Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples

Raj Rao Nadakuditi and Jack W. Silverstein

***Abstract*—The detection problem in statistical signal processing can be succinctly formulated: given m (possibly) signal bearing, n-dimensional signal-plus-noise snapshot vectors (samples) and N statistically independent n-dimensional noise-only snapshot vectors, can one reliably infer the presence of a signal? This problem arises in the context of applications as diverse as radar, sonar, wireless communications, bioinformatics, and machine learning and is the critical first step in the subsequent signal parameter estimation phase. The signal detection problem can be naturally posed in terms of the sample generalized eigenvalues. The sample generalized eigenvalues correspond to the eigenvalues of the matrix formed by "whitening" the signal-plus-noise sample covariance matrix with the noise-only sample covariance matrix. In this paper, we prove a fundamental asymptotic limit of sample generalized eigenvalue-based detection of signals in arbitrarily colored noise when there are relatively few signal bearing and noise-only samples. Specifically, we show why when the (eigen) signal-to-noise ratio (SNR) is below a critical value, that is a simple function of n, m, and N, then reliable signal detection, in an asymptotic sense, is not possible. If, however, the eigen-SNR is above this critical value then a simple, new random matrix theory-based algorithm, which we present here, will reliably detect the signal even at SNRs close to the critical value. Numerical simulations highlight the accuracy of our analytical prediction, permit us to extend our heuristic definition of the *effective number of identifiable signals in colored noise* and display the dramatic improvement in performance relative to the classical estimator by Zhao *et al.* We discuss implications of our result for the detection of weak and/or closely spaced signals in sensor array processing, abrupt change detection in sensor networks, and clustering methodologies in machine learning.**

***Index Terms*—Multivariate F distribution, random matrices, sample covariance matrix, signal detection, Wishart distribution.**

I. Introduction

THE observation vector, in many signal processing applications, can be modeled as a superposition of a finite number of signals embedded in additive noise. The model order selection problem of inferring the number of signals present is the critical first step in the subsequent signal parameter estimation problem. We consider the class of estimators that determine the model order, i.e., the number of signals, in colored noise from the sample generalized eigenvalues of the signal-plus-noise sample covariance matrix and the noise-only sample covariance matrix pair. The sample generalized eigenvalues [1] precisely correspond to the eigenvalues of the matrix formed by "whitening" the signal-plus-noise sample covariance matrix with the noise-only sample covariance matrix (assuming that the number of noise-only samples is greater than the dimensionality of the system so that the noise-only sample covariance matrix is invertible).

Such estimators are used in settings where it is possible to find a portion of the data that contains only noise fields and does not contain any signal information. This is a realistic assumption for many practical applications such as evoked neuromagnetic experiments [2]–[4], geophysical experiments that employ a "thumper" or in underwater experiments with a wideband acoustic signal transducer where such a portion can be found in a data portion taken before a stimulus is applied. In applications such as radar or sonar where the signals of interest are narrowband and located in a known frequency band, snapshot vectors collected at a frequency just outside this band can be justified as having the same noise covariance characteristics assuming that we are in the stationary-process-long-observation-time (SPLOT) regime [5].

Our main objective in this paper is to shed new light on this age old problem of detecting signal in noise from finite samples using the sample eigenvalues alone [6]–[8]. We bring into sharp focus a fundamental statistical limit that explains precisely when and why, in high-dimensional, sample size limited settings underestimation of the model order is unavoidable. This is in contrast to works in the literature that use simulations, as in [9], to highlight the chronically reported symptom of model order estimators underestimating the number of signals without providing insight into whether a fundamental limit of detection is being encountered.

In recent work [10], we examined this problem in the white noise scenario. The main contribution of this paper is the extension of the underlying idea to the arbitrary (or colored) noise scenario. Extending the result requires us to perform careful mathematical analysis to extend the results in the literature on spiked Wishart random matrices [11]–[17] to the setting where the covariance matrix (corresponding here to the inverse of the noise-only covariance matrix) is also random. What emerges

Manuscript received February 24, 2009; revised October 31, 2009; accepted November 11, 2009. First published April 15, 2010; current version published May 14, 2010. This work was supported in part by the Office of Naval Research Special Postdoctoral Award under Grant N00014-07-1-0269, in part by the U.S. Army Research Office under Grant W911NF-05-1-0244, and in part by the National Science Foundation under Agreement DMS-0112069. Portions of this work were presented at the 2007 Asilomar Conference on Signals and Systems. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Are Hjørungnes.

R. R. Nadakuditi is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor MI 48109 USA (e-mail: rajnrao@umich.edu).

J. W. Silverstein is with the Department of Mathematics, North Carolina State University, Raleigh, NC 27695 USA (e-mail: jack@unity.ncsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2009.2038310

is an appealing generalization of in [10], which allows us to mathematically justify our analogous definition of the *effective number of identifiable signals in colored noise* as the number of the generalized eigenvalues of the population (true) signal-plus-noise covariance matrix and noise-only covariance matrix pair that are greater than a (deterministic) threshold that is a simple function of the number of signal-plus-noise samples, noise-only samples and the dimensionality of the system. Analogous to the white noise case, increasing the dimensionality of the system, by say adding more sensors, raises the detectability threshold so that the effective number of identifiable signals might actually decrease.

An additional contribution of this paper is the development of a simple, new, optimal algorithm for estimating the number of signals based on the recent work of Johnstone [18] and insights on the behavior of spiked Wishart matrices from the work of Baik, *et al.* [11] and El Karoui [14]. Numerical results are used to illustrate the performance of the estimator around the detectability threshold alluded to earlier. Specifically, we observe that if the eigen-SNR of a signal is above a critical value then reliable detection using the new algorithm is possible. Conversely, if the eigen-SNR is below the critical value then the algorithm is unable to distinguish the signal from noise. When we comparing this new algorithm with the classical (ZKB) algorithm by Zhao *et al.* [8] we are able to reliably detect signals that are 8–9 dB lower than the ZKB algorithm—this is to expected given the discussion in Johnstone [15, Sec. 2.4] about the inherent biases in classical small-system-size-large-sample-size asymptotics (that the ZKB algorithm employs) that the new large-system-size-relatively-large-sample-size asymptotics accounts for (as in [15]).

The paper is organized as follows. We formulate the problem in Section II and carefully analyze the properties and the phase transition of the extreme sample (generalized) eigenvalues in the signal-free and signal-bearing setting in Section III and Section IV, respectively. Section V discusses the implications of this phase transition for applications such as array processing, sensor networks, and machine learning. A new algorithm for detecting the number of signals is presented in Section VI. Concluding remarks are offered in Section VII. The Appendix contains that mathematical proofs.

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, the snapshot vector has a (real or complex) multivariate normal distribution, i.e., $\mathbf{x}_i \sim \mathcal{N}_n(0, \mathbf{R})$ and the $\mathbf{x}_i$'s are mutually independent. The snapshot vectors are modeled as

$$\mathbf{x}_i = \mathbf{A}\mathbf{s}_i + \mathbf{z}_i \quad \text{for } i = 1, \ldots, m \tag{1}$$

where $\mathbf{z}_i \sim \mathcal{N}_n(0, \Sigma)$, denotes an n-dimensional (real or complex) Gaussian noise vector where the noise covariance Σ may be known or unknown, $\mathbf{s}_i \sim \mathcal{N}_k(\mathbf{0}, \mathbf{R}_s)$ denotes a k-dimensional (real or complex) Gaussian signal vector with covariance $\mathbf{R}_s$, and $\mathbf{A}$ is a $n \times k$ unknown nonrandom matrix. Since the signal and noise vectors are independent of each other, the covariance matrix of $\mathbf{x}_i$ can hence be decomposed a

$$\mathbf{R} = \mathbf{\Psi} + \mathbf{\Sigma} \tag{2}$$

where

$$\mathbf{\Psi} = \mathbf{A}\mathbf{R}_s\mathbf{A}' \tag{3}$$

with $'$ denoting the complex conjugate or real 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 the noise covariance matrix $\mathbf{\Sigma}$ were known *a priori* and was nonsingular, a "noise whitening" transformation may be applied to the snapshot vector $\mathbf{x}_i$ to obtain the vector

$$\widetilde{\mathbf{x}}_i = \mathbf{\Sigma}^{-1/2}\mathbf{x}_i \tag{4}$$

which will also be normally distributed with covariance

$$\mathbf{R}_{\mathbf{\Sigma}} := \mathbf{\Sigma}^{-1/2}\mathbf{R}\mathbf{\Sigma}^{-1/2} = \mathbf{\Sigma}^{-1/2}\mathbf{\Psi}\mathbf{\Sigma}^{-1/2} + \mathbf{I}. \tag{5}$$

Here, $\mathbf{\Sigma}^{1/2}$ is the Hermitian nonnegative definite square root of $\mathbf{\Sigma}$. Denote the eigenvalues of $\mathbf{R}_{\mathbf{\Sigma}}$ by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Recalling the formulation of the generalized eigenvalue problem [1Sec. 8.7], we note that the eigenvalues of $\mathbf{R}_{\mathbf{\Sigma}}$ are exactly the generalized eigenvalues of the regular matrix pair $(\mathbf{R}, \mathbf{\Sigma})$. Then, assuming that the rank of $\mathbf{\Sigma}^{-1}\mathbf{\Psi}$ is also k, it follows that the smallest $n-k$ eigenvalues of $\mathbf{R}_{\mathbf{\Sigma}}$ or, equivalently, the generalized eigenvalues of the matrix pair $(\mathbf{R}, \mathbf{\Sigma})$), are all equal to 1 so that

$$\lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_n = \lambda = 1 \tag{6}$$

while the remaining k eigenvalues of $\mathbf{R}_{\mathbf{\Sigma}}$ will be strictly greater than one.

Thus, if the true signal-plus-noise covariance matrix $\mathbf{R}$ and the noise-only covariance matrix $\mathbf{\Sigma}$ were known *a priori*, the number of signals k could be trivially determined from the multiplicity of the eigenvalues of $\mathbf{R}_{\mathbf{\Sigma}}$ equalling one.

