Rapport WP 3-2 - ANR SESAME Eigenvalue Estimation of Parametrized Covariance Matrices of Large Dimensional Data

Jianfeng Yao, Abla Kammoun, Jamal Najim

May 11, 2012

Abstract

This article deals with the problem of estimating the covariance matrix of a series of independent multivariate observations, in the case where the dimension of each observation is of the same order as the number of observations. Although such a regime is of interest for many current statistical signal processing and wireless communication issues, traditional methods fail to produce consistent estimators and only recently results relying on large random matrix theory have been unveiled.

In this paper, we develop the parametric framework proposed by Mestre, and consider a model where the covariance matrix to be estimated has a (known) finite number of eigenvalues, each of it with an unknown multiplicity. The main contributions of this work are essentially threefold with respect to existing results, and in particular to Mestre's work: To relax the (restrictive) separability assumption, to provide joint consistent estimates for the eigenvalues and their multiplicities, and to study the variance error by means of a Central Limit Theorem.

I. Introduction

Estimating the covariance matrix of a series of independent multivariate observations is a crucial issue in many signal processing applications. A reliable estimate of the covariance matrix is for instance needed

Yao, Kammoun, Najim are with Télécom Paristech, France.

{yao,kammoun,najim}@telecom-paristech.fr .

Najim is also with the Centre National de la Recherche Scientifique (CNRS).

in principal component analysis [1], direction of arrival estimation for antenna arrays [2], blind subspace estimation [3], capacity estimation [4], estimation/detection procedures [2], [5], etc.

In the case where the dimension N of the observations is small compared to the number M of observations, the empirical covariance matrix based on the observations often provides a good estimate for the unknown covariance matrix. This estimate becomes however much less accurate, and even not consistent with the dimension N getting higher (see for instance [6, Theorem 2]).

An interesting theoretical framework for modern estimation of multi-dimensional variables occurs whenever the number of available samples M grows at the same pace as the dimension N of the considered variables. Shifting to this new assumption induces fundamental differences in the behavior of the empirical covariance matrix as analyzed in Mestre's work [6], [7]. Recently, several attempts have been done to address this problem (cf. [6], [7], [8], [9], [10], [11]) using large random matrix theory which proposed powerful tools, mainly spurred by Girko's G-estimators [12], to cope with this new context. In [6], [7], Mestre considers the eigenvalue estimation of a parametrized model of covariance matrices similar to the model we shall study in this article. In [8] and [11], grid-based techniques for inverting the Marčenko-Pastur equation are proposed. In [10], the problem of estimating a specific linear functional of the eigenvalues of an unknown covariance matrix is addressed. In [9], the eigenvalues of an unknown parametrized covariance matrix are estimated by resorting on the empirical moments of the observations. This technique, which goes back to Pisarenko's ideas [13], will be also combined to large random matrix theory in the present article.

We shall consider the case where the dimension of each observation N together with the number of samples M go to infinity at the same pace, *i.e.* their ratio converges to some nonnegative constant $c > 0$. In order to present the contribution provided in this paper, let us describe the model under study.

Consider an $N \times M$ matrix $\mathbf{X}_N = (X_{ij})$ whose entries are independent and identically distributed (i.i.d.) random variables. Let $\mathbf{R}_N$ be an $N \times N$ Hermitian matrix with L (L being fixed and known) distinct eigenvalues $0 < \rho_1 < \cdots < \rho_L$ 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, 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 covariance matrix of each observation $\mathbf{y}_i$ is $\mathbf{R}_N = \mathbb{E} \mathbf{y}_i \mathbf{y}_i^H$ (matrix $\mathbf{R}_N$ is sometimes called the population covariance matrix).

In this article, we consider the problem of estimating individually the eigenvalues ρ_i as well as their multiplicities N_i in the case where the total number of eigenvalues is fixed and known.

Such a scenario is customary in applications for wireless communications. A relevant example concerns uplink CDMA systems operating over flat fading channels, where users are arranged into L classes, each class corresponding to a distinct power amount. In this case, matrix $\mathbf{Y}_N$ can be modeled as:

$$\mathbf{Y}_N = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{V}_N$$

where $\mathbf{W}$ are $\mathbf{V}_N$ represent respectively the signature matrix assumed to be orthogonal and the noise matrix, while $\mathbf{P}$ is diagonal with diagonal elements taking distinct values among the finite set $\{\rho_1, \cdots, \rho_L\}$. In a decentralized context, where each user selects its own power from this finite set according to a defined control energy strategy, the base station which stands for the receiver can have to estimate the number of users in each class as well as their corresponding powers. Obviously, this problem amounts to estimating the eigenvalues of the theoretical covariance matrix as well as their corresponding multiplicities. Similar scenarios are studied in [14], [15].

Among the proposed parametric techniques, we cite the one developed by Mestre [7] and taken up by Vallet *et al* [16] and Couillet *et al* [17] for more elaborated models. Although being computationally efficient, this technique requires a *separability condition*, namely the assumption that the number of samples is large compared to the dimension of each sample (small limiting ratio $c = \lim \frac{N}{M} > 0$). In such a case, the limiting spectrum of the empirical covariance matrix possesses as many clusters[1]as there are eigenvalues to be estimated, and each eigenvalue can be estimated by a contour integral surrounding the related cluster. Mestre's technique cannot be applied any more in the case where c is larger (which reflects a higher dimension of the observations relatively to the sample dimension). In fact, the dimension of the clusters may grow and neighbouring clusters may merge, violating the one-to-one correspondence between clusters and eigenvalues to be estimated (see for instance Fig. 1 and 2).

A way to circumvent the separability condition has recently been proposed by Bai, Chen and Yao [9], based on the use of the empirical asymptotic moments:

$$\hat{\alpha}_k = \frac{1}{M}\text{Tr}\,\left(\mathbf{Y}_N\mathbf{Y}_N\right)^k, k \in \{1, \cdots, 2L\},$$

which can be shown to be a sufficient statistic to estimate $\left(\frac{N_1}{N}, \cdots, \frac{N_L}{N}, \rho_1, \cdots, \rho_L\right)$. Although being robust to separability condition, this technique suffers from numerical difficulties, since the proposed estimator has no closed-form expression and thus should be determined numerically. An interesting

[1]By cluster, we mean a connected component of the support of the limiting probability distribution of the spectrum.

contribution, although not directly focused on estimating the covariance of the observations is the work by Rubio and Mestre [18], where an alternative way to estimate the moments

$$\gamma_k = \frac{1}{N}\mathrm{Tr}(\mathbf{R}_N^k),$$

for all $k \in \mathbb{N}$ is proposed, yielding an explicit (yet lengthy) formula.

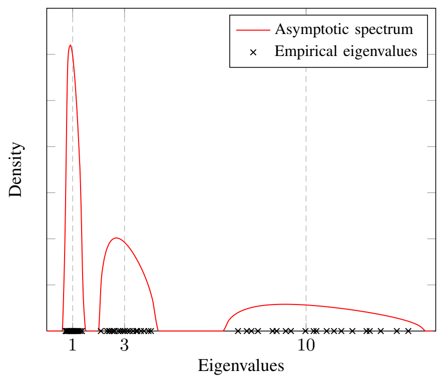

Fig. 1. Empirical and asymptotic eigenvalue distribution of $\hat{\mathbf{R}}_N$ for $L = 3$, $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 10$, $N/M = c = 0.1$, $N = 60$, $N_1 = N_2 = N_3 = 20$. In this case, there are 3 clusters in the limiting eigenvalues distribution and the separability assumption holds true.

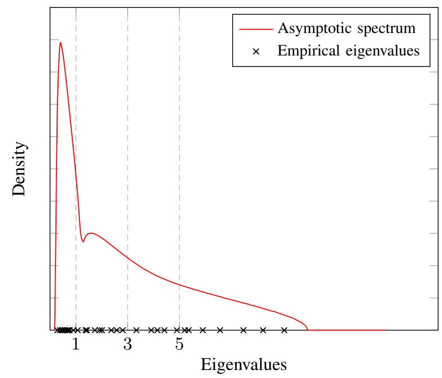

Fig. 2. Empirical and asymptotic eigenvalue distribution of $\hat{\mathbf{R}}_N$ for $L = 3$, $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 5$, $N/M = c = 3/8$, $N = 30$, $N_1 = N_2 = N_3 = 10$. In this case, there is only one cluster in the limiting eigenvalue distribution while there are 3 underlying eigenvalues to be estimated: The separability assumption does not hold true any more.

In this article, we improve existing work in several directions: With respect to Mestre's seminal papers [6], [7], we propose a joint estimation of the eigenvalues and their multiplicities, and drop the separability assumption. The proposed estimator is close in spirit to the one developed by Bai *et al.* [2] in [9], although we carefully establish the existence and uniqueness of the estimator, which is not explicit in [9]. Comparisons on the relative numerical efficiency of both procedures is provided in the simulations section. Finally, we study the fluctuations of the estimator and establish a Central Limit Theorem (CLT).

The remainder of the paper is organized as follows. In Section II, the main assumptions are provided and Mestre's estimator [7] is briefly reviewed. In Section III, the proposed estimator is described. Its fluctuations are studied in Section IV, where a CLT is stated. Simulations are presented in Section V, and a conclusion ends the paper in Section VI. Finally, the remaining technical details are postponed to

[2]We shall also mention an ongoing work by Li and Yao, not yet disclosed to our knowledge.

the Appendix.

