

# Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case.

Pascal Vallet <sup>1</sup>   Philippe Loubaton <sup>2</sup>   Xavier Mestre <sup>3</sup>

## Abstract

We consider the problem of subspace estimation in situations where the number of available snapshots and the observation dimension are comparable in magnitude. In this context, traditional subspace methods tend to fail because the eigenvectors of the sample correlation matrix are heavily biased with respect to the true ones. It has recently been suggested that this situation (where the sample size is small compared to the observation dimension) can be very accurately modeled by considering the asymptotic regime where the observation dimension  $M$  and the number of snapshots  $N$  converge to  $+\infty$  at the same rate. Using large random matrix theory results, it can be shown that traditional subspace estimates are not consistent in this asymptotic regime. Furthermore, new consistent subspace estimate can be proposed, which outperform the standard subspace methods for realistic values of  $M$  and  $N$ . The work carried out so far in this area has always been based on the assumption that the observations are random, independent and identically distributed in the time domain. The goal of this paper is to propose new consistent subspace estimators for the case where the source signals are modelled as unknown deterministic signals. In practice, this allows to use the proposed approach regardless of the statistical properties of the source signals. In order to construct the proposed estimators, new technical results concerning the almost sure location of the eigenvalues of sample covariance matrices of Information plus Noise complex Gaussian models are established. These results are believed to be of independent interest.

<sup>1</sup> Université de Paris-Est IGM LabInfo, UMR-CNRS 8049, 5 Boulevard Descartes, 77454 Marne la Vallée Cedex 2, France, E-mail: pascal.vallet@univ-mlv.fr

<sup>2</sup> Université de Paris-Est IGM LabInfo, UMR-CNRS 8049, 5 Boulevard Descartes, 77454 Marne la Vallée Cedex 2, France, E-mail: philippe.loubaton@univ-mlv.fr

<sup>3</sup> Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Av. Carl Friedrich Gauss 7, 08860 Castelldefels (Barcelona), Spain, E-mail: xavier.mestre@cttc.cat

## Index Terms

Subspace-based estimation, random matrix theory, information plus noise model, limit eigenvalue distribution.

**Notation:** Matrix (resp. vectors) quantities are denoted by boldfaced capital (resp. lower case) letters. The  $N \times N$  identity matrix is denoted as  $\mathbf{I}_N$ . Trace and spectral norm will be denoted  $\text{Tr}[\cdot]$  and  $\|\cdot\|$  respectively, and  $[\cdot]^T$  and  $[\cdot]^H$  represent the transpose and the conjugate transpose. For a set  $\mathcal{U}$ , we denote by  $\text{Int}(\mathcal{U})$  and  $\partial\mathcal{U}$  its interior and boundary respectively. Given a complex number  $z$ ,  $\text{Re}(z)$  and  $\text{Im}(z)$  denote its real and imaginary parts respectively,  $(\cdot)^*$  stands for complex conjugation and  $i$  denotes the imaginary unit. The upper complex half plane is denoted by  $\mathbb{C}_+$ , i.e.  $\mathbb{C}_+ = \{z \in \mathbb{C} : \text{Im}(z) > 0\}$ , and equivalently  $\mathbb{C}_-$  will denote the lower complex half plane. Similarly,  $\mathbb{R}_+$  and  $\mathbb{R}_-$  represent the set of all positive real numbers and the set of all negative real numbers respectively. We will also write  $\mathbb{R}^* \equiv \mathbb{R} \setminus \{0\}$  and  $\mathbb{C}^* \equiv \mathbb{C} \setminus \{0\}$ . For a given contour  $\mathcal{C}$  on the complex plane,  $\text{Ind}_{\mathcal{C}}(\xi)$  will denote the index of the contour with respect to a point  $\xi \in \mathbb{C}$ . The support of a particular function  $\phi$  will be denoted as  $\text{supp}(\phi)$ , and  $\mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$  will represent the set of compactly supported real-valued smooth functions defined on  $\mathbb{R}$ .

## I. INTRODUCTION

Subspace estimation methods have been widely proposed in multiple applications of communications and signal processing, such as direction of arrival (DoA) estimation [1], beamforming [2], channel identification [3], waveform estimation [4], and many other general parametric estimation problems based on multivariate observations [5]. In general terms, these algorithms are applicable to the situation where a number of parameters needs to be extracted from a set of multivariate observations, which are composed of a noise part, with full-rank empirical correlation matrix, plus a signal contribution that has low-rank empirical correlation matrix. By exploiting the inherent orthogonality between the signal subspace (i.e. the subspace spanned by the columns of the signal empirical correlation matrix) and the noise subspace, one can try to extract the original parameters from the set of noisy observations. In general terms, the resulting estimators are computationally much more affordable and hence are generally preferred over other estimators such as those based in the Maximum Likelihood (ML) principle, which generally perform better but unfortunately involve an exhaustive search in a multi-dimensional parametric space.

In order to formulate a generic subspace estimator, one must first infer the eigenvectors of the correlation matrix of the observation. This is generally difficult, because the correlation matrix of the multivariate

observation is generally unknown. In 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 observations (denoted by  $N$ ) tends to infinity while the observation dimension (denoted by  $M$ ) remains constant. Indeed, under certain ergodicity assumptions, when  $N \rightarrow \infty$  for a fixed  $M$ , the sample correlation matrix of the observation converges almost surely to the true one, and consequently when  $N \gg M$  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 what is usually referred to as the breakdown effect of subspace-based techniques [6].

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  $N \gg M$ . However, it has been recently suggested [7] 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.  $M, N \rightarrow +\infty$ , whereas  $c_N \equiv \frac{M}{N}$  converges towards a strictly positive constant. Using Large Random Matrix Theory (LRMT) results, it was shown in [7] 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, LRMT can be very useful (1) to characterize how the sample eigenvectors differ from the true ones in a scenario where  $M$  and  $N$  are comparable in magnitude and (2) to derive alternative estimators of the eigenvectors that converge, not only when  $N \rightarrow +\infty$  for a fixed  $M$ , but also when  $M, N \rightarrow +\infty$  at the same rate. This was more extensively demonstrated in [8] and [9], which respectively considered the characterization of the sample eigenvectors when  $M, N \rightarrow +\infty$  at the same rate, and proposed alternative consistent estimators for these quantities in the new asymptotic regime.

Unfortunately, the work in [8] and [9] 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 observation does not present this structure, and is better modelled as a deterministic component (corresponding to the signal part) plus some additive noise, that is generally Gaussian distributed. This model is usually referred to as the “information plus noise model” in the LRMT literature [10], as opposed

to the more classical “sample covariance matrix model” [11], which was the one used in [7], [8], [9]. The main objective of this paper is to propose improved subspace estimators for the information plus noise model, which will represent the case where the source signals are as non-observable deterministic sequences. In order to obtain these estimators, new mathematical results related to the almost sure location of the eigenvalues of the empirical covariance matrix of a Gaussian information plus noise model are derived. These results are believed to be of independent interest.

The rest of the paper is organized as follows. Section II introduces the information plus noise model associated with the specific application addressed here: the determination of multiple directions of arrival (DoA) using an array of antennas. The main objectives of the paper in mathematical terms are also formulated. Section III provides some general facts related to the convergence of the eigenvalues of the empirical correlation matrix for the information plus noise model. It is further explained in Section IV that the eigenvalues of the sample correlation matrix tend to concentrate around some clusters when both  $M, N \rightarrow +\infty$  at the same rate. A very simple description of the position of these asymptotic eigenvalue clusters is also provided. It is in particular shown that each cluster is associated with a set of consecutive eigenvalues of true covariance matrix of the observation. Section V presents an intermediate result that has its own interest. In brief, it is shown that, for sufficiently large  $M, N$ , with probability one no eigenvalues of the sample correlation matrix will be located outside the asymptotic eigenvalue clusters. Furthermore, the number of sample eigenvalues that are located in each of these clusters is directly related to the dimensionality of the corresponding eigenspace of the true covariance matrix. In order to focus on the applicative context of the paper, this claim is proved for the cluster associated with the noise subspace, but it can be extended easily to the other clusters. This fact generalizes the results derived in [12] and [13] in the context of source signals independent identically distributed in the time domain. In contrast with [12] and [13], the results presented in this paper, inspired by the approach developed in [14], are only valid in the complex Gaussian case. The above mentioned results are then used in Section VI in order to derive an estimator of the localization function of the subspace estimate that is consistent not only when  $N \rightarrow +\infty$  for fixed  $M$ , but also when  $M, N \rightarrow +\infty$  at the same rate. Section VII provides some numerical examples that illustrate the effectiveness of the proposed estimators. Finally Section VIII concludes the paper. Most of the technical derivations have been relegated to the appendices.

The results of this paper have been partly presented in the short conference paper [15].

## II. PROBLEM STATEMENT

In order to motivate and illustrate the signal model that is used in this paper, we consider the following DoA estimation problem. Assume that  $K$  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 mathematically described as

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

where  $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$  is an  $M \times K$  matrix that contains the steering vectors of the  $K$  sources,  $\mathbf{s}_n$  is a  $K \times 1$  column vector containing the transmitted signals from the  $K$  sources at time instant  $n$ , and where  $\mathbf{v}_n$  is an additive Gaussian white noise component with zero mean and covariance matrix  $\mathbb{E}[\mathbf{v}_n \mathbf{v}_n^H] = \sigma^2 \mathbf{I}_M$ . We assume that  $\mathbf{y}_n$  is available from  $n = 1$  to  $n = N$ , and that  $M < N$ , or equivalently that  $c_N = \frac{M}{N}$  is strictly less than 1. It is possible to generalize our results to the situation where  $c_N > 1$ , although the presentation of the corresponding results would however complicate the developments of the present paper.

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

$$\mathbf{Y}_N = \mathbf{A}\mathbf{S}_N + \mathbf{V}_N \quad (1)$$

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, we can define the empirical spatial correlation matrix of the observation as  $\hat{\mathbf{R}}_N \equiv \frac{1}{N} \mathbf{Y}_N \mathbf{Y}_N^H$ , whereas the empirical spatial correlation matrix associated with the noiseless observation will take the form  $\frac{1}{N} \mathbf{A}\mathbf{S}_N \mathbf{S}_N^H \mathbf{A}^H$ . It is worth pointing out here that, since the number of signals is assumed to be lower than the number of antennas ( $K < M$ ), the steering matrix  $\mathbf{A}$  will always be a tall matrix and therefore the empirical spatial correlation matrix of the noiseless observation will never be full rank. In other words, the minimum eigenvalue of the matrix  $\frac{1}{N} \mathbf{A}\mathbf{S}_N \mathbf{S}_N^H \mathbf{A}^H$  will always be zero and will have multiplicity equal to  $M - K$ .

In order to simplify the notation in the subsequent exposition, we define the matrices  $\mathbf{\Sigma}_N$ ,  $\mathbf{B}_N$ ,  $\mathbf{W}_N$  as

$$\mathbf{\Sigma}_N = \frac{\mathbf{Y}_N}{\sqrt{N}}, \quad \mathbf{B}_N = \frac{\mathbf{A}\mathbf{S}_N}{\sqrt{N}}, \quad \mathbf{W}_N = \frac{\mathbf{V}_N}{\sqrt{N}} \quad (2)$$

so that (1) can be equivalently formulated as

$$\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{W}_N \quad (3)$$

where  $\mathbf{\Sigma}_N$  is the (normalized) matrix of observations,  $\mathbf{B}_N$  is a deterministic matrix containing the signals contribution and  $\mathbf{W}_N$  is a complex Gaussian white noise matrix with i.i.d. entries that have zero mean and

variance  $\sigma^2/N$ . We denote by  $\mathbf{\Pi}_N$  the orthogonal projection matrix on the “noise subspace”, which in our context is defined as the orthogonal complement of the column space of matrix  $\mathbf{A}$ . In the following, we assume that the empirical correlation matrix of  $\mathbf{S}_N$  defined by  $\frac{1}{N}\mathbf{S}_N\mathbf{S}_N^H$  is full rank. Therefore, the noise subspace coincides with the kernel of the empirical correlation matrix of the noiseless signal, namely  $\mathbf{B}_N\mathbf{B}_N^H$ .

Let  $\{\gamma_k^{(N)}\}_{k=1,\dots,M}$  denote the eigenvalues of the empirical correlation matrix of the signal component, namely  $\mathbf{B}_N\mathbf{B}_N^H$ , arranged in increasing order and let  $\{\mathbf{e}_k^{(N)}\}_{k=1,\dots,M}$  denote the corresponding unit norm eigenvectors. We note in particular that  $\gamma_1^{(N)} = \dots = \gamma_{M-K}^{(N)} = 0$  while the remaining eigenvalues are strictly positive and that  $\mathbf{\Pi}_N = \sum_{k=1}^{M-K} \mathbf{e}_k^{(N)} \left(\mathbf{e}_k^{(N)}\right)^H$ . The subspace method for the determination of the  $K$  directions of arrival (commonly referred to as MUSIC algorithm) is based on the observation that the angles  $\{\theta_k\}_{k=1,\dots,K}$  coincide with the  $K$  solutions of the equation  $\mathbf{a}(\theta)^H \mathbf{\Pi}_N \mathbf{a}(\theta) = 0$ . In order to be able to use this last observation, it is in practice necessary to estimate the function  $\mathbf{a}(\theta)^H \mathbf{\Pi}_N \mathbf{a}(\theta)$  (usually referred to as the “localization function”) for each  $\theta \in [-\pi, \pi]$ , or more generically to estimate the quantity

$$\eta_N(\mathbf{b}) = \mathbf{b}^H \mathbf{\Pi}_N \mathbf{b}$$

for each deterministic  $M$ -dimensional vector  $\mathbf{b}$ .

If  $N \rightarrow +\infty$  while  $M$  is fixed, the empirical correlation matrix of the observations  $\hat{\mathbf{R}}_N = \Sigma_N \Sigma_N^H$  of  $\mathbf{Y}_N$  converges towards the matrix  $\mathbf{R}_N = \mathbf{B}_N\mathbf{B}_N^H + \sigma^2\mathbf{I}_M$  in the sense that

$$\|\hat{\mathbf{R}}_N - (\mathbf{B}_N\mathbf{B}_N^H + \sigma^2\mathbf{I}_M)\| \rightarrow 0 \quad \text{a.s.} \quad (4)$$

where a.s. represents the almost sure convergence. We will denote by  $\{\hat{\lambda}_k^{(N)}\}_{k=1,\dots,M}$  the eigenvalues of  $\hat{\mathbf{R}}_N$  arranged in increasing order and by  $\{\hat{\mathbf{e}}_k^{(N)}\}_{k=1,\dots,M}$  the corresponding eigenvectors. The convergence result in (4) implies that for each  $\theta$ ,  $\hat{\eta}_N^{trad}(\mathbf{a}(\theta)) - \eta_N(\mathbf{a}(\theta)) \rightarrow 0$  a.s. where  $\hat{\eta}_N^{trad}(\mathbf{a}(\theta))$  is the traditional estimator of the localization function defined as

$$\hat{\eta}_N^{trad}(\mathbf{a}(\theta)) = \sum_{k=1}^{M-K} \mathbf{a}^H(\theta) \hat{\mathbf{e}}_k^{(N)} \left(\hat{\mathbf{e}}_k^{(N)}\right)^H \mathbf{a}(\theta). \quad (5)$$

In practice, predictions provided by the asymptotic regime corresponding to letting  $N \rightarrow +\infty$  for fixed  $M$  are reliable only if  $N$  is much larger than  $M$ . However, this assumption may be quite restrictive in a number of important application contexts. If  $M$  and  $N$  are comparable in magnitude, then the asymptotic regime described by letting  $M, N \rightarrow +\infty$  in such a way that  $c_N = \frac{M}{N}$  converges towards a non zero constant appears to be more relevant. In this regime, the behavior of various classical estimates are more complicated, and have to be studied carefully. In particular, it can be shown that  $\hat{\eta}_N^{trad}(\mathbf{b}) - \eta_N(\mathbf{b})$

does not converge to 0 when  $M, N \rightarrow +\infty$ , which implies that the standard MUSIC estimates are not consistent under this new asymptotic regime. The purpose of this paper is to introduce an improved subspace estimate  $\hat{\eta}_N^{new}(\mathbf{b})$  of  $\eta_N(\mathbf{b})$  for each deterministic vector  $\mathbf{b}$ . The main feature of  $\hat{\eta}_N^{new}(\mathbf{b})$  is to be consistent if  $M, N \rightarrow +\infty$  in such a way that  $c_N = \frac{M}{N}$  converges towards a non zero constant value. In order to achieve this, we will heavily rely on results related to the asymptotic behavior of the eigenvalue distribution of the empirical correlation matrix  $\hat{\mathbf{R}}_N$ . It is however useful to mention that it is not established that

$$\sup_{\theta \in [-\pi, \pi]} |\hat{\eta}_N^{new}(\mathbf{a}(\theta)) - \eta_N(\mathbf{a}(\theta))| \rightarrow 0 \quad (6)$$

almost surely, a useful, but stronger property. We feel that the proof of (6) would need mathematical technics different from those which are used in the present paper.

### III. PROPERTIES OF THE ASYMPTOTIC EIGENVALUE DISTRIBUTION OF MATRIX $\hat{\mathbf{R}}_N$

In this section, we will review some of the important properties related to the asymptotic behavior of the eigenvalue distribution of the empirical correlation matrix  $\hat{\mathbf{R}}_N$  when  $M, N \rightarrow +\infty$  in such a way that  $c_N = \frac{M}{N}$  converges towards a non zero constant, which will be denoted as  $c_*$ . This implies that the observation dimension  $M$  in principle depends on  $N$ , and should be denoted  $M(N)$ . We will however drop this dependence on  $N$  in order to simplify the exposition. Whenever it is clear from the context, we will also drop the dependence on the number of snapshots  $N$  in matrices  $\Sigma_N$ ,  $\mathbf{B}_N$ ,  $\hat{\mathbf{R}}_N$ , eigenvalues  $\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_M^{(N)}$  and  $\gamma_1^{(N)}, \dots, \gamma_M^{(N)}$ , as well as eigenvectors.

**Remark 1.** *From now on,  $N \rightarrow \infty$  will implicitly denote the limit as both  $M, N \rightarrow +\infty$  such that  $\frac{M}{N}$  converges towards a non zero constant  $c_*$ , where it is assumed that  $0 < c_* < 1$ .*

**Remark 2.** *All results that are presented in this paper are equally valid regardless of the behavior of the number of sources  $K$  when  $N$  increases. In other words,  $K$  may scale up with  $N$ , or it may stay constant regardless of  $N$ .*

From now on, we assume that the spectral norms of matrices  $(\mathbf{B}_N)_{N \geq 1}$  remain bounded when  $N \rightarrow \infty$ , i.e. it exists  $b_{max} > 0$  such that

$$\sup_{N \geq 1} \|\mathbf{B}_N\| < b_{max} < \infty \quad (7)$$

The eigenvalue distribution of  $\hat{\mathbf{R}}_N$  is characterized by the empirical distribution function of its eigenvalues, namely

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

where  $\text{card}$  denotes the cardinality of a set. For each  $\lambda \in \mathbb{R}$ , the function  $\hat{F}_N(\lambda)$  gives the proportion of the eigenvalues of  $\hat{\mathbf{R}}_N$  which are lower than or equal to  $\lambda$ . Its associated probability measure, denoted  $\hat{\mu}_N$ , is given by  $d\hat{\mu}_N(\lambda) = \frac{1}{M} \sum_{k=1}^M \delta(\lambda - \hat{\lambda}_k^{(N)})$  and is carried by  $\mathbb{R}_+$ . In order to characterize the asymptotic behavior of  $\hat{\mu}_N$ , it is in practice quite common to characterize the asymptotic behavior of its Stieltjès transform. If  $\mu$  is a positive finite measure (i.e.  $\mu(\mathbb{R}) < \infty$ ), the Stieltjès transform of  $\mu$  is the function  $\Psi_\mu$  of complex variable defined as

$$\Psi_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(\lambda)}{\lambda - z} \quad (8)$$

We recall the following well-known properties of the Stieltjès transform, which will be useful in the mathematical developments throughout the paper.

**Lemma 1.** *Let  $\Psi_\mu$  be the Stieltjès transform of some positive finite measure  $\mu$  (i.e.  $\mu(\mathbb{R}) < \infty$ ), and let us denote as  $\mathcal{S}_\mu$  its support. Then,*

- 1)  $\Psi_\mu$  is holomorphic on  $\mathbb{C} \setminus \mathcal{S}_\mu$ .
- 2)  $\lim_{y \rightarrow +\infty} -iy\Psi_\mu(iy) = \mu(\mathbb{R})$
- 3)  $\forall z \in \mathbb{C} \setminus \mathbb{R}$ ,

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

where  $\text{Im}(z)$  denotes the imaginary part of  $z$ . Moreover,  $\forall z \in \mathbb{C} \setminus \mathcal{S}_\mu$  it holds that

$$|\Psi_\mu(z)| \leq \frac{\mu(\mathbb{R})}{\text{dist}(z, \mathcal{S}_\mu)} \quad (9)$$

- 4)  $\Psi_\mu \in \mathbb{C}_+$  if  $z \in \mathbb{C}_+$ , where  $\mathbb{C}_+$  is the upper complex half plane.
- 5) If  $\mu$  is carried by  $\mathbb{R}_+$ , then  $z\Psi_\mu(z) \in \mathbb{C}_+$  if  $z \in \mathbb{C}_+$ .
- 6) Conversely, if  $\Psi$  is a function analytic in  $\mathbb{C}_+$  satisfying
  - $\Psi(z)$  and  $z\Psi(z)$  belong to  $\mathbb{C}_+$  if  $z \in \mathbb{C}_+$
  - $\sup_{y>1} |iy\Psi(iy)| < +\infty$

then,  $\Psi$  is the Stieltjès transform of a positive finite measure carried by  $\mathbb{R}_+$ .

- 7)  $\forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$ , (the set of compactly supported real-valued smooth functions defined on  $\mathbb{R}$ ), we have

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

Having recalled these basic properties of the Stieltjès transform of a positive finite measure, let us now go back to the asymptotic characterization of the empirical measure  $\hat{\mu}_N$  or, quite equivalently, its Stieltjès transform, which is defined for  $z \in \mathbb{C} - \mathbb{R}_+$  as

$$\hat{m}_N(z) = \int_{\mathbb{R}_+} \frac{d\hat{\mu}_N(\lambda)}{\lambda - z} = \frac{1}{M} \sum_{m=1}^M \frac{1}{\hat{\lambda}_m - z}. \quad (10)$$

It is worth pointing out that  $\hat{m}_N(z)$  can be expressed as the normalized trace of the resolvent matrix, which is a matrix-valued function defined as

$$\mathbf{Q}_N(z) = \left( \hat{\mathbf{R}}_N - z\mathbf{I}_M \right)^{-1} = \left( \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H - z\mathbf{I}_M \right)^{-1} \quad (11)$$

namely  $\hat{m}_N(z) = \frac{1}{M} \text{Tr} [\mathbf{Q}_N(z)]$ . Except (16), the following results can be more or less immediately derived from [10] (see also [16])

**Theorem 1.** *There exists a deterministic probability distribution  $\mu_N$  carried by  $\mathbb{R}_+$  such that  $\hat{\mu}_N - \mu_N$  converges in distribution almost surely towards 0 when  $N \rightarrow \infty$ . The measure  $\mu_N$ , referred to in what follows as the asymptotic eigenvalue distribution of matrix  $\hat{\mathbf{R}}_N$ , is characterized by its Stieltjès transform  $m_N(z)$  as*

$$m_N(z) = \int_{\mathbb{R}_+} \frac{d\mu_N(\lambda)}{\lambda - z} \quad (12)$$

which is a solution of the equation

$$m_N(z) = \frac{1}{M} \text{Tr} \left[ -z(1 + \sigma^2 c_N m_N(z)) \mathbf{I}_M + \sigma^2 (1 - c_N) \mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^H}{1 + \sigma^2 c_N m_N(z)} \right]^{-1} \quad (13)$$

for each  $z \in \mathbb{C} - \mathbb{R}_+$ . Let  $\mathbf{T}_N(z)$  be the  $M \times M$  matrix valued function defined on  $\mathbb{C} - \mathbb{R}_+$  by

$$\mathbf{T}_N(z) = \left[ -z(1 + \sigma^2 c_N m_N(z)) \mathbf{I}_M + \sigma^2 (1 - c_N) \mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^H}{1 + \sigma^2 c_N m_N(z)} \right]^{-1}. \quad (14)$$

Then,  $\mathbf{T}_N(z)$  is holomorphic on  $\mathbb{C} - \mathbb{R}_+$ . Moreover, almost surely,

$$\lim_{N \rightarrow \infty} (\hat{m}_N(z) - m_N(z)) = 0 \quad (15)$$

for each  $z \in \mathbb{C} - \mathbb{R}_+$ . Finally, for each  $M$ -dimensional deterministic vectors  $\mathbf{u}_N, \mathbf{v}_N$  such that  $\sup_N \|\mathbf{u}_N\| < \infty$  and  $\sup_N \|\mathbf{v}_N\| < \infty$ , it holds that almost surely

$$\lim_{N \rightarrow \infty} \mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N = 0 \quad (16)$$

for each  $z \in \mathbb{C} - \mathbb{R}_+$ .

*Proof:* Convergence of  $\hat{\mu}_N - \mu_N$  towards 0 as well as the fact that  $m_N(z)$  is a solution to (13) is due to [10]. As for the result in (15), it is a well known consequence of the convergence of  $\hat{\mu}_N - \mu_N$  towards 0. (16) is proved in the Appendix F. ■

Theorem 1 is pointing out that the entries of the resolvent  $\mathbf{Q}_N(z)$  are almost surely asymptotically close to the entries of the deterministic matrix function  $\mathbf{T}_N(z)$  (this statement follows from (16) by selecting  $\mathbf{u}_N$  and  $\mathbf{v}_N$  as two columns of  $\mathbf{I}_M$ ); and that its normalized trace,  $\hat{m}_N(z)$  as defined in (10), is almost surely asymptotically close to  $m_N(z)$ , one of the solutions to the polynomial equation in (13). Furthermore, the random measure  $\hat{\mu}_N$  is also almost surely equivalent (in distribution) to the deterministic measure  $\mu_N$  in this asymptotic regime.

We denote by  $\mathcal{S}_N$  the support of this measure  $\mu_N$ , which will play a very important role in the following. The characterization of  $\mathcal{S}_N$  has been first presented in [17], and is based on the study of the properties of function  $m_N(z)$  which, since it is a Stieltjès transform, is holomorphic on  $\mathbb{C} \setminus \mathcal{S}_N$  and real-valued on  $\mathbb{R} \setminus \mathcal{S}_N$ . In order to characterize  $\mathcal{S}_N$ , we will also consider the function  $w_N(z)$ , introduced in [17], defined from  $m_N(z)$  as follows

$$w_N(z) = z(1 + \sigma^2 c_N m_N(z))^2 - \sigma^2(1 - c_N)(1 + \sigma^2 c_N m_N(z)). \quad (17)$$

It will be seen later on that the function  $w_N(z)$  has very interesting properties that will be crucial for the derivations in this paper. In particular, we will show in the following that the support of  $\mu_N$ , namely  $\mathcal{S}_N$ , is in fact equal to the support of the imaginary part of  $w_N(z)$  when  $z$  approaches the real axis. Thanks to this fact, we will be able to characterize the support  $\mathcal{S}_N$  by studying the properties of  $w_N(z)$  for  $z$  on the real axis.

The next proposition provides some preliminary properties of  $m_N(z)$  and  $w_N(z)$  that will become useful in the following sections. Most of these properties are established in [17]. We will denote by  $f_N(w)$  the function on  $\mathbb{C} - \{\gamma_1, \dots, \gamma_M\}$  defined by

$$f_N(w) = \frac{1}{M} \text{Tr} \left[ (\mathbf{B}_N \mathbf{B}_N^H - w \mathbf{I}_M)^{-1} \right]$$

which coincides with the Stieltjès transform of the eigenvalue distribution  $\nu_N(d\lambda) = \frac{1}{M} \sum_{k=1}^M \delta(\lambda - \gamma_k)$  associated with the signal matrix  $\mathbf{B}_N \mathbf{B}_N^H$ .

