12

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 article 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 SNR's close to the critical value. Numerical simulations highlight the accuracy of our analytical prediction and permit us to extend our heuristic definition of the *effective number of identifiable signals in colored noise*. 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

signal detection, random matrices, sample covariance matrix, Wishart distribution, multivariate F distribution

EDICS Category: SSP-DETC Detection; SAM-SDET Source detection

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. 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

Department of Mechanical Engineering, Massachusetts Institute of Technology, Email: raj@mit.edu, Phone: (857) 891 8303, Fax: (617) 253-4358

Department of Mathematics, North Carolina State University, Email: jack@unity.ncsu.edu, Phone: (857) 891 8303, Fax: (617) 253-4358

Supported by an Office of Naval Research Special Postdoctoral Award under grant N00014-07-1-0269. Portions of this work were presented at the 2007 Asilomar Conference on Signals and Systems.

Supported by the U.S. Army Research Office under Grant W911NF-05-1-0244.

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], [7]. 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 [8], 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 [9], 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. Analogous to the definition in [9], we define 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, algorithm for estimating the number of signals based on the recent work of Johnstone [10]. 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, correctly for the reason described earlier, is unable to distinguish the signal from noise.

The paper is organized as follows. We formulate the problem in Section II and state the main result in Section III. The effective number of signals is defined in Section III-A along with a discussion on its implications for applications such as array processing, sensor networks and machine learning. A new algorithm for detecting the number of signals is presented in Section IV. Concluding remarks are offered in Section V. The mathematical proofs of the main result are provided in Section VI.

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 modelled as

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

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

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

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 apriori and was non-singular, 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

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

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 [1][Section 8.7], we note that the eigenvalues of $\mathbf{R}_{\mathbf{\Sigma}}$ are exactly the generalized eigenvalues of the regular matrix pair $(\widehat{\mathbf{R}}, \widehat{\mathbf{\Sigma}})$. 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 $\mathbf{R}_{\mathbf{\Sigma}}$ of 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 apriori, 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 straight-forward algorithm cannot be used. Instead we have an estimate 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

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

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 $\widehat{\mathbf{\Sigma}}$, which has the Wishart distribution [11], is non-singular and hence invertible with probability 1 [12, Chapter 3, pp. 97], [13, Chapter 7.7, pp. 272-276]. Following (5), we then form the matrix

$$\widehat{\mathbf{R}}_{\widehat{\Sigma}} = \widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{R}}, \tag{9}$$

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 $(\widehat{\mathbf{R}}, \widehat{\mathbf{\Sigma}})$. Note that whenever $N < n$, the signal-plus-noise sample covariance matrix $\mathbf{R}$ will be singular so that the $n-N$ generalized eigenvalues will equal zero, *i.e.*, $\hat{\lambda}_{N+1} = \hat{\lambda}_{N+2} = \ldots = \hat{\lambda}_n = 0$. Figure 1 illustrates why the blurring of the sample eigenvalues relative to the population eigenvalues makes the problem more challenging.

In this paper, we are interested in the class of algorithms that infer the number of signals buried in arbitrary noise from the eigenvalues of $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Sigma}}}$ or $\widehat{\mathbf{R}}_{\mathbf{\Sigma}}$ alone. Such algorithms are widely used in practice and arise naturally from classical multivariate statistical theory [10] where the matrix $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ is referred to as the multivariate F matrix [12], [14]. The information theoretical approach to model order estimation, first introduced by Wax and Kailath [6], was extended to the colored noise setting by Zhao et al in [15] who prove consistency of their estimator in the large sample size regime. Tam and Wu [16] performed a rate of convergence analysis of Zhao et al's algorithm for different choices of the penalty function; their does not yield any insight into the high dimensional, finite sample setting.

Consequently, research has focussed on developing sophisticated techniques for improving performance of eigenvalue based methods in the finite sample setting. Zhu et al [17] improve the performance of their eigenvalue estimator by assuming a model for the noise covariance matrix. Stoica and Cedervall [18] 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

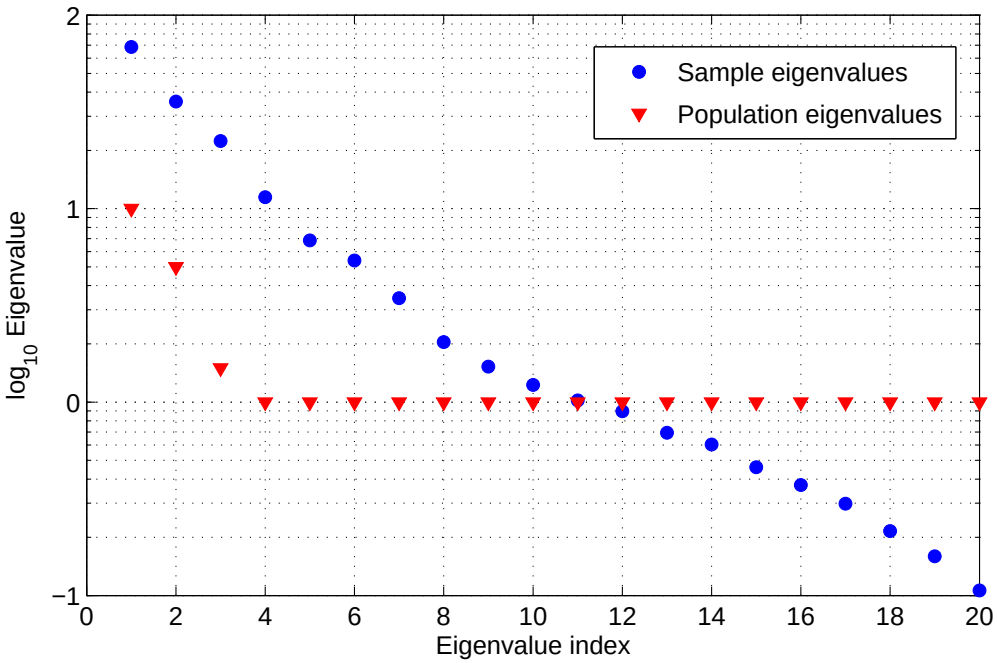


Fig. 1: The dimension of the "noise" subspace is equal to the multiplicity of the population eigenvalue equal to one. When the population eigenvalues are known, then detecting the number of signals becomes trivial. However, estimating the number of signals from the sample generalized eigen-spectrum is considerably more challenging because of the finite sample effects. Specifically, the finite number of noise-only and signal-plus-noise samples induces a blurring in the sample eigenspectrum relative to the population eigenspectrum makes discrimination of the "signal" from the "noise" challenging. The figure shows one random instance generated for a $n = 20$ dimensional system with $N = 25$ noise-only samples and $m = 40$ signal-plus-noise bearing samples.

signals. Other techniques in the literature 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 example [19], [20] and the references in [9].

Informally speaking, it is evident that performance of such model order estimation algorithms is coupled to the "quality" of the estimated signal-plus-noise and noise-only covariance matrices which in turn are dependent on the number of snapshots used to estimate them, respectively. Researchers applying these techniques have noted the absence of a mathematically rigorous, general purpose formula in the literature for predicting the minimum number of samples needed to obtain "good enough" detection accuracy (see, for example [3][pp. 846]. A larger, more fundamental question that has remained unanswered, till now, is whether there is a statistical limit being encountered.

