Fluctuations of an Improved Population Eigenvalue Estimator in Sample Covariance Matrix Models

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

August 26, 2011

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 known and 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. A specific application to wireless sensor networks is developed.

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], array processing [5], etc.

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

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.

$\|\hat{\mathbf{R}}_N - \mathbf{R}_N\| \to 0$, 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, especially in wireless communications where the number of available samples has often a size comparable to the dimension of each sample. In order to cope with this issue, random matrix theory [6], [7] has proposed new tools, mainly spurred by the *G-estimators* of Girko [8]. Other works include convex optimization methods [9], [10] and free probability tools [11], [12]. 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 $\mathbf{R}$. The main contributor is Mestre [13]-[14] who studies the case where $\mathbf{R}_N = \mathbf{U}_N\mathbf{D}_N\mathbf{U}_N^H$ with $\mathbf{D}_N$ diagonal with entries of large multiplicities and $\mathbf{U}_N$ with i.i.d. entries. For this model, he provides an estimator for every eigenvalue of $\mathbf{R}$ with large multiplicity under some separability condition, see also Vallet *et al.* [15], Couillet *et al.* [4] for more elaborate models.

These estimators, although proven asymptotically unbiased, have nonetheless not been fully characterized in terms of performance statistics. 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 fluctuations of the population eigenvalue estimator of [14] 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 approximation, asymptotically accurate is also derived.

The results are applied in a cognitive radio context in which we assume the co-existence of a licensed (primary) network and an opportunistic (secondary) network aiming at reusing the bandwidth resources left unoccupied by the primary network. The eigenvalue estimator is used here by secondary users to estimate the transmit power of primary users, while the fluctuations are used to provide a confidence margin on the estimate.

The remainder of the article is structured as follows: In Section II, the system model is introduced and the main results from [13], [14] are recalled. In Section III, the CLT for the estimator in [14] is stated with the asymptotic variance. In Section IV, an empirical approximation for the variance is derived. A cognitive radio application of these results is provided in Section V, with comparative Monte Carlo simulations. Finally, Section VI concludes this article. Technical proofs are postponed to the appendix.

II. Estimation of the population eigenvalues

A. Notations

In this paper, the notations $s, \mathbf{x}, \mathbf{M}$ stand for scalars, vectors and matrices, respectively. As usual, $\|\mathbf{x}\|$ 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 a $N \times 1$ vector, then $\mathrm{diag}(\mathbf{x})$ is the $N \times N$ matrix with diagonal elements 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 $\delta_{k\ell}$ is denoted as Kronecker's symbol (whose value is 1 if $k = \ell$, 0 otherwise).

If the support $\mathcal{S}$ of a probability measure over $\mathbb{R}$ is the finite union of 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 nonnegative Hermitian matrix with eigenvalues $(\xi_i;\ 1 \leq i \leq N)$, we denote in the sequel 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 a $N \times M$ matrix $\mathbf{X}_N = (X_{ij})$ whose entries are independent and identically distributed (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 a $N \times N$ Hermitian matrix with L (L being fixed) distinct eigenvalues $\rho_1 < \cdots < \rho_L$ with respective multiplicities $N_1, \cdots, N_L$ (notice 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 $\mathbf{R}_N = \mathbb{E}\mathbf{y}_i\mathbf{y}_i^H$. In this article, we are interested in recovering information on $\mathbf{R}_N$ based on the observation

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

which is referred to as the *sample covariance matrix*.

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 1 (A1):

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

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

Assumption 2 (A2):

We assume that the limiting support $\mathcal{S}$ of the eigenvalue distribution of $\hat{\mathbf{R}}_N$ is formed of L compact disjoint subsets (cf. Figure 1). Following [14], one can also reformulate this condition in a more analytic manner: The limiting support of $\hat{\mathbf{R}}_N$ is formed of L clusters if and only if for $i \in \{1, .., L\}$, $\inf_N\{\frac{M}{N} - \Psi_N(i)\} > 0$, 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 $L-1$ different ordered solutions to 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 $N = 60$ with equal multiplicity.

C. *Mestre's Estimator of the population eigenvalues*

In [14], an estimator of the population covariance matrix eigenvalues $(\rho_k; \ 1 \leq k \leq L)$ based on the observations $\hat{\mathbf{R}}_N$ is proposed.

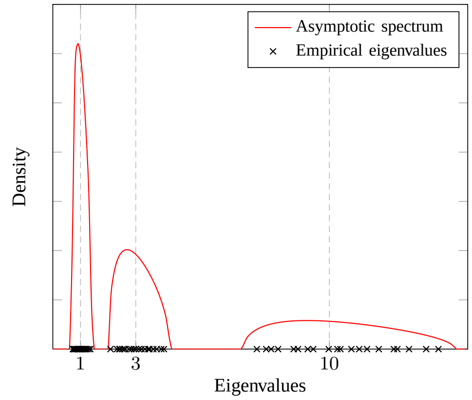

Fig. 1. Empirical and asymptotic eigenvalue distribution of $\hat{\mathbf{R}}_N$ for $L = 3$, $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 10$, $N/M = c = 0.1$, $N = 60$, $N_1 = N_2 = N_3 = 20$.

Theorem 1: [14] Denote by $\hat{\lambda}_1 \leq \cdots \leq \hat{\lambda}_N$ the ordered eigenvalues of $\hat{\mathbf{R}}_N$. Let $M, N \to \infty$ in the sense of the assumption (A1). Under the assumptions (A1)-(A2), 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 $\hat{\mu}_1 \leq \cdots \leq \hat{\mu}_N$ the (real and) ordered solutions of:

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

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

The proof of Theorem 1 relies on random matrix theory, and in particular, [16], [17] use as a key ingredient the *Stieltjes transform*.

