Estimating the eigenvalues and associated subspaces of correlation matrices from a small number of observations

Xavier Mestre

Abstract—In this paper, a new method for estimating the eigenvalues and associated subspaces of covariance matrices using their sample estimates is presented. The method is based on random matrix theory and provides consistent estimates when both the sample size and the observation dimension tend to infinity at the same rate, provided that an eigenvalue asymptotic splitting condition is fulfilled. The proposed estimators are shown to have an excellent performance in low sample size scenarios, where the observation dimension and the sample size have the same order of magnitude.

I. INTRODUCTION

The eigenvalues and eigenvectors of covariance matrices are extensively used in multiple signal processing applications. In a number of problems, the inherent properties of the stochastic processes can be inferred from the eigenvalues and eigenvectors of the underlying covariance matrix. For example, in array processing applications, it is common practice to the number of signals impinging on the array by using the AIC/MDL criteria based on statistical tests built from the eigenvalues of the covariance matrix [1]. Also, in the same type of application, the noise power is usually estimated by averaging the smallest eigenvalues of the covariance matrix. On the other hand, subspace-based spectral estimation techniques (such as MUSIC [2]) are based on the projection of a known signature vector onto the column span of some eigenvectors of the covariance matrix.

In practice the covariance matrix is unknown and so are therefore both their eigenvalues and associated eigenvectors. These quantities are usually estimated from the sample covariance matrix, which is constructed directly from the observations as follows. Consider a collection of \( N \) independent, identically distributed (i.i.d.) \( M \)-dimensional observations of an underlying stochastic process. These observations will be denoted by \( \{ y(1), \ldots, y(N) \} \), and we will assume that they have zero mean and covariance matrix \( \Sigma_m \). The sample covariance matrix is constructed as

\[
\hat{\Sigma}_m = \frac{1}{N} \sum_{n=1}^{N} y(n)y^H(n).
\]

If \( \hat{M} \) is the number of eigenvalues of \( \Sigma_m \) that are pairwise different \( (\hat{M} \leq M) \), the strict inequality holding when there are eigenvalues with multiplicity higher than one), the set of distinct eigenvalues will be denoted as \( \gamma_1 < \gamma_2 < \ldots < \gamma_{\hat{M}} \). Each of the eigenvalues \( \gamma_m, m = 1, \ldots, \hat{M} \), is assumed to have multiplicity \( K_m \) (hence, we will have \( M = \sum_{m=1}^{\hat{M}} K_m \)). Associated with each eigenvalue \( \gamma_m, m = 1 \ldots \hat{M} \), there is a complex subspace of dimension \( K_m \). This subspace is determined by an \( \hat{M} \times K_m \) matrix of eigenvectors, denoted by \( E_m \), such that \( E_m^H E_m = I_{K_m} \).

The objective of this paper is to provide optimized estimators for the eigenvalues \( \gamma_1, \gamma_2, \ldots, \gamma_{\hat{M}} \) and associated subspaces of the true covariance matrix from the sample covariance matrix \( \hat{\Sigma}_m \), assuming that their multiplicities \( (K_1, \ldots, K_{\hat{M}}) \) are known (otherwise, one can simply take \( K_1 = \ldots = K_{\hat{M}} = 1 \)). The eigenvalues and associated eigenvectors of the sample covariance matrix \( \hat{\Sigma}_m \) will be denoted by \( \hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \ldots \leq \hat{\lambda}_{\hat{M}} \) and \( \hat{e}_1, \hat{e}_2, \ldots, \hat{e}_{\hat{M}} \) respectively. Throughout the paper, we will refer to these quantities as the sample eigenvalues and eigenvectors.

II. TRADITIONAL ESTIMATION OF EIGENVALUES AND EIGENVECTORS

Let us assume that we want to estimate the \( m \)th eigenvalue of the true covariance matrix \( \Sigma_m, \gamma_m \), which has multiplicity \( K_m \). Let \( K_m \) be the set of indexes

\[
K_m = \left\{ \sum_{r=1}^{m-1} K_r + 1, \ldots, \sum_{r=1}^{m} K_r \right\}
\]

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

\[
\hat{\gamma}_m^{\text{trad}} = \frac{1}{K_m} \sum_{k \in K_m} \hat{\lambda}_k.
\]