We tackle this problem head on in this paper by employing sophisticated techniques from random matrix theory in [21]. We show that in an asymptotic sense, to be made precise later, that only the "signal" eigenvalues of $\mathbf{R}_\Sigma$ that are above a deterministic threshold can be reliably distinguished from the "noise" eigenvalues. The threshold is a simple, deterministic function of the the dimensionality of the system, the number of noise-only and signal-plus-noise snapshots, and the noise and signal-plus noise covariance, and described explicitly next. Note the applicability of the results to the situation when the signal-plus-noise covariance matrix is singular.

III. Main result

For a Hermitian matrix $\mathbf{A}$ with n real eigenvalues (counted with multiplicity), the empirical distribution function (e.d.f.) is defined as

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

Of particular interest is the convergence of the e.d.f. of $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ in the signal-free case, which is described next.

Theorem 3.1: Let $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ denote the matrix in (9) formed from m (complex Gaussian) noise-only snapshots and N independent noise-only (complex Gaussian) snapshots. Then the e.d.f. $F^{\widehat{\mathbf{R}}_{\widehat{\Sigma}}}(x) \to F^{R_\Sigma}(x)$ almost surely for every x, as $m, n(m) \to \infty$, $m, N(m) \to \infty$ and $c_m = n/m \to c > 0$ and $c_N^1 = n/N \to c_1 < 1$ where

$$dF(x) = \max\left(0, \left(1 - \frac{1}{c}\right)\right)\delta(x) + \frac{(1-c_1)\sqrt{(x-b_1)(b_2-x)}}{2\pi x(xc_1+c)}\mathbb{I}_{[b_1,b_2]}(x)\,dx, \tag{11}$$

where

$$b_1 = \left(\frac{1-\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2, \quad b_2 = \left(\frac{1+\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2, \tag{12}$$

$\mathbb{I}_{[b_1,b_2]}(x) = 1$ when $b_1 \le x \le b_2$ and zero otherwise, and $\delta(x)$ is the Dirac delta function.

Proof: This result was proved in [14]. When $c_1 \to 0$ we recover the famous Marčenko-Pastur density [22]. ■

The following result exposes when the "signal" eigenvalues are asymptotically distinguishable from the "noise" eigenvalues.

Theorem 3.2: Let $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ denote the matrix in (9) formed from m (real or complex Gaussian) signal-plus-noise snapshots and N independent (real or complex Gaussian) noise-only snapshots. Denote the eigenvalues of $\mathbf{R}_\Sigma$ by $\lambda_1 \ge \lambda_2 > \ldots \ge \lambda_k > \lambda_{k+1} = \ldots \lambda_n = 1$. Let l_j denote the j-th largest eigenvalue of $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$. Then as $n, m(n) \to \infty$, $n, N(n) \to \infty$ and $c_m = n/m \to c > 0$ and $c_N^1 = n/N \to c_1 < 1$ we have

$$l_j \to \begin{cases} \lambda_j\left(1 - c - c\dfrac{-c_1\,\lambda_j - \lambda_j + 1 + \sqrt{{c_1}^2{\lambda_j}^2 - 2\,c_1\,{\lambda_j}^2 - 2\,c_1\,\lambda_j + {\lambda_j}^2 - 2\,\lambda_j + 1}}{2c_1\,\lambda_j}\right), & \lambda_j > \tau(c,c_1) \\ \dfrac{-\,c_1\,c + \,c + 1 + \,c_1 + 2\sqrt{c + c_1 - c_1 c}}{{c_1}^2 + 1 - 2\,c_1}, & \lambda_j \le \tau(c,c_1) \end{cases}$$

for $j = 1, \ldots, k$ and the convergence is almost surely and the threshold $\mathrm{T}(c, c_1)$ is given by

$$\mathrm{T}(c,c_1) = \frac{1 + \tau - \tau c_1 + \sqrt{(1+\tau-\tau c_1)^2 - 4\tau}}{2}, \tag{13}$$

where

$$\tau = \frac{(1+c_1)\alpha + \sqrt{\alpha}\sqrt{4\alpha - c_1 + (1-c_1)^2c^2}}{(1-c_1)^2\alpha} = \frac{(1+c_1)\alpha + \sqrt{\alpha}(2c_1 + c(1-c_1))}{(1-c_1)^2\alpha} \tag{14}$$

and $\alpha = c + c_1 - c_1 c$.

Proof: The result follows from Theorem 6.5. The threshold $\mathrm{T}(c, c_1)$ is obtained by solving the inequality

$$t' > \tau$$

where for $j = 1, \ldots, k$, t', from [9], [23]–[25], is given by

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

and τ is given by (31).

Note that when $c_1 \to 0$, $\mathrm{T}(c, c_1) \to (1 + \sqrt{c})$ so that we recover the results of Baik and Silverstein [24]. ■

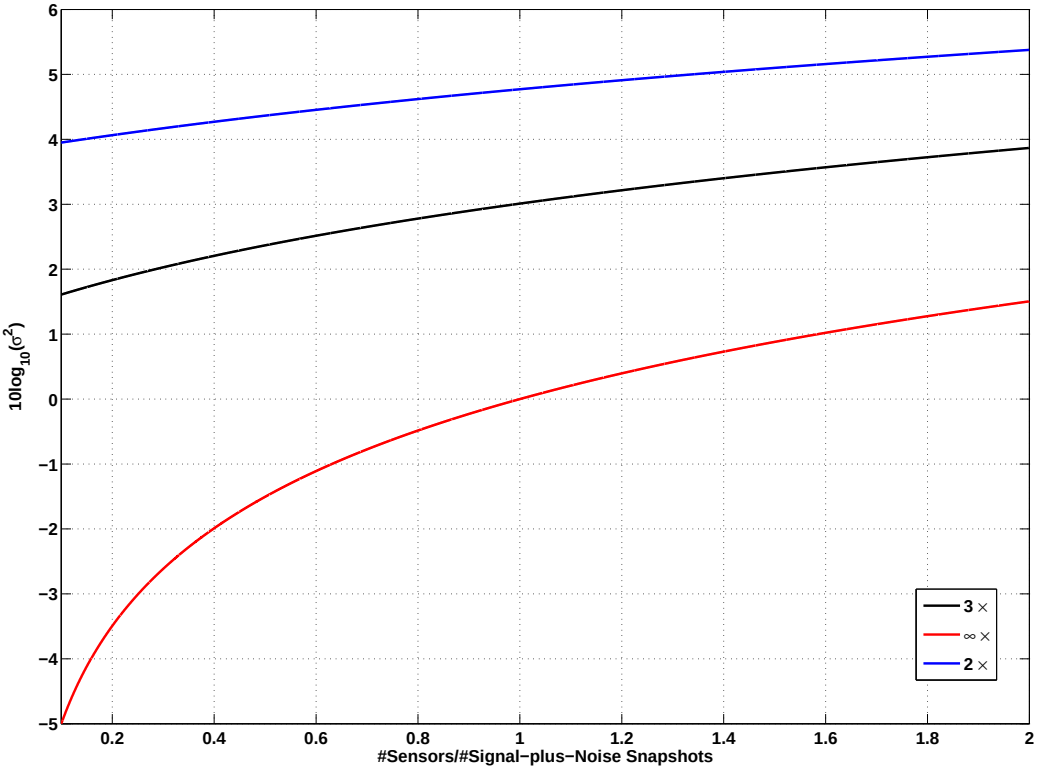


Fig. 2: Plot of the minimum (generalized) Eigen-SNR required (equal to $\mathrm{T}(c, c_1) - 1$ where $\mathrm{T}(c, c_1)$ is given by (13)) to be able to asymptotically discriminate between the "signal" and "noise" eigenvalue of the matrix $\widehat{\mathbf{R}}_{\widehat{\Sigma}}$ constructed as in (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 upper two lines and the bottom most line represents the SNR loss due to noise covariance matrix estimation.

