Abstract—The problem of estimating the eigenvalues and eigenvectors of the covariance matrix associated with a multivariate stochastic process is considered. The focus is on finite sample size situations, whereby the number of observations is limited and comparable in magnitude to the observation dimension. Using tools from random matrix theory, and assuming a certain eigenvalue splitting condition, new estimators of the eigenvalues and eigenvectors of the covariance matrix are derived, that are shown to be consistent in a more general asymptotic setting than the traditional one. Indeed, these estimators are proven to be consistent, not only when the sample size increases without bound for a fixed observation dimension, but also when the observation dimension increases to infinity at the same rate as the sample size. Numerical evaluations indicate that the estimators have an excellent performance in small sample size scenarios, where observation dimension and the sample size are comparable in magnitude.

Index Terms—Eigenvalues, eigenvectors, G-estimation, random matrix theory, sample covariance matrix.

I. INTRODUCTION

EIGENVALUES and eigenvectors of covariance matrices are extensively used in multiple applications of signal processing, pattern classification, econometrics, decision theory, and statistical inference, among other fields. There exist multiple estimation and detection procedures, such as direction of arrival detection [1], channel estimation [2], multiuser detection [3], or model order selection [4], which are precluded by the estimation of the eigenvalues and eigenvectors of the covariance matrix of the observation. In all these procedures, it is important to depart from a good estimation of the eigenvalues and/or the eigenvectors of the observation covariance matrix, which is in practice unknown.

The problem of estimating the eigenvalues and eigenvectors of a covariance matrix has usually been solved using the sample covariance matrix constructed from the observations. The eigenvalues and eigenvectors of the true covariance matrix are usually estimated as the eigenvalues and eigenvectors of the sample covariance matrix, which in what follows will be referred to as sample eigenvalues and sample eigenvectors, respectively. Although these estimators are consistent for large sample sizes, it turns out that they have serious deficiencies in the small sample size regime, where the sample size is comparable in magnitude to the observation dimension. It is well known, for instance, that the sample eigenvalues tend to be more spread out than the original ones, so that the largest (resp., the smallest) sample eigenvalue tends to overestimate (resp., underestimate) the corresponding eigenvalue of the true covariance matrix.

Let us briefly introduce the problem formulation in its full generality. To that effect, we consider a collection of $N$ independent and identically distributed (i.i.d.) observations of a certain $M$-dimensional stochastic process, denoted by $y(1), \ldots, y(N)$, where $y(n) \in \mathbb{C}^M$, $n = 1 \ldots N$. We assume that these observations have zero mean and covariance matrix $R_M \in \mathbb{C}^{M \times M}$. We will denote by $\gamma_1 < \gamma_2 < \cdots < \gamma_M$ the set of pairwise different eigenvalues of the covariance matrix $R_M$, where here $M$ is the number of distinct true eigenvalues ($1 \leq M \leq M$). Each of the eigenvalues $\gamma_m$, $m = 1, \ldots, M$, is known to have multiplicity $K_m$, $m = 1, \ldots, M$, so that $M = \sum_{m=1}^{M} K_m$. Associated with each eigenvalue $\gamma_m$, $m = 1, \ldots, M$, there is a complex subspace of dimension $K_m$. This subspace is determined by an $M \times K_m$ matrix of eigenvectors, denoted by $E_m$, such that $E_m E_m^H = I_{K_m}$. Note that this specification is unique up to right multiplication by an orthogonal matrix. Hence, we can write

$$R_M = \sum_{i=1}^{M} \gamma_i E_i E_i^H = E \begin{bmatrix} \gamma_1 I_{K_1} & & \\ & \ddots & \\ & & \gamma_M I_{K_M} \end{bmatrix} E^H,$$

where $E \in \mathbb{C}^{M \times M}$ contains all the eigenvector matrices, namely, $E = [E_1, \ldots, E_M]$. Now, since the eigenvector matrices associated to a particular eigenvalue are defined up to right multiplication by an orthogonal matrix, it is more convenient to formulate the problem in terms of eigenvector orthogonal projection matrices, defined as

$$P_m = E_m E_m^H, \quad m = 1 \ldots M.$$ 

Contrary to the eigenvector matrices, the entries of these projection matrices are always univocally specified. Furthermore, one can easily recover $E_i$ from $P_i$ using standard algebraic methods.

Problem statement: Assuming that the multiplicities of the eigenvalues $\{K_1, \ldots, K_M\}$ are known, estimate the true eigenvalues $\gamma_1, \ldots, \gamma_M$ and their associated eigenvector projection matrices $P_1, \ldots, P_M$ using the observations $y(1), \ldots, y(N)$. 

Manuscript received July 20, 2006; revised July 15, 2008. This work was supported in part by the Catalan Government under Grant SGR2005-00690 and the European Commission under Project NEWCOM++ 7FP-ICT-216715. The material in this paper was presented in part at the 2006 Second International Symposium on Communications and Control, Marrakech, Morocco, March 2006.

The author is with the Centre Tecnològic de Telecomunicacions de Catalunya Parc Mediterrani de la Tecnologia, Castelldefels (Barcelona), Spain (e-mail: xavier.mestre@cttc.cat).

Communicated by X. Wang, Associate Editor for Detection and Estimation.

Color versions of Figures 1 and 4–7 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.929938
The classical method to estimate these quantities is based on the sample covariance matrix, which is constructed as

$$\hat{R}_M = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}(n)\mathbf{y}^H(n).$$

The eigenvalues and associated eigenvectors of the sample covariance matrix $\hat{R}_M$ will be denoted by $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$ and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_M$ respectively, so that

$$\hat{R}_M = \sum_{i=1}^{M} \lambda_i \mathbf{e}_i \mathbf{e}_i^H.$$ 

Throughout the paper, we will refer to $\{\lambda_i\}$ and $\{\mathbf{e}_i\}$ as sample eigenvalues and sample eigenvectors, respectively.

Let us first concentrate on the estimation of the $m$th eigenvalue of the true covariance matrix $R_M$, $\gamma_m$, which has known multiplicity $K_m$. Let $K_m$ be the set of indices

$$K_m = \{1, 2, \ldots, m\}.$$ 

(The cardinality of $K_m$ is the equal to the multiplicity of the eigenvalue $\gamma_m$, namely, $K_m$). Now, the classical estimator of $\gamma_m$ eigenvalue of the true covariance matrix $R_M$, is given by

$$\hat{\gamma}_m^{\text{cl}} = \frac{1}{K_m} \sum_{k \in K_m} \lambda_k.$$ 

(1)

This is, indeed, the maximum-likelihood estimator of $\gamma_m$ when the observations $\{\mathbf{y}(n), n = 1 \cdots N\}$ follow a multivariate Gaussian distribution (see [5, Theorem 2] and [6, Theorem 9.3.1]). In the particular case where all the true eigenvalues have multiplicity one ($K_1 = \cdots = K_M = 1$), the estimated eigenvalues are the sample eigenvalues.

On the other hand, the eigenvectors of the true covariance matrix are traditionally estimated as the sample eigenvectors. In terms of eigenvector orthogonal projection matrices, the classical estimator is given by

$$\hat{P}_m^{\text{cl}} = \sum_{k \in K_m} \mathbf{e}_k \mathbf{e}_k^H.$$ 

(2)

Under some standard statistical assumptions, the two estimators in (1) and (2) are strongly $N$-consistent (i.e., almost surely consistent as the number of samples goes to infinity, $N \to \infty$), and their joint asymptotic distribution has been extensively studied in the literature (again, as $N \to \infty$) [5], [7], [8].

Now, the main flaw of these traditional estimators in (1) and (2) is the fact that they are designed to provide good estimates whenever the sample size $N$ is sufficiently high in terms of the observation dimension $M$. In the small sample size regime (i.e., when the number of samples $N$ is comparable in magnitude to the observation dimension $M$), these estimators do not need to provide the optimum behavior, and better estimators can be obtained using other approaches.

A great deal of research has been devoted to improving the estimation of the eigenvalues and eigenvectors when only a finite collection of observations are available. Regarding the estimation of the true eigenvalues, most of the work has been done under the assumption that the observation follows a Gaussian distribution, or assuming that the sample size is large compared to the observation dimension. For example, in [9], Anderson proposed an estimator that corrects the asymptotic ($N \to \infty$) bias of the sample eigenvalues. Assuming that all the true eigenvalues have multiplicity one, this estimator takes the form

$$\hat{\gamma}_m^{A} = \lambda_m \left(1 - \frac{1}{N} \sum_{k \in K_m} \lambda_k \right).$$

Other estimators of the eigenvalues are derived as the minimizers of a particular risk function (see [10] for an extensive review of the most classical eigenvalue estimators of this type). Within this category of eigenvalue estimators one can find the original Stein estimator [10] which, assuming again that all the true eigenvalues have multiplicity one, takes the form

$$\hat{\gamma}_m^{S} = \lambda_m \left(1 + \frac{1}{N} \sum_{k \in K_m} \frac{\lambda_k + 0.5}{\lambda_m - \lambda_k}\right).$$

Other similar examples are the Haff estimator [11] or the estimator proposed by Dey and Srinivasan [12]. In all these cases, the estimators tend to move the sample eigenvalues in an intuitively appealing way and revert to the traditional sample eigenvalue estimators when the sample size $N$ is sufficiently high compared to the observation dimension ($M$). More recent approaches include diverse methods based on large sample estimates [13], estimators under a quadratic loss function [14], [15], or a method based on the estimation of the characteristic polynomial [16]. As for the improved estimation of the eigenvectors of the covariance matrix from a finite number of samples, results are much harder to come by. The work in the literature has concentrated on estimation of eigenvectors under prior information [13] and on test hypothesis constructed from the sample eigenvectors [17].

