

Parametric Joint Detection-Estimation of the Number of Sources in Array Processing

Noam Arkind and Boaz Nadler*

Abstract

Detection of the number of signals and estimation of their directions of arrival (DOAs) are fundamental problems in signal array processing. In this paper we present three main contributions to this problem. First, we show that there is an explicit relation between model selection and the breakdown phenomena of the Maximum Likelihood estimator (MLE). Second, we provide a simple formula for the location of the breakdown of the MLE, using tools from extreme value theory. This gives an explicit formula for the signal strength required for reliable detection. Third, we apply these results and propose a new joint detection-estimation algorithm with state-of-the-art performance. We present computer simulations that demonstrate the performance of our algorithm, compared to other popular methods.

Index Terms

joint detection-estimation, statistical hypothesis tests, extreme value theory, MLE breakdown phenomena.

I. INTRODUCTION

Detection of the number of sources impinging on an array of sensors and estimation of their parameters, such as directions of arrival (DOA), source strength, etc., are important and well studied problems in signal array processing (see for example [17], [20], [7]). Some popular algorithms for parameter estimation, such as MUSIC and ESPRIT, assume that the number of signals is a-priori known. These methods rely on non-parametric or semi-parametric algorithms to detect the number of sources [8], [14], [15], [27], [31],

N. Arkind is a student and B. Nadler is a faculty member at the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, 76100, Israel. Phone: (972)-8-9343540, Fax: (972)-8-9346023. E-mail: {noam.arkind,boaz.nadler}@weizmann.ac.il.

[33], [16], [22], [18]. Other approaches, often more computationally intensive, perform joint detection and estimation using the full knowledge of the array geometry [20], [2], [1], [6].

In this paper we consider the problem of *joint detection-estimation* of the number of signals and their DOAs. We propose a method based on a series of hypotheses tests, where for each test, a generalized likelihood ratio test (GLRT) approach is used. Using tools from extreme value theory, we present a statistical analysis of this technique. Our analysis reveals an intimate connection between the model order selection problem, the distribution of extreme values, and the performance breakdown phenomena of the maximum likelihood estimator (MLE). We then derive a simple explicit expression for the location of the MLE breakdown, which is used to set the threshold in the GLRT. Finally, we analyze the model order overestimation probability of the new proposed algorithm. Using computer simulations, we demonstrate the superior detection performance of our algorithm, compared to other popular methods.

II. PROBLEM FORMULATION

We consider the problem of detecting the number of sources impinging on an array of p sensors. While our analysis is quite general, for simplicity we focus on the case of a uniform linear array (ULA), where we derive explicit results.

We thus consider a one-dimensional ULA consisting of p equally-spaced sensors with half-wavelength inter-sensor distance. The time-dependent observations received at the array are assumed to have the following form:

$$\mathbf{x}(t) = \sum_{i=1}^K s_i(t) \mathbf{a}(\theta_i) + \sigma \eta(t) \quad (1)$$

where t is time, K is the number of signals, s_i are the signal waveforms, θ_i are the corresponding directions of arrival (DOA) relative to the array axis, η is a $p \times 1$ complex valued noise vector with distribution $\mathcal{CN}(0, I_p)$, and σ is the noise level. The steering vector $\mathbf{a}(\theta) \in \mathbb{C}^p$ is the response of the array to a unit strength signal emitted from direction θ and depends on the array geometry. In the case of a ULA,

$$\mathbf{a}(\theta)_j = e^{i\pi(j-1)\sin(\theta)}, \quad j = 1, \dots, p. \quad (2)$$

The model (1) can also be written in matrix notation as follows

$$\mathbf{x}(t) = A(\boldsymbol{\Theta})\mathbf{s}(t) + \sigma \eta(t), \quad (3)$$

where $A(\boldsymbol{\Theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$, and $\mathbf{s}(t) = [s_1(t), \dots, s_K(t)]^T$.

We assume that the signals $s_i(t)$ are narrow-band stochastic complex Gaussian processes, with amplitude and phase that vary slowly relative to the propagation time. For simplicity, we also assume that the

noise level is known and without loss of generality $\sigma = 1$. Note that no further assumptions are made on the sources. In particular, the results presented in this paper hold even for fully correlated signals.

The problem we address is as follows: At discrete times t_1, \dots, t_n , we observe n samples $\mathbf{x}(t_1), \dots, \mathbf{x}(t_n)$ from the model (1). Given these n noisy observations, the task is to estimate the unknown number of signals K and their DOAs $\theta_1, \dots, \theta_K$.

III. PREVIOUS WORK

As outlined above, there are two key tasks in signal array processing: i) *Detection* of the number of sources. ii) *Estimation* of the signals parameters. Some methods differentiate the two tasks, solving each separately. Other algorithms address both tasks simultaneously, via joint detection-estimation methods. For stand-alone estimation algorithms, detecting the correct number of sources is crucial, since the success of the estimation step may be very sensitive to detection errors [21].

The various detection algorithms can be divided into several groups, see for example Table 1 in [16]. One group consists of *non-parametric* detection methods, which do not assume any knowledge of the array structure. These methods use only the eigenvalues of the sample covariance matrix

$$\hat{R} = \frac{1}{n} \sum_j \mathbf{x}_j \mathbf{x}_j^H. \quad (4)$$

Two popular non-parametric algorithms to detect the number of signals, based on information theoretic criteria, are AIC (Akaike's information criterion) and MDL (minimum description length) [28], [30]. Statistical analysis of these non-parametric methods reveals a close connection to random matrix theory [18], [19]. In fact, this connection is used in [18] to devise a new non-parametric detection algorithm, with significantly improved performance over the MDL estimator.

A second type consists of *semi-parametric* detection algorithms, which utilize partial knowledge on the sensor array geometry. A common assumption is that sensors can be divided into pairs, with the sensors in each pair separated by a fixed known translation vector. These methods detect the number of signals using the resulting rotational invariance property of the signal subspace ([16], [22] for example).

Another example is the method by Abramovich, Spencer, and Gorokhov [2], [3], which assumes the array is fully augmentable, i.e. the set of all inter-sensor distances is complete $\mathcal{D} = \{1, \dots, p-1\}$. In this approach, a series of (sphericity) tests is used,

$$\mathcal{H}_0 : R_k^{-\frac{1}{2}} \hat{R} R_k^{-\frac{1}{2}} = \sigma^2 I \quad \text{vs.} \quad \mathcal{H}_1 : R_k^{-\frac{1}{2}} \hat{R} R_k^{-\frac{1}{2}} \neq \sigma^2 I,$$

where $R_k \in \Omega_k$, the set of positive definite Toeplitz covariance matrices having the smallest $p-k$ eigenvalues all equal to σ^2 . To decide between the two hypotheses, a GLRT is employed, taking the

following form

$$\Lambda = \left(\frac{\det(\hat{R}_k^{-1} \hat{R})}{[\frac{1}{p} \text{Tr}(\hat{R}_k^{-1} \hat{R})]^p} \right)^n > threshold$$

where $\hat{R}_k \in \Omega_k$ maximizes Λ .

The main reason for the popularity of the various non-parametric and semi-parametric methods is their low computational complexity, which makes them useful for real-time applications. However, it was realized early on that non-parametric methods are typically not optimal for detection, as they do not utilize the full *known* structure of the array manifold. Hence, a third group are *parametric* detection methods which take advantage of full knowledge of the array geometry. Many of these methods perform a series of hypotheses tests to decide between different possible number of sources. For example, Otterstern, Viberg, Stoica and Nehorai [20] proposed a joint detection-estimation scheme, based on a series of decision problems, each solved by a generalized likelihood ratio test (GLRT). For increasing values of the number of signals k , their test decides between the following two hypotheses,

$$\mathcal{H}_0 : R = A(\Theta_k)SA(\Theta_k)^H + \sigma^2 I \quad \text{vs.} \quad \mathcal{H}_1 : R \text{ is arbitrary}$$

where R is the population covariance matrix of the input $\mathbf{x}(t)$, and $S \in \mathbb{C}^{k \times k}$ is the signal covariance matrix. The GLRT takes the form $\Lambda > threshold$, where

$$\Lambda = 2n(\max_{\Theta} \log \det(A\hat{S}A - \hat{\sigma}^2 I) - \log \det(\hat{R})), \quad (5)$$

$A = A(\Theta)$ is the matrix composed of the steering vectors, $\hat{\sigma} = \hat{\sigma}(\Theta)$ is the MLE of the noise power, and $\hat{S} = \hat{S}(\Theta)$ is the MLE of the signal covariance matrix. In [20] it is shown that at the correct model order, the random variable Λ in Eq. (5) follows a Chi-square distribution with $p^2 - k^2 - k - 1$ degrees of freedom, hence the threshold in Eq.(5) can be easily set given any target false alarm rate.

While these parametric methods typically have better detection performance than various non-parametric methods, we claim that their performance may still be significantly improved, in particular for arrays with a large number of sensors. The reason is that the tests employed by these methods compare at each step a specific *parametric* hypothesis against a *completely general* alternative on the covariance matrix, in the sense that their union is the complete parameter space. Hence, these methods have low statistical power in any specific direction and in particular, in directions corresponding to possible additional signals.

