Rapport WP 3-1 - ANR SESAME Fluctuations of an Improved Population Eigenvalue Estimator in Sample Covariance Matrix Models

J. Yao, R. Couillet, J. Najim, M. Debbah

May 11, 2012

Abstract

This article provides a central limit theorem for a consistent estimator of population eigenvalues with large multiplicities based on sample covariance matrices. The focus is on limited sample size situations, whereby the number of available observations is comparable in magnitude to the observation dimension. An exact expression as well as an empirical, asymptotically accurate, approximation of the limiting variance is derived. Simulations are performed that corroborate the theoretical claims.

I. Introduction

Problems of statistical inference based on M independent observations of an N-variate random variable $\mathbf{y}$, with $\mathbb{E}[\mathbf{y}] = 0$ and $\mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{R}_N$ have drawn the attention of researchers from many fields for years: Portfolio optimization in finance [1], gene coexistence in biostatistics [2], channel capacity in wireless communications [3], power estimation in sensor networks [4], distance of targets in array processing [5], etc.

This work was partially supported by Agence Nationale de la Recherche (France), program ANR-07-MDCO-012-01 'Sesame'.

Yao and Najim are with Télécom Paristech, France. {yao,najim}@telecom-paristech.fr ;

Yao is also with Ecole Normale Supérieure and Najim, with CNRS.

Couillet and Debbah are with Supélec, France. {romain.couillet, merouane.debbah}@supelec.fr,

Debbah also holds Alcatel-Lucent/Supélec Flexible Radio chair, France.

In particular, retrieving spectral properties of the *population covariance matrix* $\mathbf{R}_N$, based on the observation of M independent and identically distributed (i.i.d.) samples $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}$, is paramount to many questions of general science. If M is large compared to N, then it is known that almost surely $\|\hat{\mathbf{R}}_N - \mathbf{R}_N\| \to 0$, as $M \to \infty$, for any standard matrix norm, where $\hat{\mathbf{R}}_N$ is the *sample covariance matrix* $\hat{\mathbf{R}}_N \triangleq \frac{1}{M}\sum_{m=1}^{M} \mathbf{y}^{(m)}\mathbf{y}^{(m)H}$. However, one cannot always afford a large number of samples. In order to cope with this issue, random matrix theory [6], [7] has proposed new estimators, mainly spurred by the *G-estimators* of Girko [8]. Other works include convex optimization methods [9], [10], free probability tools [11], [12], and regularized estimation (banding, tapering, thresholding, etc.) [13], [14], [15], when the structure of $\mathbf{R}_N$ is known. Many of those estimators are consistent in the sense that they are asymptotically unbiased as M, N grow large at the same rate. Nonetheless, only recently have techniques been unveiled which allow to estimate individual eigenvalues and functionals of eigenvectors of $\mathbf{R}_N$. The first contributor is Mestre [16]-[17] who studied the case where $\hat{\mathbf{R}}_N = \mathbf{R}_N^{1/2}\mathbf{U}_N\mathbf{U}_N^H\mathbf{R}_N^{1/2}$ with $\mathbf{R}_N$ having eigenvalues with large multiplicities and unknown eigenvectors, and $\mathbf{U}_N$ with i.i.d. entries. For this model, he provides an estimator for every eigenvalue of $\mathbf{R}_N$ with large multiplicity under some separability condition, see also Vallet *et al.* [18], Couillet *et al.* [4] for more elaborate models.

These estimators, although proven asymptotically unbiased, have nonetheless not been fully characterized in terms of their asymptotic performances. It is in particular fundamental to evaluate the variance of these estimators for not-too-large M, N. The purpose of this article is to study the asymptotic fluctuations of the population eigenvalue estimator of [17] in the case of structured population covariance matrices. A central limit theorem (CLT) is provided to describe the asymptotic fluctuations of the estimators with exact expression for the variance as M, N tend to infinity. An empirical, asymptotically accurate, approximation is also derived. For an application of these results in a cognitive radio context, see for instance [19].

The remainder of the article is structured as follows: In Section II, the system model is introduced and the main results from [16], [17] are recalled. In Section III, the CLT for the estimator in [17] is stated and the asymptotic variance derived. In Section IV, an empirical approximation for the variance is derived. Finally, Section V concludes this article. Technical proofs are postponed to the appendix.

II. Estimation of the population eigenvalues

A. Notations

In this article, lowercase (resp. boldface lowercase, boldface uppercase) symbols stand for scalars (resp. vectors, matrices); $\|\mathbf{x}\|$ represents the Euclidean norm of vector $\mathbf{x}$ and $\|\mathbf{M}\|$ stands for the spectral norm of $\mathbf{M}$. The superscripts $(\cdot)^T$ and $(\cdot)^H$ respectively stand for the transpose and transpose conjugate; the trace of $\mathbf{M}$ is denoted by $\mathrm{Tr}(\mathbf{M})$; the mathematical expectation operator, by $\mathbb{E}$. If $\mathbf{x}$ is an $N \times 1$ vector, then $\mathrm{diag}(\mathbf{x})$ is the $N \times N$ matrix with diagonal elements constituted from the components of $\mathbf{x}$. If $z \in \mathbb{C}$, then $\Re(z)$ and $\Im(z)$ respectively stand for z's real and imaginary parts, while $\mathbf{i}$ stands for $\sqrt{-1}$; $\overline{z}$ stands for z's conjugate and $\delta_{k\ell}$ denotes Kronecker's symbol (whose value is 1 if $k = \ell$, 0 otherwise). For two sequences a_n, b_n such that $b_n \neq 0$, $a_n = \mathcal{O}(b_n)$ if $\max_{n\geq 1} \frac{a_n}{b_n} < \infty$ and $a_n = o(b_n)$ if $\frac{a_n}{b_n} \to 0$ when $n \to \infty$.

If the support $\mathcal{S}$ of a probability measure over $\mathbb{R}$ is the finite union of disjoint closed compact intervals $\mathcal{S}_k$ for $1 \leq k \leq L$, we will refer to each compact interval $\mathcal{S}_k$ as a *cluster* of $\mathcal{S}$.

If $\mathbf{Z} \in \mathbb{C}^{N\times N}$ is a Hermitian matrix with eigenvalues $(\xi_i;\ 1 \leq i \leq N)$, we denote by $\mathrm{eig}(\mathbf{Z}) = \{\xi_i, 1 \leq i \leq N\}$ the set of its eigenvalues and by $F^{\mathbf{Z}}$ the empirical distribution of its eigenvalues (also called *spectral distribution* of $\mathbf{Z}$), *i.e.*:

$$F^{\mathbf{Z}}(d\,\lambda) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_i}(d\,\lambda) \ ,$$

where δ_x stands for the Dirac probability measure at x.

Convergence in distribution will be denoted by $\xrightarrow{\mathcal{D}}$, in probability by $\xrightarrow{\mathcal{P}}$, and almost sure convergence, by $\xrightarrow{a.s.}$.

B. Matrix Model

Consider an $N \times M$ matrix $\mathbf{X}_N = (X_{ij})$ whose entries are i.i.d. random variables, with distribution $\mathcal{CN}(0,1)$, i.e. $X_{ij} = U + \mathbf{i}V$, where U, V are both i.i.d. real Gaussian random variables $\mathcal{N}(0, \frac{1}{2})$. Let $\mathbf{R}_N$ be an $N \times N$ Hermitian matrix with L (L being fixed) distinct eigenvalues $\rho_1 < \cdots < \rho_L$ with respective multiplicities $N_1, \cdots, N_L$ (so that $\sum_{i=1}^{L} N_i = N$). Consider now

$$\mathbf{Y}_N = \mathbf{R}_N^{1/2}\mathbf{X}_N \ .$$

The matrix $\mathbf{Y}_N = [\mathbf{y}_1, \cdots, \mathbf{y}_M]$ is the concatenation of M independent observations $[\mathbf{y}_1, \cdots, \mathbf{y}_M]$, where each observation writes $\mathbf{y}_i = \mathbf{R}_N^{1/2}\mathbf{x}_i$ with $\mathbf{X}_N = [\mathbf{x}_1, \cdots, \mathbf{x}_M]$. In particular, the (population) covariance matrix of each observation $\mathbf{y}_i$ is $\mathbf{R}_N = \mathbb{E}\left(\mathbf{y}_i\mathbf{y}_i^H\right)$. In this article, we are interested in recovering information on $\mathbf{R}_N$ based on the observation

$$\hat{\mathbf{R}}_N = \frac{1}{M}\mathbf{R}_N^{1/2}\mathbf{X}_N\mathbf{X}_N^H\mathbf{R}_N^{1/2} \ ,$$

commonly referred to as the *sample covariance matrix* of the $\mathbf{y}_i$'s.

It is in general a complicated task to infer the spectral properties of $\mathbf{R}_N$ based on $\hat{\mathbf{R}}_N$ for all finite N, M. Instead, in the following, we assume that N and M are large, and consider the following asymptotic regime:

Assumption A1. The dimensions N, M and $(N_i)_{1\leq i\leq L}$ satisfy the following conditions:

$$N, M, N_i \to \infty \ , \quad \frac{N}{M} \to c \in (0,\infty) \quad \text{and} \quad \frac{N_i}{M} \to c_i \in (0,\infty) \ , \ 1 \leq i \leq L. \tag{1}$$

This assumption will be shortly referred to as $N, M \to \infty$.

Assumption A2. The limiting support $\mathcal{S}$ of the eigenvalue distribution of $\hat{\mathbf{R}}_N$ is formed of L compact disjoint subsets $(\mathcal{S}_k; 1 \leq k \leq L)$, often referred to as *clusters* in the sequel.

From [17], one can also reformulate this condition in mathematical terms: The limiting support of $\hat{\mathbf{R}}_N$ is formed of L disjoint clusters if and only if for $1 \leq i \leq L$, $\inf_N\{\frac{M}{N} - \Psi_N(i)\} > 0$, where

$$\Psi_N(i) = \begin{cases} \frac{1}{N}\sum_{r=1}^{L} N_r \left(\frac{\rho_r}{\rho_r-\alpha_1}\right)^2 & m = 1, \\ \max\left\{\frac{1}{N}\sum_{r=1}^{L} N_r \left(\frac{\rho_r}{\rho_r-\alpha_{m-1}}\right)^2, \frac{1}{N}\sum_{r=1}^{L} N_r \left(\frac{\rho_r}{\rho_r-\alpha_m}\right)^2\right\} & 1 < m < L, \\ \frac{1}{N}\sum_{r=1}^{L} N_r \left(\frac{\rho_r}{\rho_r-\alpha_{L-1}}\right)^2 & m = L \end{cases}$$

where $\alpha_1 \leq \cdots \leq \alpha_{L-1}$ are the $L-1$ distinct real ordered solutions of the equation:

$$\frac{1}{N}\sum_{r=1}^{L} N_r \frac{\rho_r^2}{(\rho_r - x)^3} = 0.$$

This condition is also called the *separability* condition.

Figure 1 depicts the eigenvalues of a realization of the random matrix $\hat{\mathbf{R}}_N$ and the associated limiting distribution as N, M grow large, for $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 10$ and $N = 60, M = 600$ with $N_1 = N_2 = N_3 = 20$. The separability condition is illustrated there. Figure 2 shows another situation where the separability condition is not satisfied for $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 5$ and $N = 30, M = 80$ with $N_1 = N_2 = N_3 = 10$.

C. Mestre's Estimator of the population eigenvalues

In [17], an estimator of the population eigenvalues $(\rho_k;\ 1 \leq k \leq L)$ based on the observations $\hat{\mathbf{R}}_N$ is proposed.

Theorem 1 ([17, Th. 3]): Let Assumptions **A1** and **A2** hold true and denote by $\hat{\lambda}_1 \leq \cdots \leq \hat{\lambda}_N$ the ordered eigenvalues of $\hat{\mathbf{R}}_N$. Then the following convergence holds true:

$$\hat{\rho}_k - \rho_k \xrightarrow[M,N\to\infty]{a.s.} 0 \ , \tag{2}$$

where

$$\hat{\rho}_k = \frac{M}{N_k} \sum_{m \in \mathcal{N}_k} \left(\hat{\lambda}_m - \hat{\mu}_m \right) \ , \tag{3}$$

with $\mathcal{N}_k = \{\sum_{j=1}^{k-1} N_j + 1, \ldots, \sum_{j=1}^{k} N_j\}$ and the $\hat{\mu}_i$'s defined[1] as follows:

- If $N \leq M$, then $\hat{\mu}_1 \leq \cdots \leq \hat{\mu}_N$ are the real ordered solutions of

$$\frac{1}{N} \sum_{m=1}^{N} \frac{\hat{\lambda}_m}{\hat{\lambda}_m - \mu} = \frac{M}{N} \ . \tag{4}$$

- If $N > M$, $\hat{\mu}_i = 0$ for $1 \leq i \leq N - M$ and $\hat{\mu}_{N-M+1}, \cdots, \hat{\mu}_N$ are the real solutions of the above equation.

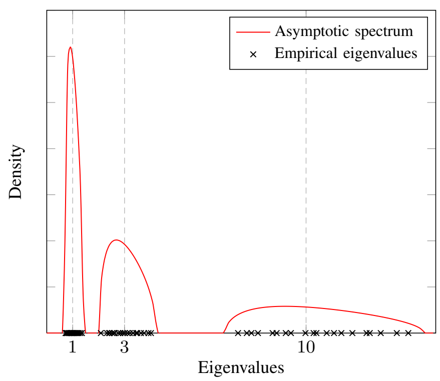

Fig. 1. Empirical and asymptotic eigenvalue distribution of $\hat{\mathbf{R}}_N$ for $L = 3$, $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 10$, $N/M = c = 0.1$, $N = 60$, $N_1 = N_2 = N_3 = 20$.

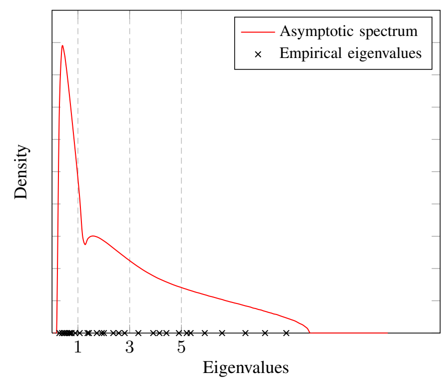

Fig. 2. Empirical and asymptotic eigenvalue distribution of $\hat{\mathbf{R}}_N$ for $L = 3$, $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 5$, $N/M = c = 3/8$, $N = 30$, $N_1 = N_2 = N_3 = 10$.