In this paper, we take a different approach in order to derive improved estimators of the eigenvalues and associated eigenvectors. We focus on a realistic situation where the sample size $N$ and the observation dimension $M$ are comparable in magnitude. However, instead of focusing on the finite situation (as in the examples above), we consider the asymptotic situation where these two quantities are large but their quotient converges to a fixed finite quantity. In other words, we assume that the sample size is a function of the observation dimension, i.e., $N = N(M)$, so that $\lim_{M \to \infty} N(M) = +\infty$ but $\lim_{M \to \infty} N(M)/M = 1/c$, where $0 < c < \infty$ is a fixed quantity as in a real situation. In practice, estimators that are consistent in this generalized asymptotic regime are more robust to the presence of a finite sample size than other estimators which are only guaranteed to converge for $N \to \infty$. In what follows, and for the sake of simplicity we will drop the argument of the sample size, which will be denoted by $N$, and we will also let $c$ denote both the quotient $M/N$ and its asymptotic limit (it will be clear from the context which of the two definitions is being used).

Hence, the objective of this paper is to derive estimators of the eigenvalues and the eigenvectors of the covariance matrix that converge, not only when $N \to \infty$ (as it is the case of the traditional sample estimators) but also when the observation dimension $M$ grows to infinity at the same rate as the number

Authorized licensed use limited to: IEEE Xplore. Downloaded on December 13, 2008 at 19:17 from IEEE Xplore. Restrictions apply.
of observations, i.e., \( M, N \to \infty \) at the same rate. This guarantees that the estimators are consistent even when the sample size and the observation dimension are comparable in magnitude. In practice, it is observed that estimators that are designed under these more generic conditions outperform their traditional counterparts in the small sample size regime, where by definition the number of samples per observation dimension is a finite quantity.

The first question that we need to answer is whether the traditional estimators in (1) and (2) are \( M, N \)-consistent (i.e., consistent as \( M, N \to \infty \) at the same rate). Intuition suggests that, given the fact that these estimators perform quite poorly in the small sample size regime, they are not \( M, N \)-consistent. This was confirmed in [18], where the asymptotic behavior of these two estimators \( M, N \to \infty \) at the same rate was investigated. Under some assumptions that will be specified later, two asymptotic deterministic equivalents of the traditional estimators in (1) and (2) were derived in [18], showing that these estimators fail to provide consistent estimates as \( M, N \to \infty \) at the same rate. This in turn raises the question of whether \( M, N \)-consistent estimators exist and, if so, how to calculate them. Answering these questions will be the main objective of this paper. We will see that, under some circumstances, \( M, N \)-consistent estimators of the eigenvalues and eigenvectors exist, and we will present a novel method to calculate them in closed analytical form.

Before proceeding any further, we need to clarify the operative meaning of \( M, N \)-consistency for eigenvector orthogonal projection matrices. Note that it makes little sense to try to estimate the eigenprojection matrices \( P_m = E_mE_m^H \) because their dimension increases to infinity as \( M \) grows large. Instead, we will try to estimate quadratic forms of the type

\[
\eta_m = s_1^H E_mE_m^H s_2
\]

where \( s_1, s_2 \) are two generic \( M \times 1 \) deterministic column vectors. In particular, if \( s_1, s_2 \) are selected as two columns of an \( M \times M \) identity matrix, \( \eta_m \) will be the corresponding entry of the projection matrix.

The rest of the paper is organized as follows. Section II presents some properties of the spectrum of large sample covariance matrices that are used in the rest of the paper and that are needed to understand the assumptions made in the statement of the main result. Section III presents the main result of the paper, namely, the proposed alternative \( M, N \)-consistent estimators. The result is proven in Section IV and numerically evaluated in Section V. Finally, Section VI concludes the paper.

II. PRELIMINARIES

The asymptotic characterization of the spectrum of sample correlation matrices when both \( M \) and \( N \) increase without bound at the same rate has received a lot of attention in the past decades. Random matrix theory has studied the asymptotic behavior of the eigenvalues and eigenvectors of different random matrix models, including the sample covariance matrix. Traditionally, random matrix theory focuses on the empirical distribution functions of the sample eigenvalues, defined as

\[
\hat{Q}_M(\lambda) = \frac{1}{M} \# \{ k : \hat{\lambda}_k \leq \lambda \}
\]

where \( \# \{ \cdot \} \) denotes the cardinality of a set. It turns out that the asymptotic characterization of \( \hat{Q}_M(\lambda) \) is in practice too complex, so that instead mathematicians have concentrated on the convergence of the so-called Stieltjes transform of \( \hat{Q}_M(\lambda) \), which is defined as

\[
\hat{b}_M(z) = \int \frac{1}{\lambda - z} d\hat{Q}_M(\lambda) = \frac{1}{M} \sum_{r=1}^{M} \frac{1}{\hat{\lambda}_r - z} = \frac{1}{M} \text{tr} \left[ \left( \hat{R}_M - zI_M \right)^{-1} \right]
\]

for \( z \in \mathbb{C}^+ \equiv \{ z \in \mathbb{C} : \text{Im}(z) > 0 \} \). It turns out that the distribution function \( \hat{Q}_M(\lambda) \) can be recovered from the corresponding Stieltjes transform \( \hat{b}_M(z) \) using the Stieltjes inversion formula

\[
\hat{Q}_M(x) = \lim_{y \to 0^+} \int_{-\infty}^{\infty} \frac{1}{\pi} \text{Im} \left[ \hat{b}_M(\lambda + jy) \right] d\lambda
\]

and is valid for every \( x \in \mathbb{R} \). Hence, in order to characterize the asymptotic distribution of the sample eigenvalues, one can alternatively characterize the asymptotic behavior of the corresponding Stieltjes transform \( \hat{b}_M(z) \), and then use the Stieltjes inversion formula in (5).

The Stieltjes transform in (4) is appropriate to characterize the asymptotic behavior of the sample eigenvalues, but it is not useful to describe the asymptotic properties of the eigenvectors (note that \( \hat{b}_M(z) \) depends on the sample eigenvalues only, and not on the sample eigenvectors). Hence, for our purposes, it is also convenient to consider the following function:

\[
\hat{m}_M(z) = \sum_{r=1}^{M} \frac{s_1^H \hat{v}_r \hat{v}_r^H s_2}{\hat{\lambda}_r - z} = s_1^H \left( \hat{R}_M - zI_M \right)^{-1} s_2, \quad z \in \mathbb{C}^+
\]

where \( s_1, s_2 \) are two \( M \times 1 \) deterministic vectors. Note that this function depends on both the sample eigenvalues and the sample eigenvectors. They are a class of Stieltjes transforms that were introduced by Girko in e.g., [19], [20], and are very useful in order to characterize the asymptotic behavior of quadratic forms of sample eigenvectors.

Next, we present the convergence result that will be the basis for the development of this paper. It is formulated under the following assumptions.

(A1) The covariance matrix \( \hat{R}_M \) has uniformly bounded spectral norm for all \( M \).

(A2) The two deterministic vectors \( s_1, s_2 \) have uniformly bounded norm for all \( M \).

(A3) The sample covariance matrix takes the form \( \hat{R}_M = \hat{R}_M^{1/2} \hat{U}_M \hat{U}_M^H \hat{R}_M^{1/2} \), where \( \hat{R}_M^{1/2} \) is the \( M \times M \) Hermitian positive definite square root of the true covariance matrix \( R_M \), and where \( \hat{U}_M \) is an \( M \times N \) matrix with complex i.i.d. absolutely continuous random entries, each one of them having i.i.d. real and imaginary parts with zero mean, variance \( 1/(2N) \), and finite eighth order moments.

Theorem 1: For all \( z \in \mathbb{C}^+ \), under (A1–A3) and as \( M, N \to \infty \) at the same rate (\( M/N \to c \), \( 0 < c < \infty \))

\[
\begin{align*}
\left| \hat{b}_M(z) - \hat{b}_M(z) \right| \to 0
\end{align*}
\]
almost surely for all \( z \in \mathbb{C}^+ \), where \( \bar{b}_M(z) \) is the unique solution to the following equation in the set \( \{b \in \mathbb{C} : -(1-c)/z + \bar{b} \in \mathbb{C}^+ \} \):

\[
\bar{b} = \frac{1}{M} \sum_{r=1}^{M} K_r \frac{1}{\gamma_r (1-c - czb)} - z. 
\tag{8}
\]

Furthermore

\[
|\hat{m}_M(z) - \bar{m}_M(z)| \to 0
\tag{9}
\]
also almost surely for all \( z \in \mathbb{C}^+ \), where

\[
\bar{m}_M(z) = \frac{1}{M} \sum_{r=1}^{M} \frac{s_r^H E_r E_r^H s_2}{\gamma_r (1-c - czb)} - z
\tag{10}
\]
and \( \bar{m}_M(z) \) is as defined above.

Proof: The result in (7) is well known in the random matrix theory literature. For instance, in [21] this result was derived under more generic statistical assumptions, where the assumption concerning the finite moments in (A3) was not needed. On the other hand, the convergence of \( \hat{m}_M(z) \) in (9) was presented in [20], although part of the proof was omitted. A similar result has recently appeared in [22], again under a slightly different set of assumptions. In [23], we provide an alternative proof under the assumptions of this theorem.

An immediate consequence of Theorem 1 is the fact that, assuming that the empirical eigenvalue distribution of \( \bar{R}_M \) converges as \( M, N \to \infty \) at the same rate, the empirical eigenvalue distribution of \( \bar{R}_M \) tends almost surely to a nonrandom distribution that can be retrieved from the limiting \( \bar{b}_M(z) \) using the Stieltjes inversion formula in (5). For finite \( M \), the distribution associated to the Stieltjes transform \( \bar{b}_M(z) \) is a deterministic approximation to the actual sample eigenvalue distribution.