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

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

In the case where $F^{\mathbf{Z}}$ is the spectral distribution associated to a nonnegative Hermitian matrix $\mathbf{Z} \in \mathbb{C}^{N\times N}$ with eigenvalues $(\xi_i;\ 1 \leq i \leq N)$, 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 can be seen as the normalized trace of the resolvent $(\mathbf{Z} - z\mathbf{I}_N)^{-1}$. Since the seminal paper of Marčenko and Pastur [16], 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 first important result is due to Bai and Silverstein [17] (see also [16]).

Theorem 2: [17] Denote by $F^{\mathbf{R}}$ the limiting spectral distribution of $\mathbf{R}_N$, *i.e.* $F^{\mathbf{R}}(d\,\lambda) = \sum_{k=1}^{L} \frac{c_k}{c} \delta_{\rho_k}(d\,\lambda)$. Under the assumption (A1), 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 $z \in \mathbb{C}^+ = \{z \in \mathbb{C},\ \Im(z) > 0\}$, where $\underline{m}(z)$ is defined as the unique solution in $\mathbb{C}^+$ of:

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

Note that $\underline{m}(z)$ is also a Stieltjes transform whose associated distribution function will be denoted $\underline{F}$, 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 $F^{\underline{\hat{\mathbf{R}}}_N}$. Notice 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 $m_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i) = 0$. Otherwise stated, the $\hat{\mu}_i$'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:

$$
\begin{aligned}
\underline{m}_N(z) &= -\left(z - \frac{N}{M}\int \frac{t}{1+t\underline{m}_N(z)} dF^{\mathbf{R}_N}(t)\right)^{-1} \quad , \\
m_N(z) &= \frac{M}{N}\underline{m}_N(z) - \left(1-\frac{M}{N}\right)\frac{1}{z} \; .
\end{aligned}
$$

It can be shown that m_N and $\underline{m}_N$ are Stieltjes transforms of given probability measures F_N and $\underline{F}_N$, respectively (cf. [7, Theorem 3.2]).

With these notations at hand, we can now derive 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 $w = -\frac{1}{\underline{m}_M(z)}$ and the condition that the limiting support $\mathcal{S}$ of the eigenvalue distribution of $\mathbf{R}_N$ is formed of L distinct clusters $(\mathcal{S}_k, 1\le k\le L)$ (cf. Figure 1), we can write:

$$
\rho_k = \frac{M}{2i\pi N_k}\oint_{\mathcal{C}_k} z\frac{\underline{m}'_N(z)}{\underline{m}_N(z)}dz \; , \quad 1\le k\le L \tag{5}
$$

where $\mathcal{C}_k$ and $\mathcal{C}_\ell$ denote negatively oriented contours which enclose the corresponding clusters $\mathcal{S}_k$ and $\mathcal{S}_\ell$ respectively. Defining

$$
\hat{\rho}_k \triangleq \frac{M}{2\pi i N_k}\oint_{\mathcal{C}_k} z\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z)}{m_{\underline{\hat{\mathbf{R}}}_N}(z)}dz \; , \quad 1\le k\le L \; , \tag{6}
$$

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).

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

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: Under the assumptions (A1)-(A2) and with the same notations:

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

where $\mathcal{N}_L$ refers to a real L-dimensional Gaussian distribution, and $\mathbf{\Theta}$ is a $L\times L$ matrix whose entries $\Theta_{k\ell}$ are given by (7), where $\mathcal{C}_k$ and $\mathcal{C}_\ell$ are defined as before (cf. Formula (5)).

$$\Theta_{k\ell} \quad = \quad -\frac{1}{4\pi^2 c_k c_\ell} \oint_{\mathcal{C}_k} \oint_{\mathcal{C}_\ell} \left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \right] \frac{1}{\underline{m}(z_1)\underline{m}(z_2)} dz_1 dz_2 \ . \tag{7}$$

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: Almost surely,

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

Denote by $C(\mathcal{C}_k, \mathbb{C})$ the set of continuous functions from $\mathcal{C}_k$ to $\mathbb{C}$ endowed with the supremum norm $\|u\|_\infty = \sup_{\mathcal{C}_k} |u|$. Consider the process:

$$(X_N, X'_N, u_N, u'_N) : \mathcal{C}_k \to \mathbb{C}^4$$

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 due to 'no eigenvalue' result (cf. [18], see also Proposition 1), (X_N, X'_N, u_N, u'_N) almost surely belongs to $C(\mathcal{C}_k, \mathbb{C})$ and $M(\hat{\rho}_k - \rho_k)$ writes:

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

where

$$\Upsilon_N(x, x', u, u') \ = \ \frac{M}{2\pi i N_k} \oint_{\mathcal{C}_k} z \left(\frac{\underline{m}_N(z) x'(z) - u'(z) x(z)}{\underline{m}_N(z) u(z)} \right) dz \ . \tag{8}$$

If needed, we shall explicitly indicate the dependence in the contour $\mathcal{C}_k$ and write $\Upsilon_N(x, x', u, u', \mathcal{C}_k)$. 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 $\Upsilon_N(X_N, X'_N, u_N, u'_N)$ with the help of the continuous mapping theorem [19].

(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 $\Upsilon_N(X_N, X'_N, u_N, u'_N, \mathcal{C}_k)$ and $\Upsilon_N(X_N, X'_N, u_N, u'_N, \mathcal{C}_\ell)$.

Remark 2: 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 $\mathcal{C}_k$. Tightness turns out to be difficult to establish due to the lack of control over the eigenvalues of $\underline{\hat{\mathbf{R}}}_N$ whenever the contour crosses the real line. In order to circumvent this issue, we shall introduce, following Bai and Silverstein [20], a process that approximates X_N and X'_N.

Let us now start the proof of Theorem 3.