In this paper we present a new joint detection-estimation method. It is also based on a series of hypothesis tests, but with a key difference that at each step we test one parametric hypothesis against a parametric alternative. Note that our method is closely related to the algorithm proposed in [6], though our analysis is more extensive, including an appropriate threshold for the corresponding GLRT, and a

theoretical analysis of the asymptotic properties of the algorithm. Our algorithm is also related to the parametric MDL [12], which takes the form

$$\hat{k}_{MDL} = \arg \min_k \left\{ -L(\hat{\Theta}_{ML}^{(k)}) + \frac{1}{2} |\Theta_k| \log n \right\},$$

where $L = \log f(\mathbf{x}|\Theta^{(k)})$ is the log-likelihood function, $\hat{\Theta}_{ML}^{(k)}$ is the MLE and $|\Theta_k|$ is the minimal number of parameters need to characterize the distribution function f completely (for a ULA with K sources, $|\Theta_k| = k(n+1)$). A key result of this paper is that the penalty of the parametric MDL estimator is un-necessarily too large, and can be reduced significantly, leading to better detection performance. Our findings are similar to those in the non-parametric setting, where the $O(\log(n))$ penalty term of MDL is also too large. There, an appropriate penalty is derived from Random Matrix Theory considerations [18], whereas here it is derived using extreme value theory.

As illustrated in section VI, the new proposed method has excellent detection performance compared to the non-parametric method by Kritchman and Nadler [18], the semi-parametric method by Silverstein [22] and the parametric method by Otterstern *et. al.* [20].

IV. MAIN RESULTS

We present three main contributions on the problem of joint detection-estimation of the number of sources and their parameters, under the parametric assumption of a fully known array geometry.

First, we examine the connection between the detection of the number of sources, which is a model selection problem, and the MLE breakdown phenomena. Second, we investigate the properties of the GLRT when a single source is present. Finally we present a joint detection-estimation algorithm with state-of-the-art performance, and a statistical analysis of its asymptotic properties. Proofs of the various theorems appear in section V.

A. Model Selection and The MLE Breakdown Phenomena

For simplicity we first consider the problem of detecting a *single* signal. Later on we extend this analysis to an arbitrary number of signals. We thus consider the case where either no source is present ($K = 0$), or only a single source is present ($K = 1$). Detecting the presence of a signal can be formulated as an hypotheses testing problem as follows:

$$\begin{aligned} \mathcal{H}_0 : & \text{no sources are present } (K = 0) \quad vs. \\ \mathcal{H}_1 : & \text{one source is present } (K = 1). \end{aligned} \tag{6}$$

A standard method to decide between these two hypotheses is via the generalized likelihood ratio test (GLRT). Under the assumption that the noise level σ is known, the GLRT takes the following form [23]

$$T_{n,p}(\hat{\theta}_{ML}) = \frac{1}{n\sigma^2} \sum_{j=1}^n |\langle \mathbf{x}_j, \frac{\mathbf{a}(\hat{\theta}_{ML})}{\|\mathbf{a}(\hat{\theta}_{ML})\|} \rangle|^2 > threshold \quad (7)$$

where $\hat{\theta}_{ML}$ is the maximum likelihood estimator of the unknown DOA. Notice that Eq. (7) holds for any array geometry. The key question is thus how to set the threshold in (7).

To find an appropriate threshold, it is instructive to analyze the behavior of $\hat{\theta}_{ML}$ and of $T_{n,p}(\hat{\theta}_{ML})$ under the different hypotheses \mathcal{H}_0 and \mathcal{H}_1 . First, notice that under \mathcal{H}_0 , when no signals are present, the DOA θ is a *non-existent nuisance parameter* which cannot be estimated. In statistical terms, this is a singular situation, where the GLRT does not follow a χ^2 distribution. Rather, as we show below, its distribution is closely related to the supremum of a certain stochastic process and can be approximated using tools from extreme value theory. This situation is typical in parametric model selection problems, see [9], [10] for similar examples.

In contrast, under \mathcal{H}_1 the DOA θ is an actual physical parameter, and its MLE is asymptotically consistent [26]. However, it is well known that for small sample sizes or weak signals, the MLE suffers from a "breakdown phenomena" in its performance ([1], [5], [25] for example). In more detail, below a certain threshold of either the SNR or of the number of samples, the MLE error increases rapidly, starting to significantly deviate from the Cramer-Rao lower bound. The reason for the large estimation error is the emergence of a random outlier θ' , unrelated to the actual unknown θ , yet having a higher likelihood due to noise. Hence, for sample size too small or signal strength too weak, even if a signal is present, its DOA cannot be reliably estimated.

In typical applications where the DOA is the parameter of interest, it is thus meaningless to declare that a signal is present if its DOA cannot be reliably estimated. This key observation leads to the following model selection or joint detection-estimation principle:

Detect only what can be reliably estimated

(8)

For our analysis we shall call $\hat{\theta}_{ML}$ a reliable estimate of θ if it is located inside the main lobe of the beamformer spectrum (the lobe that corresponds to the true DOA θ), see figure 1 for an example. We remark that if the MLE $\hat{\theta}_{ML}$ is inside the main lobe, then its mean square error is close to the Cramer-Rao lower bound.

Note that the model selection principle in Eq. (8) is quite general and applicable to many parametric model selection problems, where it is of interest not only to detect the presence of signals but also

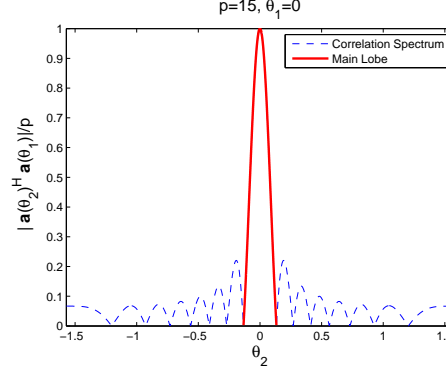


Fig. 1. The correlation $\frac{|\langle \mathbf{a}(\theta_1), \mathbf{a}(\theta_2) \rangle|}{p}$ as a function of θ_2 for fixed θ_1

to estimate their corresponding parameters. This principle suggests that the threshold in Eq. (7) should correspond to the location of the MLE breakdown. Hence it is necessary to estimate this location, and devise a diagnostic test, that for given observed data, assesses if the ML estimate is above or below this threshold. In other words, without knowledge of θ , we need to test whether $\hat{\theta}_{ML}$ is a reliable estimate of θ , and indeed a signal is present, or if the data consists of pure noise (or contains a very weak signal), and thus $\hat{\theta}_{ML}$ is random. This close relation between the model selection problem, the MLE breakdown phenomena and the statistical properties of extremes, is illustrated in figure 2.

B. Single Source Detection and The MLE Breakdown location

Following the reasoning above, we begin by introducing a simple diagnostic test for the presence of a single signal.

Theorem 1. Consider n observations from the model (1) under the null hypothesis \mathcal{H}_0 of no sources ($K = 0$). Let $\alpha \ll 1$ be a desired false alarm rate, and let

$$th(\alpha) = 1 + \frac{C(p, \alpha)}{\sqrt{n}} \quad (9)$$

where

$$C(p, \alpha) = \sqrt{2 \ln \frac{1}{\alpha} + \ln \left(\frac{p^2 - 1}{6} \right)}. \quad (10)$$

Then, as $n \rightarrow \infty$

$$\Pr \left[\max_{-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}} T_{n,p}(\theta) > th(\alpha) | \mathcal{H}_0 \right] \leq \alpha + \Phi^c(C(p, \alpha)), \quad (11)$$

where $\Phi^c(x) = 1 - \Phi(x) = \int_x^\infty \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$.

Note that while Eq. (11) is proved only in the limit $n \rightarrow \infty$, in practice, as demonstrated in section VI, expressions (9), (10) are quite accurate even for small values of n . Furthermore, notice that the term $\Phi^c(C(p, \alpha))$ in Eq. (11) is negligible for practical values of p and α . For example, for $p = 5$ and $\alpha = 0.01$, this term is about 0.0005. From Eq. (10), we see that for small p and $\alpha \ll 1$, the first term is dominant, and the threshold $th(\alpha)$ has a very weak dependence on p .

Theorem 1 implies that if

$$\max_{\theta} T_{n,p}(\theta) > th(\alpha) \quad (12)$$

with $\alpha \ll 1$, then the probability that the data consists only of pure noise is small. Hence condition (12) is a strong statistical indication for the presence of a signal. Consider thus Eq (12) as a diagnostic test for signal detection. There are two key questions associated with this test. The first is which signal strengths can be detected (with high probability) by this test. The second question is whether $\hat{\theta}_{ML}$ is a reliable estimate of θ , assuming that Eq. (12) holds, and that a signal is present. The following theorem provides an answer to both of these questions.

Theorem 2. *Let $\{\mathbf{x}_j\}_{j=1}^n$ be n observations from the model (1) with a single source $s(t)$ ($K = 1$). Consider detection of the signal using Eq. (12) with $th(\alpha)$ given by (9) and $\alpha \ll 1$.*

If the signal strength satisfies

$$\sigma_s^2 = \mathbb{E}[|s|^2] > \frac{C(p, \alpha) + \Phi^{-1}(1 - \epsilon)}{p\sqrt{n}} \quad (13)$$

with $p \gg 1$, then as $n \rightarrow \infty$, the signal $s(t)$ will be detected with probability at least $1 - \epsilon$, and $\hat{\theta}_{ML}$ will be a reliable estimate of θ with probability at least

$$1 - \delta = 1 - 2\Phi^c\left(\frac{1}{2}\left(1 - \frac{1}{\pi^2}\right)(C(p, \alpha) + S_{1-\epsilon})\right). \quad (14)$$

In section VI we show that in practice this theorem is quite accurate even for a relative small number of samples.

From theorem 2 it follows that the right hand side of Eq. (13) is an approximate expression for the location of the MLE breakdown. In particular, for signal strength below this threshold, the probability that $\hat{\theta}_{ML}$ is a reliable estimate of θ is significantly smaller than 1. However, with high probability, such signals will not be detected, since they will not satisfy condition (12). The accuracy of the estimated location of the MLE breakdown is shown in figure 6, see also section VI.