D. *Integral representation of estimator $\hat{\rho}_k$ - Stieltjes transforms*

The proof of Theorem 1 relies on large random matrix theory, and in particular on the *Stieltjes transform* of a probability distribution. The Stieltjes transform $m_{\mathbb{P}}$ of a probability distribution $\mathbb{P}$ over $\mathbb{R}^+$ is a $\mathbb{C}$-valued function defined by:

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

There also exists an inverse formula to recover the probability distribution associated to a Stieljes transform: Let $a < b$ be two continuity points of the cumulative distribution function associated to $\mathbb{P}$, then

[1]Another characterization of interest of the $\hat{\mu}_i$'s is the fact that they are the eigenvalues of $\mathrm{diag}(\hat{\boldsymbol{\lambda}}) - \frac{1}{M}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^T$, where $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^T$, see for instance [7, Chapter 8].

$$\mathbb{P}([a,b]) = \frac{1}{\pi}\lim_{y\downarrow 0}\Im\left[\int_a^b m_{\mathbb{P}}(x+iy)dx\right].$$

In the case where $F^{\mathbf{Z}}$ is the spectral distribution associated to a Hermitian matrix $\mathbf{Z}\in\mathbb{C}^{N\times N}$ with eigenvalues $(\xi_i;\ 1\le i\le N)$, the Stieltjes transform $m_{\mathbf{Z}}$ of $F^{\mathbf{Z}}$ takes the particular form:

$$\begin{aligned} m_{\mathbf{Z}}(z) &= \int \frac{F^{\mathbf{Z}}(d\,\lambda)}{\lambda - z} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\xi_i - z} \;=\; \frac{1}{N}\mathrm{Tr}\,(\mathbf{Z}-z\mathbf{I}_N)^{-1}\ , \end{aligned}$$

which is the normalized trace of the resolvent $(\mathbf{Z}-z\mathbf{I}_N)^{-1}$. Since the seminal paper of Marčenko and Pastur [20], the Stieltjes transform has proved to be extremely efficient to describe the limiting spectrum of large dimensional random matrices.

In the following, we recall some elements of the proof of Theorem 1, necessary for the remainder of the article. The following important result is due to Bai and Silverstein [21] (see also [20]).

Theorem 2 ([21]): Let Assumption **A1** hold true and denote by F^R the limiting spectral distribution of $\mathbf{R}_N$, *i.e.*

$$F^R(d\,\lambda) = \sum_{k=1}^{L}\frac{c_k}{c}\delta_{\rho_k}(d\,\lambda)\ .$$

Then, the spectral distribution $F^{\hat{\mathbf{R}}_N}$ of the sample covariance matrix $\hat{\mathbf{R}}_N$ converges (weakly and almost surely) to a probability distribution F as $M,N\to\infty$, whose Stieltjes transform $m(z)$ satisfies:

$$m(z) = \frac{1}{c}\underline{m}(z) - \left(1-\frac{1}{c}\right)\frac{1}{z}\ ,$$

for $z\in\mathbb{C}^+ = \{z\in\mathbb{C},\ \Im(z)>0\}$ and where $\underline{m}(z)$ is defined as the unique solution in $\mathbb{C}^+$ of:

$$\underline{m}(z) = -\left(z - c\int\frac{t}{1+t\underline{m}(z)}F^R(dt)\right)^{-1}.$$

Note that $\underline{m}(z)$ is also the Stieltjes transform of a probability distribution $\underline{F}$, which turns out to be the limiting spectral distribution of $F^{\underline{\hat{\mathbf{R}}}_N}$ where $\underline{\hat{\mathbf{R}}}_N$ is defined as:

$$\underline{\hat{\mathbf{R}}}_N \triangleq \frac{1}{M}\mathbf{X}_N^H\mathbf{R}_N\mathbf{X}_N\ .$$

Denote by $m_{\hat{\mathbf{R}}_N}(z)$ and $m_{\underline{\hat{\mathbf{R}}}_N}(z)$ the Stieltjes transforms of $F^{\hat{\mathbf{R}}_N}$ and $F^{\underline{\hat{\mathbf{R}}}_N}$. Note in particular that

$$m_{\hat{\mathbf{R}}_N}(z) = \frac{M}{N}m_{\underline{\hat{\mathbf{R}}}_N}(z) - \left(1-\frac{M}{N}\right)\frac{1}{z}\ .$$

Remark 1: This relation associated to (4) readily implies that for $\hat{\mu}_i\neq 0$, $m_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)=0$. Otherwise stated, the (non null) $\hat{\mu}_i$'s are the zeros of $m_{\underline{\hat{\mathbf{R}}}_N}$. This fact will be of importance in the sequel.

Denote by $m_N(z)$ and $\underline{m}_N(z)$ the finite-dimensional counterparts of $m(z)$ and $\underline{m}(z)$, respectively, defined by the relations:

$$\begin{aligned} \underline{m}_N(z) &= -\left(z - \frac{N}{M}\int \frac{t}{1+t\underline{m}_N(z)} F^{\mathbf{R}_N}(dt)\right)^{-1} , \qquad (5) \\ m_N(z) &= \frac{M}{N}\underline{m}_N(z) - \left(1-\frac{M}{N}\right)\frac{1}{z} , \end{aligned}$$

where $\underline{m}_N(z)$ is the unique solution of (5) satisfying $\underline{m}_N(z) \in \mathbb{C}^+$ if $z \in \mathbb{C}^+$. It can be shown that m_N and $\underline{m}_N$ are Stieltjes transforms of probability measures F_N and $\underline{F}_N$, respectively (cf. [7, Theorem 3.2]).

With these notations at hand, we can now provide some elements of the proof of Theorem 1.

Elements of proof for Theorem 1: By Cauchy's formula, write:

$$\rho_k = \frac{N}{N_k}\frac{1}{2i\pi}\oint_{\Gamma_k}\left(\frac{1}{N}\sum_{r=1}^{L} N_r \frac{w}{\rho_r - w} dw\right) ,$$

where Γ_k is a negatively oriented contour taking values on $\mathbb{C} \setminus \{\rho_1, \cdots, \rho_L\}$ and only enclosing ρ_k. With the change of variable $w = -\frac{1}{\underline{m}_M(z)}$, the condition that the limiting support $\mathcal{S}$ of the eigenvalue distribution of $\mathbf{R}_N$ is formed of L distinct clusters $(\mathcal{S}_k, 1 \leq k \leq L)$ (cf. Assumption **A2**), and standard properties of contour integrals, we can write:

$$\rho_k = \frac{M}{2i\pi N_k}\oint_{\mathcal{R}_k} z\frac{\underline{m}_N'(z)}{\underline{m}_N(z)} dz , \quad 1 \leq k \leq L \qquad (6)$$

where $\mathcal{R}_k$ denotes a negatively oriented, rectangular and symmetric with respect to the abcissa axis, contour which only encloses the corresponding cluster $\mathcal{S}_k$. Defining

$$\hat{\rho}_k \triangleq \frac{M}{2\pi i N_k}\oint_{\mathcal{R}_k} z\frac{m_{\underline{\hat{\mathbf{R}}}_N}'(z)}{m_{\underline{\hat{\mathbf{R}}}_N}(z)} dz , \quad 1 \leq k \leq L , \qquad (7)$$

dominated convergence arguments ensure that $\rho_k - \hat{\rho}_k \to 0$, almost surely. The integral form of $\hat{\rho}_k$ can then be explicitly computed thanks to residue calculus, and this finally yields (3).

■

Remark 2 (About the contour integrals): If $\mathcal{R}_k'$ is another (rectangular and symmetric with respect to the abcissa axis) contour which only encloses the k-th cluster, then the value of the contour integrals in (6) and (7) remains unchanged. In particular, we can arbitrarily choose two non-overlapping contours $\mathcal{R}_k$ and $\mathcal{R}_k'$ of the same cluster $\mathcal{S}_k$ in the sequel.

The main objective of this article is to study the performance of the estimators $(\hat{\rho}_k,\ 1 \leq k \leq L)$. More precisely, we will establish a CLT for $(M(\hat{\rho}_k - \rho_k),\ 1 \leq k \leq L)$ as $M, N \to \infty$, explicitly characterize the limiting covariance matrix $\mathbf{\Theta} = (\Theta_{k\ell})_{1 \leq k,\ell \leq L}$, and finally provide an estimator for $\mathbf{\Theta}$.

$$\Theta_{k\ell} \quad = \quad -\frac{1}{4\pi^2 c_k c_\ell} \oint_{\mathcal{R}_k} \oint_{\mathcal{R}'_\ell} \left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} - \frac{1}{(z_1-z_2)^2} \right] \frac{1}{\underline{m}(z_1)\underline{m}(z_2)} dz_1 dz_2 \ . \tag{8}$$

III. Fluctuations of the population eigenvalue estimators

A. The Central Limit Theorem

The main result of this article is the following CLT which expresses the fluctuations of $(\hat{\rho}_k,\ 1 \le k \le L)$.

Theorem 3: Let Assumptions **A1** and **A2** hold true and recall the definitions of the $\hat{\rho}_k$'s and ρ_k's. Then:

$$(M(\hat{\rho}_k - \rho_k),\ 1 \le k \le L) \xrightarrow[M,N\to\infty]{\mathcal{D}} \mathbf{x} \sim \mathcal{N}_L(0, \mathbf{\Theta}) \ ,$$

where $\mathcal{N}_L$ refers to a real L-dimensional Gaussian distribution, and $\mathbf{\Theta}$ is an $L \times L$ matrix whose entries $\Theta_{k\ell}$ are given by (8). The contours in (8) are defined as follows. The contours $(\mathcal{R}_k; 1 \le k \le L)$ and $(\mathcal{R}'_k; 1 \le k \le L)$ are negatively oriented rectangles, symmetric with respect to the abcissa axis, and only enclosing the cluster $\mathcal{S}_k$. They also verify:

$$\begin{aligned} \mathcal{R}_k \cap \mathcal{R}_\ell \ &= \ \mathcal{R}'_k \cap \mathcal{R}'_\ell \ = \ \emptyset \quad \text{for } k \ne \ell \ , \\ \mathcal{R}_k \cap \mathcal{R}'_\ell \ &= \ \emptyset \qquad \text{for all } k, \ell. \end{aligned}$$

In particular, the families $(\mathcal{R}_k)$ and $(\mathcal{R}'_k)$ are non-overlapping.

Remark 3: In Theorem 3, the separability assumption **A2** can be relaxed to some extent. For example, if only the cluster associated to ρ_k satisfies the separability condition, one can study the fluctuations of $\hat{\rho}_k$ by relying on the same techniques.

B. Proof of Theorem 3

We first outline the main steps of the proof and then provide the details.

Using the integral representation of $\hat{\rho}_k$ and ρ_k, we get:

$$M(\hat{\rho}_k - \rho_k) \quad = \quad \frac{M^2}{2\pi i N_k} \oint_{\mathcal{R}_k} z \left(\frac{m'_{\hat{\underline{\mathbf{R}}}_N}(z)}{m_{\hat{\underline{\mathbf{R}}}_N}(z)} - \frac{\underline{m}'_N(z)}{\underline{m}_N(z)} \right) dz \ .$$

Let $\mathcal{K}$ be the union of the $\mathcal{R}_k$'s and the $\mathcal{R}'_k$'s; denote by $C(\mathcal{K}, \mathsf{X})$ the set of continuous functions from $\mathcal{K}$ to a Banach space X endowed with the supremum norm $\|u\|_\infty = \sup_{\mathcal{K}} |u|$. Consider the process:

$$\begin{aligned} &(X_N, X'_N, u_N, u'_N) : \mathcal{K} \to \mathbb{C}^4 \\ &\qquad\qquad z \mapsto (X_N(z), X'_N(z), u_N(z), u'_N(z)) \end{aligned}$$

where

$$
\begin{aligned}
X_N(z) &= M\left(m_{\hat{\underline{\mathbf{R}}}_N}(z) - \underline{m}_N(z)\right) ,\\
X'_N(z) &= M\left(m'_{\hat{\underline{\mathbf{R}}}_N}(z) - \underline{m}'_N(z)\right) ,\\
u_N(z) &= m_{\hat{\underline{\mathbf{R}}}_N}(z) , \quad u'_N(z) \;=\; m'_{\hat{\underline{\mathbf{R}}}_N}(z) .
\end{aligned}
$$

Then from [22] (see also Proposition 1), (X_N, X'_N, u_N, u'_N) almost surely belongs to $C(\mathcal{K}, \mathbb{C}^4)$ for N, M large enough and $M(\hat{\rho}_k - \rho_k)$ writes:

$$
\begin{aligned}
M(\hat{\rho}_k - \rho_k) &= \frac{M}{2\pi i N_k}\oint_{\mathcal{R}_k} z\left(\frac{\underline{m}_N(z)X'_N(z) - u'_N(z)X_N(z)}{\underline{m}_N(z)u_N(z)}\right) dz\\
&\triangleq \Upsilon_N(X_N, X'_N, u_N, u'_N) ,
\end{aligned}
$$

where

$$
\Upsilon_N(x, x', u, u') \;=\; \frac{M}{2\pi i N_k}\oint_{\mathcal{R}_k} z\left(\frac{\underline{m}_N(z)x'(z) - u'(z)x(z)}{\underline{m}_N(z)u(z)}\right) dz . \tag{9}
$$

Remark 4: Note that Υ_N is a real random variable; if needed, we shall explicitly indicate the dependence on the contour $\mathcal{R}_k$ and write $\Upsilon_N(x, x', u, u', \mathcal{R}_k)$.

Remark 5: Note that, due to formulas (6) and (7) and to Remark 2, the following equality holds true:

$$
\Upsilon_N(x, x', u, u', \mathcal{R}_k) = \Upsilon_N(x, x', u, u', \mathcal{R}'_k),
$$

if $\mathcal{R}_k$ and $\mathcal{R}'_k$ are two contours which only contain the k-th cluster. This fact will be of importance later.

