Statistical Inference in Large Antenna Arrays under Unknown Noise Pattern

Julia Vinogradova, Romain Couillet, and Walid Hachem

***Abstract*—In this article, a general information-plus-noise transmission model is assumed, the receiver end of which is composed of a large number of sensors and is unaware of the noise pattern. For this model, and under reasonable assumptions, a set of results is provided for the receiver to perform statistical eigen-inference on the information part. In particular, we introduce new methods for the detection, counting, and the power and subspace estimation of multiple sources composing the information part of the transmission. The theoretical performance of some of these techniques is also discussed. An exemplary application of these methods to array processing is then studied in greater detail, leading in particular to a novel MUSIC-like algorithm assuming unknown noise covariance.**

***Index Terms*—Random matrix theory, sensor arrays, correlated noise, source detection, power estimation, MUSIC algorithm.**

I. Introduction

A. Motivation

Consider the very general information-plus-noise transmission model with multivariate output $y_t \in \mathbb{C}^N$ at time t

$$y_t = Hx_t + v_t \tag{1}$$

where $x_t \in \mathbb{C}^K$ is the vector of transmitted symbols at time t, $H \in \mathbb{C}^{N\times K}$ is the linear communication medium, and $v_t \in \mathbb{C}^N$ the noise experienced by the receiver at time t.

Array processing consists in a set of tools to perform statistical inference on the information part composing y_t. The first tool is the mere detection of this information (called then a signal source), that is the question whether $K > 0$. Once source signals are detected, the next operation consists in the evaluation of their number, *i.e.* estimating K. When the existence of these sources is guaranteed, several of their parameters can then be retrieved. One of these parameters is the transmission power of the source, or alternatively, the distance from the source to the receiver. Denoting $H = [h_1,\ldots,h_K]$, it is also of interest to retrieve information from the individual h_k vectors. In wireless communications, these represent channel beams which the receiver may want to identify in order to decode the entries of x_t. In array processing, they stand for steering vectors parameterized by the angle-of-arrival of the source signals.

In order to perform these tasks, one assumes the observation of T (non-necessarily independent) samples $y_1,\ldots,y_T$ of the process y_t. Denoting $Y_T = T^{-1/2}[y_1,\ldots,y_T]$, the first mentioned estimators are often based on the eigenvalues of $Y_TY_T^{\mathsf{H}}$. When it comes to vector identification, the interest is rather on the eigenvectors of $Y_TY_T^{\mathsf{H}}$. The standard eigen-inference approaches in the literature often rely on two strong assumptions: (i) T is large compared to N and (ii) the statistics of v_t are partially or perfectly known due to independent (information-free) observations of the process v_t. This article revisits these methods by proposing alternative algorithms to perform eigen-inference for the model (1) accounting for the aforementioned limitations (i) and (ii).

B. Literature review

Assuming $T \to \infty$, N fixed, and v_t white Gaussian with known variance, the energy detection procedure [1] allows for the detection of signal sources by evaluating the total received power which is compared to a threshold that ensures a maximum false alarm rate. If the signal structure is known, the parameters composing H can be recovered from the eigenvalues and eigenvectors of $\mathrm{E}[y_ty_t^{\mathsf{H}}]$, which can be estimated through the sample covariance matrix $Y_TY_T^{\mathsf{H}}$, $Y_T = T^{-1/2}[y_1,\ldots,y_T] \in \mathbb{C}^{N\times T}$. To estimate the number of sources K, the Akaike information criterion (AIC) [2] and the minimum description length (MDL) [3], [4] were historically proposed, which rely on functions of the eigenvalues of $Y_TY_T^{\mathsf{H}}$. The MDL is T-consistent while the AIK tends to overestimate the number of sources as $T \to \infty$. In terms of power estimation, since $Y_TY_T^{\mathsf{H}} \xrightarrow{\text{a.s.}} \mathrm{E}[y_ty_t^{\mathsf{H}}]$, a T-consistent estimate of the powers is easily obtained by mapping the eigenvalues of $Y_TY_T^{\mathsf{H}}$ to those of $\mathrm{E}[y_ty_t^{\mathsf{H}}]$. When the vectors $h_k = h(\theta_k)$ are steering vectors and the one aims at retrieving θ_k for $k = 1,\ldots,K$, the multiple signal classification (MUSIC) algorithm [5] allows for a T-consistent estimation of the angles $\theta_1,\ldots,\theta_K$ by determining the local maxima of the quadratic forms $\gamma(\theta) = h(\theta)^{\mathsf{H}}\Pi h(\theta)$ where Π is a projector on the eigenspace of $\mathrm{E}[y_ty_t^{\mathsf{H}}]$ corresponding to its K largest eigenvalues (assuming $\|h(\theta)\|$ constant with θ).

Due to the increase of the antenna array sizes and the need for faster detection and estimation dynamics, modern antenna array technologies have to deal with the scenario where the condition $T \gg N$ is no longer met. Under this condition, since $Y_TY_T^{\mathsf{H}}$ becomes a poor estimator for $\mathrm{E}[y_ty_t^{\mathsf{H}}]$, most of the above techniques collapse. New methods, based on the field of large dimensional random matrix theory, have therefore emerged, which assume that both N and T are large and that the ratio N/T is non-trivial. The AIC and MDL algorithms are in particular improved in [6] using better estimators for

The first and third authors are with CNRS LTCI; Telecom ParisTech, France (e-mail: julia.vinogradova@telecom-paristech.fr; walid.hachem@telecom-paristech.fr). The second author is with Supelec, France (e-mail: romain.couillet@supelec.fr). This work is supported in part by the French Ile-de-France region, DIM LSC fund, Digiteo project DESIR, and in part by the ANR DIONISOS.

functionals of the eigenvalues of $\mathrm{E}[y_t y_t^{\mathsf{H}}]$. In terms of power estimation, N, T-consistent techniques were proposed in [7]. The MUSIC algorithm was improved on the same grounds in [8] into the so-called G-MUSIC estimator.

A second difficulty faced by antenna array technologies is that the interfering environment may be far from white Gaussian. The $v_1, \ldots, v_T$ may not be independent or the spatial correlation of v_t may not be white. When the noise is not white, the energy detection procedure is not valid as no false alarm threshold can be set. When the noise is close-to-white Gaussian with unknown variance, the generalized likelihood-ratio test (GLRT) [9] copes with the indetermination of the variance. Similar schemes are analyzed in the large N, T regime in [10], [11], [12], [13]. If the noise is not white, it is difficult to derive any test for detection. The power and direction estimation techniques equally suffer from this indetermination, because too little is a priori known of the eigenstructure of $V_T V_T^{\mathsf{H}}$ with $V_T = T^{-1/2}[v_1, \ldots, v_T]$. To circumvent this issue, one generally assumes the existence of a sequence of T' pure-noise test samples which are used to "whiten" the observations. For T' large compared to N, after whitening, the noise becomes white Gaussian with unit variance, leading back to traditional schemes. For N, T' simultaneously large, the whitening procedure gives rise to a noise matrix of the F-matrix type [14], [15].

However, the requirement to possess observations purely composed of noise may be impractical in real systems. As such, in this article, we address the problems of detection, counting, and parameter estimation of multiple sources without resorting to a pre-whitening of the received data matrix Y_T. Since the problem may not be well-posed in its generality, we assume a set of reasonable conditions:

- $N, T \to \infty$, $N/T \to \mathbf{c} > 0$, K constant. This allows for $Y_T Y_T^{\mathsf{H}}$ to be seen as a small rank perturbation of $V_T V_T^{\mathsf{H}}$.
- $V_T = W_T R_T^{1/2}$ (*i.e.* white in space, correlated in time), where $W_T \in \mathbb{C}^{N \times T}$ is standard complex Gaussian and R_T is a deterministic *unknown* Hermitian nonnegative, or $V_T = R_T^{1/2} W_T$ (*i.e.* white in time, correlated in space).[1]
- As $N/T \to \mathbf{c}$, the eigenvalues of $V_T V_T^{\mathsf{H}}$ tend to cluster in a compact interval. This assumption is satisfied by most noise models used in practice, *e.g.* auto-regressive moving average (ARMA) noise processes (see Section III-B).
- The source signals in x_t are random, independent and identically distributed (i.i.d.), even though this assumption can be relaxed in many cases.

Under these assumptions, we show that a maximum of K isolated eigenvalues of $Y_T Y_T^{\mathsf{H}}$ can be found for all large N, T beyond the right edge of the limiting eigenvalue distribution support of $V_T V_T^{\mathsf{H}}$. This phenomenon is at the origin of the detection and estimation procedures developed in this paper. Precisely, we show that the isolated eigenvalues of $Y_T Y_T^{\mathsf{H}}$ can be uniquely mapped to individual signal sources. The presence of these eigenvalues will be used to detect signal sources as well as to estimate their number K while their values will be exploited to estimate the source powers. The associated eigenvectors will then be used to retrieve information on the vectors h_k.

The remainder of the article is structured as follows. In Section II, we introduce the system model and recall important results from the random matrix literature. In Section III, we introduce the source detector and parameter estimators for the generic model (1) and for a specific array processing scenario with an ARMA noise process. In Section IV, we study the second order statistics of some of these estimators. Simulations are then provided in Section V. The article is concluded by Section VI. Some technical lemmas are proved in the appendix.

Notations: The superscript $(\cdot)^{\mathsf{H}}$ is the Hermitian transpose of a matrix and $\|\cdot\|$ denotes the spectral norm. The symbols $\xrightarrow{\text{a.s.}}$, $\xrightarrow{\mathcal{P}}$, and $\xrightarrow{\mathcal{L}}$ stand respectively for the almost sure convergence, the convergence in probability, and the convergence in law, while "w.p. 1" means "with probability one". We denote by $\mathcal{N}(a, \sigma^2)$ the real Gaussian distribution with mean a and variance σ^2 and by $\mathcal{CN}(a, \sigma^2)$ the complex circular Gaussian distribution with mean a and variance σ^2. We denote by $\boldsymbol{\delta}_{k\ell}$ the Kronecker delta function ($= 1$ if $k = \ell$ and 0 otherwise) and by δ_x the Dirac measure at x.

II. Assumptions and known results

Consider a sequence of integers $N = N(T)$, $T = 1, 2, \ldots$ and matrices $Y_T = A_T + W_T R_T^{1/2} \in \mathbb{C}^{N \times T}$ where A_T stands for the signal matrix and $V_T = W_T R_T^{1/2}$ for the noise matrix.[2] We assume the following asymptotic regime:

Assumption 1. *As $T \to \infty$, $c_T \triangleq N/T \to \mathbf{c} > 0$.*

A. Hypotheses on the noise matrix

We first characterize the assumptions on $V_T \triangleq W_T R_T^{1/2}$.

Assumption 2. *$W_T = T^{-1/2}[w_{n,t}]_{n,t=1}^{N,T}$, with $(w_{n,t})_{n,t \geq 1}$ an infinite array of independent $\mathcal{CN}(0, 1)$ variables.*

Assumption 3. *$R_T \in \mathbb{C}^{T \times T}$ is Hermitian nonnegative with eigenvalues $\sigma_{1,T}^2, \ldots, \sigma_{T,T}^2$ satisfying:*

1) *With $\nu_T = T^{-1} \sum_{t=1}^{T} \delta_{\sigma_{t,T}^2}$, $\nu_T \xrightarrow{\mathcal{L}} \nu$, a probability measure with support $\mathrm{supp}(\nu) = [a_\nu, b_\nu] \subset \mathbb{R}_+ \triangleq [0, \infty)$. Moreover, $\nu(\{0\}) = 0$.*
2) *The distances from the $\sigma_{t,T}^2$ to $\mathrm{supp}(\nu)$ satisfy:*

$$\max_{t \in \{1, \ldots, T\}} \mathbf{d}\left(\sigma_{t,T}^2, \mathrm{supp}(\nu)\right) \xrightarrow[T \to \infty]{} 0.$$

Let $\lambda_{1,T} \geq \ldots \geq \lambda_{N,T}$ be the eigenvalues of $V_T V_T^{\mathsf{H}} = W_T R_T W_T^{\mathsf{H}}$ and let $\tau_T = N^{-1} \sum_{i=1}^{N} \delta_{\lambda_{i,T}}$ be its spectral measure. The asymptotic behavior of τ_T is of prime importance in this paper. We recall some well known results describing this behavior (see [16], [17] for Items 1–6, [18] for Item 4, and [19] for Item 5):

Theorem 1. *Under Assumptions 1–3, the following hold true:*

[1] Assuming the general correlated noise in both time and space would lead to too much indetermination and is so far too difficult to address.

[2] Up to studying Y_T^{H} instead of Y_T, the noise correlation can be either in time or in space.

