Sample Eigenvalue Based Detection of High-Dimensional Signals in White Noise Using Relatively Few Samples

Raj Rao Nadakuditi, *Member, IEEE*, and Alan Edelman

Abstract—The detection and estimation of signals in noisy, limited data is a problem of interest to many scientific and engineering communities. We present a mathematically justifiable, computationally simple, sample-eigenvalue-based procedure for estimating the number of high-dimensional signals in white noise using relatively few samples. The main motivation for considering a sample-eigenvalue-based scheme is the computational simplicity and the robustness to eigenvector modelling errors which can adversely impact the performance of estimators that exploit information in the sample eigenvectors. There is, however, a price we pay by discarding the information in the sample eigenvectors; we highlight a fundamental asymptotic limit of sample-eigenvalue-based detection of weak or closely spaced high-dimensional signals from a limited sample size. This motivates our heuristic definition of the *effective number of identifiable signals* which is equal to the number of "signal" eigenvalues of the population covariance matrix which exceed the noise variance by a factor strictly greater than $1 + \sqrt{\text{Dimensionality of the system/Sample size}}$. The fundamental asymptotic limit brings into sharp focus why, when there are too few samples available so that the effective number of signals is less than the actual number of signals, underestimation of the model order is unavoidable (in an asymptotic sense) when using any sample-eigenvalue-based detection scheme, including the one proposed herein. The analysis reveals why adding more sensors can only exacerbate the situation. Numerical simulations are used to demonstrate that the proposed estimator, like Wax and Kailath's MDL-based estimator, consistently estimates the true number of signals in the dimension fixed, large sample size limit and the effective number of identifiable signals, unlike Wax and Kailath's MDL-based estimator, in the large dimension, (relatively) large sample size limit.

Index Terms—Detection, random matrices, sample covariance matrix.

I. INTRODUCTION

THE observation vector, in many signal processing applications, can be modelled as a superposition of a finite number of signals embedded in additive noise. Detecting the number of signals present becomes a key issue and is often the starting point for the signal parameter estimation problem in diverse applications such as biomedical signal processing [1], wireless communications [2], geophysical signal processing [3], array processing [4], and finance [5], [6] to list a few. When the signals and the noise are assumed, as we do in this paper, to be samples of a stationary, ergodic Gaussian vector process, the sample covariance matrix formed from m observations has the Wishart distribution [7]. This paper uses an information theoretic approach, inspired by the seminal work of Wax and Kailath [8], for determining the number of signals in white noise from the eigenvalues of the Wishart distributed empirical covariance matrix formed from *relatively few samples*.

The reliance of the Wax and Kailath estimator and their successors [9]–[13], to list a few, on the distributional properties of the eigenvalues of *nonsingular* Wishart matrices render them inapplicable in high-dimensional, sample-starved settings where the empirical covariance matrix is singular. Ad hoc modifications to such estimators are often not mathematically justified, and it is seldom clear, even using simulations as in [11], whether a fundamental limit of detection is being encountered *vis à vis* the chronically reported symptom of underestimating the number of signals.

This paper addresses both of these issues using relevant results [14]–[17] from large random matrix theory. The main contributions of this paper are: 1) the development of a mathematically justifiable, computationally simple, sample-eigenvalue-based signal detection algorithm that operates effectively in sample-starved settings and 2) the introduction of the concept of *effective number of (identifiable) signals* which brings into sharp focus a fundamental limit in the identifiability, under sample size constraints, of closely spaced or weak signals using sample-eigenvalue-based detection techniques of the sort developed in this paper.

The proposed estimator exploits the distributional properties of the trace of powers, i.e., the moments of the eigenvalues, of (singular and nonsingular) Wishart distributed large-dimensional sample covariance matrices. The definition of *effective number of identifiable signals* is based on the mathematically rigorous results of Baik and Silverstein [18], Paul [19], and Baik *et al.* [20], the physically rigorous derivation by Hoyle and Rattray [21], and the heuristic derivation of the first author in [22]. This concept captures the fundamental limit of sample-eigenvalue-based detection by explaining why, in the large-system relatively large sample-size limit, *if the signal level is below a threshold* that depends on the noise variance, sample size, and the dimensionality of the system, *then reliable sample-eigenvalue-based detection is not possible*. This brings into sharp

focus the fundamental undetectability of weak or closely spaced signals using sample-eigenvalue-based schemes when too few samples are available. Adding more sensors will only exacerbate the problem by raising the detectability threshold.

Conversely, if the signal level is above this threshold, and the dimensionality of the system is large enough, then reliable detection using the proposed estimator is possible. We demonstrate this via numerical simulations that illustrate the superiority of the proposed estimator with respect to the Wax–Kailath MDL-based estimator. Specifically, simulations reveal that while both the new estimator and the Wax–Kailath MDL estimator are consistent estimators of the number of signals k in the dimensionality n fixed, sample size $m \to \infty$ sense, the MDL estimator is an inconsistent estimator of the *effective number of signals* in the large-system relatively large sample-size limit, i.e., in $n, m(n) \to \infty$ limit where the ratio $n/m(n) \to c \in (0, \infty)$ sense. Simulations suggest that the new estimator is a consistent estimator of the effective number of signals in the $n, m(n) \to \infty$ with $n/m(n) \to c \in (0, \infty)$ sense. We note that simulations will demonstrate the applicability of the proposed estimator in moderate dimensional settings as well.

The paper is organized as follows. The problem formulation in Section II is followed by a summary in Section III of the relevant properties of the eigenvalues of large-dimensional Wishart distributed sample covariance matrices. An estimator for the number of signals present that exploits these results is derived in Section IV. An extension of these results to the frequency domain is discussed in Section V. Consistency of the proposed estimator and the concept of *effective number of signals* is discussed in Section VI. Simulation results that illustrate the superior performance of the new method in high-dimensional, sample-starved settings are presented in Section VII. Some concluding remarks are presented in Section VIII.

II. PROBLEM FORMULATION

We observe m samples ("snapshots") of possibly signal bearing n-dimensional snapshot vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ where for each i, $\mathbf{x}_i \sim \mathcal{N}_n(0, \mathbf{R})$ and $\mathbf{x}_i$ are mutually independent. The snapshot vectors are modelled as

$$\mathbf{x}_i = \begin{cases} \mathbf{z}_i & \text{No Signal} \\ \mathbf{A}\,\mathbf{s}_i + \mathbf{z}_i & \text{Signal Present} \end{cases} \quad \text{for } i = 1, \ldots, m \quad (1)$$

where $\mathbf{z}_i \sim \mathcal{N}_n(0, \sigma^2 \mathbf{I})$, denotes an n-dimensional (real or circularly symmetric complex) Gaussian noise vector where σ^2 is assumed to be unknown, $\mathbf{s}_i \sim \mathcal{N}_k(\mathbf{0}, \mathbf{R}_s)$ denotes a k-dimensional (real or circularly symmetric complex) Gaussian signal vector with covariance $\mathbf{R}_s$, and $\mathbf{A}$ is a $n \times k$ unknown nonrandom matrix. In array processing applications, the jth column of the matrix $\mathbf{A}$ encodes the parameter vector associated with the jth signal whose magnitude is described by the jth element of $\mathbf{s}_i$.

Since the signal and noise vectors are independent of each other, the covariance matrix of $\mathbf{x}_i$ can be decomposed as

$$\mathbf{R} = \mathbf{\Psi} + \sigma^2 \mathbf{I} \quad (2)$$

where

$$\mathbf{\Psi} = \mathbf{A}\mathbf{R}_s\mathbf{A}' \quad (3)$$

with $'$ denoting the conjugate transpose. Assuming that the matrix $\mathbf{A}$ is of full column rank, i.e., the columns of $\mathbf{A}$ are linearly independent, and that the covariance matrix of the signals $\mathbf{R}_s$ is nonsingular, it follows that the rank of $\mathbf{\Psi}$ is k. Equivalently, the $n - k$ smallest eigenvalues of $\mathbf{\Psi}$ are equal to zero.

If we denote the eigenvalues of $\mathbf{R}$ by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ then it follows that the smallest $n - k$ eigenvalues of $\mathbf{R}$ are all equal to σ^2 so that

$$\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n = \lambda = \sigma^2. \quad (4)$$

Thus, if the true covariance matrix $\mathbf{R}$ were known *a priori*, the dimension of the signal vector k can be determined from the multiplicity of the smallest eigenvalue of $\mathbf{R}$. When there is no signal present, all the eigenvalues of $\mathbf{R}$ will be identical. The problem in practice is that the covariance matrix $\mathbf{R}$ is unknown so that such a straightforward algorithm cannot be used. The signal detection and estimation problem is hence posed in terms of an inference problem on m samples of n-dimensional multivariate real or complex Gaussian snapshot vectors.

A. Prior Work and Motivation for New Sample-Eigenvalue-Based Estimator

Inferring the number of signals from these m samples reduces the signal detection problem to a model selection problem for which there are many approaches. A classical approach to this problem, developed by Bartlett [23] and Lawley [24], uses a sequence of hypothesis tests. Though this approach is sophisticated, the main problem is the subjective judgement needed by the practitioner in selecting the threshold levels for the different tests.

Information theoretic criteria for model selection such as those developed by Akaike [25], [26], Schwartz [27], and Rissanen [28] address this problem by proposing the selection of the model which gives the minimum *information criteria*. The criteria for the various approaches is generically a function of the log-likelihood of the maximum-likelihood estimator of the parameters of the model and a term which depends on the number of parameters of the model that penalizes overfitting of the model order.

For the problem formulated above, Kailath and Wax [8] propose an estimator for the number of signals (*assuming $m > n$ and $\mathbf{x}_i \in \mathbb{C}^n$*) based on the eigenvalues $l_1 \geq l_2 \geq \cdots \geq l_n$ of the sample covariance matrix (SCM) defined by

$$\widehat{\mathbf{R}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i' = \frac{1}{m} \mathbf{X}\mathbf{X}' \quad (5)$$

where $\mathbf{X} = [\mathbf{x}_1 | \ldots | \mathbf{x}_m]$ is the matrix of observations (samples). The Akaike Information Criteria (AIC) form of the estimator is given by

$$\hat{k}_{\text{AIC}} = \arg\min_{k} -2(n-k)m \log \frac{g(k)}{a(k)} + 2k(2n-k)$$
$$\text{for } k \in \mathbb{N} : 0 \leq k < n \quad (6)$$

while the mnimum descriptive length (MDL) criterion is given by

$$\hat{k}_{\text{MDL}} = \arg\min_{k} -(n-k)m \log \frac{g(k)}{a(k)} + \frac{1}{2}k(2n-k) \log m$$
$$\text{for } k \in \mathbb{N} : 0 \leq k < n \quad (7)$$

where $g(k) = \prod_{j=k+1}^{n} l_j^{1/(n-k)}$ is the geometric mean of the $n - k$ smallest sample eigenvalues and $a(k) = (1/(n - k)) \sum_{j=k+1}^{n} l_j$ is their arithmetic mean.