The problem in practice is that the signal-plus-noise and the noise covariance matrices $\mathbf{R}$ are unknown so that such a straightforward algorithm cannot be used. Instead we have an estimate of the signal-plus-covariance matrix obtained as

$$\widehat{\mathbf{R}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i\mathbf{x}_i' \tag{7}$$

and an estimate of the noise-only sample covariance matrix obtained as

$$\widehat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{z}_j\mathbf{z}_j' \tag{8}$$

TABLE I
ZKB ALGORITHM [8]

ZKB Algorithm
Input: Eigenvalues $\widehat{\lambda}_j$ for $j=1,\ldots,n$ of $\widehat{\mathbf{R}}_{\boldsymbol{\Sigma}}$, penalty function C_{m+N}
1. Initialization: Compute $\kappa_{\widehat{\lambda}} = \{\#\widehat{\lambda}_j > 1\}$,
2. Set q = 0
3. Compute $L(q, C_{m+N}) = \sum_{j=1+\min(q,\kappa_{\widehat{\lambda}})}^{n} \frac{n}{2}\left[\frac{m}{m+N}\log\widehat{\lambda}_j - \log\left(\frac{N}{m+N} + \frac{m}{m+N}\widehat{\lambda}_j\right)\right]$
4. Compute $I(q, C_{m+N}) = L(q, C_{m+N}) + \frac{C_{m+N}}{2}q(2n-q+1)$
5. Increment q
6. If $q < n$, then go to step 2.
7. Return $\widehat{k} = \min_q I(q, C_{m+N})$

where $\mathbf{x}_i$ for $i = 1,\ldots,m$ are (possibly) signal-bearing snapshots and $\mathbf{z}_j$ for $j = 1,\ldots,N$ are independent noise-only snapshots. We assume here that the number of noise-only snapshots exceeds the dimensionality of the system, i.e., $N > n+1$, so that the noise-only sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$, which has the Wishart distribution [19], is non-singular and hence invertible with probability 1 [20, Ch. 3, p. 97], [21, Ch. 7.7, pp. 272–276]. Following (5), we then form the matrix

$$\widehat{\mathbf{R}}_{\widehat{\Sigma}} = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\mathbf{R}}\widehat{\boldsymbol{\Sigma}}^{-1/2} \tag{9}$$

and compute its eigen-decomposition to obtain the eigenvalues of $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$, which we denote by $\hat{\lambda_1} \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_n$. We note, once again, that the eigenvalues of $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ are simply the generalized eigenvalues of the regular matrix pair $(\widehat{\mathbf{R}}, \widehat{\boldsymbol{\Sigma}})$. Note that whenever $N < n$, the signal-plus-noise sample covariance matrix $\widehat{\mathbf{R}}$ will have rank at most N, so that the $n-N$ generalized eigenvalues will equal zero, i.e., $\hat{\lambda}_{N+1} = \hat{\lambda}_{N+2} = \ldots = \hat{\lambda}_n = 0$.

In this paper, we are interested in algorithms that infer the number of signals from the eigenvalues of $\widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Sigma}}}$ or $\widehat{\mathbf{R}}_{\boldsymbol{\Sigma}} := \boldsymbol{\Sigma}^{-1/2}\widehat{\mathbf{R}}\boldsymbol{\Sigma}^{-1/2}$ alone. Such algorithms are widely used in practice and arise naturally from classical multivariate statistical theory [18] where the matrix $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ is referred to as the multivariate F matrix [20], [22]. The information-theoretic approach to model order estimation, first introduced in [6], was extended to the colored noise setting by Zhao *et al.* in [8] who prove large sample consistency of estimator (see Table I). Tam and Wu [23] performed a rate of convergence analysis of the ZKB algorithm for different choices of the penalty function but provide no insight into the performance in the large system, finite sample setting. A larger, more fundamental question that remained unanswered is whether there are any fundamental statistical limits to detection.

Consequently, research has focused refining the performance of eigenvalue based methods in the finite sample setting. Zhu *et al.* [24] improve the performance of their eigenvalue estimator by assuming a model for the noise covariance matrix. Stoica and Cedervall [25] improve the performance of their estimator in two reasonable settings: one, where it is reasonable to assume that the noise covariance matrix is block diagonal or banded and two, where the temporal correlation of the noise has a shorter length than the signals. Other techniques exploit other characteristics of the signal or noise to effectively reduce the dimensionality of the signal subspace and improve model order estimation given finite samples. See for, e.g. [26], [27] and the references in [10].

We tackle the statistical limit issue head on in this paper by employing tools from random matrix theory [28]. What we find that is that there is a fundamental statistical limit of detection, which allows one to predict how many samples are "good enough" [3 p. 846]. Consequently, the phase transition that the ZKB algorithm encounters (see Fig. 1) is unavoidable but it turns out that the ZKB algorithm is far from that limit.

III. ANALYZING THE SIGNAL-FREE CASE

We begin our investigation with a careful analysis of the eigenvalues of sample covariance matrices. Since our analysis extends beyond the setting where the snapshot vectors are modeled as Gaussian (as in Section II), we shall employ different notation to avoid any confusion. This allows us to treat the specific problem in Section II as a special case and to extend our results for a variety of related generalize eigenvalue based detection scenarios.

A. Empirical Distribution Function and Its Transform

Let for $i, j = 1, 2, \ldots$, X_{ij} be a collection of complex valued independent and identically distributed (i.i.d.) random variables with $\mathsf{E}X_{1\,1} = 0$ and $\mathsf{E}|X_{1\,1}|^2 = 1$. For positive integers n and m let $\mathbf{X}_n = (X_{ij})$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. Assume for each n, $\mathbf{T}_n$ is an $n \times n$ Hermitian nonnegative definite matrix. The matrix $\mathbf{B}_n \equiv (1/m)\mathbf{T}_n^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_n^{1/2}$, where $\mathbf{T}_n^{1/2}$ is any Hermitian square root of $\mathbf{T}_n$, can be viewed as a sample covariance matrix, formed from m samples of the random vector $\mathbf{T}_n^{1/2}\mathbf{X}_{\cdot 1}$ with $\mathbf{X}_{\cdot 1}$ denoting the first column of $\mathbf{X}_n$, which has $\mathbf{T}_n$ for its population covariance matrix. When n and m are both large and on the same order of magnitude, $\mathbf{B}_n$ will not be near $\mathbf{T}_n$, due to an insufficient number of samples required for such a large dimensional random vector. However, there exist results on the eigenvalues of $\mathbf{B}_n$. They are limit theorems as $n \to \infty$ with $m = m(n)$ and $c_n \equiv n/m \to c$, which provide information on the eigenvalues of $\mathbf{T}_n$. One result [29] is on the *empirical distribution function* (e.d.f.), F^{B_n}, of the eigenvalues of $\mathbf{B}_n$, which throughout the paper, is defined for any Hermitian $n \times n$ matrix $\mathbf{A}$ as

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

The limit theorem is expressed in terms of the *Stieltjes transform* (s.t.) of the limiting e.d.f. of the F^{B_n}'s, where for any distribution function (d.f.) G its s.t. m_G is defined to be

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}.$$

There exists a one-to-one correspondence between the distribution functions (d.f.'s) and their s.t.'s ([30]).

The limit theorem allows the $\mathbf{T}_n$ to be random, but independent of $\mathbf{X}_n$, and assuming as $n \to \infty$, the convergence of F^{T_n} to a nonrandom proper probability distribution function H_n, i.e., $H_n \equiv F^{T_n} \xrightarrow{a.s.} H$. The theorem states that with probability one, as $n \to \infty$, $F^{B_n} \to_D F$, where F is nonrandom, with s.t. $m = m_F(z)$, $z \in \mathbb{C}^+$ satisfying the equation

$$m = \int \frac{1}{t(1-c-czm) - z} dH(t) \tag{11}$$

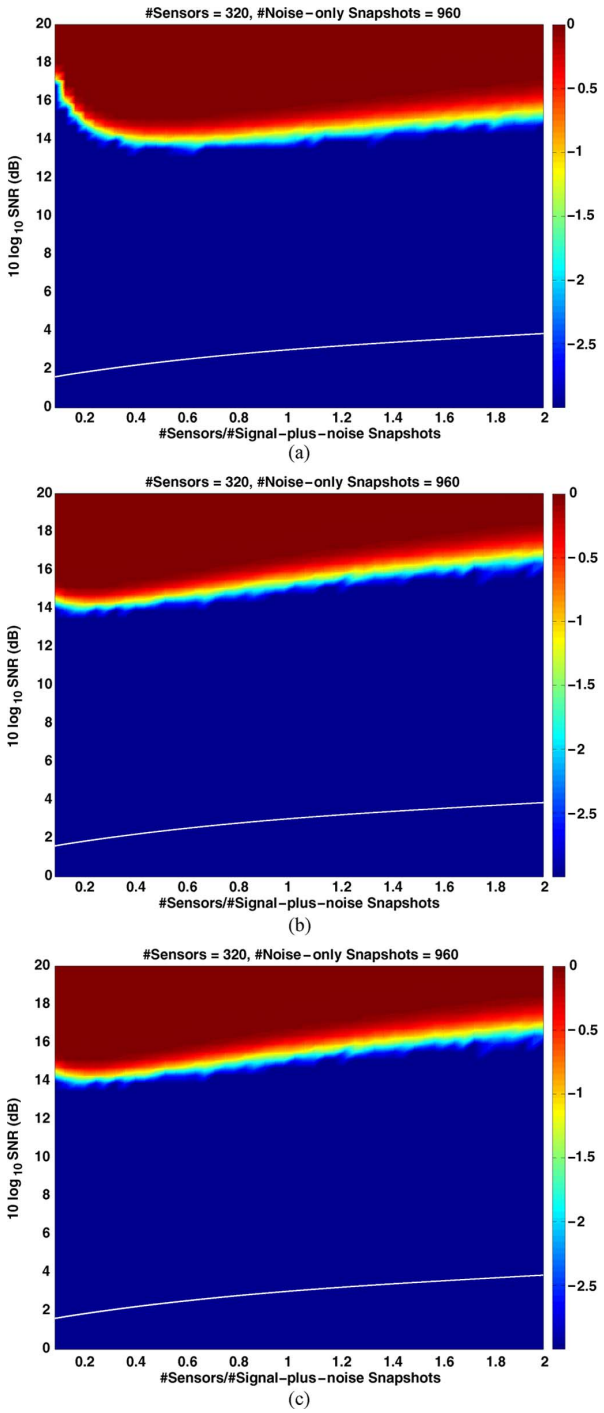


Fig. 1. Heat map of the log probability of signal detection using the ZKB algorithm described in Table I, for different choices of the penalty function C_{m+N}, in (eigen) SNR versus number of sensors to number of signal-plus-noise snapshots phase space. We set $\mathbf{\Sigma} = \mathbf{I}$, $\mathbf{R} = \mathrm{diag}(\lambda_1 = 1 + \mathrm{SNR}, 1, \ldots, 1)$ and evaluate $\mathrm{Prob}(\hat{k} = 1)$ over 1000 Monte-Carlo trials and a grid of 50 equally spaced points in the 0 to 20 dB (eigen) SNR range and 50 equally spaced points in the $c_1 = n/m$ space by setting $m = n/c_1$. The values of the colormap at each of the 50 × 50 faces were interpolated across each line segment and face to obtain the above plot. The superimposed solid white line demarcates the theoretically predicted threshold and shows the suboptimality of the ZKB algorithm. The new algorithm gets within 0.5 dB of the threshold (see Fig. 4) and is thus able to reliably detect 8–9 dB weaker signals. (a) $C_{m+N} = 0.65\sqrt{m+N}/\log(m+N)$; (b) $C_{m+N} = 0.5\log(m+N)$; (c) $C_{m+N} = 0.5\log((m+N)/2)$.

which is unique in the set $\{m \in \mathbb{C} : -(1-c)z^{-1} + cm \in \mathbb{C}^+\}$.

It is more convenient to work with the e.d.f. of the $m \times m$ matrix $\underline{\mathbf{B}}_n = (1/m)\mathbf{X}_n'\mathbf{T}_n\mathbf{X}_n$, whose eigenvalues differ from those of $\mathbf{B}_n$ by $|n-m|$ zero eigenvalues. Indeed, with $F^{c,H}$ denoting the limiting e.d.f. of the eigenvalues of $\underline{\mathbf{B}}_n$, we have, with $\mathbb{I}_A$ denoting the indicator function on the set A the exact relationship

$$\begin{aligned} F^{(1/m)X_n^{*}T_nX_n}(x) &= (1-c_n))\, I_{[0,\infty)}(x) + c_n F^{B_n}(x) \\ &\rightarrow_D = (1-c))\, I_{[0,\infty)}(x) + cF(x) \equiv F^{c,H}(x) \end{aligned}$$

almost surely, implying

$$m_{F^{c,H}}(z) = -(1-c)z^{-1} + cm_F(z). \tag{12}$$

Upon substituting $m_{F^{c,H}}$ into (11) we find that for $z \in \mathbb{C}^+$ $m = m_{F^{c,H}}(z)$ solves the equation

$$z = -\frac{1}{m} + c\int \frac{\lambda}{1+\lambda m} dH(\lambda) \tag{13}$$

and is unique in $\mathbb{C}^+$. Thus we have an explicit inverse for $m_{F^{c,H}}$.