A. Effective number of identifiable signals

Theorem 3.2 brings into sharp focus the reason why, in the large-system-relatively-large-sample-size limit, model order underestimation is sometimes unavoidable. This motivates our heuristic definition of the *effective number of identifiable signals* below:

$$k_{eff}(\mathbf{R}, \mathbf{\Sigma}) = \#\text{ Eigs. of } \mathbf{\Sigma}^{-1}\mathbf{R} > \mathrm{T}(c, c_1) \approx \mathrm{T}\left(\frac{n}{m}, \frac{n}{N}\right). \tag{15}$$

If we denote the eigenvalues of $\mathbf{R}_{\mathbf{\Sigma}} \equiv \mathbf{\Sigma}^{-1}\mathbf{R}$ by $\lambda_1 \geq \lambda_2 > \ldots \geq \lambda_k > \lambda_{k+1} = \ldots \lambda_n = 1$ then we define the eigen-SNR of the j-th signal as $\lambda_j - 1$ then (15) essentially states that signals with eigen-SNR's smaller than $\mathrm{T}(n/m, n/N)$ will be asymptotically undetectable.

Figure 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 $1/c_1$. Such an analytical prediction was not possible before the results presented in this paper. Note the fundamental limit of detection in the situation when the noise-only covariance matrix is known apriori (solid line) and increase in the threshold eigen-SNR needed as the number of snapshots available to estimate the noise-only covariance matrix decreases.

B. Implications for array processing

Suppose there are two uncorrelated (hence, independent) signals so that $\mathbf{R}_s = \mathrm{diag}(\sigma^2_{\mathrm{S1}}, \sigma^2_{\mathrm{S2}})$. 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 σ^2_{S1} and σ^2_{S2}, located at θ_1 and θ_2, respectively. The signal-plus-noise covariance matrix is given by

$$\mathbf{R} = \sigma^2_{\mathrm{S1}}\mathbf{v}_1\mathbf{v}_1' + \sigma^2_{\mathrm{S2}}\mathbf{v}_2\mathbf{v}_2' + \mathbf{\Sigma} \tag{16}$$

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

$$\mathbf{R}_\Sigma = \mathbf{\Sigma}^{-1}\mathbf{R} = \sigma^2_{\mathrm{S1}}\mathbf{\Sigma}^{-1}\mathbf{v}_1\mathbf{v}_1' + \mathbf{\Sigma}^{-1}\sigma^2_{\mathrm{S2}}\mathbf{v}_2\mathbf{v}_2' + \mathbf{I}$$

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

$$\lambda_1 = 1 + \frac{\left(\sigma^2_{\mathrm{S1}} \parallel \mathbf{u}_1 \parallel^2 + \sigma^2_{\mathrm{S2}} \parallel \mathbf{u}_2 \parallel^2\right)}{2} + \frac{\sqrt{\left(\sigma^2_{\mathrm{S1}} \parallel \mathbf{u}_1 \parallel^2 - \sigma^2_{\mathrm{S2}} \parallel \mathbf{u}_2 \parallel^2\right)^2 + 4\sigma^2_{\mathrm{S1}}\sigma^2_{\mathrm{S2}}|\langle \mathbf{u}_1, \mathbf{u}_2\rangle|^2}}{2} \tag{17a}$$

$$\lambda_2 = 1 + \frac{\left(\sigma^2_{\mathrm{S1}} \parallel \mathbf{u}_1 \parallel^2 + \sigma^2_{\mathrm{S2}} \parallel \mathbf{u}_2 \parallel^2\right)}{2} - \frac{\sqrt{\left(\sigma^2_{\mathrm{S1}} \parallel \mathbf{u}_1 \parallel^2 - \sigma^2_{\mathrm{S2}} \parallel \mathbf{u}_2 \parallel^2\right)^2 + 4\sigma^2_{\mathrm{S1}}\sigma^2_{\mathrm{S2}}|\langle \mathbf{u}_1, \mathbf{u}_2\rangle|^2}}{2} \tag{17b}$$

respectively, where $\mathbf{u}_1 := \mathbf{\Sigma}^{-1/2}\mathbf{v}_1$ and $\mathbf{u}_2 := \mathbf{\Sigma}^{-1/2}\mathbf{v}_2$. Applying the result in Theorem 3.2 allows us to express the effective number of signals as

$$k_{\mathrm{eff}} = \begin{cases} 2 & \text{if } \ \mathrm{T}\left(\frac{n}{m}, \frac{n}{N}\right) < \lambda_2 \\ 1 & \text{if } \ \lambda_2 \leq \mathrm{T}\left(\frac{n}{m}, \frac{n}{N}\right) < \lambda_1 \\ 0 & \text{if } \ \lambda_1 \leq \mathrm{T}\left(\frac{n}{m}, \frac{n}{N}\right). \end{cases} \tag{18}$$

Equation (18) 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 $\mathbf{v}_2$. 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.

C. 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 [26], [27], then the conditional independence property of GMRF's [28] 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 [29]. Generalized eigenvalue based techniques for clustering have been investigated in [30], [31]. Our results might provide insight when spectral clustering algorithms are likely to fail. In particular, we note that the results of Theorem 3.2 hold even in situations where the data is not Gaussian (see Theorem 6.5) as is commonly assumed in machine learning applications.

IV. An algorithm for reliable detection of signals in noise

In [10], Johnstone proves that in the signal-free case, 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 Tracy-Widom law [32]–[34]. In the setting where there are signals present, we expect, after appropriate centering and scaling, the distribution of the signal eigenvalues of $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Sigma}}}$ above the detectability threshold will obey a Gaussian law whereas those below the detectability threshold will obey the Tracy-Widom law as in the signal-free case. An analogous results for the signal bearing eigenvalues of $\widehat{\mathbf{R}}_{\mathbf{\Sigma}}$ was proved by Baik et al [23] and El Karoui [35]. Numerical investigations for (see Figure 3) corroborate the accuracy of our asymptotic predictions and form the basis of Algorithm 1 presented below for estimating the number of signals at (asymptotic) significance level α. Theoretical support for this observation remains incomplete.

Algorithm 1

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

Figure 4 illustrates the accuracy of the predicted statistical limit and the ability of the proposed algorithm to reliably detect the presence of the signal at this limit. Figure 5 highlights how far away from this limit the classical algorithm of Zhao et al [15] is for various choices of the penalty functions considered in [36].

In the special setting where the noise covariance matrix is known apriori, the results of Baik et al [23], El Karoui [35] and Ma [37] form the basis of Algorithm 2 presented below for estimating the number of signals at (asymptotic) significance level α.