**Proposition 1.** *The following properties hold:*

- 1) *The condition  $c_N < 1$  implies that 0 does not belong to  $\mathcal{S}_N$ .*
- 2) *For each  $x \in \mathbb{R}$ ,  $\lim_{z \in \mathbb{C}_+, z \rightarrow x} m_N(z)$  exists, and will be denoted  $m_N(x)$ . The function  $m_N(z)$  thus defined is continuous on  $\mathbb{C}_+ \cup \mathbb{R}$ , and continuously differentiable on  $\mathbb{C}_+ \cup \mathbb{R} - \partial \mathcal{S}_N$ . Moreover, for each  $x \in \mathbb{R}$ ,  $\lim_{z \in \mathbb{C}_-, z \rightarrow x} m_N(z)$  exists, and is equal to  $(m_N(x))^*$ . The measure  $\mu_N$  is absolutely continuous, its density is  $\frac{1}{\pi} \text{Im}(m_N(x))$ , and the interior  $\text{Int}(\mathcal{S}_N)$  of  $\mathcal{S}_N$  is given by*

$$\text{Int}(\mathcal{S}_N) = \{x > 0 : \text{Im}(m_N(x)) > 0\} \quad (18)$$

- 3) For each  $x \in \mathbb{R}$ ,  $\lim_{z \in \mathbb{C}_+, z \rightarrow x} w_N(z)$  exists, and is still denoted by  $w_N(x)$ . The function  $z \rightarrow w_N(z)$  is continuous on  $\mathbb{C}_+ \cup \mathbb{R}$ , and is continuously differentiable on  $\mathbb{C}_+ \cup \mathbb{R} - \partial\mathcal{S}_N$ . Moreover,  $w_N(x) = x(1 + \sigma^2 c_N m_N(x))^2 - \sigma^2(1 - c_N)(1 + \sigma^2 c_N m_N(x))$ . Finally,  $\lim_{z \in \mathbb{C}_-, z \rightarrow x} w_N(z) = w_N(x)^*$ .
- 4)  $w_N(x)$  does not belong to the set  $\{\gamma_1^{(N)}, \dots, \gamma_M^{(N)}\}$  if  $x \in \mathbb{R} - \mathcal{S}_N$ .
- 5)  $\text{Im}[w_N(z)] > 0$  if  $\text{Im}z > 0$ .
- 6)  $\text{Re}[1 + c_N \sigma^2 m_N(z)] > 0$  for each  $z \in \mathbb{C}$ .
- 7) For any  $x \in \mathbb{R} - \partial\mathcal{S}_N$ , the function  $m_N(x)$  is solution of the equation in (13)
- 8) For any  $x \in \mathbb{R} - \partial\mathcal{S}_N$ , the function  $w_N(x)$  is a solution of the equation

$$\phi_N(w_N(x)) = x \quad (19)$$

where  $\phi_N(w)$  is defined by

$$\phi_N(w) = w(1 - c_N \sigma^2 f_N(w))^2 + (1 - c_N) \sigma^2 (1 - c_N \sigma^2 f_N(w)) \quad (20)$$

*Proof:* Property 1 is not established in [17], and is proved in Appendix A. As for Property 2, the existence of the limit of  $m_N(x + iy)$  is proved in [17] for  $x \neq 0$  because [17] did not assume that  $c_N < 1$ . However, Property 1 implies immediately that the limit exists if  $x = 0$  because  $m_N(z)$  is holomorphic in a neighborhood of the origin. The continuity and the differentiability of  $x \rightarrow m_N(x)$  is established in [17] on  $\mathbb{R}^*$  and  $\mathbb{R}^* \setminus \partial\mathcal{S}_N$  respectively, but it also holds on  $\mathbb{R}$  and  $\mathbb{R} \setminus \partial\mathcal{S}_N$  by Property 1 and the fact that  $m_N(z)$  is holomorphic  $\mathbb{C} \setminus \mathcal{S}_N$ . Since  $m_N(z)$  is the Stieltjès transform of a positive measure, it is clear that  $m_N(z^*)$  coincides with  $m_N^*(z)$ . This implies immediately that  $\lim_{y < 0, y \rightarrow 0} m_N(x + iy) = m_N^*(x)$ . Finally, (18) is a direct consequence of the continuity of  $x \rightarrow m_N(x)$ . Property 3 follows directly from Property 2. Properties 4 and 5 are established in [17]. As for Property 6, it was initially proven in [17] for  $z \in \mathbb{C}^*$ , but it can be shown easily that it holds for  $z = 0$  using Property 1 as well as the proof of Lemma 2-1 of [17]. Finally, [17] established that  $m_N(x)$  is solution of (13) if  $x \in \text{int}(\mathcal{S}_N)$ . This also holds if  $x \in \mathbb{C} \setminus \mathcal{S}_N$  because by Properties 4 and 6, the right hand side of (13) is holomorphic on  $\mathbb{C} \setminus \mathcal{S}_N$ . Since  $m_N(z)$  is itself holomorphic on  $\mathbb{C} \setminus \mathcal{S}_N$ , the equality in (13) must hold not only on  $\mathbb{C} \setminus \mathbb{R}_+$  but also on  $\mathbb{C} \setminus \mathcal{S}_N$ . Recalling that  $\mathcal{S}_N$  is a closed set, all this implies that  $m_N(x)$  is solution of equation (13) for  $x \in \mathbb{R} \setminus \partial\mathcal{S}_N$ .

Let us finally establish Property 8. Thanks to Properties 6 and 7 and to (13), we can write

$$\frac{m_N(x)}{1 + \sigma^2 c_N m_N(x)} = f_N(w_N(x)) \quad (21)$$

for each  $x \in \mathbb{R} \setminus \partial \mathcal{S}_N$ . This last equality can be rewritten as

$$1 - \sigma^2 c_N f_N(w_N(x)) = \frac{1}{1 + \sigma^2 c_N m_N(x)} \quad (22)$$

where the right hand side is well defined thanks to Property 6. Now, plugging (22) into (17), we obtain that, for  $x \in \mathbb{R} \setminus \partial \mathcal{S}_N$ ,  $w_N(x)$  is a solution of the equation

$$\phi_N(w) = x \quad (23)$$

where function  $\phi_N(w)$  is defined in (20). In other words, the function  $w_N(x)$  satisfies (19) for each  $x \in \mathbb{R} \setminus \partial \mathcal{S}_N$ . ■

Proposition 1 is establishing the fact that both  $m_N(z)$  and  $w_N(z)$  are well defined when  $z$  approaches the real axis, and that  $m_N(x)$  and  $w_N(x)$  can be determined as one of the solutions to (13) and (19) respectively for any  $x \in \mathbb{R} \setminus \partial \mathcal{S}_N$ . In the next section we will establish some properties that characterize  $w_N(x)$  out of the set of all the solutions of (19), and this will in turn help us in the characterization of the support  $\mathcal{S}_N$ .

#### IV. AN ALTERNATIVE CHARACTERIZATION OF $\mathcal{S}_N$

In this section we will provide a characterization of the support  $\mathcal{S}_N$  as a simpler alternative to the study provided in [17]. It must be pointed out that [17] assumed that the eigenvalue distribution of matrix  $\mathbf{B}_N \mathbf{B}_N^H$  converges to a limit distribution  $\nu_\infty(d\lambda)$ , and showed that  $\mu_N$  converges towards a probability distribution  $\mu_\infty$ . Its Stieltjès transform  $m_\infty$  is solution of (13), but in which the discrete measure  $\nu_N(d\lambda) = \frac{1}{M} \sum_{j=1}^M \delta(\lambda - \gamma_k^{(N)})$  is replaced by measure  $\nu_\infty(d\lambda)$ , i.e.

$$m_\infty(z) = \int \left[ -z(1 + \sigma^2 c_N m_N(z)) + \sigma^2(1 - c_N) + \frac{\lambda}{1 + \sigma^2 c_N m_N(z)} \right]^{-1} \nu_\infty(d\lambda).$$

In [17], a detailed analysis of the support  $\mathcal{S}_\infty$  of  $\mu_\infty$  was presented. The corresponding results provide of course a characterization of  $\mathcal{S}_N$  by replacing the general probability distribution  $\nu_\infty(d\lambda)$  by the discrete measure  $\nu_N(d\lambda) = \frac{1}{M} \sum_{j=1}^M \delta(\lambda - \gamma_k^{(N)})$ . However, we show in the following that it is possible to reformulate the results of [17] in a more explicit manner by taking into account immediately that  $\frac{1}{M} \sum_{j=1}^M \delta(\lambda - \gamma_k^{(N)})$  is a discrete measure. We hope that the following analysis, based on quite elementary technics, is easier to follow than the general approach of [17].

Our approach is based on the study of the function  $w_N(z)$  that has been introduced in (17). We have established in Proposition 1 that  $w_N(x)$  is well defined in the real axis, and that it can be expressed as one of the roots of the polynomial equation in (19). Let us now see how this function can help us in the characterization of the support  $\mathcal{S}_N$ .

**Proposition 2.** *The function  $w_N(z)$  defined in (17) satisfies the following properties:*

- 1)  $\text{Int}(\mathcal{S}_N) = \{x \in \mathbb{R}_+ : \text{Im}\{w_N(x)\} > 0\}$
- 2)  $w'_N(x) > 0$ , for  $x \in \mathbb{R} \setminus \mathcal{S}_N$ .
- 3)  $1 - \sigma^2 c_N f_N(w_N(x)) > 0 \quad \forall x \in \mathbb{R} \setminus \mathcal{S}_N$ .

*Proof:* See Appendix B. ■

**Remark 3.** *By taking derivatives with respect to  $x$  on both sides of the equation  $\phi_N(w_N(x)) = x$ , we see that  $w'_N(x)\phi'_N(w_N(x)) = 1$  holds for  $x \in \mathbb{R} - \partial\mathcal{S}_N$ . Property 2 of the above proposition is thus equivalent to*

$$\phi'_N(w_N(x)) > 0 \text{ if } x \in \mathbb{R} \setminus \mathcal{S}_N. \quad (24)$$

Property 1 in Proposition 2 is basically stating the fact that the interior of the support  $\mathcal{S}_N$  coincides the region of values of  $\mathbb{R}_+$  for which the imaginary part of  $w_N(x)$  is strictly positive. Hence, it suffices to study the behavior of  $\text{Im}[w_N(x)]$  in order to characterize the interior of the support  $\mathcal{S}_N$ . On the other hand, we know from Property 8 in Proposition 1 that, for any  $x \in \mathbb{R} \setminus \partial\mathcal{S}_N$ ,  $w_N(x)$  is one of the solutions to the polynomial equation in (19). Proposition 2 is helping us to identify which one of the roots is in fact  $w_N(x)$ . More specifically, we will later show that:

- If  $x \in \text{Int}(\mathcal{S}_N)$ , then  $w_N(x)$  will be the unique root of (19) with positive imaginary part<sup>1</sup>, thanks to Property 1.
- If  $x \in \mathbb{R} \setminus \mathcal{S}_N$ , then  $w_N(x)$  will be the unique root of (19) such that Properties 2 and 3 hold.

In order to establish the fact that these properties completely determine the value of  $w_N(x)$  out of the set of roots of the equation in (19), we need to study the form of the function  $\phi_N$  in (20) more closely. The analysis of the roots of the corresponding equation in (19) will allow us to determine the intervals of  $\mathbb{R}$  for which  $w_N(x)$  is real-valued and the intervals in which it has a strictly positive imaginary part.

#### A. Characterization of the function $\phi_N(w)$

In the following, we assume that the  $K$  non-zero eigenvalues of the matrix  $\mathbf{B}_N \mathbf{B}_N^H$ , namely  $\{\gamma_{M-K+1}^{(N)}, \dots, \gamma_M^{(N)}\}$ , have multiplicity 1. Under this hypothesis, the equation in (19) is in fact equivalent to a polynomial equation of degree  $2(K+1)$ . This can be readily seen by using the expression of  $f_N(w)$  in (20), so that

<sup>1</sup>The existence and unicity of such root will be established in what follows.

we can express  $\phi_N(w)$  as sums of quotients of polynomials in  $w$ , i.e.

$$\begin{aligned} \phi_N(w) = & w \left( 1 + \sigma^2 \frac{M-K}{M} \frac{c_N}{w} - \sigma^2 \frac{c_N}{M} \sum_{m=M-K+1}^M \frac{1}{\gamma_m - w} \right)^2 \\ & + (1 - c_N) \sigma^2 \left( 1 + \sigma^2 \frac{M-K}{M} \frac{c_N}{w} - \sigma^2 \frac{c_N}{M} \sum_{m=M-K+1}^M \frac{1}{\gamma_m - w} \right). \end{aligned} \quad (25)$$

Hence, multiplying both sides of equation  $\phi_N(w) = x$  by  $w \prod_{m=M-K+1}^M (\gamma_m - w)^2$  we end up with a polynomial equation of degree  $2(K+1)$ . If certain eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$  are multiple,  $\phi_N(w) = x$  will be a polynomial equation of degree  $2(\overline{K}+1)$  where  $\overline{K}$  represents the number of distinct non zero eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$ . The following results can thus be immediately adapted by replacing  $K$  by  $\overline{K}$ . The assumption  $K = \overline{K}$  allows to avoid the introduction of new notations representing the distinct eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$  in the forthcoming analysis.

1) **Zeros of  $\phi_N(w)$ :** It is easily seen that the function  $\phi_N$  has exactly  $2K+2$  different real zeros, which will be denoted as  $z_0^{(N)-} < z_0^{(N)+} < \dots < z_K^{(N)-} < z_K^{(N)+}$ . An elementary analysis of the function  $\phi_N$  determines the position of these zeros, as well as the behavior of the function  $\phi_N(w)$  in their neighborhood:

- The lowest couple of zeros are located on the negative real axis, namely  $z_0^{(N)-}, z_0^{(N)+} \in ]-\infty, 0[$ . Furthermore, the function  $\phi_N$  is increasing at  $z_0^{(N)-}$  and decreasing at  $z_0^{(N)+}$ , namely  $\phi'_N(z_0^{(N)-}) > 0$  and  $\phi'_N(z_0^{(N)+}) < 0$ , where  $\phi'_N$  denotes the derivative of  $\phi_N$ .
- The next couple of zeros are located between zero and the first positive eigenvalue of  $\mathbf{B}_N \mathbf{B}_N^H$ , i.e.  $z_1^{(N)-}, z_1^{(N)+} \in ]0, \gamma_{M-K+1}^{(N)}[$ , and it turns out that the function  $\phi_N$  is decreasing at  $z_1^{(N)-}$  and increasing at  $z_1^{(N)+}$ , namely  $\phi'_N(z_1^{(N)-}) < 0$  and  $\phi'_N(z_1^{(N)+}) > 0$ .
- Each one of the remaining couples of zeros is located between two positive eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$ , i.e.  $z_k^{(N)-}, z_k^{(N)+} \in ]\gamma_{M-K+k-1}^{(N)}, \gamma_{M-K+k}^{(N)}[$ ,  $\forall k = 2, \dots, K$ , and the function  $\phi_N$  is always decreasing at the first zero and increasing at the second, i.e.  $\phi'_N(z_k^{(N)-}) < 0$  and  $\phi'_N(z_k^{(N)+}) > 0$ ,  $\forall k = 2, \dots, K$ .

In order to obtain these results, one only needs to factor  $\phi_N(w)$  as the product of two terms, namely

$$\phi_N(w) = [1 - c_N \sigma^2 f_N(w)] [w (1 - c_N \sigma^2 f_N(w)) + (1 - c_N) \sigma^2] \quad (26)$$

and therefore  $\phi_N(w) = 0$  if and only if one of these two terms is zero. Out of the  $2K+2$  zeros of the function  $\phi_N(w)$ , a total of  $K+1$  are the zeros of the first term in (26). More formally:

- The second zero, namely  $z_0^{(N)+}$ , is solution of the equation  $1 - \sigma^2 c_N f_N(w) = 0$ .



2) **Local extrema and monotonicity intervals of  $\phi_N(w)$ :** Next, we investigate the local extrema of the function  $\phi_N$ . The following proposition summarizes the most interesting properties of the positive local extrema.

**Proposition 3.** 1) *The function  $\phi_N$  admits  $2Q$  positive local extrema counting multiplicities (with  $1 \leq Q \leq K + 1$ ) whose preimages, denoted  $w_1^{(N)-} < 0 < w_1^{(N)+} \leq w_2^{(N)-} \dots \leq w_Q^{(N)-} < w_Q^{(N)+}$ , belong to the set  $\{w \in \mathbb{R} : 1 - \sigma^2 c_N f_N(w) > 0\}$*

2) *If we denote by  $x_k^{(N)-} = \phi_N(w_k^{(N)-})$  and  $x_k^{(N)+} = \phi_N(w_k^{(N)+})$  these positive extrema, then*

$$0 < x_1^{(N)-} < x_1^{(N)+} \leq x_2^{(N)-} \dots \leq x_Q^{(N)-} < x_Q^{(N)+} \quad (28)$$

3) *Each eigenvalue  $\gamma_l^{(N)}$  of  $\mathbf{B}_N \mathbf{B}_N^H$  belongs to one and only one of the intervals  $]w_q^{(N)-}, w_q^{(N)+}[$ ,  $q = 1 \dots Q$ .*

4) *The function  $\phi_N$  is increasing on the intervals  $] -\infty, w_1^{(N)-}[$ ,  $\{ [w_q^{(N)+}, w_{q+1}^{(N)-}] \}_{q=1, Q-1}$ , and  $[w_Q^{(N)+}, +\infty[$ . Moreover,*

$$\begin{aligned} \phi_N \left( ] -\infty, w_1^{(N)-}[ \right) &= ] -\infty, x_1^{(N)-}[ \\ \phi_N \left( [w_q^{(N)+}, w_{q+1}^{(N)-}] \right) &= [x_q^{(N)+}, x_{q+1}^{(N)-}] \text{ for each } q = 1, \dots, Q-1, \text{ and} \\ \phi_N \left( [w_Q^{(N)+}, +\infty[ \right) &= [x_Q^{(N)+}, +\infty[. \end{aligned}$$

*Proof:* Except for the inequalities in (28), which are proved in Appendix C, the statements of Proposition 3 follow directly from an elementary analysis of the function  $\phi_N$ .  $\blacksquare$

We see from Proposition 3 that the local extrema always appear in groups of two, and the actual number of extremum couples ( $Q$ ) will generally depend on  $\sigma^2$ ,  $c_N$  and on the positive eigenvalues of the matrix  $\mathbf{B}_N \mathbf{B}_N^H$ . For example, in the situation represented in Figure 1, the number of positive local extrema was equal to four, which implies that  $Q = 2$ . In Figures 2 and 3 we depict other equivalent examples of  $\phi_N$ , for which we had  $Q = 1$  and  $Q = 3$  respectively.

### B. Characterization of $w_N(x)$ out of the roots of $\phi_N(w) = x$

We know from Proposition 1 that  $w_N(x)$  for real valued  $x$  will be a solution of the equation  $\phi_N(w) = x$ . In this section, we will characterize which one of these roots is actually  $w_N(x)$ . First of all, observe that, since the equation  $\phi_N(w) = x$  is equivalent to a polynomial equation of degree  $2(K + 1)$ , the number of solutions (counting multiplicities) will always be equal to  $2(K + 1)$ . Out of these solutions, we can graphically find the real-valued ones by exploring the crossings between the graph of  $\phi_N(w)$

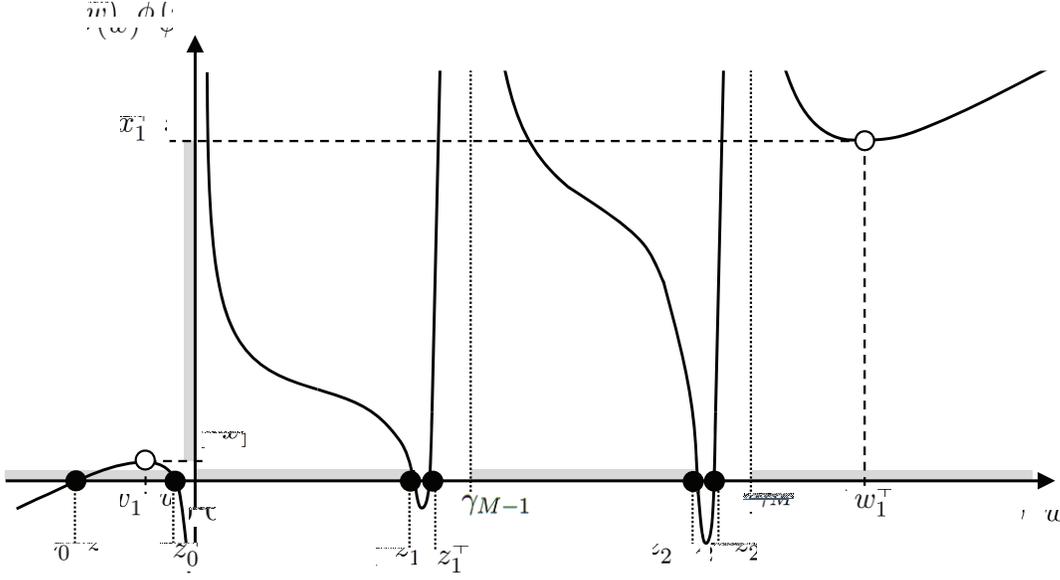


Figure 2. Typical representation of  $\phi_N(w)$  as a function of  $w$  for  $K = 2$  and  $Q = 1$  (we drop the dependence on  $N$  for clarity). The shaded region in the horizontal axis represents the set of points for which  $1 - \sigma^2 c_N f_N(w) > 0$ . The shaded region in the vertical axis represents  $\mathcal{S}_N$ .

and a horizontal line at  $x$ . This is further illustrated in Figure 4. By the properties of the function  $\phi_N(w)$  presented in Section IV-A, we can clearly differentiate between two different situations:

- If  $x \notin \bigcup_{k=1}^Q [x_k^{(N)-}, x_k^{(N)+}]$ , it is easily shown that the equation  $\phi_N(w) = x$  presents exactly  $2(K + 1)$  different real-valued solutions (cf. upper horizontal line in Figure 4). Since the original equation has degree  $2(K + 1)$ , there are no complex-valued solutions. In particular,  $w_N(x)$  will be real-valued.
- If  $x \in \bigcup_{k=1}^Q ]x_k^{(N)-}, x_k^{(N)+}[$ , in what follows, it will be shown that the equation  $\phi_N(w) = x$  has exactly  $2K$  different real-valued solutions (cf. lower horizontal line in Figure 4). This implies that there is a couple of complex conjugated solutions to the equation  $\phi_N(w) = x$ .

Let us now see how we can completely characterize  $w_N(x)$  in these two different situations:

1) **Case**  $x \in \mathbb{R} \setminus \bigcup_{k=1}^Q [x_k^{(N)-}, x_k^{(N)+}]$ : From (24) and Property 3 of Proposition 2, we know that  $w_N(x)$  is a root of the equation  $\phi_N(w) = x$  such that  $\phi'_N(w_N(x)) > 0$  and that  $1 - \sigma^2 c_N f_N(w_N(x)) > 0$ . We now prove that this completely characterizes  $w_N(x)$  out of the set of all roots of  $\phi_N(w) = x$ , in the sense that there is only one root of  $\phi_N(w) = x$  that has these two properties. We first consider

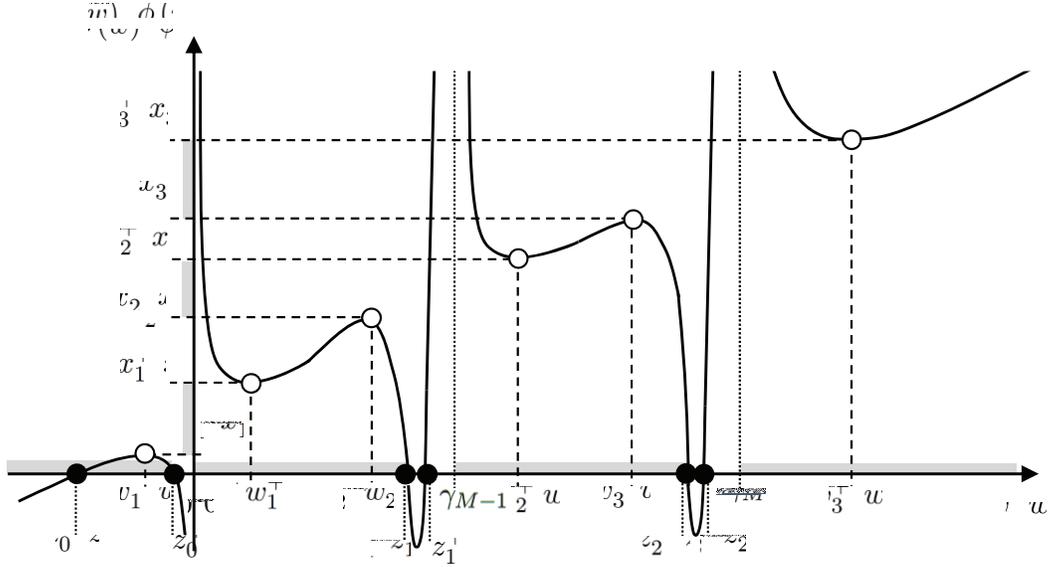


Figure 3. Typical representation of  $\phi_N(w)$  as a function of  $w$  for  $K = 2$  and  $Q = 3$  (we drop the dependence on  $N$  for clarity). The shaded region in the horizontal axis represents the set of points for which  $1 - \sigma^2 c_N f_N(w) > 0$ . The shaded region in the vertical axis represents  $\mathcal{S}_N$ .

the case  $x < x_1^{(N)-}$ . By Property 4 of Proposition 3,  $\phi_N$  is an increasing one to one correspondence from  $]-\infty, w_1^{(N)-}[$  onto  $]-\infty, x_1^{(N)-}[$ . Its inverse  $\phi_N^{-1}$  is thus a well defined increasing function from  $]-\infty, x_1^{(N)-}[$  onto  $]-\infty, w_1^{(N)-}[$ . We claim that  $w_N(x)$  coincides with  $\phi_N^{-1}(x)$ . Indeed, observe that since  $\phi_N^{-1}(x) < w_1^{(N)-}$ , we automatically have  $\phi'_N(\phi_N^{-1}(x)) > 0$  and that  $1 - \sigma^2 c_N f_N(\phi_N^{-1}(x)) > 0$ . On the other hand, the behavior of  $\phi_N$  established in Propositions 2 and 3 implies that the other real-valued solutions of  $\phi_N(w) = x$  do not satisfy either  $1 - \sigma^2 c_N f_N(w) > 0$  or  $\phi'_N(w) > 0$  (see further Figures 1 to 3). Therefore,  $w_N(x)$  can be expressed as  $\phi_N^{-1}(x)$ , and is the only root of  $\phi_N(w) = x$  such that  $1 - \sigma^2 c_N f_N(w) > 0$  and  $\phi'_N(w) > 0$ .

The above analysis can be extended if  $x$  belongs to  $]x_k^{(N)+}, x_{k+1}^{(N)-}[$  for  $k = 1, \dots, Q - 1$  or if  $x > x_Q^{(N)+}$ . Indeed, observe first that  $\phi_N$  is a bijection between  $]w_k^{(N)+}, w_{k+1}^{(N)-}[$  and  $]x_k^{(N)+}, x_{k+1}^{(N)-}[$  for  $k = 1, \dots, Q - 1$  and between  $]w_Q^{(N)+}, +\infty[$  and  $]x_Q^{(N)+}, +\infty[$ . Hence,  $\phi_N^{-1}$  is well defined on  $]x_k^{(N)+}, x_{k+1}^{(N)-}[$  for  $k = 1, \dots, Q - 1$  and on  $]x_Q^{(N)+}, +\infty[$ . Thanks to the form of the function  $\phi_N$ , we see that  $\phi_N^{-1}(x)$  is the only root that verifies  $1 - \sigma^2 c_N f_N(w) > 0$  and  $\phi'_N(w) > 0$  (see further Figures 1 to 3), and this implies that  $w_N(x) = \phi_N^{-1}(x)$ . Since  $w_N(x)$  is continuous on  $\mathbb{R}$ , we also get that  $w_N(x_k^{(N)-}) = w_k^{(N)-}$  as well as  $w_N(x_k^{(N)+}) = w_k^{(N)+}$  for  $k = 1, \dots, Q$ .



### C. Characterization of the support $\mathcal{S}_N$

As the interior of  $\mathcal{S}_N$  coincides with  $\{x \in \mathbb{R}_+, \text{Im}(w_N(x)) > 0\}$  (see Property 1 of Proposition 2), we have shown the following Theorem.

**Theorem 2.** *The support  $\mathcal{S}_N$  is given by*

$$\mathcal{S}_N = \bigcup_{k=1}^Q [x_k^{(N)-}, x_k^{(N)+}]. \quad (29)$$

The above analysis shows that  $x_1^{(N)-} < x_1^{(N)+} \leq x_2^{(N)-} < \dots < x_{Q-1}^{(N)+} \leq x_Q^{(N)-} < x_Q^{(N)+}$  coincide with the set of all positive extrema of  $\phi_N$ . Theorem 2 thus establishes a very simple method to determine the support  $\mathcal{S}_N$ . First, one needs to determine all the local extrema of  $\phi_N(w)$ , namely the solutions to the polynomial equation  $\phi'_N(w) = 0$ . The solutions will be  $\{w_1^{(N)-}, w_1^{(N)+}, \dots, w_Q^{(N)-}, w_Q^{(N)+}\}$  with possible repetitions if one of these roots has multiplicity two, plus  $K$  additional ones (it is easily seen that  $\phi_N$  has exactly  $K$  negative local minima). By evaluating the function  $\phi_N$  at these points, and selecting those for which  $\phi_N$  is positive, we are determining the values  $\{x_1^{(N)-}, x_1^{(N)+}, \dots, x_Q^{(N)-}, x_Q^{(N)+}\}$  that characterize the support in (29). Observe that the support  $\mathcal{S}_N$  is a disjoint reunion of compact intervals, which will be referred to as clusters. Each of these clusters  $[x_q^{(N)-}, x_q^{(N)+}]$  will be associated to an interval of the type  $[w_q^{(N)-}, w_q^{(N)+}]$ ,  $q = 1 \dots Q$ , in the sense that  $x_q^{(N)-} = \phi_N(w_q^{(N)-})$  and  $x_q^{(N)+} = \phi_N(w_q^{(N)+})$ . On the other hand, we can also clearly see that a specific eigenvalue  $\gamma_k^{(N)}$ ,  $k = 1, \dots, M$ , always belongs to one, and only one of the intervals  $[w_q^{(N)-}, w_q^{(N)+}]$ . This motivates the following definition.

**Definition 1.** *We say that the eigenvalue  $\gamma_k^{(N)}$ ,  $k = 1, \dots, M$ , of the matrix  $\mathbf{B}_N \mathbf{B}_N^H$  is associated with the cluster  $[x_q^{(N)-}, x_q^{(N)+}]$  if  $\gamma_k^{(N)} \in [w_q^{(N)-}, w_q^{(N)+}]$ .*

Observe that this is not a one-to-one correspondence, in the sense that multiple consecutive eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$  may be associated with the same cluster. For instance, in Figure 2 the three eigenvalues  $(0, \gamma_{M-1}^{(N)}$  and  $\gamma_M^{(N)})$  are associated with the same eigenvalue cluster, while in Figure 3 each eigenvalue is associated with its own different cluster.

The first cluster  $[x_1^{(N)-}, x_1^{(N)+}]$  plays a special role because it is always associated with the eigenvalue 0 of matrix  $\mathbf{B}_N \mathbf{B}_N^H$ . As seen below, the main results of this paper will be valid under the assumption that the strictly positive eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$  are not associated to the cluster  $[x_1^{(N)-}, x_1^{(N)+}]$ . Intuitively, this means that the eigenvalues corresponding to the noise subspace are separated from the eigenvalues of the signal subspace. Both Figure 1 and Figure 3 satisfy this property, but not Figure 2.