This is, indeed, the maximum likelihood estimator of \( \gamma_m \) when the observations \( \{ y(n) \} \) follow a multivariate Gaussian distribution [3], [4]. In the particular case where the true eigenvalue is assumed to have multiplicity one, the estimated eigenvalues are taken to be the eigenvalues of the sample correlation matrix, namely \( \hat{\gamma}_m^{\text{trad}} = \hat{\lambda}_m \). On the other hand, the classical estimator of the associated subspace is given by the sample eigenvectors, namely

\[
\hat{E}_m^{\text{trad}} \left( \hat{E}_m^{\text{trad}} \right)^H = \hat{E}_k \hat{E}_k^H
\]

where \( \hat{E}_k \) is an \( \hat{M} \times K_m \) matrix containing the eigenvectors \( \{ \hat{e}_k, k \in K_m \} \) as columns.

We will assume throughout the paper that the sample covariance matrix has the following structure:

This work was not supported by the European Commission under the project NEWCOM IST-2002-507525, the Spanish Ministry of Science and Education under projects FIT-330220-2005-108 and FIT-330210-2005-23, and the Generalitat de Catalunya SGR2005-00690.

X. Mestre is with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Avinguda del Canal Olímpic s/n, Parc Mediterrani de la Tecnologia, Castelldefels 08860, Barcelona, Catalonia (Spain), xavier.mestre@cttc.es.
The sample covariance matrix takes the form

\( \hat{R}_M = R_M^{1/2} U_M H R_M^{1/2} \),

where \( U_M \) 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 \( 1/2(2N) \), and finite sixteenth moments.

On the other hand, \( R_M^{1/2} \) is an \( M \times M \) Hermitian square root of the true covariance matrix \( R_M \), assumed Hermitian and positive definite.

Under \((\text{As}1)\), the two estimators in (1) and (2) are strongly \( N \)-consistent (i.e., consistent with probability one as the number of samples goes to infinity, \( N \to \infty \)). This is reasonable because, under \((\text{As}1)\), \( \hat{R}_M \to R_M \) almost surely as \( N \to \infty \). However, in the low sample size regime (i.e., when the number of samples \( N \) has the same order of magnitude as the observation dimension \( M \)), these estimators do not provide the optimum behavior, and much better estimators can be obtained using other approaches. For example, it is well known that \( \gamma_m \) is unbiased and that, in fact, the eigenvalues of \( R_M \) tend to be more spread out than those of \( R_M \). This has motivated a number works aimed at shrinking or expanding the sample eigenvalues according to their magnitude, see e.g. [5], [6], [7] and references therein.

In order to improve the performance of the traditional estimators in the low sample size regime, it is useful to consider two-dimensional asymptotics in the definition of consistency. More specifically, an estimator of a certain function of the true covariance matrix can be considered to perform well in the low sample size regime if it provides consistent estimates when both \( M \) and \( N \) tend to infinity at the same rate (i.e., are both large but have the same order of magnitude). These type of asymptotics are very appropriate for capturing a finite sample size situation, where \( M \) and \( N \) have the same order of magnitude. Furthermore, results are usually very representative of the non-asymptotic reality, even for moderate values of \( M \) and \( N \). Thus, from now on, we will state that an estimator is \( M, N \)-consistent if it provides consistent estimates when both the number of observations \( N \) and their dimension \( (M) \) grow without bound at the same rate, i.e. \( N, M \to \infty, M/N \to c, 0 < c < \infty \).

We will assume that the eigenvalue density of \( R_M \) is essentially the same regardless of \( M \), namely

\( \text{(As2)} \): \( R_M \) has always the same eigenvalue distribution for all values of \( M \) under consideration.

This particular way of fixing the convergence of the eigenvalue density of \( R_M \) is very appropriate in order to analyze estimators of quantities that depend on the spectrum of \( R_M \). Since the eigenvalue distribution of \( R_M \) is not altered with the observation dimension \( M \), consistency as \( N, M \to \infty \) will indicate closeness to the spectral structure of \( R_M \) in the finite dimensional situation.

Regarding the estimation of the subspaces associated with the eigenvalues of the covariance matrix, it makes little sense to consider the estimation of the matrices \( E_m E_m^H \) because their dimension increases to infinity as \( M \) grows large. Instead of that, we will try to estimate quadratic forms of the type

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

where \( s_1, s_2 \) are two \( M \times 1 \) deterministic column vectors such that:

\( \text{(As3)} \): The two \( M \times 1 \) column vectors \( s_1, s_2 \) have uniformly bounded norm for all \( M \), that is

\[ \sup_M ||s_1|| < +\infty, \quad \sup_M ||s_2|| < +\infty \]

where \( || \cdot || \) here denotes Euclidean norm.