We will denote by \( q_M(x) \) the corresponding density, which can be retrieved from \( \bar{b}_M(z) \) as

\[
q_M(x) = \lim_{y \to 0} \frac{1}{\pi} \text{Im}\left[\bar{b}_M(z)\right]
\tag{11}
\]
for any \( x \in \mathbb{R}^N \equiv \mathbb{R} \setminus \{0\} \).

Remark 1: Observe that (8) and (10) provide a relationship between the sample eigenvalues/eigenvectors (related asymptotically to \( \bar{b}_M(z) \) and \( \bar{m}_M(z) \)) and the true ones, which appear explicitly in these two equations. This relationship is exploited in this paper in order to obtain estimators of the true quantities in terms of the sample estimates. After the submission of this paper, we became aware of an alternative method proposed in [24], where the author proposes to invert the equation in (8) by considering a discretized version of the equation and then applying convex optimization methods. The approach followed here is more analytical in the sense that, although it is derived under more restrictive assumptions, it will provide estimators in closed analytical form. It is also worth mentioning that, as pointed out by a reviewer of this paper, it is also possible to obtain \( M \), \( N \)-consistent estimators of the eigenvalues \( \gamma_r \) by exploiting the relationship between the moments of the sample covariance matrix, namely, \( \frac{1}{M} \text{tr}\left[\bar{R}_M^k\right], k \in \mathbb{N} \), and those of the true covariance matrix, \( \frac{1}{M} \text{tr}\left[R_M^k\right], k \in \mathbb{N} \), see [28], [29], and then solving the resulting nonlinear equations that relate the moments \( \frac{1}{M} \text{tr}\left[\bar{R}_M^k\right] \) with the eigenvalues. Even though the computational complexity of this method is bound to be higher than the one proposed herein, it would be interesting to compare the performance of both estimators; this is, however, out of the scope of this paper.

For the purposes of this paper, it is important to understand the behavior of \( q_M(x) \) as a function of \( c \). In Fig. 1, we represent the form of \( q_M(x) \) for different values of \( c \) in a situation where the true eigenvalues were \( \{1, 2, 3, 7\} \), all with the same multiplicity. Observe that, as the number of samples per observation dimension increases \((1/c \to \infty)\), or equivalently, \( c \to 0 \), \( q_M(x) \) tends to concentrate around the four true eigenvalues. This is reasonable, because as \( N \to \infty \) for a fixed \( M \), the entries of \( \bar{R}_M \) tend almost surely to those of \( R_M \). Conversely, when the observation dimension is high compared to the number of samples (high \( c \)), \( q_M(x) \) presents only one cluster. In the example we are considering, for \( c = 0.01 \) (100 samples per observation dimension), we can clearly distinguish a different cluster for each true eigenvalue. However, for the case where \( c = 1 \) (one sample per observation dimension), \( q_M(x) \) consists of a single cluster.

In order to introduce the main result of the paper, it is important to have a full characterization of the support of \( q_M(x) \). This has been analyzed by a number of authors (see, e.g., [25]–[27]). The result can be summarized in the following proposition, which shows that the support of \( q_M(x) \) is composed of the union of a set of clusters associated with the true eigenvalues.

Proposition 1: The support of \( q_M(x) \) in \( \mathbb{R}^N, x \in \mathbb{R}^N \equiv \mathbb{R} \setminus \{0\} \), is given by the union of \( Q \) disjoint compact intervals, namely

\[
S_M = [x_Q^-, x_Q^+] \cup \cdots \cup [x_Q^+, x_Q^-]
\tag{12}
\]
where \( x_Q^- = \Phi \left( f_q^- \right) \), \( x_Q^+ = \Phi \left( f_q^+ \right) \), and

\[
\Phi(f) = f \left( 1 - c \frac{1}{M} \sum_{r=1}^{M} K_r \frac{\gamma_r}{\gamma_r - f} \right)
\tag{13}
\]
and where \( 2Q \) is the total number of real-valued solutions counting multiplicities of the equation

\[
\frac{1}{M} \sum_{r=1}^{M} K_r \left( \frac{\gamma_r}{\gamma_r - f} \right)^2 = \frac{1}{c}
\tag{14}
\]
which are denoted by \( f_1^- < f_2^+ \leq f_2^- \leq \cdots \leq f_Q^- < f_Q^+ \) (where equality is understood to hold when there exists a solution with double multiplicity).

Proof: See [18], [27].

This result can be interpreted as follows. In Fig. 2, we give a typical representation of the function on the left-hand side of (14). The solutions of the equation can be graphically represented as the crossing points with a horizontal line at \( 1/c \). The number of real-valued solutions counting multiplicities is always even (counting multiplicities), and each pair of solutions \( [f_q^-, f_q^+] \), \( q = 1 \ldots Q \), determines the position of one of
in a situation where the distribution of eigenvalues of $\mathbf{R}_M$ is de

more particularly, each cluster is sup-

always consists of a single cluster, and

with the corresponding interval

are needed

in

anyway follows that the cluster of the asymptotic eigenvalue dis-

this cluster splits up into smaller ones as $c$ decreases to zero. For

a sufficiently low $c$, the asymptotic distribution of eigenvalues

will consist of exactly $\bar{M}$ different clusters (each one associated

with one of the distinct eigenvalues $\gamma_m, m = 1, \ldots, \bar{M}$).

For the purposes of this paper, it is useful to establish how

low the value of $c$ must—or, equivalently, how many samples

$N$ for a fixed observation dimension $M$ are needed—in order for

a cluster associated with a particular eigenvalue of the true
covariance matrix to separate from the rest of the distribution in

$q_M(x)$. From the description given in Proposition 1, it triv-

ialy follows that the cluster of the asymptotic eigenvalue dis-

tribution corresponding to the eigenvalue $\gamma_m$ is separated from

the clusters associated with adjacent eigenvalues if and only if

$1/c > \xi_M(m)$, where $\xi_M(m)$ is defined in (15) at the top of

the following page, and $\bar{j}_k, k = 1, \ldots, \bar{M} - 1$, being the $\bar{M} - 1$
different real-valued solutions to the equation

$$
\frac{1}{M} \sum_{r=1}^{\bar{M}} K_r \frac{\gamma_m^2}{(\gamma_r - f)^3} = 0
$$

(16)

ordered as $\bar{j}_1 < \bar{j}_2 < \cdots < \bar{j}_{\bar{M} - 1}$. This condition will be

essential in order to guarantee the consistency of the proposed

estimators as $M, N \to \infty$ at the same rate. For this reason, we

will further assume the following.

Fig. 1. Asymptotic eigenvalue distribution of $\mathbf{R}_M$ in a situation where the distribution of eigenvalues of $\mathbf{R}_M$ is given by four different eigenvalues 1, 2, 3, 7 with the same multiplicity.

Fig. 2. Typical representation of the function on the left-hand side of (14).

the clusters of $q_M(x)$. More particularly, each cluster is supported by an interval of the type $[x_q^-, x_q^+]$, $q = 1, \ldots, Q$, where $x_q^- = \Phi(f_q^-), x_q^+ = \Phi(f_q^+)$. Observing the position of the roots of (14), it can readily be seen that, given a particular eigenvalue of the true covariance matrix $\gamma_m$, there exists a unique $q \in \{1, \ldots, Q\}$ such that $\gamma_m \in (f_q^-, f_q^+)$, see further Fig. 2. We can therefore associate $\gamma_m$ with the corresponding interval $[x_q^-, x_q^+]$ of the support of $q_M(x)$. In general, however, this correspondence is not bijective, and different true eigenvalues can be associated with the same asymptotic sample eigenvalue cluster. The number of eigenvalue clusters increases with decreasing $c$ (or, equivalently, increasing number of samples per observation dimension). In particular, for sufficiently high $c$, the asymptotic distribution of the eigenvalues of $\mathbf{R}_M$ always consists of a single cluster, and this cluster splits up into smaller ones as $c$ decreases to zero. For a sufficiently low $c$, the asymptotic distribution of eigenvalues will consist of exactly $\bar{M}$ different clusters (each one associated with one of the distinct eigenvalues $\gamma_m, m = 1, \ldots, \bar{M}$).
(As4) If $\gamma_m, m \in \{1, \ldots, \tilde{M}\}$, is the eigenvalue that is to be estimated, then

$$\inf_{\tilde{M}} \left\{ \frac{N}{\tilde{M}} - \xi_M(m) \right\} > 0$$

where $\xi_M(m)$ is as defined in (15).

### III. MAIN RESULT

As pointed out in the Introduction, in order to motivate the search for an $M, N$-consistent estimator of the quantities $\gamma_m$ and $\eta_m$, one should first prove that the traditional estimators in (1) and (2) are inconsistent in this asymptotic regime, that is, they fail to provide consistent estimates when the observation dimension goes to infinity at the same rate as the sample size ($M, N \to \infty, M/N \to c$). The following result, proven in [18] using the same techniques as the ones introduced in Section IV, reveals this inconsistency.

**Theorem 2:** Consider the two traditional sample estimators of the eigenvalues and associated eigenprojection matrices, namely

$$\xi^\text{trad}_{m} = \frac{1}{K_m} \sum_{k \in K_m} \lambda_k, \quad \eta^\text{trad}_{m} = \frac{1}{K_m} \left( \sum_{k \in K_m} \hat{e}_k \hat{e}_k^H \right) s_2.$$ 

Under (As1–As4), and as $M, N \to \infty$ at the same rate ($M/N \to c$, $0 < c < \infty$), $\xi^\text{trad}_{m} \to \xi^\text{m} \to 0$ and $\eta^\text{trad}_{m} \to \eta^\text{m} \to 0$ almost surely, where $\xi^\text{trad}_{m}$ and $\eta^\text{trad}_{m}$ are defined as

$$\xi^\text{trad}_{m} = \sum_{k=1}^{M} w_m(k) s_k^H E_k E_k^H s_2,$$

$$w_m(k) = \begin{cases} \frac{1}{K_m} \sum_{k \in K_m} K_r \left( \frac{\gamma_r}{\gamma_r - \gamma_m} \right), & k = m \\ \frac{1 - \frac{1}{K_m} \sum_{k \in K_m} K_r \left( \frac{\mu_m - \mu_k}{\gamma_m - \gamma_m} \right)}{\gamma_m - \gamma_m - \mu_k - \mu_m}, & k \neq m \end{cases}$$

where $\mu_m$ is the $m$th solution to the following equation in $\mu$:

$$\frac{1}{M} \sum_{r=1}^{M} K_r \frac{\gamma_r}{\gamma_r - \mu_m} = \frac{1}{c}$$

under the convention $\mu_1 < \cdots < \mu_{\tilde{M}}$.