1) For any $z \in \mathbb{C}_+ \triangleq \{z \in \mathbb{C}, \Im z > 0\}$, the equation

$$\mathbf{m} = \left(-z + \int \frac{t}{1+\mathbf{c}\mathbf{m}t}\nu(dt)\right)^{-1} \quad (2)$$

has a unique solution $\mathbf{m} \in \mathbb{C}_+$. The function $\mathbf{m}(z) = \mathbf{m}$ so defined on $\mathbb{C}_+$ is the Stieltjes transform (ST)[3] of a probability measure μ.

2) For every bounded and continuous real function f,

$$\int f(t)\tau_T(dt) \xrightarrow[T\to\infty]{a.s.} \int f(t)\mu(dt)$$

and therefore μ, defined by (2), is the limiting spectral measure of $V_T V_T^{\mathsf{H}}$.

3) The function

$$\tilde{\mathbf{m}}(z) = \int \frac{-1}{z(1+\mathbf{c}\mathbf{m}(z)t)}\nu(dt)$$

is defined on $\mathbb{C}_+$ and is the ST of the probability measure $\tilde{\mu} = \mathbf{c}\mu + (1-\mathbf{c})\delta_0$, limiting spectral measure of $V_T^{\mathsf{H}} V_T$. As such, $\tilde{\mathbf{m}}(z) = \mathbf{c}\mathbf{m}(z) - (1-\mathbf{c})/z$.

4) μ is of the form $\mu(dt) = \max(0, 1-\mathbf{c}^{-1})\delta_0 + f(t)dt$ where $f(t)$ is a continuous density on $(0,\infty)$. The support of $f(t)dt$ is a compact interval $[a,b] \subset \mathbb{R}_+$, and $f(t) > 0$ on (a,b).

5) For any interval $[x_1, x_2] \subset (0,a) \cup (b,\infty)$,

$$\sharp\{i : \lambda_{i,T} \in [x_1,x_2]\} = 0 \text{ w.p. } 1 \text{ for large } T.$$

6) The function $\underline{m}_T(x) = N^{-1}\sum_{n=1}^{N}(\lambda_{n,T} - x)^{-1}$ converges w.p. 1 to $\mathbf{m}(x)$, and uniformly so on the compact subsets of (b,∞).

A procedure for determining the interval $[a,b]$ from the knowledge of $\mathbf{c}$ and ν is provided in [18]. We are interested here in the determination of the upper bound b, to which $\lambda_{1,T}$ converges. This can be done with the help of the following proposition. Observe that $\mathbf{m}(z)$ can be extended to $\mathbb{C} - (\{0\} \cup [a,b])$ and that $\mathbf{m}(x) = \int (t-x)^{-1}\mu(dt)$, its restriction to $\mathbb{R}$, is negative and increases to zero on (b,∞). Recall that $\mathrm{supp}(\nu) = [a_\nu, b_\nu] \subset \mathbb{R}_+$.

Proposition 1 (see [18])**.** *The point b defined in Theorem 1-4) coincides with the infimum of the function*

$$\mathbf{x}(m) = -\frac{1}{m} + \int \frac{t}{1+\mathbf{c}mt}\,\nu(dt)$$

on the interval $(-(\mathbf{c}b_\nu)^{-1}, 0)$. On this interval, there is a unique m_b ($m_b < 0$) such that $\mathbf{x}(m) \to b$ as $m \downarrow m_b$. The restriction of $\mathbf{x}(m)$ to the interval $(m_b, 0)$ coincides with the inverse with respect to composition of the restriction of $\mathbf{m}(x)$ to (b,∞).

In order to easily characterize the value of b, it will be convenient to make an assumption on the measure ν which will not be restrictive in practice:

Assumption 4. *If $\nu(\{b_\nu\}) = 0$, then there exists $\varepsilon > 0$ and a function $f_\nu(t) \geq C(b_\nu - t)$ on $[b_\nu - \varepsilon, b_\nu]$ with $C > 0$ such that for any Borel set A of $[a_\nu, b_\nu]$,*

$$\nu(A \cap [b_\nu - \varepsilon, b_\nu]) = \int_{A\cap[b_\nu-\varepsilon,b_\nu]} f_\nu(t)\,dt.$$

This assumption leads to the following corollary to Proposition 1, proven in Appendix A:

Corollary 1. *Under Assumption 4,*

$$b = -\frac{1}{m_b} + \int \frac{t}{1+\mathbf{c}m_b t}\nu(dt)$$

where m_b is the unique solution in $(-(\mathbf{c}b_\nu)^{-1}, 0)$ to the equation in m

$$\int \left(\frac{mt}{1+\mathbf{c}mt}\right)^2 \nu(dt) = \frac{1}{\mathbf{c}}. \quad (3)$$

B. Hypotheses on the signal matrix

We now turn to the hypotheses on the signal matrix A_T:

Assumption 5. *Let $K \geq 0$ be a fixed integer. The matrix $A_T \in \mathbb{C}^{N\times T}$ is random, independent of W_T, with rank $\mathrm{rank}(A_T) = K$ w.p. 1 for all large T. In addition, $\sup_T \|A_T\| < \infty$ w.p. 1.*

In the remainder of the paper, when $K \leq \min(N,T)$, the notation $A_T = U_T B_T^{\mathsf{H}}$ refers to any factorization of A_T where $U_T \in \mathbb{C}^{N\times K}$ satisfies $U_T^{\mathsf{H}} U_T = I_K$. By Assumption 5, the rank of $B_T \in \mathbb{C}^{T\times K}$ is equal to K, w.p. 1.

Assumption 6. *There exists a factorization $A_T = U_T B_T^{\mathsf{H}}$ such that, for any $z \in \mathbb{C} - \mathrm{supp}(\nu)$,*

$$B_T^{\mathsf{H}}\left(R_T - zI_T\right)^{-1} B_T \xrightarrow[T\to\infty]{a.s.} m_\nu(z)P$$

for some $P = \mathrm{diag}(p_1 I_{j_1}, \ldots, p_t I_{j_t})$, $p_1 > \ldots > p_t > 0$, $j_1 + \ldots + j_t = K$ and where it is recalled that $m_\nu(z)$ is the ST of the probability measure ν.

Remark 1. *Assumption 6 is in general very strong. It basically requires that either the right eigenvectors of A_T or the eigenvectors of R_T be sufficiently isotropic. It is however often met in practice:*

- <u>Array Processing</u>: *Consider the model $A_T = H_T P^{1/2} S_T^{\mathsf{H}}$, with $H_T = [h(\theta_1), \cdots, h(\theta_K)]$ (θ_k distinct) the matrix of steering vectors, $P = \mathrm{diag}(a_1^2, \ldots, a_K^2)$ the matrix of source powers, and $S_T = T^{-1/2}[s_{t,k}^*]_{t,k=1}^{T,K}$ the matrix of source signals, and take $V_T = W_T R_T^{1/2}$. Assume the $s_{k,t}$ i.i.d. of zero mean and unit variance and $[\sqrt{N}h(\theta)]_n = e^{-2\pi\imath n\sin(\theta)}$ as in a uniform linear array. In this setting, decomposing $A_T = U_T B_T^{\mathsf{H}}$ with $U_T = H_T(H_T H_T^{\mathsf{H}})^{-1/2}$ and $B_T = S_T P^{1/2}(H_T H_T^{\mathsf{H}})^{1/2}$, we can show $(H_T H_T^{\mathsf{H}})^{-1/2} \xrightarrow{\text{a.s.}} I_K$ while $S_T^{\mathsf{H}}(R_T - zI_T)^{-1} S_T \xrightarrow{\text{a.s.}} m_\nu(z)P$ so that Assumption 6 holds. See the proof of Lemma 1 for details.*
- <u>MIMO Communication</u>: *Let $A_T = H_T P^{1/2} S_T^{\mathsf{H}}$, with now $H_T = [h_1, \ldots, h_K]$ the wireless channels with i.i.d. zero mean and unit variance entries of K transmitters, P their diagonal power matrix and S_T their matrix of transmitted zero mean unit variance i.i.d. signals. Taking now $V_T = R_T^{1/2} W_T$, i.e. spatially correlated noise, and considering*

[3] We recall that the ST m_μ of a probability measure μ with support in $\mathbb{R}$ is defined by $m_\mu(z) = \int (t-z)^{-1}\mu(dt)$. It is analytic on $\mathbb{C} - \mathrm{supp}(\mu)$ and completely characterizes the measure μ.

Y_T^{H} *instead of* Y_T*, we may write* $A_T^{\mathsf{H}} = U_T B_T^{\mathsf{H}}$ *with* $U_T = S_T(S_T S_T^{\mathsf{H}})^{-1/2}$ *and* $B_T = H_T P^{1/2}(S_T S_T^{\mathsf{H}})^{1/2}$ *to obtain* $B_T^{\mathsf{H}}(R_T - zI_N)^{-1}B_T \xrightarrow{\text{a.s.}} m_\nu(z)P$.

Section III-A will introduce the main results of the article, and in particular the new detection and estimation procedures under the general hypothesis of Assumption 6. Since the results of Section III-A may be difficult to grasp in the full generality of the set of hypotheses, we will then devote Section III-B to the specific study of Item 1) in Remark 1 with v_t an ARMA process for which an improved MUSIC algorithm to estimate the angles θ_k will be proposed.

C. Results on the information-plus-noise matrix

Before moving to these applications, we first recall the main results concerning the eigenvalue distribution of $Y_T Y_T^{\mathsf{H}}$. Since $Y_T Y_T^{\mathsf{H}}$ is at most a rank $2K$ perturbation of $V_T V_T^{\mathsf{H}}$ with K fixed, Weyl's interlacing inequalities [20, Th. 4.3.6] show, in conjunction with Theorem 1, that the spectral measure of $Y_T Y_T^{\mathsf{H}}$ also converges to μ in the sense of Theorem 1-2). However, a finite number of eigenvalues of $Y_T Y_T^{\mathsf{H}}$ might stay isolated away from the support of μ [21, Th. 2.2]:

Theorem 2. *Under Assumptions 1–6, let* μ *and* $[a,b]$ *be as in Theorem 1. Let* $\hat\lambda_{1,T} \geq \cdots \geq \hat\lambda_{N,T}$ *be the eigenvalues of* $Y_T Y_T^{\mathsf{H}}$ *with spectral measure* $\hat\tau_T = N^{-1}\sum_{i=1}^N \hat\lambda_{i,T}$*. Then:*

1) *For every bounded and continuous real function* f*,*
$$\int f(t)\hat\tau_T(dt) \xrightarrow[T\to\infty]{\text{a.s.}} \int f(t)\mu(dt).$$
2) *For any interval* $[x_1,x_2] \subset (0,a)$
$$\sharp\{i \,:\, \hat\lambda_{i,T} \in [x_1,x_2]\} = 0 \text{ w.p. } 1 \text{ for all large } T.$$
3) *The function* $\mathbf{g}(x) \triangleq x\mathbf{m}(x)\tilde{\mathbf{m}}(x)$ *is positive and decreases from* $\mathbf{g}(b^+)$ *to zero on* (b,∞)*. If* $p_1\mathbf{g}(b^+) \leq 1$*, then* $\hat\lambda_{1,T} \xrightarrow{\text{a.s.}} b$*. Otherwise, let* $s \in \{1,\ldots,t\}$ *be the largest index for which* $p_s\mathbf{g}(b^+) > 1$ *(see Assumption 6). For* $k = 1,\ldots,s$*, let* $\boldsymbol{\rho}_k$ *be the unique solution* x *in* (b,∞) *of* $p_k\mathbf{g}(x) = 1$*. Then, for* $i = 1,\ldots,s$ *and with* $j_0 = 0$*,*
$$\hat\lambda_{j_1+\cdots+j_{i-1}+1,T},\ldots,\hat\lambda_{j_1+\cdots+j_i,T} \xrightarrow[T\to\infty]{\text{a.s.}} \boldsymbol{\rho}_i$$
$$\hat\lambda_{j_1+\cdots+j_s+1,T} \xrightarrow[T\to\infty]{\text{a.s.}} b.$$
4) *The condition* $p_k\mathbf{g}(b^+) > 1$ *is equivalent to*
$$p_k > \left(\int \frac{-m_b}{1+\mathbf{c}m_b t}\nu(dt)\right)^{-1} \tag{4}$$
with m_b *the solution in* $(-(\mathbf{c}b_\nu)^{-1},0)$ *to Equation* (3).

Proof: The first two items in this theorem are proved in [21] in a more general setting than in this paper. To obtain the last item, observe that $\mathbf{g}(x) = -\int \mathbf{m}(x)(1+\mathbf{c}\mathbf{m}(x)t)^{-1}\nu(dt)$ from the definition of $\tilde{\mathbf{m}}$ in Theorem 1-3) and recall that $\mathbf{m}(x) \downarrow m_b$ as $x \downarrow b$, where m_b is defined in Corollary 1. ■

This theorem shows in particular that the number of isolated eigenvalues of $Y_T Y_T^{\mathsf{H}}$ is upper bounded by the rank K of A_T and it reaches this rank if p_t is large enough.