In particular, one can choose \( s_1 \) and \( s_2 \) as the \( i \)th and \( j \)th column vectors of the \( M \times M \) identity matrix. In this case, \( \eta_m \) in (3) would correspond to the \((i,j)\)th entry of \( E_m E_m^H \).

III. ASYMPTOTIC CHARACTERIZATION OF THE SPECTRUM OF \( \hat{R}_M \)

Using traditional random matrix theory one can show that, under the statistical assumptions stated above, the distribution of eigenvalues of \( \hat{R}_M \) tends almost surely (as \( M, N \to \infty \) at the same rate) to a non-random distribution function associated with a compactly supported density.

Furthermore, one can characterize the asymptotic distribution of the eigenvalues of \( \hat{R}_M \) in terms of the asymptotic distribution of the eigenvalues of \( R_M \). That characterization is usually established in terms of different scalar functions of the resolvent of the sample covariance matrix, defined as

\[ \hat{Q}_M(z) = \left( \hat{R}_M - z I_M \right)^{-1} \].

For our purposes, it is useful to consider

\[ \hat{b}_M(z) = \frac{1}{M} \text{tr} \left[ \hat{Q}_M(z) \right] \], \quad \hat{m}_M(z) = s_1^H \hat{Q}_M(z) s_2 \]

(4)

where \( \hat{b}_M(z) \) has traditionally been referred to as the Stieltjes transform of the sample eigenvalues.

It is shown in [8], [9] that, for all \( z \in \mathbb{C}^+ \equiv \{ z \in \mathbb{C} : \text{Im}(z) > 0 \} \), under \((\text{As1} - \text{As3})\) and as \( M, N \to \infty \) at the same rate \((c = M/N, 0 < c < \infty)\), \( \hat{b}_M(z) \) and \( \hat{m}_M(z) \) have the same asymptotic behavior as two different deterministic functions \( \bar{b}(z) \) and \( \bar{m}_M(z) \), in the sense that

\[ \hat{b}_M(z) \to \bar{b}(z), \quad |\hat{m}_M(z) - \bar{m}_M(z)| \to 0 \]

(5)

almost surely for all \( z \in \mathbb{C}^+ \), where \( \bar{b}(z) = \bar{b} \) is the unique solution to the following equation in the set \( \{ \bar{b} \in \mathbb{C} : -(1-c)/z + c \bar{b} \in \mathbb{C}^+ \} \):

\[ \bar{b} = \frac{1}{M} \sum_{m=1}^{M} K_m \frac{1}{\gamma_m (1-c-c\bar{b})} - z \]

(6)

and where

\[ \bar{m}_M(z) = \sum_{m=1}^{M} \frac{s_1^H E_m E_m^H s_2}{\gamma_m (1-c-c\bar{b}(z)) - z} \].

An immediate consequence of all this is the fact that, as \( M, N \to \infty \) at the same rate, the empirical eigenvalue distribution of \( \hat{R}_M \) tends almost surely to a non-random distribution with Stieltjes transform \( \bar{b}(z) \). The actual asymptotic density, denoted as \( q(x) \), can be retrieved from \( \bar{b}(z) \) using the Stieltjes inversion formula:

\[ q(x) = \lim_{y \to 0^+} \frac{1}{\pi} \text{Im} \left[ \bar{b}(x + j y) \right] \].
It turns out that the support of \( q(x) \) is compact and presents a number of connected components (excluding \( x = 0 \) when \( c > 1 \)) lower than or equal to \( M \), the total number of different eigenvalues of \( R_M \). Each of the eigenvalues of \( R_M \) is univocally associated with a particular connected component or cluster of the support of \( q(x) \). Moreover, if \( c \) is sufficiently low or, equivalently, the number of samples per observation dimension is sufficiently high, each connected component of the support of \( q(x) \) is associated with a different eigenvalue of \( R_M \).

