. From the identity $\frac{\tilde{\mathbf{T}}_N(z)\mathbf{B}_N^*}{1+\sigma\tilde{\delta}_N(z)} = \frac{\mathbf{B}_N^*\mathbf{T}_N(z)}{1+\sigma\delta_N(z)}$, we deduce $u_{1,N}(z) = \tilde{u}_{1,N}(z)$, and thus we have the following system

$$\begin{bmatrix}\mathrm{Im}\left(\delta_N(z)\right)\\ \mathrm{Im}\left(z\tilde{\delta}_N(z)\right)\end{bmatrix} = \begin{bmatrix}u_{1,N}(z) & v_{1,N}(z)\\ |z|^2\tilde{v}_{1,N}(z) & u_{1,N}(z)\end{bmatrix}\begin{bmatrix}\mathrm{Im}\left(\delta_N(z)\right)\\ \mathrm{Im}\left(z\tilde{\delta}_N(z)\right)\end{bmatrix} + \frac{\mathrm{Im}(z)}{\sigma}v_N(z)\begin{bmatrix}v_{1,N}(z)\\ u_{1,N}(z)\end{bmatrix}. \tag{2.35}$$

whose determinant is exactly $\Delta_{1,N}(z)$. From Cauchy-Schwarz inequality,

$$v_{1,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{T}_N(z)\mathbf{T}_N(z)^* \geq \frac{\sigma^2}{MN}\left|\mathrm{Tr}\,\mathbf{T}_N(z)\right|^2 \geq \frac{1}{c_N}\left|\delta_N(z)\right|^2 \geq \frac{1}{c_N}\mathrm{Im}\left(\delta_N(z)\right)^2 > 0,$$

for $z \in \mathbb{C}^+$, since δ_N is a Stieltjes transform. Thus, we can rewrite the first equation of the system (2.35) as

$$\mathrm{Im}\left(z\tilde{\delta}_N(z)\right) = \left(1-u_{1,N}(z)\right)\frac{\mathrm{Im}\left(\delta_N(z)\right)}{v_{1,N}(z)} - \frac{\mathrm{Im}(z)}{\sigma}v_{1,N}(z).$$

Mixing with the second equation, straightforward computations yields

$$\Delta_{1,N}(z) = \left(1-u_{1,N}(z)\right)^2 - z^2 v_{1,N}(z)\tilde{v}_{1,N}(z) = \frac{v_{1,N}(z)}{\sigma}\frac{\mathrm{Im}(z)}{\mathrm{Im}\left(\delta_N(z)\right)}.$$

Since $|1+\sigma\delta_N(z))| \geq \sigma\mathrm{Im}(\delta_N(z)) > 0$ for $z \in \mathbb{C}^+$, it is easy to see that

$$\begin{aligned}\frac{\delta_N(z)}{1+\sigma\delta_N(z)} - \frac{\delta_N(z)^*}{1+\sigma\delta_N(z)^*} &= \frac{\sigma}{N}\mathrm{Tr}\left(\left[\mathbf{B}_N\mathbf{B}_N^* - w_N(z)\right]^{-1} - \left[\mathbf{B}_N\mathbf{B}_N^* - w_N(z)^*\right]^{-1}\right)\\ &= \left(w_N(z) - w_N(z)^*\right)\frac{v_{1,N}(z)}{\sigma\left|1+\sigma\delta_N(z)\right|^2},\end{aligned}$$

with $w_N(z) \triangleq z(1+\sigma\delta_N(z))(1+\sigma\tilde{\delta}_N(z))$. This implies $\mathrm{Im}\left(\delta_N(z)\right) = \mathrm{Im}\left(w_N(z)\right)\frac{v_{1,N}(z)}{\sigma}$. Consequently, $\mathrm{Im}(w_N(z)) > 0$ and

$$\Delta_{1,N}(z) = \frac{\mathrm{Im}(z)}{\mathrm{Im}\left(w_N(z)\right)}. \tag{2.36}$$

With this expression, it is easy to compute an upperbound on $\Delta_{1,N}(z)^{-1}$. Indeed, from the relation between $\delta_N(z)$ and $\tilde{\delta}_N(z)$, we can rewrite $w_N(z) = z(1+\sigma\delta_N(z))^2 - \sigma^2(1-c_N)(1+\sigma\delta_N(z))$. Since $\delta_N(z)$ is the Stieltjes transform of the finite measure $\sigma c_N \mu_N$, we have $|\delta_N(z)| \le \frac{\sigma c_N}{\mathrm{Im}(z)}$ for $z \in \mathbb{C}^+$, which implies that $|w_N(z)| \le \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{\mathrm{Im}(z)}\right)$. Using this bound in (2.36), we finally get

$$\Delta_{1,N}(z)^{-1} \le \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{\mathrm{Im}(z)}\right).$$

□

We now turn to a polynomial bound concerning $\Delta_{2,N}(z)$.

Lemma 2.7.3. *There exists two polynomials* $\mathrm{Q}_1, \mathrm{Q}_2$ *with positive coefficients, independent of* N, z *and a set*

$$\mathscr{E}_N = \left\{ z \in \mathbb{C}^+ : 1 - \frac{1}{N^2}\mathrm{Q}_1(|z|)\,\mathrm{Q}_2\left(\frac{1}{\mathrm{Im}(z)}\right) > 0 \right\}.$$

such that $u_{2,N}(z) < 1$ *and*

$$\Delta_{2,N}(z)^{-1} \le \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{\mathrm{Im}(z)}\right),$$

for all large N *and* $z \in \mathscr{E}_N$.

Proof. From (2.22) and (2.20), we have

$$\alpha_N(z) = \frac{\sigma}{N}\mathrm{Tr}\,\mathbf{R}_N(z) + \epsilon_N(z) \text{ and } \tilde{\alpha}_N(z) = \frac{\sigma}{N}\mathrm{Tr}\,\tilde{\mathbf{R}}_N(z) + \tilde{\epsilon}_N(z),$$

with $|\epsilon_N(z)|, |\tilde{\epsilon}_N(z)| \le \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)$. Using $\mathbf{R}_N(z) - \mathbf{R}_N(z)^* = \mathbf{R}_N(z)(\mathbf{R}_N(z)^{-*} - \mathbf{R}_N(z)^{-1})\mathbf{R}_N(z)^*$ and $z\mathbf{R}_N(z) - z^*\mathbf{R}_N(z)^* = |z|\mathbf{R}_N(z)(z^{-*}\mathbf{R}_N(z)^{-*} - z^{-1}\mathbf{R}_N(z)^{-1})\mathbf{R}_N(z)^*$, we get

$$\begin{bmatrix} \mathrm{Im}(\alpha_N(z)) \\ \mathrm{Im}(z\tilde{\alpha}_N(z)) \end{bmatrix} = \begin{bmatrix} u_{2,N}(z) & v_{2,N}(z) \\ |z|^2\tilde{v}_{2,N}(z) & \tilde{u}_{2,N}(z) \end{bmatrix} \begin{bmatrix} \mathrm{Im}(\alpha_N(z)) \\ \mathrm{Im}(z\tilde{\alpha}_N(z)) \end{bmatrix} + \frac{\mathrm{Im}(z)}{\sigma}\begin{bmatrix} v_{2,N}(z) \\ u_{2,N}(z) \end{bmatrix} + \frac{1}{N^2}\begin{bmatrix} \mathrm{Im}(\epsilon_N(z)) \\ \mathrm{Im}(z\tilde{\epsilon}_N(z)) \end{bmatrix} \tag{2.37}$$

with $\tilde{u}_{2,N} \triangleq \frac{\tilde{\mathbf{R}}_N(z)\mathbf{B}_N^*\mathbf{B}_N\tilde{\mathbf{R}}_N(z)^*}{|1+\sigma\tilde{\alpha}_N(z)|^2}$. From the identity $\frac{\tilde{\mathbf{R}}_N(z)\mathbf{B}_N^*}{1+\sigma\tilde{\alpha}_N(z)} = \frac{\mathbf{B}_N^*\mathbf{R}_N(z)}{1+\sigma\alpha_N(z)}$, we deduce $u_{2,N}(z) = \tilde{u}_{2,N}(z)$. The system (2.37) is thus equivalent to

$$\left(1 - u_{2,N}(z)\right)\mathrm{Im}(\alpha_N(z)) = v_{2,N}(z)\mathrm{Im}(z\tilde{\alpha}_N(z)) + \frac{\mathrm{Im}(z)}{\sigma}v_{2,N}(z) + \frac{1}{N^2}\mathrm{Im}(\epsilon_N(z)) \tag{2.38}$$

$$\left(1 - u_{2,N}(z)\right)\mathrm{Im}(z\tilde{\alpha}_N(z)) = |z|^2\tilde{v}_{2,N}(z)\mathrm{Im}(\alpha_N(z)) + \frac{\mathrm{Im}(z)}{\sigma}u_{2,N}(z) + \frac{1}{N^2}\mathrm{Im}(z\tilde{\epsilon}_N(z)) \tag{2.39}$$

Since $\alpha_N(z)$ and $\tilde{\alpha}_N(z)$ are Stieltjes transform of finite measure with support included in $\mathbb{R}^+$, and $|v_{2,N}(z)| \ge \frac{1}{c_N}\left|\frac{\sigma}{N}\mathrm{Tr}\,\mathbf{R}_N(z)\right|^2 > 0$ (similarly to $v_{1,N}(z)$ since $\frac{\sigma}{N}\mathrm{Tr}\,\mathbf{R}_N(z)$ is also the Stieltjes transform of a finite positive measure), we obtain from (2.38),

$$\left(1 - u_{2,N}(z)\right)\mathrm{Im}(\alpha_N(z)) > \frac{\mathrm{Im}(z)}{\sigma}v_{2,N}(z) - \frac{1}{N^2}|\epsilon_N(z)|. \tag{2.40}$$

We now prove that there exists $\eta > 0$ such that for all large N,

$$v_{2,N}(z) \ge \frac{\mathrm{Im}(z)^2\sigma^2 c_N}{64(\eta^2 + |z|^2)^2}. \tag{2.41}$$

We rely on theorem 2.4.1 and remark that the sequence of probability measures (μ_N) is tight. Therefore, it exists $\eta > 0$ such that $\inf_N \mu_N([0,\eta]) > \frac{1}{2}$. From (2.22) and the convergence (2.33), we deduce $\frac{1}{M}\mathrm{Tr}\,\mathbf{R}_N(z) - m_N(z) \to_N 0$ for all $z \in \mathbb{C}^+$. Since $\frac{1}{M}\mathrm{Tr}\,\mathbf{R}_N(z)$ is the Stieltjes transform of a probability measure ξ_N carried by $\mathbb{R}^+$, this implies $\xi_N - \mu_N \xrightarrow[N\to\infty]{w} 0$ and thus $\xi_N([0,\eta]) > \frac{1}{4}$ for all large N. Therefore,

$$\left|\frac{\sigma}{N}\mathrm{Tr}\,\mathbf{R}_N(z)\right| = \sigma c_N\left|\int_{\mathbb{R}^+}\frac{\mathrm{d}\xi_N(\lambda)}{\lambda - z}\right| \ge \sigma c_N \mathrm{Im}(z)\int_{\mathbb{R}^+}\frac{\mathrm{d}\xi_N(\lambda)}{|\lambda - z|^2} \ge \mathrm{Im}(z)\frac{\sigma c_N \xi_N([0,\eta])}{2(\eta^2 + |z|^2)},$$

which implies (2.41). Going back to (2.40), we get the following bound

$$\left(1-u_{2,N}(z)\right)\mathrm{Im}\left(\alpha_N(z)\right) > \frac{\mathrm{Im}(z)^3\sigma c_N}{64(\eta^2+|z|^2)^2} - \frac{1}{N^2}\mathrm{P}_1\left(|z|\right)\mathrm{P}_2\left(\frac{1}{\mathrm{Im}(z)}\right).$$

Define

$$\mathscr{E}_{1,N} = \left\{ z\in\mathbb{C}^+ : \frac{\mathrm{Im}(z)^3\sigma c_N}{64(\eta^2+|z|^2)^2} - \frac{1}{N^2}\mathrm{P}_1\left(|z|\right)\mathrm{P}_2\left(\frac{1}{\mathrm{Im}(z)}\right) > 0 \right\}.$$

Remark that this set can be written as

$$\mathscr{E}_{1,N} = \left\{ z\in\mathbb{C}^+ : 1 - \frac{1}{N^2}\mathrm{S}_1\left(|z|\right)\mathrm{S}_2\left(\frac{1}{\mathrm{Im}(z)}\right) > 0 \right\}.$$

with $\mathrm{S}_1, \mathrm{S}_2$ two polynomials with positive coefficients independent of N, z. Notice also that $1-u_{2,N}(z) > 0$ on $\mathscr{E}_{1,N}$, for all large N. Rewriting equation (2.38) as

$$\mathrm{Im}\left(z\tilde{\alpha}_N(z)\right) = \frac{1}{v_{2,N}(z)}\left(\left(1-u_{2,N}(z)\right)\mathrm{Im}(\alpha_N(z)) - \frac{\mathrm{Im}(z)}{\sigma} - \frac{\mathrm{Im}(\epsilon_N(z))}{N^2}\right)$$

and inserting into (2.39), we get

$$\Delta_{2,N}(z) = \frac{1}{\mathrm{Im}(\alpha_N(z))}\left(\frac{v_{2,N}(z)\mathrm{Im}(z)}{\sigma} + \left(1-u_{2,N}(z)\right)\frac{\mathrm{Im}(\epsilon_N(z))}{N^2} + v_{2,N}(z)\frac{\mathrm{Im}(z\tilde{\epsilon}_N(z))}{N^2}\right),$$

which implies the bound

$$\Delta_{2,N}(z) \geq \frac{v_{2,N}(z)}{\sigma}\frac{\mathrm{Im}(z)}{\mathrm{Im}(\alpha_N(z))} - \frac{1}{\mathrm{Im}(\alpha_N(z))}\left(\frac{|\epsilon_N(z)|}{N^2} + v_{2,N}(z)\frac{|z\tilde{\epsilon}_N(z)|}{N^2}\right)$$

for $z\in\mathscr{E}_{1,N}$ and for all large N. Using the bounds $\frac{\mathrm{Im}(z)}{\mathrm{Im}(\alpha_N(z))} \geq \frac{\mathrm{Im}(z)^2}{\sigma c_N}$, $v_{2,N}(z) \geq \frac{\mathrm{Im}(z)^2\sigma^2 c_N}{64(\eta^2+|z|^2)^2}$ and the fact that $\epsilon_N(z)$ and $\tilde{\epsilon}_N(z)$ are polynomially bounded ((2.22) and (2.20)), this lead us to the bound

$$\Delta_{2,N}(z) \geq \frac{\mathrm{Im}(z)^4\sigma c_N}{64(\eta^2+|z|^2)^2}\left(1 - \frac{1}{N^2}\tilde{\mathrm{S}}_1\left(|z|\right)\tilde{\mathrm{S}}_2\left(\frac{1}{\mathrm{Im}(z)}\right)\right).$$

for all large N, with $\tilde{\mathrm{S}}_1$, $\tilde{\mathrm{S}}_2$ some polynomials with positive coefficients independent of N, z. Define the set

$$\mathscr{E}_{2,N} = \left\{ z\in\mathbb{C}^+ : 1 - \frac{1}{N^2}\tilde{\mathrm{S}}_1\left(|z|\right)\tilde{\mathrm{S}}_2\left(\frac{1}{\mathrm{Im}(z)}\right) > \frac{1}{2} \right\}.$$

Define also the polynomials $\mathrm{Q}_i = \mathrm{S}_i + \sqrt{2}\tilde{\mathrm{S}}_i$ for $i = 1, 2$ and the set

$$\mathscr{E}_N = \left\{ z\in\mathbb{C}^+ : 1 - \frac{1}{N^2}\mathrm{Q}_1\left(|z|\right)\mathrm{Q}_2\left(\frac{1}{\mathrm{Im}(z)}\right) > 0 \right\}.$$

Therefore, one can easily check that $\mathscr{E}_N \subset \mathscr{E}_{1,N}\cap\mathscr{E}_{2,N}$, which concludes the proof. □

End of the proof of (2.17)

We finally use the above results to tackle the proof of the convergence (2.17). We recall that

$$\Delta_{1,N}(z) = \left(1-u_{1,N}(z)\right)^2 - z^2 v_{1,N}(z)\tilde{v}_{1,N}(z) \quad \text{and} \quad \Delta_{2,N}(z) = \left(1-u_{2,N}(z)\right)^2 - z^2 v_{2,N}(z)\tilde{v}_{2,N}(z),$$

by definition and that the bound (2.34) states that

$$|\Delta_N(z)| \geq \left(1 - \left|u_{1,N}(z)\right|^{1/2}\left|u_{2,N}(z)\right|^{1/2}\right)^2 - |z|^2\left|v_{1,N}(z)\right|^{1/2}\left|v_{2,N}(z)\right|^{1/2}\left|\tilde{v}_{1,N}(z)\right|^{1/2}\left|\tilde{v}_{2,N}(z)\right|^{1/2}. \tag{2.42}$$

It is easy to check that for nonnegative real numbers x_i, s_i, t_i $(i = 1, 2)$ with $x_i \leq 1$ and $(1-x_i)^2 - s_i t_i \geq 0$ for $i = 1, 2$, we have

$$(1-\sqrt{x_1 x_2})^2 - \sqrt{s_1 s_2 t_1 t_2} \geq \sqrt{(1-x_1)^2 - s_1 t_1}\sqrt{(1-x_2)^2 - s_2 t_2}.$$

From lemmas 2.7.2 and 2.7.3, for $z \in \mathcal{E}_N$, we have $u_{1,N}(z) < 1$, $u_{2,N}(z) < 1$, $\Delta_{1,N}(z) > 0$ and $\Delta_{2,N}(z) > 0$. Thus, from (2.34), we get $|\Delta_N(z)| \geq \sqrt{\Delta_{1,N}(z)}\sqrt{\Delta_{2,N}(z)}$ and consequently, using the polynomial bounds in lemmas 2.7.2 and 2.7.3, we immediately deduce that

$$|\Delta_N(z)|^{-1} \leq \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{\mathrm{Im}(z)}\right). \tag{2.43}$$

for all large N and $z \in \mathcal{E}_N$. By inverting the system (2.21) and using the bound $|u_N(z)| \leq \frac{\sigma^2 B_{\max}|z|^2}{|\mathrm{Im}(z)|^4}$, $|v_N(z)| \leq \frac{\sigma^2}{\mathrm{Im}(z)^2}$ and $|\tilde{v}_N(z)| \leq \frac{\sigma^2}{\mathrm{Im}(z)^2}$, we obtain

$$|\alpha_N(z) - \delta_N(z)| \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{\mathrm{Im}(z)}\right) \tag{2.44}$$

for $z \in \mathcal{E}_N$. On the other hand, for $z \in \mathbb{C}^+ \backslash \mathcal{E}_N$, we use the trick of Haagerup & Thorbjornsen [22] and we have $1 \leq \frac{1}{N^2}\mathrm{Q}_1(|z|)\,\mathrm{Q}_2\left(\frac{1}{\mathrm{Im}(z)}\right)$, and consequently

$$|\alpha_N(z) - \delta_N(z)| \leq \frac{2}{|\mathrm{Im}(z)|} \leq \frac{1}{N^2}\frac{2\sigma}{\mathrm{Im}(z)}\mathrm{Q}_1(|z|)\,\mathrm{Q}_2\left(\frac{1}{\mathrm{Im}(z)}\right).$$

This concludes the proof of (2.17).

2.7.3 Proof of theorem 2.2.3

The proof borrows several results from the proof of theorem 2.2.2, given in appendix 2.7.2. In this appendix, $\mathrm{P}_1, \mathrm{P}_2$ will be a generic notation for polynomials independent of N wit positive coefficients. Their value may change from one line to another.

We first begin with the equivalent of lemma 2.7.1.

Lemma 2.7.4. *For $z \in \mathbb{C}\backslash\mathbb{R}$, let $(\mathbf{M}_N(z))$ a sequence of deterministic matrix such that*

$$\|\mathbf{M}_N(z)\| \leq \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$

and $(\mathbf{d}_{1,N})$, $(\mathbf{d}_{2,N})$ two sequences of deterministic vectors such that $\sup_N \|\mathbf{d}_{1,N}\|, \sup_N \|\mathbf{d}_{2,N}\| < \infty$. Then,

$$\mathrm{Var}\left[\mathbf{d}_{1,N}^*\mathbf{Q}_N(z)\mathbf{M}_N(z)\mathbf{d}_{2,N}\right] \leq \frac{1}{N}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$
$$\mathrm{Var}\left[\mathbf{d}_{1,N}^*\boldsymbol{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{M}_N(z)\mathbf{d}_{2,N}\right] \leq \frac{1}{N}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

and

$$\mathrm{Var}\left[\left(\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N\right)^2\right] \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

Proof. The proof is based on straightforward computations similar to lemma 2.7.1, and is therefore omitted. □

We now prove that

$$\left|\mathbb{E}\left[\mathbf{d}_{1,N}^*\mathbf{Q}_N(z)\mathbf{d}_{2,N}\right] - \mathbf{d}_{1,N}^*\mathbf{T}_N(z)\mathbf{d}_{2,N}\right| \leq \frac{1}{N^{3/2}}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right), \tag{2.45}$$

Using equations (2.25) and (2.30) in appendix 2.7.2, we have for $z \in \mathbb{C}^+$,

$$\mathbb{E}\left[\mathbf{d}_{1,N}^*\mathbf{Q}_N(z)\mathbf{d}_{2,N}\right] = \mathbf{d}_{1,N}^*\mathbf{R}_N(z)\mathbf{d}_{2,N} + \mathbf{d}_{1,N}^*\boldsymbol{\Delta}_N(z)\mathbf{R}_N(z)\mathbf{d}_{2,N} + \mathbb{E}\left[\mathbf{d}_{1,N}^*\mathbf{Q}_N(z)\mathbf{R}_N(z)\mathbf{d}_{2,N}\right]\frac{\sigma^2}{N}\mathrm{Tr}\,\boldsymbol{\Delta}_N(z). \tag{2.46}$$

The arguments used in appendix 2.7.2 to bound the terms involving matrix $\boldsymbol{\Delta}_N(z)$ lead to

$$\left|\frac{1}{N}\mathrm{Tr}\,\boldsymbol{\Delta}_N(z)\right| \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$

and therefore the third term in the righthandside of (2.46) is also bounded by $N^{-2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)$. The same arguments also allow to handle the second term. Indeed,

$$\left|\mathbf{d}_{1,N}^*\mathbf{\Delta}_{1,N}(z)\mathbf{R}_N(z)\mathbf{d}_{2,N}\right| \leq \mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)\mathrm{Var}\left[\mathbf{d}_{1,N}^*\mathbf{Q}_N(z)\mathbf{R}_N(z)\mathbf{d}_{2,N}\right]^{1/2}\mathrm{Var}\left[\frac{1}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\right]^{1/2},$$

and using Cauchy-Schwarz inequality, lemmas 2.7.4 and 2.7.1, we obtain

$$\left|\mathbf{d}_{1,N}^*\mathbf{\Delta}_{1,N}(z)\mathbf{R}_N(z)\mathbf{d}_{2,N}\right| \leq \frac{1}{N^{3/2}}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

The terms $\mathbf{d}_{1,N}^*\mathbf{\Delta}_{i,N}(z)\mathbf{R}_N(z)\mathbf{d}_{2,N}$ for $i=2,3$ are bounded with similar arguments. Therefore,

$$\left|\mathbb{E}\left[\mathbf{d}_{1,N}^*\mathbf{Q}_N(z)\mathbf{d}_{2,N}\right] - \mathbf{d}_{1,N}^*\mathbf{R}_N(z)\mathbf{d}_{2,N}\right| \leq \frac{1}{N^{3/2}}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right). \tag{2.47}$$

We now handle the difference $\mathbf{d}_{1,N}^*\left(\mathbf{R}_N(z)-\mathbf{T}_N(z)\right)\mathbf{d}_{2,N}$. For this, we write the usual identity $\mathbf{R}_N(z)-\mathbf{T}_N(z) = \mathbf{R}_N(z)\left(\mathbf{T}_N(z)^{-1}-\mathbf{R}_N(z)^{-1}\right)\mathbf{T}_N(z)$ and obtain

$$\begin{aligned}&\mathbf{d}_{1,N}^*\left(\mathbf{R}_N(z)-\mathbf{T}_N(z)\right)\mathbf{d}_{2,N} = \\ &\quad\frac{\sigma(\alpha_N(z)-\delta_N(z))}{(1+\sigma\delta_N(z))(1+\sigma\alpha_N(z))}\mathbf{d}_{1,N}^*\mathbf{R}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)\mathbf{d}_{2,N} + \sigma z(\tilde{\alpha}_N(z)-\tilde{\delta}_N(z))\mathbf{d}_{1,N}^*\mathbf{R}_N(z)\mathbf{T}_N(z)\mathbf{d}_{2,N}.\end{aligned}$$

From theorem 2.2.2, it holds that

$$|\alpha_N(z)-\delta_N(z)| \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$

for all large N. Using $\tilde{\alpha}_N(z) = \alpha_N(z) - z^{-1}\sigma(1-c_N)$ and $\tilde{\delta}_N(z) = \delta_N(z) - z^{-1}\sigma(1-c_N)$, we also have

$$\left|\tilde{\alpha}_N(z)-\tilde{\delta}_N(z)\right| \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$

which implies

$$\left|\mathbf{d}_{1,N}^*\left(\mathbf{R}_N(z)-\mathbf{T}_N(z)\right)\mathbf{d}_{2,N}\right| \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

This concludes the proof of (2.45).

The last part of the proof is dedicated to prove that

$$\left|\mathbf{d}_{1,N}^*\mathbf{Q}_N(z)\mathbf{d}_{2,N} - \mathbf{d}_{1,N}^*\mathbf{T}_N(z)\mathbf{d}_{2,N}\right| \xrightarrow[N\to\infty]{\text{a.s.}} 0,$$

for all $z\in\mathbb{C}\backslash\mathbb{R}$. From (2.45), it is sufficient to prove that,

$$\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N} \xrightarrow[N\to\infty]{\text{a.s.}} 0,$$

for all $z\in\mathbb{C}\backslash\mathbb{R}$, or equivalently for all $z\in\mathbb{C}^+$ since $\mathbf{Q}_N(z^*) = \mathbf{Q}_N(z)^*$. Obviously, we have

$$\begin{aligned}\mathbb{E}\left|\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N}\right|^4 &= \left|\mathbb{E}\left[\left(\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N}\right)^2\right]\right|^2 + \mathrm{Var}\left[\left(\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N}\right)^2\right],\\ &\leq \mathrm{Var}\left[\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N}\right]^2 + \mathrm{Var}\left[\left(\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N}\right)^2\right].\end{aligned}$$

Using lemma 2.7.4 and Borel-Cantelli lemma, we deduce that for $z\in\mathbb{C}^+$, the convergence

$$\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N} \xrightarrow[N\to\infty]{} 0$$

holds on a event of probability one Ω_z, depending on z. We now prove that it exists an event of probability one Ω on which the previous convergence holds for all $z\in\mathbb{C}^+$. Let (z_k) a dense sequence in $\mathbb{C}^+$, and define $\Omega = \bigcup_k \Omega_{z_k}$. Fix a realization in Ω. For $z\in\mathbb{C}^+$, there exists a subsequence (z_{k_l}) such that $|z-z_{k_l}|\leq\frac{1}{l}$, and thus

$$\begin{aligned}&\left|\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N}\right| \\ &\quad\leq \left|\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z)-\mathbf{Q}_N(z_{k_l})\right)\mathbf{d}_{2,N}\right| + \left|\mathbf{d}_{1,N}^*\left(\mathbb{E}\left[\mathbf{Q}_N(z_{k_l})\right]-\mathbb{E}\left[\mathbf{Q}_N(z)\right]\right)\mathbf{d}_{2,N}\right| + \left|\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z_{k_l})-\mathbb{E}\left[\mathbf{Q}_N(z_{k_l})\right]\right)\mathbf{d}_{2,N}\right| \\ &\quad\leq \left|\mathbf{d}_{1,N}^*\left(\mathbf{Q}_N(z_{k_l})-\mathbb{E}\left[\mathbf{Q}_N(z_{k_l})\right]\right)\mathbf{d}_{2,N}\right| + \|\mathbf{d}_{1,N}\|\|\mathbf{d}_{2,N}\|\frac{|z-z_{k_l}|}{\mathrm{Im}(z)\mathrm{Im}(z_{k_l})}.\end{aligned}$$

Therefore,

$$\limsup_N \left| \mathbf{d}_{1,N}^* \left(\mathbf{Q}_N(z) - \mathbb{E}\left[\mathbf{Q}_N(z) \right] \right) \mathbf{d}_{2,N} \right| \leq \frac{1}{l} \frac{\sup_N \|\mathbf{d}_{1,N}\| \|\mathbf{d}_{2,N}\|}{\mathrm{Im}(z)\mathrm{Im}(z_{k_l})}, \tag{2.48}$$

which goes to 0 by taking the limit in l.

2.7.4 Proof of theorem 2.3.1: 0 does not belong to the support if $c_N < 1$

In order to establish that 0 does not belong to the support $\mathscr{S}_N$, we show that it exists $\epsilon > 0$ for which $\mu_N([0,x]) = 0$ for each $x \in]0,\epsilon[$. To show this, we will consider the function $h(m,z)$ defined as

$$h(m,z) = \frac{1}{M} \mathrm{Tr} \left(-z(1+\sigma^2 c_N m)\mathbf{I}_M + \sigma^2(1-c_N)\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma^2 c_N m} \right)^{-1}.$$

Observe that the equation $m = h(m,0)$ is equivalent to

$$m = \frac{1}{M} \mathrm{Tr} \left[\sigma^2(1-c_N)\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma^2 c_N m} \right]^{-1}.$$

Now, the condition $c_N < 1$ implies that the function $m \to \frac{h(m,0)}{m}$ is decreasing on $\mathbb{R}_+$. Therefore, the equation $m = h(m,0)$ has a unique strictly positive solution denoted m_0. Next, we will check that

$$1 - \left. \frac{\partial h}{\partial m} \right|_{(m_0,0)} > 0.$$

Indeed, observe that

$$\left. \frac{\partial h}{\partial m} \right|_{(m_0,0)} = \frac{\sigma^2 c_N}{1+\sigma^2 c_N m_0} \frac{1}{M} \mathrm{Tr} \left(\frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma^2 c_N m_0} \left(\sigma^2(1-c_N)\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma^2 c_N m_0} \right)^{-2} \right),$$

so that

$$\left. \frac{\partial h}{\partial m} \right|_{(m_0,0)} < \frac{\sigma^2 c_N}{1+\sigma^2 c_N m_0} \frac{1}{M} \mathrm{Tr} \left(\sigma^2(1-c_N)\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma^2 c_N m_0} \right)^{-1} = \frac{\sigma^2 c_N m_0}{1+\sigma^2 c_N m_0} < 1,$$

as required. Hence, the implicit function theorem implies that there exists an open disk centered at zero with radius $\eta > 0$ denoted $\mathscr{D}(0,\eta)$, and a unique function $\overline{m}(z)$, holomorphic on $\mathscr{D}(0,\eta)$, satisfying $\overline{m}(0) = m_0$ and such that

$$\overline{m}(z) = h(\overline{m}(z), z)$$

for $|z| < \eta$. Evaluating the successive derivatives of function $z \to h(\overline{m}(z), z)$ at the origin, one can check that for each $l \geq 0$, $\overline{m}^{(l)}(0)$ is real-valued. Since $m_0 > 0$, there exists a positive quantity ϵ, $0 < \epsilon \leq \eta$ such that $\overline{m}(x)$ is real-valued and $\overline{m}(x) > 0$ if $x \in]-\epsilon,\epsilon[$. On the other hand, it can be readily checked that if $x < 0$, the equation $m = h(m,x)$ has a unique strictly positive solution. Now, for $x < 0$, $m_N(x)$ is strictly positive, and satisfies this equation. Therefore, it holds that $m_N(x) = \overline{m}(x)$ for $-\epsilon < x < 0$. Since the two functions m_N and $\overline{m}$ are holomorphic on $\mathscr{D}(0,\epsilon) \setminus [0,\epsilon]$ and coincide on a set of values with an accumulation point, they must coincide on the whole domain of analicity, namely $\mathscr{D}(0,\epsilon) \setminus [0,\epsilon]$. We recall that for $0 \leq x < \epsilon$, $\mu_N([0,x])$ can be expressed as

$$\mu_N([0,x]) = \frac{1}{\pi} \lim_{y\to 0, y>0} \int_0^x \mathrm{Im}(m_N(s+iy))\mathrm{d}s.$$

Therefore,

$$\mu_N([0,x]) = \frac{1}{\pi} \lim_{y\to 0, y>0} \int_0^x \mathrm{Im}(\overline{m}(s+iy))\mathrm{d}s.$$

As $\overline{m}$ is holomorphic on $D(0,\epsilon)$, the dominated convergence theorem implies that

$$\frac{1}{\pi} \lim_{y\to 0, y>0} \int_0^x \mathrm{Im}(\overline{m}(s+iy))\mathrm{d}s = \frac{1}{\pi} \int_0^x \mathrm{Im}(\overline{m}(s))\mathrm{d}s = 0.$$

because $\overline{m}(s) \in \mathbb{R}$ if $s \in [0,x]$. This establishes that $\mu_N([0,x]) = 0$.

2.7.5 Proof of property 2.3.1: lower bound for $m_N(z)$

It is shown in [15] that $\mathrm{Re}(1+\sigma m_N(z)) > 0$. We rewrite here the proof and improve the bound. We use here the notations introduced in the proof of theorem 2.2.2 in appendix 2.7.2, as well as some results. In particular, doing similar computations as for the system of equations (2.35), it is straightforward to obtain the following 2×2 system

$$\begin{bmatrix}\mathrm{Re}(1+\sigma\delta_N(z))\\ \mathrm{Im}(1+\sigma\delta_N(z))\end{bmatrix} = \begin{bmatrix} u_{1,N}(z)-\mathrm{Re}(z)v_{1,N}(z) & \mathrm{Im}(z)v_{1,N}(z)\\ \mathrm{Im}(z)v_{1,N}(z) & u_{1,N}(z)+\mathrm{Re}(z)v_{1,N}(z)\end{bmatrix}\begin{bmatrix}\mathrm{Re}(1+\sigma\delta_N(z))\\ \mathrm{Im}(1+\sigma\delta_N(z))\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix}, \tag{2.49}$$

where $u_{1,N}$ and $v_{1,N}$ are defined in appendix 2.7.2 by

$$u_{1,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\frac{\mathbf{T}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)^*}{|1+\sigma\delta_N(z)|^2} \quad \text{and} \quad v_{1,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{T}_N(z)\mathbf{T}_N(z)^*.$$

The determinant of the above system is $(1-u_{1,N}(z))^2 - |z|^2 v_{1,N}(z)^2$. In appendix 2.7.2, we have defined

$$\Delta_{1,N}(z) = \left(1-u_{1,N}(z)\right)^2 - |z|^2 v_{1,N}(z)\tilde{v}_{1,N}(z),$$

with $\tilde{v}_{1,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\tilde{\mathbf{T}}_N(z)\tilde{\mathbf{T}}_N(z)^*$. Moreover, with the usual identity

$$z\mathbf{T}_N(z) - z^*\mathbf{T}_N(z)^* = z\mathbf{T}_N(z)\left(\frac{\mathbf{T}_N(z)^*}{z^*} - \frac{\mathbf{T}_N(z)}{z}\right)z^*\mathbf{T}_N(z)^*,$$

it follows that

$$\mathrm{Im}(z\delta_N(z)) = \mathrm{Im}\left(\tilde{\delta}_N(z)\right)|z|^2 v_{1,N}(z) + \mathrm{Im}\left(z\tilde{\delta}_N(z)\right)u_{1,N}(z) + \frac{\mathrm{Im}(z)}{\sigma}u_{1,N}(z).$$

But from the system (2.35) in appendix 2.7.2,

$$\mathrm{Im}\left(z\tilde{\delta}_N(z)\right) = \mathrm{Im}\left(\tilde{\delta}_N(z)\right)|z|^2 \tilde{v}_{1,N}(z) + \mathrm{Im}\left(z\tilde{\delta}_N(z)\right)u_{1,N}(z) + \frac{\mathrm{Im}(z)}{\sigma}u_{1,N}(z).$$

From the very definition of $\tilde{\delta}_N(z)$, we have $\mathrm{Im}(z\tilde{\delta}_N(z)) = \mathrm{Im}(z\delta_N(z))$ and the two previous equations give $\tilde{v}_{1,N}(z) = \frac{\mathrm{Im}(\tilde{\delta}_N(z))}{\mathrm{Im}(\delta_N(z))}v_{1,N}(z)$. Using the relation between $v_{1,N}(z)$ and $\tilde{v}_{1,N}(z)$, we get

$$\Delta_{1,N}(z) = \left(1-u_{1,N}(z)\right)^2 - |z|^2 v_{1,N}(z)\tilde{v}_{1,N}(z) = \left(1-u_{1,N}(z)\right)^2 - |z|^2 v_{1,N}(z)^2\left(1+\frac{\sigma(1-c_N)\mathrm{Im}(z)}{|z|^2\mathrm{Im}(\delta_N(z))}\right),$$

and from lemma 2.7.2 in appendix 2.7.2, we have $\Delta_{1,N}(z) > 0$ for $z\in\mathbb{C}^+$, which implies

$$(1-u_{1,N}(z))^2 - |z|^2 v_{1,N}(z)^2 > \frac{\sigma(1-c_N)\mathrm{Im}(z)v_{1,N}^2}{|z|^2\mathrm{Im}(\delta_N(z))},$$

for all $z\in\mathbb{C}^+$. Since $v_{1,N}, \mathrm{Im}(\delta_N(z)) > 0$ for $z\in\mathbb{C}^+$ (see appendix 2.7.2), we deduce

$$(1-u_{1,N}(z))^2 - |z|^2 v_{1,N}(z)^2 > 0.$$

Therefore, by inverting the above system (2.49), we obtain

$$\mathrm{Re}(1+\sigma\delta_N(z)) = \frac{1-u_{1,N}(z)-\mathrm{Re}(z)v_{1,N}(z)}{(1-u_{1,N}(z))^2-|z|^2v_{1,N}(z)^2} \geq \frac{1}{1-u_{1,N}(z)+|z|v_{1,N}(z)} \geq \frac{1}{2},$$

the last inequality following from $0 < |z|v_{1,N}(z) < 1-u_{1,N}(z) < 1$. The extension to $\mathbb{C}^-$ comes as usual from $\delta_N(z)^* = \delta_N(z^*)$ and the extension to the real axis is straightforward by the continuity of $m_N(z)$ when $z\to x\in\mathbb{R}$, described in section 2.3.

2.7.6 Proof of property 2.4.1: function w_N

In this appendix, we prove the different items stated in property 2.4.1.

- Item 1 is trivial from the properties of m_N stated in section 2.3.

- Item 2 is proved in [15, Th.3.2].
- Since $\mathrm{Re}(1+\sigma^2 c_N m_N(z)) > 0$ for all $z \in \mathbb{C}$ (see property 2.3.1) and $m_N(z)$ satisfies equation (2.4) for all $z \in \mathbb{C}\backslash\partial\mathscr{S}_N$, it is easy to see that

$$\frac{m_N(z)}{1+\sigma^2 c_N m_N(z)} = f_N(w_N(z)), \tag{2.50}$$

  or equivalently that $1-\sigma^2 c_N f_N(w_N(z)) = \frac{1}{1+\sigma^2 c_N m_N(z)}$. Plugging this equality into the expression of $w_N(z)$, we obtain $\phi_N(w_N(z)) = z$ which proves item 3.
- For item 4, by differentiating (2.50) on both sides, we obtain for $x \in \mathbb{R}\backslash\partial\mathscr{S}_N$,

$$w_N'(x) f_N'(w_N(x)) = \frac{m_N'(x)}{(1+\sigma^2 c_N m_N(x))^2},$$

  and $w_N'(x) > 0$ follows by noticing that $f_N'(w) > 0$ on $\mathbb{R}\backslash\{\lambda_{1,N},\dots,\lambda_{M,N}\}$, $m_N(x) \in \mathbb{R}$ and $m_N'(x) > 0$ (by differentiating the integral representation of m_N) for $x \in \mathbb{R}\backslash\mathscr{S}_N$. By taking derivatives with respect to $x \in \mathbb{R}\backslash\mathscr{S}_N$ on both sides of the equation $\phi_N(w_N(x)) = x$, we see that

$$w_N'(x)\phi_N'(w_N(x)) = 1.$$

  Thus, $\phi_N'(w_N(x)) > 0$. Finally, item 3 and property 2.3.1 implies $1-\sigma^2 c_N f_N(w_N(x)) > 0$, for $x \in \mathbb{R}\backslash\mathscr{S}_N$.
- We now handle the last item 5. Let $x \in \mathrm{Int}(\mathscr{S}_N)$. Using (2.50) and taking imaginary part on both sides, we obtain

$$\frac{\mathrm{Im}(m_N(x))}{|1+\sigma^2 c_N m_N(x)|^2} = \mathrm{Im}(w_N(x))\frac{1}{M}\mathrm{Tr}\,\left(\mathbf{B}_N\mathbf{B}_N^* - w_N(x)\mathbf{I}_M\right)^{-1}\left(\mathbf{B}_N\mathbf{B}_N^* - w_N(x)^*\mathbf{I}_M\right)^{-1},$$

  But,

$$\begin{aligned}\frac{1}{M}\mathrm{Tr}\,&\left(\mathbf{B}_N\mathbf{B}_N^* - w_N(x)\mathbf{I}_M\right)^{-1}\left(\mathbf{B}_N\mathbf{B}_N^* - w_N(x)^*\mathbf{I}_M\right)^{-1}\\ &\le \left\|\left(\mathbf{B}_N\mathbf{B}_N^* - w_N(x)\mathbf{I}_M\right)^{-1}\right\|^2 = \frac{1}{\min_{k=1,\dots,M}|\lambda_{k,N} - w_N(x)|^2}.\end{aligned}$$

  Equality (2.50) also shows that $w_N(x) \notin \{\lambda_{1,N},\dots,\lambda_{M,N}\}$ and thus $\mathrm{Im}(w_N(x)) > 0$ if $\mathrm{Im}(m_N(x)) > 0$. On the other hand, assume $\mathrm{Im}(w_N(x)) > 0$, then $m_N(x) \notin \mathbb{R}$, otherwise we would have $w_N(x) = x(1+\sigma^2 c_N m_N(x))^2 - \sigma^2(1-c_N)(1+\sigma^2 c_N m_N(x)) \in \mathbb{R}$.

2.7.7 Proof of property 2.4.2: increase of the local extrema of ϕ_N

We will prove a more general property, namely if w_1, w_2 are two critical points of ϕ_N (i.e $\phi_N'(w_{1,2}) = 0$) such that $1-\sigma^2 c_N f_N(w_{1,2}) > 0$, then $\left(\phi_N(w_1) - \phi_N(w_2)\right)/(w_2 - w_1)$ is always positive. We denote by ϕ_1 and ϕ_2 the quantities $\phi_N(w_1)$ and $\phi_N(w_2)$, and we define $h_n = 1-\sigma^2 c_N f_N(w_n)$ so that we can write $\phi_n = w_n h_n^2 + \sigma^2(1-c_N)h_n$, $n \in \{1,2\}$.

