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

Noam Arkind, Boaz Nadler
Department of Computer Science and Applied Mathematics
Weizmann Institute of Science
Rehovot, Israel 76100
Email: arkind.noam@gmail.com, boaz.nadler@weizmann.ac.il

***Abstract*—Detection of the number of signals and estimation of their directions of arrival (DOAs) are fundamental problems in array processing. We present three main contributions to these problems, under the conditional model, where signal amplitudes are assumed deterministic unknown. First, we show that there is an explicit relation between model selection and the breakdown phenomena of the Maximum Likelihood estimator (MLE). Second, for the case of a single source, we provide a simple approximate formula for the location of the breakdown of the MLE, using tools from the maxima of stochastic processes. This gives an explicit formula for the source 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 demonstrate via simulations the improved detection performance of our algorithm, compared to other popular source enumeration methods.**

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 [6], [9]. Most 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.

In this paper, in contrast, we focus on the often more computationally intensive though more accurate *joint detection-estimation* methods, that use full knowledge of the array geometry. We consider the conditional model [13], where signal amplitudes are assumed deterministic unknown. Further, we assume that the estimation of the signal amplitudes and DOA's is done by Maximum Likelihood (ML). The key questions we address are thus the following: i) Given the likelihood values for different model orders, detect the number of sources; ii) determine which source strengths can be reliably detected together with an accurate estimate of their DOA's. Our statistical analysis, based on the theory of maxima of stochastic processes, reveals an intimate connection between the model order selection problem, the distribution of extreme values of certain stochastic processes, and the performance breakdown phenomena of the maximum likelihood estimator (MLE). This yields a simple explicit expression for the location of the MLE breakdown for the case of a single source, as well as a new joint detection-estimation algorithm based on a sequence of GLRT-type hypothesis tests. Using simulations, we demonstrate the superior detection performance of our algorithm, compared to other source enumeration methods.

II. Problem Set-Up

We consider the following standard model for observations received at an array of p sensors,

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

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), 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.

We focus on the conditional model, where the signals $s_i(t)$ are assumed deterministic unknown processes. For simplicity, we also assume that the noise level is known and w.l.g. $\sigma = 1$.

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

While our analysis is general and applies to arrays of arbitrary geometry, for simplicity we analyze in detail the case of a uniform linear array (ULA), where we derive explicit results. For a ULA with half-wavelength inter sensor spacing,

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

III. Previous Work

As outlined above, two key tasks in array processing are: 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.

Source detection methods can be divided into several groups [5]. One group consists of *non-parametric* methods, which do not assume any knowledge of the array structure, and use only the eigenvalues of the sample covariance matrix

$$\hat{R} = \frac{1}{n}\sum_j \mathbf{x}_j \mathbf{x}_j^H. \tag{3}$$

Popular methods include the AIC and MDL estimators [14]. These methods can be both analyzed and significantly improved using results from random matrix theory, see [7], [8].

A second group of *semi-parametric* detection algorithms, utilize partial knowledge on the array geometry, such as rotational invariance of the signal subspace [5], [11]. For example, for a fully augmentable array, Abramovich et al. [2] proposed to detect the number of sources by the following series of (sphericity) tests,

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

where $R_k \in \Omega_k$, the set of positive definite Toeplitz covariance matrices with the smallest $p-k$ eigenvalues all equal to σ^2.

The popularity of non-parametric and semi-parametric methods is due to their low computational complexity, which makes them useful for real-time applications. These methods, however, are not optimal for detection, as they do not utilize the full *known* structure of the array manifold. In this paper we thus focus on the third group of *parametric* detection methods which take advantage of full knowledge of the array geometry. Many of these methods determine the number of sources via a series of hypotheses tests. For example, Ottersten et al. [9] considered a stochastic signal model and proposed a joint detection-estimation scheme, by testing the following hypotheses for increasing values of k,

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

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.

While these parametric methods 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 at each step, these methods compare 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. Our method is closely related to [4], though there the authors considered a stochastic signal model and did not present a theory for setting the appropriate threshold for the corresponding GLRT. Our method is also related to the parametric MDL estimator [15],

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

where $L = \log f(\mathbf{x}|\mathbf{\Theta}^{(k)})$ is the log-likelihood function, $\hat{\mathbf{\Theta}}^{(k)}$ is the MLE and $|\mathbf{\Theta}_k|$ is the number of parameters needed to fully characterize the data's distribution function (for a ULA with K sources, $|\mathbf{\Theta}_k| = k(n+1)$). A key result of this paper is that the penalty of the parametric MDL estimator is unnecessarily too large, and can be reduced significantly, leading to improved detection performance.

