Detection of Signals by Information Theoretic Criteria: General Asymptotic Performance Analysis

Eran Fishler, *Member, IEEE*, Michael Grosmann, and Hagit Messer, *Fellow, IEEE*

***Abstract*—Detecting the number of sources is a well-known and a well-investigated problem. In this problem, the number of sources impinging on an array of sensors is to be estimated. The common approach for solving this problem is to use an information theoretic criterion like the minimum description length (MDL), or the Akaike information criterion (AIC). Although it has been gaining much popularity and has been used in a variety of problems, the performance of information theoretic criteria-based estimators for the unknown number of sources has not been sufficiently studied, yet. In the context of array processing, the performance of such estimators were analyzed only for the special case of Gaussian sources where no prior knowledge of the array structure, if given, is used. Based on the theory of misspecified models, this paper presents a general asymptotic analysis of the performance of any information theoretic criterion-based estimator, and especially of the MDL estimator. In particular, the performance of the MDL estimator, which assumes Gaussian sources and structured array when applied to Gaussian sources, is analyzed. In addition, it is shown that the performance of a certain MDL estimator is not very sensitive to the actual distribution of the source signals. However, appropriate use of prior knowledge about the array geometry can lead to significant improvement in the performance of the MDL estimator. Simulation results show good fit between the empirical and the theoretical results.**

***Index Terms*—Array processing, asymptotic analysis, MDL.**

I. INTRODUCTION

MOST parametric bearing estimation techniques assume that the number of sources impinging on the array is known or pre-estimated using some technique. The problem of estimating the number of sources impinging on a passive array of sensors has received a considerable amount of attention during the last two decades (see, among many others, [1]–[6]). The most common approach for estimating this number is to apply information theoretic criteria, like the minimum description length (MDL) or Akaike information criterion (AIC) [7]. Since 1985 [3], when first suggested for estimating the number of narrowband sources impinging on an array of sensors, the MDL estimator practically became the standard tool for accomplishing this task.

Manuscript received March 20, 2001; revised January 2, 2002. The associate editor coordinating the review of this paper and approving it for publication was Dr. Alex B. Gershman.

E. Fishler was with the Department of Electrical Engineering—Systems, Tel Aviv University, Tel Aviv, Israel. He is now with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: efishler@ee.princeton.edu).

M. Grossman was with the Department of Electrical Engineering—Systems, Tel Aviv University, Tel Aviv, Israel. He is now with the Advanced Technology Group, AudioCodes Ltd., Tel Aviv, Israel.

H. Messer is with the Department of Electrical Engineering—Systems, Tel Aviv University, Tel Aviv, Israel (e-mail: messer@eng.tau.ac.il).

Publisher Item Identifier S 1053-587X(02)03222-1.

A. Problem Formulation

Assume an array of p sensors, and denote by $\mathbf{x}(t)$ the received, p-dimensional, signal vector at time instance t. In addition, denote by q the number of signals impinging on the array. A common model for the received signal vector is

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{A} = [\mathbf{a}(\boldsymbol{\theta}_1), \mathbf{a}(\boldsymbol{\theta}_2), \ldots, \mathbf{a}(\boldsymbol{\theta}_q)]$ is a $p \times q$ matrix composed of q p-dimensional vectors, where $\mathbf{a}(\boldsymbol{\theta})$ lies on the array manifold $\{\mathcal{A} = \mathbf{a}(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. $\mathbf{a}(\boldsymbol{\theta})$ is called the array response vector or the steering vector, and $\mathbf{A}$ is referred to as the steering matrix. $\mathbf{s}(t) = [s_1(t) \ \cdots \ s_q(t)]^T$ is a $q \times 1$ signals vector, and $\mathbf{n}(t)$ is a $p \times 1$ vector of the additive noise. $\boldsymbol{\theta}_i$ is the unknown parameter vector associated with the ith source. Finally, assume also that $\mathbf{n}(t)$ and $\mathbf{s}(t)$ are ergodic, independent random vector processes.

Many problems may be formulated using this simple, linear model (see [8], [9], and references therein). These problems differ by the structure of the mixing matrix $\mathbf{A}$ by the assumed knowledge about the unknown parameters or by the statistical modeling. For example, in bearing estimation [9], $\boldsymbol{\theta}_i$ is a scalar (namely the bearing of the ith source). In other problems, $\boldsymbol{\theta}_i$ are all the elements of the ith column of $\mathbf{A}$, which represent a complete lack of knowledge about the steering matrix [3]. Denote by $\boldsymbol{\theta}^{(q)} = [\boldsymbol{\theta}_0^T, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \ldots, \boldsymbol{\theta}_q^T]$ the complete unknown parameter vector, given that q sources exist. $\boldsymbol{\theta}_0$ represents the vector of unknown parameters that do not belong to any specific source, for example the noise level. Denote by $\boldsymbol{\Theta}_q$ the parameter space of $\boldsymbol{\theta}^{(q)}$. The problem is to estimate the number of sources q, given N independent snapshots of the array output $\mathbf{x}(t_1), \ldots, \mathbf{x}(t_N)$.

B. MDL Approach

The information theoretic criteria approach is a general approach for choosing a model that fits the data mostly from a family of possible models [7], [10]. That is, given a parameterized family of probability densities, $f_{\mathbf{X}}(\mathbf{X}|\boldsymbol{\theta}^{(q)})$, $\boldsymbol{\theta}^{(q)} \in \boldsymbol{\Theta}_q$ for various q, select $\hat{q}$ such that

$$\hat{q} = \arg\min_q -L\left(\hat{\boldsymbol{\theta}}^{(q)}\right) + p(q) \tag{2}$$

where $L(\boldsymbol{\theta}^{(q)}) \triangleq \log f_{\mathbf{X}}(\mathbf{X}|\boldsymbol{\theta}^{(q)})$ is the log-likelihood of the measurements denoted by $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, $p(q)$ is a general penalty function associated with the qth family, and $\hat{\boldsymbol{\theta}}^{(q)} = \arg\max_{\boldsymbol{\theta}^{(q)} \in \boldsymbol{\Theta}_q} f_{\mathbf{X}}(\mathbf{X}|\boldsymbol{\theta}^{(q)})$. $\hat{\boldsymbol{\theta}}^{(q)}$ is usually referred to as the maximum likelihood (ML) estimate of the unknown parameters given the qth family of distributions.

The MDL estimator is a special case of (2) with a certain penalty function. It is given by minimizing the MDL metric [7], that is

$$\hat{q}_{\text{MDL}} = \arg\min_q \text{MDL}(q) = \arg\min_q \left\{ -\log\left(f_{\mathbf{X}}\left(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)}\right)\right) + 0.5|\boldsymbol{\Theta}_q|\log(N)\right\} \tag{3}$$

where $\text{MDL}(q) \triangleq \{-\log(f(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})) + 0.5|\boldsymbol{\Theta}_q|\log(N)\}$, and $|\boldsymbol{\Theta}_q|$ is the number of free parameters in $\boldsymbol{\Theta}_q$. Since many problems are overparameterized, $|\boldsymbol{\Theta}_q|$ represents the minimum number of parameters that characterize $f(\mathbf{X}|\boldsymbol{\theta}^{(q)})$ completely. It is well known that asymptotically, under certain regularity conditions, the MDL estimator minimizes the description length (measured in bits) of both the measurements $\mathbf{X}$ and the model $\hat{\boldsymbol{\theta}}^{(q)}$ [7]. Note that the MDL estimator is a function of the second-order statistics, only when $f_{\mathbf{X}}(\mathbf{X}|\boldsymbol{\theta}^{(q)})$ is the Gaussian distribution.

Assume the following.

1) The sources $s_1(t), \ldots, s_q(t)$ are white, complex, stationary, and ergodic Gaussian random process, with zero mean and positive definite covariance matrix.
2) The noise vector $\mathbf{n}(t)$ is a white, complex, stationary, and ergodic Gaussian vector process, independent of the signals, with zero mean and covariance matrix given by $\sigma^2\mathbf{I}$.
3) All the elements of the steering matrix $\mathbf{A}$ are assumed unknown with the only restriction that $\mathbf{A}$ is of full rank.

When applying these assumptions to (3), the MDL estimator, which is referred to as the GMDL, becomes [3]

$$\hat{q}_{\text{GMDL}} = \arg\min_{q=0,\ldots,p-1} \text{GMDL}(q) = \arg\min_{q=0,\ldots,p-1} \left\{ -N\log\left(\frac{\prod_{i=q+1}^{p} l_i}{\left(\frac{1}{p-q}\sum_{i=q+1}^{p} l_i\right)^{p-q}}\right) + 0.5(q(2p-q)+1)\log N \right\} \tag{4}$$

where

$$\text{GMDL}(q) \triangleq \left\{ -N\log\left(\frac{\prod_{i=q+1}^{p} l_i}{\left(\frac{1}{p-q}\sum_{i=q+1}^{p} l_i\right)^{p-q}}\right) + 0.5(q(2p-q)+1)\log N \right\}$$

and $l_1 > l_2 > \cdots > l_p$ are the eigenvalues of the empirical correlation matrix $\hat{\mathbf{R}} = (1/N)\sum_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i^H$, and $\cdot^H$ denotes the conjugate transpose operation.

The GMDL estimator is analyzed in [10], and it is shown to be a consistent estimator of the number of sources, i.e., the probability of error decreases to zero as the number of measurements increases to infinity. Moreover, under mild regularity conditions, like finite second moment, it is a consistent estimator of the number of sources, even if the measurements are non–Gaussian. The GMDL is further analyzed in [11] and [12]. This analysis is carried for the case of both Gaussian sources and Gaussian noise, and it uses the asymptotic distribution of the eigenvalues of the empirical correlation matrix.