We compare our algorithm to the classical information theoretic estimator proposed by Zhao et al in [15]. Define $\kappa_{\hat{\lambda}} = \{\#\hat{\lambda}_j > 1\}$. Then for a choice of the penalty function C_{m+N} that satisfies the conditions $\lim_{m+N\to\infty} C_{m+N}/(m+N) \to 0$ and $\lim_{m+N\to\infty} C_{m+N}/\log\log(m+n) \to \infty$, the algorithm follows:

$$\widehat{k} = \min(I(0, C_{m+N}), I(1, C_{m+N}), \ldots, I(n, C_{m+N})), \tag{19a}$$

where

$$I(q, C_{m+N}) = -\sum_{j=1+\min(q,\kappa_{\hat{\lambda}})}^{n} \frac{n}{2}\left[\frac{m}{m+N}\log\widehat{\lambda}_j - \log\left(\frac{N}{m+N} + \frac{m}{m+N}\widehat{\lambda}_j\right)\right] + \frac{C_{m+N}}{2}q(2n-q+1), \tag{19b}$$

$$\widehat{k} = \min\{I(0, C_{m+N}), I(1, C_{m+N}), \ldots, I(n, C_{m+N})\}. \tag{19c}$$

Scenario \ Arguments	$\mu_{\{\cdot\}}[n,m,N]$	$\sigma_{\{\cdot\}}[n,m,N]$	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) Algorithm 1

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

(b) Algorithm 2

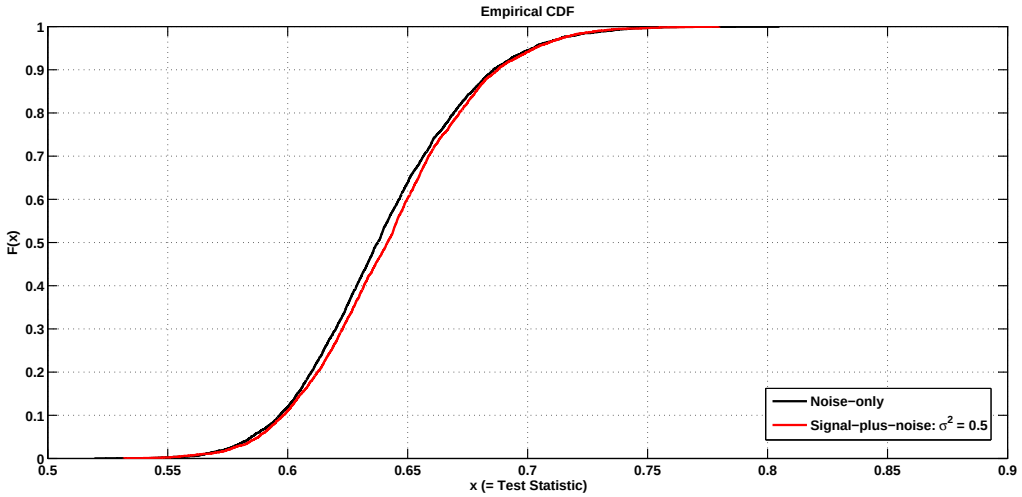
TABLE I: Parameters for signal detection algorithms.

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

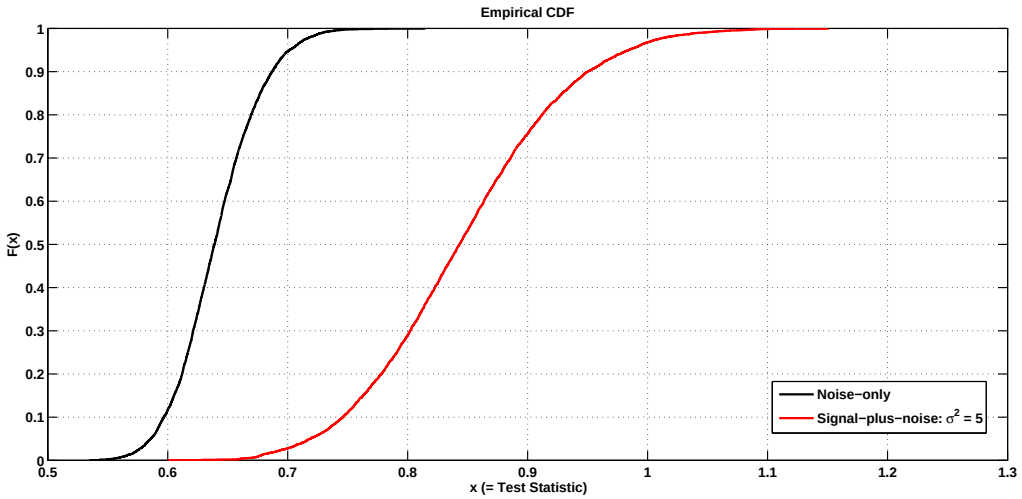
TABLE II: The third and fourth column show the percentiles of the Tracy-Widom real and complex distribution respectively corresponding to fractions in the second column. The percentiles were computed in MATLAB using software provided by Folkmar Bornemann for the efficient evaluation of the real and complex Tracy-Widom distribution functions $F_{\{\mathbb{R},\mathbb{C}\}}^{TW}(x)$. The percentiles are computed using the `fzero` command in MATLAB . The accuracy of the computed percentiles is about $\pm 5 \times 10^{-15}$ in absolute error terms.

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 $\tau_\alpha := TW_{\{\mathbb{R},\mathbb{C}\}}^{-1}(1-\alpha)$ from Table II
3. Set k = 0
4. Compute $\mu_{\{\mathbb{R},\mathbb{C}\}}[n-k, m]$ and $\sigma_{\{\mathbb{R},\mathbb{C}\}}[n-k, m]$ from Table I(b)
5. Is $\dfrac{m\hat{\lambda}_{k+1} - \mu_{\{\mathbb{R},\mathbb{C}\}}[n-k, m]}{\sigma_{\{\mathbb{R},\mathbb{C}\}}[n-k, m]} \geq \tau_\alpha$?
6. If yes, then go to step 9
7. Otherwise, increment k.
8. If $k < \min(n, m)$, go to step 3. Else go to step 9.
9. Return $\widehat{k} = k$



(a) Here $\sigma^2 = 0.5$, so that $\lambda_1 = 1 + \sigma^2 = 1.5 < \mathrm{T}(320160, 320960) = 3.4365$.



(b) Here $\sigma^2 = 5$, so that $\lambda_1 = 1 + \sigma^2 = 6 > \mathrm{T}(\frac{320}{160}, \frac{320}{960}) = 3.4365$

Fig. 3: In (a), for the setting described in Theorem 3.2 we set $n = 320$, $m = 160$, $N = 960$, $\sigma^2 = 0.5$, and w.l.o.g. $\mathbf{\Sigma} = \mathbf{I}$, $\mathbf{R} = \mathrm{diag}(\lambda_1 = 1 + \sigma^2, 1, \ldots, 1)$ and compare the the empirical cdf of the largest eigenvalue of $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Sigma}}}$ with the largest eigenvalue of $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Sigma}}}$ with $\mathbf{R} = \mathbf{I}$, *i.e.*, in the noise-only case, over 1000 Monte-Carlo trials. In (b), we plot the empirical cdf but now with $\sigma^2 = 5$.