Remark 2. *In the white noise setting, i.e.* $R_T = I_T$ *(hence,* $\nu = \delta_1$*),* μ *is the celebrated Marchenko-Pastur law, and Equation* (4) *boils down to* $p_k > \sqrt{\mathbf{c}}$ *(see e.g. [22]). The source detection approaches studied in [11], [12], [13] rely on this condition.*

III. Source detection and parameter estimation

We start by stating the results in the general context of Assumptions 1–6. We shall then deal more specifically with the model of Remark 1-1).

A. General results

Theorem 2 gives the following signal dimension estimator:

Theorem 3. *Under Assumptions 1–6 hold true, let* $s \geq 0$ *be the largest integer for which Equation* (4) *holds. Let* $0 < \varepsilon < (\boldsymbol{\rho}_s/b) - 1$ *(take* $\boldsymbol{\rho}_0 = \infty$*). Given* $L \geq K$*, define*
$$\hat k_T = \arg\max_{k\in\{0,\ldots,L\}} \frac{\hat\lambda_{k,T}}{\hat\lambda_{k+1,T}} > 1+\varepsilon$$
(take $\hat\lambda_{0,T} = \infty$*). Then, for all* T *large, w.p. 1,*
$$\hat k_T = j_1 + \ldots + j_s.$$

Proof: Writing $\boldsymbol{k} = j_1 + \ldots + j_s$, Items 1) and 3) of Theorem 2 ensure $\hat\lambda_{\boldsymbol{k},T} \xrightarrow{\text{a.s.}} \boldsymbol{\rho}_s > b$ and $\hat\lambda_{\ell,T} \xrightarrow{\text{a.s.}} b$ for $\ell = \boldsymbol{k}+1,\ldots,L$. ■

Theorem 3 allows in practice to evaluate the number of strong sources when T is large. This however requires ε to be taken such that $\varepsilon < (\boldsymbol{\rho}_s/b)-1$, a value which is practically not known. An empirical approach consists in taking ε sufficiently small (but not too small to avoid counting noise eigenvalues), see Section V. Theorem 3 also assumes that the receiver knows an upper bound L on K, which is less problematic in practice.

In the sequel, for $i \in \{1,\ldots,K\}$, we let $\mathcal{K}(i) = 1$ if $1 \leq i \leq j_1$, $\mathcal{K}(i) = 2$ if $j_1+1 \leq i \leq j_1+j_2$, …, $\mathcal{K}(i) = t$ if $j_1+\cdots+j_{t-1}+1 \leq i \leq K$. The following theorem provides a means for estimating consistently $p_1,\ldots,p_s$:

Theorem 4. *In the setting of Theorem 3, let*
$$\hat m_T(x) \triangleq \frac{1}{N-\hat k_T}\sum_{n=\hat k_T+1}^N \frac{1}{\hat\lambda_{n,T}-x}$$
$$\hat g_T(x) \triangleq \hat m_T(x)(xc_T\hat m_T(x) + c_T - 1)$$
$$\hat p_{i,T} \triangleq \frac{1}{\hat g_T(\hat\lambda_{i,T})},\ i = 1,\ldots,\hat k_T.$$
Then
$$\hat p_{i,T} - p_{\mathcal{K}(i)} \xrightarrow[T\to\infty]{\text{a.s.}} 0.$$

Proof: Recall that $\lambda_{1,T} \geq \ldots \geq \lambda_{N,T}$ are the eigenvalues of $W_T R_T W_T^{\mathsf{H}}$. In the proof, we restrict the elementary events to belong to the probability one set where $\lambda_{1,T} \to b$, $\underline{m}_T(x) \to \mathbf{m}(x)$ uniformly on the compact subsets of (b,∞) (see Theorem 1–6), $\hat\lambda_{i,T} \to \boldsymbol{\rho}_{\mathcal{K}(i)}$ for $i = 1,\ldots,j_1+\cdots+j_s$, $\hat\lambda_{j_1+\cdots+j_s+1,T} \to b$, and $\hat k_T \to j_1+\cdots+j_s$ from Theorems 1, 2, and 3. Observe that $Y_T Y_T^{\mathsf{H}}$ is at most a (nonnegative) rank $2K$ perturbation of $V_T V_T^{\mathsf{H}}$. In these conditions, Weyl's inequalities [20, Th. 4.3.6] ensure $\hat{\lambda}_{n,T} \leq \lambda_{n-2K,T}$ and $\lambda_{n,T} \leq \hat{\lambda}_{n-2K,T}$ for $= 2K+1,\ldots,N$. For any $x > b$ and T large, we then obtain

$$\begin{aligned}\hat{m}_T(x) &\geq \frac{1}{N-\hat{k}_T}\left(\sum_{n=1}^{N-2K}\frac{1}{\lambda_{n,T}-x}+\sum_{n=\hat{k}_T+1}^{2K}\frac{1}{\hat{\lambda}_{n,T}-x}\right)\\ &\triangleq \underline{m}_T(x)+e_T(x)\end{aligned}$$

where $e_T(x) \to 0$ uniformly on compact sets of (b,∞), and

$$\begin{aligned}&\hat{m}_T(x)\\ &\leq \frac{1}{N-\hat{k}_T}\left(\sum_{n=\hat{k}_T+1+2K}^{N}\frac{1}{\lambda_{n,T}-x}+\sum_{n=N-2K+1}^{N}\frac{1}{\hat{\lambda}_{n,T}-x}\right)\\ &\triangleq \underline{m}_T(x)+e'_T(x)\end{aligned}$$

where $e'_T(x) \to 0$ uniformly on compact sets of (b,∞). Consequently, $\hat{g}_T(\hat{\lambda}_{i,T}) - \mathbf{g}(\hat{\lambda}_{i,T}) \to 0$ for $i = 1,\ldots,\hat{k}_T$. Clearly, $\mathbf{g}(\hat{\lambda}_{i,T}) - \mathbf{g}(\boldsymbol{\rho}_{\mathcal{K}(i)}) \to 0$ so that $\hat{g}_T(\hat{\lambda}_{i,T}) - \mathbf{g}(\boldsymbol{\rho}_{\mathcal{K}(i)}) \to 0$ which, along with $\mathbf{g}(\boldsymbol{\rho}_{\mathcal{K}(i)}) = 1/p_{\mathcal{K}(i)}$, gives the result. ■

Let now $A_T = U_T B_T^{\mathsf{H}}$ following Assumption 6 and write $U_T = [U_{1,T},\ldots,U_{t,T}]$, $U_{\ell,T} \in \mathbb{C}^{N\times j_\ell}$. We introduce the orthogonal projection matrix $\Pi_{\ell,T} = U_{\ell,T}U_{\ell,T}^{\mathsf{H}} \in \mathbb{C}^{N\times N}$. Similarly, we denote $\hat{\Pi}_{\ell,T}$ the orthogonal projection matrix on the eigenspace corresponding to the set of eigenvalues $\{\hat{\lambda}_{j_1+\ldots+j_{\ell-1}+1,T},\ldots,\hat{\lambda}_{j_1+\ldots+j_\ell}\}$ in $Y_TY_T^{\mathsf{H}}$, for $\ell = 1,\ldots,t$ ($j_0 = 0$). With these notations, we have the following estimate of bilinear forms of the type $a_T^{\mathsf{H}}\Pi_{\ell,T}b_T$:

Theorem 5. *Under Assumptions 1–6, let $a_T, b_T \in \mathbb{C}^N$ be two sequences of deterministic vectors with bounded norms and let $\mathcal{K}(i) \leq s$ with s the largest integer for which* (4) *holds. Then:*

$$a_T^{\mathsf{H}}\Pi_{\mathcal{K}(i),T}b_T - \frac{\hat{g}'_T(\hat{\lambda}_{i,T})}{\hat{m}_T(\hat{\lambda}_{i,T})\hat{g}_T(\hat{\lambda}_{i,T})}a_T^{\mathsf{H}}\hat{\Pi}_{\mathcal{K}(i),T}b_T \xrightarrow[T\to\infty]{a.s.} 0.$$

Proof: From Assumption 6, $B_T^{\mathsf{H}}B_T \xrightarrow{\text{a.s.}} P$ (multiply each side of the convergence by $-z$ and take z large). Therefore, $p_1,\ldots,p_t$ are the limiting positive eigenvalues of $A_TA_T^{\mathsf{H}}$. For $R_T = I_N$, the theorem thus coincides with [22, Theorem 2] since then $V_T = W_T$ is a bi-unitarily invariant (here Gaussian) matrix as requested by [22, Assumption 2]. We now reproduce the steps of [22, Theorem 2] under our set of assumptions. [22, Equation (8)] remains valid in our setting which, under the present notations, reads

$$\begin{aligned}a_T^{\mathsf{H}}\hat{\Pi}_{\ell,T}b_T &= -\frac{1}{\imath\pi}\oint_{\mathcal{C}_{\ell,T}}\tilde{a}_T^{\mathsf{H}}\underline{Q}_T(z)\tilde{b}_Tdz\\ &\quad+\frac{1}{\imath\pi}\oint_{\mathcal{C}_{\ell,T}}\hat{a}_T^{\mathsf{H}}\hat{H}_T(z)^{-1}\hat{b}_Tdz \end{aligned}\qquad (5)$$

for $\mathcal{C}_{\ell,T}$ a complex positively oriented contour enclosing only the eigenvalues $\hat{\lambda}_{j_1+\ldots+j_{\ell-1}+1,T},\ldots,\hat{\lambda}_{j_1+\ldots+j_\ell,T}$, with

$$\begin{aligned}&\tilde{a}_T^{\mathsf{T}} = [a_{k,T}^{\mathsf{T}},0,\ldots,0],\ \tilde{b}_T^{\mathsf{T}} = [b_{k,T}^{\mathsf{T}},0,\ldots,0]\\ &Q_T(z) = (V_TV_T^{\mathsf{H}}-zI_N)^{-1},\ \tilde{Q}_T(z) = (V_T^{\mathsf{H}}V_T-zI_T)^{-1}\\ &\underline{Q}_T(z) = \begin{bmatrix} zQ_T(z^2) & V_T\tilde{Q}_T(z^2)\\ \tilde{Q}_T(z^2)V_T^{\mathsf{H}} & z\tilde{Q}_T(z^2)\end{bmatrix}\end{aligned}$$

$$\begin{aligned}&\hat{a}_T = \begin{bmatrix} zU_T^{\mathsf{H}}\tilde{Q}_T(z^2)\\ B_T^{\mathsf{H}}\tilde{Q}_T(z^2)V_T^{\mathsf{H}}\end{bmatrix}a_T,\ \hat{b}_T = \begin{bmatrix} zU_T^{\mathsf{H}}\tilde{Q}_T(z^2)\\ B_T^{\mathsf{H}}\tilde{Q}_T(z^2)V_T^{\mathsf{H}}\end{bmatrix}b_T\\ &\hat{H}_T(z) = \begin{bmatrix} zU_T^{\mathsf{H}}Q_T(z^2)U_T & U_T^{\mathsf{H}}V_T\tilde{Q}_T(z^2)B_T+I_K\\ B_T^{\mathsf{H}}\tilde{Q}_T(z^2)V_T^{\mathsf{H}}U_T+I_K & zB_T^{\mathsf{H}}\tilde{Q}_T(z^2)B_T\end{bmatrix}.\end{aligned}$$

Let $\ell \leq s$. From Theorem 2-2), for all large T w.p. 1, the first term on the right-hand side of (5) is null (no pole of $\underline{Q}_T$ lies in $\mathcal{C}_{\ell,T}$ for large T), while in the second term $\mathcal{C}_{\ell,T}$ can be replaced by a contour $\mathcal{C}_\ell$ enclosing $\boldsymbol{\rho}_\ell$ but no $\boldsymbol{\rho}_k$, $k \neq \ell$. We must now prove $\hat{a}_T^{\mathsf{H}}\hat{H}_T(z)\hat{b}_T - \bar{a}_T^{\mathsf{H}}\bar{H}_T(z)\bar{b}_T \xrightarrow{\text{a.s.}} 0$ where

$$\begin{aligned}&\bar{a}_T = \begin{bmatrix} z\mathbf{m}(z^2)U_T^{\mathsf{H}}\\ 0\end{bmatrix}a_T,\ \bar{b}_T = \begin{bmatrix} z\mathbf{m}(z^2)U_T^{\mathsf{H}}\\ 0\end{bmatrix}b_T\\ &\bar{H}_T(z) = \begin{bmatrix} z\mathbf{m}(z^2)I_K & I_K\\ I_K & z\tilde{\mathbf{m}}(z^2)P\end{bmatrix}.\end{aligned}$$