It is known [8] that the AIC form inconsistently estimates the number of signals, while the MDL form estimates the number of signals consistently in the classical n fixed, $m \to \infty$ sense. The simplicity of the estimator and the large sample consistency are among the primary reasons why the Kailath-Wax MDL estimator continues to be employed in practice [4]. In the two decades since the publication of the WK paper, researchers have come up with many innovative solutions for making the estimators more "robust" in the sense that estimators continue to work well in settings where the underlying assumptions of snapshot and noise Gaussianity and intersnapshot independence (e.g., in the presence of multipath) can be relaxed as in the work of Zhao *et al.* [29], [30], Xu *et al.* [31], and Stoica-Cedervall [32] among others [33].

Despite its obvious practical importance, the robustness of the algorithm to model mismatch is an issue that we shall not address in this paper. Instead we aim to revisit the original problem considered by Wax and Kailath with the objective of designing an estimator that is robust to high-dimensionality and sample size constraints. We are motivated by the observation that the most important deficiency of the Wax–Kailath estimator and its successors that is yet to be satisfactorily resolved occurs in the setting where the sample size is smaller than the number of sensors, i.e., when $m < n$, which is increasingly the case in many state-of-the-art radar and sonar systems where the number of sensors exceeds the sample size by a factor of $10 - 100$. In this situation, the SCM is singular and the estimators become degenerate, as seen in (6) and (7). Practitioners often overcome this in an *ad hoc* fashion by, for example, restricting k in (7) to integer values in the range $0 \le k < \min(n, m)$ so that

$$\hat{k}_{\overline{\text{MDL}}} = \arg\min -(n-k)m \log \frac{g(k)}{a(k)} + \frac{1}{2}k(2n-k)\log m$$

$$\text{for } k \in \mathbb{N} : 0 \le k < \min(n, m). \quad (8)$$

Since large sample, i.e., $m \gg n$, asymptotics [34] were used to derive the estimators in [8], there is no rigorous theoretical justification for such a reformulation even if the simulation results suggest that the WK estimators are working "well enough."

This is true for other sample-eigenvalue-based solutions found in the literature that exploit the sample eigenvalue order statistics [12], [13], employ a Bayesian framework by imposing priors on the number of signals [35], involve solving a set of possibly high-dimensional nonlinear equations [9], or propose sequential hypothesis testing procedures [10]. The fact that these solutions are computationally more intensive or require the practitioner to set subjective threshold levels makes them less attractive than the WK MDL solution; more importantly, they do not address the sample-starved setting in their analysis or their simulations either.

For example, in [13], Fishler *et al.* use simulations to illustrate the performance of their algorithm with $n = 7$ sensors and a sample size $m > 500$ whereas in a recent paper [33], Fishler and Poor illustrate their performance with $n = 10$ sensors and $m > 50$ samples. Van Trees discusses the various techniques for estimating the number of signals in [4, Sec. 7.8]; the sample-starved setting where $m = O(n)$ or $m < n$ is not treated in the simulations either.

There is, however, some notable work on detecting the number of signals using short data records. Particle filter-based techniques [36], have proven to be particularly useful in such short data record settings. Their disadvantage, from our perspective, is that they require the practitioner to model the eigenvectors of the underlying population covariance matrix as well; this makes them especially sensitive to model mismatch errors that are endemic to challenging real-world high-dimensional settings such as those encountered in large-array signal processing. This observation also applies to ℓ_1 minimization-based techniques [37], [38] that exploit the sparse representation of observations in an known overcomplete basis (e.g., composed of samples from the array manifold as in [39]).

This motivates our development of a mathematically justifiable, sample-eigenvalue-based estimator with a computational complexity comparable to that of the modified WK estimator in (8) that remains robust to high dimensionality and sample size constraints. The new estimator given by (26) is derived in Section IV. An implicit assumption in the derivation of (26) is that the number of signals is much smaller than the system size, i.e., $k \ll n$. A rather important shortcoming of our sample-eigenvalue-based detection scheme, as we shall elaborate in Section VI, is that it just might not be possible to detect low-level or closely spaced signals when there are too few samples available. In other words, if the signals are not strong enough or not spaced far enough part, then not only will proposed estimator consistently underestimate the number of signals but so will any other sample-eigenvalue-based detector.

The concept of the effective number of signals provides insight into the fundamental limit, due to snapshot constraints in high-dimensional settings, of reliable signal detection by eigen-inference , i.e., by using the sample eigenvalues alone. This helps identify scenarios where algorithms that exploit any structure in the eigenvectors of the signals, such as the MUSIC and the Capon-MVDR algorithms in sensor array processing [4], particle filter-based techniques [36] or sparsity (in a known basis) exploiting techniques [37]–[39], might be better able to tease out lower level signals from the background noise. It is worth noting that the proposed approach remain relevant in situations where the eigenvector structure has been identified. This is because eigen-inference methodologies are inherently robust to eigenvector modelling errors that occur in high-dimensional settings. Thus the practitioner may use the proposed methodologies to complement and "robustify" the inference provided by algorithms that exploit the eigenvector structure.

III. PERTINENT RESULTS FROM RANDOM MATRIX THEORY

Analytically characterizing the distribution of the sample eigenvalues, as a function of the population eigenvalues, is the first step in designing a sample-eigenvalue-based estimator that is robust to high-dimensionality and sample size constraints. For arbitrary covariance $\mathbf{R}$, the joint density function of the eigenvalues $l_1, \ldots, l_n$ of the SCM $\hat{\mathbf{R}}$ *when* $m > n + 1$ is shown to be given by [34]

$$f(l_1, \ldots, l_n) = \tilde{Z}_{n,m}^{\beta} \sum \prod_{i=1}^{n} l_i^{\beta(m-n+1)/2-1} \prod_{i<j}^{n} |l_i - l_j|^{\beta} \prod_{i=1}^{n} dl_i$$

$$\times \int_{\mathbf{Q}} \exp\left(-\frac{m\beta}{2} \text{Tr}\left(\mathbf{R}^{-1}\mathbf{Q}\hat{\mathbf{R}}\mathbf{Q}'\right)\right) d\mathbf{Q} \quad (9)$$
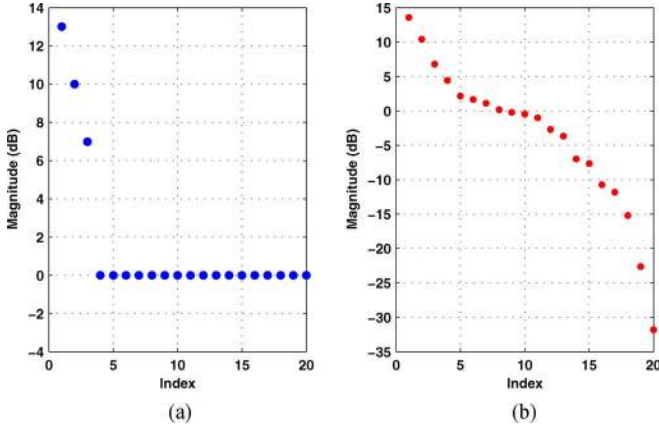
Fig. 1. Blurring of sample eigenvalues relative to the population eigenvalues when are there a finite number of snapshots. (a) True eigenvalues. (b) Sample eigenvalues formed from 20 snapshots.

where $l_1 > \cdots > l_n > 0$, $\widetilde{Z}_{n,m}^\beta$ is a normalization constant, and $\beta = 1$ (or 2) when $\widehat{\mathbf{R}}$ is real (respectively, complex). In (9), $\mathbf{Q} \in \mathbf{O}(n)$ when $\beta = 1$ while $\mathbf{Q} \in \mathbf{U}(n)$ when $\beta = 2$ where $\mathbf{O}(n)$ and $\mathbf{U}(n)$ are, respectively, the set of $n \times n$ orthogonal and unitary matrices with Haar measure.

Note that the exact characterization of the joint density of the eigenvalues in (9) involves a multidimensional integral over the orthogonal (or unitary) group. This makes it intractable for analysis without resorting to asymptotics. Anderson's landmark paper [34] does just that by characterizing the distribution of the sample eigenvalues using large sample asymptotics. A classical result due to Anderson establishes the consistency of the sample eigenvalues in the dimensionality n fixed, sample size $m \to \infty$ asymptotic regime [34]. When the dimensionality is small and there are plenty of samples available, Anderson's analysis suggests that the sample eigenvalues will be (roughly) symmetrically centered around the population eigenvalues. When the dimensionality is large, and the sample size is relatively small, Anderson's prediction of sample eigenvalue consistency is in stark contrast to the asymmetric spreading of the sample eigenvalues that is observed in numerical simulations. This is illustrated in Fig. 1 where the $n = 20$ eigenvalues of a SCM formed from $m = 20$ samples are compared with the eigenvalues of the underlying population covariance matrix.

The role of random matrix theory comes in because of new analytical results that are able to precisely describe the spreading of the sample eigenvalues exhibited in Fig. 1. Since our new estimator explicitly exploits these analytical results, the use of our estimator in high-dimensional, sample-starved settings is more mathematically justified than other sample-eigenvalue-based approaches found in the literature that explicitly use Anderson's sample eigenvalue consistency results. See, for example [8, eq. (13a), pp. 389] and [12, eq. (5), pp. 2243].

We argue that this is particularly so in settings where $m < n$, where practitioners (though, not the original authors!) have often invoked the equality between the nonzero eigenvalues of the $\widehat{\mathbf{R}} = (1/m)\mathbf{X}\mathbf{X}'$ and the matrix $(1/m)\mathbf{X}'\mathbf{X}$ to justify *ad hoc* modifications to estimators that use only the nonzero eigenvalues of the SCM, as in (8). We contend that such an *ad hoc* modification to any estimator that explicitly uses Anderson's sample eigenvalue consistency results is mathemati-

cally unjustified because the sample eigen-spectrum blurring, which is only further exacerbated in the $m < n$ regime, remains unaccounted for.

Before summarizing the pertinent results, we note that the analytical breakthrough alluded to is a consequence of considering the large-system relatively large sample-size asymptotic regime as opposed to "classical" fixed system-size large-sample-size asymptotic regime. Mathematically speaking, the new results describe the distribution of the eigenvalues in $n, m \to \infty$ with $n/m \to c \in (0, \infty)$ asymptotic regime as opposed to the n fixed, $m \to \infty$ regime *à la* Anderson. We direct the reader to Johnstone's excellent survey for a discussion on these asymptotic regimes [40, p. 9] and much more.

A. Eigenvalues of the Signal-Free SCM

A central object in the study of large random matrices is the empirical distribution function (e.d.f.) of the eigenvalues, which for an arbitrary matrix $\mathbf{A}$ with n real eigenvalues (counted with multiplicity) is defined as

$$F^{\mathbf{A}}(x) = \frac{\text{Number of eigenvalues of } \mathbf{A} \le x}{n}. \quad (10)$$

For a broad class of random matrices, the sequence of e.d.f.'s can be shown to converge in the $n \to \infty$ limit to a nonrandom distribution function [41]. Of particular interest is the convergence of the e.d.f. of the signal-free SCM which is described next.