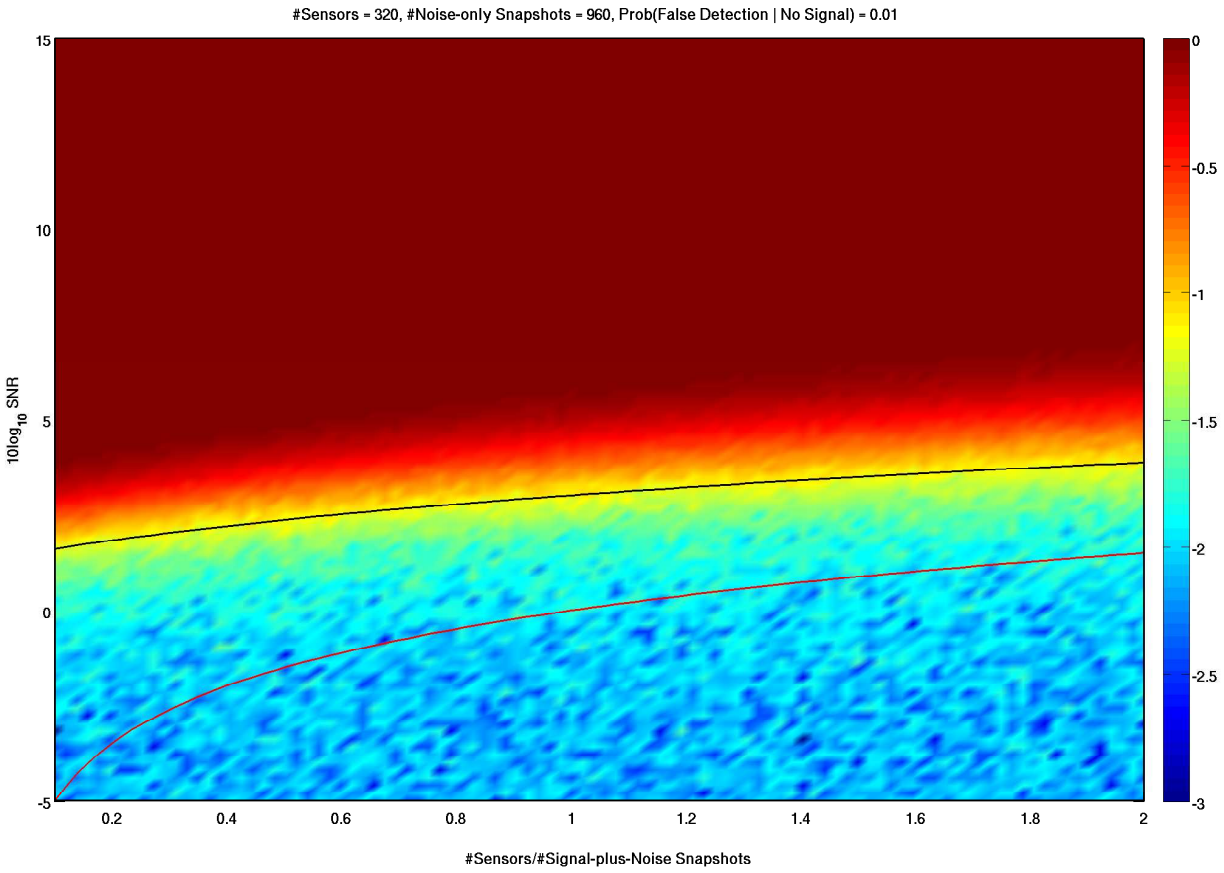
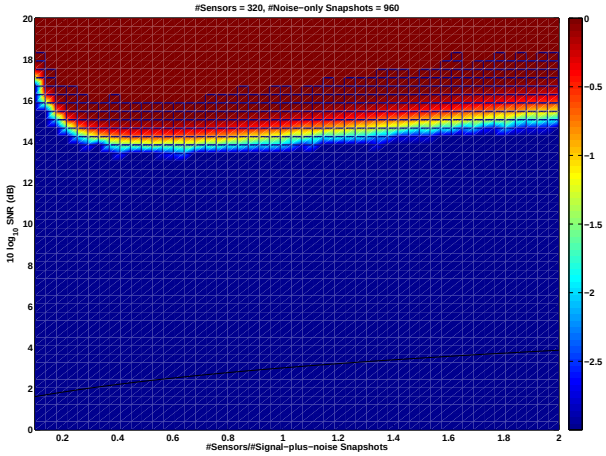
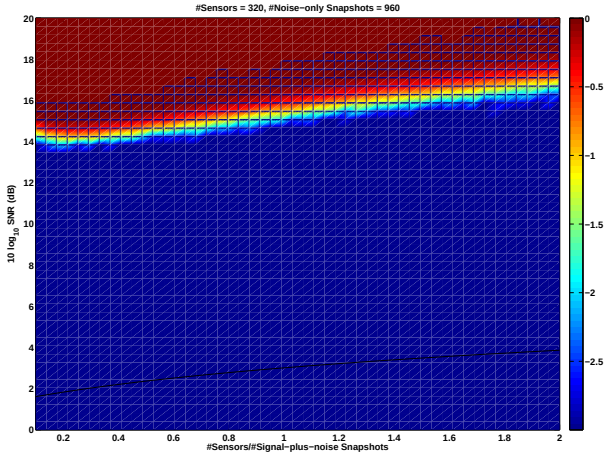


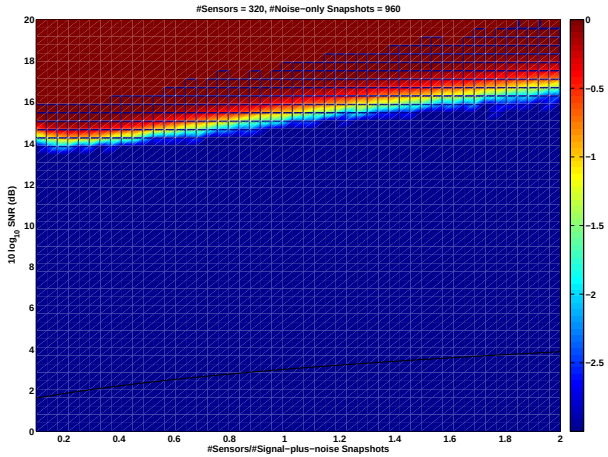
Fig. 4: A heat map of the log probability of signal detection using Algorithm 1 in Section IV, with the significance level α set at 0.01, in (eigen) SNR versus number of sensors to number of signal-plus-noise snapshots phase space. In this example, for the setting described in Theorem 3.2 we set $n = 320$, $N = 960$ and w.l.o.g. $\mathbf{\Sigma} = \mathbf{I}$, $\mathbf{R} = \text{diag}(\lambda_1 = 1 + \text{SNR}, 1, \ldots, 1)$ and evaluated Prob($\hat{k} = 1$) over 1000 Monte-Carlo trials and a grid of 100 equally spaced points in the -5 dB 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.



(a) $C_{m+N} = 0.65\sqrt{m+N}/\log(m+N)$



(b) $C_{m+N} = 0.5\log(m+N)$



(c) $C_{m+N} = 0.5\log(\frac{m+N}{2})$

Fig. 5: A heat map of the log probability of signal detection using the algorithm (19) in Section IV, 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 are in the same setting as in Figure 4 except that we plot Prob($\hat{k} \geq 1$) over 1000 Monte-Carlo trials and a grid of 50 equally spaced points in the 0 dB 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 black line demarcates the theoretically predicted threshold as in Figure 4.

V. Conclusion

Figure 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 Figure 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.