Using direct substraction of the expressions of ϕ_1 and ϕ_2 we can write

$$\frac{\phi_2-\phi_1}{w_2-w_1} = (h_1+h_2)\frac{(w_2h_2 - w_1h_1)}{w_2-w_1} + \sigma^2(1-c_N)\frac{h_2-h_1}{w_2-w_1} - h_1h_2.$$

Consider now the following inequality

$$\frac{2}{M}\sum_{k=1}^{M}\frac{\lambda_{k,N}}{(\lambda_{k,N}-w_1)(\lambda_{k,N}-w_2)} \le \frac{1}{M}\sum_{k=1}^{M}\frac{\lambda_{k,N}}{(\lambda_{k,N}-w_1)^2} + \frac{1}{M}\sum_{k=1}^{M}\frac{\lambda_{k,N}}{(\lambda_{k,N}-w_2)^2}, \tag{2.51}$$

which can be readily obtained by noting that

$$\frac{1}{M}\sum_{k=1}^{M}\left(\frac{\left(\lambda_{k,N}\right)^{1/2}}{(\lambda_{k,N}-w_1)} - \frac{\left(\lambda_{k,N}\right)^{1/2}}{(\lambda_{k,N}-w_2)}\right)^2 \ge 0. \tag{2.52}$$

Using the definition of h_1 and h_2 we can readily write

$$\frac{w_2 h_2 - w_1 h_1}{w_2 - w_1} = 1 - \frac{\sigma^2 c_N}{M} \sum_{k=1}^{M} \frac{\lambda_{k,N}}{(\lambda_{k,N} - w_1)(\lambda_{k,N} - w_2)},$$

and hence the inequality in (2.51) is giving us

$$\begin{aligned}\frac{\phi_2 - \phi_1}{w_2 - w_1} \geq (h_1 + h_2)\left[1 - \frac{\sigma^2 c_N}{2}\left(f_N(w_1) + f_N(w_2) + w_1 f_N'(w_1) + w_2 f_N'(w_2)\right)\right] \\ - h_1 h_2 + \sigma^2(1 - c_N)\frac{h_2 - h_1}{w_2 - w_1}, \end{aligned} \tag{2.53}$$

where $f_N'(w)$ denotes the derivative of $f_N(w)$. Using again the definition of h_1 and h_2, we can rewrite the last term of the previous expression as

$$\frac{h_2 - h_1}{w_2 - w_1} = -\frac{\sigma^2 c_N}{2}\left[f_N'(w_1) + f_N'(w_2) - \frac{1}{M}\sum_{k=1}^{M}\frac{(w_2 - w_1)^2}{(\lambda_{k,N} - w_1)^2(\lambda_{k,N} - w_2)^2}\right].$$

By inserting this last equality into (2.53) and replacing $f_N(w_1)$ with $\sigma^{-2}(1 - h_1)$, we obtain the expression

$$\begin{aligned}\frac{\phi_2 - \phi_1}{w_2 - w_1} \geq \frac{\sigma^4 c_N(1 - c_N)}{2}\frac{1}{M}\sum_{k=1}^{M}\frac{(w_2 - w_1)^2}{(\lambda_{k,N} - w_1)^2(\lambda_{k,N} - w_2)^2} + \frac{h_1^2 + h_2^2}{2} \\ - \frac{\sigma^4 c_N(1 - c_N)}{2}\left[f_N'(w_1) + f_N'(w_2)\right] - \frac{\sigma^2}{2}\frac{h_1 + h_2}{\left(w_1 f_N'(w_1) + w_2 f_N'(w_2)\right)}.\end{aligned} \tag{2.54}$$

Now, both w_1 and w_2 are critical points ϕ_N, so that for $n = 1, 2$, we have $\phi_N'(w_n) = h_n^2 - 2\sigma^2 w_n f_N'(w_n) h_n - \sigma^4(1 - c_N) f_N'(w_n) = 0$. Thus, we can write

$$\frac{h_1^2 + h_2^2}{2} = \sigma^2\left[w_1 h_1 f_N'(w_1) + w_2 h_2 f_N'(w_2)\right] + \frac{\sigma^4 c_N(1 - c_N)}{2}\left[f_N'(w_1) + f_N'(w_2)\right],$$

and by inserting the last equality into (2.54), we obtain

$$\frac{\phi_2 - \phi_1}{w_2 - w_1} \geq \frac{\sigma^4 c_N(1 - c_N)}{2}\frac{1}{M}\sum_{k=1}^{M}\frac{(w_2 - w_1)^2}{(\lambda_{k,N} - w_1)^2(\lambda_{k,N} - w_2)^2} + \frac{\sigma^2}{2}(h_1 - h_2)(w_1 f_N'(w_1) - w_2 f_N'(w_2)). \tag{2.55}$$

Using again the fact that $\phi_N'(w_n) = 0$, we can write $w_n f_N'(w_n) = \frac{h_n}{2\sigma^2} - \frac{\sigma^2(1 - c_N)}{2}\frac{f_N'(w_n)}{h_n}$ and thus (2.55) becomes

$$\begin{aligned}\frac{\phi_2 - \phi_1}{w_2 - w_1} \geq \frac{\sigma^4 c_N(1 - c_N)}{2}\frac{1}{M}\sum_{k=1}^{M}\frac{(w_2 - w_1)^2}{(\lambda_{k,N} - w_1)^2(\lambda_{k,N} - w_2)^2} + \frac{(h_1 - h_2)^2}{4} \\ - \frac{\sigma^4(1 - c_N)}{4}\left(f_N'(w_1) - f_N'(w_2)\right) + \frac{\sigma^4(1 - c_N)}{4}c_N\left[\frac{h_1}{h_2}f_N'(w_2) + \frac{h_2}{h_1}f_N'(w_1)\right].\end{aligned} \tag{2.56}$$

Clearly, we have

$$\frac{1}{M}\sum_{k=1}^{M}\frac{(w_2 - w_1)^2}{(\lambda_{k,N} - w_1)^2(\lambda_{k,N} - w_2)^2} - \left[f_N'(w_1) + f_N'(w_2)\right] = -\frac{2}{M}\sum_{k=1}^{M}\frac{1}{(\lambda_{k,N} - w_1)(\lambda_{k,N} - w_2)},$$

and thus by multiplying the previous equality with $h_1 h_2$ and adding $h_2^2 f_N'(w_1) + h_1^2 f_N'(w_2)$, we can also write

$$\begin{aligned}\frac{1}{M}\sum_{k=1}^{M}\left(\frac{h_2}{\lambda_{k,N} - w_1} - \frac{h_1}{\lambda_{k,N} - w_2}\right)^2 = \\ h_2^2 f_N'(w_1) + h_1^2 f_N'(w_2) + \frac{1}{M}\sum_{k=1}^{M}\frac{h_1 h_2(w_2 - w_1)^2}{(\lambda_{k,N} - w_1)^2(\lambda_{k,N} - w_2)^2} - h_1 h_2\left[f_N'(w_1) + f_N'(w_2)\right].\end{aligned}$$

The left hand side of the previous equality appears in (2.56) as a common factor on the last two terms of the right hand side of that equation. Hence, plugging it into (2.56), we obtain

$$\begin{aligned}\frac{\phi_2 - \phi_1}{w_2 - w_1} \geq \frac{\sigma^4 c_N(1 - c_N)}{4}\frac{1}{M}\sum_{k=1}^{M}\frac{(w_2 - w_1)^2}{(\lambda_{k,N} - w_1)^2(\lambda_{k,N} - w_2)^2} \\ + \frac{(h_1 - h_2)^2}{4} + \frac{\sigma^4 c_N(1 - c_N)}{4 h_1 h_2}\frac{1}{M}\sum_{k=1}^{M}\left(\frac{h_2}{\lambda_{k,N} - w_1} - \frac{h_1}{\lambda_{k,N} - w_2}\right)^2.\end{aligned}$$

Finally, noting that all the terms of the above equation are non-negative, we have established the desired property.

2.7.8 Proof of property 2.4.3: mass of the clusters

In this section, we evaluate the mass of any clusters by μ_N, i.e the quantity $\mu_N([x_{q,N}^-, x_{q,N}^+])$.

We first give some additional properties on function w_N. In particular, we show that it can be used as a proper integration contour. The idea is to use the curve $\{w_N(x) : x \in [x_{q,N}^-, x_{q,N}^+]\} \cup \{w_N(x)^* : x \in [x_{q,N}^-, x_{q,N}^+]\}$ to perform contour integration. The main difficulty is that $w_N'(x)$ becomes unbounded when $x \to x_{q,N}^-, x_{q,N}^+$ (see remark 2.4.4 in section 2.4). However, it holds that w_N' is still integrable in a neighborhood of $x_{q,N}^-, x_{q,N}^+$ as stated in the following property.

Property 2.7.1. *Let $q = 1, \ldots, Q$. Then, there exists a constant $C > 0$ and neighborhoods $\mathcal{V}(x_{q,N}^-)$, $\mathcal{V}(x_{q,N}^+)$ of respectively $x_{q,N}^-$ and $x_{q,N}^+$ such that,*

$$|w_N'(x+iy)| \leq C\left|x - x_{q,N}^-\right|^{-1/2} \quad \forall x+iy \in \mathcal{V}(x_{q,N}^-)\backslash\{x_{q,N}^-\}, \tag{2.57}$$

$$|w_N'(x+iy)| \leq C\left|x - x_{q,N}^+\right|^{-1/2} \quad \forall x+iy \in \mathcal{V}(x_{q,N}^+)\backslash\{x_{q,N}^+\}. \tag{2.58}$$

Proof. This property is a straightforward consequence of the results obtained in Dozier & Silverstein [15, Sec. 4]. We prove the result only for $x_{1,N}^-$, the arguments being similar in the other cases. Moreover, since $w_N'(z^*) = w_N'(z)^*$, we only consider $z \in \mathbb{C}^+ \cup \mathbb{R}$.

An elementary analysis of function ϕ_N (see property 2.4.2) shows that the equation $\phi_N(w) = x_{1,N}^-$ has at least two solutions in each interval $(0, \lambda_{M-K+1,N})$ and $(\lambda_{M-K+k,N}, \lambda_{M-K+k+1,N})$ for $k = 1, \ldots, K-1$, and one solution at point $w_{1,N}^- < 0$. Since the equation $\phi_N(w) = x_{1,N}^-$ is a polynomial equation with degree $2K+2$ (see remark 2.4.3), it remains one real solution. The solutions of $\phi_N(w) = x$ being continuous functions of x (see theorem 1.3.2 in chapter 1), we deduce that $w_{1,N}^-$ is necessarily a double solution, i.e $w_{1,N}^-$ is a zero of $\phi_N(w) - x_{1,N}^-$ with multiplicity 2. Since $w_{1,N}^-$ is a the preimage by ϕ_N of the local maximum $x_{1,N}^-$, it follows that $\phi''(w_{1,N}^-) < 0$.

Therefore, local inversion theorem implies the existence of a biholomorphism ψ_N (i.e. an holomorphic bijection with holomorphic inverse) defined on a neighborhood of $w_{1,N}^-$ to a neighborhood of 0 such that $\psi_N(w_{1,N}^-) = 0$ and

$$\phi_N(w) - x_{1,N}^- = \psi_N(w)^2.$$

Since $w_N(z) \to w_{1,N}^-$ when $z \in \mathbb{C}^+ \to x_{1,N}^-$, we obtain $z - x_{1,N}^- = \psi_N(w_N(z))^2$ for $z \in \mathbb{C}^+$ in a neighborhood of $x_{1,N}^-$. Consider the principal branch of the square root. Without loss of generality, we set $\sqrt{z - x_{1,N}^-} = \psi_N(w_N(z))$ for $z \notin \mathbb{C}^+$ in a neighborhood of $x_{1,N}^-$. Since ψ_N is invertible in a neighborhood of $w_{1,N}^-$, the last equality rewrites $w_N(z) = \psi_N^{-1}\left(\sqrt{z - x_{1,N}^-}\right)$ and by taking derivative, we obtain

$$w_N'(z) = \frac{1}{2\sqrt{z - x_{1,N}^-}\,\psi_N'\left(\psi_N^{-1}\left(\sqrt{z - x_{1,N}^-}\right)\right)}.$$

for all $z \notin \mathbb{C}^+$ in a neighborhood of $x_{1,N}^-$. Moreover, $\psi_N'(w_{1,N}^-) \neq 0$ since ψ_N is invertible in a neighborhood of $w_{1,N}^-$. Therefore, we obtain the following bound

$$\left|w_N'(z)\right| \leq \frac{C}{\sqrt{\left|z - x_{1,N}^-\right|}},$$

for all $z \notin \mathbb{C}^+$ in a neighborhood of $x_{1,N}^-$, with $C > 0$ a constant. For $z = x + iy$, we get

$$\left|w_N'(x+iy)\right| \leq \frac{C}{\sqrt{\left|x - x_{1,N}^-\right|}},$$

and since $\lim_{y \downarrow 0} w_N'(x+iy) \to w_N'(x)$ for $x \notin \partial\mathcal{S}_N$ (see remark 2.4.2 in section 2.4), the bound is also valid for $y = 0$ and $x \neq x_{1,N^-}$ in a neighborhood of x_{1,N^-}. This concludes the proof. □

We now use this property to perform contour integration by using w_N. We first define the set

$$\mathscr{C}_{q,N} \triangleq \left\{w_N(x) : x \in [x_{q,N}^-, x_{q,N}^+]\right\} \cup \left\{w_N(x)^* : x \in [x_{q,N}^-, x_{q,N}^+]\right\}.$$

Lemma 2.7.5. *$\mathscr{C}_{q,N}$ is a continuous closed path such that $\mathscr{C}_{q,N} \cap \mathbb{R} = \left\{w^-_{q,N}, w^+_{q,N}\right\}$. Moreover, the integral*

$$\oint_{\mathscr{C}^+_{q,N}} g(\lambda)\mathrm{d}\lambda \triangleq \int_{x^-_{q,N}}^{x^+_{q,N}} g(w_N(x))\, w'_N(x)\mathrm{d}x - \int_{x^-_{q,N}}^{x^+_{q,N}} g(w_N(x)^*)\, w'_N(x)^*\mathrm{d}x \tag{2.59}$$

is well defined for all functions g, continuous in a neighborhood of $\mathscr{C}_{q,N}$, where $\mathscr{C}^+_{q,N}$ means that $\mathscr{C}_{q,N}$ is counter-clockwise oriented. In particular,

$$\mathrm{Ind}_{\mathscr{C}_{q,N}}(\xi) \triangleq \frac{1}{2\pi i}\oint_{\mathscr{C}^+_{q,N}} \frac{\mathrm{d}\lambda}{\xi - \lambda} = \begin{cases} 1 & \text{if } \xi \in (w^-_{q,N}, w^+_{q,N}) \\ 0 & \text{if } \xi \notin [w^-_{q,N}, w^+_{q,N}]. \end{cases} \tag{2.60}$$

Proof. From the results of section 2.4, we know that

- w_N is continuous on $\mathbb{R}$,
- $w_N(x^-_{q,N}) = w^-_{q,N}$ and $w_N(x^+_{q,N}) = w^+_{q,N}$,
- $\mathrm{Im}(w_N(x)) > 0$ for $x \in (x^-_{q,N}, x^+_{q,N})$.

This implies that $\mathscr{C}_{q,N}$ is a continuous closed path enclosing $(w^-_{q,N}, w^+_{q,N})$. The integral (2.59) is well defined from property 2.7.1. Finally, one can easily check that basic properties concerning the winding number are still valid in the context of contour $\mathscr{C}_{q,N}$, which implies (2.60). □

Lemma 2.7.5 is basically pointing out the fact that w_N defines a valid parametrization of a contour that will not intersect any eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$. In particular, we can check that all the results concerning integrals over piece-wise continuously differentiable paths are still valid with $\mathscr{C}_{q,N}$, especially the residue theorem. If g is a continuous function in a neighborhood of $\mathscr{C}_{q,N}$ such that $g(\lambda^*) = g(\lambda)^*$, we notice that

$$\frac{1}{2\pi i}\oint_{\mathscr{C}^-_{q,N}} g(\lambda)\mathrm{d}\lambda = \frac{1}{\pi}\int_{x^-_{q,N}}^{x^+_{q,N}} \mathrm{Im}\left(g(w_N(x))\, w'_N(x)\right)\mathrm{d}x.$$

where $\mathscr{C}^-_{q,N}$ means that the contour is clockwise oriented.

With the previous results, we are now able to prove the result of property 2.4.3. From section 2.3, μ_N is absolutely continuous with density $\pi^{-1}\mathrm{Im}(m_N(x))$. Therefore, it holds that

$$\mu_N([x^-_{q,N}, x^+_{q,N}]) = \frac{1}{\pi}\mathrm{Im}\left(\int_{x^-_{q,N}}^{x^+_{q,N}} m_N(x)\mathrm{d}x\right). \tag{2.61}$$

In order to evaluate the righthandside of (2.61), we rely on lemma 2.7.5. From property 2.4.1, we deduce

$$m_N(x) = \frac{f_N(w_N(x))}{1 - \sigma^2 c_N f_N(w_N(x))} \quad \forall x \in \mathbb{R}\backslash\partial\mathscr{S}_N.$$

Moreover, from remark 2.4.2, we have $w'_N(x)\phi'_N(w_N(x)) = 1$ on $(x^-_{q,N}, x^+_{q,N})$. Consequently, we get

$$\mu_N([x^-_{q,N}, x^+_{q,N}]) = \frac{1}{\pi}\mathrm{Im}\left(\int_{x^-_{q,N}}^{x^+_{q,N}} g_N(w_N(x))\, w'_N(x)\mathrm{d}x\right), \tag{2.62}$$

where $g_N(\lambda)$ is the rational function defined by

$$g_N(\lambda) = \frac{f_N(\lambda)\phi'_N(\lambda)}{1 - \sigma^2 c_N f_N(\lambda)} = f_N(\lambda)\left(1 - \sigma^2 c_N f_N(\lambda) - 2\sigma^2 c_N w f'_N(\lambda) - \frac{\sigma^4 c_N(1-c_N) f'_N(\lambda)}{1 - \sigma^2 c_N f_N(\lambda)}\right).$$

In order to justify the existence of the integral at the righthandside of (2.62), we prove that $g_N(w)$ is continuous in a neighborhood of $\mathscr{C}_{q,N}$. This is a consequence of the properties of function w_N described in property 2.4.1 in section 2.4. We first note that the poles of $g_N(w)$ coincide with the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ and the real zeros $z^+_{0,N}, z^-_{1,N}, \ldots, z^-_{K,N}$ of $1 - \sigma^2 c_N f_N(w)$ (see property 2.4.2). As $w_N(x)$ is not real on $(x^-_{q,N}, x^+_{q,N})$, $x \to g_N(w_N(x))$ is continuous on $(x^-_{q,N}, x^+_{q,N})$. The continuity at $x^-_{q,N}$ and $x^+_{q,N}$ follows from $w^-_{q,N} = w_N(x^-_{q,N})$ and

$w^+_{q,N} = w_N(x^+_{q,N})$, which do not coincide with one the poles of $g_N(w)$ (see again property 2.4.2). Therefore, it is clear that $\mu_N([x^-_{q,N}, x^+_{q,N}])$ can also be written as

$$\mu_N([x^-_{q,N}, x^+_{q,N}]) = \frac{1}{2\pi i} \oint_{\mathscr{C}^-_{q,N}} g_N(\lambda) \mathrm{d}\lambda.$$

The integral can be evaluated using residue theorem and we give here the main steps of calculation. Since $\mathscr{C}_{q,N}$ only encloses $(w^-_{q,N}, w^+_{q,N})$ (lemma 2.7.5), we will have residues at the following points:

- for $q = 1$: residues at $z_{0,N}$, 0 and $z_{k,N}, \lambda_{M-K+k,N}$ for $k \in \mathscr{I}_1$,
- for $q \geq 2$: residues at $z_{k,N}, \lambda_{M-K+k,N}$ for $k \in \mathscr{I}_q$,

where $\mathscr{I}_q$ is defined in (2.14). We consider the decomposition $g_N(\lambda) = g_{1,N}(\lambda) + g_{2,N}(\lambda) + g_{3,N}(\lambda)$, with

$$\begin{aligned}
g_{1,N}(\lambda) &= f_N(\lambda)\left(1 - \sigma^2 c_N f_N(\lambda)\right), \\
g_{2,N}(\lambda) &= -2\sigma^2 c_N \lambda f_N(\lambda) f'_N(\lambda), \\
g_{3,N}(\lambda) &= -\sigma^4 c_N (1 - c_N) \frac{f_N(\lambda) f'_N(\lambda)}{1 - \sigma^2 c_N f_N(\lambda)}.
\end{aligned}$$

These three functions admit poles at $0, \left(\lambda_{M-K+k,N}\right)_{k=1,\ldots,K}$, and $g_{3,N}$ has moreover poles at $(z_{k,N})_{k=0,\ldots,K}$. After tedious but straightforward calculations, we finally find that for $k \in \{1, 2, \ldots, K\}$,

$$\begin{aligned}
\operatorname{Res}\left(g_{1,N}, \lambda_{M-K+k,N}\right) &= -\frac{1}{M} + \frac{2\sigma^2 c_N}{M^2} \sum_{l \neq k} \frac{1}{\lambda_{l,N} - \lambda_{k,N}}, \\
\operatorname{Res}\left(g_{2,N}, \lambda_{M-K+k,N}\right) &= -\frac{2\sigma^2 c_N}{M^2} \sum_{l \neq k} \frac{1}{\lambda_{l,N} - \lambda_{k,N}}, \\
\operatorname{Res}\left(g_{3,N}, \lambda_{M-K+k,N}\right) &= -\frac{1 - c_N}{c_N}.
\end{aligned}$$

For the residues at 0, we get

$$\begin{aligned}
\operatorname{Res}\left(g_{1,N}, 0\right) &= -\frac{M-K}{M} + 2\sigma^2 c_N \frac{M-K}{M} \frac{1}{M} \sum_{l=1}^{K} \frac{1}{\lambda_{l,N}}, \\
\operatorname{Res}\left(g_{2,N}, 0\right) &= -2\sigma^2 c_N \frac{M-K}{M} \frac{1}{M} \sum_{l=1}^{K} \frac{1}{\lambda_{l,N}}, \\
\operatorname{Res}\left(g_{3,N}, 0\right) &= -\frac{1 - c_N}{c_N}.
\end{aligned}$$

Finally, the residues at $z_{k,N}$ for $k = 0, \ldots, K$ are given by $\operatorname{Res}(g_{3,N}, z_{k,N}) = \frac{1-c_N}{c_N}$. Using these evaluations, we obtain immediately that if $q \geq 2$, then,

$$\begin{aligned}
\mu_N([x^-_{q,N}, x^+_{q,N}]) &= -\sum_{k \in \mathscr{I}_q} \left[\operatorname{Res}\left(g_{1,N}, \lambda_{M-K+k,N}\right) + \operatorname{Res}\left(g_{2,N}, \lambda_{M-K+k,N}\right) + \operatorname{Res}\left(g_{3,N}, \lambda_{M-K+k,N}\right) + \operatorname{Res}\left(g_{3,N}, z_{k,N}\right)\right] \\
&= \frac{|\mathscr{I}_q|}{M}.
\end{aligned}$$

If $q = 1$,

$$\begin{aligned}
\mu_N([x^-_{1,N}, x^+_{1,N}]) = &-\sum_{k \in \mathscr{I}_1} \left[\operatorname{Res}\left(g_{1,N}, \lambda_{M-K+k,N}\right) + \operatorname{Res}\left(g_{2,N}, \lambda_{M-K+k,N}\right) + \operatorname{Res}\left(g_{3,N}, \lambda_{M-K+k,N}\right) + \operatorname{Res}\left(g_{3,N}, z_{k,N}\right)\right] \\
&- \left[\operatorname{Res}\left(g_{1,N}, 0\right) + \operatorname{Res}\left(g_{2,N}, 0\right) + \operatorname{Res}\left(g_{3,N}, 0\right) + \operatorname{Res}\left(g_{3,N}, z_{0,N},\right)\right] \\
= &\frac{|\mathscr{I}_1|}{M} + \frac{M-K}{M},
\end{aligned}$$

which concludes the proof.

2.7.9 Proof theorem 2.5.1: support in the spiked model case

Preliminary results on perturbed equations

We first state two useful lemmas related to the solutions of perturbed equations. They can be interpreted as extensions of lemmas 3.2 and 3.3 of [7]. In the following, we denote respectively by $\mathscr{D}_o(z,r)$, $\mathscr{D}_c(z,r)$ and $\mathscr{C}(z,r)$ the open disk, closed disk and circle of radius $r>0$ with center z. Moreover, in this paragraph, the notation $o(1)$ denotes a term that converges towards 0 when the variable ϵ converges towards 0. The first result is a straightforward modification of [7, lemma 3.2]. Its proof is thus omitted.

Lemma 2.7.6. *For each $\epsilon>0$, we consider $h_\epsilon(z)=h(z)+\chi_\epsilon(z)$ with h,χ_ϵ two holomorphic functions in a disk $\mathscr{D}_o(z_0,r_0)$. We assume that $\sup_{z\in\mathscr{D}_o(z_0,r_0)}|\chi_\epsilon(z)|=o(1)$. We consider $z_{0,\epsilon}=z_0+\delta_\epsilon$ with $\delta_\epsilon=o(1)$. Then, $\exists\,\epsilon_0>0$ and $r>0$ such that for each $0<\epsilon\leq\epsilon_0$, $z_{0,\epsilon}\in\mathscr{D}_o(z_0,r)$ and the equation*

$$z-z_{0,\epsilon}-\epsilon h_\epsilon(z)=0,$$

admits a unique solution in $\mathscr{D}_o(z_0,r)$ given by

$$z_\epsilon=z_{0,\epsilon}+\epsilon h(z_0)+o(\epsilon).$$

Moreover, if we assume that $z_0\in\mathbb{R}$, $h(z)\in\mathbb{R}$ for $z\in\mathbb{R}$, and that for ϵ small enough, $z_{0,\epsilon}\in\mathbb{R}$, $h_\epsilon(z)\in\mathbb{R}$ for $z\in\mathbb{R}$, then $z_\epsilon\in\mathbb{R}$.

The second result is an extension of [7, Lem.3.3] to certain third degree equations. The proof is given at the end of the section.

Lemma 2.7.7. *For each $\epsilon>0$ and $i=1,2$, we consider $h_{i,\epsilon}(z)=h_i(z)+\chi_{i,\epsilon}(z)$ with $h_i,\chi_{i,\epsilon}$ holomorphic functions in a disk $\mathscr{D}_o(z_0,r_0)$. We assume that $h_1(z_0)\neq 0$ and that $\sup_{z\in\mathscr{D}_o(z_0,r_0)}|\chi_{i,\epsilon}(z)|=o(1)$ for $i=1,2$. We consider $z_{0,\epsilon}=z_0+\delta_\epsilon$ with $\delta_\epsilon=o(1)$. Then, $\exists\,\epsilon_0>0$ and $r>0$ such that $z_{0,\epsilon}\subset\mathscr{D}_o(z_0,r)$ $\forall\epsilon\in(0,\epsilon_0)$ and the equation*

$$\left(z-z_{0,\epsilon}\right)^3-\epsilon\left(z-z_{0,\epsilon}\right)h_{1,\epsilon}(z)+\epsilon^2h_{2,\epsilon}(z)=0$$

has 3 solutions in $\mathscr{D}_o(z_0,r)$ given by

$$z_\epsilon^-=z_{0,\epsilon}-\sqrt{\epsilon}\sqrt{h_1(z_0)}+o(\sqrt{\epsilon}),\qquad z_\epsilon^+=z_{0,\epsilon}+\sqrt{\epsilon}\sqrt{h_1(z_0)}+o(\sqrt{\epsilon})$$

and

$$z_\epsilon=z_{0,\epsilon}+\epsilon\frac{h_2(z_0)}{h_1(z_0)}+o(\epsilon),$$

where $\sqrt{.}$ is an arbitrary branch of the square root, analytic in a neighborhood of $h_1(z_0)$. Moreover, if we assume that

- *$z_0\in\mathbb{R}$, $h_i(z)\in\mathbb{R}$ for $z\in\mathbb{R}$,*
- *for ϵ small enough, $z_{0,\epsilon}\in\mathbb{R}$, $h_{i,\epsilon}(z)\in\mathbb{R}$ for $z\in\mathbb{R}$,*

then z_ϵ is real. Moreover, if $h_1(z_0)>0$ then z_ϵ^-, z_ϵ^+ and z_ϵ are real while z_ϵ^-, z_ϵ^+ are non real if $h_1(z_0)<0$.

Study of the support

We now identify the clusters of the support $\mathscr{S}_N$, and evaluate the points $x_{q,N}^-,x_{q,N}^+$ for $q=1,\ldots,Q$. From theorem 2.4.1 in section 2.4, these points coincide with the positive extrema of function ϕ_N (defined in (2.10) section 2.4). Therefore, we first evaluate the real zeros of $\phi_N'(w)=(1-\sigma^2c_Nf_N(w))^2-2\sigma^2c_Nwf_N'(w)(1-\sigma^2c_Nf_N(w))-\sigma^4c_N(1-c_N)f_N'(w)$. Straightforward calculations give

$$\phi_N'(w)=\frac{1}{w^2\prod_{k=1}^K(\lambda_{k,N}-w)^3}\left[\chi_{1,N}(w)+\frac{1}{M}\chi_{2,N}(w)+\frac{1}{M^2}\chi_{3,N}(w)\right],$$

with

$$\chi_{1,N}(w) = (w^2 - \sigma^4 c_N) \prod_{k=1}^{K} (\lambda_{M-K+k,N} - w)^3,$$

$$\chi_{2,N}(w) = -2\sigma^2 c_N \prod_{k=1}^{K} (\lambda_{M-K+k,N} - w) \sum_{j=1}^{K} \left[\lambda_{j,N} \left(w^2 + \sigma^2(1+c_N)w - \frac{\sigma^2(1+c_N)\lambda_{M-K+j,N}}{2} \right) \prod_{\substack{l=1 \\ l\neq j}}^{K} (\lambda_{M-K+l,N} - w)^2 \right],$$

$$\chi_{3,N}(w) = \sigma^4 c_N^2 \left(\sum_{k=1}^{K} \lambda_{M-K+k,N} \prod_{\substack{l=1 \\ l\neq k}}^{K} (\lambda_{M-K+l,N} - w) \right) \left(\sum_{k=1}^{K} \lambda_{M-K+k,N} (3w - \lambda_{M-K+k,N}) \prod_{\substack{l=1 \\ l\neq k}}^{K} (\lambda_{M-K+l,N} - w)^2 \right).$$

Therefore, $\phi_N^{'}(w) = 0$ if and only if

$$\chi_{1,N}(w) + \frac{1}{M}\chi_{2,N}(w) + \frac{1}{M^2}\chi_{3,N}(w) = 0. \tag{2.63}$$

Note that $c_N < 1$ for N large enough since $c < 1$. Observe that the zeros of ϕ_N' are included into a compact interval $\mathscr{I}$ independent of N (see the proof of property 2.3.1 section 2.4). Next, we claim that for each $\alpha > 0$, it exists $\beta > 0$ such that

$$\left| \chi_{1,N}(w) + \frac{1}{M}\chi_{2,N}(w) + \frac{1}{M^2}\chi_{3,N}(w) \right| > \beta,$$

for N large enough, if $|w - \sigma^2\sqrt{c}| > \alpha$, $|w + \sigma^2\sqrt{c}| > \alpha$, $|w - \gamma_k| > \alpha$, for $k = 1,\ldots,K$ and $w \in \mathscr{I}$. This follows immediately from the inequality

$$\left| \chi_{1,N}(w) + \frac{1}{M}\chi_{2,N}(w) + \frac{1}{M^2}\chi_{3,N}(w) \right| \geq |\chi_{1,N}(w)| - \frac{1}{M}\chi_{2,max} - \frac{1}{M^2}\chi_{3,max},$$

where $\chi_{i,max} = \max_{w\in\mathscr{I}} |\chi_{i,N}(w)|$ for $i = 2,3$. This shows that the solutions of eq. (2.63) are located around the points $\sigma^2\sqrt{c}, -\sigma^2\sqrt{c}, \gamma_k, k = 1,\ldots,K$.

In a disk $\mathscr{D}_o(\sigma^2\sqrt{c}, r)$, (2.63) is equivalent to

$$w - \sigma^2\sqrt{c_N} + \frac{1}{M}\frac{w - \sigma^2\sqrt{c_N}}{\chi_{1,N}(w)} \left(\chi_{2,N}(w) + \frac{1}{M}\chi_{3,N}(w) \right) = 0. \tag{2.64}$$

We use lemma 2.7.6 with $\epsilon = M^{-1}$, $z_0 = \sigma^2\sqrt{c}$, $z_{0,\epsilon} = \sigma^2\sqrt{c_N}$, and the functions

$$h_\epsilon(w) = -\frac{(w - \sigma^2\sqrt{c_N})}{\chi_{1,N}(w)} \left[\chi_{2,N}(w) + \frac{1}{M}\chi_{3,N}(w) \right]$$

and $h(w) = \lim_{M\to+\infty} h_\epsilon(w)$. $h(w)$ is obtained by replacing c_N and the $(\lambda_{M-K+k,N})_{k=1,\ldots,K}$ by c and the $(\gamma_k)_{k=1,\ldots,K}$ in the expression of h_ϵ. Lemma 2.7.6 implies that it exists r for which equation (2.64), or equivalently equation (2.63), has a unique solution in $\mathscr{D}_o(\sigma^2\sqrt{c}, r)$ for M large enough. This solution is given by $\sigma^2\sqrt{c_N} + \mathscr{O}(\frac{1}{M})$. It is easy to check that

$$\phi_N\left(\sigma^2\sqrt{c_N} + \mathscr{O}\left(\frac{1}{M}\right) \right) = \sigma^2(1 + \sqrt{c_N})^2 + \mathscr{O}\left(\frac{1}{M}\right).$$

This quantity is positive, and it is easily seen that ϕ_N' has a change of sign in $\mathscr{D}_o(\sigma^2\sqrt{c}, r)$ for M large enough, thus showing that $\sigma^2\sqrt{c_N} + \mathscr{O}(M^{-1})$ is the pre-image of a positive extremum of ϕ_N. Exchanging $\sigma^2\sqrt{c}$ with $-\sigma^2\sqrt{c}$, we obtain similarly that it exists a neighborhood of $-\sigma^2\sqrt{c}$ in which equation (2.63) has a unique solution given by $-\sigma^2\sqrt{c_N} + \mathscr{O}(\frac{1}{M})$. Moreover,

$$\phi_N\left(-\sigma^2\sqrt{c_N} + \mathscr{O}\left(\frac{1}{M}\right) \right) = \sigma^2(1 - \sqrt{c_N})^2 + \mathscr{O}\left(\frac{1}{M}\right),$$

so that $-\sigma^2\sqrt{c_N} + \mathscr{O}(\frac{1}{M})$ is also the pre-image of a positive extremum of ϕ_N.

We now consider $i \in \{1,\dots,K\}$, and study the equation (2.63) in a neighborhood $\mathscr{D}_o(\gamma_i, r)$ of γ_i. In order to use lemma 2.7.7, we put $\epsilon = \frac{1}{M}$, $z_0 = \gamma_i$, $z_{0,\epsilon} = \lambda_{M-K+i,N}$. It is easily seen that in $\mathscr{D}_o(\gamma_i, r)$, eq. (2.63) is equivalent to

$$(w - \lambda_{M-K+i,N})^3 - \frac{1}{M}(w - \lambda_{M-K+i,N})h_{1,\epsilon}(w) + \frac{1}{M^2}h_{2,\epsilon}(w) = 0,$$

where

$$h_{1,\epsilon}(w) = \frac{2\sigma^2 c_N \sum_{k=1}^K \left[\lambda_{M-K+k,N}\left(w^2 + \sigma^2(1+c_N)w - \frac{\sigma^2(1+c_N)\lambda_{M-K+k,N}}{2}\right)\prod_{\substack{l=1\\ l\neq k}}^K (\lambda_{M-K+l,N} - w)^2\right]}{(w^2 - \sigma^4 c_N)\prod_{\substack{k=1\\ k\neq i}}^K (\lambda_{M-K+k,N} - w)^2},$$

$$h_{2,\epsilon}(w) = -\frac{\chi_{3,N}(w)}{(w^2 - \sigma^4 c_N)\prod_{k\neq i}^K (\lambda_{M-K+k,N} - w)^3}.$$

We denote by $h_1(w)$ and $h_2(w)$ the limits of $h_{1,\epsilon}(w)$ and $h_{2,\epsilon}(w)$ when $\epsilon \to 0$, i.e. the functions obtained by replacing c_N and the $(\lambda_{M-K+k,N})_{k=1,\dots,K}$ by c and the $(\gamma_k)_{k=1,\dots,K}$ respectively in the expressions of $h_{1,\epsilon}, h_{2,\epsilon}$. After some algebra, we obtain that

$$h_1(\gamma_i) = \frac{2\sigma^2 c\gamma_i^2(\gamma_i + \frac{\sigma^2(1+c)}{2})}{\gamma_i^2 - \sigma^4 c},$$

while $h_2(\gamma_i)$ is equal to

$$h_2(\gamma_i) = -\frac{2\sigma^4 c^2 \gamma_i^3}{\gamma_i^2 - \sigma^4 c}.$$

Lemma 2.7.7 implies that it exists r such that

$$\lambda_{M-K+i,N} - \frac{1}{M}\frac{\sigma^2 c\gamma_i}{\gamma_i + \sigma^2\frac{1+c}{2}} + o\left(\frac{1}{M}\right) \tag{2.65}$$

is solution of (2.63) contained in $\mathscr{D}_o(\gamma_i, r)$. It is however easy to check that

$$\phi_N\left(\lambda_{M-K+i,N} - \frac{1}{M}\frac{\sigma^2 c\gamma_i}{\gamma_i + \sigma^2\frac{1+c}{2}} + o\left(\frac{1}{M}\right)\right) = -\frac{\sigma^4(1-c)^2}{2\gamma_i}\left(1 - \frac{c}{2}\right) < 0.$$

Therefore, (2.65) cannot be one the points $w_{q,N}^-, w_{q,N}^+$. Moreover, if $\gamma_i < \sigma^2\sqrt{c}$, then $h_1(\gamma_i) < 0$ and (2.63) has no extra real solution in $\mathscr{D}_o(\gamma_i, r)$. If $\gamma_i > \sigma^2\sqrt{c}$, then $h_1(\gamma_i) > 0$, and the quantities

$$\lambda_{M-K+i,N} - \frac{1}{\sqrt{M}}\sqrt{h_1(\gamma_i)} + o\left(\frac{1}{\sqrt{M}}\right) \quad \text{and} \quad \lambda_{M-K+i,N} + \frac{1}{\sqrt{M}}\sqrt{h_1(\gamma_i)} + o\left(\frac{1}{\sqrt{M}}\right) \tag{2.66}$$

are the 2 other real solutions of (2.63) contained in $\mathscr{D}_o(\gamma_i, r)$. After some algebra, we get that

$$\begin{aligned}
&\phi_N\left(\lambda_{M-K+i,N} - \frac{1}{\sqrt{M}}\sqrt{h_1(\gamma_i)} + o\left(\frac{1}{\sqrt{M}}\right)\right) = \\
&\qquad \frac{(\lambda_{M-K+i,N} + \sigma^2 c_N)(\lambda_{M-K+i,N} + \sigma^2)}{\lambda_{M-K+i,N}} - \frac{1}{\sqrt{M}}\frac{2\sqrt{h_1(\gamma_i)}(\gamma_i^2 - \sigma^4 c)}{\gamma_i^2} + o\left(\frac{1}{\sqrt{M}}\right), \\
&\phi_N\left(\lambda_{M-K+i,N} + \frac{1}{\sqrt{M}}\sqrt{h_1(\gamma_i)} + o\left(\frac{1}{\sqrt{M}}\right)\right) = \\
&\qquad \frac{(\lambda_{M-K+i,N} + \sigma^2 c_N)(\lambda_{M-K+i,N} + \sigma^2)}{\lambda_{M-K+i,N}} + \frac{1}{\sqrt{M}}\frac{2\sqrt{h_1(\gamma_i)}(\gamma_i^2 - \sigma^4 c)}{\gamma_i^2} + o\left(\frac{1}{\sqrt{M}}\right),
\end{aligned}$$

are both positive. It is easy to check that if $k = K - K_s + 1, \dots, K$, then, $\sigma^2\sqrt{c_N} < \lambda_{M-K+k,N}$ for N large enough. By noticing that ϕ_N' changes sign around the points (2.66), one can deduce as above that the critical points (2.66) are necessarily local extrema. Thus, $\mathscr{S}_N$ has K_s+1 clusters, and for $k=K-K_s+1,\dots,K$, $q=k-K+K_s+1$

$$
\begin{aligned}
x_{1,N}^- &= \sigma^2(1-\sqrt{c_N})^2+\mathcal{O}\left(\frac{1}{M}\right),\\
x_{1,N}^+ &= \sigma^2(1+\sqrt{c_N})^2+\mathcal{O}\left(\frac{1}{M}\right),\\
x_{q,N}^- &= \frac{(\lambda_{M-K+k,N}+\sigma^2c_N)(\lambda_{M-K+k,N}+\sigma^2)}{\lambda_{M-K+k,N}}-\frac{1}{\sqrt{M}}\frac{2\sqrt{h_1(\gamma_k)}(\gamma_k^2-\sigma^4c)}{\gamma_k^2}+o\left(\frac{1}{\sqrt{M}}\right),\\
x_{q,N}^+ &= \frac{(\lambda_{M-K+k,N}+\sigma^2c_N)(\lambda_{M-K+k,N}+\sigma^2)}{\lambda_{M-K+k,N}}+\frac{1}{\sqrt{M}}\frac{2\sqrt{h_1(\gamma_k)}(\gamma_k^2-\sigma^4c)}{\gamma_k^2}+o\left(\frac{1}{\sqrt{M}}\right),
\end{aligned}
$$

which proves theorem 2.5.1.

Proof of lemma 2.7.7

We begin by choosing $r>0$ and $\epsilon_1>0$ such that $r<r_0$, $z_{0,\epsilon}\in\mathscr{D}_c(z_0,r)$ and $\mathscr{D}_c(z_{0,\epsilon},r)\subset\mathscr{D}_0(z_0,r_0)$, for each $0<\epsilon<\epsilon_1$. Let $f_\epsilon(z)=(z-z_{0,\epsilon})^3-\epsilon(z-z_{0,\epsilon})h_{1,\epsilon}(z)+\epsilon^2h_{2,\epsilon}(z)$ and $g_\epsilon(z)=(z-z_{0,\epsilon})^3$. Moreover, define $\frac{K_i}{2}=\sup_{\mathscr{D}_c(z_0,r)}|h_i(z)|$ (for $i=1,2$).