B. Recovering Eigenvalue Support From the Integral Transform

Qualitative properties of $F^{c,H}$ have been obtained in [30], most notably the fact that on $(0,\infty)$ $F^{c,H}$ has a continuous derivative. The paper [30] also shows how intervals outside the support of $F^{c,H}$ can be determined from the graph of (13) for $m \in \mathbb{R}$.

Let S_G denote the support of the d.f. G, S_G' its complement, and define $x_{c,H} = x_{c,H}(m)$ to be (13) with $m \in \mathbb{R}$. Intuitively, on $S'_{F^{c,H}}$, by definition of the Stieltjes transform, $m_{F^{c,H}}$ is well defined and increasing. Therefore, it is invertible on each interval in $S'_{F^{c,H}}$, its inverse, namely $x_{c,H}$, is also increasing. The details are herewith stated more formally.

Lemma 3.1 (Theorems 4.1, 4.2 of [30]): If $x \in S'_{F^{c,H}}$, then $m = m_{F^{c,H}}$ satisfies (1) $m \in \mathbb{R} \setminus \{0\}$, (2) $-m^{-1} \in S'_H$, and (3) $(d/dm)x_{c,H}(m) > 0$. Conversely, if m satisfies (1)–(3), then $x = x_{c,H}(m) \in S'_{F^{c,H}}$.

In simple terms $S'_{F^{c,H}}$ is comprised of the range of values where $x_{c,H}$ is increasing.

Another result which will be needed later is the following.

Lemma 3.2 (Theorem 4.3 of [30]): Suppose each m contained in the interval $[m_1, m_2]$ satisfies (1) and (2) of Lemma 3.1, and $(d/dm)x_{c,H}(m_i) \geq 0$ for $i = 1, 2$. Then $(d/dm)x_{c,H}(m) > 0$ for all $m \in (m_1, m_2)$.

Since, by definition of the e.d.f in (10), the convergence in distribution of F^{B_n} only addresses how *proportions of eigenvalues behave*, understanding the possible appearance or non-appearance of individual eigenvalues in $S'_{F^{c,H}}$ requires further work.

The extreme eigenvalues when $\mathbf{T}_n = \mathbf{I}$ has been answered by Yin, Bai, and Krishnaiah in [31], and Bai and Yin in [32], respectively, under the additional assumption $\mathsf{E}|\mathbf{X}_{1\,1}|^4 < \infty$, the largest eigenvalue and $\min(n,m)$th largest eigenvalue of

$(1/m)\mathbf{X}_n\mathbf{X}_n^*$ converge a.s. to $(1+\sqrt{c})^2$ and $(1-\sqrt{c})^2$, respectively, matching the support, $[(1-\sqrt{c})^2,(1+\sqrt{c})^2]$ of F on $(0,\infty)$. More on this case will be given later.

For general $\mathbf{T}_n$, restricted to being bounded in spectral norm, the nonappearance of eigenvalues in $S'_{F^{c,H}}$ has been proven by Bai and Silverstein in [33]. Moreover, the separation of eigenvalues across intervals in $S'_{F^{c,H}}$, mirrors exactly the separation of eigenvalues over corresponding intervals in S'_H [34]. The results are summarized below.

Theorem 3.1 ([33], [34]) Assume additionally $\mathsf{E}|\mathbf{X}_{1\,1}|^4<\infty$ and the $\mathbf{T}_n$ are nonrandom and are bounded in spectral norm for all n. Let F^{c_n,H_n} denote the "limiting" e.d.f. associated with $(1/m)\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$, in other words, F^{c_n,H_n} is the d.f. having s.t. with inverse (13), where c,H are replaced by c_n,H_n. This provides a mechanism for accounting for the finite matrix size behavior associated in how c_n and H_n converge to c and H, respectively.

Assume the following condition:

- $(*)$ Interval $[a,b]$ with $a>0$ lies in an open interval outside the support of F^{c_n,H_n} for all large n.

Then $\mathsf{P}(\text{no eigenvalue of } \mathbf{B}_n \text{ appears in } [a,b] \text{ for all large } n)=1$

For $n\times n$ Hermitian non-negative definite matrix $\mathbf{A}$, let λ_k^A denote the kth largest eigenvalue of A. For notational convenience, define $\lambda_0^A=\infty$ and $\lambda_{n+1}^A=0$.

1) If $c(1-H(0))>1$, then x_0, the smallest value in the support of $F^{c,H}$, is positive, and with probability 1, $\lambda_m^{B_n}\to x^0$ as $n\to\infty$.
2) If $c(1-H(0))\le 1$, or $c(1-H(0))>1$ but $[a,b]$ is not contained in $[0,x_0]$, then $m_{F^{c,H}}(b)<0$, and for all n large there is an index i_n for which

$$\lambda_{i_n}^{T_n}>-1/m_{F^{c,H}}(b) \quad \text{and} \quad \lambda_{i_n+1}^{T_n}<-1/m_{F^{c,H}}(a). \tag{14}$$

Then $\mathsf{P}(\lambda_{i_n}^{B_n}>b \text{ and } \lambda_{i_n+1}^{B_n}<a \text{ for all large } n)=1$.

The behavior of the extreme eigenvalues of $(1/m)\mathbf{X}_n\mathbf{X}_n^*$ leads to the following corollary of Theorem 3.1.

Corollary 3.2: If $\lambda_1^{T_n}$ converges to the largest number in the support of H, then $\lambda_1^{B_n}$ converges a.s. to the largest number in the support of F. If $\lambda_n^{T_n}$ converges to the smallest number in the support of H, then $c\le 1$ $(c>1)$ implies $\lambda_n^{B_n}$ $(\lambda_n^{(1/m)X_n^*T_nX_n})$ converges a.s. to the smallest number in the support of F $(F^{c,H})$.

In Theorem 3.1, Case 1) applies when $n>m$, whereby the rank of $\mathbf{B}_n$ would be at most m, the conclusion asserting, that with probability 1, for all n large, the rank is equal to m. From Lemma 3.1, Case 2) of Theorem 3.1 covers all intervals in $S'_{F^{c,H}}$ on $(0,\infty)$ resulting from intervals on $(-\infty,0)$ where $x_{c,H}$ is increasing. For all n large x_{c_n,H_n} is increasing on $[m_{F^{c_n,H_n}}(a),m_{F^{c_n,H_n}}(b)]$, which, from inspecting the vertical asymptotes of x_{c_n,H_n} and Lemma 3.1, must be due to the existence of $\lambda_{i_n}^{T_n}$, $\lambda_{i_n+1}^{T_n}$ satisfying (14).

We now extend Theorem 3.1 to random $\mathbf{T}_n$, independent of $\{\mathbf{X}_{ij}:i,j\ge 1\}$ with the aid of Tonelli's Theorem [35, pp. 234], provided the condition $(*)$ on $[a,b]$ is strengthened to the following:

- $(**)$ With probability 1 for all n large $[a,b]$ (nonrandom) lies in an open interval outside the support of F^{c_n,H_n}.

Proof: See Appendix. Note the novelty of this Theorem lies in the fact that we allow $\mathbf{T}_n$ to be random unlike other results in the literature [11]–[17]. ■

Consider now case 2) of Theorem 3.1 in terms of the corresponding interval outside the support of H and the H_n's. Let $m_a=m_{F^{c,H}}(a)$, $m_b=m_{F^{c,H}}(b)$. By Lemma 3.1 and condition $(*)$, we have the existence of an $\epsilon>0$ such that $0\notin[m_a-\epsilon,m_b+\epsilon]$, and for all n large

$$\frac{d}{dm}x_{c_n,H_n}(m)=\frac{1}{m^2}\left(1-c_n\int\frac{(\lambda m)^2}{(1+\lambda m)^2}dH_n(\lambda)\right)>0 \tag{15}$$

$m\in[m_a-\epsilon,m_b+\epsilon]$. Let $t_a=-1/m_a$, $t_b=-1/m_b$. Then by Lemma 3.1 we have the existence of an $\epsilon'>0$ for which $t_a-\epsilon'>0$ and $[t_a-\epsilon',t_b+\epsilon']\subset S'_{H_n}$ for all n large. Moreover, by (15) we have for all n large

$$c_n\int\lambda^2(\lambda-t)^{-2}dH_n(\lambda)<1, \quad t\in[t_a-\epsilon',t_b+\epsilon']. \tag{16}$$

Necessarily, $\lambda_{i_n}^{T_n}>t_b+\epsilon'$ and $\lambda_{i_n+1}^{T_n}<t_a-\epsilon'$.

Notice the steps can be completely reversed, that is, beginning with an interval $[t_a,t_b]$, with $t_a>0$, lying in an open interval in S'_{H_n} for all n large and satisfying (16) for some $\epsilon'>0$, will yield $[a,b]$, with $a=x_{c,H}(-1/t_a)$, $b=x_{c,H}(-1/t_b)$, satisfying condition $(*)$. Case 2) applies, since $[a,b]$ is within the range of $x_{c,H}(m)$ for $m<0$. If $c(1-H(0))>1$, then we would have $a>x_0$.

C. Behavior of the Eigenvalues in the Signal-Free Case

Let $\mathbf{Y}_{ij}$ be another collection of i.i.d. random variables (not necessarily having the same distribution as the $\mathbf{X}_{ij}$'s), with $\mathsf{E}\mathbf{Y}_{1\,1}=0$, $\mathsf{E}|\mathbf{Y}_{1\,1}|=1$, $\mathsf{E}|\mathbf{Y}_{1\,1}|^4<\infty$, and independent of the $\mathbf{X}_{ij}$'s.

We form the $n\times N$ matrix $\mathbf{Y}_n=(Y_{ij})$, $i=1,2,\ldots,n$, $j=1,2,\ldots,N$ with $N=N(n)$, $n<N$, and $c_n^1\equiv n/N\to c_1\in(0,1)$ as $n\to\infty$.

Let now $\mathbf{T}_n=((1/N)\mathbf{Y}_n\mathbf{Y}_n^*)^{-1}$, whenever the inverse exists.

From Bai and Yin's work [32] we know that with probability 1, for all n large, $\mathbf{T}_n$ exists with $\lambda_1^{T_n}\to(1-\sqrt{c_1})^{-2}$. Whenever $\lambda_n^{(1/N)\mathbf{Y}_n\mathbf{Y}_n^*}=0$ define $\mathbf{T}_n$ to be $\mathbf{I}$. The matrix $\mathbf{T}_n(1/N)\mathbf{X}_n\mathbf{X}_n^*$, typically called a multivariate F matrix, has the same eigenvalues as $\mathbf{B}_n$. Its limiting e.d.f. has density on $(0,\infty)$ given by

$$dF(x)=\frac{(1-c_1)\sqrt{(x-b_1)(b_2-x)}}{2\pi x(xc_1+c)}\mathbb{I}_{[b_1,b_2]}(x)dx + \max\left(0,\left(1-\frac{1}{c}\right)\right)\delta(x) \tag{17}$$

where

$$b_1=\left(\frac{1-\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2 \tag{18}$$

$$b_2=\left(\frac{1+\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2 \tag{19}$$

and $\delta(x)$ is the Dirac delta function. When $c \in (0,1]$, there is no mass at 0, whereas for $c > 1$ F has mass $(1-(1/c))$ at 0 [22].