Acknowledgements

Raj Rao was supported by an Office of Naval Research Post-Doctoral Fellowship Award under grant N00014-07-1-0269. Jack Silverstein was supported by the U.S. Army Research Office under Grant W911NF-05-1-0244. R. R. thanks Arthur Baggeroer for encouragement and invaluable feedback. This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. 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.

We thank Alan Edelman for his encouragement and support, Interactive Supercomputing, Inc. for providing access to the Star-P parallel computing software and Sudarshan Raghunathan of Interactive Supercomputing, Inc. for his patience and support in answering our multitude of Star-P programming queries. We remain grateful to Al Davis and Chris 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 Figures 4 and Figure 5 in 4 days! Without their gracious help, the computation would have taken 5 months on the latest single processor laptop. We thank Folkmar Bornemann for providing the MATLAB code for computing the percentiles in Table II.

VI. Appendix

A. Mathematical preliminaries

Let for $i, j = 1, 2, \ldots$, X_{ij} be a collection of complex valued i.i.d. random variables with $\mathsf{E}X_{11} = 0$ and $\mathsf{E}|X_{11}|^2 = 1$. 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 [38] 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

$$F^A(x) \equiv (1/n)(\text{number of eigenvalues of A} \leq x).$$

The limit theorem is expressed in terms of the *Stieltjes transform* of the limiting e.d.f. of the F^{B_n}'s, where for any distribution function (d.f.) G its Stieltjes transform, 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 Stieltjes transforms, due to the inversion formula

$$G(b) - G(a) = \lim_{v \to 0} \int_a^b \Im m_G(x + iv) dx,$$

for a, b continuity points of G.

The limit theorem allows the $\mathbf{T}_n$ to be random, only 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 Stieltjes transform $m = m_F(z)$, $z \in \mathbb{C}^+$ satisfying the equation

$$m = \int \frac{1}{t(1 - c - czm) - z} dH(t), \tag{20}$$

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

It is more convenient to work with the eigenvalues of the $m \times m$ matrix $(1/m)\mathbf{X}_n^{'}\mathbf{T}_n\mathbf{X}_n$, whose eigenvalues differ from those of $\mathbf{B}_n$ by $|n - m|$ zero eigenvalues. Indeed, with I_A denoting the indicator function on the set A we have the exact relationship

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

almost surely, implying

$$m_{F^{c,H}}(z) = -(1 - c)/z + cm_F(z). \tag{21}$$

Upon substituting $m_{F^{c,H}}$ into (20) we find that for $z \in \mathbb{C}^+$ $m = m_{F^{c,H}}(z)$ solves the equation

$$z = -\frac{1}{m} + c \int \frac{\lambda}{1 + \lambda m} dH(\lambda), \tag{22}$$

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

Qualitative properties of $F^{c,H}$ have been obtained in [39], most notably the fact that on $(0, \infty)$ $F^{c,H}$ has a continuous derivative. The paper [39] also shows how intervals outside the support of $F^{c,H}$ can be determined from the graph of (22) 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 (22) with $m \in \mathbb{R}$. Intuitively, on $S'_{F^{c,H}}$ $m_{F^{c,H}}$ 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 stated in the following.

Lemma 6.1 (Theorems 4.1, 4.2 of [39]): If $x \in S'_{F^{c,H}}$, then $m = m_{F^{c,H}}$ satisfies (1) $m \in \mathbb{R}\backslash\{0\}$, (2) $-m^{-1} \in S'_H$, and (3) $\frac{d}{dm} x_{c,H}(m) > 0$. 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 6.2 (Theorem 4.3 of [39]): Suppose each m contained in the interval $[m_1, m_2]$ satisfies (1) and (2) of Lemma 6.1, and $\frac{d}{dm} x_{c,H}(m_i) \geq 0$ for $i = 1, 2$. Then $\frac{d}{dm} x_{c,H}(m) > 0$ for all $m \in (m_1, m_2)$.

Limiting eigenvalue mass at zero is also derived in [39]. It is shown that

$$F(0) = \begin{cases} H(0), & c(1 - H(0)) \leq 1, \\ 1 - c^{-1}, & c(1 - H(0)) > 1. \end{cases} \tag{23}$$

B. Support of eigenvalues

Since the convergence in distribution of F^{B_n} only addresses how proportions of eigenvalues behave, understanding the possible appearance or non-appearance of eigenvalues in $S'_{F^{c,H}}$ requires further work.

The question of the behavior of the largest and smallest eigenvalues when $\mathbf{T}_n = \mathbf{I}$ has been answered by Yin, Bai, and Krishnaiah in [40], and Bai and Yin in [41], respectively, under the additional assumption $\mathsf{E}|\mathbf{X}_{11}|^4 < \infty$: the largest eigenvalue and $\min(n, m)^{\text{th}}$ 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 F when $\mathbf{T}_n = \mathbf{I}$ will be given later.

For general $\mathbf{T}_n$, restricted to being bounded in spectral norm, the non-appearance of eigenvalues in $S'_{F^{c,H}}$ has been proven by Bai and Silverstein in [42]. Moreover, the separation of eigenvalues across intervals in $S'_{F^{c,H}}$, mirrors exactly the separation of eigenvalues over corresponding intervals in S'_H [43]. The results are summarized below.

Theorem 6.1: Assume additionally $\mathsf{E}|\mathbf{X}_{11}|^4 < \infty$ 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 Stieltjes transform with inverse (22), where c, H are replace by c_n, H_n.

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

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

(i) 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 $n \to \infty$.

(ii) If $c(1 - H(0)) \leq 1$, 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

$$\lambda_{i_n}^{T_n} > -1/m_{F^{c,H}}(b) \quad \text{and} \quad \lambda_{i_n+1}^{T_n} < -1/m_{F^{c,H}}(a). \tag{24}$$

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

Proof: See proof of Theorems 1.1 in [42], [43]). ■

The behavior of the extreme eigenvalues of $(1/m)\mathbf{X}_n\mathbf{X}_n^*$ leads to the following corollary of Theorem 6.1.

Corollary 6.2: 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 $c \leq 1$ ($c > 1$) 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 6.1, Case (i) 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 6.1, Case (ii) of Theorem 6.1 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 6.1, must be due to the existence of $\lambda_{i_n}^{T_n}$, $\lambda_{i_n+1}^{T_n}$ satisfying (24).

Theorem 6.1 easily extends to random $\mathbf{T}_n$, independent of $\{\mathbf{X}_{ij} : i,j \geq 1\}$ with the aid of Tonelli's Theorem [44, pp. 234], provided the condition (*) on $[a,b]$ is strengthened to:

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

$$A = \{\lambda_{i_n}^{B_n} > b \text{ and } \lambda_{i_n+1}^{B_n} < a \text{ for all large } n\}.$$

Let $t \in T$ be an element of the event defined in (**). Then by Theorem 6.1 $I_A((t,x)) = 1$ for all x contained in a subset of X having probability 1. Therefore, by Tonelli's theorem

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

Consider now case (ii) of Theorem 6.1 in terms of the corresponding interval outside the support of H and the H_n's. By Lemma 6.1 and condition (*), we have the existence of an $\epsilon > 0$ such that $0 \notin [m_{F^{c,H}}(a) - \epsilon, m_{F^{c,H}}(b) + \epsilon]$, and for all n large

$$\frac{d}{dm} x_{c_n,H_n}(m) = \frac{1}{m^2}\left(1 - c_n \int \frac{(\lambda m)^2}{(1+\lambda m)^2} dH_n(\lambda)\right) > 0, \quad m \in [m_{F^{c,H}}(a) - \epsilon, m_{F^{c,H}}(b) + \epsilon]. \tag{25}$$

Let $t_a = -1/m_{F^{c,H}}(a)$, $t_b = -1/m_{F^{c,H}}(b)$. Then by Lemma 6.1 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 (25) we have for all n large

$$c_n \int \frac{\lambda^2}{(\lambda - t)^2} dH_n(\lambda) < 1, \quad t \in [t_a - \epsilon', t_b + \epsilon']. \tag{26}$$

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 (26) 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 (ii) 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 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 6.1, with probability one, eigenvalues of the new $\mathbf{B}_n$ will appear in $S'_{F^{c,H}}$ for all n large.