As $\sup_{z\in\mathscr{D}_o(z_0,r_0)}|\chi_{i,\epsilon}(z)|=o(1)$, it exists $\epsilon_2\le\epsilon_1$ such that $\sup_{\mathscr{D}_c(z_0,r)}|h_{i,\epsilon}(z)|\le K_i$ (for $i=1,2$) for each $\epsilon\le\epsilon_2$. For $z\in\mathscr{D}_c(z_0,r)$, it holds that

$$\left|f_\epsilon(z)-g_\epsilon(z)\right|\le\epsilon\left|z-z_{0,\epsilon}\right|\left|h_{1,\epsilon}(z)\right|+\epsilon^2\left|h_{2,\epsilon}(z)\right|.$$

As $z_{0,\epsilon}-z_0=o(1)$, it exists $\epsilon_3\le\epsilon_2$ such that, for each $\epsilon\le\epsilon_3$, $|z-z_{0,\epsilon}|<2r$ on $\mathscr{D}_c(z_0,r)$. Hence, for each $\epsilon\le\epsilon_3$, it holds that $\left|f_\epsilon(z)-g_\epsilon(z)\right|\le 2\epsilon rK_1+\epsilon^2K_2$ on $\mathscr{D}_c(z_0,r)$. We now restrict z to $\mathscr{C}(z_0,r)$, the boundary of $\mathscr{D}_c(z_0,r)$. It exists $\epsilon_4\le\epsilon_3$ for which $2\epsilon rK_1+\epsilon^2K_2<\frac{r^3}{2}<r^3=|z-z_0|^3$ holds on $\mathscr{C}(z_0,r)$ for each $\epsilon\le\epsilon_4$. Therefore, $\forall z\in\mathscr{C}(z_0,r)$, we have $|f_\epsilon(z)-g_\epsilon(z)|<|g_\epsilon(z)|$ for $\epsilon\le\epsilon_4$. It follows from Rouché's theorem that these values of ϵ, then f_ϵ and g_ϵ have the same number of zeros inside $\mathscr{D}_o(z_0,r)$. Thus, for $\epsilon\le\epsilon_4$, the equation

$$(z-z_{0,\epsilon})^3-\epsilon(z-z_{0,\epsilon})h_{1,\epsilon}(z)+\epsilon^2h_{2,\epsilon}(z)=0 \tag{2.67}$$

has three solutions in $\mathscr{D}_o(z_0,r)$. Using the the same procedure to functions $f_\epsilon(z)=(z-z_{0,\epsilon})^2-\epsilon h_{1,\epsilon}(z)$ and $g_\epsilon(z)=(z-z_{0,\epsilon})^2$, we deduce that if $\epsilon\le\epsilon_5\le\epsilon_4$, the equation

$$(z-z_{0,\epsilon})^2-\epsilon h_{1,\epsilon}(z)=0 \tag{2.68}$$

has two solutions $\hat{z}_\epsilon^-,\hat{z}_\epsilon^+$ in $\mathscr{D}_o(z_0,r)$. We clearly have $|z_{0,\epsilon}-\hat{z}_\epsilon^-|=\mathcal{O}(\epsilon^{1/2})$ and $|z_0-\hat{z}_\epsilon^-|=o(1)$. Therefore, $h_{1,\epsilon}(\hat{z}_\epsilon^-)-h_1(z_0)=o(1)$. As $h_1(z_0)\neq0$, it exists $\epsilon_6\le\epsilon_5$ and a neighborhood of $h_1(z_0)$, containing $h_{1,\epsilon}(\hat{z}_\epsilon^-),h_{1,\epsilon}(z_0)$ for each $\epsilon\le\epsilon_6$, in which a suitable branch of the square-root $\sqrt{.}$ is analytic. We assume that solution $\hat{z}_\epsilon^-$ is given by $z_{0,\epsilon}-\hat{z}_\epsilon^-=-\sqrt{\epsilon}\sqrt{h_{1,\epsilon}(\hat{z}_\epsilon^-)}$. As $|h_1(z_0)-h_{1,\epsilon}(\hat{z}_\epsilon^-)|=o(1)$, we have $z_{0,\epsilon}-\hat{z}_\epsilon^-=-\sqrt{\epsilon}\sqrt{h_1(z_0)}+o(\sqrt{\epsilon})$. We obtain similarly that $z_{0,\epsilon}-\hat{z}_\epsilon^+=\sqrt{\epsilon}\sqrt{h_1(z_0)}+o(\sqrt{\epsilon})$.

Considering again $\hat{z}_\epsilon^-$, it follows that it exists $\epsilon_7\le\epsilon_6$ such that for each $\epsilon\le\epsilon_7$, it holds that

$$\left|z_{0,\epsilon}-\hat{z}_\epsilon^-\right|>\frac{\sqrt{\epsilon}\sqrt{h_1(z_0)}}{2}>\sqrt{\epsilon}\sqrt{r'}, \tag{2.69}$$

with $r'<\frac{|h_1(z_0)|}{4}$. For $\epsilon\le\epsilon_8\le\epsilon_7$, we have $\sqrt{\epsilon r'}<r$ and for $z\in\mathscr{D}_c(z_{0,\epsilon},\sqrt{\epsilon r'})$, we get

$$\left|(z-z_{0,\epsilon})^2-\epsilon h_{1,\epsilon}(z)\right|>\epsilon|h_{1,\epsilon}(z)|-|z-z_{0,\epsilon}|^2>\epsilon\left(|h_{1,\epsilon}(z)|-r'\right).$$

It is easy to check that for each $\epsilon\le\epsilon_9\le\epsilon_8$, then $|h_{1,\epsilon}(z)|>\frac{|h_1(z_0)|}{2}$ for $z\in\mathscr{D}_c(z_{0,\epsilon},\sqrt{\epsilon r'})$. Therefore,

$$\left|(z-z_{0,\epsilon})^2-\epsilon h_{1,\epsilon}(z)\right|>\epsilon\left(\frac{|h_1(z_0)|}{2}-r'\right)>\epsilon r'. \tag{2.70}$$

The inequalities (2.69) and (2.70) prove that in $\mathscr{D}_c(z_{0,\epsilon},\sqrt{\epsilon r'})$, the equation (2.68) has no solution and that the equation $(z-z_{0,\epsilon})^3-\epsilon(z-z_{0,\epsilon})h_{1,\epsilon}(z)=0$ has only one solution there.

We now study the number of solutions in $\mathcal{D}_c(z_{0,\epsilon}, \sqrt{\epsilon r'})$ of the equation (2.67). Consider

$$f_\epsilon(z) = (z - z_{0,\epsilon})^3 - \epsilon(z - z_{0,\epsilon})h_{1,\epsilon}(z) + \epsilon^2 h_{2,\epsilon}(z),$$
$$g_\epsilon(z) = (z - z_{0,\epsilon})^3 - \epsilon(z - z_{0,\epsilon})h_{1,\epsilon}(z).$$

We have $|f_\epsilon(z) - g_\epsilon(z)| = \epsilon^2 |h_{2,\epsilon}(z)|$. We consider $z \in \mathscr{C}(z_{0,\epsilon}, \sqrt{\epsilon r'})$. From (2.70), $|g_\epsilon(z)| > (\epsilon r')^{3/2}$. Therefore, for each $\epsilon \le \epsilon_{10} \le \epsilon_9$, it holds that $|g_\epsilon(z)| > \epsilon^2 |h_{2,\epsilon}(z)| = |f_\epsilon(z) - g_\epsilon(z)|$. Thus, from Rouché's theorem, the equation (2.67) has only one solution in $\mathcal{D}_o(z_{0,\epsilon}, \sqrt{\epsilon r'})$, denoted by z_ϵ. To obtain z_ϵ, we write

$$z_\epsilon - z_{0,\epsilon} = \frac{-\epsilon^2 h_{2,\epsilon}(z_\epsilon)}{(z_\epsilon - z_{0,\epsilon})^2 - \epsilon h_{1,\epsilon}(z_\epsilon)}.$$

Since $|(z - z_{0,\epsilon})^2 - \epsilon h_{1,\epsilon}(z)| > \epsilon r'$ on $\mathcal{D}_c(z_{0,\epsilon}, \sqrt{\epsilon r'})$ (see (2.70)), we get that

$$\left| z_\epsilon - z_{0,\epsilon} \right| \le \frac{\epsilon^2 K_2}{\epsilon r'} = \mathcal{O}(\epsilon).$$

But from equation (2.67), we also have $\epsilon(z_\epsilon - z_{0,\epsilon})h_{1,\epsilon}(z_\epsilon) = (z_\epsilon - z_{0,\epsilon})^3 + \epsilon^2 h_{2,\epsilon}(z_\epsilon)$ which leads to

$$z_\epsilon - z_{0,\epsilon} = \epsilon \frac{h_{2,\epsilon}(z_\epsilon)}{h_{1,\epsilon}(z_\epsilon)} + \frac{(z_\epsilon - z_{0,\epsilon})^3}{\epsilon h_{1,\epsilon}(z_\epsilon)}.$$

It is clear that

$$\frac{h_{2,\epsilon}(z_\epsilon)}{h_{1,\epsilon}(z_\epsilon)} - \frac{h_2(z_0)}{h_1(z_0)} = o(1),$$

so that

$$z_\epsilon - z_{0,\epsilon} = \epsilon \frac{h_2(z_0)}{h_1(z_0)} + o(\epsilon).$$

We now evaluate the two remaining solutions of (2.67) located in the set $\mathcal{D}_o(z_0, r) \backslash \mathcal{D}_o(z_{0,\epsilon}, \sqrt{\epsilon r'})$, denoted z_ϵ^-, z_ϵ^+. As $|z_\epsilon^- - z_{0,\epsilon}| > \sqrt{r'\epsilon}$, we can write

$$\left(z_\epsilon^- - z_{0,\epsilon}\right)^2 = \epsilon h_{1,\epsilon}(z_\epsilon^-) - \epsilon^2 \frac{h_{2,\epsilon}(z_\epsilon^-)}{z_\epsilon^- - z_{0,\epsilon}} \tag{2.71}$$

This implies that $|z_\epsilon^- - z_{0,\epsilon}| = \mathcal{O}(\sqrt{\epsilon})$ and that $|z_\epsilon^- - z_0| = o(1)$. Taking a suitable branch of the square root, (2.71) implies that

$$z_\epsilon^- - z_{0,\epsilon} = -\sqrt{\epsilon h_{1,\epsilon}(z_\epsilon^-)} + o(\sqrt{\epsilon}) = -\sqrt{\epsilon h_1(z_0)} + o(\sqrt{\epsilon}).$$

We obtain similarly that $z_\epsilon^+ - z_{0,\epsilon} = \sqrt{\epsilon h_1(z_0)} + o(\sqrt{\epsilon})$.

We finally verify that if z_0 and $z_{0,\epsilon}$ belong to $\mathbb{R}$ for each ϵ, and that $h_i(z)$ and $h_{i,\epsilon}(z)$ belong to $\mathbb{R}$ for each ϵ if $z \in \mathbb{R}$ for $i = 1, 2$, then z_ϵ is real while $z_\epsilon^-, z_\epsilon^+$ are real if $h_1(z_0) > 0$.

If z_ϵ is not real, it is clear that z_ϵ^* is also solution of (2.67) because functions $h_{i,\epsilon}$ verifies $(h_{i,\epsilon}(z))^* = h_{i,\epsilon}(z^*)$. As $|z_\epsilon^* - z_{0,\epsilon}| = |z_\epsilon - z_{0,\epsilon}| = \mathcal{O}(\epsilon)$, and that (2.67) has a unique solution in the disk $\mathcal{D}_o(z_{0,\epsilon}, \sqrt{\epsilon r'})$, this implies that $z_\epsilon^* = z_\epsilon$. On the other hand, assume that $h_1(z_0) > 0$ and the $z_\epsilon^-, z_\epsilon^+$ are non-real. Then, $z_\epsilon^{+\,*}$ and $z_\epsilon^{-\,*}$ are also solution of (2.68). Since equation (2.68) has only two solutions outside the disk $\mathcal{D}_o(z_{0,\epsilon}, \sqrt{\epsilon r'})$, it follows that $\hat{z}_\epsilon^+$ and $\hat{z}_\epsilon^-$ are complex conjuguate. But as their real parts have opposite sign for ϵ small enough, this leads to a contradiction. Therefore $\hat{z}_\epsilon^+$ and $\hat{z}_\epsilon^-$ are real. We finally note that if $h_1(z_0) < 0$, then $\hat{z}_\epsilon^+$ and $\hat{z}_\epsilon^-$ are non real.

Chapter 3

Spectrum localization in the Gaussian information plus noise model

This short chapter is dedicated to study properties concerning the localization of the eigenvalues $\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{M,N}$ of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$. Roughly speaking, we will show that no eigenvalue appears outside $\mathscr{S}_N$ a.s. for all large N, and that the number of eigenvalues in each cluster $[x_{q,N}^-, x_{q,N}^+]$ coincides with the number of eigenvalues of $\mathbf{B}_N \mathbf{B}_N^*$ associated with $[x_{q,N}^-, x_{q,N}^+]$. This last property will be referred to as the exact separation of the eigenvalues of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$.

Historically, the first results concerning localization of the eigenvalues have been given by Bai & Silverstein in a couple of papers [4] [5], in the context of zero-mean correlated random matrix model (possibly non-Gaussian). The technics used appear to be rather complicated, and more recently, Haagerup & Thorbjornsen [22] introduced, in the context of Gaussian Wigner matrices, a much more simple method fully exploiting the properties of the Gaussian model to study the almost sure absence of sample eigenvalues outside the support of the limiting spectral distribution. Following this work, Capitaine et al. used this method and extended this result to the so-called Wigner deformed model (i.e the sum of a Wigner matrix plus a deterministic Hermitian matrix with finite rank), and moreover proved an exact separation property, but using the method of [5].

In this chapter, we use the method of [22] to prove that for $\epsilon > 0$ such that $(a-\epsilon, b+\epsilon)$ is outside $\mathscr{S}_N$ for all large N, then no eigenvalue of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ lives in (a, b) almost surely for all large N. We moreover show that it is still possible to use the results of [22] to prove an exact separation property. We will prove here that almost surely, for N large enough, the number of eigenvalues of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ less than a (resp. greater than b) coincides with the number of eigenvalues of $\mathbf{B}_N \mathbf{B}_N^*$ associated to the clusters included into $[0, a]$ (resp. included into $[b, \infty)$).

In this chapter, we will also study the limiting behaviour of the eigenvalues of information plus noise matrices, in the context of spiked model, i.e when the rank of the deterministic matrix is independent of the dimensions, i.e $K = \text{rank}\,(\mathbf{B}_N)$ is independent of N and the K non zero eigenvalues of $\mathbf{B}_N \mathbf{B}_N^*$ $\lambda_{M-K+1,N}, \dots, \lambda_{M,N}$ converge to respective limits $\gamma_1 < \dots < \gamma_K$ when $N \to \infty$. Using the characterization of the support in this context (theorem 2.5.1 in chapter 2), and the above general results on almost sure localization, we prove that if $\gamma_k > \sigma^2 \sqrt{c}$, the corresponding eigenvalue of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$, i.e $\hat{\lambda}_{M-K+k}$, splits from the other eigenvalues and have a deterministic limit, a phenomenon called "phase transition" in the literature.

First results of this type have once again first been discovered for the zero-mean correlated model. In this context, Johnstone [28], on the basis of several applications examples, proposed the scenario where the eigenvalues of $\mathbf{H}_N$ are all equal to 1 except a few ones, and outlined the problem of the behaviour the eigenvalues of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ in this context. The first study was given by Baik et al. [6] in the case where the entries of $\mathbf{W}_N$ are Gaussian. Using extensively the explicit form of the joint probability distribution of the entries of $\mathbf{\Sigma}_N$, [6] established the almost sure convergence of the largest eigenvalues of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ as well as central limit theorems. Later, Baik & Silverstein [7] studied completely the almost sure convergence in the non Gaussian case. Their method heavily relies on the results of Bai-Silverstein [4][5] on the localization of the eigenvalues of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$, as well as the characterization of the support of the deterministic limiting eigenvalue distribution of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ provided in Silverstein-Choi [40] (their technic was especially used in the previous chapter for the proof of theorem 2.5.1). More recently, Bai-Yao [3] addressed central limit theorems in the non Gaussian case. It is worth noticing that these works have been unified by Benaych & Nadakuditi [9] who proposed a common method to study several spiked random matrices model. The ideas of [9] were to reduce all of these models to the spiked Wigner case, and to study the characteristic polynomial of such matrices. This work was shortly followed by [8] where central limit theorems are established.

The chapter is organized as follows. In section 3.1, we introduce the main tool of the approach of Haagerup &

Thorbjornsen [22], which consists in a "smooth indicator function" whose purpose is to count the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ inside a compact interval outside $\mathscr{S}_N$, and we establish the main related properties. In section 3.2, we prove that the spectrum of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ is almost surely included in $\mathscr{S}_N$ for all large N. We also prove a further result by evaluating the rate of convergence to 0 of the probability that the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ escape from a neighborhood of the support $\mathscr{S}_N$. Finally, we prove the property of almost sure separation in the spectrum of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. In section 3.3, we use the previous results in the special case of the spiked models, and describe the limiting behaviour of the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. In section 3.4, we provide some discussions about the results of this chapter and give some numerical examples.

3.1 Preliminary results

In this section, we derive results concerning the quantity

$$\frac{1}{M}\operatorname{Tr}\varphi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*) \triangleq \frac{1}{M}\sum_{k=1}^{M}\varphi\left(\hat{\lambda}_{k,N}\right),$$

where $\varphi \in \mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$. More precisely, we compute an approximation for its expectation and evaluate the rate of convergence of its variance. Such results will be of crucial importance to prove that $\hat{\lambda}_{1,N},\ldots,\hat{\lambda}_{M,N}$ are located inside $\mathscr{S}_N$ for all large N and the separation of the eigenvalues.

Remark 3.1.1. *If φ is equal to 1 on an interval $[a,b]$ and 0 on $\mathbb{R}\backslash(a-\epsilon,b+\epsilon)$, then $\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)$ plays the role of a smooth counting function of the set $[a,b]$, i.e. it is equal to the number of eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ inside $[a,b]$, provided that no eigenvalue of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ belongs to $(a-\epsilon,a)\cup(b,b+\epsilon)$.*

Before stating the main results of this section, we recall the useful property derived in theorem 2.2.2, i.e for all $z \in \mathbb{C}\backslash\mathbb{R}$,

$$\mathbb{E}\left[\hat{m}_N(z)\right] = m_N(z) + \frac{\chi_N(z)}{N^2}. \tag{3.1}$$

for all large N, with χ_N holomorphic on $\mathbb{C}\backslash\mathbb{R}$ satisfying $\left|\chi_N(z)\right| \le \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)$, with P_1 and P_2 two polynomials with positive coefficients independent of N. The following lemma will be also useful.

Lemma 3.1.1. *Let $\varphi \in \mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ independent of N and (h_N) a sequence of holomorphic functions on $\mathbb{C}\backslash\mathbb{R}$ satisfying the polynomial bound $|h_N(z)| \le \mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)$, with P_1 and P_2 two polynomials with positive coefficients independent of N, z. Then, we have*

$$\limsup_{y\downarrow 0}\int_{\mathbb{R}}\left|\varphi(x)h_N(x+iy)\right|\mathrm{d}x \le C < \infty,$$

with C a positive constant independent of N.

Proof. Lemma 3.1.1 is proved in Capitaine et al. [11], and relies essentially on the ideas of Haagerup & Thorbjornsen [22]. □

The next two results are the extensions to the information plus noise model of the results of [22, Prop.4.7] and the proofs use exactly the same arguments.

Lemma 3.1.2. *Let $\varphi \in \mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ independent of N. Then,*

$$\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right] = \int_{\mathbb{R}}\varphi(\lambda)\mathrm{d}\mu_N(\lambda) + \mathscr{O}\left(\frac{1}{N^2}\right).$$

Proof. From property 1.2.2 in chapter 1,

$$\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right] = \frac{1}{\pi}\lim_{y\downarrow 0}\mathrm{Im}\left(\int_{\mathbb{R}}\varphi(x)\mathbb{E}\left[\hat{m}_N(x+iy)\right]\right)\mathrm{d}x$$

as well as

$$\int_{\mathbb{R}}\varphi(\lambda)\mathrm{d}\mu_N(\lambda) = \frac{1}{\pi}\lim_{y\downarrow 0}\mathrm{Im}\left(\int_{\mathbb{R}}\varphi(x)m_N(x+iy)\right)\mathrm{d}x.$$

Using (3.1), it follows that

$$\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right]-\int_{\mathbb{R}}\varphi(\lambda)\mathrm{d}\mu_N(\lambda)=\frac{1}{\pi N^2}\lim_{y\downarrow 0}\operatorname{Im}\left(\int_{\mathbb{R}}\varphi(x)\chi_N(x+iy)\right)\mathrm{d}x. \tag{3.2}$$

The result of the proposition then follows from a direct application of lemma 3.1.1. □

Lemma 3.1.3. *Let $\varphi\in\mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ independent of N and constant on each cluster $[x_{q,N}^-,x_{q,N}^+]$ of $\mathscr{S}_N$ for all large N. Then,*

$$\operatorname{Var}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right]=\mathscr{O}\left(\frac{1}{N^4}\right).$$

Proof. The application of Poincaré inequality gives

$$\begin{aligned}&\operatorname{Var}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right]\\&\le\frac{\sigma^2}{N}\sum_{k,l}\mathbb{E}\left[\left|\frac{\partial}{\partial W_{k,l,N}}\left\{\frac{1}{M}\operatorname{Tr}\varphi\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right\}\right|^2+\left|\frac{\partial}{\partial W_{k,l,N}^*}\left\{\frac{1}{M}\operatorname{Tr}\varphi\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right\}\right|^2\right].\end{aligned}$$

Using theorem 1.3.1 in chapter 1, we obtain

$$\frac{\partial}{\partial W_{k,l,N}}\left\{\frac{1}{M}\operatorname{Tr}\varphi\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right\}=\frac{1}{M}\left[\boldsymbol{\Sigma}_N^*\varphi'\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right]_{l,k}, \tag{3.3}$$

and the derivative w.r.t $W_{k,l,N}^*$ is the complex conjuguate of (3.3) since φ is real-valued. This yields

$$\operatorname{Var}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)\right]\le\frac{C}{N^2}\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi'\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)^2\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right],$$

for some constant $C>0$ independent of N. Applying lemma 3.1.2 with the $\mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ function $\lambda\mapsto\varphi'(\lambda)^2\lambda$, we get

$$\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi'\left(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right)^2\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\right]=\int_{\mathbb{R}}\lambda\varphi'(\lambda)^2\mathrm{d}\mu_N(\lambda)+\mathscr{O}\left(\frac{1}{N^2}\right).$$

But from the assumptions, $\operatorname{supp}(\varphi')\cap\mathscr{S}_N=\emptyset$ and the result of the proposition follows. □

3.2 Localization of the eigenvalues

In this section, we study the almost-sure location of the eigenvalues of $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$ for N large enough. We first prove the almost-sure absence of the eigenvalues of $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$ outside the support $\mathscr{S}_N$ of the limiting spectral measure μ_N, and we further show that the probability that an eigenvalue escapes from any neighborhood containing $\mathscr{S}_N$ for all large N decreases at rate $\frac{1}{N^p}$ for all $p\in\mathbb{N}$. The third part of this section is dedicated to study the eigenvalue separation phenomenon, namely that for N large enough, the eigenvalues of $\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*$ split into groups following the clusters of $\mathscr{S}_N$.

3.2.1 Absence of eigenvalues outside the support

The arguments we use to prove the almost-sure absence of the eigenvalues outside $\mathscr{S}_N$ are due to Haagerup & Thorbjornsen [22] and have been used by Capitaine et al. [11].

Theorem 3.2.1. *Let $a,b\in\mathbb{R}$, and $\epsilon>0$ such that $(a-\epsilon,b+\epsilon)\cap\mathscr{S}_N=\emptyset$ for all large N. Then, with probability one,*

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

for N large enough.

Proof. The result is in fact a simple consequence of the results derived in section 3.1. Indeed, consider a function $\varphi \in \mathscr{C}_c^\infty(\mathbb{R},[0,1])$ and

$$\varphi(\lambda) = \begin{cases} 1 & \text{for } \lambda \in [a,b] \\ 0 & \text{for } \lambda \in \mathbb{R}\backslash(a-\epsilon, b+\epsilon). \end{cases}$$

Then we have

$$\begin{aligned} \mathbb{P}\left(\frac{1}{M}\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right) > \frac{1}{N^{4/3}}\right) &\leq N^{8/3}\mathbb{E}\left|\frac{1}{M}\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right|^2 \\ &= N^{8/3}\left(\left|\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right]\right|^2 + \operatorname{Var}\left[\frac{1}{M}\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right]\right) \\ &= \mathscr{O}\left(\frac{1}{N^{4/3}}\right), \end{aligned}$$

where the last inequality follows from lemmas 3.1.2 and 3.1.3. Therefore, with probability one,

$$\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right) \leq \frac{M}{N^{4/3}},$$

for all large N. From the definition of φ, the number of eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ in $[a,b]$ is upper-bounded by $\operatorname{Tr}\varphi\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)$, and is therefore $\mathscr{O}(N^{-1/3})$ with probability one for all large N. Since this number has to be an integer, it is equal to 0 and consequently we deduce that no eigenvalue belongs to $[a,b]$ with probability one for all large N. This concludes the proof of theorem 3.2.1. □

3.2.2 Escape probability of the eigenvalues

In this section, we improve the result obtained in theorem 3.2.1 by evaluating the probability that an eigenvalue of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ escapes from a compact neighborhood containing $\mathscr{S}_N$ for all large N. For a compact set $\mathscr{K} \subset \mathbb{R}$, we denote by $\mathscr{K}_\epsilon$ the closed ϵ-neighborhood of $\mathscr{K}$, i.e the compact set

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

Theorem 3.2.2. *Fix $\epsilon > 0$ and let $\mathscr{K}$ be a compact set containing $\mathscr{S}_N$ for all large N, and $\mathscr{K}_\epsilon$ the closed ϵ-neighborhood of $\mathscr{K}$. Then, it holds that*

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

for all $l \in \mathbb{N}$.

To prove this result, we consider a function $\varphi \in \mathscr{C}^\infty(\mathbb{R},[0,1])$ and

$$\varphi(\lambda) = \begin{cases} 1 & \text{for } \lambda \in \mathscr{K}_\epsilon^c, \\ 0 & \text{for } \lambda \in \mathscr{K}. \end{cases} \tag{3.4}$$

From this definition, we clearly have

$$\mathbb{P}\left(\exists k : \hat{\lambda}_{k,N} \in \mathscr{K}_\epsilon^c\right) \leq \mathbb{P}\left(\operatorname{Tr}\varphi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*) \geq 1\right) \leq \mathbb{E}\left[\left(\operatorname{Tr}\varphi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l}\right]$$

for $l \in \mathbb{N}$. Therefore, to establish theorem 3.2.2, it is sufficient to prove the following lemma.

Lemma 3.2.1. *For all function $\psi \in \mathscr{C}^\infty(\mathbb{R},\mathbb{R})$ constant outside a compact set and vanishing on $\mathscr{S}_N$ for N large enough, it holds that*

$$\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l}\right] = \mathscr{O}\left(\frac{1}{N^{2l}}\right), \tag{3.5}$$

for each $l \in \mathbb{N}$.

Proof. We prove lemma 3.2.1 by induction on l. Consider $l = 1$. Denote by b the constant value taken by ψ over the complementary of a certain compact set. Write $\psi = \tilde{\psi} + b$, where $\tilde{\psi} \in \mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ satisfying $\tilde{\psi} = -b$ on $\mathscr{S}_N$ for N large enough. Using lemmas 3.1.2 and 3.1.3, we obtain

$$\begin{aligned}\operatorname{Var}\left[\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right] &= \operatorname{Var}\left[\operatorname{Tr}\tilde{\psi}(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right] = \mathscr{O}\left(\frac{1}{N^2}\right),\\ \mathbb{E}\left[\operatorname{Tr}\tilde{\psi}(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right] &= M\int_{\mathbb{R}}\tilde{\psi}(\lambda)\mathrm{d}\mu_N(\lambda) + \mathscr{O}\left(\frac{1}{N}\right) = -Mb + \mathscr{O}\left(\frac{1}{N}\right).\end{aligned}$$

As $\mathbb{E}\left[\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right] = bM + \mathbb{E}\left[\operatorname{Tr}\tilde{\psi}(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right]$, this leads to $\mathbb{E}\left[\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right] = \mathscr{O}\left(\frac{1}{N}\right)$. From the equality,

$$\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^2\right] = \left(\mathbb{E}\left[\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right]\right)^2 + \operatorname{Var}\left[\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right],$$

we finally obtain that (3.5) holds for $l = 1$.

We now assume that (3.5) holds until the order $l-1$ for each function of $\mathscr{C}^\infty(\mathbb{R},\mathbb{R})$ vanishing on $\mathscr{S}_N$ for N large enough and constant outside a compact set. We consider such a function ψ and evaluate the behaviour of the $2l$-th order moment of $\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)$. We have

$$\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l}\right] = \left(\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^l\right]\right)^2 + \operatorname{Var}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^l\right]. \tag{3.6}$$

The first term of the righthandside of (3.6) can be upperbounded as follows

$$\left(\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^l\right]\right)^2 \leq \mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^2\right]\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2(l-1)}\right] = \mathscr{O}\left(\frac{1}{N^{2l}}\right),$$

using that (3.5) holds until the order $l-1$. The second term of the righthandside of (3.6) can be evaluated with the Poincaré inequality. Using that the partial derivative of $\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)$ w.r.t. $W_{i,j,N}$ and $W_{i,j,N}^*$ are equal respectively to $\mathbf{e}_j^*\mathbf{\Sigma}_N^*\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\mathbf{e}_i$ and $\mathbf{e}_i^*\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\mathbf{\Sigma}_N\mathbf{e}_j$ (theorem 1.3.1 in chapter 1), we obtain immediately that

$$\operatorname{Var}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^l\right] \leq C\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}\left(\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)^2\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l-2}\right],$$

with $C > 0$ a generic constant independent of N. Using Hölder's inequality, it follows that

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}\left(\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)^2\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l-2}\right]\\ &\quad\leq C\left(\mathbb{E}\left|\frac{1}{N}\operatorname{Tr}\left(\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)^2\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right|^l\right)^{\frac{1}{l}}\left(\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l}\right]\right)^{\frac{l-1}{l}}.\end{aligned} \tag{3.7}$$

Since the function $\lambda \to \psi'(\lambda)^2\lambda$ which belongs to $\mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ vanishes on $\mathscr{S}_N$ and is constant outside a compact set for N large enough,

$$\begin{aligned}&\mathbb{E}\left|\frac{1}{N}\operatorname{Tr}\left(\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)^2\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right|^l\\ &\quad\leq \sqrt{\mathbb{E}\left|\frac{1}{N}\operatorname{Tr}\left(\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)^2\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right|^2}\sqrt{\mathbb{E}\left|\frac{1}{N}\operatorname{Tr}\left(\psi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)^2\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right|^{2(l-1)}}\\ &\quad= \mathscr{O}\left(\frac{1}{N^{2l}}\right).\end{aligned}$$

Plugging the previous estimates into (3.7), we get

$$\operatorname{Var}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^l\right] \leq \frac{C}{N^2}\left(\mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l}\right]\right)^{\frac{l-1}{l}}.$$

Define $x_N = \mathbb{E}\left[\left(\operatorname{Tr}\psi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\right)^{2l}\right]$ and $u_N = N^{2l}x_N$. From (3.6), we have the inequalities $x_N \leq \frac{C_1}{N^2}x_N^{\frac{l-1}{l}} + \frac{C_2}{N^{2l}}$ and $u_N \leq C_1 u_N^{\frac{l-1}{l}} + C_2$. We claim that the sequence (u_N) is bounded. If this is not the case, there exists a subsequence u_{k_N} extracted from u_N which converges towards $+\infty$. However, the inequality $\frac{C_1}{u_{k_N}^{1/l}} + \frac{C_2}{u_{k_N}} \geq 1$ must holds for N large enough. As $u_{k_N} \to +\infty$, this leads to a contradiction. Therefore, u_N is bounded and $x_N \leq \frac{C}{N^{2l}}$ for N large enough. This proves lemma 3.2.1. □

3.2.3 Separation of the eigenvalues

In this section, we show that the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ splits into several groups, related to the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$. We refer the reader to chapter 2 section 2.4 for an exposition of function w_N and its link with the support $\mathscr{S}_N$.

Theorem 3.2.3. *Let $a,b\in\mathbb{R}$ and $\epsilon>0$ such that $(a-\epsilon,b+\epsilon)\cap\mathscr{S}_N=\emptyset$ for all large N. Then, under Assumption* **A-1**, *with probability one,*

$$\operatorname{card}\{k:\hat\lambda_{k,N}<a\}=\operatorname{card}\{k:\lambda_{k,N}<w_N(a)\}, \tag{3.8}$$

$$\operatorname{card}\left\{k:\hat\lambda_{k,N}>b\right\}=\operatorname{card}\left\{k:\lambda_{k,N}>w_N(b)\right\}, \tag{3.9}$$

for N large enough.

We first prove (3.8) and assume that $a>0$ because (3.8) is obvious if $a\le 0$. We consider $\eta<\epsilon$ and assume without restriction that $0<\eta<a$. We consider a function $\varphi_a\in\mathscr{C}_c^\infty(\mathbb{R},[0,1])$, independent of N, and such that

$$\varphi_a(\lambda)=\begin{cases}1 & \forall\lambda\in[0,a-\eta]\\ 0 & \forall\lambda\in(-\infty,-\eta)\cup(a,\infty).\end{cases}$$

By lemma 3.1.2, we have

$$\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right]-\int_{\mathbb{R}^+}\varphi_a(\lambda)\mathrm{d}\mu_N(\lambda)=\mathscr{O}\left(\frac{1}{N^2}\right),$$

or equivalently

$$\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right]=\mu_N\left([0,a-\eta]\right)+\mathscr{O}\left(\frac{1}{N^2}\right).$$

Lemma 3.1.3 also implies that

$$\operatorname{Var}\left[\frac{1}{M}\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right]=\mathscr{O}\left(\frac{1}{N^4}\right).$$

Therefore,

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{M}\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)-\mu_N\left([0,a-\eta]\right)\right|>\frac{1}{N^{4/3}}\right)\\ &\le N^{8/3}\operatorname{Var}\left[\frac{1}{M}\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)\right]+N^{8/3}\left|\mathbb{E}\left[\frac{1}{M}\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)-\mu_N\left([0,a-\eta]\right)\right]\right|^2\\ &=\mathscr{O}\left(\frac{1}{N^{4/3}}\right),\end{aligned}$$

which implies that with probability one,

$$\frac{1}{M}\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)=\mu_N\left([0,a-\eta]\right)+\mathscr{O}\left(\frac{1}{N^{4/3}}\right). \tag{3.10}$$

By defining $Q_a=\max\{q:x_{q,N}^+<a\}$, it is clear that $\mu_N([0,a-\eta])=\sum_{q=1}^{Q_a}\mu_N([x_{q,N}^-,x_{q,N}^+])$ because $\mu_N((a-\eta,a))=0$. From property 2.4.3,

$$\mu_N([0,a-\eta])=\frac{M-K}{M}+\sum_{q=1}^{Q_a}|\mathscr{I}_q|,$$

with $\mathscr{I}_q=\{k\in\{1,\dots,K\}:\lambda_{M-K+k,N}\in(w_{q,N}^-,w_{q,N}^+)\}$, and thus

$$\mu_N([0,a-\eta])=\operatorname{card}\{k:\lambda_{k,N}<w_N(a)\}.$$

Therefore, using (3.10), we get that

$$\left|\operatorname{Tr}\varphi_a\left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right)-\operatorname{card}\{k:\lambda_{k,N}<w_N(a)\}\right|=\mathscr{O}\left(\frac{1}{N^{1/3}}\right). \tag{3.11}$$

But almost surely, for N large enough, Tr $\varphi_a(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)$ is exactly the number of eigenvalues contained in $[0,a]$ because no eigenvalue of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ belong to $[a-\eta,a]$ (use theorem 3.2.1 with $a-\eta$ in place of a). The left handside of (3.11) is thus an integer. Since this integer decreases at rate $N^{-1/3}$, it is equal to zero for N large enough (for further details, see the properties of w_N in section 2.4 chapter 2). To evaluate the number of eigenvalues in the interval $(b,+\infty)$, we use that no eigenvalue belongs to $[a,b]$ (theorem 3.2.1). Therefore,

$$\mathrm{card}\{k:\hat{\lambda}_{k,N}>b\} = M-\mathrm{card}\{k:\hat{\lambda}_{k,N}<a\}.$$

which coincides with the number of eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ in interval $(w_N(b),+\infty)$, for all large N, a.s. This concludes the proof of theorem 3.2.3.

3.3 Applications to the spiked models

In this section, we use the results of the previous section on localization of the eigenvalues, to compute the limits of the largest eigenvalues when dealing with spiked models. We refer the reader to chapter 2 section 2.5 for an exposition on the spiked model. Recall that under Assumption **A-2**, K_s was defined as the number of limiting eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ greater than $\sigma^2\sqrt{c}$, and that we defined function ψ by $\psi(\lambda,c)=\frac{(\lambda+\sigma^2c)(\lambda+c)}{\lambda}$. The main result is stated as follows.

Theorem 3.3.1. *Under Assumption* **A-2**,

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

Proof. We first assume $\gamma_k\neq\sigma^2\sqrt{c}$ for $k=1,\dots,K$. To prove theorem 3.3.1, we use theorem 3.2.3 in section 3.2.3. Let $k\in\{1,\dots,K_s\}$. From theorem 2.5.1, the eigenvalue $\lambda_{M-K_s+k,N}$ is the unique eigenvalue of $\mathbf{B}_N\mathbf{B}_N^*$ associated with the interval $(x_{q,N}^-,x_{q,N}^+)$, $q=k+1$, for N large enough. Moreover, the number of clusters of $\mathcal{S}_N$ is equal to K_s+1 for N large enough and the sequences $x_{q,N}^-$ and $x_{q,N}^+$ converge towards limits equal to $\sigma^2(1-\sqrt{c})^2$ and $\sigma^2(1+\sqrt{c})^2$ for $q=1$, and equal to $\psi(\gamma_l,c)$ for $q\geq 2$ and $l=K-K_s+q-1$. Theorem 3.2.3 implies that for each $\epsilon>0$, almost surely for N large enough, then $\hat{\lambda}_{M-K_s+k,N}\in(\psi(\gamma_k,c)-\epsilon,\psi(\gamma_k,c)+\epsilon)$ for $k=1,\dots,K_s$ and that $\hat{\lambda}_{M-K_s,N}\in(\sigma^2(1-\sqrt{c})^2-\epsilon,\sigma^2(1+\sqrt{c})^2+\epsilon)$. This shows that $\hat{\lambda}_{M-K_s+k,N}\to\psi(\gamma_k,c)$ for $k=1,\dots,K_s$.

We now prove the convergence of $\hat{\lambda}_{M-K_s,N}$ to $\sigma^2(1+\sqrt{c})^2$. We have already shown $\limsup_N\hat{\lambda}_{M-K_s,N}\leq\sigma^2(1+\sqrt{c})^2$ almost surely. It remains to prove $\liminf_N\hat{\lambda}_{M-K_s,N}\geq\sigma^2(1+\sqrt{c})^2$. Assume the converse is true. Then it exists $\epsilon>0$ such that $\liminf_N\hat{\lambda}_{M-K_s,N}<\sigma^2(1+\sqrt{c})^2-\epsilon$. We can thus extract a subsequence $\hat{\lambda}_{M-K_s,\varphi(N)}$ converging towards a limit less than $\sigma^2(1+\sqrt{c})^2-\epsilon$. Let $\hat{\mu}_{\phi(N)}$ be the empirical spectral measure associated with matrix $\mathbf{\Sigma}_{\varphi(N)}\mathbf{\Sigma}_{\varphi(N)}^*$. We deduce that

$$\hat{\mu}_{\varphi(N)}\left((\sigma^2(1+\sqrt{c})^2-\epsilon,\sigma^2(1+\sqrt{c})^2]\right)=0\quad\text{a.s for all large N.}\tag{3.12}$$

Theorem 2.2.1 in chapter 2 implies that $\hat{\mu}_{\varphi(N)}$ converges towards the Marcenko-Pastur distribution whose support is exactly $(\sigma^2(1-\sqrt{c})^2,\sigma^2(1+\sqrt{c})^2]]$. This contradicts (3.12).

We now handle the case where $\gamma_{K-K_s}=\sigma^2\sqrt{c}$. For this, we will use the Fan inequality (see theorem 1.3.1 in chapter 1 section 1.3.3). For a $M\times N$ matrix $\mathbf{A}$, we will denote by $\kappa_1(\mathbf{A}),\dots,\kappa_M(\mathbf{A})$ its singular values. We define $\mathbf{u}_{k,N}$ and $\mathbf{v}_{k,N}$ the left and right singular vector of $\mathbf{B}_N$ associated with $\kappa_k(\mathbf{B}_N)$. Fan inequality (chapter 1 section 1.3.3) gives, for $\epsilon>0$,

$$\kappa_{M-K_s}(\mathbf{B}_N+\sigma\mathbf{W}_N)\leq\kappa_{M-K_s}(\mathbf{B}_N+\sigma\mathbf{W}_N+\epsilon\mathbf{u}_{M-K_s,N}\mathbf{v}_{M-K_s,N}^*)+\kappa_M(\epsilon\mathbf{u}_{M-K_s,N}\mathbf{v}_{M-K_s,N}^*),$$

and

$$\kappa_{M-K_s}(\mathbf{B}_N+\sigma\mathbf{W}_N+\epsilon\mathbf{u}_{M-K_s,N}\mathbf{v}_{M-K_s,N}^*)\leq\kappa_{M-K_s}(\mathbf{B}_N+\sigma\mathbf{W}_N)+\kappa_M(\epsilon\mathbf{u}_{M-K_s,N}\mathbf{v}_{M-K_s,N}^*).$$

From the results of the previous section, it is clear that, almost surely,

$$\kappa_{M-K_s}(\mathbf{B}_N+\sigma\mathbf{W}_N+\epsilon\mathbf{u}_{M-K_s,N}\mathbf{v}_{M-K_s,N}^*)=\sqrt{\psi\left(\left(\sqrt{\gamma_{K-K_s}}+\epsilon\right)^2,c\right)}+o(1).$$

Therefore, we end up with

$$\sqrt{\psi\left(\left(\sqrt{\gamma_{K-K_s}}+\epsilon\right)^2,c\right)}-\epsilon\leq\liminf_N\kappa_{M-K_s}(\mathbf{B}_N+\sigma\mathbf{W}_N)\leq\limsup_N\kappa_{M-K_s}(\mathbf{B}_N+\sigma\mathbf{W}_N)\leq\sqrt{\psi\left(\left(\sqrt{\gamma_{K-K_s}}+\epsilon\right)^2,c\right)}+\epsilon.$$

Since $\psi(\lambda,c)\to\sigma^2(1+\sqrt{c})^2$ when $\lambda\to\sigma^2\sqrt{c}$, this completes the results of theorem 3.3.1. $\square$

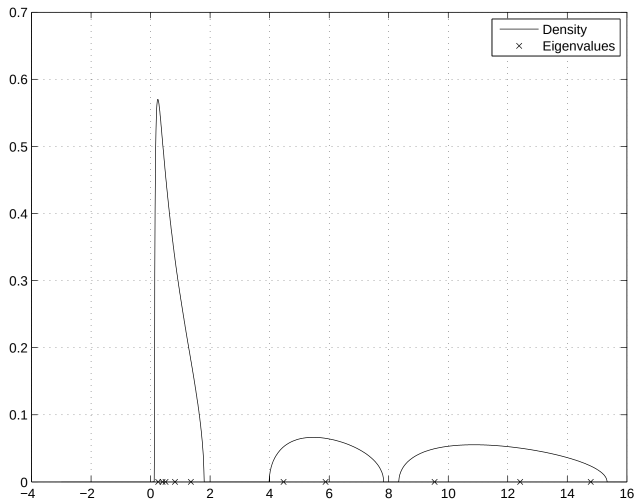


Figure 3.1: Density of μ_N and locations of the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$

3.4 Discussions and numerical examples

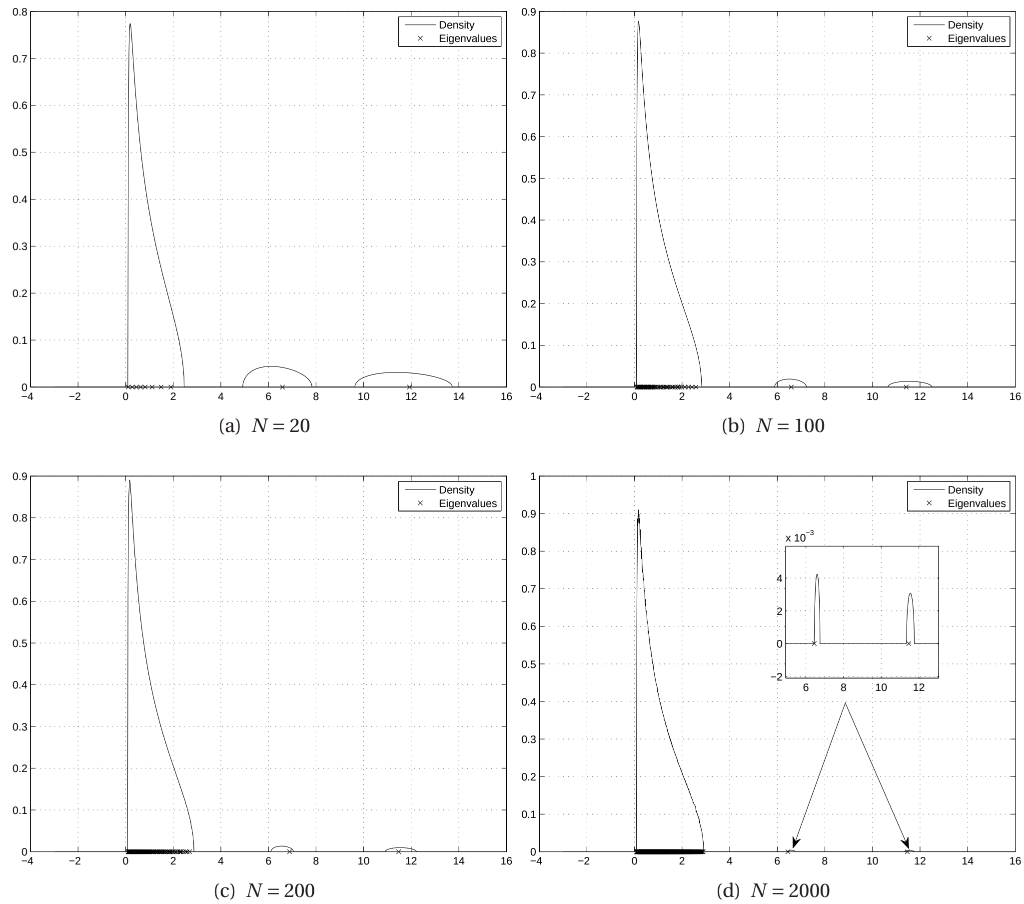
In this section, we discuss the results on localization and separation of the eigenvalues (theorems 3.2.1, 3.2.2 and 3.2.3), as well as the application to the spiked model (theorem 3.3.1), and give some numerical illustrations of the phenomena described.