We note that although various papers discuss the MLE breakdown phenomena (see [1], [5], [13], [25] for example), to the best of our knowledge, our analysis is the first one to provide a simple explicit

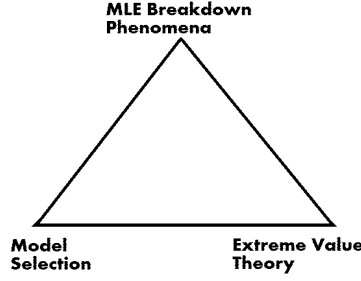


Fig. 2. This figure illustrates the relation between the MLE breakdown, model selection and extreme value theory.

expression for its location, Eq. (13).

It is also interesting to compare the signal strength required for detection in the parametric setting σ_{par}^2 , given by Eq. (13), and in the non-parametric setting, denoted $\sigma_{non-par}^2$. According to [18] Eq. (25), in the non-parametric setting a single signal can be detected with probability at least $\epsilon = \frac{1}{2}$, if the corresponding signal population eigenvalue satisfies

$$\lambda_1 \gtrsim \sqrt{\frac{p-1}{n}} \left(1 + \frac{\sqrt{s(\alpha)(1 - \sqrt{(p-1)/n})}}{(p-1)^{1/3}} \right),$$

where $s(\alpha) = F_2^{-1}(1-\alpha)$ is the inverse of the Tracy-Widom distribution. For a ULA, we have $\lambda_1 = \sigma_s^2 p$. Therefore, for $n \gg 1$, $\sigma_{non-par}^2 = O(\frac{1}{\sqrt{np}})$. Hence, when the array geometry is known, we gain a factor of $\frac{\sigma_{non-par}^2}{\sigma_{par}^2} = O(\sqrt{p})$ on the minimal signal strength required for reliable detection.

C. A Joint Detection-Estimation Scheme

In this section we extend our analysis to the detection of multiple sources and estimation of their DOAs. Let $\mathcal{L}^{(k)}$ be the maximum of the log-likelihood function under the assumption that k signals are present and divided by n ,

$$\mathcal{L}^{(k)} = - \min_{\mathbf{S}^{(k)}, \theta^{(k)}} \frac{1}{n} \sum_{j=1}^n \left\| \mathbf{x}_j - \sum_{r=1}^k s_{r,j} \mathbf{a}(\theta_r) \right\|^2. \quad (15)$$

Note that $\mathcal{L}^{(k)}$ is a monotonic decreasing function with respect to k . Therefore, as is well known, we cannot use the likelihood values themselves for model selection, and an external criteria is required to avoid over-fitting. We propose to use a series of hypotheses decision problems of the form

$$\begin{aligned} \mathcal{H}_k : & \text{at most } k \text{ sources are present } (K \leq k) \quad vs. \\ \mathcal{H}_{k+1} : & \text{at least } k+1 \text{ sources are present } (K \geq k+1). \end{aligned} \quad (16)$$

For each k , we use a generalized likelihood ratio test to distinguish between the two hypotheses. As before, we need to choose an appropriate threshold for each test. We show below that the threshold $th(\alpha)$ in Eq. (9), derived for the detection of a single source, can be also used in the multiple sources scenario. We thus propose the following joint detection-estimation algorithm (denoted *EVT* for extreme value theory):

Algorithm 1 Joint Detection-Estimation

1. Set $k = 0$ and compute $\mathcal{L}^{(0)} = -\frac{1}{n} \sum_{j=1}^n \|\mathbf{x}_j\|^2$.
 2. **while** $k < p$ **do**
 3. Compute $\mathcal{L}^{(k+1)}$.
 4. **if** $\mathcal{L}^{(k+1)} - \mathcal{L}^{(k)} < th(\alpha)$, for some user defined false alarm rate α **then**
 5. return $\hat{k}_{EVT} = k$ and stop.
 6. **else**
 7. set $k = k + 1$.
 8. **end if**
 9. **end while**
 10. **if** all of the tests passed **then**
 11. return $\hat{k}_{EVT} = p$.
 12. **end if**
-

Note that in our algorithm, we need to compute $\mathcal{L}^{(k)}$ for each value of k . This task requires a k dimensional search over the parameter space, and hence it may be computationally intensive. Therefore, various optimization methods were derived which approximate the ML ([20], [24] for example). In our implementation the k -dimensional maxima is computed via an alternating projection technique [29]. First, we use the estimated signals from the previous step as an initial guess for the first $k - 1$ signals, and estimate the k -th signal from the residual. We repeat this procedure, where at each iteration, the parameters of $k - 1$ sources are kept fixed and reduced from the input while a single source is estimated from the residual, via a one dimensional search, until a local maxima is found. Our method is thus computationally fast, and may be used in practical applications.

Next we investigate the asymptotic properties of the proposed algorithm. We introduce the following notation, let $P = A(A^H A)^{-1} A^H$ denote the projection matrix onto $span\{\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)\}$, and let P^\perp be its orthogonal complement, $P \oplus P^\perp = I$. Finally, denote $\mathbf{a}^\perp(\theta) = P^\perp \mathbf{a}(\theta)$ and $\eta^\perp = P^\perp \eta$. The

following proposition gives a bound on the probability to over-estimate the correct number of sources, as $n \rightarrow \infty$.

Proposition 3. *Consider n samples from the model (1) with $K < p - 1$ sources from distinct directions $\Theta = \{\theta_1, \dots, \theta_K\}$. Let $\tilde{T}_{n,p}(\theta) = \frac{1}{n} \sum_{j=1}^n \left| \langle \eta_j, \frac{P^\perp \mathbf{a}(\theta)}{\|P^\perp \mathbf{a}(\theta)\|} \rangle \right|^2$, and let $\tilde{r}(\theta_1, \theta_2)$ be the covariance function of $\tilde{U}_{n,p}(\theta) = \sqrt{n}(\tilde{T}_{n,p}(\theta) - 1)$. Then*

$$\begin{aligned} \tilde{r}_{1,1}(\theta, \theta, P^\perp) &= \frac{\partial^2 \tilde{r}}{\partial \theta_1 \partial \theta_2}(\theta, \theta) = \\ &= 2 \frac{\|P^\perp \mathbf{a}(\theta)'\|^2 \|P^\perp \mathbf{a}(\theta)\|^2 - |\langle \mathbf{a}(\theta)', P^\perp \mathbf{a}(\theta) \rangle|^2}{\|P^\perp \mathbf{a}(\theta)\|^4}. \end{aligned} \quad (17)$$

If

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\tilde{r}_{1,1}(\theta, \theta, P^\perp)} d\theta \leq \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\tilde{r}_{1,1}(\theta, \theta, I)} d\theta, \quad (18)$$

then, as $n \rightarrow \infty$,

$$\Pr \left[\mathcal{L}^{(K+1)} - \mathcal{L}^{(K)} > th(\alpha) | \mathcal{H}_K \right] \leq \alpha + \Phi^c(C(p, \alpha)). \quad (19)$$

In the proof of proposition 3, we show that as $n \rightarrow \infty$, the logarithm of the likelihood ratio is distributed as the maximum of the random process $\tilde{T}_{n,p}(\theta)$. Notice that when no signal is present ($K = 0$), we have $P^\perp = I$ and $\tilde{T}_{n,p}(\theta) = T_{n,p}(\theta)$. Furthermore, this also means that $r_{1,1}(\theta, \theta, I)$ is the derivative of the covariance function of $U_{n,p}(\theta) = \sqrt{n}(T_{n,p}(\theta) - 1)$, which is derived in the proof of theorem 1. Therefore assumption (18) can be used to bound the probability to over-estimate the correct number of sources K , by the probability that $T_{n,p}$ will exceed $th(\alpha)$, given explicitly by Eq. (11) in theorem 1.

We remark that although it is difficult to prove mathematically that Eq. (18) holds, it is possible to verify it numerically. In figure 3, we show $\sqrt{\tilde{r}_{1,1}(\theta, \theta, I)}$, which according to Eq. (27) is just a constant, and $\sqrt{\tilde{r}_{1,1}(\theta, \theta, P^\perp)}$, in the interval $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, where $\Theta = \{-\frac{\pi}{4}, \frac{\pi}{4}\}$. Although there are several locations for which the second function is slightly larger than the first, there is a large region near each θ_i for which $\sqrt{\tilde{r}_{1,1}(\theta, \theta, I)} \gg \sqrt{\tilde{r}_{1,1}(\theta, \theta, P^\perp)}$, hence the integral of $\sqrt{\tilde{r}_{1,1}(\theta, \theta, I)}$ over the entire interval is larger. One can check numerically that this phenomena occurs for any number of angles θ_i in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Finally, we complete this section by showing the weak consistency of our new proposed algorithm.