Theorem 6.3 below shows this will happen when a "sprinkled", or "spiked" eigenvalue lies in (t_a, t_b). Theorem 6.4 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 (ii) of Theorem 6.1.

Theorem 6.3, below, allows the number of spiked eigenvalues to grow with n, provided it remains $o(n)$.

Theorem 6.3: Assume in additon to the assumptions in Theorem 6.1 on the $\mathbf{X}_{ij}$ and $\mathbf{T}_n$:

(a) 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 $\mathbf{T}_n$.

(b) 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 $t = t_a, t_b$.

(c) $t' \in (t_a, t_b)$.

Suppose $\lambda_{i_n}^{T_n}, \ldots, \lambda_{i_n+\ell-1}^{T_n}$ are the eigenvalues stated in (a). Then, with probability one

$$\lim_{n\to\infty} \lambda_{i_n}^{B_n} = \cdots = \lim_{n\to\infty} \lambda_{i_n+\ell-1}^{B_n} = t' \left(1 + c \int \frac{\lambda}{t' - \lambda} dH(\lambda)\right). \tag{27}$$

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

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

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 (b) we have $\frac{d}{dm} x_{c,H}(m) \geq 0$ for $m = -1/t_a, -1/t_b$ (recall (25),(26)).

By Lemma 6.2 we therefore have $\frac{d}{dm} x_{c,H}(m) > 0$ 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 $\frac{d}{dm} x_{c_n,\hat{H}_n}(m) \geq \delta$ 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) $\frac{d}{dm} x_{c_n,H_n}(m) > 0$ for all m contained in these two intervals.

Therefore, by Lemma 6.1, 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 $a_L = x_{c,H}(-1/t' - \epsilon - \frac{2}{3}\delta')$, $b_L = x_{c,H}(-1/t' - \epsilon - \frac{1}{3}\delta')$, $a_R = x_{c,H}(-1/t' + \epsilon + \frac{1}{3}\delta')$, and $b_R = x_{c,H}(-1/t' + \epsilon + \frac{2}{3}\delta')$. Then for all n large

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

and

$$[a_R, b_R] \subset (x_{c,H}(-1/t' + \epsilon + \tfrac{1}{6}\delta'), x_{c,H}(-1/t' + \epsilon + \tfrac{5}{6}\delta')) \subset [x_{c_n,H_n}(-1/t' + \epsilon), x_{c_n,H_n}(-1/t' + \epsilon + \delta')]. \quad (29)$$

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 (ii) of Theorem 6.1 applies and we have

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

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

$$\lim_{n\to\infty} \lambda_{i_n}^{B_n} = \lim_{n\to\infty} \lambda_{i_n+\ell-1}^{B_n} = x_{c,H}(-1/t') = (27).$$

■

Theorem 6.4: Assume, besides the assumptions in Theorem 6.1, 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 6.1). Then $0 < t_a < t' < t_b$, implying (c) of Theorem 6.3. 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 (b) of Theorem 6.3 is true. If ℓ remains bounded, then (a) of Theorem 6.3 also holds.

Proof: By Lemma 6.1, $[t_a, t_b] \in S'_H$, 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 (i) of Theorem 6.1 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 $\frac{d}{dm}x_{c_n,\hat{H}_n}(m) \to \frac{d}{dm}x_{c,H}(m)$ for all $m \in [m_{c,H}(a) - \epsilon, m_{c,H}(b) + \epsilon]$, implying for all n large $\frac{d}{dm}x_{c_n,\hat{H}_n}(m) > 0$ for all $m \in [m_{c,H}(a), m_{c,H}(b)]$. Therefore (b) is true

We assume now that ℓ is bounded. Suppose (a) 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 6.1, (27), 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 (a) holds. ■

D. Behavior of extreme eigenvalues

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 $g_n(t) \equiv c_n \int \frac{\lambda^2}{(\lambda-t)^2} d\hat{H}_n(t)$ 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 (27) 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 (27) 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 $g(\tau) = c \int \frac{\lambda^2}{(\lambda-t)^2} dH(t) = 1$, 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 (b) and (c) of Theorem 6.3 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 6.4, 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 $[\lambda_H^{\max},\infty)$), $\lambda_{i_n}^{B_n},\ldots,\lambda_{i_n+\ell-1}^{T_n}$ 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 6.1 Tonelli's Theorem can easily be applied to establish equivalent results when $\mathbf{T}_n$'s are random and independent of $\mathbf{X}$.

E. The eigenvalues of the multivariate F matrix

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 $\mathsf{E}\mathbf{Y}_{11}=0$, $\mathsf{E}|\mathbf{Y}_{11}|=1$, $\mathsf{E}|\mathbf{Y}_{11}|^4<\infty$, 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 [41] 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

$$f_{c,c_1}(x)=\frac{(1-c_1)\sqrt{(x-b_1)(b_2-x)}}{2\pi x(xc_1+c)}\quad b_1<x<b_2,$$

where

$$b_1=\left(\frac{1-\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2,\quad b_2=\left(\frac{1+\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2.$$

When $c\in(0,1]$, there is no mass at 0, whereas for $c>1$ F has mass $(1-(1/c))$ at 0 [14].

We are interested in the effect on spikes on the right side of the support of the H_n.

Because of the corollary to Theorem 6.1, we know $\lambda_1^{B_n}\to b_2$ a.s. as $n\to\infty$. We proceed in computing the function

$$g(t)=c\int\frac{\lambda^2}{(\lambda-t)^2}dH(t).$$

We will see that it is unnecessary to compute the limiting e.d.f. of $\mathbf{T}_n$. It suffices to know the limiting Stieltjes transform 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

$$g(t)=c\int\frac{(1/\lambda)^2}{(t-1/\lambda)^2}dH_1(\lambda)=c\int\frac{1}{(\lambda t-1)^2}dH_1(\lambda)=ct^{-2}\int\frac{1}{(\lambda-(1/t))^2}dH_1(\lambda)$$
$$=t^{-2}\frac{d}{dx}m_{H_1}(x)\Big|_{x=(1/t)}.$$

We use (22) to find $m_{F^{c_1,I_{[1,\infty)}}}$:

$$z=-\frac{1}{m}+c_1\frac{1}{1+m}\quad\Leftrightarrow\quad zm^2+(z+1-c_1)m+1=0.$$

$$\Leftrightarrow\quad m=\frac{-z-1+c_1\pm\sqrt{(z+1-c_1)^2-4z}}{2z}$$

(the sign depending on with branch of the square root is taken).