To get some insights on theorem 3.2.1, assume for example that for all large N, the number of clusters Q of $\mathscr{S}_N$ does not depend on N, and that for each $q = 1, \ldots, Q$, the sequences of boundary points $(x_{q,N}^-)$ and $(x_{q,N}^+)$ converge torwards limits x_q^- and x_q^+, satisfying $x_1^- < x_1^+ < x_2^- < x_2^+ < \ldots < x_Q^- < x_Q^+$. In this context, theorem 3.2.1 implies that almost surely, for all $\epsilon > 0$, each eigenvalue belongs to one of the intervals $[x_q^- - \epsilon, x_q^+ + \epsilon]$ for N large enough.

To interpret theorem 3.2.3, we keep the same simplified assumptions and use the terminology introduced in section 2.6 chapter 2. In this case, the result of theorem 3.2.3 means that almost surely for N large enough, the number of sample eigenvalues that belong to each interval $[x_q^- - \epsilon, x_q^+ + \epsilon]$ coincides with the number of eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ that are associated with the cluster $[x_{q,N}^-, x_{q,N}^+]$ for all large N.

These two facts are illustrated in figure 3.1 where we have plotted the density of μ_N and the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. The parameters are $N = 20$, $M = 10$, $\sigma = 1$ and matrix $\mathbf{B}_N\mathbf{B}_N^*$ is diagonal with eigenvalues 0 (with multiplicity 5), 5 (with multiplicity 2) and 10 (with multiplicity 3). On figure 3.1, we clearly see that no eigenvalue of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ is located outside the support. This property is verified in practice for small values of N, which confirms that the escape of eigenvalues outside the support is a rare event (theorem 3.2.2). Moreover, we also see the clear separation between eigenvalues. Indeed, the three clusters contain respectively 5, 2 and 3 eigenvalues. Note that in the assumptions made in this manuscript (Assumption **A-1** in chapter 2 section 2.4), the non zero eigenvalues are supposed to have multiplicity one. It is not difficult to see that the statement of theorem 3.2.3 also holds in the general case, by counting each eigenvalue of $\mathbf{B}_N\mathbf{B}_N^*$ with its respective multiplicity.

In figure 3.2, we have represented the evolution of the density of μ_N in the spiked model assumption when $N = 20, 100, 200, 2000$. We have kept the same settings as in figure 3.1, except that eigenvalues 5 and 10 have always multiplicity 1 (and thus eigenvalue 0 have multiplicity $M-2$). The density of μ_N is given for different values of N, and the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ are also represented. This example agrees with the statements of theorems 2.5.1 and 3.3.1.



(a) $N = 20$ (b) $N = 100$

(c) $N = 200$ (d) $N = 2000$

Figure 3.2: Effects of the spiked model assumption on the density of μ_N and spectrum of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$

Chapter 4

Subspace and DoA estimation in large sensor networks

This chapter is devoted to the application of the random matrix results given in the previous chapters to the problem of "subspace estimation" in array processing.

In general, one has to infer on K parameters from a set of N multivariate observations with dimension M, which are composed of a signal and a noise part. Usually, the signal part has a low rank correlation matrix while the noise part has a full rank correlation matrix. To extract the K parameters of interest from the observations, a classical procedure is to exploit the fact that the observation space of the correlation matrix of the observations splits into two orthogonal subspaces: a signal subspace of dimension K (corresponding to the eigenspace of the signal correlation matrix) and a noise subspace with dimension $M-K$. In general, the resulting estimators are computationally much more affordable than other estimators such as those based on the maximum likelihood (M.L.) principle, which generally perform better but unfortunately need an exhaustive search in a multi-dimensional parameter space.

In order to formulate a generic subspace estimator, one must first infer the eigenvectors of the correlation matrix of the observations, which is not available. As a consequence, classical subspace estimation methods make use of the empirical correlation matrix, and approximate the eigenvectors of the true correlation matrix as the eigenvectors of the sample estimate. This procedure is clearly optimal when the number of samples N tends to infinity while the observation dimension M remains constant. Indeed, under certain ergodicity assumptions, when $N \to \infty$ for a fixed M, the sample correlation matrix of the observations converges almost surely to the true one, and consequently when $N >> M$ the sample eigenvectors (i.e. the eigenvectors of the sample correlation matrix) tend to be very good representations of the true ones. In practical applications, however, the number of available observations N and the observation dimension M are comparable in magnitude, which leads to strong discrepancies between the sample eigenvectors and the true ones. This originates from what is usually referred to as the breakdown effect of subspace-based techniques (see e.g. Tufts et al.[46]).

The fact that sample eigenvectors are not the best estimators of the true ones has been known for decades, although the study of valid alternatives to the classical estimators has been limited by the fact that investigations basically concentrated on the regime where $N >> M$. However, it has been recently suggested (see Mestre [51]) that finite sample size situations (whereby N and M are comparable in magnitude) can be better examined by investigating the asymptotic regime in which M and N converge to $+\infty$ at the same rate, i.e. $M, N \to +\infty$, whereas $c_N = \frac{M}{N}$ converges towards a strictly positive constant. Using random matrix theory, Mestre [51] showed that traditional subspace estimators are asymptotically biased in this asymptotic regime. Furthermore, consistent estimators for this regime can be found, which outperform the traditional ones for realistical values of M and N. In this context, random matrix theory can be very useful to characterize how the sample eigenvectors differ from the true ones in a scenario where M and N are comparable in magnitude and to derive alternative estimators of the eigenvectors that converge, not only when $N \to +\infty$ for fixed M, but also when $M, N \to +\infty$ at the same rate. This was more extensively demonstrated in [50] and [49], which respectively considered the characterization of the sample eigenvectors when $M, N \to +\infty$ at the same rate, and proposed alternative consistent estimators for these quantities in the new asymptotic regime.

However, [50] and [49] cannot be applied to the signal plus noise model considered here, unless the observations are random multivariate quantities that are Gaussian, independent and identically distributed in the time domain. In practice, however, there are multiple applications in which the observations do not present this structure, and are better modelled as a deterministic component (corresponding to the signal part) plus some additive noise, that is generally Gaussian distributed. This corresponds to the information plus noise matrix model studied in chapters 2 and 3. [51, 50, 49] used the classical zero mean correlated matrix model, developped in Silverstein [39], The purpose of this chapter is thus to propose improved subspace estimators for the information plus noise model, which will represent the case where the source signals are modelled as non-observable deterministic sequences, by using the previous results developped in chapters 2 and 3.

The chapter is organized as follows. In section 4.1, we introduce the model of observations which will be used throughout the chapter, and state the subspace estimation problem in this context. We also present the traditional estimation procedure used for the subspace estimation problem. In section 4.2, we derive a consistent subspace estimator in the regime $N, M \to \infty$ while M/N converges to $c \in (0,1)$. In section 4.3, we apply the previous results of subspace estimation to the problem of Direction of Arrival (DoA) estimation, and in particular we derive an improved MUSIC type algorithm and prove the consistency of the source angle of arrival estimates. In section 4.4, we provide some numerical examples illustrating the performance of the improved estimator.

4.1 Statistical model and classical subspace estimation

In this section, we introduce the classical statistical model associated with the subspace estimation problem in signal processing. We consider the context where K narrow band deterministic source signals $(s_k)_{k=1,\dots,K}$ are received by an antenna array of M elements, $K < M$. The corresponding M-dimensional observation signal $\mathbf{y}_n$ (at discrete time n) can be modelled as

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

where $\mathbf{A}$ is a $M \times K$ complex matrix whose entries represent the attenuation between the K source signals and the M receive antennas, $\mathbf{s}_n$ is a K-dimensional complex vector containing the transmitted signals from the K sources at time instant n, and where $\mathbf{v}_n$ is an additive white complex Gaussian noise with zero mean and covariance matrix $\mathbb{E}[\mathbf{v}_n\mathbf{v}_n^*] = \sigma^2\mathbf{I}_M$. We assume that $\mathbf{y}_n$ is available from $n = 1$ to $n = N$, and that $c_N = \frac{M}{N} < 1$. From now on, we adopt the conventions of chapter 2, i.e $M = M(N) < N$ and $K = K(N) < M$ are functions of N.

We denote by $\mathbf{Y}_N = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ the $M \times N$ observation matrix, which can be readily written as

$$\mathbf{Y}_N \triangleq \mathbf{A}\mathbf{S}_N + \mathbf{V}_N, \tag{4.1}$$

where $\mathbf{S}_N = [\mathbf{s}_1, \dots, \mathbf{s}_N]$ and $\mathbf{V}_N = [\mathbf{v}_1, \dots, \mathbf{v}_N]$. From this matrix, the empirical spatial correlation matrix of the observation is given by $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$, whereas the empirical spatial correlation matrix associated with the noiseless observation is given by $\frac{1}{N}\mathbf{A}\mathbf{S}_N\mathbf{S}_N^*\mathbf{A}^*$.

In order to simplify the notation in the subsequent exposition, we set $\mathbf{\Sigma}_N = N^{-1/2}\mathbf{Y}_N$, $\mathbf{B}_N = N^{-1/2}\mathbf{A}\mathbf{S}_N$ and $\mathbf{W}_N = N^{-1/2}\mathbf{V}_N$ so that (4.1) can be equivalently formulated as

$$\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{W}_N, \tag{4.2}$$

Under the assumption that $\mathbf{A}$ and $\mathbf{S}_N$ have full rank K, we retrieve the main properties of the information plus noise model defined in chapter 2 section 2.1 , namely

- $\mathbf{B}_N$ is a rank K deterministic matrix,
- $\mathbf{W}_N$ is a complex Gaussian matrix with i.i.d. entries having zero mean and variance σ^2/N.

Remark 4.1.1. *In the context of source localization (estimation of the direction of arrival of the K sources), a typical model for matrix* $\mathbf{A}$ *is given by* $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$ *where* $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ *are the angles of arrival of the K sources impinging on the array of receive antennas. The column* $\mathbf{a}(\theta_k)$ *is called in this context "steering vector" of the k-th source and depends on the angle of arrival* θ_k *and the geometry of the antennas. This model will be used in section 4.3 in the context of the so-called MUSIC algorithm.*

We will assume without loss of generality that the non null eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ have multiplicity one. We denote, as in chapter 2, by $\hat{\lambda}_{1,N} \leq \dots \leq \hat{\lambda}_{M,N}$ and $0 = \lambda_{1,N} = \dots = \lambda_{M-K,N} < \lambda_{M-K+1} < \dots < \lambda_{M,N}$ the respective eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ and $\mathbf{B}_N\mathbf{B}_N^*$. The associated eigenvectors are $\hat{\mathbf{u}}_{1,N}, \dots, \hat{\mathbf{u}}_{M,N}$ and $\mathbf{u}_{1,N}, \dots, \mathbf{u}_{M,N}$.

In the terminology of subspace estimation, we call "noise subspace" the subspace $\text{span}\{\mathbf{u}_{1,N}, \dots, \mathbf{u}_{M-K,N}\}$, i.e the eigenspace associated with eigenvalue 0 of $\mathbf{B}_N\mathbf{B}_N^*$ and "signal subspace" the orthogonal complement, i.e the eigenspace associated with the non null eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$. The goal of subspace estimation is to infer on one of these two subspaces, in terms of the projection matrices. Therefore, the purpose is to estimate the projection matrix on the noise subspace, i.e $\mathbf{\Pi}_N = \sum_{k=1}^{M-K} \mathbf{u}_{k,N}\mathbf{u}_{k,N}^*$. This estimation problem may involve a high number of parameters, if M, N are large, therefore one usually prefer to estimate bilinear forms of this projector, i.e quantities as $\mathbf{d}_{1,N}^*\mathbf{\Pi}_N\mathbf{d}_{2,N}$. By a classical polarization identity, this reduces to estimating any quadratic form of $\mathbf{\Pi}_N$. The subspace estimation problem we consider here is to find a consistent estimator of

$$\eta_N = \mathbf{d}_N^*\mathbf{\Pi}_N\mathbf{d}_N, \text{ when } N \to \infty, \tag{4.3}$$

where $(\mathbf{d}_N)$ represents a sequence of deterministic vectors such that $\sup_N \|\mathbf{d}_N\| < \infty$. Traditionally, η_N is estimated by $\hat{\eta}_N = \mathbf{d}_N^*\hat{\mathbf{\Pi}}_N\mathbf{d}_N = \sum_{k=1}^{M-K} \mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}\mathbf{d}_N$, i.e by replacing the eigenvectors of $\mathbf{B}_N\mathbf{B}_N^*$ with their empirical estimates. This estimator makes sense in the regime where M does not depend on N (thus $c_N \to_N 0$), because from the classical law of large numbers,

$$\left\| \mathbf{\Sigma}_N\mathbf{\Sigma}_N^* - (\mathbf{B}_N\mathbf{B}_N^* + \sigma^2\mathbf{I}_M) \right\| \xrightarrow[N\to\infty]{\text{a.s.}} 0.$$

However, the latter convergence is not true in general, if $c_N \to_N c > 0$ (see the general results in chapter 2, or the results concerning the spiked model chapter 3 section 3.3 for an immediate counterexample). In particular, it can be shown that $\eta_N - \hat{\eta}_N$ does not converge to 0.

The purpose of the next section is to provide a consistent estimate of η_N by using the results concerning the convergence of bilinear forms of the resolvent of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$, provided in chapter 2 theorem 2.2.3.

4.2 Generalized subspace estimation

In this section, we will make the two additionnal assumptions, basically expressing the fact that the eigenvalues associated with the noise subspace are separated from the eigenvalues associated with the signal subspace, for all large N. We refer the reader to chapter 2, for an exposition of function w_N and the characterization of the support $\mathcal{S}_N$ of measure μ_N.

Assumption A-3: *For all large N, $\lambda_{M-K+1,N} > w_N(x_{2,N}^-)$, i.e the non zero eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ are not associated with the first cluster $[x_{1,N}^-, x_{1,N}^+]$ of $\mathcal{S}_N$.*

Assumption A-4: *There exists $t_{i,N}^-, t_{i,N}^+ > 0$ ($i = 1,2$) such that*

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

It should be noticed that since $\sup_N x_{Q,N}^+ < \infty$ (chapter 2 section 2.4 theorem 2.4.1), we always can find $t_2^+ > 0$ satisfying Assumption **A-4**. Assumption **A-4** is mainly technical and the most important is Assumption **A-3**, whose purpose will be fully revealed during the derivation of a consistent estimator of η_N. Roughly speaking, it will allow not to take into account any contribution of the signal subspace.

4.2.1 Preliminary results

We first give the following useful result, which will be of constant use in the sequel, and which is proved in appendix 4.5.1.

Lemma 4.2.1. *Let $\mathcal{K} \subset \mathbb{C}\setminus\left([t_1^-, t_1^+] \cup [t_2^-, \infty)\right)$ a compact set. Then, under Assumption **A-4**,*

$$\begin{aligned} &\sup_{z\in\mathcal{K}} |\hat{m}_N(z) - m_N(z)| \xrightarrow[N\to\infty]{a.s.} 0, \\ &\sup_{z\in\mathcal{K}} \left|\hat{m}_N'(z) - m_N'(z)\right| \xrightarrow[N\to\infty]{a.s.} 0, \end{aligned} \tag{4.4}$$

and for $(\mathbf{d}_{1,N})$, $(\mathbf{d}_{2,N})$ as in theorem 2.2.3

$$\sup_{z\in\mathcal{K}} \left|\mathbf{d}_{1,N}^*(\mathbf{Q}_N(z) - \mathbf{T}_N(z))\mathbf{d}_{2,N}\right| \xrightarrow[N\to\infty]{a.s.} 0. \tag{4.5}$$

Finally, if $0 \notin \mathcal{K}$,

$$\sup_{z\in\mathcal{K}} \left| \frac{1}{1+\sigma^2 c_N \hat{m}_N(z)} - \frac{1}{1+\sigma^2 c_N m_N(z)} \right| \xrightarrow[N\to\infty]{a.s.} 0. \tag{4.6}$$

We now introduce certain new quantities. We define the $M \times M$ matrix

$$\hat{\mathbf{\Omega}}_N = \hat{\mathbf{\Lambda}}_N + \frac{\sigma^2 c_N}{M}\mathbf{1}\mathbf{1}^T, \tag{4.7}$$

where $\hat{\mathbf{\Lambda}}_N = \mathrm{Diag}(\hat{\lambda}_{1,N},\dots,\hat{\lambda}_{M,N})$ and $\mathbf{1} = [1,\dots,1]^T$. We denote by $\hat{\omega}_{1,N},\dots,\hat{\omega}_{M,N}$ the eigenvalues of $\hat{\mathbf{\Omega}}_N$ (in increasing order). With this definition, the solutions to the equation $1+\sigma^2 c_N \hat{m}_N(z) = 0$ are included in the set $\{\hat{\omega}_{1,N},\dots,\hat{\omega}_{M,N}\}$. This can be seen by writing the characteristic polynomial of $\hat{\mathbf{\Omega}}_N$, i.e

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

More precisely, if $\hat{\lambda}_{k,N}$ has multiplicity i, i.e. $\hat{\lambda}_{k-1,M} < \hat{\lambda}_{k,N} = \ldots = \hat{\lambda}_{k+i-1,N} < \hat{\lambda}_{k+i,N}$, then we have

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

Remark 4.2.1. *The eigenvalues $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$ have multiplicity 1 almost surely (see section 1.4.2). This readily implies that a.s. $\hat{\omega}_{1,N} < \ldots < \hat{\omega}_{M,N}$, and therefore the equation $1+\sigma^2 c_N \hat{m}_N(z) = 0$ has M solutions satisfying*

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

Consequently, from the previous remark, when the eigenvalues $\hat{\omega}_{1,N},\dots,\hat{\omega}_{M,N}$ occur in a statement where a set of probability one is used, one can always assume that they have multiplicity one. The main result concerning $\hat{\omega}_{1,N},\dots,\hat{\omega}_{M,N}$ is that they satisfy a similar separation property than in theorem 3.2.3 (chapter 3 section 3.2.3) Applying the result of theorem 3.2.3, Assumptions **A-3** and **A 4** immediately imply that with probability one,

$$\hat{\lambda}_{1,N},\dots,\hat{\lambda}_{M-K,N} \in [t_1^-, t_1^+] \quad \text{and} \quad \hat{\lambda}_{M-K+1,N},\dots,\hat{\lambda}_{M,N} \in [t_2^-, t_2^+], \tag{4.8}$$

for all large N. Therefore, we also have the following result.

Corollary 4.2.1. *Under Assumptions* **A-1**, **A-3** *and* **A-4**, *with probability one, it holds that*

$$\hat{\omega}_{1,N},\dots,\hat{\omega}_{M-K,N} \in [t_1^-, t_1^+] \quad \textit{and} \quad \hat{\omega}_{M-K+1,N} \geq t_2^- \tag{4.9}$$

for all large N.

Proof. Fix a realization in the probability one event $\Omega = \Omega_1 \cap \Omega_2$, with Ω_1 and Ω_2 the respective probability one event on which Theorem 3.2.3 and Lemma 4.2.1 hold. The interlacement of $\hat{\lambda}_{k,N}$ and $\hat{\omega}_{k,N}$ implies that

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

for all large N. Therefore, we only need to prove that $\hat{\omega}_{M-K,N} \leq t_1^+$ for all large N. Let $\delta > 0$ such that $t_1^- - \delta > 0$ and $t_1^+ + \delta < t_2^-$, $y > 0$ and

$$\mathscr{R} = \left\{u + \mathrm{i}v : u \in [t_1^- - \delta, t_1^+ + \delta], v \in [-y, y]\right\}.$$

We first establish that

$$\sup_{z\in\partial\mathscr{R}} \left|\hat{w}_N(z)^{-1} - w_N(z)^{-1}\right| \to_N 0. \tag{4.10}$$

From Lemma 4.2.1, $\sup_{z\in\partial\mathscr{R}} |\hat{w}_N(z) - w_N(z)| \to_N 0$ and thus we just need to prove that

$$\inf_N \inf_{z\in\partial\mathscr{R}} |w_N(z)| > 0. \tag{4.11}$$

Matrix $\mathbf{T}_N(z)$ can be written as

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

Therefore, it holds that

$$\|\mathbf{T}_N(z)\| = \frac{|1+\sigma^2 c_N m_N(z)|}{\min_{k=1,\dots,M} |\lambda_{k,N} - w_N(z)|}.$$

Since 0 is eigenvalue of $\mathbf{B}_N\mathbf{B}_N^*$, we get for $z \in \partial\mathscr{R}$

$$\begin{aligned}|1+\sigma^2 c_N m_N(z)| &= \min_{k=1,\ldots,M} |\lambda_{k,N} - w_N(z)| \|\mathbf{T}_N(z)\| \\ &\leq \|\mathbf{T}_N(z)\| |w_N(z)|\end{aligned}$$

and the bounds in property 2.3.1 and (2.8) imply

$$|w_N(z)| \geq \frac{\mathrm{dist}(z, \mathscr{S}_N)}{2},$$

which proves (4.11). From the properties of function w_N (see property 2.4.1 in chapter 2), the set $\mathscr{W}_{1,N} = \{w_N(z) : z \in \partial\mathscr{R}\}$ is a closed $\mathscr{C}^1$ path intersecting the real axis at points $w_N(t_1^- - \delta)$, $w_N(t_1^+ + \delta)$, enclosing the interval $(w_{1,N}^-, w_{1,N}^+)$ and leaving $[w_N(t_2^-), +\infty)$ outside, for all large N. Assumption **A-3** implies that 0 is the unique eigenvalue of $\mathbf{B}_N\mathbf{B}_N^*$ enclosed by $\mathscr{W}_{1,N}$ for all large N. As the contour $\mathscr{W}_{1,N} = w_N(\partial\mathscr{R})$ encloses 0, it holds that

$$1 = \frac{1}{2\pi \mathrm{i}} \oint_{\mathscr{W}_{1,N}^+} \lambda^{-1} \mathrm{d}\lambda = \frac{1}{2\pi\mathrm{i}} \oint_{\partial\mathscr{R}^+} \frac{w_N'(z)}{w_N(z)} \mathrm{d}z,$$

where the superscript + means that the contours are counterclockwise oriented. From (4.10) and Lemma 4.2.1, we have

$$\sup_{z\in\partial\mathscr{R}} \left| \frac{\hat{w}_N'(z)}{\hat{w}_N(z)} - \frac{w_N'(z)}{w_N(z)} \right| \xrightarrow[N\to\infty]{} 0,$$

where

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

Thus it follows that

$$\frac{1}{2\pi\mathrm{i}} \oint_{\partial\mathscr{R}^+} \frac{\hat{w}_N'(z)}{\hat{w}_N(z)} \mathrm{d}z \xrightarrow[N\to\infty]{} 1. \tag{4.12}$$

An elementary analysis of the function $\hat{w}_N$ shows that apart from $\hat{\omega}_{1,N}, \ldots, \hat{\omega}_{M,N}$, $\hat{w}_N(z)$ admits $M+1$ real additional zeros denoted $\hat{z}_{0,N}, \ldots, \hat{z}_{M,N}$ such that $\hat{z}_{0,N} \in (0, \hat{\lambda}_{1,N})$, $\hat{\lambda}_{M,N} < \hat{\omega}_{M,N} < \hat{z}_{M,N}$, and $\hat{\lambda}_{k,N} < \hat{\omega}_{k,N} < \hat{z}_{k,N} < \hat{\lambda}_{k+1,N}$. Since we have $\inf_N \inf_{z\in\partial\mathscr{R}} |w_N(z)| > 0$ and $\sup_{z\in\partial\mathscr{R}} |\hat{w}_N(z) - w_N(z)| \to_N 0$, it holds that $\hat{\omega}_{M-K,N}, \hat{z}_{M-K,N} \notin \partial\mathscr{R}$ for N large enough. The argument principle states that the integral in (4.12) is the number of zeros minus the number of poles of $\hat{w}_N$ (counting multiplicities) contained in $\mathscr{R}$. The poles of $\hat{w}_N$ in $\mathscr{R}$ are $\hat{\lambda}_{1,N}, \ldots, \hat{\lambda}_{M-K,N}$ (with multiplicity 2) and consequently $1 = \mathrm{card}\{z \in \mathscr{R} : \hat{w}_N(z) = 0\} - 2(M-K)$ for all large N. We already know that $\hat{\omega}_{k,N}, \hat{z}_{k,N} \in \mathscr{R}$ for $k = 1, \ldots, M-K-1$, and therefore 3 more zeros are contained in $\mathscr{R}$. These 3 zeros are necessarily $\hat{z}_{0,N}, \hat{\omega}_{M-K,N}$ and $\hat{z}_{M-K,N}$ because $\hat{\lambda}_{M-K+1} \geq t_2^-$ does not belong to $\mathscr{R}$. Since in the definition of $\mathscr{R}$, $\delta > 0$ can be made arbitrarily small, this concludes the proof. □

4.2.2 The general case

We are now in position to introduce a new consistent estimator of $\eta_N = \sum_{k=1}^{M-K} \mathbf{d}_N^* \mathbf{u}_{k,N} \mathbf{u}_{k,N}^* \mathbf{d}_N$ with $(\mathbf{d}_N)_{N\geq 1}$ a sequence of deterministic vector such that $\sup_N \|\mathbf{d}_N\| < \infty$.

Theorem 4.2.1. *Define*

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

where $\mathscr{R}$ is the rectangle $\mathscr{R} = \left\{u + \mathrm{i}v : u \in [t_1^- - \delta, t_1^+ + \delta], v \in [-y, y]\right\}$, with $\delta > 0$ such that $t_1^- - \delta > 0$ and $t_1^+ + \delta < t_2^-$, and $y > 0$. Then, under Assumptions **A-1**, **A-3** *and* **A-4**,

$$\hat{\eta}_{\mathrm{new},N} - \eta_N \xrightarrow[N\to\infty]{\mathrm{a.s}} 0,$$

and $\hat{\eta}_{\mathrm{new},N} = \sum_{k=1}^M \hat{\xi}_{k,N} \mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N$ with probability one for N large enough, where, for $k = 1, \ldots, M-K$,

$$\hat{\xi}_{k,N} = 1 + \frac{\sigma^2 c_N}{M} \sum_{l=M-K+1}^{M} \frac{\hat{\lambda}_{k,N} + \hat{\lambda}_{l,N}}{(\hat{\lambda}_{k,N} - \hat{\lambda}_{l,N})^2} + \sigma^2(1-c_N) \sum_{l=M-K+1}^{M} \left(\frac{1}{\hat{\lambda}_{k,N} - \hat{\lambda}_{l,N}} - \frac{1}{\hat{\lambda}_{k,N} - \hat{\omega}_{l,N}} \right), \tag{4.13}$$

and for $k = M-K+1,\dots,M$,

$$\hat{\xi}_{k,N} = -\frac{\sigma^2 c_N}{M}\sum_{l=1}^{M-K}\frac{\hat{\lambda}_{k,N}+\hat{\lambda}_{l,N}}{(\hat{\lambda}_{k,N}-\hat{\lambda}_{l,N})^2} - \sigma^2(1-c_N)\sum_{l=1}^{M-K}\left(\frac{1}{\hat{\lambda}_{k,N}-\hat{\lambda}_{l,N}} - \frac{1}{\hat{\lambda}_{k,N}-\hat{\omega}_{l,N}}\right). \tag{4.14}$$

Proof. From the properties of function w_N (see property 2.4.1 in chapter 2), the set $\mathscr{W}_{1,N} = \{w_N(z) : z \in \partial\mathscr{R}\}$ is a closed $\mathscr{C}^1$ path intersecting the real axis at points $w_N(t_1^- - \delta)$, $w_N(t_1^+ + \delta)$, enclosing the interval $(w_{1,N}^-, w_{1,N}^+)$ and leaving $[w_N(t_2^-), +\infty)$ outside, for all large N. Assumption **A-3** implies that 0 is the unique eigenvalue of $\mathbf{B}_N\mathbf{B}_N^*$ enclosed by $\mathscr{W}_{1,N}$ for all large N (note that these remarks have already been made in the proof of Corollary 4.2.1). Thus, from residue theorem

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

where $\mathscr{W}_{1,N}^-$ means that $\mathscr{W}_{1,N}$ is clockwise oriented. This leads directly to

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

Using the equality $(1+\sigma^2 c m_N(z))(\mathbf{B}_N\mathbf{B}_N^* - w_N(z)\mathbf{I}_M)^{-1} = \mathbf{T}_N(z)$, which follows easily from the definition in (2.6), we can write

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

with $g_N(z) = \mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N\frac{w_N'(z)}{1+\sigma^2 c m_N(z)}$. Now, the key point of the proof is based on the observation that $g_N(z)$ can be estimated consistently (and uniformly) from the elements of the sample covariance matrix $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. Indeed, we have

$$\sup_{z\in\partial\mathscr{R}}\left|\hat{g}_N(z) - g_N(z)\right| \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{4.15}$$

where

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

with $\hat{w}_N'(z)$ the derivative of $\hat{w}_N(z) = z(1+\sigma^2 c_N\hat{m}_N(z))^2 - \sigma^2(1-c_N)(1+\sigma^2 c_N\hat{m}_N(z))$. The convergence (4.15) is a straightforward consequence of lemma 4.2.1. This directly implies that,

$$\left|\frac{1}{2\pi\mathrm{i}}\oint_{\partial\mathscr{R}^-}\left(\hat{g}_N(z) - g_N(z)\right)\mathrm{d}z\right| \xrightarrow[N\to\infty]{\text{a.s.}} 0.$$

To conclude the proof, it remains to prove that $\hat{\eta}_{\text{new},N}$ defined in Theorem 4.2.1 verifies

$$\hat{\eta}_{\text{new},N} = \frac{1}{2\pi\mathrm{i}}\oint_{\partial\mathscr{R}^-}\hat{g}_N(z)\mathrm{d}z. \tag{4.16}$$

$\hat{g}_N$ is a rational function and its poles are

- $\hat{\lambda}_{1,N},\dots,\hat{\lambda}_{M,N}$ the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$,
- $\hat{\omega}_{1,N},\dots,\hat{\omega}_{M,N}$ the zeros of $z\mapsto 1+\sigma^2 c_N\hat{m}_N(z)$.

Consequently, from the location of the poles in (4.8) and (4.9), we obtain by the residue theorem

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

where $\mathrm{Res}(\hat{g}_N,\lambda)$ denotes the residue of function $\hat{g}_N$ at point λ and $\mathrm{Ind}_{\partial\mathscr{R}^-}(\lambda)$ the winding number of $\partial\mathscr{R}^-$ around λ. After tedious, but straightforward computations (see Appendix 4.5.2), we eventually check that (4.16) holds. □

The new consistent estimator introduced in Theorem 4.2.1 can be seen as an extension of the work of Mestre [51], which assumes that the useful signals are Gaussian mutually independent random i.i.d. sequences.

We remark that the consistent estimator $\hat{\eta}_{\text{new},N}$ is a linear combination of the terms $\mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N$ for $k = 1,\dots,M$ and in contrast to the traditional estimator $\hat{\eta}_{\text{trad},N} = \sum_{k=1}^{M-K} \mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N$, it contains contributions of both the noise subspace and the signal subspace of the sample covariance matrix.

We also note that Assumptions **A-3** and **A-4** are intuitively important because the various sums on the right hand side of (4.13) and (4.14) remain bounded thanks to (4.8) and (4.9): in (4.13) and (4.14), the terms $\left|\hat{\lambda}_{k,N} - \hat{\lambda}_{l,N}\right|$ and $\left|\hat{\lambda}_{k,N} - \hat{\omega}_{l,N}\right|$ are greater than $t_2^- - t_1^+$.

Remark 4.2.2. *It is worth pointing out that whenever the number of samples is forced to be much larger than the observation dimension ($N >> M$ or equivalently $c_N \to_N 0$), the proposed estimator converges to the classical sample eigenvector estimate $\hat{\eta}_N$. This can be readily seen by taking the limit as $c_N \to_N 0$ in the coefficients of* (4.13) *and* (4.14)*, and noticing that $\hat{\omega}_{l,N} \to_N \hat{\lambda}_{l,N}$ when $c_N \to_N 0$. Hence, as $c_N \to_N 0$, we have $\hat{\xi}_{k,N} \to_N 1$ for $k = 1,\dots,M-K$, and $\hat{\xi}_{k,N} \to_N 0$ for $k = M-K+1,\dots,M$, implying that $\hat{\eta}_{\text{new},N} - \hat{\eta}_{\text{trad},N} \to_N 0$. This shows that the proposed estimator is in fact a generalization of the classical sample eigenvector estimate.*

4.2.3 The spiked model case

In this section, we consider the special case of spiked models, already introduced in chapter 2 section 2.5 and whose main consequences have been given in chapter 3 section 3.3. Using Assumption **A-2**, we derive from theorem 4.2.1 a simplified estimator.

Theorem 4.2.2. *Let*

$$\hat{\eta}_{\text{spike},N} = \mathbf{d}_N^* \hat{\mathbf{\Pi}}_N \mathbf{d}_N + \sum_{k=M-K+1}^{M} \mathbf{d}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{d}_N \left(1 - \frac{\Gamma'(\hat{\lambda}_{k,N})}{\Gamma(\hat{\lambda}_{k,N}) m(\hat{\lambda}_{k,N})}\right), \tag{4.17}$$

where $\Gamma(x) = x m(x) \tilde{m}(x)$, and $m(x)$ is the Stieltjes transform of the Marcenko-Pastur law, defined in (2.15) *and $\tilde{m}(x) = c m(x) - \frac{1-c}{x}$. Then, under Assumption* **A-2**, *if $\lim_N \lambda_{M-K+1,N} = \gamma_1 > \sigma^2\sqrt{c}$, it holds that*

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

The result of theorem was also derived in [25], using a different method. The proof of theorem 4.2.2 relies on the following lemma (proved in appendix 4.5.3), which is similar to lemma 4.2.1.

Lemma 4.2.2. *Let*

$$\mathcal{K} \subset \mathbb{C} \setminus \left([0, \sigma^2(1+\sqrt{c})^2] \cup \{\psi(\gamma_1),\dots,\psi(\gamma_K)\} \right),$$

a compact set. Then under Assumption **A-2**, *if $\lim_N \lambda_{M-K+1,N} = \gamma_1 > \sigma^2\sqrt{c}$, then it holds that*

$$\sup_{z\in\mathcal{K}} |\hat{m}_N(z) - m(z)| \xrightarrow[N\to\infty]{\text{a.s.}} 0,$$
$$\sup_{z\in\mathcal{K}} \left|\hat{m}'_N(z) - m'(z)\right| \xrightarrow[N\to\infty]{\text{a.s.}} 0.$$

Moreover, if $\hat{m}_{t,N}(z) = \frac{1}{M}\sum_{k=1}^{M-K} \frac{1}{\hat{\lambda}_{k,N} - z}$ denotes the truncated sum associated with $\hat{m}_N(z)$, and $\mathcal{K} \subset \mathbb{C} \setminus [0, \sigma^2(1+\sqrt{c})^2]$ a compact set, then,

$$\sup_{z\in\mathcal{K}} \left|\hat{m}_{t,N}(z) - m(z)\right| \xrightarrow[N\to\infty]{\text{a.s.}} 0,$$
$$\sup_{z\in\mathcal{K}} \left|\hat{m}'_{t,N}(z) - m'(z)\right| \xrightarrow[N\to\infty]{\text{a.s.}} 0.$$

We now state that the $(\hat{\omega}_{M-K+k,N})_{k=0,\dots,K}$ have the same behaviour than the $(\hat{\lambda}_{M-K+k,N})_{k=0,\dots,K}$. More precisely:

Lemma 4.2.3. *Under Assumption* **A-2**, *if $\lim_N \lambda_{M-K+1,N} = \gamma_1 > \sigma^2\sqrt{c}$, we have*

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

for $k = 1,\dots,K$ and $\hat{\omega}_{M-K,N} \xrightarrow[N\to\infty]{\text{a.s.}} \sigma^2(1+\sqrt{c})^2$.