More rigorously, we assume from now on that the following hypotheses hold.

**(As 1):**  $\exists N_0 \in \mathbb{N}$  such that  $\forall N \in \mathbb{N}, N \geq N_0$ , the non zero eigenvalues  $\{\gamma_k^{(N)}\}_{k=M-K+1, \dots, M}$  of  $\mathbf{B}_N \mathbf{B}_N^H$  are not associated to the first cluster  $[x_1^{(N)-}, x_1^{(N)+}]$ .

**(As 2):**  $\exists t_1^- > 0, t_1^+, t_2^- \in \mathbb{R}$  independent of  $N$  such that

$$t_1^- < \inf_{N \geq N_0} \{x_1^{(N)-}\} < \sup_{N \geq N_0} \{x_1^{(N)+}\} < t_1^+ < t_2^- < \inf_{N \geq N_0} \{x_2^{(N)-}\} \quad \forall N \geq N_0. \quad (30)$$

These two assumptions imply that for each  $N \geq N_0$ , the eigenvalue 0 of  $\mathbf{B}_N \mathbf{B}_N^H$  belongs to the interval  $]w_1^{(N)-}, w_1^{(N)+}[$  and thus to  $]w_N(t_1^-), w_N(t_1^+)[$  because  $w_N(t_1^-) < w_1^{(N)-}$  and  $w_N(t_1^+) > w_1^{(N)+}$ . Similarly, the non zero eigenvalues  $\{\gamma_{M-K+l}^{(N)}\}_{l=1, \dots, K}$  of  $\mathbf{B}_N \mathbf{B}_N^H$  satisfy  $\gamma_{M-K+l}^{(N)} > w_N(t_2^-)$ .

## V. CONVERGENCE AND LOCALIZATION OF THE SAMPLE EIGENVALUES

The previous results are related to the properties of the limit deterministic distribution  $\mu_N$ . The almost sure convergence of  $\hat{\mu}_N - \mu_N$  towards 0 does not mean by itself that the eigenvalues of  $\hat{\mathbf{R}}_N$  belong almost surely to  $\mathcal{S}_N$ , or to an interval containing  $\mathcal{S}_N$ . As one may imagine, it is important to be able to locate the eigenvalues  $(\hat{\lambda}_k^{(N)})_{k=1, \dots, M}$  of matrix  $\hat{\mathbf{R}}_N$  with respect to  $\mathcal{S}_N$  for  $N$  large enough. Bai and Silverstein established in [12], [13] powerful related results in the context of correlated zero-mean, possibly non Gaussian, random matrices. In the following, we establish similar results for the Information plus Noise model. However, the mathematical approach we use in the present paper has no connection with the techniques used in [12], [13] also valid in the non Gaussian case. Since  $\Sigma$  is assumed Gaussian, we rather adapt to the Information plus Noise model the ideas developed in [14] in the context of Gaussian Wigner matrices. We prove in the following two theorems which are believed to be of independent interest.

**Theorem 3.** *Assume that there exists a positive quantity  $\epsilon > 0$ , two real values  $a, b \in \mathbb{R}$ , and an integer  $N_0$  such that*

$$]a - \epsilon, b + \epsilon[ \cap \mathcal{S}_N = \emptyset \quad \forall N \in \mathbb{N}, N \geq N_0 \quad (31)$$

*where  $\mathcal{S}_N$  denotes the support of  $\mu_N$ . Then, with probability one, no eigenvalue of  $\hat{\mathbf{R}}_N$  appears in  $[a, b]$  for all  $N$  large enough.*

**Theorem 4.** *If Assumptions 1 and 2 hold, then, for all  $N$  large enough, with probability one,*

$$\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_{M-K}^{(N)} \in ]t_1^-, t_1^+ [ \quad (32)$$

$$\hat{\lambda}_{M-K+1}^{(N)} > t_2^- \quad (33)$$

Although Assumptions 1 and 2 depend on the deterministic distributions  $\mu_N$ , Theorem 4 shows that almost surely, the smallest  $M - K$  eigenvalues of  $\hat{\mathbf{R}}_N$  are always separated from the others for all  $N$  large enough.

#### A. Proof of Theorem 3

We first state the following proposition, the proof of which is demanding, and is detailed in Appendix E. The result will play a fundamental role in the proof of Theorem 3.

**Proposition 4.**  $\forall z \in \mathbb{C} \setminus \mathbb{R}_+$ , we have for  $N$  large enough,

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} [\mathbf{Q}_N(z)] \right] = \frac{1}{M} \text{Tr} [\mathbf{T}_N(z)] + \frac{1}{N^2} \chi_N(z)$$

with  $\chi$  is analytic in  $\mathbb{C} - \mathbb{R}_+$  and satisfies

$$|\chi_N(z)| \leq (|z| + C)^k \mathbb{P}(|\text{Im}(z)|^{-1}) \quad (34)$$

for each  $z \in \mathbb{C}_+$  where  $C$  is a constant,  $k$  is an integer independent of  $N$  and  $\mathbb{P}$  is a polynomial with positive coefficients independent of  $N$ .

We now follow [18] and [14] and prove the Lemma:

**Lemma 2.** *Let  $\phi$  be a compactly supported real-valued smooth function defined on  $\mathbb{R}$ , i.e.  $\phi \in \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$ .*

*Then<sup>2</sup>,*

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} [\phi(\mathbf{\Sigma}_N \mathbf{\Sigma}_N^H)] \right] - \int_{\mathcal{S}_N} \phi(\lambda) d\mu_N(\lambda) = \mathcal{O}\left(\frac{1}{N^2}\right) \quad (35)$$

*Proof:* We first note that, by Property 7 in Lemma 1, we can write

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} [\phi(\mathbf{\Sigma}_N \mathbf{\Sigma}_N^H)] \right] = \frac{1}{\pi} \lim_{y \downarrow 0} \text{Im} \left\{ \int_{\mathbb{R}_+} \phi(x) \mathbb{E} \left[ \frac{1}{M} \text{Tr} [\mathbf{Q}_N(x + iy)] \right] dx \right\}$$

as well as

$$\left[ \int_{\mathcal{S}_N} \phi(\lambda) d\mu_N(\lambda) \right] = \frac{1}{\pi} \lim_{y \downarrow 0} \text{Im} \left\{ \int_{\mathbb{R}_+} \phi(x) \left[ \frac{1}{M} \text{Tr} [\mathbf{T}_N(x + iy)] \right] dx \right\}$$

<sup>2</sup>By applying the function  $\phi$  to a Hermitian matrix, we implicitly represent the action of  $\phi$  on the corresponding eigenvalues.

Therefore, using Proposition 4, we can express the right hand side of (35) as

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} [\phi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] - \int_{\mathcal{S}_N} \phi(\lambda) \mu_N(d\lambda) = \frac{1}{N^2} \frac{1}{\pi} \lim_{y \downarrow 0} \text{Im} \left\{ \int_{\mathbb{R}_+} \phi(x) \chi_N(x + iy) dx \right\} \quad (36)$$

Since the function  $\chi_N(z)$  satisfies the inequality (34), the Appendix of [19] implies that

$$\limsup_{y \downarrow 0} \left| \int_{\mathbb{R}} \varphi(x) \chi_N(x + iy) dx \right| \leq C < +\infty$$

where  $C$  is a constant independent of  $N$ . Hence, (36) readily implies (35).  $\blacksquare$

In order to establish Theorem 3, we consider a function  $\psi \in \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$  satisfying  $0 \leq \psi \leq 1$  and

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

Condition (31) implies that  $\int_{\mathcal{S}_N} \psi(\lambda) d\mu_N(\lambda) = 0$  if  $N$  is large enough. Therefore, (35) implies that

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] = \mathcal{O} \left( \frac{1}{N^2} \right).$$

We now establish that

$$\text{Var} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] = \mathcal{O} \left( \frac{1}{N^4} \right) \quad (37)$$

In order to prove (37), we use the Nash-Poincaré inequality [20], [21], [22], [14] which implies that

$$\text{Var} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \leq \frac{\sigma^2}{N} \sum_{i,j} \mathbb{E} \left[ \left| \frac{\partial}{\partial \mathbf{W}_{ij}} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \right|^2 + \left| \frac{\partial}{\partial \mathbf{W}_{ij}^*} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \right|^2 \right] \quad (38)$$

where  $\mathbf{W}_{ij}$  denotes the  $(i, j)$ th entry of matrix  $\mathbf{W}$  defined in (2). Now, applying e.g. [18, Lemma 4.6]

we can readily see that

$$\frac{\partial}{\partial \mathbf{W}_{ij}} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] = \frac{1}{M} [\boldsymbol{\Sigma}_N^H \psi' (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)]_{j,i} \quad (39)$$

$$\frac{\partial}{\partial \mathbf{W}_{ij}^*} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] = \frac{1}{M} [\psi' (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H) \boldsymbol{\Sigma}_N]_{i,j} \quad (40)$$

where  $\psi'$  denotes the derivative of  $\psi$ . Consequently, the sum on the right hand side of (38) can be written as

$$\begin{aligned} \sum_{i,j} \mathbb{E} \left[ \left| \frac{\partial}{\partial \mathbf{W}_{ij}} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \right|^2 + \left| \frac{\partial}{\partial \mathbf{W}_{ij}^*} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \right|^2 \right] &= \\ &= \frac{2}{M^2} \mathbb{E} \left[ \text{Tr} \left[ [\psi' (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)]^2 \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H \right] \right]. \end{aligned}$$

This yields

$$\text{Var} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \leq C \frac{1}{N^2} \mathbb{E} \left[ \frac{1}{M} \text{Tr} \left[ [\psi' (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)]^2 \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H \right] \right] \quad (41)$$

for some constant  $C$  independent of  $N$ . Next, consider the function  $h(\lambda)$ , defined as  $h(\lambda) = \lambda [\psi'(\lambda)]^2$ , which clearly belongs to  $\mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$ . Lemma 2 implies that

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} \left[ [\psi'(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)]^2 \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H \right] \right] = \int_{\mathcal{S}_N} h(\lambda) d\mu_N(\lambda) + \mathcal{O} \left( \frac{1}{N^2} \right).$$

But it is clear from (31) that  $\int_{\mathcal{S}_N} h(\lambda) d\mu_N(\lambda) = 0$  if  $N$  is large enough. Therefore, (41) gives (37).

We are now in position to complete the proof of Theorem 3 as in [14]. Applying the classical Markov inequality together with the above results, we can write (for  $N$  large enough)

$$\begin{aligned} \mathbb{P} \left( \frac{1}{M} \text{Tr} [\psi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] > \frac{1}{N^{4/3}} \right) &\leq N^{8/3} \mathbb{E} \left[ \left| \frac{1}{M} \text{Tr} [\psi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right|^2 \right] \\ &= N^{8/3} \left( \left| \mathbb{E} \left[ \frac{1}{M} \text{Tr} [\psi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \right|^2 + \text{Var} \left[ \frac{1}{M} \text{Tr} [\psi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] \right) = \mathcal{O} \left( \frac{1}{N^{4/3}} \right) \end{aligned} \quad (42)$$

Then, by Borel-Cantelli lemma, for  $N$  large enough, we have with probability one,

$$\frac{1}{M} \text{Tr} [\psi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \leq \frac{1}{N^{4/3}}$$

By the very definition of  $\psi$ , the number of eigenvalues of  $\hat{\mathbf{R}}_N = \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H$  in  $[a, b]$  is upper-bounded by  $\text{Tr} [\psi(\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)]$  and is therefore a  $\mathcal{O}(N^{-\frac{1}{3}})$  with probability one. Since this number has to be an integer, we deduce that for  $N$  large enough, there is no eigenvalue in  $[a, b]$ . This completes the proof of Theorem 3.

### B. Proof of Theorem 4

The approach we use to establish Theorem 4 differs from the method of [14] which is inspired by [13]. The first part our proof is similar to the proof of Theorem 3, and thus we will omit certain details. For the second part, we will need a certain result that we summarize in the following proposition:

**Proposition 5.** *Consider the curve  $\mathcal{C}$  defined by the complex valued function  $w_N(x)$  in (17) on the complex plane as  $x$  moves from  $t_1^-$  to  $t_1^+$ , concatenated with the function  $w_N^*(x)$  as  $x$  moves back from  $t_1^+$  to  $t_1^-$ , namely*

$$\mathcal{C} = \{w_N(x) : x \in [t_1^-, t_1^+]\} \cup \{w_N^*(x) : x \in [t_1^-, t_1^+]\}. \quad (43)$$

*This is a closed curve that encloses the points of  $]w_1^-, w_1^+[$  (see further Figure 5). Let  $\psi(z)$  be a function holomorphic in a neighborhood of  $\mathcal{C}$ . Then, the contour integral  $\int_{\mathcal{C}^-} \psi(\lambda) d\lambda$  is well defined by*

$$\oint_{\mathcal{C}^-} \psi(\lambda) d\lambda = 2i \text{Im} \left[ \int_{[t_1^-, t_1^+]} \psi(w_N(x)) w_N'(x) dx \right]. \quad (44)$$

where  $w'_N(z)$  denotes the derivative of  $w_N(z)$  and where the symbol  $\mathcal{C}^-$  means that  $\mathcal{C}$  is oriented clockwise.

Finally, let  $\xi \in \mathbb{R}$  a point that does not belong to  $[w_N(t_1^-), w_1^{(N)-}] \cup [w_1^{(N)+}, w_N(t_1^+)]$ . Then,

$$\text{Ind}_{\mathcal{C}}(\xi) = \frac{1}{2i\pi} \int_{\mathcal{C}^-} \frac{d\lambda}{\xi - \lambda} = \begin{cases} 1 & \text{if } \xi \in ]w_1^{(N)-}, w_1^{(N)+}[ \\ 0 & \text{if } \xi < w_N(t_1^-) \text{ or } \xi > w_N(t_1^+), \end{cases}$$

*Proof:* According to the discussion in Section IV-B, if  $x \in [t_1^-, x_1^{(N)-}]$ , then  $w_N(x)$  is real-valued, and increases from  $w_N(t_1^-)$  to  $w_N(x_1^{(N)-}) = w_1^{(N)-}$ . For  $x \in ]x_1^{(N)-}, x_1^{(N)+}[$ , the point  $w_N(x)$  belongs to  $\mathbb{C}_+$ . Finally, if  $x \in [x_1^{(N)+}, t_1^+]$ ,  $w_N(x)$  is again real-valued, and increases from  $w_N(x_1^{(N)+}) = w_1^{(N)+}$  to  $w_N(t_1^+)$ . The contour  $\mathcal{C}$  is therefore well defined and encloses the points of  $]w_1^{(N)-}, w_1^{(N)+}[$ .

Let us now prove (44). Observe that the function  $x \rightarrow w_N(x)$  is not exactly a piecewise continuously differentiable function on  $[t_1^-, t_1^+]$  because  $|w'_N(x)|$  increases without bound when  $x \rightarrow x_1^{(N)-}, x_1^{(N)+}$ . To see that  $w_N(x)$  can indeed be used as a valid parametrization of  $\mathcal{C}$ , we need to see that the integral in (44) is well defined. It is thus necessary to study the behavior of  $w'_N$  around the points  $\{x_1^{(N)-}, x_1^{(N)+}\}$ . The following lemma is an immediate consequence of the analysis of the behavior of the density of measure  $\mu_N$  near a point of  $\partial\mathcal{S}_N$  provided in [17] (see Appendix D for a proof).

**Lemma 3.** *There exists neighborhoods  $\mathcal{V}(x_1^{(N)-})$  and  $\mathcal{V}(x_1^{(N)+})$  of  $x_1^{(N)-}$  and  $x_1^{(N)+}$  such that*

$$|w'_N(x + iy)| \leq \frac{C}{\sqrt{|x - x_1^{(N)-}|}} \text{ for } y \geq 0, x + iy \in \mathcal{V}(x_1^{(N)-}), \text{ and } x \neq x_1^{(N)-} \quad (45)$$

and

$$|w'_N(x + iy)| \leq \frac{C}{\sqrt{|x - x_1^{(N)+}|}} \text{ for } y \geq 0, x + iy \in \mathcal{V}(x_1^{(N)+}) \text{ and } x \neq x_1^{(N)+} \quad (46)$$

In particular, Lemma 3 implies that  $\int_{[t_1^-, t_1^+]} |\psi(w_N(x))| |w'_N(x)| dx < +\infty$  so that the right hand side of (44) is well defined. The reader may check that it is possible to use the usual results related to integrals over piecewise continuously differentiable contours. In particular, as  $\text{Im}(w_N(x)) > 0$  if  $x \in ]x_1^{(N)-}, x_1^{(N)+}[$ , the index of a point  $\xi \in \mathbb{R}$  which does not belong to  $[w_N(t_1^-), w_1^{(N)-}] \cup [w_1^{(N)+}, w_N(t_1^+)]$  is equal to 1 is  $\xi \in ]w_1^{(N)-}, w_1^{(N)+}[$  and to 0 if either  $\xi < w_N(t_1^-)$  or  $w > w_N(t_1^+)$ . ■

Proposition 5 is basically pointing out that the function  $w_N(x)$  defines a valid parametrization of a contour that will not intersect with any eigenvalue of  $\mathbf{B}_N \mathbf{B}_N^H$ . Furthermore, Assumptions 1 and 2 imply that

$$\text{Ind}_{\mathcal{C}}(0) = 1 \quad (47)$$

and

$$\text{Ind}_{\mathcal{C}}(\gamma_{M-K+l}) = 0 \quad (48)$$

for  $l = 1, \dots, K$ . This means that the contour will only enclose the zero eigenvalue, and none of the positive eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$ , which will be of crucial importance in the following development. Figure 5 gives a schematic representation of the form of the contour  $\mathcal{C}$ .

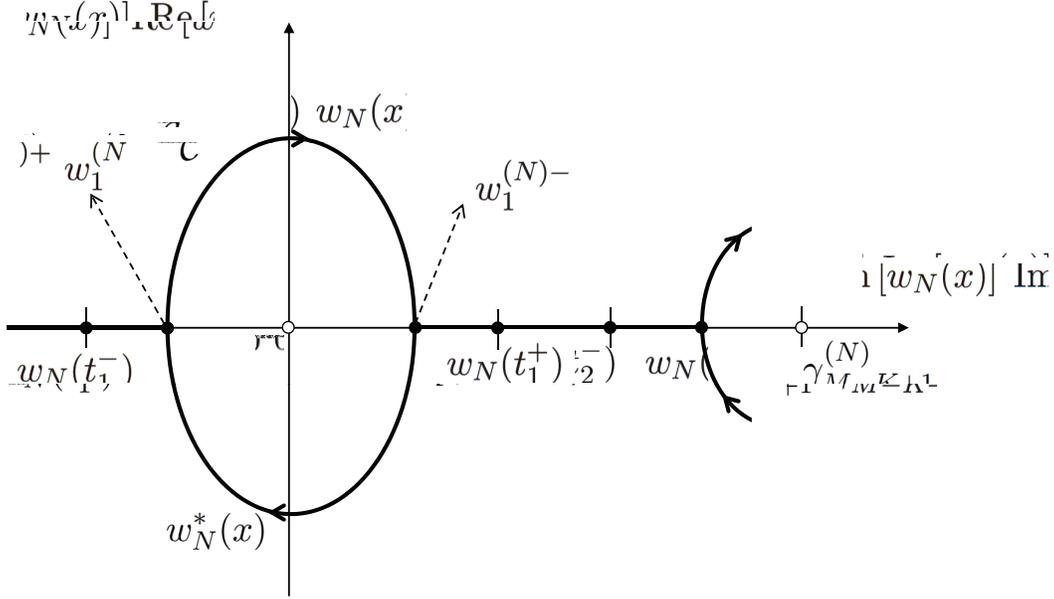


Figure 5. Representation of the contour  $\mathcal{C}$  on the complex plane.

Having introduced the result in Proposition 5, we are now in the position of establishing the proof of Theorem 4. Let  $\psi \in \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$  such that  $0 \leq \psi \leq 1$  and

$$\psi(\lambda) = \begin{cases} 1 & \forall \lambda \in [t_1^-, t_1^+] \\ 0 & \forall \lambda \in \mathbb{R} - [t_1^- - \epsilon, t_1^+ + \epsilon] \end{cases}$$

with  $\epsilon$  chosen in such a way that  $t_1^+ + \epsilon < t_2^-$ . Since  $\psi \in \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$ , we can use Lemma 2 to get

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} [\psi (\Sigma_N \Sigma_N^H)] \right] = \int_{\mathbb{R}_+} \psi(\lambda) d\mu_N(\lambda) + \mathcal{O} \left( \frac{1}{N^2} \right).$$

Assumptions 1 and 2 imply that

$$\int_{\mathbb{R}_+} \psi(\lambda) d\mu_N(\lambda) = \mu_N \left( [x_1^{(N)-}, x_1^{(N)+}] \right) = \mu_N([t_1^-, t_1^+])$$

for  $N$  large enough. This leads to

$$\mathbb{E} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] = \mu_N([t_1^-, t_1^+]) + \mathcal{O} \left( \frac{1}{N^2} \right)$$

As established in (37), we also have

$$\text{Var} \left[ \frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] \right] = \mathcal{O} \left( \frac{1}{N^4} \right)$$

because  $\text{supp}(\psi') \cap \mathcal{S}_N = \emptyset$  for  $N$  large enough. Therefore, using again the proof of theorem 3 (inequality (42)), we get that

$$\frac{1}{M} \text{Tr} [\psi (\boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^H)] - \mu_N ([t_1^-, t_1^+]) = \mathcal{O} \left( \frac{1}{N^{4/3}} \right) \quad \text{a.s.} \quad (49)$$

Let us now find a closed form expression for  $\mu_N([t_1^-, t_1^+])$ . Noting that  $\mu_N$  is absolutely continuous with density  $\frac{1}{\pi} \text{Im}(m_N(x))$ , we can write

$$\mu_N ([t_1^-, t_1^+]) = \frac{1}{\pi} \text{Im} \left[ \int_{t_1^-}^{t_1^+} m_N(x) dx \right].$$

By expressing the Stieltjès transform as  $m_N(x) = \frac{f_N(w_N(x))}{1 - \sigma^2 c_N f_N(w_N(x))}$  (see further (22)),  $\mu_N ([t_1^-, t_1^+])$  can be written as

$$\mu_N ([t_1^-, t_1^+]) = \frac{1}{\pi} \text{Im} \left[ \int_{t_1^-}^{t_1^+} \frac{f_N(w_N(x))}{1 - \sigma^2 c_N f_N(w_N(x))} dx \right]$$

In order to express  $\mu_N ([t_1^-, t_1^+])$  in terms of an integral over the contour  $\mathcal{C}$ , we can use the relation  $w'_N(x) \phi'_N(w_N(x)) = 1$  for each  $x \in \mathbb{R} - \partial \mathcal{S}_N$  (see further (19)). Now, using Proposition 5, we see that

$$\begin{aligned} \mu_N ([t_1^-, t_1^+]) &= \frac{1}{\pi} \text{Im} \left[ \int_{[t_1^-, t_1^+]} \frac{f_N(w_N(x)) \phi'_N(w_N(x))}{1 - \sigma^2 c_N f_N(w_N(x))} w'_N(x) dx \right] = \frac{1}{2\pi i} \oint_{\mathcal{C}^-} \frac{f_N(\lambda) \phi'_N(\lambda)}{1 - \sigma^2 c_N f_N(\lambda)} d\lambda \\ &= \frac{1}{2\pi i} \oint_{\mathcal{C}^-} f_N(\lambda) \frac{(1 - c_N \sigma^2 f_N(\lambda))^2 - 2c_N \sigma^2 \lambda f'_N(\lambda) (1 - c_N \sigma^2 f_N(\lambda)) - c_N \sigma^4 (1 - c_N) f'_N(\lambda)}{1 - \sigma^2 c_N f_N(\lambda)} d\lambda \end{aligned} \quad (50)$$

The integrand of the right hand side of (50) is a meromorphic function. The contour integral can be thus evaluated using the residue theorem. The poles of the integrand are the eigenvalues of  $\mathbf{B}_N \mathbf{B}_N^H$  as well as the solutions of the equation  $1 - \sigma^2 c_N f_N(\lambda) = 0$ . This equation has  $K + 1$  real-valued solutions that we have denoted  $z_0^{(N)+}$ , and  $\{z_l^{(N)-}\}_{l=1, \dots, K}$  (see further Figures 1 to 3). Assumptions 1 and 2 imply that only the poles  $\{0\}$  and  $\{z_0^{(N)+}\}$  of the integrand are in fact enclosed by  $\mathcal{C}$ . Using the residue theorem, and after some straightforward calculations, we obtain a closed form for the above integral, namely

$$\mu_N ([t_1^-, t_1^+]) = \frac{M - K}{M} \alpha_1^{(N)} + \frac{1}{M} \sum_{k=M-K+1}^M \alpha_k^{(N)}$$

with

$$\alpha_1^{(N)} = \frac{N-K}{M-K} \left( 1 - \frac{\sigma^2 c_N}{M} \sum_{l=M-K+1}^M \frac{1}{\gamma_l^{(N)}} \right) + \frac{\sigma^2(1-c_N)}{z_0^{(N)+}} \quad (51)$$

$$\alpha_k^{(N)} = \left( 1 - \frac{K}{N} \right) \frac{\sigma^2}{\gamma_k^{(N)}} + \frac{\sigma^2(1-c_N)}{z_0^{(N)+} - \gamma_k^{(N)}} \quad (52)$$

Therefore, we can write

$$\mu_N([t_1^-, t_1^+]) = \frac{N-K}{M} + \sigma^2 \frac{1-c_N}{M} \left( \frac{M-K}{M} \frac{1}{z_0^{(N)+}} + \sum_{k=M-K+1}^M \frac{1}{z_0^{(N)+} - \gamma_k^{(N)}} \right) \quad (53)$$

$$= \frac{N-K}{M} - \sigma^2(1-c_N) f_N(z_0^{(N)+}) \quad (54)$$

but, using the fact that  $1 - \sigma^2 c_N f_N(z_0^{(N)+}) = 0$ , we obtain that  $\mu_N([t_1^-, t_1^+]) = \frac{M-K}{M}$ . Inserting this into (49), we get

$$\text{Tr}[\psi(\Sigma_N \Sigma_N^H)] - (M-K) = \mathcal{O}\left(\frac{1}{N^{1/3}}\right)$$

with probability 1. Moreover, thanks to theorem 3, no eigenvalue of  $\Sigma_N \Sigma_N^H$  appears in  $[t_1^- - \epsilon, t_1^-] \cup [t_1^+, t_1^+ + \epsilon]$  almost surely for  $N$  large enough. Therefore, almost surely for  $N$  large enough,  $\text{Tr}[\psi(\Sigma_N \Sigma_N^H)]$  coincides with the number of eigenvalues of  $\Sigma_N \Sigma_N^H$  contained in the interval  $]t_1^-, t_1^+[$ . This number is thus equal to  $M-K$ . These eigenvalues are moreover the  $M-K$  smallest ones: otherwise the smallest eigenvalue of  $\Sigma_N \Sigma_N^H$  would belong to  $[0, t_1^-]$ , a contradiction by Theorem 3. Finally, Theorem 3 again implies that  $\hat{\lambda}_{M-K+1}^{(N)} > t_2^-$ . This completes the proof of Theorem 4.

## VI. CONSISTENT ESTIMATION OF THE LOCALIZATION FUNCTION

We now present a consistent estimator  $\eta_N = \mathbf{b}_N^H \mathbf{\Pi}_N \mathbf{b}_N$  of the subspace method localization function. Here,  $\mathbf{b}_N$  represents a  $M$ -dimensional deterministic vector, and we assume that  $\sup_N \|\mathbf{b}_N\| < \infty$ . The new consistent estimator presented in this section can be seen as an extension of the work in [7], which implicitly assumes that the useful signals are Gaussian random i.i.d. sequences. In order to simplify the notation, we drop the dependence on  $N$  from all the sample eigenvalues and sample eigenvectors.

**Theorem 5.** *Under Assumptions 1 and 2, we have with probability one,*

$$\hat{\eta}_N^{new} - \eta_N \longrightarrow 0$$

where  $\hat{\eta}_N^{new}$  is defined by

$$\hat{\eta}_N^{new} = \sum_{k=1}^M \hat{\xi}_k \mathbf{b}_N^H \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^H \mathbf{b}_N \quad (55)$$

Here, the coefficients  $\{\hat{\xi}_k\}_{k=1,\dots,M-K}$  are given by

$$\hat{\xi}_k = 1 + \frac{\sigma^2 c_N}{M} \sum_{l=M-K+1}^M \frac{\hat{\lambda}_k + \hat{\lambda}_l}{(\hat{\lambda}_k - \hat{\lambda}_l)^2} + \sigma^2(1 - c_N) \sum_{l=M-K+1}^M \left( \frac{1}{\hat{\lambda}_k - \hat{\lambda}_l} - \frac{1}{\hat{\lambda}_k - \hat{\omega}_l} \right) \quad (56)$$

and  $\{\hat{\xi}_k\}_{k=M-K+1,\dots,M}$  by

$$\hat{\xi}_k = -\frac{\sigma^2 c_N}{M} \sum_{l=1}^{M-K} \frac{\hat{\lambda}_k + \hat{\lambda}_l}{(\hat{\lambda}_k - \hat{\lambda}_l)^2} - \sigma^2(1 - c_N) \sum_{l=1}^{M-K} \left( \frac{1}{\hat{\lambda}_k - \hat{\lambda}_l} - \frac{1}{\hat{\lambda}_k - \hat{\omega}_l} \right) \quad (57)$$

and where  $\{\hat{\omega}_l\}_{l=1,\dots,M}$  represent the solutions (arranged in increasing order) of the equation

$$1 + \frac{\sigma^2 c_N}{M} \text{Tr} \left[ (\mathbf{\Sigma}_N \mathbf{\Sigma}_N^H - x \mathbf{I}_M)^{-1} \right] = 0. \quad (58)$$

We remark that the consistent estimator is a linear combination of the terms  $(\mathbf{b}_N^H \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^H \mathbf{b}_N)_{k=1,\dots,M}$ . In contrast to the traditional estimator  $\eta_{trad} = \sum_{k=1}^{M-K} \mathbf{b}_N^H \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^H \mathbf{b}_N$ , it contains contributions of both the noise subspace and the signal subspace. We also note that the assumptions 1 and 2 and Theorem 4 are intuitively important because the various sums on the right hand side of (56) and (57) remain bounded: in (56) and (57), the terms  $|\hat{\lambda}_k - \hat{\lambda}_l|$  are greater than  $t_2^- - t_1^+$ , and it will be shown that a similar property holds for the terms  $|\hat{\lambda}_k - \hat{\omega}_l|$ .