Lemma 1: Under the assumptions (A1)-(A2), the process

$$(X_N, X'_N) : \mathcal{C}_k \to \mathbb{C}^4$$

converges in distribution to a Gaussian process (X, Y) with mean function zero and covariance function:

$$\begin{aligned}
\mathrm{cov}(X(z), X(\tilde{z})) &= \frac{\underline{m}'(z)\underline{m}'(\tilde{z})}{(\underline{m}(z) - \underline{m}(\tilde{z}))^2} - \frac{1}{(z-\tilde{z})^2} \triangleq \kappa(z, \tilde{z}) \ , \qquad (9)\\
\mathrm{cov}(Y(z), X(\tilde{z})) &= \frac{\partial}{\partial z}\kappa(z, \tilde{z}) \ ,\\
\mathrm{cov}(X(z), Y(\tilde{z})) &= \frac{\partial}{\partial \tilde{z}}\kappa(z, \tilde{z}) \ ,\\
\mathrm{cov}(Y(z), Y(\tilde{z})) &= \frac{\partial^2}{\partial z \partial \tilde{z}}\kappa(z, \tilde{z}) \ .
\end{aligned}$$

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

Proposition 1: Under the assumptions (A1)-(A2) 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).

A direct consequence of Lemma 1 yields that $(X_N, X'_N, u_N, u'_N) : \mathcal{C}_k \to \mathbb{C}^4$ converges in distribution to the Gaussian process $(X, Y, \underline{m}, \underline{m}')$ defined as before. 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 as expressed in [19] is reminded below.

Proposition 2 (cf. [19, 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 2 to the process (X_N, X_N', u_N, u_N') and to the function Υ_N as defined in (8). Denote by[1]

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

and consider the set

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

Then, it is shown in [6, Section 9.12.1] that $\inf_{\mathcal{C}_k} |\underline{m}| > 0$, 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 2 applies to $\Upsilon_N(x_N, y_N, v_N, w_N)$ and the following convergence holds true:

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

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

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

It remains to compute the covariance between $\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{C}_k)$ and $\Upsilon(X, Y, \underline{m}, \underline{m}', \mathcal{C}_\ell)$ for possibly different contours $\mathcal{C}_k$ and $\mathcal{C}_\ell$. We shall therefore evaluate, for $1 \leq k, \ell \leq L$:

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

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

$$\hat{\Theta}_{k\ell} = -\frac{M^2}{4\pi^2 N_k N_\ell} \oint_{\mathcal{C}_k} \oint_{\mathcal{C}_\ell} \left(\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z_1) m'_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{(m_{\underline{\hat{\mathbf{R}}}_N}(z_1) - m_{\underline{\hat{\mathbf{R}}}_N}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \right) \times \frac{1}{m_{\underline{\hat{\mathbf{R}}}_N}(z_1) m_{\underline{\hat{\mathbf{R}}}_N}(z_2)} d\,z_1 d\,z_2 \ . \tag{10}$$

---

where (a) follows from the fact that $\inf_{z\in\mathcal{C}_k} |\underline{m}(z)| > 0$ 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 $\partial^2/\partial z_1 \partial z_2$.

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,

$$\begin{aligned} & \oint \frac{z_1 z_2 \underline{m}(z_2) \partial_{1,2} \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}^2(z_2)} dz_1 \\ &= -\oint \frac{z_2 \partial_2 \kappa(z_1, z_2)}{\underline{m}(z_1) \underline{m}(z_2)} dz_1 + \oint \frac{z_1 z_2 \underline{m}'(z_1) \partial_2 \kappa(z_1, z_2)}{\underline{m}^2(z_1) \underline{m}(z_2)} dz_1 \ . \end{aligned}$$

Hence

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

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:

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

and (7) 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 observations $\hat{\mathbf{R}}_N$.

$$\hat{\Theta}_{k\ell} = \frac{M^2}{N_k N_\ell}\left[\sum_{(i,j)\in\mathcal{N}_k\times\mathcal{N}_\ell,\ i\neq j}\frac{-1}{(\hat{\mu}_i-\hat{\mu}_j)^2 m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_j)}\right.$$
$$\left.+\delta_{k\ell}\sum_{i\in\mathcal{N}_k}\left(\frac{m'''_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)}{6m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)^3}-\frac{m''_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)^2}{4m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)^4}\right)\right]. \quad (11)$$

Theorem 4: Assume that the assumptions (A1)-(A2) hold true, and recall the definition of $\Theta_{k\ell}$ given in (7). Let $\hat{\Theta}_{k\ell}$ be defined by (11), 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 4 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 (7), and 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 (7) 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 (10).

The proof of Theorem 4 now breaks down into two steps: The convergence of $\hat{\Theta}_{k\ell}$ to $\Theta_{k\ell}$, which relies on the definition (10) of $\hat{\Theta}_{k\ell}$ and on a dominated convergence argument, and the effective computation of the integral in (10) which relies on Cauchy's residue theorem [21], and yields (11).

We first address the convergence of $\hat{\Theta}_{k\ell}$ to $\Theta_{k\ell}$. Due to [18], [22], 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 $z \in \mathcal{C}_k$ will be eventually uniformly lower-bounded. By [14, Lemma 1], the same result 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 $\mathcal{C}_k$ (if not, then by compacity, there would exist $z \in \mathcal{C}_k$ 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 the contour). With these arguments at hand, one can easily apply the dominated convergence theorem and conclude that a.s. $\hat{\Theta}_{k\ell} \to \Theta_{k\ell}$.

We now evaluate the integral (10) by computing the residues of the integrand within $\mathcal{C}_k$ and $\mathcal{C}_\ell$. 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 (10), that is:

$$h(z_1,z_2)=\left(\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z_1)m'_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{(m_{\underline{\hat{\mathbf{R}}}_N}(z_1)-m_{\underline{\hat{\mathbf{R}}}_N}(z_2))^2}-\frac{1}{(z_1-z_2)^2}\right)\times\frac{1}{m_{\underline{\hat{\mathbf{R}}}_N}(z_1)m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}. \tag{12}$$

We first consider the case where $k\neq\ell$.