1) Proposition 3.1: Let $\widehat{\mathbf{R}}$ denote a signal-free sample covariance matrix formed from an $n \times m$ matrix of observations with i.i.d. Gaussian samples of mean zero and variance $\lambda = \sigma^2$. Then the e.d.f. $F^{\widehat{\mathbf{R}}}(x) \to F^W(x)$ almost surely for every x, as $m, n \to \infty$ and $c_m = n/m \to c$ where

$$dF^W(x) = \max\left(0, \left(1 - \frac{1}{c}\right)\right)\delta(x)$$
$$+ \frac{\sqrt{(x - a_-)(a_+ - x)}}{2\pi\lambda xc}\mathbb{1}_{[a_-, a_+]}(x)\,dx \quad (11)$$

with $a_\pm = \lambda(1 \pm \sqrt{c})^2$, $\mathbb{1}_{[a,b]}(x) = 1$ when $a \le x \le b$ and zero otherwise, and $\delta(x)$ is the Dirac delta function.

Proof: This result was proved in [42] and [43] in very general settings. Other proofs include [15], [44], and [45]. The probability density in (11) is often referred to as the Marčenko–Pastur density. ∎

Fig. 2 plots the Marčenko–Pastur density in (11) for $\lambda = 1$ and different values of $c = \lim n/m$. Note that as $c \to 0$, the eigenvalues are increasingly clustered around $x = 1$ but for modest values of c, the spreading is quite significant.

The almost sure convergence of the e.d.f. of the signal-free SCM implies that the moments of the eigenvalues converge almost surely, so that

$$\frac{1}{n}\sum_{i=1}^{n} l_i^k \xrightarrow{a.s.} \int x^k dF^W(x) =: M_k^W. \quad (12)$$

The moments of the Marčenko–Pastur density are given by [15], [46]

$$M_k^W = \lambda^k \sum_{j=0}^{k-1} c^j \frac{1}{j+1}\binom{k}{j}\binom{k-1}{j}. \quad (13)$$
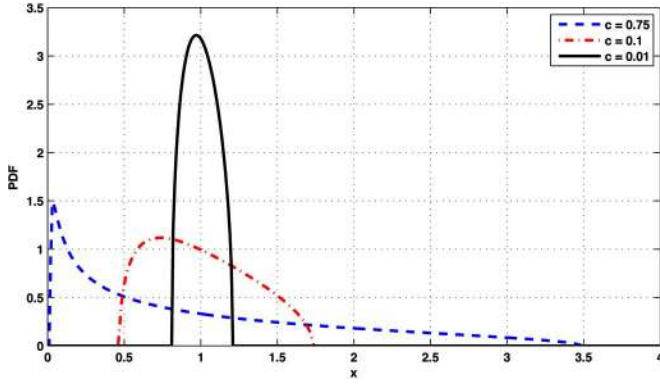
Fig. 2. Marčenko–Pastur density, given by (11), for the eigenvalues of the signal-free sample covariance matrix with noise variance 1 and $c = \lim n/m$.

For finite n and m, the sample moments, i.e., $(1/n) \sum_{i=1}^{n} l_i^k$, will fluctuate about these limiting values. The precise nature of the fluctuations is described next.

2) Proposition 3.2: If $\widehat{\mathbf{R}}$ satisfies the hypotheses of Proposition 3.1 for some λ then as $m, n \to \infty$ and $c_m = n/m \to c \in (0, \infty)$, we have

$$n \left(\begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} l_i \\ \frac{1}{n} \sum_{i=1}^{n} l_i^2 \end{bmatrix} - \begin{bmatrix} \lambda \\ \lambda^2(1+c) \end{bmatrix} \right)$$

$$\xrightarrow{\mathcal{D}} \mathcal{N} \left(\underbrace{\begin{bmatrix} 0 \\ \left(\frac{2}{\beta} - 1\right)\lambda^2 c \end{bmatrix}}_{=\boldsymbol{\mu}_Q}, \frac{2}{\beta} \underbrace{\begin{bmatrix} \lambda^2 c & 2\lambda^3 c\,(c+1) \\ 2\lambda^3 c\,(c+1) & 2\lambda^4 c\,(2c^2+5c+2) \end{bmatrix}}_{=\mathbf{Q}} \right)$$
(14)

where the convergence is in distribution.

Proof: This result appears in [14] and [15] for the real case and in [16] for the real and complex cases. The result for general β appears in Dumitriu and Edelman [17]. ∎

We now use the result in Proposition 3.2 to develop a test statistic q_n whose distribution is independent of the unknown noise variance λ. The distributional properties of this test statistic are described next.

3) Proposition 3.3: Assume $\widehat{\mathbf{R}}$ satisfies the hypotheses of Proposition 3.1 for some λ. Consider the statistic

$$q_n = \frac{\frac{1}{n} \sum_{i=1}^{n} l_i^2}{\left(\frac{1}{n} \sum_{i=1}^{n} l_i \right)^2}.$$

Then as $m, n \to \infty$ and $c_m = n/m \to c \in (0, \infty)$, we have

$$n \left[q_n - (1+c) \right] \xrightarrow{\mathcal{D}} \mathcal{N} \left(\left(\frac{2}{\beta} - 1 \right) c, \frac{4}{\beta} c^2 \right) \quad (15)$$

where the convergence is in distribution.

Proof: Define the function $g(x, y) = y/x^2$. Its gradient vector $\nabla g(x, y)$ is given by

$$\nabla g(x,y) := \begin{bmatrix} \partial_x g(x,y) & \partial_y g(x,y) \end{bmatrix}^T = \begin{bmatrix} \frac{-2y}{x^3} & \frac{1}{x^2} \end{bmatrix}^T. \quad (16)$$

The statistic q_n can be written in terms of $g(x, y)$ as simply $q_n = g((1/n) \sum_{i=1}^{n} l_i, (1/n) \sum_{i=1}^{n} l_i^2)$. The limiting distribution of q_n can be deduced from the distributional properties of $(1/n) \sum_{i=1}^{n} l_i$ and $(1/n) \sum_{i=1}^{n} l_i^2$ established in Proposition 3.2. Specifically, by an application of the delta method [47], we obtain that as $n, m \to \infty$ with $n/m \to c \in (0, \infty)$, then

$$n \left[q_n - g(\lambda, \lambda^2(1+c)) \right] \xrightarrow{\mathcal{D}} \mathcal{N}(\mu_q, \sigma_q^2)$$

where the mean μ_q and the variance σ_q^2 are given by

$$\mu_q = \boldsymbol{\mu}_Q^T \nabla g \left(\lambda, \lambda^2(1+c) \right) \quad (17a)$$

$$\sigma_q^2 = \frac{2}{\beta} \nabla g \left(\lambda, \lambda^2(1+c) \right)^T \mathbf{Q} \nabla g \left(\lambda, \lambda^2(1+c) \right). \quad (17b)$$

Substituting, the expressions for $\boldsymbol{\mu}_Q$ and $\mathbf{Q}$ given in (14), in (17a), and (17b) gives us the required expressions for the mean and the variance of the normal distribution on the right-hand side of (15). ∎

B. Eigenvalues of the Signal Bearing SCM

When there are k signals present then, in the $n \to \infty$ limit, where k is kept fixed, the *limiting* e.d.f. of $\widehat{\mathbf{R}}$ will still be given by Proposition 3.1. This is because the e.d.f., defined as in (10), weights the contribution of every eigenvalue equally so that effect of the k/n fraction of "signal" eigenvalues vanishes in the $n \to \infty$ limit.

Note, however, that in the signal-free case, i.e., when $k = 0$, Proposition 3.1 and the result in [48] establish the almost sure convergence of the largest eigenvalue of the SCM to $\lambda(1+\sqrt{c})^2$. In the signal bearing case, a so-called phase transition phenomenon is observed, in that the largest eigenvalue will converge to a limit different from that in the signal-free case only if the "signal" eigenvalues are above a certain threshold. This is described next.

1) Proposition 3.4: Let $\widehat{\mathbf{R}}$ denote a sample covariance matrix formed from an $n \times m$ matrix of Gaussian observations whose columns are independent of each other and identically distributed with mean $\mathbf{0}$ and covariance $\mathbf{R}$. Denote the eigenvalues of $\mathbf{R}$ by $\lambda_1 \geq \lambda_2 > \cdots \geq \lambda_k > \lambda_{k+1} = \cdots \lambda_n = \lambda$. Let l_j denote the jth largest eigenvalue of $\widehat{\mathbf{R}}$. Then as $n, m \to \infty$ with $c_m = n/m \to c \in (0, \infty)$

$$l_j \to \begin{cases} \lambda_j \left(1 + \frac{\lambda c}{\lambda_j - \lambda} \right) & \text{if } \lambda_j > \lambda \left(1 + \sqrt{c}\right) \\ \lambda \left(1 + \sqrt{c}\right)^2 & \text{if } \lambda_j \leq \lambda \left(1 + \sqrt{c}\right) \end{cases} \quad (18)$$

for $j = 1, \ldots, k$ and the convergence is almost surely.

Proof: This result appears in [18] for very general settings. A matrix theoretic proof for the real-valued SCM case may be found in [19] while a determinental proof for the complex case may be found in [20]. A heuristic derivation that relies on an interacting particle system interpretation of the sample eigenvalues appears in [22]. ∎