Minimizing the description length is an information theoretic criterion, whereas in detection problems, minimizing the probability of error is the desired optimization criterion. Define the probability of misdetection as the probability that the estimate is less than the true number of sources, that is, $P_{MD} = P(\hat{q} < q)$, where $\hat{q}$ denotes the estimate for the number of sources. The probability of false alarm is defined as the probability that the estimate is greater than the true number of sources, that is, $P_{FA} = P(\hat{q} > q)$. The probability of error is the sum of the two $P_e = P_{MD} + P_{FA}$. The aim of any estimator for the number of sources is to minimize the probability of error.

C. Summary of the Paper's Main Results

This paper presents a novel asymptotic analysis of any information theoretic criterion for estimating the number of sources. The analysis includes simple closed-form expressions for the probability of misdetection $P(\hat{q} < q)$ for any information theoretic criteria-based estimator (2). This analysis can be used even in cases of mismatches, where, for example, one uses the GMDL estimator (4), but the sources are binary phase shifting key (BPSK) signals. Moreover, a proof wherein, for every MDL-based estimator (3), the probability of misdetection approaches zero as the number of snapshots approaches infinity $P(\hat{q}_{\text{MDL}} < q) \xrightarrow{N\to\infty} 0$, is given. For the MDL estimator (3), as seen in many simulations studies [11]–[13], misdetection is the main source of error, and therefore, the resulting analysis is practically an approximated complete performance (error) analysis of the MDL estimator.

In addition to the general analysis of the MDL estimator (3), special attention is given to three specific cases:

1) probability of misdetection of the MDL estimator of (3), when the measurements are Gaussian and some prior information about the array is utilized;
2) probability of misdetection of the GMDL estimator, when the measurements are Gaussian (note that in this case, no information about the array is utilized);
3) probability of misdetection of the GMDL estimator when the noise is normally distributed while the sources are digital signals, like BPSK.

The general analysis of the MDL estimator is used to quantify the improvement of using the MDL estimator instead of the GMDL estimator. Specifically, cases 1) and 2) are compared, and the improvement in performance due to the use of prior information about the array geometry in the MDL estimator (3) instead of using the GMDL estimator is quantified.

The paper is organized as follows. Section II analyzes the performance of the MDL estimator under a very general set of conditions. Section III applies the analysis for two important cases: Gaussian sources and digital modulated sources. Section IV provides simulations that demonstrate the applicability of the theoretical analysis for a wide variety scenarios. Section V provides a discussion and some concluding remarks.

II. Performance of the MDL Estimator

The performance of the GMDL estimator (4) is analyzed in [11] and [12] for Gaussian measurements. The main result of this analysis is that the random variable $\mathrm{GMDL}(q-1)-\mathrm{GMDL}(q)$ is normally distributed. In addition, a simple closed-form expression for the mean and the variance of the random variable $\mathrm{GMDL}(q-1)-\mathrm{GMDL}(q)$ are given. By approximating the probability of misdetection by the probability of the event $\mathrm{GMDL}(q-1)-\mathrm{GMDL}(q)<0$, i.e., $P_{MD}\approx P(\mathrm{GMDL}(q-1)-\mathrm{GMDL}(q)<0)$, this analysis provides a complete analysis of the probability of misdetection of the GMDL estimator (4), assuming Gaussian measurements.

In this section, the results of [11] and [12] are shown to be applicable to the general MDL estimator (3). It is proven that asymptotically the random variable $\mathrm{MDL}(q-1)-\mathrm{MDL}(q)$ is normally distributed. In addition, simple closed-form expressions for the mean and variance of this random variable are given. These results hold even when the assumed probability density function (pdf) of the measurements differs from $f_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta}^{(k)})$ for every $\boldsymbol{\theta}^{(k)}\in\boldsymbol{\Theta}_k$ and $k=0,1,\ldots,$. As in [11] and [12], by using the approximation $P_{MD}\approx P(\mathrm{MDL}(q-1)-\mathrm{MDL}(q)<0)$, this analysis provides complete analysis of the probability of misdetection of every MDL estimator (3). For the special case where the measurements are normally distributed and $\mathbf{A}$ is assumed unknown, the MDL estimator (3) becomes the GMDL estimator (4), and our analysis results unite with those of [11] and [12].

In order to state and prove the main theorem, some assumptions and definitions are required. Denote by $g(\mathbf{x})$ the actual pdf of the received vector $\mathbf{x}$. Recall that $f(\mathbf{x}|\boldsymbol{\theta}^{(q-1)})$ and $f(\mathbf{x}|\boldsymbol{\theta}^{(q)})$ are the assumed pdfs of the received data given $q-1$ and q sources, respectively. Note that introducing $g(\mathbf{x})$ allows us to consider the case of statistical modeling errors, e.g., the case where the sources are assumed Gaussian, whereas, in fact, they are BPSK signals. In this case, as will be seen later, $g(\mathbf{x})$ is a mixture of complex Gaussians, whereas $f(\mathbf{x}|\boldsymbol{\theta}^{(k)})$ is Gaussian. Denote by $D(f\|g)$ the Kullback–Leibler divergence between two pdfs f and g:

$$D(f\|g)\triangleq\int_{-\infty}^{\infty}\log\left(\frac{g(x)}{f(x)}\right)g(x)\,dx. \tag{5}$$

Assume that $D(f(\mathbf{x}|\boldsymbol{\theta}^{(q)})\|g(\mathbf{x}))$ and $D(f(\mathbf{x}|\boldsymbol{\theta}^{(q-1)})\|g(\mathbf{x}))$ exist for every $\boldsymbol{\theta}^{(q)}\in\boldsymbol{\Theta}_q$ and $\boldsymbol{\theta}^{(q-1)}\in\boldsymbol{\Theta}_{q-1}$, respectively. In addition, assume that there exist two unique, interior points (one in $\boldsymbol{\Theta}_q$ and the other in $\boldsymbol{\Theta}_{q-1}$ denoted by $\boldsymbol{\theta}_q^*$ and $\boldsymbol{\theta}_{q-1}^*$, which minimize $D(f(\mathbf{x}|\boldsymbol{\theta}^{(q)})\|g(\mathbf{x}))$ and $D(f(\mathbf{x}|\boldsymbol{\theta}^{(q-1)})\|g(\mathbf{x}))$, respectively. That is

$$\boldsymbol{\theta}_q^*\triangleq\arg\min_{\boldsymbol{\theta}^{(q)}\in\boldsymbol{\Theta}_q}D\left(f\left(\mathbf{x}|\boldsymbol{\theta}^{(q)}\right)\|g(\mathbf{x})\right) \tag{6}$$

$$\boldsymbol{\theta}_{q-1}^*\triangleq\arg\min_{\boldsymbol{\theta}^{(q-1)}\in\boldsymbol{\Theta}_{q-1}}D\left(f\left(\mathbf{x}|\boldsymbol{\theta}^{(q-1)}\right)\|g(\mathbf{x})\right). \tag{7}$$

Denote by $J(\boldsymbol{\theta})$ and by $H(\boldsymbol{\theta})$ the following matrices:

$$J(\boldsymbol{\theta})=\mathrm{E}_g\left\{\frac{\partial\log f(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\,\frac{\partial\log f(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}^T\right\}\quad\forall\boldsymbol{\theta}\in\boldsymbol{\Theta}_{q-1}\cup\boldsymbol{\Theta}_q \tag{8}$$

$$H(\boldsymbol{\theta})=\mathrm{E}_g\left\{\frac{\partial^2\log f(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}(\partial\boldsymbol{\theta})^T}\right\};\qquad\forall\boldsymbol{\theta}\in\boldsymbol{\Theta}_{q-1}\cup\boldsymbol{\Theta}_q \tag{9}$$

where $\mathrm{E}_g\{f(\mathbf{x})\}\triangleq\int f(\mathbf{x})g(\mathbf{x})\,d\mathbf{x}$ is the expectation of $f(\mathbf{x})$ with respect to $g(\mathbf{x})$. Assume that (8) and (9) exist for every $\boldsymbol{\theta}^{(q)}\in\boldsymbol{\Theta}_q$ and $\boldsymbol{\theta}^{(q-1)}\in\boldsymbol{\Theta}_{q-1}$. Note that when $g(\mathbf{x})=f(\mathbf{x}|\boldsymbol{\theta}_0^{(q)})$, where $\boldsymbol{\theta}_0^{(q)}$ is some interior point in $\boldsymbol{\Theta}_q$, i.e., the true distribution of the data is contained in the assumed family of distributions, no mismatch exists. Finally, assume that $f(\mathbf{x}|\boldsymbol{\theta}^{(q)})$ and $f(\mathbf{x}|\boldsymbol{\theta}^{(q-1)})$ are bounded, continuous families of distribution functions and twice differentiable in $\boldsymbol{\theta}^{(q)}$ and in $\boldsymbol{\theta}^{(q-1)}$, respectively. The regularity conditions stated ensure that $\hat{\boldsymbol{\theta}}^{(q)}$, $\hat{\boldsymbol{\theta}}^{(q-1)}$ are asymptotically normally distributed, with mean equal $\boldsymbol{\theta}_q^*$ and $\boldsymbol{\theta}_{q-1}^*$, respectively [14].

Theorem 1: Under the conditions stated, $\mathrm{MDL}(q-1)-\mathrm{MDL}(q)$ is asymptotically normally distributed with mean

$$\mu=N\cdot\mathrm{E}_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}|\boldsymbol{\theta}_q^*)\}+0.5\log N(|\boldsymbol{\Theta}_{q-1}|-|\boldsymbol{\Theta}_q|) \tag{10}$$

and variance

$$\begin{aligned}\sigma^2&=N\cdot Var_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}|\boldsymbol{\theta}_q^*)\}\\&=N\mathrm{E}_g\{(-\log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}|\boldsymbol{\theta}_q^*)\\&\qquad-\mathrm{E}_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}|\boldsymbol{\theta}_q^*)\})^2\}.\end{aligned} \tag{11}$$

Proof of Theorem 1: See Appendix A.