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

From the identity in (21) we find that

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

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

$$m_{H_1}(x) = \frac{-x+1-c_1-\sqrt{(x-(1-\sqrt{c_1})^2)(x-(1+\sqrt{c_1})^2)}}{2c_1 x}, \tag{30}$$

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

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

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

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

We see that the equation $g(t) = 1$ leads to the following quadratic equation in t:

$$(1-c_1)^2\alpha t^2 - 2(1+c_1)\alpha t + \alpha - c^2 = 0, \quad \text{where } \alpha = c_1 + c - cc_1,$$

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

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

Reducing further we find the threshold, τ, to be

$$\tau = \frac{(1+c_1)\alpha + \sqrt{\alpha}\sqrt{4\alpha - c_1 + (1-c_1)^2c^2}}{(1-c_1)^2\alpha} = \frac{(1+c_1)\alpha + \sqrt{\alpha}(2c_1 + c(1-c_1))}{(1-c_1)^2\alpha}. \tag{31}$$

We now compute the right hand side of (27). We have for $t' \geq \tau$

$$\begin{aligned} t'\left(1 + c\int \frac{\lambda}{t'-\lambda} dH(\lambda)\right) &= t'\left(1 + c\int \frac{1/\lambda}{t'-1/\lambda} dH_1(\lambda)\right) = t'(1 + ct'^{-1} m_{H_1}(1/t')) \\ &= \frac{t'(2c_1 + c(1-c_1)) - c - c\sqrt{(1-t'(1-\sqrt{c_1})^2)(1-t'(1+\sqrt{c_1})^2)}}{2c_1} \equiv \lambda(t'). \end{aligned} \tag{32}$$

A straightforward (but tedious) calculation will yield $\lambda(\tau) = b_2$.

Using the results from the previous section, we have proved the following:

Theorem 6.5: Assume in addition to the assumptions in Theorem 6.1 on the $\mathbf{X}_{ij}$

(a) 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 $F^{((1/N)Y_n Y_n^*)^{-1}}$, defined above.

(b) 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 $\mathbf{T}_n$.

(c) 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 (31).

Suppose $\lambda_{i_n}^{T_n}, \ldots, \lambda_{i_n+\ell-1}^{T_n}$ are the eigenvalues stated in (b) of Theorem 6.3. Then, with the function $\lambda(\cdot)$ defined in (32), with probability one

$$\lim_{n\to\infty} \lambda_{i_n}^{B_n} = \cdots = \lim_{n\to\infty} \lambda_{i_n+\ell-1}^{B_n} = \begin{cases} \lambda(t'), & \text{if } t' > \tau \\ b_2, & \text{if } t' \in (\limsup_n \hat{\lambda}_n^{\max}, \tau]. \end{cases}$$

Note: From Theorem 6.3, when $t' > \tau$ the result can allow $\ell = o(n)$.

REFERENCES

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

[2] E. Maris, "A resampling method for estimating the signal subspace of spatio-temporal EEG/MEG data," *Biomedical Engineering, IEEE Transactions on*, 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," *Biomedical Engineering, IEEE Transactions on*, vol. 44, no. 9, pp. 839–847, 1997.

[4] ——, "MEG spatio-temporal analysis using a covariance matrix calculated from nonaveraged multiple-epoch data," *Biomedical Engineering, IEEE Transactions on*, vol. 46, no. 5, pp. 515–521, 1999.

[5] H. L. V. Trees, *Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing*. new York: John wiley and Sons, Inc., 2002.

[6] 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, 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] 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, August 2001.

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

[10] I. M. Johnstone, "Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy–Widom limits and rates of convergence," *Annals of Statistics*, vol. 36, no. 6, pp. 2638–2716, 2008.

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

[12] R. J. Muirhead, *Aspects of multivariate statistical theory*. New York: John Wiley & Sons Inc., 1982, wiley Series in Probability and Mathematical Statistics.

[13] T. W. Anderson, *An introduction to multivariate statistical analysis*, 3rd ed., ser. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2003.

[14] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate F matrix," *SIAM Journal on Math. Anal.*, vol. 16, no. 3, pp. 641–646, 1985.

[15] 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.

[16] K.-W. Tam and Y. Wu, "On rates of convergence of general information criteria in signal processing when the noise covariance matrix is arbitrary," *Information Theory, IEEE Transactions on*, vol. 37, no. 6, pp. 1667–1671, Nov 1991.

[17] Z. Zhu, S. Haykin, and X. Huang, "Estimating the number of signals using reference noise samples," *IEEE Trans. on Aero. and Elec. Systems*, vol. 27, no. 3, pp. 575–579, May 1991.

[18] P. Stoica and M. Cedervall, "Detection tests for array procesing in unknown correlated noise fields," *IEEE Trans. Signal Process.*, vol. 45, pp. 2351–2362, September 1997.

[19] G. Xu, R. H. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Process.*, vol. SP-42, pp. 102–112, January 1994.

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

[21] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. Beijing: Science Press, 2006.

[22] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues in certain sets of random matrices," *Mat. Sb. (N.S.)*, vol. 72 (114), pp. 507–536, 1967.

[23] 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.

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

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

[26] A. Anandkumar, L. Tong, and A. Swami, "Detection of Gauss-Markov Random Fields with nearest-neighbor dependency," *Information Theory, IEEE Transactions on*, vol. 55, no. 2, pp. 816–827, Feb. 2009.

[27] Y. Sung, L. Tong, and H. Poor, "Neyman–Pearson Detection of Gauss–Markov Signals in Noise: Closed-Form Error Exponent and Properties," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1354–1365, 2006.

[28] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, 2005.

[29] S. Vempala, "Spectral Algorithms for Learning and Clustering," *LECTURE NOTES IN COMPUTER SCIENCE*, vol. 4539, p. 3, 2007.

[30] O. Mangasarian and E. Wild, "Multisurface Proximal Support Vector Machine Classification via Generalized Eigenvalues," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 69–74, 2006.

[31] M. Guarracino, C. Cifarelli, O. Seref, and P. Pardalos, "A classification method based on generalized eigenvalue problems," *Optimization Methods and Software*, vol. 22, no. 1, pp. 73–81, 2007.

[32] C. Tracy and H. Widom, "Level-spacing distribution and Airy kernel," *Communications in Mathematical Physics*, vol. 159, pp. 151–174, 1994.

[33] ——, "On orthogonal and symplectic matrix ensembles," *Communications in Mathematical Physics*, vol. 177, pp. 727–754, 1996.

[34] I. M. Johnstone, "High dimensional statistical inference and random matrices," in *International Congress of Mathematicians. Vol. I*. Eur. Math. Soc., Zürich, 2007, pp. 307–333.
[35] 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.
[36] Y. Wu and K.-W. Tam, "On determination of the number of signals in spatially correlated noise," *Signal Processing, IEEE Transactions on*, vol. 46, no. 11, pp. 3023–3029, Nov 1998.
[37] Z. Ma, "Accuracy of the Tracy–Widom limit for the largest eigenvalue in white Wishart matrices," 2008, `http://arxiv.org/abs/0810.1329`.
[38] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. of Multivariate Anal.*, vol. 55(2), pp. 331–339, 1995.
[39] 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.
[40] 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.
[41] 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.
[42] 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.
[43] ——, "Exact separation of eigenvalues of large dimensional sample covariance matrices," *Ann. Probab.*, vol. 27, No. 3, pp. 1536–1555, 1999.
[44] P. Billingsley, *Probability and measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc., 1995, a Wiley-Interscience Publication.