Theorem 4. *Consider n samples from the model (1) with $K < p - 1$ sources. Then under the assumptions*

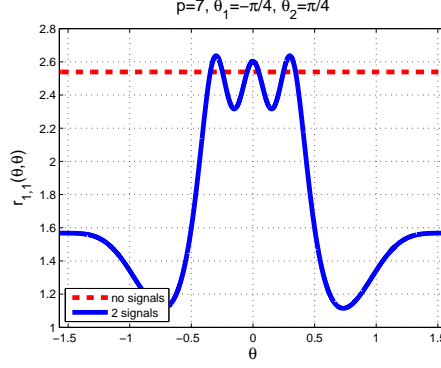


Fig. 3. Comparison between the two covariance function derivatives $r_{1,1}(\theta, \theta, P^\perp)$, and $r_{1,1}(\theta, \theta, I)$, as a function of θ , for a ULA with $p = 7$ sensors. In this simulation, P^\perp is defined by two signals from directions $\theta_1 = 0$ and $\theta_2 = \frac{\pi}{4}$.

of proposition 3, the algorithm above has the following asymptotic properties:

$$\Pr[\hat{k}_{EVT} < K] \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (20)$$

$$\Pr[\hat{k}_{EVT} > K] \leq \alpha + \Phi^c(C(p, \alpha)) \text{ as } n \rightarrow \infty. \quad (21)$$

V. DETAILED ANALYSIS

A. Proof of Theorem 1

To prove theorem 1 we estimate the probability that $\max_{\theta} T_{n,p}(\theta)$ will exceed $th(\alpha)$ under \mathcal{H}_0 as $n \rightarrow \infty$. Note that as a function of $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, the beamformer statistic $T_{n,p}(\theta)$ is a continuous random process. In general, computing the exact probability that a random process exceeds a given threshold is a difficult mathematical problem, and exact solutions are known only for a handful of cases.

However, under quite general assumptions, there are accurate asymptotic approximations for the probability of exceedance ([11], [4], [32] for example). For our purposes, we will use the following well known result for the case of a stationary Gaussian process (proposition 4.1 in [32]).

Proposition 5. Let $\{X(t)\}_{t \geq 0}$ be a real-valued zero mean Gaussian process with unit variance and continuously differentiable sample paths. Let $r(s, t)$ be its covariance function. Denote $M_T = \sup_{t \in [0, T]} X(t)$, then for any value of u ,

$$\Pr[M_T > u] \leq \frac{e^{-\frac{u^2}{2}}}{2\pi} \int_0^T \sqrt{r_{1,1}(t, t)} dt + \Phi^c(u), \quad (22)$$

where $r_{1,1}(s, t) = \frac{\partial^2 r}{\partial s \partial t}(s, t)$.

Next, we apply proposition 5 to the process $U_{n,p}(\theta) = \sqrt{n}(T_{n,p}(\theta) - 1)$, which has zero mean and unit variance. From its definition as a finite sum of smooth functions (Eq. (7)), $T_{n,p}(\theta)$ clearly is continuously differentiable, and hence also $U_{n,p}(\theta)$. Note that under the null hypothesis \mathcal{H}_0 , for any fixed θ , $T_{n,p}(\theta)$ is not Gaussian, rather it is distributed as a $\frac{\chi_{2n}^2}{2n}$ random variable. However, for any fixed θ , as $n \rightarrow \infty$, $U_{n,p}(\theta)$ converges in distribution to a $\mathcal{N}(0, 1)$ Gaussian variable. Finally, its covariance function is given in the following claim.

Claim 6. *The covariance function $r(\theta_1, \theta_2)$ of the random process $U_{n,p}(\theta)$ is given by*

$$r(\theta_1, \theta_2) = \left| \left\langle \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|}, \frac{\mathbf{a}(\theta_2)}{\|\mathbf{a}(\theta_2)\|} \right\rangle \right|^2. \quad (23)$$

In particular, for the case of a ULA, if we consider the following change of variables

$$\theta(\tau) = \arcsin\left(\frac{2\tau}{p} - 1\right) \quad \text{for } \tau \in [0, p], \quad (24)$$

then the process $U_{n,p}(\theta(\tau))$ is stationary. Hence $r(\tau_1, \tau_2) = \rho(\tau_1 - \tau_2)$ and as $\tau_1 \rightarrow \tau_2$,

$$\rho(\tau_1 - \tau_2) = 1 - \frac{\pi^2}{3} \left(1 - \frac{1}{p^2}\right) (\tau_1 - \tau_2)^2 + O(\tau_1 - \tau_2)^4. \quad (25)$$

Proof: Since $\mathbb{E}[U_{n,p}(\theta)] = 0$ for any $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$,

$$\begin{aligned} \text{Cov}(U_{n,p}(\theta_1), U_{n,p}(\theta_2)) &= \\ &= n \mathbb{E}_{\mathbf{x}_1, \dots, \mathbf{x}_n} [(T_{n,p}(\theta_1) - 1)(T_{n,p}(\theta_2) - 1)] \\ &= n \mathbb{E}_{\mathbf{x}_1, \dots, \mathbf{x}_n} [T_{n,p}(\theta_1)T_{n,p}(\theta_2)] - n. \end{aligned}$$

Using the definition of $T_{n,p}$ (Eq. (7)), we obtain

$$T_{n,p}(\theta_1)T_{n,p}(\theta_2) = \frac{1}{n^2} \sum_{j=1}^n \sum_{l=1}^n \left| \left\langle \mathbf{x}_j, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \right\rangle \left\langle \mathbf{x}_l, \frac{\mathbf{a}(\theta_2)}{\|\mathbf{a}(\theta_2)\|} \right\rangle \right|^2.$$

Under the null hypothesis \mathcal{H}_0 , the n random variables $\left| \left\langle \mathbf{x}_j, \frac{\mathbf{a}(\theta)}{\|\mathbf{a}(\theta)\|} \right\rangle \right|^2$, for $j = 1, \dots, n$ are all i.i.d with mean 1. Hence the $n(n-1)$ terms in the double sum above with $j \neq l$, all have mean 1. This gives

$$\begin{aligned} \text{Cov}(U_{n,p}(\theta_1), U_{n,p}(\theta_2)) &= \\ &= \frac{1}{n} \left(n \mathbb{E}_{\mathbf{x}} \left[\left| \left\langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \right\rangle \left\langle \mathbf{x}, \frac{\mathbf{a}(\theta_2)}{\|\mathbf{a}(\theta_2)\|} \right\rangle \right|^2 \right] + n(n-1) \right) - n \\ &= \mathbb{E}_{\mathbf{x}} \left[\left| \left\langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \right\rangle \left\langle \mathbf{x}, \frac{\mathbf{a}(\theta_2)}{\|\mathbf{a}(\theta_2)\|} \right\rangle \right|^2 \right] - 1. \end{aligned}$$

Next, we write $\frac{\mathbf{a}(\theta_2)}{\|\mathbf{a}(\theta_2)\|} = \gamma \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} + \mathbf{b}_1^\perp$, where $\gamma = \langle \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|}, \frac{\mathbf{a}(\theta_2)}{\|\mathbf{a}(\theta_2)\|} \rangle$ and \mathbf{b}_1^\perp is a vector perpendicular to $\mathbf{a}(\theta_1)$. Therefore, the expression above becomes

$$\begin{aligned} &= \mathbb{E}_{\mathbf{x}} \left[\left| \langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \rangle^2 \gamma + \langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \rangle \langle \mathbf{x}, \mathbf{b}_1^\perp \rangle \right|^2 \right] - 1 \\ &= |\gamma|^2 \mathbb{E}_{\mathbf{x}} \left[\left| \langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \rangle \right|^4 \right] + \mathbb{E}_{\mathbf{x}} \left[\left| \langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \rangle \langle \mathbf{x}, \mathbf{b}_1^\perp \rangle \right|^2 \right] + \\ &\quad 2 \operatorname{Re} \left(\gamma \mathbb{E}_{\mathbf{x}} \left[\langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \rangle^3 \langle \mathbf{x}, \mathbf{b}_1^\perp \rangle \right] \right) - 1. \end{aligned}$$

Since \mathbf{b}_1^\perp is perpendicular to $\mathbf{a}(\theta_1)$, the random variables $\langle \mathbf{x}, \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|} \rangle \sim \mathcal{CN}(0, 1)$ and $\langle \mathbf{x}, \mathbf{b}_1^\perp \rangle \sim \mathcal{CN}(0, \|\mathbf{b}_1^\perp\|)$ are independent. Hence, the third term in the sum above equals zero, and we get

$$\operatorname{Cov}(U_{n,p}(\theta_1), U_{n,p}(\theta_2)) = 2|\gamma|^2 + \|\mathbf{b}_1^\perp\|^2 - 1.$$

Finally, one can easily check that $\|\mathbf{b}_1^\perp\|^2 = \|\frac{\mathbf{a}(\theta_2)}{\|\mathbf{a}(\theta_2)\|} - \gamma \frac{\mathbf{a}(\theta_1)}{\|\mathbf{a}(\theta_1)\|}\|^2 = 1 - |\gamma|^2$, and hence Eq. (23) follows.