Theorem 1 states that under very general conditions, the asymptotic pdf of the random variable $\mathrm{MDL}(q-1)-\mathrm{MDL}(q)$ is Gaussian. Moreover, a simple way to compute the asymptotic mean and variance of $\mathrm{MDL}(q-1)-\mathrm{MDL}(q)$ are given. Note that the conditions under which the theorem holds include the case where the possible parametric models $f(\mathbf{x}|\boldsymbol{\theta}^{(q-1)})$, $f(\mathbf{x}|\boldsymbol{\theta}^{(q)})$ do not include $g(\mathbf{x})$. Thus, the theorem enables one to compute the asymptotic performance of the MDL estimator under mismodeling.

The asymptotic probability of misdetection is approximated by

$$P_{MD}\approx P(\mathrm{MDL}(q-1)-\mathrm{MDL}(q)<0)=1-Q\left(\frac{-\mu}{\sigma}\right) \tag{12}$$

where μ (10) and σ (11) are given in the theorem, and $Q(x)=\int_x^{\infty}(e^{-(1/2)\alpha^2}/\sqrt{2\pi})\,d\alpha$.

Proposition 1: Under the conditions of Theorem 1, the probability of misdetection of *any* MDL estimator approaches zero as the number of snapshots goes to infinity.

Proof of Proposition 1: By using (12), proving Proposition 1 is equivalent to proving that μ/σ approaches infinity as the number of snapshots goes to infinity. It is easily seen that

$$\begin{aligned}\frac{\mu}{\sigma} =& \sqrt{N}\, \frac{\mathrm{E}_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1}) + \log f(\mathbf{x}|\boldsymbol{\theta}^*_q)\}}{Var_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1}) + \log f(\mathbf{x}|\boldsymbol{\theta}^*_q)\}} \\ &+ 0.5\, \frac{|\boldsymbol{\Theta}_{q-1}| - |\boldsymbol{\Theta}_q|}{Var_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1}) + \log f(\mathbf{x}|\boldsymbol{\theta}^*_q)\}}\, \frac{\log N}{N}.\end{aligned} \tag{13}$$

Since the hypothesis that $q-1$ sources exist is nested in the hypothesis that q sources exist, $\mathrm{E}_g\{\log f(\mathbf{x}|\boldsymbol{\theta}^*_q)\} \geq \mathrm{E}_g\{\log f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1})\}$, and thus

$$\sqrt{N}\, \frac{\mathrm{E}_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1}) + \log f(\mathbf{x}|\boldsymbol{\theta}^*_q)\}}{Var_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1}) + \log f(\mathbf{x}|\boldsymbol{\theta}^*_q)\}} = O\left(\sqrt{N}\right).$$

The second term in (13)

$$0.5\, \frac{|\boldsymbol{\Theta}_{q-1}| - |\boldsymbol{\Theta}_q|}{Var_g\{-\log f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1}) + \log f(\mathbf{x}|\boldsymbol{\theta}^*_q)\}}\, \frac{\log N}{N}$$

approaches zero as N approaches infinity. Thus, $\mu/\sigma = O(\sqrt{N})$ and therefore, $P_{MD} \approx 1 - Q(-\mu/\sigma) \rightarrow 1 - Q(-\infty) = 0$.

III. Applications

This section demonstrates the use of Theorem 1 for obtaining new approximations for the probability of misdetection of the MDL estimator (3) in a wide variety of scenarios. The general technique is demonstrated for two important scenarios. The first is the case of Gaussian sources with structured arrays. The second is the case of digital sources where the GMDL estimator is used.

A. MDL Estimator (3) for Gaussian Sources and Structured Arrays

In this subsection, we analyze the performance of the MDL estimator (3) under the assumption of both Gaussian sources and Gaussian noise. Recall that the model for the received signal is given by

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t). \tag{14}$$

Assume that both the sources and the additive noise are white, independent, and zero mean complex Gaussian random processes. That is $\mathbf{s}(t) \sim \mathcal{CN}_q(0, \mathbf{R_s})$ and $\mathbf{n}(t) \sim \mathcal{CN}_p(0, \sigma_n^2\mathbf{I})$, where $\mathbf{R_s} \triangleq E\{\mathbf{s}(t)\mathbf{s}^H(t)\}$ is the correlation matrix of the q sources. It is easy to verify that $\mathbf{x}(t)$ is a white, zero mean, complex Gaussian random vector, and the log likelihood of $\mathbf{x}(t)$ given $\boldsymbol{\theta}^{(q)}$, is $L(\boldsymbol{\theta}^{(q)}) = \log f(\mathbf{x}(t)|\boldsymbol{\theta}^{(q)}) = C - \log|\mathbf{R_x}(\boldsymbol{\theta}^{(q)})| - \mathbf{x}^H(t)[\mathbf{R_x}(\boldsymbol{\theta}^{(q)})]^{-1}\mathbf{x}(t)$, where again, $\boldsymbol{\theta}^{(q)}$ is the unknown parameter vector assuming q sources, and C is a constant independent of $\boldsymbol{\theta}^{(q)}$. Note that the MDL estimator (3) is used to estimate the number of sources and not the GMDL estimator (4), and $\boldsymbol{\theta}^{(q)}$ represents the amount of *a priori* knowledge.

Denote by $\boldsymbol{\theta}_0^{(q)}$ the true value of the unknown parameter vector, for example, the unknown bearings of the different sources. Note that since no mismodeling exists, $g(\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta}_0^{(q)})$ and $\boldsymbol{\theta}^*_q = \boldsymbol{\theta}_0^{(q)}$. Recall that $\boldsymbol{\theta}^*_{q-1} \in \boldsymbol{\Theta}_{q-1}$ is the parameter vector that minimizes the Kullback–Leibler divergence between $g(\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta}_0^{(q)})$ and $f(\mathbf{x}|\boldsymbol{\theta}^{(q-1)})$, $\boldsymbol{\theta}^{(q-1)} \in \boldsymbol{\Theta}_{q-1}$, that is

$$\begin{aligned}\boldsymbol{\theta}^*_{q-1} &= \arg \min_{\boldsymbol{\theta}^{(q-1)} \in \boldsymbol{\Theta}_{q-1}} D\left(f\left(\mathbf{x}|\boldsymbol{\theta}^{(q-1)}\right) \| f\left(\mathbf{x}|\boldsymbol{\theta}_0^{(q)}\right)\right) \\ &= \arg \min_{\boldsymbol{\theta}^{(q-1)} \in \boldsymbol{\Theta}_{q-1}} \log\left|\mathbf{R_x}\left(\boldsymbol{\theta}^{(q-1)}\right)\right| \\ &\quad \cdot \mathrm{Tr}\left\{\mathbf{R_x}\left(\boldsymbol{\theta}_0^{(q)}\right)\mathbf{R_x}^{-1}\left(\boldsymbol{\theta}^{(q-1)}\right)\right\}\end{aligned} \tag{15}$$

where $\mathbf{R_x}(\boldsymbol{\theta})$ is the correlation matrix of the received vector given $\boldsymbol{\theta}$, $\mathbf{R_x}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})\mathbf{R_S}\mathbf{A}^H(\boldsymbol{\theta}) + \sigma_n^2\mathbf{I}$. We also used the fact that for the normally distributed measurements, $D(f(\mathbf{x}|\boldsymbol{\theta}^{(q-1)})\|f(\mathbf{x}|\boldsymbol{\theta}_0^{(q)}))$ is equal to $\log|\mathbf{R_x}(\boldsymbol{\theta}^{(q-1)})| + \mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_0^{(q)})\mathbf{R_x}^{-1}(\boldsymbol{\theta}^{(q-1)})\} - \log|\mathbf{R_x}(\boldsymbol{\theta}_0^{(q)})| - p$.

Thus, by using (10) and (11), the mean μ and the variance σ^2 of the random variable $\mathrm{MDL}(q-1) - \mathrm{MDL}(q)$ for this case satisfy (the full derivation is in Appendix B)

$$\begin{aligned}\frac{1}{N}\mu =& -\log|\mathbf{R_x}(\boldsymbol{\theta}^*_q)| + \log|\mathbf{R_x}(\boldsymbol{\theta}^*_{q-1})| - p \\ &+ \mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}^*_q)[\mathbf{R_x}(\boldsymbol{\theta}^*_{q-1})]^{-1}\} \\ &- 0.5\, \frac{\log N}{N}\, (|\boldsymbol{\Theta}_q| - |\boldsymbol{\Theta}_{q-1}|)\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{1}{N}\sigma^2 =& \mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}^*_q)[\mathbf{R_x}(\boldsymbol{\theta}^*_{q-1})]^{-1}\mathbf{R_x}(\boldsymbol{\theta}^*_q)[\mathbf{R_x}(\boldsymbol{\theta}^*_{q-1})]^{-1}\} \\ &- 2\mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}^*_q)[\mathbf{R_x}(\boldsymbol{\theta}^*_{q-1})]^{-1}\} + p.\end{aligned} \tag{17}$$

1) Unstructured Arrays: The case of unstructured arrays corresponds to complete lack of knowledge about the steering matrix, that is, $\mathbf{A}$ is completely unknown but assumed to be full rank [3]. Assuming q sources and using the spectral representation theorem [3], $\mathbf{R_x} = \mathbf{A}\mathbf{R_s}\mathbf{A}^H + \sigma_n^2\mathbf{I} = \sum_{i=1}^q(\lambda_i - \sigma_n^2)\mathbf{v}_i\mathbf{v}_i^H + \sigma_n^2\mathbf{I}$, where $\lambda_1, \ldots, \lambda_q, \mathbf{v}_1, \ldots, \mathbf{v}_q$ are the q largest eigenvalues of $\mathbf{R_x}$ and their corresponding eigenvectors, and σ_n^2 is the unknown noise level. Thus, the unknown parameters, assuming q sources, are $\boldsymbol{\theta}^{(q)} = [\lambda_1, \ldots, \lambda_q, \sigma_n^2, \mathbf{v}_1, \ldots, \mathbf{v}_q]$.