We note that (17) is a probabilistic description of the eigenvalues of $\widehat{R}_{\widehat{\Sigma}}$ in the signal-free case, with $\mathbf{T}_n = \widehat{\mathbf{\Sigma}}^{-1}$, $\widehat{\mathbf{R}} = (1/m)(\mathbf{X}\mathbf{X}^T)$ and $k = 0$.

IV. Analyzing the Signal-Bearing Case

With these equivalences in mind, we now turn to the signal-bearing setting.

A. Phase Transition of Spiked Eigenvalues

Suppose now the $\mathbf{T}_n$'s are altered, where a finite number of eigenvalues are interspersed between the previously adjacent eigenvalues $\lambda_{i_n+1}^{T_n}$ and $\lambda_{i_n}^{T_n}$. It is clear that the limiting F will remain unchanged. However, the graph of x_{c_n,H_n} on $(-1/\lambda_{i_n+1}^{T_n}, -1/\lambda_{i_n}^{T_n})$ will now contain vertical asymptotes. If the graph remains increasing on two intervals for all n large, each one between successive asymptotes, then because of Theorem 3.1, with probability one, eigenvalues of the new $\mathbf{B}_n$ will appear in $S'_{F^{c,H}}$ for all n large.

Theorem 4.1, stated next, shows this will happen when a "sprinkled", or "spiked" eigenvalue lies in (t_a, t_b).

Theorem 4.1: Assume in addition to the assumptions in Theorem 3.1 on the $\mathbf{X}_{ij}$ and $\mathbf{T}_n$:

1) There are $\ell = o(n)$ positive eigenvalues of $\mathbf{T}_n$ all converging uniformly to t', a positive number. Denote by $\hat{H}_n$ the e.d.f. of the $n-\ell$ other eigenvalues of $\mathbf{T}_n$.
2) There exists positive $t_a < t_b$ contained in an interval (α, β) with $\alpha > 0$ which is outside the support of $\hat{H}_n$ for all large n, such that for these n
$$c_n \int \frac{\lambda^2}{(\lambda - t)^2} d\hat{H}_n(\lambda) \leq 1$$
for $t = t_a, t_b$.
3) $t' \in (t_a, t_b)$.

Suppose $\lambda_{i_n}^{T_n}, \ldots, \lambda_{i_n+\ell-1}{}^{T_n}$ are the eigenvalues stated in a). Then, with probability one
$$\lim_{n\to\infty} \lambda_{i_n}^{B_n} = \cdots = \lim_{n\to\infty} \lambda_{i_n+\ell-1}^{B_n} = t'\left(1 + c\int \frac{\lambda}{t'-\lambda} dH(\lambda)\right). \tag{20}$$

Proof: See the Appendix. ∎

Note that Theorem 4.1 allows the number of spiked eigenvalues to grow with n, provided it remains $o(n)$. Theorem 4.2, stated next, provides a converse, in the sense that any isolated eigenvalue of B_n must be due to a spiked eigenvalue, the absence of which corresponds to case 2) of Theorem 3.1.

Theorem 4.2: Assume, besides the assumptions in Theorem 3.1, there is an eigenvalue of $\mathbf{B}_n$ which converges in probability to a nonrandom positive number, $\lambda' \in S'_F$. Let interval $[a,b] \in S'_F$, with $a > 0$, be such that $\lambda' \in (a,b)$, and let $t_a = -1/m_{c,H}(a)$, $t' = -1/m_{c,H}(\lambda')$, $t_b = -1/m_{c,H}(b)$ (finite by Lemma 3.1). Then $0 < t_a < t' < t_b$, implying 3) of Theorem 4.1. Let $\ell = \ell(n)$ denote the number of eigenvalues of $\mathbf{T}_n$ contained in $[t_a, t_b]$ and let $\hat{H}_n$ denote the e.d.f. of the other $n-\ell$ eigenvalues of $\mathbf{T}_n$. Then $\ell = o(n)$ and 2) of Theorem 4.1 is true. If ℓ remains bounded, then 1) of Theorem 4.1 also holds.

Proof: See the Appendix. ∎

B. Threshold Where Phase Transition Manifests

Theorem 4.1 proves almost sure convergence of the spiked eigenvalues that lie *in* (t_a, t_b).

Consider now t' lying on either side of the support of $\hat{H}$. Let $\hat{\lambda}_n^{\min}$ and $\hat{\lambda}_n^{\max}$ denote, respectively, the smallest and largest numbers in the support of $\hat{H}_n$ Notice that $g_n(t) \equiv c_n \int (\lambda^2/(\lambda - t)^2) d\hat{H}_n(t)$ is decreasing for $t > \hat{\lambda}_n^{\max}$, and if $\hat{\lambda}_n^{\min} > 0$, g_n is increasing on $(0, \hat{\lambda}_n^{\min})$. Therefore, if for all n large, $t' > \hat{\lambda}_n^{\max}$, it is necessary and sufficient to find a $t_a \in (\hat{\lambda}_n^{\max}, t')$ for which $g(t_a) \leq 1$ in order for (20) to hold.

Similarly, if for all n large $t' \in (0, \hat{\lambda}_n^{\min})$, then it is necessary and sufficient to find a $t_b \in (0, t')$ for which $g_n(t_b) \leq 1$ in order for (20) to hold. Notice if $c(1 - H(0)) > 1$ then $g_n(t) > 1$ for all $t \leq \hat{\lambda}_n^{\min}$ and all n large.

Let for d.f. G with bounded support, $\lambda_G^{\max}$ denote the largest number in S_G. If there is a $\tau > \lambda_H^{\max}$ for which $g(\tau) = c\int(\lambda^2/(\lambda - t)^2) dH(t) = 1$, and if $\limsup_n \hat{\lambda}_n^{\max} < \tau$, then τ can be used as a threshold for $t' \in (\limsup_n \hat{\lambda}_n^{\max}, \infty)$. Indeed, by the dominated convergence theorem, $\lim_{n\to\infty} g_n(t') = g(t')$. Therefore, if $t' > \tau$, conditions 2) and 3) of Theorem 4.1 hold, with $t_a = \tau$, and t_b any arbitrarily large number.

On the other hand, suppose $\lambda_{i_n}^{T_n}, \ldots, \lambda_{i_n+\ell-1}^{T_n}$, where ℓ remains bounded, are the eigenvalues of $\mathbf{T}_n$ approaching the interval $(\limsup_n \hat{\lambda}_n^{\max}, \tau]$. Then by Theorem 4.2, for any $\epsilon > 0$ with probability one, none of $\lambda_{i_n}^{B_n}, \ldots, \lambda_{i_n+\ell-1}^{B_n}$ can remain in $(\lambda_F^{\max} + \epsilon, \infty)$ with for all n large.

Also, since the largest $i_n + \ell - 1$ eigenvalues of $\mathbf{T}_n$ must be $o(n)$ (otherwise, H would have additional mass on $[\lambda_H^{\max}, \infty)$), $\lambda_{i_n}^{B_n}, \ldots, \lambda_{i_n+\ell-1}^{B_n}$ must all converge a.s. to $\lambda_F^{\max}$. Similar results can be obtained for the interval to the left of S_F. As in Theorem 3.1 Tonelli's Theorem can easily be applied to establish equivalent results when $\mathbf{T}_n$'s are random and independent of $\mathbf{X}$.

Thus, we have shown that the spiked eigenvalues will separate from the bulk only if t' is greater than some threshold τ which only depends on the limiting e.d.f. H and is the solution of the equation
$$g(\tau) = c\int \frac{\lambda^2}{(\lambda - \tau)^2} dH(\lambda) = 1.$$

C. Finite Signal-Bearing Case

We now apply the above result to the problem in Section II where we are interested in the effect on spikes on the right side of the support of the H_n, corresponding to the largest eigenvalues. Because of the corollary to Theorem 3.1, we know $\lambda_1^{B_n} \to b_2$ a.s. as $n \to \infty$. To compute the threshold we must first compute the function $g(t) = c\int(\lambda^2/(\lambda - t)^2) dH(\lambda)$. We will see that it is unnecessary to compute the limiting e.d.f. $H(\lambda)$ of $\mathbf{T}_n$. It suffices to know the limiting s.t. of $F^{(1/N)Y_nY_n^*}$.

Let H_1 denote the limiting e.d.f. of $F^{(1/N)Y_nY_n^*}$. We have
$$\begin{aligned} g(t) &= c\int \frac{1}{(\lambda(t - 1/\lambda))^2} dH_1(\lambda) \\ &= c\frac{1}{t^2}\int \frac{1}{(\lambda - \frac{1}{t})^2} dH_1(\lambda) = c\frac{1}{t^2}\frac{d}{dx} m_{H_1}(x)|_{x=(1/t)}. \end{aligned}$$

We use (13) to find $m_{F^{c_1,I_{[1,\infty)}}}$

$$\begin{aligned} z &= -m^{-1} + c_1(1+m)^{-1} \\ &\Leftrightarrow zm^2 + (z+1-c_1)m + 1 = 0 \\ &\Leftrightarrow m = \frac{-z-1+c_1 \pm \sqrt{(z+1-c_1)^2 - 4z}}{2z} \\ &= \frac{-z-1+c_1 \pm \sqrt{\left(z-(1-\sqrt{c_1})^2\right)\left(z-(1+\sqrt{c_1})^2\right)}}{2z}. \end{aligned}$$

(the sign depending on which branch of the square root is taken). From the identity in (12) we find that

$$m_{H_1}(z) = \frac{-z+1-c_1 \pm \sqrt{\left(z-(1-\sqrt{c_1})^2\right)\left(z-(1+\sqrt{c_1})^2\right)}}{2c_1 z}.$$

As mentioned earlier the support of H_1 is $[(1-\sqrt{c_1})^2, (1+\sqrt{c_1})^2]$. We need $g(t)$ for $t > (1-\sqrt{c_1})^{-2}$, so we need $m_{H_1}(x)$ for $x \in (0, (1-\sqrt{c_1})^2)$.