As illustrated in section V, the new proposed method has excellent detection performance compared to the non-parametric method of [7], the semi-parametric method of [11] and the parametric GLRT method by Ottersten *et al.* [9].

IV. Main Results

A. Model selection and maxima of random processes

Let $\hat{\mathbf{\Theta}}^{(k)} = \{\hat{\theta}_1, \ldots, \hat{\theta}_k\}$ be the ML estimates of the k DOA's assuming a model order k. Our proposed procedure for joint detection estimation of the number of sources is the following: For increasing values of k, decide between

$$\mathcal{H}_0 : k \text{ sources} \qquad vs. \quad \mathcal{H}_1 : (k+1) \text{ sources}. \tag{4}$$

via the following GLRT,

$$G_k = \frac{1}{n} \ln \left(\frac{\mathcal{L}(\mathbf{x}_i, \hat{\mathbf{\Theta}}^{(k+1)})}{\mathcal{L}(\mathbf{x}_i, \hat{\mathbf{\Theta}}^{(k)})} \right) > threshold \tag{5}$$

where $\mathcal{L}(\mathbf{x}_i, \hat{\mathbf{\Theta}}^{(k)})$ is the likelihood for model order k,

$$\mathcal{L}(\mathbf{x}_i, \hat{\mathbf{\Theta}}^{(k)}) = \max_{\mathbf{S^{(k)}}, \mathbf{\Theta^{(k)}}} \exp\big(-\sum_{j=1}^{n} \|\mathbf{x}_j - \sum_{r=1}^{k} s_{r,j}\mathbf{a}(\theta_r)\|^2/\sigma^2\big)$$

Our estimator for the number of sources, denoted $\hat{k}_{EVT}$, is the maximal k such that $G_k > threshold$. EVT stands for extreme value theory, as will become apparent below.

For example, for $k=0$ the GLRT is equivalent to

$$\max_{\theta_1} T_{n,p}(\hat{\theta}_1) = \frac{1}{n\sigma^2} \sum_{j=1}^{n} \big|\langle \mathbf{x}_j, \frac{\mathbf{a}(\hat{\theta}_1)}{\|\mathbf{a}(\hat{\theta}_1)\|}\rangle\big|^2 > threshold, \tag{6}$$

Note that Eq. (6) holds for any array geometry. The key question is thus how to set the threshold in Eqs. (6) or (5). Whereas in principle the threshold may depend on the model order k, our analysis shows that a single threshold provides a nearly constant false alarm probability for all $k \ll n$.

To this end, we thus first analyze the case $k=0$, namely the behavior of $\hat{\theta}_1$ and of $T_{n,p}(\hat{\theta}_1)$ under the two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. The key observation is that under the null hypothesis of no signals ($\mathcal{H}_0$), 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, $T_{n,p}(\theta)$ is a χ^2_n/n random field and the GLRT distribution depends on the maxima of this field.

Theorem 1. *Consider n observations from the model (1) with no sources ($K=0$). For any false alarm rate $\alpha \ll 1$ let*

$$th(\alpha) = 1 + \tfrac{C(p,\alpha)}{\sqrt{n}} \tag{7}$$

where

$$C(p,\alpha) = \sqrt{2\ln\tfrac{1}{\alpha} + \ln\left(\tfrac{p^2-1}{6}\right)}. \tag{8}$$

Then, as $n \to \infty$

$$\Pr\left[\max_{-\frac{\pi}{2}\leqslant\theta\leqslant\frac{\pi}{2}} T_{n,p}(\theta) > th(\alpha)|\mathcal{H}_0\right] \leqslant \alpha + \Phi^c(C(p,\alpha)), \tag{9}$$

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

While Eq. (9) holds only as $n \to \infty$, it is quite accurate even for small values of n. Furthermore, the term $\Phi^c(C(p,\alpha))$ in Eq. (9) is negligible for practical values of p and α.