We now complete the proof of theorem 4.2.2. For this, we will use the formula of the estimator given in Theorem 4.2.1. Write

$$\sum_{k=1}^{M-K}\hat{\xi}_{k,N}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N = \mathbf{d}_N^*\hat{\mathbf{\Pi}}_N\mathbf{d}_N^* + \frac{\sigma^2 c_N}{M}\sum_{l=M-K+1}^{M}\sum_{k=1}^{M-K}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N\left(\frac{1}{\hat{\lambda}_{k,N}-\hat{\lambda}_{l,N}}+\frac{2\hat{\lambda}_{l,N}}{(\hat{\lambda}_{k,N}-\hat{\lambda}_{l,N})^2}\right)$$
$$+\sigma^2(1-c_N)\sum_{l=M-K+1}^{M}(\hat{\lambda}_{l,N}-\hat{\omega}_{l,N})\sum_{k=1}^{M-K}\frac{1}{(\hat{\lambda}_{k,N}-\hat{\lambda}_{l,N})(\hat{\lambda}_{k,N}-\hat{\omega}_{l,N})}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N. \tag{4.18}$$

From Theorem 3.3.1 and Lemma 4.2.3, for all $k=1,\ldots,K$, $\hat{\omega}_{M-K+k,N}$ and $\hat{\lambda}_{M-K+k,N}$ converge to $\psi(\gamma_k,c)$ and thus there exists $\epsilon>0$ such that $|\hat{\lambda}_{M-K,N}-\hat{\lambda}_{M-K+k,N}|\geq\epsilon$ and $|\hat{\lambda}_{M-K,N}-\hat{\omega}_{M-K+k,N}|\geq\epsilon$, w.p.1 for all large N. Consequently, we have the inequality

$$\left|\mathbf{d}_N^*\hat{\mathbf{\Pi}}_N\mathbf{d}_N^*-\sum_{k=1}^{M-K}\hat{\xi}_{k,N}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N\right|\leq\mathbf{d}_N^*\hat{\mathbf{\Pi}}_N\mathbf{d}_N\sum_{l=M-K+1}^{M}\left[\frac{\sigma^2c_N}{M}\left(\frac{1}{\epsilon}+\frac{2\hat{\lambda}_{l,N}}{\epsilon^2}\right)+\sigma^2(1-c_N)\frac{\left|\hat{\lambda}_{l,N}-\hat{\omega}_{l,N}\right|}{\epsilon^2}\right],$$

which implies

$$\mathbf{d}_N^*\hat{\mathbf{\Pi}}_N\mathbf{d}_N^*-\sum_{k=1}^{M-K}\hat{\xi}_{k,N}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N\xrightarrow[N\to\infty]{\text{a.s.}}0,$$

Next, we write

$$\sum_{k=M-K+1}^{M}\hat{\xi}_{k,N}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N=-\sigma^2c_N\sum_{k=M-K+1}^{M}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N\frac{1}{M}\sum_{l=1}^{M-K}\left(\frac{1}{\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}}+\frac{2\hat{\lambda}_{k,N}}{(\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N})^2}\right)$$
$$-\sigma^2(1-c_N)\sum_{k=M-K+1}^{M}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N\sum_{l=1}^{M-K}\left(\frac{1}{\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}}-\frac{1}{\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}}\right). \tag{4.19}$$

Using Lemma 4.2.2 and Theorem 3.3.1, it is easy to see that

$$\frac{1}{M}\sum_{l=1}^{M-K}\left(\frac{1}{\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}}+\frac{2\hat{\lambda}_{k,N}}{(\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N})^2}\right)\xrightarrow[N\to\infty]{\text{a.s.}}m(\psi(\gamma_k,c))+2\psi(\gamma_k,c)m'(\psi(\gamma_k,c)).$$

To handle the second term on the righthandside of (4.19), we first notice that $\hat{\omega}_{1,N},\ldots,\hat{\omega}_{M,N}$ are the eigenvalues of matrix $\hat{\mathbf{\Lambda}}_N+\frac{\sigma^2c_N}{M}\mathbf{1}\mathbf{1}^T$, with $\hat{\mathbf{\Lambda}}_N=\mathrm{Diag}(\hat{\lambda}_{1,N},\ldots,\hat{\lambda}_{M,N})$ and $\mathbf{1}=[1,\ldots,1]^T$. Using the matrix inversion lemma, we obtain for $x\notin\{\hat{\lambda}_{1,N},\ldots,\hat{\lambda}_{M,N},\hat{\omega}_{1,N},\ldots,\hat{\omega}_{M,N}\}$,

$$\begin{aligned}\frac{1}{M}\sum_{k=1}^{M}\frac{1}{\hat{\omega}_{k,N}-x}&=\frac{1}{M}\mathrm{Tr}\left(\left(\hat{\mathbf{\Lambda}}_N-x\mathbf{I}_M\right)^{-1}-\frac{\sigma^2c_N}{M}\frac{\left(\hat{\mathbf{\Lambda}}_N-x\mathbf{I}_M\right)^{-1}\mathbf{1}\mathbf{1}^T\left(\hat{\mathbf{\Lambda}}_N-x\mathbf{I}_M\right)^{-1}}{1+\sigma^2c_N\hat{m}_N(x)}\right)\\&=\hat{m}_N(x)-\frac{\sigma^2c_N}{M}\frac{\hat{m}_N'(x)}{1+\sigma^2c_N\hat{m}_N(x)}.\end{aligned} \tag{4.20}$$

Next, from Lemma 4.2.3 and Theorem 3.3.1, for $x\notin[\sigma^2(1-\sqrt{c})^2,\sigma^2(1-\sqrt{c})^2]\cup\{\psi(\lambda_1,c),\ldots,\psi(\lambda_K,c)\}$, w.p.1 for all large N, we have the following decomposition,

$$\sum_{l=1}^{M-K}\left(\frac{1}{\hat{\omega}_{l,N}-x}-\frac{1}{\hat{\lambda}_{l,N}-x}\right)=\sum_{l=M-K+1}^{M}\left(\frac{1}{\hat{\lambda}_{l,N}-x}-\frac{1}{\hat{\omega}_{l,N}-x}\right)+\left(\sum_{l=1}^{M}\frac{1}{\hat{\omega}_{l,N}-x}-\sum_{l=1}^{M}\frac{1}{\hat{\lambda}_{l,N}-x}\right). \tag{4.21}$$

From the limits of $\hat{\lambda}_{M-K+k,N}$ and $\hat{\omega}_{M-K+k,N}$ for $k=1,\ldots,K$, the first term of the righthanside of (4.21) satisfy

$$\sum_{l=M-K+1}^{M}\left(\frac{1}{\hat{\lambda}_{l,N}-x}-\frac{1}{\hat{\omega}_{l,N}-x}\right)\xrightarrow[N\to\infty]{\text{a.s.}}0.$$

From (4.20), the second term satisfies

$$\sum_{l=1}^{M}\left(\frac{1}{\hat{\omega}_{l,N}-x}-\frac{1}{\hat{\lambda}_{l,N}-x}\right)=-\sigma^2c_N\frac{\hat{m}_N'(x)}{1+\sigma^2c_N\hat{m}_N(x)}\xrightarrow[N\to\infty]{\text{a.s.}}-\sigma^2c\frac{m'(x)}{1+\sigma^2cm(x)}.$$

Let

$$\hat{h}_N(x) = \sum_{l=1}^{M-K}\left(\frac{1}{\hat{\omega}_{l,N}-x} - \frac{1}{\hat{\lambda}_{l,N}-x}\right) + \sigma^2 c\frac{m'(x)}{1+\sigma^2 cm(x)}.$$

With probability one, for N large enough, the sequence $(\hat{h}_N)$ is a normal family on $]\sigma^2(1+\sqrt{c})^2,\infty[$ by Montel's Theorem. Consequently, $\hat{h}_N \to_N 0$ uniformly on each compact subset of $]\sigma^2(1+\sqrt{c})^2,\infty[$. Getting back to (4.19), we finally end up with

$$\sum_{k=1}^{K}(\hat{\xi}_{M-K+k,N} - \alpha_k)\mathbf{d}_N^*\hat{\mathbf{u}}_{M-K+k,N}\hat{\mathbf{u}}_{M-K+k,N}^*\mathbf{d}_N \xrightarrow[N\to\infty]{} 0.$$

with probability one, where $\alpha_k = \sigma^4 c(1-c)\frac{m'(\psi(\gamma_k,c))}{1+\sigma^2 cm(\psi(\gamma_k,c))} - \sigma^2 c(m(\psi(\gamma_k,c)) + 2\psi(\gamma_k,c)m'(\psi(\gamma_k,c)))$.

From the Marcenko-Pastur canonical equation (2.15), we have

$$\tilde{m}(z) = \frac{-1}{z(1+\sigma^2 cm(z))} \quad \text{and} \quad m(z) = \frac{-1}{z(1+\sigma^2 c\tilde{m}(z))},$$

where $\tilde{m}(z) = cm(z) - \frac{1-c}{z}$. This implies that the identity

$$\psi(\gamma_k,c)\tilde{m}(\psi(\gamma_k,c))(1+\sigma^2 cm(\psi(\gamma_k,c))) = -1, \tag{4.22}$$

holds. Using twice (4.22) leads to

$$\alpha_k = 1 - \left(2\sigma^2 c\psi(\gamma_k,c)m'(\psi(\gamma_k,c)) + \sigma^4 c(1-c)\psi(\gamma_k,c)\tilde{m}(\psi(\gamma_k,c))m'(\psi(\gamma_k,c)) - \frac{1}{\psi(\gamma_k,c)\tilde{m}(\psi(\gamma_k,c))}\right).$$

We therefore end up with $\alpha_k = 1 - \frac{\beta_k}{\Gamma(\psi(\gamma_k,c))m(\psi(\gamma_k,c))}$ with

$$\beta_k = -\Gamma(\psi(\gamma_k,c))\left(-\sigma^4 c(1-c)m'(\psi(\gamma_k,c)) - 2\sigma^2 c\frac{m(\psi(\gamma_k,c))}{\Gamma(\psi(\gamma_k,c))}\psi(\gamma_k,c)m'(\psi(\gamma_k)) + \frac{m(\psi(\gamma_k,c))^2}{\Gamma(\psi(\gamma_k,c))^2}\right).$$

It is easy to see that for $x > \sigma^2(1+\sqrt{c})^2$, $\Gamma(x)w(x) = 1$, with $w(x) = x(1+\sigma^2 cm(x))^2 - \sigma^2(1-c)(1+\sigma^2 cm(x))$. Therefore, $\Gamma'(\psi(\gamma_k,c)) = -\frac{\Gamma(\psi(\gamma_k,c))w'(\psi(\gamma_k,c))}{w(\psi(\gamma_k,c))}$. From (4.22), we have

$$\frac{m(\psi(\gamma_k,c))}{\Gamma(\psi(\gamma_k,c))} = -(1+\sigma^2 cm(\psi(\gamma_k,c))), \tag{4.23}$$

which implies $\beta_k = \Gamma'(\psi(\gamma_k,c))$. Therefore, $\alpha_k = 1 - \frac{\Gamma'(\psi(\gamma_k,c))}{\Gamma(\psi(\gamma_k,c))m(\psi(\gamma_k,c))}$, which finally implies

$$\hat{\eta}_{\text{new}} - \hat{\eta}_{\text{spike}} \xrightarrow[N\to\infty]{\text{a.s.}} 0.$$

4.3 DoA estimation

In section 4.1, we introduced the $M\times N$ matrix of observations

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

where $\mathbf{S}_N$ represent the deterministic source signals, $\mathbf{V}_N$ the white gaussian noise and $\mathbf{A}$ the $M\times K$ matrix of steering vectors, containing the transmission coefficients between the K sources and the M receive antennas. As stated in remark 4.1.1, a typical model for matrix $\mathbf{A}$ is

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

where $\boldsymbol{\theta} = [\theta_1,\ldots,\theta_K]$ contains the angles of arrival of the K sources.

The classical source localization problem in signal processing consists in estimating the vector $\boldsymbol{\theta}$ from the N samples collected in the matrix $\mathbf{Y}_N$. The so-called subspace-based estimator of $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_K)^T$ relies on the observation that if matrices $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{S}_N$ have both full rank K, then the angles $\theta_1,\ldots,\theta_K$ are solutions [1] of the equation $\mathbf{a}(\theta)^*\mathbf{\Pi}_N\mathbf{a}(\theta)=0$, where we recall that $\mathbf{\Pi}_N$ represents the orthogonal projection matrix on the kernel of matrix $\mathbf{A}(\boldsymbol{\theta})\mathbf{S}_N\mathbf{S}_N^*\mathbf{A}(\boldsymbol{\theta})^*$. In this context, the MUSIC (MUltiple SIgnal Classification) algorithm, derived by Schmidt [38] consists in estimating for each θ the quadratic form $\eta_N(\theta)=\mathbf{a}(\theta)^*\mathbf{\Pi}_N\mathbf{a}(\theta)$ of $\mathbf{\Pi}_N$ by a certain term $\hat{\eta}_N(\theta)$, and then to estimate the K angles as the argument of the K most significant local minima of function $\theta\to\hat{\eta}_N(\theta)$.

This approach has been extensively developed in the usual regime where $N\to+\infty$ and M is fixed. As stated in section 4.1, $\eta_N(\theta)$ can be estimated consistently in this context, for each θ, by $\hat{\eta}_N(\theta)=\mathbf{a}(\theta)^*\hat{\mathbf{\Pi}}_N\mathbf{a}(\theta)$ with $\hat{\mathbf{\Pi}}_N$ the orthogonal projection matrix on the eigenspace associated to the $M-K$ smallest eigenvalues of the empirical covariance matrix $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$. It clearly holds that

$$\sup_{\theta\in[-\pi,\pi]}\left|\hat{\eta}_N(\theta)-\eta_N(\theta)\right|\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{4.24}$$

and (4.24) immediately implies that the corresponding estimators of the direction of arrivals are consistent. Of course, as stated in section 4.1, this convergence does not hold under the assumptions given in section 2.1.

In this section, we prove that the DoA estimates obtained from the improved subspace estimator introduced in theorem 4.2.1 are consistent as $N\to\infty$. We consider a uniform linear array of antennas the elements of which are located at half the wavelength. The steering vector $\mathbf{a}(\theta)$ is thus given by

$$\mathbf{a}(\theta)=\frac{1}{\sqrt{M}}\left[1,e^{\mathrm{i}\pi\sin(\theta)},\ldots,e^{\mathrm{i}(M-1)\pi\sin(\theta)}\right]^T, \tag{4.25}$$

for $\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]$. However, we will consider the model

$$\mathbf{a}(\theta)=\frac{1}{\sqrt{M}}\left[1,e^{\theta},\ldots,e^{\mathrm{i}(M-1)\theta}\right]^T, \tag{4.26}$$

for $\theta\in[-\pi,\pi]$, which is equivalent to (4.25), but simpler for the computations.

The consistency of the DoA requires to have a uniform consistency result as (4.24).

Proposition 4.3.1. *Under Assumptions* **A-1**, **A-3** *and* **A-4**, *and if* $\mathbf{a}(\theta)$ *is given by* (4.26),

$$\sup_{\theta\in[-\pi,\pi]}\left|\hat{\eta}_{\mathrm{new},N}(\theta)-\eta_N(\theta)\right|\xrightarrow[N\to\infty]{\text{a.s.}}0.$$

Remark 4.3.1. *Proposition 4.3.1 hold of course for more general function* $\mathbf{a}(\theta)$. *In particular, the proof of Proposition 4.3.1 can be easily adapted for any function* $\mathbf{a}(\theta):[-\pi,\pi]\to\mathbb{C}^M$ *such that*

$$\left\|\mathbf{a}(\theta)-\mathbf{a}(\theta')\right\|\le CN^r\left|\theta-\theta'\right|^s \text{ and } \sup_N\sup_{\theta\in[-\pi,\pi]}\|\mathbf{a}(\theta)\|. \tag{4.27}$$

with $s>0$, $r\ge0$ *and* $C>0$ *a positive constant.*

The uniform consistency in proposition 4.3.1 can be transfered to the angles estimates, as follows. In order to define the estimators of $\theta_1,\ldots,\theta_K$ properly, we need to assume that the number of sources K is constant with N, and we consider K disjoint intervals $\mathscr{I}_1,\ldots,\mathscr{I}_K$, such that $\theta_k\in\mathscr{I}_k$, and define for each k the estimator $\hat{\theta}_{k,N}$ of θ_k by $\hat{\theta}_{k,N}=\arg\min_{\theta\in\mathscr{I}_k}|\hat{\eta}_{\mathrm{new},N}(\theta)|$. We deduce the following result.

Theorem 4.3.1. *Under Assumptions* **A-3** *and* **A-4**, *and if* $K>0$ *is independent of* N, *then,*

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

for $k=1,\ldots,K$

Proposition 4.3.1 and Theorem 4.3.1 are proved in the following subsections.

[1] This is due to the fact that $\mathrm{span}\{\mathbf{a}(\theta_1),\ldots,\mathbf{a}(\theta_K)\}=\mathrm{span}\{\mathbf{u}_{M-K+1,N},\ldots,\mathbf{u}_{M,N}\}$, or equivalently that the orthogonal complement of the column space of matrix $\mathbf{A}$ coincides with the noise subspace. Note that the K angles may be unique solutions under certain assumptions on function $\theta\to\mathbf{a}(\theta)$.

4.3.1 Regularization of the spectrum

We first introduce a "regularization trick" which will be fundamental to prove Proposition 4.3.1. It will be shown in the next subsection that it is sufficient to establish that for each $\alpha > 0$ and for each $\theta \in [-\pi, \pi]$, $\mathbb{P}(|\hat{\eta}_{\text{new},N}(\theta) - \eta_N(\theta)| > \alpha)$ decreases fast enough towards 0. For this, a tempting choice is to use the Markov inequality, and to establish that the moments of $|\hat{\eta}_{\text{new},N}(\theta) - \eta_N(\theta)|$ decrease fast enough.

From theorem 4.2.1, we have

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

and we choose $y > 0$ and $\epsilon > 0$, $\epsilon < \frac{y}{3}$ small enough such that the rectangle $\mathscr{R}$ is given by

$$\mathscr{R} = \{u + \mathrm{i}v, 0 < t_1^- - 3\epsilon \le u \le t_1^+ + 3\epsilon < t_2^- - 3\epsilon, -y \le v \le y\}. \tag{4.28}$$

Since $\operatorname{dist}\left(\partial\mathscr{R}, \{(\hat{\lambda}_{k,N}, \hat{\omega}_{k,N})_{k=1,\dots,M}\}\right) > 3\epsilon$ with probability one for all large N (see (4.8) and (4.9)), the estimator is well-defined for N greater than a random integer, but this fact does not imply the existence of the moments of $|\hat{\eta}_{\text{new},N}(\theta)|$. In order to solve this technical problem, it is necessary to prove that the probability that at least one element of $\{\hat{\lambda}_{k,N}, \hat{\omega}_{k,N} : k = 1, \dots, M\}$ escapes from $[t_1^- - 2\epsilon, t_1^+ + 2\epsilon] \cup [t_2^- - 2\epsilon, t_2^+ + 2\epsilon]$ decreases at rate $\frac{1}{N^l}$ for any $l \in \mathbb{N}$, and that the moments of a convenient regularized version of $|\hat{\eta}_{\text{new},N}(\theta) - \eta_N(\theta)|$ converge fast enough towards 0.

In the following, we denote by $\mathscr{T}_\epsilon$ the set

$$\mathscr{T}_\epsilon = [t_1^- - \epsilon, t_1^+ + \epsilon] \cup [t_2^- - \epsilon, t_2^+ + \epsilon],$$

and define the events $\mathscr{A}_{1,N}$ and $\mathscr{A}_{2,N}$

$$\mathscr{A}_{1,N} = \{\exists k : \hat{\lambda}_{k,N} \notin \mathscr{T}_\epsilon\} \quad \text{and} \quad \mathscr{A}_{2,N} = \{\exists k : \hat{\omega}_{k,N} \notin \mathscr{T}_\epsilon\}. \tag{4.29}$$

From theorem 3.2.2, we know that

$$\mathbb{P}\left(\mathscr{A}_{1,N}\right) = \mathscr{O}\left(\frac{1}{N^l}\right), \tag{4.30}$$

for all $l \in \mathbb{N}$. A similar property also holds for the $\hat{\omega}_{k,N}$.

Lemma 4.3.1. *Under Assumptions* **A-1**, **A-3** *and* **A-4**, *it holds that*

$$\mathbb{P}\left(\mathscr{A}_{2,N}\right) = \mathscr{O}\left(\frac{1}{N^l}\right), \tag{4.31}$$

for all $l \in \mathbb{N}$.

Lemma 4.3.1 is proved in appendix 4.5.5. Using this result, we now introduce a regularization term, denoted χ_N, defined as follows. We consider a function $\varphi \in \mathscr{C}_c^\infty(\mathbb{R}, [0,1])$ satisfying

$$\varphi(\lambda) = \begin{cases} 1 & \text{for } \lambda \in \mathscr{T}_\epsilon \\ 0 & \text{for } \lambda \in \mathbb{R} \backslash \left([t_1^- - 2\epsilon, t_1^+ + 2\epsilon] \cup [t_2^- - 2\epsilon, t_2^+ + 2\epsilon]\right), \end{cases} \tag{4.32}$$

and define the random variable

$$\chi_N = \det \varphi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*) \det \varphi\left(\hat{\boldsymbol{\Omega}}_N\right), \tag{4.33}$$

which verifies $\mathbb{1}_{\mathscr{A}_N^c} \le \chi_N$ where $\mathscr{A}_N = \mathscr{A}_{1,N} \cup \mathscr{A}_{2,N}$. It turns out that, considered as a function of the real and imaginary part of the entries of $\mathbf{W}_N$, χ_N is a $\mathscr{C}^1$ function, and using Poincaré inequality, we establish in appendix 4.5.6

Lemma 4.3.2. *Under Assumptions* **A-1**, **A-3** *and* **A-4**, *if* $(\mathbf{d}_N)$ *is a sequence of uniformly bounded deterministic vectors, i.e.* $\sup_N \|\mathbf{d}_N\| < \infty$, *then, for each integer* l, *it holds that*

$$\sup_{z \in \partial\mathscr{R}_y} \mathbb{E}\left[\left|\mathbf{d}_N^* \left(\mathbf{Q}_N(z) \frac{\hat{w}_N'(z)}{1+\sigma^2 c_N \hat{m}_N(z)} - \mathbf{T}_N(z) \frac{w_N'(z)}{1+\sigma^2 c_N m_N(z)}\right) \mathbf{d}_N \chi_N^2\right|^{2l}\right] \le \frac{C}{N^l} \tag{4.34}$$

where the constant C *does not depend on the sequence* $(\mathbf{d}_N)$.

The above mentioned property eventually allow to prove the uniform consistency of estimator $\hat{\eta}_{\text{new},N}(\theta)$.

4.3.2 Uniform consistency of the subspace estimate

We now handle the proof of Theorem 4.3.1, and when function $\theta \to \mathbf{a}(\theta)$ satisfies the conditions (4.27). We recall that from the bounds (2.7), (2.8) and property 2.3.1, we have

$$\sup_{z\in\partial\mathscr{R}_y} \|\mathbf{T}_N(z)\|, \sup_{z\in\partial\mathscr{R}_y} \left|\frac{1}{1+\sigma^2 c_N m_N(z)}\right|, \sup_{z\in\partial\mathscr{R}_y} \left|\frac{w_N'(z)}{1+\sigma^2 c_N m_N(z)}\right| \le C, \tag{4.35}$$

where $C>0$ is independent of N. Moreover, since $\hat{m}_N(z), \|\mathbf{Q}_N(z)\| \le \operatorname{dist}\left(z, \{\hat{\lambda}_{1,N},\ldots,\hat{\lambda}_{M,N}\}\right)^{-1}$ and from the very definition of event $\mathscr{A}_{2,N}$, it also holds

$$\sup_{z\in\partial\mathscr{R}_y} \|\mathbf{Q}_N(z)\|\chi_N, \sup_{z\in\partial\mathscr{R}_y} \left|\frac{\chi_N}{1+\sigma^2 c_N \hat{m}_N(z)}\right|, \sup_{z\in\partial\mathscr{R}_y} \left|\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N \hat{m}_N(z)}\right|\chi_N \le C. \tag{4.36}$$

We consider the set

$$\vartheta_N = \left\{-\pi + \frac{2(k-1)\pi}{N^2} : k=1,\ldots,N^2\right\},$$

whose elements are denoted $\nu_{k,N}$ for $k=1,\ldots,N^2$, and remark that for each $\theta\in[-\pi,\pi]$ and for each N, there exists an integer k_N such that $|\theta-\nu_{k_N,N}| \le \frac{2\pi}{N^2}$. For each $\theta\in[-\pi,\pi]$, it holds that

$$\hat{\eta}_{\mathrm{new},N}(\theta) - \eta_N(\theta) = \left[\hat{\eta}_{\mathrm{new},N}(\theta) - \hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N})\right] + \left[\hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N}) - \eta_N(\nu_{k_N,N})\right] + \left[\eta_N(\nu_{k_N,N}) - \eta_N(\theta)\right]. \tag{4.37}$$

It is easy to check that the third term of the r.h.s. of (4.37) satisfies

$$\sup_{\theta\in[-\pi,\pi]} \left|\eta_N(\nu_{k_N,N}) - \eta_N(\theta)\right| \le 2 \sup_{\theta\in[-\pi,\pi]} \left\|\mathbf{a}(\theta) - \mathbf{a}(\nu_{k_N,N})\right\| = \mathscr{O}\left(\frac{1}{N}\right). \tag{4.38}$$

In order to evaluate the behaviour of the supremum over θ of the first term of the r.h.s. of (4.37), we prove that for each $\alpha>0$,

$$\mathbb{P}\left(\sup_{\theta\in[-\pi,\pi]} |\hat{\eta}_{\mathrm{new},N}(\theta) - \hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N})| > \alpha\right) = \mathscr{O}\left(\frac{1}{N^{1+\beta}}\right),$$

where $\beta>0$. We first remark that for each $l\in\mathbb{N}$, it holds that

$$\begin{aligned}
\mathbb{P}\left(\sup_{\theta\in[-\pi,\pi]} \left|\hat{\eta}_{\mathrm{new},N}(\theta) - \hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N})\right| > \alpha\right) &\le \mathbb{P}\left(\sup_{\theta\in[-\pi,\pi]} \left|\hat{\eta}_{\mathrm{new},N}(\theta) - \hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N})\right| \mathbb{1}_{\mathscr{A}_N^c} > \alpha\right) + \mathbb{P}\left(\mathscr{A}_N\right) \\
&\le \frac{1}{\alpha^l}\mathbb{E}\left[\sup_{\theta\in[-\pi,\pi]} \left|\hat{\eta}_{\mathrm{new},N}(\theta) - \hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N})\right|^l \mathbb{1}_{\mathscr{A}_N^c}\right] + \mathscr{O}\left(\frac{1}{N^l}\right).
\end{aligned}$$

Moreover,

$$\left|\hat{\eta}_{\mathrm{new},N}(\theta) - \hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N})\right|^l \mathbb{1}_{\mathscr{A}_N^c} \le C \oint_{\partial\mathscr{R}_y^-} \left|\left(\mathbf{a}(\theta) - \mathbf{a}(\nu_{k_N,N})\right)^* \mathbf{Q}_N(z) \frac{\hat{w}_N'(z)}{1+\sigma^2 c_N \hat{m}_N(z)} \mathbf{a}(\nu_{k_N,N})\right|^l \mathbb{1}_{\mathscr{A}_N^c} |\mathrm{d}z|.$$

The inequalities (4.36) and $\mathbb{1}_{\mathscr{A}_N^c} \le \chi_N$ imply that

$$\sup_{z\in\partial\mathscr{R}_y} \|\mathbf{Q}_N(z)\| \left|\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N \hat{m}_N(z)}\right| \mathbb{1}_{\mathscr{A}_N^c} < C, \tag{4.39}$$

for some constant term C. Inequality (4.38) thus implies that

$$\sup_{\theta\in[-\pi,\pi]} \left|\left(\mathbf{a}(\theta) - \mathbf{a}(\nu_{k_N,N})\right)^* \mathbf{Q}_N(z) \frac{\hat{w}_N'(z)}{1+\sigma^2 c_N \hat{m}_N(z)} \mathbf{a}(\nu_{k_N,N})\right|^l \mathbb{1}_{\mathscr{A}_N^c} \le \frac{C}{N^l}$$

thus showing that

$$\mathbb{P}\left(\sup_{\theta\in[-\pi,\pi]} \left|\hat{\eta}_{\mathrm{new},N}(\theta) - \hat{\eta}_{\mathrm{new},N}(\nu_{k_N,N})\right| > \alpha\right) = \mathscr{O}\left(\frac{1}{N^l}\right)$$

for each integer l. Borel-Cantelli's lemma eventually implies that

$$\sup_{\theta\in[-\pi,\pi]} \left|\hat{\eta}_{\text{new},N}(\theta)-\hat{\eta}_{\text{new},N}(\nu_{k_N,N})\right| \to 0$$

almost surely.

We finally study the supremum of the second term of (4.37).

Let $\alpha > 0$, then

$$\begin{aligned}\mathbb{P}\left(\sup_{\theta\in[-\pi,\pi]} \left|\hat{\eta}_{\text{new},N}(\nu_{k_N,N})-\eta_N(\nu_{k_N,N})\right| > \alpha\right) &\le \mathbb{P}\left(\sup_{k=1,\dots,N^2} \left|\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right| > \alpha\right)\\ &\le \sum_{k=1}^{N^2} \mathbb{P}\left(\left|\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right| > \alpha\right)\\ &\le \sum_{k=1}^{N^2} \left[\mathbb{P}\left(\left\{\left|\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right| > \alpha\right\}\cap\mathscr{A}_N^c\right)\right] + \mathcal{O}\left(\frac{1}{N^l}\right)\end{aligned}$$

for each integer l. We now introduce in the above term the regularization term $\chi_N = \det\,\varphi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)\det\,\varphi(\hat{\mathbf{\Omega}}_N)$ defined in (4.33). As χ_N is equal to 1 on $\mathscr{A}_N^c$, it holds that

$$\begin{aligned}\mathbb{P}\left(\left\{\left|\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right| > \alpha\right\}\cap\mathscr{A}_N^c\right) &= \mathbb{P}\left(\left\{\left|\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right|\chi_N^2 > \alpha\right\}\cap\mathscr{A}_N^c\right)\\ &\le \mathbb{P}\left(\left|\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right|\chi_N^2 > \alpha\right)\\ &\le \frac{1}{\alpha^{2l}}\mathbb{E}\left|\left(\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right)\chi_N^2\right|^{2l}.\end{aligned}$$

The introduction of χ_N is in part motivated by the observation that the moments of $\hat{\eta}_{\text{new},N}(\nu_{k,N})\chi_N^2$. are finite. Moreover, it holds that

$$\begin{aligned}&\mathbb{E}\left|\left(\hat{\eta}_{\text{new},N}(\nu_{k,N})-\eta_N(\nu_{k,N})\right)\chi_N^2\right|^{2l}\\ &\le C\oint_{\partial\mathscr{R}_y^-}\mathbb{E}\left[\left|\mathbf{a}(\nu_{k,N})^*\left(\mathbf{Q}_N(z)\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N\hat{m}_N(z)}-\mathbf{T}_N(z)\frac{w_N'(z)}{1+\sigma^2 c_N m_N(z)}\right)\mathbf{a}(\nu_{k,N})\chi_N^2\right|^{2l}\right]|\mathrm{d}z| \qquad (4.40)\end{aligned}$$

and the proof of proposition 4.3.1 follows by applying lemma 4.3.2 to the above expression.

4.3.3 Consistency of the angles estimates

We now adress the proof of theorem 4.3.1, by considering the model of steering vectors

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

for $\theta \in [-\pi,\pi]$, which is equivalent to the model (4.25). Recall that we assumed the number of sources K is fixed, i.e. that K does not scale with N. In other words, model $\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{W}_N$ corresponds to a finite rank perturbation of the complex Gaussian i.i.d. matrix $\mathbf{W}_N$.

Remark 4.3.2. *In this context, we derived in section 4.2.3 an alternative consistent estimator, $\hat{\eta}_{\text{spike},N}(\theta)$ of $\eta_N(\theta)$. However, as we will in section 4.4, estimator $\hat{\eta}_{\text{new},N}(\theta)$ always leads in practice to the same performance as $\hat{\eta}_{\text{spike},N}(\theta)$ if $\frac{K}{M} << 1$ but outperforms $\hat{\eta}_{\text{spike},N}(\theta)$ for greater values of $\frac{K}{M}$. Therefore, the use of estimator $\hat{\eta}_{\text{new},N}(\theta)$ appears in practice more relevant than $\hat{\eta}_{\text{spike},N}(\theta)$.*

In order to establish the proposition, we follow a classical approach initiated by Hannan [26] to study sinusoid frequency estimates. The approach is based on the following result whose proof is straightforward.

Lemma 4.3.3. *Let (α_M) a real-valued sequence of a compact subset of $(-0.5, 0.5]$, and converging to α as $M\to\infty$. Define $q_M(\alpha_M) = \frac{1}{M}\sum_{k=1}^M e^{-i2\pi k\alpha_M}$. If $\alpha\neq 0$ or if $\alpha = 0$ and $M|\alpha_M|\to\infty$, then $q_M(\alpha_M)\to 0$. If $\alpha = 0$ and $M\alpha_M \xrightarrow[M\to\infty]{} \beta\in\mathbb{R}$, then $q_M(\alpha_M)\to e^{i\frac{\beta}{2}}\operatorname{sinc}\left(\frac{\beta}{2}\right)$.*

We denote by $\mathbf{A}$ the matrix $\mathbf{A}(\boldsymbol{\theta})$ corresponding the true angles $\boldsymbol{\theta} = (\theta_1,\dots,\theta_K)^T$. It is clear that $\eta_N(\theta) = 1 - \mathbf{a}(\theta)^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{a}(\theta)$. By the very definition of $\hat{\theta}_{k,N}$, $|\hat{\eta}_{\mathrm{new},N}(\hat{\theta}_{k,N})| \le |\hat{\eta}_{\mathrm{new},N}(\theta_k)|$. From proposition 4.3.1 and the equality $\eta_N(\theta_k) = 0$, we have $|\hat{\eta}_{\mathrm{new},N}(\hat{\theta}_{k,N})| \to 0$ w.p.1., as $N \to \infty$. Consequently,

$$\begin{aligned}|\eta_N(\hat{\theta}_{k,N})| &\le |\eta_N(\hat{\theta}_{k,N}) - \hat{\eta}_{\mathrm{new},N}(\hat{\theta}_{k,N})| + |\hat{\eta}_{\mathrm{new},N}(\hat{\theta}_{k,N})| \\ &\le \sup_{\theta\in[-\pi,\pi]} |\eta_N(\theta) - \hat{\eta}_{\mathrm{new},N}(\theta)| + |\hat{\eta}_{\mathrm{new},N}(\hat{\theta}_{k,N})| \\ &\xrightarrow[N\to\infty]{a.s.} 0. \end{aligned} \tag{4.41}$$

From Lemma 4.3.3, $(\mathbf{A}^*\mathbf{A})^{-1}$ converges to $\mathbf{I}_K$ as $N \to \infty$. Since $(\hat{\theta}_{k,N})$ is bounded, we can extract a converging subsequence $\left(\hat{\theta}_{k,\varphi(N)}\right)$. Let $\alpha_N = \hat{\theta}_{k,\varphi(N)} - \theta_k$. From Lemma 4.3.3, if $\alpha_N \to \alpha \neq 0$ as $N \to \infty$, then

$$\mathbf{a}(\hat{\theta}_{k,\varphi(N)})^*\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{a}(\hat{\theta}_{k,\varphi(N)}) \xrightarrow[N\to\infty]{a.s.} 0, \tag{4.42}$$

and thus $\eta_N(\hat{\theta}_{k,\varphi(N)}) \to 1$, a contradiction with (4.41). This implies that the whole sequence $(\hat{\theta}_{k,N})$ converges towards θ_k. If $N|\hat{\theta}_{k,N} - \theta_k|$ is not bounded, we can extract a subsequence such that $N|\hat{\theta}_{k,\phi(N)} - \theta_k| \to +\infty$ and Lemma 4.3.3 again implies that (4.42) holds, a contradiction. $N|\hat{\theta}_{k,N} - \theta_k|$ is thus bounded, and we consider a subsequence such that $N(\hat{\theta}_{k,\varphi(N)} - \theta_k) \to \beta$ where $\beta \in [-\pi,\pi]$. From Lemma 4.3.3, if $\beta \neq 0$, we get

$$\eta_{\varphi(N)}(\hat{\theta}_{k,\varphi(N)}) \xrightarrow[N\to\infty]{a.s.} 1 - \mathrm{sinc}\left(\frac{\beta}{2}\right)^2 > 0,$$

which is again in contradiction with (4.41). Therefore, $\beta = 0$ and all converging subsequences of $\left(N|\hat{\theta}_{k,\varphi(N)} - \theta_k|\right)$ converge to 0, which of course implies that the whole sequence $(N|\hat{\theta}_{k,N} - \theta_k|)$ converges to 0. We finally end up with $N(\hat{\theta}_{k,N} - \theta_k) \to 0$ w.p.1., as $N \to \infty$.

4.4 Discussion and numerical examples

We consider a uniform linear array of antennas the elements of which are located at half the wavelength whose steering vector model $\mathbf{a}(\theta)$ is given by (4.25). In the following numerical experiments, source signals are realizations of mutually independent unit variance AR(1) sequences with correlation coefficient 0.9. The SNR is defined as $10\log(\sigma^{-2})$. The additive noise varies from trials to trials.

Comparison with the traditional estimator and the unconditional estimator We first compare the results provided by the traditional subspace estimate, the new estimate $\hat{\eta}_{\mathrm{new},N}(\theta)$ defined as $\hat{\eta}_{\mathrm{new},N}$ for $\mathbf{d}_N = \mathbf{a}(\theta)$ (referred to in the figure as the "conditional estimator"), and the improved estimate of [51] derived under the assumption that the source signals are i.i.d. sequences (referred to as the "unconditional estimator"), and denoted $\hat{\eta}_{\mathrm{unc},N}(\theta)$. The corresponding angular estimates (defined as the preimages of the K deepest local minima of the estimated localization function) are denoted respectively "Trad-MUSIC", "Conditional G-MUSIC" and "Unconditional G-MUSIC" in this section. We refer as "separation condition" the property that the eigenvalue 0 of matrix $\mathbf{B}_N\mathbf{B}_N^*$ is separated from the clusters corresponding to its non zero eigenvalues, i.e. for each σ^2, M and N, $0 < w_{1,N}^+ < w_{2,N}^- < \lambda_{M-K+1,N}$.

We mention that the estimate of [51] is supposed to be unconsistent in the context of the following experiments because the source signals are not i.i.d. sequences. However, we will see that the performance of the conditional and the unconditional estimates are quite close, a property which will need further work (see the remarks below).

- In experiment 1, we consider two closely spaced sources, such that $\theta_1 = 16°$ and $\theta_2 = 18°$. The number of antennas is $M = 20$ and the number of snapshots is $N = 40$. The separation condition is verified if the SNR is larger than 10 dB. In order to evaluate the performance of the estimates of the localization function, for each improved estimator (conditional and unconditional), we plot versus θ in figure 4.1 the ratio of the MSE of the traditional estimator of $\mathbf{a}(\theta)^H\mathbf{\Pi}\mathbf{a}(\theta)$, i.e. $\hat{\eta}_{\mathrm{trad},N}(\theta) = \mathbf{a}(\theta)^H\hat{\mathbf{\Pi}}\mathbf{a}(\theta)$ over the MSE of the improved estimator $\hat{\eta}_{\mathrm{new},N}(\theta)$. The SNR is equal to 16 dB. Figure 4.1 shows that the two improved estimates have nearly the same performance, and that they outperform significantly the traditional approach around the 2 angles. We however notice that the three estimates have nearly the same performance if θ is far away from $\theta_1 = 16°$ and $\theta_2 = 18°$. In order to evaluate more precisely the improvements provided by the conditional and the unconditional estimators around θ_1 and θ_2, we plot the quantities $\frac{1}{2}\sum_{k=1}^{2}\mathbb{E}|\hat{\eta}_{\mathrm{trad},N}(\theta_k) - \eta_N(\theta_k)|^2$, $\frac{1}{2}\sum_{k=1}^{2}\mathbb{E}|\hat{\eta}_{\mathrm{new},N}(\theta_k) - \eta_N(\theta_k)|^2$ and $\frac{1}{2}\sum_{k=1}^{2}\mathbb{E}|\hat{\eta}_{\mathrm{unc},N}(\theta_k) - \eta_N(\theta_k)|^2$ vs SNR (note that $\eta_N(\theta_k) = 0$), in figure 4.2.

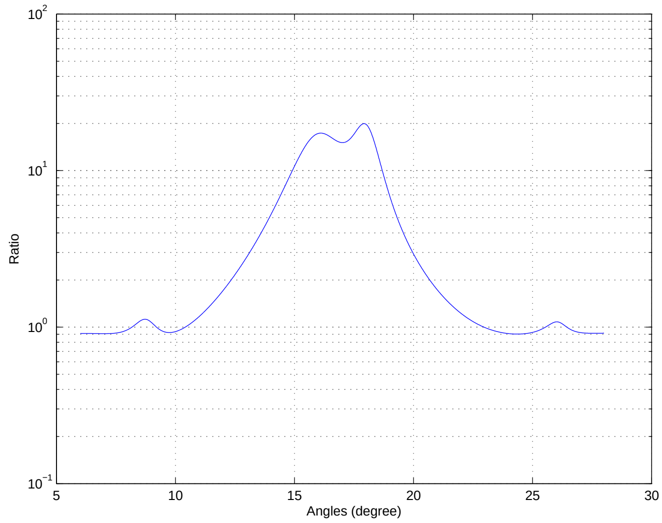


Figure 4.1: Ratio (in dB) of the MSE of $\hat{\eta}_{\mathrm{trad},N}(\theta)$ over the MSE of $\hat{\eta}_{\mathrm{new},N}(\theta)$ vs angles θ.

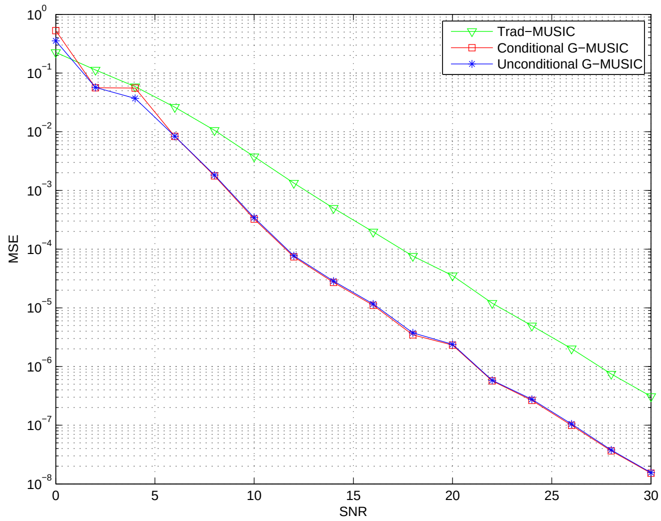


Figure 4.2: Mean of the MSE of the estimates of $\eta_N(\theta_1) = \mathbf{a}(\theta_1)^H \mathbf{\Pi a}(\theta_1)$ and $\eta_N(\theta_2) = \mathbf{a}(\theta_2)^H \mathbf{\Pi}_N \mathbf{a}(\theta_2)$.

In figure 4.3, we plot for each method the mean of the MSE of the two estimated angles versus the SNR, i.e $0.5(\mathbb{E}|\hat{\theta}_1 - \theta_1|^2 + \mathbb{E}|\hat{\theta}_2 - \theta_2|^2)$, where $\hat{\theta}_{1,2}$ denote the estimated angles. The estimates of θ_1 and θ_2 are defined as usual by taking the arguments of the two deepest local minima of the estimated localization function. The mean of the two Cramer-Rao bounds is also represented. The performance of the two improved estimates are again quite similar, and they provide an improvement of 4 dB w.r.t the traditional estimator in the range 15dB-25dB. We now plot the probability of outlier, defined here as the probability that one of the two estimated angles is separated from the true one by more than half of the separation between the two true sources. In figure 4.4, we compare the outlier probability of the three approaches versus the SNR of the three estimators. For a target probability of error of 0.5, the two improved estimators provide a gain of 8 dB over the traditional estimate. We finally evaluate the influence of M and N on the performance. N varies from 20 to 200 while the ratio c_N is kept constant to 0.5, and SNR = 15 dB. In figure 4.5 we have plotted the mean of the MSEs on the estimates of $\eta_N(\theta_k) = \mathbf{a}(\theta_k)^H \mathbf{\Pi a}(\theta_k)$ for $i = 1,2$. The separation condition occurs for $N \geq 32$. Figure 4.5 illustrates clearly the unconsistency of that the traditional estimate.

- In experiment 2, we now assume that the number of sources K is of the same order of magnitude that M and N, i.e. $K = 10$, $M = 20$, $N = 40$. The ten angles $(\theta_k)_{k=1,\ldots,10}$ are equal to $\theta_k = -40° + (k-1)10°$ for $k = 1,\ldots,10$. The separation condition holds if SNR is greater than 15 dB. We again plot versus θ in figure 4.6 the ratio of

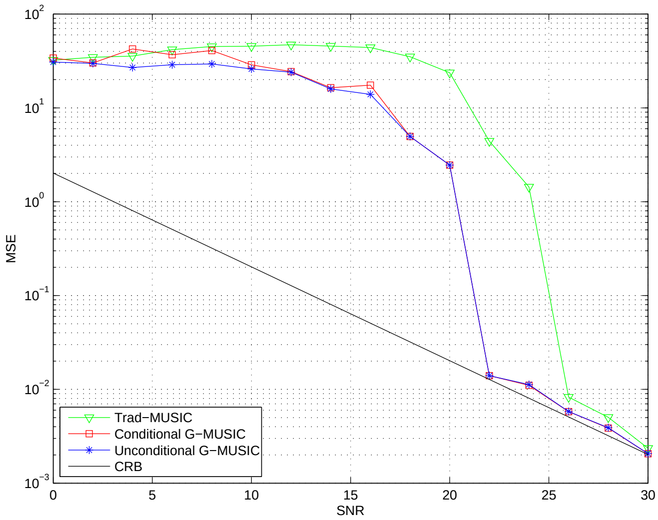


Figure 4.3: Mean of the MSE of the angles estimates versus SNR

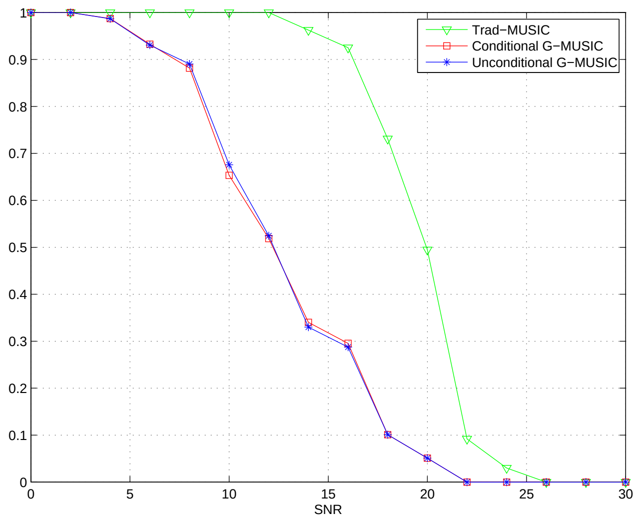


Figure 4.4: Outlier Probability vs the SNR

the MSE of the traditional estimator of the localization function over the MSE of its conditional and unconditional estimators. SNR is equal to 16 dB. Figure 4.6 shows again that the performance improvement of the conditional and unconditional estimates is optimum around the angles $(\theta_k)_{k=1,\dots,10}$. Figure 4.7 represents the mean of the MSEs of the various estimates of $\eta_N(\theta_k) = \mathbf{a}(\theta_k)^H \mathbf{\Pi} \mathbf{a}(\theta_k)$ for $k = 1, \dots, 10$ w.r.t. the SNR, and confirms the superiority of the two improved estimates when the separation condition holds.