Since $0 \in S'_{H_1}$, $m_{H_1}(0)$ exists and is real, which dictates what sign is taken on $(0, (1-\sqrt{c_1})^2)$. We find that, on this interval

$$m_{H_1}(x) = \frac{-x+1-c_1 - \sqrt{\left(x-(1-\sqrt{c_1})^2\right)\left(x-(1+\sqrt{c_1})^2\right)}}{2c_1 x} \tag{21}$$

and using the fact that the discriminant equals $x^2 - 2x(1+c_1) + (1-c_1)^2$

$$\frac{d}{dx} m_{H_1}(x) = -\frac{1}{2c_1 x^2} \times \left((1-c_1) + \frac{x(1+c_1) - (1-c_1)^2}{\sqrt{\left(x-(1-\sqrt{c_1})^2\right)\left(x-(1+\sqrt{c_1})^2\right)}} \right).$$

We therefore find that for $t > (1-\sqrt{c_1})^{-2}$

$$g(t) = \frac{c}{2c_1} \times \left(-(1-c_1) + \frac{t(1-c_1)^2 - (1+c_1)}{\sqrt{\left(1-t(1-\sqrt{c_1})^2\right)\left(1-t(1+\sqrt{c_1})^2\right)}} \right).$$

We see that the equation $g(t) = 1$ leads to the quadratic equation in t: $(1-c_1)^2\alpha t^2 - 2(1+c_1)\alpha t + \alpha - c^2 = 0$, where $\alpha = c_1 + c - cc_1$, giving us

$$t = \frac{(1+c_1)\alpha + \sqrt{(1+c_1)^2\alpha^2 - (1-c_1)^2\alpha(\alpha - c^2)}}{(1-c_1)^2\alpha}$$

the positive sign in front of the square root being correct due to

$$\frac{1+c_1}{(1-c_1)^2} = \frac{1+c_1}{\left((1-\sqrt{c_1})(1+\sqrt{c_1})\right)^2} < \frac{1}{(1-\sqrt{c_1})^2}.$$

Reducing further we find the threshold, τ, to be

$$\begin{aligned} \tau &= \frac{(1+c_1)\alpha + \sqrt{\alpha}\sqrt{4\alpha - c_1 + (1-c_1)^2c^2}}{(1-c_1)^2\alpha} \\ &= \frac{(1+c_1)\alpha + \sqrt{\alpha}\,(2c_1 + c(1-c_1))}{(1-c_1)^2\alpha}. \end{aligned} \tag{22}$$

We now compute the right-hand side of (20). We have for $t' \geq \tau$, the expression in (23), shown in the equation at bottom of page.

A straightforward (but tedious) calculation will yield $\lambda(\tau) = b_2$. Using the results from the previous section, we have proved the following.

Theorem 4.3: Assume in addition to the assumptions in Theorem 3.1 on the $\mathbf{X}_{ij}$

1) the $\mathbf{T}_n$, possibly random, are independent of the $\mathbf{X}_{ij}$, with $F^{T_n} \to_D H$, a.s. as $n \to \infty$, H being the limiting e.d.f. of $F^{((1/N)Y_nY_n^*)^{-1}}$, defined above.
2) Almost surely, there are ℓ (remaining finite for each realization) eigenvalues of $\mathbf{T}_n$ converging to nonrandom $t' > (1-\sqrt{c_1})^{-2}$, as $n \to \infty$. Denote by $\hat{H}_n$ the e.d.f. of the $n-\ell$ other eigenvalues of $\mathbf{T}_n$.
3) With $\hat{\lambda}_n^{\max}$ defined to be the largest number in the support of $\hat{H}_n$, with probability one, $\limsup_n \hat{\lambda}_n^{\max} < \tau$ the threshold defined in (22).

Suppose $\lambda_{i_n}^{T_n}, \ldots, \lambda_{i_n+\ell-1}^{T_n}$ are the eigenvalues stated in 2) of Theorem 4.1. Then, with the function $\lambda(\cdot)$ defined in (23), with probability one

$$\lim_{n\to\infty} \lambda_{i_n}^{B_n} = \cdots = \lim_{n\to\infty} \lambda_{i_n+\ell-1}^{B_n},$$

and

$$\lim_{n\to\infty} \lambda_{i_n}^{B_n} = \begin{cases} \lambda(t'), & \text{if } t' > \tau \\ b_2, & \text{if } t' \in \left(\limsup_n \hat{\lambda}_n^{\max}, \tau\right]. \end{cases} \tag{24}$$

Note: From Theorem 4.1, when $t' > \tau$ the result can allow $\ell = o(n)$. Recall that we had $\mathbf{B}_n = \widehat{\mathbf{R}}_{\widehat{\mathbf{\Sigma}}}$,

$$\begin{aligned} & t'\left(1 + c\int \lambda \frac{1}{t'-\lambda} dH(\lambda)\right) \\ &= t'\left(1 + c\frac{1}{t'} m_{H_1}(1/t')\right) \\ &= \frac{t'\,(2c_1 + c(1-c_1)) - c - c\sqrt{\left(1-t'(1-\sqrt{c_1})^2\right)\left(1-t'(1+\sqrt{c_1})^2\right)}}{2c_1} \\ &\equiv \lambda(t'). \end{aligned} \tag{23}$$

$\mathbf{T}_n = ((1/N)Y_nY_n^*)^{-1} = \widehat{\mathbf{\Sigma}}^{-1}$, $\widehat{\mathbf{R}} = (1/m)(\mathbf{X}\mathbf{X}')$ and $k = l$ is the number of signals. In Theorem 4.3, we have, from [11], [12] that

$$t' = \frac{1}{\lambda_j\left(1+\frac{c_1}{\lambda_j-1}\right)}$$

so that solving $t' > \tau$ gives us an expression for the threshold $\mathrm{T}(c, c_1)$ that the "signal" eigenvalue of $\mathbf{R}_\Sigma$ must exceed to be detectable. This expression is

$$\mathrm{T}(c, c_1) = \frac{1+\tau-\tau c_1+\sqrt{(1+\tau-\tau c_1)^2-4\tau}}{2}. \quad (25)$$

Note that the generality of Theorem 4.1 allows us to analyze the setting where $\widehat{\mathbf{\Sigma}}$ is obtained from noise-only snapshots that are temporally correlated with some covariance. Of course, in that setting, we would not get such clean analytical expressions.

V. Discussion

We have seen that if there are k signals than the "signal" eigenvalues $\lambda_1, \ldots, \lambda_k$ of $\mathbf{R}_\Sigma$ will have almost sure limit that is different from the almost sure limit of the largest eigenvalue in noise-only setting if and only if $\lambda_i > \mathrm{T}(c, c_1)$. This is the reason why, in the large-system-relatively-large-sample-size limit, model order underestimation when using sample eigenvalue-only-based detection is unavoidable. This motivates our heuristic definition of the *effective number of identifiable signals* as follows:

$$k_{eff}(\mathbf{R}, \mathbf{\Sigma}) = \#\text{ Eigs. of } \mathbf{\Sigma}^{-1}\mathbf{R} > \mathrm{T}\left(\frac{n}{m}, \frac{n}{N}\right). \quad (26)$$

Fig. 2 shows the eigen-SNR threshold $\mathrm{T}(c, c_1) - 1$ needed for reliable detection for different values as a function of c for different values of $1/c_1$. Note the fundamental limit of detection in the situation when the noise-only covariance matrix is known *a priori* (solid line) and increase in the threshold eigen-SNR needed as the number of snapshots available to estimate the noise-only covariance matrix decreases.

A. Implications for Array Processing

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 $\mathbf{v}_1 \equiv \mathbf{v}(\theta_1)$ 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 signal-plus-noise covariance matrix is given by

$$\mathbf{R} = \sigma_{\mathrm{S}1}^2\mathbf{v}_1\mathbf{v}_1' + \sigma_{\mathrm{S}2}^2\mathbf{v}_2\mathbf{v}_2' + \mathbf{\Sigma} \quad (27)$$

where $\mathbf{\Sigma}$ is the noise-only covariance matrix. The matrix $\mathbf{R}_\Sigma$ defined in (5) can be decomposed as

$$\mathbf{R}_\Sigma = \mathbf{\Sigma}^{-1}\mathbf{R} = \sigma_{\mathrm{S}1}^2\mathbf{\Sigma}^{-1}\mathbf{v}_1\mathbf{v}_1' + \mathbf{\Sigma}^{-1}\sigma_{\mathrm{S}2}^2\mathbf{v}_2\mathbf{v}_2' + \mathbf{I}$$

so we that we can readily note that $\mathbf{R}_\Sigma$ has the $n-2$ smallest eigenvalues $\lambda_3 = \ldots = \lambda_n = 1$ and the two largest eigenvalues can be computed explicitly.

What emerges as a result of extending this computation as in [36] is the ability to succinctly, yet precisely, describe 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 $\mathbf{v}_2$. Note that since the effective number of signals depends on the structure of the theoretical signal and noise covariance matrices (via the eigenvalues of $\mathbf{R}_\Sigma$), different assumed noise covariance structures (AR(1) versus white noise, for example) will impact the signal level SNR needed for reliable detection in different ways.

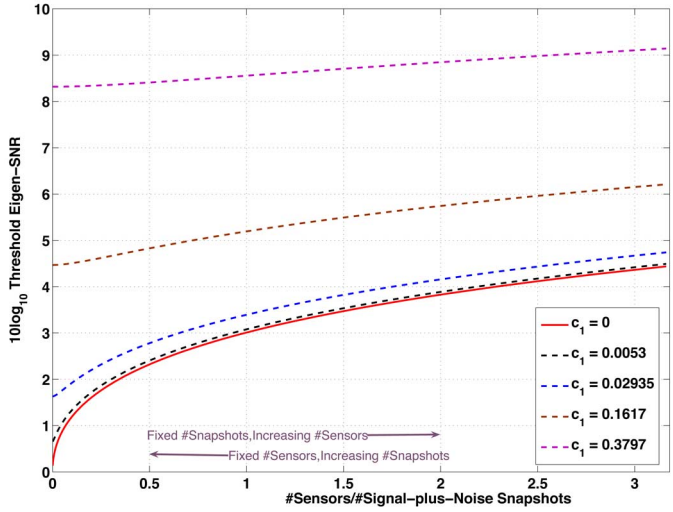


Fig. 2. Plot of the minimum (generalized) Eigen-SNR required [equal to $\mathrm{T}(c, c_1) - 1$ where $\mathrm{T}(c, c_1)$ is given by (25)] to be able to asymptotically discriminate between the "signal" and "noise" eigenvalue of the matrix $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ constructed as in (9) as a function of the ratio of the number of sensors to snapshots for different values of $1/c_1$, where $c_1 \approx$ Number of sensors/Number of noise-only snapshots. The gap between the dashed line and the bottom-most solid line and the bottom-most line represents the SNR loss due to noise covariance matrix estimation.

B. Other Applications

There is interest in detecting abrupt change in a system based on stochastic observations of the system using a network of sensors. When the observations made at various sensors can be modeled as Gauss–Markov random field (GMRF), as in [37], [38], then the conditional independence property of GMRFs [39] is a useful assumption. The assumption states that conditioned on a particular hypothesis, the observations at sensors are independent. This assumption results in the precision matrix, i.e., the inverse of the covariance matrix, having a sparse structure with many entries identically equal to zero.

Our results might be used to provide insight into the types of systemic changes, reflected in the structure of the signal-plus-noise covariance matrix, that are undetectable using sample generalized eigenvalue-based estimators. Specifically, the fact that the inverse of the noise-only covariance matrix will have a sparse structure means that one can experiment with different (assumed) conditional independence structures and determine how "abrupt" the system change would have to be in order to be reliably detected using finite samples.

Spectral methods are popular in machine learning applications such as unsupervised learning, image segmentation, and information retrieval [40]. Generalized eigenvalue-based tech-

niques for clustering have been investigated in [41], [42]. Our results might provide insight when spectral clustering algorithms are likely to fail. In particular, we note that the results of Theorem 4.3 hold even in situations where the data is not Gaussian (see Theorem 4.3) as is commonly assumed in machine learning applications.