The "signal" eigenvalues strictly below the threshold described in Proposition 3.4 exhibit, on rescaling, fluctuations described by the Tracy–Widom distributions [49]–[51]. There is an interesting phenomenon described in [20] that occurs for "signal" eigenvalues exactly equal to the threshold. For "signal" eigenvalues above the threshold described in Proposition 3.4, the fluctuations about the asymptotic limit are described next.
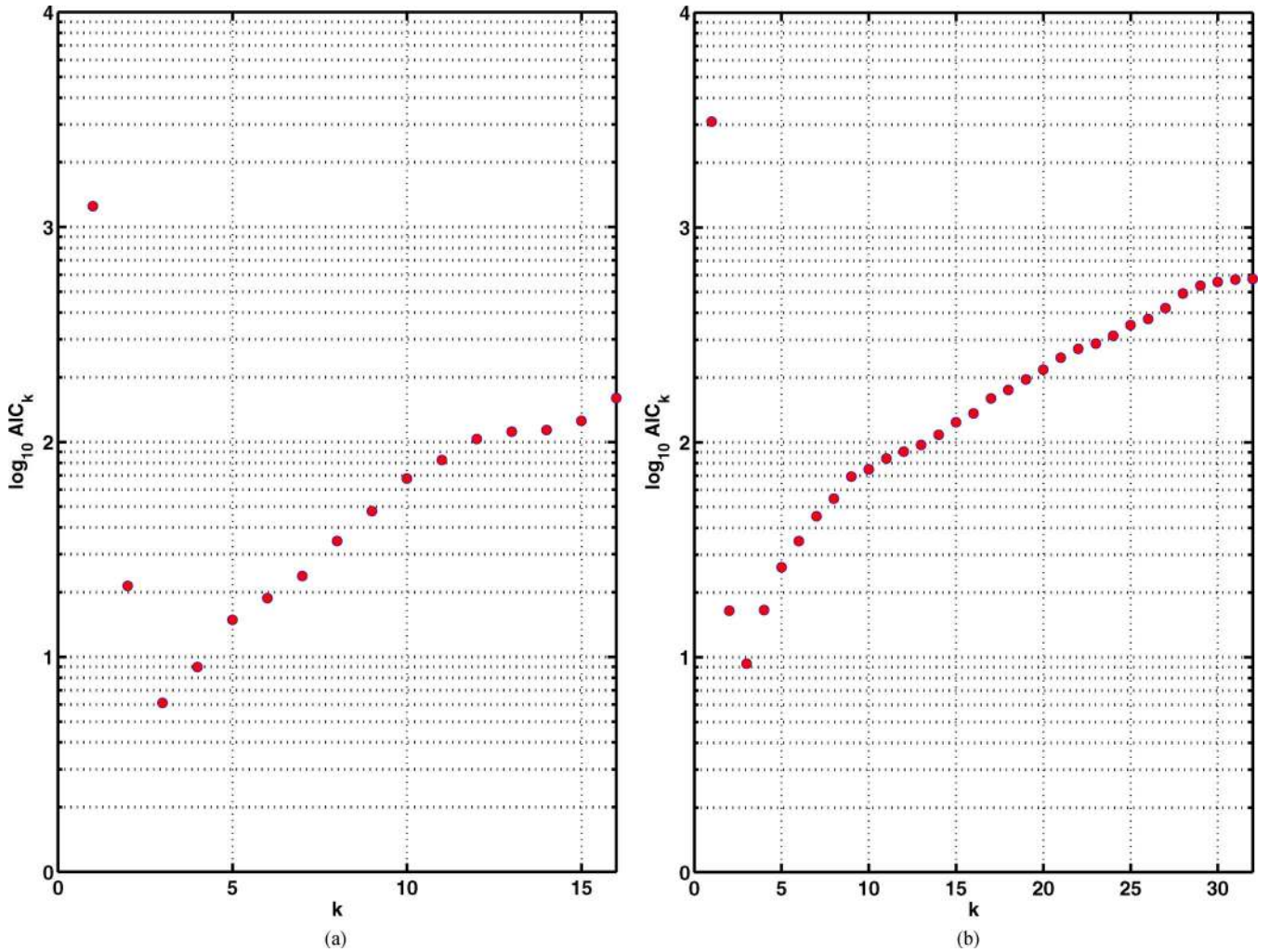
Fig. 3. Sample realizations of the proposed criterion when there are $k = 2$ complex-valued signals and $\lambda_1 = 10$, $\lambda_2 = 3$, and $\lambda_3 = \cdots = \lambda_n = 1$. (a) Complex signals: $n = 16$, $m = 32$. (b) Complex signals: $n = 32$, $m = 64$.

2) Proposition 3.5: Assume that $\widehat{\mathbf{R}}$ and $\mathbf{R}$ satisfy the hypotheses of Proposition 3.4. If $\lambda_j > \lambda(1 + \sqrt{c})$ has multiplicity 1 and if $\sqrt{m}|c - n/m| \to 0$ then

$$\sqrt{n}\left[l_j - \lambda_j\left(1 + \frac{\lambda c}{\lambda_j - \lambda}\right)\right]$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{2}{\beta}\lambda_j^2\left(1 - \frac{c}{(\lambda_j - \lambda)^2}\right)\right) \quad (19)$$

where the convergence in distribution is almost surely.

Proof: A matrix theoretic proof for the real case may be found in [19] while a determinantal proof for the complex case may be found in [20]. The result has been strengthened for non-Gaussian situations by Baik and Silverstein for general $c \in (0, \infty)$ [52]. ∎

IV. ESTIMATOR FOR THE NUMBER OF SIGNALS

We derive an information theoretic estimator for the number of signals by exploiting the distributional properties of the moments of eigenvalues of the (signal-free) SCM given by Propositions 3.2 and 3.3, as follows. The overarching principle used is that, given an observation $\mathbf{y} = [y(1), \ldots, y(N)]$ and a family

of models, or equivalently a parameterized family of probability densities $f(\mathbf{y}|\boldsymbol{\theta})$ indexed by the parameter vector $\boldsymbol{\theta}$, we select the model which gives the minimum AIC [26] defined by

$$\text{AIC}_k = -2\log f(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + 2k \quad (20)$$

where $\widehat{\boldsymbol{\theta}}$ is the maximum-likelihood estimate of $\boldsymbol{\theta}$, and k is the number of free parameters in $\boldsymbol{\theta}$. Since the noise variance is unknown, the parameter vector of the model, denoted by $\boldsymbol{\theta}_k$, is given by

$$\boldsymbol{\theta}_k = [\lambda_1, \ldots, \lambda_k, \sigma^2]^T. \quad (21)$$

There are thus $k + 1$ free parameters in $\boldsymbol{\theta}_k$. Assuming that there are $k < \min(n, m)$ signals, the maximum-likelihood estimate of the noise variance is given by [34] (which Proposition 3.2 corroborates in the $m < n$ setting)

$$\widehat{\sigma}^2_{(k)} = \frac{1}{n - k}\sum_{i=k+1}^{n} l_i \quad (22)$$

where $l_1 \geq \cdots \geq l_n$ are the eigenvalues of $\widehat{\mathbf{R}}$.

Consider the test statistic

$$q_k = n \left[\frac{\frac{1}{n-k} \sum_{i=k+1}^{n} l_i^2}{\left(\widehat{\sigma}_{(k)}^2\right)^2} - (1+c) \right] - \left(\frac{2}{\beta} - 1\right) c \qquad (23)$$

$$= n \left[\frac{\frac{1}{n-k} \sum_{i=k+1}^{n} l_i^2}{\left(\frac{1}{n-k} \sum_{i=k+1}^{n} l_i\right)^2} - (1+c) \right] - \left(\frac{2}{\beta} - 1\right) c \qquad (24)$$

for a constant $c > 0$. This choice is motivated by Proposition 3.3 where we note that the limiting distribution of the test statistic q_0 in the signal-free case does not depend on the (unknown) noise variance. When $k > 0$ signals are present and assuming $k \ll n$, then the distributional properties of the $n - k$ "noise" eigenvalues are closely approximated by the distributional properties of the eigenvalues given by Proposition 3.2 of the signal-free SCM, i.e., when $k = 0$. It is hence reasonable to approximate the distribution of the statistic q_k with the normal distribution whose mean and variance, for some $c > 0$, given in Proposition 3.3. The log-likelihood function $\log f(q_k | \widehat{\boldsymbol{\theta}})$, for large n, m can hence be approximated by

$$-\log f(q_k | \widehat{\boldsymbol{\theta}}) \approx \frac{q_k^2}{2\frac{4}{\beta}c^2} + \underbrace{\frac{1}{2} \log 2\pi \frac{4}{\beta} c^2}_{\text{Constant}}. \qquad (25)$$

In (23) and (25), it is reasonable (Bai and Silverstein provide an argument in [16]) to use $c_m = n/m$ for the (unknown) limiting parameter $c = \lim n/m$. Plugging in $c \approx c_m = n/m$ into (23) and (25), ignoring the constant term on the right-hand side of (25) when the log-likelihood function is substituted into (20) yields the estimator (26a)–(26b), shown at the bottom of the page. In (26), the l_i's are the eigenvalues of the sample covariance matrix $\widehat{\mathbf{R}}$. Fig. 3 plots sample realizations of the score function.

V. EXTENSION TO FREQUENCY DOMAIN AND VECTOR SENSORS

When the m snapshot vectors $\mathbf{x}_i(w_j)$ for $i = 1, \dots, m$ represent Fourier coefficients vectors at frequency w_j then the sample covariance matrix

$$\widehat{\mathbf{R}}(w_j) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i(w_j) \mathbf{x}_i(w_j)' \qquad (27)$$

(a) When n and m large enough, so that the (largest) "signal" eigenvalue is sufficiently separated from the "noise" eigenvalues, then reliable detection is possible.

(b) When n and m are not quite large enough so that the (largest) "signal" eigenvalue is *not* sufficiently separated from the largest "noise" eigenvalue, then reliable detection becomes challenging.
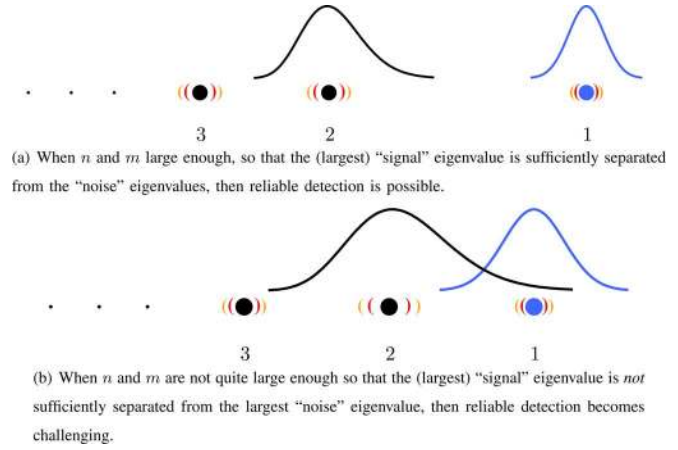
Fig. 4. Finite system dimensionality and sample size induced fluctuations of the single "signal" eigenvalue on the far right and the "noise" eigenvalues about their limiting positions are shown. The magnitude of the fluctuations impacts the ability to discriminate the "signal" eigenvalue from the largest "noise" eigenvalue.

is the periodogram estimate of the spectral density matrix at frequency w_j. The time-domain approach carries over to the frequency domain so that the estimator in (26) remains applicable with $l_i \equiv l_i(w_j)$ where $l_1(w_j) \geq l_2(w_j) \geq \cdots \geq l_n(w_j)$ are the eigenvalues of $\widehat{\mathbf{R}}(w_j)$.

When the signals are wideband and occupy M frequency bins, denoted by $w_1, \dots, w_M$, then the information on the number of signals present is contained in all the bins. The assumption that the observation time is much larger than the correlation times of the signals (sometimes referred to as the SPLOT assumption—stationary process, long observation time) ensures that the Fourier coefficients corresponding to the different frequencies are statistically independent.

