

Determining the Number of Discrete Alphabet Sources from Sensor Data

Eran Fishler

*Stern School of Business, New York University, NY 10012, USA
Email: eran.fishler@stern.nyu.edu*

Michael Grosman

*School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel
Email: michael.grosman@audiocodes.com*

Hagit Messer

*School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel
Email: messer@eng.tau.ac.il*

Received 22 December 2003; Revised 10 June 2004

Determining the number of sources in a received wave field 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. Under the assumption that the transmitted signals are Gaussian, the MDL estimator takes both a simple and an intuitive form. Therefore this estimator is commonly used even when the signals are known to be non-Gaussian communication signals. However, its ability to resolve signals (resolution capacity) is limited by the number of sensors minus one. In this paper, we study the MDL estimator that is based on the correct, non-Gaussian signal distribution of digital signals. We show that this approach leads to both improved performance and improved resolution capacity, that is, the number of signals that can be detected by the resulting MDL processor is larger than the number of array sensors. In addition, a novel asymptotic performance analysis, which can be used to predict the performance of the MDL estimator analytically, is presented. Simulation results support the theoretical conclusions.

Keywords and phrases: MDL, asymptotic analysis, communication signals.

1. INTRODUCTION

One of the fundamental problems in the field of array processing is to determine the number of signals that impinge on a passive sensors array. The problem arises in several fields of array processing such as radar, and seismology (see, e.g., [1, 2, 3, 4, 5, 6]). This question received considerable interest not only because of its nature, but also because most algorithms for direction-of-arrival (DOA) estimation assume a prior knowledge of the number of signals. Moreover, with the advent of new communication systems and the growing lack of frequency spectrum, efficient techniques of communication are being developed. One of these approaches is to use an antenna array to generate a narrow beam focused at the receiver. This, in turn, enables us to reuse the frequency spectrum for communication with other receivers. The method is being developed for cellular communication where the users move and the cells should sense the

advent of new users. New users are detected by monitoring the change in the number of users that the cell senses. The ability to detect the number of transmitters and their location is also required in spectrum monitoring and control applications.

The most common approach for estimating this number is to apply information theoretic criteria like the minimum description length (MDL) or the Akaike information criterion (AIC) [7]. Since 1985 [3], when a Gaussian version of the MDL was first suggested for estimating the number of narrowband sources impinging on an array of sensors, this MDL-based estimator became a standard tool for accomplishing this task.

1.1. Problem formulation

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

for the received signal vector is

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_q]$ is a $p \times q$ matrix composed of q p -dimensional vectors, $\mathbf{a}_1, \dots, \mathbf{a}_q$. \mathbf{A} is referred to as the steering matrix. $\mathbf{s}(t) = [s_1(t) \ \dots \ s_q(t)]^T$ is a $q \times 1$ signals vector and $\mathbf{n}(t)$ is a $p \times 1$ vector of the additive noise.

Many problems may be formulated using this simple, linear model (see [8, 9] and references therein). These problems differ in the structure of the mixing matrix, \mathbf{A} , in the assumed knowledge about the unknown parameters, or in the statistical modeling of both the signal and the noise. For example, \mathbf{n} and/or \mathbf{s} can be assumed to be Gaussian or non-Gaussian; the additive noise correlation matrix can be assumed to be white or not; or the mixing matrix \mathbf{A} can be assumed to be of full rank or not.

In this paper we are concerned with the problem of estimating q when the sources are digital signals. Denote by $\mathcal{D} = [d_1, \dots, d_{|\mathcal{D}|}]$ an arbitrary signal constellation, where $d_i \in \mathcal{C}$, and $|\mathcal{D}|$ is the size of the constellation. In the sequel, we make the following assumptions.

- (1) The additive noise is zero mean, both spatially and temporal white, complex Gaussian random process, with correlation matrix $\sigma_n^2 \mathbf{I}$.
- (2) The signal vector is uniformly distributed over \mathcal{D}^q , independent from snapshot to snapshot.
- (3) The additive noise and the signal vector are independent.

Based on these assumptions, the probability distribution function (pdf) of the received signal is

$$f(\mathbf{x}_1, \dots, \mathbf{x}_N | \mathbf{A}, \sigma_n^2) = \prod_{i=1}^N \frac{1}{|\mathcal{D}|^q} \sum_{\mathbf{s} \in \mathcal{D}^q} \mathcal{C} \mathcal{N}(\mathbf{x}_i - \mathbf{A}\mathbf{s}, \sigma_n^2 \mathbf{I}), \quad (2)$$

where $\mathcal{C} \mathcal{N}(\mathbf{x}, \mathbf{R}) = (1/\pi^{|\mathbf{R}|})e^{-\mathbf{x}^H \mathbf{R}^{-1} \mathbf{x}}$, and H denotes the complex transpose. We note that, opposed to the common problem formulation [3], we do not assume that \mathbf{A} is a full-rank matrix. As will be shown in the sequel, the full-rank requirement is not necessary for identifiability when the signals are digital signals.

Note that our model, (1), is also used to represent the reception of any multiple access system, for example, code-division-multiple-access (CDMA) systems transmitting over flat fading channels [10]. Our problem is to estimate the number of sources, q , given N independent snapshots of the array output, $\mathbf{x}(t_1), \dots, \mathbf{x}(t_N)$, described by the pdf of (2).

1.2. The MDL approach

The information theoretic criteria approach is a general one for choosing a model that fits the data mostly from a family of possible models [7, 11]. That is, given a parameterized family of probability densities, $f_{\mathbf{X}}(\mathbf{X} | \boldsymbol{\theta}^{(q)})$, $\boldsymbol{\theta}^{(q)} \in \Theta_q$, for various q ,

select \hat{q} such that

$$\hat{q} = \arg \min_q \{ -L(\hat{\boldsymbol{\theta}}^{(q)}) + P(q) \}, \quad (3)$$

where $L(\boldsymbol{\theta}^{(q)}) = \log f_{\mathbf{X}}(\mathbf{X} | \boldsymbol{\theta}^{(q)})$ is the log likelihood of the measurements, denoted by $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, $P(q)$ is a general penalty function associated with the q th family, and $\hat{\boldsymbol{\theta}}^{(q)} = \arg \max_{\boldsymbol{\theta}^{(q)} \in \Theta_q} \{ f_{\mathbf{X}}(\mathbf{X} | \boldsymbol{\theta}^{(q)}) \}$. $\hat{\boldsymbol{\theta}}^{(q)}$ is the maximum likelihood estimate of the unknown parameters, given the q th family of distributions.