The main idea of the proof of the theorem lies in three steps:

(i) To prove the convergence in distribution of the process (X_N, X'_N, u_N, u'_N) to a Gaussian process.

(ii) To transfer this convergence to the quantity $\Upsilon_N(X_N, X'_N, u_N, u'_N, \mathcal{R}_k)$ with the help of the continuous mapping theorem [23].

(iii) To check that the limit (in distribution) of $\Upsilon_N(X_N, X'_N, u_N, u'_N)$ is Gaussian and to compute the limiting covariance between $\Upsilon_N(X_N, X'_N, u_N, u'_N, \mathcal{R}_k)$ and $\Upsilon_N(X_N, X'_N, u_N, u'_N, \mathcal{R}_\ell)$.

Remark 6: Note that the convergence in step (i) is a distribution convergence at a process level, hence one has to first establish the finite dimensional convergence of the process and then to prove that the process is tight over $\mathcal{C}_k$ (see for instance [24, Theorem 13.1]). Tightness turns out to be difficult to establish due to the lack of control over the eigenvalues of $\hat{\underline{\mathbf{R}}}_N$ whenever the contour crosses the real line. In order to circumvent this issue, we shall introduce, following Bai and Silverstein [25], a process that approximates X_N and X'_N.

Let us now start the proof of Theorem 3.

We begin by simple considerations on complex Gaussian random vectors. Consider a $\mathbb{C}^2$-valued, centered, random vector (U,V). If (U,V) is, as an $\mathbb{R}^4$-valued vector, Gaussian, then its distribution is fully characterized with the quantities:

$$\mathbb{E}U^2,\ \mathbb{E}V^2,\ \mathbb{E}U\bar{U},\ \mathbb{E}V\bar{V},\ \mathbb{E}UV \text{ and } \mathbb{E}U\bar{V}\ .$$

Lemma 1: Let $\underline{\mathcal{S}}$ be the support of the distribution $\underline{F}$.

1) The function $\kappa : (\mathbb{C} \setminus \underline{\mathcal{S}})^2 \to \mathbb{C}$ defined by

$$\kappa(z,\tilde{z}) = \begin{cases} \frac{\underline{m}'(z)\underline{m}'(\tilde{z})}{(\underline{m}(z)-\underline{m}(\tilde{z}))^2} - \frac{1}{(z-\tilde{z})^2} & \text{if } z \neq \tilde{z}\ , \\ \frac{\underline{m}'''(z)}{6\underline{m}'(z)} - \frac{\underline{m}''(z)^2}{4\underline{m}'(z)^2} & \text{if } z = \tilde{z}\ , \end{cases}$$

is continuous and admits partial derivatives up to order 2 over $(\mathbb{C} \setminus \underline{\mathcal{S}})^2$.

2) Let Assumptions **A1** and **A2** hold true and consider a compact set $\mathcal{K} \subset \mathbb{C}$, symmetric with respect to the real axis (i.e. $z \in \mathcal{K} \Rightarrow \bar{z} \in \mathcal{K}$) which does not intersect $\underline{\mathcal{S}}$. Then, the process

$$(X_N, X_N') : \mathcal{K} \to \mathbb{C}^2$$
$$z \mapsto (X_N(z), X_N'(z))$$

converges in distribution to a stochastic process (X,Y) satisfying $\overline{X(z)} = X(\bar{z})$ and $\overline{Y(z)} = Y(\bar{z})$. As an $\mathbb{R}^4$-valued real process, the process (X,Y) is a centered Gaussian process with mean function zero and covariance function defined as follows, for $z, \tilde{z} \in \mathcal{K}$:

$$\begin{aligned} \mathbb{E}\, X(z)X(\tilde{z}) &= \kappa(z,\tilde{z})\ , \\ \mathbb{E}\, Y(z)X(\tilde{z}) &= \frac{\partial \kappa}{\partial z}(z,\tilde{z})\ , \\ \mathbb{E}\, X(z)Y(\tilde{z}) &= \frac{\partial \kappa}{\partial \tilde{z}}(z,\tilde{z})\ , \\ \mathbb{E}\, Y(z)Y(\tilde{z}) &= \frac{\partial^2}{\partial z \partial \tilde{z}}\kappa(z,\tilde{z})\ . \end{aligned} \tag{10}$$

Remark 7: Due to the properties of the process (X,Y), $\mathbb{E}\, X(z)\overline{X(w)} = \kappa(z,\bar{w})$ (and similarly for the other cross-conjugate quantities $\mathbb{E}\, X(z)\overline{Y(w)}$, etc.); moreover, the quantities $\mathbb{E}\, X(z)Y(z)$ and $\mathbb{E}\, Y(z)Y(z)$ can be computed by considering the limits $\lim_{\tilde{z}\to z} \frac{\partial \kappa}{\partial z}(z,\tilde{z})$ and $\lim_{\tilde{z}\to z} \frac{\partial^2 \kappa}{\partial z \partial \tilde{z}}(z,\tilde{z})$. The covariance structure of the process (X,Y) is hence fully described.

Lemma 1 is the cornerstone to the proof of Theorem 3; its proof is postponed to Appendix B and relies on the following proposition, of independent interest:

Proposition 1: Assume that **A1** and **A2** hold true and denote by $\mathcal{S}$ the support of the probability distribution associated to the Stieltjes transform m. Then, for every $\varepsilon > 0$, $\ell \in \mathbb{N}^*$:

$$\mathbb{P}\left(\sup_{\lambda \in \mathrm{eig}(\hat{\mathbf{R}}_{\mathbf{N}})} d(\lambda, \mathcal{S}) > \varepsilon \right) = \mathcal{O}\left(\frac{1}{N^\ell} \right) ,$$

where $d(\lambda, \mathcal{S}) = \inf_{x \in \mathcal{S}} |\lambda - x|$.

The proof of Proposition 1 is postponed to Appendix A.

As $(u_N, u'_N) \xrightarrow[N,M\to\infty]{a.s.} (\underline{m}, \underline{m}')$, a straightforward corollary of Lemma 1 yields the convergence in distribution of (X_N, X'_N, u_N, u'_N) to $(X, Y, \underline{m}, \underline{m}')$. This concludes the proof of step (i).

Consider two families of contours $(\mathcal{R}_k)$ and $(\mathcal{R}'_k)$ as described in Theorem 3. Denote by

$$\mathcal{K} = \bigcup_{k=1:L} \mathcal{R}_k \cup \bigcup_{k=1:L} \mathcal{R}'_k \ . \tag{11}$$

A direct consequence of Lemma 1 yields that $(X_N, X'_N, u_N, u'_N) : \mathcal{K} \to \mathbb{C}^4$ converges in distribution to the Gaussian process $(X, Y, \underline{m}, \underline{m}')$ with mean $(0, 0, \underline{m}, \underline{m}')$ and covariance structure inherited from the Gaussian process (X, Y). We are now in position to transfer the convergence of (X_N, X'_N, u_N, u'_N) to $\Upsilon_N(X_N, X'_N, u_N, u'_N)$ via the continuous mapping theorem, whose statement is reminded below.

Theorem 4 ([23, Th. 4.27]): For any metric spaces S_1 and S_2, let ξ, $(\xi_n)_{n\geq 1}$ be random elements in S_1 with $\xi_n \xrightarrow[n\to\infty]{\mathcal{D}} \xi$ and consider some measurable mappings f, $(f_n)_{n\geq 1}$: $S_1 \mapsto S_2$ and a measurable set $\Gamma \subset S_1$ with $\xi \in \Gamma$ a.s. such that $f_n(s_n) \to f(s)$ as $s_n \to s \in \Gamma$. Then $f_n(\xi_n) \xrightarrow[n\to\infty]{\mathcal{D}} f(\xi)$.

It remains to apply Theorem 4 to the process (X_N, X'_N, u_N, u'_N) and to the function Υ_N as defined in (9). Denote by[2]

$$\Upsilon(x, y, v, w) \ = \ \frac{1}{2\pi i c_k} \oint_{\mathcal{R}_k} z \left(\frac{\underline{m}(z) y(z) - w(z) x(z)}{\underline{m}(z) v(z)} \right) dz \ ,$$

and consider the set

$$\Gamma = \left\{ (x, y, v, w) \in C^4(\mathcal{K}, \mathbb{C}) \ , \ \inf_{\mathcal{K}} |v| > 0, x, y, w \text{ are continuous over } \mathcal{K} \right\} \ .$$

It is obvious that X, Y and $\underline{m}'$ are continuous over $\mathcal{K}$. Then, it is shown in [6, Section 9.12.1] that $\inf_{\mathcal{K}} |\underline{m}| > 0$, and, by a dominated convergence theorem argument, that $(x_N, y_N, v_N, w_N) \to (x, y, v, w) \in \Gamma$ implies that $\Upsilon_N(x_N, y_N, v_N, w_N) \to \Upsilon(x, y, v, w)$. Therefore, Theorem 4 applies to $\Upsilon_N(x_N, y_N, v_N, w_N)$ and the following convergence holds true:

$$\Upsilon_N(X_N, X'_N, u_N, u'_N) \xrightarrow[M,N\to\infty]{\mathcal{D}} \Upsilon(X, Y, \underline{m}, \underline{m}') \ ,$$

[2]As previously, we shall explicitly indicate the dependence on the contour $\mathcal{R}_k$ if needed and write $\Upsilon(x, x', u, u', \mathcal{R}_k)$.

and step (ii) is established.

It now remains to prove step (iii), *i.e.* to check the Gaussianity of the random variable $\Upsilon(X, Y, \underline{m}, \underline{m}')$ and to compute the covariance between $\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{C}_k)$ and $\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{C}_\ell)$.

In order to propagate the Gaussianity of the deviations in the integrands of (7) to the fluctuations of the integral which defines $\hat{\rho}_k$, it suffices to notice that the integral can be written as the limit of a finite Riemann sum and that a finite Riemann sum of Gaussian random variables is still Gaussian. Therefore $M(\hat{\rho}_k - \rho_k)$ converges to a real Gaussian distribution (notice that $\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_k)$, being the limiting distribution of the real random variable Υ_N, is real as well). The same argument applies to the whole vector $(M(\hat{\rho}_k - \rho_k); 1 \le k \le L)$, which hence converges toward a Gaussian vector $\boldsymbol{\Upsilon}$:

$$\begin{pmatrix} M(\hat{\rho}_1 - \rho_1) \\ \vdots \\ M(\hat{\rho}_L - \rho_L) \end{pmatrix} \xrightarrow[M,N\to\infty]{\mathcal{D}} \boldsymbol{\Upsilon} \triangleq \begin{pmatrix} \Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_1) \\ \vdots \\ \Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_L) \end{pmatrix} \sim \mathcal{N}_L(\mathbf{m}, \boldsymbol{\Theta}) \ ,$$

where $\mathbf{m}$ is a $L \times 1$ vector and $\boldsymbol{\Theta} = (\Theta_{k\ell})$ is a $L \times L$ covariance matrix.

As $\inf_{z\in\mathcal{K}} |\underline{m}(z)| > 0$, a straightforward application of Fubini's theorem together with the fact that $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ yields:

$$\mathbb{E} \oint \left(z \frac{\underline{m}'(z) X(z)}{\underline{m}^2(z)} - z \frac{Y(z)}{\underline{m}(z)} \right) dz = 0 \ ,$$

hence $\mathbf{m} = 0$.

It remains to compute the covariance between $\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_k)$ and $\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_\ell)$. As $\boldsymbol{\Theta} = \mathbb{E}(\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^T)$, write:

$$\begin{aligned} \Theta_{k\ell} &= \mathbb{E}\left(\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_k) \Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_\ell) \right) \ , \\ &\overset{(a)}{=} \mathbb{E}\left(\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_k) \Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}'_\ell) \right) \ , \\ &= -\frac{1}{4\pi^2 c_k c_l} \mathbb{E} \oint_{\mathcal{R}_k} z_1 \left(\frac{\underline{m}'(z_1) X(z_1)}{\underline{m}^2(z_1)} - \frac{Y(z_1)}{\underline{m}(z_1)} \right) dz_1 \oint_{\mathcal{R}'_\ell} z_2 \left(\frac{\underline{m}'(z_2) X(z_2)}{\underline{m}^2(z_2)} - \frac{Y(z_2)}{\underline{m}(z_2)} \right) dz_2 \ , \end{aligned}$$

where (a) follows from Remark 2 and enforces the fact that the contours are non-overlapping. Choosing non-overlapping contours will help us to compute the $\Theta_{k\ell}$'s by evaluating contour integrals with no singularities on the contours.