II. Main assumptions and general background

A. Notations

In this paper, the notations $s, \mathbf{x}, \mathbf{M}$ stand for scalars, vectors and matrices, respectively. Superscripts $(\cdot)^T$ and $(\cdot)^H$ respectively stand for the transpose and transpose conjugate; trace of $\mathbf{M}$ is denoted by $\mathrm{Tr}(\mathbf{M})$; determinant of $\mathbf{M}$, by $\det(\mathbf{M})$; the mathematical expectation operator, by $\mathbb{E}$. 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.

If $\mathbf{Z} \in \mathbb{C}^{N\times N}$ is a nonnegative Hermitian matrix with eigenvalues $(\xi_i;\ 1 \le i \le N)$, we denote in the sequel 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. Main assumptions

Consider the model

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

and

$$\hat{\mathbf{R}}_N = \frac{1}{M}\mathbf{Y}_N\mathbf{Y}_N^H.$$

At first, an assumption about the matrix $\mathbf{R}_N$ is needed:

Assumption 1: $\mathbf{R}_N$ is an $N \times N$ Hermitian non-negative definite matrix with L (L being fixed and known) distinct eigenvalues $0 < \rho_1 < \cdots < \rho_L$ with respective multiplicities $N_1, \cdots, N_L$ (notice that $\sum_{i=1}^L N_i = N$).

As mentioned earlier, we consider the asymptotic regime where the number of samples M and the dimension N grow to infinity at the same pace, together with the multiplicities of each eigenvalue of $\mathbf{R}_N$.

Assumption 2: Let M, N be integers such that:

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

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

The following assumption is standard and is sufficient for estimation purposes.

Assumption 3: Let $\mathbf{X}_N = (X_{ij})$ be a $N \times M$ matrix whose entries are i.i.d. random variables in $\mathbb{C}$ such that $\mathbb{E}(\mathbf{X}_{1,1}) = 0$, $\mathbb{E}(|\mathbf{X}_{1,1}|^2) = 1$ with finite fourth moment: $\mathbb{E}(|\mathbf{X}_{1,1}|^4) < \infty$.

Remark 1: In order to establish the fluctuations of the estimators, the Gaussianity of the entries of $\mathbf{X}_N$ is needed (although this technical condition may be removed with substantial extra work).

Assumption 3b: The entries of the $N \times M$ matrix $\mathbf{X}_N = (X_{ij})$ are i.i.d. standard complex Gaussian variables, *i.e.* $X_{ij} = U + \mathbf{i}V$, where U, V are both independent real Gaussian random variables $\mathcal{N}(0, \frac{1}{2})$.

It is well-known in large random matrix theory that under Assumptions 1, 2 and 3, $F^{\hat{\mathbf{R}}_N}$ converges to a limiting probability distribution. In Mestre's paper [7], a *separability condition*[3] is needed in order to derive the estimator of $\mathbf{R}_N$'s eigenvalues:

Assumption 4: The support $\mathcal{S}$ of the limiting probability distribution of $F^{\hat{\mathbf{R}}_N}$ is composed of L compact connex disjoint subsets, and not reduced to a singleton.

Remark 2: Note that when $M < N$, matrix $\hat{\mathbf{R}}_N$ is singular and thus admits $(N - M)$ eigenvalues equal to zero. Hence, the limiting spectrum of $\hat{\mathbf{R}}_N$ has an additional mass in zero with weight $1 - \frac{1}{c}$, which will not be considered among the L clusters.

The separability condition is illustrated in Fig. 1 and 2. In both figures, the limiting distribution of $F^{\hat{\mathbf{R}}_N}$ is drawn (red line). In Fig. 1, $\mathbf{R}_N$'s eigenvalues are $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 10$, they have the same multiplicity and the ratio c is equal to 0.1. In this case, the separability condition is satisfied as the limiting distribution exhibits 3 clusters. The separability condition is no longer satisfied in Fig. 2, when $\rho_1 = 1$, $\rho_2 = 3$, $\rho_3 = 5$ and $c = 0.375$. In this case, the limiting distribution only exhibits a single cluster.

C. Background on Large Random Matrices, Mestre's estimators and their fluctuations

The *Stieltjes transform* has proved since Marčenko and Pastur's seminal paper [19] to be extremely efficient to describe the limiting spectrum of large random matrices. Given a probability distribution $\mathbb{P}$

[3]The precise technical statement of the separability condition together with a mathematical interpretation are available in [7], but are not necessary here.

defined over $\mathbb{R}^+$, its Stieltjes transform is a $\mathbb{C}$-valued function defined by:

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

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 is exactly the normalized trace of the resolvent $(\mathbf{Z} - z\mathbf{I}_N)^{-1}$.

An important result associated to the model presently under investigation is Bai and Silverstein's description of the limiting spectral distribution of $\hat{\mathbf{R}}_N$ [20] (see also [19]):

Theorem 1: [20] Assume that Assumptions 1, 2, 3 hold true and denote by $F^{\mathbf{R}}$ the limiting spectral distribution of $\mathbf{R}_N$, *i.e.* $F^{\mathbf{R}}(d\,\lambda) = \sum_{k=1}^{L} c_k \delta_{\rho_k}(d\,\lambda)$. 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:

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

Remark 3: Note that $\underline{m}(z)$ is also a Stieltjes transform whose associated probability 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:

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

Remark 4: 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

$$m_{\hat{\mathbf{R}}_N}(z) = \frac{M}{N} m_{\underline{\hat{\mathbf{R}}}_N}(z) - \left(1 - \frac{M}{N}\right)\frac{1}{z} \ . \tag{2}$$

Remark 5: 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:

$$\begin{cases} \underline{m}_N(z) = -\left(z - \frac{N}{M}\int \frac{t}{1+t\underline{m}_N(z)} F^{\mathbf{R}_N}(dt)\right)^{-1} \ , \\ m_N(z) = \frac{M}{N}\underline{m}_N(z) - \left(1 - \frac{M}{N}\right)\frac{1}{z} \ . \end{cases} \tag{3}$$

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. [21, Theorem 3.2]).

In [7], Mestre proposes a novel approach to estimate the eigenvalues $(\rho_k;\ 1 \le k \le L)$ of the population covariance matrix based on the observations $\hat{\mathbf{R}}_N$ under the additional Assumption 4. His approach relies on large random matrix theory and the separability condition presented above plays a major role in the mere definition of the estimators. As it will be a useful background in the sequel, we provide hereafter a brief description of Mestre's results:

Theorem 2: [7] Denote by $\hat{\lambda}_1 \le \cdots \le \hat{\lambda}_N$ the ordered eigenvalues of $\hat{\mathbf{R}}_N$. Under Assumptions 1, 2, 3, 4 and assuming moreover that the multiplicities $N_1, \cdots, N_L$ are known, the following convergence holds true:

$$\tilde{\rho}_k - \rho_k \xrightarrow[M,N\to\infty]{a.s.} 0 \ , \tag{4}$$

where

$$\tilde{\rho}_k = \frac{M}{N_k} \sum_{m \in \mathcal{N}_k} \left(\hat{\lambda}_m - \hat{\mu}_m \right) \ , \tag{5}$$

with $\mathcal{N}_k = \{\sum_{j=1}^{k-1} N_j + 1, \ldots, \sum_{j=1}^{k} N_j\}$ and $\hat{\mu}_1 \le \cdots \le \hat{\mu}_N$ the (real and) ordered solutions of:

$$\frac{1}{N} \sum_{m=1}^{N} \frac{\hat{\lambda}_m}{\hat{\lambda}_m - \mu} = \frac{M}{N} \tag{6}$$

repeated with their multiplicites. When $N > M$, we use the convention $\hat{\mu}_1 = \cdots = \hat{\mu}_{N-M+1} = 0$, whereas $\hat{\mu}_{N-M+2}, \cdots, \hat{\mu}_N$ are the positive solutions of the above equation.

Remark 6: Notice that (6) associated to (2) readily implies that for non null $\hat{\mu}_i$, $m_{\underline{\hat{\mathbf{R}}}_N}(\hat{\mu}_i) = 0$. 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.

Sketch of proof: We can now describe the main steps of Theorem 2. 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 positively oriented (clockwise) contour taking values in $\mathbb{C} \backslash \{\rho_1, \cdots, \rho_L\}$ 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:

$$\rho_k = \frac{M}{2i\pi N_k} \oint_{\mathcal{C}_k} z \frac{\underline{m}'_N(z)}{\underline{m}_N(z)} dz \ , \quad 1 \le k \le L, \tag{7}$$

where $\mathcal{C}_k$ denotes positively oriented contours which enclose the corresponding clusters $\mathcal{S}_k$. Defining

$$\tilde{\rho}_k \triangleq \frac{M}{2\pi i N_k} \oint_{\mathcal{C}_k} z \frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z)}{m_{\underline{\hat{\mathbf{R}}}_N}(z)} dz \ , \quad 1 \le k \le L \ , \tag{8}$$

dominated convergence arguments ensure that $\tilde{\rho}_k - \rho_k \to 0$, almost surely. The integral form of $\tilde{\rho}_k$ can then be explicitly computed thanks to residue calculus, and this finally yields (5). $\square$

Recently, a CLT has been derived [15] for this estimator under the extra assumption that the entries of $\mathbf{X}_N$ are Gaussian:

Theorem 3: [15] With the same notations as before, under Assumptions 1, 2, 3b, 4 and with known multiplicities $N_1, \cdots, N_L$, then:

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

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,

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

where $\mathcal{C}_k$ (resp. $\mathcal{C}_\ell$) is a closed counterclockwise oriented contour which only contains the k-th cluster (resp. ℓ-th) .