Thus the AIC-based criterion for detecting the number of wideband signals that occupy the frequency bands $w_1, \dots, w_M$ is obtained by summing the corresponding criterion in (26) over the frequency range of interest

$$t_{j,k} = \left[(n-k) \frac{\sum_{i=k+1}^{n} l_i(w_j)^2}{\left(\sum_{i=k+1}^{n} l_i(w_j)\right)^2} - \left(1 + \frac{n}{m}\right) \right] n - \left(\frac{2}{\beta} - 1\right) \frac{n}{m} \qquad (28a)$$

$$t_k = \left[(n-k) \frac{\sum_{i=k+1}^{n} l_i^2}{\left(\sum_{i=k+1}^{n} l_i\right)^2} - \left(1 + \frac{n}{m}\right) \right] n - \left(\frac{2}{\beta} - 1\right) \frac{n}{m} \qquad (26a)$$

$$\hat{k}_{\text{NEW}} = \arg\min_k \left\{ \frac{\beta}{4} \left[\frac{m}{n}\right]^2 t_k^2 \right\} + 2(k+1) \quad \text{for } k \in \mathbb{N} : 0 \leq k < \min(n, m) \qquad (26b)$$

Here $\beta = 1$ if $\mathbf{x}_i \in \mathbb{R}^n$, and $\beta = 2$ if $\mathbf{x}_i \in \mathbb{C}^n$

Fig. 5. Comparison of the performance of the new estimator in (26) with the MDL estimator in (8) for various n (system size) and m (sample size). (a) Empirical probability that $\hat{k} = 2$ for various n and m. (b) Empirical probability that $\hat{k} = 1$ for various n and m. (c) Empirical probability that $\hat{k} = 0$ for various n and m.

$$\hat{k}_{\text{NEW}} = \underset{k \in \mathbb{N} : 0 \leq k < \min(n, m)}{\arg\min} \sum_{j=1}^{M} \frac{\beta}{4} \left[\frac{m}{n}\right]^2 t_{j,k}^2 + 2M(k+1). \tag{28b}$$

When the number of snapshots is severely constrained, the SPLOT assumption is likely to be violated so that the Fourier coefficients corresponding to different frequencies will not be statistically independent. This will likely degrade the performance of the proposed estimators.

When the measurement vectors represent quaternion-valued narrowband signals, then $\beta = 4$ so that the estimator in (26) can be used. Quaternion-valued vectors arise when the data collected from vector sensors is represented using quaternions as in [53].

Fig. 6. Effective number of identifiable signals, computed using (32) for the values of n (system size) and m (sample size) considered in Fig. 5 when the population covariance matrix has two signal eigenvalues $\lambda_1 = 10$ and $\lambda_2 = 3$ and $n - 2$ noise eigenvalues $\lambda_3 = \cdots = \lambda_n = \sigma^2 = 1$.
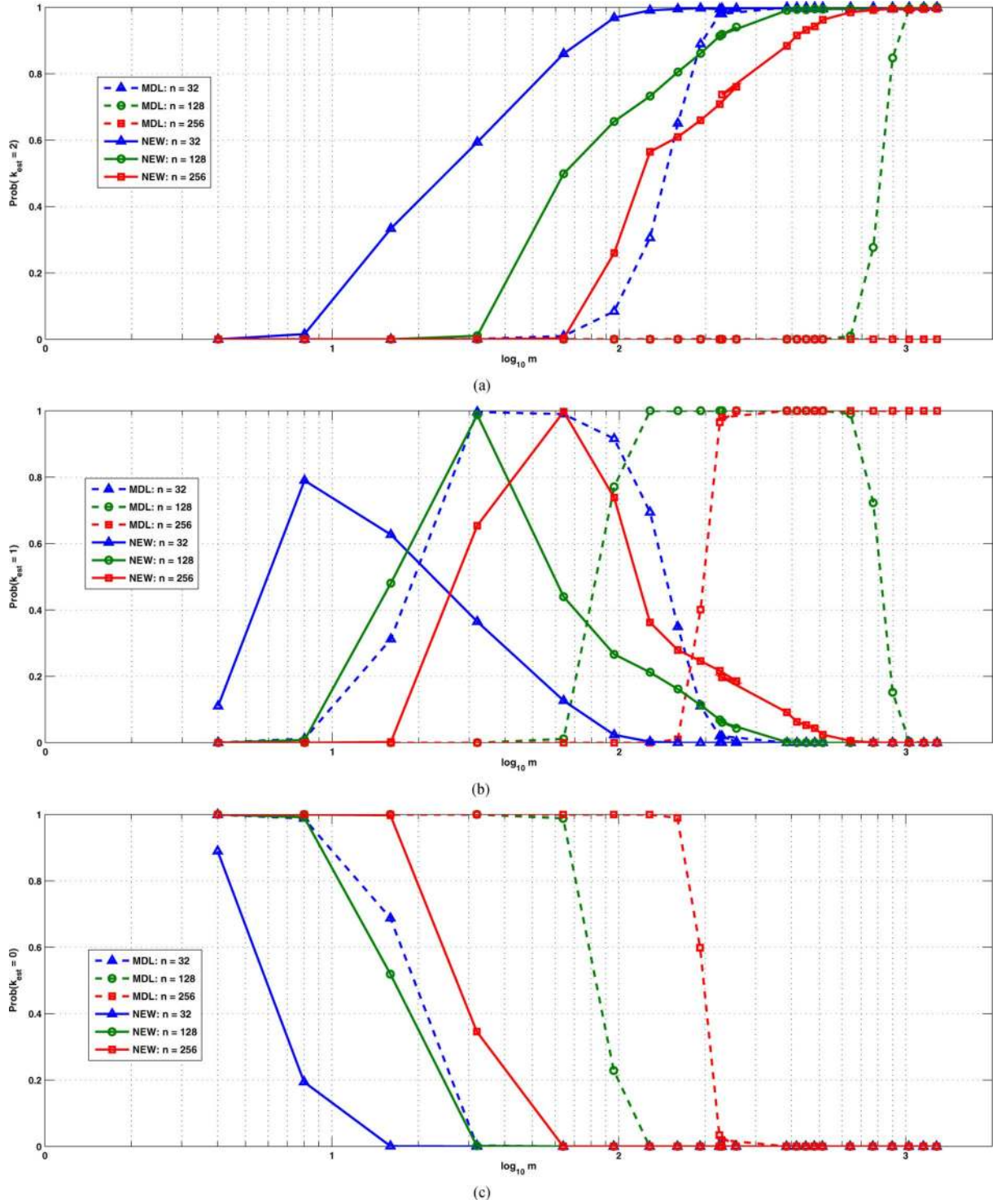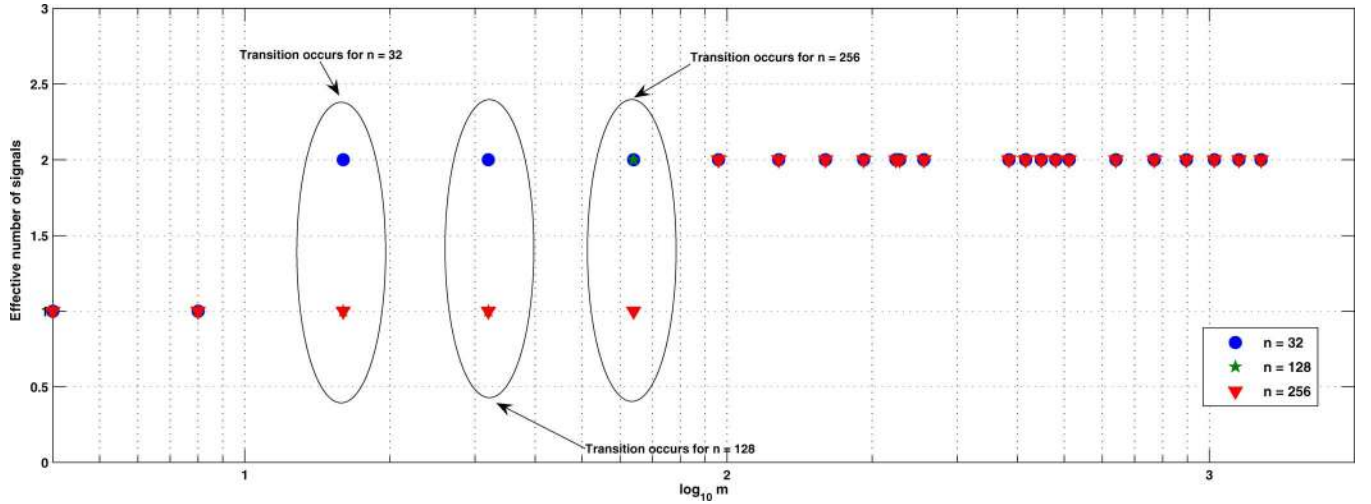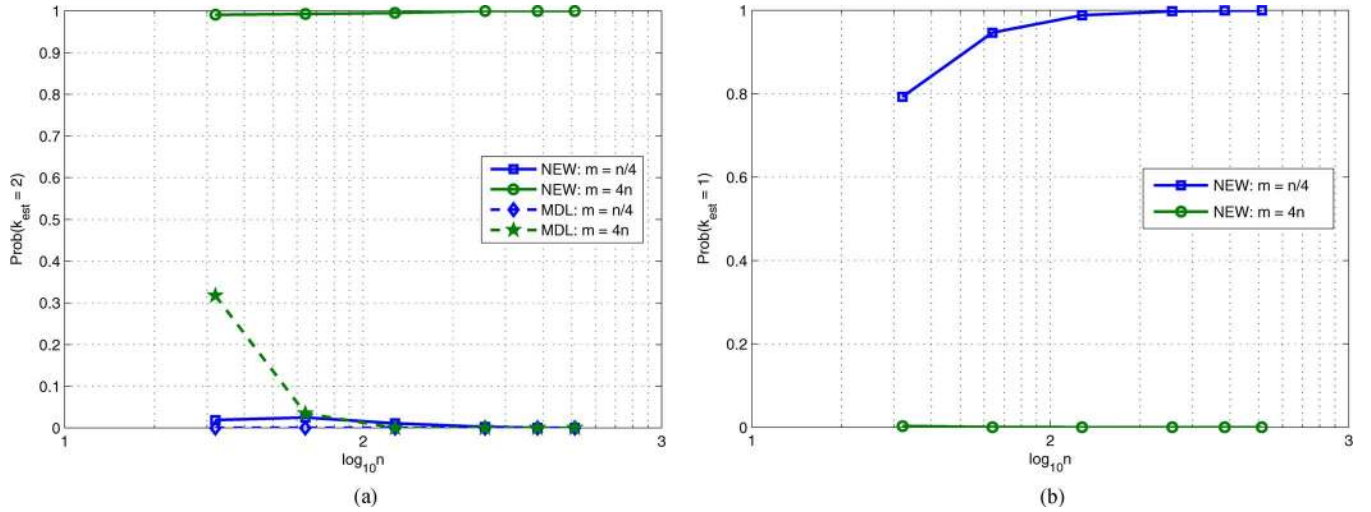

Fig. 7. Comparison of the performance of the new estimator in (26) with the MDL estimator in (8) for various values of n (system size) with m (sample size) such that n/m is fixed. (a) Empirical probability that $\hat{k} = 2$ for various n and fixed n/m. (b) Empirical probability that $\hat{k} = 1$ for various n and fixed n/m.

VI. CONSISTENCY OF THE ESTIMATOR AND THE EFFECTIVE NUMBER OF IDENTIFIABLE SIGNALS

For a fixed sample size, and system dimensionality, the probability of detecting a signal is the most practically useful criterion for comparing the performance of different estimators. For theoretical purposes, however, the large sample consistency of the estimator is (usually) more analytically tractable and hence often supplied as the justification for employing an estimator. We conjecture that the proposed algorithm is a consistent estimator of the true number of signals in the "classical" large sample asymptotic regime in the sense made explicit next.