By [21, Lemmas 4.1–4.6], $\|\hat{a}_T-\bar{a}_T\| \xrightarrow{\text{a.s.}} 0$, $\|\hat{b}_T-\bar{b}_T\| \xrightarrow{\text{a.s.}} 0$,

$$\left\|\hat{H}_T(z) - \begin{bmatrix} z\mathbf{m}(z^2)I_K & I_K\\ I_K & \frac{B_T^{\mathsf{H}}\left(I_T+\mathbf{c}\mathbf{m}(z^2)R_T\right)^{-1}B_T}{-z}\end{bmatrix}\right\| \xrightarrow{\text{a.s.}} 0.$$

Assumption 6 and the definition of $\tilde{\mathbf{m}}(z)$ then imply $\|\frac{-1}{z}B_T^{\mathsf{H}}\left(I_T+\mathbf{c}\mathbf{m}(z^2)R_T\right)^{-1}B_T - z\tilde{\mathbf{m}}(z^2)P\| \xrightarrow{\text{a.s.}} 0$, which finally gives $\hat{a}_T^{\mathsf{H}}\hat{H}_T(z)\hat{b}_T - \bar{a}_T^{\mathsf{H}}\bar{H}_T(z)\bar{b}_T \xrightarrow{\text{a.s.}} 0$. For $z \in \mathcal{C}_\ell$, $z\mathbf{m}(z^2)$ and $z\tilde{\mathbf{m}}(z^2)$ are bounded by $[\mathbf{d}(\mathcal{C}_\ell,\operatorname{supp}(\mu))]^{-1}$. Take $0 < \varepsilon < \mathbf{d}(\mathcal{C}_\ell,\operatorname{supp}(\mu))$. Then, for all large T, $zQ_T(z^2)$ and $z\tilde{Q}(z^2)$ are bounded by ε^{-1} w.p. 1. The dominated convergence theorem therefore ensures that

$$a_T^{\mathsf{H}}\hat{\Pi}_{\ell,T}b_T - \frac{1}{\imath\pi}\oint_{\mathcal{C}_\ell}\bar{a}_T^{\mathsf{H}}\bar{H}_T(z)^{-1}\bar{b}_Tdz \xrightarrow{\text{a.s.}} 0.$$

Residue calculus of the right-hand side integrand as in [22, Equations (10)-(11)] then gives

$$a_T^{\mathsf{H}}\hat{\Pi}_{\ell,T}b_T - \frac{\mathbf{m}(\boldsymbol{\rho}_\ell)\mathbf{g}(\boldsymbol{\rho}_\ell)}{\mathbf{g}'(\boldsymbol{\rho}_\ell)}a_T^{\mathsf{H}}\Pi_{\ell,T}b_T \xrightarrow[T\to\infty]{\text{a.s.}} 0.$$

Take i such that $\mathcal{K}(i) = \ell$. Using $\hat{\lambda}_{i,T} \xrightarrow{\text{a.s.}} \boldsymbol{\rho}_\ell$, $\hat{m}_T(x) \xrightarrow{\text{a.s.}} \mathbf{m}(x)$, $\hat{g}_T(x) \xrightarrow{\text{a.s.}} \mathbf{g}(x)$, and $\hat{g}'_T(x) \xrightarrow{\text{a.s.}} \mathbf{g}'(x)$ for x outside the support of μ then concludes the proof. ■

B. Narrowband array processing

We now apply the results of Section III-A to the array processing model of Remark 1. Consider a uniform linear array of N antennas which captures T successive realizations $y_1,\ldots,y_T$ of the random process:

$$y_t = \sum_{k=1}^{K}a_kh(\theta_k)s_{k,t}+v_t \qquad (6)$$

with $a_1 \geq \ldots \geq a_K > 0$ the amplitude of sources $1,\ldots,K$, $h(\theta) \in \mathbb{C}^N$ the steering-vector function

$$h(\theta) = \frac{1}{\sqrt{N}}\left[1, e^{-2\imath\pi\sin\theta},\ldots,e^{-2\imath\pi(N-1)\sin\theta}\right]^{\mathsf{T}} \qquad (7)$$

with θ_k the angle-of-arrival of the signal from source k (the θ_k are assumed distinct), $s_{k,t} \in \mathbb{C}$ the signal emitted by source k at time t such that $(s_{t,k})_{t,k=1}^{\infty,K}$ is an infinite array of circular complex i.i.d. random variables with $\mathbb{E}s_{1,1} = 0$, $\mathbb{E}|s_{1,1}|^2 = 1$, and $\mathbb{E}|s_{1,1}|^8 < \infty$, and $v_t \in \mathbb{C}^N$ the noise received at the sensor array at time t.

Denoting $Y_T = [y_1, \ldots, y_T] \in \mathbb{C}^{N\times T}$, (6) takes the form

$$Y_T = H_T P^{1/2} S_T^{\mathsf{H}} + V_T \tag{8}$$

where $H_T = [h(\theta_1), h(\theta_2), \ldots, h(\theta_K)] \in \mathbb{C}^{N\times K}$, $S_T = T^{-1/2}[s^*_{t,k}]_{t,k=1}^{T,K} \in \mathbb{C}^{T\times K}$, $P = \mathrm{diag}(a_1^2, \ldots, a_K^2)$, and $V_T = T^{-1/2}[v_1, \ldots, v_T] \in \mathbb{C}^{N\times T}$. We assume the rows of $\sqrt{T}V_T$ to be independent snapshots of a complex Gaussian circular causal ARMA(m,n) stationary process. This process can be represented as the output of a filter with transfer function $\mathbf{p}(z) = (1+\alpha_1 z^{-1}+\ldots+\alpha_m z^{-m})/(1+\beta_1 z^{-1}+\ldots+\beta_n z^{-n})$ driven by a standard complex Gaussian circular white noise. For $|z| \geq 1$, $\mathbf{p}(z) = \sum_{\ell=0}^{\infty} \psi_\ell z^{-\ell}$ where $\sum |\psi_\ell| < \infty$, and we can write $V_T = W_T R_T^{1/2}$ with W_T as in Assumption 2 and

$$R_T = \begin{bmatrix} r_0 & r_1 & \ldots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \ldots & r_{-1} & r_0 \end{bmatrix}$$

with $r_k = \sum_{\ell\geq 0} \psi_{\ell+k}\psi^*_\ell$ for any $k \in \mathbb{N}$, the matrix being nonnegative.

Lemma 1. *Under Assumption 1, the model* (8) *satisfies Assumptions 2–6 with ν defined by*

$$\int g(t)\nu(dt) = \int_0^1 g(|\mathbf{p}(\exp(2\imath\pi u))|^2)\, du \tag{9}$$

for every positive measurable function g, and with P in Assumption 6 the matrix of the source powers a_k^2.

Proof: We start with Assumptions 3 and 4. If $m = n = 0$, then $\nu = \delta_1$ and these assumptions are trivially satisfied. Assume $\min(m,n) > 0$. Then Assumption 3–1) is a well known result on the spectral behavior of large Toeplitz matrices [23], [24]. The support of ν is the compact interval $[a_\nu, b_\nu] = [\min_u q(u), \max_u q(u)]$, $q(u) \triangleq |\mathbf{p}(\exp(2\imath\pi u))|^2$. It is also well known [23, §4.2] that $a_\nu \leq \sigma^2_{t,T} \leq b_\nu$, so that Assumption 3–2) is satisfied. Since $\mathbf{p}(z)$ is ARMA, for $g(t)$ the indicator function on a set of Lebesgue measure zero, the right hand side of (9) is zero. Hence ν has a density f_ν with respect to the Lebesgue measure. Let us provide the expression of f_ν at a point $s \in (a_\nu, b_\nu)$ such that for any u for which $q(u) = s$, $q'(u) \neq 0$. In a neighborhood of any of these u, q has a local inverse that we denote $q_u^{(-1)}$. Then, for $\varepsilon > 0$ small enough,

$$\begin{aligned}\nu(s-\varepsilon, s+\varepsilon) &= \int_{t\,:\,q(t)\in[s-\varepsilon,s+\varepsilon]} dt \\ &= \sum_{u\,:\,q(u)=s} \int_{[s-\varepsilon,s+\varepsilon]} \frac{1}{\left|q'(q_u^{(-1)}(v))\right|}\, dv\end{aligned}$$

by the variable change $q(t) = v$. Letting $\varepsilon \downarrow 0$, we obtain

$$\lim_{\varepsilon\downarrow 0} \frac{\nu(s-\varepsilon, s+\varepsilon)}{2\varepsilon} = \sum_{u\,:\,q(u)=s} \frac{1}{|q'(u)|} = f_\nu(s).$$

This proves $f_\nu(s) \to \infty$ as $s \uparrow b_\nu$, implying Assumption 4.

We now turn to Assumptions 5 and 6. Since the θ_i are distinct (modulo π), $H_T^{\mathsf{H}} H_T \to I_K$. By the law of large numbers, $S_T^{\mathsf{H}} S_T \xrightarrow[T\to\infty]{\text{a.s.}} I_K$. Hence $\mathrm{rank}(A_T) = K$ w.p. 1 for all large T, and $\sup_T \|A_T\| < \infty$ w.p. 1. Let us write $A_T = U_T B_T^{\mathsf{H}}$ where $U_T = H_T(H_T^{\mathsf{H}} H_T)^{-1/2}$ and where $B_T = S_T P^{1/2}(H_T^{\mathsf{H}} H_T)^{1/2}$. By [19, Lemma 2.7] and $\mathbb{E}|s_{1,1}|^8 < \infty$, for any $z \in \mathbb{C}_+$ and any $1 \leq i, j \leq K$,

$$\mathbb{E}\left|\left[S_T^{\mathsf{H}}(R_T - zI_T)^{-1}S_T - \frac{\mathrm{Tr}[(R_T - zI_T)^{-1}]}{T} I_K\right]_{i,j}\right|^4 \leq \frac{C}{T^2}$$

for some $C > 0$. By Markov's inequality, the argument of $\mathbb{E}|\cdot|^4$ converges to zero w.p. 1, and this convergence can be extended to $\mathbb{C} - \mathrm{supp}(\mu)$. Since $T^{-1}\,\mathrm{Tr}[(R_T - zI_T)^{-1}] \to m_\nu(z)$ for $z \in \mathbb{C} - \mathrm{supp}(\nu)$, Assumption 6 is satisfied. ■

With these results, Lemma 1 and Theorems 3 and 4 lead to the following inference methods:

Proposition 2. *Consider the model* (8)*. Let $k \geq 0$ be the largest integer for which (take $a_0 = \infty$)*

$$a_k^2 > \left(\int_0^1 \frac{-m_b}{1 + \mathbf{c} m_b |\mathbf{p}(\exp(2\imath\pi u))|^2}\, du\right)^{-1} \tag{10}$$

with $m_b \in (-(\mathbf{c}\max_u |\mathbf{p}(\exp(2\imath\pi u))|^2)^{-1}, 0)$ the solution of

$$\int_0^1 \left(\frac{m\,|\mathbf{p}(\exp(2\imath\pi u))|^2}{1 + \mathbf{c} m\,|\mathbf{p}(\exp(2\imath\pi u))|^2}\right)^2 du = \frac{1}{\mathbf{c}}.$$

Given $L \geq K$ and $\varepsilon > 0$, define (with $\hat\lambda_{0,T} = \infty$)

$$\hat{k}_T = \arg\max_{m\in\{0,\ldots,L\}} \frac{\hat\lambda_{m,T}}{\hat\lambda_{m+1,T}} > 1 + \varepsilon.$$

Then $\hat{k}_T = k$ w.p. 1 for all large T and ε small enough. Moreover, for $i = 1, \ldots, \hat{k}_T$ let $\hat{a}^2_{i,T} \triangleq (\hat{g}_T(\hat\lambda_{i,T}))^{-1}$ with $\hat{g}_T(\hat\lambda_{i,T})$ as in Theorem 4. Then

$$\hat{a}^2_{i,T} \xrightarrow{\text{a.s.}} a_i^2.$$

Based on Theorem 5, we now provide a source localization method based on MUSIC [5]. Recall that MUSIC exploits the fact that $h(\theta_i)^{\mathsf{H}}(I_N - \Pi^\ell_{1,T})h(\theta_i) = 0$ with $\Pi^\ell_{1,T}$ a projector on the subspace generated by $h(\theta_1), \ldots, h(\theta_\ell)$ for any $i \leq \ell \leq K$. Since $\|h(\theta)\| = 1$, $\theta_1, \ldots, \theta_\ell$ are the arguments of the local maxima of

$$\gamma^\ell_T(\theta) \triangleq h(\theta)^{\mathsf{H}} \Pi^\ell_{1,T} h(\theta).$$