The proof of this theorem is based on [22] and the continuous mapping theorem. Details are available in [15].

The main objective of this article is to provide estimators for the ρ_k's without relying any more on the separability condition (*i.e.* to remove Assumption 4). A Central Limit Theorem will be established as well for the proposed estimator. As a by-product, the knowledge of the multiplicities will no longer be needed, and they will be estimated as well.

III. ESTIMATION OF THE EIGENVALUES ρ_i

In this section, we provide a method to estimate consistently the eigenvalues of the population covariance matrix and their multiplicities without the need of the separability condition (cf. Fig. 2). Our method is based on the asymptotic evaluation of the moments of the eigenvalues of $\mathbf{R}_N$,

$$\gamma_i \triangleq \frac{1}{N} \mathrm{Tr}\, \mathbf{R}_N^i = \sum_{k=1}^{L} \frac{N_k}{N} \rho_k^i, \quad 1 \leq i \leq 2L-1. \tag{9}$$

If $(\widehat{m}_i)_{1\leq i \leq 2L-1}$ are the empirical moments of the sample eigenvalues, then it is well-known that except for $i = 1$, γ_i cannot be approximated by $\widehat{m}_i$. Consistent estimators for γ_i are provided in [18], where it has been proved that:

$$\gamma_i - \tilde{\gamma}_i \xrightarrow[N,M\to+\infty]{} 0,$$

where

$$\tilde{\gamma}_i = \sum_{l=1}^{i} \mu_S(l,i)\widehat{m}_l, \tag{10}$$

$\mu_S(l,i)$ being some given coefficients that depend on the system dimensions and on the empirical moments $\widehat{m}_i$ [18]. An alternative is to use the Stieltjes transform:

Lemma 1: Let Assumptions 1, 2 and 3 hold true. Let $\hat{\gamma}_i$ be the real quantities given by:

$$\begin{cases} \hat{\gamma}_0 &= 1, \\ \hat{\gamma}_1 &= -\frac{M}{2N\mathbf{i}\pi}\oint_{\mathcal{C}} \frac{zm'_{\hat{\underline{\mathbf{R}}}_N}(z)}{m_{\hat{\underline{\mathbf{R}}}_N}(z)}dz, \\ \hat{\gamma}_k &= \frac{M(-1)^k}{2Nk\mathbf{i}\pi}\oint_{\mathcal{C}} \frac{dz}{m^k_{\hat{\underline{\mathbf{R}}}_N}(z)}, \quad \text{for } 2 \le k \le 2L-1 \end{cases}$$

where $\mathcal{C}$ is a counterclockwise oriented contour which encloses the support $\mathcal{S}$ of the limiting distribution of the eigenvalues of $\hat{\underline{\mathbf{R}}}_N$. Let γ_i be the theoretical moments as given in (9). Then, for $1 \le i \le 2L-1$,

$$\hat{\gamma}_i - \gamma_i \xrightarrow[N,M\to\infty]{a.s.} 0\ .$$

The proof of this lemma is postponed to Appendix A. While the estimates proposed by [18] are better in practice, estimates $(\hat{\gamma}_i)$ will be of interest in order to establish the Central Limit Theorem, and to obtain a closed-form expression of the asymptotic variance.

An interesting remark is that the map that links the eigenvalues and their multiplicities to their first $2L-1$ moments is invertible. Retrieving the eigenvalues from the estimates of the $2L-1$ moments is thus possible. This is the basic idea on which our method is founded.

The main result is stated as below:

Theorem 4: Let Assumptions 1, 2, 3 hold true and let $(\hat{\gamma}_k, 1 \le k \le 2L-1)$ be as in Lemma 1. Consider the following system of equations:

$$\begin{cases} \sum_{i=1}^{L} x_i = 1, \\ \sum_{i=1}^{L} x_i y_i^k = \hat{\gamma}_k \quad \text{for } 1 \le k \le 2L-1, \end{cases} \tag{11}$$

where $(x_i)_{1\le i\le L}$ and $(y_i)_{1\le i\le L}$ are $2L$ unknown parameters. Then for N, M large enough, the system of equations (11) has one and only one real solution $(\hat{c}_1, \cdots, \hat{c}_L, \hat{\rho}_1, \cdots, \hat{\rho}_L)$ with $\hat{\rho}_1 < \cdots < \hat{\rho}_L$. Moreover, $(\hat{c}_1, \cdots, \hat{c}_L, \hat{\rho}_1, \cdots, \hat{\rho}_L)$ is a consistent estimator of $(c_1, \cdots, c_L, \rho_1, \cdots, \rho_L)$, *i.e.*,

$$\hat{c}_\ell - c_\ell \xrightarrow[N,M\to\infty]{a.s.} 0 \quad \text{and} \quad \hat{\rho}_\ell - \rho_\ell \xrightarrow[N,M\to\infty]{a.s.} 0,$$

with $c_\ell = \lim \frac{N_\ell}{N}$ for $1 \le \ell \le L$.

Remark 7: The condition of separability is not required in the previous theorem. Moreover, the multiplicities are assumed to be unknown and thus have to be estimated. Fig. 2 represents a case where the three clusters are merged into one cluster. In such a situation, the estimator in [7] is biased whereas the proposed one is asymptotically consistent.

Remark 8: We use the estimator proposed in Lemma 1. However, the proof below does not depend on the estimator of the moments we choose. In fact, for any consistent estimator of the moments γ_i, the above theorem always holds true.

A. *Proof of Theorem 4*

The proof can be split into two main steps. By using the inverse function theorem, we can prove the almost sure existence of a real solution. Then, the uniqueness is ensured by a matrix inversion argument.

1) Existence of a real solution of the system: The first task is to show that the system of equations (11) admits, for N sufficiently large, one real solution $(\hat{c}_1, \cdots, \hat{c}_L, \hat{\rho}_1 \cdots, \hat{\rho}_L)$ satisfying $\hat{\rho}_1 < \hat{\rho}_2 < \cdots < \hat{\rho}_L$. We shall also establish the consistency of the obtained solution. The proof of the existence of a real solution follows in the same way as in [9]. It is merely based on the use of the inverse function theorem which ensures the existence as soon as the Jacobian matrix of the considered transformation is invertible. We recall below the inverse function theorem [23]:

Theorem 5: [23] Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function. Let $\mathbf{a}$ and $\mathbf{b}$ be vectors of $\mathbb{R}^n$ such that $f(\mathbf{a}) = \mathbf{b}$. If the Jacobian matrix of f at $\mathbf{a}$ is invertible, then there exists a neighborhood U containing $\mathbf{a}$ such that $f : U \to f(U)$ is a diffeomorphism, i.e, for every $\mathbf{y} \in f(U)$ there exists a unique $\mathbf{x}$ such that $f(\mathbf{x}) = \mathbf{y}$. In particular, f is invertible in U.

Consider the functional f defined as:

$$f(x_1, \cdots, x_L, y_1, \cdots, y_L) = \left(\sum_{\ell=1}^{L} x_\ell \ , \ \sum_{\ell=1}^{L} x_\ell y_\ell \ , \ \cdots \ , \ \sum_{\ell=1}^{L} x_\ell y_\ell^{2L-1} \right).$$

Consider $\mathbf{z} = (x_1, \cdots, x_L, y_1, \cdots, y_L)$ and denote by $\mathbf{c} = (c_1, \cdots, c_L, \rho_1, \cdots, \rho_L)$; we then have:

$$\mathbf{M} \triangleq \left. \frac{\partial f}{\partial \mathbf{z}} \right|_{\mathbf{z}=\mathbf{c}} = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ \rho_1 & \cdots & \rho_L & c_1 & \cdots & c_L \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \rho_1^{2L-1} & \cdots & \rho_L^{2L-1} & (2L-1)c_1\rho_1^{2L-2} & \cdots & (2L-1)c_L\rho_L^{2L-2} \end{bmatrix}. \tag{12}$$

As proven in [9, Proposition 1], matrix $\mathbf{M}$ is invertible. The inverse function theorem then applies. Denote by $\psi_i = \sum_{k=1}^{L} c_k \rho_k^i$ for $0 \leq i \leq 2L-1$. There exists a neighborhood U of $(c_1, \cdots, c_L, \rho_1, \cdots, \rho_L)$ and a neighborhood V of $(\psi_0, \cdots, \psi_{2L-1})$ such that f is a diffeomorphism from U onto V. On the other hand, we have:

$$\hat{\gamma}_i - \gamma_i \xrightarrow{a.s.} 0.$$

As $\gamma_i - \psi_i \to 0$, therefore, almost surely, $(\hat{\gamma}_0, \cdots, \hat{\gamma}_{2L-1}) \in V$ for N and M large enough. Hence, a real solution

$$(\hat{c}_1, \cdots, \hat{c}_L, \hat{\rho}_1, \cdots, \hat{\rho}_L) = f^{-1}(\hat{\gamma}_0, \cdots, \hat{\gamma}_{2L-1}) \in U$$

exists. And by the continuity, one can get easily that:

$$\hat{c}_\ell - c_\ell \xrightarrow[N,M\to\infty]{a.s.} 0 \quad \text{and} \quad \hat{\rho}_\ell - \rho_\ell \xrightarrow[N,M\to\infty]{a.s.} 0 \quad \text{for} \quad 1 \leq \ell \leq L \ .$$

2) Uniqueness of the solution of the system: Consider the polynomial Q with degree L defined as:

$$\mathrm{Q(X)} = \prod_{\ell=0}^{L}(X - \hat{\rho}_\ell) \triangleq \sum_{\ell=0}^{L} s_\ell \mathrm{X}^\ell$$

where $s_L = 1$. Denote by $\mathbf{s} = [s_0, \cdots, s_{L-1}]^T$. It is clear that $g : (\hat{\rho}_1, \cdots, \hat{\rho}_L) \to \mathbf{s}$ is a homeomorphism. It remains thus to show that vector $\mathbf{s}$ is uniquely determined by $(\hat{\gamma}_0, \cdots, \hat{\gamma}_{2L-1})$.