Notice that Eq. (23), is a very simple expression for the covariance of the stochastic process $U_{n,p}(\theta)$, that holds for any array geometry. For the case of a ULA, the process turns out to be stationary after the change of variables in Eq. (24), as we show below. We insert the definition of $\mathbf{a}(\theta)$ (Eq. (2)) into Eq. (23). Since $\|\mathbf{a}(\theta)\|^2 = p$ for any θ , we get

$$\begin{aligned} \operatorname{Cov}(U_{n,p}(\theta(\tau_1)), U_{n,p}(\theta(\tau_2))) &= \\ &= \frac{1}{p^2} \sum_{j=0}^{p-1} e^{-i \frac{2\pi}{p} j(\tau_1 - \tau_2)} \sum_{k=0}^{p-1} e^{i \frac{2\pi}{p} k(\tau_1 - \tau_2)} \\ &= \frac{1 - e^{-i 2\pi(\tau_1 - \tau_2)}}{1 - e^{-i \frac{2\pi}{p}(\tau_1 - \tau_2)}} \frac{1 - e^{i 2\pi(\tau_1 - \tau_2)}}{1 - e^{i \frac{2\pi}{p}(\tau_1 - \tau_2)}} \\ &= \frac{1}{p^2} \frac{1 - \cos(2\pi(\tau_1 - \tau_2))}{1 - \cos(\frac{2\pi}{p}(\tau_1 - \tau_2))} = \rho(\tau_1 - \tau_2). \end{aligned} \quad (26)$$

Eq. (26) shows that the process $U_{n,p}(\tau)$ is indeed stationary as its covariance depends only on the difference $\tau_1 - \tau_2$. Eq. (25) follows from a Taylor expansion of (26), when $\tau_1 \rightarrow \tau_2$. \blacksquare

Eq. (25) allows us to compute $r_{1,1}(\tau, \tau)$ in the ULA case,

$$r_{1,1}(\tau, \tau) = \frac{\partial^2 r}{\partial s \partial t}(\tau, \tau) = \frac{2\pi^2}{3} \frac{p^2 - 1}{p^2}. \quad (27)$$

Finally, combining Eq. (27) with proposition 5 we bound the probability of exceedance as follows

$$\begin{aligned} \Pr \left[\max_{-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}} \sqrt{n}(T_{n,p}(\theta) - 1) > u | \mathcal{H}_0 \right] &\leq \\ &e^{\frac{-u^2}{2}} \sqrt{\frac{1}{6}(p^2 - 1)} + \Phi^c(u). \end{aligned} \quad (28)$$

Eq. (11) now follows with $u = C(p, \alpha)$ given in Eq. (10). \square

B. Proof of Theorem 2

Let $\{\mathbf{x}_j\}_{j=1}^n$ be n i.i.d. observations from the model (1) with a single source from an unknown direction θ , that is

$$\mathbf{x}_j = s_j \mathbf{a}(\theta) + \eta_j. \quad (29)$$

Since the source is detected only if Eq. (12) is satisfied, we analyze the probability of signal detection, defined as

$$P_d(p, n) = \Pr\left[\max_{-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}} T_{n,p}(\theta) > th(\alpha) \mid \mathcal{H}_1\right]. \quad (30)$$

In particular, we are interested in deriving a sufficient condition on the signal strength, such that $P_d > 1 - \epsilon$, for some $\epsilon \ll 1$. For this purpose, we consider the probability that $T_{n,p}$ exceeds the threshold at some fixed $\tilde{\theta}$,

$$P_d(p, n, \tilde{\theta}) = \Pr[T_{n,p}(\tilde{\theta}) > th(\alpha) \mid \mathcal{H}_1]. \quad (31)$$

We use this probability as a lower bound on (30). Hence we look for a condition on the signal strength such that

$$P_d(p, n) \geq P_d(p, n, \tilde{\theta}) > 1 - \epsilon. \quad (32)$$

Let $\gamma = \langle \mathbf{a}(\tilde{\theta}), \mathbf{a}(\theta) \rangle$. Then, inserting (29) into Eq. (7) gives

$$T_{n,p}(\tilde{\theta}) = \frac{1}{np} \sum_{j=1}^n \left| \langle \mathbf{a}(\tilde{\theta}), \mathbf{x}_j \rangle \right|^2 = \frac{1}{np} \sum_{j=1}^n \left| \gamma s_j + \langle \mathbf{a}(\tilde{\theta}), \eta_j \rangle \right|^2.$$

Since $\eta_j \sim \mathcal{CN}(0, I_p)$ and $\|\mathbf{a}(\theta)\|^2 = p$, we get $\langle \mathbf{a}(\tilde{\theta}), \eta_j \rangle = \sqrt{\frac{p}{2}}(a_j + ib_j)$ where $a_j, b_j \sim \mathcal{N}(0, 1)$.

Thus,

$$\begin{aligned} T_{n,p}(\tilde{\theta}) &= \frac{1}{np} \sum_{j=1}^n \left| \gamma s_j + \sqrt{\frac{p}{2}}(a_j + ib_j) \right|^2 = \\ &= \frac{1}{2n} \sum_{j=1}^n \left\{ (a_j + \sqrt{2} \operatorname{Re}(\frac{\gamma}{\sqrt{p}} s_j))^2 + (b_j + \sqrt{2} \operatorname{Im}(\frac{\gamma}{\sqrt{p}} s_j))^2 \right\}, \end{aligned}$$

and therefore

$$\mathbb{E}[T_{n,p}(\tilde{\theta}) \mid \mathcal{H}_1] = 1 + \frac{|\gamma|^2}{p} \sigma_s^2 \quad (33)$$

where $\sigma_s^2 = \mathbb{E}[|s_j|^2]$. Recall our assumption that the signal $s(t)$ is a narrow-band complex Gaussian random process, hence s_j , for $j = 1, \dots, n$ are all i.i.d, $\mathcal{CN}(0, \sigma_s)$ random variables. Under \mathcal{H}_1 it readily follows that $2nT_{n,p}(\tilde{\theta})$ is distributed according to a non-central chi-square distribution with parameters $k = 2n, \lambda = 2n \frac{|\gamma|^2}{p} \sigma_s^2$. Therefore

$$\mathbb{V}ar[T_{n,p}(\tilde{\theta})] = \frac{2}{4n^2} (k + 2\lambda) = \frac{1}{n} (1 + 2 \frac{|\gamma|^2}{p} \sigma_s^2). \quad (34)$$

Since for $n \gg 1$ it is possible to detect very weak signals, we may assume that the signal strength is small, in particular, $2\frac{|\gamma|^2}{p}\sigma_s^2 \ll 1$. Hence, as $n \rightarrow \infty$, the leading order term in Eq. (34) is $\frac{1}{n}$. Finally, since $T_{n,p}$ is the average of n i.i.d. random variables, for $n \gg 1$, it can be approximated by a Gaussian variable as follows

$$T_{n,p}(\tilde{\theta}) = 1 + \frac{|\gamma|^2}{p}\sigma_s^2 + \frac{1}{\sqrt{n}}\eta \quad (35)$$

where $\eta \sim \mathcal{N}(0, 1)$.

Next, consider the value of $T_{n,p}$ at the true DOA θ . In this case $|\gamma| = p$, and hence

$$T_{n,p}(\theta) \approx 1 + p\sigma_s^2 + \frac{1}{\sqrt{n}}\eta. \quad (36)$$

The approximation in Eq. (36) is useful as it allows us to rewrite Eq. (31), with $\tilde{\theta} = \theta$, as follows,

$$\begin{aligned} P_d(p, n, \theta) &= \Pr \left[1 + p\sigma_s^2 + \frac{1}{\sqrt{n}}\eta > 1 + \frac{C(p, \alpha)}{\sqrt{n}} \right] \\ &= \Pr \left[\eta > -p\sigma_s^2\sqrt{n} + C(p, \alpha) \right] \\ &= \Phi^c \left(-p\sigma_s^2\sqrt{n} + C(p, \alpha) \right). \end{aligned}$$

Therefore Eq. (32) holds, if

$$-p\sigma_s^2\sqrt{n} + C(p, \alpha) < \Phi^{-1}(\epsilon) = -\Phi^{-1}(1 - \epsilon), \quad (37)$$

and after simple manipulations we can obtain Eq. (13). Hence if Eq. (13) holds, then for $n \gg 1$, a signal $s(t)$ will be detected by the test (12) with probability at least $1 - \epsilon$. This proves the first part of the theorem.

Next, we show that under condition (13), with high probability, $\hat{\theta}_{ML}$ is a reliable estimate of the unknown DOA θ . That is, the probability that $T_{n,p}(\theta)$ attains its maximum at an angle outside the main lobe region in the beamformer spectrum (the probability for an outlier) is small. To prove the second part of the theorem, we shall use the following simple claim.

Claim 7. *Let $\tilde{\theta}$ be an angle corresponding to an outlier outside the main lobe of the true DOA θ . Then*

$$|\gamma| = |\langle \mathbf{a}(\tilde{\theta}), \mathbf{a}(\theta) \rangle| \leq \left| \frac{2}{1 - e^{i2\pi/p}} \right|. \quad (38)$$

In particular, for $p \geq 4$, $|\gamma| \leq \frac{p}{\pi} \approx 0.32p$.

Proof: Consider the following change of variables, $\tau = \sin(\theta)$ and $\tilde{\tau} = \sin(\tilde{\theta})$. Then, by definition,

$$\gamma = \langle \mathbf{a}(\tilde{\theta}), \mathbf{a}(\theta) \rangle = \sum_{j=0}^{p-1} e^{i\pi j(\tau - \tilde{\tau})} = \frac{1 - e^{i\pi p(\tau - \tilde{\tau})}}{1 - e^{i\pi(\tau - \tilde{\tau})}}. \quad (39)$$

Notice that for $|\tau - \tilde{\tau}| = \frac{2}{p}m$ where $m \in [1, \dots, p]$, we get $\gamma = 0$. Hence the condition that $\tilde{\theta}$ is outside the main lobe can be written as $|\tau - \tilde{\tau}| > \frac{2}{p}$. Inserting this into (39) yields Eq. (38). For $p \gg 1$, $\frac{i2\pi}{e^p} \approx 1 + \frac{i2\pi}{p}$, hence $|\gamma| \lesssim \frac{p}{\pi}$. We remark that in practice, this bound holds for any $p \geq 4$. ■