The ML estimate of the unknown parameters for the case of k sources $\hat{\boldsymbol{\theta}}^{(k)}$ is given by [3]: $[l_1, \ldots, l_k, (1/(p-k))\sum_{i=k+1}^p l_i, \mathbf{c}_1, \ldots, \mathbf{c}_k]$, where $l_1 > l_2 > \cdots > l_p$ are the eigenvalues of the empirical correlation matrix $\hat{\mathbf{R}}_\mathbf{x} \triangleq \sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^H$, and $\mathbf{c}_1, \ldots, \mathbf{c}_p$ are their corresponding eigenvectors.

The MDL estimator is evaluated by plugging back the ML estimator into the likelihood function and adding the appropriate penalty function. The resulting estimator for the number of the sources is the GMDL estimator given by (4). The value of $\boldsymbol{\theta}^*_{q-1}$ can be obtained simply by computing the ML estimate for the case of $q-1$ sources, whereas $\hat{\mathbf{R}}_\mathbf{x}$ is replaced by its true value $\mathbf{R_x}(\boldsymbol{\theta}^*_q)$. In this case, we obtain

$$\boldsymbol{\theta}^*_{q-1} = \left[\lambda_1, \ldots, \lambda_{q-1}, \frac{(p-q)\sigma_n^2 + \lambda_q}{p-q+1}, \mathbf{v}_1, \ldots, \mathbf{v}_{q-1}\right]. \tag{18}$$

Plugging back this result into (16) and (17), we show in Appendix C that

$$\mu = N \log \left\{ \frac{\sigma_n^2}{\lambda_q} \left[1 + \frac{1}{p-q+1} \left(\frac{\lambda_q}{\sigma_n^2} - 1 \right) \right]^{p-q+1} \right\} + 0.5 \log N(|\Theta_{q-1}| - |\Theta_q|) \tag{19}$$

$$\sigma^2 = N \left(1 + \frac{1}{p-q} \right) \frac{(p-q)^2(\lambda_q - \sigma_n^2)^2}{(\lambda_q + (p-q)\sigma_n^2)^2}. \tag{20}$$

This result is also given in [11] and [12], where the invariance principal of the maximum likelihood estimator and the asymptotic distribution of the eigenvalues of the empirical correlation matrix are used.

B. GMDL With Digital Sources

As mentioned, the GMDL estimator (4) is a consistent estimator of the number of sources, even if the sources are non-Gaussian [10]. In this subsection, we evaluate its explicit performance when the sources are digital modulated sources.

Assume that each signal at each time instant $s_i(t_j)$ is a discrete random variable taking values in some signal constellation, which is denoted by $\mathcal{Z}$, of size $|\mathcal{Z}|$. The signal vector $\mathbf{s}(t_i)$ is a discrete vector taking values in Z^q. $\mathbf{x}(t_j) = \mathbf{A}\mathbf{s}(t_j) + \mathbf{n}(t_j)$ is the discrete time baseband equivalent model, after down conversion, match filtering, and sampling at the symbol rate. The pdf of $\mathbf{x}(t)$, given $\mathbf{s}(t)$, is complex Gaussian, with mean $\mathbf{A}\mathbf{s}(t)$ and covariance matrix $\sigma_n^2\mathbf{I}$. The pdf of $\mathbf{x}(t)$ is a finite, mixed, complex Gaussian distribution. This mixture has $|\mathcal{Z}|^q$ components, each component corresponds to a different value of the signal vector, and the covariance matrix is $\sigma_n^2\mathbf{I}$ [15]. That is

$$g(\mathbf{x}) \triangleq f(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{|\mathcal{Z}|^q} \sum_{i=1}^{|\mathcal{Z}|^q} \mathcal{CN}(\mathbf{A}\mathbf{s}_i, \sigma_n^2\mathbf{I}) \tag{21}$$

where $\mathbf{s}_i$ is one of the $|\mathcal{Z}|^q$ possible values of the vector $\mathbf{s}(t)$. $\boldsymbol{\theta}$ is the unknown parameter vector. We assume, without loss of generality, that the signals are independent, and thus, any scaling can be absorbed in the mixing matrix $\mathbf{A}$, i.e., $\mathbf{R_s} = \mathbf{I}$.

Computing the performance of the GMDL estimator with digital signals requires a similar procedure to the one done in Section III-A1. The sole difference is that now, the expectations in (16) and (17) are done with respect to the true distribution of the received vector given by (21) and not with respect to the Gaussian distribution.

The correlation matrix of $\mathbf{x}(t)$ is given by $\mathbf{R_x} = \mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I}$. According to (6) and (7), we first have to find the closest parameter vectors to the true distribution $g(\mathbf{x})$, $\boldsymbol{\theta}_q^* \in \boldsymbol{\Theta}_q$, and $\boldsymbol{\theta}_{q-1}^* \in \boldsymbol{\Theta}_{q-1}$. Straightforward evaluation yields

$$\boldsymbol{\theta}_q^* = [\lambda_1, \ldots, \lambda_q, \sigma_n^2, \mathbf{v}_1, \ldots, \mathbf{v}_q] \tag{22}$$

$$\boldsymbol{\theta}_{q-1}^* = \left[\lambda_1, \ldots, \lambda_{q-1}, \frac{(p-q)\sigma_n^2 + \lambda_q}{p-q+1}, \mathbf{v}_1, \ldots, \mathbf{v}_{q-1} \right] \tag{23}$$

where $\lambda_1 > \cdots > \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = \sigma_n^2$ are the eigenvalues of $\mathbf{R_x}$, and $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are their corresponding eigenvectors.

The derivation of μ, σ^2 is carried out in Appendix D. The final result is

$$\mu = -N \log |\mathbf{R}(\boldsymbol{\theta}_q^*)| + N \log |\mathbf{R}(\boldsymbol{\theta}_{q-1}^*)| + 0.5 \log N(|\boldsymbol{\Theta}_{q-1}| - |\boldsymbol{\Theta}_q|) \tag{24}$$

$$\begin{aligned} \sigma^2 = \frac{N}{|Z|^q} \sum_{i=1}^{|Z|^q} & \sigma_n^4 tr\{(\mathbf{R}^{-1}(\boldsymbol{\theta}_q^*) - \mathbf{R}^{-1}(\boldsymbol{\theta}_{q-1}^*))^2\} \\ & + \sigma_n^4(\mathbf{s}_i^H \mathbf{A}^H (\mathbf{R}^{-1}(\boldsymbol{\theta}_q^*) - \mathbf{R}^{-1}(\boldsymbol{\theta}_{q-1}^*))\mathbf{A}\mathbf{s}_i)^2 \\ & + 2\mathbf{s}_i \mathbf{A}^H (\mathbf{R}^{-1}(\boldsymbol{\theta}_q^*) - \mathbf{R}^{-1}(\boldsymbol{\theta}_{q-1}^*))^2 \mathbf{A}\mathbf{s}_i. \end{aligned} \tag{25}$$

The results of (24) and (25) provide a way to assess the performance of the GMDL estimator in almost any scenario. Since almost any sources distribution $g(\mathbf{x})$ can be approximated by a Gaussian mixture, a similar procedure to the one resulting in (24) and (25) can be carried out. This procedure will result in closed-form expressions for μ and for σ^2, similarly to (24) and (25).

IV. Simulations

This section compares the theoretical performance predicted by the proposed method for approximating the probability of misdetection, with the probability of misdetection obtained by simulations. In all the simulations, a uniform linear array of seven omni-directional elements and half-wavelength spacing is used. The sources are taken to be far-field, narrowband, zero-mean, and temporally white random processes. In all the experiments, the noise is a white, zero mean, complex Gaussian random process. For each point on the various graphs, 5000 Monte Carlo runs were carried out, and the probability of misdetection was computed as the number of runs in which the estimated number of sources was smaller than the true number of sources normalized by 5000.

In the first set of simulations, the improvement in performance, due to the use of prior information about the array geometry, of the MDL estimator (3) over the GMDL estimator (4) is examined. Consider the case of two equipower Gaussian sources at bearings $-5°$ and $10°$. Assume that the MDL estimator (3) is used. In this case, the unknown parameters are the sources bearings, the sources correlation matrix, and the noise level. $N = 900$ snapshots are taken. The signal-to-noise ratio per element $SNR \triangleq E\{s^2\}/\sigma^2$ is varied between -16.5 dB and -10 dB. P_{MD} as a function of the SNR is depicted in Fig. 1.