VI. An Improved Algorithm for Reliable Detection of Signals in Noise

Fig. 1 shows the suboptimality of the ZKB algorithm in the high-dimensional, relatively small sample size setting relative to the fundamental limit of sample eigenvalue-based detection. This is not surprising because the ZKB algorithm was derived in [8] using large-sample-size asymptotics that did not account for the dimensionality of the underlying system. In modern parlance, they used small n, large m asymptotics. Much of the recent explosion in activity at the interface of multivariate statistics, statistical signal processing and random matrices has to do with the fact that we are now able to make incredibly precise computations in the large n, large m asymptotic regime [15, Sec. 2.4]. Consequently, the shortcomings of algorithms like the ZKB algorithm that are hard-wired based on the small n, large m asymptotics can be largely overcome by deriving new algorithms based on the large n, large m asymptotics.

For the problem at hand, it is an open problem to fully characterize the fluctuations of the eigenvalues of the matrix $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Sigma}}}$ for the spiked setting. We do, however, have a precise description due to Johnstone [18] for the signal-free setting, i.e., when $k = 0$ where we know that the distribution of the largest eigenvalue of $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Sigma}}}$, on appropriate centering and scaling, can be approximated to order $O(n^{-2/3})$ by the (real or complex) Tracy–Widom distribution [43]–[45].

In the spiked Wishart setting, corresponding to the signal bearing eigenvalues of $\widehat{\mathbf{R}}_{\mathbf{\Sigma}}$, Baik *et al.* [11] and El Karoui [14] showed that after appropriate centering and scaling, the distribution of the signal eigenvalues of $\widehat{\mathbf{R}}_{\mathbf{\Sigma}}$ above the detectability threshold will obey a Gaussian law with fluctuations on the order $n^{-1/2}$ whereas the signal eigenvalues below the detectability threshold will obey the Tracy–Widom law as in the signal-free case with fluctuations on the order $n^{-2/3}$.

Specifically, we have that for eigenvalues above threshold we get Gaussian fluctuations about $\lambda(t')$ in (23), so that

$$n^{1/2}\frac{\widehat{\lambda}_j - \lambda(t')}{\sigma} \xrightarrow{D} N(0,1)$$

for some σ. When below threshold we have that

$$n^{2/3}\frac{\widehat{\lambda}_j - b_2}{\sigma'} \xrightarrow{D} TW_{\{\mathbb{R},\mathbb{C}\}}$$

for some explicitly computable $\sigma' \neq \sigma$. Thus, for every $q = 0, 1, \ldots, n-1$, we can sequentially test the null hypothesis that there are "q signals present" against the alternative hypothesis that there are "at least $q+1$ signals present" on $\widehat{\lambda}_{q+1}$ by computing the test statistic

$$S(q) = n^{2/3}\frac{\widehat{\lambda}_{q+1} - b_2}{\sigma'}$$

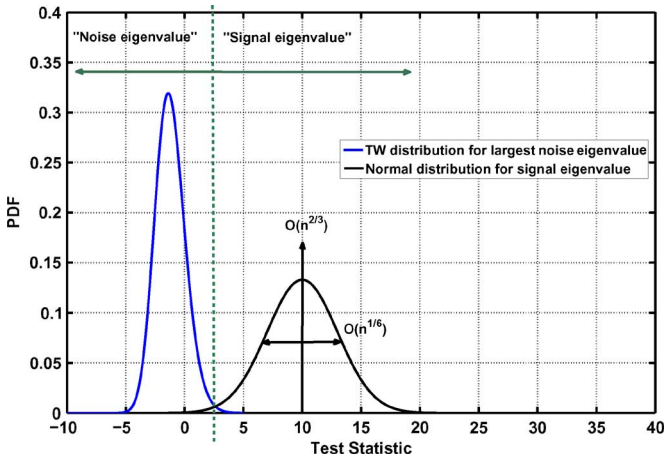


Fig. 3. Distribution of the test statistic for the algorithms in Table II under the null and alternative hypothesis when $q = k_{\text{eff}} - 1$. Consequently, we have $\text{Prob}(\widehat{k} = k_{\text{eff}-1}) \to 0$ as $n \to \infty$.

and accepting the null hypothesis whenever $S[q] < TW^{-1}_{\mathbb{R},\mathbb{C}}(1-\alpha) =: \tau_\alpha$ for some significance level α. Consequently, the first value of $q = 0, \ldots, n-1$ for which the "q signals present" hypothesis is accepted will be our estimate $\widehat{k}$ of the number of signals.

To see why we expect this algorithm to be asymptotically reliable, consider the setting where $q = k_{\text{eff}} - 1 \geq 0$ so that $\widehat{\lambda}_{q+1}$ corresponds to a signal eigenvalue that is above the threshold. Hence, we have

$$\widehat{\lambda}_{q+1} \approx \lambda(t') + \frac{O_p(1)}{\sqrt{n}}\mathcal{N}(0,1)$$

so that since $\lambda(t') > b_2$

$$n^{2/3}\frac{\widehat{\lambda}_{q+1} - b_2}{\sigma'} \approx n^{2/3}\frac{\lambda(t') - b_2}{\sigma'} + n^{1/3}O_p(1)\mathcal{N}(0,1)$$

so that for any fixed threshold τ_α (that does not depend on n), the probability that the test statistic exceeds τ_α is

$$\text{P}\left[S(q) > \tau_\alpha\right] = \text{P}\left(n^{2/3}\frac{\widehat{\lambda}_{q+1} - b_2}{\sigma'} > \tau_\alpha\right) \longrightarrow 1$$

as $n \longrightarrow \infty$. Thus, the alternative hypothesis that there are at least "$q = k_{\text{eff}} - 1$ signals present" will be accepted with exceedingly high probability. This situation is depicted in Fig. 3.

When $q = k_{\text{eff}}$, then $\widehat{\lambda}_{q+1}$ will be distributed like the largest signal-free eigenvalue. Hence, we have

$$\widehat{\lambda}_j \approx b_2 + \frac{O_p(1)}{n^{2/3}}TW$$

so that

$$n^{2/3}\frac{\widehat{\lambda}_{q+1} - b_2}{\sigma'} \approx TW.$$

Consequently, as $n \longrightarrow \infty$, the null hypothesis that there are "$q = k_{\text{eff}}$ signals present" will be accepted with probability $1 - \alpha$, for some significance level α, as desired.

TABLE II
NEW SIGNAL DETECTION ALGORITHMS. (a) NOISE COVARIANCE ESTIMATED. (b) NOISE COVARIANCE KNOWN

(a)

Algorithm 1

Input: Eigenvalues $\widehat{\lambda}_j$ for $j = 1, \ldots, n$ of $\widehat{\mathbf{R}}_{\mathbf{\Sigma}}$
1. Initialization: Set significance level $\alpha \in (0, 1)$
2. Compute $\tau_\alpha := TW^{-1}_{\{\mathbb{R},\mathbb{C}\}}(1-\alpha)$ from Table III
3. Set q $= 0$
4. Compute $\mu_{\{\mathbb{R},\mathbb{C}\}}[n-q, m]$ and $\sigma_{\{\mathbb{R},\mathbb{C}\}}[n-q, m]$ from Table IV-(a)
5. Is $\dfrac{\log m\hat{\lambda}_{q+1}/N - \mu_{\{\mathbb{R},\mathbb{C}\}}[n-q, m-q, N]}{\sigma_{\{\mathbb{R},\mathbb{C}\}}[n-q, m-q, N]} < \tau_\alpha$?
6. If yes, then go to step 9
7. Otherwise, increment q.
8. If $q < \min(n, m)$, go to step 4. Else go to step 9.
9. Return $\widehat{k} = q$

(b)

Algorithm 2

Input: Eigenvalues $\widehat{\lambda}_j$ for $j = 1, \ldots, n$ of $\widehat{\mathbf{R}}_{\mathbf{\Sigma}}$
1. Initialization: Set significance level $\alpha \in (0, 1)$
2. Compute $\tau_\alpha := TW^{-1}_{\{\mathbb{R},\mathbb{C}\}}(1-\alpha)$ from Table III
3. Set $q = 0$
4. Compute $\mu_{\{\mathbb{R},\mathbb{C}\}}[n-q, m]$ and $\sigma_{\{\mathbb{R},\mathbb{C}\}}[n-q, m]$ from Table IV-(b)
5. Is $\dfrac{m\hat{\lambda}_{q+1} - \mu_{\{\mathbb{R},\mathbb{C}\}}[n-q, m]}{\sigma_{\{\mathbb{R},\mathbb{C}\}}[n-q, m]} < \tau_\alpha$?
6. If yes, then go to step 9
7. Otherwise, increment q.
8. If $q < \min(n, m)$, go to step 4. Else go to step 9.
9. Return $\widehat{k} = q$

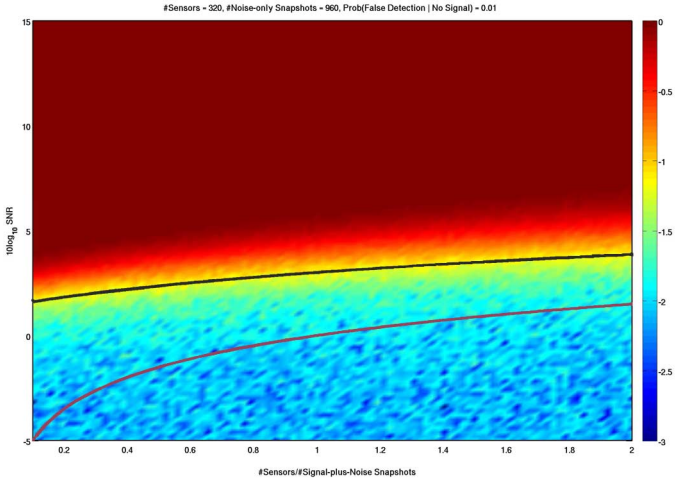


Fig. 4. Heat map of the log probability of signal detection using Algorithm 1 in Section VI, with the significance level α set at 0.01, in (eigen) SNR versus number of sensors to number of signal-plus-noise snapshots phase space. In this example, we set $n = 320$, $N = 960$ and w.l.o.g. $\mathbf{\Sigma} = \mathbf{I}$, $\mathbf{R} = \mathrm{diag}(\lambda_1 = 1 + \mathrm{SNR}, 1, \ldots, 1)$ and evaluated $\mathrm{Prob}(\hat{k} = 1)$ over 1000 Monte-Carlo trials and a grid of 100 equally spaced points in the -5 to 15 dB (eigen) SNR range and 100 equally spaced points in the $c_1 = n/m$ space by setting $m = n/c_1$. The values of the colormap at each of the 100×1000 faces were interpolated across each line segment and face to obtain the above plot. In the dark zone (upper half of the plot) a signal can be reliably detected whereas in the lighter zone (lower half of the plot) the signal is statistically indistinguishable from noise as evidenced from the probability of detection being close to the significance level. The superimposed solid black line demarcates the theoretically predicted threshold while the superimposed solid red line is the theoretically predicted threshold in the setting where the noise covariance matrix is perfectly known. The gap between the two lines thus represents the SNR loss due to noise covariance matrix estimation.