Next, we show that if $\tilde{\theta}$ is outside the main lobe of θ , and condition (13) holds, then

$$\Pr[T_{n,p}(\tilde{\theta}) > T_{n,p}(\theta)] \ll 1.$$

Using the approximation of Eq. (35) and (36), the previous probability is

$$\Pr\left[\frac{|\gamma|^2}{p}\sigma_s^2 + \frac{1}{\sqrt{n}}\tilde{\eta} > p\sigma_s^2 + \frac{1}{\sqrt{n}}\eta\right] = \Pr\left[\frac{1}{\sqrt{n}}(\tilde{\eta} - \eta) > \sigma_s^2\left(p - \frac{|\gamma|^2}{p}\right)\right],$$

and using condition (13) and claim 7, we can bound this probability by

$$< \Pr[\tilde{\eta} - \eta > (1 - \frac{1}{\pi^2})(C(p, \alpha) + S_{1-\epsilon})].$$

We bound this probability even further using a union bound

$$\begin{aligned} &\leq \Pr\left[\left\{\tilde{\eta} > \frac{1}{2}\left(1 - \frac{1}{\pi^2}\right)(C(p, \alpha) + S_{1-\epsilon})\right\} \cup \right. \\ &\quad \left.\left\{-\eta > \frac{1}{2}\left(1 - \frac{1}{\pi^2}\right)(C(p, \alpha) + S_{1-\epsilon})\right\}\right] \\ &\leq \Pr\left[\tilde{\eta} > \frac{1}{2}\left(1 - \frac{1}{\pi^2}\right)(C(p, \alpha) + S_{1-\epsilon})\right] + \\ &\quad \Pr\left[-\eta > \frac{1}{2}\left(1 - \frac{1}{\pi^2}\right)(C(p, \alpha) + S_{1-\epsilon})\right] = \delta, \end{aligned}$$

where δ is as in Eq. (14). Note that if $\alpha \ll 1$ and $\epsilon \ll 1$, then $\delta \ll 1$ as well. For example, for $p = 5$, $\alpha = \epsilon = 0.001$, we get $\delta \approx 0.001$.

C. Proof of Proposition 3

Consider the model (1) with K sources impinging on the array from distinct angles $\theta_1, \dots, \theta_K$. We analyze the probability that our proposed algorithm described in section IV will over estimate the true number of sources ($\hat{K} > K$).

Claim 8. *Consider the setting of proposition 3. Then as $n \rightarrow \infty$, the likelihood ratio $\mathcal{L}^{(K+1)} - \mathcal{L}^{(K)}$ is distributed as the maximum of the following stochastic process*

$$\max_{\theta} \tilde{T}_{n,p}(\theta) = \max_{\theta} \frac{1}{n} \sum_{j=1}^n \left| \langle \eta_j^\perp, \frac{\mathbf{a}^\perp(\theta)}{\|\mathbf{a}^\perp(\theta)\|} \rangle \right|^2. \quad (40)$$

Proof: First, recall that if K is known, then $\hat{\Theta}_{ML}$ is a consistent estimate of $\theta_1, \dots, \theta_K$ [26]. Therefore, as $n \rightarrow \infty$,

$$\mathcal{L}^{(K)} \rightarrow -\min_{\hat{\mathbf{S}}^{(K)}} \frac{1}{n} \sum \|A(\Theta)\mathbf{s}_j + \eta_j - A(\Theta)\tilde{\mathbf{s}}_j\|^2. \quad (41)$$

It is easy to see that the minimum value in Eq. (41) is achieved for $\hat{\mathbf{s}}_j = \mathbf{s}_j + A^+ \eta_j$, where $A^+ = (A^H A)^{-1} A^H$ is the Moore-Penrose pseudo inverse matrix of A . Hence

$$\mathcal{L}^{(K)} \rightarrow -\frac{1}{n} \sum_{j=1}^n \|\eta_j^\perp\|^2. \quad (42)$$

This implies that when we test for the correct number of sources K , the likelihood depends only on the part of the noise vector, orthogonal to $\text{span}\{\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)\}$.

Next, we claim that as $n \rightarrow \infty$, the maximum of $\mathcal{L}^{(K+1)}$ is achieved when $\hat{\theta}_l = \theta_l$ for $l = 1, \dots, K$, and for $\hat{\mathbf{s}}_{j,1:K} = \mathbf{s}_j + A^+ \eta_j + A^+ \mathbf{a}(\hat{\theta}_{K+1}) \hat{s}_{j,K+1}$, where $\hat{\mathbf{s}}_{j,1:K}$ are the first K coordinates of $\hat{\mathbf{s}}_j$. This fact is clear by the same proof in [26] for the consistency of $\hat{\Theta}_{ML}$, when the number of sources K is known. Thus, by inserting these values into Eq. (15), we get that as $n \rightarrow \infty$,

$$\mathcal{L}^{(K+1)} \rightarrow -\min_{\tilde{\theta}, \tilde{\mathbf{s}}} \frac{1}{n} \sum \|\eta_j^\perp - \mathbf{a}^\perp(\tilde{\theta}) \tilde{s}_j\|^2. \quad (43)$$

Finally, we follow the same derivation of $T_{n,p}$ that appears in [23], and obtain Eq. (40). \blacksquare

Observe that since P^\perp is an orthogonal projection matrix, then $(P^\perp)^H P^\perp = P^\perp$, and hence $\tilde{T}_{n,p}(\theta)$ can be written as

$$\tilde{T}_{n,p}(\theta) = \frac{1}{n} \sum_j \left| \langle \eta_j, \frac{\mathbf{a}^\perp(\theta)}{\|\mathbf{a}^\perp(\theta)\|} \rangle \right|^2. \quad (44)$$

Note that for any fixed $\theta \notin \{\theta_1, \dots, \theta_K\}$, $\tilde{T}_{n,p}(\theta)$ is distributed as a $\frac{\chi_{2n}^2}{2n}$ random variable. Hence, by the same argument as in the proof of theorem 1, the random process $\tilde{U}_{n,p}(\theta) = \sqrt{n}(\tilde{T}_{n,p}(\theta) - 1)$ tends to a centered Gaussian process. Clearly $\tilde{U}_{n,p}(\theta)$ has continuously differentiable sample paths for $\theta \notin \{\theta_1, \dots, \theta_K\}$. Hence by proposition 5,

$$\begin{aligned} \Pr\left[\max_{-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}} \sqrt{n}(\tilde{T}_{n,p}(\theta) - 1) > u\right] &\leq \\ &\frac{e^{-\frac{u^2}{2}}}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{r_{1,1}(\theta, \theta, P^\perp)} d\theta + \Phi^c(u). \end{aligned} \quad (45)$$

As in Eq. (23), the covariance function of the process $\tilde{U}_{n,p}$ is

$$\tilde{r}(\theta_1, \theta_2) = \left| \left\langle \frac{\mathbf{a}^\perp(\theta_1)}{\|\mathbf{a}^\perp(\theta_1)\|}, \frac{\mathbf{a}^\perp(\theta_2)}{\|\mathbf{a}^\perp(\theta_2)\|} \right\rangle \right|^2. \quad (46)$$

Next, we show that the derivative of the covariance function $\tilde{r}_{1,1}$, can be computed using Eq. (17).

For ease of notation we let $g_1 = \mathbf{a}^\perp(\theta_1)$ and $g_2 = \mathbf{a}^\perp(\theta_2)$. The covariance function is then $\tilde{r}(\theta_1, \theta_2) = \frac{|\langle g_1, g_2 \rangle|^2}{\|g_1\|^2 \|g_2\|^2}$ and its derivative is

$$\begin{aligned} \frac{\partial \tilde{r}}{\partial \theta_1} &= \frac{\langle g'_1, g_2 \rangle \langle g_2, g_1 \rangle + \langle g_1, g_2 \rangle \langle g_2, g'_1 \rangle}{\|g_1\|^2 \|g_2\|^2} \\ &\quad - 2 \frac{\langle g_1, g_2 \rangle \langle g_2, g_1 \rangle}{\|g_1\|^4 \|g_2\|^2} \text{Re}(\langle g'_1, g_1 \rangle), \end{aligned}$$

where $g'_i = \frac{\partial g_i}{\partial \theta_i}$, for $i = 1, 2$. Next,

$$\begin{aligned}
\tilde{r}_{1,1}(\theta_1, \theta_2) &= \frac{\partial^2 \tilde{r}}{\partial \theta_1 \partial \theta_2} = \\
&\frac{2\text{Re}(\langle g'_1, g'_2 \rangle \langle g_2, g_1 \rangle + \langle g'_1, g_2 \rangle \langle g'_2, g_1 \rangle)}{\|g_1\|^2 \|g_2\|^2} \\
&- 2 \frac{\langle g'_1, g_2 \rangle \langle g_2, g_1 \rangle + \langle g_1, g_2 \rangle \langle g_2, g'_1 \rangle}{\|g_1\|^2 \|g_2\|^4} \text{Re}(\langle g'_2, g_2 \rangle) \\
&- 2 \frac{\langle g_1, g'_2 \rangle \langle g_2, g_1 \rangle + \langle g_1, g_2 \rangle \langle g'_2, g_1 \rangle}{\|g_1\|^4 \|g_2\|^2} \text{Re}(\langle g'_1, g_1 \rangle) \\
&+ 4 \frac{\langle g_1, g_2 \rangle \langle g_2, g_1 \rangle}{\|g_1\|^4 \|g_2\|^4} \text{Re}(\langle g'_1, g_1 \rangle) \text{Re}(\langle g'_2, g_2 \rangle).
\end{aligned} \tag{47}$$

Finally, we evaluate the above at $\theta_1 = \theta_2 = \theta$ (denote $g_1 = g_2 = g$):

$$\begin{aligned}
\tilde{r}_{1,1}(\theta, \theta) &= \\
&= 2 \frac{\|g'\|^2}{\|g\|^2} + 2 \frac{\text{Re}(\langle g', g \rangle^2)}{\|g\|^4} - 4 \frac{\text{Re}(\langle g', g \rangle)^2}{\|g\|^4} \\
&= 2 \frac{\|g'\|^2}{\|g\|^2} + \frac{\langle g', g \rangle^2 + \langle g, g' \rangle^2}{\|g\|^4} - \frac{(\langle g', g \rangle + \langle g, g' \rangle)^2}{\|g\|^4} \\
&= 2 \left(\frac{\|g'\|^2}{\|g\|^2} - \frac{|\langle g', g \rangle|^2}{\|g\|^4} \right).
\end{aligned}$$

Notice that by the linearity of the derivative, $\mathbf{a}^\perp(\theta)' = (P^\perp \mathbf{a}(\theta))' = P^\perp \mathbf{a}(\theta)'$, and therefore Eq. (17) follows.