As can be seen from the graph, when using the MDL estimator, a significant advantage of about 3 dB is gained over the GMDL estimator. Moreover, the theoretical results predict very accurately the empirical ones. Computing the MDL estimator (3) requires huge computational efforts. Each point on the graph is based on 5000 Monte Carlo trials. In each trial, the ML estimates of the sources bearings assuming 0, 1, 2, and 3 sources were obtained using multidimensional searches. The estimates were plugged into the log-likelihood function, and after proper computations, an estimate for the number of sources was obtained. By noting that computing the ML estimates of the bearings of q source requires a q-dimensional search, a total of 5000 single-dimensional and 10 000 multidimensional searches

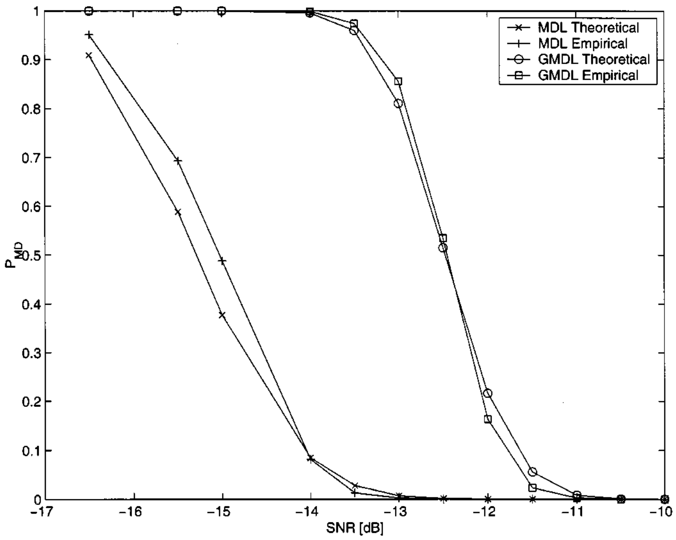

Fig. 1. P_{MD} as a function of the SNR.

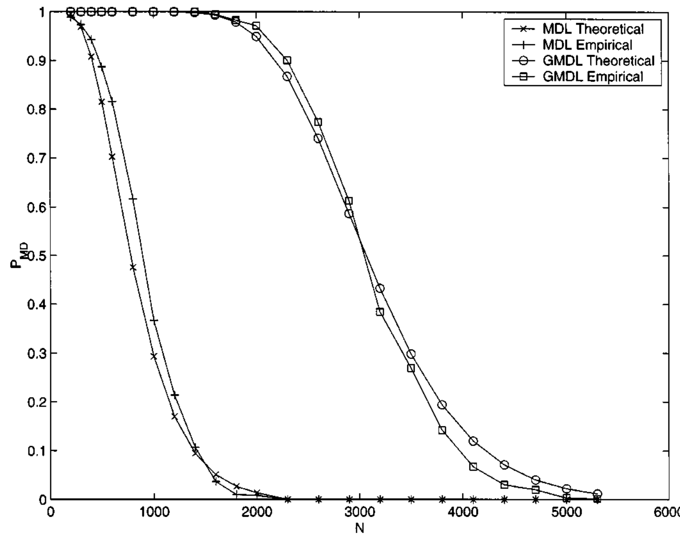

Fig. 2. P_{MD} as a function of the number of snapshots.

were required for evaluating each point on the graph. For computing the theoretical graph, the following general approach is taken. First, $\boldsymbol{\theta}^*_{q-1}$ is evaluated with the aid of (15). Note that computing $\boldsymbol{\theta}^*_{q-1}$ requires one $q-1$-dimensional search. The result $\boldsymbol{\theta}^*_{q-1}$ is then plugged into (16) and (17), which results in the asymptotic mean and variance of the random variable $\mathrm{MDL}(q-1)-\mathrm{MDL}(q)$. Note that this requires only very simple operations like trace and determinant. The asymptotic mean and variance are used in conjunction with (12) for obtaining the desired approximation for the probability of misdetection. The difference in the complexity of computing the empirical graph and the theoretical graph is clearly evident. The theoretical approximation may be used as a simple mean for estimating the performance of the MDL estimator, instead of using simulations.

Fig. 2 depicts the same scenario, where the SNR is held fixed at -15 dB, whereas the number of snapshots is varied between 200 to 5400. Again, it is seen that theoretical performance fits well the empirical performance of both the MDL estimator and the GMDL estimator. It is also seen that the MDL estimator has significant advantage over the GMDL estimator. The GMDL estimator needs four times more snapshots in order to perform as well as the MDL estimator.

In the next set of simulations, we examine the performance of the GMDL estimator under various sources signals distributions. Again, consider the case of two equipower sources at bearing $-5°$ and $10°$. Assume that the signal-to-noise ratio per element $SNR \triangleq E\{s^2\}/\sigma^2$ is -10 dB. We use the GMDL estimator for estimating the number of sources. Table I describes the probability of misdetection predicted by the theoretical analysis (19) and (20) as well as (24) and (25) and the empirical probability of misdetection based on simulations as a function of the number of snapshots. We repeat this experiment for Gaussian and for BPSK, QPSK, and 16QAM source signals.

We first note that for all the cases the theoretical approximations are close to the empirical values obtained by simulations. The difference between the theoretical approximation and the empirical results is no more than 5%. Table I also shows that the GMDL estimator is very robust. Its performance is hardly

TABLE I
P_{MD} OF THE GMDL AS A FUNCTION OF THE NUMBER OF SNAPSHOTS FOR VARIOUS SOURCES DISTRIBUTIONS

	N	100	200	300	400	500
Gaussian	Theoretical	0.9968	0.8257	0.4142	0.1356	0.0336
	Empirical	0.998	0.8715	0.4115	0.095	0.0113
BPSK	Theoretical	0.9974	0.8319	0.4120	0.1290	0.0302
	Empirical	0.9962	0.8765	0.3940	0.0823	0.0075
QPSK	Theoretical	0.9975	0.8335	0.4114	0.1279	0.0293
	Empirical	0.9982	0.8688	0.3992	0.0778	0.0067
16QAM	Theoretical	0.9973	0.8310	0.4124	0.1304	0.0307
	Empirical	0.9972	0.8740	0.4080	0.0825	0.0063

affected by changes in the sources distribution from Gaussian to the digital sources. In all the experiments conducted, the probability of false alarm was very small (less than 1%), and thus, the theoretical results can be used to predict the total probability of error.

Table II describes the performance in a similar scenario, where the number of snapshots is held fixed at 900 for various *SNR* values. Again, it is seen that the theoretical performance fits well the empirical performance and that the performance of the GMDL estimator is hardly affected by the exact distribution of the sources.

V. Discussion

The simulations of the GMDL estimator show that it is a consistent estimator for the number of sources, insensitive to their actual pdf. Its robustness stems from the fact that it only makes use of second-order statistics. Moreover, the simulations show an interesting phenomenon. Whenever the probability of misdetection is lower (higher) than 50%, the detector performs better

TABLE II
P_{MD} OF THE GMDL AS A FUNCTION OF THE SNR FOR VARIOUS SOURCE DISTRIBUTIONS

	SNR [dB]	-14	-13.5	-13	-12.5	-12	-11.5
Gaussian	Theoretical	0.996	0.9599	0.8105	0.515	0.2164	0.0556
	Empirical	0.9975	0.9748	0.849	0.5303	0.1668	0.0243
BPSK	Theoretical	0.9963	0.9611	0.8128	0.5152	0.2132	0.0527
	Empirical	0.998	0.9775	0.8465	0.5192	0.1643	0.0203
QPSK	Theoretical	0.9963	0.9614	0.8134	0.5152	0.2124	0.052
	Empirical	0.9982	0.9770	0.8492	0.5050	0.1643	0.0238
16QAM	Theoretical	0.9962	0.9610	0.8125	0.5152	0.2137	0.0531
	Empirical	0.998	0.9768	0.8538	0.5145	0.1518	0.0222

(worse) with digital sources than Gaussian sources. In a companion paper [16], this phenomenon is investigated thoroughly. In estimation problems, robustness means that the estimator's performance is not degraded considerably due to changes in the assumed statistical model. As can be seen from both the simulation results and the theoretical analysis, the performance of the GMDL estimator improves if some specific types of mismatch exist. This phenomena, where mismatch can cause an improvement in the detector performance, is, as far as we know, reported here for the first time.

One of the main questions that arises with any asymptotic analysis is the following: What is the minimum number of snapshots required to the analysis results to hold? The proof of Theorem 1 provides the answer for this question. The analysis results hold when the ML estimate is unbiased and is normally distributed. When these conditions do not hold, the analysis results cannot be guaranteed to predict accurately the performance of the MDL estimator. For high *SNR,* the probability of misdetection decreases to zero for smaller N than predicted by the asymptotic analysis. This happens since the ML estimates are biased, and the bias "helps" the detection procedure to discriminate between the possible number of sources. Since this "bias" is not considered by the asymptotic analysis, the results of this analysis can be used as an upper bound for the probability of misdetection.

Although the GMDL estimator performs quite well when the sources are digital or not all the *a priori* information about the array is utilized, this performance is worse than that of the MDL estimator designed for the given source distribution. As can be seen in the simulations, in the case of Gaussian sources, the MDL estimator (3) exhibits a 3-dB gain over the GMDL estimator when the array geometry is properly used. In [13], even larger performance gains are exhibit by the MDL estimator (3) designed for digital sources over the GMDL estimator. Another disadvantage in using the GMDL estimator for digital signals is related to the number of sources that can be detected. Moreover, the GMDL estimator for the number of sources can detect up to $p-1$ noncoherent sources. In many practical scenarios, the number of sources is much larger than the number of sensors. For example, in the IS-95 cellular system [17], the number of sources received by an array of no more than four sensors, located at the base station, can be up to 20 different sources. In [13], it was proven that the maximal number of digital sources of a given modulation that can be detected by the MDL designed for that digital sources is not limited to $p-1$. The MDL estimator has an improved performance and capabilities over the GMDL estimator at the price of an increase in the estimator complexity. This increase might be quite high, for example, if the MDL estimator that utilizes the array geometry requires a series of multidimensional searches. Thus, if one does not require this performance improvement, or the number of sources to be detected is less than $p-1$, the GMDL will usually suffice.