All the previous plots clearly show that the conditional estimator outperforms the traditional one, while its difference with the unconditional one is negligible. This is a quite surprising fact. To explain this, we recall that the unconditional estimator has been derived in [51] under the assumption that matrix $\mathbf{S}_N$ is a Gaussian matrix with unit variance i.i.d. entries. The unconditional estimator of [51] is based on the observation that if $\mathbf{S}_N$ is an i.i.d. Gaussian matrix, then the entries of $(\mathbf{\Sigma}_N \mathbf{\Sigma}_N^* - z\mathbf{I}_M)^{-1}$ have the same behaviour as the entries of matrix $\mathbf{T}_{\mathrm{iid},N}(z)$ defined by the following equation

$$
\begin{aligned}
m_{\mathrm{iid},N}(z) &= \frac{1}{M} \mathrm{Tr}\, \mathbf{T}_{\mathrm{iid},N}(z), \\
\mathbf{T}_{\mathrm{iid},N}(z) &= \\
&\left[\left(\mathbf{A}\mathbf{A}^* + \sigma^2 \mathbf{I}_M \right) (1 - c_N - c_N z m_{\mathrm{iid},N}(z)) - z\mathbf{I}_M \right]^{-1}.
\end{aligned}
$$

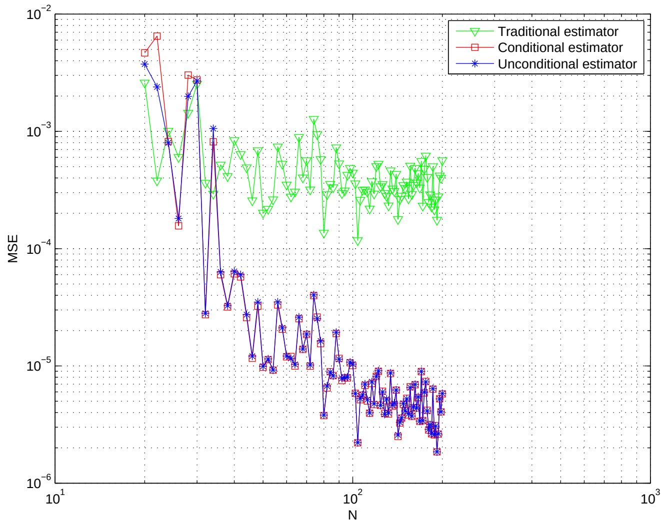


Figure 4.5: MSE for the estimators of the localization function vs N

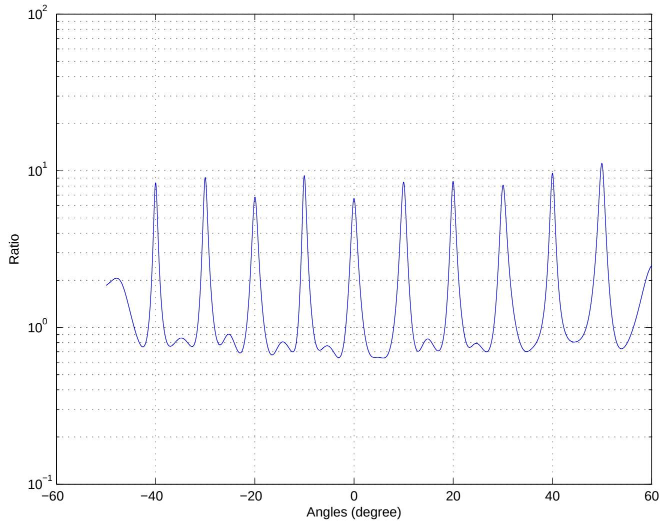


Figure 4.6: Ratio (in dB) of the MSE of the traditional estimate of the localization function over the MSE of its improved estimates versus θ

One can verify that the entries of $\mathbf{T}_N(z)$ defined by (2.6), which depend on $\mathbf{S}_N$, have the same asymptotic behaviour as the entries of $\mathbf{T}_{\text{iid},N}(z)$ when $\mathbf{S}_N$ is a realization of an i.i.d. matrix. In this case, the conditional and unconditional estimators have of course the same behaviour. If however $\mathbf{S}_N$ is not an i.i.d. matrix, then the entries of $(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^* - z\mathbf{I}_M)^{-1}$ do not behave like the entries of $\mathbf{T}_{\text{iid},N}(z)$ so that the unconditional estimator should become unconsistent. The previous simulation results tend to indicate that it is not the case. The explanation of this phenomenon is a topic for further research.

Comparison with the estimator (4.17) We now present some numerical comparisons between the subspace estimators $\hat{\eta}_{\text{new},N}(\theta)$ and $\hat{\eta}_{\text{spike},N}(\theta)$ defined by (4.17) for $\mathbf{d}_N = \mathbf{a}(\theta)$.

- We consider again the parameters of experiment 1 above, i.e. we consider two sources located at $\theta_1 = 16°$ and $\theta_1 = 18°$. The number of antennas is $M = 20$ and the number of snapshots is $N = 40$. In figure 4.8, we evaluate by Monte-Carlo simulations the quantity $0.5(\mathbb{E}|\hat{\theta}_1 - \theta_1|^2 + \mathbb{E}|\hat{\theta}_2 - \theta_2|^2)$, which is the mean of the MSE of the two estimated angles, versus the SNR. The performance of the SPIKE-MUSIC and G-MUSIC estimators are very close. The mean of the two Cramer-Rao bounds is also represented. In figure 4.9, we compute by Monte-Carlo simulations $\frac{1}{2}\sum_{k=1}^{2}\mathbb{E}|\hat{\eta}(\theta_k) - \eta(\theta_k)|^2$, i.e the mean over the MSE of the localization function,

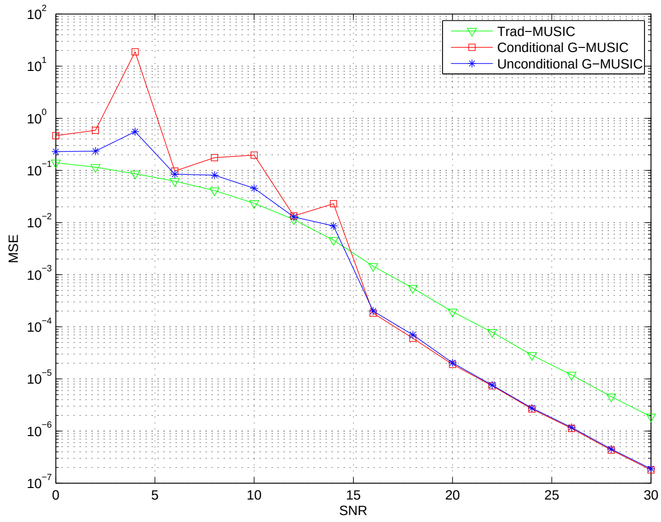


Figure 4.7: Mean of the MSE of the estimates of $\mathbf{a}(\theta_k)^H \mathbf{\Pi} \mathbf{a}(\theta_k)$ for $k = 1, \ldots, 10$ versus SNR

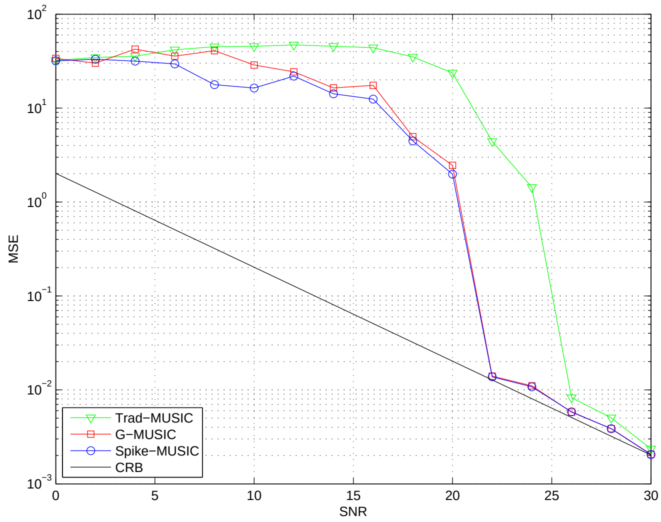


Figure 4.8: Mean of the MSE of the two estimated angles

evaluated at the true angles. For an SNR greater that 10 dB, the performance of $\hat{\eta}_{\text{spike},N}$ is close once again to $\hat{\eta}_{\text{new},N}$.

- In experiment 3, we consider $K = 5$ sources located at $-20°$, $-10°$, $0°$, $10°$ and $20°$ and M and N are still equal to 20 and 40. The separation condition is verified for all values of SNR between 10 dB and 30 dB. In figure 4.10, we plot the same graph as in figure 4.9. We notice that the spike estimator is not close anymore to the G-MUSIC because the ratio K/N is not small enough. However, it still outperforms the traditional MUSIC estimator.

4.5 Appendix

4.5.1 Proof of lemma 4.2.1: some uniform convergences

Define the event $\Omega = \Omega_1 \cap \Omega_2$, where Ω_1 and Ω_2 are the probability one events over which the results of Theorems 2.2.1 (chapter 2) and 3.2.3 (chapter 3) hold, and fix a realization in Ω. Therefore, for N large enough, $(\hat{m}_N - m_N)$ is

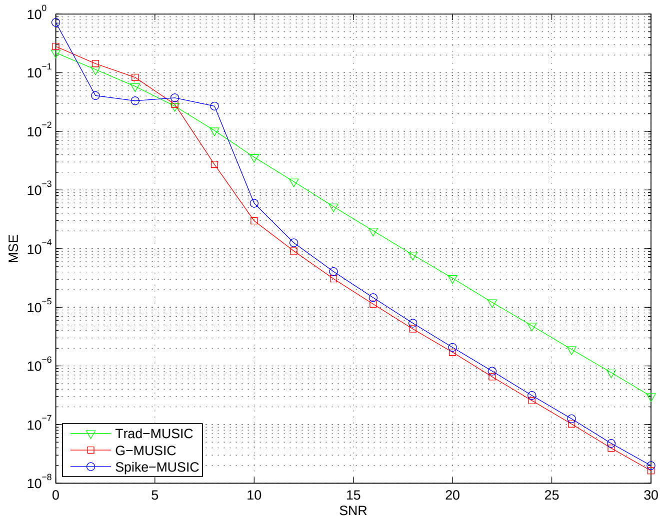


Figure 4.9: Mean of the MSE (localization function), $K = 2$.

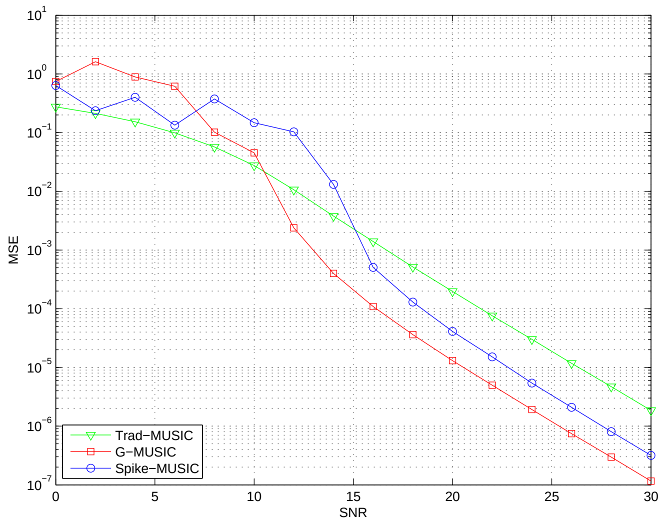


Figure 4.10: Mean of the MSE (localization function), $K = 5$.

a sequence of holomorphic functions on $\mathbb{C}\setminus\left([t_1^-, t_1^+]\cup[t_2^-,\infty)\right)$ and from the bound in property 1.2.1,

$$|\hat{m}_N(z) - m_N(z)| \leq \frac{2}{\operatorname{dist}\left(z, [t_1^-, t_1^+]\cup[t_2^-,\infty)\right)},$$

This implies that $(\hat{m}_N - m_N)$ is a normal family on $\mathbb{C}\setminus\left([t_1^-, t_1^+]\cup[t_2^-,\infty)\right)$ by Montel's theorem. Let $(\hat{m}_{\varphi(N)} - m_{\varphi(N)})$ be a subsequence which converges uniformly on each compact subset of $\mathbb{C}\setminus\left([t_1^-, t_1^+]\cup[t_2^-,\infty)\right)$ to a holomorphic function ξ. Theorem 2.2.1 implies that $\hat{m}_{\varphi(N)}(z) - m_{\varphi(N)}(z) \to_N 0$ for all $z \in \mathbb{C}\setminus\mathbb{R}^+$, and thus $\xi = 0$ on $\mathbb{C}\setminus\mathbb{R}^+$. By analytic continuation, we deduce that $\xi = 0$ on $\mathbb{C}\setminus\left([t_1^-, t_1^+]\cup[t_2^-,\infty)\right)$, and consequently all converging subsequence extracted from the normal family $(\hat{m}_N - m_N)$ converges uniformly to 0 on each compact subsets of $\mathbb{C}\setminus\left([t_1^-, t_1^+]\cup[t_2^-,\infty)\right)$ which in turn implies the uniform convergence of the whole sequence $(\hat{m}_N - m_N)$ to 0. The uniform convergence of the derivatives $\hat{m}'_N - m'_N$ is a straightforward consequence.

Using the bounds

$$\|\mathbf{Q}_N(z)\|, \|\mathbf{T}_N(z)\| \leq \operatorname{dist}\left(z, [t_1^-, t_1^+]\cup[t_2^-,\infty)\right)$$

(Theorem 3.2.3 in chapter 3 and bound (2.8) in chapter 2), (4.5) is obtained similarly.

Finally, from property 2.3.1 (chapter 2),

$$\inf_{z\in\mathcal{K}}\left|1+\sigma^2 c_N m_N(z)\right|\geq 2,$$

for all large N. Therefore, using (4.4), we have for all $z\in\mathcal{K}$ $\left|1+\sigma^2 c_N \hat{m}_N(z)\right|\geq 1$ for all large N, and thus,

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

This concludes the proof.

4.5.2 Computations of the residues and formula of the estimator

In order to evaluate these residues, we first remark that

$$\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N=\sum_{k=1}^{M}\frac{\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N}{\hat{\lambda}_{k,N}-z},$$

and $\hat{g}_N(z)$ can thus be written as

$$\hat{g}_N(z)=\sum_{k=1}^{M}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N\left(\hat{\alpha}_{k,N}(z)+\hat{\beta}_{k,N}(z)+\hat{\gamma}_{k,N}(z)\right),$$

where we have defined

$$\begin{aligned}
\hat{\alpha}_{k,N}(z)&=\frac{1+\sigma^2 c_N\hat{m}_N(z)}{\hat{\lambda}_{k,N}-z},\\
\hat{\beta}_{k,N}(z)&=\frac{2\sigma^2 c_N z\hat{m}_N'(z)}{\hat{\lambda}_{k,N}-z},\\
\hat{\gamma}_{k,N}(z)&=-\sigma^4 c_N(1-c_N)\frac{\hat{m}_N'(z)}{(\hat{\lambda}_{k,N}-z)(1+\sigma^2 c_N\hat{m}_N(z))}.
\end{aligned}$$

Consequently, with probability one

$$\begin{aligned}
&\frac{1}{2\pi\mathrm{i}}\oint_{\partial\mathcal{R}^-}\hat{g}_N(z)\mathrm{d}z=\\
&\quad-\sum_{k=1}^{M}\mathbf{d}_N^*\hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*\mathbf{d}_N\sum_{m=1}^{M-K}\left(\mathrm{Res}\left(\hat{\alpha}_{k,N},\hat{\lambda}_{m,N}\right)+\mathrm{Res}\left(\hat{\beta}_{k,N},\hat{\lambda}_{m,N}\right)+\mathrm{Res}\left(\hat{\gamma}_{k,N},\hat{\lambda}_{m,N}\right)+\mathrm{Res}\left(\hat{\gamma}_{k,N},\hat{\omega}_{m,N}\right)\right),
\end{aligned}$$

for all large N and classical residue calculus gives

$$\begin{aligned}
\mathrm{Res}\left(\hat{\alpha}_{k,N},\hat{\lambda}_{m,N}\right)&=\begin{cases}-\frac{\sigma^2 c_N}{M}\frac{1}{\hat{\lambda}_{k,N}-\hat{\lambda}_{m,N}} & \text{for } k\neq m,\\ -\left(1+\sigma^2 c_N\frac{1}{M}\sum_{\substack{i=1\\ i\neq k}}^{M}\frac{1}{\hat{\lambda}_{i,N}-\hat{\lambda}_{k,N}}\right) & \text{for } k=m,\end{cases}\\
\mathrm{Res}\left(\hat{\beta}_{k,N},\hat{\lambda}_{m,N}\right)&=\begin{cases}\frac{2\sigma^2 c_N}{M}\frac{\hat{\lambda}_{k,N}}{(\hat{\lambda}_{k,N}-\hat{\lambda}_{m,N})^2} & \text{for } k\neq m,\\ -\frac{2\sigma^2 c_N}{M}\sum_{\substack{i=1\\ i\neq k}}^{M}\frac{\hat{\lambda}_{k,N}}{(\hat{\lambda}_{i,N}-\hat{\lambda}_{k,N})^2} & \text{for } k=m,\end{cases}\\
\mathrm{Res}\left(\hat{\gamma}_{k,N},\hat{\lambda}_{m,N}\right)&=\begin{cases}\sigma^2(1-c_N)\frac{1}{\hat{\lambda}_{k,N}-\hat{\lambda}_{m,N}} & \text{for } k\neq m,\\ -M\frac{1-c_N}{c_N}\left(1+\frac{\sigma^2 c_N}{M}\sum_{\substack{i=1\\ i\neq k}}^{M}\frac{1}{\hat{\lambda}_{i,N}-\hat{\lambda}_{k,N}}\right) & \text{for } k=m,\end{cases}\\
\mathrm{Res}\left(\hat{\gamma}_{k,N},\hat{\omega}_{m,N}\right)&=-\sigma^2\frac{1-c_N}{\hat{\lambda}_{k,N}-\hat{\omega}_{m,N}}.
\end{aligned}$$

Next, we define $\hat{\xi}_{k,N}$ as

$$\hat{\xi}_{k,N}=-\sum_{m=1}^{M-K}\left(\mathrm{Res}\left(\hat{\alpha}_{k,N},\hat{\lambda}_{m,N}\right)+\mathrm{Res}\left(\hat{\beta}_{k,N},\hat{\lambda}_{m,N}\right)+\mathrm{Res}\left(\hat{\gamma}_{k,N},\hat{\lambda}_{m,N}\right)+\mathrm{Res}\left(\hat{\gamma}_{k,N},\hat{\omega}_m\right)\right).$$

We obtain, for $k = 1, \ldots, M-K$

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

and for $k = M-K+1, \ldots, M$

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

To retrieve the final form of $\hat{\xi}_{k,N}$ given in the statement of Theorem 4.2.1, we notice that

$$1 + \sigma^2 c_N \frac{1}{M} \sum_{i=1}^{M} \frac{1}{\hat{\lambda}_{i,N} - \hat{\omega}_{k,N}} = 0,$$

and use the following lemma

Lemma 4.5.1. *The following identity holds for any $k = 1 \ldots M$,*

$$\frac{1}{M} \sum_{\substack{i=1\\i\neq k}}^{M} \frac{1}{\hat{\lambda}_{i,N} - \hat{\omega}_{k,N}} = \frac{2}{M} \sum_{\substack{i=1\\i\neq k}}^{M} \frac{1}{\hat{\lambda}_{i,N} - \hat{\lambda}_{k,N}} - \frac{1}{M} \sum_{\substack{i=1\\i\neq k}}^{M} \frac{1}{\hat{\omega}_{i} - \hat{\lambda}_{k,N}}.$$

Proof. We first write the equation in ω, $1 + \sigma^2 c_N \hat{m}_N(\omega) = 0$ as

$$\frac{\sigma^2 c_N}{M} \sum_{j=1}^{M} \frac{1}{\hat{\lambda}_{j,N} - \omega} + 1 = 0, \tag{4.43}$$

and by multiplying the left-hand side by $\prod_{i=1}^{M}\left(\hat{\lambda}_{j,N} - \omega\right)$, we define a new polynomial $Q(\omega)$, by

$$Q(\omega) = \frac{\sigma^2 c_N}{M} \sum_{j=1}^{M} \prod_{\substack{l=1\\l\neq j}}^{M} \left(\hat{\lambda}_{l,N} - \omega\right) + \prod_{l=1}^{M} \left(\hat{\lambda}_{l,N} - \omega\right).$$

As the polynomial function Q has M roots at $\hat{\omega}_{1,N}, \ldots, \hat{\omega}_{M,N}$, we can write

$$Q(\omega) = \prod_{l=1}^{M} \left(\hat{\omega}_{l,N} - \omega\right).$$

Therefore,

$$Q(\hat{\lambda}_{k,N}) = \prod_{l=1}^{M} \left(\hat{\omega}_{l,N} - \hat{\lambda}_{k,N}\right) = \frac{\sigma^2 c_N}{M} \prod_{\substack{l=1\\l\neq k}}^{M} \left(\hat{\lambda}_{l,N} - \hat{\lambda}_{k,N}\right) \tag{4.44}$$

which will be useful later on. Let us now consider the derivative of Q given by

$$Q'(\omega) = -\sum_{j=1}^{M} \prod_{\substack{l=1\\l\neq j}}^{M} \left(\hat{\omega}_{l,N} - \omega\right) = -\sum_{j=1}^{M} \prod_{\substack{l\neq j\\l=1}}^{M} \left(\hat{\lambda}_{l,N} - \omega\right) - \frac{\sigma^2 c_N}{M} \sum_{m=1}^{M} \sum_{\substack{l=1\\l\neq m}}^{M} \prod_{\substack{j=1\\j\neq m,l}}^{M} \left(\hat{\lambda}_{j,N} - \omega\right).$$

Evaluating again this function at point $\hat{\lambda}_{k,N}$, we obtain

$$Q'(\hat{\lambda}_{k,N}) = -\sum_{j=1}^{M} \prod_{\substack{l=1\\l\neq j}}^{M} \left(\hat{\omega}_{l,N} - \hat{\lambda}_{k,N}\right) = -\prod_{\substack{l=1\\l\neq k}}^{M} \left(\hat{\lambda}_{l,N} - \hat{\lambda}_{k,N}\right) - \frac{2\sigma^2 c_N}{M} \sum_{\substack{l=1\\l\neq k}}^{M} \prod_{\substack{j=1\\j\neq k,l}}^{M} \left(\hat{\lambda}_{j,N} - \hat{\lambda}_{k,N}\right) \tag{4.45}$$

or, dividing both sides by the first term on the right hand side of the equation,

$$\frac{\sum_{j=1}^{M}\prod_{\substack{l=1\\l\neq j}}^{M}\left(\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}\right)}{\prod_{\substack{l=1\\l\neq k}}^{M}\left(\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}\right)}=1+\frac{2\sigma^2 c_N}{M}\sum_{\substack{l=1\\l\neq k}}^{M}\frac{1}{\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}},$$

Going back to equation (4.44), one can also write

$$\frac{\sum_{j=1}^{M}\prod_{\substack{l=1\\l\neq j}}^{M}\left(\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}\right)}{\prod_{\substack{l=1\\l\neq k}}^{M}\left(\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}\right)}=\frac{\sigma^2 c_N}{M}\frac{\sum_{j=1}^{M}\prod_{\substack{l=1\\l\neq j}}^{M}\left(\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}\right)}{\prod_{l=1}^{M}\left(\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}\right)}=\frac{\sigma^2 c_N}{M}\sum_{l=1}^{M}\frac{1}{\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}}.$$

Consequently, we see that we can write

$$1+\frac{2\sigma^2 c_N}{M}\sum_{\substack{l=1\\l\neq k}}^{M}\frac{1}{\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}}=\frac{\sigma^2 c_N}{M}\frac{1}{\hat{\omega}_{k,N}-\hat{\lambda}_{k,N}}+\frac{\sigma^2 c_N}{M}\sum_{\substack{l=1\\l\neq k}}^{M}\frac{1}{\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}},$$

or, reorganizing the terms of this expression in a convenient way,

$$1+\frac{\sigma^2 c_N}{M}\frac{1}{\hat{\lambda}_{k,N}-\hat{\omega}_{k,N}}+\frac{\sigma^2 c_N}{M}\sum_{\substack{l=1\\l\neq k}}^{M}\frac{1}{\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}}=\frac{\sigma^2 c_N}{M}\sum_{\substack{l=1\\l\neq k}}^{M}\frac{1}{\hat{\omega}_{l,N}-\hat{\lambda}_{k,N}}-\frac{\sigma^2 c_N}{M}\sum_{\substack{l=1\\l\neq k}}^{M}\frac{1}{\hat{\lambda}_{l,N}-\hat{\lambda}_{k,N}}. \tag{4.46}$$

But from the equation in ω (4.43), we obtain

$$1+\frac{\sigma^2 c_N}{M}\frac{1}{\hat{\lambda}_{k,N}-\hat{\omega}_{k,N}}+\frac{\sigma^2 c_N}{M}\sum_{\substack{l=1\\l\neq k}}^{M}\frac{1}{\hat{\lambda}_{l,N}-\hat{\omega}_{k,N}}=0 \tag{4.47}$$

and by inserting this expression into (4.46), we finally get the expression in the lemma. □

4.5.3 Proof of lemma 4.2.2: another uniform convergences

We only prove the second assertion of the theorem concerning $\hat{m}_{t,N}$, since the technic is similar for $\hat{m}_N$. We define the domains $\mathscr{U}=\mathbb{C}\backslash[0,\sigma^2(1+\sqrt{c})^2]$ and $\mathscr{U}_t=\mathscr{U}\backslash\{\psi(\lambda_1),\ldots,\psi(\lambda_K)\}$, and denote by Ω_1 the probability one event on which the convergences in theorem 3.3.1 holds. We easily deduce that $\hat{m}_{t,N}(z)-\hat{m}_N(z)\to_N 0$ on Ω_1, for all $z\in\mathscr{U}_t$. Moreover, $\hat{m}_N(z)-m(z)\to_N 0$ on an event $\Omega_{2,z}$ depending on $z\in\mathscr{U}_t$, and s.t. $\mathbb{P}(\Omega_{2,z})=1$. Let (z_k) a dense sequence of $\mathbb{C}\backslash\mathbb{R}$ and fix a realization in the probability 1 event $\Omega=\Omega_1\cap\Omega_2$ with $\Omega_2=\bigcap_k\Omega_{2,z_k}$. For, $z\in\mathscr{U}_t$, let (z_{k_l}) a subsequence s.t. $|z-z_{k_l}|\leq\frac{1}{l}$. Then,

$$|\hat{m}_N(z)-m(z)|\leq\left|\hat{m}_N(z)-\hat{m}_N(z_{k_l})\right|+\left|\hat{m}_N(z_{k_l})-m(z_{k_l})\right|+\left|m(z_{k_l})-m(z)\right|.$$

But from theorem 3.3.1, there exists $\epsilon>0$ s.t. $|\hat{\lambda}_{m,N}-z|\geq\epsilon$ for N large enough and all m. This implies that for l large enough, $|\hat{\lambda}_{m,N}-z_{k_l}|\geq\frac{\epsilon}{2}$ and

$$\left|\hat{m}_N(z)-\hat{m}_N(z_{k_l})\right|\leq\frac{|z_{k_l}-z|}{M}\sum_{m=1}^{M}\frac{1}{|\hat{\lambda}_{m,N}-z||\hat{\lambda}_{m,N}-z_{k,l}|}\leq\frac{2|z_{k_l}-z|}{\epsilon^2}.$$

Consequently, $\limsup_N|\hat{m}_N(z)-m(z)|\leq\frac{2}{l\epsilon^2}+\left|m(z_{k_l})-m(z)\right|$, which goes to 0 by taking the limit in l. Therefore, we have shown that on the event Ω, $\hat{m}_{t,N}(z)-m(z)\to_N 0$ for all $z\in\mathscr{U}_t$.

To conclude the proof, we fix $\epsilon>0$ and consider again a realization in the probability 1 event Ω, and notice that $(\hat{m}_{t,N}-m)$ is a normal family on $\mathscr{U}_\epsilon=\mathbb{C}\backslash[0,\sigma^2(1+\sqrt{c})^2+\epsilon]$ by Montel's theorem, for N large enough. For a subsequence converging uniformly on each compact subset of $\mathscr{U}_\epsilon$, we have shown that the limit is 0 on the set $\mathscr{U}_\epsilon\backslash\{\psi(\lambda_1),\ldots,\psi(\lambda_K)\}$ and by analytic continuation, also on $\mathscr{U}_\epsilon$. Consequently, the whole sequence $(\hat{m}_{t,N}-m)$ converges to 0 uniformly on each compact subset of $\mathscr{U}_\epsilon$. This of course implies the same convergence for the derivative $(\hat{m}'_{t,N}-m')$. Since ϵ can be made arbitrarily small, this concludes the proof.

4.5.4 Proof of lemma 4.2.3: Behaviour of $\hat{\omega}_{k,N}$ under the spiked model assumption

In this proof, we denote by $\mathscr{C}(\lambda, r)$ and $\mathscr{D}(\lambda, r)$ the circle and open disk centered at λ with radius r.

Let Ω_1 and Ω_2 the probability one events on which the convergence of Theorem 3.3.1 and Lemma 4.2.2 hold respectively. Fix a realization in $\Omega = \Omega_1 \cap \Omega_2$.

Let $1 \leq i \leq K$. Since m is continuous on $\mathbb{R}$ and $\operatorname{Re}(1 + \sigma^2 cm(x)) > 0$ for all $x \in \mathbb{R}$ (see theorem 2.3.1 and section 2.5 in chapter 2), we can find $\epsilon > 0$ small enough such that $\inf_{x\in\mathbb{R}} \operatorname{Re}(1 + \sigma^2 cm(x)) > 4\epsilon$. This also ensures that we can choose $r > 0$ small enough such that

$$\mathscr{D}(\sigma^2(1+\sqrt{c})^2, r) \cap \mathscr{D}(\psi(\gamma_1, c), r) \cap \ldots \cap \mathscr{D}(\psi(\gamma_K, c), r) = \emptyset, \tag{4.48}$$

and

$$\begin{aligned} \sup_{z\in\mathscr{D}(\psi(\gamma_i,c),r)} \operatorname{Re}(m(z)) &< 0, \\ \inf_{z\in\mathscr{D}(\psi(\gamma_i,c),r)} \operatorname{Re}\left(1+\sigma^2 cm(z)\right) &\geq 3\epsilon, \\ \sup_{z\in\mathscr{D}(\psi(\gamma_i,c),r)} |\operatorname{Im}\left(\sigma^2 cm(z)\right)| &\leq \frac{\epsilon}{2}. \end{aligned}$$

Let $0 < r' < r$. Thus, Theorem 3.3.1 implies

$$\hat{\lambda}_{M-K,N} \in \mathscr{D}(\sigma^2(1+\sqrt{c})^2, r'),$$

and $\hat{\lambda}_{M-K+k,N} \in \mathscr{D}(\psi(\gamma_k, c), r')$ for $k = 1, \ldots, K$ and N large enough. Since $\hat{m}_N$ converges to m uniformly on each compact of $\mathbb{C} \setminus \left([0, \sigma^2(1+\sqrt{c})^2] \cup \{\psi(\gamma_1, c), \ldots, \psi(\gamma_K, c)\}\right)$, we deduce that for N large enough,

- $\inf_{z\in\mathscr{C}(\psi(\gamma_i,c),r')} \operatorname{Re}\left(1+\sigma^2 c\hat{m}_N(z)\right) \geq 2\epsilon$,
- $\sup_{z\in\mathscr{C}(\psi(\gamma_i,c),r')} |\operatorname{Im}\left(\sigma^2 c\hat{m}_N(z)\right)| \leq \epsilon$,
- $\sup_{z\in\mathscr{C}(\psi(\gamma_i,c),r')} \operatorname{Re}(\hat{m}_N(z)) < 0$.

In particular, the three points above imply

$$\sup_{z\in\mathscr{C}(\psi(\gamma_i,c),r)} |\sigma^2 c_N \hat{m}_N(z)| \leq 1-\epsilon.$$

Consequently, by defining $h(z) = \hat{\lambda}_{M-K+i,N} - z$ and $\tilde{h}(z) = \left(\hat{\lambda}_{M-K+i,N} - z\right)\left(1+\sigma^2 c_N \hat{m}_N(z)\right)$, it follows that

$$|h(z) - \tilde{h}(z)| = \left|\hat{\lambda}_{M-K+i,N} - z\right| \left|\sigma^2 c_N \hat{m}_N(z)\right| < \left|\hat{\lambda}_{M-K+i,N} - z\right| = |h(z)|,$$

for all $z \in \mathscr{C}(\psi(\gamma_i, c), r')$. Therefore, as $1+\sigma^2 c_N \hat{m}_N(z)$ has one pole in $\mathscr{D}(\psi(\gamma_i, c), r')$, i.e. $\hat{\lambda}_{M-K+i}$, Rouché's Theorem implies that $1 + \sigma^2 c_N \hat{m}_N(z)$ has exactly one zero inside $\mathscr{D}(\psi(\gamma_i, c), r')$. Since $r' > 0$ can be made arbitrarily small, we deduce that $\hat{\omega}_{M-K+i,N} \to_N \psi(\gamma_i, c)$ for $i = 1, \ldots, K$.

We now prove that $\hat{\omega}_{M-K,N} \to_N \sigma^2(1+\sqrt{c})^2$. Assume this is not the case. For $\epsilon > 0$, we have from Theorem 3.3.1 $\sigma^2(1+\sqrt{c})^2 - \epsilon \leq \hat{\omega}_{M-K,N} \leq \psi(\gamma_1, c) + \epsilon$ for N large enough. Therefore we can extract a subsequence $\hat{\omega}_{M-K,\varphi(N)} \to_N \theta \neq \sigma^2(1+\sqrt{c})^2$. From the above discussion, we also have $\theta \neq \psi(\gamma_1, c)$. Thus,

$$1 + \sigma^2 c_{\varphi(N)} \hat{m}_{\varphi(N)}(\hat{\omega}_{M-K,\varphi(N)}) = 0,$$

for all N and Lemma 4.2.2 yields $1 + \sigma^2 cm(\theta) = 0$, which leads to a contradiction since $\operatorname{Re}(1 + \sigma^2 cm(x)) > 0$ for all $x \in \mathbb{R}$.

4.5.5 Proof of lemma 4.3.1: Escape of $\hat{\omega}_{1,N}, \ldots, \hat{\omega}_{M,N}$

To prove lemma 4.3.1, we follow the same approach than in Section 3.2.2, and therefore it is sufficient to prove

Lemma 4.5.2. *For all function $\psi \in \mathscr{C}^\infty(\mathbb{R}, \mathbb{R})$ constant over the complementary of a compact interval and vanishing on $\mathscr{S}_N$ for N large enough, it holds that*

$$\mathbb{E}\left[\left(\operatorname{Tr} \psi(\hat{\mathbf{\Omega}}_N)\right)^{2l}\right] = \mathscr{O}\left(\frac{1}{N^{2l}}\right) \tag{4.49}$$

for each $l \in \mathbb{N}$.

For this, we study the behaviour of the Stieltjes transform $\hat{n}_N(z)$ of the distribution $\frac{1}{M}\sum_{k=1}^{M}\delta_{\hat{\omega}_{k,N}}$ defined by

$$\hat{n}_N(z) = \frac{1}{M}\mathrm{Tr}\left(\hat{\mathbf{\Omega}}_N - z\mathbf{I}\right)^{-1}.$$

and use Lemma 3.1.1 as well as the inverse Stieltjes transform formula (property 1.2.2). Our starting point is the following result showing that the empirical eigenvalue distribution of $\hat{\mathbf{\Omega}}_N$ is very similar to the distribution of the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$. In the following, P_1, P_2 are two polynomials independent of N with positive coefficients, whose values may change from one line to another.

A rate of convergence for $\mathbb{E}\left[\hat{m}_N'(z)\right] - m_N'(z)$

We first prove, as an auxiliary result, that

$$\left|\mathbb{E}\left[\hat{m}_N'(z)\right] - m_N'(z)\right| = \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right), \tag{4.50}$$

a result similar to theorem 2.2.2.

For the quantities used below, as well as their properties, we refer the reader to the proof of theorem 2.2.2 (appendix 2.7.2). We first notice that (4.50) is equivalent to

$$\left|\alpha_N'(z) - \delta_N'(z)\right| \le \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right). \tag{4.51}$$

where we recall that $\alpha_N(z) = \mathbb{E}[\sigma^2 c_N \hat{m}_N(z)]$ and $\delta_N(z) = \sigma^2 c_N m_N(z)$. In order to prove (4.51), we first show that $\epsilon_N'(z)$, the derivative of $\epsilon(z) = \alpha_N(z) - \frac{\sigma}{N}\mathrm{Tr}\,\mathbf{R}_N(z)$, where we recall that $\mathbf{R}_N(z) = \left(-z(1+\sigma\tilde{\alpha}_N(z))\mathbf{I}_M + \frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma\alpha_N(z)}\right)^{-1}$, satisfies

$$\left|\epsilon_N'(z)\right| \le \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right), \tag{4.52}$$

and deduce from this that (4.51) holds. In order to evaluate the behaviour of $\epsilon_N'(z)$, we recall from (2.31) that

$$\alpha_N(z) - \frac{1}{N}\mathrm{Tr}\,\mathbf{R}_N(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbb{E}\left[\mathbf{Q}_N(z)\right]\mathbf{R}_N(z)\frac{\sigma}{N}\mathrm{Tr}\,\mathbf{\Delta}_N(z) + \frac{\sigma}{N}\mathrm{Tr}\,\mathbf{\Delta}_N(z)\mathbf{R}_N(z), \tag{4.53}$$

where $\mathbf{\Delta}_N(z)$ is defined in (2.26), and differentiate (4.53) with respect to z. It is easily checked that the derivative of the righthandside of (2.26) is bounded by $\mathrm{P}_1(|z|)\mathrm{P}_2(|\mathrm{Im}(z)|^{-1})$, using lemma 2.7.1 [2] and the fact that $\|\mathbf{R}_N'(z)\| \le \mathrm{P}_1(|z|)\mathrm{P}_2(|\mathrm{Im}(z)|^{-1})$. The details are omitted. It is also possible to prove

$$\left|\tilde{\epsilon}_N'(z)\right| = \left|\tilde{\alpha}_N'(z) - \frac{\sigma}{N}\mathrm{Tr}\,\tilde{\mathbf{R}}_N'(z)\right| \le \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right), \tag{4.54}$$

with $\tilde{\alpha}_N(z) = \alpha_N(z) - \frac{\sigma(1-c_N)}{z}$ and $\tilde{\mathbf{R}}_N(z) = \left(-z(1+\sigma\alpha_N(z))\mathbf{I}_N + \frac{\mathbf{B}_N^*\mathbf{B}_N}{1+\sigma\tilde{\alpha}_N(z)}\right)^{-1}$. In order to complete the proof of 4.51, we observe that

$$\alpha_N'(z) - \delta_N'(z) = \sigma\frac{1}{N}\mathrm{Tr}\,(\mathbf{R}_N'(z)) - \delta_N'(z) + \epsilon_N'(z), \tag{4.55}$$

$$\tilde{\alpha}_N'(z) - \tilde{\delta}_N'(z) = \sigma\frac{1}{N}\mathrm{Tr}\,(\tilde{\mathbf{R}}_N'(z)) - \tilde{\delta}_N'(z) + \tilde{\epsilon}_N'(z). \tag{4.56}$$

From the system of equations (4.109), we have

$$\frac{\sigma}{N}\mathrm{Tr}\,(\mathbf{R}_N(z) - \mathbf{T}_N(z)) = \left(\tilde{\alpha}_N(z) - \tilde{\delta}_N(z)\right)zv_N(z) + \left(\alpha_N(z) - \delta_N(z)\right)u_N(z), \tag{4.57}$$

$$\frac{\sigma}{N}\mathrm{Tr}\,\left(\tilde{\mathbf{R}}_N(z) - \tilde{\mathbf{T}}_N(z)\right) = \left(\tilde{\alpha}_N(z) - \tilde{\delta}_N(z)\right)u_N(z) + \left(\alpha_N(z) - \delta_N(z)\right)z\tilde{v}_N(z), \tag{4.58}$$

By differentiating (4.57), (4.58)) w.r.t. z, we use (4.55), (4.56) and (2.22), and recall that both $|\alpha_N(z) - \delta_N(z)|$ and $|\tilde{\alpha}_N(z) - \tilde{\delta}_N(z)|$ are bounded that $\frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(|\mathrm{Im}(z)|^{-1}\right)$ (see theorem 2.2.2). We check that $u_N(z)$, $zv_N(z)$, $z\tilde{v}_N(z)$ and their derivatives are bounded by $\mathrm{P}_1(|z|)\mathrm{P}_2\left(|\mathrm{Im}(z)|^{-1}\right)$, and obtain eventually that

$$\begin{bmatrix}\alpha_N'(z) - \delta_N'(z)\\ \tilde{\alpha}_N'(z) - \tilde{\delta}_N'(z)\end{bmatrix} = \begin{bmatrix}u_N(z) & zv_N(z)\\ z\tilde{v}_N(z) & u_N(z)\end{bmatrix}\begin{bmatrix}\alpha_N'(z) - \delta_N'(z)\\ \tilde{\alpha}_N'(z) - \tilde{\delta}_N'(z)\end{bmatrix} + \begin{bmatrix}\xi_N(z)\\ \tilde{\xi}_N(z)\end{bmatrix},$$

[2] In fact, it is necessary to adapt the statement of lemma 2.7.1 by replacing $\mathbf{Q}_N(z)$ with $\mathbf{Q}_N(z)^2$. The proof is similar.

with $|\xi_N(z)|$, $|\tilde{\xi}_N(z)|$ bounded by $N^{-2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(|\mathrm{Im}(z)|^{-1}\right)$. The determinant of the above system is the same as (4.109), i.e.

$$\Delta_N(z) = (1 - u_N(z))^2 - z v_N(z)\tilde{v}_N(z). \tag{4.59}$$

From (2.43), $\left|\Delta_N(z)^{-1}\right| \le \mathrm{P}_1(|z|)\mathrm{P}_2\left(|\mathrm{Im}(z)|^{-1}\right)$ for all z in subset $\mathscr{A}_N$ of $\mathbb{C}$ defined as

$$\mathscr{A}_N = \left\{z \in \mathbb{C}\backslash\mathbb{R}, \frac{1}{N^2}\mathrm{Q}_1(|z|)\mathrm{Q}_2\left(|\mathrm{Im}(z)|^{-1}\right) < 1\right\},$$

where Q_1 and Q_2 are 2 polynomials independent of N. Thus, we can invert the previous system on $\mathscr{A}_N$ to get

$$\begin{bmatrix}\alpha'_N(z) - \delta'_N(z)\\ \tilde{\alpha}'_N(z) - \tilde{\delta}'_N(z)\end{bmatrix} = \frac{1}{\Delta_N(z)}\begin{bmatrix}1 - u_N(z) & z v_N(z)\\ z\tilde{v}_N(z) & 1 - u_N(z)\end{bmatrix}\frac{1}{N^2}\begin{bmatrix}\xi_N(z)\\ \tilde{\xi}_N(z)\end{bmatrix}.$$

This implies that $|\alpha'_N(z) - \delta'_N(z)|$ is bounded by $\frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(|\mathrm{Im}(z)|^{-1}\right)$ on $\mathscr{A}_N$. If $z \in \mathbb{C}\backslash\{\mathbb{R}\cup\mathscr{A}_N\}$, we proceed as usual: we remark that

$$|\alpha'_N(z) - \delta'_N(z)| \le |\alpha'_N(z)| + |\delta'_N(z)| \le \frac{C}{|\mathrm{Im}z|},$$

for each z with $C > 0$ independent of N, and $1 \le \frac{1}{N^2}\mathrm{Q}_1(|z|)\mathrm{Q}_2\left(|\mathrm{Im}(z)|^{-1}\right)$ for $z \in \mathbb{C}\backslash\{\mathbb{R}\cup\mathscr{A}_N\}$. Therefore,

$$|\alpha'_N(z) - \delta'_N(z)| \le \frac{C}{|\mathrm{Im}z|}\frac{1}{N^2}\mathrm{Q}_1(|z|)\mathrm{Q}_2\left(|\mathrm{Im}(z)|^{-1}\right) \le \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(|\mathrm{Im}(z)|^{-1}\right),$$

on $\mathbb{C}\backslash\{\mathbb{R}\cup\mathscr{D}_N\}$. This in turn shows that (4.51) holds on $\mathbb{C}\backslash\mathbb{R}$.