**Proof:** See [18].

This theorem establishes the fact that traditional sample estimators of both the eigenvalues and the associated subspaces are not $M, N$-consistent, because in general $\xi^\text{trad}_{m} \neq \gamma_m$ and $\eta^\text{trad}_{m} \neq \eta_m$. They are, as is well known, $N$-consistent estimators (this can be observed from the above expressions, taking limits as $c \to 0$ and noting that $\mu_m \to \gamma_m$). The lack of $M, N$-consistency explains the presence of heavy biases when $M, N$ are both finite. In what follows, we present two alternative estimators of these quantities that are not only $N$-consistent, but also $M, N$-consistent.

**Theorem 3:** Under (As1–As4), the following quantities are strongly $M, N$-consistent estimators of $\gamma_m$ and $\eta_m$, respectively:

$$\gamma_m = \frac{N}{R_m} \sum_{k \in K_m} \left( \lambda_k - \hat{\mu}_k \right), \quad \eta_m = \sum_{k=1}^{M} \theta_m(k) s_k^H E_k E_k^H s_2,$$

where

$$\theta_m(k) = \begin{cases} \phi_m(k), & k \notin K_m \\ \eta_m(k), & k \in K_m \end{cases}$$

$$\phi_m(k) = \sum_{r \in K_m} \left( \frac{\lambda_r - \hat{\mu}_r}{\lambda_k - \lambda_r} \right) \frac{\lambda_r - \hat{\mu}_r}{\lambda_k - \hat{\mu}_k}$$

$$\psi_m(k) = \sum_{r \notin K_m} \left( \frac{\lambda_r - \hat{\mu}_r}{\lambda_k - \lambda_r} \right) \frac{\lambda_r - \hat{\mu}_r}{\lambda_k - \hat{\mu}_k}$$

and where $\hat{\mu}_1 \leq \hat{\mu}_2 \leq \cdots \leq \hat{\mu}_M$ are the real-valued solutions to the following equation in $\hat{\mu}$:

$$\frac{1}{M} \sum_{k=1}^{M} \frac{\lambda_k - \hat{\mu}}{\lambda_k - \hat{\mu}} = \frac{1}{c}$$

repeated according to their multiplicity. When $c > 1$, we use the convention $\hat{\mu}_1 = \cdots = \hat{\mu}_{[M-N]+1} = 0$, whereas $\hat{\mu}_{[M-N]+2}, \ldots, \hat{\mu}_M$ contain the positive solutions to the above equation.

**Proof:** See Section IV.

Let us now analyze in more detail the form of the proposed estimators $\gamma_m$ and $\eta_m$. In the case of the $m$th eigenvalue $\gamma_m$, instead of using an average of the sample eigenvalues, the proposed estimator takes the sample eigenvalues $\hat{\lambda}_k$ and subtracts the corresponding solutions to the equation in $\hat{\mu}$ (19). It can be shown that $\hat{\lambda}_k > \hat{\mu}_k$ whenever these quantities are different from zero, and hence the result $\hat{\lambda}_k - \hat{\mu}_k$ is always positive. The resulting quantities are averaged and multiplied by the sample size $N$.

As for the proposed estimate of the subspace associated with the $m$th eigenvalue, $\eta_m$, the proposed estimator makes use of all the eigenvectors of the sample covariance matrix and not only

\(^1\)In the expression of $\phi_m(k)$ and $\psi_m(k)$ we use the convention that any term of the form $0/0$ is identically zero.
those associated with the eigenvalue in question (as it is the case in the traditional estimator). The estimator applies a different weight to the eigenvectors depending on whether they are associated with the $m$th true eigenvalue or any other eigenvalue. It can easily be shown that $\phi_m(k) > 0$ and $\psi_m(k) > 0$, so that the eigenprojection estimator is always based on a proper combination of all the different subspaces associated with the sample covariance matrix.

Remark 2: When the sample size is high compared to the observation dimension, these two estimators revert to their traditional counterparts. This can be easily shown by noting that, for any fixed $M$

$$\hat{\lambda}_k \longrightarrow \tilde{\lambda}_k, \quad N(\tilde{\lambda}_k - \hat{\lambda}_k) \longrightarrow \tilde{\lambda}_k$$

as $N \rightarrow \infty$. This way

$$\hat{\gamma}_m \longrightarrow \frac{1}{K_m} \sum_{k \in K_m} \tilde{\lambda}_k = \tilde{\gamma}_m$$

as $N \rightarrow \infty$. On the other hand, also $\phi_m(k) \longrightarrow 0$ and $\psi_m(k) \longrightarrow 0$, which implies that $\tilde{\gamma}_m \longrightarrow \tilde{\gamma}_m$ as $N \rightarrow \infty$.

IV. PROOF OF THEOREM 3

In order to prove Theorem 3, we start by noticing that we can express the two quantities that need to be estimated in integral form as follows:

$$\gamma_m = \frac{M}{K_m} \int_{C^-} \left( \frac{1}{M} \sum_{r=1}^{M} K_r \frac{\omega}{\gamma_r - \omega} \right) d\omega$$

(20)

$$\eta_m = \frac{1}{2 M} \int_{C^-} \left( \sum_{r=1}^{M} s_r^H E_r E_r^H s_2 \gamma_r - \omega \right) d\omega$$

(21)

where $C^- = \{ \gamma_1, \ldots, \gamma_{33} \}$ is a negatively oriented contour taking values on $\mathbb{C} \setminus \{ \gamma_1, \ldots, \gamma_{33} \}$ and enclosing only $\gamma_m$. Now, the main idea behind the proof is based on carefully choosing the integration contour $C^- = \{ \gamma_1, \ldots, \gamma_{33} \}$ in the above integrals.

The integration contour is chosen as follows. Consider once again the function $\tilde{b}_m(z)$ defined in Theorem 1 for $z \in \mathbb{C}^+$. We can trivially extend this definition to $z \in \mathbb{C}^+ \equiv \{ z \in \mathbb{C} : \text{Im}(z) < 0 \}$ as $\tilde{b}_m(z) = \tilde{b}_m(z^*)$. On the other hand, it can be shown [26] that $\lim_{y \to 0^+} \tilde{b}_m(x + iy)$ exists for all $x \neq 0$, and therefore we can define

$$\tilde{b}_m(x) = \lim_{y \to 0^+} \tilde{b}_m(x + iy), \quad x \in \mathbb{R}^p \equiv \mathbb{R} \setminus \{0\}.$$  

With this, we have defined $\tilde{b}_m(z)$ on $\mathbb{C}^* \equiv \mathbb{C} \setminus \{0\}$. Next, consider the following function:

$$f_M(z) = \frac{z}{1 - c - cz M(z)}$$

(22)

valid for any $z \in \mathbb{C}^*$. For $z = 0$, we further define $f_M(0) = 0$ for $c \leq 1$ and $f_M(0) = \phi$ for $c > 1$, where $\phi$ is the unique positive solution to

$$\frac{1}{M} \sum_{r=1}^{M} K_r \frac{\gamma_r}{\gamma_r - \phi} = \frac{1}{c}.$$  

Remark 3: The function $f_M(z)$ can equivalently be defined as follows. For $z \in \mathbb{C}^+$ (respectively, $z \in \mathbb{C}^-$), $f_M(z)$ is chosen as the unique solution to the following equation in $f$:

$$f \left( 1 - c \frac{1}{M} \sum_{r=1}^{M} K_r \frac{\gamma_r}{\gamma_r - f} \right) = z$$

(23)

on the set $\mathbb{C}^+$ (respectively, the set $\mathbb{C}^-$). For $z \in \mathbb{R}$, we have two possible options. If

$$z \in \left( \mathfrak{x}_1^+, \mathfrak{x}_2^+ \right) \cup \cdots \cup \left( \mathfrak{x}_Q^+, \mathfrak{x}_Q^+ \right)$$

then there exists a unique solution $f$ to (23) on the set $\mathbb{C}^+$ and we define $f_M(z) = f$. Otherwise, if

$$z \in \mathbb{R} \setminus \left( \left( \mathfrak{x}_1^+, \mathfrak{x}_2^+ \right) \cup \cdots \cup \left( \mathfrak{x}_Q^+, \mathfrak{x}_Q^+ \right) \right)$$

then all the solutions to (23) are positive, but there exists a single solution $f$ with the property

$$\frac{1}{M} \sum_{r=1}^{M} K_r \left( \frac{\gamma_r}{\gamma_r - f} \right)^2 \leq \frac{1}{c}$$

(24)

and we choose $f_M(z) = f$.

The function $f_M(z)$ defined above has very interesting properties, which we summarize in the following proposition.

Proposition 2: The function $f_M(z)$ defined above has the following properties.

1) As a function of complex variable, $f_M(z)$ is holomorphic on $\mathbb{C} \setminus S_M$, where $S_M = \left( x_1^+, x_2^+ \right) \cup \cdots \cup \left( x_Q^+, x_Q^+ \right)$, defined in Proposition 1, is the support of $q_M(x)$ on $\mathbb{R}^p$.