In this case, the two integration contours are different and it can be assumed that they never intersect (so it can always be assumed that $z_1\neq z_2$). 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 $\mathcal{C}_k$ is straightforward and yields:

$$r(z_2)\overset{\triangle}{=}\mathrm{Res}(h(\cdot,z_2),\hat{\mu}_i)=\left(\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)\,m'_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{m^2_{\underline{\hat{\mathbf{R}}}_N}(z_2)}-\frac{1}{(\hat{\mu}_i-z_2)^2}\right)\frac{1}{m'_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i)m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}\ . \tag{13}$$

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 $\mathcal{C}_k$, one obtains:

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

Then we need to consider the residue ξ on the set $\mathcal{R}_{z_2}=\{z_1:m_{\underline{\hat{\mathbf{R}}}_N}(z_1)=m_{\underline{\hat{\mathbf{R}}}_N}(z_2)\neq 0,\ z_1\neq z_2\}$. (If this set is empty, then the residue is zero.) Notice that ξ is not a residue of $\frac{1}{(z_1-z_2)^2}\frac{1}{m_{\underline{\hat{\mathbf{R}}}_N}(z_1)m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}$, hence one needs to compute

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

for the residue ξ. By integration by parts, one gets

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

Let $k=\min\{i\in\mathbb{N}^*:m^{(i)}_{\underline{\hat{\mathbf{R}}}_N}(\xi)\neq 0\}$, then by a Taylor expansion

$$m_{\underline{\hat{\mathbf{R}}}_N}(z_1)=m_{\underline{\hat{\mathbf{R}}}_N}(z_2)+\frac{(z_1-\xi)^k}{k!}m^{(k)}_{\underline{\hat{\mathbf{R}}}_N}(\xi)+o(z_1-\xi)^k,$$

and

$$m'_{\underline{\hat{\mathbf{R}}}_N}(z_1)=\frac{(z_1-\xi)^{k-1}}{(k-1)!}m^{(k)}_{\underline{\hat{\mathbf{R}}}_N}(\xi)+o(z_1-\xi)^{k-1}.$$

Hence

$$\mathrm{Res}(g,\xi)=-\frac{km'_{\underline{\hat{\mathbf{R}}}_N}(z_2)}{m^3_{\underline{\hat{\mathbf{R}}}_N}(z_2)}.$$

As it is the derivative function of $\frac{k}{2m^2_{\underline{\hat{\mathbf{R}}}_N}(z_2)}$, the integration with respect to z_2 is zero.

It remains to count the number of zeros within each contour. By [14, Lemma 1], eventually, there are exactly as many zeros as eigenvalues within each contour. It has been proved that the contribution of the residues of ξ on $\mathcal{R}_{z_2}$ is null, hence the result in the case $k \neq \ell$:

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

We now compute the integral (10) in the case where $k = \ell$, and begin by the computation of the residues at $\hat{\mu}_i$. The definition (13) of r and the computation of $\operatorname{Res}(r, \hat{\mu}_j)$ still hold true in the case where $\hat{\mu}_j$ is within $\mathcal{C}_k$ but different from $\hat{\mu}_i$. It remains to compute $\operatorname{Res}(r, \hat{\mu}_i)$. Taking $z_2 \to \mu_i$, we get:

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

Finally,

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

Hence the residue:

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

There are two other residues that should be taken into account for the computation of the integral: The residues of ξ on $\mathcal{R}_{z_2}$, 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 $\frac{1}{(z_1 - z_2)^2} \frac{1}{m_{\underline{\hat{\mathbf{R}}}_N}(z_1) m_{\underline{\hat{\mathbf{R}}}_N}(z_2)}$ for the residue $z_1 = z_2$. The integration by parts yields that:

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

Then the residue for $z_1 = z_2$ is:

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

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

Finally both have a null contribution, hence the formula:

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

■

V. Performance in the context of cognitive radios

We introduce below a practical application of the above result to the telecommunication field of cognitive radios. Consider a communication network implementing orthogonal code division multiple access (CDMA) in the uplink, which we refer to as the *primary* network. The primary network is composed of K transmitters. The data of transmitter k are modulated by the n_k orthogonal N-chip codes $\mathbf{w}_{k,1},\ldots,\mathbf{w}_{k,n_k} \in \mathbb{C}^N$. Consider also a secondary network, in sensor mode, that we assume time-synchronized with the primary network, and whose objective is to determine the distances of the primary transmitters in order to optimally reuse the frequencies used by the primary transmitters[2]. From the viewpoint of the secondary network, primary user k has power P_k. Then, at symbol time m, any secondary user receives the N-dimensional data vector

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K}\sqrt{P_k}\sum_{j=1}^{n_k}\mathbf{w}_{k,j}x_{k,j}^{(m)} + \sigma\mathbf{n}^{(m)} \tag{14}$$

with $\sigma\mathbf{n}^{(m)} \in \mathbb{C}^N$ an additive white Gaussian noise $\mathcal{CN}(0,\sigma^2\mathbf{I})$ received at time m and $x_{k,j}^{(m)} \in \mathbb{C}$ the signal transmitted by user k on the carrier code j at time m, which we assume $\mathcal{CN}(0,1)$ as well. The propagation channel is considered frequency flat on the CDMA transmission bandwidth. We do not assume that the sensor knows σ^2 neither the vectors $\mathbf{w}_{k,j}$. The secondary users may or may not be aware of the number of codewords employed by each user.

Equation (14) can be compacted under the form

$$\mathbf{y}^{(m)} = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{x}^{(m)} + \sigma\mathbf{n}^{(m)}$$

with $\mathbf{W} = [\mathbf{w}_{1,1},\ldots,\mathbf{w}_{1,n_1},\mathbf{w}_{2,1},\ldots,\mathbf{w}_{K,n_K}] \in \mathbb{C}^{N\times n}$, $n \triangleq \sum_{k=1}^K n_k$, $\mathbf{P} \in \mathbb{C}^{n\times n}$ the diagonal matrix with entry P_1 of multiplicity n_1, P_2 of multiplicity n_2, etc. and P_K of multiplicity n_K, and $\mathbf{x}^{(m)} = [\mathbf{x}_1^{(m)T},\ldots,\mathbf{x}_K^{(m)T}]^T \in \mathbb{C}^n$ where $\mathbf{x}_k^{(m)} \in \mathbb{C}^{n_k}$ is a column vector with j-th entry $x_{k,j}^{(m)}$.