**Remark 4.** It is worth pointing out that whenever the number of samples is forced to be much larger than the observation dimension ( $N \gg M$  or equivalently  $c_N \rightarrow 0$ ), the proposed estimator converges to the classical sample eigenvector estimate. This can be readily seen by taking the limit as  $c_N \rightarrow 0$  in the coefficients of (56) and (57) and noticing that  $\hat{\omega}_l \rightarrow \hat{\lambda}_l$  when  $c_N \rightarrow 0$ . Hence, as  $c_N \rightarrow 0$  we have  $\hat{\xi}_k \rightarrow 1$  for  $k = 1, \dots, M-K$ , and  $\hat{\xi}_k \rightarrow 0$  for  $k = M-K+1, \dots, M$ , implying that  $\hat{\eta}_N^{new} - \hat{\eta}_N^{trad} \rightarrow 0$ . This shows that the proposed estimator is in fact a generalization of the classical sample eigenvector estimate.

The remaining of this section is devoted to presenting the main points of the proof of Theorem 5. The starting point consists in remarking that Assumptions 1 and 2 imply that

$$\eta_N = \frac{1}{2\pi i} \oint_{\mathcal{C}^-} \mathbf{b}_N^H (\mathbf{B}_N \mathbf{B}_N^H - \lambda \mathbf{I}_M)^{-1} \mathbf{b}_N d\lambda$$

where  $\mathcal{C}^-$  is the closed path defined by (43). This leads to

$$\begin{aligned} \eta_N &= \frac{1}{2\pi i} \int_{t_1^-}^{t_1^+} \mathbf{b}_N^H (\mathbf{B}_N \mathbf{B}_N^H - w_N(x) \mathbf{I}_M)^{-1} \mathbf{b}_N w'_N(x) dx + \\ &\quad - \frac{1}{2\pi i} \int_{t_1^-}^{t_1^+} \mathbf{b}_N^H (\mathbf{B}_N \mathbf{B}_N^H - w_N^*(x) \mathbf{I}_M)^{-1} \mathbf{b}_N (w'_N(x))^* dx = \\ &\quad \frac{1}{\pi} \operatorname{Im} \left( \int_{t_1^-}^{t_1^+} \mathbf{b}_N^H (\mathbf{B}_N \mathbf{B}_N^H - w_N(x) \mathbf{I}_M)^{-1} \mathbf{b}_N w'_N(x) dx \right). \end{aligned} \quad (59)$$

Let  $g_N(x + iy) = \mathbf{b}_N^H (\mathbf{B}_N \mathbf{B}_N^H - w_N(x + iy) \mathbf{I}_M)^{-1} \mathbf{b}_N w'_N(x + iy)$ . The function  $y \rightarrow g_N(x + iy)$  is continuous on  $\mathbb{R}_+$  for each  $x \in \mathbb{R} \setminus \partial \mathcal{S}_N$  thanks to Proposition 1. Lemma 3 and the dominated convergence theorem imply that

$$\eta_N = \lim_{y \downarrow 0} \frac{1}{\pi} \operatorname{Im} \left( \int_{t_1^-}^{t_1^+} \mathbf{b}_N^H (\mathbf{B}_N \mathbf{B}_N^H - w_N(x + iy) \mathbf{I}_M)^{-1} \mathbf{b}_N w'_N(x + iy) dx \right) \quad (60)$$

$$= \lim_{y \downarrow 0} \left[ \frac{1}{2\pi i} \oint_{\partial \mathcal{R}_y^-} g_N(z) dz - \frac{1}{2\pi} \int_{-y}^y g_N(t_1^- + ih) dh + \frac{1}{2\pi} \int_{-y}^y g_N(t_1^+ - ih) dh \right] \quad (61)$$

where  $\partial \mathcal{R}_y^-$  is the boundary (clockwise oriented) of the rectangle  $\mathcal{R}_y$  defined for  $y > 0$  by

$$\mathcal{R}_y = \{u + iv : u \in [t_1^-, t_1^+], v \in [-y, y]\}. \quad (62)$$

Notice that the last two integrands vanish as  $y \downarrow 0$  (since the function  $v \mapsto g_N(t_1^- + iv)$  is continuous on  $[-y, y]$ ), and thus

$$\eta_N = \lim_{y \downarrow 0} \frac{1}{2\pi i} \oint_{\partial \mathcal{R}_y^-} g_N(z) dz.$$

Moreover, since  $g_N(z)$  is holomorphic in  $\mathbb{C} \setminus [x_1^{(N)-}, x_1^{(N)+}]$ , the value of the contour integral does not depend on  $y > 0$ , and therefore the limit can be dropped, namely

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

Using the equality  $(1 + \sigma^2 c m_N(z)) (\mathbf{B}_N \mathbf{B}_N^H - w_N(z) \mathbf{I}_M)^{-1} = \mathbf{T}_N(z)$ , which follows easily from the definition in (14), we can write

$$g_N(z) = \mathbf{b}_N^H \mathbf{T}_N(z) \mathbf{b}_N \frac{w'_N(z)}{1 + \sigma^2 c m_N(z)}.$$

Now, the key point of the proof is based on the observation that  $g_N(z)$  can be estimated consistently from the elements of matrix  $\hat{\mathbf{R}}_N$ . We recall that  $\hat{m}_N(z)$  is defined by

$$\hat{m}_N(z) = \frac{1}{M} \operatorname{Tr} [\mathbf{Q}_N(z)] = \frac{1}{M} \sum_{k=1}^M \frac{1}{\hat{\lambda}_k - z} \quad (63)$$

and we define  $\hat{w}_N(z)$  as the function obtained by replacing function  $m_N(z)$  with  $\hat{m}_N(z)$  in the definition of  $w_N(z)$ , i.e.

$$\hat{w}_N(z) = z \left(1 + \sigma^2 c_N \hat{m}_N(z)\right)^2 - \sigma^2 (1 - c_N) \left(1 + \sigma^2 c_N \hat{m}_N(z)\right) \quad (64)$$

We define the corresponding random asymptotic equivalent of  $g_N(z)$  by

$$\hat{g}_N(z) = \mathbf{b}_N^H \mathbf{Q}_N(z) \mathbf{b}_N \frac{\hat{w}'_N(z)}{1 + \sigma^2 c_N \hat{m}_N(z)}.$$

Observe from the definition of  $\hat{m}_N$  and of  $\mathbf{Q}_N$  that the function  $\hat{g}_N$  is meromorphic with poles at  $\hat{\lambda}_1, \dots, \hat{\lambda}_M$  and at  $\hat{\omega}_1, \dots, \hat{\omega}_M$ , the  $M$  real-valued solutions to the polynomial equation (of degree  $M$ )  $1 + \sigma^2 c_N \hat{m}_N(x) = 0$ . In the following, it is important to locate the  $(\hat{\omega}_l)_{l=1, \dots, M}$ .

**Lemma 4.** *For  $N$  large enough, with probability one*

$$\hat{\lambda}_1, \dots, \hat{\lambda}_{M-K}, \hat{\omega}_1, \dots, \hat{\omega}_{M-K} \in ]t_1^-, t_1^+ [ \quad (65)$$

$$\hat{\lambda}_{M-K+1}, \dots, \hat{\lambda}_M, \hat{\omega}_{M-K+1}, \dots, \hat{\omega}_M \text{ are greater than } t_2^- \quad (66)$$

Theorem 1 implies that almost surely,  $g_N(z) - \hat{g}_N(z) \rightarrow 0$  on  $\partial\mathcal{R}_y \setminus \{t_1^-, t_1^+\}$ . In order to be able to use the dominated convergence theorem, we first state the following inequalities proven in Appendix H: there exists  $N_0 \in \mathbb{N}$  such that

$$\sup_{N \geq N_0} \sup_{z \in \partial\mathcal{R}_y} |g_N(z)| < +\infty \quad (67)$$

and

$$\sup_{N \geq N_0} \sup_{z \in \partial\mathcal{R}_y} |\hat{g}_N(z)| < +\infty \quad (68)$$

almost surely. The dominated convergence theorem thus implies that

$$\left| \frac{1}{2\pi i} \oint_{\partial\mathcal{R}_y^-} g_N(z) - \hat{g}_N(z) dz \right| \longrightarrow 0 \quad \text{a.s.}$$

We now establish that the integral

$$\hat{\eta}_N^{new} = \frac{1}{2\pi i} \oint_{\partial\mathcal{R}_y^-} \hat{g}_N(z) dz$$

is equal to  $\hat{\eta}_N^{new}$  defined by (55). This can be shown using residue Theorem.

Lemma 4 implies that for  $N$  large enough

$$\hat{\eta}_N^{new} = \sum_{k=1}^{M-K} \left[ \text{Ind}_{\partial\mathcal{R}_y^-}(\hat{\lambda}_k) \text{Res}(\hat{g}_N, \hat{\lambda}_k) + \text{Ind}_{\partial\mathcal{R}_y^-}(\hat{\omega}_k) \text{Res}(\hat{g}_N, \hat{\omega}_k) \right]$$

where  $\text{Res}(\hat{g}_N, \lambda)$  denotes the residue of function  $\hat{g}_N$  at point  $\lambda$ .

In order to evaluate these residues, we first remark that

$$\mathbf{b}_N^H \mathbf{Q}_N(z) \mathbf{b}_N = \sum_{k=1}^M \frac{\mathbf{b}_N^H \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^H \mathbf{b}_N}{\hat{\lambda}_k - z}$$

$\hat{g}_N(z)$  can thus be written as

$$\hat{g}_N(z) = \sum_{k=1}^M \mathbf{b}_N^H \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^H \mathbf{b}_N \left[ \hat{\alpha}_k(z) + \hat{\beta}_k(z) + \hat{\gamma}_k(z) \right]$$

where we have defined

$$\hat{\alpha}_k(z) = \frac{1 + \sigma^2 c_N \hat{m}_N(z)}{\hat{\lambda}_k - z} \quad (69)$$

$$\hat{\beta}_k(z) = \frac{2\sigma^2 c_N z \hat{m}'_N(z)}{\hat{\lambda}_k - z} \quad (70)$$

$$\hat{\gamma}_k(z) = -\sigma^4 c_N (1 - c_N) \frac{\hat{m}'_N(z)}{(\hat{\lambda}_k - z)(1 + \sigma^2 c_N \hat{m}_N(z))} \quad (71)$$

and consequently with probability one for  $N$  large enough

$$\hat{\eta}_N^{new} = - \sum_{k=1}^M \mathbf{b}_N^H \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^H \mathbf{b}_N \sum_{m=1}^{M-K} \left[ \text{Res}(\hat{\alpha}_k, \hat{\lambda}_m) + \text{Res}(\hat{\beta}_k, \hat{\lambda}_m) + \text{Res}(\hat{\gamma}_k, \hat{\lambda}_m) + \text{Res}(\hat{\gamma}_k, \hat{\omega}_m) \right].$$

Classical residue calculus gives

$$\text{Res}(\hat{\alpha}_k, \hat{\lambda}_m) = \begin{cases} -\frac{\sigma^2 c_N}{M} \frac{1}{\hat{\lambda}_k - \hat{\lambda}_m} & k \neq m \\ -\left(1 + \sigma^2 c_N \frac{1}{M} \sum_{\substack{i=1 \\ i \neq k}}^M \frac{1}{\hat{\lambda}_i - \hat{\lambda}_k}\right) & k = m \end{cases} \quad (72)$$

$$\text{Res}(\hat{\beta}_k, \hat{\lambda}_m) = \begin{cases} \frac{2\sigma^2 c_N}{M} \frac{\hat{\lambda}_k}{(\hat{\lambda}_k - \hat{\lambda}_m)^2} & k \neq m \\ -\frac{2\sigma^2 c_N}{M} \sum_{\substack{i=1 \\ i \neq k}}^M \frac{\hat{\lambda}_k}{(\hat{\lambda}_i - \hat{\lambda}_k)^2} & k = m \end{cases} \quad (73)$$

$$\text{Res}(\hat{\gamma}_k, \hat{\lambda}_m) = \begin{cases} \sigma^2 (1 - c_N) \frac{1}{\hat{\lambda}_k - \hat{\lambda}_m} & k \neq m \\ -M \frac{1 - c_N}{c_N} \left(1 + \frac{\sigma^2 c_N}{M} \sum_{\substack{i=1 \\ i \neq k}}^M \frac{1}{\hat{\lambda}_i - \hat{\lambda}_k}\right) & k = m \end{cases} \quad (74)$$

$$\text{Res}(\hat{\gamma}_k, \hat{\omega}_m) = -\sigma^2 \frac{1 - c_N}{\hat{\lambda}_k - \hat{\omega}_m}. \quad (75)$$

Next, we define  $\hat{\xi}_k$  as

$$\hat{\xi}_k = - \sum_{m=1}^{M-K} \text{Res}(\hat{\alpha}_k, \hat{\lambda}_m) + \text{Res}(\hat{\beta}_k, \hat{\lambda}_m) + \text{Res}(\hat{\gamma}_k, \hat{\lambda}_m) + \text{Res}(\hat{\gamma}_k, \hat{\omega}_m).$$

We obtain, for  $k = 1, \dots, M - K$

$$\hat{\xi}_k = 1 - \frac{\sigma^2 c_N}{M} \sum_{i=M-K+1}^M \frac{1}{\hat{\lambda}_k - \hat{\lambda}_i} + \frac{2\sigma^2 c_N}{M} \sum_{i=M-K+1}^M \frac{\hat{\lambda}_k}{(\hat{\lambda}_k - \hat{\lambda}_i)^2} + M \frac{1 - c_N}{c_N} \quad (76)$$

$$+ \sigma^2(1 - c_N) \left( \sum_{\substack{i=1 \\ i \neq k}}^{M-K} \frac{1}{\hat{\lambda}_i - \hat{\lambda}_k} - \sum_{\substack{i=1 \\ i \neq k}}^{M-K} \frac{1}{\hat{\omega}_i - \hat{\lambda}_k} + \sum_{\substack{i=1 \\ i \neq k}}^M \frac{1}{\hat{\lambda}_i - \hat{\lambda}_k} \right) \quad (77)$$

and for  $k = M - K + 1, \dots, M$

$$\hat{\xi}_k = -\frac{\sigma^2 c_N}{M} \sum_{i=1}^{M-K} \frac{1}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{2\sigma^2 c_N}{M} \sum_{i=1}^{M-K} \frac{\hat{\lambda}_k}{(\hat{\lambda}_k - \hat{\lambda}_i)^2} \quad (78)$$

$$+ \sigma^2(1 - c_N) \sum_{i=1}^{M-K} \frac{\hat{\omega}_i - \hat{\lambda}_i}{(\hat{\lambda}_i - \hat{\lambda}_k)(\hat{\omega}_i - \hat{\lambda}_k)}. \quad (79)$$

To retrieve the final form of  $\hat{\xi}_k$  given in the statement of the theorem, we notice that

$$1 + \sigma^2 c_N \frac{1}{M} \sum_{i=1}^M \frac{1}{\hat{\lambda}_i - \hat{\omega}_k} = 0$$

and use the following lemma proved in Appendix I:

**Lemma 5.** *The following identity holds for any  $k = 1 \dots M$*

$$\frac{1}{M} \sum_{\substack{i=1 \\ i \neq k}}^M \frac{1}{\hat{\lambda}_i - \hat{\omega}_k} = \frac{2}{M} \sum_{\substack{i=1 \\ i \neq k}}^M \frac{1}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{1}{M} \sum_{\substack{i=1 \\ i \neq k}}^M \frac{1}{\hat{\omega}_i - \hat{\lambda}_k}$$

This establishes that  $\hat{\eta}_{new} = \hat{\eta}_{new}$  and completes the proof of Theorem 5.

## VII. NUMERICAL RESULTS

In this section, we compare the results provided by the traditional subspace estimate, the new estimate (55) (referred to in the figure as the "conditional estimator"), and the improved estimate of [7] derived under the assumption that the source signals are i.i.d. sequences (referred to as the "unconditional estimator").

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

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \left[ 1, e^{i\pi \sin(\theta)}, \dots, e^{i(M-1)\pi \sin(\theta)} \right]^T \quad (80)$$

In the following numerical experiments, source signals are realizations of mutually independent unit variance AR(1) sequences with correlation coefficient 0.9. In order to evaluate the performance of the

various estimators, we use Monte Carlo simulations. The additive noise varies from trials to trials, but, for fixed  $M$  and  $N$ , matrix  $\mathbf{S}$  remains unchanged. Finally, unless otherwise stated, the cluster associated to the eigenvalue 0 of matrix  $\mathbf{AS}$  is assumed to be separated from the clusters corresponding to its non zero eigenvalues, i.e. for each  $\sigma^2$ ,  $M$  and  $N$ , it holds that

$$0 < w_1^{(N)+} < w_2^{(N)-} < \gamma_{M-K+1}^{(N)} \quad (81)$$

We finally mention that the estimate of [7] is supposed to be inconsistent 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 Remark 5).

**Experiment 1:** We first consider two closely spaced sources, i.e.  $\theta_1 = 16^\circ$  and  $\theta_2 = 18^\circ$ . The number of antennas is  $M = 20$  and the number of snapshots is  $N = 40$ . The separation condition (81) 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  $\theta$  in figure 6 the ratio of the MSE of the traditional estimator of  $\mathbf{a}(\theta)^H \mathbf{\Pi} \mathbf{a}(\theta)$  over the MSE of the improved estimator. The SNR is equal to 16 dB. Figure 6 shows that the 2 improved estimates have nearly the same performance, and that they outperform significantly the traditional approach around the 2 angles. We however notice that the 3 estimates have nearly the same performance if  $\theta$  is far away from  $\theta_1 = 16^\circ$  and  $\theta_2 = 18^\circ$ . In order to evaluate more precisely the improvements provided by the conditional and the unconditional estimators around  $\theta_1$  and  $\theta_2$ , we plot vs SNR in figure 7 the mean of the MSEs of the estimates of  $\mathbf{a}(\theta_1)^H \mathbf{\Pi} \mathbf{a}(\theta_1)$  and  $\mathbf{a}(\theta_2)^H \mathbf{\Pi} \mathbf{a}(\theta_2)$ .

In figure 8, we plot for each method the mean of the MSE of the two estimated angles versus the SNR. The estimates of  $\theta_1$  and  $\theta_2$  are defined as 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 2 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, i.e. 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 9, 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 2 improved estimators provide a gain of 8 dB over the traditional

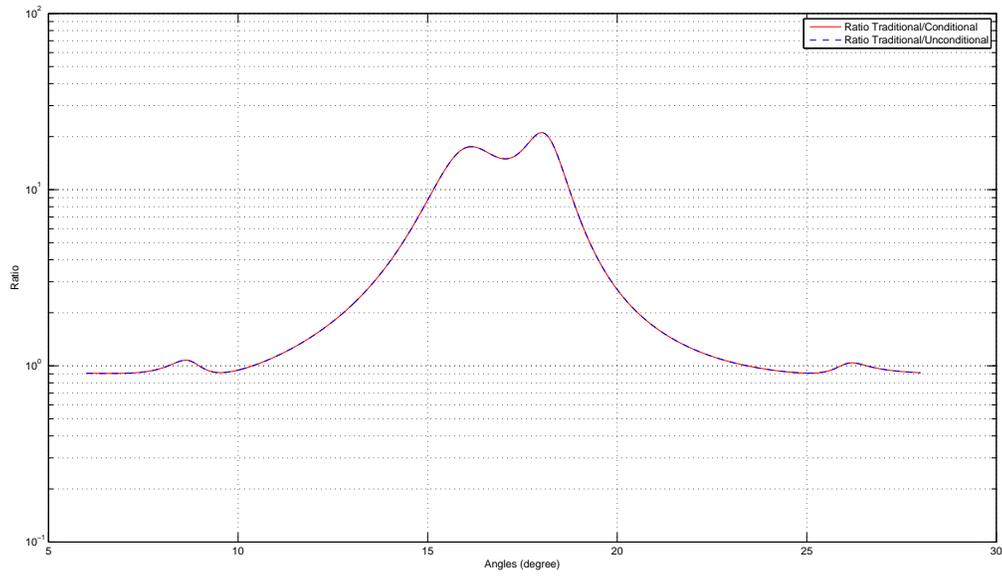


Figure 6. Ratio (in dB) of the MSE of the traditional estimate of  $\mathbf{a}(\theta)^H \mathbf{\Pi} \mathbf{a}(\theta)$  over the improved estimates vs angles.

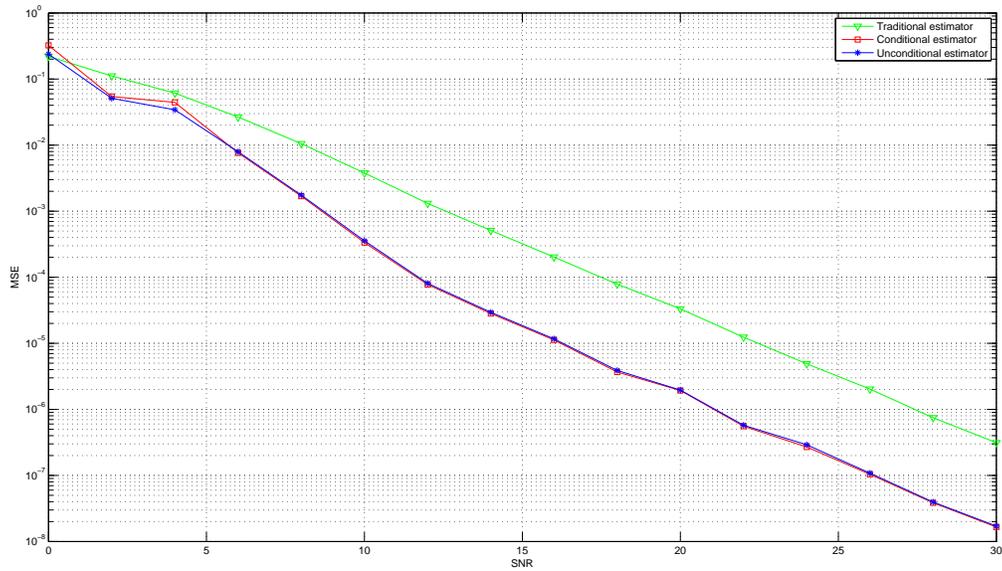


Figure 7. Mean of the MSE of the estimates of  $\mathbf{a}(\theta_1)^H \mathbf{\Pi} \mathbf{a}(\theta_1)$  and  $\mathbf{a}(\theta_2)^H \mathbf{\Pi} \mathbf{a}(\theta_2)$ .

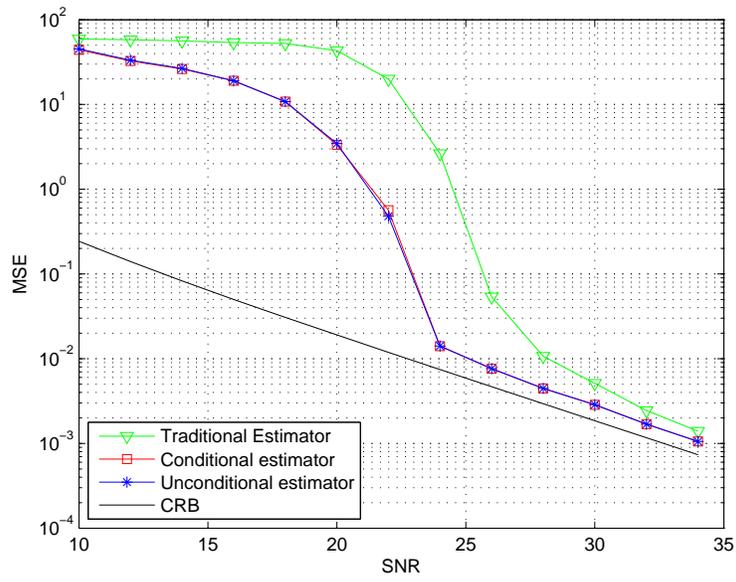


Figure 8. Mean of the MSE of the angles estimates versus SNR

estimate.

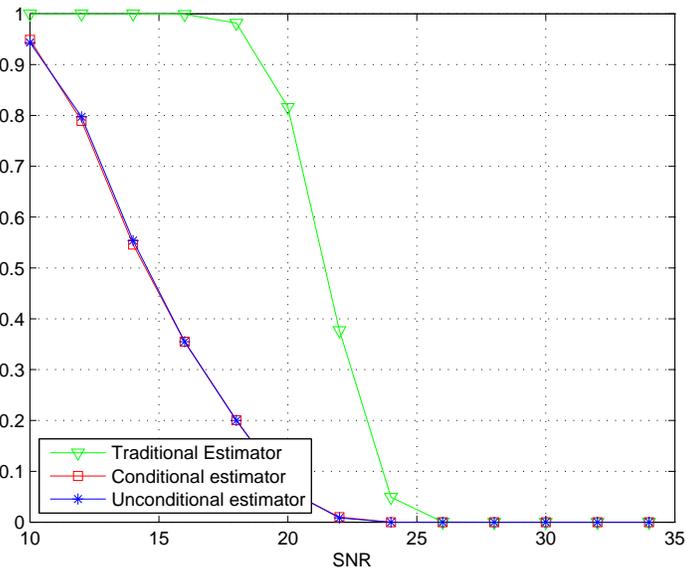


Figure 9. Outlier Probability vs the SNR

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  $\text{SNR} = 15$  dB. In figure 10 we have plotted the mean of the MSEs on the estimates of  $\mathbf{a}(\theta_i)^H \mathbf{\Pi} \mathbf{a}(\theta_i)$  for  $i = 1, 2$ . The separation condition (81) occurs for  $N \geq 32$ . Figure 10 illustrates clearly the inconsistency of that the traditional estimate.

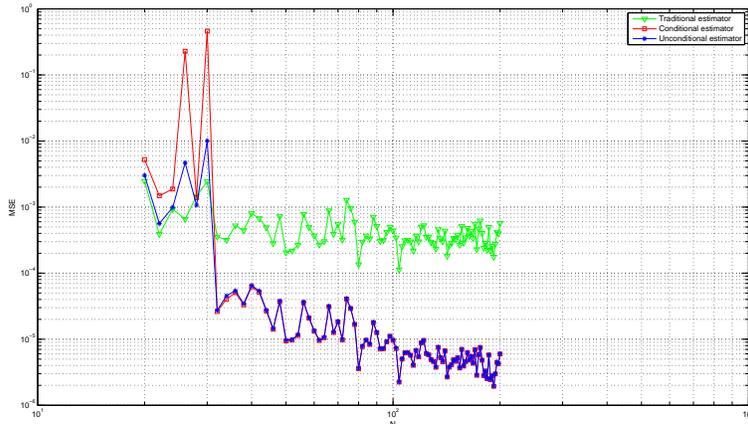


Figure 10. MSE for the estimators of the localization function vs  $N$

**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  $(\theta_i)_{i=1,\dots,10}$  are equal to  $\theta_i = -40^\circ + (i-1)10^\circ$  for  $i = 1, \dots, 10$ . The separation condition holds if  $\text{SNR}$  is greater than 15 dB. We again plot versus  $\theta$  in figure 11 the ratio of the MSE of the traditional estimator of the localization function over the MSE of its conditional and unconditional estimators.  $\text{SNR}$  is equal to 16 dB. Figure 11 shows again that the performance improvement of the conditional and unconditional estimates is optimum around the angles  $(\theta_i)_{i=1,\dots,10}$ .