The MDL estimator is a special case of (3) with a certain penalty function. It is given by minimizing the MDL metric [7], that is,

$$\begin{aligned} \hat{q}_{\text{MDL}} &= \arg \min_q \text{MDL}(q) \\ &= \arg \min_q \left\{ -\log \left(f_{\mathbf{X}}(\mathbf{X} | \hat{\boldsymbol{\theta}}^{(q)}) \right) + 0.5 |\Theta_q| \log(N) \right\}, \end{aligned} \quad (4)$$

where $\text{MDL}(q) = \{ -\log(f_{\mathbf{X}}(\mathbf{X} | \hat{\boldsymbol{\theta}}^{(q)})) + 0.5 |\Theta_q| \log(N) \}$, and $|\Theta_q|$ is the minimum number of free parameters in Θ_q . 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].

Although in many problems associated with array processing, for example, direction-of-arrival (DOA) estimation, one has some prior knowledge on the signals' statistical properties or on the array geometry, when the number of sources is estimated, this prior knowledge is usually ignored. The reason for this is that by assuming Gaussian sources and regardless of the array geometry, the resulting MDL estimator (4), termed the Gaussian minimum description length (GMDL) estimator, has a simple closed-form expression given by

$$\begin{aligned} \hat{q}_{\text{GMDL}} &= \arg \min_{q=0, \dots, p-1} \left\{ -N \log \frac{\prod_{i=q+1}^p l_i}{\left((1/(p-q)) \sum_{i=q+1}^p l_i \right)^{p-q}} \right. \\ &\quad \left. + \frac{1}{2} (q(2p-q) + 1) \log N \right\}, \end{aligned} \quad (5)$$

where $l_1 \geq l_2 \geq \dots \geq l_p$ are the eigenvalues of the empirical received signal's correlation matrix, $\hat{\mathbf{R}} = (1/N) \sum \mathbf{x}_i \mathbf{x}_i^H$. It is well known that the GMDL estimator is a consistent estimator of the number of sources [12].

In our problem the unknown parameters are the matrix \mathbf{A} and the noise level. Since we do not restrict our attention to full-rank mixing matrices, when the number of sources is assumed to be q , the number of unknown parameters is $2pq + 1$. We denote by $\hat{\mathbf{A}}_q$ and σ_n^2 the MLEs of the unknown parameters assuming q sources. The MDL estimator (4) for

our problem (2) becomes

$$\begin{aligned} \hat{q}|_{\text{MDL}} &= \arg \min_{q=0,1,\dots} -\log f(\mathbf{x}_1, \dots, \mathbf{x}_N | \hat{\mathbf{A}}_q, \hat{\sigma}_{n_q}^2) \\ &\quad + 0.5(2pq + 1) \log N \\ &= \arg \min_{q=0,1,\dots} \left\{ -\sum_{i=1}^N \log \frac{1}{|\mathcal{D}|^q} \sum_{\mathbf{s} \in \mathcal{D}^q} \mathcal{C} \mathcal{N}(\mathbf{x}_i - \hat{\mathbf{A}}_q \mathbf{s}, \hat{\sigma}_{n_q}^2 \mathbf{I}) \right. \\ &\quad \left. + 0.5(2pq + 1) \log N \right\}. \end{aligned} \quad (6)$$

The reader can already spot an important difference between the commonly used GMDL (5) and the exact MDL solution (6). While the GMDL can detect up to $p - 1$ sources, the MDL estimator can detect any number of sources. This property of the MDL estimator will be discussed in Section 2. One should note that in practice, although the MDL can detect any number of sources, this number will be limited by the overall system resources, for example, the number of distinct receivers.

1.3. Paper organization

The paper is organized as follows. Section 2 discusses the resolution capacity of this problem and the asymptotic characteristics of the MDL estimator. Section 3 is devoted to numerical study of the performance of the MDL estimator; whereas Section 4 is devoted to analytical study of the performance of the MDL estimator. Section 5 provides a summary and some concluding remarks.

2. IDENTIFIABILITY AND CONSISTENCY

2.1. Identifiability

Consider a parameterized family of probability density functions $f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. This family of densities is said to be *identifiable* if, for every $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, the divergence between $f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})$ and $f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}')$ is greater than zero, that is, $D(f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})||f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}')) > 0$, where $D(f(\mathbf{x})||g(\mathbf{x})) = \int f \log(f/g) d\mathbf{x}$ is the Kullback-Leibler divergence between $f(\mathbf{x})$ and $g(\mathbf{x})$. This condition ensures that there is a one-to-one relationship between the parameter space and the statistical properties of the measurements.

The model order selection problem [13] discussed in Section 1.1 is unidentifiable if it is possible to find, for some $k \neq l$, two points in the parameter space, $\boldsymbol{\theta}_k \in \Theta_k$ and $\boldsymbol{\theta}_l \in \Theta_l$, such that $f(\cdot|\boldsymbol{\theta}_k) = f(\cdot|\boldsymbol{\theta}_l)$. If this is the case, one cannot distinguish between k sources and l sources because the statistical distribution of the received measurements is insensitive to whether $\boldsymbol{\theta}_k$ or $\boldsymbol{\theta}_l$ was transmitted. The following theorem establishes the identifiability of our problem.

Theorem 1. *Let \mathbf{A} and $\tilde{\mathbf{A}}$ be two arbitrary $p \times m$ and $p \times n$ matrices, respectively. In addition, let $\sigma_n^2 > 0$ and $\tilde{\sigma}_n^2 > 0$. The two pdfs $f(\mathbf{x}|\mathbf{A}, \sigma_n^2)$ and $f(\mathbf{x}|\tilde{\mathbf{A}}, \tilde{\sigma}_n^2)$ are equal if and only if $m = n$, $\sigma_n^2 = \tilde{\sigma}_n^2$, and $\mathbf{A} = \tilde{\mathbf{A}}$ up to a permutation of the rows.*

See Appendix A for the proof.

Theorem 1 points to very important difference between the problem of detecting Gaussian sources and detecting digital sources. For Gaussian sources, the problem is identifiable only when the number of sources is smaller than the number of sensors and the mixing matrix is of full rank. For the case of digital sources, on the other hand, the problem is identifiable for every number of sources and for every mixing matrix \mathbf{A} , whether full rank or not. This fact has a significant importance in communication systems since we can estimate the number of users utilizing a given channel (say using CDMA scheme) with only one antenna. It should be noted that the identifiability was proved under the assumption that all the transmitters use the same constellation. This result can be extended to the case of unequal signal constellations, as long as any linear combination of the received signal is not equal to any other received signal.