[2]the rationale being that far transmitters will not be interfered with by low power communications within the secondary network.

Gathering M successive independent observations, we obtain the matrix $\mathbf{Y} = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N \times M}$ given by

$$\mathbf{Y} = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{N} = \begin{bmatrix} \mathbf{W}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{N} \end{bmatrix}$$

where $\mathbf{X} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}]$ and $\mathbf{N} = [\mathbf{n}^{(1)}, \dots, \mathbf{n}^{(M)}]$.

The $\mathbf{y}^{(m)}$ are therefore independent Gaussian vectors of zero mean and covariance $\mathbf{R} \triangleq \mathbf{W}\mathbf{P}\mathbf{W}^H + \sigma^2\mathbf{I}_N$. Since the objective is to retrieve the powers P_k, while σ^2 is known, the problem boils down to finding the eigenvalues of $\mathbf{W}\mathbf{P}\mathbf{W}^H + \sigma^2\mathbf{I}_N$. However, the sensors only have access to $\mathbf{Y}$, or equivalently to the sample covariance matrix

$$\mathbf{R}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^H = \frac{1}{M}\sum_{m=1}^{M} \mathbf{y}^{(m)}\mathbf{y}^{(m)H}.$$

Assuming $\mathbf{R}_N$ conveys a good appreciation of the eigenvalue clustering to the secondary user (as in Figure 1), Theorem 1 enables the detection of primary transmitters and the estimation of their transmit powers $P_1, \dots, P_K$; this boils down to estimating the largest K eigenvalues of $\mathbf{W}\mathbf{P}\mathbf{W}^H + \sigma^2\mathbf{I}_N$, i.e. the $P_k + \sigma^2$, and to subtract σ^2 (optionally estimated from the smallest eigenvalue of $\mathbf{W}\mathbf{P}\mathbf{W}^H + \sigma^2\mathbf{I}_N$ if $n < N$). Call $\hat{P}_k$ the estimate of P_k.

Based on these power estimates, the secondary user can determine the optimal coverage for secondary communications that ensures no interference with the primary network. A basic idea for instance is to ensure that the closest primary user, i.e. that with strongest received power, is not interfered with. Our interest is then cast on P_K. Now, since the power estimator is imperfect, it is hazardous for the secondary network to state that K has power $\hat{P}_K$ or to add some empirical security margin to $\hat{P}_K$. The results of Section III partially answer this problem.

Theorems 3 and 4 enable the secondary sensor to evaluate the accuracy of $\hat{P}_k$. In particular, assume that the cognitive radio protocol allows the secondary network to interfere the primary network with probability q and denote A the value

$$A \triangleq \inf_a \{\mathbb{P}(P_K - \hat{P}_K > a) \leq q\}.$$

According to Theorem 3, for N, M large, A is well approximated by $\hat{\Theta}_{K,K}Q^{-1}(q)$, with Q the Gaussian cumulative distribution function. If the secondary users detect a user with power P_K, estimated by $\hat{P}_K$, $\mathbb{P}(\hat{P}_K + A < P_K) < q$ and then it is safe for the secondary network to assume the worst case scenario where user K transmits at power $\hat{P}_K + A \simeq \hat{P}_K + \hat{\Theta}_K Q^{-1}(q)$.

In Figure 2, the performance of Theorem 3 is compared against $10,000$ Monte Carlo simulations of a scenario of three users, with $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3 = 20$, $N = 60$ and $M = 600$.

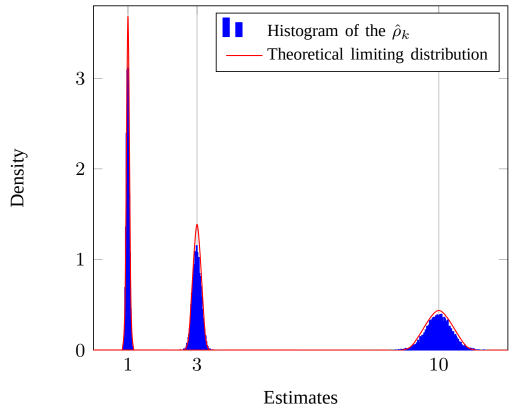

Fig. 2. Comparison of empirical against theoretical variances, based on Theorem 3, for three users, $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3 = 20$ codes per user, $N = 60$, $M = 600$ and SNR= 20 dB.

It appears that the limiting distribution is very accurate for these values of N, M. We also performed simulations to obtain empirical estimates $\hat{\Theta}_k$ of Θ_k from Theorem 4, which suggest that $\hat{\Theta}_k$ is an accurate estimator as well.

VI. Conclusion

In this article, we derived an exact expression and an approximation of the limiting performance of a statistical inference method that estimates the population eigenvalues of a class of sample covariance matrices. These results are applied in the context of cognitive radios to optimize secondary network coverage based on measures of the primary network activity.

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 clusters:

$$\mathcal{S} = (a_1, b_1) \cup \cdots \cup (a_L, b_L) \ .$$

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

Proposition 3: 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:

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

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 3: If the support $\mathcal{S}_N$ contains zero, (ex: $N > M$), then zero is also in the support $\mathcal{S}$, the conclusion is still true.

Proof of Proposition 3: Recall the relations:

$$\underline{m}_N(z) = -\left(z - \frac{N}{M}\int \frac{t}{1+t\underline{m}_N(z)} dF^{\mathbf{R}_N}(t)\right)^{-1} \tag{15}$$

and

$$m_N(z) = \frac{M}{N}\underline{m}_N(z) - \left(1 - \frac{M}{N}\right)\frac{1}{z}. \tag{16}$$

As the inverse of Stieltjes transform of $-\frac{1}{z}$ is δ_0 (the Dirac mass on 0) 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 $\underline{\mathcal{S}}_N$ associated to $\underline{F}_N$.