It is clear that each $\hat{\rho}_k$ is also the zero of the polynomial functions $\mathrm{R}_\ell(X)$ given by:

$$\mathrm{R}_\ell(\mathrm{X}) = \sum_{i=0}^{L} s_i \mathrm{X}^{i+\ell} \ ,$$

where $0 \leq \ell \leq L-1$. In other words, for $1 \leq k \leq L$, we get:

$$\sum_{i=0}^{L} s_i \hat{\rho}_k^{\ell+i} = 0,$$

or equivalently:

$$\sum_{i=0}^{L} s_i \hat{c}_k \hat{\rho}_k^{\ell+i} = 0. \tag{13}$$

Summing (13) over k, we obtain:

$$\sum_{i=0}^{L} \hat{\gamma}_{i+\ell} s_i = 0 \ , \tag{14}$$

for $0 \leq \ell \leq L-1$. Since $s_L = 1$, (14) becomes:

$$\hat{\gamma}_{L+\ell} + \sum_{i=0}^{L-1} s_i \hat{\gamma}_{i+\ell} = 0 \ , \tag{15}$$

for $0 \leq \ell \leq L-1$.

Writing (15) in a matrix form, we get: $\mathbf{\Gamma s} = -\mathbf{b}$, where

$$\mathbf{\Gamma} = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{L-1} \\ \hat{\gamma}_1 & \hat{\gamma}_2 & \cdots & \hat{\gamma}_L \\ \vdots & \ddots & \ddots & \vdots \\ \hat{\gamma}_{L-1} & \hat{\gamma}_L & \cdots & \hat{\gamma}_{2L-2} \end{bmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{bmatrix} \hat{\gamma}_L \\ \vdots \\ \hat{\gamma}_{2L-1} \end{bmatrix}. \tag{16}$$

On the other hand, we have $\mathbf{\Gamma} = \mathbf{A}\mathbf{D}\mathbf{A}^T$, where $\mathbf{D} = \mathrm{diag}(\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_L)$ and

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_L \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_1^{L-1} & \hat{\rho}_2^{L-1} & \cdots & \hat{\rho}_L^{L-1} \end{bmatrix}. \tag{17}$$

Then,

$$\det(\mathbf{\Gamma}) = \prod_{k=1}^{L} \hat{c}_k \prod_{1 \leq i < j \leq L} (\hat{\rho}_i - \hat{\rho}_j)^2 > 0.$$

Therefore, the vector $\mathbf{s}$ is then uniquely determined by $\mathbf{\Gamma}$ and $\mathbf{b}$ and is given by:

$$\mathbf{s} = -\mathbf{\Gamma}^{-1}\mathbf{b}.$$

Hence the unicity. Proof of Theorem 4 is completed.

B. Summary of the main steps of the estimation procedure

The proof of the unicity shows that the solutions of the system of equations (11) can be directly obtained from the estimates of the first $2L-1$ moments. More precisely, the estimation of the eigenvalues and their corresponding multiplicities can be performed through the following steps:

---

1) Set $\hat{\gamma}_0$ to 1. Estimate the first $2L-1$ moments using (10). Coefficients $\mu_S(l,i)$ are computed using Eq. (46) in [18].
2) Construct the matrix $\mathbf{\Gamma}$ and $\mathbf{b}$ using (16).
3) Compute the vector $\mathbf{s}$ as $\mathbf{s} = -\mathbf{\Gamma}^{-1}\mathbf{b}$.
4) Determine (by using for instance function roots of MATLAB) the roots $\hat{\rho}_1, \cdots, \hat{\rho}_L$ of the polynomial whose coefficients are given by vector $\mathbf{s}$.
5) Construct matrix $\mathbf{A}$ as specified by (17), and vector $\mathbf{d} = [\hat{\gamma}_0, \cdots, \hat{\gamma}_{L-1}]^{\mathrm{T}}$.
6) The coefficient estimates $\hat{\mathbf{c}} = [\hat{c}_1, \cdots, \hat{c}_L]^{\mathrm{T}}$ are thus given by:

$$\hat{\mathbf{c}} = \mathbf{A}^{-1}\mathbf{d}.$$

Remark 9: Note that while the existence of a real solution is only proven for M and N large enough, the previous algorithm always yield a solution, even for very small dimensions. However, in such scenarios, the validity of the obtained solution is not ensured. In fact, if N and M are not large enough, the moment estimates are not accurate, and the solution of the algorithm may yield complex or negative eigenvalues. This event completely disappears when N and M or only M take higher values. In practice, getting such inadequate solutions should warn that more samples are required.

IV. Fluctuations of the estimator

In this section, we shall study the fluctuations of the multiplicities and eigenvalues estimators $(\hat{c}_1, \cdots, \hat{c}_L, \hat{\rho}_1, \cdots, \hat{\rho}_L)$ introduced in Theorem 4. In particular, we establish a Central Limit Theorem for the whole vector in the case where the entries of matrix $\mathbf{X}_N$ are Gaussian.

Theorem 6: Let Assumptions 1, 2, 3b hold true. Let $(\hat{c}_1, \cdots, \hat{c}_L, \hat{\rho}_1, \cdots, \hat{\rho}_L)$ be the estimators obtained in Theorem 4. Then

$$M\left[\hat{c}_1 - \frac{N_1}{N}, \cdots, \hat{c}_L - \frac{N_L}{N}, \hat{\rho}_1 - \rho_1, \cdots, \hat{\rho}_L - \rho_L\right] \xrightarrow[N,M\to\infty]{\mathcal{D}} \mathcal{N}_{2L}(0, \mathbf{\Theta})$$

where $\mathbf{\Theta}$ is a $2L \times 2L$ matrix admitting the decomposition $\mathbf{\Theta} = \mathbf{M}^{-1}\mathbf{W}\mathbf{M}^{-1^T}$ and matrix $\mathbf{M}$ is the Jacobian matrix of f evaluated for $z = c$ and is defined in (12) and

$$\mathbf{W} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{bmatrix},$$

where $\mathbf{V}$ is a $(2L-1) \times (2L-1)$ matrix whose entries are given by (for $1 \le k, \ell \le 2L-1$):

$$V_{k,\ell} = -\frac{(-1)^{k+\ell}}{4\pi^2 c^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \left(\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \right) \times \frac{1}{\underline{m}^k(z_1)\underline{m}^\ell(z_2)} d\,z_1 d\,z_2$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are two closed contours non-overlapping which contain the support $\mathcal{S}$ of F and are counterclockwise oriented.

Proof: The proof relies on the same techniques as developed in [15]. We outline hereafter the main steps and then provide the details.

By Theorem 4, the estimate vector $(\hat{c}_1, \cdots, \hat{c}_L, \hat{\rho}_1, \cdots, \hat{\rho}_L)$ verifies the following system of equations:

$$\begin{cases} \sum_{i=1}^L \hat{c}_i = 1, \\ \sum_{i=1}^L \hat{c}_i \hat{\rho}_i = \hat{\gamma}_1, \\ \sum_{i=1}^L \hat{c}_i \hat{\rho}_i^k = \hat{\gamma}_k \text{ for } 2 \le k \le 2L-1, \end{cases}$$

where the $\hat{\gamma}_i$'s are the moment estimates provided by Lemma 1.