Another interesting question that arises is whether utilizing *a priori* information always improves the performance of the MDL estimator. In problems associated with parameter estimation, it is well known that by reducing the number of unknown parameters, there is no degradation, and in some cases even improvement, in the performance of the estimator for the remaining unknown parameters. Although it is widely believed that this is the case in detection problem as well, no formal proof exists. In what follows, we demonstrate that for the cases of Gaussian sources by reducing the number of unknown parameters, performance improvement occurs. Consider the case of two parametric models, where one is nested in the other, i.e., $\boldsymbol{\Theta}'_{q-1} \subseteq \boldsymbol{\Theta}_{q-1}$. For example, such two parametric models can be found when the noise level is known in advance, and thus, $\boldsymbol{\Theta}'_{q-1}$ does not contain the noise level as an unknown parameter, whereas one prefers to treat the noise level as unknown and, thus, to use the parametric model $\boldsymbol{\Theta}_{q-1}$ that contains the noise level as unknown parameter. In such cases, the asymptotic mean and variance of $\mathrm{MDL}(q-1) - \mathrm{MDL}(q)$ satisfy

$$\mu(\boldsymbol{\theta}) \leq \mu(\boldsymbol{\theta}') \qquad \sigma^2(\boldsymbol{\theta}) \leq \sigma^2(\boldsymbol{\theta}') \tag{26}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\theta}' \in \boldsymbol{\Theta}'$. For the asymptotic mean, this is easily seen since

$$\begin{aligned}\mu &= 0.5 \log N(|\boldsymbol{\Theta}_{q-1}| - |\boldsymbol{\Theta}_q|) + ND\left(f(\mathbf{x}|\boldsymbol{\theta}^*_{q-1}) \| f\left(\mathbf{x}|\boldsymbol{\theta}_0^{(q)}\right)\right)\\ &= 0.5 \log N(|\boldsymbol{\Theta}_q| - |\boldsymbol{\Theta}_{q-1}|) + N \min_{\boldsymbol{\theta}^{(q-1)} \in \boldsymbol{\Theta}_{q-1}}\\ &\quad \cdot D\left(f\left(\mathbf{x}|\boldsymbol{\theta}^{(q-1)}\right) \| f\left(\mathbf{x}|\boldsymbol{\theta}_0^{(q)}\right)\right) \geq 0\end{aligned} \tag{27}$$

and thus

$$\begin{aligned}&\mu(\boldsymbol{\theta}') - \mu(\boldsymbol{\theta})\\ &= C \log N + N\left(\min_{\boldsymbol{\theta}^{(q-1)} \in \boldsymbol{\Theta}'_{q-1}} D\left(f\left(\mathbf{x}|\boldsymbol{\theta}^{(q-1)}\right) \| f\left(\mathbf{x}|\boldsymbol{\theta}_0^{(q)}\right)\right)\right.\\ &\quad \left. - \min_{\boldsymbol{\theta}^{(q-1)} \in \boldsymbol{\Theta}_{q-1}} D\left(f\left(\mathbf{x}|\boldsymbol{\theta}^{(q-1)}\right) \| f\left(\mathbf{x}|\boldsymbol{\theta}_0^{(q)}\right)\right)\right) \geq 0\end{aligned} \tag{28}$$

where the last term is positive since $\boldsymbol{\Theta}'_{q-1} \subseteq \boldsymbol{\Theta}_{q-1}$.

The same can be seen for the asymptotic variance. This implies that the probability of error *increases* whenever a more general parameterization is used. In particular, the probability of misdetection is smaller when the model of structured array is applied instead of an unstructured array. This intuitive property is simply because a structured array model implies the use of additional *a priori* knowledge.

APPENDIX A
PROOF THEOREM 1

In this Appendix, Theorem 1 is proven. This proof is based on the derivations done in [14] and on the proof of [18, Th. 1]. Denote by T the random variable MDL$(q-1)-$ MDL(q). It can be easily seen that T is equal

$$\begin{aligned} T =& -\log f\left(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q-1)}\right)+\log f\left(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)}\right) \\ &+0.5\log N(|\boldsymbol{\Theta}_{q-1}|-|\boldsymbol{\Theta}_q|) \\ =& -\sum_{i=1}^{N}\log f\left(\mathbf{x}_i|\hat{\boldsymbol{\theta}}^{(q-1)}\right)+\sum_{i=1}^{N}\log f\left(\mathbf{x}_i|\hat{\boldsymbol{\theta}}^{(q)}\right) \\ &+0.5\log N\left(|\boldsymbol{\Theta}_{q-1}|-|\boldsymbol{\Theta}_q|\right) \end{aligned} \tag{29}$$

where $\mathbf{x}_i$ denotes the measurement at instance t_i, $\mathbf{x}(t_i)$.

The ML estimators $\hat{\boldsymbol{\theta}}^{(q)}$ and $\hat{\boldsymbol{\theta}}^{(q-1)}$ are consistent estimators for $\boldsymbol{\theta}_q^*$ and $\boldsymbol{\theta}_{q-1}^*$, respectively [14]. The first two elements of the Taylor's expansion of the functions $\log f(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})$ and $\log f(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q-1)})$, around $\boldsymbol{\theta}_q^*$ and $\boldsymbol{\theta}_{q-1}^*$, respectively, can be used as an approximation for $\log f(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})$ and $\log f(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q-1)})$ [18]

$$\begin{aligned} &\sum_{i=1}^{N}\log f\left(\mathbf{x}_i|\hat{\boldsymbol{\theta}}^{(q-1)}\right) \\ &\quad\approx\sum_{i=1}^{N}\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)+\frac{N}{2}\left(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)}\right)^T \\ &\quad\cdot H(\boldsymbol{\theta}_{q-1}^*)\left(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)}\right) \end{aligned} \tag{30}$$

$$\begin{aligned} &\sum_{i=1}^{N}\log f\left(\mathbf{x}_i|\hat{\boldsymbol{\theta}}^{(q)}\right) \\ &\quad\approx\sum_{i=1}^{N}\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q}^*)+\frac{N}{2}\left(\boldsymbol{\theta}_{q}^*-\hat{\boldsymbol{\theta}}^{(q)}\right)^T \\ &\quad\cdot H(\boldsymbol{\theta}_{q}^*)\left(\boldsymbol{\theta}_{q}^*-\hat{\boldsymbol{\theta}}^{(q)}\right). \end{aligned} \tag{31}$$

The asymptotic distribution of $\sqrt{N}(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)})$, $\sqrt{N}(\boldsymbol{\theta}_q^*-\hat{\boldsymbol{\theta}}^{(q)})$ is [14]

$$\begin{aligned} &\sqrt{N}\left(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)}\right) \\ &\quad\sim\mathcal{N}(0,\,H^{-1}(\boldsymbol{\theta}_{q-1}^*)J(\boldsymbol{\theta}_{q-1}^*)H^{-1}(\boldsymbol{\theta}_{q-1}^*)) \end{aligned} \tag{32}$$

$$\sqrt{N}\left(\boldsymbol{\theta}_{q}^*-\hat{\boldsymbol{\theta}}^{(q)}\right)\sim\mathcal{N}(0,\,H^{-1}(\boldsymbol{\theta}_{q}^*)J(\boldsymbol{\theta}_{q}^*)H^{-1}(\boldsymbol{\theta}_{q}^*)). \tag{33}$$

Plugging (30) and (31) back into (29) results in the following approximation:

$$\begin{aligned} T \approx& -\sum_{i=1}^{N}\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)-\frac{N}{2}\left(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)}\right)^T \\ &\cdot H\left(\boldsymbol{\theta}_{q-1}^*\right)\left(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)}\right)+\sum_{i=1}^{N}\log f(\mathbf{x}_i|\boldsymbol{\theta}_q^*) \\ &+\frac{N}{2}\left(\boldsymbol{\theta}_q^*-\hat{\boldsymbol{\theta}}^{(q)}\right)^T H(\boldsymbol{\theta}_q^*)\left(\boldsymbol{\theta}_q^*-\hat{\boldsymbol{\theta}}^{(q)}\right) \\ &+0.5\log N(|\boldsymbol{\Theta}_{q-1}|-|\boldsymbol{\Theta}_q|) \\ =& \sum_{i=1}^{N}[-\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}_i|\boldsymbol{\theta}_q^*)] \\ &-\frac{N}{2}\left(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)}\right)^T H(\boldsymbol{\theta}_{q-1}^*)\left(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)}\right) \\ &+\frac{N}{2}\left(\boldsymbol{\theta}_q^*-\hat{\boldsymbol{\theta}}^{(q)}\right)^T H(\boldsymbol{\theta}_q^*)\left(\boldsymbol{\theta}_q^*-\hat{\boldsymbol{\theta}}^{(q)}\right) \\ &+0.5\log N(|\boldsymbol{\Theta}_{q-1}|-|\boldsymbol{\Theta}_q|). \end{aligned} \tag{34}$$

The exact asymptotic distribution of T (34) is very complicated. This approximation can be further simplified by noting that $(N/2)(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)})^T H(\boldsymbol{\theta}_{q-1}^*)(\boldsymbol{\theta}_{q-1}^*-\hat{\boldsymbol{\theta}}^{(q-1)})$ and $(N/2)(\boldsymbol{\theta}_q^*-\hat{\boldsymbol{\theta}}^{(q)})^T H(\boldsymbol{\theta}_q^*)(\boldsymbol{\theta}_q^*-\hat{\boldsymbol{\theta}}^{(q)})$ are Gaussian quadratic forms, as such it can be shown that to be distributed as a weighted sum of chi square random variables with one degree of freedom, that is, $\sum_i \nu_i\chi_{(1)}^2$, $\sum_i \nu_i'\chi_{(1)}^2$, respectively. Noting that the mean and variance of these quadratic forms are bounded and that as N approaches infinity, the mean and variance of $\sum_{i=1}^N[-\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}_i|\boldsymbol{\theta}_q^*)]$ are unbounded, the approximation of T can be further simplified by neglecting the quadratic forms. As a result, T is asymptotically equal:

$$\begin{aligned} T =& \,\mathrm{MDL}(q-1)-\mathrm{MDL}(q) \\ &\approx\sum_{i=1}^{N}[-\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}_i|\boldsymbol{\theta}_q^*)] \\ &\quad+0.5\log N(|\boldsymbol{\Theta}_{q-1}|-|\boldsymbol{\Theta}_q|). \end{aligned} \tag{35}$$