Write:

$$\begin{aligned} \Theta_{k\ell} &= \mathbb{E}\left(\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}_k) \Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{R}'_\ell) \right) \ , \\ &\overset{(a)}{=} -\frac{1}{4\pi^2 c_k c_l} \oint_{\mathcal{R}_k} \oint_{\mathcal{R}'_\ell} z_1 z_2 \Big(\frac{\underline{m}'(z_1) \underline{m}'(z_2) \kappa(z_1, z_2)}{\underline{m}^2(z_1) \underline{m}^2(z_2)} - \frac{\underline{m}'(z_1) \partial_2 \kappa(z_1, z_2)}{\underline{m}^2(z_1) \underline{m}(z_2)} \\ &\qquad - \frac{\underline{m}'(z_2) \partial_1 \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}^2(z_2)} + \frac{\partial^2_{12} \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}(z_2)} \Big) dz_1 dz_2 \ , \end{aligned}$$

$$\hat{\Theta}_{k\ell} = -\frac{M^2}{4\pi^2 N_k N_\ell} \oint_{\mathcal{R}_k} \oint_{\mathcal{R}'_\ell} \left(\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z_1) m'_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{(m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - m_{\underline{\hat{\mathbf{R}}}_N}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \right) \times \frac{1}{m_{\underline{\hat{\mathbf{R}}}_N}(z_1) m_{\underline{\hat{\mathbf{R}}}_N}(z_2)} d\, z_1 d\, z_2 \ . \tag{12}$$

---

where (a) follows from Lemma 1 and the fact that $\inf_{z \in \mathcal{K}} |\underline{m}(z)| > 0$ together with Fubini's theorem, and $\partial_1, \partial_2, \partial_{12}^2$ respectively stand for $\partial/\partial z_1$, $\partial/\partial z_2$ and $\partial^2/\partial z_1 \partial z_2$. The above double integral is also well-defined as $\kappa(z_1, z_2)$ is well-defined and C^∞-differentiable over $\mathcal{R}_k \times \mathcal{R}'_\ell$. By integration by parts, we obtain

$$\begin{aligned} &\oint \frac{z_1 z_2 \underline{m}'(z_2) \partial_1 \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}^2(z_2)} dz_1 \\ &= \oint \left(-\frac{z_2 \underline{m}'(z_2) \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}^2(z_2)} + \frac{z_1 z_2 \underline{m}'(z_1) \underline{m}'(z_2) \kappa(z_1, z_2)}{\underline{m}^2(z_1) \underline{m}^2(z_2)} \right) dz_1 \ . \end{aligned}$$

Similarly,

$$\begin{aligned} &\oint \frac{z_1 z_2 \underline{m}(z_2) \partial_{12} \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}^2(z_2)} dz_1 \\ &= -\oint \frac{z_2 \partial_2 \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}(z_2)} dz_1 + \oint \frac{z_1 z_2 \underline{m}'(z_1) \partial_2 \kappa(z_1, z_2)}{\underline{m}^2(z_1) \underline{m}(z_2)} dz_1 \ . \end{aligned}$$

Hence

$$\Theta_{k\ell} = -\frac{1}{4\pi^2 c_k c_l} \left\{ \oint_{\mathcal{R}_k} \oint_{\mathcal{R}'_\ell} \frac{z_2 \underline{m}'(z_2) \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}^2(z_2)} dz_1 dz_2 - \oint_{\mathcal{R}_k} \oint_{\mathcal{R}'_\ell} \frac{z_2 \partial_2 \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}(z_2)} dz_1 dz_2 \right\} \ .$$

Another integration by parts yields

$$\oint \frac{z_2 \partial_2 \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}(z_2)} dz_2 = -\oint \frac{\kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}(z_2)} dz_2 + \oint \frac{z_2 \underline{m}'(z_2) \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}^2(z_2)} dz_2 \ .$$

Finally, we obtain:

$$\Theta_{k\ell} = -\frac{1}{4\pi^2 c_k c_\ell} \oint_{\mathcal{R}_k} \oint_{\mathcal{R}'_\ell} \frac{\kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}(z_2)} d\, z_1 d\, z_2 \ ,$$

and (8) is established.

IV. Estimation of the covariance matrix

Theorem 3 describes the limiting performance of the estimator of Theorem 1, with an exact characterization of its variance. Unfortunately, the variance $\mathbf{\Theta}$ depends upon unknown quantities. We provide hereafter consistent estimates $\hat{\mathbf{\Theta}}$ for $\mathbf{\Theta}$ based on the observation $\hat{\mathbf{R}}_N$.

$$\hat{\Theta}_{k\ell} = \frac{M^2}{N_k N_\ell}\left[\sum_{(i,j)\in\mathcal{N}_k\times\mathcal{N}_\ell,\ i\neq j} \frac{-1}{(\hat{\mu}_i-\hat{\mu}_j)^2 m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i) m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_j)} + \delta_{k\ell}\sum_{i\in\mathcal{N}_k}\left(\frac{m'''_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)}{6m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)^3} - \frac{m''_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)^2}{4m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)^4}\right)\right]. \quad (13)$$

Theorem 5: Assume that Assumptions **A1** and **A2** hold true, and recall the definition of $\Theta_{k\ell}$ given in (8) and Theorem 3. Let $\hat{\Theta}_{k\ell}$ be defined by (13), where $(\mathcal{N}_k)$ and $(\hat{\mu}_k)$ are defined in Theorem 1, then:

$$\hat{\Theta}_{k\ell} - \Theta_{k\ell} \xrightarrow{a.s.} 0$$

as $N, M \to \infty$.

Theorem 5 is useful in practice as one can obtain simultaneously an estimate $\hat{\rho}_k$ of the values of ρ_k as well as an estimation of the degree of confidence for each $\hat{\rho}_k$.

Proof: In view of formula (8), taking into account the fact that $m_{\underline{\hat{\mathbf{R}}}_N}$ and $m'_{\underline{\hat{\mathbf{R}}}_N}$ are consistent estimates for $\underline{m}$ and $\underline{m}'$, it is natural to define $\hat{\Theta}_{k\ell}$ by replacing the unknown quantities $\underline{m}$ and $\underline{m}'$ in (8) by their empirical counterparts $m_{\underline{\hat{\mathbf{R}}}_N}$ and $m'_{\underline{\hat{\mathbf{R}}}_N}$, hence the definition of $\hat{\Theta}_{k\ell}$ in (12).

The proof of Theorem 5 now breaks down into two steps: The convergence of $\hat{\Theta}_{k\ell}$ to $\Theta_{k\ell}$, which relies on the definition (12) of $\hat{\Theta}_{k\ell}$ and on a dominated convergence argument, and the effective computation of the integral in (12) which relies on Cauchy's residue theorem [26], and yields (13).

We first address the convergence of $\hat{\Theta}_{k\ell}$ to $\Theta_{k\ell}$. Due to [22], [27], almost surely, the eigenvalues of $\underline{\hat{\mathbf{R}}}_N$ will eventually belong to any ε-blow-up of the support $\underline{\mathcal{S}}$ of the probability measure associated to $\underline{m}$, *i.e.* the set $\{x \in \mathbb{R} : d(x, \underline{\mathcal{S}}) < \varepsilon\}$. Hence, if ε is small enough, the distance between these eigenvalues and any $z \in \mathcal{R}_k$ will be eventually uniformly lower-bounded. By [17, Lemma 1], the same holds true for the zeros of $m_{\underline{\hat{\mathbf{R}}}_N}$ (which are real). In particular, this implies that $m_{\underline{\hat{\mathbf{R}}}_N}$ is eventually uniformly lower-bounded on $\mathcal{R}_k$ (if not, then by compacity, there would exist $z \in \mathcal{K}$ such that $m_{\underline{\hat{\mathbf{R}}}_N}(z) = 0$ which yields a contradiction because all the zeroes of $m_{\underline{\hat{\mathbf{R}}}_N}$ are strictly within any contour). With these arguments at hand, one can easily apply the dominated convergence theorem and conclude that $\hat{\Theta}_{k\ell} \to \Theta_{k\ell}$ a.s.

We now evaluate the integral (12) by computing the residues of the integrand within $\mathcal{R}_k$ and $\mathcal{R}'_\ell$. There are two cases to discuss depending on whether $k \neq \ell$ and $k = \ell$. Denote by $h(z_1, z_2)$ the integrand in (12), that is:

$$h(z_1,z_2)=\left(\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z_1)m'_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{(m_{\underline{\hat{\mathbf{R}}}_N}(z_1)-m_{\underline{\hat{\mathbf{R}}}_N}(z_2))^2}-\frac{1}{(z_1-z_2)^2}\right)\times\frac{1}{m_{\underline{\hat{\mathbf{R}}}_N}(z_1)m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}. \tag{14}$$

Note that, when z_2 is fixed, for $z_1 \to \hat{\lambda}_i$,

$$\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z_1)}{(m_{\underline{\hat{\mathbf{R}}}_N}(z_1)-m_{\underline{\hat{\mathbf{R}}}_N}(z_2))^2}\xrightarrow[z_1\to\hat{\lambda}_i]{a.s.}-M.$$

Then $h(z_1,z_2)\xrightarrow{a.s.}0$ when $z_1\to\hat{\lambda}_i$. Same result holds for $z_1\to 0$. That is to say, $\hat{\lambda}_i$ and 0 are not poles of $h(z_1,z_2)$.

To apply the residue theorem, we first consider the case where $k\neq\ell$.

In this case, the two integration contours are different and never intersect (in particular, z_1 is always different from z_2). Let z_2 be fixed, and denote by $\hat{\mu}_i$ the zeroes (labeled in increasing order) of $m_{\underline{\hat{\mathbf{R}}}_N}$, then the computation of the residue $\mathrm{Res}(h(\cdot,z_2),\hat{\mu}_i)$ of $h(\cdot,z_2)$ at a zero $\hat{\mu}_i$ of $m_{\underline{\hat{\mathbf{R}}}_N}$ which is located within $\mathcal{R}_k$ is straightforward and yields:

$$r(z_2)\stackrel{\triangle}{=}\mathrm{Res}(h(\cdot,z_2),\hat{\mu}_i)=\left(\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)\,m'_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{m^2_{\underline{\hat{\mathbf{R}}}_N}(z_2)}-\frac{1}{(\hat{\mu}_i-z_2)^2}\right)\frac{1}{m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}\ . \tag{15}$$

Similarly, if one computes $\mathrm{Res}(r,\hat{\mu}_j)$ at a zero $\hat{\mu}_j$ of $m_{\underline{\hat{\mathbf{R}}}_N}$ located within $\mathcal{R}_k$, one obtains:

$$\mathrm{Res}(r,\hat{\mu}_j)=-\frac{1}{(\hat{\mu}_i-\hat{\mu}_j)^2\,m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_j)}\ .$$

As stated in the following proposition, let $z_2\in\mathcal{R}_k\backslash\mathbb{R}$, the set $\mathcal{R}_{z_2}=\{z_1\in\mathbb{C}:z_1\neq z_2, m_{\underline{\hat{\mathbf{R}}}_N}(z_1)=m_{\underline{\hat{\mathbf{R}}}_N}(z_2)\neq 0\}$ is eventually empty a.s. for all N,M large, and if $z_2\in\mathcal{R}_k\cap\mathbb{R}$, this set is not empty, however, the integration with respect to z_2 for this residue is zero because the set $\mathcal{R}_k\cap\mathbb{R}$ only contains two points, hence the residue in this set has not to be counted.

Proposition 2: Let Assumptions **A1** and **A2** hold true, then for $z_2\in\mathcal{R}_k\backslash\mathbb{R}$,

$$\mathcal{R}_{z_2}=\{z_1\in\mathbb{C}:z_1\neq z_2, m_{\underline{\hat{\mathbf{R}}}_N}(z_1)=m_{\underline{\hat{\mathbf{R}}}_N}(z_2)\neq 0\}=\varnothing\quad a.s.$$

for all N,M large.

The proof of Proposition 2 is postponed to Appendix C.

It remains to count the number of $\hat{\mu}_i$ within each contour. By [17, Lemma 1], eventually, there are exactly as many $\hat{\mu}_i$ as eigenvalues within each contour, hence the result in the case $k\neq\ell$:

$$\hat{\Theta}_{k\ell}=\frac{M^2}{N_kN_\ell}\sum_{(i,j)\in\mathcal{N}_k\times\mathcal{N}_\ell}-\frac{1}{(\hat{\mu}_i-\hat{\mu}_j)^2m'_{\underline{\hat{\mathbf{R}}}_N}(\mu_i)m'_{\underline{\hat{\mathbf{R}}}_N}(\mu_j)}\ .$$

We now compute the integral (12) in the case where $k=\ell$, and begin by the computation of the residues at $\hat{\mu}_i$. The definition (15) of r and the computation of $\mathrm{Res}(r,\hat{\mu}_j)$ still hold true in the case where $\hat{\mu}_j$ is within $\mathcal{R}_k$ but different from $\hat{\mu}_i$. It remains to compute $\mathrm{Res}(r,\hat{\mu}_i)$. Taking $z_2 \to \mu_i$, we get:

$$\lim_{z_2\to\hat{\mu}_i}(z_2-\hat{\mu}_i)^3\left(\frac{1}{m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)m_{\hat{\underline{\mathbf{R}}}_N}(z_2)(\hat{\mu}_i-z_2)^2}\right) = \frac{1}{m'^2_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)},$$

$$\lim_{z_2\to\hat{\mu}_i}(z_2-\hat{\mu}_i)^2\left(\frac{1}{m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)m_{\hat{\underline{\mathbf{R}}}_N}(z_2)(\hat{\mu}_i-z_2)^2}-\frac{1}{{m'_{\hat{\underline{\mathbf{R}}}_N}}^2(\hat{\mu}_i)(z_2-\hat{\mu}_i)^3}\right) = -\frac{m''_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)}{2{m'_{\hat{\underline{\mathbf{R}}}_N}}^3(\hat{\mu}_i)}.$$

Finally,

$$\lim_{z_2\to\hat{\mu}_i}(z_2-\hat{\mu}_i)\left(\frac{1}{m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)m_{\hat{\underline{\mathbf{R}}}_N}(z_2)(\hat{\mu}_i-z_2)^2}-\frac{1}{m'^2_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)(z_2-\hat{\mu}_i)^3}+\frac{m''_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)}{2m'^3_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)(z_2-\hat{\mu}_i)^2}\right)$$
$$=\frac{m'''{}_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)}{6m'{}_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^3}-\frac{m''{}_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^2}{4m'{}_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^4}.$$

Hence the residue:

$$\mathrm{Res}(r,\hat{\mu}_i)=\frac{m'''_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)}{6m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^3}-\frac{m''_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^2}{4m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^4}.$$

There are two other cases that should be taken into account for the computation of the integral: The set $\mathcal{R}_{z_2}$, and the residue for $z_1=z_2$. The first case can be handled as before. For $z_1=z_2$, the calculus of $g(z_1,z_2)$ for the residue $z_1=z_2$ is exactly the same as before. It remains to compute $\frac{1}{(z_1-z_2)^2}\frac{1}{m_{\hat{\underline{\mathbf{R}}}_N}(z_1)m_{\hat{\underline{\mathbf{R}}}_N}(z_2)}$ for the residue $z_1=z_2$. The integration by parts formula yields that:

$$\oint\frac{1}{(z_1-z_2)^2}\frac{dz_1}{m_{\hat{\underline{\mathbf{R}}}_N}(z_1)m_{\hat{\underline{\mathbf{R}}}_N}(z_2)}=\oint-\frac{m'_{\hat{\underline{\mathbf{R}}}_N}(z_1)}{(z_1-z_2)}\frac{dz_1}{m^2_{\hat{\underline{\mathbf{R}}}_N}(z_1)m_{\hat{\underline{\mathbf{R}}}_N}(z_2)}.$$

Then the residue for $z_1=z_2$ is:

$$-\frac{m'_{\hat{\underline{\mathbf{R}}}_N}(z_2)}{m^3_{\hat{\underline{\mathbf{R}}}_N}(z_2)}.$$

Again, this is the derivative function of $\frac{1}{2m^2_{\hat{\underline{\mathbf{R}}}_N}(z_2)}$, then the integration is zero.