Using the integral representation of $\sum_{i=1}^L c_i \rho_i$ and $\sum_{i=1}^L c_i \rho_i^k$ (cf. Section A in the Appendix and Formula (22)), we get:

$$\begin{cases} \sum_{i=1}^L M\left(\hat{c}_i - \frac{N_i}{N}\right) = 0, \\ \sum_{i=1}^L M\left(\hat{c}_i \hat{\rho}_i - \frac{N_i}{N}\rho_i\right) = -\frac{M^2}{2N\mathbf{i}\pi} \oint_{\mathcal{C}} z \left(\frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z)}{m_{\underline{\hat{\mathbf{R}}}_N}(z)} - \frac{\underline{m}'_N(z)}{\underline{m}_N(z)} \right) dz, \\ \sum_{i=1}^L M\left(\hat{c}_i \hat{\rho}_i^k - \frac{N_i}{N}\rho_i^k\right) = \frac{M^2(-1)^k}{2\mathbf{i}(k-1)N\pi} \oint_{\mathcal{C}} \left(\frac{1}{m_{\underline{\hat{\mathbf{R}}}_N}(z)^{k-1}} - \frac{1}{\underline{m}_N(z)^{k-1}} \right) dz, \ 2 \le k \le 2L-1. \end{cases}$$

Denote by $C(\mathcal{C},\mathbb{C})$ the set of continuous functions from $\mathcal{C}$ to $\mathbb{C}$ endowed with the supremum norm $\|u\|_\infty = \sup_{\mathcal{C}} |u|$. In the same way as in [15], consider the process: $(X_N, X'_N, u_N, u'_N) : \mathcal{C} \to \mathbb{C}$, where

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

Then, $M\sum_{i=1}^{L}\left(\hat{c}_i\hat{\rho}_i - \frac{N_i}{N}\rho_i\right)$ can be written as:

$$\begin{aligned} M\sum_{i=1}^{L}\left(\hat{c}_i\hat{\rho}_i - \frac{N_i}{N}\rho_i\right) &= -\frac{M}{2\mathbf{i}N\pi}\oint_{\mathcal{C}} z\left(\frac{\underline{m}_N(z)X'_N(z) - u'_N(z)X_N(z)}{\underline{m}_N(z)u_N(z)}\right)dz, \\ &\triangleq \Upsilon_N(X_N, X'_N, u_N, u'_N), \end{aligned}$$

where

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

On the other hand, using the decomposition $a^k - b^k = (a-b)\sum_{\ell=0}^{k-1} a^\ell b^{k-1-\ell}$, we can prove that:

$$\begin{aligned} \sum_{i=1}^{L} M\left(\hat{c}_i\hat{\rho}_i^k - \frac{N_i}{N}\rho_i^k\right) &= \frac{M^2(-1)^k}{2\mathbf{i}N\pi(k-1)}\oint_{\mathcal{C}}\sum_{\ell=0}^{k-2} -\frac{m_{\underline{\hat{\mathbf{R}}}_N}(z) - \underline{m}_N(z)}{m^{\ell+1}_{\underline{\hat{\mathbf{R}}}_N}(z)\underline{m}_N^{k-1-\ell}(z)}dz \\ &= \frac{M(-1)^{k+1}}{2\mathbf{i}N(k-1)\pi}\oint_{\mathcal{C}}\sum_{\ell=0}^{k-2} X_N(z)u_N(z)^{-\ell-1}\underline{m}_N(z)^{-k+1+\ell}dz \\ &\triangleq \Phi_{N,k}(X_N, u_N), \end{aligned}$$

for $2 \le k \le 2L-1$, where

$$\Phi_{N,k}(x,u) = \frac{M(-1)^{k+1}}{2\mathbf{i}N(k-1)\pi}\oint_{\mathcal{C}}\sum_{\ell=0}^{k-2} x(z)u(z)^{-\ell-1}\underline{m}_N(z)^{-k+1+\ell}dz.$$

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

1) Prove the convergence of the processes (X_N, X'_N, u_N, u'_N) and (X_N, u_N) over the contour $\mathcal{C}$ by using Bai and Silverstein's theorem [22].
2) Prove the convergence of $[\Upsilon_N(X_N, X'_N, u_N, u'_N), \Phi_{N,2}(X_N, u_N), \cdots, \Phi_{N,L}(X_N, u_N)]^T$ to a Gaussian random vector with the help of the continuous mapping theorem (cf. Theorem 7).
3) Compute the limiting covariance between $M\sum_{i=1}^{L}\left(\hat{c}_i\hat{\rho}_i^k - \frac{N_i}{N}\rho_i^k\right)$ and $M\sum_{i=1}^{L}\left(\hat{c}_i\hat{\rho}_i^\ell - \frac{N_i}{N}\rho_i^\ell\right)$.
4) Conclude by expressing $M\left[\hat{c}_1 - \frac{N_1}{N}, \cdots, \hat{c}_L - \frac{N_L}{N}, \hat{\rho}_1 - \rho_1, \cdots, \hat{\rho}_L - \rho_L\right]^T$ as a linear function of $M\left[\hat{\gamma}_0 - \gamma_0, \cdots, \hat{\gamma}_{2L-1} - \gamma_{2L-1}\right]^T$.

A. Convergence of the processes (X_N, X'_N, u_N, u'_N) and (X_N, u_N)

The cornerstone of the first step is the convergence of the process

$$X_N : \mathcal{C} \to \mathbb{C}$$
$$z \to X_N(z)$$

to a Gaussian process $X(z)$ which is ensured in [24, Lemma 9.11].

For the process (X_N, X'_N, u_N, u'_N), it has been proved in [15, Lemma 1] that it indeed converges to the process $(X, Y, \underline{m}, \underline{m}')$ where (X, Y) is a Gaussian process with mean function zero and covariance function given by:

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

For the process (X_N, u_N), since $u_N \xrightarrow[N,M\to+\infty]{} \underline{m}$, (X_N, u_N) converges in distribution to $(X, \underline{m})$. The convergence of the process (X_N, u_N) is achieved.

B. Fluctuations of the moments

The next step is to prove the convergence of the vector $[\Upsilon_N(X_N, X'_N, u_N, u'_N), \Phi_{N,2}(X_N, u_N), \cdots, \Phi_{N,L}(X_N, u_N)]^T$.

The convergence of $\Upsilon_N(X_N, X'_N, u_N, u'_N)$ to a Gaussian random variable has been established in [15] where it has been proved that:

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

where

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

The next task is to prove the convergence in distribution of $\Phi_{N,k}(X_N, u_N)$ over the contour $\mathcal{C}$, for $2 \le k \le L$. Let $\Phi_k(x, u)$ be defined as:

$$\Phi_k(x, u) = \frac{(-1)^k}{2\mathbf{i}c\pi}\oint_{\mathcal{C}} x(z)u(z)^{-k}dz.$$

We want to show that $\Phi_k(X_N, u_N)$ converges in distribution to a Gaussian vector. The continuous mapping theorem is useful to transform one convergence to another.

Theorem 7 (cf. [25, 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, $(f_n)_{n\geq 1}$: $S_1 \to S_2$ 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 $f_n(\xi_n) \xrightarrow[n\to\infty]{\mathcal{D}} f(\xi)$.

Consider the set:

$$\Gamma = \left\{ (x,u) \in C^2\left(\mathcal{C}, \mathbb{C}\right), \inf_{\mathcal{C}} |u| > 0 \right\}.$$

Then, since $\inf_{\mathcal{C}} |\underline{m}| > 0$ (see [24, Section 9.12]), the dominated convergence theorem implies that the convergence of $(x_N, y_N) \to (x,y) \in \Gamma$ leads to $\Phi_{N,k}(x_N,y_N) \to \Phi_k(x,y)$. The continuous mapping theorem applies, thus giving:

$$\Phi_{N,k}(X_N,u_N) \xrightarrow[M,N\to\infty]{\mathcal{D}} \Phi_k(X,u).$$

It now remains to prove that the limit law $\Phi_k(X,u)$ is Gaussian. For that, 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 the elements of a Gaussian random vector is still Gaussian.

The convergence of $\Upsilon_N(X_N, X'_N, u_N, u'_N)$ and $\Phi_{N,k}(X_N,u_N)$ to Gaussian random variables is not sufficient to establish a CLT for the whole vector. It remains to prove that any linear combination of $[\Upsilon_N(X_N, X'_N, u_N, u'_N), \Phi_{N,2}(X_N,u_N), \cdots, \Phi_{N,L}(X_N,u_N)]^T$ converges toward a Gaussian distribution, which can easily be established in the same way as before. It implies that this vector converges to a Gaussian vector. This ends the proof of the fluctuations of the moments.

C. Computation of the variance

We now come to the third step. We shall therefore evaluate the quantities:

$$\begin{aligned}
\mathbf{V}_{1,1} &= \mathbb{E}\left[\Upsilon(X,Y,\underline{m},\underline{m}')\Upsilon(X,Y,\underline{m},\underline{m}')\right], \\
\mathbf{V}_{1,k} = \mathbf{V}_{k,1} &= \mathbb{E}\left[\Upsilon(X,Y,\underline{m},\underline{m}')\Phi_k(X,\underline{m})\right], \quad 2 \leq k \leq L, \\
\mathbf{V}_{k,\ell} &= \mathbb{E}\left[\Phi_k(X,\underline{m})\Phi_\ell(X,\underline{m})\right], \quad 2 \leq k,\ell \leq 2L-1.
\end{aligned}$$

The details of the calculations are in Appendix B and yield: For $1 \leq k,\ell \leq 2L-1$

$$\mathbf{V}_{k,\ell} = -\frac{(-1)^{k+\ell}}{4\pi^2 c^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} - \frac{1}{(z_1-z_2)^2} \right] \frac{1}{\underline{m}^k(z_1)\underline{m}^\ell(z_2)} dz_1 dz_2 \,. \tag{18}$$

Let $\mathbf{w}_M = M\left[\hat{\gamma}_0 - \gamma_0, \cdots, \hat{\gamma}_{2L-1} - \gamma_{2L-1}\right]^T$. We have just proved that the vector $\mathbf{w}_M$ converges asymptotically to:

$$\mathbf{w}_M \xrightarrow[N,M\to+\infty]{\mathcal{D}} \mathcal{N}_{2L}(0, \mathbf{W}),$$

where

$$\mathbf{W} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{bmatrix}$$

and $\mathbf{V}$ is the $(2L-1) \times (2L-1)$ matrix whose entries $V_{k,l}$ are given by (18).