Since the first term is the sum of independent and identically distributed random variables, the central limit theory can be invoked, and thus, for large N, $\sum_{i=1}^N[-\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}_i|\boldsymbol{\theta}_q^*)]$ is asymptotically distributed as a Gaussian random variable. Thus, the asymptotic distribution of MDL$(q-1)-$ MDL(q) is given in the following:

$$\mathrm{MDL}(q-1)-\mathrm{MDL}(q)\sim\mathcal{N}(\mu,\,\sigma^2) \tag{36}$$

$$\begin{aligned} \mu =& N\mathrm{E}_g\{-\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)-\log f(\mathbf{x}_i|\boldsymbol{\theta}_q^*)\} \\ &+0.5\log N(|\boldsymbol{\Theta}_{q-1}|-|\boldsymbol{\Theta}_q|) \end{aligned} \tag{37}$$

$$\sigma^2 = N\,Var_g\{-\log f(\mathbf{x}_i|\boldsymbol{\theta}_{q-1}^*)+\log f(\mathbf{x}_i|\boldsymbol{\theta}_q^*)\}. \tag{38}$$

APPENDIX B
ASYMPTOTIC MEAN AND VARIANCE OF MDL$(q-1)-$MDL(q)

In this Appendix, we evaluate the asymptotic mean and variance of the random variable MDL$(q-1)-$ MDL(q)

$$\begin{aligned} &\frac{1}{N}\left(\mu+0.5\log N(|\boldsymbol{\Theta}_q|-|\boldsymbol{\Theta}_{q-1}|)\right) \\ &\quad=E\{\log f(\mathbf{x}|\boldsymbol{\theta}_q^*)-\log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*)\} \\ &\quad=-\log|\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_q^*)|+\log|\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_{q-1}^*)| \\ &\qquad-E\{\mathbf{x}^H([\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_q^*)]^{-1}-[\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_{q-1}^*)]^{-1})\mathbf{x}\} \\ &\quad=-\log|\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_q^*)|+\log|\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_{q-1}^*)| \\ &\qquad-\mathrm{Tr}\{\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_q^*)([\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_q^*)]^{-1}-[\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_{q-1}^*)]^{-1})\} \\ &\quad=-\log|\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_q^*)|+\log|\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_{q-1}^*)|-p \\ &\qquad+\mathrm{Tr}\{\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_q^*)[\mathbf{R}_{\mathbf{x}}(\boldsymbol{\theta}_{q-1}^*)]^{-1}\} \end{aligned} \tag{39}$$

$$\begin{aligned}\frac{1}{N}\sigma^2 &= E\{(\log f(\mathbf{x}|\boldsymbol{\theta}_q^*) - \log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*)\\ &\qquad - 0.5\log N(|\Theta_q| - |\Theta_{q-1}|) - \mu)^2\}\\ &= E\{(\mathbf{x}^H([\mathbf{R_x}(\boldsymbol{\theta}_q^*)]^{-1} - [\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1})\mathbf{x}\\ &\qquad - \mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_q^*)([\mathbf{R_x}(\boldsymbol{\theta}_q^*)]^{-1} - [\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1})\})^2\}\\ &= Var\{\mathbf{x}^H([\mathbf{R_x}(\boldsymbol{\theta}_q^*)]^{-1} - [\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1})\mathbf{x}\}\\ &= \mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_q^*)([\mathbf{R_x}(\boldsymbol{\theta}_q^*)]^{-1} - [\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1})\\ &\qquad \cdot \mathbf{R_x}(\boldsymbol{\theta}_q^*)([\mathbf{R_x}(\boldsymbol{\theta}_q^*)]^{-1} - [\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1})\}\\ &= \mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_q^*)[\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1}\mathbf{R_x}(\boldsymbol{\theta}_q^*)[\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1}\}\\ &\qquad - 2\mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_q^*)[\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)]^{-1}\} + p \end{aligned} \tag{40}$$

where we used the fact that for a zero mean, complex, Gaussian, random vector $\mathbf{x}$ with correlation matrix $\mathbf{R_x}$, the mean of the quadratic form $\mathbf{x}^H\mathbf{P}\mathbf{x}$ equals $\mathrm{Tr}\{\mathbf{R_x}\mathbf{P}\}$, and the variance of the same quadratic form is $\mathrm{Tr}\{\mathbf{R_x}\mathbf{P}\mathbf{R_X}\mathbf{P}\}$.

Appendix C
Gaussian Sources

In this Appendix, we compute (16) and (17) for the case of unstructured arrays.

Using the spectral representation, we get

$$\mathbf{R_x}(\boldsymbol{\theta}_q^*) = \sum_{i=1}^{q}\lambda_i\mathbf{v}_i\mathbf{v}_i^H + \sum_{i=q+1}^{p}\sigma_n^2\mathbf{v}_i\mathbf{v}_i^H \tag{41}$$

$$\begin{aligned}&\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)\\ &\quad= \sum_{i=1}^{q-1}\lambda_i\mathbf{v}_i\mathbf{v}_i^H + \sum_{i=q-1}^{p}\frac{(p-q)\sigma_n^2+\lambda_q}{p-q+1}\mathbf{v}_i\mathbf{v}_i^H\end{aligned} \tag{42}$$

$$\begin{aligned}&\mathbf{R}_\mathbf{x}^{-1}(\boldsymbol{\theta}_{q-1}^*)\\ &\quad= \sum_{i=1}^{q-1}\frac{1}{\lambda_i}\mathbf{v}_i\mathbf{v}_i^H + \sum_{i=q-1}^{p}\frac{p-q+1}{(p-q)\sigma_n^2+\lambda_q}\mathbf{v}_i\mathbf{v}_i^H\end{aligned} \tag{43}$$

$$\begin{aligned}&\mathbf{R_x}(\boldsymbol{\theta}_q^*)\mathbf{R}_\mathbf{x}^{-1}(\boldsymbol{\theta}_{q-1}^*)\\ &\quad= \sum_{i=1}^{q-1}\mathbf{v}_i\mathbf{v}_i^H + \frac{(p-q+1)\lambda_q}{(p-q)\sigma_n^2+\lambda_q}\mathbf{v}_{q-1}\mathbf{v}_{q-1}^H\\ &\qquad + \sum_{i=q+1}^{p}\frac{(p-q+1)\sigma_n^2}{(p-q)\sigma_n^2+\lambda_q}\mathbf{v}_i\mathbf{v}_i^H.\end{aligned} \tag{44}$$

Thus, we get that the eigenvalues of $\mathbf{R_x}(\boldsymbol{\theta}_q^*)\mathbf{R}_\mathbf{x}^{-1}(\boldsymbol{\theta}_{q-1}^*)$ are equal to

$$\left[\mathbf{1}_{q-1},\ \frac{(p-q+1)\lambda_q}{(p-q)\sigma_n^2+\lambda_q},\ \frac{(p-q+1)\sigma_n^2}{(p-q)\sigma_n^2+\lambda_q}\mathbf{1}_{p-1}\right]$$

where $\mathbf{1}_m$ is the all-1s vector of length m. Note that $\mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_q^*)\mathbf{R}_\mathbf{x}^{-1}(\boldsymbol{\theta}_{q-1}^*)\}$ is equal to the sum of the eigenvalues; thus, $\mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_q^*)\mathbf{R}_\mathbf{x}^{-1}(\boldsymbol{\theta}_{q-1}^*)\} = p$.

The evaluation of μ and σ^2 is given hereafter:

$$\begin{aligned}\mu &= -\log|\mathbf{R_x}(\boldsymbol{\theta}_q^*)| + \log|\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)| - p\\ &\quad + \mathrm{Tr}\{\mathbf{R_x}(\boldsymbol{\theta}_q^*)\mathbf{R}_\mathbf{x}^{-1}(\boldsymbol{\theta}_{q-1}^*)\}\\ &= -\sum_{i=1}^{q}\log\lambda_i - (p-q)\log\sigma_n^2 + \sum_{i=1}^{q-1}\log\lambda_i\\ &\quad + (p-q+1)\log\frac{(p-q)\sigma_n^2+\lambda_q}{p-q+1} - p + p\\ &= (p-q+1)\log\frac{(p-q)\sigma_n^2+\lambda_q}{p-q+1} - \log\lambda_q - (p-q)\log\sigma^2\\ &= \log\left\{\frac{\sigma_n^2}{\lambda_q}\left[1+\frac{1}{p-q+1}\left(\frac{\lambda_q}{\sigma_n^2}-1\right)\right]^{p-q+1}\right\}.\end{aligned} \tag{45}$$

Using the same procedure, the variance σ^2 is equal to

$$\sigma^2 = \left(1+\frac{1}{p-q}\right)\frac{1}{N}\frac{(p-q)^2(\lambda_q-\sigma_n^2)^2}{(\lambda_q+(p-q)\sigma_n^2)^2}. \tag{46}$$