The study of the asymptotic behavior of the traditional estimators turns out to be especially simple when the asymptotic eigenvalue cluster associated with the eigenvalue to be estimated is separated from the clusters associated with the other true eigenvalues. This condition can be shown to be equivalent to:

\[(\text{As4}) \text{ (Eigenvalue splitting condition)} \quad \text{If } \gamma_m, \ m \in \{1 \ldots M\}, \ \text{is the eigenvalue that is to be estimated, then } 1/c > \xi(m), \ \text{where } \xi(m) = \max_{j \in \{m-1, m\}} \{\beta(j)\}, \]

\[
\beta(j) = \frac{1}{M} \sum_{r=1}^{M} K_r \left( \frac{\gamma_r}{\gamma_r - j} \right)^2
\]

for \( 0 < j < \bar{M} \), \( \beta(0) = 0 \), and \( \bar{f}_j, j = 1 \ldots \bar{M} - 1 \), denoting the \( \bar{M} - 1 \) different real-valued solutions to the equation

\[
\frac{1}{M} \sum_{m=1}^{\bar{M}} K_m \frac{\gamma_m^2}{(\gamma_m - f)^2} = 0
\]

ordered as \( \bar{f}_1 < \bar{f}_2 < \ldots < \bar{f}_{\bar{M}-1} \).

IV. MAIN RESULTS

In this section, we state the main results of the paper. Due to space limitations, the proofs are omitted. The interested reader is referred to [10] for further details.

A. Asymptotic behavior of the traditional estimators

We now present a result that characterizes the asymptotic behavior of the traditional eigenvalue/subspace estimators in (1) and (2) when both \( M \) and \( N \) increase without bound at the same rate.

**Theorem 1:** Under (As1 - As4), and as \( M, N \rightarrow \infty \) at the same rate \( (M/N \rightarrow c, \ 0 < c < \infty) \),

\[
\hat{\gamma}_m \rightarrow \gamma_m \left( 1 - \frac{c}{M} \sum_{r=1}^{\bar{M}} K_r \frac{\gamma_r}{\gamma_r - \gamma_m} \right) \quad (7)
\]

almost surely, where

\[
\hat{\eta}_m = \sum_{k=1}^{\bar{M}} w_m(k) s^H_k E_{\hat{\lambda}_k} E_k^H s_2
\]

\[
w_m(k) = \left\{ \begin{array}{ll} 1 - \frac{1}{\bar{M}} \sum_{r=1}^{\bar{M}} K_r \hat{\eta}_m(r) & k = m \\ \hat{\eta}_m(k) & k \neq m \end{array} \right.
\]

\[
\hat{\eta}_m(k) = \frac{\gamma_m}{\gamma_k - \gamma_m} - \frac{\mu_m}{\gamma_k - \mu_m}
\]

and \( \mu_1 < \ldots < \mu_M \), are the \( \bar{M} \) real-valued solutions to the following equation in \( \mu \),

\[
\frac{1}{\bar{M}} \sum_{r=1}^{\bar{M}} K_r \frac{\gamma_r}{\gamma_r - \mu} = \frac{1}{c}
\]

As a direct consequence of this theorem, we observe that the traditional estimators of both the eigenvalues and the associated subspaces are not \( M, N \)-consistent. In fact, they are only \( N \)-consistent (this can be observed from the above expressions, taking limits as \( c \rightarrow 0 \) and noting that \( \mu_m \rightarrow \gamma_m \)). This justifies the poor performance of these estimators when \( M \) and \( N \) have the same order of magnitude.

B. Alternative \( M, N \)-consistent estimators

The equation in (6) provides an intrinsic relationship between the eigenvalues of the true covariance matrix \( \{\gamma_m\} \) and \( \hat{b}(z) \), the almost sure limit of \( \hat{b}_M(z) \), which in turn depends on the eigenvalues of the sample covariance matrix only \( \{\hat{\lambda}_m\} \). The theory of G-estimation developed by V.L. Girko [11], [12] exploits that intrinsic relationship in order to derive estimators of functions of the true eigenvalues in terms of their corresponding sample estimates. Using techniques similar to G-estimation, we are able to derive the following \( M, N \)-consistent estimators of the eigenvalues and associated subspaces of the covariance matrix.