Remark 10: The zeros in the variance simply follow from the fact that $\hat{\gamma}_0 - \gamma_0 = 0$.

D. Fluctuations of the eigenvalues estimates

To transfer this convergence to $\mathbf{q}_M \triangleq M\left[\hat{c}_1 - \frac{N_1}{N}, \cdots, \hat{c}_L - \frac{N_L}{N}, \hat{\rho}_1 - \rho_1, \cdots, \hat{\rho}_L - \rho_L\right]^T$, we shall use Slutsky's lemma which is as below:

Lemma 2 (cf. [26]): Let $\mathbf{X}_n$, $\mathbf{Y}_n$ be sequences of vector or matrix random elements. If $\mathbf{X}_n$ converges in distribution to a random element $\mathbf{X}$, and $\mathbf{Y}_n$ converges in probability to a constant $\mathbf{C}$, then

$$\mathbf{Y}_n^{-1}\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{C}^{-1}\mathbf{X}$$

provided that $\mathbf{C}$ is invertible.

We will show that $\mathbf{w}_M$ satisfies the following linear system:

$$\mathbf{w}_M = \hat{\mathbf{M}}_M \mathbf{q}_M \tag{19}$$

where we will try to find a matrix $\hat{\mathbf{M}}_M$ who converges in probability to $\mathbf{M}$ which is given by (12).

To this end, let us work out the expression of $w_{k,M}$, the k-th element of $\mathbf{w}_M$.

If $k = 1$, it is easy to see that $w_{1,M} = 0$.

For $k \geq 2$, $w_{k,M}$ is given by:

$$\begin{aligned} w_{k,M} &= M \sum_{i=1}^{L} \left(\hat{c}_i \hat{\rho}_i^{k-1} - \frac{N_i}{N} \rho_i^{k-1} \right) \\ &= M \sum_{i=1}^{L} \left(\hat{c}_i \hat{\rho}_i^{k-1} - \frac{N_i}{N} \hat{\rho}_i^{k-1} + \frac{N_i}{N} \hat{\rho}_i^{k-1} - \frac{N_i}{N} \rho_i^{k-1} \right) \\ &= M \sum_{i=1}^{L} \left(\left(\hat{c}_i - \frac{N_i}{N} \right) \hat{\rho}_i^{k-1} + \frac{N_i}{N} (\hat{\rho}_i - \rho_i) \sum_{\ell=0}^{k-2} \hat{\rho}_i^{\ell} \rho_i^{k-2-\ell} \right). \end{aligned}$$

Then define

$$\hat{\mathbf{M}}_M = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ \hat{\rho}_1 & \cdots & \hat{\rho}_L & \frac{N_1}{N} & \cdots & \frac{N_L}{N} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \hat{\rho}_1^{2L-1} & \cdots & \hat{\rho}_L^{2L-1} & \frac{N_1}{N} \sum_{\ell=0}^{2L-2} \hat{\rho}_1^{\ell} \rho_1^{2L-2-\ell} & \cdots & \frac{N_L}{N} \sum_{\ell=0}^{2L-2} \hat{\rho}_L^{\ell} \rho_L^{2L-2-\ell} \end{pmatrix}.$$

We can easily check that Eq. (19) is satisfied and $\hat{\mathbf{M}}_M$ converges in probability to $\mathbf{M}$. It remains to check that $\mathbf{M}$ is invertible. Note that the non-singularity of matrix $\mathbf{M}$ has already been established in Section III, where this property was required to prove the existence of an estimator. As a consequence, using Slutsky's lemma, we deduce that:

$$\hat{\mathbf{M}}_M \mathbf{q}_M \xrightarrow[M,N\to+\infty]{\mathcal{D}} \mathcal{N}_{2L}(0, \mathbf{W})$$

and

$$\mathbf{q}_M \xrightarrow[M,N\to+\infty]{\mathcal{D}} \mathcal{N}_{2L}\left(0, \mathbf{M}^{-1}\mathbf{W}(\mathbf{M}^{-1})^T\right).$$

This ends the proof for the fluctuation.

■

V. Simulations

In this section, we compare the performance of the proposed estimator with Mestre's estimator [7] in Section V-A; we then compare the proposed estimator with the estimator proposed by Bai et al. [9] in Section V-B. We finally verify by simulations the accuracy of the Gaussian approximation stated by the CLT in Section V-C.

A. Comparison with Mestre's estimator - with and without separability

As will be seen below, the separability assumption is compulsory for Mestre's method to be effective. If this assumption holds true, a simple clustering procedure enables to estimate the unknown multiplicities and Mestre's method outperforms our moment estimator (see Fig. 3).

If, however, the separability assumption is not met, then it is not clear how to directly estimate (even roughly) the multiplicities; and even if those were known, Mestre's estimation method has no methodological foundations (as the estimator is not even consistent in this case!) and the computation of Mestre's estimator yields a systematic error (see Fig. 4 for instance).

A final remark is in order with respect to the separability assumption: Although it is easy in simulations to generate data fulfilling or violating the separability assumption, it is not an easy task, while facing real data, to decide whether the separability assumption holds true or not. Building such a test remains an open problem, advocating for our procedure by default - unless any extra argument emerges to support a separability assumption. Otherwise stated, the non-separability assumption is much more realistic in practical cases.

In the first experiment, we consider the case where the separability condition holds true. We assume also that the covariance matrix has three different eigenvalues $(\rho_1, \rho_2, \rho_3) = (1, 3, 7)$, which are distributed as : $\frac{N_1}{N} = 0.5, \frac{N_2}{N} = \frac{N_3}{N} = \frac{1}{4}$. The ratio $\frac{N}{M}$ is set to $\frac{30}{200} = \frac{3}{20}$. The separability condition being met, the clusters are well separated so that the multiplicities can be estimated in a heuristic way based on the difference of the ordered eigenvalues. More precisely, an empirical method for estimating the multiplicities consists in the following steps:

- Arrange the eigenvalues of the covariance matrix in increasing order: $\hat{\lambda}_1 \leq \cdots, \leq \hat{\lambda}_N$.
- Take L indexes $i_1, \cdots, i_L$ satisfying:

$$\begin{aligned}
i_1 &= \arg\max_{i} \left(\hat{\lambda}_{i+1} - \hat{\lambda}_i\right) \ , \\
i_2 &= \arg\max_{i \neq i_1} \left(\hat{\lambda}_{i+1} - \hat{\lambda}_i\right) \ , \\
&\vdots \\
i_L &= \arg\max_{i \notin \{i_1, \cdots, i_{L-1}\}} \left(\hat{\lambda}_{i+1} - \hat{\lambda}_i\right) \ .
\end{aligned}$$

- Arrange these indexes in the increasing order: $i_{[1]} \leq \cdots \leq i_{[L]}$. Empirical estimates of the multiplicities are thus given by:

$$\begin{aligned}
\hat{N}_1 &= i_{[1]} \\
\hat{N}_2 &= i_{[2]} - i_{[1]} \\
&\vdots \\
\hat{N}_L &= N - i_{[L-1]}.
\end{aligned}$$

This empirical method has proved to be efficient in the asymptotic regime[4].

Fig. 3 compares the performance of the Mestre's estimator using the aforementioned method for estimating the multiplicities with that of the proposed estimator, in terms of MSE:

$$\mathrm{MSE} \triangleq \frac{1}{1000} \sum_{k=1}^{1000} \sum_{i=1}^{3} |\hat{\rho}_i^k - \rho_i|^2,$$

In this case, Mestre's estimator outperforms the proposed estimator. This can be attributed to numerical difficulties which will be discussed in the next section.

[4]Applying exact separation results from Bai and Silverstein [27], [28], it can be proved that the estimates of the normalized multiplicities $(\hat{N}_k/N)$ are asymptotically consistent.

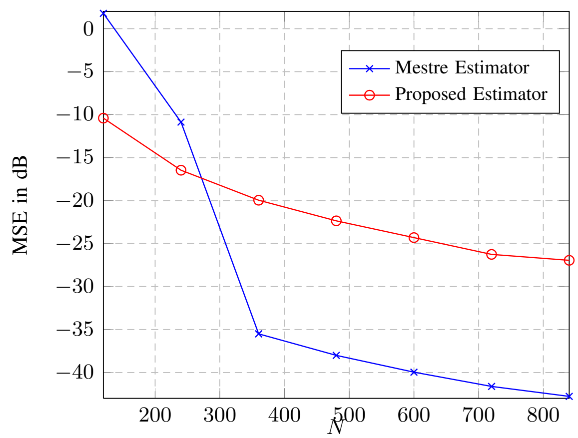

Fig. 3. Experienced MSE with N when $\frac{N}{M} = \frac{3}{20}$ and $(\rho_1, \rho_2, \rho_3) = (1, 3, 7)$

In the second experiment, we consider the case where the separability condition does not hold. In particular, we assume that the covariance matrix $\mathbf{R}_N$ has three different eigenvalues $(\rho_1, \rho_2, \rho_3) = (1, 2, 3)$, each with the same multiplicity, *i.e.* $\frac{N_1}{N} = \frac{N_2}{N} = \frac{N_3}{N} = \frac{1}{3}$. We also set the ratio between the dimension of variables and the number of samples $\frac{N}{M}$ to $3/8$, a ratio which is too high for the separability condition to hold true. We assume for our estimator that the multiplicities are not known, a hypothesis that obviously cannot be used for Mestre's estimator. We thus favour Mestre's estimator by assuming that it knows perfectly the multiplicities. Fig. 4 compares the obtained results in terms of MSE: for different values of M and N satisfying a constant ratio $c = N/M = 3/8$ and 1000 realizations. We note that as M and N increase, the estimator in [7] exhibits an error floor, underlying the fact that without the separability assumption, Mestre's estimators are no longer consistent.