B. Single source detection and the MLE breakdown location

Theorem 1 implies that if

$$\max_\theta T_{n,p}(\theta) > th(\alpha) \tag{10}$$

with $\alpha \ll 1$, then the probability that the data consists only of pure noise is small. We thus consider Eq (10) as a diagnostic test for signal detection, and now study the behavior of $\hat{\theta}_1$ and of $T_{n,p}(\hat{\theta}_1)$ under $\mathcal{H}_1$. Here the DOA θ is an actual physical parameter, and its MLE is asymptotically consistent [13]. However, it is well known that for small sample sizes or weak signals, the MLE suffers from a performance "breakdown phenomena", whereby below a certain threshold of either SNR or number of samples, the MLE error starts to increase rapidly, and significantly deviate from the Cramer-Rao lower bound.

There are thus two key questions associated with the signal detection test (10). The first is which signal strengths can be detected (with high probability) by this test. The second question is whether $\hat{\theta}_1$ is a reliable estimate of θ, assuming that a signal is present and it is detected by Eq. (10). For simplicity in our analysis we call $\hat{\theta}_1$ a reliable estimate of θ if it is located inside the main lobe of the beamformer spectrum corresponding to the true DOA θ. 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* $s(t)$ $(K = 1)$*. Assume that the signal amplitudes* $s(t_j)$ *are i.i.d. realizations from a zero mean random variable with finite fourth moment. Consider signal detection using Eq. (10) with* $th(\alpha)$ *given by (7) and* $\alpha \ll 1$*.*

Assume $n \gg 1$*. If the signal strength satisfies*

$$\sigma_s^2 = \mathbb{E}[|s|^2] > \frac{C(p,\alpha) + \Phi^{-1}(1-\epsilon)}{p\sqrt{n}} \tag{11}$$

with $p \gg 1$*, then the signal* $s(t)$ *will be detected with probability at least* $1-\epsilon$*, and* $\hat{\theta}_1$ *will be a reliable estimate of* θ *with probability at least*

$$1 - \delta = 1 - 2\Phi^c\left(\frac{1}{2}(1 - \frac{1}{\pi^2})(C(p,\alpha) + S_{1-\epsilon})\right). \tag{12}$$

The right hand side of Eq. (11) is thus an approximate expression for the location of the MLE breakdown. In particular, for signal strength below this threshold, the probability that $\hat{\theta}_1$ 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 (10). The accuracy of the estimated location of the MLE breakdown is shown in figure 2, see also section V.

We note that although various papers discuss the MLE breakdown phenomena (see [1], [3], [12], [10] for example), to the best of our knowledge, our analysis is the first one to provide a simple explicit expression for its location, Eq. (11). In [3] for example, an approximation for the MSE is derived by analyzing only the single signal case, and the probability of an outlier. This analysis results in a complicated expression which involves an infinite integral of the Bessel function. Finally, we note that a similar analysis can be done for the case of multiple signals. Of course, the SNR needed to detect two closely spaced sources depends on their spatial separation and may be much higher than Eq. (11).

C. The distribution of the GLRT with multiple sources

The above analysis can be extended to higher order models as follows: let P denote the projection onto $Span\{\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)\}$, and $P^\perp$ its orthogonal complement,

Proposition 3. *Consider* n *samples from the model (1) with* $K < p-1$ *sources from distinct directions* $\Theta = \{\theta_1, \ldots, \theta_K\}$*. As* $n \to \infty$ *the GLRT at the correct model order* K *is distributed as the maxima of the following random field,*

$$G_K \sim \max_\theta \tilde{T}_{n,p}(\theta)$$

where $\tilde{T}_{n,p}(\theta) = \frac{1}{n}\sum_{j=1}^n \left|\langle \eta_j, \frac{P^\perp \mathbf{a}(\theta)}{\|P^\perp \mathbf{a}(\theta)\|}\rangle\right|^2$*. Next, let* $\tilde{r}(\theta_1,\theta_2)$ *be the covariance function of* $\sqrt{n}(\tilde{T}_{n,p}(\theta) - 1)$*. Then*

$$\tilde{r}_{1,1}(\theta,\theta,P^\perp) = \frac{\partial^2 \tilde{r}}{\partial\theta_1 \partial\theta_2}(\theta,\theta) = \tag{13}$$
$$2\frac{\|P^\perp \frac{d}{d\theta}\mathbf{a}(\theta)\|^2 \|P^\perp \mathbf{a}(\theta)\|^2 - |\langle \frac{d}{d\theta}\mathbf{a}(\theta), P^\perp \mathbf{a}(\theta)\rangle|^2}{\|P^\perp \mathbf{a}(\theta)\|^4}.$$