Appendix D
Digital Sources

In this Appendix, we compute the mean and the variance of the random variable $\mathrm{MDL}(q) - \mathrm{MDL}(q-1)$ when the sources are digital but were assumed to be Gaussian sources

$$\begin{aligned}&\frac{1}{N}\mu - 0.5\log N(|\Theta_{q-1}| - |\Theta_q|)\\ &\quad= \mathrm{E}_g\{\log f(\mathbf{x}|\boldsymbol{\theta}_q^*) - \log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*)\}\\ &\quad= -\log|\mathbf{R_x}(\boldsymbol{\theta}_q^*)| + \log|\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)|\\ &\qquad - \mathrm{E}_g\{\mathbf{x}^H(\mathbf{R_x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)^{-1})\mathbf{x}\}\\ &\quad= -\log|\mathbf{R_x}(\boldsymbol{\theta}_q^*)| + \log|\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)|\\ &\qquad + \frac{1}{|Z|^q}\sum_{i=1}^{|Z|^q}\int_{-\infty}^{\infty}\mathbf{x}^H(\mathbf{R_x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)^{-1})\\ &\qquad \cdot \mathbf{x}\frac{1}{\pi^p\sigma_n^{2p}}e^{-(1/\sigma_n^2)(\mathbf{x}-\mathbf{A}\mathbf{s}_i)^H(\mathbf{x}-\mathbf{A}\mathbf{s}_i)}\,d\mathbf{x}.\end{aligned} \tag{47}$$

Each term in the sum is the mean of Gaussian quadratic form, where $\mathbf{x} \sim \mathcal{CN}(\mathbf{A}\mathbf{s}_i,\ \sigma_n^2\mathbf{I})$, and using the fact that for $\mathbf{X} \sim \mathcal{CN}(\boldsymbol{\mu},\ \mathbf{C})$ the mean and the second moment of the quadratic form $Y = \mathbf{X}^H\mathbf{P}\mathbf{X}$ is given by $E\{Y\} = \mathrm{Tr}\{\mathbf{C}\mathbf{P}\} + \boldsymbol{\mu}^H\mathbf{P}\boldsymbol{\mu}$ and $E\{Y^2\} = \mathrm{Tr}\{\mathbf{C}\mathbf{P}\mathbf{C}\mathbf{P}\} + (\mathrm{Tr}\{\mathbf{C}\mathbf{P}\})^2 + 2\mathrm{Tr}\{\mathbf{C}\mathbf{P}\}\boldsymbol{\mu}^H\mathbf{P}\boldsymbol{\mu} + (\boldsymbol{\mu}^H\mathbf{P}\boldsymbol{\mu})^2 + 2\boldsymbol{\mu}^H\mathbf{P}\mathbf{C}\mathbf{P}\boldsymbol{\mu}$, the mean μ is given by

$$\begin{aligned}&\frac{1}{N}\mu - 0.5\log N(|\Theta_{q-1}| - |\Theta_q|)\\ &\quad= -\log|\mathbf{R_x}(\boldsymbol{\theta}_q^*)| + \log|\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)|\\ &\qquad + \frac{1}{|Z|^q}\sum_{i=1}^{|Z|^q}\mathrm{Tr}\{\sigma_n^2\mathbf{I}(\mathbf{R_x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)^{-1})\}\\ &\qquad + \mathbf{s}_i^H\mathbf{A}^H(\mathbf{R_x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)^{-1})\mathbf{A}\mathbf{s}_i\\ &\quad= -\log|\mathbf{R_x}(\boldsymbol{\theta}_q^*)| + \log|\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)|\\ &\qquad + \frac{1}{|Z|^q}\sum_{i=1}^{|Z|^q}\mathbf{s}_i^H\mathbf{A}^H(\mathbf{R_x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)^{-1})\mathbf{A}\mathbf{s}_i\\ &\quad= -\log|\mathbf{R_x}(\boldsymbol{\theta}_q^*)| + \log|\mathbf{R_x}(\boldsymbol{\theta}_{q-1}^*)|.\end{aligned} \tag{48}$$

The variance of the random variable $\mathrm{MDL}(q-1) - \mathrm{MDL}(q)$ is computed by the following:

$$\begin{aligned}\sigma^2 &= \mathrm{E}_g\{(\log f(\mathbf{x}|\boldsymbol{\theta}_q^*) - \log f(\mathbf{x}|\boldsymbol{\theta}_{q-1}^*) - \mu)^2\} \\ &= \mathrm{E}_g\{(\mathbf{x}^H(\mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_{q-1}^*)^{-1})\mathbf{x})^2\}. \qquad (49)\end{aligned}$$

Using the same method that was used for computing μ

$$\begin{aligned}\sigma^2 = \frac{1}{|Z|^q}\sum_{i=1}^{|Z|^q} &\mathrm{Tr}\{(\mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_{q-1}^*)^{-1})^2\} \\ &+ (\mathbf{s}_i^H\mathbf{A}^H(\mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_{q-1}^*)^{-1})\mathbf{A}\mathbf{s}_i)^2 \\ &+ 2\mathbf{s}_i\mathbf{A}^H(\mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_q^*)^{-1} - \mathbf{R}_\mathbf{x}(\boldsymbol{\theta}_{q-1}^*))^2\mathbf{A}\mathbf{s}_i. \qquad (50)\end{aligned}$$

ACKNOWLEDGMENT

The authors wish to thank the associate editor Prof. A. Gershman and the anonymous reviewer for their constructive comments.

REFERENCES

[1] C. M. Cho and P. M. Djuric, “Detection and estimation of doa’s via Bayesian predictive densities,” *IEEE Trans. Signal Processing*, vol. 42, pp. 3051–3060, Nov. 1994.

[2] E. Fishler and H. Messer, “On the use of order statistics for improved detection of signals by the mdl criterion,” *IEEE Trans. Signal Processing*, vol. 48, pp. 2242–2247, Aug. 2000.

[3] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Feb. 1985.

[4] 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 Trans. Signal Processing*, vol. 38, pp. 1959–1971, Nov. 1990.

[5] H. T. Wu, J. F. Yang, and F. K. Chen, “Source number estimation using transformed gerschgorin radii,” *IEEE Trans. Signal Processing*, vol. 43, pp. 1325–1333, June 1995.

[6] A. M. Zoubir, “Bootstrap methods for model selection,” *AEU-Int. J. Electron. Commun.*, vol. 53, pp. 386–392, 1999.

[7] J. Rissanen, “A universal prior for integers and estimation by minimum description length,” *Ann. Stat.*, vol. 11, no. 2, pp. 431–466, 1983.

[8] Y. Bresler and A. Macovski, “Exact maximum likelihood parameter estimation of superimposed exponential signals in noise,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1081–1089, Oct. 1986.

[9] A. M. Bruckstein, T.-J. Shan, and T. Kailath, “The resolution of overlapping echos,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1357–1367, Dec. 1985.

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

[11] W. Xu and M. Kaveh, “Analysis of the performance and sensitivity of eigendecomposition-based detectors,” *IEEE Trans. Signal Processing*, vol. 43, pp. 1413–1426, June 1995.

[12] Q.-T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, “Statistical analysis of the performance of information theoretic criteria in the detection of the number of sources in array processing,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1557–1566, Oct. 1989.

[13] M. Grosman, “Constellation based detection of the number of sources,” M.S. thesis, Tel Aviv Univ., Tel Aviv, Israel, Feb. 2001.

[14] H. White, “Maximum likelihood estimation of misspecified models,” *Econometrica*, vol. 50, pp. 1–26, 1982.

[15] E. Moulines, J.-F. Cardoso, and E. Gassiat, “Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models,” in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 5, 1997, pp. 3617–3620.

[16] J. Friedmann, E. Fishler, and H. Messer, “General asymptotic analysis of the glrt detector for a Gaussian point source with statistical or spatial mismodeling,” *IEEE Trans. Signal Processing*, submitted for publication.

[17] T. W. Rappaport, *Wireless Communications*, First ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[18] J. T. Kent, “Robust properties of likelihood ratio tests,” *Biometrika*, vol. 69, pp. 19–27, 1982.

Eran Fishler (M’02) was born in Tel Aviv, Israel, in 1972. He received the B.Sc. degree in mathematics and the M.Sc. and Ph.D. degrees in electrical engineering in 1995, 1997, and 2001, respectively, from Tel Aviv University.

From 1993 to 1995, he was with the Israeli Navy as a Research Engineer. From 1998 to 2000, he has been with Tadiran Electronic System, Holon, Israel. He is currently a Lecturer with Tel Aviv University. His research interests include statistical signal processing and communication theory with emphasis on ultra wideband systems.

Michael Grosmann was born in Tel Aviv, Israel, in 1971. He received the B.Sc. (summa cum laude) and M.Sc. (magna cum laude) degrees in electrical engineering in 1994 and 2000, respectively, from Tel Aviv University.

From 1994 to 2000, he was with the Israeli Military Technology Center, first as a DSP Research Engineer and later as the Head of the DSP Research Group. Since 2000, he has been with AudioCodes Ltd. as the System Architect of the Advanced Technology Group.

Hagit Messer (F’01) received the B.Sc, M.Sc., and Ph.D. degrees in electrical engineering from Tel Aviv University, Tel Aviv, Israel, in 1977, 1979, and 1984, respectively.

Since 1977, she has been with the Department of Electrical Engineering—Systems at the Faculty of Engineering, Tel Aviv University, first as a Research and Teaching Assistant and then, after a one-year post-doctoral appointment at Yale University (from 1985 to 1986), as a Faculty Member. Her research interests are in the field of signal processing and statistical signal analysis. Her main research projects are related to tempo-spatial and non-Gaussian signal detection and parameter estimation.

Dr. Messer was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994 to 1996, and she is now an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. She is a member of the Technical Committee of the IEEE Signal Processing Society for Signal Processing, Theory, and Methods and was a co-organizer of the IEEE Workshop HOS’99. Since 2000, she has been the Chief Scientist of the Israeli Ministry of Science, Culture, and Sport.