2.2. Consistency of the MDL estimator

In the previous subsection it was proven that the estimation problem defined in Section 1.1 is identifiable. Once the estimation problem has been shown to be identifiable, it is possible to infer the number of sources from the measurements. However for a specific estimator, the issue of consistency must be considered. In model order selection problems, the common performance measure is the probability of error, that is, $P_e = P\{\hat{q} \neq q\} = 1 - P_c$. In what follows the MDL estimator is proven to be a consistent estimator, that is $\lim_{N \rightarrow \infty} P_e = 0$.

Theorem 2. *The MDL estimator is a consistent estimator of the number of sources.*

See Appendix B for the proof.

Both the proof of Theorem 2 and its practical implications deserve special attention. The proof of Theorem 2 is divided into two parts. In the first part it is argued that asymptotically the probability of underestimation, that is, $P\{\hat{q} < q\}$, approaches zero as $N \rightarrow \infty$; whereas in the second part it is argued that, asymptotically, the probability of overestimation, that is, $P\{\hat{q} > q\}$, approaches zero as $N \rightarrow \infty$. The first part is easily proven based on a more general theorem which appears in [12]. The second part, however, is proven using Wilks' theorem. The reader might recall that the first published consistency proof of the GMDL was based on Wilks' theorem [3]. Later, it was demonstrated that Wilks' theorem cannot be used in this problem [14], and an alternate proof, which is based on Taylor's expansion, was proposed in [11]. Since then, Taylor's expansion is the only tool used for this type of proofs. In what follows, we briefly explain why Wilks' theorem can be used in our proof.

Wilks' theorem asserts that under certain regularity conditions, the log likelihood ratio is distributed as a chi-square random variable with a certain number of degrees of freedom. The main regularity condition in Wilks' theorem is that the unknown parameters under the null hypothesis are not on the border of the parameter space of the unknown parameters under the alternative hypothesis. Simple example will

clarify this point. Consider a random variable x . Assume that under the null hypothesis $x \sim \mathcal{N}(0, 1)$, and under the alternate hypothesis $x \sim \mathcal{N}(\theta, 1)$, $\theta \in \Theta$. Now, if $\Theta = (-\infty, \infty)$, the null hypothesis is, with some abuse of notation, “inside” the parameter space of the alternate hypothesis. However, if $\Theta = (0, \infty)$, the null hypothesis is, with some abuse of notation, “on the boarder” of the parameter space of the alternate hypothesis (the alternate hypothesis does not “wrap” the null hypothesis as required).

Recall that for Gaussian signals it is required that the matrix $\mathbf{AR}_s\mathbf{A}^H$ be a rank- q matrix. Therefore, assuming that we have q sources, the unknown parameters are $[\lambda_1, \dots, \lambda_q, \mathbf{v}_1, \dots, \mathbf{v}_q, \sigma_n^2]$, where σ_n^2 is the unknown noise level, $\lambda_1 > 0, \dots, \lambda_q > 0$ are the eigenvalues of the matrix $\mathbf{AR}_s\mathbf{A}^H$, and $\mathbf{v}_1, \dots, \mathbf{v}_q$ are the eigenvectors corresponding to these eigenvalues. It is easily proven that the parameter space that corresponds to q sources lies on the border of the parameter space that corresponds to $q + 1$ sources, and not inside it. The intuition for this is very simple. The same way the point 0 is on the boarder of the parameter space $(0, T)$, the points $[\lambda_1, \dots, \lambda_q, 0, \mathbf{v}_1, \dots, \mathbf{v}_{q+1}, \sigma_n^2]$, which compose the parameter space corresponding to q sources, are on the border of the points $[\lambda_1, \dots, \lambda_q, \lambda_{q+1}, \mathbf{v}_1, \dots, \mathbf{v}_{q+1}, \sigma_n^2]$, which compose the parameter space corresponding to $q + 1$ sources. The regularity conditions in Wilks’ theorem require that the parameter space corresponding to q sources be inside the parameter space corresponding to $q + 1$ sources and not on its boarder. Since in the Gaussian case this condition does not hold, Wilks’ theorem cannot be used.

Consider the case of digital signals. From Theorem 1 it is clear that assuming q sources, the parameter space is $[a_{11}, \dots, a_{p1}, a_{21}, \dots, a_{pq}, \sigma_n^2 > 0]$, where $a_{ij} \in \mathcal{C}$ for every $1 \leq i \leq p$, $1 \leq j \leq q$. Now, the same way the point 0 is inside the parameter space $[-T, T]$, the points $[a_{11}, \dots, a_{pq}, 0, \dots, 0, \sigma_n^2 > 0]$, which compose the parameter space corresponding to q sources, are inside the points $[a_{11}, \dots, a_{p(q+1)}, \sigma_n^2 > 0]$, which compose the parameter space corresponding to $q + 1$ sources. Therefore, Wilks’ theorem can be used in our problem for proving the consistency of the MDL estimator.

Theorem 1 demonstrates that one can estimate the number of digital sources even when the number of receivers is one. The importance of Theorem 2 lies in demonstrating that indeed one can estimate the number of sources (transmitters) with diminishing probability of error even with one receiving antenna. This is of special importance in CDMA systems that use blind multiuser detection. In these systems it is assumed that the number of users is smaller than the processing gain. However, the MDL estimator proposed in this paper can be used to estimate any number of users and hence can be used to increase the range of operation of blind multiuser detectors.

3. SIMULATION STUDY

In what follows we compare the performance of the GMDL estimator (5) and the MDL estimator (6), which exploits the special structure of the transmitted signals. We compare the

performance of the two processors when either the number of sensors is larger than the number of sources or the number of sensors is smaller than the number of sources. Note that in the latter, the GMDL cannot be used and the MDL estimator is the only existing option. For simplicity we use BPSK signals; hence $\mathcal{D} = \{\pm 1\}$. In addition, in the MDL simulation, we calculated the MDL metric for $k = 0, \dots, q + 2$, where q is the true number of sources in the scenario. The reason is that the probability to overestimate the number of true sources by more than one is negligible.

It is obvious that the GMDL estimator is much more simple than the MDL estimator. The complexity of the GMDL estimator is $\mathcal{O}(p^3)$, which is the complexity of computing the eigenvalues. However, the complexity of the MDL estimator grows exponentially with p . In order to have a fair compression, assume that $q = p$. In this case the complexity of the MDL estimator is $\mathcal{O}(N\mathcal{D}^p)$. The actual complexity of the MDL estimator is even larger because of the lack of closed form expressions for the unknown parameters. In order to overcome this complexity issue, in the simulations, we used the EM algorithm for computing the MLEs of the unknown parameters [15].