Figure 12 represents the mean of the MSEs of the various estimates of  $\mathbf{a}(\theta_i)^H \mathbf{\Pi} \mathbf{a}(\theta_i)$  for  $i = 1, \dots, 10$  w.r.t. the  $\text{SNR}$ , and confirms the superiority of the 2 improved estimates when the separation condition (81). We note that

**Remark 5.** *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 [7] under the assumption that matrix  $\mathbf{S}_N$  is a Gaussian matrix with unit variance i.i.d. entries. The unconditional estimator of [7] is based on the observation that if  $\mathbf{S}_N$  is an i.i.d. Gaussian matrix, then the entries of  $(\hat{\mathbf{R}}_N - z\mathbf{I})^{-1}$  have the same*

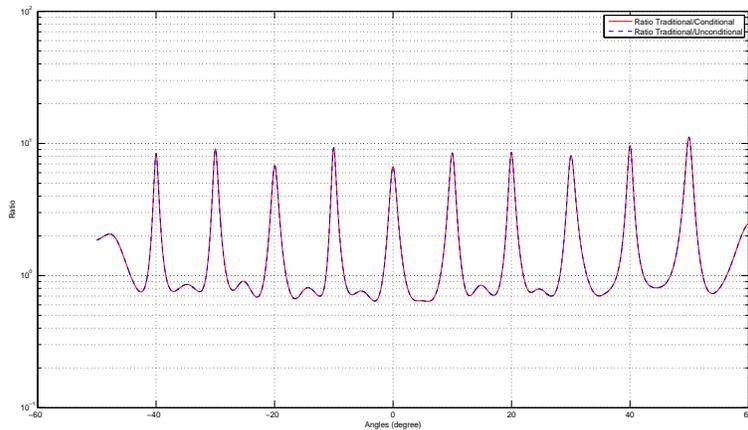


Figure 11. Ratio (in dB) of the MSE of the traditional estimate of the localization function over the MSE of its improved estimates versus  $\theta$

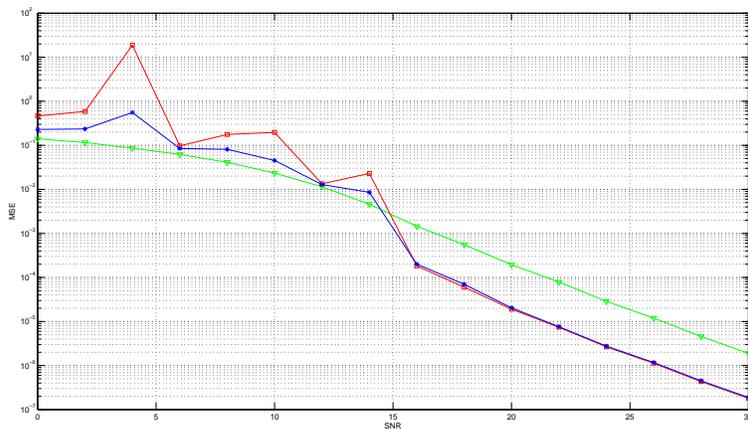


Figure 12. Mean of the MSE of the estimates of  $\mathbf{a}(\theta_i)^H \mathbf{\Pi} \mathbf{a}(\theta_i)$  for  $i = 1, \dots, 10$  versus SNR

behaviour than the entries of matrix  $\mathbf{T}_{N,iid}(z)$  defined by the following equation

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

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

One can verify that the entries of  $\mathbf{T}_N(z)$  defined by (14), which depend on  $\mathbf{S}_N$ , have the same asymptotic behaviour than the entries of  $\mathbf{T}_{N,iid}(z)$  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  $(\hat{\mathbf{R}}_N - z\mathbf{I})^{-1}$  do not behave like the entries of  $\mathbf{T}_{N,iid}(z)$  so that the unconditional estimator should become inconsistent. The previous simulation results tend to indicate that it is not the case. The explanation of this phenomenon is a topic for further researchs.

## VIII. CONCLUSIONS

This paper has considered the use of subspace estimation algorithms in situations where the number of available samples and the observation dimension are comparable in magnitude. We have considered the information plus noise signal model, according to which the received signals are deterministic unknowns whose empirical spatial correlation matrix is low-rank. We have derived an estimator of the noise subspace of the spatial correlation matrix that is consistent, not only when the number of samples tends to infinity for a fixed observation dimension, but also when these two quantities increase to infinity at the same rate. This guarantees that the estimator will present a good performance even when these two quantities are comparable in magnitude. In order to establish the consistency of the estimator, we have proven new results concerning the almost sure location of the eigenvalues of the sample covariance matrix of an Information plus Noise Gaussian model.

## APPENDIX A

### PROOF OF PROPERTY 1 OF PROPOSITION 1

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

$$h(m, z) = \frac{1}{M} \text{Tr} \left[ -z(1 + \sigma^2 c_N m) \mathbf{I}_M + \sigma^2(1 - c_N) \mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^H}{1 + \sigma^2 c_N m} \right]^{-1}. \quad (82)$$

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

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

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

$$1 - \frac{\partial h}{\partial m} \Big|_{(m_*, 0)} > 0. \quad (83)$$

Indeed, observe that

$$\frac{\partial h}{\partial m} \Big|_{(m_*, 0)} = \frac{\sigma^2 c_N}{1 + \sigma^2 c_N m_*} \frac{1}{M} \text{Tr} \left[ \frac{\mathbf{B}_N \mathbf{B}_N^H}{1 + \sigma^2 c_N m_*} \left( \sigma^2(1 - c_N) \mathbf{I}_M + \frac{\mathbf{B}_N \mathbf{B}_N^H}{1 + \sigma^2 c_N m_*} \right)^{-2} \right]$$

so that

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

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

$$\overline{m}(z) = h(\overline{m}(z), z) \quad (84)$$

for  $|z| < \eta$ . Evaluating the successive derivatives of function  $z \rightarrow h(\overline{m}(z), z)$  at the origin, one can check that for each  $l \geq 0$ ,  $\overline{m}^{(l)}(0)$  is real-valued. Since  $m_* > 0$ , there exists a positive quantity  $\epsilon$ ,  $0 < \epsilon \leq \eta$  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  $D(0, \epsilon) \setminus \{[0, \epsilon]\}$  and coincide on a set of values with an accumulation point, they must coincide on the whole domain of analyticity, namely  $D(0, \epsilon) \setminus \{[0, \epsilon]\}$ . We recall that for  $0 \leq x < \epsilon$ ,  $\mu_N([0, x])$  can be expressed as

$$\mu_N([0, x]) = \frac{1}{\pi} \lim_{y \rightarrow 0, y > 0} \int_0^x \text{Im}(m_N(s + iy)) ds$$

Therefore,

$$\mu_N([0, x]) = \frac{1}{\pi} \lim_{y \rightarrow 0, y > 0} \int_0^x \text{Im}(\overline{m}(s + iy)) ds$$

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

$$\frac{1}{\pi} \lim_{y \rightarrow 0, y > 0} \int_0^x \text{Im}(\overline{m}(s + iy)) ds = \frac{1}{\pi} \int_0^x \text{Im}(\overline{m}(s)) ds = 0$$

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

## APPENDIX B

### PROOF OF PROPOSITION 2

In order to prove Property 1, we establish that  $\text{Im}(w_N(x)) > 0$  if and only if  $\text{Im}(m_N(x)) > 0$ . Assume that  $\text{Im}(m_N(x)) > 0$ , i.e. that  $x \in \text{Int}(\mathcal{S}_N)$ , which in particular implies that  $x > 0$ , and consider  $z = x + iy$  with  $y > 0$ . Equation (13) can be written in terms of  $w_N(z)$  as

$$\frac{m_N(z)}{1 + c_N \sigma^2 m_N(z)} = f_N(w_N(z)). \quad (85)$$

Taking the imaginary part from both sides yields the identity

$$\frac{\operatorname{Im}(m_N(z))}{|1 + \sigma^2 c_N m_N(z)|^2} = \operatorname{Im}(w_N(z)) \frac{1}{M} \operatorname{Tr} [(\mathbf{B}_N \mathbf{B}_N^H - w_N(z) \mathbf{I}_M)^{-1} (\mathbf{B}_N \mathbf{B}_N^H - w_N^*(z) \mathbf{I}_M)^{-1}]$$

or equivalently,

$$\operatorname{Im}(m_N(z)) = \operatorname{Im}(w_N(z)) |1 + \sigma^2 c_N m_N(z)|^2 \frac{1}{M} \operatorname{Tr} [(\mathbf{B}_N \mathbf{B}_N^H - w_N(z))^{-1} (\mathbf{B}_N \mathbf{B}_N^H - w_N^*(z))^{-1}] \quad (86)$$

$$= \operatorname{Im}(w_N(z)) \frac{1}{M} \operatorname{Tr} [\mathbf{T}_N(z) \mathbf{T}_N^H(z)] \quad (87)$$

It is shown in [17] (see Eq. (2.6)) that

$$\frac{\sigma^2}{N} \operatorname{Tr} [\mathbf{T}_N(z) \mathbf{T}_N^H(z)] \leq \frac{1}{|z|} \leq \frac{1}{x}$$

which implies

$$\operatorname{Im}(m_N(z)) \leq \operatorname{Im}(w_N(z)) \frac{1}{\sigma^2 c_N |x|}. \quad (88)$$

If  $y \rightarrow 0$ , we get that

$$0 < \operatorname{Im}(m_N(x)) \leq \operatorname{Im}(w_N(x)) \frac{1}{\sigma^2 c_N |x|}$$

which implies that  $\operatorname{Im}(w_N(x)) > 0$ . Conversely, assume that  $\operatorname{Im}(w_N(x)) > 0$ . Then,  $m_N(x)$  cannot be real-valued, otherwise,  $w_N(x) = x(1 + \sigma^2 c_N m_N(x))^2 - \sigma^2(1 - c_N)(1 + \sigma^2 c_N m_N(x))$  would be also real-valued.

Next, we prove Property 2. Since  $x \rightarrow m_N(x)$  is differentiable on  $\mathbb{R} - \partial \mathcal{S}_N$ ,  $x \rightarrow w_N(x)$  is differentiable on the same subset. By Property 4 of Proposition 1,  $w_N(x)$  does not belong to the spectrum of matrix  $\mathbf{B}_N \mathbf{B}_N^H$  if  $x \in \mathbb{R} \setminus \mathcal{S}_N$ . Therefore, the function  $x \rightarrow f_N(w_N(x))$  is differentiable for  $x \in \mathbb{R} \setminus \mathcal{S}_N$ . Since (85) holds on  $x \in \mathbb{R} \setminus \mathcal{S}_N$ , we can differentiate it with respect to  $x$  on  $x \in \mathbb{R} \setminus \mathcal{S}_N$ . This gives

$$w'_N(x) f'_N(w_N(x)) = \frac{m'_N(x)}{(1 + c_N \sigma^2 m_N(x))^2}$$

for  $x \in \mathbb{R} \setminus \mathcal{S}_N$ . Now, observe that  $m'_N(x) > 0$  on  $\mathbb{R} \setminus \mathcal{S}_N$  because  $m_N(z)$  is the Stieltjès transform of a probability measure carried by  $\mathcal{S}_N$ . On the other hand, the function  $f'_N$  is of course strictly positive on  $\mathbb{R}$ . This in turn shows that  $w'_N(x) > 0$  on  $x \in \mathbb{R} \setminus \mathcal{S}_N$ .

To establish the last property, we use (13) at point  $x \in \mathbb{R} \setminus \mathcal{S}_N$ , and get that

$$1 - c_N \sigma^2 f_N(w(x)) = \frac{1}{1 + c_N \sigma^2 m_N(x)}. \quad (89)$$

The conclusion follows from the inequality  $1 + c_N \sigma^2 m_N(x) > 0$  if  $x \in \mathbb{R} \setminus \mathcal{S}_N$  (see Proposition 1).

## APPENDIX C

## PROOF OF (28) IN PROPOSITION 3

We consider  $w_1, w_2 \in \{w_1^{(N)-}, w_1^{(N)+}, \dots, w_Q^{(N)-}, w_Q^{(N)+}\}$ , and denote by  $\phi_1$  and  $\phi_2$  the quantities  $\phi_N(w_1)$  and  $\phi_N(w_2)$  respectively. We define  $h_n = 1 - \sigma^2 c_N f_N(w_n)$  so that we can write  $\phi_n = w_n h_n^2 + \sigma^2(1 - c_N)h_n$ ,  $n \in \{1, 2\}$ . Our objective is to show that the quantity  $(\phi_2 - \phi_1)/(w_2 - w_1)$  is always positive. Note that, by definition,  $w_1$  and  $w_2$  are inflexion points of  $\phi_N(w)$  such that  $h_1 \geq 0$  and  $h_2 \geq 0$ .

Using direct subtraction of the expressions of  $\phi_1$  and  $\phi_2$  we can write

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

Consider now the following inequality

$$\frac{2}{M} \sum_{k=1}^M \frac{\gamma_k^{(N)}}{(\gamma_k^{(N)} - w_1)(\gamma_k^{(N)} - w_2)} \leq \frac{1}{M} \sum_{k=1}^M \frac{\gamma_k^{(N)}}{(\gamma_k^{(N)} - w_1)^2} + \frac{1}{M} \sum_{k=1}^M \frac{\gamma_k^{(N)}}{(\gamma_k^{(N)} - w_2)^2} \quad (90)$$

which can be readily obtained by noting that

$$\frac{1}{M} \sum_{k=1}^M \left( \frac{(\gamma_k^{(N)})^{1/2}}{(\gamma_k^{(N)} - w_1)} - \frac{(\gamma_k^{(N)})^{1/2}}{(\gamma_k^{(N)} - w_2)} \right)^2 \geq 0.$$

Using the definition of  $h_1$  and  $h_2$  we can readily write

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

and hence the inequality in (90) is giving us

$$\begin{aligned} \frac{\phi_2 - \phi_1}{w_2 - w_1} \geq (h_1 + h_2) \left[ 1 - \frac{\sigma^2 c_N}{2} (f_N(w_1) + f_N(w_2) + w_1 f'_N(w_1) + w_2 f'_N(w_2)) \right] + \\ - h_1 h_2 + \sigma^2(1 - c_N) \frac{h_2 - h_1}{w_2 - w_1} \quad (91) \end{aligned}$$

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

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

By inserting this last equality into (91) and replacing  $f_N(w_1)$  with  $\sigma^{-2}(1 - h_1)$ , we obtain the expression

$$\begin{aligned} \frac{\phi_2 - \phi_1}{w_2 - w_1} \geq \frac{\sigma^4 c_N (1 - c_N)}{2} \frac{1}{M} \sum_{k=1}^M \frac{(w_2 - w_1)^2}{(\gamma_k^{(N)} - w_1)^2 (\gamma_k^{(N)} - w_2)^2} + \frac{h_1^2 + h_2^2}{2} + \\ - \frac{\sigma^4 c_N (1 - c_N)}{2} [f'_N(w_1) + f'_N(w_2)] - \frac{\sigma^2}{2} \frac{h_1 + h_2}{(w_1 f'_N(w_1) + w_2 f'_N(w_2))}. \quad (92) \end{aligned}$$

Now, both  $w_1$  and  $w_2$  are preimages of local extrema of  $\phi_N$ , 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

$$\frac{h_1^2 + h_2^2}{2} = \sigma^2 [w_1 h_1 f'_N(w_1) + w_2 h_2 f'_N(w_2)] + \frac{\sigma^4 c_N (1 - c_N)}{2} [f'_N(w_1) + f'_N(w_2)]$$

and by inserting the last equality into (92), we obtain

$$\frac{\phi_2 - \phi_1}{w_2 - w_1} \geq \frac{\sigma^4 c_N (1 - c_N)}{2} \frac{1}{M} \sum_{k=1}^M \frac{(w_2 - w_1)^2}{(\gamma_k^{(N)} - w_1)^2 (\gamma_k^{(N)} - w_2)^2} + \frac{\sigma^2}{2} (h_1 - h_2) (w_1 f'_N(w_1) - w_2 f'_N(w_2)). \quad (93)$$

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 (93) becomes

$$\begin{aligned} \frac{\phi_2 - \phi_1}{w_2 - w_1} \geq & \frac{\sigma^4 c_N (1 - c_N)}{2} \frac{1}{M} \sum_{k=1}^M \frac{(w_2 - w_1)^2}{(\gamma_k^{(N)} - w_1)^2 (\gamma_k^{(N)} - w_2)^2} + \frac{(h_1 - h_2)^2}{4} + \\ & - \frac{\sigma^4 (1 - c_N)}{4} (f'_N(w_1) - f'_N(w_2)) + \frac{\sigma^4 (1 - c_N)}{4} c_N \left[ \frac{h_1}{h_2} f'_N(w_2) + \frac{h_2}{h_1} f'_N(w_1) \right] \end{aligned} \quad (94)$$

Clearly, we have

$$\frac{1}{M} \sum_{k=1}^M \frac{(w_2 - w_1)^2}{(\gamma_k^{(N)} - w_1)^2 (\gamma_k^{(N)} - w_2)^2} - [f'_N(w_1) + f'_N(w_2)] = -\frac{2}{M} \sum_{k=1}^M \frac{1}{(\gamma_k^{(N)} - w_1)(\gamma_k^{(N)} - w_2)}$$

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

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

The left hand side of the previous equality appears in (94) as a common factor on the last two terms of the right hand side of that equation. Hence, plugging it into (94), we obtain

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

Finally, noting that all the terms of the above equation are non-negative, we have established (28).

APPENDIX D  
PROOF OF LEMMA 3

The proof of this Lemma is a direct consequence of [17, Section 4]. Next, we provide some details on how to obtain (45); the same procedure can be applied in order to obtain (46). As in [17], we define in this section function  $b_N(z)$  by  $b_N(z) = 1 + \sigma^2 c_N m_N(z)$  for  $z \in \mathbb{C}$ , and denote by  $b_1^-$  the quantity  $b_N(x_1^-)$  (note that we drop the dependence on  $N$  in  $x_1^-$ ). Since  $x_1^-$  belongs to  $\partial\mathcal{S}_N$ , both  $m_N(x_1^-)$  and  $b_1^-$  are real-valued. Proposition 1 thus implies that  $z \rightarrow b_N(z)$  is continuous at the point  $x_1^-$ . Similarly,  $w_N(x_1^-) = w_1^-$  is real-valued so that the function  $z \rightarrow w_N(z)$  is also continuous at  $x_1^-$ .

Since  $f'_N(w_1^-) > 0$ , there exists a neighborhood  $\mathcal{V}(w_1^-)$  of  $w_1^-$  on which  $f_N$  is biholomorphic. For  $z \in \mathbb{C}_+ \cup \mathbb{R}$ , it follows from (21) that we can write

$$f_N(w_N(z)) = \frac{m_N(z)}{1 + \sigma^2 c_N m_N(z)} = \frac{1}{\sigma^2 c_N} \left( 1 - \frac{1}{b_N(z)} \right). \quad (95)$$

Since  $w_N$  is continuous at  $x_1^-$  and since  $w_N(z) \in \mathbb{C}_+$  if  $z \in \mathbb{C}_+$  (see Property 5 of Proposition 1), there exists a neighborhood  $\mathcal{V}(x_1^-)$  of  $x_1^-$  such that

$$w_N(\mathcal{V}(x_1^-) \cap \mathbb{C}_+) \subset \mathcal{V}(w_1^-) \cap \mathbb{C}_+.$$

Therefore, applying the holomorphic inverse of  $f_N$ , denoted as  $f_N^{-1}$ , to both sides of (95) we get, for any  $z \in \mathcal{V}(x_1^-) \cap \mathbb{C}_+$ ,

$$w_N(z) = f_N^{-1} \left( \frac{1}{\sigma^2 c_N} \left( 1 - \frac{1}{b_N(z)} \right) \right).$$

Using the fact that  $w_N(z) = z b_N^2(z) - \sigma^2(1 - c_N)b_N(z)$  and solving with respect to  $z$ , we get that

$$z = Z_N(b_N(z)) \quad z \in \mathcal{V}(x_1^-) \cap \mathbb{C}_+ \quad (96)$$

where  $Z_N$  is the function defined in an appropriate neighborhood of  $b_1^-$  by

$$Z_N(b) = \frac{1}{b^2} f_N^{-1} \left( \frac{1}{\sigma^2 c_N} \left( 1 - \frac{1}{b} \right) \right) + \frac{\sigma^2(1 - c_N)}{b}.$$

Next, we recall the following result from [17].

**Lemma 6.** *There exists a neighborhood  $\mathcal{V}(b_1^-)$  of  $b_1^-$  and a function  $\Psi_N$ , biholomorphic from  $\mathcal{V}(b_1^-)$  onto a neighborhood of the origin  $\mathcal{V}(0)$  such that  $\forall b \in \mathcal{V}(b_1^-)$*

$$Z_N(b) - x_1^- = \Psi_N^2(b).$$

Since the function  $b_N$  is continuous at the point  $x_1^-$ , and since  $b_N(z) \in \mathbb{C}_+$  if  $z \in \mathbb{C}_+$  (which follows from the definition of  $b_N$ ), there exist two smaller neighborhoods  $\mathcal{V}'(x_1^-) \subset \mathcal{V}(x_1^-)$  and  $\mathcal{V}'(b_1^-) \subset \mathcal{V}(b_1^-)$  of  $x_1^-$  and  $b_1^-$  respectively, such that

$$b_N(z) \in \mathcal{V}'(b_1^-) \cap \mathbb{C}_+ \quad \forall z \in \mathcal{V}'(x_1^-) \cap \mathbb{C}_+$$

Therefore, using (96), we can write

$$(\Psi_N(b_N(z)))^2 = z - x_1^-$$

$\forall z \in \mathcal{V}'(x_1^-) \cap \mathbb{C}_+$ . Let us now choose,  $\forall z \in \mathcal{V}'(x_1^-) \cap \mathbb{C}_+$ ,

$$\Psi_N(b_N(z)) = \sqrt{z - x_1^-}$$

where  $\sqrt{\cdot}$  represents any determination of the complex square root that is holomorphic<sup>3</sup> on  $\mathbb{C}_+$  and such that  $\sqrt{1} = 1$  (the following reasoning applies verbatim to the square root determination for which  $\sqrt{1} = -1$ ). We denote by  $\Psi_N^{-1}$  the holomorphic inverse function of  $\Psi_N$  defined on  $\mathcal{V}(0)$ . We have

$$b_N(z) = \Psi_N^{-1}\left(\sqrt{z - x_1^-}\right) \quad \forall z \in \mathcal{V}'(x_1^-) \cap \mathbb{C}_+$$

Taking derivatives with respect to  $z$  at both sides of the previous equality, we obtain

$$b'_N(z) = \frac{1}{2\sqrt{z - x_1^-}} [\Psi_N^{-1}]' \left( \sqrt{z - x_1^-} \right).$$

Now, since  $\Psi_N^{-1}$  is holomorphic on  $\mathcal{V}(0)$  by Lemma 6, the function  $[\Psi_N^{-1}]'$  will be bounded on the same neighborhood of 0 and thus we will have

$$|b'_N(z)| \leq \frac{C}{\left| \sqrt{z - x_1^-} \right|}$$

for some constant  $C$  independent of  $z$ . Therefore, for  $z = x + iy \in \mathcal{V}'(x_1^-) \cap \mathbb{C}_+$ , we can write

$$|w'_N(x + iy)| = |b_N(z)^2 + 2zb'_N(z) - \sigma^2(1 - c_N)b'_N(z)| \leq \frac{C}{\sqrt{|x - x_1^- + iy|}}. \quad (97)$$

The inequality

$$\frac{C}{\sqrt{|x - x_1^- + iy|}} \leq \frac{C}{\sqrt{|x - x_1^-|}}$$

for  $x \neq x_1^-$  completes the proof of (45) for  $y > 0$ . (45) for  $y = 0$  follows from the observation that  $w'_N(x) = \lim_{y \downarrow 0} w'_N(x + iy)$ .

<sup>3</sup>This property must hold for all possible choices of  $\Psi_N$  because, by definition,  $\Psi_N$  is holomorphic on  $\mathcal{V}(b_1^-)$  and  $b_N(z) \in \mathbb{C}_+$  if  $z \in \mathbb{C}_+$ . Since  $b_N(z)$  is holomorphic on  $\mathcal{V}'(x_1^-) \cap \mathbb{C}_+$ ,  $\Psi_N(b_N(z))$  must be holomorphic on the same set.

## APPENDIX E

## PROOF OF PROPOSITION 4

In this section, we drop as much as possible the subscript  $N$  for an easier reading. In the following,  $P_1(|z|)$  and  $P_2(\frac{1}{|\text{Im}(z)|})$  represent generic positive coefficients polynomials of the variables  $|z|$  and  $\frac{1}{|\text{Im}(z)|}$  whose mean feature is to be independent of  $N$ . The values of  $P_1$  and  $P_2$  can change from one line to another.

We rely extensively on the results of the Appendix II of [23] related to the properties of matrix  $(\mathbf{B} + \mathbf{D}^{1/2} \mathbf{W} \tilde{\mathbf{D}}^{1/2})(\mathbf{B} + \mathbf{D}^{1/2} \mathbf{W} \tilde{\mathbf{D}}^{1/2})^H$  where  $\mathbf{D}$  and  $\tilde{\mathbf{D}}$  are deterministic diagonal matrix. We thus use [23] in the case where  $\mathbf{D} = \sigma \mathbf{I}_M$  and  $\tilde{\mathbf{D}} = \sigma \mathbf{I}_N$  which corresponds to the context of the present paper. In order to help the reader, we use the same notations as in [23] all along this section. More precisely, we define

$$\delta(z) = \sigma c m(z) \quad (98)$$

$$\tilde{\delta}(z) = \delta(z) - \sigma \frac{1-c}{z} \quad (99)$$

$$\alpha(z) = \mathbb{E} \left[ \frac{\sigma}{N} \text{Tr} \mathbf{Q}(z) \right] \quad (100)$$

$$\tilde{\alpha}(z) = \alpha(z) - \sigma \frac{1-c}{z} \quad (101)$$

We remark that  $\alpha(z)$  is the Stieltjès transform of measure  $c\sigma\omega$  where  $\omega$  is the probability measure carried by  $\mathbb{R}_+$  defined by

$$\omega(\mathcal{B}) = \mathbb{E}(\hat{\mu}(\mathcal{B})) \quad (102)$$

for each Borel set  $\mathcal{B}$ . We recall that  $\hat{\mu}$  represents the empirical eigenvalue distribution of  $\hat{\mathbf{R}}_N = \Sigma_N \Sigma_N^H$ . Finally, it is easily seen that  $\tilde{\delta}$  is the Stieltjès transform of measure  $\sigma c \mu + \sigma(1-c)\delta_0$  ( $\delta_0$  represents the Dirac distribution at 0), and that  $\tilde{\alpha}(z)$ , which can be expressed by

$$\tilde{\alpha}(z) = \mathbb{E} \left[ \sigma \frac{1}{N} \text{Tr} \tilde{\mathbf{Q}}(z) \right] \quad (103)$$

where  $\tilde{\mathbf{Q}}(z)$  is defined by

$$\tilde{\mathbf{Q}}(z) = (\Sigma^H \Sigma - z \mathbf{I})^{-1} \quad (104)$$

coincides with the Stieltjès transform of measure  $\sigma c \omega + \sigma(1-c)\delta_0$ .

Matrix  $\mathbf{T}(z)$  defined by (14) can be written as

$$\mathbf{T}(z) = \left[ -z(1 + \sigma \tilde{\delta}(z)) \mathbf{I}_M + \frac{\mathbf{B} \mathbf{B}^H}{1 + \sigma \tilde{\delta}(z)} \right]^{-1}$$

and  $\delta(z)$  is equal to

$$\delta(z) = \sigma \frac{1}{N} \text{Tr} \mathbf{T}(z) \quad (105)$$

We also define matrix  $\tilde{\mathbf{T}}(z)$  by

$$\tilde{\mathbf{T}}(z) = \left[ -z(1 + \sigma\delta(z))\mathbf{I}_N + \frac{\mathbf{B}^H \mathbf{B}}{1 + \sigma\tilde{\delta}(z)} \right]^{-1} \quad (106)$$

and remark, after simple calculations, that

$$\tilde{\delta}(z) = \sigma \frac{1}{N} \text{Tr} \tilde{\mathbf{T}}(z) \quad (107)$$

We finally denote by  $\mathbf{R}(z)$  and  $\tilde{\mathbf{R}}(z)$  the matrices defined by

$$\mathbf{R}(z) = \left[ -z(1 + \sigma\tilde{\alpha}(z))\mathbf{I}_M + \frac{\mathbf{B}\mathbf{B}^H}{1 + \sigma\alpha(z)} \right]^{-1} \quad (108)$$

$$\tilde{\mathbf{R}}(z) = \left[ -z(1 + \sigma\alpha(z))\mathbf{I}_N + \frac{\mathbf{B}^H \mathbf{B}}{1 + \sigma\tilde{\alpha}(z)} \right]^{-1} \quad (109)$$

Using Property 6 of Lemma 1, it is easily checked that functions  $(-z(1 + \sigma\delta(z)))^{-1}$ ,  $(-z(1 + \sigma\tilde{\delta}(z)))^{-1}$ ,  $(-z(1 + \sigma\alpha(z)))^{-1}$ ,  $(-z(1 + \sigma\tilde{\alpha}(z)))^{-1}$  are Stieltjès transforms of probability measures carried by  $\mathbb{R}_+$ . Proposition 5.1 of [16] thus implies that matrix valued functions  $\mathbf{T}(z)$ ,  $\tilde{\mathbf{T}}(z)$ ,  $\mathbf{R}(z)$ ,  $\tilde{\mathbf{R}}(z)$  are holomorphic in  $\mathbb{C} - \mathbb{R}_+$ , coincide with the Stieltjès transforms of positive matrix valued measures carried by  $\mathbb{R}_+$ , the mass of which are equal to  $\mathbf{I}$ , and their spectral norms are bounded by  $\frac{1}{|\text{Im}(z)|}$  on  $\mathbb{C}_+$  (see [16] for more details).

We finally recall that matrices  $\mathbf{Q}(z)$  and  $\tilde{\mathbf{Q}}(z)$  satisfy  $\|\mathbf{Q}\| \leq (\text{Im}(z))^{-1}$  and  $\|\tilde{\mathbf{Q}}\| \leq (\text{Im}(z))^{-1}$  for  $z \in \mathbb{C}_+$  (see e.g. [11], [18], [14], [16]).

In order to establish Proposition 4, we have first to study the term

$$\mathbb{E} \left( \frac{1}{N} \text{Tr} \mathbf{Q}(z) \right) - \frac{1}{N} \text{Tr} \mathbf{R}(z)$$

A. *Study of*  $\mathbb{E} \left( \frac{1}{N} \text{Tr} \mathbf{Q}(z) \right) - \frac{1}{N} \text{Tr} \mathbf{R}(z)$

Let  $\tilde{\tau}(z)$  and  $\tilde{\Delta}(z)$  defined by

$$\tilde{\tau}(z) = \frac{-\sigma}{z(1 + \sigma\alpha(z))} \left[ 1 - \frac{1}{N} \text{Tr} \left( \frac{\mathbf{B}^H \mathbb{E}[\mathbf{Q}(z)] \mathbf{B}}{1 + \sigma\alpha(z)} \right) \right] \quad (110)$$

and

$$\mathbf{\Delta}(z) := \mathbf{\Delta}_1(z) + \mathbf{\Delta}_2(z) + \mathbf{\Delta}_3(z) \quad (111)$$

$$\mathbf{\Delta}_1(z) := -\frac{\sigma}{1 + \sigma\alpha(z)} \mathbb{E} \left[ \mathbf{Q}(z) \mathbf{\Sigma} \mathbf{\Sigma}^H \frac{\sigma}{N} \text{Tr} (\mathbf{Q}(z) - \mathbb{E}[\mathbf{Q}(z)]) \right] \quad (112)$$

$$\mathbf{\Delta}_2(z) := -\frac{\sigma^2}{1 + \sigma\alpha(z)} \mathbb{E} \left[ (\mathbf{Q}(z) - \mathbb{E}[\mathbf{Q}(z)]) \frac{\sigma}{N} \text{Tr} \mathbf{\Sigma}^H \mathbf{Q}(z) \mathbf{B} \right] \quad (113)$$

$$\mathbf{\Delta}_3(z) := \frac{\sigma^2}{(1 + \sigma\alpha(z))^2} \mathbb{E} [\mathbf{Q}(z)] \mathbb{E} \left[ \frac{\sigma}{N} \text{Tr} (\mathbf{Q}(z) - \mathbb{E}[\mathbf{Q}(z)]) \frac{\sigma}{N} \text{Tr} \mathbf{\Sigma}^H \mathbf{Q}(z) \mathbf{B} \right] \quad (114)$$

As it will become apparent below, the entries of matrix  $\mathbf{\Delta}(z)$  converge towards 0.