Finally both have a null contribution, hence the formula:

$$\hat{\Theta}_{kk}=\frac{M^2}{N_k^2}\left[\sum_{(i,j)\in\mathcal{N}_k^2,\ i\neq j}\frac{-1}{(\hat{\mu}_i-\hat{\mu}_j)^2m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_j)}+\sum_{i\in\mathcal{N}_k}\left(\frac{m'''_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)}{6m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^3}-\frac{m''_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^2}{4m'_{\hat{\underline{\mathbf{R}}}_N}(\hat{\mu}_i)^4}\right)\right].$$

$\blacksquare$

V. Conclusion

This article provides a central limit theorem to describe the fluctuations of consistent estimators of a large covariance matrix. The emphasis is put in the case where the dimension of each variable is comparable to the number of available samples and a key assumption is a separability condition which assesses that the number of clusters of the limiting spectral distribution is the same as the number of (population matrix) eigenvalues to be estimated. Moreover, an estimation of the limiting covariance matrix, based on the observations, is provided.

The results presented here are mainly based on large random matrix and probability theory. They are part of a recent effort to develop statistical results for large dimensional/small sample size dataset, a context of growing interest with the spectacular evolution of data acquisition, and the recent issues in sensor networks, cognitive radio, and wireless communications at large.

Appendix

A. Proof of Proposition 1

Let us first begin by considerations related to the supports of the probability distributions associated to $m(z)$ and $m_N(z)$. Denote by $\mathcal{S}$ and $\mathcal{S}_N$ these supports and recall that $\mathcal{S}$ is the union of L disjoint clusters: For $a_1 \leq b_1 < \cdots < a_L \leq b_L$,

$$\mathcal{S} = [a_1, b_1] \cup \cdots \cup [a_L, b_L] \ .$$

The following lemma clarifies the relations between $\mathcal{S}_N$ and $\mathcal{S}$.

Lemma 2: Let $N, M \to \infty$, then for N large enough, the support $\mathcal{S}_N$ of the probability distribution associated to the Stieltjes transform $m_N(z)$ is the union of L clusters: For $a_1^N \leq b_1^N < \cdots < a_L^N \leq b_L^N$,

$$\mathcal{S}_N = [a_1^N, b_1^N] \cup \cdots \cup [a_L^N, b_L^N] \ .$$

Moreover, the following convergence holds true:

$$a_\ell^N \xrightarrow[N,M\to\infty]{} a_\ell \ , \quad b_\ell^N \xrightarrow[N,M\to\infty]{} b_\ell \ ,$$

for $1 \leq \ell \leq L$.

Remark 8: If the support $\mathcal{S}_N$ contains zero, (ex: $N > M$), $a_1^N = b_1^N = 0$. By Assumption **A1**, the multiplicity N_1 corresponding to zero satisfies $\frac{N_1}{N} \to c_1 > 0$, hence zero is also in the support $\mathcal{S}$. In this case, we will get that $a_1 = b_1 = 0$, and the conclusion still holds true.

Proof of Lemma 2: Recall the relations:

$$\underline{m}_N(z) = -\left(z - \frac{N}{M}\int \frac{t}{1+t\underline{m}_N} F^{\mathbf{R}_N}(dt)\right)^{-1} \tag{16}$$

and

$$m_N(z) = \frac{M}{N}\underline{m}_N(z) - \left(1 - \frac{M}{N}\right)\frac{1}{z}. \tag{17}$$

As the Stieltjes transform of δ_0 (the Dirac mass at 0) is $-\frac{1}{z}$ and $\underline{m}_N(z)$ is a continuous function over $\mathbb{R}_+^*$, for a,b with $0 < a < b$, by the inverse formula of Stieltjes transform, one gets:

$$F_N([a,b]) = \frac{M}{N}\underline{F}_N([a,b]).$$

So it suffices to study the support $\underline{\mathbb{S}}_N$ associated to $\underline{F}_N$.

From the definition of $\underline{m}_N(z)$ (see formula (16)), we obtain:

$$z_{\mathbf{R}_N}(\underline{m}_N) = -\frac{1}{\underline{m}_N} + \frac{N}{M}\int \frac{tdF^{\mathbf{R_N}}(t)}{1+t\underline{m}_N}.$$

Denote by $B = \{m \in \mathbb{R} : m \neq 0, -m^{-1} \notin \{\rho_1, \cdots, \rho_L\}\}$. In [28, Theorem 4.1 and Theorem 4.2], Silverstein and Choi show that for a real number x, $x \in \underline{\mathbb{S}}_N^c \iff \underline{m}_x \in B$ and $z'_{\mathbf{R_N}}(\underline{m}_x) = \frac{1}{\underline{m}_x^2} - \frac{N}{M}\int \frac{t^2 dF^{\mathbf{R_N}}(t)}{(1+t\underline{m}_x)^2} > 0$ with $\underline{m}_N(x) = \underline{m}_x$ and $z_{\mathbf{R_N}}(\underline{m}_x) = x$.

Then if $a \in \partial\underline{\mathbb{S}}_N$, $m_a \notin B$ or $z'_{\mathbf{R_N}}(m_a) \leq 0$ with $m_a = \underline{m}_N(a)$. Now we will show that $m_a \in B$. In [28, Theorem 5.1], $m_a \neq 0$. If $-m_a^{-1} \in S_{F^{\mathbf{R_N}}}$, as $F^{\mathbf{R_N}}$ is discrete, we get that $\lim_{m\to m_a}\int \frac{t^2 dF^{\mathbf{R_N}}(t)}{(1+tm)^2} \longrightarrow \infty$. So on the neighborhood to the left and to the right of m_a, $z'_{\mathbf{R_N}} < 0$ which contradicts [28, Theorem 5.1].

Hence $z'_{\mathbf{R_N}}(m_a) \leq 0$. By the continuity, we get

$$z'_{\mathbf{R_N}}(m_a) = \frac{1}{m_a^2} - \frac{N}{M}\int \frac{t^2 dF^{\mathbf{R}_N}(t)}{(1+tm_a)^2} = 0.$$

This is equivalent to the following equation:

$$z'_{\mathbf{R_N}}(m_a) = \frac{1}{m_a^2} - \frac{1}{M}\sum_{i=1}^{L} N_i \frac{\rho_i^2}{(1+\rho_i m_a)^2} = 0. \tag{18}$$

By multiplying the common denominator, one gets a polynomial of the degree 2L in m_a. Let us now prove that these 2L roots are real. At first, note that:

$$\frac{1}{m^2} - \frac{N}{M}\int \frac{t^2 dF^{\mathbf{R}_N}(t)}{(1+tm)^2} \xrightarrow[m\to -\frac{1}{\rho_i}]{} -\infty,$$

and

$$z''_{\mathbf{R_N}}(m) = -\frac{2}{m^3} + \frac{N}{M}\int \frac{2t^3 dF^{\mathbf{R}_N}(t)}{(1+tm)^3}.$$

So $z''_{\mathbf{R_N}}(m)$ has one and only one zero in the open set $(-\frac{1}{\rho_i}, -\frac{1}{\rho_{i+1}})$ for $1 \leq i \leq L-1$. Then for $\beta_i \in (-\frac{1}{\rho_i}, -\frac{1}{\rho_{i+1}})$ such that $z''_{\mathbf{R_N}}(\beta_i) = 0$, it suffices to show that $z'_{\mathbf{R_N}}(\beta_i) > 0$ in order to prove that there will be two zeros for $z'_{\mathbf{R_N}}(m)$ in the set $(-\frac{1}{\rho_i}, -\frac{1}{\rho_{i+1}})$. From the separability condition (cf. Assumption **A2**), $\inf_N\{\frac{M}{N} - \Psi_N(i)\} > 0$, and

$$
\begin{aligned}
z'_{\mathbf{R_N}}\left(-\frac{1}{\alpha_i}\right) &= \alpha_i^2 - \frac{N}{M}\int \frac{t^2 dF^{\mathbf{R_N}}(t)}{(1-\frac{t}{\alpha_i})^2} \ , \\
&= \alpha_i^2 \left(1 - \frac{1}{M}\sum_{r=1}^{L} N_i \frac{\rho_i^2}{(\alpha_i - \rho_i)^2}\right) > 0 \ .
\end{aligned}
$$

Thus we obtain $2(L-1)$ roots. Besides, in the open set $(-\rho_L^{-1}, 0)$,

$$
\frac{1}{m^2} - \frac{N}{M}\int \frac{t^2 dF^{\mathbf{R}_N}(t)}{(1+tm)^2} \xrightarrow[m_a \to 0^-]{} +\infty,
$$

there exists another root in this set. In the open set $(-\infty, -\rho_1^{-1})$,

$$
\frac{1}{m^2} - \frac{N}{M}\int \frac{t^2 dF^{\mathbf{R}_N}(t)}{(1+tm)^2} \xrightarrow[m \to -\infty]{} 0
$$

and

$$
\frac{1}{m^2} - \frac{N}{M}\int \frac{t^2 dF^{\mathbf{R}_N}(t)}{(1+tm)^2} \underset{m \to -\infty}{\sim} \frac{1}{m^2}(1 - \frac{L}{M}) > 0.
$$

Hence the last root in this open set. This proves that $\mathcal{S}_N = [a_1^N, b_1^N] \cup \cdots \cup [a_L^N, b_L^N]$.

To prove $a_\ell^N \xrightarrow[N,M\to\infty]{} a_\ell$ and $b_\ell^N \xrightarrow[N,M\to\infty]{} b_\ell$, note that a_i b_i satisfy the same type of equation by replacing $\frac{N}{M}$ by c and $F^{\mathbf{R_N}}$ by F^R. As $\frac{N}{M} \to c$ and $\frac{K_i}{M} \to c_i$, the roots of Eq. (18) converge to those of the limiting equation (see [29] for instance). Hence the conclusion.

∎

We are now in position to establish the proof of Proposition 1.

Denote by $\mathcal{S}(\varepsilon)$ the ε-blow-up of $\mathcal{S}$, i.e. $\mathcal{S}(\varepsilon) = \{x \in \mathbb{R},\ d(x, S) < \varepsilon\}$. Let $\varepsilon > 0$ be small enough and consider a smooth function ϕ equal to zero on $\mathcal{S}(\varepsilon/3)$, equal to 1 if $x \notin \mathcal{S}(\varepsilon)$, equal to zero again if $|x| \geq \tau$ (as we shall see, τ will be chosen to be large), and smooth in-between with $0 \leq \phi \leq 1$:

$$
\phi(x) = \begin{cases} 0 & \text{if } d(x, \mathcal{S}) < \varepsilon/3 \ , \\ 1 & \text{if } d(x, \mathcal{S}) > \varepsilon \ , |x| \leq \tau - \epsilon \\ 0 & \text{if } |x| > \tau \ . \end{cases}
$$

Notice that if $N, M \to \infty$ and N is large enough, then by Lemma 2, $\phi(x) = 0$ for all $x \in \mathcal{S}_N$. Now if $\mathbf{Z}$ is a $M \times M$ hermitian matrix with spectral decomposition $\mathbf{Z} = \mathbf{U}\,\mathrm{diag}\,(\gamma_i;\ 1 \le i \le M))\ \mathbf{U}^H$, where $\mathbf{U}$ is unitary and $(\gamma_i;\ 1 \le i \le M) = \mathrm{eig}(\mathbf{Z})$, write $\phi(\mathbf{Z}) = \mathbf{U}\,\mathrm{diag}\,(\phi(\gamma_i);\ 1 \le i \le M))\ \mathbf{U}^H$.

We have:

$$\begin{aligned}
\mathbb{P}(\sup_n d(\lambda_n, S) > \varepsilon) &\le \mathbb{P}(\|\hat{\mathbf{R}}_N\| > \tau - \varepsilon) \ + \ \mathbb{P}(\mathrm{Tr}\,\phi(\hat{\mathbf{R}}_N) \ge 1) \\
&= \mathbb{P}(\|\hat{\mathbf{R}}_N\| > \tau - \varepsilon) \ + \ \mathbb{P}([\mathrm{Tr}\,\phi(\hat{\mathbf{R}}_N)]^p \ge 1) \\
&\overset{(a)}{\le} \mathbb{P}(\|\hat{\mathbf{R}}_N\| > \tau - \varepsilon) \ + \ \mathbb{E}[\mathrm{Tr}\,\phi(\hat{\mathbf{R}}_N)]^p \ ,
\end{aligned}$$

for every $p \ge 1$, where (a) follows from Markov's inequality. The fact that $\mathbb{P}(\|\hat{\mathbf{R}}_N\| > \tau) = \mathcal{O}(N^{-\ell})$ for τ large enough and every $\ell \in \mathbb{N}^*$ is well-known (see for instance [6, Section 9.7]). We shall therefore establish estimates over $\mathbb{E}[\mathrm{Tr}\,\phi(\hat{\mathbf{R}}_N)]^p$. Take $p = 2^k$; we prove the following statement by induction: For $k \ge 1$ and for every integer $\beta < 2^k$ and for every smooth function f with compact support whose value on $\mathcal{S}(\varepsilon/3)$ is zero ,

$$\mathbb{E}\left(\mathrm{Tr} f(\hat{\mathbf{R}}_N)\right)^{2^k} \quad = \quad \mathcal{O}\left(\frac{1}{N^\beta}\right) \ .$$

First notice that, due to Lemma 2, $\int_{\mathcal{S}_N} f(\lambda)\, F_N(d\lambda) = 0$ (where F_N is the probability distribution associated to m_N) for N, M large enough ($N, M \to \infty$). A minor modification of [30, Lemma 2] (whose model is slightly different) with the help of [31, Proposition 5] yields that for $N, M \to \infty$ and N large enough, $\mathbb{E}\,\mathrm{Tr}\, f(\hat{\mathbf{R}}_N) = \mathcal{O}(N^{-1})$, and the property is verified for $k = 0$.