1) Conjecture 6.1: Let $\mathbf{R}$ be a $n \times n$ covariance matrix that satisfies the hypothesis of Proposition 3.4. Let $\hat{\mathbf{R}}$ be a sample covariance matrix formed from m snapshots. Then in the n fixed, $m \to \infty$ limit, $\hat{k}$ is a consistent estimator of k where $\hat{k}$ is the estimate of the number of signals obtained using (26).

The "classical" notion of large sample consistency does not adequately capture the suitability of an estimator in high-dimensional, sample-starved settings when $m < n$ or $m = O(n)$. In

such settings, it is more natural to investigate the consistency properties of the estimator in the large-system relatively large sample-size limit instead. We can use Proposition 3.4 to establish an important property of the proposed estimator in such a limit.

2) Theorem 6.2: Let $\mathbf{R}$ and $\widetilde{\mathbf{R}}$ be two $n \times n$ sized covariance matrices whose eigenvalues are related as

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p, \lambda_{p+1}, \ldots, \lambda_k, \lambda, \ldots, \lambda) \quad (29a)$$

$$\widetilde{\mathbf{\Lambda}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p, \lambda, \ldots, \lambda) \quad (29b)$$

where for some $c \in (0, \infty)$, and all $i = p+1, \ldots, k$, $\lambda < \lambda_i \leq \lambda(1 + \sqrt{c})$. Let $\hat{\mathbf{R}}$ and $\widehat{\widetilde{\mathbf{R}}}$ be the associated sample covariance matrices formed from m snapshots. Then for every $n, m(n) \to \infty$ such that $c_m = n/m \to c$

$$\mathrm{Prob}(\hat{k} = j \mid \mathbf{R}) \to \mathrm{Prob}(\hat{k} = j \mid \widetilde{\mathbf{R}}) \qquad \text{for } j = 1, \ldots, p \quad (30a)$$

TABLE I

COMPARISON OF THE EMPIRICAL PERFORMANCE OF THE NEW ESTIMATOR IN (26) WITH THE WAX–KAILATH MDL ESTIMATOR IN (8) WHEN THE POPULATION COVARIANCE MATRIX HAS TWO SIGNAL EIGENVALUES $\lambda_1 = 10$ AND $\lambda_2 = 3$ AND $n - 2$ NOISE EIGENVALUES $\lambda_3 = \cdots = \lambda_n = \sigma^2 = 1$. THE EFFECTIVE NUMBER OF IDENTIFIABLE SIGNALS IS COMPUTED USING (32) WHILE THE SEPARATION METRIC $\mathcal{Z}_j^{\mathrm{SEP}}$ IS COMPUTED FOR $j = k_{\mathrm{eff}}$ USING (37). HERE n DENOTES THE SYSTEM SIZE, m DENOTES THE SAMPLE SIZE, AND THE SNAPSHOT VECTORS, MODELLED IN (1), ARE TAKEN TO BE COMPLEX-VALUED

n	m	n/m	k_{eff}	$\mathcal{Z}_2^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	128	0.25	2	2.1909	-	0.0035	0.9898	0.0067
64	256	0.25	2	3.0984	-	-	0.9922	0.0077
128	512	0.25	2	4.3818	-	-	0.9930	0.0070
256	1024	0.25	2	6.1968	-	-	0.9938	0.0063
512	2048	0.25	2	8.7636	-	-	0.9948	0.0053
1024	4096	0.25	2	12.3935	-	-	0.9952	0.0047

(a) The performance of the new estimator when $m = 4n$.

n	m	n/m	k_{eff}	$\mathcal{Z}_2^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	128	0.25	2	2.1909	-	0.6920	0.3080	-
64	256	0.25	2	3.0984	-	0.9663	0.0338	-
128	512	0.25	2	4.3818	-	1.0000	-	-
256	1024	0.25	2	6.1968	-	1.0000	-	-
512	2048	0.25	2	8.7636	-	1.0000	-	-
1024	4096	0.25	2	12.3935	-	1.0000	-	-

(b) The performance of the Wax-Kailath MDL based estimator when $m = 4n$.

n	m	n/m	k_{eff}	$\mathcal{Z}_1^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	8	4	1	3.1588	0.1867	0.7920	0.0213	-
64	16	4	1	4.4673	0.0260	0.9537	0.0203	-
128	32	4	1	6.3177	0.0013	0.9898	0.0090	-
256	64	4	1	8.9345	-	0.9972	0.0027	-
512	128	4	1	12.6353	-	0.9998	0.0003	-
1024	256	4	1	17.8690	-	-	-	-

(c) The performance of the new estimator when $m = n/4$.

n	m	n/m	k_{eff}	$\mathcal{Z}_1^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	8	4	1	3.1588	0.9900	0.0100	-	-
64	16	4	1	4.4673	0.9998	0.0002	-	-
128	32	4	1	6.3177	1.0000	-	-	-
256	64	4	1	8.9345	1.0000	-	-	-
512	128	4	1	12.6353	1.0000	-	-	-
1024	256	4	1	17.8690	1.0000	-	-	-

(d) The performance of the Wax-Kailath MDL based estimator when $m = n/4$.

and

$$\mathrm{Prob}(\hat{k} > p \,|\, \mathbf{R}) \to \mathrm{Prob}(\hat{k} > p \,|\, \widetilde{\mathbf{R}}) \qquad (30b)$$

where the convergence is almost surely and $\hat{k}$ is the estimate of the number of signals obtained using the algorithm in (26).

Proof: The result follows from Proposition 3.4. The $(p + 1)$th sample eigenvalue of both $\widehat{\mathbf{R}}$ and $\widehat{\widetilde{\mathbf{R}}}$ will converge almost surely to the same limit equal to $\lambda(1 + \sqrt{c})^2$. Furthermore, for every $i = 1, \ldots, p$, the ith largest eigenvalue of $\widehat{\mathbf{R}}$ will converge to the same limit, given by Proposition 3.4, as the ith largest eigenvalue of $\widehat{\widetilde{\mathbf{R}}}$. The fluctuations about this limit will hence be identical by Proposition 3.5 so that (30) follows in the asymptotic limit. ∎

Note that the rate of convergence to the asymptotic limit for $\mathrm{Prob}(\hat{k} > p \,|\, \mathbf{R})$ and $\mathrm{Prob}(\hat{k} > p \,|\, \widetilde{\mathbf{R}})$ will, in general, depend on the eigenvalue structure of $\mathbf{R}$ and may be arbitrarily slow. Thus, Theorem 6.2 yields *no insight into rate of convergence-type issues* which are important in practice. Rather, the theorem is a statement on the asymptotic equivalence, from an identifiability point of view, of sequences of sample covariance matrices which are related in the manner described. At this point, we are unable to prove the consistency of the proposed estimator as this would require more a refined analysis that characterizes the fluctuations of subsets of the (ordered) "noise" eigenvalues. The statement regarding consistency of the proposed estimator, in the sense of large-system relatively large sample-size limit, is presented as a conjecture with numerical simulations used as nondefinitive yet corroborating evidence.

3) Conjecture 6.3: Let $\mathbf{R}$ be a $n \times n$ covariance matrix that satisfies the hypothesis of Proposition 3.4. Let $\widehat{\mathbf{R}}$ be a sample covariance matrix formed from m snapshots. Define

$$k_{\mathrm{eff}}(c \,|\, \mathbf{R}) := \text{Number of eigenvalues of } \mathbf{R} > \lambda(1 + \sqrt{c}). \quad (31)$$

Then in the $m, n \to \infty$ limit with $c_m = n/m \to c$, $\hat{k}$ is a consistent estimator of $k_{\mathrm{eff}}(c)$ where $\hat{k}$ is the estimate of the number of signals obtained using the algorithm in (26).

Motivated by Proposition 3.4, we (heuristically) define the *effective number of (identifiable) signals* as

$$k_{\mathrm{eff}}(\mathbf{R}) = \text{Number of eigenvalues of } \mathbf{R} > \sigma^2 \left(1 + \sqrt{\frac{n}{m}} \right). \quad (32)$$

Conjecture 6.3 then simply states that the proposed estimator is a consistent estimator of the effective number of (identifiable) signals in the large-system relatively large sample-size limit.

TABLE II
COMPARISON OF THE EMPIRICAL PERFORMANCE OF THE NEW ESTIMATOR IN (26) WITH THE WAX–KAILATH MDL ESTIMATOR IN (8) WHEN THE POPULATION COVARIANCE MATRIX HAS TWO SIGNAL EIGENVALUES $\lambda_1 = 10$ AND $\lambda_2 = 3$ AND $n - 2$ NOISE EIGENVALUES $\lambda_3 = \cdots = \lambda_n = \sigma^2 = 1$. THE EFFECTIVE NUMBER OF IDENTIFIABLE SIGNALS IS COMPUTED USING (32) WHILE THE SEPARATION METRIC $\mathcal{Z}_j^{\mathrm{SEP}}$ IS COMPUTED FOR $j = k_{\mathrm{eff}}$ USING (37). HERE n DENOTES THE SYSTEM SIZE, m DENOTES THE SAMPLE SIZE, AND THE SNAPSHOT VECTORS, MODELLED IN (1), ARE TAKEN TO BE COMPLEX-VALUED

n	m	n/m	k_{eff}	$\mathcal{Z}_1^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	4	8	1	2.5218	0.8822	0.1177	-	-
64	8	8	1	3.5663	0.6937	0.3063	-	-
128	16	8	1	5.0435	0.5178	0.4823	-	-
256	32	8	1	7.1326	0.3450	0.6550	-	-
512	64	8	1	10.0871	0.1993	0.8007	-	-
1024	128	8	1	14.2653	0.0745	0.9255	-	-

(a) The performance of the new estimator when $m = n/8$.

n	m	n/m	k_{eff}	$\mathcal{Z}_1^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	4	8	1	2.5218	0.9995	0.0005	-	-
64	8	8	1	3.5663	1.0000	-	-	-
128	16	8	1	5.0435	1.0000	-	-	-
256	32	8	1	7.1326	1.0000	-	-	-
512	64	8	1	10.0871	1.0000	-	-	-
1024	128	8	1	14.2653	1.0000	-	-	-

(b) The performance of the Wax-Kailath MDL based estimator when $m = n/8$.

n	m	n/m	k_{eff}	$\mathcal{Z}_2^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	32	1	2	1.0887	-	0.3608	0.5972	0.0395
64	64	1	2	1.5396	-	0.2727	0.6805	0.0445
128	128	1	2	2.1773	-	0.2130	0.7352	0.0505
256	256	1	2	3.0792	-	0.1895	0.7592	0.0485
512	512	1	2	4.3546	-	0.1812	0.7678	0.0495
1024	1024	1	2	6.1584	-	0.1665	0.7780	0.0537

(c) The performance of the new estimator when $m = n$.

n	m	n/m	k_{eff}	$\mathcal{Z}_2^{\mathrm{SEP}}$	$P(\hat{k}=0)$	$P(\hat{k}=1)$	$P(\hat{k}=2)$	$P(\hat{k}=3)$
32	32	1	2	1.0887	0.0040	0.9955	0.0005	-
64	64	1	2	1.5396	0.0008	0.9992	-	-
128	128	1	2	2.1773	-	1.0000	-	-
256	256	1	2	3.0792	-	1.0000	-	-
512	512	1	2	4.3546	0.0003	0.9998	-	-
1024	1024	1	2	6.1584	0.0030	0.9970	-	-

(d) The performance of the Wax-Kailath MDL based estimator when $m = n$.