Proposition 3. *Let k and $\hat{k}_T$ be as in Proposition 2 and denote $\hat{u}_{1,T}, \ldots, \hat{u}_{\hat{k}_T,T}$ the eigenvectors of $Y_T Y_T^{\mathsf{H}}$ with respective eigenvalues $\hat\lambda_{1,T}, \ldots, \hat\lambda_{\hat{k}_T,T}$. Then, for $\theta \in [-\pi/2, \pi/2]$,*

$$\gamma^k_T(\theta) - \hat\gamma^{\hat{k}_T}_T(\theta) \xrightarrow{\text{a.s.}} 0$$

where

$$\begin{aligned}\gamma^k_T(\theta) &\triangleq h(\theta)^{\mathsf{H}} \Pi^k_{1,T} h(\theta) \\ \hat\gamma^{\hat{k}_T}_T(\theta) &\triangleq \sum_{j=1}^{\hat{k}_T} \frac{\hat{g}'_T(\hat\lambda_{j,T})}{\hat{m}_T(\hat\lambda_{j,T})\hat{g}_T(\hat\lambda_{j,T})} h(\theta)^{\mathsf{H}} \hat{u}_{j,T}\hat{u}^{\mathsf{H}}_{j,T} h(\theta).\end{aligned}$$

Proof: Lemma 1 ensures that Assumptions 1–6 are satisfied, so Theorem 5 can be applied for each $i \leq k$. Taking $a_T =$

$b_T = h(\theta)$ and $U_T = H_T(H_T^{\mathsf{H}} H_T)^{-1/2}$ as in Theorem 5, we obtain the desired result for $U_T J U_T^{\mathsf{H}}$, $J = \mathrm{diag}(I_k, 0)$, instead of $\Pi_{1,T}^k$. As $(H_T^{\mathsf{H}} H_T)^{-1/2} J (H_T^{\mathsf{H}} H_T)^{-1/2} \xrightarrow{\text{a.s.}} J$ and $H_T J H_T^{\mathsf{H}}$ is the same projector as $\Pi_{1,T}^k$, we have $h(\theta)^{\mathsf{H}} \Pi_{1,T}^k h(\theta) - h(\theta)^{\mathsf{H}} U_T J U_T^{\mathsf{H}} h(\theta) \xrightarrow{\text{a.s.}} 0$, completing the proof. ■

Proposition 3 ensures that $\hat{\gamma}_T^{\hat{k}_T}(\theta)$ is a consistent estimator of the localization function $\gamma_T^k(\theta)$. The alternative MUSIC algorithm we therefore propose consists in estimating $\theta_1, \ldots, \theta_k$ as the arguments of the $\hat{k}_T$ highest maxima of $\hat{\gamma}_T^{\hat{k}_T}(\theta)$. Observe that, although the system models differ in both articles, the MUSIC estimator proposed here exactly corresponds to that provided in [22]. This remark would not hold if it were not for Assumption 6.

IV. Second order performance analysis

In this section, we discuss the asymptotic (second order) performance of the detection and estimation schemes derived in Section III. The model of Section III-B is considered. Following the notations of Section III-A, we gather the source powers a_k^2 in groups of equal powers $p_1 > \ldots > p_t$ with respective multiplicities $j_1, \ldots, j_t$.

A. Main results

We start by studying the fluctuations of the isolated eigenvalues of $Y_T Y_T^{\mathsf{H}}$. Recall the definition of ν_T in Assumption 3 and recall that $c_T = N/T$. Replacing ν and $\mathbf{c}$ with ν_T and c_T, respectively, in Theorem 1, we obtain that

$$m_T(z) = \left(-z + \int \frac{t}{1 + c_T m_T(z) t} \nu_T(dt) \right)^{-1} \tag{11}$$

uniquely defines the ST $m_T(z)$ of a probability measure μ_T supported by $\mathbb{R}_+$. In addition, μ_T converges weakly to μ as $T \to \infty$; the Hausdorff distance between the supports of these two measures converges to zero [17], [19] and, for each $b' > b$, $m_T(z)$ is analytic on $\mathbb{C} - [0, b']$ for all large T. Let

$$\begin{aligned} \tilde{m}_T(z) &= \int \frac{-1}{z(1 + c_T m_T(z) t)} \nu_T(dt) \\ &= \frac{-1}{zT} \operatorname{Tr}(I_T + c_T m_T(z) R_T)^{-1}. \end{aligned}$$

Similarly to Theorem 1-3), $\tilde{m}_T(z)$ satisfies $\tilde{m}_T(z) = c_T m_T(z) - (1 - c_T)/z$. Consequently, for all T large, $g_T(x) \triangleq x m_T(x) \tilde{m}_T(x)$ is defined on (b', ∞), $b' > b$, and, for any k such that $p_k \mathbf{g}(b^+) > 1$, $p_k g_T(x) = 1$ has a unique solution $\rho_{k,T}$ in (b, ∞).

The main result of this section (Theorem 6) describes the fluctuations of $\hat{\lambda}_{i,T} - \rho_{\mathcal{K}(i),T}$, $i \le s$, with s the largest integer satisfying (4).

Lemma 2. *Consider the model* (8)*. Then the function*

$$\boldsymbol{\Delta}(x) = 1 - \mathbf{c} \int \left(\frac{\mathbf{m}(x) t}{1 + \mathbf{c}\mathbf{m}(x) t} \right)^2 \nu(dt)$$

is defined and positive on (b, ∞)*. Furthermore,* $\boldsymbol{\Delta}(x) \to 0$ *as* $x \downarrow b$ *and* $\boldsymbol{\Delta}(x) \to 1$ *as* $x \to \infty$.

Proof: See Appendix B. ■

Theorem 6. *Consider* (8) *with the assumptions of Section III-B. Assume in addition* $\mathbb{E}[s_{1,1}^u (s_{1,1}^*)^v] = 0$ *for* $u + v \le 4$ *and* $u \ne v$*, and let* $\kappa \triangleq \mathbb{E}|s_{1,1}|^4 - 2$*. Let* s *be the largest integer (assumed* ≥ 1*) for which* (10) *holds. For* $k = 1, \ldots, s$ *and all* T *large, let* $\rho_{k,T}$ *be the unique solution in* (b, ∞) *of* $p_k g_T(x) = 1$*. Define (with* $j_0 = 0$*)*

$$\eta_{k,T} = \sqrt{T} \left(\begin{bmatrix} \hat{\lambda}_{j_1 + \cdots + j_{k-1} + 1, T} \\ \vdots \\ \hat{\lambda}_{j_1 + \cdots + j_k, T} \end{bmatrix} - \rho_{k,T} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right),$$

$$\begin{aligned} \alpha_k &= \frac{\mathbf{m}^2(\boldsymbol{\rho}_k)}{\boldsymbol{\Delta}(\boldsymbol{\rho}_k)} \Bigg[\int \frac{t^2 + 2 p_k t}{(1 + \mathbf{c}\mathbf{m}(\boldsymbol{\rho}_k) t)^2} \nu(dt) \\ &\qquad + \mathbf{c} \Big(\int \frac{p_k \mathbf{m}(\boldsymbol{\rho}_k) t}{(1 + \mathbf{c}\mathbf{m}(\boldsymbol{\rho}_k) t)^2} \nu(dt) \Big)^2 \Bigg], \\ \beta_k &= \int \frac{p_k^2 \mathbf{m}(\boldsymbol{\rho}_k)^2}{(1 + \mathbf{c}\mathbf{m}(\boldsymbol{\rho}_k) t)^2} \nu(dt), \quad \textit{and} \\ \phi_k &= \left(\int \frac{p_k \mathbf{m}(\boldsymbol{\rho}_k)}{1 + \mathbf{c}\mathbf{m}(\boldsymbol{\rho}_k) t} \nu(dt) \right)^2. \end{aligned}$$

Let $M_1, \ldots, M_s$*,* $M_k = [M_{\ell,m,k}]_{1 \le \ell, m \le j_k}$*, be random independent Hermitian matrices such that* $\{M_{\ell,m,k}\}_{\ell \le m}$ *are independent,* $M_{\ell,\ell,k} \sim \mathcal{N}(0, \alpha_k + \beta_k + \kappa\phi_k)$*, and* $M_{\ell,m,k} \sim \mathcal{CN}(0, \alpha_k + \beta_k)$ *for* $1 \le \ell < m \le j_k$*. Let* χ_k *be the* $\mathbb{R}^{j_k}$*−valued vector of the decreasingly ordered eigenvalues of* $(p_k \mathbf{g}'(\boldsymbol{\rho}_k))^{-1} M_k$*. Then*

$$(\eta_{1,T}, \ldots, \eta_{s,T}) \xrightarrow[T \to \infty]{\mathcal{L}} (\chi_1, \ldots, \chi_s).$$

Proof: The proof of the theorem is given in Section IV-B. ■

Theorem 6 shows that, after appropriate centering and scaling, the vector of the isolated eigenvalues of $Y_T Y_T^{\mathsf{H}}$ that converge to $\boldsymbol{\rho}_k > b$ tends to fluctuate like the eigenvalues of a certain Hermitian matrix with Gaussian elements. If $\kappa = 0$, this matrix is a scaled Gaussian Unitary Ensemble (GUE) matrix.[4] When $K = 0$, $\boldsymbol{s} T^{2/3}(\hat{\lambda}_{1,T} - b_T)$ converges in law to the Tracy-Widom probability distribution $\mathsf{TW}(\cdot)$, where b_T is the finite horizon equivalent to b and $\boldsymbol{s}$ is a scaling parameter that depends on $\mathbf{c}$ and ν [25]. This result can be generalized to show that for any fixed integer r, the vector $T^{2/3}(\hat{\lambda}_{1,T} - b_T, \ldots, \hat{\lambda}_{r,T} - b_T)$ converges in distribution to a multidimensional version of the Tracy-Widom law. These results and Theorem 6 can then be used to evaluate the error probabilities of the source detection schemes described in Theorem 3 and Proposition 2.

Remark 3. *We note without proof that for the specific ARMA model considered here, the measure* ν_T *can be freely replaced with* ν *in Equation (11). The error incurred on* $m_T(z)$ *by this replacement is negligible in the ARMA context.*

Theorem 6 can also be used to characterize the fluctuations of the source power estimates:

[4] We recall that a GUE matrix is a random Hermitian matrix $M = [M_{ij}]$ such that $M_{ii} \sim \mathcal{N}(0, 1)$, $M_{ij} \sim \mathcal{CN}(0, 1)$ for $i < j$, these random variables being independent.

Theorem 7. *Consider the setup of Theorem 6 and let* $\hat{p}_{i,T} = (\hat{g}_T(\hat{\lambda}_{i,T}))^{-1}$ *for* $i = 1, \ldots, j_1 + \cdots + j_s$. *For* $k = 1, \ldots, s$, *define (with* $j_0 = 0$*)*

$$\xi_{k,T} = \sqrt{T}\left(\begin{bmatrix}\hat{p}_{j_1+\cdots+j_{k-1}+1,T}\\ \vdots \\ \hat{p}_{j_1+\cdots+j_k,T}\end{bmatrix} - p_k\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}\right).$$

Let M_k *be defined as in Theorem 6 and let* $\check{\chi}_k$ *be the* $\mathbb{R}^{j_k}-$*valued vector of the decreasingly ordered eigenvalues of* $p_k M_k$. *Then*

$$(\xi_{1,T}, \ldots, \xi_{s,T}) \xrightarrow[T\to\infty]{\mathcal{L}} (\check{\chi}_1, \ldots, \check{\chi}_s).$$

Proof: A sketch of the proof is provided in Appendix C. ■

As a corollary of Theorem 7, the following proposition provides the behavior of the power estimates for extreme values of p_k, *i.e.* for $p_k \to \infty$ and for p_k close to the detectability limit given by (10):

Proposition 4. *Consider the setting of Theorem 7. Let* $p_{\lim}$ *be the infimum of the* p_k *satisfying* (10), M_k *be defined as in Theorem 6, and* $\psi_k \triangleq \alpha_k + \beta_k + \kappa\phi_k$, $\breve{\psi}_k \triangleq \alpha_k + \beta_k$. *Then*

$$\psi_k \xrightarrow[p_k \downarrow p_{\lim}]{} \infty, \quad \breve{\psi}_k \xrightarrow[p_k \downarrow p_{\lim}]{} \infty$$
$$\psi_k \xrightarrow[p_k \to \infty]{} 1+\kappa, \quad \breve{\psi}_k \xrightarrow[p_k \to \infty]{} 1.$$