Finally, if

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\tilde{r}_{1,1}(\theta,\theta,P^\perp)}d\theta \leqslant \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\tilde{r}_{1,1}(\theta,\theta,I)}d\theta, \tag{14}$$

then, as $n \to \infty$*,*

$$\Pr\left[G_K > th(\alpha) | \mathcal{H}_K\right] \leqslant \alpha + \Phi^c(C(p,\alpha)). \tag{15}$$

Proposition 3 shows that if Eq. (14) holds, then the threshold $th(\alpha)$ in Eq. (7), derived for detection of a single source, can also be used in the multiple sources scenario to prevent overestimation. While we have not been able to mathematically prove that Eq. (14) indeed holds, it is easy to verify it numerically for any specific number of sources, DOA's etc. Next we show the weak asymptotic consistency of this estimator.

Theorem 4. *Consider* n *samples from the model (1) with* $K < p-1$ *sources. Then under the assumptions of Proposition 3,*

$$\Pr[\hat{k}_{EVT} < K] \;\to\; 0 \;\textit{as}\; n \to \infty, \tag{16}$$
$$\Pr[\hat{k}_{EVT} > K] \;\leqslant\; \alpha + \Phi^c(C(p,\alpha)) \;\textit{as}\; n \to \infty. \tag{17}$$

V. SIMULATIONS

We illustrate our analysis and source detection algorithm by a series of simulations, all with a ULA of $p = 7$ or $p = 15$ sensors, $\sigma = 1$ and Gaussian distributed signals (even though we consider them as deterministic unknown).

Fig. 1 shows the probability that the test (10) fails e.g. that a signal is not detected despite its presence. The vertical lines are the detection thresholds from Eq. (11) with $\epsilon = 0.01$. For

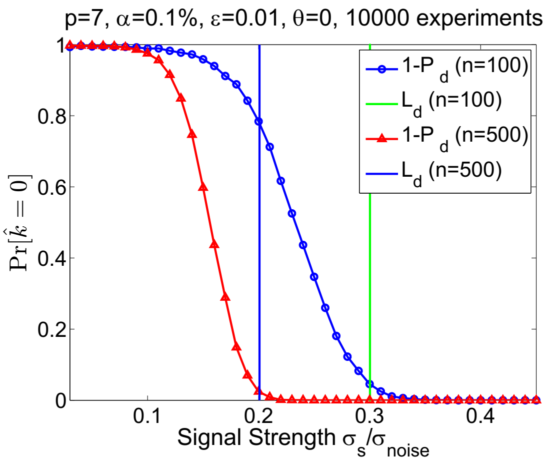

Fig. 1. The probability that the test (10) failed, when a single source is present, as a function of the signal strength. P_d is the probability of detection, and L_d is the right hand side of Eq. (11).

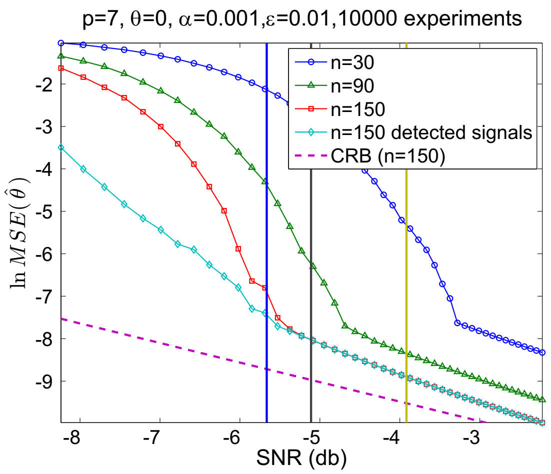

Fig. 2. MSE of $\hat{\theta}_1$ as a function of SNR, for different values of n. L_d is the right hand side of Eq. (11). For $n = 150$ we show the MSE given that Eq. (10) holds, as well as the Cramer-Rao bound.