It is proved in [23] that for each  $z \in \mathbb{R}_*^*$ , the following equality holds true

$$\mathbf{I}_M + \mathbf{\Delta}(z) = \mathbb{E} [\mathbf{Q}(z)] \left( -z(1 + \sigma\tilde{\tau}(z)) \mathbf{I}_M + \frac{\mathbf{B}\mathbf{B}^H}{1 + \sigma\alpha(z)} \right) \quad (115)$$

As the lefthandside and the righthandside of (115) are analytic on  $\mathbb{C} - \mathbb{R}_+$ , Eq. (115) holds not only on  $\mathbb{R}_*^-$ , but on  $\mathbb{C} - \mathbb{R}_+$ . It is shown in [23] that  $\tilde{\alpha}(z) - \tilde{\tau}(z)$  converges towards 0 for each  $z \in \mathbb{C} - \mathbb{R}_+$  when  $N \rightarrow +\infty$ . The general expression of  $\tilde{\alpha}(z) - \tilde{\tau}(z)$  given in [23] is complicated. However, the simplicity of the model considered in this paper (matrices  $\mathbf{D}$  and  $\tilde{\mathbf{D}}$  in [23] are reduced to  $\sigma\mathbf{I}$ ) allows to derive the following Lemma.

**Lemma 7.** *For each  $z \in \mathbb{C} - \mathbb{R}_+$ , it holds that*

$$z(\tilde{\alpha}(z) - \tilde{\tau}(z)) = -\sigma \frac{1}{N} \text{Tr} \mathbf{\Delta}(z) \quad (116)$$

*Proof:* Multiplying (115) from both sides by  $\sigma$  and taking the trace, we obtain

$$\sigma \frac{1}{N} \text{Tr} \left( \frac{\mathbf{B}^H \mathbb{E}[\mathbf{Q}(z)] \mathbf{B}}{1 + \sigma\alpha(z)} \right) = \sigma \frac{M}{N} + \sigma \frac{1}{N} \text{Tr} \mathbf{\Delta}(z) + z(1 + \sigma\tilde{\tau}(z)) \alpha(z) \quad (117)$$

From the definition of  $\tilde{\tau}(z)$  (equation (110)), we also have

$$\sigma \frac{1}{N} \text{Tr} \left( \frac{\mathbf{B}^H \mathbb{E}[\mathbf{Q}(z)] \mathbf{B}}{1 + \sigma\alpha(z)} \right) = z\tilde{\tau}(z)(1 + \sigma\alpha(z)) + \sigma \quad (118)$$

The two above equalities imply that

$$\alpha(z) - \tilde{\tau}(z) = \frac{\sigma(1 - c)}{z} - \frac{\sigma}{z} \frac{1}{N} \text{Tr} \mathbf{\Delta}(z) \quad (119)$$

Using (101), we get that

$$\tilde{\alpha}(z) - \tilde{\tau}(z) = -\frac{\sigma}{z} \frac{1}{N} \text{Tr} \mathbf{\Delta}(z) \quad (120)$$

and (116). ■

Writing the righthandside of (115) as

$$\mathbb{E}(\mathbf{Q}(z))\mathbf{R}(z)^{-1} + z\sigma(\tilde{\alpha}(z) - \tilde{\tau}(z))\mathbb{E}(\mathbf{Q}(z))$$

and using (116), we obtain immediately that

$$\mathbb{E}(\mathbf{Q}(z)) - \mathbf{R}(z) = \mathbf{\Delta}(z)\mathbf{R}(z) + \sigma^2 \frac{1}{N} [\text{Tr } \mathbf{\Delta}(z)] \mathbb{E}(\mathbf{Q}(z))\mathbf{R}(z) \quad (121)$$

and that

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr } \mathbf{Q}(z) \right] - \frac{1}{N} \text{Tr } \mathbf{R}(z) = \frac{\sigma}{N} \text{Tr} (\mathbb{E}[\mathbf{Q}(z)] \mathbf{R}(z)) \frac{\sigma}{N} \text{Tr } \mathbf{\Delta}(z) + \frac{1}{N} \text{Tr } \mathbf{\Delta}(z)\mathbf{R}(z) \quad (122)$$

The above expression of  $\mathbb{E} \left[ \frac{1}{N} \text{Tr } \mathbf{Q}(z) \right] - \frac{1}{N} \text{Tr } \mathbf{R}(z)$  allows to prove the following Proposition.

**Proposition 6.**  $\forall z \in \mathbb{C}_+$ , we have

$$\left| \mathbb{E} \left[ \frac{1}{N} \text{Tr } \mathbf{Q}(z) \right] - \frac{1}{N} \text{Tr } \mathbf{R}(z) \right| \leq \frac{1}{N^2} P_1(|z|) P_2(|\text{Im}(z)|^{-1}) \quad (123)$$

*Proof:* We first prove the following preliminary result.

**Lemma 8.** Consider  $M \times M$  matrices  $\mathbf{U}_N$  and  $M \times N$  matrices  $\mathbf{U}'_N$  satisfying  $\sup_N \|\mathbf{U}_N\| < \infty, \sup_N \|\mathbf{U}'_N\| < \infty$ . Then, we have  $\forall z \in \mathbb{C}_+$

$$\text{Var} \left[ \frac{1}{N} \text{Tr } \mathbf{Q}(z) \mathbf{U} \right] \leq C \|\mathbf{U}\|^2 \frac{1}{N^2} P_1(|z|) P_2\left(\frac{1}{|\text{Im}(z)|}\right) \quad (124)$$

$$\text{Var} \left[ \frac{1}{N} \text{Tr } \Sigma^H \mathbf{Q}(z) \mathbf{U}' \right] \leq C \frac{1}{N^2} \|\mathbf{U}'\|^2 P_1(|z|) P_2\left(\frac{1}{|\text{Im}(z)|}\right) \quad (125)$$

where the polynomials  $P_1$  and  $P_2$  and constant  $C$  are independent of  $M, N$  and  $\mathbf{U}, \mathbf{U}'$ .

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

$$\frac{\partial[\mathbf{Q}(z)]_{pq}}{\partial \mathbf{W}_{ij}} = -\mathbf{Q}_{pi} (\Sigma^H \mathbf{Q})_{jq} \quad (126)$$

$$\frac{\partial[\mathbf{Q}(z)]_{pq}}{\partial \mathbf{W}_{ij}^*} = -\mathbf{Q}_{iq} (\mathbf{Q} \Sigma)_{pj} \quad (127)$$

The Nash-Poincaré inequality gives

$$\text{Var} \left[ \frac{1}{N} \text{Tr} \mathbf{Q}(z) \mathbf{U} \right] \leq \frac{\sigma^2}{N} \sum_{i,j} \left[ \mathbb{E} \left| \frac{1}{N} \sum_{p,q} \frac{\partial[\mathbf{Q}(z)]_{pq}}{\partial \mathbf{W}_{ij}} \mathbf{U}_{qp} \right|^2 + \mathbb{E} \left| \frac{1}{N} \sum_{p,q} \frac{\partial[\mathbf{Q}(z)]_{pq}}{\partial \mathbf{W}_{ij}^*} \mathbf{U}_{qp} \right|^2 \right] \quad (128)$$

$$\leq C \frac{1}{N^3} \sum_{i,j} \left[ \mathbb{E} \left| [\boldsymbol{\Sigma}^H \mathbf{Q}(z) \mathbf{U} \mathbf{Q}(z)]_{ji} \right|^2 + \mathbb{E} \left| [\mathbf{Q}(z) \mathbf{U} \mathbf{Q}(z) \boldsymbol{\Sigma}^H]_{ij} \right|^2 \right] \quad (129)$$

$$\leq C \frac{1}{N^3} \sum_j \mathbb{E} \left[ (\boldsymbol{\Sigma}^H \mathbf{Q}(z) \mathbf{U} \mathbf{Q}(z) \mathbf{Q}(z)^H \mathbf{U}^H \mathbf{Q}(z)^H \boldsymbol{\Sigma})_{jj} \right] + \quad (130)$$

$$C \frac{1}{N^3} \sum_j \mathbb{E} \left[ (\boldsymbol{\Sigma}^H \mathbf{Q}(z)^H \mathbf{U}^H \mathbf{Q}(z)^H \mathbf{Q}(z) \mathbf{U} \mathbf{Q}(z) \boldsymbol{\Sigma})_{jj} \right] \quad (131)$$

$$\leq C \frac{1}{N^3} \mathbb{E} \left[ \text{Tr} (\mathbf{Q}(z) \mathbf{U} \mathbf{Q}(z) \mathbf{Q}(z)^H \mathbf{U}^H \mathbf{Q}(z)^H \boldsymbol{\Sigma} \boldsymbol{\Sigma}^H) \right] + \quad (132)$$

$$C \frac{1}{N^3} \mathbb{E} \left[ \text{Tr} (\mathbf{Q}(z)^H \mathbf{U}^H \mathbf{Q}(z)^H \mathbf{Q}(z) \mathbf{U} \mathbf{Q}(z) \boldsymbol{\Sigma} \boldsymbol{\Sigma}^H) \right] \quad (133)$$

We use the resolvent identity

$$\mathbf{Q}(z) \boldsymbol{\Sigma} \boldsymbol{\Sigma}^H = \boldsymbol{\Sigma} \boldsymbol{\Sigma}^H \mathbf{Q}(z) = \mathbf{I} + z \mathbf{Q}(z) \quad (134)$$

Therefore,

$$\text{Var} \left[ \frac{1}{N} \text{Tr} \mathbf{Q}(z) \mathbf{U} \right] \leq C \frac{1}{N^3} \mathbb{E} \left| \text{Tr} (\mathbf{Q}(z) \mathbf{U} \mathbf{Q}(z) \mathbf{Q}(z)^H \mathbf{U}^H (\mathbf{I} + z^* \mathbf{Q}(z)^H)) \right| + \quad (135)$$

$$C \frac{1}{N^3} \mathbb{E} \left| \text{Tr} (\mathbf{Q}(z)^H \mathbf{U}^H \mathbf{Q}(z)^H \mathbf{Q}(z) \mathbf{U} (\mathbf{I} + z \mathbf{Q}(z))) \right| \quad (136)$$

$$\leq C \|\mathbf{U}\|^2 \frac{1}{N^2} \left( \frac{|z|}{|\text{Im}(z)|^4} + \frac{1}{|\text{Im}(z)|^3} \right) \quad (137)$$

$$\leq C \|\mathbf{U}\|^2 \frac{1}{N^2} (|z| + 1) \left( \frac{1}{|\text{Im}(z)|^4} + \frac{1}{|\text{Im}(z)|^3} \right) \quad (138)$$

which establishes the first statement of Lemma 8. ■

We now complete the proof of Proposition 6. For this, we use the inequalities  $\|\mathbf{Q}(z)\| \leq \frac{1}{|\text{Im}(z)|}$  and  $\|\mathbf{R}(z)\| \leq \frac{1}{|\text{Im}(z)|}$  for  $z \in \mathbb{C} - \mathbb{R}$ . This leads to

$$\left| \frac{\sigma}{N} \text{Tr} (\mathbb{E} [\mathbf{Q}(z)] \mathbf{R}(z)) \frac{\sigma}{N} \text{Tr} \boldsymbol{\Delta}(z) \right| \leq C \frac{1}{|\text{Im}(z)|^2} \left| \frac{1}{N} \text{Tr} \boldsymbol{\Delta}_1(z) + \frac{1}{N} \text{Tr} \boldsymbol{\Delta}_2(z) + \frac{1}{N} \text{Tr} \boldsymbol{\Delta}_3(z) \right| \quad (139)$$

We establish that

$$\left| \frac{1}{N} \text{Tr} (\boldsymbol{\Delta}_i(z)) \right| \leq \frac{1}{N^2} \mathbf{P}_1(|z|) \mathbf{P}_2(|\text{Im}(z)|^{-1}) \quad (140)$$

for  $i = 1, 2, 3$ . In order to evaluate  $\frac{1}{N} \text{Tr} (\boldsymbol{\Delta}_i(z))$  for  $i = 1, 2, 3$ , we first remark that

$$\frac{1}{|z(1 + \sigma \alpha(z))|} < \frac{1}{|\text{Im}(z)|}$$

because  $-\frac{1}{z(1+\sigma\alpha(z))}$  is the Stieltjès transform of a probability measure. Therefore, we have

$$\frac{1}{|1 + \sigma\alpha(z)|} < \frac{|z|}{|\operatorname{Im}(z)|} \quad (141)$$

The resolvent identity (134) implies that

$$\frac{1}{N} \operatorname{Tr}(\mathbf{\Delta}_1(z)) = -\frac{\sigma}{1 + \sigma\alpha(z)} \mathbb{E} \left[ z \frac{1}{N} \operatorname{Tr} \mathbf{Q}(z) \frac{\sigma}{N} \operatorname{Tr} (\mathbf{Q}(z) - \mathbb{E} \mathbf{Q}(z)) \right] \quad (142)$$

$$= -\frac{\sigma}{1 + \sigma\alpha(z)} \mathbb{E} \left[ \frac{z}{N} (\operatorname{Tr} (\mathbf{Q}(z) - \mathbb{E} \mathbf{Q}(z))) \frac{\sigma}{N} (\operatorname{Tr} (\mathbf{Q}(z) - \mathbb{E} \mathbf{Q}(z))) \right] \quad (143)$$

(141) and the first statement of Lemma 8 give immediately (140) for  $i = 1$ . Similarly,  $\frac{1}{N} \operatorname{Tr}(\mathbf{\Delta}_2(z))$  can be written as

$$\frac{1}{N} \operatorname{Tr}(\mathbf{\Delta}_2(z)) = -\frac{\sigma^2}{1 + \sigma\alpha(z)} \mathbb{E} \left[ \left( \frac{1}{N} \operatorname{Tr} \mathbf{Q}(z) - \mathbb{E} \left( \frac{1}{N} \operatorname{Tr} \mathbf{Q}(z) \right) \right) \left( \frac{\sigma}{N} \operatorname{Tr} \Sigma^H \mathbf{Q}(z) \mathbf{B} - \mathbb{E} \left( \frac{\sigma}{N} \operatorname{Tr} \Sigma^H \mathbf{Q}(z) \mathbf{B} \right) \right) \right]$$

Using again (141), the Schwartz inequality, Lemma 8, and the identity  $(xy)^{1/2} \leq (\frac{x+y}{2})$  for  $x \geq 0, y \geq 0$ , we get (140) for  $i = 2$ . (140) for  $i = 3$  is obtained similarly. This and (139) imply that

$$\left| \frac{\sigma}{N} \operatorname{Tr} (\mathbb{E} [\mathbf{Q}(z)] \mathbf{R}(z)) \frac{\sigma}{N} \operatorname{Tr} \mathbf{\Delta}(z) \right| \leq \frac{1}{N^2} P_1(|z|) P_2(|\operatorname{Im}(z)|^{-1})$$

Using the same approach and the identity  $\|\mathbf{R}(z)\| \leq (|\operatorname{Im}(z)|)^{-1}$ , we obtain easily that

$$\left| \frac{1}{N} \operatorname{Tr} \mathbf{\Delta}(z) \mathbf{R}(z) \right| \leq \frac{1}{N^2} P_1(|z|) P_2(|\operatorname{Im}(z)|^{-1})$$

(122) thus implies Proposition 6. ■

**Remark 6.** *It is also possible to establish that  $\forall z \in \mathbb{C}_+$ , we have*

$$\left| \mathbb{E} \left[ \frac{1}{N} \operatorname{Tr} \tilde{\mathbf{Q}}(z) \right] - \frac{1}{N} \operatorname{Tr} \tilde{\mathbf{R}}(z) \right| \leq \frac{1}{N^2} P_1(|z|) P_2(|\operatorname{Im}(z)|^{-1}) \quad (144)$$

because it is shown in [23] that a relation similar to (115) holds for  $\mathbb{E}(\tilde{\mathbf{Q}}(z))$ . Following the derivation of (121), we obtain an expression of  $\mathbb{E} \left[ \frac{1}{N} \operatorname{Tr} \tilde{\mathbf{Q}}(z) \right] - \frac{1}{N} \operatorname{Tr} \tilde{\mathbf{R}}(z)$  similar to (122) which allows to establish (144).

*B. Study of  $\mathbb{E} \left( \frac{1}{N} \operatorname{Tr} \mathbf{Q}(z) \right) - \frac{1}{N} \operatorname{Tr} \mathbf{T}(z)$*

In order to complete the proof of Proposition 4, we show in this paragraph that

$$\sigma \left| \mathbb{E} \left( \frac{1}{N} \operatorname{Tr} \mathbf{Q}(z) \right) - \frac{1}{N} \operatorname{Tr} \mathbf{T}(z) \right| = |\alpha(z) - \delta(z)| \leq \frac{1}{N^2} P_1(|z|) P_2(|\operatorname{Im}(z)|^{-1}) \quad (145)$$

for each  $z \in \mathbb{C}_+$ . For this, we denote by  $\epsilon(z)$  and  $\tilde{\epsilon}(z)$  the terms defined by

$$\epsilon(z) = \alpha(z) - \sigma \frac{1}{N} \text{Tr}(\mathbf{R}(z)) = \sigma \left( \mathbb{E} \frac{1}{N} \text{Tr}(\mathbf{Q}(z)) - \frac{1}{N} \text{Tr}(\mathbf{R}(z)) \right) \quad (146)$$

$$\tilde{\epsilon}(z) = \tilde{\alpha}(z) - \sigma \frac{1}{N} \text{Tr}(\tilde{\mathbf{R}}(z)) = \sigma \left( \mathbb{E} \frac{1}{N} \text{Tr}(\tilde{\mathbf{Q}}(z)) - \frac{1}{N} \text{Tr}(\tilde{\mathbf{R}}(z)) \right) \quad (147)$$

Proposition 6 and Remark 6 imply that

$$|\epsilon(z)| \leq \frac{1}{N^2} P_1(|z|) P_2(|\text{Im}(z)|^{-1}) \quad (148)$$

$$|\tilde{\epsilon}(z)| \leq \frac{1}{N^2} P_1(|z|) P_2(|\text{Im}(z)|^{-1}) \quad (149)$$

for each  $z \in \mathbb{C}_+$ . In order to study  $\alpha(z) - \delta(z)$ , we express  $\alpha(z)$  as  $\alpha(z) = \sigma \frac{1}{N} \text{Tr}(\mathbf{R}(z)) + \epsilon(z)$ . Therefore,  $\alpha(z) - \delta(z) = \sigma \frac{1}{N} \text{Tr}(\mathbf{R}(z) - \mathbf{T}(z)) + \epsilon(z)$ . We have similarly  $\tilde{\alpha}(z) - \tilde{\delta}(z) = \sigma \frac{1}{N} \text{Tr}(\tilde{\mathbf{R}}(z) - \tilde{\mathbf{T}}(z)) + \tilde{\epsilon}(z)$ . We remark that  $\mathbf{R}(z) - \mathbf{T}(z)$  can be written as  $\mathbf{R}(z) (\mathbf{T}^{-1}(z) - \mathbf{R}^{-1}(z)) \mathbf{T}(z)$ , and that  $\tilde{\mathbf{R}}(z) - \tilde{\mathbf{T}}(z)$  is equal  $\tilde{\mathbf{R}}(z) (\tilde{\mathbf{T}}^{-1}(z) - \tilde{\mathbf{R}}^{-1}(z)) \tilde{\mathbf{T}}(z)$ . Using the expression of  $\mathbf{R}(z)^{-1}$ ,  $\mathbf{T}(z)^{-1}$ ,  $\tilde{\mathbf{R}}(z)^{-1}$  and  $\tilde{\mathbf{T}}(z)^{-1}$ , we obtain that

$$\begin{pmatrix} \alpha(z) - \delta(z) \\ \tilde{\alpha}(z) - \tilde{\delta}(z) \end{pmatrix} = \mathbf{D}_0(z) \begin{pmatrix} \alpha(z) - \delta(z) \\ \tilde{\alpha}(z) - \tilde{\delta}(z) \end{pmatrix} + \begin{pmatrix} \epsilon(z) \\ \tilde{\epsilon}(z) \end{pmatrix} \quad (150)$$

where

$$\mathbf{D}_0(z) = \begin{pmatrix} u_0(z) & z v_0(z) \\ z \tilde{v}_0(z) & \tilde{u}_0(z) \end{pmatrix} \quad (151)$$

with  $u_0, \tilde{u}_0, v_0, \tilde{v}_0$  defined by

$$u_0(z) = \frac{1}{N} \text{Tr} \frac{\sigma^2 \mathbf{R}(z) \mathbf{B} \mathbf{B}^H \mathbf{T}(z)}{(1 + \sigma \alpha(z))(1 + \sigma \delta(z))} \quad (152)$$

$$\tilde{u}_0(z) = \frac{1}{N} \text{Tr} \frac{\sigma^2 \tilde{\mathbf{R}}(z) \mathbf{B}^H \mathbf{B} \tilde{\mathbf{T}}(z)}{(1 + \sigma \tilde{\alpha}(z))(1 + \sigma \tilde{\delta}(z))} \quad (153)$$

$$v_0(z) = \frac{1}{N} \text{Tr} \sigma^2 \mathbf{R}(z) \mathbf{T}(z) \quad (154)$$

$$\tilde{v}_0(z) = \frac{1}{N} \text{Tr} \sigma^2 \tilde{\mathbf{R}}(z) \tilde{\mathbf{T}}(z) \quad (155)$$

Using the matrix inversion lemma and the observation that matrices  $\mathbf{R}, \mathbf{T}, \mathbf{B} \mathbf{B}^H$  commute, the reader can check easily that  $u_0(z) = \tilde{u}_0(z)$ .

In order to establish (145), we remark that (150) is equivalent to the linear system

$$(\mathbf{I} - \mathbf{D}_0(z)) \begin{pmatrix} \alpha(z) - \delta(z) \\ \tilde{\alpha}(z) - \tilde{\delta}(z) \end{pmatrix} = \begin{pmatrix} \epsilon(z) \\ \tilde{\epsilon}(z) \end{pmatrix} \quad (156)$$

In the following, we show matrix  $(\mathbf{I} - \mathbf{D}_0(z))$  is invertible for  $z \in \mathbb{C}_+$ , and that the entries of its inverse can be bounded by terms such as  $P_1(|z|)P_2(|\text{Im}(z)|^{-1})$ . Proposition 4 will follow immediately from (148) and (149).

We first evaluate a lower bound of  $\det(\mathbf{I} - \mathbf{D}_0(z))$  for  $z \in \mathbb{C}_+$ . For this, we introduce matrix  $\mathbf{D}(z)$  defined by

$$\mathbf{D}(z) = \begin{pmatrix} u(z) & v(z) \\ |z|^2 \tilde{v}(z) & \tilde{u}(z) \end{pmatrix} \quad (157)$$

with  $u, \tilde{u}, v, \tilde{v}$  defined by

$$u(z) = \frac{1}{N} \text{Tr} \frac{\sigma^2 \mathbf{T}(z) \mathbf{B} \mathbf{B}^H \mathbf{T}(z)^H}{|1 + \sigma \delta(z)|^2} \quad (158)$$

$$\tilde{u}(z) = \frac{1}{N} \text{Tr} \frac{\sigma^2 \tilde{\mathbf{T}}(z) \mathbf{B}^H \mathbf{B} \tilde{\mathbf{T}}(z)^H}{|1 + \sigma \tilde{\delta}(z)|^2} \quad (159)$$

$$v(z) = \frac{1}{N} \text{Tr} \sigma^2 \mathbf{T}(z) \mathbf{T}(z)^H \quad (160)$$

$$\tilde{v}(z) = \frac{1}{N} \text{Tr} \sigma^2 \tilde{\mathbf{T}}(z) \tilde{\mathbf{T}}(z)^H \quad (161)$$

and define matrix  $\mathbf{D}'(z)$  as the analogue of  $\mathbf{D}(z)$  but in which  $\mathbf{T}, \tilde{\mathbf{T}}, \delta, \tilde{\delta}$  are replaced by  $\mathbf{R}, \tilde{\mathbf{R}}, \alpha, \tilde{\alpha}$  respectively. The entries of  $\mathbf{D}'(z)$  are denoted by  $u', v', |z|^2 \tilde{v}', \tilde{u}'$ . We note that the entries of  $\mathbf{D}(z)$  and  $\mathbf{D}'(z)$  are positive, and that, using the matrix inversion lemma, it is easily seen that  $u = \tilde{u}$  and that  $u' = \tilde{u}'$ . These matrices are useful because we have the following proposition.

**Proposition 7.** *There exists a strictly positive constant  $\eta$  such that*

$$\det(\mathbf{I} - \mathbf{D}(z)) \geq \frac{1}{(16)^2} \frac{|\text{Im}(z)|^8}{(\eta^2 + |z|^2)^4} \quad (162)$$

for each  $z \in \mathbb{C}_+$  and for each  $N$ . Moreover, there exist an integer  $N_0$  and 2 polynomials  $Q_1$  and  $Q_2$ , independent of  $N$ , with positive coefficients, such that for each  $N > N_0$ ,

$$\det(\mathbf{I} - \mathbf{D}'(z)) \geq \frac{1}{(64)^2} \frac{|\text{Im}(z)|^8}{2(\eta^2 + |z|^2)^4} \quad (163)$$

for each element  $z$  of the set  $\mathbb{E}_N$  defined by

$$\mathbb{E}_N = \{z \in \mathbb{C}_+, 1 - \frac{1}{N^2} Q_1(|z|) Q_2(\text{Im}(z)^{-1}) > 0\} \quad (164)$$

Finally, for each  $N > N_0$ ,

$$|\det(\mathbf{I} - \mathbf{D}_0(z))| > \sqrt{\det(\mathbf{I} - \mathbf{D}(z))} \sqrt{\det(\mathbf{I} - \mathbf{D}'(z))} > \frac{1}{(32)^2} \frac{|\text{Im}(z)|^8}{\sqrt{2}(\eta^2 + |z|^2)^4} \quad (165)$$

if  $z \in \mathbb{E}_N$ .