A. Asymptotic Identifiability of Two Closely Spaced Signals

Suppose there are two uncorrelated (hence, independent) signals so that $\mathbf{R}_s = \mathrm{diag}(\sigma_{\mathrm{S}1}^2, \sigma_{\mathrm{S}2}^2)$. In (1) let $\mathbf{A} = [\mathbf{v}_1 \mathbf{v}_2]$. In a sensor array processing application, we think of $\mathbf{v}_1 \equiv \mathbf{v}_1(\theta_1)$ and $\mathbf{v}_2 \equiv \mathbf{v}_2(\theta_2)$ as encoding the array manifold vectors for a source and an interferer with powers $\sigma_{\mathrm{S}1}^2$ and $\sigma_{\mathrm{S}2}^2$, located at θ_1 and θ_2, respectively. The covariance matrix given by

$$\mathbf{R} = \sigma_{\mathrm{S}1}^2 \mathbf{v}_1 \mathbf{v}_1' + \sigma_{\mathrm{S}2}^2 \mathbf{v}_2 \mathbf{v}_2' + \sigma^2 \mathbf{I} \qquad (33)$$

has the $n - 2$ smallest eigenvalues $\lambda_3 = \cdots = \lambda_n = \sigma^2$ and the two largest eigenvalues

$$\lambda_1 = \sigma^2 + \frac{\left(\sigma_{\mathrm{S}1}^2 \|\mathbf{v}_1\|^2 + \sigma_{\mathrm{S}2}^2 \|\mathbf{v}_2\|^2\right)}{2}$$
$$+ \frac{\sqrt{\left(\sigma_{\mathrm{S}1}^2 \|\mathbf{v}_1\|^2 - \sigma_{\mathrm{S}2}^2 \|\mathbf{v}_2\|^2\right)^2 + 4\sigma_{\mathrm{S}1}^2 \sigma_{\mathrm{S}2}^2 |\langle \mathbf{v}_1, \mathbf{v}_2 \rangle|^2}}{2}$$
$$(34a)$$

$$\lambda_2 = \sigma^2 + \frac{\left(\sigma_{\mathrm{S}1}^2 \|\mathbf{v}_1\|^2 + \sigma_{\mathrm{S}2}^2 \|\mathbf{v}_2\|^2\right)}{2}$$
$$- \frac{\sqrt{\left(\sigma_{\mathrm{S}1}^2 \|\mathbf{v}_1\|^2 - \sigma_{\mathrm{S}2}^2 \|\mathbf{v}_2\|^2\right)^2 + 4\sigma_{\mathrm{S}1}^2 \sigma_{\mathrm{S}2}^2 |\langle \mathbf{v}_1, \mathbf{v}_2 \rangle|^2}}{2}$$
$$(34b)$$

respectively. Applying the result in Proposition 3.4 allows us to express the effective number of signals as

$$k_{\mathrm{eff}} = \begin{cases} 2, & \text{if } \sigma^2 \left(1 + \sqrt{\frac{n}{m}}\right) < \lambda_2 \\ 1, & \text{if } \lambda_2 \leq \sigma^2 \left(1 + \sqrt{\frac{n}{m}}\right) < \lambda_1 \\ 0, & \text{if } \lambda_1 \leq \sigma^2 \left(1 + \sqrt{\frac{n}{m}}\right). \end{cases} \qquad (35)$$

In the special situation when $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \|\mathbf{v}\|$ and $\sigma_{\mathrm{S}1}^2 = \sigma_{\mathrm{S}2}^2 = \sigma_{\mathrm{S}}^2$, we can (in an asymptotic sense) reliably detect the presence of *both signals* from the sample eigenvalues alone whenever we have the conditions in (36), shown at the bottom of the next page. Equation (36) captures the tradeoff between the identifiability of two closely spaced signals, the dimensionality of the system, the number of available snapshots, and the cosine of the angle between the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

We note that the concept of the effective number of signals is an asymptotic concept for large dimensions and relatively large sample sizes. For moderate dimensions and sample sizes, the fluctuations in the "signal" and "noise" eigenvalues affect the reliability of the underlying detection procedure as illustrated in Fig. 4. From Proposition 3.4, we expect that the largest "noise" eigenvalue will, with high probability, be found in a neighborhood around $\sigma^2(1 + \sqrt{n/m})^2$ while the "signal" eigenvalues will, with high probability, be found in

a neighborhood around $\lambda_j \left(1 + \sigma^2 n/(m(\lambda_j - \sigma^2))\right)$. From Proposition 3.5, we expect the "signal" eigenvalues to exhibit Gaussian fluctuations with a standard deviation of approximately $\sqrt{(2/\beta n)\lambda_j^2 (1 - n/m(\lambda_j - \sigma^2)^2)}$. This motivates our definition of the metric Z_j^{Sep} given by

$$Z_j^{\mathrm{Sep}} := \frac{\lambda_j \left(1 + \frac{\sigma^2 n}{m(\lambda_j - \sigma^2)}\right) - \sigma^2 \left(1 + \sqrt{\frac{n}{m}}\right)^2}{\sqrt{\frac{2}{\beta n}\lambda_j^2 \left(1 - \frac{n}{m(\lambda_j - \sigma^2)^2}\right)}} \qquad (37)$$

which measures the (theoretical) separation of the jth "signal" eigenvalue from the largest "noise" eigenvalue in standard deviations of the jth signal eigenvalue's fluctuations. Simulations suggest that reliable detection (with an empirical probability greater than 90%) of the effective number of signals is possible if Z_j^{Sep} lies between 5 and 15. This large range of values for the minimum Z_j^{Sep}, which we obtained from the results in Section VII, suggests that a more precise characterization of the finite system, finite sample performance of the estimator will have to take into account the more complicated-to-analyze interactions between the "noise" and the "signal" eigenvalues that are negligible in the large-system relatively large sample-size limit. Nonetheless, because of the nature of the random matrix results on which our guidance is based, we expect our heuristics to be more accurate in high-dimensional, relatively large sample size settings than the those proposed in [54] and [12] which rely on Anderson's classical large sample asymptotics.

We note that in using estimator in (26) it is implicit that n is the dimension of the "spatial" variable whose covariance is modelled by (2) while m is the dimension of the "temporal" variable. Flipping the role of "space" and "time" when forming the sample covariance matrix will rescale the eigenvalues but *will not affect the threshold below which the "signal" eigenvalues will be indistinguishable from the "noise" eigenvalues.* The noninterchangeability of the "space" and "time" dimensions in the determination of the thresholds in (31) and (36) is a consequence of the asymmetricity of our initial assumption that the observations have spatial but not temporal correlations.

VII. NUMERICAL SIMULATIONS

We now illustrate the performance of our estimator using Monte Carlo simulations. The results obtained provide evidence for the consistency properties conjectured in Section VI. In all of our simulations, we use a population covariance matrix $\mathbf{R}$ that has arbitrarily fixed, yet unknown, eigenvectors, $k = 2$ "signal" eigenvalues with $\lambda_1 = 10$ and $\lambda_2 = 3$, and $n - 2$ "noise" eigenvalues with $\lambda_3 = \cdots = \lambda_n = \lambda = \sigma^2 = 1$. We assume that the snapshot vectors $\mathbf{x}_i$ modelled as in (1) are complex-valued so that we must plug in $\beta = 2$ in (26); the choice of complex-valued signals is motivated by applications in array signal processing and wireless communications.

Using 4000 Monte Carlo simulations, and various values of n and m, we obtain an estimate of the number of signals from the eigenvalues of the sample covariance matrix using our new estimator and the modified Wax–Kailath estimator, described in (26) and (8) respectively. We do not consider the Wax–Kailath AIC estimator in (6) in our simulations because of its proven [8] inconsistency in the fixed system size, large sample limit—we are interested in estimators that exhibit the consistency conjectured in Section VI in *both* asymptotic regimes. A thorough comparison of the performance of our estimator with other estimators (and their *ad hoc* modifications) found in the literature is beyond the scope of this paper.

We first investigate the large sample consistency in the classical sense of n fixed and $m \to \infty$. For a choice of n and different values of m we compute the empirical probability of detecting two signals. For large values of m we expect both the new and the Wax–Kailath MDL estimator to detect both signals with high probability. Fig. 5 plots the results obtained in the numerical simulations.

Fig. 5(a) shows that for $n = 32$ and $n = 128$, if m is large enough then either estimator is able to detect both signals with high probability. However, the new estimator requires significantly fewer samples to do so than the Wax–Kailath MDL estimator.

Fig. 5(b) and (c) plot the empirical probability of detecting one and zero signals, respectively, as a function of m for various values of n. The results exhibit the chronically reported symptom of estimators underestimating the number of signals—this is not surprising given the discussion in Section VI. Fig. 6 plots the effective number of identifiable signals k_{eff}, determined using (32) for the various values of n and m considered. We observe that the values of n and m for which the empirical probability of the new estimator detecting one signal is high also correspond to regimes where $k_{\mathrm{eff}} = 1$. This suggests that the asymptotic concept of the effective number of signals remains relevant in a nonasymptotic regime as well. At the same time, however, one should not expect the signal identifiability/unidentifiability predictions in Section VI to be accurate in the severely sample-starved settings where $m \ll n$. For example, Fig. 5(c) reveals that the new estimator detects zero signals with high empirical probability when there are less than ten samples available even though $k_{\mathrm{eff}} = 1$ in this regime from Fig. 6. In the large-system relatively large sample-size asymptotic limit, however, these predictions are accurate—we discuss this next.

When $m = 4n$ samples are available, Fig. 7(a) shows that the proposed estimator consistently detects two signals while the Wax–Kailath MDL estimator does not. However, when $m = n/4$ samples are available, Fig. 7(a) suggests that neither estimator is able to detect both the signals present. A closer examination of the empirical data presents a different picture. The population covariance has two signal eigenvalues $\lambda_1 = 10$

$$\text{Asymptotic identifiability condition :} \qquad \sigma_S^2 \|\mathbf{v}\|^2 \left(1 - \frac{|\langle \mathbf{v}_1, \mathbf{v}_2 \rangle|}{\|\mathbf{v}\|}\right) > \sigma^2 \sqrt{\frac{n}{m}} \qquad (36)$$

and $\lambda_2 = 3$ with the noise eigenvalues $\sigma^2 = 1$. Hence, when $m = n/4$, from (31), the effective number of signals $k_{\text{eff}} = 1$. Fig. 7(b) shows that for large n and $m = n/4$, the new estimator consistently estimates one signal, as expected. We remark that the signal eigenvalue λ_2, which is asymptotically unidentifiable, falls exactly on the threshold in (31). The consistency of the new estimator with respect to the effective number of signals corroborates the asymptotic tightness of the fundamental limit of sample-eigenvalue-based detection. On inspecting Table I(b) and (d) it is evident that the Wax–Kailath MDL estimator consistently underestimates the effective number of signals in the large-system relatively large sample-size limit.