**Theorem 2:** Under (As1 - As4), the following quantities are strongly \( M, N \)-consistent estimators of \( \gamma_m \) and \( \eta_m \), respectively:

\[
\hat{\gamma}_m = \sum_{k \in \mathcal{K}_m} \left( \hat{\lambda}_k - \hat{\mu}_k \right)
\]

\[
\hat{\eta}_m = \sum_{k=1}^{\bar{M}} \theta_m(k) s^H_k E_{\hat{\lambda}_k} E_k^H s_2
\]

where \( \hat{\mu}_k, k = 1 \ldots \bar{M} \), are the real-valued solutions to the following equation in \( \hat{\mu} \) (counting multiplicities, if \( c > 1 \))

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

and

\[
\theta_m(k) = \left\{ \begin{array}{ll} -\phi_m(k) & k \notin \mathcal{K}_m \\ 1 - \phi_m(k) + \psi_m(k) & k \in \mathcal{K}_m \end{array} \right.
\]

where

\[
\phi_m(k) = \sum_{r \in \mathcal{K}_m, \ r \neq k} \left( \frac{\lambda_r - \hat{\lambda}_k}{\lambda_r - \hat{\lambda}_k} - \frac{\hat{\mu}_r}{\lambda_r - \hat{\mu}_k} \right)
\]

\[
\psi_m(k) = \sum_{r \in \mathcal{K}_m, \ r \neq k} \left( \frac{\hat{\lambda}_k - \lambda_r}{\lambda_r - \hat{\mu}_k} - \frac{\hat{\mu}_k}{\lambda_r - \hat{\mu}_k} \right).
\]

In the above definition of \( \phi_m(r) \) and \( \psi_m(r) \), we use the convention that any terms of the form \( 0/0 \) are equal to zero.

Even though these estimators are known to be \( M, N \)-consistent under (As4), it turns out that they have excellent performance even when this eigenvalue separation condition does not hold. Furthermore, the behavior is excellent even for low values of \( M, N \), see further Section V.
acovariance matrix are low and have the same order of magnitude. We consider namely \( \{ M_{case} \} \) for the proposed estimators in a real situation, where both traditional subspace estimator and the solution proposed in this paper, Fig. 2. Empirical densities of the traditional estimator and the proposed estimator of the extremal eigenvalues of the covariance matrix.

Fig. 1. Empirical distribution of the orthogonality factor achieved by the traditional subspace estimator and the solution proposed in this paper, \( M = 5, N = 15 \).

Fig. 2. Empirical densities of the traditional estimator and the proposed estimator of the extremal eigenvalues of the covariance matrix, \( M = 5, N = 4 \).

V. SIMULATIONS

In this section, we evaluate the performance of the proposed estimators in a real situation, where both \( M \) and \( N \) are low and have the same order of magnitude. We consider a covariance matrix \( R_{M} \) with three different eigenvalues, namely \( \{ 1, 4, 7 \} \), with respective multiplicities \( \{ 3, 1, 1 \} \). In this situation, the minimum number of samples per observation dimension in order to achieve the asymptotic splitting condition in (As4) is \( \xi (1) = 1.7857 \) for the smallest eigenvalue and \( \xi (2) = \xi (3) = 5.2765 \) for the other two. The observations are generated according to a Gaussian distribution, and the results are obtained from a total of \( 10^4 \) independent runs. In Figure 1 we represent, for the case \( M = 5, N = 15 \), the orthogonality factor of the estimated subspaces corresponding to the traditional and proposed estimator. This factor is defined as

\[
O(m) = \frac{\text{tr}\left[ E_m E_m^H P_m \right]}{\text{tr}\left[ (I_m - E_m E_m^H) P_m \right]}
\]

where \( P_m = \hat{E}_m \hat{E}_m^H \) for the traditional estimator, and \( P_m = \sum_{k=1}^{M} \theta_m(k) \hat{E}_k \hat{E}_k^H \) for the proposed estimator. Observe that we obtain a much better orthogonality factor with the proposed approach, even when the splitting condition (As4) is not fulfilled (as it is the case for the largest eigenvalue eigenvector, i.e. \( m = 3 \)). In Figure 2 we represent the empirical distribution of the two extreme eigenvalues in the case \( M = 5, N = 4 \) (i.e., the number of samples was lower than the observation dimension). Observe that our estimator is able to correct part of the bias of the sample eigenvalues, providing more reliable estimates, even when the number of samples is lower than the observation dimension and much lower than the minimum required by (As4).

VI. CONCLUSIONS

We have presented new estimators for the eigenvalues and associated subspaces of the sample covariance matrix. 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, provided that an asymptotic eigenvalue splitting condition is fulfilled. In any case, it is shown via simulations that the performance of the proposed estimators is excellent, even in the non-asymptotic regime and when the asymptotic eigenvalue splitting condition does not hold.

REFERENCES