First we consider the performance of the MDL and GMDL estimators in the presence of one source. Consider a uniform linear array (ULA) with three sensors and one transmitter located at the array boresight. In the first simulation we study the performance of the GMDL and MDL estimators as a function of the number of snapshots, whereas in the second simulation we examine the performance of the GMDL and the MDL estimators as a function of the source signal-to-noise ratio (SNR). Figures 1 and 2 depict the probability of correct decision as a function of the number of snapshots and as a function of the source SNR, respectively. For the first simulation the SNR per element was set to -6 dB, and for the second simulation the number of snapshots was set to 60. For each point in the graph, 2000 Monte Carlo runs were made.

In the graphs, we see the clear advantage of the MDL over the GMDL estimator. The GMDL requires about 2 dB additional power or about 3 times the number of snapshots in order to achieve the same performance as the MDL estimator that exploits the special signal structure. The uniform performance improvement of the MDL over the GMDL should not come as a surprise. In [16] it was demonstrated that by using additional a priori information, one can improve the estimator performance considerably.

In the next set of simulations we examine the case of detecting more than one source. Figure 3 depicts the probability of correct detection of two sources as a function of their spatial separation. For the first scenario, one source is located at angle $\theta = 0^\circ$ and the second at angle ρ where ρ is varied between 0° and 30° . We consider two scenarios. In the first, the sources’ SNR was set to -3 dB and the number of snapshots was set to 50, while in the second, the sources’ SNR was set to 0 dB and the number of snapshots was set to 100. Here again the MDL demonstrates uniform performance improvement over the GMDL. More surprisingly, however, is that with a sufficient SNR and or number

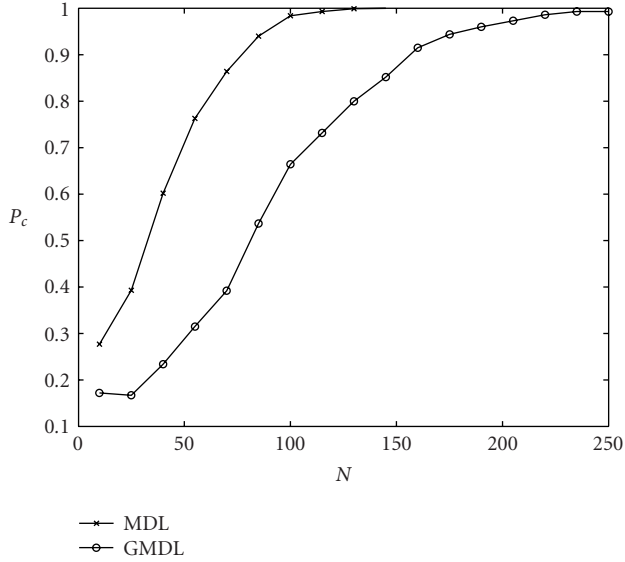


FIGURE 1: The probability of a correct decision of the GMDL and MDL estimators for a single source as a function of the number of snapshots N . ULA with $p = 3$, $\theta = 0^\circ$, and SNR = -6 dB.

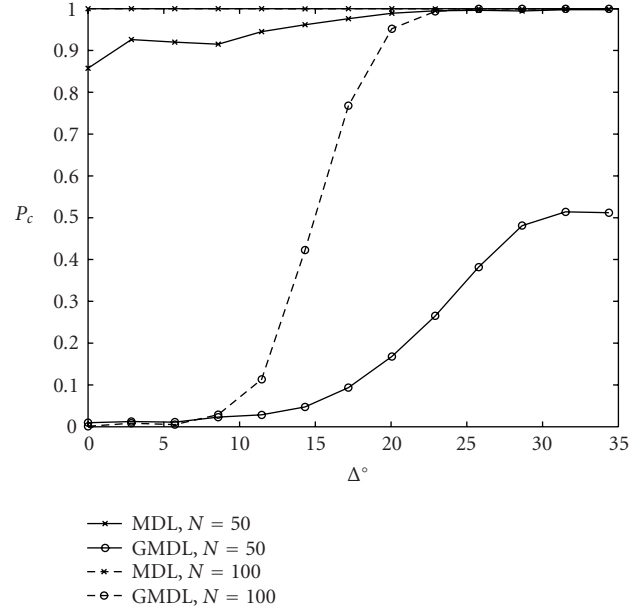


FIGURE 3: The probability of a correct decision of the GMDL and MDL estimators for two equal-power sources as a function of the source separation. ULA with $p = 3$, $\theta_1 = 0^\circ$, $N = 50$, SNR = -3 dB, and $N = 100$, SNR = 0 dB.

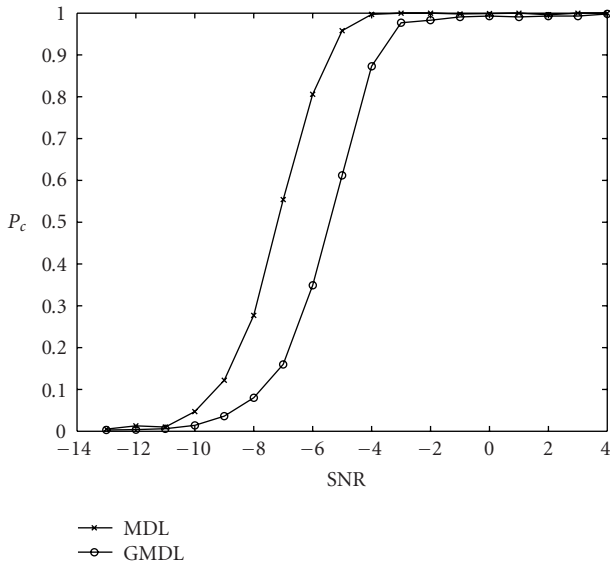


FIGURE 2: The probability of a correct decision of the GMDL and MDL estimators for a single source as a function of the SNR. ULA with $p = 3$, $\theta = 0^\circ$, and $N = 60$.