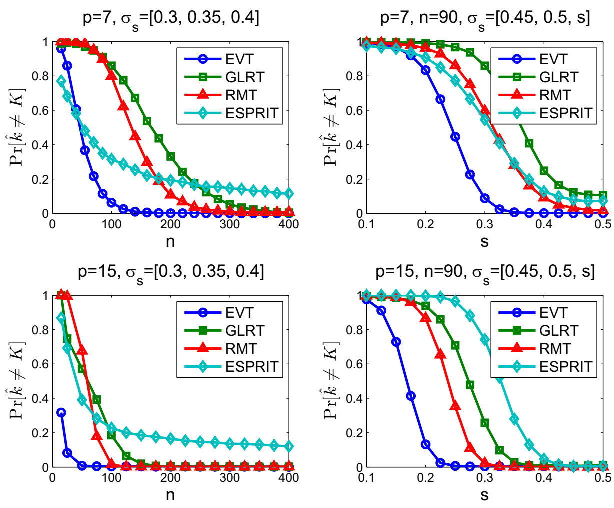

Fig. 3. The probability to detect an incorrect number of signals, when three signals are present from directions $\Theta = \{-\frac{\pi}{4}, 0, \frac{\pi}{4}\}$, as a function of number of samples (left) or signal strength (right) for $p = 7$ or $p = 15$ sensors.

signal strengths *above* this threshold the mis-detection rate is close to $\epsilon = 0.01$, and drops below ϵ as n increases.

Fig. 2 shows the mean square error $\mathbb{E}(\hat{\theta}_1 - \theta)^2$, when a single signal is present, as a function of signal strength σ_s^2. The vertical lines are the limit of detection (Eq. (11)). In accordance to theorem 2 these provide an approximation for the location of the MLE performance breakdown, which becomes increasingly more accurate for larger n. In addition, we show the MSE computed only from those experiments where a signal was detected by Eq. (10). Note that here there is no breakdown phenomena, in accordance to theorem 2.

Finally, in fig. 3 we compare the detection performance of our algorithm to the non-parametric method of [7], based on random matrix theory (RMT), the semi-parametric ESPRIT method of [11], and the parametric $GLRT$ method of [9]. As shown in the figure our method has better detection performance at a comparable false alarm rate.

Acknowledgment. The authors thank Yonina Eldar, Anthony Weiss and Hagit Messer-Yaron 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 Arrays Part I: Fully Augmentable Arrays, *IEEE Trans. Sig. Proc.*, vol. 49, no. 5, May 2001, pp. 959–971.

[3] F. Athley, Threshold region performance of maximum likelihood direction of arrival estimators, *IEEE Tran. Sig. Proc.*, vol. 53, pp. 1359–1373, 2005.

[4] 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.

[5] J.A. Jiang, M.A. Ingram - Robust Detection of Number of Sources Using the Transformed Rotational Matrix, *IEEE Wir. Comm. Net. Conf.*, 2004.

[6] H. Krim ,M. Viberg, Two Decades of Signal Array Processing Research, *IEEE Sig. Proc. Mag.*, vol. 13, pp. 67–94, 1996.

[7] 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.

[8] B. Nadler, Nonparametric detection of signals by information theoretic criteria, vol. 58, pp. 2746–2756, *IEEE Trans. Sig. Proc.*, 2010.

[9] B. Ottersten, M. Viberg, P. Stoica, A. Nehorai, Exact and large sample ML techniques for parameter estimation and detection in array processing, Ch. 4 in "Radar Array Processing" Simon Haykin, Springer-Verlag, 1993.

[10] I. Reuven, H. Messer, The use of the Barankin bound for determining the threshold SNR in estimating the bearing of a source in the presence of another, *Proc. IEEE Int. Conf. Acoustics, Speech, Sig. Proc.*, vol. 3, pp. 1645-1648, 1995.

[11] 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.

[12] P. Stoica, M. Hawkes, A. Nehorai, Performance Breakdown of Subspace Based Methods: Prediction and Cure, *ICASSP*, pp. 4005–4008, 2001.

[13] 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.

[14] 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, 1985.

[15] M. Wax, I. Ziskind, Detection of the Number of Coherent Signals by the MDL Principle, *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 37, no. 8, pp. 1190-1196, 1989.