Decomposition of $\mathbb{E}[\hat{n}_N(z)]$

We now prove that

$$\mathbb{E}[\hat{n}_N(z)] = \int_{\mathscr{S}_N}\frac{\mathrm{d}\mu_N(\lambda)}{\lambda - z} - \frac{1}{M}\int_{\mathscr{S}_N}\frac{d\kappa_N(\lambda)}{\lambda - z} + r_N(z), \tag{4.60}$$

where κ_N a finite signed measure carried by $\mathscr{S}_N$ such that $\kappa_N\left([x^-_{q,N}, x^+_{q,N}]\right) = 0$ for $q = 1,\ldots,Q_N$, and r_N a holomorphic function on $\mathbb{C}\backslash\mathbb{R}$ satisfying

$$|r_N(z)| \le \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right),$$

To prove (4.60), we use that $\hat{\mathbf{\Omega}}_N$ is a rank 1 perturbation of $\hat{\mathbf{\Lambda}}_N$ (see section 4.2.1), and obtain immediately that

$$\hat{n}_N(z) = \hat{m}_N(z) - \frac{1}{M}\hat{h}_N(z)$$

where $\hat{h}_N(z) = \frac{\sigma^2 c_N \hat{m}'_N(z)}{1+\sigma^2 c_N \hat{m}_N(z)}$. Therefore, for $z \in \mathbb{C}\backslash\mathbb{R}$, it holds that

$$\mathbb{E}[\hat{n}_N(z)] = \mathbb{E}[\hat{m}_N(z)] - \frac{1}{M}\mathbb{E}\left[\hat{h}_N(z)\right]. \tag{4.61}$$

We first establish that

$$\left|\mathbb{E}\left[\hat{h}_N(z)\right] - h_N(z)\right| \le \frac{1}{N}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right), \tag{4.62}$$

where $h_N(z) = \frac{\sigma^2 c_N m'_N(z)}{1+\sigma^2 c_N m_N(z)}$. For this, we write

$$\begin{aligned}&\hat{h}_N(z) - h_N(z) = \\ &\frac{\sigma^2 c_N(\hat{m}'_N(z) - m'_N(z))}{(1+\sigma^2 c_N\hat{m}_N(z))(1+\sigma^2 c_N m_N(z))} + \frac{(\sigma^2 c_N)^2\left(m_N(z)(\hat{m}'_N(z) - m'_N(z)) + m'_N(z)(m_N(z) - \hat{m}_N(z))\right)}{(1+\sigma^2 c_N\hat{m}_N(z))(1+\sigma^2 c_N m_N(z))}.\end{aligned} \tag{4.63}$$

From theorem 2.2.2 and lemma 2.7.1, we get

$$\mathbb{E}|\hat{m}_N(z) - m_N(z)|^2 \leq 2\mathrm{Var}[\hat{m}_N(z)] + 2|\mathbb{E}[\hat{m}_N(z)] - m_N(z)|^2 \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

Therefore,

$$\mathbb{E}|\hat{m}_N(z) - m_N(z)| \leq \frac{1}{N}\left(\mathrm{P}_1(|z|) + \mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)\right).$$

Since lemma 2.7.1 also holds with $\mathbf{Q}_N(z)^2$ instead of $\mathbf{Q}_N(z)$, this implies similarly with (4.50) that

$$\mathbb{E}\left|\hat{m}'_N(z) - m'_N(z)\right|^2 \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

Therefore, it holds that

$$\mathbb{E}\left|\hat{m}'_N(z) - m'_N(z)\right| \leq \frac{1}{N}\left(\mathrm{P}_1(|z|) + \mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)\right).$$

Since $-z(1+\sigma^2 c_N \hat{m}_N(z))^{-1}$ is the Stieltjes transform of a probability measure carried by $\mathbb{R}^+$, we have

$$\frac{1}{|1+\sigma^2 c_N m_N(z)|} \leq \frac{|z|}{|\mathrm{Im}(z)|},$$

and using the bounds $|m_N(z)| \leq \frac{1}{|\mathrm{Im}(z)|}$, $|m'_N(z)| \leq \frac{1}{|\mathrm{Im}(z)|^2}$, $|1+\sigma^2 c_N m_N(z)|^{-1} \leq 2$ (property 2.3.1), we eventually get from (4.63) that

$$\mathbb{E}\left|\hat{h}_N(z) - h_N(z)\right| \leq \frac{1}{N}\mathrm{P}_1(|z|)\,\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right).$$

This immediately implies (4.62). Using property 2.3.1, we obtain that $|h_N(z)| \leq 2\sigma^2 c_N |m'_N(z)|$. As $|m'_N(z)| \leq \frac{1}{\mathrm{dist}(z,\mathscr{S}_N)^2}$, it holds that $|h_N(z)| \leq C\frac{1}{\mathrm{dist}(z,\mathscr{S}_N)^2}$, with C a positive constant. Property 1.2.3 implies that $h_N(z)$ is the Stieltjes transform of a finite signed measure κ_N, the support of which is the set of singularities of $h_N(z)$, i.e. $\mathscr{S}_N$. In order to evaluate $\kappa_N([x^-_{q,N}, x^+_{q,N}])$, we use the inverse Stieltjes transform formula,

$$\kappa_N([x^-_{q,N}, x^+_{q,N}]) = \lim_{y\downarrow 0}\frac{1}{\pi}\mathrm{Im}\left(\int_{[x^-_{q,N}, x^+_{q,N}]} h_N(x+iy)\mathrm{d}x\right).$$

It is clear that $h_N(x+iy) = \frac{\partial \log(1+\sigma^2 c_N m_N(x+iy))}{\partial x}$, where the complex logarithm corresponds to the principal determination defined on $\mathbb{C}\backslash\mathbb{R}^-$. We note that property 2.3.1 justifies the use of the principal determination. Therefore,

$$\int_{[x^-_{q,N}, x^+_{q,N}]} h_N(x+iy)\mathrm{d}x = \log\left(1+\sigma^2 c_N m_N(x^+_{q,N}+iy)\right) - \log\left(1+\sigma^2 c_N m_N(x^-_{q,N}+iy)\right).$$

When $y \to 0$, this converges towards $\log(1+\sigma^2 c_N m_N(x^+_{q,N})) - \log(1+\sigma^2 c_N m_N(x^-_{q,N}))$, a real quantity because $x^-_{q,N}$ and $x^+_{q,N}$ belong to $\partial\mathscr{S}_N$ and $1+\sigma^2 c_N m_N(x^\pm_{q,N}) > 0$ (property 2.3.1). This shows that $\kappa_N([x^-_{q,N}, x^+_{q,N}]) = 0$. Consequently,

$$\mathbb{E}\left[\hat{h}_N(z)\right] = \int_{\mathscr{S}_N}\frac{d\kappa_N(\lambda)}{\lambda - z} + M r_N(z),$$

where $r_N(z)$ is holomorphic on $\mathbb{C}\backslash\mathbb{R}$ such that $|r_N(z)| \leq \frac{1}{N^2}\mathrm{P}_1(|z|)\mathrm{P}_2\left(\frac{1}{|\mathrm{Im}(z)|}\right)$. This shows (4.60).

Proof of lemma 4.5.2 for $l = 1$

We now handle the proof of lemma 4.5.2 in the special case $l = 1$, and in the next paragraph, extend the result for any $l \in \mathbb{N}$ by induction. For this, we use the Poincaré inequality (theorem 1.4.2) as in the proof of lemma 3.1.3. However, in the present case, the entries of $\hat{\mathbf{\Omega}}_N$, considered as functions of the real and imaginary parts of the entries of $\mathbf{W}_N$, are not continuously differentiable on $\mathbb{R}^{2MN}$ because function $\mathbf{W}_N \to \hat{\lambda}_{k,N}$ is not differentiable at points for which eigenvalue $\hat{\lambda}_{k,N}$ is multiple. The use of Poincaré inequality has therefore to be justified carefully, which is the purpose of the following result.

Lemma 4.5.3. *Let $\tilde{\psi}$ be a function of $\mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$. Then,* $\mathrm{Tr}\ \tilde{\psi}(\hat{\mathbf{\Omega}}_N)$*, considered as a function of the real and imaginary parts of the entries of $\mathbf{W}_N$, is continuously differentiable. Moreover, if the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ have multiplicity 1, it holds that*

$$\frac{\partial}{\partial W_{i,j,N}}\left\{\frac{1}{M}\mathrm{Tr}\ \left(\tilde{\psi}(\hat{\mathbf{\Omega}}_N)\right)\right\} = \frac{1}{M}\left[\mathbf{\Sigma}_N^*\sum_{l=1}^M [\tilde{\psi}'(\hat{\mathbf{\Omega}}_N)]_{l,l}\hat{\mathbf{u}}_{l,N}\hat{\mathbf{u}}_{l,N}^*\right]_{j,i}. \tag{4.64}$$

Lemmas 4.5.3 is proved at the end of this section. The bound

$$\sup_N \mathbb{E}\left[\|\mathbf{W}_N\mathbf{W}_N^*\|^p\right] < +\infty, \tag{4.65}$$

valid for $p\in\mathbb{N}$, which have been given in section 1.4.2, will be also useful for the following.

Denote by b the constant value taken by ψ over the complementary of a certain compact interval, and by $\tilde{\psi}$ the function of $\mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ defined by $\tilde{\psi}(\lambda) = \psi(\lambda) - b$, which is equal to $-b$ on $\mathscr{S}_N$. Using Lemma 3.1.1 and (4.60), we obtain

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\ \tilde{\psi}(\hat{\mathbf{\Omega}}_N)\right] = \int_{\mathscr{S}_N}\tilde{\psi}(\lambda)\,d\mu_N(\lambda) - \frac{1}{M}\int_{\mathscr{S}_N}\tilde{\psi}(\lambda)\mathrm{d}\kappa_N(\lambda) + \mathscr{O}\left(\frac{1}{N^2}\right). \tag{4.66}$$

Using that $\kappa_N([x_{q,N}^-, x_{q,N}^+]) = 0$ for each $q = 1,\dots,Q_N$, we get that $\int_{\mathscr{S}_N}\tilde{\psi}(\lambda)d\kappa_N(\lambda) = 0$ and that

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\ \tilde{\psi}(\hat{\mathbf{\Omega}}_N)\right] = -b + \mathscr{O}\left(\frac{1}{N^2}\right).$$

Therefore, it holds that

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\ \psi(\hat{\mathbf{\Omega}}_N)\right] = \mathscr{O}\left(\frac{1}{N^2}\right).$$

To prove lemma 4.5.2 for $l = 1$, it thus remains to prove that

$$\mathrm{Var}\left[\frac{1}{M}\mathrm{Tr}\ \left(\psi\left(\hat{\mathbf{\Omega}}_N\right)\right)\right] = \mathscr{O}\left(\frac{1}{N^4}\right). \tag{4.67}$$

We first note that, considered as a function of $(\mathrm{Re}(W_{i,j,N}),\mathrm{Im}(W_{i,j,N}))_{1\le i\le M, 1\le j\le N}$, $\frac{1}{M}\mathrm{Tr}\ \tilde{\psi}(\hat{\mathbf{\Omega}}_N)$ is continuously differentiable by Lemma 4.5.3. Therefore, function $\frac{1}{M}\mathrm{Tr}\ \psi(\hat{\mathbf{\Omega}}_N)$ is continuously differentiable as well. It is thus possible to use the Poincaré inequality to evaluate the lefthandside of (4.67). Furthermore, as the eigenvalues $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$ have multiplicity 1 a.s. (see remark 4.2.1), it is sufficient to evaluate the partial derivatives of function $\frac{1}{M}\mathrm{Tr}\ \psi(\hat{\mathbf{\Omega}}_N)$ when $\mathbf{W}_N$ is such that the $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$ have multiplicity 1. As the derivative of ψ coincides with $\tilde{\psi}'$, (4.64) and Poincaré inequality lead to

$$\mathrm{Var}\left[\frac{1}{M}\mathrm{Tr}\ \left(\psi(\hat{\mathbf{\Omega}}_N)\right)\right] \le \frac{C}{N^2}\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\ \left(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\sum_{l=1}^M\left|[\psi'(\hat{\mathbf{\Omega}}_N)]_{l,l}\right|^2\hat{\mathbf{u}}_{l,N}\hat{\mathbf{u}}_{l,N}^*\right)\right],$$

or equivalently,

$$\mathrm{Var}\left[\frac{1}{M}\mathrm{Tr}\ \left(\psi(\hat{\mathbf{\Omega}}_N)\right)\right] \le \frac{C}{N^2}\mathbb{E}\left[\frac{1}{M}\sum_{l=1}^M\hat{\lambda}_{l,N}\left|[\psi'(\hat{\mathbf{\Omega}}_N)]_{l,l}\right|^2\right].$$

From property 1.3.3, we get

$$\left|[\psi'(\hat{\mathbf{\Omega}}_N)]_{l,l}\right|^2 \le \left([\psi'(\hat{\mathbf{\Omega}}_N)]^2\right)_{l,l}. \tag{4.68}$$

Therefore, it holds that

$$\mathrm{Var}\left[\frac{1}{M}\mathrm{Tr}\ \left(\psi(\hat{\mathbf{\Omega}}_N)\right)\right] \le \frac{C}{N^2}\mathbb{E}\left[\left\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right\|\frac{1}{M}\mathrm{Tr}\ \psi'(\hat{\mathbf{\Omega}}_N)^2\right]. \tag{4.69}$$

As $\sup_N\|\mathbf{B}_N\mathbf{B}_N^*\| < +\infty$, (4.65) yields

$$\sup_N\mathbb{E}\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\|^p < +\infty.$$

We remark that $\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\| < t_2^+ + \epsilon$ on the set $\mathscr{A}_{1,N}^c$ (see (4.29)), and thus

$$\mathbb{E}\left[\left\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right\| \mathbb{1}_{\mathscr{A}_{1,N}^c} \frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right] \le \left(t_2^+ + \epsilon\right)\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right].$$

Function ψ'^2 belongs to $\mathscr{C}_c^\infty(\mathbb{R},\mathbb{R})$ and vanishes on $\mathscr{S}_N$. Therefore, (4.60) implies that $\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right] = \mathscr{O}\left(\frac{1}{N^2}\right)$. Moreover, as $\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2 \le \sup_\lambda \psi'(\lambda)^2 < C$ with C a positive constant, we have

$$\mathbb{E}\left[\left\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right\| \mathbb{1}_{\mathscr{A}_{1,N}} \frac{1}{M}\sum_{l=1}^M \psi'(\hat{\omega}_l)^2\right] < C\mathbb{E}\left[\left\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\right\| \mathbb{1}_{\mathscr{A}_{1,N}}\right], \le C\left(\mathbb{E}\left[\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\|^2\right]\right)^{1/2}\mathbb{P}(\mathscr{A}_{1,N})^{1/2} = \mathscr{O}\left(\frac{1}{N^p}\right),$$

for each integer p. This completes the proof of (4.67).

Proof of lemma 4.5.2 for any l

Assume that (4.49) holds until integer $l-1$. We write as previously that

$$\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^{2l}\right] = \left(\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^l\right]\right)^2 + \mathrm{Var}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^l\right]. \tag{4.70}$$

The Cauchy-Schwarz inequality leads immediately to

$$\left(\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^l\right]\right)^2 \le \mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^2\right]\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^{2l-2}\right] = \mathscr{O}\left(\frac{1}{N^{2l}}\right). \tag{4.71}$$

As for the second term of the r.h.s. of (4.70), we use Poincaré inequality and Hölder's inequality to obtain

$$\begin{aligned}\mathrm{Var}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^l\right] &\le C\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^{2(l-1)}\frac{1}{M}\sum_{k=1}^M \hat{\lambda}_{k,N}\left([\psi'(\hat{\mathbf{\Omega}}_N)]_{k,k}\right)^2\right], \\ &\le C\left(\mathbb{E}\left|\frac{1}{M}\sum_{k=1}^M \hat{\lambda}_{k,N}\left([\psi'(\hat{\mathbf{\Omega}}_N)]_{k,k}\right)^2\right|^l\right)^{\frac{1}{l}}\left(\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^{2l}\right]\right)^{\frac{l-1}{l}},\end{aligned}$$

where C is a positive constant. Property 1.3.3 leads again to

$$\mathbb{E}\left|\frac{1}{M}\sum_{k=1}^M \hat{\lambda}_{k,N}\left([\psi'(\hat{\mathbf{\Omega}}_N)]_{k,k}\right)^2\right|^l \le \mathbb{E}\left|\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\|\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right|^l.$$

We write again that

$$\mathbb{E}\left|\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\|\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right|^l = \mathbb{E}\left|\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\|\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\mathbb{1}_{\mathscr{A}_{1,N}}\right|^l + \mathbb{E}\left|\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\|\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\mathbb{1}_{\mathscr{A}_{1,N}^c}\right|^l,$$

and obtain as previously that

$$\mathbb{E}\left|\|\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*\|\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right|^l \le C\left(\mathbb{E}\left|\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right|^l + \left(\mathbb{P}(\mathscr{A}_{1,N})\right)^{1/2}\right).$$

But, applying Cauchy-Schwarz inequality as in (4.71) to $\mathbb{E}\left|\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right|^l$ leads to $\mathbb{E}\left|\frac{1}{M}\mathrm{Tr}\,\psi'(\hat{\mathbf{\Omega}}_N)^2\right|^l = \mathscr{O}\left(\frac{1}{N^{2l}}\right)$. Gathering all the previous inequalities, we find that

$$\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^{2l}\right] \le \frac{C}{N^2}\left(\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^{2l}\right]\right)^{\frac{l-1}{l}} + \mathscr{O}\left(\frac{1}{N^{2l}}\right),$$

and in the same way as in the proof of lemma 3.2.1, we obtain $\mathbb{E}\left[\left(\mathrm{Tr}\,\psi(\hat{\mathbf{\Omega}}_N)\right)^{2l}\right] = \mathscr{O}\left(\frac{1}{N^{2l}}\right)$. This concludes the proof of Lemma 4.5.2.

Proof of auxiliary lemma 4.5.3

We first establish the following auxiliary result.

Lemma 4.5.4. *Given an integer $D > 0$, let f be a continuous real function on $\mathbb{R}^D$. Let $\mathcal{O}$ be an open set of $\mathbb{R}^D$ such that $\mathbb{R}^D \backslash \mathcal{O}$ has a zero Lebesgue measure. Assume that f is a $\mathcal{C}^1$ function on $\mathcal{O}$ and that its gradient f' on $\mathcal{O}$ can be continuously extended to $\mathbb{R}^D$. Then f is $\mathcal{C}^1$ on the whole $\mathbb{R}^D$ with gradient f'.*

Proof. We only need to prove that for any $x \in \mathbb{R}^D - \mathcal{O}$ and any sequence $x_n \to x$,

$$f(x_n) - f(x) = \langle f'(x), x_n - x \rangle + o(d_n).$$

where $d_n = \|x_n - x\|$. Since f is uniformly continuous on any small neighborhood of x, there exists a sequence δ_n such that for every y and y' in this neighborhood for which $\|y - y'\| < \delta_n$, $|f(y) - f(y')| \leq d_n^2$. Since $\mathbb{R}^D - \mathcal{O}$ has a zero Lebesgue measure, there exists y_n and z_n in $\mathcal{O}$ such that

$$\|x_n - y_n\| < \min(\delta_n, d_n^2) \text{ and } \|x - z_n\| < \min(\delta_n, d_n^2).$$

Therefore, it holds that $\max(|f(x_n) - f(y_n)|, |f(z_n) - f(x)|) < d_n^2$. Writing $f(x_n) - f(x) = f(x_n) - f(y_n) + f(y_n) - f(z_n) + f(z_n) - f(x)$, we obtain that $f(x_n) - f(x) = f(y_n) - f(z_n) + o(d_n)$. By differentiability of f on $\mathcal{O}$ and continuity of f' at x,

$$f(y_n) - f(z_n) = \langle f'(z_n), y_n - z_n \rangle + o(\|y_n - z_n\|) = \langle f'(x), x_n - x \rangle + o(d_n)$$

which proves the lemma. □

We now complete the proof of the Lemma. We consider $\tilde{\psi} \in \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$, and establish that, considered as a function of the real and imaginary parts of $\mathbf{W}_N$, function $\frac{1}{M} \mathrm{Tr}\, \tilde{\psi}(\hat{\mathbf{\Omega}}_N)$ is continuously differentiable on $\mathbb{R}^{2MN}$, i.e. that for each pair (i, j), the partial derivatives $\frac{\partial}{\partial W_{i,j,N}} \left\{ \frac{1}{M} \mathrm{Tr}\, \tilde{\psi}\left(\hat{\mathbf{\Omega}}_N\right) \right\}$ exist, and are continuous [3]. We denote by $\mathcal{O}$ the open subset of $\mathbb{R}^{2MN}$ for which the eigenvalues $(\hat{\lambda}_{l,N})_{l=1,\ldots,M}$ of $\mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$ have multiplicity 1. It is clear that $\mathbb{R}^{2MN} \backslash \mathcal{O}$ has a zero Lebesgue measure. On $\mathcal{O}$, it is standard that the eigenvalues $(\hat{\lambda}_{l,N})_{l=1,\ldots,M}$ are $\mathcal{C}^1$ functions (see lemma 1.3.2) and that

$$\frac{\partial \hat{\lambda}_{l,N}}{\partial W_{i,j,N}} = \left[\mathbf{\Sigma}_N^* \hat{\mathbf{u}}_{l,N} \hat{\mathbf{u}}_{l,N}^* \right]_{j,i}. \tag{4.72}$$

Using again lemma 1.3.2 and the very definition of $\hat{\mathbf{\Omega}}_N$ (see (4.7)), we obtain

$$\frac{\partial}{\partial W_{i,j,N}} \left\{ \mathrm{Tr}\, \tilde{\psi}(\hat{\mathbf{\Omega}}_N) \right\} = \mathrm{Tr} \left(\tilde{\psi}'(\hat{\mathbf{\Omega}}_N) \frac{\partial}{\partial W_{i,j,N}} \{\hat{\mathbf{\Omega}}_N\} \right) = \left[\mathbf{\Sigma}_N^* \sum_{l=1}^{M} [\tilde{\psi}'(\hat{\mathbf{\Omega}}_N)]_{ll} \hat{\mathbf{u}}_{l,N} \hat{\mathbf{u}}_{l,N}^* \right]_{j,i} \tag{4.73}$$

and get that $\frac{1}{M} \mathrm{Tr}\, \tilde{\Psi}(\hat{\mathbf{\Omega}}_N)$ is a $\mathcal{C}^1$ on $\mathcal{O}$. By Lemma 4.5.4, it remains to establish that the righthandside of (4.73) can be continuously extended to any point $\mathbf{W}_N^0$ of $\mathbb{R}^{2MN} \backslash \mathcal{O}$. For this, we first prove the following useful result.

Lemma 4.5.5. *If $\hat{\lambda}_{k,N} = \hat{\lambda}_{l,N}$, then $[\tilde{\psi}(\hat{\mathbf{\Omega}}_N)]_{kk} = [\tilde{\psi}(\hat{\mathbf{\Omega}}_N)]_{ll}$.*

Proof. We start by observing that for any integers $m_1, m_2, \ldots, m_t$, matrix $\mathbf{A} = \hat{\mathbf{\Lambda}}_N^{m_1} \mathbf{1}\mathbf{1}^T \hat{\mathbf{\Lambda}}_N^{m_2} \cdots \mathbf{1}\mathbf{1}^T \hat{\mathbf{\Lambda}}_N^{m_t}$ writes

$$\mathbf{A} = \begin{bmatrix} \hat{\lambda}_{1,N}^{m_1} & \cdots & \hat{\lambda}_{1,N}^{m_1} \\ \vdots & \vdots & \vdots \\ \hat{\lambda}_{M,N}^{m_1} & \cdots & \hat{\lambda}_{M,N}^{m_1} \end{bmatrix} \cdots \begin{bmatrix} \hat{\lambda}_{1,N}^{m_{t-1}} & \cdots & \hat{\lambda}_{1,N}^{m_{t-1}} \\ \vdots & \vdots & \vdots \\ \hat{\lambda}_{M,N}^{m_{t-1}} & \cdots & \hat{\lambda}_{M,N}^{m_{t-1}} \end{bmatrix} \hat{\mathbf{\Lambda}}_N^{m_t}$$

hence $[\mathbf{A}]_{kk} = [\mathbf{A}]_{ll}$ if $\hat{\lambda}_{k,N} = \hat{\lambda}_{l,N}$. The same can be said about $\mathbf{1}\mathbf{1}^T\mathbf{A}$ and $\mathbf{A}\mathbf{1}\mathbf{1}^T$. Consequently, the result of the lemma is true when $\tilde{\psi}$ is a polynomial. Since any continuous function $\tilde{\psi}$ is the uniform limit of a sequence of polynomials on compact subsets of $\mathbb{R}$, the result is true for such $\tilde{\psi}$. □

[3] $\tilde{\psi}$ is real valued, the partial derivatives w.r.t. $\overline{W}_{i,j,N}$ thus coincide with the complex conjugate of the partial derivative w.r.t. $W_{i,j,N}$. It is therefore sufficient to consider these derivatives.

We consider an element $\mathbf{W}_N^0$ of $\mathbb{R}^{2MN}\backslash\mathscr{O}$, and denote by $m_1,\dots,m_L$, with $M=\sum_{l=1}^L m_l$, the respective multiplicities of the eigenvalues of $\boldsymbol{\Sigma}_N^0\boldsymbol{\Sigma}_N^{0*}$ where $\boldsymbol{\Sigma}_N^0=\mathbf{B}_N+\mathbf{W}_N^0$. We also denote by $(\overline{\boldsymbol{\Pi}}_{l,N})_{l=1,\dots,L}$ the orthogonal projection matrices over the corresponding eigenspaces. Lemma 4.5.5 implies that for each $i=1,\dots,L$,

$$\left[\tilde{\psi}'(\hat{\boldsymbol{\Omega}})\right]_{m_1+\dots+m_i,m_1+\dots+m_i}=\dots=\left[\tilde{\psi}'(\hat{\boldsymbol{\Omega}})\right]_{m_1+\dots+m_i+m_{i+1}-1,m_1+\dots+m_i+m_{i+1}-1}=\kappa_i.$$

Therefore, for any sequence $(\mathbf{W}_{N,n})_{n\in\mathbb{N}}$ converging torward $\mathbf{W}_N^0$, it holds that

$$\lim_{n\to\infty}\frac{\partial}{\partial W_{i,j,N}}\left\{\frac{1}{M}\mathrm{Tr}\,\tilde{\psi}\left(\hat{\boldsymbol{\Omega}}_N\right)\right\}\Big|_{\mathbf{W}_N=\mathbf{W}_{N,n}}=\left[\boldsymbol{\Sigma}_N^*\sum_{l=1}^L\kappa_l\overline{\boldsymbol{\Pi}}_{l,N}\right]_{j,i}.$$

This completes the proof of Lemma 4.5.3.

4.5.6 Proof of lemma 4.3.2

In order to shorten the notations, we denote by $\hat{g}_N(z)$ and $g_N(z)$ the functions defined by

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

and

$$g_N(z)=\mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N\frac{w_N'(z)}{1+\sigma^2c_Nm_N(z)}.$$

In order to evaluate $\mathbb{E}|\hat{g}_N(z)-g_N(z)\chi_N^2|^{2l}$, we use the Poincaré inequality. For this, we need first to differentiate the regularization factor χ_N.

Differentiation of the regularization factor

We recall that if $\mathbf{H}$ a hermitian matrix with a spectral decomposition $\mathbf{H}=\sum_l\gamma_l\mathbf{x}_l\mathbf{x}_l^*$, its adjoint (i.e. the transpose of its cofactor matrix) denoted by adj($\mathbf{H}$) is given by $\mathrm{adj}(\mathbf{H})=\sum_l(\prod_{k\neq l}\gamma_k)\mathbf{x}_l\mathbf{x}_l^*$. When $\mathbf{H}$ is invertible, $\mathrm{adj}(\mathbf{H})=\det(\mathbf{H})\mathbf{H}^{-1}$.

Lemma 4.5.6. *Assume assumptions **A-1** to **A-6** hold. Considered as functions of the real and imaginary parts of the entries of* $\mathbf{W}_N$*, functions* $\det\phi(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*)$ *and* $\det\phi(\hat{\boldsymbol{\Omega}}_N)$ *belong to* $\mathscr{C}^1(\mathbb{R}^{2MN})$*, and their partial derivatives w.r.t.* $W_{i,j,N}$ *denoted by*

$$[\mathbf{D}_1]_{i,j,N}:=\frac{\partial}{\partial W_{i,j,N}}\left\{\det\phi(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*)\right\},\quad\textit{and}\quad[\mathbf{D}_2]_{i,j,N}:=\frac{\partial}{\partial W_{i,j,N}}\left\{\det\phi(\hat{\boldsymbol{\Omega}}_N)\right\},$$

are given almost surely by

$$[\mathbf{D}_1]_{i,j,N}=\mathbf{e}_j^*\boldsymbol{\Sigma}_N^*\mathrm{adj}\left(\phi(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*)\right)\phi'(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*)\mathbf{e}_i,\tag{4.74}$$

$$[\mathbf{D}_2]_{i,j,N}=\left[\boldsymbol{\Sigma}_N^*\sum_{l=1}^M\left[\mathrm{adj}\left(\phi(\hat{\boldsymbol{\Omega}}_N)\right)\phi'(\hat{\boldsymbol{\Omega}}_N)\right]_{ll}\hat{\boldsymbol{\Pi}}_{l,N}\right]_{ji}\tag{4.75}$$

If we denote by $\mathscr{B}_{1,N}$ *and* $\mathscr{B}_{2,N}$ *the events defined by*

$$\mathscr{B}_{1,N}=\left\{\exists k:\hat{\lambda}_{k,N}\notin\mathscr{T}_\epsilon\right\}\cap\left\{\hat{\lambda}_{1,N},\dots,\hat{\lambda}_{M,N}\in\mathrm{supp}(\varphi)\right\},\quad\textit{and}\quad\mathscr{B}_{2,N}=\left\{\exists k:\hat{\omega}_{k,N}\notin\mathscr{T}_\epsilon\right\}\cap\left\{\hat{\omega}_{1,N},\dots,\hat{\omega}_{M,N}\in\mathrm{supp}(\varphi)\right\}.$$

then $[\mathbf{D}_1]_{i,j,N}=0$ *on* $\mathscr{B}_{1,N}^c$ *and* $[\mathbf{D}_2]_{i,j,N}=0$ *on* $\mathscr{B}_{2,N}^c$.

Proof. We first establish that $\det\varphi(\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*)$ is a $\mathscr{C}^1$ function, and that (4.74) holds. We use the same approach as in Haagerup & Thorbjornsen [22, Lem. 4.6]. We start begin by showing that the differential of $\det\phi(\mathbf{X})$ is given by

$$\det\varphi(\mathbf{X})'.\mathbf{H}=\mathrm{Tr}\left(\mathrm{adj}(\varphi(\mathbf{X}))\varphi'(\mathbf{X})\mathbf{H}\right).\tag{4.76}$$

As $\det(\mathbf{X})'.\mathbf{H}=\mathrm{Tr}\,(\mathrm{adj}(\mathbf{X})\mathbf{H})$ and $(\mathbf{X}^n)'.\mathbf{H}=\sum_{i=0}^{n-1}\mathbf{X}^i\mathbf{H}\mathbf{X}^{n-1-i}$ for any $n\in\mathbb{N}$, we have

$$\det(\mathbf{X}^n)'.\mathbf{H}=\mathrm{Tr}\left(\mathrm{adj}(\mathbf{X}^n)(n\mathbf{X}^{n-1})\mathbf{H}\right)$$

since $\mathrm{adj}(\mathbf{X}^n)$ and $\mathbf{X}$ commute. So (4.76) is true when φ is a polynomial. By choosing a sequence of polynomials P_n such that $P_n \to \varphi$ and $P_n' \to \varphi'$ uniformly on compact subsets of $\mathbb{R}$, we generalize (4.76) to any $\varphi \in \mathscr{C}^1$. Now one can check that

$$\frac{\partial(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*)}{\partial W_{i,j,N}} = \mathbf{e}_i\mathbf{e}_j^*\mathbf{\Sigma}_N^*,$$

and it remains to apply the composition formula for differentials to obtain (4.74).

We also remark that at a point $\mathbf{W}_N$ for which there exists a $\hat{\lambda}_{l,N} \notin \mathrm{supp}(\varphi)$, we have

$$\mathrm{adj}\left(\varphi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N{}^*)\right)\varphi'(\mathbf{\Sigma}_N\mathbf{\Sigma}_N{}^*) = \sum_{l=1}^{M}\Big(\prod_{k\neq l}\varphi(\hat{\lambda}_{k,N})\Big)\varphi'(\hat{\lambda}_{l,N})\hat{\mathbf{u}}_{l,N}\hat{\mathbf{u}}_{l,N}^* = \mathbf{0} \tag{4.77}$$

hence the derivative (4.74) is zero on $\mathscr{B}_{1,N}^c$.

It is easy to check that $\det\varphi(\hat{\mathbf{\Omega}}_N)$ is a $\mathscr{C}^1$ function on the open set $\mathscr{O}$ of all matrices $\mathbf{W}_N$ for which the eigenvalues of $\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*$ are simple, and that (4.75) holds if $\mathbf{W}_N \in \mathscr{O}$, i.e. on a set of probability 1. In order to show that $\det\varphi(\hat{\mathbf{\Omega}}_N)$ is a $\mathscr{C}^1$ function on $\mathbb{R}^{2MN}\backslash\mathscr{O}$, we use again lemma 4.5.4, and verify that (4.75) can be continuously extended to $\mathbb{R}^{2MN}\backslash\mathscr{O}$. For this, we claim that

$$\left[\mathrm{adj}(\varphi(\hat{\mathbf{\Omega}}_N))\varphi'(\hat{\mathbf{\Omega}}_N)\right]_{k,k} = \left[\mathrm{adj}(\varphi(\hat{\mathbf{\Omega}}_N))\varphi'(\hat{\mathbf{\Omega}}_N)\right]_{l,l} \tag{4.78}$$

if $\hat{\lambda}_{k,N} = \hat{\lambda}_{l,N}$. Indeed, given $\varepsilon > 0$, let $\varphi_\varepsilon(x) = \varphi(x) + \varepsilon$. Since $\varphi_\varepsilon(\hat{\mathbf{\Omega}}_N) > 0$,

$$\mathrm{adj}(\varphi_\varepsilon(\hat{\mathbf{\Omega}}_N))\varphi_\varepsilon'(\hat{\mathbf{\Omega}}_N) = \det\left(\varphi_\varepsilon(\hat{\mathbf{\Omega}}_N)\right)\varphi_\varepsilon^{-1}(\hat{\mathbf{\Omega}}_N)\varphi_\varepsilon'(\hat{\mathbf{\Omega}}_N).$$

Applying lemma 4.5.5 to $\tilde{\psi} = \varphi_\varepsilon^{-1} \times \varphi_\varepsilon'$, we obtain that

$$\left[\mathrm{adj}(\varphi_\varepsilon(\hat{\mathbf{\Omega}}_N))\varphi_\varepsilon'(\hat{\mathbf{\Omega}}_N)\right]_{kk} = \left[\mathrm{adj}(\varphi_\varepsilon(\hat{\mathbf{\Omega}}_N))\varphi_\varepsilon'(\hat{\mathbf{\Omega}}_N)\right]_{ll} \quad \text{if } \hat{\lambda}_{k,N} = \hat{\lambda}_{l,N} \tag{4.79}$$

and letting $\varepsilon \to 0$, we obtain the same result for $\mathrm{adj}(\varphi(\hat{\mathbf{\Omega}}_N))\varphi'(\hat{\mathbf{\Omega}}_N)$. Similarly to the proof of lemma 4.5.3, this proves that (4.75) can be continuously extended to $\mathbb{R}^{2MN}\backslash\mathscr{O}$. □

End of the proof

We now establish (4.34) by induction on l, and first consider the case $l = 1$. We recall that from the bounds (2.7), (2.8) and property 2.3.1, we have

$$\sup_{z\in\partial\mathscr{R}_y}\|\mathbf{T}_N(z)\|,\ \sup_{z\in\partial\mathscr{R}_y}\left|\frac{1}{1+\sigma^2 c_N m_N(z)}\right|,\ \sup_{z\in\partial\mathscr{R}_y}\left|\frac{w_N'(z)}{1+\sigma^2 c_N m_N(z)}\right| \le C \tag{4.80}$$

where $C > 0$ is independent of N. Moreover, since $\hat{m}_N(z), \|\mathbf{Q}_N(z)\| \le \mathrm{dist}\left(z, \{\hat{\lambda}_{1,N},\ldots,\hat{\lambda}_{M,N}\}\right)^{-1}$ and from the very definition of event $\mathscr{A}_{2,N}$, it also holds

$$\sup_{z\in\partial\mathscr{R}_y}\|\mathbf{Q}_N(z)\|\chi_N,\ \sup_{z\in\partial\mathscr{R}_y}\left|\frac{\chi_N}{1+\sigma^2 c_N \hat{m}_N(z)}\right|,\ \sup_{z\in\partial\mathscr{R}_y}\left|\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N \hat{m}_N(z)}\right|\chi_N \le C. \tag{4.81}$$

In the following, C will be a positive constant independent of N whose value may change from one line to another.

We write the second moment of $(\hat{g}_N(z) - g_N(z))\chi_N^2$ as

$$\mathbb{E}\left|(\hat{g}_N(z) - g_N(z))\chi_N^2\right|^2 = \left|\mathbb{E}\left((\hat{g}_N(z) - g_N(z))\chi_N^2\right)\right|^2 + \mathrm{Var}\left((\hat{g}_N(z) - g_N(z))\chi_N^2\right).$$

We evaluate $\mathrm{Var}\left[(\hat{g}_N(z) - g_N(z))\chi_N^2\right]$ using the Poincaré inequality and get

$$\mathrm{Var}\left[(\hat{g}_N(z) - g_N(z))\chi_N^2\right] \le \frac{\sigma^2}{N}\sum_{i,j}\mathbb{E}\left[\chi_N^4\left(\left|\frac{\partial\hat{g}_N(z)}{\partial W_{i,j,N}}\right|^2 + \left|\frac{\partial\hat{g}_N(z)}{\overline{W}_{i,j,N}}\right|^2\right)\right] + 2\mathbb{E}\left[|\hat{g}_N(z) - g_N(z)|^2\left|\frac{\partial\chi_N^2}{\partial W_{i,j,N}}\right|^2\right]. \tag{4.82}$$

It is clear that

$$\frac{\partial\hat{g}_N(z)}{\partial W_{i,j,N}} = \mathbf{d}_N^*\frac{\partial\mathbf{Q}_N(z)}{\partial W_{i,j,N}}\mathbf{d}_N\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N\hat{m}_N(z)} + \mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N\frac{\partial}{\partial W_{i,j,N}}\left\{\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N\hat{m}_N(z)}\right\}.$$

We verify that

$$\mathbf{d}_N^* \frac{\partial \mathbf{Q}_N(z)}{\partial W_{i,j,N}} \mathbf{d}_N = -\mathbf{d}_N^* \mathbf{Q}_N(z) \mathbf{e}_i \, \mathbf{e}_j \boldsymbol{\Sigma}_N^* \mathbf{Q}_N(z) \mathbf{d}_N,$$

so that

$$\sum_{i,j} \left| \mathbf{d}_N^* \frac{\partial \mathbf{Q}_N(z)}{\partial W_{i,j,N}} \mathbf{d}_N \right|^2 = \mathbf{d}_N^* \mathbf{Q}_N(z) \mathbf{Q}_N(z)^* \mathbf{d}_N \, \mathbf{d}_N^* \mathbf{Q}_N(z) \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* \mathbf{Q}_N(z)^* \mathbf{d}_N.$$

As $\chi_N \neq 0$ implies that $\hat{\lambda}_{M,N} = \|\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*\| \leq t_2^+ + 2\epsilon$, (4.81) implies that

$$\sup_{z \in \partial \mathscr{R}_y} \chi_N^2 \, \mathbf{d}_N^* \mathbf{Q}_N(z) \mathbf{Q}_N(z)^* \mathbf{d}_N \, \mathbf{d}_N^* \mathbf{Q}_N(z) \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* \mathbf{Q}_N(z)^* \mathbf{d}_N \leq C.$$

Using again (4.81), we get that

$$\sup_{z \in \partial \mathscr{R}_y} \chi_N^4 \left| \frac{\hat{w}_N'(z)}{1 + \sigma^2 c_N \hat{m}_N(z)} \right|^2 \sum_{i,j} \left| \mathbf{d}_N^* \frac{\partial \mathbf{Q}_N(z)}{\partial W_{i,j,N}} \mathbf{d}_N \right|^2 \leq C. \tag{4.83}$$

We obtain similarly that

$$\sup_{z \in \partial \mathscr{R}_y} \chi_N^4 |\mathbf{d}_N^* \mathbf{Q}_N(z) \mathbf{d}_N|^2 \sum_{i,j} \left| \frac{\partial}{\partial W_{i,j,N}} \left\{ \frac{\hat{w}_N'(z)}{1 + \sigma^2 c_N \hat{m}_N(z)} \right\} \right|^2 \leq \frac{C}{N}. \tag{4.84}$$

The same conclusions hold when the derivatives w.r.t. variables $\overline{W}_{i,j,N}$ are considered. This shows that the first term of the r.h.s. of (4.82) is a $\mathcal{O}\left(\frac{1}{N}\right)$ term. We now evaluate the behaviour of the second term of the r.h.s. of (4.82), and establish that

$$\sup_{z \in \partial \mathscr{R}_y} \mathbb{E}\left[|\hat{g}_N(z) - g_N(z)|^2 \sum_{i,j} \left| \frac{\partial \chi_N^2}{\partial W_{i,j,N}} \right|^2 \right] = \mathcal{O}\left(\frac{1}{N^p}\right) \tag{4.85}$$

for each integer p. We express $\frac{\partial \chi_N^2}{\partial W_{i,j,N}}$ as $2\chi_N \frac{\partial \chi_N}{\partial W_{i,j,N}}$. (4.80) and (4.81) imply that $\sup_{z \in \partial \mathscr{R}_y} \chi_N^2 |\hat{g}_N(z) - g_N(z)|^2 < C$. Therefore, it is sufficient to check that

$$\mathbb{E}\left[\sum_{i,j} \left| \frac{\partial \chi_N}{W_{i,j,N}} \right|^2 \right] = \mathcal{O}\left(\frac{1}{N^p}\right)$$

for each integer p. $\frac{\partial \chi_N}{\partial W_{i,j,N}}$ can be written as

$$\frac{\partial \chi_N}{\partial W_{i,j,N}} = [\mathbf{D}_1]_{i,j,N} \det \phi(\hat{\boldsymbol{\Omega}}_N) + [\mathbf{D}_2]_{i,j,N} \det \phi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*).$$

It holds that

$$\mathbb{E}\left[\sum_{i,j} \left| [\mathbf{D}_1]_{i,j,N} \det \phi(\hat{\boldsymbol{\Omega}}_N)) \right|^2 \right] = \mathbb{E}\left[\det \phi(\hat{\boldsymbol{\Omega}}_N)^2 \operatorname{Tr}\left(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* \phi'(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*)^2 \operatorname{adj}\left(\phi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N{}^*) \right)^2 \right) \mathbb{1}_{\mathscr{B}_{1,N}} \right]$$

Moreover, we can write

$$\operatorname{Tr}\left(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* \phi'(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*)^2 \operatorname{adj}\left(\phi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N{}^*) \right)^2 \right) = \sum_k \hat{\lambda}_{k,N} \phi'(\hat{\lambda}_{k,N})^2 \prod_{l \neq k} \phi(\hat{\lambda}_{l,N})^2 \leq \operatorname{Tr}\left(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* \phi'(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*)^2 \right),$$

because $\phi(\lambda) \leq 1$ on $\mathbb{R}$. Therefore, it holds that

$$\mathbb{E}\left[\sum_{i,j} \left| [\mathbf{D}_1]_{i,j,N} \det \phi(\hat{\boldsymbol{\Omega}}_N) \right|^2 \right] \leq \mathbb{E}\left[\det \phi(\hat{\boldsymbol{\Omega}}_N)^2 \operatorname{Tr}\left(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* \phi'(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*)^2 \right) \mathbb{1}_{\mathscr{A}_{1,N}} \right] \leq C N \mathbb{P}(\mathscr{B}_{1,N}),$$

because $\det \phi(\hat{\boldsymbol{\Omega}}_N)) \leq 1$, and $\operatorname{Tr}\left(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^* \phi'(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^*)^2 \right) \leq CN$ on $\mathscr{B}_{1,N}$. As $\mathscr{B}_{1,N} \subset \mathscr{A}_{1,N}$, (4.30) implies that

$$\mathbb{E}\left[\sum_{i,j} \left| [\mathbf{D}_1]_{i,j,N} \det \phi(\hat{\boldsymbol{\Omega}}_N)) \right|^2 \right] = \mathcal{O}\left(\frac{1}{N^p}\right)$$

for each integer p. Using similar calculations and (4.31), we obtain that

$$\mathbb{E}\left[\sum_{i,j}\left|[\mathbf{D}_2]_{i,j,N}\det\phi(\mathbf{\Sigma}_N\mathbf{\Sigma}_N^*))\right|^2\right] = \mathscr{O}\left(\frac{1}{N^p}\right)$$

for each integer p. This completes the proof of (4.85) and establishes that

$$\sup_{z\in\partial\mathscr{R}_y} \operatorname{Var}\left[(\hat{g}_N(z) - g_N(z))\chi_N^2\right) = \mathscr{O}\left(\frac{1}{N}\right].$$

In order to evaluate the term $\left|\mathbb{E}\left[(\hat{g}_N(z) - g_N(z))\chi_N^2\right]\right|^2$, we also need the following auxilliary lemma proved at the end of this section.