2) When $z = x$ is restricted to the real axis, $f_M(x)$ is continuous on $\mathbb{R}$ and differentiable on $\mathbb{R}$ except for the points $\left( x_1^+, x_2^+, \ldots, x_Q^+, x_Q^+ \right)$. We denote by $f_M(x)$ the derivative of $f_M(x)$ on $\mathbb{R} \setminus \left( \left( x_1^+, x_2^+ \right) \cup \cdots \cup \left( x_Q^+, x_Q^+ \right) \right)$.

3) The real part of $f_M(x), x \in \mathbb{R}$, is monotonically increasing with $x$, going from $-\infty$ to $+\infty$ when $x$ grows from $-\infty$ to $+\infty$. Furthermore, $f_M(x_Q^-) = f_{Q}^+$ and $f_M(x_Q^+) = f_{Q}^-$. $q = 1 \ldots Q$, where $f_{Q}^+$ and $f_{Q}^-$ are defined in Proposition 1.

4) The imaginary part of $f_M(x), x \in \mathbb{R}$, is positive for $x \in \left( x_1^+, x_2^+ \right) \cup \cdots \cup \left( x_Q^+, x_Q^+ \right)$, and zero elsewhere. In other words, according to Remark 3 and Property 3, the imaginary part of $f_M(x)$ is positive only when the real part of $f_M(x)$ is in the set $\left( f_1^+ \ldots f_q^+ \right) \cup \cdots \cup \left( f_Q^+ \ldots f_Q^+ \right)$ and zero otherwise.

Proof: See [18, Proposition 2].

As a consequence of the last two properties, we can conclude that, as $x$ moves from $-\infty$ to $+\infty$, the complex function $f_M(x)$ describes a curve on the complex plane as shown in Fig. 3, where we have also represented the corresponding complex conjugate curve $f_M(x)$.

Note that, if a cluster associated with a particular eigenvalue $\gamma_m$ is separated from the rest of the
be very careful, though, because \( f'_M(x) \) is not bounded at the points \( \{ x_{-1}^{-1}, x_1^+, \ldots, x_Q^+, x_Q^- \} \).

**Proposition 3:** Under (Assumption 4), the integrals in (25) and (26) can be expressed as

\[
\begin{align*}
\gamma_m &= \frac{M}{K_m} \frac{1}{2\pi j} \oint \frac{1}{\gamma_r - f_M(z)} f'_M(z) dz \\
\eta_m &= \frac{1}{2\pi j} \oint \frac{1}{\gamma_r - f_M(z)} f'_M(z) dz
\end{align*}
\]  

(27)

(28)

where \( \partial \mathcal{R}_y(m) \) is a negatively oriented contour described by the boundary of the rectangle

\[
\mathcal{R}_y(m) = \{ z \in \mathbb{C} : \sigma_1 \leq \Re[z] \leq \sigma_2, |\Im[z]| \leq y \}
\]

\( y \) is a real positive value and \( \sigma_1, \sigma_2 \) are two real values such that \( (\sigma_1, \sigma_2) \) encloses the support of the cluster of \( \varrho_M(x) \) associated with \( \gamma_m \) and no other component of \( \varrho_M(x) \).

**Proof:** See Appendix I.

At this point, we can reformulate the problem of estimating \( \gamma_m \) and \( \eta_m \) as the problem of estimating the two equivalent integrals presented in Proposition 3. Note first that, since the two integrands are evaluated on the upper and lower complex plane, they can be estimated using the results of Theorem 1. More specifically, using the definition of \( f_M(z) \) in (22) we see that the two integrands in (27) and (28) can be expressed in terms of \( \mathcal{B}_M(z) \) and \( \mathcal{M}_M(z) \) as follows:

\[
\begin{align*}
h_M(z) &= \frac{1}{M} \sum_{r=1}^{M} \frac{f_M(z)}{\gamma_r - f_M(z)} f'_M(z) \\
&= \mathcal{B}_M(z)
\end{align*}
\]  

(29)

where in (a) we have used the fact that \( \mathcal{B}_M(z) \) is a solution to (8) and in (b) we have used the definition of \( \mathcal{M}_M(z) \) given in (10).

In the above equations, \( \mathcal{B}_M(z) \) denotes the complex derivative of \( \mathcal{B}_M(z) \) (this function is holomorphic on the upper and lower complex semiplanes).

It turns out that, by definition, \( \mathcal{B}_M(z) \) and \( \mathcal{M}_M(z) \) are (pointwise) consistently estimated by \( \mathcal{B}_M(z) \) and \( \mathcal{M}_M(z) \), respectively, for any \( z \in \mathbb{C} \setminus \mathbb{R} \) as \( M, N \to \infty \) at the same rate. Hence, it is a trivial consequence of Theorem 1 that the two integrands are pointwise strongly \( M, N \)-consistently estimated by the following two functions:

\[
\begin{align*}
h_M(z) &= z \mathcal{B}_M(z) \frac{1 - c + c^2 \mathcal{B}_M(z)}{(1 - c - c \mathcal{M}_M(z))^2} \\
\mathcal{M}_M(z) &= \mathcal{M}_M(z) \frac{1 - c + c^2 \mathcal{B}_M(z)}{1 - c - c \mathcal{M}_M(z)}
\end{align*}
\]
where \( \hat{h}_M(z) \) and \( \tilde{h}_M(z) \) are defined in (4) and (6), respectively, and where \( \hat{h}_M'(z) \) denotes the derivative\(^5\) of \( \hat{h}_M(z) \).

Although the pointwise convergence of the two integrands is easy to prove, it is not clear that we can replace the original integrands by the sample estimates in (29) without altering the asymptotic properties of the integrals. This question is answered in the following proposition.

**Proposition 4:** Under (As1–As4), the following integrals:

\[
\frac{M}{K_m} \frac{1}{2\pi i} \oint_{\mathcal{R}_m} \hat{h}_M(z) dz, \quad \frac{1}{2\pi i} \oint_{\mathcal{R}_m} \tilde{g}_M(z) dz
\]

are strongly \( M, N \)-consistent estimators of \( \gamma_m \) and \( \eta_m \), in the sense that

\[
\left| \frac{M}{K_m} \frac{1}{2\pi i} \oint_{\mathcal{R}_m} \hat{h}_M(z) dz - \gamma_m \right| \to 0, \\
\left| \frac{1}{2\pi i} \oint_{\mathcal{R}_m} \tilde{g}_M(z) dz - \eta_m \right| \to 0
\]

almost surely as \( M, N \to \infty \) at the same rate.

**Proof:** See Appendix III.

We will write

\[
\hat{\gamma}_m = \frac{M}{K_m} \frac{1}{2\pi i} \oint_{\mathcal{R}_m} \hat{h}_M(z) dz, \\
\hat{\eta}_m = \frac{1}{2\pi i} \oint_{\mathcal{R}_m} \tilde{g}_M(z) dz.
\]

In order to finish the proof of Theorem 3, we only need to find closed-form expressions for \( \hat{\gamma}_m \) and \( \hat{\eta}_m \) and show that they correspond to the ones given in the statement of the theorem. We note from (29) that \( \hat{h}_M(z) \) and \( \tilde{g}_M(z) \) are holomorphic on \( \mathbb{C}^+ \), except for two sets of poles: 1) The eigenvalues of \( \mathcal{R}_M \), namely, \( \lambda_1 < \cdots < \lambda_M \). These poles correspond to singularities of \( \hat{h}_M(z) \) on the complex plane. 2) The values that null out the denominators in (29), i.e., the real-valued solutions to the following equation in \( \mu \):

\[
\frac{1}{M} \sum_{k=1}^{M} \frac{\hat{\lambda}_k}{\lambda_k - \mu} = \frac{1}{c}.
\]

If \( c < 1 \), all the sample eigenvalues \( \hat{\lambda}_k, k = 1, \ldots, M \), are different with probability one. Consequently, there exist exactly \( M \) different solutions to (31), which will be denoted as \( \mu_1, \ldots, \mu_M \). By observing the position of the asymptotes of the expression on the left-hand side of (31), we can conclude that these solutions are all positive and they are interlaced with the sample eigenvalues as follows:

\[
0 < \mu_1 < \hat{\lambda}_1 < \mu_2 < \cdots < \hat{\lambda}_{M-1} < \mu_M < \lambda_M.
\]

This also holds for the case \( c = 1 \), except that now \( \mu_1 = 0 \). If \( c > 1 \), so that \( N < M \), we have \( \hat{\lambda}_k = 0 \) for \( 1 \leq k \leq M - N \) and only \( N - 1 \) solutions are different from zero, although for ease of notation we will additionally write \( \mu_1 = \cdots = \mu_{M-N} = \mu_{M-N+1} = 0 \). The other solutions are interlaced with the sample eigenvalues as

\[
0 = \mu_{M-N+1} < \hat{\lambda}_{M-N+1} < \cdots < \mu_M < \lambda_M.
\]

Hence, the result follows from the residue formula, namely

\[
\frac{M}{K_m} \frac{1}{2\pi i} \oint_{\mathcal{R}_m} \hat{h}_M(z) dz = -\mathbf{R}(\lambda_k, \mu_k) + \sum_{\mu_k \in \mathcal{R}_m} \text{Res}(\hat{h}_M(z), \lambda_k)
\]

\[
\frac{1}{2\pi i} \oint_{\mathcal{R}_m} \tilde{g}_M(z) dz = -\left( \sum_{\lambda_k \in \mathcal{R}_m} \text{Res}(\tilde{g}_M(z), \lambda_k) + \sum_{\mu_k \in \mathcal{R}_m} \text{Res}(\tilde{g}_M(z), \mu_k) \right).
\]

Now, it remains to determine which of these poles fall within the integration region \( \mathcal{R}_m \). To do this, we invoke two important results related to the eigenvalue separation proven by Silverstein and Bai in [27] and [30]. In [27] it was proven that, for sufficiently high \( M, N \), with probability one there are no sample eigenvalues \( \{\hat{\lambda}_1, \ldots, \hat{\lambda}_M\} \) located outside the support of the asymptotic sample eigenvalue distribution. This result was later improved in [30], where it was shown that the separation of the sample eigenvalues into the asymptotic clusters is exact. This means that the relative number of sample and true eigenvalues (counting multiplicities) contained in a particular cluster of the asymptotic sample eigenvalue distribution is exactly the same with probability one, for sufficiently high \( M, N \). The only exception to that is the lowest asymptotic sample eigenvalue cluster in the case \( c > 1 \). Indeed, when \( M > N \), there are exactly \( M - N \) sample eigenvalues equal to zero, which do not contribute to the first cluster.