*Proof:* We first establish (162). For this, we express  $\text{Im}(\delta(z))$  and  $\text{Im}(z\tilde{\delta}(z))$  as

$$\text{Im}(\delta(z)) = \frac{1}{N} \text{Tr}(\sigma \text{Im}(\mathbf{T}(z))) \quad (166)$$

$$\text{Im}(z\tilde{\delta}(z)) = \frac{1}{N} \text{Tr}\left(\sigma \text{Im}(z\tilde{\mathbf{T}}(z))\right) \quad (167)$$

where for each matrix  $\mathbf{U}$ , we define  $\text{Im}(\mathbf{U})$  by  $\text{Im}(\mathbf{U}) = \frac{\mathbf{U} - \mathbf{U}^H}{2i}$ . Writing  $\text{Im}(\mathbf{T}(z))$  as  $\frac{1}{2i} \mathbf{T}(z)(\mathbf{T}(z)^{-H} - \mathbf{T}(z)^{-1})\mathbf{T}(z)^H$  and  $\text{Im}(z\tilde{\mathbf{T}}(z))$  as  $\frac{1}{2i} z \mathbf{T}(z)((z\mathbf{T}(z))^{-H} - (z\mathbf{T}(z))^{-1})(z\mathbf{T}(z))^H$ , we get immediately that

$$\begin{pmatrix} \text{Im}(\delta(z)) \\ \text{Im}(z\tilde{\delta}(z)) \end{pmatrix} = \mathbf{D}(z) \begin{pmatrix} \text{Im}(\delta(z)) \\ \text{Im}(z\tilde{\delta}(z)) \end{pmatrix} + \begin{pmatrix} w(z) \\ \tilde{w}(z) \end{pmatrix} \text{Im}(z) \quad (168)$$

where  $w(z)$  and  $\tilde{w}(z)$  are defined by

$$w(z) = \frac{1}{N} \text{Tr}(\sigma^2 \mathbf{T}(z)\mathbf{T}(z)^H) \quad \tilde{w}(z) = \frac{1}{N} \text{Tr}\left(\frac{\sigma \tilde{\mathbf{T}}(z) \mathbf{B}^H \mathbf{B} \tilde{\mathbf{T}}(z)^H}{|1 + \sigma \tilde{\delta}|^2}\right) \quad (169)$$

This is equivalent to

$$(1 - u) \text{Im}\delta = v \text{Im}(z\tilde{\delta}) + w \text{Im}z \quad (170)$$

$$(1 - \tilde{u}) \text{Im}(z\tilde{\delta}) = |z|^2 \tilde{v} \text{Im}\delta + \tilde{w} \text{Im}z \quad (171)$$

As  $\delta$  and  $\tilde{\delta}$  are proportional to the Stieltjès transform of probability measures carried by  $\mathbb{R}_+$ ,  $\text{Im}(\delta) > 0$ ,  $\text{Im}(z\tilde{\delta}) > 0$  for  $z \in \mathbb{C}_+$  (see Property 5 of Lemma 1). Therefore, (170, 171) imply that  $1 - u = 1 - \tilde{u}$  is strictly positive. After some algebra, we also obtain that  $\det(\mathbf{I} - \mathbf{D}) = (1 - u)(1 - \tilde{u}) - |z|^2 v \tilde{v}$  coincides with

$$\det(\mathbf{I} - \mathbf{D}) = (v\tilde{w} + (1 - \tilde{u})w) \frac{\text{Im}z}{\text{Im}\delta} \quad (172)$$

Therefore,

$$\det(\mathbf{I} - \mathbf{D}) \geq (1 - \tilde{u})w \frac{\text{Im}z}{\text{Im}\delta}$$

As  $\delta(z) = \sigma c m(z)$ , Property 3 of Lemma 1 implies that  $\text{Im}(\delta(z)) \leq \frac{\sigma c}{\text{Im}(z)}$  or equivalently that  $\frac{\text{Im}z}{\text{Im}\delta} \geq (\text{Im}(z))^2 / \sigma c$ . Hence,

$$\det(\mathbf{I} - \mathbf{D}) \geq \frac{(1 - \tilde{u})w(\text{Im}(z))^2}{\sigma c}$$

(170) implies that

$$1 - u = 1 - \tilde{u} > w \frac{\text{Im}z}{\text{Im}\delta} \geq \frac{w(\text{Im}(z))^2}{\sigma c}$$

We finally get that

$$\det(\mathbf{I} - \mathbf{D}) \geq \frac{w^2(\text{Im}(z))^4}{(\sigma c)^2} \quad (173)$$

In order to obtain a lower bound of  $w = \frac{1}{N} \text{Tr} \sigma \mathbf{T} \mathbf{T}^H$ , we first remark that  $\frac{1}{M} \text{Tr} \mathbf{T} \mathbf{T}^H \geq \left| \frac{1}{M} \text{Tr} \mathbf{T} \right|^2 = |m|^2$  by the Jensen inequality. Therefore,  $w \geq \sigma c |m|^2 \geq \sigma c |\text{Im}(m)|^2$ .  $\text{Im}(m(z))$  can be written as

$$\text{Im}(m(z)) = \text{Im}(z) \int_{\mathbb{R}_+} \frac{d\mu_N(\lambda)}{|\lambda - z|^2}$$

We recall that it is shown in [16] that the sequence  $(\mu_N)_{N \geq 0}$  is tight. This implies that it exists  $\eta > 0$  for which  $\mu_N([\eta, +\infty[) \leq 1/2$  for each  $N \in \mathbb{N}$ , or equivalently for which

$$\mu_N([0, \eta]) > 1/2 \quad (174)$$

for each integer  $N$ . It is clear that

$$\int_{\mathbb{R}_+} \frac{d\mu_N(\lambda)}{|\lambda - z|^2} > \int_0^\eta \frac{d\mu_N(\lambda)}{|\lambda - z|^2} > \frac{1}{2(\eta^2 + |z|^2)} \mu_N([0, \eta]) > \frac{1}{4(\eta^2 + |z|^2)}$$

Therefore,  $w > \frac{\sigma c (\text{Im}(z))^2}{16(\eta^2 + |z|^2)^2}$  and Eq. (173) gives (162).

We now establish (163). For this, we express that  $\text{Im}(\alpha(z))$  and  $\text{Im}(z\tilde{\alpha}(z))$  as

$$\text{Im}(\alpha(z)) = \frac{1}{N} \text{Tr}(\sigma \text{Im}(\mathbf{R}(z))) + \text{Im}(\epsilon(z)) \quad (175)$$

$$\text{Im}(z\tilde{\alpha}(z)) = \frac{1}{N} \text{Tr}(\sigma \text{Im}(z\tilde{\mathbf{R}}(z))) + \text{Im}(z\tilde{\epsilon}(z)) \quad (176)$$

After some algebra, we obtain that

$$\begin{pmatrix} \text{Im}(\alpha(z)) \\ \text{Im}(z\tilde{\alpha}(z)) \end{pmatrix} = \mathbf{D}'(z) \begin{pmatrix} \text{Im}(\alpha(z)) \\ \text{Im}(z\tilde{\alpha}(z)) \end{pmatrix} + \begin{pmatrix} w'(z) \\ \tilde{w}'(z) \end{pmatrix} \text{Im}(z) + \begin{pmatrix} \text{Im}(\epsilon(z)) \\ \text{Im}(z\tilde{\epsilon}(z)) \end{pmatrix} \quad (177)$$

where  $w'(z)$  and  $\tilde{w}'(z)$  are defined as  $w(z)$  and  $\tilde{w}(z)$  by replacing  $\mathbf{T}(z)$ ,  $\tilde{\mathbf{T}}(z)$ ,  $\delta(z)$ ,  $\tilde{\delta}(z)$  by  $\mathbf{R}(z)$ ,  $\tilde{\mathbf{R}}(z)$ ,  $\alpha(z)$ ,  $\tilde{\alpha}(z)$  respectively. This is equivalent to

$$(1 - u') \text{Im}\alpha = v \text{Im}(z\tilde{\alpha}) + w' \text{Im}z + \text{Im}(\epsilon(z)) \quad (178)$$

$$(1 - \tilde{u}') \text{Im}(z\tilde{\alpha}) = |z|^2 \tilde{v}' \text{Im}\alpha + \tilde{w}' \text{Im}z + \text{Im}(z\tilde{\epsilon}(z)) \quad (179)$$

These equations are of course similar to (170, 171) except that the righthandsides of (178, 179) are corrupted by the two error terms  $\text{Im}(\epsilon(z))$  and  $\text{Im}(z\tilde{\epsilon}(z))$ . In order to prove (163), we follow the proof of (162) but take into account the presence of the error terms in (178, 179). As  $\alpha$  and  $\tilde{\alpha}$  are proportional to the Stieltjès transform of probability measures carried by  $\mathbb{R}_+$ ,  $\text{Im}(\alpha) > 0$ ,  $\text{Im}(z\tilde{\alpha}) > 0$  for  $z \in \mathbb{C}_+$ . Therefore, (178) implies that

$$(1 - u') \text{Im}\alpha > w' \text{Im}(z) - |\epsilon(z)| \quad (180)$$

In order to determine a subset of  $\mathbb{C}_+$  on which  $1 - u' = 1 - \tilde{u}'$  is strictly positive, we evaluate a lower bound of  $w'(z) = \frac{1}{N} \text{Tr}(\sigma \mathbf{R}(z) \mathbf{R}(z)^H)$ . For this, we follow what preceds. We express  $w'$  as

$w' = \sigma c \frac{1}{M} \text{Tr}(\mathbf{R}(z)\mathbf{R}(z)^H)$  and note that  $w' \geq \sigma c \left| \frac{1}{M} \text{Tr} \mathbf{R} \right|^2$ . As  $\mathbf{R}(z)$  is the Stieltjès transform of a matrix valued measure whose mass is the matrix  $\mathbf{I}_M$ ,  $\frac{1}{M} \text{Tr} \mathbf{R}(z)$  is the Stieltjès transform of a probability measure  $\xi_N$ . It is shown in [23] that  $\frac{1}{M} \text{Tr} \mathbf{R}(z) - m_N(z) \rightarrow 0$  for each  $z \in \mathbb{C} - \mathbb{R}_+$ . Therefore, the sequence  $(\xi_N - \mu_N)_{N \geq 0}$  converges weakly towards 0.  $\eta > 0$  being defined by (174), it thus exists an integer  $N_1$  for which

$$\xi_N([0, \eta]) > \frac{1}{4} \quad (181)$$

for each  $N > N_1$ . Using the same calculations as above, we obtain that  $w' > \frac{\sigma c (\text{Im}(z))^2}{64(\eta^2 + |z|^2)^2}$ . Hence, using (180) and (148), we get

$$(1 - u') \text{Im}(\alpha) > \frac{\sigma c (\text{Im}(z))^3}{64(\eta^2 + |z|^2)^2} - \frac{1}{N^2} \text{P}_1(|z|) \text{P}_1((\text{Im}(z))^{-1}) \quad (182)$$

If we denote by  $\mathbb{E}_{1,N}$  the subset of  $\mathbb{C}_+$  defined by

$$\frac{\sigma c (\text{Im}(z))^3}{64(\eta^2 + |z|^2)^2} - \frac{1}{N^2} \text{P}_1(|z|) \text{P}_1((\text{Im}(z))^{-1}) > 0 \quad (183)$$

it is clear that  $1 - u' = 1 - \tilde{u}' > 0$  for each  $N > N_1$  and each  $z \in \mathbb{E}_{1,N}$ . We note that  $\mathbb{E}_{1,N}$  can be written as

$$\left\{ z \in \mathbb{C}_+, 1 - \frac{1}{N^2} \text{S}_1(|z|) \text{S}_2((\text{Im}(z))^{-1}) > 0 \right\} \quad (184)$$

for some polynomials with positive coefficients.

Using some algebra as well as the identity  $u' = \tilde{u}'$ , we get that

$$\det(\mathbf{I} - \mathbf{D}') = \left( v' \tilde{w}' + (1 - u') w' \right) \frac{\text{Im} z}{\text{Im} \alpha} + v' \text{Im}(z \tilde{\epsilon}) + (1 - u') \text{Im}(\epsilon) \quad (185)$$

Therefore, for each  $N > N_1$  and each  $z \in \mathbb{E}_{1,N}$ , we have

$$\det(\mathbf{I} - \mathbf{D}') > (1 - u') w' \frac{\text{Im} z}{\text{Im} \alpha} - v' |z \tilde{\epsilon}| - |\epsilon|$$

Moreover, as  $\frac{\text{Im}(\alpha)}{\text{Im}(z)} \leq \frac{\sigma c}{(\text{Im}(z))^2}$ , using (180), we get

$$(1 - u') > \frac{w' (\text{Im}(z))^2}{\sigma c} - \frac{|\epsilon|}{\text{Im}(\alpha)}$$

It is shown in [23] that  $\frac{1}{M} \text{Tr}(\mathbb{E}(\mathbf{Q}(z))) - m_N(z) \rightarrow 0$  for each  $z \in \mathbb{C} - \mathbb{R}_+$ . Therefore, the sequence  $(\omega_N - \mu_N)_{N \geq 0}$  converges weakly towards 0 where measure  $\omega_N$  is defined by (102).  $\eta > 0$  being defined by (174), it thus exists an integer  $N_0 \geq N_1$  for which

$$\omega_N([0, \eta]) > \frac{1}{4} \quad (186)$$

for each  $N > N_0$ . This allows to show that  $\text{Im}(\alpha) > \frac{\sigma c \text{Im}(z)}{8(\eta^2 + |z|^2)}$  for  $N > N_0$ , and that

$$(1 - u') > \frac{w' (\text{Im}(z))^2}{\sigma c} - \frac{8(\eta^2 + |z|^2)}{\sigma c \text{Im}(z)} |\epsilon(z)|$$

As  $\|\mathbf{R}(z)\| \leq (\text{Im}(z))^{-1}$ ,  $v' = \frac{1}{N} \text{Tr} \sigma^2 \mathbf{R} \mathbf{R}^H$  verifies  $v' \leq \sigma^2 c (\text{Im}(z))^{-2}$  while  $w' = \frac{1}{N} \text{Tr} \sigma \mathbf{R} \mathbf{R}^H$  is less than  $\sigma c (\text{Im}(z))^{-2}$ . Putting all the pieces together, we obtain that

$$(1 - u') w' \frac{\text{Im} z}{\text{Im} \alpha} > \frac{\text{Im}(z)^8}{(64)^2 (\eta^2 + |z|^2)^4} - \frac{64(\eta^2 + |z|^2)^2}{\sigma c (\text{Im}(z))^4} |\epsilon(z)| \quad (187)$$

and

$$\det(\mathbf{I} - \mathbf{D}') > \frac{\text{Im}(z)^8}{(64)^2 (\eta^2 + |z|^2)^4} - \left(1 + \frac{64(\eta^2 + |z|^2)^2}{\sigma c (\text{Im}(z))^4}\right) |\epsilon(z)| - \frac{\sigma^2 c}{(\text{Im}(z))^2} |z| |\tilde{\epsilon}(z)| \quad (188)$$

for  $N > N_0$  and for  $z \in \mathbb{E}_{1,N}$ . (188) can also be written as

$$\det(\mathbf{I} - \mathbf{D}') > \frac{\text{Im}(z)^8}{(64)^2 (\eta^2 + |z|^2)^4} \left(1 - \frac{1}{N^2} S'_1(|z|) S'_2((\text{Im}(z))^{-1})\right)$$

for  $N > N_0$  and for  $z \in \mathbb{E}_{1,N}$  for some polynomials with positive coefficients independent of  $N$   $S'_1$  and  $S'_2$ . We denote by  $\mathbb{E}_{2,N}$  the set

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

We remark that

$$\left\{ z \in \mathbb{C}_+, 1 - \frac{1}{N^2} S_1(|z|) S_2((\text{Im}(z))^{-1}) - \frac{2}{N^2} S'_1(|z|) S'_2((\text{Im}(z))^{-1}) > 0 \right\} \subset \mathbb{E}_{1,N} \cap \mathbb{E}_{2,N}$$

We consider polynomials  $Q_1$  and  $Q_2$  defined by  $Q_i = S_i + \sqrt{2} S'_i$  for  $i = 1, 2$  and define the set  $\mathbb{E}_N$  by

$$\mathbb{E}_N = \left\{ z \in \mathbb{C}_+, 1 - \frac{1}{N^2} Q_1(|z|) Q_2((\text{Im}(z))^{-1}) > 0 \right\}$$

which is included into  $\mathbb{E}_{1,N} \cap \mathbb{E}_{2,N}$ . It is clear that (163) holds.

In order to verify (165), we first remark that the following inequalities hold:

$$|\det(\mathbf{I} - \mathbf{D}_0(z))| = |(1 - u_0)(1 - \tilde{u}_0) - z^2 v_0 \tilde{v}_0| \quad (189)$$

$$\geq |1 - u_0| |1 - \tilde{u}_0| - |z|^2 |v_0| |\tilde{v}_0| \quad (190)$$

$$\geq (1 - |u_0|)(1 - |\tilde{u}_0|) - |z|^2 |v_0| |\tilde{v}_0| \quad (191)$$

Using the Schwartz inequality, we get that  $|u_0| = |\tilde{u}_0| \leq |u|^{1/2} |u'|^{1/2} = |\tilde{u}|^{1/2} |\tilde{u}'|^{1/2}$ ,  $|v_0| \leq |v|^{1/2} |v'|^{1/2}$ , and  $|\tilde{v}_0| \leq |\tilde{v}|^{1/2} |\tilde{v}'|^{1/2}$ . For  $N > N_0$  and for  $z \in \mathbb{E}_N$ ,  $u = \tilde{u} < 1$  and  $u' = \tilde{u}' < 1$  hold. Therefore, we obtain that

$$|\det(\mathbf{I} - \mathbf{D}_0(z))| \geq (1 - |u|^{1/2} |u'|^{1/2})(1 - |\tilde{u}|^{1/2} |\tilde{u}'|^{1/2}) - |z|^2 |v|^{1/2} |v'|^{1/2} |\tilde{v}|^{1/2} |\tilde{v}'|^{1/2} \quad (192)$$

As  $\det(\mathbf{I} - \mathbf{D}(z)) = (1 - u)(1 - \tilde{u}) - |z|^2 v \tilde{v}$  and  $\det(\mathbf{I} - \mathbf{D}'(z)) = (1 - u')(1 - \tilde{u}') - |z|^2 v' \tilde{v}'$  are positive for  $N > N_0$  and for  $z \in \mathbb{E}_N$ , it is easy to check that the righthandside of (192) is greater than  $(\det(\mathbf{I} - \mathbf{D}(z))\det(\mathbf{I} - \mathbf{D}'(z)))^{1/2}$  for  $N > N_2$  and for  $z \in \mathbb{E}_N$ . This shows (165).  $\blacksquare$

In order to complete the proof of (145), we express  $\alpha(z) - \delta(z)$  as

$$\alpha(z) - \delta(z) = \frac{1}{\det(\mathbf{I} - \mathbf{D}_0(z))} [(1 - \tilde{u}_0(z))\epsilon(z) + z v_0(z)\tilde{\epsilon}(z)]$$

If  $N > N_2$ , and if  $z \in \mathbb{E}_N$ , (165), (148, 149),  $|v_0(z)| \leq \frac{\sigma^2 c}{(\text{Im}(z))^2}$  and  $|u_0(z)| \leq \frac{\sigma^2 b_{max}^2 |z|^2}{(\text{Im}(z))^2}$  (recall that  $b_{max}$  is defined by (7)) give immediately

$$|\alpha(z) - \delta(z)| \leq \frac{1}{N^2} P_1(|z|) P_2((\text{Im}(z))^{-1}) \quad (193)$$

for some polynomials  $P_i$ ,  $i = 1, 2$  with positive coefficients. If  $z \in \mathbb{C}_+ \setminus \mathbb{E}_N$ , we follow the trick of [18] and [14], and remark that

$$|\alpha(z) - \delta(z)| \leq |\alpha(z)| + |\delta(z)| \leq \frac{2\sigma c}{\text{Im}(z)}$$

If  $z \in \mathbb{C}_+ \setminus \mathbb{E}_N$ ,  $2 \leq \frac{2}{N^2} Q_1(|z|) Q_2((\text{Im}(z))^{-1})$  so that

$$|\alpha(z) - \delta(z)| \leq \frac{2\sigma c}{\text{Im}(z)} \frac{1}{N^2} Q_1(|z|) Q_2((\text{Im}(z))^{-1})$$

Therefore, for  $N > N_0$ , and for each  $z \in \mathbb{C}_+$ ,

$$|\alpha(z) - \delta(z)| \leq \frac{1}{N^2} \left( P_1(|z|) P_2((\text{Im}(z))^{-1}) + \frac{2\sigma c}{\text{Im}(z)} Q_1(|z|) Q_2((\text{Im}(z))^{-1}) \right) \leq \frac{1}{N^2} (|z| + C)^k Q((\text{Im}(z))^{-1})$$

where  $k$  is an integer,  $C$  is a positive constant and  $Q$  is a positive coefficients polynomial. Proposition 4 follows directly from the identity  $\alpha(z) - \delta(z) = \sigma c \left( \mathbb{E} \left( \frac{1}{M} \text{Tr} \mathbf{Q}(z) \right) - \frac{1}{M} \text{Tr} \mathbf{T}(z) \right)$ .

## APPENDIX F

### PROOF OF (16).

We first show that for each  $z \in \mathbb{C}_+$ ,  $\mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N$  converges towards 0 on a set of probability 1 which, in principle, depends on  $z$ . In order to obtain the almost sure convergence towards 0 for each  $z \in \mathbb{C} - \mathbb{R}_+$ , we use a standard argument based on Montel's theorem.

We first write

$$\mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N = \mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbb{E}(\mathbf{Q}_N(z))) \mathbf{v}_N + \mathbf{u}_N^H (\mathbb{E}(\mathbf{Q}_N(z)) - \mathbf{T}_N(z)) \mathbf{v}_N \quad (194)$$

We study the second term of the righthandside of (194) and write

$$\mathbf{u}_N^H (\mathbb{E}(\mathbf{Q}_N(z)) - \mathbf{T}_N(z)) \mathbf{v}_N = \mathbf{u}_N^H (\mathbb{E}(\mathbf{Q}_N(z)) - \mathbf{T}_N(z)) \mathbf{v}_N + \mathbf{u}_N^H (\mathbf{R}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N$$

where we recall that matrix  $\mathbf{R}_N(z)$  is defined by (108). (145) implies that  $\alpha_N(z) - \delta_N(z)$  and  $\tilde{\alpha}_N(z) - \tilde{\delta}_N(z)$  converge towards 0 ( $\alpha_N, \delta_N, \tilde{\alpha}_N, \tilde{\delta}_N$  are defined by (100, 98, 101, 99) respectively). Using the identity  $\mathbf{R}_N(z) - \mathbf{T}_N(z) = \mathbf{R}_N(z) (\mathbf{T}_N^{-1}(z) - \mathbf{R}_N^{-1}(z)) \mathbf{T}_N(z)$  allows to express  $\mathbf{u}_N^H (\mathbf{R}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N$  as a linear combination of  $\alpha_N(z) - \delta_N(z)$  and  $\tilde{\alpha}_N(z) - \tilde{\delta}_N(z)$ . As  $\|\mathbf{R}_N(z)\| \leq |\text{Im}(z)|^{-1}$ ,  $\|\mathbf{T}_N(z)\| \leq |\text{Im}(z)|^{-1}$ , the coefficients of this linear combination remain bounded when  $N \rightarrow +\infty$ . This shows that  $\mathbf{u}_N^H (\mathbf{R}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N$  converges towards 0.

In order to study  $\mathbf{u}_N^H (\mathbb{E}(\mathbf{Q}_N(z)) - \mathbf{R}_N(z)) \mathbf{v}_N$ , we use relation (121). Using the Nash-Poincaré inequality, it is easy to check that  $\mathbf{u}_N^H \mathbf{R}_N(z) \mathbf{\Delta}_N(z) \mathbf{v}_N \rightarrow 0$ . (140) implies moreover that  $\frac{1}{N} \text{Tr} \mathbf{\Delta}_N(z) \rightarrow 0$ . (121) thus shows that  $\mathbf{u}_N^H (\mathbb{E}(\mathbf{Q}_N(z)) - \mathbf{R}_N(z)) \mathbf{v}_N \rightarrow 0$ .

It remains to prove that  $x_N(z) = \mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbb{E}(\mathbf{Q}_N(z))) \mathbf{v}_N$  converges towards 0 almost surely. For this, it is sufficient to show that

$$\mathbb{E}|x_N(z)|^4 \leq \frac{C(z)}{N^2} \quad (195)$$

where  $C(z)$  does not depend on  $N$ . We express  $\mathbb{E}|x_N(z)|^4$  as

$$\mathbb{E}|x_N(z)|^4 = |\mathbb{E}(x_N(z)^2)|^2 + \text{Var}(x_N(z))^2$$

We remark that  $|\mathbb{E}(x_N(z)^2)|^2 \leq (\mathbb{E}|x_N(z)|^2)^2$ . Moreover,  $\mathbb{E}(x_N(z)) = 0$  implies that  $\mathbb{E}|x_N(z)|^2 = \text{Var}(x_N(z))$ . Therefore,

$$\mathbb{E}|x_N(z)|^4 \leq (\text{Var}(x_N(z)))^2 + \text{Var}[(x_N(z))^2]$$

Using the Nash-Poincaré inequality, it is easy to show that  $\text{Var}(x_N(z)) \leq \frac{C(z)}{N}$  and that  $\text{Var}((x_N(z))^2) \leq \frac{C(z)}{N^2}$ . This establishes (195) and that  $\mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N$  converges towards 0 on a set of probability 1 depending on  $z$ .

In order to prove the almost sure convergence for each  $z \in \mathbb{C} - \mathbb{R}_+$ , we use the following standard argument. We consider a countable subset  $\mathcal{Z}_c \subset \mathbb{C}_+$  having an accumulation point. On a set  $\Omega$  of probability 1,  $\mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N \rightarrow 0$  for each  $z \in \mathcal{Z}_c$ . We fix a realization of the set  $\Omega$ . We denote by  $y_N(z)$  the function  $y_N(z) = \mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N$ . Functions  $z \rightarrow \mathbf{u}_N^H \mathbf{Q}_N(z) \mathbf{v}_N$  and  $z \rightarrow \mathbf{u}_N^H \mathbf{T}_N(z) \mathbf{v}_N$  are Stieltjès transforms of bounded measures carried by  $\mathbb{R}_+$ . Therefore, function  $y_N$  is analytic on  $\mathbb{C} - \mathbb{R}_+$ , and for each compact subset  $\mathcal{K}$  of  $\mathbb{C} - \mathbb{R}_+$ , it holds that

$$|y_N(z)| \leq \frac{C}{\text{dist}(\mathcal{K}, \mathbb{R}_+)}$$

for some constant  $C$  (this is a trivial generalization of (9) to the Stieltjès transform of a non necessarily positive bounded measure carried by  $\mathbb{R}_+$ ). Montel's theorem ([24]) thus implies that it exists a subsequence  $y_{\psi(N)}$  extracted from  $y_N$  which converges uniformly on each compact subset of  $\mathbb{C} - \mathbb{R}_+$  towards a certain

function  $y_*$  which is analytic on  $\mathbb{C} - \mathbb{R}_+$ . However,  $y_*(z) = 0$  for each  $z \in \mathcal{Z}_c$ , thus showing that  $y_*$  is identically 0 on  $\mathbb{C} - \mathbb{R}_+$ . The limit of each converging subsequence extracted from  $y_N$  is thus identically 0. We thus obtain that the whole sequence  $y_N$  converges uniformly towards 0 on each compact subset of  $\mathbb{C} - \mathbb{R}_+$ . Therefore, for each realization of the probability 1 set  $\Omega$ , we have shown that

$$\mathbf{u}_N^H (\mathbf{Q}_N(z) - \mathbf{T}_N(z)) \mathbf{v}_N \rightarrow 0$$

for each  $z \in \mathbb{C} - \mathbb{R}_+$ . This completes the proof of (16).

## APPENDIX G

### PROOF OF LEMMA 4

An elementary study of function  $x \rightarrow \hat{m}_N(x)$  shows that  $\hat{\omega}_k \in ]\hat{\lambda}_k^{(N)}, \hat{\lambda}_{k+1}^{(N)}[$ ,  $\forall k = 1, \dots, M-1$  and that  $\hat{\omega}_M^{(N)} > \hat{\lambda}_M^{(N)}$ . Therefore, by Theorem 4, we only need to prove that  $\hat{\omega}_{M-K}^{(N)} < t_1^+$  almost surely for all sufficiently large  $N$ .

Consider the contour  $\mathcal{C}$  defined in Proposition 5. Noting that  $\mathcal{C}$  encloses  $\{0\}$  on the complex plane and that  $\text{Ind}_{\mathcal{C}}(0) = 1$ , we can write

$$1 = \frac{1}{2\pi i} \oint_{\mathcal{C}^+} \lambda^{-1} d\lambda \quad (196)$$

$$= \frac{1}{2\pi i} \int_{t_1^-}^{t_1^+} \left( \frac{w'_N(x)}{w_N(x)} \right)^* dx - \frac{1}{2\pi i} \int_{t_1^-}^{t_1^+} \frac{w'_N(x)}{w_N(x)} dx \quad (197)$$

where the notation  $\mathcal{C}^+$  means that the contour  $\mathcal{C}$  is counterclockwise oriented. Since functions  $h \mapsto w_N(x + ih)$  and  $h \mapsto w'_N(x + ih)$  are continuous at  $h = 0$  for all  $x \in ]t_1^-, t_1^+[$  (except for the points  $x \in \{x_1^{(N)-}, x_1^{(N)+}\}$ ), Lemma 3 together with the Dominated Convergence Theorem imply that

$$1 = \lim_{y \downarrow 0} \left[ \frac{1}{2\pi i} \int_{t_1^-}^{t_1^+} \left( \frac{w'_N(x + iy)}{w_N(x + iy)} \right)^* dx - \frac{1}{2\pi i} \int_{t_1^-}^{t_1^+} \frac{w'_N(x + iy)}{w_N(x + iy)} dx \right] \quad (198)$$

$$= \lim_{y \downarrow 0} \left[ \frac{1}{2\pi i} \oint_{\partial \mathcal{R}_y^+} \frac{w'_N(z)}{w_N(z)} d\lambda + \frac{1}{2\pi} \int_{-y}^y \frac{w'_N(t_1^- - ih)}{w_N(t_1^- - ih)} dh - \frac{1}{2\pi} \int_{-y}^y \frac{w'_N(t_1^+ + ih)}{w_N(t_1^+ + ih)} dh \right] \quad (199)$$

where  $\partial \mathcal{R}_y^+$  denotes the contour of the rectangle defined in (62) counterclockwise oriented. The function  $h \mapsto \frac{w'(x+ih)}{w(x+ih)}$  is a continuous function on the compact set  $[-y, y]$  for  $x = t_1^-$  or  $t_1^+$ , and therefore the two last integrals vanish as  $y \downarrow 0$ , so that we can write

$$1 = \lim_{y \downarrow 0} \frac{1}{2\pi i} \oint_{\partial \mathcal{R}_y^+} \frac{w'_N(z)}{w_N(z)} dz.$$

Since the function  $\frac{w'_N(\lambda)}{w_N(\lambda)}$  is holomorphic on  $\mathbb{C} \setminus [x_1^{(N)-}, x_1^{(N)+}]$ , the last integral does not depend on the value of  $y > 0$ , and thus we can drop the limit, i.e.

$$1 = \frac{1}{2\pi i} \oint_{\partial \mathcal{R}_y^+} \frac{w'_N(z)}{w_N(z)} dz. \quad (200)$$

This identity will be key in order to prove that  $\hat{\omega}_{M-K} < t_1^+$  almost surely for all sufficiently large  $N$ .