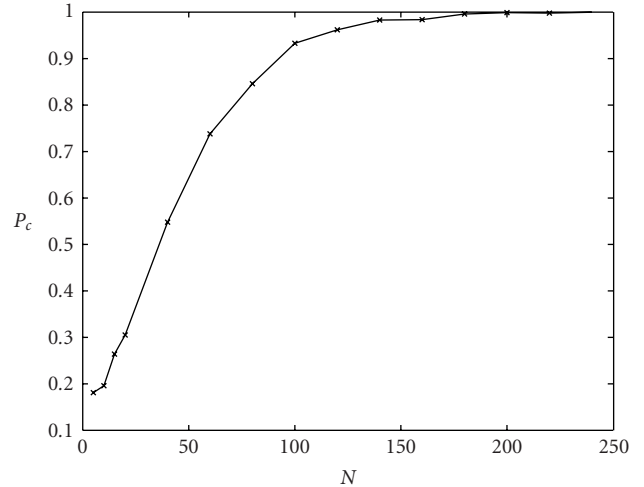


FIGURE 4: The probability of a correct decision of the MDL estimators for two equal-power sources as a function of snapshots, using one receiver only and two equal power transmitters. $\theta_1 = 0$, $\theta_2 = 5^\circ$, SNR = 0 dB.

of snapshots, the MDL can perfectly resolve two sources of no spatial separation. This is in agreement with Theorems 1 and 2 where it was shown that the MDL estimator can resolve any number of sources with only one receiver. The MDL estimator, therefore, uses the special signal structure, and not spatial diversity, to resolve multiple sources. Thus, the MDL estimator can resolve multiple sources even without any spatial separation between them, as Figure 3 demonstrates.

As already mentioned, one of the main advantages of the MDL estimator is its ability to operate even in the presence of more sources than sensors. In the next and final simulations, we explore this feature. We consider one receiving antenna and two equal-power sources, transmitting with 0 -dB SNR. Figure 4 depicts the probability of a correct decision as a function of the number of snapshots for the MDL estimator only (since the GMDL estimator is invalid in this scenario). It is evident from the graph that the MDL estimator

is a consistent estimator. Moreover, only a limited number of snapshots are required for achieving reliable probability of detection.

4. PERFORMANCE ANALYSIS OF THE MDL ESTIMATOR

The probability of error of the MDL estimator is an important figure of merit, and usually extensive simulations are carried out in order to study its performance. It is desired to have a simple tool for predicting the performance of the MDL estimator which can be used not only to study it but also to design a penalty function in (3) that meets some required performance. In what follows we provide such simple tool for predicting the performance of the MDL estimator.

Denote by P_M the probability of miss (the probability of underestimation), that is, $P_M = P(\hat{q} < q)$, and P_{FA} the probability of false alarm (the probability of overestimation), that is, $P_{FA} = P(\hat{q} > q)$. The probability of error, denoted by P_e , is the sum of the probability of miss and the probability of false alarm, that is, $P_e = P_M + P_{FA}$. The probability of miss can be approximated by $P_M \approx P(\text{MDL}(q-1) < \text{MDL}(q))$ while the probability of false alarm can be approximated by $P_{FA} = P(\text{MDL}(q+1) < \text{MDL}(q))$. These approximations are due to the fact that the MDL function, $\text{MDL}(q)$, is a convex function with a single minimum [12].

The following two lemmas establish asymptotic approximations for P_M and P_{FA} , respectively.

Lemma 1. *Asymptotically, for large N ,*

$$P_{FA} \approx 1 - F_{\chi_{2p}^2}(2p \log N), \quad (7)$$

where $F_{\chi_{2p}^2}(\cdot)$ is the cumulative distribution function of the chi-square random variable with $2p$ degrees of freedom.

See (B.2) in the proof of Theorem 2 for the proof.

Lemma 2. *Asymptotically, for large N ,*

$$P_M \approx 1 - Q\left(\frac{-\mu}{\sigma}\right), \quad (8)$$

where $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-a^2/2} da$ is the error function, $\mu = NE(-\log f(\mathbf{x}|\mathbf{A}^*, \sigma_n^{2*}) + \log f(\mathbf{x}|\mathbf{A}, \sigma_n^2) + p \log N)$, and $\sigma^2 = N \text{Var}(-\log f(\mathbf{x}|\mathbf{A}^*, \sigma_n^{2*}) + \log f(\mathbf{x}|\mathbf{A}, \sigma_n^2))$.

$(\mathbf{A}^*, \sigma_n^{2*}) = \arg \min_{[\mathbf{A}, \sigma_n^2] \in \Theta_{q-1}} \{D(-\log f(\mathbf{x}|\mathbf{A}^*, \sigma_n^{2*}) \| \log f(\mathbf{x}|\mathbf{A}, \sigma_n^2))\}$, where $D(f\|g)$ is the divergence between the two distributions.

See [12, Theorem 1] for the proof.

These lemmas provide a simple tool to approximate the performance of the MDL estimator. The approximation improves with the number of snapshots or the SNR.

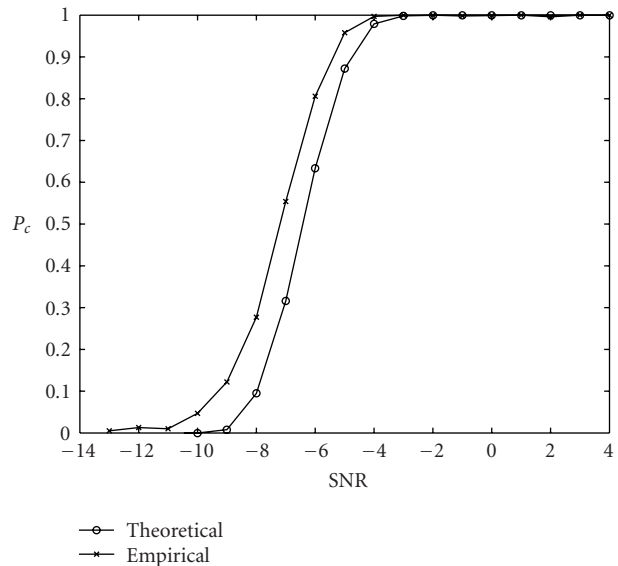


FIGURE 5: The probability of a correct decision of the MDL estimators for a single source as a function the SNR. Empirical versus theoretical results for ULA with $p = 3$, $\theta = 0^\circ$, and $N = 60$.

In order to validate our analysis we consider the following example. Assume a ULA with three sensors and one source that transmits a BPSK signal at the array boresight. Figure 5 depicts the probability of detection $P_c = 1 - P_e$, which has been derived empirically by simulations, and the probability of detection predicted by the two lemmas, as a function of the SNR. The number of snapshots was set to $N = 60$, and 2000 Monte Carlo runs were made. From the figure, one sees that even under nonasymptotic conditions ($N = 60$), our theoretical analysis predicts the probability of detection quite well, as the results of the theoretical analysis differ from the results of the simulation by no more than 1 dB. This demonstrates not only the validity of our analysis, but also its applicability as a synthesis tool.