An immediate consequence of these two results is the fact that, with probability one for all large \( M, N \), only the sample eigenvalues \( \{\hat{\lambda}_k : k \in \mathcal{K}_m\} \) will fall into the integration region. Hence, only the residues corresponding to these eigenvalues have to be computed in (32) and (33). Regarding the position of the other poles, namely, \( \mu_k \), it follows from a similar argument that only \( \{\mu_k : k \in \mathcal{K}_m\} \) fall into the integration region (with probability one, for all large \( M, N \), i.e., we have the following:

**Lemma 1:** Under (As1–As4), with probability one, for all sufficiently large \( M, N \), the values \( \{\hat{\mu}_k : k \in \mathcal{K}_m\} \) are located within the interval \((\bar{\sigma}_1, \bar{\sigma}_2)\), whereas the values \( \{\mu_k : k \not\in \mathcal{K}_m\} \) are located outside \([\bar{\sigma}_1, \bar{\sigma}_2] \).

**Proof:** See Appendix IV.\(\square\)
In order to solve the two integrals in (32) and (33), we differentiate the two cases: $c \leq 1$ and $c > 1$. Consider first the case $c \leq 1$. Since the values $\{\hat{\lambda}_k\}$ and $\{\hat{\mu}_k\}$ are all almost surely different, we can express the integrals in (32) and (33) as

$$
\frac{M}{K_m} \frac{1}{2\pi i} \oint_{\partial R_{\gamma}(m)} \hat{h}_M(z) dz
$$

for the two smaller eigenvalues, respectively, and is equal to zero.

We use the same approach as in the case $c > 1$ and is therefore omitted.

The proof is identical to the one in [18, Lemma 1].

Next, consider the case $c > 1$. In this case, if $m > 1$, one can use the same approach as in the case $c \leq 1$ and get to the same expression. If $m = 1$ (estimation of the lowest eigenvalue and associated eigenvectors) is a bit more involved because of the presence of the discontinuity at $z = 0$. However, it turns out that this discontinuity is avoidable, and hence it does not contribute to the integral. Consequently, the expression obtained above is also valid for the case $c > 1$, $m = 1$.

V. NUMERICAL EVALUATION

In this section, we evaluate the performance of the two proposed estimators in a finite sample size situation. We consider a diagonal covariance matrix $R_M$ with four different eigenvalues $\{1, 7, 15, 25\}$ with multiplicities $\{4, 2, 1, 1\}$, respectively. The minimum number of samples to guarantee separation of the asymptotic eigenvalue clusters for this problem is 10 and 23 for the two smaller eigenvalues, respectively, and 33 for the last two. Fig. 4 represents the histogram corresponding to 50,000
Fig. 4. Histogram corresponding to 50,000 realizations of the proposed estimators (solid line) and traditional estimators (dotted line), $M = 8$, $N = 25$. Vertical lines indicate the true eigenvalues.

TABLE I

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Multiplicity</th>
<th>Proposed estimator</th>
<th>Traditional (sample) estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$ = 1</td>
<td>4</td>
<td>0.9998</td>
<td>0.8221</td>
</tr>
<tr>
<td>$\gamma_2$ = 7</td>
<td>2</td>
<td>6.9859</td>
<td>6.2764</td>
</tr>
<tr>
<td>$\gamma_3$ = 15</td>
<td>1</td>
<td>14.9226</td>
<td>14.7511</td>
</tr>
<tr>
<td>$\gamma_4$ = 25</td>
<td>1</td>
<td>25.0832</td>
<td>27.3847</td>
</tr>
</tbody>
</table>

realizations of the proposed and traditional (sample) estimators of the eigenvalues of $R_M$. In this example, the number of samples was fixed to $N = 25$ and the observations were generated according to a circularly symmetric Gaussian distribution. This means that (As4) is only verified for the first two eigenvalues. Observe that the proposed eigenvalue estimator produces estimates that are centered around the true eigenvalues, whereas the traditional counterparts produce highly biased estimates. Table I provides a comparative evaluation of the proposed and traditional estimators in terms of empirical mean and standard deviation. For comparative purposes, we also introduce the asymptotic values of the traditional (sample) estimator as predicted from Theorem 2 (last column of the table). Observe that the proposed estimator provides much better values in terms of empirical mean, at the expense of a slightly higher variance. Note that this is also true for the last two eigenvalues, and this illustrates the fact that the method provides good estimates even when cluster separability (As4) does not hold. On the other hand, the asymptotic prediction of the sample eigenvalue estimates given by Theorem 2 turns out to be a very accurate approximation of the empirical mean obtained in the simulation.

In order to illustrate the convergence of the proposed eigenvalue estimators toward the true eigenvalues, we next consider a diagonal covariance matrix with the same eigenvalues as before, but now having dimension $M = 8 \times 40 = 320$. The number of samples was also multiplied by a factor of 40 with respect to the case above, so that $N = 25 \times 40 = 1000$. Consider first the case where the multiplicity of the eigenvalues is multiplied by a factor of 40, so that the minimum number of samples to guarantee cluster separation is also scaled up by the same magnitude and therefore (As4) is verified for the two smallest eigenvalues but not for the other two. Table II illustrates the performance of the proposed and the traditional methods in terms of mean and standard deviation measured from a sample size of 10,000 real-
Fig. 5. Histogram corresponding to 10,000 realizations of the proposed estimators (solid line) and traditional estimators (dotted line), $M = 360, N = 1000$. Vertical lines indicate the true eigenvalues.

Table II

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Multiplicity</th>
<th>Proposed estimator</th>
<th>Traditional (sample) estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard Dev.</td>
</tr>
<tr>
<td>$\gamma_1 = 1$</td>
<td>160</td>
<td>0.9997</td>
<td>0.0031</td>
</tr>
<tr>
<td>$\gamma_2 = 7$</td>
<td>80</td>
<td>6.9972</td>
<td>0.0325</td>
</tr>
<tr>
<td>$\gamma_3 = 15$</td>
<td>40</td>
<td>14.9771</td>
<td>0.1072</td>
</tr>
<tr>
<td>$\gamma_4 = 25$</td>
<td>40</td>
<td>25.0014</td>
<td>0.1620</td>
</tr>
</tbody>
</table>

izations. Observe that the proposed estimators converge to the true values, whereas the classical estimators exhibit high biases. Note, again, that the proposed estimator provides a substantial improvement with respect to traditional estimators even when the separability condition (As4) does not hold (as it is the case for $\gamma_3$ and $\gamma_4$). This is also illustrated in Fig. 5, where the corresponding histograms are shown.

It is worth pointing out that the $M,N$-consistency of the proposed estimators is independent of the behavior of the dimension of the subspaces as the system dimension grows without bound. In other words, the dimension of the subspaces does not need to scale up with the matrix size to guarantee the $M,N$-consistency of the estimators. To illustrate this fact, consider again the case $M = 360, N = 1000$, but assume that the multiplicity of the eigenvalues is $\{160, 80, 79, 1\}$ (so that the highest eigenvalue has multiplicity 1). In this situation, the minimum number of samples to guarantee cluster separability is 380 for the first (smallest) eigenvalue, 1152 for the second and the third eigenvalues, and 433 for the fourth (largest) eigenvalue. Table III provides a comparative evaluation of the proposed and traditional estimators in terms of empirical mean and standard deviation. The corresponding histograms are shown in Fig. 6. Observe that the proposed estimators are able to converge to the true values even when the corresponding subspace dimension does not scale up with the observation dimension (as in $\gamma_4$).

Next, we compare the performance of the traditional and the proposed estimators for the subspaces associated with each of the eigenvalues of $R_M$. We compare their performance in terms of the orthogonality factor, defined as

$$O(m) = \frac{\|\text{tr} \left[ E_m E_m^H \hat{P}_m \right]\|}{\|\text{tr} \left[ (I - E_m E_m^H) \hat{P}_m \right]\|}$$

where $\hat{P}_m = \hat{E}_m \hat{E}_m^H$ for the traditional estimator, and $\hat{P}_m = \sum_{k=1}^{5T} \theta_m(k) \hat{E}_k \hat{E}_k^H$ for the proposed estimator. Observe that high orthogonality factors indicate good estimates of the associated subspace. The cumulative distribution function of the or-
Fig. 6. Histogram corresponding to 10,000 realizations of the proposed estimators (solid line) and traditional estimators (dotted line), $M = 360$, $N = 1000$. Vertical lines indicate the true eigenvalues.