Let $k > 0$ be fixed and assume that the result holds true for $\beta < 2^k$. We want to show that $\mathbb{E}[\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}} = \mathcal{O}(N^{-2\beta})$. At step $k+1$, the expectation writes:

$$\begin{aligned}
&\left|\mathbb{E}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}\right| \qquad (19)\\
&= \left|\mathbb{E}\left([\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^k} + \mathbb{E}[\mathrm{tr} f(\hat{\mathbf{R}}_N)]^{2^k} - \mathbb{E}[\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^k}\right)^2\right| \\
&\le 2\left(\mathrm{Var}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^k} + |\mathbb{E}[\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^k}|^2\right). \qquad (20)
\end{aligned}$$

The second term of the right hand side (r.h.s.) of the equation can be handled by the induction hypothesis:

$$\left|\mathbb{E}[\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^k}\right|^2 = \mathcal{O}\left(\frac{1}{N^{2\beta}}\right).$$

We now rely on Poincaré-Nash inequality (see for instance [31, Section II-B]) to handle the first term of the r.h.s. Applying this inequality, we obtain:

$$\mathrm{Var}\left((\mathrm{Tr}\, f(\hat{\mathbf{R}}_N))^{2^k}\right) \quad \le \quad K \sum_{i,j} \mathbb{E}\left[\left|\frac{\partial [\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^k}}{\partial Y_{i,j}}\right|^2 + \left|\frac{\partial [\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^k}}{\partial \overline{Y}_{i,j}}\right|^2\right] \ , \qquad (21)$$

where K is a constant which does not depend on N, M and which is greater than $\mathbf{R}_N$'s eigenvalues. In order to compute the derivatives of the r.h.s., we rely on [32, Lemma 4.6]. This yields:

$$\begin{aligned}\frac{\partial}{\partial Y_{i,j}}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^k} &= \frac{2^k}{M}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^k-1}[\mathbf{Y}_N^* f'(\hat{\mathbf{R}}_N)]_{j,i}\ ,\\ \frac{\partial}{\partial \overline{Y_{i,j}}}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^k} &= \frac{2^k}{M}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^k-1}[f'(\hat{\mathbf{R}}_N)\mathbf{Y}_N]_{i,j}\ .\end{aligned}$$

Plugging these derivatives into (21), we obtain:

$$\begin{aligned}&\operatorname{Var}(\operatorname{Tr}[f(\hat{\mathbf{R}}_N)]^{2^k})\\ &\le \frac{K\,2^{2k+1}}{M^2}\mathbb{E}\left[(\operatorname{Tr} f(\hat{\mathbf{R}}_N))^{(2^{k+1}-2)}\operatorname{Tr}\left(f'(\hat{\mathbf{R}}_N)\mathbf{Y}_N\mathbf{Y}_N^* f'(\hat{\mathbf{R}}_N)\right)\right]\ ,\\ &= \frac{K\,2^{2k+1}}{M}\mathbb{E}\left[(\operatorname{Tr} f(\hat{\mathbf{R}}_N))^{(2^{k+1}-2)}\operatorname{Tr}\left(f'(\hat{\mathbf{R}}_N)^2\hat{\mathbf{R}}_N\right)\right]\ ,\\ &\le \frac{K\,2^{2k+1}}{M}\left|\mathbb{E}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^{k+1}}\right|^{\frac{2^{k+1}-2}{2^{k+1}}}\times\left|\mathbb{E}[\operatorname{Tr} f'(\hat{\mathbf{R}}_N)^2\hat{\mathbf{R}}_N]^{2^k}\right|^{\frac{1}{2^k}}\ ,\end{aligned}$$

where the last inequality is a consequence of Hölder's inequality.

As the function $h(\lambda)=\lambda[f'(\lambda)]^2$ satisfies the induction hypothesis, we have for every $\alpha<1$:

$$\left|\mathbb{E}\operatorname{Tr}[f'(\hat{\mathbf{R}}_N)^2\hat{\mathbf{R}}_N]^{2^k}\right|^{\frac{1}{2^k}} = \mathcal{O}(N^{-\alpha}).$$

Plugging this estimate into (19), we obtain:

$$\left|\mathbb{E}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}\right| \le K\left(\frac{1}{N^{1+\alpha}}|\mathbb{E}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}|^{\frac{2^{k+1}-2}{2^{k+1}}}\right)+\mathcal{O}(N^{-2\beta})\ , \tag{22}$$

where K is a constant independent of M, N, k. Notice that inequality (22) involves twice the quantity of interest $\mathbb{E}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}$ that we want to upper bound by $O(N^{-2\beta})$. We shall proceed iteratively.

Notice that $\operatorname{Tr}[f(\hat{\mathbf{R}}_N)] \le \sup_{x\in\mathbb{R}}|f(x)|\times N$ because f is bounded on $\mathbb{R}$; hence the rough estimate:

$$\mathbb{E}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}} = \mathcal{O}(N^{2^{k+1}}).$$

Plugging this into (22) yields:

$$\mathbb{E}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}} = \mathcal{O}(N^{a_1})\ ,$$

where $a_0 = 2^{k+1}$ and $a_1 = a_0\frac{2^{k+1}-2}{2^{k+1}} - (1+\alpha)$. Iterating the procedure, we obtain:

$$\mathbb{E}[\operatorname{Tr} f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}} = \mathcal{O}\left(N^{a_\ell\vee(-2\beta)}\right)\ ,$$

where $a_\ell = a_{\ell-1}\frac{2^{k+1}-2}{2^{k+1}} - (1+\alpha)$ and $x\vee y$ stands for $\sup(x,y)$. Now, in order to conclude the proof, it remains to prove that i) the sequence (a_ℓ) converges to some limit a_∞, ii) for some well-chosen $\alpha<1$, $a_\infty\in(-2^{k+1},-2\beta)$. Write:

$$a_{\ell+1}+2^k(1+\alpha) = \frac{2^k-1}{2^k}(a_\ell+2^k(1+\alpha))\ ,$$

hence a_ℓ converges to $-2^k(1+\alpha)$ which readily belongs to $(-2^{k+1}, -2\beta)$ for a well-chosen $\alpha \in (0,1)$. Finally $\mathbb{E}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}} = \mathcal{O}(N^{-2\beta})$ which ends the induction.

It remains to apply this estimate to $\mathbb{E}[\mathrm{Tr}\, \phi(\hat{\mathbf{R}}_N)]^\ell$ in order to get the desired result.

B. *Proof of Lemma 1*

Notice that $X_N(z) = M(m_{\hat{\underline{\mathbf{R}}}_N} - \underline{m}_N) = \overline{X_N(\overline{z})}$ for $z \in \overline{\mathbb{C}^+}$. So it suffices to verify the arguments for $z \in \mathbb{C}^+$. As $\frac{1}{\hat{\rho}_k - z}$ can converge to infinity if z is close to the real axis, the process $X_N(z)$ might be large when z is close to the real axis. Thus we begin the proof by considering a truncated version of the process X_N. More precisely, let ε_N be a real sequence decreasing to zero satisfying for some $\delta \in]0,1[$:

$$\varepsilon_N \geq N^{-\delta} \ .$$

With the same notations as in Lemma 2, denote by $\mathcal{S} = [a_1, b_1] \cup \cdots \cup [a_L, b_L]$; and take p_k, q_k such that $b_{k-1} < p_k < a_k$ and $b_k < q_k < a_{k+1}$ for $1 \leq k \leq L$ with conventions $b_0 = 0$ and $a_{L+1} = \infty$, $i.e.$ $[p_k, q_k]$ only contains the k-th cluster. Let $d > 0$. Consider:

$$\begin{aligned}
\mathcal{R}_{k,1} &= \{x + \mathbf{i}d : x \in [p_k, q_k]\} \ , \\
\mathcal{R}_{k,2} &= \left\{p_k + \mathbf{i}v : v \in \left[\frac{\varepsilon_N}{N}, d\right]\right\} \ , \\
\mathcal{R}_{k,3} &= \left\{q_k + \mathbf{i}v : v \in \left[\frac{\varepsilon_N}{N}, d\right]\right\} \ ,
\end{aligned}$$

and let $\tilde{\mathcal{R}}_k = \mathcal{R}_{k,1} \cup \mathcal{R}_{k,2} \cup \mathcal{R}_{k,3}$. The process $\hat{X}_N(\cdot)$ is defined by

$$\hat{X}_N(z) = \begin{cases} X_N(z) & \text{for } z \in \tilde{R}_k, \\ X_N(p_k + \mathbf{i}\frac{\varepsilon_N}{N}) & \text{for } x = p_k, v \in [0, \frac{\varepsilon_N}{N}], \\ X_N(q_k + \mathbf{i}\frac{\varepsilon_N}{N}) & \text{for } x = q_k, v \in [0, \frac{\varepsilon_N}{N}]. \end{cases}$$

This partition of $\tilde{\mathcal{R}}_k$ is identical to that used in [25, Section 1]. With probability one (see [22] and [27]), for all $\epsilon > 0$,

$$\lim_N \sup_{\lambda \in \mathrm{eig}(\hat{\mathbf{R}}_N)} d(\lambda, \mathcal{S}_N) < \epsilon$$

with $d(x, S)$ the Euclidean distance of x to the set S. Notice that:

$$\left| \oint_{\mathcal{R}_k} \Big(X_N(z) - \hat{X}_N(z) \Big) dz \right| \leq 2 \int_0^{\frac{\varepsilon_N}{N}} \left| X_N(p_k + \mathbf{i}x) - \hat{X}_N(p_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| + \left| X_N(q_k + \mathbf{i}x) - \hat{X}_N(q_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| dx.$$

Furthermore, with probability one, for all N large,

$$\int_0^{\frac{\varepsilon_N}{N}} \left| m_{\hat{\underline{\mathbf{R}}}_N}(p_k + \mathbf{i}x) - m_{\hat{\underline{\mathbf{R}}}_N}(p_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| dx \leq \frac{\varepsilon_N}{N} (|p_k - a_k|^{-1} + |p_k - b_{k-1}|^{-1}),$$

and

$$\int_0^{\frac{\varepsilon_N}{N}} \left| \underline{m}_N(p_k + \mathbf{i}x) - \underline{m}_N(p_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| dx \leq 2K\frac{\varepsilon_N}{N}$$

where $K = \sup_{z \in \mathcal{R}_k} |\underline{m}_N(z)|$. Thus, with probability one,

$$\left| \oint \left(X_N(z) - \hat{X}_N(z) \right) dz \right| \leq K_1 \varepsilon_N,$$

where K_1 is a constant which does not depend on N. A similar result can be achieved for the derivative functions $X'_N(z)$ and $\hat{X}'_N(z)$. One can get:

$$\left| \oint_{\mathcal{R}_k} \left(X'_N(z) - \hat{X}'_N(z) \right) dz \right| \leq 2 \int_0^{\frac{\varepsilon_N}{N}} \left| X'_N(p_k + \mathbf{i}x) - \hat{X}'_N(p_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| + \left| X'_N(q_k + \mathbf{i}x) - \hat{X}'_N(q_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| dx.$$

With probability one, for all N large,

$$\int_0^{\frac{\varepsilon_N}{N}} \left| m'_{\underline{\hat{\mathbf{R}}}_N}(p_k + \mathbf{i}x) - m'_{\underline{\hat{\mathbf{R}}}_N}(p_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| dx \leq \frac{\varepsilon_N}{N}(|p_k - a_k|^{-2} + |p_k - b_{k-1}|^{-2})$$

and

$$\int_0^{\frac{\varepsilon_N}{N}} \left| \underline{m}'_N(p_k + \mathbf{i}x) - \underline{m}'_N(p_k + \mathbf{i}\frac{\varepsilon_N}{N}) \right| dx \leq 2K'\frac{\varepsilon_N}{N}$$

where $K' = \sup_{z \in \mathcal{R}_k} |\underline{m}'_N(z)|$. So with probability one, for all N large,

$$\left| \oint_{\mathcal{R}_k} \left(X_N(z) - \hat{X}_N(z) \right) dz \right| \leq K_1 \varepsilon_N, \tag{23}$$

$$\left| \oint_{\mathcal{R}_k} \left(X'_N(z) - \hat{X}'_N(z) \right) dz \right| \leq K_2 \varepsilon_N \tag{24}$$

for some constants K_1 and K_2. Both terms converge to zero as $M \to \infty$. Then, by Slutsky's lemma [33], it suffices to establish the arguments for $\hat{X}_N(z)$ and $\hat{X}'_N(z)$.

As mentionned in Section III, there are two conditions to prove (see for instance Billingsley [24, Theorem 13.1]) to establish the convergence in distribution of the process $(\hat{X}_N, \hat{X}'_N)$ to the process (X, Y) over the compact :

- Finite-dimensional convergence of the process $(\hat{X}_N, \hat{X}'_N)$ over the compact $\mathcal{K}$.
- Tightness on the compact $\mathcal{K}$.

1) Finite-dimensional convergence: In [25], Bai and Silverstein establish a central limit theorem for $F^{\hat{\mathbf{R}}_N}$ with the complex Gaussian entries X_{ij}. We recall below their main result.

Proposition 3 (cf. [25]): With the notations introduced in Section II, for $f_1, \ldots, f_p$, analytic on an open region containing $\mathbb{R}$,

1) $\left(N \int f_i(x) d(F^{\hat{\mathbf{R}}_N} - F_N)(x) \right)_{1 \leq i \leq p}$ forms a tight sequence on N,

2)
$$\left(N\int f_i(x)d(F^{\hat{\mathbf{R}}_N}-F_N)(x)\right)_{1\leq i\leq p}\xrightarrow{\mathcal{D}}\mathcal{N}_p(0,\mathbf{V}),$$
where $\mathbf{V}=(V_{ij})$ and
$$V_{ij}=-\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}f_i(z_1)f_j(z_2)v_{ij}(z_1,z_2)dz_1dz_2,$$
with
$$v_{ij}(z_1,z_2)=\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}-\frac{1}{(z_1-z_2)^2}$$
where the integration is over positively oriented contours $\mathcal{C}_1$ and $\mathcal{C}_2$ which are supposed to be non-overlapping and both circle around the support $\mathcal{S}$.