B. Comparison with Bai, Chen and Yao's method

The estimator proposed in [9] and our proposed estimator are similar at first sight. The main difference lies in the intermediate quantities which are estimated before estimating the eigenvalues and their multiplicities. While the technique of [9] is based on the numerical computation of the empirical moments $\frac{1}{N}\operatorname{Tr}(\mathbf{Y}_N\mathbf{Y}_N^H)^k$, our technique rather relies on building consistent estimators of the theoretical moments $\frac{1}{N}\operatorname{Tr}\mathbf{R}_N^k$. This diffence induces important numerical consequences in the computation of the estimates: In [9], the functional relation between the quantities to be estimated and the empirical moments $\frac{1}{N}\operatorname{Tr}(\mathbf{Y}_N\mathbf{Y}_N^H)^k$ yields a system of equations whose resolution relies on iterative methods (based for instance on the functions `fsolve` or `fminsearch` in MATLAB) which are extremely slow.

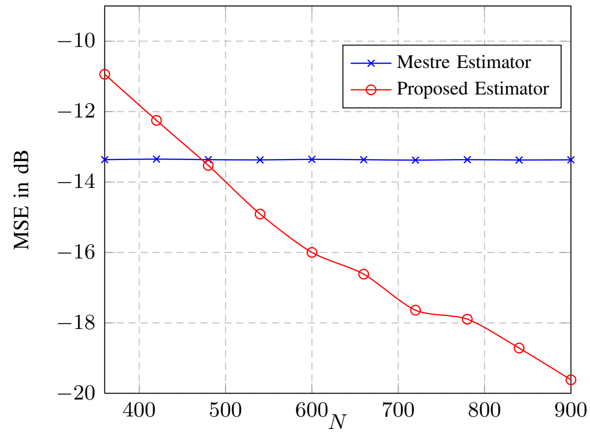

Fig. 4. Experienced MSE with N when $\frac{N}{M} = \frac{3}{8}$ and $(\rho_1, \rho_2, \rho_3) = (1, 2, 3)$

On the other hand, the method proposed in this article is based on a bijective system of equations that links the theoretical moments to the eigenvalues and their multiplicities, whose resolution relies on simple computations: A matrix inversion and solving a polynomial (see for instance end of Section III).

Simulation results indicate that our algorithm allows a great gain of complexity compared to [9], while keeping the same level of performance. Execution times for one realization are provided in the following table I for the same simulation setting as the second experiment. Note that unlike our method which exhibits low complexity, the complexity of the method of [9] tends to increase exponentially as the dimensions N and M increase.

N, M	Proposed method	Bai, Chen and Yao's method
$N = 300,\ M = 800$	0.5s	10.68s
$N = 360,\ M = 960$	0.55s	25.57s
$N = 420,\ M = 1120$	0.67s	42.62s

TABLE I

EXECUTION TIME TO OBTAIN AN ESTIMATOR FOR ONE REALIZATION (IN SECONDS)

C. Accuracy of the Gaussian approximation

Finally, we verify by simulations the accuracy of the Gaussian approximation. We consider the case where there are two different eigenvalues $\rho_1 = 1$ and $\rho_2 = 3$ that are uniformly distributed. Unlike the

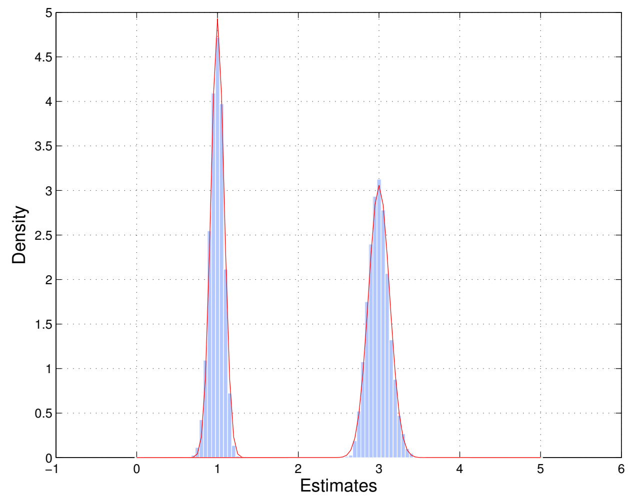

Fig. 5. Comparison of empirical against theoretical variances for $c_1 = c_2 = 0.5$ and $\rho_1 = 1$ and $\rho_2 = 3$

first experiment, we assume that the multiplicities are not knwon. We represent in Fig 5 the histogram for $\hat{\rho}_1$ and $\hat{\rho}_2$ when $N = 60$ and $M = 120$. We also represent in red line, the corresponding Gaussian distribution. We note that as it was predicted by our derived results, the histogram is similar to that of a Gaussian random variable.

VI. Conclusion

The present work is a theoretical contribution to the important problem of estimating the covariance matrices of large dimensional data. Two important assumptions (separability condition, exact knowledge of the multiplicity) have been in particular relaxed with respect to previous works. From a numerical point of view, it should be noticed however, that the situation is more contrasted: If the eigenvalues of $\mathbf{R}_N$ are far away from each other, then only the largest eigenvalue is well-estimated because in the expression of the moments, the term corresponding to the largest eigenvalue prevails. On the other hand, if the eigenvalues are too close to each other, matrix $\mathbf{\Gamma}$ is ill-conditioned, thus enlarging the induced error. These phenomenas are inherent to the moment method, and preliminary studies show that using trigonometric moments might help mitigating these numerical problems.

APPENDIX A
PROOF OF LEMMA 1

By Cauchy's formula, write:

$$\sum_{k=1}^{L}\frac{N_k}{N}\rho_k^{\ell}=\frac{1}{2\mathbf{i}\pi N}\oint_{\Gamma}\sum_{r=1}^{L}\frac{N_r\omega^{\ell}}{\omega-\rho_r}d\omega,$$

where Γ is a counterclockwise oriented contour that surrounds all the eigenvalues $\{\rho_1,\cdots,\rho_L\}$. Performing the changing variable $\omega=-\frac{1}{\underline{m}_N(z)}$ in the same manner as in [7], we get:

$$\sum_{k=1}^{L}\frac{N_k}{N}\rho_k^{\ell}=\frac{(-1)^{\ell+1}}{2\mathbf{i}\pi N}\oint_{\mathcal{C}}\sum_{r=1}^{L}\frac{N_r\underline{m}_N'(z)dz}{\underline{m}_N^{\ell+1}(z)\left(\rho_r\underline{m}_N(z)+1\right)},$$

where the contour $\mathcal{C}$ is counterclockwise oriented which contains the whole support $\mathcal{S}$.

From (3), we can establish that:

$$m_N(z)=-\frac{1}{Nz}\sum_{r=1}^{L}\frac{N_r}{1+\rho_r\underline{m}_N(z)},$$

thus yielding:

$$\sum_{k=1}^{L}\frac{N_k}{N}\rho_k^{\ell}=\frac{(-1)^{\ell}}{2\mathbf{i}\pi}\oint_{\mathcal{C}}\frac{z\underline{m}_N'(z)}{\underline{m}_N^{\ell+1}(z)}m_N(z)dz. \tag{20}$$

Plugging the relation:

$$m_N(z)=\frac{M}{N}\underline{m}_N(z)+\frac{M(1-\frac{N}{M})}{Nz}$$

into (20), we obtain:

$$\sum_{k=1}^{L}\frac{N_k}{N}\rho_k^{\ell}=\frac{(-1)^{\ell}}{2\mathbf{i}\pi}\oint_{\mathcal{C}}\frac{Mz\underline{m}_N'(z)dz}{N\underline{m}_N^{\ell}(z)}+\frac{(-1)^{\ell}}{2\mathbf{i}\pi}\oint_{\mathcal{C}}\frac{M(1-\frac{N}{M})\underline{m}_N'(z)}{N\underline{m}_N^{\ell+1}(z)}dz. \tag{21}$$

Since $\frac{\underline{m}_N'(z)}{\underline{m}_N^{\ell+1}(z)}$ is the derivative of $-\frac{1}{\ell\underline{m}_N^{\ell}(z)}$,

$$\oint_{\mathcal{C}}\frac{\underline{m}_N'(z)}{\underline{m}_N^{\ell+1}(z)}dz=0.$$

The second term on the right hand side of (21) is then equal to zero. It remains thus to deal with $\oint_{\mathcal{C}}\frac{z\underline{m}_N'(z)}{\underline{m}_N^{\ell}(z)}$. If $\ell\geq 2$, by integration by parts, we obtain:

$$\oint_{\mathcal{C}}\frac{z\underline{m}_N'(z)}{\underline{m}_N^{\ell}(z)}dz=\frac{1}{\ell-1}\oint_{\mathcal{C}}\frac{dz}{\underline{m}_N^{\ell-1}(z)}.$$