TABLE III

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Multiplicity</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>1</td>
<td>0.9997</td>
<td>0.0032</td>
<td>0.8210</td>
<td>0.0023</td>
<td>0.8210</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>7</td>
<td>6.9942</td>
<td>0.0343</td>
<td>6.1400</td>
<td>0.0208</td>
<td>6.1401</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>15</td>
<td>14.9956</td>
<td>0.0681</td>
<td>16.1835</td>
<td>0.0514</td>
<td>16.1839</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>25</td>
<td>24.9892</td>
<td>1.0713</td>
<td>28.9104</td>
<td>0.7110</td>
<td>28.9069</td>
</tr>
</tbody>
</table>

thogonality factor is plotted in Fig. 7 for the first considered case, namely, $M = 8$, $N = 25$, and eigenvalue multiplicities \{4, 2, 1, 1\}. Observe that the proposed estimators achieve a much higher orthogonality factor than the traditional sample estimators, indicating a higher capability to estimate the associated subspaces. Note also that the results are better than the traditional counterparts even when the splitting condition (As4) is not fulfilled (as it is the case for the two largest eigenvalue eigenvectors).

VI. CONCLUSION

This paper has considered the estimation of the eigenvalues and the eigenvectors of covariance matrices. Our approach is based under the assumption that the asymptotic sample eigenvalue cluster associated with the eigenvalue/eigenvector to be estimated is separated from the rest of the asymptotic sample eigenvalue distribution. Using tools from random matrix theory, it can be shown that the traditional sample estimators are inconsistent in this asymptotic regime, indicating a very poor performance in finite sample size situations. New consistent estimators of the eigenvalues and associated subspaces of the sample covariance matrix have then been derived. These estimators are based on random matrix theory, and are consistent when both the sample size and the observation dimension tend to infinity at the same rate. It is shown via simulations that the performance of the proposed estimators is much better than the traditional counterparts in finite sample size situations, even in the nonasymptotic regime and when the asymptotic eigenvalue splitting condition does not hold.

APPENDIX I

PROOF OF PROPOSITION 3

Note first that we can write the expressions in (25) and (26) as

\[
\gamma_m = \frac{M}{K_m} \frac{1}{2\pi j} \int_{\sigma_1}^{\sigma_2} h_M(x) dx = \frac{M}{K_m} \frac{1}{2\pi j} \int_{\sigma_1}^{\sigma_2} h^*_M(x) dx \\
\eta_m = \frac{1}{2\pi j} \int_{\sigma_1}^{\sigma_2} g_M(x) dx = \frac{1}{2\pi j} \int_{\sigma_1}^{\sigma_2} g^*_M(x) dx
\]
where we have defined, for general $z \in \mathbb{C}$

$$h_M(z) = \frac{1}{M} \sum_{r=1}^{M} K_r \frac{f_M(z)}{\gamma_r - f_M(z)} f'_M(z),$$

$$g_M(z) = \sum_{r=1}^{M} s^H_r E_r E^H_r s^*_2 f'_M(z)$$

where here $f'_M(z)$ is the complex derivative of $f_M(z)$. We stress that thanks to Property 1 of Proposition 2, and since $f_M(z) \neq \gamma_r$ for all $z$, one can ensure that both $h_M(z)$ and $g_M(z)$ are holomorphic on the set $\mathbb{C} \setminus S_M$, where $S_M = \bigcup_{i=1}^{\infty} \left[ x_i^{-1}, x_i^{-1} \right]$. The proof is a bit involved, because the functions $g_M(z)$ and $h_M(z)$ are not bounded on the real axis (since $f'(z)$ is not bounded).

**Lemma 3:** The quantities $\gamma_m$ and $\eta_m$ can be expressed as shown in the equations at the bottom of the page.

**Proof:** See Appendix II.

Let $\mathbb{R}_y(m)$ denote a rectangular region on the complex plane with vertices $\{\sigma_1 + jy, \sigma_1 - jy, \sigma_2 + jy, \sigma_2 - jy\}$ and $\partial \mathbb{R}_y(m)$ its boundary, negatively oriented. Now, observe that the integrals presented in Lemma 3 can be expressed in terms of the contour integral along $\partial \mathbb{R}_y(m)$, i.e.,

$$\int_{\sigma_1}^{\sigma_2} h_M(x + jy)dx - \int_{\sigma_1}^{\sigma_2} h_M(x - jy)dx$$

$$= \int_{\sigma_1}^{\sigma_2} h_M(z)dz - j \int_{-y}^{y} h_M(\sigma_1 + jt)dt$$

and, equivalently, for $g_M(z)$. On the other hand, $|h_M(\sigma_1 + jt)|$, $|h_M(\sigma_2 + jt)|$, $|g_M(\sigma_1 + jt)|$, and $|g_M(\sigma_2 + jt)|$ are continuous functions of $t$ on the compact set $[-y, y]$, so that in
particular they are bounded. But this implies that the last two integrals in the preceding equation tend to zero as \( y \to 0^+ \), and therefore

\[
\gamma_m = \frac{M}{K_m} \lim_{y \to 0^+} \frac{1}{2\pi i} \oint_{\mathcal{R}_g} h(z)dz
\]

\[
\eta_m = \lim_{y \to 0^+} \frac{1}{2\pi i} \oint_{\mathcal{R}_g} g(z)dz
\]

as we wanted to show. However, both \( h_M(z) \) and \( g_M(z) \) are holomorphic on \( \mathbb{C} \setminus \mathbb{R} \); consequently, these integrals do not depend on the value of \( y \), and the limits can be dropped. This concludes the proof of Proposition 3.

**APPENDIX II**

**PROOF OF LEMMA 3**

In this appendix, we present the proof of Lemma 3 used in Appendix I. The lemma is a direct consequence of the Lebesgue dominated convergence theorem. We only need to show that the two integrands, \( h_M(z) \) and \( g_M(z) \), are absolutely bounded by a complex-valued function that is integrable over any axis parallel to the real one. The difficulty comes from the fact that the derivative of \( f_M(z) \), namely, \( f_M'(z) \), is not bounded on the real axis. More specifically, the function \( f_M(z) \) grows without bound at the boundary points of the support of \( q_M(x) \), namely \( \{x_1^+, x_2^+, \ldots, x_Q^+, \ldots, x_M^+\} \).

Observing the form of \( h_M(z) \) and \( g_M(z) \) in (35) and noting that \( f_M(z) \) is bounded on \( \mathbb{R} \) and that \( \gamma - f_M(z) \) can never be zero, we only need to show that \( |f_M'(z)| \) is bounded by a complex-valued function that is integrable over any axis parallel to the real one. It is sufficient to see that \( |f_M'(z)| \) can be bounded by an appropriate integrable function near a singularity. Let \( \tilde{x} \) denote one of these singularities, namely, \( \tilde{x} \in \{x_1^+, x_2^+, \ldots, x_Q^+, \ldots, x_M^+\} \), and let \( f = \{f_1^+, \ldots, f_Q^+, \ldots, f_M^+\} \) denote the corresponding root of the equation in (14).

We first note that the module of the derivative \( f_M'(z) \) can be expressed as (this can be seen by taking derivatives on both sides of (23))

\[
|f_M'(z)|^2 = \frac{1}{\left(1 - \frac{\sum K_r}{\gamma - f_M(z)}\right)^2}
\]

\[
= \frac{1}{\left(\tilde{f} - f_M(z)\right)^2}
\]

\[
\times \left\{\sum_{r=1}^K K_r^2 \left(\gamma - f_M(z)\right) + \left(\gamma - \tilde{f}\right)^2\right\}^{-2}
\]

where we have used the fact that \( \tilde{f} \) is a solution to the equation in (14). Using this together in (36), we get to

\[
\lim_{z \to \tilde{f}} |f_M'(z)|^2 |z - \tilde{f}| = \frac{1}{4} \left(\sum_{r=1}^K K_r^2 \left(\gamma - f_M(z)\right)^{-2}\right)^{-1}
\]

where we have implicitly used the continuity of \( f_M(z) \). We note here that this limit is always bounded, because \( \tilde{f} \) is not a solution to the equation in (16).

Now, it follows from all the above that, given a sufficiently small quantity \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that

\[
|f_M'(z)| < \frac{1}{|z - \tilde{f}|^{\varepsilon/2}} \left(\frac{1}{\sum_{r=1}^K K_r^2 \left(\gamma - f_M(z)\right)^{-2}\right)^{-1} + \varepsilon\right)
\]

for every \( z \) such that \( |z - \tilde{f}| < \delta \). Consequently, on a small neighborhood of \( \tilde{f} \), \( |f_M'(z)| \) is upper-bounded by a complex-valued function that is integrable along any axis parallel to the real one. The proof of the lemma follows from the Lebesgue dominated convergence theorem.

**APPENDIX III**

**PROOF OF PROPOSITION 4**

Observe that we can write

\[
\gamma_m - \frac{1}{2\pi i} \oint_{\mathcal{R}_g} h_M(z)dz \leq \frac{M}{K_m} \frac{1}{2\pi i} \oint_{\mathcal{R}_g} |h_M(z)| |dz| \leq \frac{M}{K_m} \frac{1}{2\pi i} \oint_{\mathcal{R}_g} |h_M(z) - \hat{h}_M(z)| |dz|
\]

\[
\eta_m - \frac{1}{2\pi i} \oint_{\mathcal{R}_g} g_M(z)dz \leq \frac{M}{K_m} \frac{1}{2\pi i} \oint_{\mathcal{R}_g} |g_M(z) - \hat{g}_M(z)| |dz|.
\]

Hence, since \( \sigma_1 \) and \( \sigma_2 \) are uniformly bounded for all \( M \) (because (A1) ensures that the spectral norm of \( \mathcal{R}_M \) is uniformly bounded for all \( M \) using the dominated convergence theorem) it is sufficient to prove that

\[
\sup_{\zeta \in \mathcal{R}_g} |h_M(z) - \hat{h}_M(z)| < +\infty \quad (37)
\]

for all \( M, N \) sufficiently large (and, equivalently, for \( \hat{h}_M(z) \)). Now, the main problem with (37) is the fact that \( \hat{h}_M(z) \) presents discontinuities on the real axis. Hence, we need to show that these discontinuities are bounded away from \( \sigma_1 \) and \( \sigma_2 \) for all sufficiently large \( M, N \).