Lemma 4.5.7. *It holds that*

$$\sup_{z\in\partial\mathscr{R}_y}\left|\mathbb{E}\left[\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N\chi_N - \mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N\right]\right| = \mathscr{O}\left(\frac{1}{N^{3/2}}\right), \tag{4.86}$$

$$\sup_{z\in\partial\mathscr{R}_y}\left|\mathbb{E}\left[\hat{m}_N(z)\chi_N - m_N(z)\right]\right| = \mathscr{O}\left(\frac{1}{N^2}\right), \tag{4.87}$$

$$\sup_{z\in\partial\mathscr{R}_y}\left|\mathbb{E}\left[\hat{m}_N'(z)\chi_N - m_N'(z)\right]\right| = \mathscr{O}\left(\frac{1}{N^2}\right). \tag{4.88}$$

We express $(\hat{g}_N(z) - g_N(z))\chi_N^2$ as $\beta_{1,N}(z) + \beta_{2,N}(z)$ where

$$\beta_{1,N}(z) = \chi_N\left(\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N - \mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N\right)\frac{\hat{w}_N'(z)\chi_N}{1+\sigma^2 c_N\hat{m}_N(z)} \tag{4.89}$$

and

$$\beta_{2,N}(z) = \chi_N^2\mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N\left(\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N\hat{m}_N(z)} - \frac{w_N'(z)}{1+\sigma^2 c_N m_N(z)}\right),$$

and establish that

$$\sup_{z\in\partial\mathscr{R}_y}\mathbb{E}|\beta_{1,N}|^2 = \mathscr{O}\left(\frac{1}{N}\right) \quad \text{and} \quad \sup_{z\in\partial\mathscr{R}_y}\mathbb{E}|\beta_{2,N}|^2 = \mathscr{O}\left(\frac{1}{N^2}\right). \tag{4.90}$$

Using (4.81), (4.90) for $\beta_{1,N}$ will be established if we show that

$$\sup_{z\in\partial\mathscr{R}_y}\mathbb{E}\left|\chi_N\left(\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N - \mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N\right)\right|^2 = \mathscr{O}\left(\frac{1}{N}\right).$$

For this, we write that

$$\mathbb{E}\left|\chi_N(\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N - \mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N)\right|^2 = \operatorname{Var}(\chi_N\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N) + \left|\mathbb{E}\left(\chi_N(\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N - \mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N)\right)\right|^2.$$

The above calculations prove that $\sup_{z\in\partial\mathscr{R}_y}\operatorname{Var}[\chi_N\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N] = \mathscr{O}\left(\frac{1}{N}\right)$, while (4.86) and $1-\mathbb{E}(\chi_N) = \mathscr{O}(\frac{1}{N^p})$ for each p imply that

$$\mathbb{E}\left[\chi_N(\mathbf{d}_N^*\mathbf{Q}_N(z)\mathbf{d}_N - \mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N)\right] = \mathscr{O}\left(\frac{1}{N^{3/2}}\right).$$

This completes the proof of (4.90) for $\beta_{1,N}$. In order to show (4.90) for $\beta_{2,N}$, we first remark that from (4.80), $|\mathbf{d}_N^*\mathbf{T}_N(z)\mathbf{d}_N|$ is uniformly bounded on $\partial\mathscr{R}_y$, and write that

$$\chi_N^2\left(\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N\hat{m}_N(z)} - \frac{w_N'(z)}{1+\sigma^2 c_N m_N(z)}\right) =$$
$$\sigma^2 c_N\chi_N^2\left(\hat{m}_N(z) - m_N(z)\right) + 2z\sigma^2 c_N\chi_N^2\left(\hat{m}_N'(z) - m_N'(z)\right) - \sigma^4 c_N(1-c_N)\chi_N^2\left(\frac{\hat{m}_N'(z)}{1+\sigma^2 c_N\hat{m}_N(z)} - \frac{m_N'(z)}{1+\sigma^2 c_N m_N(z)}\right),$$

or equivalently that

$$\begin{aligned}
&\chi_N^2\left(\frac{\hat{w}_N'(z)}{1+\sigma^2 c_N \hat{m}_N(z)} - \frac{w_N'(z)}{1+\sigma^2 c_N m_N(z)}\right) = \\
&\quad \sigma^2 c_N \chi_N^2 \left(\hat{m}_N(z) - m_N(z)\right) + 2z\sigma^2 c_N \chi_N^2 \left(\hat{m}_N'(z) - m_N'(z)\right) \\
&\quad -(\sigma^2 c_N)^2 \sigma^2 (1-c_N) \frac{\chi_N}{(1+\sigma^2 c_N \hat{m}_N(z))(1+\sigma^2 c_N m_N(z))} \left[m_N(z)\chi_N(\hat{m}_N'(z) - m_N'(z)) - m_N'(z)\chi_N(\hat{m}_N(z) - m_N(z))\right] \\
&\quad -\sigma^2 c_N \sigma^2 (1-c_N) \frac{\chi_N}{(1+\sigma^2 c_N \hat{m}_N(z))(1+\sigma^2 c_N m_N(z))} \left[\chi_N\left(\hat{m}_N'(z) - m_N'(z)\right)\right].
\end{aligned}$$

The Poincaré inequality and Lemma 4.5.7 imply that

$$\sup_{z\in\partial\mathscr{R}_y} \mathbb{E}\left|\chi_N(\hat{m}_N(z) - m_N(z))\right|^2 = \mathscr{O}\left(\frac{1}{N^2}\right)$$

and

$$\sup_{z\in\partial\mathscr{R}_y} \mathbb{E}\left|\chi_N(\hat{m}_N'(z) - m_N'(z))\right|^2 = \mathscr{O}\left(\frac{1}{N^2}\right).$$

Eq. (4.90) follows immediately from

$$\sup_{z\in\partial\mathscr{R}_y} \left|\frac{\chi_N}{(1+\sigma^2 c_N \hat{m}_N(z))(1+\sigma^2 c_N m_N(z))}\right| \le C,$$

(see (4.80) and (4.81)). This completes the proof of (4.34) for $l = 1$.

We now assume that (4.34) holds until integer $l-1$ and write that

$$\mathbb{E}\left|\chi_N^2\left(\hat{g}_N(z) - g_N(z)\right)\right|^{2l} = \left|\mathbb{E}\left[\left(\chi_N^2(\hat{g}_N(z) - g_N(z))\right)^l\right]\right|^2 + \mathrm{Var}\left[\left(\chi_N^2(\hat{g}_N(z) - g_N(z))\right)^l\right].$$

The Cauchy-Schwarz inequality implies that

$$\left|\mathbb{E}\left[\left(\chi_N^2(\hat{g}_N(z) - g_N(z))\right)^l\right]\right|^2 \le \mathbb{E}\left|\chi_N^2(\hat{g}_N(z) - g_N(z))\right|^2 \mathbb{E}\left|\chi_N^2(\hat{g}_N(z) - g_N(z))\right|^{2(l-1)},$$

and shows that

$$\sup_{z\in\partial\mathscr{R}_y} \left|\mathbb{E}\left(\chi_N^2(\hat{g}_N(z) - g_N(z))\right)^l\right|^2 = \mathscr{O}\left(\frac{1}{N^l}\right).$$

The Poincaré inequality gives

$$\begin{aligned}
\mathrm{Var}\left[\left(\chi_N^2(\hat{g}_N(z) - g_N(z))\right)^l\right] &\le \frac{\sigma^2 l^2}{N}\mathbb{E}\left[\left|\chi_N^2(\hat{g}_N(z) - g_N(z))\right|^{2(l-1)} \chi_N^4 \sum_{i,j}\left(\left|\frac{\partial \hat{g}_N(z)}{\partial W_{i,j,N}}\right|^2 + \left|\frac{\partial \hat{g}_N(z)}{\partial \overline{W}_{i,j,N}}\right|^2\right)\right] \\
&\quad + \frac{8\sigma^2 l^2}{N}\mathbb{E}\left[\left|\hat{g}_N(z) - g_N(z)\right|^{2l} \chi_N^{4l-2} \sum_{i,j}\left|\frac{\partial \chi_N}{\partial W_{i,j,N}}\right|^2\right].
\end{aligned}$$

Finally, (4.83) and (4.84) imply that

$$\sup_{z\in\partial\mathscr{R}_y} \chi_N^4 \sum_{i,j}\left(\left|\frac{\partial \hat{g}_N(z)}{\partial W_{i,j,N}}\right|^2 + \left|\frac{\partial \hat{g}_N(z)}{\partial \overline{W}_{i,j,N}}\right|^2\right) \le C,$$

for some deterministic constant C. Therefore, the supremum over $z \in \partial\mathscr{R}_y$ of first term of the r.h.s. of (4.91) is a $\mathscr{O}\left(\frac{1}{N^l}\right)$. Moreover, it can be shown as in the case $l = 1$ that the supremum over $z \in \partial\mathscr{R}_y$ of the second term of the righthandside of (4.91) is a $\mathscr{O}\left(\frac{1}{N^p}\right)$ for each integer p. This completes the proof of lemma 4.3.2.

Proof of auxiliary lemma 4.5.7

In this section, we denote by $\alpha_{r,N}(z), \tilde{\alpha}_{r,N}(z), \mathbf{R}_{r,N}(z)$ and $\tilde{\mathbf{R}}_{r,N}(z)$ the regularized versions of the respective functions $\alpha_N(z), \tilde{\alpha}_N(z), \mathbf{R}_N(z)$ and $\tilde{\mathbf{R}}_N(z)$ (introduced in appendix 2.7.2), i.e.

$$\alpha_{r,N}(z) = \sigma\mathbb{E}\left(\frac{1}{N}\mathrm{Tr}\,(\mathbf{Q}_N(z))\chi_N\right) \quad \text{and} \quad \tilde{\alpha}_{r,N}(z) = \sigma\mathbb{E}\left(\frac{1}{N}\mathrm{Tr}\,(\tilde{\mathbf{Q}}_N(z))\chi_N\right),$$

and

$$\mathbf{R}_{r,N}(z) = \left(\frac{\mathbf{B}_N\mathbf{B}_N^*}{1+\sigma\alpha_{r,N}(z)} - z(1+\sigma\tilde{\alpha}_{r,N}(z))\right)^{-1}, \tilde{\mathbf{R}}_{r,N}(z) = \left(\frac{\mathbf{B}_N^*\mathbf{B}_N}{1+\sigma\tilde{\alpha}_{r,N}(z)} - z(1+\sigma\alpha_{r,N}(z))\right)^{-1}.$$

It is clear that $\alpha_{r,N}$ and $\tilde{\alpha}_{r,N}$ are the Stieltjes transforms of positive measures carried by $\mathbb{C}\backslash\mathrm{supp}(\varphi)$ and $\mathbb{C}^*\backslash\mathrm{supp}(\varphi)$ respectively and with mass $\sigma c_N \mathbb{E}[\chi_N]$ and $\sigma\mathbb{E}[\chi_N]$. This implies that the following uniform bounds hold: Let $\mathscr{K}$ and $\tilde{\mathscr{K}}$ be compact subsets of $\mathbb{C}\backslash\mathrm{supp}(\varphi)$ and $\mathbb{C}^*\backslash\mathrm{supp}(\varphi)$ respectively, then we have

$$\sup_{z\in\mathscr{K}} |\alpha_{r,N}(z)| < C \quad \text{and} \quad \sup_{z\in\tilde{\mathscr{K}}} |\tilde{\alpha}_{r,N}(z)| < C. \tag{4.91}$$

In order to establish lemma 4.5.7, it is necessary to show that similar bounds hold for functions $\frac{1}{1+\sigma\alpha_{r,N}(z)}$, $\|\mathbf{R}_{r,N}(z)\|$ and $\|\tilde{\mathbf{R}}_{r,N}(z)\|$. For this, we introduce function $w_{r,N}(z) = z(1+\sigma\alpha_{r,N}(z))(1+\sigma\tilde{\alpha}_{r,N}(z))$ and prove the following lemma

Lemma 4.5.8. *For any compact subset $\mathscr{K}$ of $\mathbb{C}\backslash\mathrm{supp}(\varphi)$, it holds that*

$$\sup_{z\in\mathscr{K}} \left|\alpha_{r,N}(z) - \delta_N(z)\right| \xrightarrow[N\to\infty]{} 0, \tag{4.92}$$

$$\inf_{z\in\mathscr{K}} \min_{k=1,\dots,M} \left|\lambda_{k,N} - w_{r,N}(z)\right| > C > 0. \tag{4.93}$$

Proof. Define $\kappa_N(z) := \alpha_{r,N}(z) - \delta_N(z)$ where we recall that $\delta_N(z) = \sigma c_N m_N(z) = \frac{\sigma}{N}\mathrm{Tr}\,(\mathbf{T}_N(z))$. Since $\delta_N(z)$ and $\alpha_{r,N}(z)$ are Stieltjes transforms of positive measures carried by $\mathbb{C}\backslash\mathrm{supp}(\varphi)$, κ_N is holomorphic on $\mathbb{C}\backslash\mathrm{supp}(\varphi)$ and satisfies

$$|\kappa_N(z)| \le \frac{C}{\mathrm{d}(z,\mathrm{supp}(\varphi))}.$$

This implies that the sequence (κ_N) is uniformly bounded on each compact subset of $\mathbb{C}\backslash\mathrm{supp}(\varphi)$. By Montel's theorem, (κ_N) is a normal family. Let $(\kappa_{\psi(N)})$ a subsequence of (κ_N) which converges uniformly to κ on each compact subset of $\mathbb{C}\backslash\mathrm{supp}(\varphi))$. Then κ is holomorphic on $\mathbb{C}\backslash\mathrm{supp}(\varphi)$. From theorem 2.2.2, $\mathbb{E}\left[\frac{1}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\right] - \frac{1}{N}\mathrm{Tr}\,\mathbf{T}_N(z) \xrightarrow[N]{} 0$ for $z\in\mathbb{C}\backslash\mathbb{R}^+$ and since $\chi_N \to_N 1$ a.s., dominated convergence theorem implies

$$\kappa_N(z) = \mathbb{E}\left[\frac{\sigma}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\chi_N\right] - \frac{\sigma}{N}\mathrm{Tr}\,\mathbf{T}_N(z) \xrightarrow[N]{} 0$$

for $z\in\mathbb{C}\backslash\mathbb{R}^+$. Thus, $\kappa(z)=0$ for $z\in\mathbb{C}\backslash\mathbb{R}^+$, and by analytic continuation, $\kappa(z)=0$ for all $z\in\mathbb{C}\backslash\mathrm{supp}(\varphi)$. Therefore, all converging subsequences extracted from the normal family $(\kappa_N(z))$ converge to 0 uniformly on each compact subset of $\mathbb{C}\backslash\mathrm{supp}(\varphi)$. Consequently, the whole sequence (κ_N) converges uniformly to 0 on each compact subset of $\mathbb{C}\backslash\mathrm{supp}(\varphi)$. This completes the proof of (4.92). We also notice that

$$\tilde{\alpha}_{r,N}(z) = \alpha_{r,N}(z) - \frac{\sigma(1-c_N)}{z} + \frac{\sigma(1-c_N)}{z}\left(1-\mathbb{E}(\chi_N)\right) \tag{4.94}$$

and recall that $\tilde{\delta}_N(z) = \delta_N(z) - \frac{\sigma(1-c_N)}{z}$. As $1-\mathbb{E}(\chi_N) = \mathcal{O}\left(\frac{1}{N^p}\right)$ for each integer p, (4.92) implies

$$\sup_{z\in\mathscr{K}} \left|z(\tilde{\alpha}_{r,N}(z) - \tilde{\delta}_N(z))\right| \to 0.$$

Hence, it holds that

$$\sup_{z\in\mathscr{K}} \left|w_{r,N}(z) - w_N(z)\right| \to 0.$$

Thus, (4.93) follows immediately from

$$\min_{k=1,\dots,M} \left|\lambda_{k,N} - w_N(z)\right| \ge \frac{1}{2}\mathrm{dist}(z,\mathscr{S}_N), \tag{4.95}$$

as a consequence of

$$\|\mathbf{T}_N(z)\| = \frac{|1+\sigma^2 c_N m_N(z)|}{\min_{k=1,\dots,M}\left|\lambda_{k,N} - w_N(z)\right|} \le \frac{1}{\mathrm{dist}(z,\mathscr{S}_N)}.$$

(see (2.8)) and property 2.3.1. □

Lemma 4.5.8 immediately implies that the following uniform bounds hold.

Lemma 4.5.9. *Let $\mathscr{K}$ and $\tilde{\mathscr{K}}$ be compact subsets of $\mathbb{C}\backslash\mathrm{supp}(\varphi)$ and $\mathbb{C}^*\backslash\mathrm{supp}(\varphi)$ respectively. For N large enough, we have*

$$\sup_{z\in\mathscr{K}}\left|\frac{1}{1+\sigma\alpha_{r,N}(z)}\right| < C, \tag{4.96}$$

$$\sup_{z\in\mathscr{K}}\|\mathbf{R}_{r,N}(z)\| < C, \tag{4.97}$$

$$\sup_{z\in\tilde{\mathscr{K}}}\|\tilde{\mathbf{R}}_{r,N}(z)\| < C, \tag{4.98}$$

$$\sup_{z\in\mathscr{K}}\|\mathbf{R}_{r,N}(z)-\mathbf{T}_N(z)\| \to 0, \tag{4.99}$$

$$\sup_{z\in\tilde{\mathscr{K}}}\|\tilde{\mathbf{R}}_{r,N}(z)-\tilde{\mathbf{T}}_N(z)\| \to 0. \tag{4.100}$$

Proof. The uniform convergence result (4.92) together with property 2.3.1 implies that

$$\inf_{z\in\mathscr{K}}\left|1+\sigma\alpha_{r,N}(z)\right| > \frac{1}{4}$$

for N large enough. This establishes (4.96) that holds for N large enough. In order to prove (4.97), we express $\mathbf{R}_{r,N}(z)$ as

$$\mathbf{R}_{r,N}(z) = \left(1+\sigma\alpha_{r,N}(z)\right)\left(\mathbf{B}_N\mathbf{B}_N^* - w_{r,N}(z)\right)^{-1}$$

and use (4.91) and (4.93). The proof of (4.98) is similar, and is based on the identity

$$\tilde{\mathbf{R}}_{r,N}(z) = \left(1+\sigma\tilde{\alpha}_{r,N}(z)\right)\left(\mathbf{B}_N^*\mathbf{B}_N - w_{r,N}(z)\right)^{-1}.$$

We remark that function $\tilde{\alpha}_{r,N}(z)$ has a pole at $z=0$. Hence, any compact $\tilde{\mathscr{K}}$ over which $\|\tilde{\mathbf{R}}_{r,N}(z)\|$ is supposed to be uniformly bounded should not contain 0. The proof of (4.99)s follows immediately from (4.92) and from (4.96), (4.97), (4.98). Finally, to establish (4.100), we remark that

$$\tilde{\mathbf{R}}_{r,N}(z) = \frac{\mathbf{B}_N^*\mathbf{R}_N(z)\mathbf{B}_N}{w_{r,N}(z)} - \frac{\mathbf{I}_N}{1+\sigma\alpha_{r,N}(z)},$$

$$\tilde{\mathbf{T}}_{r,N}(z) = \frac{\mathbf{B}_N^*\mathbf{T}_N(z)\mathbf{B}_N}{w_N(z)} - \frac{\mathbf{I}_N}{1+\sigma\delta_N(z)},$$

and that $|w_{r,N}(z)|$ and $|w_N(z)|$ are uniformly bounded from below from (4.95) and (4.93) (recall that 0 is one of the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$). □

We now establish (4.86) and (4.87). In order to prove that $\alpha_N(z)-\delta_N(z)=\mathscr{O}\left(\frac{1}{N^2}\right)$ on $\mathbb{C}\backslash\mathbb{R}^+$, [17] used the integration by parts formula (theorem 1.4.1) and the Poincaré inequality to show that the entries of $\mathbb{E}[\mathbf{Q}_N(z)]$ are close from the entries of $\mathbf{R}_N(z)$ (see the fundamental equation (2.25) in the proof of theorem 2.2.2). Then, $\alpha_N(z)-\delta_N(z)$ was evaluated by solving a linear system whose determinant $\Delta_N(z)$ given by (4.59) was shown to be bounded from below. Lemma 4.5.9 allows to follow exactly the same approach to establish (4.86) and (4.87). However, functions $\alpha_N, \tilde{\alpha}_N, \mathbf{R}_N, \tilde{\mathbf{R}}_N$ have to be replaced by their regularized versions. The following results show that the presence of the regularization term χ_N does not modify essentially the calculations of [17]. We first indicate how the integration by parts formula is modified. Vec(.) denotes the column by column vectorization operator of a matrix.

Lemma 4.5.10. *Let $(f_N)_{N\geq 1}$ be a sequence of continuously differentiable functions defined on $\mathbb{C}^{M(M+N)}$ with polynomially bounded partial derivatives satisfying the condition*

$$\sup_{z\in\partial\mathscr{R}_y}\left|f_N\left(\mathrm{Vec}\left(\mathbf{Q}_N(z)\right),\mathrm{Vec}(\mathbf{\Sigma}_N)\right)\chi_N\right| < C.$$

Then, for all $p\in\mathbb{N}$, we have

$$\mathbb{E}\left[f\left(\mathrm{Vec}\left(\mathbf{Q}_N(z)\right),\mathrm{Vec}(\mathbf{\Sigma}_N)\right)\chi_N\right] = \mathbb{E}\left[f\left(\mathrm{Vec}\left(\mathbf{Q}_N(z)\right),\mathrm{Vec}(\mathbf{\Sigma}_N)\right)\chi_N^k\right] + \frac{\epsilon_{1,N}(z)}{N^p}. \tag{4.101}$$

for all $k \in \mathbb{N}^*$, and

$$\mathbb{E}\left[W_{ij,N} f\left(\mathrm{Vec}\left(\mathbf{Q}_N(z)\right), \mathrm{Vec}(\boldsymbol{\Sigma}_N)\right)\chi_N\right] = \frac{\sigma^2}{N}\mathbb{E}\left[\frac{\partial f\left(\mathrm{Vec}\left(\mathbf{Q}_N(z)\right), \mathrm{Vec}(\boldsymbol{\Sigma}_N)\right)}{\partial \overline{W}_{ij,N}}\chi_N\right] + \frac{\epsilon_{2,N}(z)}{N^p}, \tag{4.102}$$

$$\mathbb{E}\left[\overline{W}_{ij,N} f\left(\mathrm{Vec}\left(\mathbf{Q}_N(z)\right), \mathrm{Vec}(\boldsymbol{\Sigma}_N)\right)\chi_N\right] = \frac{\sigma^2}{N}\mathbb{E}\left[\frac{\partial f\left(\mathrm{Vec}\left(\mathbf{Q}_N(z)\right), \mathrm{Vec}(\boldsymbol{\Sigma}_N)\right)}{\partial W_{ij,N}}\chi_N\right] + \frac{\epsilon_{3,N}(z)}{N^p}, \tag{4.103}$$

with $\sup_{z\in\partial\mathcal{R}_y} |\epsilon_{i,N}(z)| \le C < \infty$.

As for the use of the Poincaré inequality, we have:

Lemma 4.5.11. *Let $(\mathbf{M}_N(z))$ a sequence of deterministic complex $M \times M$ matrix-valued functions defined on $\mathbb{C}\backslash\mathbb{R}$ such that*

$$\sup_{z\in\partial\mathcal{R}_y} \|\mathbf{M}_N(z)\| \le C.$$

Then,

$$\sup_{z\in\partial\mathcal{R}_y} \mathrm{Var}\left[\frac{1}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\mathbf{M}_N(z)\chi_N\right] \le \frac{C}{N^2},$$

and for $\mathbf{a}_N \in \mathbb{C}^M$ such that $\sup_N \|\mathbf{a}_N\| < \infty$,

$$\sup_{z\in\partial\mathcal{R}_y} \mathrm{Var}\left[\mathbf{a}_N^*\mathbf{Q}_N(z)\mathbf{M}_N(z)\mathbf{a}_N\chi_N\right] \le \frac{C}{N}.$$

Moreover, the same kind of uniform bounds still hold when $\mathbf{Q}_N(z)$ is replaced by $\mathbf{Q}_N(z)^2$.

The proofs of these results are based on elementary arguments, and are thus omitted. Following the calculations of [17], we obtain that

$$\mathbb{E}\left[\mathbf{Q}_N(z)\chi_N\right] = \mathbf{R}_{r,N}(z) + \boldsymbol{\Delta}_{r,N}(z)\mathbf{R}_{r,N}(z) + \mathbb{E}\left[\mathbf{Q}_N(z)\chi_N\right]\mathbf{R}_{r,N}(z)\frac{\sigma^2}{N}\mathrm{Tr}\,\boldsymbol{\Delta}_{r,N}(z) + \boldsymbol{\Theta}_N(z)\mathbf{R}_{r,N}(z) \tag{4.104}$$

for each $z \in \mathbb{C}\backslash\mathrm{supp}(\varphi)$ where $\boldsymbol{\Theta}_N(z)$ is a matrix whose elements are uniformly bounded on $\partial\mathcal{R}_y$ by $\frac{C}{N^p}$ for each p, and where $\boldsymbol{\Delta}_{r,N}(z)$ is the regularized version of matrix $\boldsymbol{\Delta}_N(z)$ (see (2.26)) defined by

$$\begin{aligned}\boldsymbol{\Delta}_{r,N}(z) = &-\frac{1}{(1+\sigma\alpha_{r,N}(z))^2}\mathbb{E}\left[\mathbf{Q}_N(z)\chi_N\right]\mathbb{E}\left[\left(\frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\chi_N - \mathbb{E}\left[\frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\chi_N\right]\right)\frac{\sigma^2}{N}\mathrm{Tr}\,\boldsymbol{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{B}_N\chi_N\right]\\ &+\frac{1}{1+\sigma\alpha_{r,N}(z)}\mathbb{E}\left[\left(\frac{\sigma^2}{N}\mathrm{Tr}\,\boldsymbol{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{B}_N\chi_N - \mathbb{E}\left[\frac{\sigma^2}{N}\mathrm{Tr}\,\boldsymbol{\Sigma}_N^*\mathbf{Q}_N(z)\mathbf{B}_N\chi_N\right]\right)\mathbf{Q}_N(z)\chi_N\right]\\ &+\frac{1}{1+\sigma\alpha_{r,N}(z)}\mathbb{E}\left[\left(\frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\chi_N - \mathbb{E}\left[\frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{Q}_N(z)\chi_N\right]\right)\mathbf{Q}_N(z)\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^*\chi_N\right],\end{aligned} \tag{4.105}$$

After some calculations using lemmas 4.5.9, 4.5.10, 4.5.11, we eventually obtain that

$$\sup_{z\in\partial\mathcal{R}_y} \left|\mathbf{a}_N^*\left(\mathbb{E}[\mathbf{Q}_N(z)\chi_N] - \mathbf{R}_{r,N}(z)\right)\mathbf{a}_N\right| \le \frac{C}{N^{3/2}},$$

$$\sup_{z\in\partial\mathcal{R}_y} \left|\alpha_{r,N}(z) - \frac{\sigma}{N}\mathrm{Tr}\,(\mathbf{R}_{r,N}(z))\right| \le \frac{C}{N^2}, \tag{4.106}$$

$$\sup_{z\in\partial\mathcal{R}_y} \left|\tilde{\alpha}_{r,N}(z) - \frac{\sigma}{N}\mathrm{Tr}\,(\tilde{\mathbf{R}}_{r,N}(z))\right| \le \frac{C}{N^2}. \tag{4.107}$$

for all large N. In order to prove (4.86) and (4.87), it remains to handle the terms involving the difference $\mathbf{R}_{r,N}(z) - \mathbf{T}_N(z)$. We show in the following that

$$\sup_{z\in\partial\mathcal{R}_y} \left|\mathbf{a}_N^*\left(\mathbf{R}_{r,N}(z) - \mathbf{T}_N(z)\right)\mathbf{a}_N\right| \le \frac{C}{N^2} \tag{4.108}$$

for all large N. We start as usual with the identity $\mathbf{R}_{r,N}(z) - \mathbf{T}_N(z) = \mathbf{R}_{r,N}(z)\left(\mathbf{T}_N(z)^{-1} - \mathbf{R}_{r,N}(z)^{-1}\right)\mathbf{T}_N(z)$, to get

$$\mathbf{a}_N^*\left(\mathbf{R}_{r,N}(z) - \mathbf{T}_N(z)\right)\mathbf{a}_N = \sigma\frac{\alpha_{r,N}(z) - \delta_N(z)}{\left(1+\sigma\alpha_{r,N}(z)\right)(1+\sigma\delta_N(z))}\mathbf{a}_N^*\mathbf{R}_{r,N}(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)\mathbf{a}_N$$
$$+ z\sigma\left(\tilde{\alpha}_{r,N}(z) - \tilde{\delta}_N(z)\right)\mathbf{a}_N^*\mathbf{R}_{r,N}(z)\mathbf{T}_N(z)\mathbf{a}_N.$$

The expression (4.94) of $\tilde{\alpha}_{r,N}$ implies that $z(\tilde{\alpha}_{r,N}(z) - \tilde{\delta}_N(z)) = z(\alpha_{r,N}(z) - \delta_N(z)) + \mathcal{O}\left(\frac{1}{N^p}\right)$ for each integer p. Thus, to prove (4.86) and (4.87), it is sufficient to check that

$$\sup_{z\in\partial\mathcal{R}_y}\left|\alpha_{r,N}(z) - \delta_N(z)\right| \le \frac{C}{N^2}.$$

We will use the same ideas as in the proof of theorem 2.2.2 and remark that $(\alpha_{r,N}(z) - \delta_N(z), \tilde{\alpha}_{r,N}(z) - \tilde{\delta}_N(z))$ can be interpreted as the solution of a 2×2 linear system whose determinant is a regularized version of (4.59), and appears uniformly bounded away from zero on $\partial\mathcal{R}_y$.

Using again the previous expression of $\mathbf{R}_{r,N}(z) - \mathbf{T}_N(z)$ together with (4.106), (4.107) and repeating the procedure for $\tilde{\mathbf{R}}_{r,N}(z) - \tilde{\mathbf{T}}_N(z)$, we obtain

$$\begin{bmatrix}\alpha_{r,N}(z) - \delta_N(z)\\ \tilde{\alpha}_{r,N}(z) - \tilde{\delta}_N(z)\end{bmatrix} = \begin{bmatrix}u_{r,N}(z) & zv_{r,N}(z)\\ z\tilde{v}_{r,N}(z) & u_{r,N}(z)\end{bmatrix}\begin{bmatrix}\alpha_{r,N}(z) - \delta_N(z)\\ \tilde{\alpha}_{r,N}(z) - \tilde{\delta}_N(z)\end{bmatrix} + \frac{1}{N^2}\begin{bmatrix}\epsilon_{r,N}(z)\\ \tilde{\epsilon}_{r,N}(z)\end{bmatrix}, \tag{4.109}$$

with $u_{r,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\frac{\mathbf{R}_{r,N}(z)\mathbf{B}_N^*\mathbf{B}_N\mathbf{T}_N(z)}{(1+\sigma\alpha_{r,N}(z))(1+\sigma\delta_N(z))}$, $v_{r,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{R}_{r,N}(z)\mathbf{T}_N(z)$ and $\tilde{v}_{r,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\tilde{\mathbf{R}}_{r,N}(z)\tilde{\mathbf{T}}_N(z)$. The quantities $\epsilon_{r,N}(z)$, $\tilde{\epsilon}_{r,N}(z)$ satisfy $\sup_{z\in\partial\mathcal{R}_y}|\epsilon_{r,N}(z)| < C, \sup_{z\in\partial\mathcal{R}_y}|\tilde{\epsilon}_{r,N}(z)| < C$. The determinant of the system is given by

$$\Delta_{r,N}(z) = \left(1 - u_{r,N}(z)\right)^2 - z^2 v_{r,N}(z)\tilde{v}_{r,N}(z).$$

Lemma 4.5.9 implies that for all large N, $u_{r,N}(z)$, $v_{r,N}(z)$ and $\tilde{v}_{r,N}(z)$ are uniformly bounded on $\partial\mathcal{R}_y$. Therefore, to conclude the proof of (4.108), it remains to check that for all large N,

$$\inf_{z\in\partial\mathcal{R}_y}\left|\Delta_{r,N}(z)\right| \ge C > 0.$$

Consider the function $\check{\Delta}_N(z)$ where we have replaced the matrix $\mathbf{R}_{r,N}(z)$ and $\tilde{\mathbf{R}}_{r,N}(z)$ by $\mathbf{T}_N(z)$ and $\tilde{\mathbf{T}}_N(z)$, i.e

$$\check{\Delta}_N(z) = (1 - \check{u}_N(z))^2 - z^2\check{v}_N(z)\check{\tilde{v}}_N(z),$$

with $\check{u}_N(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\frac{\mathbf{T}_N(z)\mathbf{B}_N^*\mathbf{B}_N\mathbf{T}_N(z)}{(1+\sigma\delta_N(z))(1+\sigma\delta_N(z))}$, $\check{v}_N(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{T}_N(z)^2$, and $\check{\tilde{v}}_N(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\tilde{\mathbf{T}}_N(z)^2$. Lemmas 4.5.8, 4.5.9 imply that $\left|\check{u}_N(z) - u_{r,N}(z)\right|$, $\left|v_{r,N}(z) - \check{v}_N(z)\right|$ and $\left|\tilde{v}_{r,N}(z) - \check{\tilde{v}}_N(z)\right|$ converge to 0 uniformly on $\partial\mathcal{R}_y$ which of course implies

$$\sup_{z\in\partial\mathcal{R}_y}|\Delta_N(z) - \check{\Delta}_N(z)| \xrightarrow[N\to\infty]{} 0. \tag{4.110}$$

Using Cauchy-Schwarz inequality, we get

$$\left|\check{\Delta}_N(z)\right| \ge \Delta_{1,N}(z) := \left(1 - u_{1,N}(z)\right)^2 - |z|^2 v_{1,N}(z)\tilde{v}_{1,N}(z),$$

with $u_{1,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\frac{\mathbf{T}_N(z)\mathbf{B}_N\mathbf{B}_N^*\mathbf{T}_N(z)^*}{|1+\sigma\delta_N(z)|^2}$, $v_{1,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\mathbf{T}_N(z)\mathbf{T}_N(z)^*$ and $\tilde{v}_{1,N}(z) = \frac{\sigma^2}{N}\mathrm{Tr}\,\tilde{\mathbf{T}}_N(z)\tilde{\mathbf{T}}_N(z)^*$. From the proof of lemma 2.7.2, we know that

$$\Delta_{1,N}(z) = \frac{\mathrm{Im}(z)}{\mathrm{Im}\left(w_N(z)\right)}.$$

which implies

$$\inf_{z\in\partial\mathcal{R}_y}\Delta_{1,N}(z) \ge C.$$

We deduce from this that $\inf_{z\in\partial\mathcal{R}_y}|\Delta_N(z)| \ge C > 0$ for all large N. Therefore, we can invert the system (4.109) and obtain

$$\sup_{z\in\partial\mathcal{R}_y}\left|\alpha_{r,N}(z) - \delta_N(z)\right| \le \frac{C}{N^2},$$

for all large N. This establishes (4.87) and completes the proof of (4.86).

The proof of (4.88) is similar to the proof of (4.50) (appendix 4.5.5), but as above, $\alpha_N(z), \tilde{\alpha}_N(z), \mathbf{R}_N(z)$ and $\tilde{\mathbf{R}}_N(z)$ have to be replaced by their regularized versions $\alpha_{r,N}(z), \tilde{\alpha}_{r,N}(z), \mathbf{R}_{r,N}(z)$ and $\tilde{\mathbf{R}}_{r,N}N(z)$. The reader can check that the properties of these regularized functions allow to follow the various steps of the proof of (4.50).

Bibliography

[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. Bai and J. Yao. Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, 44(3):447–474, 2008.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] F. Benaych-Georges, A. Guionnet, and Mylène Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *arXiv:1009.0145*, 2010.

[9] F. Benaych-Georges and R.R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 2011.

[10] 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.

[11] M. Capitaine and C. Donati-Martin. Strong asymptotic freeness of Wigner and Wishart matrices. *Indiana Univ. Math. Journal*, 56:295–309, 2007.

[12] L.H.Y. Chen. An inequality for the multivariate normal distribution. *Journal of Multivariate Analysis*, 12(2):306–315, 1982.

[13] 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.

[14] K.R. Davidson and S.J. Szarek. Local operator theory, random matrices and Banach spaces. *Handbook of the geometry of Banach spaces*, 1:317–366, 2001.

[15] 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.

[16] 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.

[17] 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.

[18] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, 36(6):2757–2790, 2008.

[19] K. Fan. Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proceedings of the National Academy of Sciences of the United States of America*, 37(11):760, 1951.

[20] V. L. Girko. *Theory of stochastic canonical equations. Vol. I*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[21] V.L. Girko. *An introduction to statistical analysis of random arrays*. Vsp, 1998.

[22] U. Haagerup and S. Thorbjornsen. A new application of random matrices: $\mathrm{Ext}(C^*_{red}(F_2))$ is not a group. *Annals of Mathematics*, 162(2):711, 2005.

[23] 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.

[24] 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.

[25] W. Hachem, P. Loubaton, X.Mestre, J. Najim, and P. Vallet. A subspace estimator for fixed rank perturbations of large random matrices. *submitted*, 2011. arXiv:1106.1497.

[26] E.J. Hannan. The estimation of frequency. *Journal of Applied probability*, 10(3):510–519, 1973.

[27] A.T. James. Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics*, 35(2):475–501, 1964.

[28] I.M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

[29] M. Krein and A. Nudelman. The markov moment problem and extremal problems. translations of mathematical monographs 50. *American Mathematical Society. Providence*, 1977.

[30] 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.

[31] V.A. Marcenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457, 1967.

[32] 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.

[33] R.J. Muirhead. *Aspects of multivariate statistical theory*. Wiley Online Library, 1982.

[34] 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.

[35] 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.

[36] EA Novikov. Functionals and the random-force method in turbulence theory(euler velocity field described by random forces method, using lagrangian representation of turbulence). *Soviet Physics-JETP*, 20:1290–1294, 1965.

[37] L. Pastur. A simple approach to the global regime of Gaussian ensembles of random matrices. *Ukrainian Math. J.*, 57(6):936–966, 2005.

[38] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] E. Telatar. Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6):585–595, 1999.

[44] H.G. Tillmann. Randverteilungen analytischer funktionen und distributionen. *Mathematische Zeitschrift*, 59(1):61–83, 1953.

[45] 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.

[46] DW Tufts, AC Kot, and RJ Vaccaro. The threshold effect in signal processing algorithms which use an estimated subspace, 1991.

[47] S. Verdú and S. Shamai. Spectral efficiency of cdma with random spreading. *IEEE Transactions on Information Theory*, 45(2):622–640, 1999.

[48] E.P. Wigner. On the distribution of the roots of certain symmetric matrices. *The Annals of Mathematics*, 67(2):325–327, 1958.

[49] 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.

[50] 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.

[51] X.Mestre and M.A.Lagunas. Modified subspace algorithms for DoA estimation with large arrays. *IEEE Transactions on Signal Processing*, 56(2):598, 2008.