We thus obtain:

$$\sum_{k=1}^{L}\frac{N_k}{N}\rho_k^{\ell}=\frac{M(-1)^{\ell}}{2\mathbf{i}\pi N(\ell-1)}\oint_{\mathcal{C}}\frac{dz}{\underline{m}_N^{\ell-1}(z)}. \tag{22}$$

This proves that the theoretical moments admit the following integral representation:

$$\gamma_1 = -\frac{M}{2N\mathbf{i}\pi}\oint_{\mathcal{C}} \frac{z\underline{m}_N'(z)}{\underline{m}_N(z)}dz$$
$$\gamma_l = \frac{M(-1)^\ell}{2\mathbf{i}N(\ell-1)}\oint_{C} \frac{dz}{\underline{m}_N^{l-1}(z)}, \qquad 2 \le \ell \le 2L-1.$$

Finally, we show that consistent estimates of γ_i can be obtained by substituting the unknown term $\underline{m}_N(z)$ by its asymptotic equivalent $m_{\underline{\hat{\mathbf{R}}}_N}(z)$. Let $\hat{\gamma}_0, \cdots, \hat{\gamma}_{2L-1}$ the real quantities given by:

$$\begin{aligned}
\hat{\gamma}_0 &= 1, \\
\hat{\gamma}_1 &= -\tfrac{M}{2N\mathbf{i}\pi}\oint_{\mathcal{C}} \tfrac{zm'_{\underline{\hat{\mathbf{R}}}_N}(z)}{m_{\underline{\hat{\mathbf{R}}}_N}(z)}dz, \\
&\vdots \\
\hat{\gamma}_{2L-1} &= \tfrac{M(-1)^{2L-1}}{2N(2L-1)\mathbf{i}\pi}\oint_{\mathcal{C}} \tfrac{dz}{m^{2L-1}_{\underline{\hat{\mathbf{R}}}_N}(z)}.
\end{aligned}$$

Then, by the dominated convergence theorem and the fact that with probability one [24, Section 9.12], for all N, M large enough,

$$\inf_{z\in\mathcal{C}} |\underline{m}_N(z)| > 0$$

and

$$\inf_{z\in\mathcal{C}} |m_{\underline{\hat{\mathbf{R}}}_N}(z)| > 0,$$

one obtains: for all $k \ge 2$,

$$\left| \int_{\mathcal{C}} \frac{dz}{\underline{m}_N^{k-1}(z)} - \int_{\mathcal{C}} \frac{dz}{m^{k-1}_{\underline{\hat{\mathbf{R}}}_N}(z)} \right| \xrightarrow{a.s.} 0$$

and

$$\left| \int_{\mathcal{C}} \frac{m'_{\underline{\hat{\mathbf{R}}}_N}(z)dz}{m_{\underline{\hat{\mathbf{R}}}_N}(z)} - \int_{\mathcal{C}} \frac{\underline{m}_N'(z)dz}{\underline{m}_N(z)} \right| \xrightarrow{a.s.} 0.$$

Consequently:

$$\hat{\gamma}_i - \gamma_i \xrightarrow[N,M\to\infty]{a.s.} 0.$$

APPENDIX B
CALCULATION OF THE VARIANCE

In this section, we will show the calculations of the variance matrix $\mathbf{V}$. The computation of $\mathbf{V}_{1,1}$ has been carried out in [15] where it was shown that:

$$\mathbf{V}_{1,1} = -\frac{1}{4\pi^2c^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} - \frac{1}{(z_1-z_2)^2}\right]\frac{1}{\underline{m}(z_1)\underline{m}(z_2)}dz_1dz_2,$$

with $\mathcal{C}_1$ and $\mathcal{C}_2$ defined in the theorem. Using the fact that $\inf_{z\in\mathcal{C}} |\underline{m}(z)| > 0$ together with Fubini's theorem, the quantity $\mathbf{V}_{k,\ell}$ for $k \geq 2, \ell \geq 2$, becomes:

$$\mathbf{V}_{k,\ell} = -\frac{(-1)^{k+\ell}}{4\pi^2 c^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \mathbb{E}\left[X(z_1)X(z_2)\right] \underline{m}^{-k}(z_1)\underline{m}^{-\ell}(z_2) dz_1 dz_2.$$

Substituting $\mathbb{E}\left[X(z_1)X(z_2)\right]$ by $\kappa(z_1, z_2)$, we obtain:

$$\mathbf{V}_{k,\ell} = -\frac{(-1)^{k+\ell}}{4\pi^2 c^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2}\right] \frac{1}{\underline{m}^k(z_1)\underline{m}^\ell(z_2)} dz_1 dz_2.$$

Finally, it remains to compute $\mathbf{V}_{k,1}$. Expanding $\Upsilon(X, Y, \underline{m}, \underline{m}')$ and $\Phi_k(X, \underline{m})$, we obtain:

$$\begin{aligned}\mathbf{V}_{k,1} &= -\frac{(-1)^{k+1}}{4\pi^2 c^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \left[\frac{z_2}{\underline{m}(z_2)\underline{m}^k(z_1)} \mathbb{E}\left[X(z_1)X'(z_2)\right] dz_1 dz_2 - \frac{\underline{m}'(z_2)}{\underline{m}(z_2)^2 \underline{m}^k(z_1)} \mathbb{E}\left[X(z_1)X(z_2)\right]\right] dz_1 dz_2 \\ &= -\frac{(-1)^{k+1}}{4\pi^2 c^2} \left(\oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{z_2 \partial_2 \kappa(z_1, z_2)}{\underline{m}(z_2)\underline{m}(z_1)^k} dz_1 dz_2 - \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{\underline{m}'(z_2)\kappa(z_1, z_2)}{\underline{m}^2(z_2)\underline{m}^k(z_1)} dz_1 dz_2\right).\end{aligned}$$

By integration by parts, we obtain:

$$\oint_{\mathcal{C}_2} \frac{z_2 \partial_2 \kappa(z_1, z_2)}{\underline{m}(z_2)\underline{m}^k(z_1)} dz_2 = -\oint_{\mathcal{C}_2} \frac{\kappa(z_1, z_2)}{\underline{m}(z_2)\underline{m}^k(z_1)} dz_2 + \oint_{\mathcal{C}_2} \frac{\underline{m}'(z_2)\kappa(z_1, z_2)}{\underline{m}(z_2)^2 \underline{m}^k(z_1)} dz_2.$$

Hence,

$$\mathbf{V}_{k,1} = -\frac{(-1)^{k+1}}{4\pi^2 c^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{\kappa(z_1, z_2) dz_1 dz_2}{\underline{m}(z_2)\underline{m}^k(z_1)}.$$

This extends the expression of $\mathbf{V}_{k,\ell}$ for any $k, \ell \in \{1, \cdots, L-1\}$, thus yielding:

$$\mathbf{V}_{k,\ell} = -\frac{(-1)^{k+\ell}}{4\pi^2 c^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2}\right] \frac{1}{\underline{m}^k(z_1)\underline{m}^\ell(z_2)} dz_1 dz_2. \tag{23}$$

References

[1] I. Jolliffe, *Principal Component Analysis*. New York: Springer Verlag, 1986.

[2] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[3] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR Filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, Feb. 1995.

[4] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[5] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Signal, Speech and Signal Processing*, vol. 33, no. 2, p. 387392, 1985.

[6] 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.

[7] ——, "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.

[8] 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.

[9] Z. Bai, J. Chen, and J. Yao, "On estimation of the population spectral distribution from a high-dimensional sample covariance matrix," *Australian and New Zealand Journal of Statistics*, vol. 52, no. 4, pp. 423–437, 2010.

[10] A. Kammoun, R. Couillet, J. Najim, and M. Debbah, "Performance of capacity inference methods under colored interference," *submitted to IEEE Trans. Inf. Th.*, 2011, [Online] `http://front.math.ucdavis.edu/1105.5305` .

[11] O. Ledoit and M. Wolf, "Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices ," *Institute for Empirical Research in Economics University of Zurich*, October 2010. [Online]. Available: http://ssrn.com/abstract=1693836

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

[13] V. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysics J. Roy. Astron. Soc.*, vol. 33, pp. 347–366, Nov. 1973.

[14] 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.

[15] J. Yao, R. Couillet, J. Najim, and M. Debbah, "Fluctuations of an improved population eigenvalue estimator in sample covariance matrix model," 2011. [Online]. Available: http://arxiv.org/abs/1108.5266

[16] P. Vallet, P. Loubaton, and X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1043 –1068, feb. 2012.

[17] R. Couillet, J. Silverstein, Z. Bai, and M. Debbah, "Eigen-inference for energy estimation of multiple sources," *Information Theory, IEEE Transactions on*, vol. 57, no. 4, pp. 2420 –2439, april 2011.

[18] F. Rubio and X. Mestre, "Generalized Consistent Estimation on Low-rank Krylov Subspaces of Arbitrarily High Dimension," *IEEE Trans. Signal Process.*, vol. 57, no. 10, 2009.

[19] 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.

[20] 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.

[21] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge: Cambridge University Press, 2011.

[22] 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.

[23] S. G. Krantz, *Function Theory of Several Complex Variables*. AMS Chelsea Publishing, 1992.

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

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

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

[27] 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.

[28] ——, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," *The Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.