Now we apply this proposition to establish the finite-dimensional convergence. For all $z_i\in\mathcal{C}_k\backslash\mathbb{R}$, note that
$$m_{\hat{\mathbf{R}}_N}(z)-m_N(z)=\frac{1}{2i\pi}\oint\frac{1}{x-z}d(F^{\hat{\mathbf{R}}_N}-F_N)(x)$$
with the contour which contains the support $\mathcal{S}$ and $X_N(z)=M(m_{\underline{\hat{\mathbf{R}}}_N}(z)-\underline{m}_N(z))$. Then Proposition 3 directly implies that for every finite $p\in\mathbb{N}$, the random vector
$$\left(\hat{X}_N(z_1),\hat{X}'_N(z_1),\cdots,\hat{X}_N(z_p),\hat{X}'_N(z_p)\right)$$
converges to a centered Gaussian vector by considering the functions:
$$\left(f_1(x)=\frac{1}{x-z_1},f_2(x)=\frac{1}{(x-z_1)^2},\cdots,f_{2p-1}(x)=\frac{1}{x-z_p},f_{2p}(x)=\frac{1}{(x-z_p)^2}\right).$$
Hence the finite dimensional convergence.

The proof of the tightness is based on Poincaré-Nash inequality (see for instance [30] and [31]). In Appendix A, it is proved that for all $\epsilon>0$ and all $\ell\in\mathbb{N}$,
$$\mathbb{P}\left(\sup_{\lambda\in\mathrm{eig}(\hat{\mathbf{R}}_N)}d(\lambda,\mathcal{S})>\epsilon\right)=o(N^{-\ell}).$$
Following the same idea as Bai and Silverstein [25, Section 3 and 4], it is indeed a tight sequence. The details of the proof are in Appendix B2. Thus Lemma 1 is proved.

2) Tightness: We will show the tightness of the sequence $M(m_{\underline{\hat{\mathbf{R}}}_N}-\underline{m}_N)$ and $M(m'_{\underline{\hat{\mathbf{R}}}_N}-\underline{m}'_N)$ by using Poincaré-Nash's inequality [31] on the compact $\mathcal{K}$. As the compact $\mathcal{K}$ is the union of $2L$ contours $\mathcal{R}_k$ and $\mathcal{R}'_\ell$, it is sufficient to prove the tightness on every contour $\mathcal{R}_k$ (or equivalently $\tilde{\mathcal{R}}_k$). First, denote by $M(m_{\underline{\hat{\mathbf{R}}}_N}(z)-\underline{m}_N(z))=X_N^1(z)+X_N^2(z)$ with $X_N^1(z)=M(m_{\underline{\hat{\mathbf{R}}}_N}(z)-\mathbb{E}[m_{\underline{\hat{\mathbf{R}}}_N}(z)])$ and $X_N^2(z)=M(\mathbb{E}[m_{\underline{\hat{\mathbf{R}}}_N}(z)]-\underline{m}_N(z))$.

We now prove tightness based on [24, Theorem 13.1], i.e.

1) Tightness at any point of the contour (here $\tilde{\mathcal{R}}_k$).
2) Satisfaction of the condition

$$\sup_{N, z_1, z_2 \in \tilde{\mathcal{R}}_k} \frac{\mathbb{E}|(\hat{X}_N^1(z_1) - \hat{X}_N^1(z_2))|^2}{|z_1 - z_2|^2} \leq K.$$

Condition 1) is achieved by an immediate application of Proposition 3. We now verify the second condition.

We evaluate $\frac{\mathbb{E}|(\hat{X}_N^1(z_1) - \hat{X}_N^1(z_2))|^2}{|z_1-z_2|^2}$. Denote by $\hat{\lambda}'_1 \leq \cdots \leq \hat{\lambda}'_M$, the eigenvalues of $\underline{\hat{\mathbf{R}}}_N$. Note that

$$\begin{aligned} m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - m_{\underline{\hat{\mathbf{R}}}_N}(z_2) &= \frac{z_1 - z_2}{M} \sum_{i=1}^{M} \frac{1}{(\hat{\lambda}'_i - z_1)(\hat{\lambda}'_i - z_2)} \\ &= \frac{z_1 - z_2}{M} \mathrm{Tr}(\mathbf{D}_N^{-1}(z_1)\mathbf{D}_N^{-1}(z_2)) \end{aligned}$$

with $\mathbf{D}_N(z) = \underline{\hat{\mathbf{R}}}_N - z\mathbf{I}_M$. We have

$$\begin{aligned} &\frac{\partial}{\partial Y_{i,j}} \left(\frac{m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{z_1 - z_2} \right) \\ &= \frac{\partial}{\partial Y_{i,j}} \mathrm{Tr}(\underline{\hat{\mathbf{R}}}_N - z_1\mathbf{I}_M)^{-1}(\underline{\hat{\mathbf{R}}}_N - z_2\mathbf{I}_M)^{-1} \\ &= \frac{1}{M}\left[-\mathbf{Y}_N^*\mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-1}(z_2) - \mathbf{Y}_N^*\mathbf{D}_N^{-1}(z_1)\mathbf{D}_N^{-2}(z_2) \right]_{j,i}, \end{aligned}$$

and

$$\begin{aligned} &\frac{\partial}{\partial \bar{Y}_{i,j}} \left(\frac{m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{z_1 - z_2} \right) \\ &= \frac{1}{M}[-\mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-1}(z_2)\mathbf{Y}_N - \mathbf{D}_N^{-1}(z_1)\mathbf{D}_N^{-2}(z_2)\mathbf{Y}_N]_{i,j}. \end{aligned}$$

Then by the Poincaré-Nash inequality and the fact that $\underline{\hat{\mathbf{R}}}_N$ is uniformly bounded in spectral norm almost surely, one gets

$$\begin{aligned} \frac{\mathbb{E}|\hat{X}_N^1(z_1) - \hat{X}_N^1(z_2)|^2}{|z_1 - z_2|^2} &\leq \frac{C_1}{N}\mathbb{E}\Big[\mathrm{Tr}(\mathbf{L}_{N,1})\Big] \\ &= \frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_{N,1})\mathrm{I}_{\sup_n d(\hat{\lambda}'_n, \underline{\mathcal{S}}) \leq \varepsilon}) + \frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_{N,1})\mathrm{I}_{\sup_n d(\hat{\lambda}'_n, \underline{\mathcal{S}}) > \varepsilon}) \end{aligned}$$

with

$$\mathbf{L}_{N,1} = \underline{\hat{\mathbf{R}}}_N\mathbf{D}_N^{-4}(z_1)\mathbf{D}_N^{-2}(z_2) + 2\underline{\hat{\mathbf{R}}}_N\mathbf{D}_N^{-3}(z_1)\mathbf{D}_N^{-3}(z_2) + \underline{\hat{\mathbf{R}}}_N\mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-4}(z_2)$$

and C_1 a constant which does not depend on N or M. For the first term, $\mathrm{Tr}(\mathbf{L}_{N,1})$ is bounded on the set $\sup_n d(\hat{\lambda}'_n, \underline{\mathcal{S}}) \leq \varepsilon$. For the second term, since for all $i \in \mathbb{N}$ and all $z \in \tilde{\mathcal{R}}_k$, $\frac{1}{|\hat{\lambda}'_n - z|^i} \leq \frac{N^i}{\varepsilon_N^i}$, it leads to

$$\sum_{n=1}^{N} \frac{1}{|\hat{\lambda}'_n - z|^i} \leq \frac{N^{i+1}}{\varepsilon_N^i}.$$

Then

$$|\mathrm{Tr}(\mathbf{L}_{N,1})| \leq \mathcal{O}\left(\frac{N^7}{\varepsilon_N^6}\right).$$

As $\mathbb{P}(\sup d(\hat{\lambda}'_n, \underline{\mathcal{S}}) \geq \varepsilon) = \mathbb{P}(\sup d(\hat{\lambda}_n, \mathcal{S}) \geq \varepsilon) = o(N^{-16})$, take $\varepsilon_N = N^{-0.01}$, one obtains

$$\begin{aligned}\left|\mathbb{E}(\mathrm{Tr}(\mathbf{L}_{N,1})\mathrm{I}_{\sup_n d(\hat{\lambda}'_n, \underline{\mathcal{S}}) > \varepsilon})\right| &\leq \mathbb{E}\left|\mathrm{Tr}(\mathbf{L}_{N,1})I_{\sup d(\hat{\lambda}'_n, \underline{\mathcal{S}}) > \varepsilon}\right| \\ &\leq \mathcal{O}\left(\frac{N^7}{\epsilon_N^6}\mathbb{P}(\sup d(\hat{\lambda}'_n, \underline{\mathcal{S}}) > \varepsilon)\right) \\ &\leq \mathcal{O}\left(N^{7-0.06-16}\right) \to 0.\end{aligned}$$

The second condition to establish the tightness is achieved.

For $\hat{X}_N^2(z)$, following exactly the same method as in [6, Section 9.11], one can prove that $\hat{X}_N^2(z)$ is bounded and forms an equicontinuous family that converges to 0. Hence the tightness for $M(m_{\hat{\underline{\mathbf{R}}}_N}(z) - \underline{m}_N(z))$.

The next step is to prove the tightness of $M(m'_{\hat{\underline{\mathbf{R}}}_N}(z) - \underline{m}'_N(z))$. We have

$$\begin{aligned}&m'_{\hat{\underline{\mathbf{R}}}_N}(z_1) - m'_{\hat{\underline{\mathbf{R}}}_N}(z_2) \\ &= \frac{z_1 - z_2}{M}\sum_{i=1}^{M}\frac{2\hat{\lambda}'_i - z_1 - z_2}{(\hat{\lambda}'_i - z_1)^2(\hat{\lambda}'_i - z_2)^2} \\ &= \frac{z_1 - z_2}{M}\mathrm{Tr}\left(\mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-2}(z_2)(\mathbf{D}_N(z_1) + \mathbf{D}_N(z_2))\right).\end{aligned}$$

Following the same method as derived before, one obtains

$$\begin{aligned}&\frac{\partial}{\partial Y_{ij}}\mathbf{D}_N^{-1}(z_1)\mathbf{D}_N^{-2}(z_2) \\ &= -\frac{1}{M}\left[\mathbf{Y}_N^*\mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-2}(z_2) + 2\mathbf{Y}_N^*\mathbf{D}_N^{-1}(z_1)\mathbf{D}_N^{-3}(z_2)\right]_{j,i},\end{aligned}$$

and

$$\left|\frac{\partial}{\partial Y_{ij}}\mathrm{Tr}\mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-2}(z_2)(\mathbf{D}(z_1) + \mathbf{D}_N(z_2))\right|^2 = \frac{1}{M}\mathrm{Tr}(\mathbf{L_2})$$

with

$$\begin{aligned}\mathbf{L}_{N,2} =& 4\hat{\underline{\mathbf{R}}}_N\Big(3\mathbf{D}_N^{-4}(z_1)\mathbf{D}_N^{-4}(z_2) + 2\mathbf{D}_N^{-3}(z_1)\mathbf{D}_N^{-5}(z_2) + 2\mathbf{D}_N^{-5}(z_1)\mathbf{D}_N^{-3}(z_2) \\ &+ \mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-6}(z_2) + \mathbf{D}_N^{-6}(z_1)\mathbf{D}_N^{-2}(z_2)\Big).\end{aligned}$$

Then Poincaré-Nash inequality yields that

$$\begin{aligned}&\mathrm{Var}\frac{|\hat{X}_N^{1'}(z_1)-\hat{X}_N^{1'}(z_2)|}{|z_1-z_2|}\\&\le \frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_{N,2})\mathrm{I}_{\sup_n d(\hat{\lambda}'_n,\underline{\mathcal{S}})\le\varepsilon})+\frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_{N,2})\mathrm{I}_{\sup_n d(\hat{\lambda}'_n,\underline{\mathcal{S}})>\varepsilon})\end{aligned}$$

with C_1 the same constant defined as before. The term $\mathrm{Tr}(\mathbf{L}_{N,2})$ is bounded on the set $\sup d(\hat{\lambda}'_n,\underline{\mathcal{S}})\le\varepsilon$. For the second term, $|\mathrm{Tr}(\mathbf{L}_{N,2})|\le\mathcal{O}\left(\frac{N^9}{\varepsilon_N^8}\right)$. As $\mathbb{P}(\sup d(\hat{\lambda}'_n,\underline{\mathcal{S}})\ge\varepsilon)=o(N^{-16})$ and $\varepsilon_N=N^{-0.01}$,

$$\left|\mathbb{E}(\mathrm{Tr}(\mathbf{L}_{N,2})\mathrm{I}_{\sup_n d(\hat{\lambda}'_n,\underline{\mathcal{S}})>\varepsilon})\right|\to 0.$$

The proof of the tightness of $\hat{X}_N^{1'}(z)$ is achieved as before.

The proof of the tightness is completed if one verifies that $\hat{X}_N^{2'}(z)$ for $z\in\tilde{\mathcal{R}}_k$ is bounded and forms an equicontinuous family, and converges to 0. We will use the same method for the process $\hat{X}_N^2(z)$ (see [6, Section 9.11]).