From the definition of $\underline{m}_N(z)$ (see formula (15)), we obtain:

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

Denote by $B = \{m \in \mathbb{R} : m \neq 0, -m^{-1} \notin \{\rho_1, \cdots, \rho_L\}\}$. In [23, Theorem 4.1 and Theorem 4.2], Silverstein and Choi show that for a real number x, $x \in \underline{\mathcal{S}}_N^c \iff \underline{m}_x \in B$ 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 $a \in \partial\underline{\mathcal{S}}_N$, $m_a \notin B$ 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 [23, 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 [23, Theorem 5.1].

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

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

It is equivalent to the following equation:

$$z'_{\mathbf{R_N}}(m_a) = \frac{1}{m_a^2} - \frac{1}{M}\sum_{i=1}^{L} N_i \frac{\rho_i^2}{(1+\rho_i m_a)^2} = 0. \tag{17}$$

By multiplying the common denominator, one will get a polynomial of the degree 2L in m_a. Now we will show that these 2L roots are real. At first, notice 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 $i \in \{1, \cdots, L-1\}$. 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

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

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 $b_\ell^N \xrightarrow[N,M \to \infty]{} b_\ell$, notice that a_i b_i satisfy the same type of the equation by replacing $\frac{N}{M}$ by c and $F^{\mathbf{R_N}}$ by $F^{\mathbf{R}}$. As $\frac{N}{M} \to c$ and $\frac{K_i}{M} \to c_i$, the roots of Equation (17) converge to those of the limit equation (see [24]). Thus we achieve the second 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 Proposition 3, $\phi(x) = 0$ for all $x \in \mathcal{S}_N$. Now if $\mathbf{Z}$ is a $M \times M$ hermitian matrix with spectral decomposition $\mathbf{Z} = \mathbf{U}\,\mathrm{diag}\,(\gamma_i;\ 1 \leq i \leq M))\ \mathbf{U}^H$, where $\mathbf{U}$ is unitary and $\mathrm{diag}\,(\gamma_i;\ 1 \leq i \leq M))$ stands for the $M \times M$ diagonal matrix whose entries are $\mathbf{Z}$'s eigenvalues, write $\phi(\mathbf{Z}) = \mathbf{U}\,\mathrm{diag}\,(\phi(\gamma_i);\ 1 \leq i \leq M))\ \mathbf{U}^H$.

We have:

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

for every $p \geq 1$, 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 $k \geq 1$ 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 Proposition 3, $\int_{\mathcal{S}_N} f(\lambda)\, F_N(d\lambda) = 0$ (where F_N is the probability distribution associated to m_N) for N, M large enough ($N, M \to \infty$). A minor modification of [25, Lemma 2] (whose model is slightly different) with the help of [26, 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:

$$\begin{aligned} & \left|\mathbb{E}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}\right| && (18)\\ &= \left|\mathbb{E}\left([\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^k} + \mathbb{E}[\mathrm{tr} f(\hat{\mathbf{R}}_N)]^{2^k} - \mathbb{E}[\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^k}\right)^2\right| \\ &\leq 2\left(\mathrm{Var}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^k} + |\mathbb{E}[\mathrm{Tr} f(\hat{\mathbf{R}}_N)]^{2^k}|^2\right). && (19)\end{aligned}$$

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 [26, Section II-B]) to handle the first term of the r.h.s. Applying this inequality, we obtain:

$$\mathrm{Var}\left((\mathrm{Tr}\, f(\hat{\mathbf{R}}_N))^{2^k}\right) \leq K\sum_{i,j}\mathbb{E}\left[\left|\frac{\partial[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^k}}{\partial Y_{i,j}}\right|^2 + \left|\frac{\partial[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^k}}{\partial \overline{Y}_{i,j}}\right|^2\right], \tag{20}$$

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 [27, Lemma 4.6]. This yields:

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

Plugging these derivatives into (20), we obtain:

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

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 $\alpha < 1$:

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

Plugging this estimate into (18), we obtain:

$$\left|\mathbb{E}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}\right| \leq K\left(\frac{1}{N^{1+\alpha}}|\mathbb{E}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}|^{\frac{2^{k+1}-2}{2^{k+1}}}\right) + \mathcal{O}(N^{-2\beta}), \tag{21}$$

where K is a constant independent of M, N, k. Notice that inequality (21) involves twice the quantity of interest $\mathbb{E}[\mathrm{Tr}\, f(\hat{\mathbf{R}}_N)]^{2^{(k+1)}}$ that we want to upper bound by $O(N^{-2\beta})$. We shall proceed iteratively.

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

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

Plugging this into (21) yields:

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

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:

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

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

As explained in Section III, there are two conditions to prove (Billingsley [28, Theorem 13.1]):

- Finite-dimensional convergence of the process (X_N, X'_N).
- Tightness on the contour $\mathcal{C}_k$.

Remark 4: As u_N (resp. u'_N) converges almost surely to $\underline{u}$ (resp. $\underline{u}'$) (see Silverstein and Bai [17]), the convergence of the process (X_N, X'_N, u_N, u'_N) is achieved as soon as the convergence of the process (X_N, X'_N) is proved.

In [20], Bai and Silverstein establish a central limit theorem for $F^{\mathbf{R}_N}$ with the complex Gaussian entries X_{ij}. We recall below their main result.

Proposition 4: [20] With the notations introduced in Section II, for $f_1, \dots, f_p$, analytic on an open region containing $\mathbb{R}$,

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

2)
$$\left(N\int f_i(x)d(F^{\hat{\mathbf{R}}_N}-F_N)(x)\right)_{1\le i\le p}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\mathbf{V}),$$
where $\mathbf{V}=(V_{ij})$ and
$$V_{ij}=-\frac{1}{4\pi^2}\oint\oint f_i(z_1)f_j(z_2)v_{ij}(z_1,z_2)dz_1dz_2,$$
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 that circle around the support $\mathcal{S}$.