To do this, let us first analyze the discontinuities two complex-valued functions \( \hat{h}_M(z) \) and \( \hat{g}_M(z) \). It is shown in Section IV, that \( \hat{h}_M(z) \) and \( \hat{g}_M(z) \) are holomorphic on \( \mathbb{C} \), except for two sets of poles: the sample eigenvalues \( \{\lambda_k\} \) and the values \( \{b_k\} \) defined in the statement of Theorem 3. On the other hand, it follows from the eigenvalue separation arguments in [27] and [30] that, for all large \( M, N \) and with probability one, the only sample eigenvalues located inside \( \mathbb{R}_y(m) \) will be
\( \{ \tilde{\lambda}_k : k \in K \} \) and there will be no sample eigenvalues outside the support \( \mathcal{S}_M \). On the other hand, Lemma 1 in Section IV establishes the equivalent result for the values \( \tilde{\mu}_1 \leq \tilde{\mu}_2 \leq \cdots \leq \tilde{\mu}_M \), showing that only the values \( \{ \mu_k : k \in K \} \) are located inside \( \mathbb{R}_y(m) \) for all large \( M, N \) (with probability one), whereas the rest of these poles are located outside this region.

Now, it follows from all this that

\[
\sup_{z \in \mathbb{R}_y(m)} \left| h_M(z) - \tilde{h}_M(z) \right| 
\leq \sup_{z \in \mathbb{R}_y(m)} \left| h_M(z) \right| + \sup_{z \in \mathbb{R}_y(m)} \left| \tilde{h}_M(z) \right|
\leq +\infty
\]

almost surely for all large \( M, N \). Consequently, by virtue of the dominated convergence theorem

\[
\frac{1}{2\pi} \int_{\mathbb{C}^+} \left| h_M(z) - \tilde{h}_M(z) \right| dz \to 0
\]

almost surely, from where the final result follows. The proof that \( |h_m - \tilde{h}_m| \to 0 \) is almost identical, and is therefore omitted. This finishes the proof of Proposition 4.

**APPENDIX IV**

**PROOF OF LEMMA 1**

It is obvious from the fact that \( \tilde{\mu}_1 \leq \tilde{\lambda}_1 \leq \tilde{\mu}_2 \leq \cdots \leq \tilde{\lambda}_M \leq \tilde{\mu}_M \) together with the eigenvalue separation arguments in [27] and [30] that, with probability one for all large \( M, N \), the values

\[
\left\{ \tilde{\mu}_k : \sum_{r=1}^{m-1} K_r + 1 < k < \sum_{r=1}^{m} K_r \right\}
\]

will be located inside \( (\sigma_1, \sigma_2) \), whereas the values

\[
\left\{ \tilde{\mu}_k : k < \sum_{r=1}^{m-1} K_r + 1 \right\} \cup \left\{ \tilde{\mu}_k : k > \sum_{r=1}^{m} K_r \right\}
\]

will be located outside \( [\sigma_1, \sigma_2] \). We only need to show that, with probability one for all large \( M, N \), \( \tilde{\mu}_i \in (\sigma_1, \sigma_2) \), where \( i = \sum_{r=1}^{m-1} K_r + 1 \), whereas \( \tilde{\lambda}_j \notin [\sigma_1, \sigma_2] \), where \( j = \sum_{r=1}^{m} K_r \) (in both cases, provided that these values exist).

Assume now that either \( m > 1 \) or \( c < 1 \) (extension to the case \( m = 1 \) with \( c \geq 1 \) is trivial noting the lowest values of \( \tilde{\mu}_k \) are zero and always fall into \( (\sigma_1, \sigma_2) \)). Consider a positively oriented contour \( C^+(m) \) taking values on \( \mathbb{C}^+ \setminus \{ \gamma_1, \ldots, \gamma_M \} \) and enclosing only \( \gamma_M \) and not \( \{0\} \). Consider also the following integral:

\[
\frac{1}{2\pi} \int_{C^+(m)} \frac{1}{\omega} d\omega = 0.
\]

The integral is equal to zero because \( \omega = 0 \) is not enclosed by the contour \( C^+(m) \). Now, observe that we can choose \( C^+(m) \) as the contour described by \( f_M(z) \) when \( x \) moves from \( \sigma_1 \) to \( \sigma_2 \) concatenated with the contour described by \( f_M(x) \) as \( x \) moves back from \( \sigma_2 \) to \( \sigma_1 \). In this case

\[
\frac{1}{2\pi} \int_{C^+(m)} \frac{1}{\omega} d\omega = \frac{1}{2\pi} \int_{\sigma_1}^{\sigma_2} s_M(x) dx - \frac{1}{2\pi} \int_{\sigma_1}^{\sigma_2} s_M(x) dx
\]

where \( s_M(x) \) is the restriction to the real axis of \( \frac{f_M'(z)}{f_M(z)} \)

well defined for all \( z \in \mathbb{C}^+ \) and holomorphic for all \( z \in \mathbb{C}^+ \setminus \mathcal{S} \), where \( \mathcal{S} \) is the support of the limiting sample eigenvalue distribution excluding zero when \( c > 1 \). Now, using the proof of Lemma 3 given in Appendix II one can easily show that, since \( m > 1 \) by assumption

\[
\frac{1}{2\pi} \int_{C^+(m)} \frac{1}{\omega} d\omega = \lim_{y \to 0^+} \frac{1}{2\pi} \int_{\sigma_1}^{\sigma_2} s_M(x - jy) dx
\]

\[
- \lim_{y \to 0^+} \frac{1}{2\pi} \int_{\sigma_1}^{\sigma_2} s_M(x + jy) dx
\]

where we have used the property \( s_M(z^*) = s_M^*(z) \). Finally, since \( \{s_M(\sigma_1 + j\ell)\} \) and \( \{s_M(\sigma_2 + j\ell)\} \) are continuous functions of \( \ell \) on the compact set \( [-y, y] \), we can write

\[
0 = \frac{1}{2\pi} \int_{C^+(m)} \frac{1}{\omega} d\omega = \lim_{y \to 0^+} \frac{1}{2\pi} \int_{\mathbb{R}_y(m)} s(z) dz
\]

\[
= \frac{1}{2\pi} \int_{\mathbb{R}_y(m)} s(z) dz
\]

with \( \partial \mathbb{R}_y(m) \) denoting the boundary of \( \mathbb{R}_y(m) \) as defined above, positively oriented and where we have used the fact that \( s_M(z) \) is holomorphic on \( \mathbb{C} \setminus \mathbb{R} \) and hence the integral does not depend on \( y \).

The idea now is to construct \( M, N \)-consistent estimators of the quantity on the right-hand side of the preceding equality, and study its asymptotic behavior. We note first that the function \( s_M(z), z \in \mathbb{C}^+ \), is pointwise \( M, N \)-consistently estimated with probability one by

\[
\hat{s}_M(z) = \frac{1 - c + cz^2 \hat{\beta}_M(z)}{z \left( 1 - c - cz \hat{\beta}_M(z) \right)} = \frac{1 - c + cz \sum_{i=1}^{M} \frac{\lambda_i}{\lambda_i - z}}{z \left( 1 - c - cz \lambda_M \sum_{i=1}^{M} \frac{1}{\lambda_i - z} \right)}
\]

\[
= \frac{\hat{\mu}_M(z)}{f_M(z)}
\]

where we have defined

\[
\hat{f}_M(z) = \frac{z}{1 - c \lambda_M \sum_{i=1}^{M} \frac{1}{\lambda_i - z}}
\]

with \( \hat{\mu}_M(z) \) denoting the corresponding derivative. Observe that complex function \( \hat{f}_M(z) \) presents zeros at the sample eigenvalues \( \{\lambda_k, k = 1 \ldots M\} \) and is holomorphic everywhere except for the poles at the values \( \{\hat{\mu}_k, k = 1 \ldots M\} \).
Now, by virtue of the Dominated Convergence Theorem
\[
\frac{1}{2\pi j} \int_{\mathcal{R}_{1}^{M}(m)} \hat{s}_M(z)dz = \frac{1}{2\pi j} \int_{\mathcal{R}_{1}^{M}(m)} s_M(z)dz \to 0
\]
almost surely as \(M, N \to \infty\) at the same rate. Therefore, it follows from the argument principle [31, Theorem 6.2.3] that, almost surely for all sufficiently large \(M, N\)
\[
\frac{1}{2\pi j} \int_{\mathcal{R}_{1}^{M}(m)} \hat{s}_M(z)dz = \frac{1}{2\pi j} \int_{\mathcal{R}_{1}^{M}(m)} \hat{f}_M(z)dz = \# \left\{ k : \hat{\lambda}_k \in \mathbb{R}_y(m) \right\} - \# \left\{ k : \hat{\mu}_k \in \mathbb{R}_y(m) \right\}
\]
where \(\# \{ \cdot \} \) denotes the cardinality of a set. This means that
\[
\# \left\{ k : \hat{\lambda}_k \in \mathbb{R}_y(m) \right\} = \# \left\{ k : \hat{\mu}_k \in \mathbb{R}_y(m) \right\}
\]
(43) with probability one for all large \(M, N\).

Using the same line of thought, one can reason that
\[
\# \left\{ k : \hat{\lambda}_k > \sigma_2 \right\} = \# \left\{ k : \hat{\mu}_k > \sigma_2 \right\}
\]
almost surely for all large \(M, N\). This implies that, with probability one for all sufficiently large \(M, N\), \(\hat{\mu}_j \notin [\sigma_1, \sigma_2]\), where \(j = \sum_{r=1}^{m} K_r\). And, from (43), \(\hat{\mu}_j \in (\sigma_1, \sigma_2)\), where \(i = \sum_{r=1}^{m} K_r + 1\) almost surely for all large \(M, N\).

REFERENCES