Proof: See Appendix D. ■

B. Proof of Theorem 6

The proof relies on two ingredients: an adaption of [21, Th. 2.3] and a result on fluctuations of quadratic forms. Let $A_T = U_T B_T^{\mathsf{H}}$ with $U_T = H_T(H_T^{\mathsf{H}}H_T)^{-1/2}$ and $B_T = S_T P^{1/2}(H_T^{\mathsf{H}}H_T)^{1/2} = [B_{1,T}, \ldots, B_{t,T}]$, $B_{k,T} \in \mathbb{C}^{T\times j_k}$. In [21], it is shown that the $\eta_{k,T}$ fluctuate like the ordered eigenvalues of the matrices $(p_k \mathbf{g}(\boldsymbol{\rho}_k)')^{-1}(\sqrt{\alpha_k}G_k + \sqrt{T}F_{k,T})$ where $F_{k,T} = m_T(\rho_{k,T})B_{k,T}^{\mathsf{H}}(I_T + c_T m_T(\rho_{k,T})R_T)^{-1}B_{k,T} + I_{j_k}$ and the G_k are GUE matrices independent of the $F_{k,T}$. Using $H_T^{\mathsf{H}}H_T \xrightarrow{\text{a.s.}} I_K$, the law of large numbers and the definition of $\rho_{k,T}$ informally give

$$F_{k,T} \simeq \left(\frac{p_k}{T}\operatorname{Tr}\left[m_T(\rho_{k,T})(I_T + c_T m_T(\rho_{k,T})R_T)^{-1}\right] + 1\right) I_{j_k} = 0.$$

We thus need to study the fluctuations of $\sqrt{T}F_{k,T}$, which is the purpose of the following lemmas, proved in Appendices E, F, and G, respectively:

Lemma 3. *Let* $D_T \in \mathbb{C}^{T\times T}$ *be a sequence of deterministic Hermitian matrices with* $\sup_T \|D_T\| < \infty$. *Assume that*

$$\frac{1}{T}\operatorname{Tr} D_T^2 \xrightarrow[T\to\infty]{} \beta \quad \text{and} \quad \frac{1}{T}\operatorname{Tr}(\operatorname{diag}(D_T))^2 \xrightarrow[T\to\infty]{} \phi.$$

Consider the matrices S_T *defined by* (8). *Then*

$$\sqrt{T}\Big(S_T^{\mathsf{H}}D_T S_T - \frac{\operatorname{Tr} D_T}{T} I_K\Big) \xrightarrow[T\to\infty]{\mathcal{L}} G$$

where $G = [G_{ij}]_{1\le i,j\le K}$ *is random Hermitian such that* $\{G_{ij}\}_{i\le j}$ *are independent,* $G_{ii} \sim \mathcal{N}(0, \beta + \kappa\phi)$ *for* $1 \le i \le K$, *and* $G_{ij} \sim \mathcal{CN}(0,\beta)$ *for* $1 \le i < j \le K$.

Lemma 4. *Let* $1 \le k \le s$ *and*

$$D_T = p_k m_T(\rho_{k,T})(I_T + c_T m_T(\rho_{k,T})R_T)^{-1}.$$

Then $\limsup_T \|D_T\| < \infty$,

$$\frac{1}{T}\operatorname{Tr}(D_T^2) \xrightarrow[T\to\infty]{} \beta_k, \text{ and } \frac{1}{T}\operatorname{Tr}(\operatorname{diag}(D_T))^2 \xrightarrow[T\to\infty]{} \phi_k$$

where β_k *and* ϕ_k *are given in Theorem 6.*

Lemma 5. *Let* $M_1, \ldots, M_t$, $M_k = [M_{\ell,m,k}]_{1\le \ell,m\le j_k}$, *be random independent Hermitian matrices such that the* $\{M_{\ell,m,k}\}_{\ell\le m}$ *are independent,* $M_{\ell,\ell,k} \sim \mathcal{N}(0, \beta_k + \kappa\phi_k)$, *and* $M_{\ell,m,k} \sim \mathcal{CN}(0,\beta_k)$ *for* $1 \le \ell < m \le j_k$. *Then*

$$(\sqrt{T}F_{k,T})_{k=1,\ldots,t} \xrightarrow[T\to\infty]{\mathcal{L}} (M_k)_{k=1,\ldots,t}.$$

The proof ends with the adapted statement of [21, Th. 2.3]:[5]

Proposition 5. *In the setting of Theorem 6, let* $G_1, \ldots, G_s$, $G_k \in \mathbb{C}^{j_k\times j_k}$, *be independent GUE matrices. Then, for any bounded and continuous* $f : \mathbb{R}^{j_1+\cdots+j_s} \to \mathbb{R}$,

$$\mathbb{E}[f(\eta_{1,T}, \ldots, \eta_{s,T})] - \mathbb{E}[f(\zeta_1, \ldots, \zeta_s)] \to 0$$

where ζ_k *is the random vector of the decreasingly ordered eigenvalues of* $(p_k\mathbf{g}(\boldsymbol{\rho}_k)')^{-1}(\sqrt{\alpha_k}G_k + \sqrt{T}F_{k,T})$.

V. Simulation results

We consider the setting of Section III-B, with signals $s_{t,k}$ drawn from a QPSK constellation for which $\kappa = -1$ and $K = 1$. The signal power a_1^2 defines the signal-to-noise ratio (SNR). The noise is issued from an autoregressive (AR) process of order 1 and parameter a, so that $[R_T]_{k,l} = a^{|k-l|}$. All other parameters are given in the figure captions.

In Figure 1, the false alarm rate (FAR) and correct detection rate (CDR) performance of the detector proposed in Proposition 2 (consisting in estimating $\hat{k}_T = 1$ among $0, \ldots, L$) is compared against the MDL and AIC detectors (consisting also in finding exactly one source). We observe that the proposed detector uniformly outperforms the MDL and the AIC detectors, consistently with the known inappropriateness of the latter. Note that the AIC particularly fails to detect any source, in spite of N growing, demonstrating the inherent inconsistency of this estimator.

In Figure 2, the receiver cooperation characteristics (ROC), parameterized by ε, for different values of a are depicted. We compare here our proposed detection scheme against an oracle method which assumes perfect knowledge of R_T that is used to whiten Y_T before applying the proposed schemes in the white noise case. We observe that the proposed detector deteriorates with growing a, which can be explained by the

[5]In fact, [21, Th. 2.3] characterizes the asymptotic fluctuations of the random variables $\sqrt{T}(\hat{\lambda}_{i,T} - \boldsymbol{\rho}_{\mathcal{K}(i)})$ instead of the $\sqrt{T}(\hat{\lambda}_{i,T} - \rho_{\mathcal{K}(i),T})$, so that the speed of convergence of ν_T towards ν and of c_T towards $\mathbf{c}$ had to be controlled through [21, Assumption 7]. By replacing $\boldsymbol{\rho}_k$ with $\rho_{k,T}$, the proof of [21, Th. 2.3] goes on without the need for that assumption. Replacing $\boldsymbol{\rho}_k$ by $\rho_{k,T}$ is enough for the present purpose.

natural spread of $\mathrm{supp}(\mu)$ with a large, implying larger inter-eigenvalue spacings within the noise subspace and therefore reduced efficiency of the detection test. On the opposite, the oracle estimator benefits from increased values of a, due to the SNR gain obtained by the whitening procedure. Observe that both approaches perform identically for $a = 0$, which is expected since the system models in both cases are identical.

Figure 3 depicts the normalized mean square error (NMSE) $\mathbb{E}[(\hat{a}_1^2 - a_1^2)^2 a_1^{-4}]$ of the power estimation of Proposition 2 against its theoretical value obtained from Theorem 6. For the purpose of analysis, we assume that the source is always detected, *i.e.* $\hat{k}_T = 1$, irrespective of the SNR. As confirmed by Proposition 4, the theoretical variance diverges as $p_k \downarrow p_{\lim}$. We however observe that in the finite N, T regime, the power estimator errors remain bounded at low SNR. This is explained by the fact that, while the theoretical error diverges due to $\mathbf{\Delta} \downarrow 0$ (see Lemma 2) as $p_k \downarrow p_{\lim}$, its estimator for each N, T (obtained by replacing $\mathbf{m}$ by $\hat{m}_T$) is always non-zero even for $p_k = p_{\lim}$. In the high SNR regime, here with $\kappa = -1$, the NMSE becomes linear (in dB scale) with slope -10 dB/decade. It is easily shown that the limiting SNR gap between the proposed and oracle estimators is exactly

$$10 \log \left(\int_0^1 |\mathbf{p}(\exp(2\imath\pi u))|^2 du \cdot \int_0^1 |\mathbf{p}(\exp(2\imath\pi u))|^{-2} du \right) \text{ dB}$$

which is merely due to a gain in SNR after whitening. In particular, the larger the correlation parameter a, the bigger the limiting gap.

In Figure 4, the mean square error $\mathbb{E}[(\hat{\gamma}(\theta_1) - \gamma(\theta_1))^2]$ of the localization function at position $\theta_1 = 10°$ is compared against the performances of the oracle estimator (which performs pre-whitening prior to using the estimator of [22] or equivalently that of Proposition 3) and of the traditional MUSIC estimator with localization function $\hat{\gamma}_{\mathrm{trad},T}(\theta) \triangleq \sum_{k=1}^{\hat{k}_T} h(\theta)^{\mathsf{H}} \hat{u}_{k,T} \hat{u}_{k,T}^{\mathsf{H}} h(\theta)$ in the notations of Proposition 2. The source is again supposed always detected so that $\hat{k}_T = 1$ throughout the experiment. The proposed estimator outperforms greatly the traditional MUSIC approach here, which is both due to the large N, T regime improvement and to the consideration of the non-white noise setting. The oracle estimator shows a huge performance improvement in the low SNR regime, which translates the fact that condition (4) (which needs to be fulfilled for either method to be valid) is extremely demanding when $a = 0.6$ (due to $\mathrm{supp}(\mu)$ being large). In the large SNR regime, a constant gap is maintained which, although we do not provide theoretical support, appears as a similar SNR-gap phenomenon as observed in Figure 3.

In Figure 5, we now take $K = 2$ sources, with $a_1 = a_2$ the amplitude of which define the SNR, and again assuming $\hat{k}_T = 2$. Here are compared the performances of resolution of two close sources located at $\theta_1 = 10°$ and $\theta_2 = 12°$ for the localization method proposed in Proposition 3, for the oracle estimator, and for the traditional MUSIC estimator. The figure of merit, referred to as resolution probability, is the probability of identifying exactly two local minima of the localization function in the window $[5°, 17°]$. We observe that the proposed algorithm performs significantly better than the traditional MUSIC method, confirming the results of [22] for the current model.

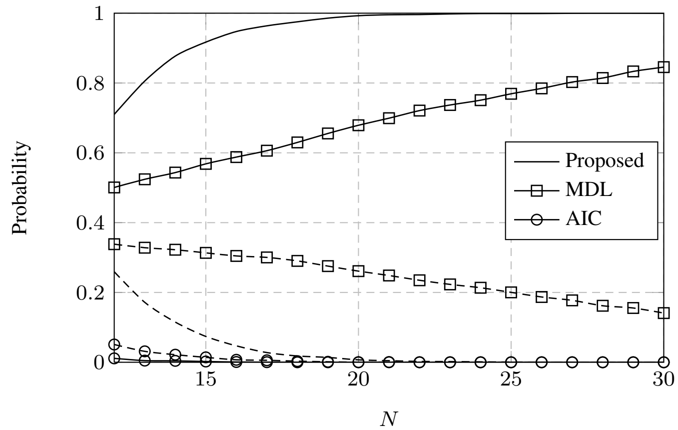

Figure 1. CDR (plain curve) and FAR (dashed curves) versus N with $K = 1$, SNR= 10 dB, $L = 5$, $\varepsilon = 0.75$, $c_T = 0.5$, and $a = 0.6$.

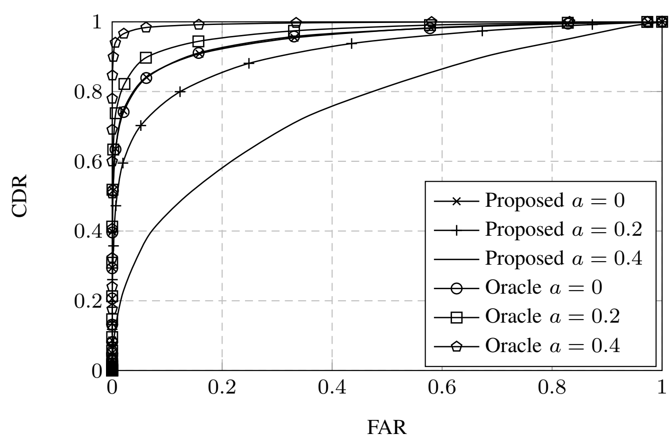

Figure 2. ROC curves with $K = 1$, SNR= 2 dB, $L = 5$, $N = 20$, and $c_T = 0.5$.