Now we apply this proposition to show the finite-dimensional convergence. For all $z_i\in\mathcal{C}_k\backslash\mathbb{R}$, notice 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 who contains the support $\mathcal{S}$ and $X_N(z)=M(m_{\underline{\hat{\mathbf{R}}}_N}(z)-\underline{m}_N(z))$. Then Proposition 4 implies directly that for all $p\in\mathbb{N}$, the random vector
$$\left(X_N(z_1),X_N^{'}(z_1),\cdots,X_N(z_p),X_N^{'}(z_p)\right)$$
converges to a centered Gaussian vector by considering the functions:
$$\left(f_1(x)=\frac{1}{x-z_1},f_2(x)=\frac{1}{(x-z_1)^2},\cdots,f_{2p-1}(x)=\frac{1}{x-z_p},f_{2p}(x)=\frac{1}{(x-z_p)^2}\right).$$
Thus the finite dimensional convergence is achieved.

The proof of the tightness is based on Nash-Poincaré inequality ([25] and [26]). In Appendix A, it is proved that for all $\epsilon>0$ and all $\ell\in\mathbb{N}$,
$$\mathbb{P}\left(\sup_{\lambda\in\operatorname{eig}(\hat{\mathbf{R}}_N)}d(\lambda,\mathcal{S})>\epsilon\right)=o(N^{-\ell}).$$
Following the same idea as Bai and Silverstein [20, Section 3 and 4], it is indeed a tight sequence. The details of the proof are in Appendix C. Thus Lemma 1 is achieved.

C. Proof of the 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 Nash-Poincaré's inequality [26]. First, denote by $M(m_{\underline{\hat{\mathbf{R}}}_N}(z)-\underline{m}_N(z))=M_N^1(z)+M_N^2(z)$ with $M_N^1(z)=M(m_{\underline{\hat{\mathbf{R}}}_N}(z)-\mathbb{E}[m_{\underline{\hat{\mathbf{R}}}_N}(z)])$ and $M_N^2(z)=M(\mathbb{E}[m_{\underline{\hat{\mathbf{R}}}_N}(z)]-\underline{m}_N(z))$.

As $\frac{1}{\hat{\rho}_k - z}$ can converge to infinite if z is close to the real axis, there will be a little trouble for the tightness. Then we need a truncated version of the process. More precisely, let ε_N be a real sequence decreasing to zero satisfying for some $\delta \in]0,1[$:

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

Remark 5: Notice that $X_N(z) = M(m_{\underline{\hat{\mathbf{R}}}_N} - \underline{m}_N) = \overline{X_N(\overline{z})}$ for $z \in \overline{\mathbb{C}^+}$. So it suffices to verify the arguments for $z \in \mathbb{C}^+$.

Denote by $([x_{2k-1}, x_{2k}], k = 1, \cdots, L)$ the k-th cluster of the support of the limiting spectral measure; and take l_{2k-1}, l_{2k} such that $x_{2k-2} < l_{2k-1} < x_{2k-1}$ and $x_{2k} < l_{2k} < x_{2k+1}$ for $k \in \{1, .., L\}$ with conventions $x_0 = 0$ and $x_{2L+1} = \infty$, $i.e.$, $[l_{2k-1}, l_{2k}]$ only contains the k-th cluster. Let $d > 0$. Consider:

$$C_u = \{x + id : x \in [l_{2k-1}, l_{2k}]\}.$$

and

$$C_r = \{l_{2k-1} + \mathbf{i}v : v \in [N^{-1}\varepsilon_N, d]\}.$$

Also

$$C_l = \{l_{2k} + \mathbf{i}v : v \in [N^{-1}\varepsilon_N, d]\}.$$

Then $C_N = C_l \cup C_u \cup C_r$. The process $\hat{M}_N^1(\cdot)$ is defined by

$$\hat{M}_N^1(z) = \begin{cases} M_N^1(z) & \text{for } z \in C_N, \\ M_N^1(l_{2k} + \mathbf{i}N^{-1}\varepsilon_N) & \text{for } x = l_{2k}, v \in [0, N^{-1}\varepsilon_N], \\ M_N^1(l_{2k-1} + \mathbf{i}N^{-1}\varepsilon_N) & \text{for } x = l_{2k-1}, v \in [0, N^{-1}\varepsilon_N]. \end{cases}$$

This partition of C_N is identical to that used in [20, Section 1]. With probability one (see [18] and [22]), for all $\epsilon > 0$,

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

with $d(x, S)$ the Euclidean distance of x to the set S. So with probability one, for all N large, ([20, page 563])

$$\left| \oint \left(M_N^1(z) - \hat{M}_N^1(z) \right) dz \right| \leq K_1 \varepsilon_N,$$

and

$$\left| \oint \left(M_N^{1\prime}(z) - \hat{M}_N^{2\prime}(z) \right) dz \right| \leq K_2 \varepsilon_N$$

for some constants K_1 and K_2. Both terms converge to zero as $M \to \infty$. Then it suffices to ensure the tightness for $\hat{M}_N^1(z)$ and $\hat{M}_N^{1\prime}(z)$.

We now prove tightness based on [28, Theorem 13.1], i.e.

1) Tightness at any point of the contour (here C_N).
2) Satisfaction of the condition

$$\sup_{N,z_1,z_2\in C_N} \frac{\mathbb{E}|(\hat{M}_N^1(z_1)-\hat{M}_N^1(z_2))|^2}{|z_1-z_2|^2} \leq K.$$

Condition 1) is achieved by an immediate application of Proposition 4. We now verify the second condition.

We evaluate $\frac{\mathbb{E}|(\hat{M}_N^1(z_1)-\hat{M}_N^1(z_2))|^2}{|z_1-z_2|^2}$. Notice that

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

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

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

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

$$\begin{aligned} \frac{\mathbb{E}|\hat{M}_1(z_1)-\hat{M}_1(z_2)|^2}{|z_1-z_2|^2} &\leq \frac{C_1}{N}\mathbb{E}\Big[\mathrm{Tr}(\mathbf{L}_N)\Big] \\ = \frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_N)\mathrm{I}_{\sup_n d(\hat{\lambda}_n,\mathcal{S})\leq\varepsilon}) &+ \frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_N)\mathrm{I}_{\sup_n d(\hat{\lambda}_n,\mathcal{S})>\varepsilon}) \end{aligned}$$

with

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

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

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

Then

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

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

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

The second condition of tightness is achieved.