Finally, from Eq. (45) and our assumption in Eq. (18), it follows that

$$\begin{aligned}
&\Pr[\max_{\theta} \tilde{T}_{n,p}(\theta) > th(\alpha)] \leq \\
&\frac{e^{-C(p,\alpha)^2/2}}{2\pi} \int_{-\pi/2}^{\pi/2} \sqrt{\tilde{r}_{1,1}(\theta, \theta, P^\perp)} d\theta + \Phi^c(C(p, \alpha)) \leq \\
&\frac{e^{-C(p,\alpha)^2/2}}{2\pi} \int_{-\pi/2}^{\pi/2} \sqrt{\tilde{r}_{1,1}(\theta, \theta, I)} d\theta + \Phi^c(C(p, \alpha)).
\end{aligned}$$

Observe that $\tilde{r}_{1,1}(\theta, \theta, I)$ corresponds to the case where no signals are present ($K = 0$), hence we can use Eq. (27). Therefore the expression above can be written as

$$\begin{aligned}
&= \frac{e^{-C(p,\alpha)^2/2}}{2\pi} \sqrt{\frac{2}{3}(p^2 - 1)\pi^2} + \Phi^c(C(p, \alpha)) \\
&= \alpha + \Phi^c(C(p, \alpha)).
\end{aligned}$$

D. Proof of Theorem 4

We investigate the probability to under-estimate the correct number of sources. Observe that as $n \rightarrow \infty$ and for any $k < K$,

$$\mathcal{L}^{(k)} - \mathcal{L}^{(k-1)} > \mathcal{L}^{(k+1)} - \mathcal{L}^{(k)},$$

hence, it suffices to show that with high probability

$$\mathcal{L}^{(K)} - \mathcal{L}^{(K-1)} > th(\alpha) \quad \text{as } n \rightarrow \infty. \quad (48)$$

First, consider Eq. (42). Since P^\perp is a projection matrix onto a $p - K$ dimensional space,

$$-\mathcal{L}^{(K)} \rightarrow \mathbb{E}[\|\eta^\perp\|^2] = \mathbb{E}[\|P^\perp \eta\|^2] = p - K.$$

Similarly, let $\hat{\theta}_1, \dots, \hat{\theta}_{K-1}$ be the minimizer of $\mathcal{L}^{(K-1)}$. It follows from the unambiguity property of the vectors $\mathbf{a}(\theta)$ (for a ULA) [20], that the dimension of the signal subspace is K . Therefore projecting the signal subspace into a $p - (K - 1)$ dimensional subspace, must necessarily leave a non-zero residual $\mathbf{r} = A(\boldsymbol{\Theta})\mathbf{s}(t) - A(\hat{\boldsymbol{\Theta}})\hat{\mathbf{s}}(t)$. Therefore, as $n \rightarrow \infty$,

$$\begin{aligned} -\mathcal{L}^{(K-1)} &\rightarrow \mathbb{E}[\|A(\boldsymbol{\Theta})\mathbf{s}(t) + \eta - A(\hat{\boldsymbol{\Theta}})\hat{\mathbf{s}}(t)\|^2] \\ &= \mathbb{E}[\|\eta^\perp + \mathbf{r}\|^2] \\ &= \mathbb{E}[\|\eta^\perp\|^2 + \|\mathbf{r}\|^2 + 2\text{Re}(\langle \mathbf{r}, \eta^\perp \rangle)]. \end{aligned}$$

Since \mathbf{r}, η^\perp are independent, the expectation of their inner product is zero. Hence, we get

$$-\mathcal{L}^{(K-1)} \rightarrow P - (K - 1) + \mathbb{E}[\|\mathbf{r}\|^2].$$

Combining these observations, we get

$$\begin{aligned} \mathcal{L}^{(K)} - \mathcal{L}^{(K-1)} &\rightarrow P - (K - 1) + \mathbb{E}[\|\mathbf{r}\|^2] - (P - K) \\ &= 1 + \mathbb{E}[\|\mathbf{r}\|^2] > 1. \end{aligned}$$

Finally, notice that the threshold $th(\alpha)$ in Eq. (9) tends to 1 as $n \rightarrow \infty$. Thus, the probability to under-estimate the correct number of sources tends to zero as $n \rightarrow \infty$.

VI. SIMULATIONS

In this section we present a series of simulations that illustrate the finite sample behavior of Theorems 1, 2, and 4, as well as the detection performance of our algorithm in comparison to some other common detection methods.

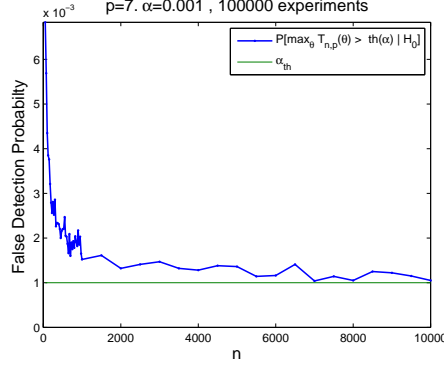


Fig. 4. The probability that the test (12) passed, when no signal is present, as a function of the number of samples.

Throughout this section, all simulations are conducted using a ULA with either $p = 7$ or $p = 15$ equally spaced sensors, and with a false alarm rate of $\alpha = 0.001$. In addition, all sources are distributed as $Ze^{i\phi}$ where ϕ is uniformly distributed in $[0, 2\pi]$ and Z is a real-valued Gaussian random variable with mean σ_s and standard deviation 0.01. In all of our simulations, the signal strengths satisfy $\sigma_s \gg 0.01$ so that $\mathbb{E}[|s|^2] \approx \sigma_s^2$.

In figure 4 we show the probability that the test in Eq. (12) passed, when no signal was present. The results are in accordance to Eq. (11) in theorem 1. That is, the empirical false alarm probability converges to α as $n \rightarrow \infty$.

Figure 5 shows the probability that the test (12) failed e.g. that a signal was not detected despite its presence. In addition we also show the limit of detection from Eq. (13) according to theorem 2. We can see that if the signal strength is *above* the threshold in Eq. (13), then the probability that the signal was not detected is close to $\epsilon = 0.01$, and drops below ϵ as n increases. For $n = 100$, the probability is $\approx 2.5\%$, while for $n = 500$, the probability is $\approx 0.8\%$.

Figure 6 shows the mean square error of $\hat{\theta}_{ML}$, when a single signal was present, as a function of the signal strength σ_s^2 . We also included the limit of detection (Eq. (13)). As explained in section IV, this expression can be used as an approximation for the location of the MLE performance breakdown. As n increases, this approximation yields a better prediction for the breakdown location. In addition, we show the log MSE computed only from the experiments where the test in Eq. (12) passed, excluding the cases when a signal was not detected by our method. Notice that in this case there is no breakdown phenomena, in accordance to theorem 2.

Finally, in fig 7 we compare the detection performance of our algorithm to the following methods: The

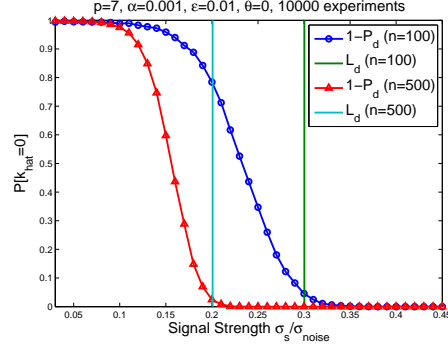


Fig. 5. The probability that the test (12) failed, when a single source is present, as a function of the signal strength. P_d is the probability of detection, and L_d is the right hand side of Eq. (13) (the limit of detection).