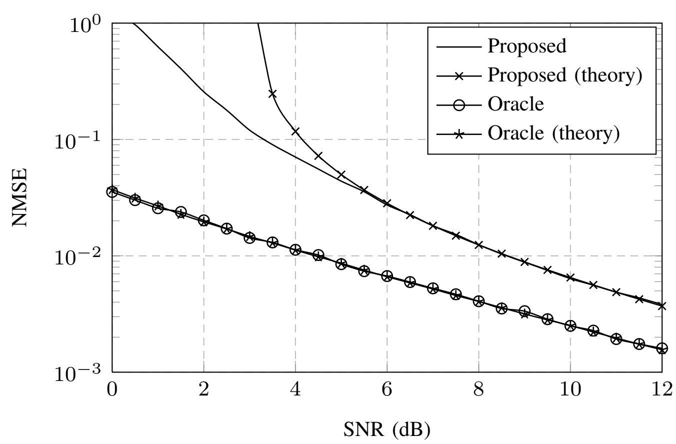

Figure 3. NMSE of the estimated power versus SNR with $K = 1$, $N = 20$, $c_T = 0.5$, and $a = 0.6$.

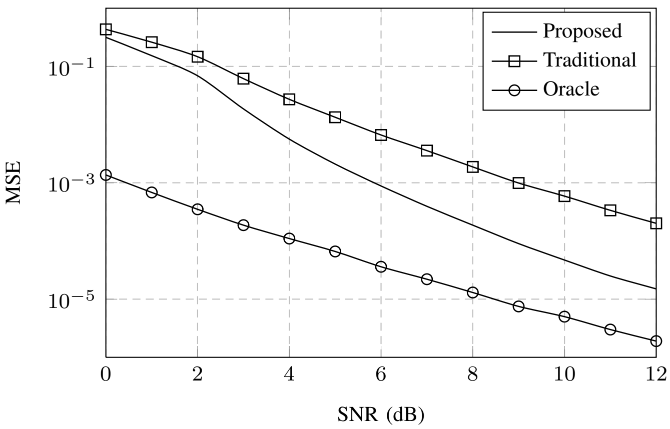

Figure 4. MSE of the localization function versus SNR with $K = 1$, $N = 20$, $c_T = 0.2$, and $a = 0.6$.

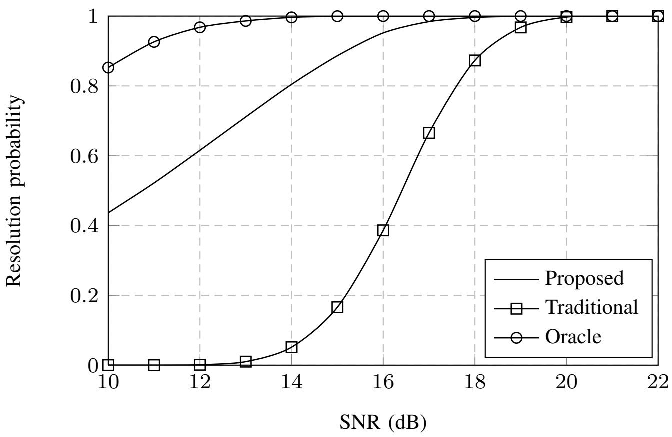

Figure 5. Resolution probability versus SNR with $K = 2$, $N = 20$, $c_T = 0.2$, and $a = 0.6$.

VI. Conclusion and research prospects

This article introduced a novel set of statistical inference methods for large dimensional information-plus-noise models with multiple sources and unknown colored noise. These techniques were proved consistent in the limiting regime where both the system size and the number of observations go large. The approach pursued here relies on the asymptotic spectral separation between noise and signal in the observed sample covariance matrix. Under the same hypotheses, using instead prior information on the noise structure, an alternative approach could consist in estimating the noise covariance in the presence of signals, similar to [26] which treats the noise-only case. It is expected that this approach performs better in the low SNR regime, resurrecting signals unseen by our current method. In the high SNR regime, the covariance estimation will instead be too degraded for this method to be beneficial. A trade-off is therefore expected between both approaches, which we shall study in a future work.

In the specific problem of signal detection, the choice of the eigenvalue "gap parameter" ε does not account for the observation of the small eigenvalues of $Y_T Y_T^{\mathsf{H}}$ as for the power and direction-of-arrival estimation techniques (through $\hat{m}_T$). It seems nonetheless natural to be able to evaluate the right-edge of $\mathrm{supp}(\mu)$ from these eigenvalues, thus resulting in a test to compare $\hat{\lambda}_{i,T}$, $i = 1, \ldots, L$, to the estimated edge. To finely tune the test, one can then use the results from [25] which proves Tracy-Widom fluctuations at the edge with scaling coefficient $\underline{x}''(m_b)$ (m_b given by Corollary 1). However, estimating both the edge and this coefficient constitute a challenging problem so far.

Appendix

A. Proof of Corollary 1

The derivative

$$\mathbf{x}'(m) = \frac{1}{m^2} - \mathbf{c} \int \left(\frac{t}{1 + \mathbf{c}mt} \right)^2 \nu(dt)$$

of $\mathbf{x}(m)$ is continuous and increasing on $(-(\mathbf{c}b_\nu)^{-1}, 0)$, and $\mathbf{x}'(m) \to \infty$ as $m \uparrow 0$. To establish the proposition, it will be enough to show that $\mathbf{x}'(m)i \to -\infty$ as $m \downarrow -(\mathbf{c}b_\nu)^{-1}$. This is obvious when $\nu(b_\nu) > 0$. Assume then $\nu(b_\nu) = 0$. When $m \downarrow -(\mathbf{c}b_\nu)^{-1}$, by the monotone convergence theorem

$$\int \frac{t^2}{(1 + \mathbf{c}mt)^2} \nu(dt) \uparrow \int \frac{t^2}{(1 - t/b)^2} \nu(dt) \geq \int_{[b_\nu - \varepsilon, b_\nu]} \frac{b^2 t^2}{(b-t)^2} f_\nu(t)\, dt = \infty$$

from the behavior of $f_\nu(t)$ near b_ν, which proves the result.

B. Proof of Lemma 2

Considering Equation (2), we obtain after some calculus that $\mathbf{m}'(x) = \mathbf{m}^2(x)/\mathbf{\Delta}(x)$ on (b, ∞). Since $\mathbf{m}(x)$ is negative and increasing on (b, ∞), both $\mathbf{m}'(x)$ and $\mathbf{m}^2(x)$ are positive on this interval so that $\mathbf{\Delta}(x) > 0$ on (b, ∞).

Proposition 1 shows that b coincides with the minimum of $\mathbf{x}(m)$ on $((-\mathbf{c}b_\nu)^{-1}, 0)$. Moreover, when Assumption 4 is satisfied (which is the case for the model (8) by Lemma 1), the proof of Corollary 1 shows that $\mathbf{x}(m)$ attains its minimum at a unique point $m_b \in ((-\mathbf{c}b_\nu)^{-1}, 0)$, and $\mathbf{x}'(m_b) = 0$. Finally, Proposition 1 shows that $\mathbf{x}(m)$ is the inverse of $\mathbf{m}(x)$ on (b, ∞). It results that $\mathbf{m}(x) \to m_b$ and $\mathbf{m}'(x) = 1/\mathbf{x}'(\mathbf{m}(x)) \to \infty$ as $x \downarrow b$. This proves $\mathbf{\Delta}(x) \to 0$ as $x \downarrow b$. When $x \to \infty$, both $(x\mathbf{m}(x))^2 = (\int x(t-x)^{-1}\mu(dt))^2$ and $x^2\mathbf{m}'(x) = \int x^2(t-x)^{-2}\mu(dt)$ converge to 1. Hence, $\mathbf{\Delta}(x) = (x\mathbf{m}(x))^2(x^2\mathbf{m}'(x))^{-1} \to 1$, concluding the proof.

C. Theorem 7: main steps of the proof

For simplicity, we focus on the fluctuations of $\sqrt{T}(\hat{p}_{1,T} - p_1)$. Recall that $\hat{p}_{1,T} = \hat{g}_T(\hat{\lambda}_{1,T})^{-1}$ and $p_1 = g_T(\rho_{1,T})^{-1}$. Define $\underline{g}_T(x) = \underline{m}_T(x)(xc_T\underline{m}_T(x) + c_T - 1)$ with $\underline{m}_T(x)$ defined in Theorem 1-6). We have

$$\begin{aligned}
\sqrt{T}(\hat{p}_{1,T} - p_1) &= \sqrt{T}(\hat{g}_T(\hat{\lambda}_{1,T})^{-1} - g_T(\rho_{1,T})^{-1}) \\
&= \sqrt{T}(\hat{g}_T(\hat{\lambda}_{1,T})^{-1} - \underline{g}_T(\hat{\lambda}_{1,T})^{-1}) \\
&\quad + \sqrt{T}(\underline{g}_T(\hat{\lambda}_{1,T})^{-1} - g_T(\hat{\lambda}_{1,T})^{-1}) \\
&\quad + \sqrt{T}(g_T(\hat{\lambda}_{1,T})^{-1} - g_T(\rho_{1,T})^{-1}) \\
&\triangleq f_{1,T}(\hat{\lambda}_{1,T}) + f_{2,T}(\hat{\lambda}_{1,T}) + f_{3,T}(\hat{\lambda}_{1,T}).
\end{aligned}$$

As $\lambda_{1,T} \xrightarrow{\text{a.s.}} \boldsymbol{\rho}_1$, we can replace $f_{1,T}(\hat{\lambda}_{1,T})$ by $f_{1,T}(\hat{\lambda}_{1,T})\mathbb{1}_I(\lambda_{1,T})$ where $\mathbb{1}_I$ is the indicator function on a small compact interval I in a neighborhood of $\boldsymbol{\rho}_1$. Mimicking the proof of Theorem 4, we can show that $\sup_{x\in I} f_{1,T}(x) \xrightarrow{\mathcal{P}} 0$. We similarly restrict $f_{2,T}$ to I. On this set, it is possible to show that the random process $T(\underline{m}_T(x) - m_T(x))$ valued in the set $C(I)$ of the continuous functions on I, converges in distribution towards a Gaussian process in $C(I)$. This result was shown in [27] for I a compact path of $\mathbb{C}_+$; this can be generalized to the interval I of interest in this proof by using the Gaussian tools used in *e.g.* [21]. As a result, $\sup_{x\in I} f_{2,T}(x) \xrightarrow{\mathcal{P}} 0$. To deal with $f_{3,T}$, we start by observing that $g_T(\rho_{k,T}) \to \mathbf{g}(\boldsymbol{\rho}_k)$ and $(1/g_T(\rho_{k,T}))' \to -\mathbf{g}'(\boldsymbol{\rho}_k)/\mathbf{g}^2(\boldsymbol{\rho}_k) = -p_k^2\mathbf{g}'(\boldsymbol{\rho}_k)$. Using the result of Theorem 6 and applying the Delta method [28, Prop. 6.1.6], we can show that $f_{3,T}(\hat{\lambda}_{1,T}) \xrightarrow{\mathcal{L}} p_1[M_1]_{11}$. The generalization to the vectors $\xi_{k,T}$ defined in the theorem shows no major difficulty.

D. Proof of Proposition 4

From Theorem 2, $\boldsymbol{\rho}_k \downarrow b$ as $p_k \downarrow p_{\lim}$. Hence, by Lemma 2, $\boldsymbol{\Delta}(\boldsymbol{\rho}_k) \to 0$ as $p_k \downarrow p_{\lim}$. Moreover, the proof of this lemma shows that $|\mathbf{m}(\rho_k)|$ remains bounded as $\boldsymbol{\rho}_k \downarrow b$. Hence, since $\nu \neq \delta_0$ by Assumption 3, the integrals in the expression of α_k are lower bounded by a positive number as $p_k \downarrow p_{\lim}$. Thus, $\alpha_k \to \infty$ which proves the first part of the lemma.
When $p_k \to \infty$, $\boldsymbol{\rho}_k/p_k \to 1$ and $\boldsymbol{\rho}_k\mathbf{m}(\boldsymbol{\rho}_k) \to -1$. Taking $p_k \to \infty$ into the expressions of the integrals on the right hand sides of the expressions of α_k, β_k, and ϕ_k and recalling that $\boldsymbol{\Delta}(\boldsymbol{\rho}_k) \to 1$, we get $\alpha_k \to 0$, $\beta_k \to 1$, and $\phi_k \to 1$, which proves the lemma.

E. Lemma 3: sketch of the proof

The fluctuations of quadratic forms of the type $s_T^H D_T s_T$ where $s_T \in \mathbb{C}^T$ has i.i.d. entries have been well studied (*e.g.* [29, Th. 2.1], [30, Th. 3]). Here, the vector s_T is replaced by the matrix $S_T \in \mathbb{C}^{T\times K}$ which introduces some differences in the proof. We follow here the lines of the proof of [30, Th. 3] and stress the main differences.