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

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

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

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

$$\begin{aligned}\mathbf{L}_2 =& 4\hat{\mathbf{R}}_N\Big(3\mathbf{D}_N^{-4}(z_1)\mathbf{D}_N^{-4}(z_2) + 2\mathbf{D}_N^{-3}(z_1)\mathbf{D}_N^{-5}(z_2) + 2\mathbf{D}_N^{-5}(z_1)\mathbf{D}_N^{-3}(z_2) \\ &+ \mathbf{D}_N^{-2}(z_1)\mathbf{D}_N^{-6}(z_2) + \mathbf{D}_N^{-6}(z_1)\mathbf{D}_N^{-2}(z_2)\Big).\end{aligned}$$

Then Nash-Poincaré inequality yields that

$$\begin{aligned}&\mathrm{Var}\frac{|\hat{M}_N^{1'}(z_1)-\hat{M}_N^{1'}(z_2)|}{|z_1-z_2|}\\ &\leq \frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_2)\mathrm{I}_{\sup_n d(\hat{\lambda}_n,\mathcal{S})\leq\varepsilon})+\frac{C_1}{N}\mathbb{E}(\mathrm{Tr}(\mathbf{L}_2)\mathrm{I}_{\sup_n d(\hat{\lambda}_n,\mathcal{S})>\varepsilon})\end{aligned}$$

with C_1 the same constant defined as before. The term $\mathrm{Tr}(\mathbf{L}_2)$ is bounded on the set $\sup d(\hat{\lambda}_n,\mathcal{S})\leq\varepsilon$. For the second term, $|\mathrm{Tr}(\mathbf{L_2})|\leq\mathcal{O}\left(\frac{N^9}{\varepsilon_N^8}\right)$. As $\mathbb{P}(\sup d(\hat{\lambda}_n,\mathcal{S})\geq\varepsilon)=o(N^{-16})$ and $\varepsilon_N=N^{-0.01}$, the proof of the tightness of $M_N^{1\prime}(z)$ is achieved as before.

The proof of the tightness is completed with the verification of $M_N^{2\prime}(z)$ for $z\in\mathcal{C}_n$ to be bounded and forms an equicontinuous family, and convergence to 0. We will use the same method for the process $M_N^2(z)$ (see [6, Section 9.11]).

By Formula (9.11.1) in [6, Section 9.11], they show that

$$(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}-\underline{m}_N)\left(1-\frac{\frac{N}{M}\int\frac{\underline{m}_N t^2 dF^{\mathbf{R_N}}(t)}{(1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N})(1+t\underline{m}_N)}}{-z+\frac{N}{M}\int\frac{tdF^{\mathbf{R_N}}}{1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}}-T_N}\right)=\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}\underline{m}_N T_N \tag{22}$$

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 derives (22) 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], they show that when N tends to infinity,

1) $\sup_{z\in\mathcal{C}_N}|\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z)-\underline{m}(z)|\to 0$ and $\sup_{z\in\mathcal{C}_N}|\underline{m}_N(z)-\underline{m}(z)|\to 0$,
2) $\dfrac{\frac{N}{M}\int\frac{t^2\underline{m}_N dF^{\mathbf{R_N}}(t)}{(1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N})(1+t\underline{m}_N)}}{-z+\frac{N}{M}\int\frac{tdF^{\mathbf{R_N}}}{1+t\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}}-T_N}$ converges ,
3) $M_N^2(z)\to 0,\quad T_N\to 0.$

With the same method, one can show easily that

4) $\sup_{z\in\mathcal{C}_N} |\mathbb{E}m'_{\underline{\hat{\mathbf{R}}}_N}(z) - \underline{m}'(z)| \to 0$,
5) $\sup_{z\in\mathcal{C}_N} |\underline{m}'_N(z) - \underline{m}'(z)| \to 0$,
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 $M_N^{2\,\prime}$ 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 B_l, $\ell = 1,..,q$, we have

$$\left|\mathbb{E}\left(\prod_{k=1}^m \mathbf{r}_t^*\mathbf{A}_k\mathbf{r_t}\prod_{\ell=\mathbf{1}}^{\mathbf{q}}(\mathbf{r_t^*R_\ell r_t} - \mathbf{M^{-1}}\mathrm{Tr}\mathbf{RB}_\ell)\right)\right| \le KM^{-(1\wedge q)}\prod_{k=1}^m \|\mathbf{A}_k\|\prod_{\ell=1}^q\|\mathbf{B}_\ell\|. \tag{23}$$

We have also that for any positive p,

$$\max(\mathbb{E}\|\mathbf{D}^{-1}(z)\|^p, \mathbb{E}\|\mathbf{D}_j^{-1}(z)\|^p, \mathbb{E}\|\mathbf{D}_{ij}^{-1}(z)\|^p) \le K_p \tag{24}$$

and

$$\sup_{n,z\in\mathcal{C}_n}\|(\mathbb{E}m_{\underline{\hat{\mathbf{R}}}_N}(z)\mathbf{R}+\mathbf{I})^{-1}\| < \infty \tag{25}$$

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 $\mathcal{C}_N$. 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 (23)-(25), $f'_M(z)$ is indeed bounded in $\mathcal{C}_N$.

Now we will show that $M_N^{2\,\prime}$ is equicontinuous. With the light work as before, it is sufficient to show that $f''_M(z)$ is bounded. Using (23), we obtain

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

Thanks to (24) and (25), the right side is indeed bounded. This ends the proof of the tightness.

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] 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.

[14] ——, “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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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

[26] 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.

[27] 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.

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