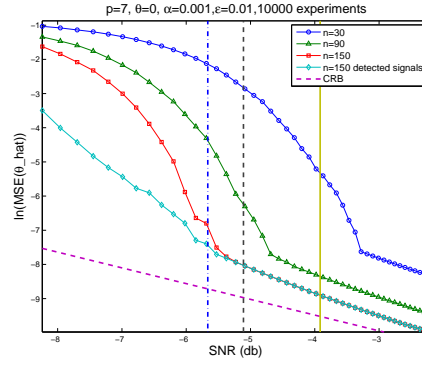


Fig. 6. MSE of the MLE $\hat{\theta}_{ML}$ as a function of the SNR, for three different values of n . L_d is the right hand side of Eq. (13). For $n = 150$ we show the MSE given that Eq. (12) holds, as well as the Cramer-Rao bound.

non-parametric detection algorithm by Kritchman and Nadler based on random matrix theory (*RMT*) [18], the parametric *GLRT* based method of Otterstern, Viberg, Stoica and Nehorai [20], and the semi-parametric method based on *ESPRIT* by Silverstein [22]. In the top panels, $p = 7$ and in the bottom panels $p = 15$. In the left panels, three signals strengths and their DOAs were fixed, while the number of samples increased from $n = 15$ to $n = 500$. In the right panels, the number of samples was fixed ($n = 90$), while three signals from fixed directions were present. Two of those signals were kept fixed, while the strength of the third varied between $s = 0.1$ and $s = 0.5$. As shown in the figure our method has better detection performance at a comparable false alarm rate. Notice that for $p = 7, n \geq 140$ and for $p = 15, n \geq 50$ the probability that $\hat{k}_{EVT} \neq K$ is less than $\alpha = 0.1\%$ in agreement with theorem 4. For comparison, the number of samples required to get the same detection probability for the non-parametric

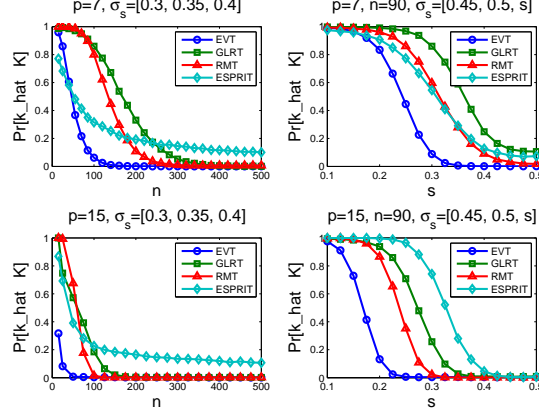


Fig. 7. The probability to detect the wrong number of signals, when three signals are present with directions $\Theta = [-\frac{\pi}{4}, 0, \frac{\pi}{4}]$, as the number of samples or the signal strength increases.

method *RMT* is $n \geq 300$ for $p = 7$, and $n \geq 125$ for $p = 15$.

VII. SUMMARY

In this paper we presented a statistical analysis of the problem of detecting the number of sources and estimating their parameters, in a parametric setting. The main contribution of this paper is an explicit connection between the model order selection problem, the MLE performance breakdown and extreme value theory. In particular, we select the maximal model order for which the MLE is a reliable estimate. For that reason, we derived a simple expression for the location of the MLE breakdown and used it in our algorithm. We used computer simulations to show that the new proposed method has improved detection performance compared to other common detection algorithms.

ACKNOWLEDGMENTS

It is our pleasure to thank Yonina Eldar and Anthony Weiss for interesting discussions.

REFERENCES

- [1] Y.I. Abramovich, B.A. Johnson, X. Mestre, MUSIC, G-MUSIC, and Maximum-Likelihood Performance Breakdown, *IEEE Tran. Sig. Proc.*, vol 56, no. 8, Aug. 2008, pp. 3944-3957.
- [2] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov, Detection-Estimation of More Uncorrelated Gaussian Sources than Sensors in Nonuniform Linear Antenna ArraysPart I: Fully Augmentable Arrays, *IEEE Trans. Sig. Proc.*, vol. 49, no. 5, May 2001, pp. 959-971.

- [3] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov, Bounds on Maximum Likelihood Ratio-Part I: Application to Antenna Array Detection-Estimation with Perfect Wavefront Coherence, *IEEE Tran. Sig. Proc.*, vol. 52, no. 6, June 2004, pp. 1524–1536.
- [4] R. J. Adler, On excursion sets, tube formulas and maxima of random fields, *The Annals of Applied Probability*, Vol. 10, No. 1, 2000.
- [5] F. Athley, Threshold region performance of maximum likelihood direction of arrival estimators, *IEEE Tran. Sig. Proc.*, vol. 53, no. 4, pp. 1359–1373, 2005.
- [6] R.E. Bethel, K.L. Bell, Maximum Likelihood Approach to Joint Array Detection/Estimation, *IEEE Tran. on Aerospace and Electronic Systems* vol. 40, no. 3, July 2004, pp. 1060–1072.
- [7] J. F. Böhme, "Array processing", in *Advances in Spectrum Analysis and Array Processing*, vol. 2, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice- Hall, 1991, pp. 1–63.
- [8] P. Chen, M.C. Wicks and R.S. Adve, Development of a statistical procedure for detecting the number of signals in a radar measurement, *IEE Proc. Radar Sonar Navig.*, vol. 148, no. 4, 2001, ch 1, pp. 219–226.
- [9] D. Dacunha-Castelle and E. Gassiat, Testing the order of a model using locally conic parameterization: population mixtures and stationary arma process, *The Annals of Statistics*, vol. 27, no. 4, 1999 pp. 1178–1209.
- [10] R. B Davies, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, vol. 74, no. 1, pp. 33–43, 1987.
- [11] M. Falk, J. Husler and R.D. Reiss, *Laws of Small Numbers: Extremes and Rare Events*, Second edition, Birkhauser, 2004.
- [12] E. Fishler, M. Grossmann, H. Messer, Detection of Signals by Information Theoretic Criteria: General Asymptotic Performance Analysis, *IEEE Tran. Sig. Proc.*, vol. 50, no. 5, May 2002, pp. 1027–1036.
- [13] P. Forster, P. Larzabal, E. Boyer, Threshold Performance Analysis of Maximum Likelihood DOA Estimation, *IEEE Tran. Sig. Proc.*, vol. 52, no. 11, Nov. 2004, pp. 3183–3191.
- [14] L. Huang, S. Wu, Low-complexity MDL method for accurate source enumeration, *IEEE Sig. Proc. Lett.*, vol. 14, no. 9, Sept. 2007, pp. 581–584.
- [15] L. Huang, S. Wu, X. Li, Reduced-rank MDL method for source enumeration in high-resolution array processing, *IEEE Tran. Sig. Proc.*, vol. 55, no. 12, Dec. 2007, pp. 5658–5667.
- [16] J.A. Jiang, M.A. Ingram - Robust Detection of Number of Sources Using the Transformed Rotational Matrix, *IEEE Wir. Comm. Net. Conf.*, Mar. 2004.
- [17] H. Krim ,M. Viberg, Two Decades of Signal Array Processing Research, The Parametric Approach, *IEEE Sig. Proc. Mag.*, vol. 13, no.4, pp. 67–94, July 1996.
- [18] S. Kritchman, B.Nadler, Non-Parametric Detection of the Number of Signals: Hypothesis Testing and Random Matrix Theory, *IEEE Trans. Sig. Proc.*, vol. 57, no. 10, Oct. 2009, pp. 3930–3941.
- [19] B. Nadler, Non-Parametric Detection of Signals by Information Theoretic Criteria: Performance Analysis and an Improved Estimator, submitted to *IEEE Tran. Sig. Proc.*.
- [20] B. Otterstern, M. Viberg, P. Stoica, A. Nehorai, Exact and large sample ML techniques for parameter estimation and detection in array processing, "Radar Array Processing" Simon Haykin, Springer-Verlag, Feb. 1993, Chapter 4.
- [21] A. Paulraj, R. Roy, T. Kailath, Estimation of Signal Parameters via Rotational Invariance Techniques - ESPRIT, *IEEE Tran. Sig. Proc.*, vol. 37, no. 7, July 1989, pp. 984–995.
- [22] S.D. Silverstein - A new use of ESPRIT for detection and model order estimation, *IEEE Sig. Proc. Lett.*, vol. 1, no. 10, Oct. 1994, pp. 147–149.

- [23] P. Stoica, O. Besson, Maximum likelihood DOA estimation for constant-modulus signal, *Electronics Letters*, vol. 36, no. 9, pp. 849–851, 2000.
- [24] P. Stoica, O. Besson, F. Vincent and A.B. Gershman, Approximate Maximum Likelihood Estimators for Array Processing in Multiplicative Noise Environments, *IEEE Tran. Sig. Proc.*, vol. 48, no. 9, pp. 2506–2518, 2000.
- [25] P. Stoica, M. Hawkes, A. Nehorai, Performance Breakdown of Subspace Based Methods: Prediction and Cure, *Proc. ICASSP* May 2001, pp. 4005-4008.
- [26] P. Stoica and A. Nehorai, MUSIC, Maximum Likelihood, and Cramer-Rao Bound *IEEE Tran. Acoust., Speech, Sig. Proc.*, vol. 17, no. 5, May 1989, pp. 720–741.
- [27] M.O. Ulfarsson, V. Solo, Dimension Estimation in noisy PCA with SURE and random matrix theory, *IEEE Tran. Sig. Proc.* vol. 54, no. 12, Dec. 2008, pp. 5804-5816.
- [28] M. Wax and T. Kailath, Detection of signals by information theoretic criteria, *IEEE Tran. Acoust., Speech, Sig. Proc.*, vol. 33, no. 2, pp. 387-392, Apr. 1985.
- [29] M. Wax, I. Ziskind, Maximum Likelihood Localization of Multiple Sources by Alternating Projection, *IEEE Tran. Acoust., Speech, Sig. Proc.*, vol. 36, no. 10, pp. 1553-1560, Oct. 1988.
- [30] M. Wax, I. Ziskind, Detection of the Number of Coherent Signals by the MDL Principle, *IEEE Tran. Acoust., Speech, Sig. Proc.*, vol. 37, no. 8, pp. 1190-1196, 1989.
- [31] K.M. Wong, Q-T. Zhang, J.P. Reilly and P.C. Yip, On information theoretic criteria for determining the number of signals in high resolution array processing, *IEEE Tran. Acoust., Speech, Sig. Proc.*, vol. 38, no. 11, Nov. 1990, pp. 1959-1971.
- [32] M. Wschebor and J.M. Azaïs, *Level Sets and Extrema of Random Processes*, Wiley, 2009.
- [33] J. Xin, N. Zheng and A. Sano, Simple and efficient nonparametric method for estimating the number of signals without eigendecomposition, *IEEE Tran. Sig. Proc.*, vol. 55, no. 4, Apr. 2007, pp. 1405-1420.