5. SUMMARY AND CONCLUDING REMARKS

In this paper, we investigated the problem of estimating the number of communication sources impinging an array of sensors. We proved that the resolution capacity is not limited by the number of sensors and that one can estimate any number of sources. This is in contrast to the usual paradigm, which assumes that the number of resolvable sources is smaller than the number of sensors. We also proved that the MDL estimator is a consistent estimator of the number of sources.

The performance of the MDL estimator is shown to be uniformly superior over the commonly used GMDL estimator. In two important scenarios, namely, of more sources than sensors and of more than one source, the MDL estimator significantly outperforms the GMDL estimator.

When more sources than sensors are to be detected, one cannot use the GMDL estimator and the MDL is the only estimator that can solve the problem. When two or more sources exist, the MDL estimator exploits both the signal structure and the spatial separation between the sources, which results in substantial performance improvement over the GMDL estimator.

APPENDICES

A. PROOF OF THEOREM 1

In this appendix we prove that the problem is identifiable. In particular we prove that there is a one-to-one correspondent between the measurements' statistical distribution and the spatial scenario. We assume, without loss of generality, that BPSK signals are transmitted, that is, we assume that $\mathcal{D} = \pm 1$.

Recall that given the mixing matrix, \mathbf{A} , and the noise level, σ_n^2 , the pdf of the received signal vector is

$$f(\mathbf{x}|\mathbf{A}, \sigma_n^2) = \frac{1}{2^q} \sum_{\mathbf{s} \in \{\pm 1\}^q} \mathcal{C} \mathcal{N}(\mathbf{x} - \mathbf{A}\mathbf{s}, \sigma_n^2 \mathbf{I}), \quad (\text{A.1})$$

where we used the assumption that the transmitted signals are BPSK signals. Let \mathbf{A} be a $p \times q$ complex matrix, and \mathbf{x} a $p \times 1$ complex vector. We next define a simple operation on the matrix \mathbf{A} and the vector \mathbf{x} . From \mathbf{A} we define $\bar{\mathbf{A}}$ as a $2p \times q$ matrix according to

$$\bar{\mathbf{A}} = [\Re[\mathbf{A}]_1: \Im[\mathbf{A}]_1: \dots \Re[\mathbf{A}]_p: \Im[\mathbf{A}]_p], \quad (\text{A.2})$$

and from \mathbf{x} we define a $2p \times 1$ vector denoted by $\bar{\mathbf{x}}$ as follows:

$$\bar{\mathbf{x}} = [\Re[\mathbf{x}]_1, \Im[\mathbf{x}]_1, \dots, \Re[\mathbf{x}]_p, \Im[\mathbf{x}]_p]^T, \quad (\text{A.3})$$

where $[\mathbf{A}]_i$ denotes the i th row of the matrix \mathbf{A} , $[\mathbf{x}]_i$ denotes the i th element of the vector \mathbf{x} , \Re is the real part of the number, and \Im is the imaginary part of the number.

It is easy to verify that

$$\begin{aligned} f(\mathbf{x}|\mathbf{A}, \sigma_n^2) &= \frac{1}{2^q} \sum_{\mathbf{s} \in \{\pm 1\}^q} \mathcal{C} \mathcal{N}(\mathbf{x} - \mathbf{A}\mathbf{s}, \sigma_n^2 \mathbf{I}) \\ &= \frac{1}{2^q} \sum_{\mathbf{s} \in \{\pm 1\}^q} \mathcal{N}(\bar{\mathbf{x}} - \bar{\mathbf{A}}\mathbf{s}, \sigma_n^2 \mathbf{I}) \\ &= \frac{1}{2^q} \sum_{\mathbf{s} \in \{\pm 1\}^q} \prod_{i=1}^{2p} \mathcal{N}([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i:} \mathbf{s}, \sigma_n^2) \\ &\triangleq f(\bar{\mathbf{x}}|\bar{\mathbf{A}}, \sigma_n^2), \end{aligned} \quad (\text{A.4})$$

where $\mathcal{N}(x, \sigma^2) = (1/\sqrt{2\pi\sigma^2})e^{-x^2/2\sigma^2}$. Furthermore, $f(\bar{\mathbf{x}}|\bar{\mathbf{A}}, \sigma_n^2)$ can be written as follows:

$$\begin{aligned} f(\bar{\mathbf{x}}|\bar{\mathbf{A}}, \sigma_n^2) &= \frac{1}{2^q} \sum_{\mathbf{s} \in \{\pm 1\}^q} \prod_{i=1}^{2p} \mathcal{N}([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i:} \mathbf{s}, \sigma_n^2) \\ &= \prod_{i=1}^{2p} \mathcal{N}([\bar{\mathbf{x}}]_i, \sigma_n^2) \\ &\quad * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,1}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,1}) \right) \\ &\quad * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,2}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,2}) \right) \\ &\quad * \dots * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,q}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,q}) \right), \end{aligned} \quad (\text{A.5})$$

where δ is the Dirac delta function, and $*$ denotes the convolution operator. Now, the moment-generating function of \mathbf{x} , given \mathbf{A} and σ_n^2 , or, equivalently, of $\bar{\mathbf{x}}$, given $\bar{\mathbf{A}}$ and σ_n^2 , is

$$\begin{aligned} \phi_{\bar{\mathbf{x}}}(\omega|\bar{\mathbf{A}}, \sigma_n^2) &= E\{e^{-j\omega^T \bar{\mathbf{x}}}\} \\ &= \frac{1}{2^q} \prod_{i=1}^{2p} e^{\sigma_n^2 \|\omega\|^2} \prod_{i=1}^q [e^{-j\omega^T [\bar{\mathbf{A}}]_{:i}} + e^{-j\omega^T [\bar{\mathbf{A}}]_{:i}}], \end{aligned} \quad (\text{A.6})$$

where $[\bar{\mathbf{A}}]_{:i}$ denotes the i th column of the matrix $\bar{\mathbf{A}}$. Note that the first product is the Fourier transform of the normal distribution, while the second product is the Fourier transform of the convolutions of the various delta functions.