Let $\sqrt{T}S_T^{\mathsf{H}} = [\mathbf{s}_1, \cdots, \mathbf{s}_T]$ where $\mathbf{s}_t = [s_{t,1}^*, \ldots, s_{t,K}^*]^{\mathsf{T}}$ and let $C = [c_{ij}] \in \mathbb{C}^{K\times K}$ Hermitian matrix. Showing that

$$\sqrt{T}\,\mathrm{Tr}\, C\Big(S_T^{\mathsf{H}} D_T S_T - \frac{1}{T}\mathrm{Tr}\, D_T I_K\Big) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \beta\,\mathrm{Tr}(C^2) + \kappa\alpha\,\mathrm{Tr}[(\mathrm{diag}(C))^2]\right)$$

and invoking the Cramér-Wold device establishes the lemma. Consider the sequence of increasing σ-fields $\mathcal{F}_t = \sigma(\mathbf{s}_1, \ldots, \mathbf{s}_t)$, $t = 1, \ldots, T$, and denote $\mathbb{E}_t$ the expectation conditional to $\mathcal{F}_t$. Then, with $\mathbb{E}_0 = \mathbb{E}$,

$$\begin{aligned}\sqrt{T}\,\mathrm{Tr}\, C\Big(S_T^{\mathsf{H}} D_T S_T - \frac{1}{T}\mathrm{Tr}\, D_T I_K\Big) \\ = \sqrt{T}\sum_{t=0}^{T-1} (\mathbb{E}_{t+1} - \mathbb{E}_t)\,\mathrm{Tr}\, C S_T^{\mathsf{H}} D_T S_T\end{aligned}$$

which is a sum of martingale increments, so that the key tool for establishing Lemma 3 is martingale CLT [31, Th. 35.12]. Writing $Z_t = (\mathbb{E}_{t+1} - \mathbb{E}_t)\,\mathrm{Tr}\, C S_T^{\mathsf{H}} D_T S_T$, we need to show:

- *Lyapunov's condition* : there exists $\delta > 0$ for which
$$T^{1+\delta/2}\sum_{t=0}^{T-1}\mathbb{E}Z_t^{2+\delta} \xrightarrow[T\to\infty]{} 0.$$
- The following convergence holds
$$T\sum_{t=0}^{T-1}\mathbb{E}_t Z_t^2 \xrightarrow[T\to\infty]{\mathcal{P}} \beta\,\mathrm{Tr}(C^2) + \kappa\alpha\,\mathrm{Tr}[(\mathrm{diag}(C))^2].$$

Taking $\delta = 2$ and mimicking the calculus of [30, page 5058] (based on Burkholder's inequality and $\mathbb{E}|s_{1,1}|^8 < \infty$) gives $T^2\sum_{t=0}^{T-1}\mathbb{E}[|(\mathbb{E}_{t+1} - \mathbb{E}_t)[S_T^{\mathsf{H}} D_T S_T]_{i,j}|^4] \to 0$, $1 \le i,j \le K$, which proves Lyapunov's condition. Denoting $D_T = [d_{ij}]$,

$$\begin{aligned}TZ_t &= d_{t+1,t+1}\,\mathrm{Tr}\, C(\boldsymbol{s}_{t+1}\boldsymbol{s}_{t+1}^{\mathsf{H}} - I_K) \\ &\quad + 2\Re\Big(\sum_{i,j=1}^K c_{i,j}\sum_{k=1}^t s_{k,j}^*\, s_{t+1,i}\, d_{k,t+1}\Big).\end{aligned}$$

Using the independence of the $s_{i,j}$ and the moments $\mathbb{E}s_{1,1} = 0$, $\mathbb{E}|s_{1,1}|^2 = 1$, and $\mathbb{E}[s_{1,1}^u(s_{1,1}^*)^v] = 0$ for $u \neq v$, we obtain

$$\begin{aligned}T^2\mathbb{E}_t Z_t^2 &= d_{t+1,t+1}^2\Big(\mathrm{Tr}\, C^2 + \kappa\sum_{k=1}^K c_{kk}^2\Big) \\ &\quad + 2\sum_{i,j,n=1}^K c_{i,j}c_{n,i}\sum_{k,\ell=1}^t s_{k,j}^*\, s_{\ell,n}\, d_{k,t+1}d_{t+1,\ell}.\end{aligned}$$

Letting $\check{D}_T = [d_{ij}\mathbb{1}_{i>j}]$, we have

$$\begin{aligned}T\sum_{t=0}^{T-1}\mathbb{E}_t Z_t^2 &= \Big(\mathrm{Tr}\, C^2 + \kappa\sum_{k=1}^K c_{kk}^2\Big)\frac{1}{T}\,\mathrm{Tr}(\mathrm{diag}(D_T))^2 \\ &\quad + \frac{2}{T}\,\mathrm{Tr}\, C S_T^{\mathsf{H}}\check{D}_T^{\mathsf{H}}\check{D}_T S_T C.\end{aligned}$$

Using [19, Lemma 2.7] and [30, Lemma 3] (or [32, P. 278]), we then get

$$\frac{1}{T}\,\mathrm{Tr}\, C S_T^{\mathsf{H}}\check{D}_T^{\mathsf{H}}\check{D}_T S_T C - \mathrm{Tr}\, C^2\frac{1}{T}\,\mathrm{Tr}\,\check{D}_T^{\mathsf{H}}\check{D}_T \xrightarrow[T\to\infty]{\mathcal{P}} 0.$$

We finally get the result by observing that

$$\frac{2}{T}\,\mathrm{Tr}\,\check{D}_T^{\mathsf{H}}\check{D}_T = \frac{1}{T}\,\mathrm{Tr}\, D_T^2 - \frac{1}{T}\,\mathrm{Tr}(\mathrm{diag}(D_T))^2.$$

F. Proof of Lemma 4

[21, Lemma 3.1] shows that for any compact $K \subset \mathbb{R} - \mathrm{supp}(\mu)$, there exists $C > 0$ such that

$$\forall T \text{ large},\ \forall t \in \mathrm{supp}(\nu_T),\ \inf_{x\in K}|1 + c_T m_T(x)t| > C$$

and hence $\liminf_T \inf_{t\in\mathrm{supp}(\nu_T)}|1 + c_T m_T(\rho_{k,T})t| > 0$. It results that $\limsup_T \|D_T\| < \infty$. Furthermore, since

$$\frac{1}{T}\,\mathrm{Tr}(D_T^2) = \int\frac{p_k^2 m_T(\rho_{k,T})^2}{(1 + c_T m_T(\rho_{k,T})t)^2}\nu_T(dt)$$

the first convergence in the statement of Lemma 4 holds true. As for the second convergence, recall that $R_T = [r_{t-n}]_{1\le t,n\le T}$, with $\sum_t |r_t| < \infty$, and define the Toeplitz matrix $\Gamma_T \triangleq [\gamma_{t-n}]_{1\le t,n\le T}$ where $\gamma_\ell = \boldsymbol{\delta}_\ell + \mathbf{c}\mathbf{m}(\boldsymbol{\rho}_k)r_\ell$. Observe that $D_T = p_k m_T(\rho_{k,T})\Gamma_T^{-1}$. Let $[\cdot]_T$ be the modulo-

T operator, and let $\widetilde{\Gamma}_T = [\gamma_{[t-n]_T}]_{1\leq t,n\leq T}$ be a circulant matrix associated with Γ_T. By [21, Lemma 3.1] again, $\liminf_T \inf_{u\in[0,1]}(1 + c_T m_T(\rho_{k,T})|\mathbf{p}(\exp(2\imath\pi u))|^2) > 0$, hence $\sup_T \|\widetilde{\Gamma}_T\| < \infty$. It results that $T^{-1}\|\Gamma_T^{-1} - \widetilde{\Gamma}_T^{-1}\|_{\text{fro}}^2 \to 0$, with $\|\cdot\|_{\text{fro}}$ the Frobenius norm [23, Th. 5.2]. On the other hand, since $\widetilde{\Gamma}_T$ is circulant, its eigenvector matrix is the Fourier $T\times T$ matrix, so that we can show

$$\operatorname{diag}(\widetilde{\Gamma}_T^{-1}) = \Big(\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{1 + c_T m_T(\rho_{k,T})|\mathbf{p}(\exp(2\imath\pi t/T))|^2}\Big) I_T.$$

The lemma is obtained by combining these last two results.

G. Proof of Lemma 5

We essentially show that we can replace the $B_{k,T}$ by $\sqrt{p_k}S_{k,T}$ with $S_T = [S_{1,T},\ldots,S_{t,T}]$, similar to B_T. Since $\theta_i \neq \theta_j$ if $i\neq j$, from the definition of the vector function $\mathbf{a}(\theta)$, we have $[H_T^{\mathsf{H}}H_T]_{k,\ell} - \boldsymbol{\delta}_{k\ell} = \mathbf{a}_T(\theta_k)^{\mathsf{H}}\mathbf{a}_T(\theta_\ell) - \boldsymbol{\delta}_{k\ell} = \mathcal{O}(1/T)$. Hence, $(H_T^{\mathsf{H}}H_T)^{1/2} \triangleq I_K + E_T$ where $\|E_T\| = \mathcal{O}(1/T)$. Given any sequence D_T of deterministic matrices such that $\sup_T\|D_T\| < \infty$, it can be seen by a moment derivation with respect to the law of S_T that $\mathbb{E}|[B_T^{\mathsf{H}}D_TB_T - P^{1/2}S_T^{\mathsf{H}}D_TS_TP^{1/2}]_{k,\ell}| = \mathcal{O}(1/T)$ for any $k,\ell\leq K$. Hence, by Markov's inequality, $\sqrt{T}(B_T^{\mathsf{H}}D_TB_T - P^{1/2}S_T^{\mathsf{H}}D_TS_TP^{1/2}) \xrightarrow{\mathcal{P}} 0$. Replacing D_T with any of the matrices $p_k m_T(\rho_{k,T})(I_T + c_T m_T(\rho_{k,T})R_T)^{-1}$, we get from Lemma 4 that $\sup_T\|D_T\| < \infty$. Therefore, the $B_{k,T}$ can be replaced with the $\sqrt{p_k}S_{k,T}$. The result is then obtained upon applying Lemmas 3 and 4 and recalling that, for $k = 1,\ldots,t$, the $S_{k,T}$ are independent.

REFERENCES

[1] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[2] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. AC-19, no. 6, pp. 716–723, 1974.

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

[4] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[5] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antenna and Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

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

[7] R. Couillet, J. W. Silverstein, Z. D. Bai, and M. Debbah, "Eigen-inference for energy estimation of multiple sources," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2420–2439, 2011.

[8] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 5113–5129, 2008.

[9] O. Besson, S. Kraut, and L. L. Scharf, "Detection of an unknown rank-one component in white noise," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2835–2839, 2006.

[10] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, "Cooperative spectrum sensing using random matrix theory," in *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*. IEEE, 2008, pp. 334–338.

[11] F. Penna, R. Garello, and M. Spirito, "Cooperative spectrum sensing based on the limiting eigenvalue ratio distribution in Wishart matrices," *Communications Letters, IEEE*, vol. 13, no. 7, pp. 507–509, 2009.

[12] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of statistical tests for single-source detection using random matrix theory," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2400–2419, 2011.

[13] B. Nadler, F. Penna, and R. Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise level," in *Communications (ICC), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–5.

[14] R. R. Nadakuditi and J. W. Silverstein, "Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 468–480, 2010.

[15] B. Nadler and I. M. Johnstone, "Detection performance of Roy's largest root test when the noise covariance matrix is arbitrary," in *Statistical Signal Processing Workshop (SSP), 2011 IEEE*. IEEE, 2011, pp. 681–684.

[16] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[17] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large-dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.

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

[19] 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.

[20] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990.

[21] F. Chapon, R. Couillet, W. Hachem, and X. Mestre, "On the isolated eigenvalues of large Gram random matrices with a fixed rank deformation," submitted. [Online] arXiv:1207.0471.

[22] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet, "A subspace estimator for fixed rank perturbations of large random matrices," *accepted for publication in the Journal of Multivariate Analysis*, 2012, [online] arXiv/1106.1497.

[23] R. M. Gray, *Toeplitz and circulant matrices: A review*, Now Pub., 2006.

[24] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, Chelsea Publishing Co., New York, second edition, 1984.

[25] 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.

[26] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[27] Z. D. Bai and Jack W. Silverstein, "CLT for linear spectral statistics of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 32, no. 1A, pp. 553–605, 2004.

[28] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*, Springer Series in Statistics. Springer, New York, 2006, Reprint of the second (1991) edition.

[29] R. J. Bhansali, L. Giraitis, and P. S. Kokoszka, "Convergence of quadratic forms with nonvanishing diagonal," *Statist. Probab. Lett.*, vol. 77, no. 7, pp. 726–734, 2007.

[30] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim, "A Central Limit Theorem for the SINR at the LMMSE estimator output for large-dimensional signals," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5048–5063, 2009.

[31] P. Billingsley, *Probability and measure*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995.

[32] N. K. Nikolski, *Operators, functions, and systems: an easy reading. Vol. 2*, vol. 93 of *Mathematical Surveys and Monographs*, American Mathematical Society, Providence, RI, 2002.