Table II summarizes an algorithm for estimating the number of signals based on this idea. Johnstone suggests using a logit transformation in [18] to $\widehat{\lambda}$ and derives the centering and scaling coefficients, listed in Table IV, associated with this transformation that results in the optimal rate of convergence. For the sake of completeness, we also include an extension to the setting where Σ is known *a priori*, in which case the correction terms that result in an optimal rate of convergence for the complex and real setting were derived by El Karoui [14] and Ma [46], respectively. In the setting where Σ is estimated, we found the analogous correction terms for the centering and rescaling coefficients that achieve the optimal rate of convergence by numerical simulations. Theoretically supporting these choices remains an open problem. Fig. 4 illustrates the performance of our algorithm and demonstrates its ability to reliably detect the presence of the signal around the derived fundamental statistical limit.

TABLE III
PERCENTILES OF THE TRACY–WIDOM REAL AND COMPLEX DISTRIBUTION

α	$1-\alpha$	$TW_{\mathbb{R}}^{-1}(1-\alpha)$	$TW_{\mathbb{C}}^{-1}(1-\alpha)$
0.990000	0.010000	-3.89543267306429	-3.72444594640057
0.950000	0.050000	-3.18037997693774	-3.19416673215810
0.900000	0.100000	-2.78242790569530	-2.90135093847591
0.700000	0.300000	-1.91037974619926	-2.26618203984916
0.500000	0.500000	-1.26857461658107	-1.80491240893658
0.300000	0.700000	-0.59228719101613	-1.32485955606020
0.100000	0.900000	0.45014328905825	-0.59685129711735
0.050000	0.950000	0.97931605346955	-0.23247446976400
0.010000	0.990000	2.02344928138015	0.47763604739084
0.001000	0.999000	3.27219605900193	1.31441948008634
0.000100	0.999900	4.35942034391365	2.03469175457082
0.000010	0.999990	5.34429594047426	2.68220732168978
0.000001	0.999999	6.25635442969338	3.27858828203370

VII. CONCLUSION

Fig. 4 captures the fundamental statistical limit encountered when attempting to discriminate signal from noise using finite samples. Simply put, a signal whose eigen-SNR is below the detectability threshold cannot be reliably detected while a signal above the threshold can be. In settings such as wireless communications and biomedical signal processing where the signal power is controllable, our results provide a prescription for how strong it needs to be so that it can be detected. If the signal level is barely above the threshold, simply adding more sensors might actually degrade the performance because of the increased dimensionality of the system. If, however, either due to clever signal design or physics-based modeling, we are able to reduce (or identify) the dimensionality of the subspace spanned by signal, then according to Fig. 4 the detectability threshold will also be lowered. With VLSI advances making sensors easier and cheaper to deploy, our results demonstrate exactly why the resulting gains in systemic performance will more than offset the effort we will have to invest in developing increasingly more sophisticated dimensionality reduction techniques. Understanding

the fundamental statistical limits of techniques for signal detection in the setting where the noise-only sample covariance matrix is singular remains an important open problem.

APPENDIX
PROOFS AND COMPUTATIONS

A. Proof of Extension of Theorem 3.1 for Random T_n

Indeed, let T denote the probability space generating $\{T_n\}$, X the probability space generating $\{X_{ij} : i,j \geq 1\}$. Let their respective measures be denoted by P_T,P_X, the product measure on $T \times X$ by $\mathsf{P}_{T\times X}$. Consider, for example in case 2), we define $A = \{\lambda_{i_n}^{B_n} > b$ and $\lambda_{i_n+1}^{B_n} < a$ for all large $n\}$. Let $t \in T$ be an element of the event defined in $(**)$. Then by Theorem 3.1 $I_A((t,x)) = 1$ for all x contained in a subset of X having probability 1.

Therefore, by Tonelli's theorem we have

$$\begin{aligned}\mathsf{P}(A) &= \int I_A(t,x) dP_{T\times X}((t,x)) \\ &= \int \left[\int I_A(t,x) dP_X(x)\right] dP_T(t) = \int 1 dP_T(t) = 1.\end{aligned}$$

B. Proof of Theorem 4.1

For $m \in [-1/t_a, -1/t_b] \cap \{-1/t'\}^c$, we have

$$x_{c_n,H_n}(m) = -m^{-1} + c_n\left(n^{-1}\sum_{j=i_n}^{i_n+\ell-1}\frac{\lambda_j^{T_n}}{1+\lambda_j^{T_n}m^{-1}} + (n-\ell)\frac{1}{n}\int \lambda(1+\lambda m)^{-1} d\hat{H}_n(\lambda)\right).$$

By considering continuity points of H in (α,β) we see that H is constant on this interval, and consequently, this interval is also contained in S'_H.

Because of 2) we have $(d/dm)x_{c,H}(m) \geq 0$ for $m = -1/t_a, -1/t_b$ [recall (15), (16)].

By Lemma 3.2 we therefore have $(d/dm)x_{c,H}(m) > 0$ for all $m \in (-1/t_a, -1/t_b)$. Thus, we can find $[\underline{t}_a, \underline{t}_b] \subset [t_a, t_b]$ and $\delta > 0$, such that $t' \in (\underline{t}_a, \underline{t}_b)$ and for all n large $(d/dm)x_{c_n,\hat{H}_n}(m) \geq \delta$ for all $m \in [-1/\underline{t}_a, -1/\underline{t}_b]$.

It follows that for any positive ϵ sufficiently small, there exist positive δ' with $\delta' \leq \epsilon$, such that, for all n large, both $[-1/t' - \epsilon - \delta', -1/t' - \epsilon]$, and $[-1/t' + \epsilon, -1/t' + \epsilon + \delta']$:

- 1) are contained in $[-1/\underline{t}_a, -/\underline{t}_b]$, and
- 2) $(d/dm)x_{c_n,H_n}(m) > 0$ for all m contained in these two intervals.

Therefore, by Lemma 3.1, for all n large, $[x_{c_n,H_n}(-1/t' - \epsilon - \delta'), x_{c_n,H_n}(-1/t' - \epsilon)]$ and $[x_{c_n,H_n}(-1/t' + \epsilon), x_{c_n,H_n}(-1/t' + \epsilon + \delta')]$ lie outside the support of F^{c_n,H_n}. Let $a_L = x_{c,H}(-1/t' - \epsilon - (2/3)\delta')$, $b_L = x_{c,H}(-1/t' - \epsilon - (1/3)\delta')$, $a_R = x_{c,H}(-1/t' + \epsilon + (1/3)\delta')$, and $b_R = x_{c,H}(-1/t' + \epsilon + (2/3)\delta')$. Then for all n large

$$\begin{aligned}[a_L, b_L] \subset \Bigg(& x_{c,H}\left(-1/t' - \epsilon - \frac{5}{6}\delta'\right), \\ & x_{c,H}\left(-1/t' - \epsilon - \frac{1}{6}\delta'\right)\Bigg) \\ \subset & [x_{c_n,H_n}(-1/t' - \epsilon - \delta'), x_{c_n,H_n}(-1/t' - \epsilon)] \quad (28)\end{aligned}$$

and

$$\begin{aligned}[a_R, b_R] \subset \Bigg(& x_{c,H}\left(-1/t' + \epsilon + \frac{1}{6}\delta'\right), \\ & x_{c,H}\left(-1/t' + \epsilon + \frac{5}{6}\delta'\right)\Bigg) \\ \subset & [x_{c_n,H_n}(-1/t' + \epsilon), x_{c_n,H_n}(-1/t' + \epsilon + \delta')]. \quad (29)\end{aligned}$$

It follows then that $[a_L, b_L]$, $[a_R, b_R]$ each lie in an open interval in $S'_{F^{c_n,H_n}}$ for all n large. Moreover $m_{F^{c,H}}(b_R) < 0$. Therefore, case 2) of Theorem 3.1 applies and we have

$$\mathsf{P}\left(\lambda_{i_n}^{B_n} < a_R \quad \text{and} \quad \lambda_{i_n+\ell-1}^{B_n} > b_L \quad \text{for all } n \text{ large}\right) = 1.$$

Therefore, considering a countable collection of ϵ's converging to zero, we conclude that, with probability 1

$$\lim_{n\to\infty} \lambda_{i_n}^{B_n} = \lim_{n\to\infty} \lambda_{i_n+\ell-1}^{B_n} = x_{c,H}(-1/t') = (20).$$

C. Proof of Theorem 4.2

Proof: By Lemma 3.1, $[t_a, t_b] \in S'_H$, and for a suitable positive ϵ, $x_{c,H}$ is increasing on $[m_{c,H}(a) - \epsilon, m_{c,H}(b) + \epsilon]$, which does not contain 0. Therefore, $t_a < t' < t_b$. If $c(1 - H(0)) > 1$, that is, case 1) of Theorem 3.1 holds, then $a > x_0$, since x_0 is the almost sure limit of $\lambda_m^{B_n}$ so λ' cannot be smaller than it, and necessarily $x_0 \in S_F$. Therefore, $m_{c,H}(b) < 0$, so that $0 < t_a$. It must be the case that only $o(n)$ eigenvalues of t_n lie in $[t_a, t_b]$, since otherwise $[t_a, t_b]$ would not be outside the support of H. We have then $\hat{H}_n \to_D H$ as $n \to \infty$, so from the dominated convergence theorem we have $(d/dm)x_{c_n,\hat{H}_n}(m) \to (d/dm)x_{c,H}(m)$ for all $m \in [m_{c,H}(a) - \epsilon, m_{c,H}(b) + \epsilon]$, implying for all n large $(d/dm)x_{c_n,\hat{H}_n}(m) > 0$ for all $m \in [m_{c,H}(a), m_{c,H}(b)]$. Therefore 2) is true.

We assume now that ℓ is bounded. Suppose 1) does not hold. Then we could find a subsequence $\{n_j\}$ of the natural numbers for which $\ell' = \ell'(n)$ of the ℓ eigenvalues converge to a $\underline{t}' \neq t'$, the remaining $\ell - \ell'$, if positive, eigenvalues remaining a positive distance d from $\underline{t}'$. Replace $\{\mathbf{T}_n\}$ with $\{\mathbf{T}'_n\}$ which matches the original sequence when $n = n_j$ and for $n \neq n_j$, $\mathbf{T}'_n$ has ℓ' eigenvalues equal to $\underline{t}'$, with the remaining $\ell - \ell'$, again, if positive, eigenvalues of $\mathbf{T}'_n$ at least d away from $\underline{t}'$. Then we have by Theorem 3.1, (20), with t' replaced by $\underline{t}'$, holding for ℓ' of the eigenvalues of $(1/m)\mathbf{T}'_n{}^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}'_n{}^{1/2}$. Thus, on $\{n_j\}$, we have the almost sure convergence of ℓ eigenvalues of $\mathbf{B}_n$ to $x_{c,H}(-1/\underline{t}') \in [a,b]$ which, because $x_{c,H}(-1/t)$ is an increasing function, does not equal $\lambda' = x_{c,H}(-1/t')$. This contradicts the assumption of convergence in probability to eigenvalues to only one number, namely λ'. Therefore 1) holds. ■