We now turn to prove the theorem. It is easily seen that if $m = n$, $\sigma_n^2 = \tilde{\sigma}_n^2$, and $\mathbf{A} = \tilde{\mathbf{A}}$ up to column permutation, $f(\mathbf{x}|\mathbf{A}, \sigma) = f(\mathbf{x}|\tilde{\mathbf{A}}, \tilde{\sigma}_n^2)$. This proves the sufficiency.

We now turn to prove the necessity. That is, if $f(\mathbf{x}|\mathbf{A}, \sigma) = f(\mathbf{x}|\tilde{\mathbf{A}}, \tilde{\sigma}_n^2)$, then $\sigma_n^2 = \tilde{\sigma}_n^2$ and $\mathbf{A} = \tilde{\mathbf{A}}$. We first prove that $\sigma_n^2 = \tilde{\sigma}_n^2$, that is, there are no two different scenarios that result in the same distribution at the output of the array.

Lemma 3. *Let \mathbf{A} and $\tilde{\mathbf{A}}$ be two arbitrary $p \times m$ and $p \times n$ matrices, respectively. In addition, let $\sigma_n^2 > 0$ and $\tilde{\sigma}_n^2 > 0$. If $f(\mathbf{x}|\mathbf{A}, \sigma_n^2) = f(\mathbf{x}|\tilde{\mathbf{A}}, \tilde{\sigma}_n^2)$, then $\sigma_n^2 = \tilde{\sigma}_n^2$.*

Proof. Assume that $f(\mathbf{x}|\mathbf{A}, \sigma_n^2) = f(\mathbf{x}|\tilde{\mathbf{A}}, \tilde{\sigma}_n^2)$. Consequently,

$$\begin{aligned} \phi_{\bar{\mathbf{x}}}(\omega|\bar{\mathbf{A}}, \sigma_n^2) &= \frac{1}{2^m} \prod_{i=1}^{2p} e^{\sigma_n^2 \|\omega\|^2} \prod_{i=1}^m [e^{-j\omega [\bar{\mathbf{A}}]_{:i}} + e^{-j\omega [\bar{\mathbf{A}}]_{:i}}] \\ &= \frac{1}{2^n} \prod_{i=1}^{2p} e^{\tilde{\sigma}_n^2 \|\omega\|^2} \prod_{i=1}^n [e^{-j\omega [\bar{\mathbf{A}}]_{:i}} + e^{-j\omega [\bar{\mathbf{A}}]_{:i}}] \\ &= \phi_{\bar{\mathbf{x}}}(\omega|\tilde{\bar{\mathbf{A}}}, \tilde{\sigma}_n^2), \end{aligned} \quad (\text{A.7})$$

where $\tilde{\bar{\mathbf{A}}}$ is the result of (A.2) applied to $\tilde{\mathbf{A}}$.

Assume that $\sigma_n^2 \neq \bar{\sigma}_n^2$, and assume, without loss of generality, that $\sigma_n^2 > \bar{\sigma}_n^2$. Dividing both sides of (A.7) by $\prod_{i=1}^{2p} e^{\bar{\sigma}_n^2 \|\omega\|^2}$ results in the following equality:

$$\begin{aligned} & \frac{1}{2^m} \prod_{i=1}^{2p} e^{(\sigma_n^2 - \bar{\sigma}_n^2) \|\omega\|^2} \prod_{i=1}^m [e^{-j\omega[\bar{\mathbf{A}}]_{i,i}} + e^{-j\omega[\bar{\mathbf{A}}]_{i,i}}] \\ &= \frac{1}{2^n} \prod_{i=1}^n [e^{-j\omega[\bar{\mathbf{A}}]_{i,i}} + e^{-j\omega[\bar{\mathbf{A}}]_{i,i}}]. \end{aligned} \quad (\text{A.8})$$

Taking the inverse Fourier transform, the left- and right-hand sides of (A.8) result in

$$\begin{aligned} & \prod_{i=1}^{2p} \mathcal{N}([\bar{\mathbf{x}}]_i, \sigma_n^2 - \bar{\sigma}_n^2) \\ & * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,1}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,1}) \right) \\ & * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,2}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,2}) \right) \\ & * \dots * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,m}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,m}) \right) \\ &= \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,1}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,1}) \right) \\ & * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,2}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,2}) \right) \\ & * \dots * \frac{1}{2} \left(\prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i - [\bar{\mathbf{A}}]_{i,n}) + \prod_{i=1}^{2p} \delta([\bar{\mathbf{x}}]_i + [\bar{\mathbf{A}}]_{i,n}) \right). \end{aligned} \quad (\text{A.9})$$

It is now obvious that $\sigma_n^2 = \bar{\sigma}_n^2$ is a necessary condition for the equivalence between the two sides of (A.9). \square

We now proceed to prove that $m = n$ is also a necessary condition for $f(\mathbf{x}|\mathbf{A}, \sigma_n^2) = f(\mathbf{x}|\bar{\mathbf{A}}, \bar{\sigma}_n^2)$. Since $\sigma_n^2 = \bar{\sigma}_n^2$, (A.9) becomes

$$\frac{1}{2^m} \sum_{\mathbf{s} \in \{\pm 1\}^m} \delta(\bar{\mathbf{x}} - \bar{\mathbf{A}}\mathbf{s}) = \frac{1}{2^n} \sum_{\mathbf{s} \in \{\pm 1\}^n} \delta(\bar{\mathbf{x}} - \bar{\mathbf{A}}\mathbf{s}). \quad (\text{A.10})$$

Consider the left-hand side of (A.10) and assume, without loss of generality, that the first row of $\bar{\mathbf{A}}$ does not contain any zeros. In this case the maximum value of $|\bar{\mathbf{x}}|_1$ is $\sum_{j=1}^q |\bar{\mathbf{A}}_{1,j}| \triangleq \alpha$. Consequently, the coefficient of product that includes $\delta([\bar{\mathbf{x}}]_1 - \alpha)$ is $1/2^m$. Using the same reasoning, we can demonstrate that there at least the coefficient of one product on the right-hand side of (A.10) is $1/2^n$. Therefore, it is necessary that $n = m$ for the equality to hold. If every row of $\bar{\mathbf{A}}$ contains at least one zero, a somewhat more complicated approach has to be taken in order to prove that $m = n$ [15].

The uniqueness of $\bar{\mathbf{A}}$ is now easily proven from (A.10) by noting that the delta functions having the coefficients have to be on the same point in \mathbf{R}^{2p} .