By Formula (9.11.1) in [6, Section 9.11], it is proved that:

$$(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}-\underline{m}_N)\left(1-\frac{\frac{N}{M}\int\frac{\underline{m}_N t^2 dF^{\mathbf{R_N}}(t)}{(1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N})(1+t\underline{m}_N)}}{-z+\frac{N}{M}\int\frac{tdF^{\mathbf{R_N}}}{1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}}-T_N}\right)=\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\underline{m}_N T_N \tag{25}$$

where

$$T_N=\frac{N}{M^2}\sum_{j=1}^M\mathbb{E}\beta_j d_j(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N})^{-1},$$

$$d_j=d_j(z)=-\mathbf{q}_j^*\mathbf{R}^{1/2}(\hat{\mathbf{R}}_{(j)}-z\mathbf{I})^{-1}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\mathbf{R}^{1/2}\mathbf{q}_j+(1/M)\mathrm{Tr}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}_N-z\mathbf{I})^{-1},$$

$$\beta_j=\frac{1}{1+\frac{1}{M}\mathbf{y}_j^*(\hat{\mathbf{R}}_{(j)}-z\mathbf{I})^{-1}\mathbf{y}_j},$$

$$\mathbf{q}_j=1/\sqrt{N}\mathbf{x}_j,$$

$$\hat{\mathbf{R}}_{(j)}=\hat{\mathbf{R}}_N-\frac{1}{M}\mathbf{y}_j\mathbf{y}_j^*.$$

If one differentiates (25) with respect to z, the equation becomes

$$\begin{aligned}&(\mathbb{E}m'_{\underline{\hat{\mathbf{R}}}_N}-\underline{m}'_N)\left(1-\frac{\frac{N}{M}\int\frac{\underline{m}_N t^2 dF^{\mathbf{R_N}}(t)}{(1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N})(1+t\underline{m}_N)}}{-z+\frac{N}{M}\int\frac{tdF^{\mathbf{R_N}}}{1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}}-T_N}\right)+(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}-\underline{m}_N)\left(1-\frac{\frac{N}{M}\int\frac{\underline{m}_N t^2 dF^{\mathbf{R_N}}(t)}{(1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N})(1+t\underline{m}_N)}}{-z+\frac{N}{M}\int\frac{tdF^{\mathbf{R_N}}}{1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}}-T_N}\right)'\\&=\mathbb{E}m'_{\underline{\hat{\mathbf{R}}}_N}\underline{m}_N T_N+\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\underline{m}'_N T_N+\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\underline{m}_N T'_N.\end{aligned}$$

In the work of [6, Section 9.11], it is proved that when N tends to infinity,

1) $\sup_{z\in\tilde{\mathcal{R}}_k}|\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z)-\underline{m}(z)|\to 0$ and $\sup_{z\in\tilde{\mathcal{R}}_k}|\underline{m}_N(z)-\underline{m}(z)|\to 0$,

2) $\dfrac{\frac{N}{M}\int \frac{t^2 \underline{m}_N dF^{\mathbf{R_N}}(t)}{(1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N})(1+t\underline{m}_N)}}{-z+\frac{N}{M}\int \frac{t dF^{\mathbf{R_N}}}{1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}} - T_N}$ converges ,

3) $\hat{X}_N^2(z) \to 0, \quad T_N \to 0.$

With the same method, one can show easily that

4) $\sup_{z\in\tilde{\mathfrak{R}}_k} |\mathbb{E}m'_{\underline{\hat{\mathbf{R}}}_N}(z) - \underline{m}'(z)| \to 0,$

5) $\sup_{z\in\tilde{\mathfrak{R}}_k} |\underline{m}'_N(z) - \underline{m}'(z)| \to 0,$

6) $\frac{N}{M}\left(\sum_{j=1}^{M} \mathbb{E}\beta_j d_j\right)'$ converges.

With these results, it suffices to show that $T'_N \to 0$, and $\hat{X}_N^{2'}$ is equicontinuous.

In [6, Section 9.9], they show that for $m, p \in \mathbb{N}$ and a non-random $N\times N$ matrix $\mathbf{A}_k$, $k = 1,..,m$ and $\mathbf{B}_\ell$, $\ell = 1,..,q$, we have

$$\left|\mathbb{E}\left(\prod_{k=1}^{m} \mathbf{r}_t^* \mathbf{A}_k \mathbf{r_t} \prod_{\ell=\mathbf{1}}^{\mathbf{q}} (\mathbf{r_t^* R}_\ell \mathbf{r_t} - \mathbf{M^{-1}}\mathrm{Tr}\mathbf{R}\mathbf{B}_\ell)\right)\right| \le KM^{-(1\wedge q)} \prod_{k=1}^{m} \|\mathbf{A}_k\| \prod_{\ell=1}^{q} \|\mathbf{B}_\ell\|. \tag{26}$$

We have also that for any positive p,

$$\max(\mathbb{E}\|\mathbf{D}^{-1}(z)\|^p, \mathbb{E}\|\mathbf{D}_j^{-1}(z)\|^p, \mathbb{E}\|\mathbf{D}_{ij}^{-1}(z)\|^p) \le K_p \tag{27}$$

and

$$\sup_{N, z\in\tilde{\mathfrak{R}}_k} \|(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z)\mathbf{R} + \mathbf{I})^{-1}\| < \infty \tag{28}$$

where K_p is a constant which depends only on p.

With all these preliminaries, as $T_N \to 0$, by the dominated convergence theorem of derivation, it suffices to show that T'_N is bounded over $\tilde{\mathfrak{R}}_k$. In [6, Section 9.11], it is sufficient to show that $(f'_M(z))$ is bounded where

$$f_M(z) = \sum_{j=1}^{M} \mathbb{E}[(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j - M^{-1}\mathrm{Tr}\mathbf{D}_j^{-1}\mathbf{R})(\mathbf{r}_j^*\mathbf{D}_j^{-1}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\mathbf{r}_j - M^{-1}\mathrm{Tr}\mathbf{D}_j^{-1}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\mathbf{R})].$$

With the help of (26)-(28), $f'_M(z)$ is indeed bounded in $\tilde{\mathfrak{R}}_k$.

Now we will show that $\hat{X}_N^{2'}$ is equicontinuous. With the light work as before, it is sufficient to show that $f''_M(z)$ is bounded. Using (26), we obtain

$$
\begin{aligned}
|f''(z)| \leq & KM^{-1}\Big[\Big(\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-3}\mathbf{R}\bar{\mathbf{D}}_1^{-3}\mathbf{R})\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}(\mathbb{E}m_{\underline{\underline{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\mathbf{R}(\mathbb{E}\bar{m}_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\bar{\mathbf{D}}_1^{-1}\mathbf{R})\Big)^{1/2} \\
& + 2\Big(\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-2}\mathbf{R}\bar{\mathbf{D}}_1^{-2}\mathbf{R})\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-2}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\mathbf{R}(\mathbb{E}\bar{m}_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\bar{\mathbf{D}}_1^{-2}\mathbf{R})\Big)^{1/2} \\
& + 2|\mathbb{E}m'_{\underline{\hat{\mathbf{R}}}_N}|\Big(\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-2}\mathbf{R}\bar{\mathbf{D}}_1^{-2}\mathbf{R})\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-2}\mathbf{R}(\mathbb{E}\bar{m}_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-\mathbf{2}}\bar{\mathbf{D}}_{\mathbf{1}}^{-\mathbf{1}}\mathbf{R})\Big)^{\mathbf{1/2}} \\
& + \Big(\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}\mathbf{R}\bar{\mathbf{D}}_1^{-1}\mathbf{R})\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-3}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\mathbf{R}(\mathbb{E}\bar{m}_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-1}\bar{\mathbf{D}}_1^{-3}\mathbf{R})\Big)^{1/2} \\
& + 2|\mathbb{E}m'_{\underline{\hat{\mathbf{R}}}_N}|\Big(\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}\mathbf{R}\bar{\mathbf{D}}_1^{-1}\mathbf{R})\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-2}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-2}\mathbf{R}(\mathbb{E}\bar{m}_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-2}\bar{\mathbf{D}}_1^{-2}\mathbf{R})\Big)^{1/2} \\
& + |\mathbb{E}m''_{\underline{\hat{\mathbf{R}}}_N}|\Big(\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}\mathbf{R}\bar{\mathbf{D}}_1^{-1}\mathbf{R})\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-2}\mathbf{R}(\mathbb{E}\bar{m}_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-2}\bar{\mathbf{D}}_1^{-1}\mathbf{R})\Big)^{1/2} \\
& + |\mathbb{E}m'{}_{\underline{\hat{\mathbf{R}}}_N}|^2\Big(\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}\mathbf{R}\bar{\mathbf{D}}_1^{-1}\mathbf{R})\mathbb{E}(\mathrm{Tr}\mathbf{D}_1^{-1}(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-3}\mathbf{R}(\mathbb{E}\bar{m}_{\underline{\hat{\mathbf{R}}}_N}\mathbf{R}+\mathbf{I})^{-3}\bar{\mathbf{D}}_1^{-1}\mathbf{R})\Big)^{1/2}\Big].
\end{aligned}
$$

Thanks to (27) and (28), the right side is indeed bounded. This ends the proof of the tightness.

C. Study of the set $\mathcal{R}_{z_2}$

For z_2 fixed, denote $\mathcal{R}_{z_2} = \{z_1 \in \mathbb{C} : z_1 \neq z_2, m_{\underline{\hat{\mathbf{R}}}_N}(z_1) = m_{\underline{\hat{\mathbf{R}}}_N}(z_2)\}$. We will show that this set is empty a.s. for all N, M large. Suppose that $z_1 \in \mathcal{R}_{z_2}$. We first use [6, Formula (9.11.4)] that

$$\sup_{z \in \mathcal{R}_{z_2}} |\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z) - \underline{m}_N(z)| \to 0.$$

By Formula (16), we get

$$z(\underline{m}_N) = -\frac{1}{\underline{m}_N} + \frac{N}{M}\int \frac{t}{1+t\underline{m}_N} F^{R_N}(dt).$$

As $z_1 \neq z_2$,

$$\underline{m}_N(z_1) \neq \underline{m}_N(z_2).$$

Take $\epsilon = |\underline{m}_N(z_1) - \underline{m}_N(z_2)|$. For N sufficiently high,

$$|\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - \underline{m}_N(z_1)| < \frac{\epsilon}{4}$$

and

$$|\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z_2) - \underline{m}_N(z_2)| < \frac{\epsilon}{4}.$$

Finally

$$
\begin{aligned}
|\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - \mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z_2)| &\geq \epsilon - |\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - \underline{m}_N(z_1)| - |\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z_2) - \underline{m}_N(z_2)| \\
&\geq \frac{\epsilon}{2},
\end{aligned}
$$

which implies, along with $m_{\underline{\hat{\mathbf{R}}}_N}(z) - \mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z) \xrightarrow{a.s.} 0$ that for all large N, $\mathcal{R}_{z_2} = \varnothing$ almost surely.

REFERENCES

[1] V. Plerous, P. Gopikrishnan, B. Rosenow, L. Amaral, T. Guhr, and H. Stanley, "Random matrix approach to cross correlations in financial data," *Phys. Rev. E*, vol. 65, no. 6, Jun. 2002.

[2] F. Luo, Y. Yang, J. Zhong, H. Gao, L. Khan, D. K. Thompson, and J. Zhou, "Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory," *BMC bioinformatics*, vol. 8, no. 1, p. 299, 2007.

[3] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim, "BER and Outage Probability Approximations for LMMSE Detectors on Correlated MIMO channels," *Information Theory*, vol. 55, no. 10, pp. 4386–4397, 2009.

[4] R. Couillet, J. W. Silverstein, and M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," *IEEE Trans. Inf. Theory*, Submitted for publication. [Online]. Available: http://arxiv.org/abs/1001.3934

[5] X. Mestre and M. Lagunas, "Modified Subspace Algorithms for DoA Estimation With Large Arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598–614, Feb. 2008.

[6] Z. Bai and J. W. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices," *Springer Series in Statistics*, 2009.

[7] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, To appear.

[8] V. L. Girko, "Ten years of general statistical analysis." [Online]. Available: http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

[9] J. W. Silverstein and P. L. Combettes, "Signal detection via spectral theory of large dimensional random matrices," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 2100–2105, 1992.

[10] N. E. Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *Annals of Statistics*, vol. 36, no. 6, pp. 2757–2790, Dec. 2008.

[11] O. . Ryan and M. Debbah, "Free deconvolution for signal processing applications," in *Proc. IEEE International Symposium on Information Theory (ISIT'07)*, Nice, France, Jun. 2007, pp. 1846–1850.

[12] R. Couillet and M. Debbah, "Free deconvolution for OFDM multicell SNR detection," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*, Cannes, France, 2008.

[13] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Statist.*, vol. 36, no. 1, pp. 199–227, 2008. [Online]. Available: http://dx.doi.org/10.1214/009053607000000758

[14] ——, "Covariance regularization by thresholding," *Ann. Statist.*, vol. 36, no. 6, pp. 2577–2604, 2008. [Online]. Available: http://dx.doi.org/10.1214/08-AOS600

[15] T. T. Cai, C. Zhang, and H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *Ann. Statist.*, vol. 38, no. 4, pp. 2118–2144, 2010. [Online]. Available: http://dx.doi.org/10.1214/09-AOS752

[16] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5353–5368, Nov. 2008.

[17] ——, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.

[18] P. Vallet, P. Loubaton, and X. Mestre, "Improved subspace DoA estimation methods with large arrays: The deterministic signals case," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)*, 2009, pp. 2137–2140.

[19] J. Yao, R. Couillet, J. Najim, E. Moulines, and M. Debbah, "CLT for Eigen-Inference Methods in Cognitive Radios," in *International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech, 2011.

[20] V. A. Marčenko and L. A. Pastur, “Distributions of eigenvalues for some sets of random matrices,” *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.

[21] J. W. Silverstein and Z. D. Bai, “On the empirical distribution of eigenvalues of a class of large dimensional random matrices,” *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[22] 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, Jan. 1998.

[23] O. Kallenberg, *Foundations of mordern Probability, 2nd edition*. Springer Verlag New York, 2002.

[24] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1995.

[25] Z. D. Bai and J. W. Silverstein, “CLT of linear spectral statistics of large dimensional sample covariance matrices,” *Annals of Probability*, vol. 32, no. 1A, pp. 553–605, 2004.

[26] J. Marsden and M. Hoffman, *Basic Complex Analysis, 3rd ed*. New York: Freeman, 1987.

[27] Z. D. Bai and J. W. Silverstein, “Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices,” *The Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.

[28] J. W. Silverstein and S. Choi, “Analysis of the limiting spectral distribution of large dimensional random matrices,” *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, 1995.

[29] F. Cucker and A. Corbalan, “An Alternate Proof of the Continuity of the Roots of a Polynomial,” *The American Mathematical Monthly*, vol. 96, no. 4, pp. 342–345, Apr. 1989.

[30] P. Vallet, P. Loubaton, and X. Mestre, “Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case,” *IEEE Trans. Inf. Theory*, Submitted for publication. [Online]. Available: http://arxiv.org/abs/1002.3234

[31] W. Hachem, O. Khorunzhy, P. Loubaton, J. Najim, and L. A. Pastur, “A new approach for capacity analysis of large dimensional multi-antenna channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 9, 2008.

[32] U. Haagerup, H. Schultz, and S. Thorbjørnsen, “A random matrix approach to the lack of projections in Cred*(F2),” *Advances in Mathematics*, vol. 204, no. 1, pp. 1–83, 2006.

[33] A. W. V. der Vaart, *Asymptotic statistics*. New York: Cambridge University Press, 2000.