TABLE IV
PARAMETERS FOR NEW SIGNAL DETECTION ALGORITHMS IN TABLE II. (a) ALGORITHM 1. (b) ALGORITHM 2

Arguments / Scenario	$\mu_{\{\cdot\}}[n,m,N]$	$\sigma_{\{\cdot\}}[n,m,N]$	Related parameters
$\mathbf{x}_i \in \mathbb{R}$	$2\log\tan\frac{\gamma+\phi}{2}$	$\left[\frac{16}{(N+m-1)^2}\,\frac{1}{\sin^2(\gamma+\phi)\,\sin\phi\sin\gamma}\right]^{1/3}$	$\gamma = 2\sin^{-1}\sqrt{\frac{\min(n,m)-0.5}{m+N-1}}$ $\phi = 2\sin^{-1}\sqrt{\frac{\max(n,m)-0.5}{m+N-1}}$
$\mathbf{x}_i \in \mathbb{C}$	$\frac{\frac{u_{\widetilde{N}}}{\tau_{\widetilde{N}}}+\frac{u_{\widetilde{N}-1}}{\tau_{\widetilde{N}-1}}}{\frac{1}{\tau_{\widetilde{N}}}+\frac{1}{\tau_{\widetilde{N}-1}}}$	$\frac{2}{\frac{1}{\tau_{\widetilde{N}}}+\frac{1}{\tau_{\widetilde{N}-1}}}$	$\widetilde{N} := \min(n,m)$ $\gamma_{\widetilde{N}} = 2\sin^{-1}\sqrt{\frac{\widetilde{N}+0.5}{2\widetilde{N}+N-n+\lvert m-n\rvert+1}}$ $\phi_{\widetilde{N}} = 2\sin^{-1}\sqrt{\frac{\widetilde{N}+\lvert m-n\rvert+0.5}{2\widetilde{N}+N-n+\lvert m-n\rvert+1}}$ $\tau_{\widetilde{N}} = \left[\frac{16}{(2\widetilde{N}+N-n+\lvert m-n\rvert+1)^2}\,\frac{1}{\sin^2(\phi_{\widetilde{N}}+\gamma_{\widetilde{N}})\,\sin\phi_{\widetilde{N}}\,\sin\gamma_{\widetilde{N}}}\right]^{1/3}$ $u_{\widetilde{N}} = 2\log\tan\frac{\phi_{\widetilde{N}}+\gamma_{\widetilde{N}}}{2}$

(a)

Arguments / Scenario	$\mu_{\{\cdot\}}[n,m]$	$\sigma_{\{\cdot\}}[n,m]$	Related parameters
$\mathbf{x}_i \in \mathbb{R}$	$\left(\sqrt{n-\frac{1}{2}}+\sqrt{m-\frac{1}{2}}\right)^2$	$\left(\sqrt{n-\frac{1}{2}}+\sqrt{m-\frac{1}{2}}\right)\left(\frac{1}{\sqrt{n-\frac{1}{2}}}+\frac{1}{\sqrt{m-\frac{1}{2}}}\right)^{1/3}$	-
$\mathbf{x}_i \in \mathbb{C}$	$\left(\frac{1}{\sigma_{n,m-1}^{1/2}}+\frac{1}{\sigma_{n-1,m}^{1/2}}\right)\left(\frac{1}{\mu_{n,m-1}\sigma_{n,m-1}^{1/2}}+\frac{1}{\mu_{n-1,m}\sigma_{n-1,m}^{1/2}}\right)^{-1}$	$(1+\gamma_{n,m})\left(\frac{1}{\sigma_{n,m-1}}+\frac{\gamma_{n,m}}{\sigma_{n-1,m}}\right)$	$\mu_{n,m} = \left(\sqrt{n+\frac{1}{2}}+\sqrt{m+\frac{1}{2}}\right)^2$ $\sigma_{n,m} = \left(\sqrt{n+\frac{1}{2}}+\sqrt{m+\frac{1}{2}}\right)\left(\frac{1}{\sqrt{n+\frac{1}{2}}}+\frac{1}{\sqrt{m+\frac{1}{2}}}\right)^{1/3}$ $\gamma_{n,m} = \mu_{n,m-1}\sigma_{n-1,m}^{1/2}/\,\mu_{n-1,m}\sigma_{n,m-1}^{1/2}$

(b)

ACKNOWLEDGMENT

R. R. Nadakuditi would like to thank A. Baggeroer for encouragement and invaluable feedback. The authors would like to thank A. Edelman for his encouragement and support, Interactive Supercomputing, Inc., for providing access to the Star-P parallel computing software and S. Raghunathan of Interactive Supercomputing, Inc., for his patience and support in answering our multitude of Star-P programming queries. The authors remain grateful to A. Davis and C. Hill of MIT for granting us access to the Darwin Project computing cluster. Thanks to their involvement we were able to program, debug, and complete the computation needed to produce Figs. 4 and Fig. 1 in 4 days! Without their gracious help, the computation would have taken 5 months on the latest single processor laptop. The authors also would like to thank F. Bornemann for providing the MATLAB code for computing the percentiles in Table III. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*, ser. Johns Hopkins Studies in the Mathematical Sciences, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[2] 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.

[3] K. Sekihara, D. Poeppel, A. Marantz, H. Koizumi, and Y. Miyashita, "Noise covariance incorporated MEG-MUSIC algorithm: A method formultiple-dipole estimation tolerant of the influence of background brainactivity," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 839–847, Sep. 1997.

[4] K. Sekihara, D. Poeppel, A. Marantz, H. Koizumi, and Y. Miyashita, "MEG spatio-temporal analysis using a covariance matrix calculated from nonaveraged multiple-epoch data," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 5, pp. 515–521, May 1999.

[5] H. L. V. Trees, *Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing*. New York: Wiley, 2002.

[6] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

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

[8] 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. Multivariate Anal.*, vol. 20, no. 1, pp. 26–49, 1986.

[9] 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.

[10] R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, Jul. 2008.

[11] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, 2005.

[12] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivariate Anal.*, vol. 97, no. 6, pp. 1382–1408, 2006.

[13] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statist. Sinica*, vol. 17, no. 4, pp. 1617–1642, 2007.

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

[15] I. M. Johnstone, "High dimensional statistical inference and random matrices," in *Proc. Int. Congr. Math.*, 2006 [Online]. Available: http://arxiv.org/abs/math/0611589

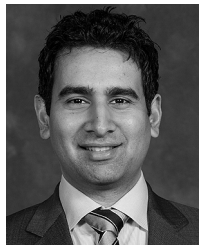
[16] D. C. Hoyle and M. Rattray, "Statistical mechanics of learning multiple orthogonal signals: Asymptotic theory and fluctuation effects," *Phys. Rev. E (3)*, vol. 75, no. 1, p. 016 101, 2007, 13.

[17] B. Nadler, "Finite sample approximation results for principal component analysis: A matrix perturbation approach," *Ann. Statist.*, vol. 36, no. 6, pp. 2791–2817, 2008.

[18] I. M. Johnstone, "Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence," *Ann. Statist.*, vol. 36, no. 6, pp. 2638–2716, 2008.

[19] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20 A, pp. 32–52, 1928.

[20] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, ser. Probability and Mathematical Statistics. New York: Wiley, 1982.
[21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, ser. Wiley Series in Probability and Statistics, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2003.
[22] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate F matrix," *SIAM J. Math. Anal.*, vol. 16, no. 3, pp. 641–646, 1985.
[23] K.-W. Tam and Y. Wu, "On rates of convergence of general information criteria in signal processing when the noise covariance matrix is arbitrary," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1667–1671, Nov. 1991.
[24] Z. Zhu, S. Haykin, and X. Huang, "Estimating the number of signals using reference noise samples," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 3, pp. 575–579, May 1991.
[25] 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.
[26] 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.
[27] 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.
[28] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. Beijing, China: Science Press, 2006.
[29] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, no. 2, pp. 331–339, 1995.
[30] J. W. Silverstein and S.-I. Choi, "Analysis of the limiting spectral distribution of large-dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 295–309, 1995.
[31] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix," *Probab. Theory Related Fields*, vol. 78, no. 4, pp. 509–521, 1988.
[32] Z. D. Bai and Y. Q. Yin, "Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix," *Ann. Probab.*, vol. 21, no. 3, pp. 1275–1294, 1993.
[33] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, 1998.
[34] Z. D. Bai and J. W. Silverstein, "Exact separation of eigenvalues of large dimensional sample covariance matrices," *Ann. Probab.*, vol. 27, no. 3, pp. 1536–1555, 1999.
[35] P. Billingsley, *Probability and Measure*, ser. Wiley Series in Probability and Mathematical Statistics, 3rd ed. New York: Wiley, 1995.
[36] R. Nadakuditi and J. Silverstein, "Fundamental limit of sample eigenvalue based detection of signals in colored noise using relatively few samples," in *Proc. Signals, Syst. Comput. ACSSC 2007. Conf. Rec. 41st Asilomar Conf.*, Nov. 2007, pp. 686–690.
[37] A. Anandkumar, L. Tong, and A. Swami, "Detection of Gauss–Markov random fields with nearest-neighbor dependency," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 816–827, Feb. 2009.
[38] Y. Sung, L. Tong, and H. Poor, "Neyman–Pearson detection of Gauss–Markov signals in noise: Closed-form error exponent and properties," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1354–1365, Apr. 2006.
[39] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: Chapman & Hall/CRC, 2005.
[40] S. Vempala, "Spectral algorithms for learning and clustering," *Lecture Notes in Comput. Sci.*, vol. 4539, p. 3, 2007.
[41] O. Mangasarian and E. Wild, "Multisurface proximal support vector machine classification via generalized eigenvalues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 69–74, Jan. 2006.
[42] M. Guarracino, C. Cifarelli, O. Seref, and P. Pardalos, "A classification method based on generalized eigenvalue problems," *Optim. Methods Software*, vol. 22, no. 1, pp. 73–81, 2007.
[43] C. Tracy and H. Widom, "Level-spacing distribution and airy kernel," *Commun. Math. Phys.*, vol. 159, pp. 151–174, 1994.
[44] C. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Commun. Math. Phys.*, vol. 177, pp. 727–754, 1996.
[45] I. M. Johnstone, "High dimensional statistical inference and random matrices," in *Proc. Int. Congr. Math. Vol. I*, Zürich, Switzerland, 2007, pp. 307–333, Eur. Math. Soc..
[46] Z. Ma, "Accuracy of the Tracy–Widom limit for the largest eigenvalue in white Wishart matrices," 2008 [Online]. Available: http://arxiv.org/abs/0810.1329



Raj Rao Nadakuditi received the B.S. degree in electrical engineering from Lafayette College, Easton, PA, in 1999, and the M.S. and Ph.D. degrees in electrical engineering and oceanographic engineering from the Massachusetts Institute of Technology and Woods Hole Oceanographic Institution Joint Program in Applied Ocean Science and Engineering in 2001 and 2007, respectively.

He joined the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, in September 2009, where he is currently an Assistant Professor. His research interests are in the general area of statistical signal processing with an emphasis on the application of random matrix theory for high-dimensional problems that arise in the context of radar, sonar, wireless communications, and machine learning.

Dr. Nadakuditi received an Office of Naval Research Ocean Acoustics Award to continue his postdoctoral work at the Massachusetts Institute of Technology.

Jack W. Silverstein received the B.A. degree in mathematics from Hofstra University, Hempstead, NY, in 1971 and the M.S. and Ph.D. degrees in applied mathematics from Brown University, Providence, RI, in 1973 and 1975, respectively.

After postdoctoring and teaching at Brown, he began in 1978 a tenure track position in the Department of Mathematics at North Carolina State University, Raleigh, where he has been Professor since 1994. His research interests are in probability theory with emphasis on the spectral behavior of large-dimensional random matrices.

Prof. Silverstein was elected Fellow of the Institute of Mathematical Statistics in 2007.