B. PROOF OF THEOREM 2

In this appendix the consistency of the MDL estimator is proven. Specifically it is shown that the probability of error of the MDL estimator converges to zero as the number of snapshots increases to infinity. An error event will occur if and only if there exists $k \neq q$ such that $\text{MDL}(q) - \text{MDL}(k) > 0$. Thus in order to prove the lemma, it suffice to prove that, for every $k \neq q$, $P(\text{MDL}(q) - \text{MDL}(k) > 0) \rightarrow 0$.

Assume that $k < q$. It was previously shown that under very weak conditions, the probability of miss of every MDL estimator converges to zero as $N \rightarrow \infty$ [12]. In particular the probability of miss of the MDL estimator considered in this paper, which satisfies the condition stated in [12], converges to zero as $N \rightarrow \infty$.

Now, assume that $k > q$. $\text{MDL}(q) - \text{MDL}(k)$ can be written as follows:

$$\begin{aligned} & \text{MDL}(q) - \text{MDL}(k) \\ &= -\log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})) + 0.5(2pq + 1) \log N \\ & \quad + \log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(k)})) - 0.5(2pk + 1) \log N \quad (\text{B.1}) \\ &= -\log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})) + \log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(k)})) \\ & \quad + 0.5(2p(q - k)) \log N. \end{aligned}$$

In what follows we are going to use Wilks' theorem to compute the asymptotic distribution of $\text{MDL}(q) - \text{MDL}(k)$.

Consider the difference between the log likelihoods, given q and $k > q$ sources, that is, $-\log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})) + 2 \log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(k)}))$. According to Wilks' theorem, under some regularity conditions, asymptotically $-\log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})) + \log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(k)}))$ is distributed as a chi-square random variable with $2p(k - q)$ degrees of freedom. The regularity conditions are the ones that insure that the MLE is unique, efficient, and asymptotically normal. In [17], it is shown that these regularity conditions hold in our problem.

Using the asymptotic distribution of $-\log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})) + \log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(k)}))$, we can compute the asymptotic probability of the event $\{\text{MDL}(q) - \text{MDL}(k) > 0\}$:

$$\begin{aligned} & P(\text{MDL}(q) - \text{MDL}(k) > 0) \\ &= P(-\log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(q)})) \\ & \quad + \log(f_{\mathbf{X}}(\mathbf{X}|\hat{\boldsymbol{\theta}}^{(k)})) > 0.5(2p(q - k)) \log N) \\ & \xrightarrow{N \rightarrow \infty} P(\chi_{2p(k-q)}^2 > (2p(q - k)) \log N) \xrightarrow{N \rightarrow \infty} 0. \end{aligned} \quad (\text{B.2})$$

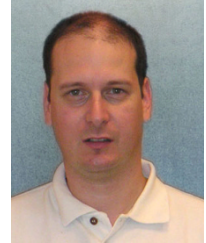
We have proven that for $k \neq q$, the probability of the event $P(\text{MDL}(q) - \text{MDL}(k) > 0) \xrightarrow{N \rightarrow \infty} 0$. Therefore, the probability of error approaches zero as well.

REFERENCES

- [1] C.-M. Cho and P. M. Djuric, "Detection and estimation of DOA's of signals via Bayesian predictive densities," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3051–3060, 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, no. 8, pp. 2242–2247, 2000.
- [3] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 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, no. 11, pp. 1959–1971, 1990.
- [5] H.-T. Wu, J.-F. Yang, and F.-K. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Trans. Signal Processing*, vol. 43, no. 6, pp. 1325–1333, 1995.
- [6] A. M. Zoubir, "Bootstrap methods for model selection," *AEÜ-International Journal of Electronics and Communications*, vol. 53, no. 6, pp. 386–392, 1999.
- [7] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Annals of Statistics*, vol. 11, no. 2, pp. 416–431, 1983.
- [8] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1081–1089, 1986.
- [9] A. M. Bruckstein, T.-J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1357–1367, 1985.
- [10] S. Verdú, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.
- [11] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in the presence of white noise," *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 1–25, 1986.
- [12] E. Fishler, M. Grosman, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1027–1036, 2002.
- [13] J. Rissanen, "Universal coding, information prediction, and estimation," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 629–636, 1984.
- [14] L.-C. Zhao, P. R. Krishnaiah, and Z.-D. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 35, no. 2, pp. 129–132, 1987.
- [15] M. Grosman, "Constellation based detection of the number of sources," M.S. thesis, Tel Aviv University, Tel Aviv, Israel, February 2001.
- [16] E. Fishler and H. Messer, "On the effect of a-priori information on performance of the MDL estimator," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '02)*, vol. 3, pp. 2981–2984, Orlando, Fla, USA, May 2002.
- [17] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '97)*, vol. 5, pp. 3617–3620, Munich, Germany, April 1997.

Eran Fishler was born in Tel Aviv, Israel, in 1972. He received his B.S. degree in mathematics, and his M.S. and Ph.D. degrees in electrical engineering in 1995, 1997, and 2001, respectively, all from Tel Aviv University. From 1993 to 1995, he was with the Israeli Navy as a Research Engineer. From 1996 to 2000, he was also with Tadiran Electronic System, Holon, Israel. During 2002 and parts of 2003, he was a Research Fellow at Princeton University, where he worked on various topics related to UWB communication systems. Currently he studies at the Stern School of Business, New York University. His research interests are statistical signal processing and communication theory.

Michael Grosman was born on December 31, 1971. He received the B.S. degree (summa com laude) and M.S. degree (cum laude) in electrical engineering from Tel Aviv university in 1995 and 2000, respectively. He worked for MOD Israel between 1994 and 2000 in the field of data communication and digital signal processing. Since 2000, he has been working at AudioCodes.



Hagit Messer received the Ph.D. degree in electrical engineering from Tel Aviv University, Israel, and after a one-year postdoctoral appointment at Yale University, she joined the Faculty of Engineering, Tel Aviv University, in 1986, where she is currently a Professor of electrical engineering. From February 1, 2000 to March 31, 2003, she was the Chief Scientist and the Head of the Directorate of Science at the Ministry of Science, Israel. Professor Hagit Messer-Yaron is an expert in statistical signal processing and applications to source localization, communication, and radar systems. She has published about 60 refereed journal papers, over 100 papers in conference proceedings, and she has supervised over 50 graduate students, all in the field of signal processing and statistical signal analysis. Professor Messer-Yaron is a Fellow of the IEEE. She is a member of technical committees of the Signal Processing Society since 1993. She has been an Associate Editor of the IEEE Transactions on Signal Processing and of the IEEE Signal Processing Letters for parameter estimation, detection, and performance analysis. She held international visiting positions and has been involved in the organization of international conferences and workshops.