Before going further into the proof of this result, let us first examine the function  $\hat{w}_N(z)$  defined by (64) when  $z \in \mathbb{R}$ . The following result follows from elementary analysis:

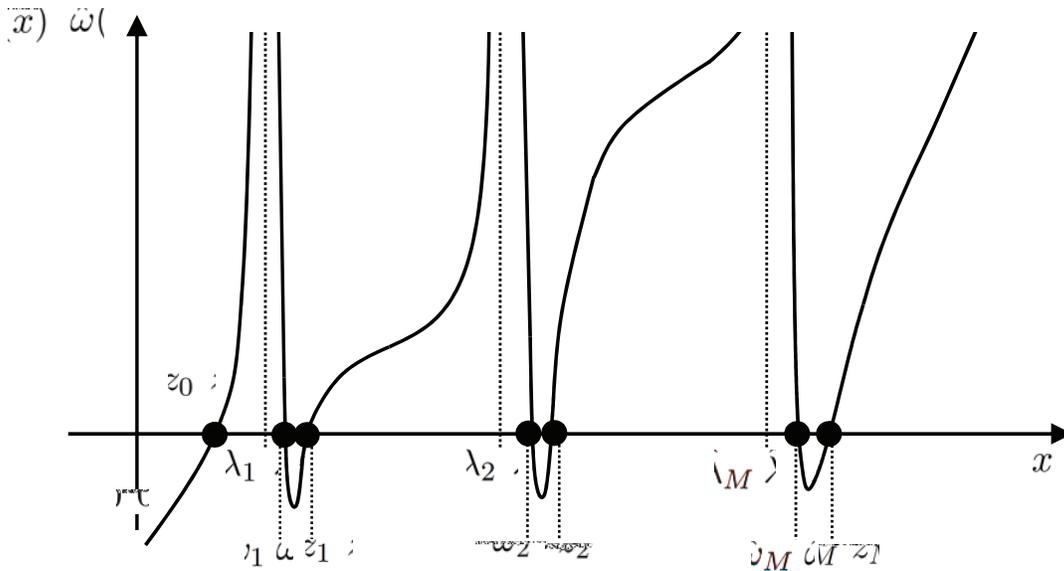


Figure 13. Typical representation of  $\hat{w}_N(x)$  as a function of  $x$  for  $M = 3$  (we drop the dependence with  $N$  from all quantities for clarity).

**Lemma 9.** *The function  $\hat{w}_N$  defined on  $\mathbb{R}$  by*

$$\hat{w}_N(x) = x \left(1 + \sigma^2 c_N \hat{m}_N(x)\right)^2 - \sigma^2 (1 - c_N) \left(1 + \sigma^2 c_N \hat{m}_N(x)\right)$$

*satisfies (see further Figure 13)*

$$\lim_{x \downarrow \hat{\lambda}_k} \hat{w}_N(x) = +\infty, \quad \lim_{x \uparrow \hat{\lambda}_k} \hat{w}_N(x) = +\infty \quad (201)$$

$$\lim_{x \rightarrow +\infty} \hat{w}_N(x) = +\infty, \quad \lim_{x \rightarrow -\infty} \hat{w}_N(x) = -\infty. \quad (202)$$

Moreover,  $\hat{w}_N(x) = 0$  is a polynomial equation with degree  $2M + 1$  with the following zeros:

- One zero in  $]0, \hat{\lambda}_1^{(N)}[$ , denoted as  $\hat{z}_0^{(N)}$ .
- Two zeros in each interval  $]\hat{\lambda}_k^{(N)}, \hat{\lambda}_{k+1}^{(N)}[$ , denoted as  $\hat{\omega}_k^{(N)}, \hat{z}_k^{(N)}$ ,  $k = 1 \dots M - 1$ .
- Two zeros in  $]\hat{\lambda}_M^{(N)}, +\infty[$ , denoted as  $\hat{\omega}_M^{(N)}, \hat{z}_M^{(N)}$ .

Furthermore, we have

$$0 < \hat{z}_0^{(N)} < \hat{\lambda}_1^{(N)} < \hat{\omega}_1^{(N)} < \hat{z}_1^{(N)} < \hat{\lambda}_2^{(N)} < \dots$$

$$\dots < \hat{\lambda}_k^{(N)} < \hat{\omega}_k^{(N)} < \hat{z}_k^{(N)} < \hat{\lambda}_{k+1}^{(N)} < \dots < \hat{\lambda}_M^{(N)} < \hat{\omega}_M^{(N)} < \hat{z}_M^{(N)}.$$

Now, the function  $z \rightarrow \hat{w}_N(z)$ , defined on  $\mathbb{C}$ , is holomorphic everywhere except at poles (of order 2)  $\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_M^{(N)}$ . Moreover, function  $z \rightarrow \frac{\hat{w}'_N(z)}{\hat{w}_N(z)}$  is holomorphic everywhere except at the zeros of  $\hat{w}_N$  and at the sample eigenvalues  $\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_M^{(N)}$ .

Figure 14 gives an schematic representation of the positions of the zeros and poles of  $\hat{w}_N(x)$  in terms of the contour  $\partial\mathcal{R}_y$ . Observe that, for sufficiently high  $N$ , Theorem 4 ensures that  $\{\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_{M-K}^{(N)}\}$  will be inside  $\partial\mathcal{R}_y$ , whereas the rest of the sample eigenvalues will be outside. Given the position of the zeros  $\hat{\omega}_k^{(N)}, \hat{z}_k^{(N)}$  established in Lemma 9, we see that the position of the sample eigenvalues determines that the zeros  $\{\hat{\omega}_k^{(N)}, \hat{z}_k^{(N)}, k = 1 \dots M - K - 1\}$  will also be inside  $\partial\mathcal{R}_y$  for all  $N$  sufficiently high. Furthermore, the remaining zeros will be outside  $\partial\mathcal{R}_y$ , except for the zeros  $\hat{z}_0^{(N)}, \hat{\omega}_{M-K}^{(N)}$  and  $\hat{z}_{M-K}^{(N)}$ , for which we can not state anything. In what follows, we will see that these three zeros are in fact located inside  $\partial\mathcal{R}_y$  with probability one for all large  $N$ , which will conclude the proof of Lemma 4. As a first step, we introduce an intermediate result that establishes that none of these zeros can converge to a the boundary point of  $\partial\mathcal{R}_y$  when  $N \rightarrow +\infty$ .

**Lemma 10.** For all  $N$  large enough,  $\hat{z}_0^{(N)} \neq t_1^-$ ,  $\hat{\omega}_{M-K}^{(N)} \neq t_1^+$  and  $\hat{z}_{M-K}^{(N)} \neq t_1^+$ .

*Proof:* We will just establish that  $\hat{\omega}_{M-K}^{(N)} \neq t_1^+$  and  $\hat{z}_{M-K}^{(N)} \neq t_1^+$ , since the proof that  $\hat{z}_0^{(N)} \neq t_1^-$  is quite similar. For this, we prove the following:

$$\inf_N \inf_{x \in [t_1^+, t_2^-]} |w_N(x)| > 0 \quad (203)$$

$$\lim_{N \rightarrow +\infty} \sup_{x \in [t_1^+, t_2^-]} |w_N(x) - \hat{w}_N(x)| = 0 \text{ a.s.} \quad (204)$$

If (203, 204) hold true, it is clear that almost surely, it exists  $N_1 \in \mathbb{N}$  for which

$$\inf_{N > N_1} \inf_{x \in [t_1^+, t_2^-]} |\hat{w}_N(x)| > 0 \text{ a.s.} \quad (205)$$

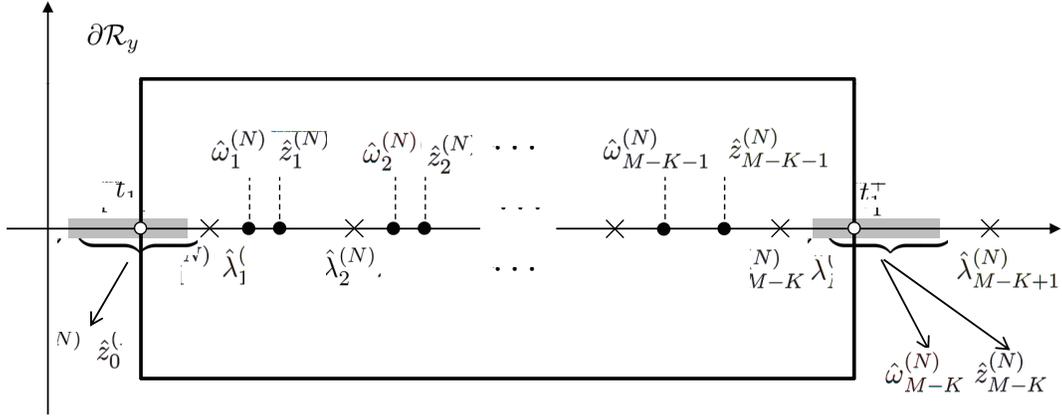


Figure 14. Schematic representation of the position of the zeros (circles) and poles (crosses) of the function  $\hat{w}_N(z)$  on the region enclosed by  $\partial\mathcal{R}_y$ .

a property which implies that  $\hat{\omega}_{M-K}^{(N)} \neq t_1^+$  and  $\hat{z}_{M-K}^{(N)} \neq t_1^+$  for  $N > N_1$ .

In order to prove (203), we note that Assumptions 1 and 2 imply the existence of  $\epsilon > 0$  such that  $w_N(x) > 0$  if  $x \in [t_1^+ - \epsilon, t_2^- + \epsilon]$  and  $N > N_0$ . Now, we write  $w_N(z)$  as

$$w_N(z) = z(1 + \sigma\delta_N(z))(1 + \sigma\tilde{\delta}_N(z)) = z(1 + \sigma^2 c_N m_N(z))(1 + \sigma^2 c_N m_N(z) - \frac{\sigma^2(1 - c_N)}{z}) \quad (206)$$

where we recall that  $\delta_N$  and  $\tilde{\delta}_N$  are defined by (98) and (99) respectively. It has been mentioned in Appendix E that function  $z \rightarrow -\frac{1}{z(1+\sigma\delta_N(z))} = -\frac{1}{z(1+\sigma^2 c_N m_N(z))}$  coincides with the Stieltjès transform of a probability measure carried by  $\mathbb{R}_+$ . We denote by  $\gamma_N$  this measure. As  $w_N(x) > 0$  if  $x \in [t_1^+ - \epsilon, t_2^- + \epsilon]$ , function  $z \rightarrow -\frac{1}{z(1+\sigma\tilde{\delta}_N(z))}$  is analytic on  $\mathbb{C}_+ \cup \mathbb{C}_- \cup [t_1^+ - \epsilon, t_2^- + \epsilon]$  and is real-valued on  $[t_1^+ - \epsilon, t_2^- + \epsilon]$ . The support of measure  $\gamma_N$  is thus included into  $\mathbb{R}_+ - ]t_1^+ - \epsilon, t_2^- + \epsilon[$ . Therefore, Property 9 of Lemma 1 implies that

$$|x(1 + \sigma^2 c_N m_N(x))|^{-1} \leq \frac{1}{\epsilon} \quad (207)$$

for each  $x \in [t_1^+, t_2^-]$ . It can also be shown that  $z \rightarrow -\frac{1}{z(1+\sigma\tilde{\delta}_N(z))} = -\frac{1}{z(1+\sigma^2 c_N m_N(z)) - \sigma^2(1 - c_N)}$  coincides with the Stieltjès transform of a probability measure carried by  $\mathbb{R}_+$ . Using the same approach as above, we obtain that

$$|x(1 + \sigma^2 c_N m_N(x)) - \sigma^2(1 - c_N)|^{-1} \leq \frac{1}{\epsilon} \quad (208)$$

for each  $x \in [t_1^+, t_2^-]$ . This, in turn, implies (203).

In order to establish (204), we note that it is sufficient to establish that

$$\lim_{N \rightarrow +\infty} \sup_{x \in [t_1^+, t_2^-]} |m_N(x) - \hat{m}_N(x)| = 0 \text{ a.s.} \quad (209)$$

Theorem 4 implies the existence of  $\epsilon > 0$  for which, almost surely, function  $z \rightarrow \hat{m}_N(z)$  is analytic on  $\mathbb{C}_+ \cup \mathbb{C}_- \cup ]t_1^+ - \epsilon, t_2^- + \epsilon[$  for  $N > N_1$  where  $N_1 > N_0$  is a certain integer. Eq. (9) implies that for each compact subset  $\mathcal{K}$  of  $\mathbb{C}_+ \cup \mathbb{C}_- \cup ]t_1^+ - \epsilon, t_2^- + \epsilon[$ , there exists a constant  $C(\mathcal{K})$  for which almost surely  $\sup_{N > N_1} \sup_{z \in \mathcal{K}} |\hat{m}_N(z)| \leq C(\mathcal{K})$ . For the same reasons, it holds that  $\sup_{N > N_1} \sup_{z \in \mathcal{K}} |m_N(z)| \leq C(\mathcal{K})$ . Montel's Theorem ([24]) thus implies that it exists a subsequence  $\hat{m}_{\psi(N)} - m_{\psi(N)}$  extracted from  $(\hat{m}_N - m_N)_{N > N_1}$  which converges uniformly on each compact subset of  $\mathbb{C}_+ \cup \mathbb{C}_- \cup ]t_1^+ - \epsilon, t_2^- + \epsilon[$  towards a function  $p_*(z)$ , analytic on  $\mathbb{C}_+ \cup \mathbb{C}_- \cup ]t_1^+ - \epsilon, t_2^- + \epsilon[$ . Proposition 1 implies that almost surely,  $\hat{m}_N(z) - m_N(z) \rightarrow 0$  for each  $z \in \mathbb{C} \setminus \mathbb{R}_+$ . This implies that  $p_*(z)$  is identically zero. As the limit of each convergent subsequence extracted from  $\hat{m}_N - m_N$  is 0, the whole sequence  $(\hat{m}_N - m_N)_{N > N_1}$  converges uniformly towards 0 on each compact subset of  $\mathbb{C}_+ \cup \mathbb{C}_- \cup ]t_1^+ - \epsilon, t_2^- + \epsilon[$ . This, of course, implies (209). This completes the proof of Lemma 10. ■

Using the same arguments as above, it is easy to show that there exists  $N_2 \in \mathbb{N}$  such that  $\inf_{N > N_2} \inf_{z \in \partial \mathcal{R}_y} |w_N(z)| > 0$  and such that, almost surely,  $\inf_{N > N_2} \inf_{z \in \partial \mathcal{R}_y} |\hat{w}_N(z)| > 0$ . It also holds that  $\sup_{N > N_2} \sup_{z \in \partial \mathcal{R}_y} |w'_N(z)| < +\infty$  and  $\sup_{N > N_2} \sup_{z \in \partial \mathcal{R}_y} |\hat{w}'_N(z)| < +\infty$  almost surely. Since almost surely the function  $\frac{\hat{w}'_N(z)}{\hat{w}_N(z)} - \frac{w'_N(z)}{w_N(z)}$  converges to 0 for each  $z \in \partial \mathcal{R}_y$ , the Dominated Convergence Theorem ensures that, with probability one,

$$\left| \frac{1}{2\pi i} \oint_{\partial \mathcal{R}_y^+} \left[ \frac{w'_N(z)}{w_N(z)} - \frac{\hat{w}'_N(z)}{\hat{w}_N(z)} \right] dz \right| \xrightarrow{N \rightarrow +\infty} 0$$

Now, according to Lemma 10,  $\hat{z}_0^{(N)} \neq t_1^-, \hat{\omega}_{M-K}^{(N)} \neq t_1^+, \hat{z}_{M-K}^{(N)} \neq t_1^+$  with probability one for all large  $N$ . Hence, it is possible to use the argument principle to function  $\frac{\hat{w}'(z)}{\hat{w}(z)}$  on contour  $\partial \mathcal{R}_y$ . More precisely,

$$\frac{1}{2\pi i} \oint_{\partial \mathcal{R}_y^+} \frac{\hat{w}'_N(z)}{\hat{w}_N(z)} dz = \text{card} \{z \in : \hat{w}_N(z) = 0\} - 2(M - K)$$

and since the previous integral is an integer, using (200), we finally have with probability one for  $N$  large enough

$$2(M - K) + 1 = \text{card} \{z \in \mathcal{R}_y : \hat{w}_N(z) = 0\}.$$

We already know that  $\hat{z}_1^{(N)}, \dots, \hat{z}_{M-K-1}^{(N)}$  and  $\hat{\omega}_1^{(N)}, \dots, \hat{\omega}_{M-K-1}^{(N)}$ , which are zeros of  $\hat{w}_N(z)$ , belong to  $\mathcal{R}_y$ . Since the total number of zeros is  $2M + 1$ , 3 other zeros of  $\hat{w}_N(z)$  belong to  $\mathcal{R}_y$  with probability one for  $N$  large enough. However, all the zeros of  $\hat{w}_N(z)$  are real-valued, which implies that the 3 additional zeros necessarily include  $\hat{\omega}_{M-K}^{(N)}$ . This concludes the proof Lemma 4.

## APPENDIX H

## PROOF OF (67) AND (68).

We first establish (67). For this, we recall that  $\mathbf{T}_N(z)$  is the Stieltjès transform of a positive matrix valued measure  $\mathbf{\Gamma}_N$  with mass  $\mathbf{I}_N$ . Therefore, function  $z \rightarrow \mathbf{b}_N^H \mathbf{T}_N(z) \mathbf{b}_N$  coincides with the Stieltjès transform of the positive measure  $\mathbf{b}_N^H \mathbf{\Gamma}_N \mathbf{b}_N$ . This measure is clearly absolutely continuous w.r.t. measure  $\text{Tr}(\mathbf{\Gamma}_N)$ , or equivalently w.r.t. measure  $\mu_N = \frac{1}{M} \text{Tr}(\mathbf{\Gamma}_N)$ . The support of  $\mathbf{b}_N^H \mathbf{\Gamma}_N \mathbf{b}_N$  is thus contained into  $\mathcal{S}_N$ . Therefore, it holds that

$$|\mathbf{b}_N^H \mathbf{T}_N(z) \mathbf{b}_N| \leq \frac{\|\mathbf{b}_N\|^2}{\text{dist}(z, \mathcal{S}_N)}$$

(see (9). We have already mentioned in Appendix E and in Appendix G that function  $z \rightarrow (-z(1 + \sigma^2 c_N m_N(z)))^{-1}$  is the Stieltjès transform of a probability measure carried by  $\mathbb{R}_+$ . This function is moreover analytic in  $\mathbb{C} - \mathcal{S}_N$  because  $1 + \sigma^2 c_N m_N(z) \neq 0$  on  $\mathbb{C} - \mathcal{S}_N$  (see Property 6 of Proposition 1), a property which implies that the support of its associated measure is included into  $\mathcal{S}_N$ . Therefore, we have

$$|-z(1 + \sigma^2 c_N m_N(z))|^{-1} \leq \frac{1}{\text{dist}(z, \mathcal{S}_N)}$$

or equivalently

$$|1 + \sigma^2 c_N m_N(z)|^{-1} \leq \frac{|z|}{\text{dist}(z, \mathcal{S}_N)}$$

Assumptions (1) and (2) imply that  $\inf_{N > N_0} \text{dist}(\partial \mathcal{R}_y, \mathcal{S}_N) > 0$ . We thus obtain that

$$\sup_{N > N_0} \sup_{z \in \partial \mathcal{R}_y} \frac{|\mathbf{b}_N^H \mathbf{T}_N(z) \mathbf{b}_N|}{|1 + \sigma^2 c_N m_N(z)|} < +\infty$$

Using again that  $\inf_{N > N_0} \text{dist}(\partial \mathcal{R}_y, \mathcal{S}_N) > 0$ , it can be checked that  $\sup_{N > N_0} \sup_{z \in \partial \mathcal{R}_y} |w'_N(z)| < +\infty$ . This in turn establishes (67).

In order to prove (68), we recall that  $\hat{m}_N(z)$  is the Stieltjès transform of the probability measure  $\hat{\mu}_N = \frac{1}{M} \sum_{k=1}^M \delta(\lambda - \hat{\lambda}_k^{(N)})$ . Assumptions (1) and (2) imply it exists  $N_0 \in \mathbb{N}$  such that the distance between  $\partial \mathcal{R}_y$  and the support of  $\hat{\mu}_N$  is lower bounded by a strictly positive term independent of  $N \geq N_0$ . It is easily seen that  $z \rightarrow \mathbf{b}_N^H \mathbf{Q}_N(z) \mathbf{b}_N$  is the Stieltjès transform of measure  $\frac{1}{M} \sum_{k=1}^M |\mathbf{b}_N^H \hat{\mathbf{e}}_k^{(N)}|^2 \delta(\lambda - \hat{\lambda}_k^{(N)})$ . The support of this measure is included into  $\{\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_M^{(N)}\}$ . Using (9) as above, we deduce from this that

$$\sup_{N \geq N_0} \sup_{z \in \partial \mathcal{R}_y} \mathbf{b}_N^H \mathbf{Q}_N(z) \mathbf{b}_N < +\infty$$

The same arguments can be used to show that  $\sup_{N \geq N_0} \sup_{z \in \partial \mathcal{R}_y} |\hat{w}'_N(z)| < +\infty$ .

Finally, using Property 6 of Lemma 1, it is easily seen that function  $z \rightarrow (-z(1 + \sigma^2 c_N \hat{m}_N(z)))^{-1}$  is the Stieltjès transform of a probability measure. Its support is included into the set  $\{\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_M^{(N)}, \hat{\omega}_1^{(N)}, \dots, \hat{\omega}_M^{(N)}\}$ .

Moreover, in the statement of Lemma 4,  $t_1^-$  and  $t_1^+$  can be replaced by  $t_1^- + \epsilon_1$  and  $t_1^+ - \epsilon_1$  where  $\epsilon_1$  is chosen in such a way that  $t_1^- + \epsilon_1 < \inf_{N > N_0} x_1^{(N)-} < \sup_{N > N_0} x_1^{(N)+} < t_1^+ - \epsilon_1$ . Therefore, the distance between  $\partial\mathcal{R}_y$  and  $\{\hat{\lambda}_1^{(N)}, \dots, \hat{\lambda}_M^{(N)}, \hat{\omega}_1^{(N)}, \dots, \hat{\omega}_M^{(N)}\}$  is lower bounded by a strictly positive term independent of  $N \geq N_0$ . This implies that

$$\sup_{N \geq N_0} \sup_{z \in \partial\mathcal{R}_y} |1 + \sigma^2 c_N \hat{m}_N(z)|^{-1} < +\infty$$

This completes the proof of (68).

## APPENDIX I

### PROOF OF LEMMA 5

We first write the equation in  $\omega$ ,  $1 + \sigma^2 c_N \hat{m}_N(\omega) = 0$  as

$$\frac{\sigma^2 c_N}{M} \sum_{j=1}^M \frac{1}{\hat{\lambda}_j - \omega} + 1 = 0 \quad (210)$$

and by multiplying the left hand side by  $\prod_{i=1}^M (\hat{\lambda}_i - \omega)$ , we define a new polynomial  $Q(\omega)$ , by

$$Q(\omega) = \frac{\sigma^2 c_N}{M} \sum_{j=1}^M \prod_{\substack{l=1 \\ l \neq j}}^M (\hat{\lambda}_l - \omega) + \prod_{l=1}^M (\hat{\lambda}_l - \omega).$$

As the monic polynomial function  $Q$  has  $M$  roots at  $\hat{\omega}_1, \dots, \hat{\omega}_M$ , we can write

$$Q(\omega) = \prod_{l=1}^M (\hat{\omega}_l - \omega)$$

Therefore,

$$Q(\hat{\lambda}_k) = \prod_{l=1}^M (\hat{\omega}_l - \hat{\lambda}_k) = \frac{\sigma^2 c_N}{M} \prod_{\substack{l=1 \\ l \neq k}}^M (\hat{\lambda}_l - \hat{\lambda}_k) \quad (211)$$

which will be useful later on. Let us now consider the derivative of  $Q$  given by

$$Q'(\omega) = - \sum_{j=1}^M \prod_{\substack{l=1 \\ l \neq j}}^M (\hat{\omega}_l - \omega) = - \sum_{j=1}^M \prod_{\substack{l=1 \\ l \neq j}}^M (\hat{\lambda}_l - \omega) - \frac{\sigma^2 c_N}{M} \sum_{m=1}^M \sum_{\substack{l=1 \\ l \neq m}}^M \prod_{\substack{j=1 \\ j \neq m, l}}^M (\hat{\lambda}_j - \omega) \quad (212)$$

Evaluating again this function at point  $\hat{\lambda}_k$ , we obtain

$$Q'(\hat{\lambda}_k) = - \sum_{j=1}^M \prod_{\substack{l=1 \\ l \neq j}}^M (\hat{\omega}_l - \hat{\lambda}_k) = - \prod_{\substack{l=1 \\ l \neq k}}^M (\hat{\lambda}_l - \hat{\lambda}_k) - \frac{2\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \prod_{\substack{j=1 \\ j \neq k, l}}^M (\hat{\lambda}_j - \hat{\lambda}_k) \quad (213)$$

or, dividing both sides by the first term on the right hand side of the equation,

$$\frac{\sum_{j=1}^M \prod_{\substack{l=1 \\ l \neq j}}^M (\hat{\omega}_l - \hat{\lambda}_k)}{\prod_{\substack{l=1 \\ l \neq k}}^M (\hat{\lambda}_l - \hat{\lambda}_k)} = 1 + \frac{2\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \frac{1}{\hat{\lambda}_l - \hat{\lambda}_k}$$

Going back to equation (211), one can also write

$$\frac{\sum_{j=1}^M \prod_{\substack{l=1 \\ l \neq j}}^M (\hat{\omega}_l - \hat{\lambda}_k)}{\prod_{\substack{l=1 \\ l \neq k}}^M (\hat{\lambda}_l - \hat{\lambda}_k)} = \frac{\sigma^2 c_N}{M} \frac{\sum_{j=1}^M \prod_{\substack{l=1 \\ l \neq j}}^M (\hat{\omega}_l - \hat{\lambda}_k)}{\prod_{l=1}^M (\hat{\omega}_l - \hat{\lambda}_k)} = \frac{\sigma^2 c_N}{M} \sum_{l=1}^M \frac{1}{\hat{\omega}_l - \hat{\lambda}_k}. \quad (214)$$

Consequently, we see that we can write

$$1 + \frac{2\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \frac{1}{\hat{\lambda}_l - \hat{\lambda}_k} = \frac{\sigma^2 c_N}{M} \frac{1}{\hat{\omega}_k - \hat{\lambda}_k} + \frac{\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \frac{1}{\hat{\omega}_l - \hat{\lambda}_k}$$

or, reorganizing the terms of this expression in a convenient way,

$$1 + \frac{\sigma^2 c_N}{M} \frac{1}{\hat{\lambda}_k - \hat{\omega}_k} + \frac{\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \frac{1}{\hat{\lambda}_l - \hat{\lambda}_k} = \frac{\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \frac{1}{\hat{\omega}_l - \hat{\lambda}_k} - \frac{\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \frac{1}{\hat{\lambda}_l - \hat{\lambda}_k}. \quad (215)$$

But from the equation in  $\omega$  (210), we obtain

$$1 + \frac{\sigma^2 c_N}{M} \frac{1}{\hat{\lambda}_k - \hat{\omega}_k} + \frac{\sigma^2 c_N}{M} \sum_{\substack{l=1 \\ l \neq k}}^M \frac{1}{\hat{\lambda}_l - \hat{\omega}_k} = 0$$

and by inserting this expression into (215), we finally get the expression in the lemma.

## REFERENCES

- [1] R.O. Schmidt, "Multiple emitter localization and signal parameter estimation," in *Proceedings of the RADC, Spectral Estimation Workshop*, Rome (NY), 1979, pp. 243–258, Reprinted in "Modern Spectral Analysis II", S.B. Kesler (ed.), IEEE Press, New York, 1986.
- [2] E.K. Hung and R.M. Turner, "A fast beamforming algorithm for large arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 19, no. 4, pp. 598–607, Jul. 1983.
- [3] K. Abed-Meraim, J.F. Cardoso, A.Y. Gorokhov, P. Loubaton, and E. Moulines, "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Trans. on Signal Processing*, vol. 45, pp. 42–55, Jan. 1997.
- [4] H. Liu and G. Xu, "A subspace method for signature waveform estimation in synchronous CDMA systems," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1346–1354, Oct. 1996.
- [5] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. on Signal Processing*, vol. 40, no. 3, pp. 590–600, March 1992.
- [6] D.W. Tufts, A.K. Kot, and R.J. Vaccaro, *The Threshold Effect in Signal Processing Algorithms Which Use an Estimated Subspace*, in *SVD and Signal Processing, II: Algorithms, Analysis and Applications*, Elsevier, New York, 1991.
- [7] X. Mestre and M.A. Lagunas, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 598, 2008.
- [8] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5353–5368, Nov. 2008.
- [9] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, 2008.

- [10] 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*, vol. 98, no. 4, pp. 678–694, 2007.
- [11] J.W. Silverstein, “Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices,” *Journal of Multivariate Analysis*, vol. 5, pp. 331–339, 1995.
- [12] 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*, vol. 26, no. 1, pp. 316–345, 1998.
- [13] Z.D. Bai and J.W. Silverstein, “Exact separation of eigenvalues of large dimensional sample covariance matrices,” *Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.
- [14] M. Capitaine, C. Donati-Martin, and D. Féral, “The largest eigenvalue of finite rank deformation of large Wigner matrices: convergence and non-universality of the fluctuations,” *Annals of Probability*, vol. 37, no. 1, pp. 1–47, 2009.
- [15] P. Vallet, P. Loubaton, and X. Mestre, “Improved subspace DoA estimation methods with large arrays: The deterministic signals case,” in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE Signal Processing Society, 2009, pp. 2137–2140.
- [16] W. Hachem, P. Loubaton, and J. Najim, “Deterministic equivalents for certain functionals of large random matrices,” *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.
- [17] 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*, vol. 98, no. 6, pp. 1099–1122, 2007.
- [18] U. Haagerup and S. Thorbjørnsen, “A new application of random matrices:  $\text{Ext}(C_{red}^*(F_2))$  is not a group,” *Annals of Mathematics*, vol. 162, no. 2, pp. 711, 2005.
- [19] M. Capitaine and C. Donati-Martin, “Strong asymptotic freeness of Wigner and Wishart matrices,” *Indiana Univ. Math. Journal*, vol. 56, pp. 295–309, 2007.
- [20] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, “A new approach for mutual information analysis of large dimensional multi-antenna channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.
- [21] L.A. Pastur, “A simple approach to the global regime of Gaussian ensembles of random matrices,” *Ukrainian Mathematical Journal*, vol. 57, no. 6, pp. 936–966, June 2005.
- [22] S. Chatterjee and A. Bose, “A new method for bounding rates of convergence of empirical spectral distributions,” *Journal of Theoretical Probability*, vol. 17, no. 4, pp. 1003–1019, 2004.
- [23] 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*, March 2010.
- [24] J.F. Conway, *Functions of one complex variable*, Springer Verlag, 2nd edition, 1978.