Table II provides additional evidence for Conjecture 6.3. We offer Table II(c) and (d) as evidence for the observation that large-system relatively large sample consistency aside, the rate of convergence can be arbitrary slow and cannot be entirely explained by the metric $\mathcal{Z}_{k_{\text{eff}}}^{\text{SEP}}$ in (37).

VIII. CONCLUDING REMARKS

We have developed an information theoretic approach for detecting the number of signals in white noise from the sample eigenvalues alone. The proposed estimator explicitly takes into account the blurring of the sample eigenvalues due to the finite size. The stated conjecture on the consistency of the algorithm, in both the n fixed, $m \to \infty$ sense and the $n, m(n) \to \infty$ with $n/m(n) \to c$ sense, remains to be proven. It would be interesting to investigate the impact of a broader class of penalty functions on the consistency, strong or otherwise, in both asymptotic regimes, in the spirit of [29].

In future work, we plan to address the problem of estimating the number of high-dimensional signals in noise with arbitrary covariance [30], using relatively few samples when an independent estimate of the noise sample covariance matrix, that is itself formed from relative few samples, is available. This estimator will also be of the form in (26) and will exploit the analytical characterization of properties of the traces of powers of random Wishart matrices with a covariance structure that is also random [55].

It remains an open question to analyze such signal detection algorithms in the Neyman–Pearson sense of finding the most powerful test that does not exceed a threshold probability of false detection. Finer properties, perhaps buried in the rate of convergence to the asymptotic results used, might be useful in this context. In the spirit of Wax and Kailath's original work, we developed a procedure that did not require us to make any subjective decisions on setting threshold levels. Thus, we did not consider largest eigenvalue tests in sample-starved settings of the sort described in [40], [51], [56], and the references therein. Nevertheless, if the performance can be significantly improved using a sequence of nested hypothesis tests, then this might be a price we might be ready to pay. This is especially true for the detection of low-level signals right around the threshold where the asymptotic results suggest that it becomes increasingly difficult, if not impossible, to detect signals using the sample eigenvalues alone.

ACKNOWLEDGMENT

The authors thank A. Baggeroer, W. Ballance, and the anonymous reviewers for their feedback and encouragement. The authors are especially grateful to the associate editor for his accommodating multiple requests for extensions so that the authors could incorporate the reviewers' excellent suggestions into the manuscript.

REFERENCES

[1] E. Maris, "A resampling method for estimating the signal subspace of spatio-temporal EEG/MEG data," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 8, pp. 935–949, Aug. 2003.
[2] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of frequency-selective fast-varying channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1337–1347, Aug. 2003.
[3] C.-I Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
[4] H. L. V. Trees, *Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing*. New York: Wiley, 2002.
[5] J. Bai and S. Ng, "Determining the number of factors in approximate factor models," *Econometrica*, vol. 70, no. 1, pp. 191–221, 2002.
[6] L. P. Middleton and S. E. Satchell, "Deriving the arbitrage pricing theory when the number of factors is unknown," *Quant. Finance*, vol. 1, no. 5, pp. 502–508, 2001.
[7] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20 A, pp. 32–52, 1928.
[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Feb. 1985.
[9] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array rocessing," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 38, no. 11, pp. 1959–1971, Nov. 1990.
[10] A. A. Shah and D. W. Tufts, "Determination of the dimension of a signal subspace from short data records," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2531–2535, Sep. 1994.
[11] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, Aug. 2001.
[12] E. Fishler and H. Messer, "On the use of order statistics for improved detection of signals by the MDL criterion," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2242–2247, Aug. 2000.
[13] E. Fishler, M. Grossman, and H. Messer, "Detection of signals by information theoretic criteria: General asympstotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1027–1036, May 2002.
[14] K. Johansson, "On fluctuations of random Hermitian matrices," *Duke Math. J.*, vol. 91, pp. 151–203, 1998.
[15] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivar. Anal.*, vol. 12, pp. 1–38, 1982.
[16] Z. D. Bai and J. W. Silverstein, "CLT for linear spectral statistics of a large dimensional sample covariance matrix," *Ann. Prob.*, vol. 32, pp. 553–605, 2004.
[17] I. Dumitriu and A. Edelman, "Global spectrum fluctuations for the β-Hermite and β-Laguerre ensembles via matrix models," *J. Math. Phys.*, vol. 47, no. 6, p. 063302, 2006.
[18] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivar. Anal.*, no. 6, pp. 1382–1408, 2006.
[19] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model Stanford Univ., Tech. Rep., 2005, [Online]. Available: http://anson.ucdavis.edu/~debashis/techrep/eigenlimit.pdf
[20] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Prob.*, vol. 33, no. 5, pp. 1643–1697, 2005.
[21] D. Hoyle and M. Rattray, "Principal component analysis eigenvalue spectra from data with symmetry breaking structure," *Phys. Rev. E*, vol. 69, p. 026124, 2004.
[22] R. R. Nadakuditi, "Applied Stochastic Eigen-Analysis," Ph.D., Massachusetts Inst. Technol., Dept. Elect. Eng. Comput. Sci., Cambridge, MA, 2007.
[23] M. S. Bartlett, "A note on the multiplying factors for various χ^2 approximations," *J. R. Stat. Soc., Ser. B*, vol. 16, pp. 296–298, 1954.
[24] D. N. Lawley, "Tests of significance of the latent roots of the covariance and correlation matrices," *Biometrica*, vol. 43, pp. 128–136, 1956.
[25] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inf. Theory (Tsahkadsor, 1971)*, Budapest, Romania, 1973, pp. 267–281.

[26] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 716–723, 1974.

[27] G. Schwartz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, pp. 461–464, 1978.

[28] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[29] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivar. Anal.*, vol. 20, no. 1, pp. 1–25, 1986.

[30] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals when the noise covariance matrix is arbitrary," *J. Multivar. Anal.*, vol. 20, no. 1, pp. 26–49, 1986.

[31] G. Xu, R. H. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 102–112, Jan. 1994.

[32] P. Stoica and M. Cedervall, "Detection tests for array procesing in unknown correlated noise fields," *IEEE Trans. Signal Process.*, vol. 45, no. 9, pp. 2351–2362, Sep. 1997.

[33] E. Fishler and H. V. Poor, "Estimation of the number of sources in unbalanced arrays via information theoretic criteria," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3543–3553, Sep. 2005.

[34] T. W. Anderson, "Asymptotic theory of principal component analysis," *Ann. Math. Stat.*, vol. 34, pp. 122–248, 1963.

[35] N. K. Bansal and M. Bhandary, "Bayes estimation of number of signals," *Ann. Inst. Stat. Math.*, vol. 43, no. 2, pp. 227–243, 1991.

[36] J.-R. Larocque, J. P. Reilly, and W. Ng, "Particle filters for tracking and unknown number of sources," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 2926–2937, Dec. 2002.

[37] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[38] D. L. Donoho and Y. Tsaig, "Recent advances in sparsity-driven signal recovery," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP '05)*, Mar. 2005, vol. 5, pp. 713–716.

[39] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[40] I. M. Johnstone, "High dimensional statistical inference and random matrices," in *Proc. Int. Congr. Mathematicians*, 2006.

[41] A. Edelman and N. R. Rao, "Random matrix theory," in *Acta Numerica*. Cambridge, U.K.: Cambridge Univ. Press, 2005, vol. 14, pp. 233–297.

[42] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues in certain sets of random matrices," *Mat. Sb. (N.S.)*, vol. 72, no. 114, pp. 507–536, 1967.

[43] K. W. Wachter, "The strong limits of random matrix spectra for sample matrices of independent elements," *Ann. Prob.*, vol. 6, pp. 1–18, 1978.

[44] J. W. Silverstein and S.-I. Choi, "Analysis of the limiting spectral distribution of large-dimensional random matrices," *J. Multivar. Anal.*, vol. 54, no. 2, pp. 295–309, 1995.

[45] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivar. Anal.*, vol. 55, no. 2, pp. 331–339, 1995.

[46] I. Dumitriu and E. Rassart, "Path counting and random matrix theory," *Electron. J. Combin.*, vol. 7, no. 7, p. r-43, 2003.

[47] G. Casella and R. L. Berger, *Statistical Inference*. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1990.

[48] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix," *Prob. Theory Related Fields*, vol. 78, no. 4, pp. 509–521, 1988.

[49] C. Tracy and H. Widom, "Level-spacing distributions and the Airy kernel," *Commun. Math. Phys.*, vol. 159, pp. 151–174, 1994.

[50] C. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Commun. Math. Phys.*, vol. 177, pp. 727–754, 1996.

[51] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Stat.*, vol. 29, no. 2, pp. 295–327, 2001.

[52] J. W. Silverstein, Private Communication.

[53] S. Miron, N. Le Bihan, and J. Mars, "Quaternion-MUSIC for vector-sensor array processing," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1218–1229, Apr. 2006.

[54] M. Kaveh, H. Wang, and H. Hung, "On the theoretical performance of a class of estimators of the number of narrow-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 9, pp. 1350–1352, Sep. 1987.

[55] N. R. Rao, J. Mingo, R. Speicher, and A. Edelman, "Statistical eigen-inference from large Wishart matrices," *Ann. Stat.* 2008 [Online]. Available: http://arxiv.org/abs/math/0701314, to be published

[56] N. El Karoui, "Tracy-widom limit for the largest eigenvalue of a large class of complex sample covariance matrices," *Ann. Prob.*, vol. 35, no. 2, pp. 663–714, 2007.

Raj Rao Nadakuditi (S'03–M'08) received the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, and the MIT-WHOI Joint Program in Oceanography and Applied Ocean Science in 2007.

He is currently a Postdoctoral Researcher with a joint appointment in the Department of Mathematics and the Department of Mechanical Engineering/Center for Ocean Engineering at MIT. He is a recipient of the 2007 Office of Naval Research Postdoctoral fellowship award. He particularly enjoys using/adapting/developing stochastic eigen-analysis (random matrix theory) methodologies to address problems that arise in high-dimensional statistical inference. An important component of his work is the development of implementable algorithms that can be used in real-world settings to tease out low-level signals from oceanographic, financial, and econometric time or frequency measurements.

Alan Edelman received the Ph.D. degree in mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1989.

He is currently a Professor of applied mathematics at MIT. Prior to that, he was an employee and consultant at Thinking Machines Corp. In 1990, he held a National Science Foundation postdoctoral fellowship to the European Center for Research and Advanced Training in Scientific Computation (CERFACS), Toulouse, France. He then went to the University of California at Berkeley as a Morrey Assistant Professor in the Department of Mathematics and also as a Lewy Fellow at the Lawrence Berkeley Laboratory. His research spans the areas of high-performance computing, numerical computation, linear algebra, and stochastic eigenanalysis (random matrix theory).

Dr. Edelman has won numerous prizes for his work in parallel scientific computing and numerical linear algebra including the Gordon Bell prize for parallel computing and a Householder prize for numerical linear algebra.