

MMSE-Based MDL Method for Accurate Source Number Estimation

Lei Huang, *Member, IEEE*, Teng Long, Erke Mao, and H. C. So, *Senior Member, IEEE*

Abstract—In civilian communication systems, the signature sequence of the desired signal in training phase is known to the receiver. In this letter, using the mutual information, we bridge the probability density function and minimum mean-square error (MMSE) between the observed data and training sequence of the desired signal, and then employ the MMSE to construct a minimum description length (MDL) criterion for accurate source enumeration. Numerical results demonstrate that the proposed method is superior to existing MDL methods in terms of detection performance particularly for small number of snapshots and/or source angular separation.

Index Terms—Eigenvalue decomposition, minimum description length, sensor array processing, source number estimation.

I. INTRODUCTION

ARRAY processing has been widely used in civilian communications. One of the features of these friendly communications systems is that *a priori* knowledge, such as the signature sequence of the desired signal in training phase, is well known to the receivers. To localize the sources by using a super-resolution method and then suppress the interference sources and noise, it is crucial to determine the number of sources first. In the literature, source enumeration (SE) and direction finding (DF) have been extensively studied, such as in [1]–[8]. To enhance the DF accuracy, a number of methods have been suggested by using *a priori* knowledge of the signals. Nevertheless, works on improving the SE accuracy by exploiting *a priori* knowledge of the signals are relatively few.

In this letter, an accurate minimum description length (MDL) method is devised. By exploiting the mutual information, we bridge the probability density function (PDF) and the minimum mean-square error (MMSE) between the observed data and training sequence of the desired signal, and then use the MMSE to construct an MMSE-based MDL (MMDL) method for source number estimation. Since the training sequence of the desired signal is used to calculate the MMSEs, the proposed method is significantly superior to the classical MDL method in detection accuracy particularly when the number of snapshots is small and/or the signal-to-noise ratio (SNR) is low.

Manuscript received March 09, 2009; revised May 19, 2009. First published June 10, 2009; current version published July 10, 2009. This work was supported by the Natural Science Foundation of China (NSFC) under Grant 60702068. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Amir Asif.

L. Huang, T. Long, and E. Mao are with the Department of Electronic Engineering, Beijing Institute of Technology, Beijing 100081, China (e-mail: lhuang8sasp@hotmail.com).

H. C. So is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR, China (e-mail: hcso@ee.cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2009.2024785

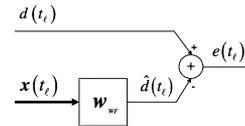


Fig. 1. Classical Wiener filter.

II. PROBLEM FORMULATION

Consider a uniform linear array (ULA) of M sensors receiving p ($p < M$) narrow-band sources $s_1(t_\ell), \dots, s_p(t_\ell)$ from distinct directions $\theta_1, \dots, \theta_p$, respectively. Assume that the sources and array are in the same plane. In the sequel, the ℓ th snapshot vector of the array output is written as

$$\mathbf{x}(t_\ell) = [x_1(t_\ell), \dots, x_M(t_\ell)]^T = \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(t_\ell) + \mathbf{n}(t_\ell) \quad (1)$$

where $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_p)]$ contains the steering vectors, $\mathbf{s}(t_\ell) = [s_1(t_\ell), \dots, s_p(t_\ell)]^T$ is the source vector, $\mathbf{n}(t_\ell) = [n_1(t_\ell), \dots, n_M(t_\ell)]^T$ is the noise vector, and $(\cdot)^T$ is the transpose operation and N denotes the number of snapshots. The steering vector is given by

$$\mathbf{a}(\theta_i) = \left[1, e^{j2\pi d \sin(\theta_i)/\lambda}, \dots, e^{j2\pi d(M-1) \sin(\theta_i)/\lambda} \right]^T \quad (2)$$

where λ denotes the wavelength, d is the inter-sensor spacing, and $i = 1, \dots, p$. The source waveforms $s_1(t_\ell), \dots, s_p(t_\ell)$ are assumed to be jointly stationary, statistically uncorrelated and zero-mean complex Gaussian random processes. The sensor noise $\mathbf{n}(t_\ell)$ is assumed to be an ergodic, zero-mean, spatially and temporally white complex Gaussian process with covariance matrix $\sigma_n^2 \mathbf{I}_M$ where \mathbf{I}_M is the $M \times M$ identity matrix. In addition, the noise is assumed to be uncorrelated with the sources. Given the received data of the array and training sequence of the desired signal, the objective of this work is to accurately estimate p .

When $\mathbf{x}(t_\ell)$ and the training sequence of the desired signal, denoted by $d(t_\ell)$, are available, we propose to use the classical Wiener filter (WF) to obtain the optimal estimate of the desired signal in MMSE sense. The Wiener filtering problem is depicted in Fig. 1 where \mathbf{w}_{WF} is the M -dimensional WF, $\hat{d}(t_\ell)$ is the scalar estimate of $d(t_\ell)$ and $e(t_\ell)$ is the scalar error signal. The classical Wiener filtering problem is to minimize the mean-square error between the reference signal $d(t_\ell)$ and its estimate

$$\hat{d}(t_\ell) = \mathbf{w}_{WF}^H \mathbf{x}(t_\ell) \quad (3)$$

where $(\cdot)^H$ is the Hermitian transpose and \mathbf{w}_{WF} has the form:

$$\mathbf{w}_{WF} = \mathbf{R}_x^{-1} \mathbf{r}_{xd} \quad (4)$$

with $\mathbf{R}_x = E[\mathbf{x}(t_\ell) \mathbf{x}^H(t_\ell)]$ and $\mathbf{r}_{xd} = E[\mathbf{x}(t_\ell) d^*(t_\ell)]$. Here $E[\cdot]$ and $(\cdot)^*$ denote the expectation operator and complex conjugate, respectively. Consequently, the error signal is given by

$$e(t_\ell) = d(t_\ell) - \hat{d}(t_\ell) = d(t_\ell) - \mathbf{w}_{WF}^H \mathbf{x}(t_\ell). \quad (5)$$

Denote the eigenvalue decomposition (EVD) of \mathbf{R}_x as

$$\mathbf{R}_x = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \quad (6)$$

where $\mathbf{\Lambda} = \text{diag}([\lambda_1, \dots, \lambda_M])$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_M]$, $\lambda_1 \geq \dots \geq \lambda_{p+1} = \dots = \lambda_M$ are the eigenvalues and \mathbf{u}_i ($i = 1, \dots, M$) are the corresponding eigenvectors. Without loss of generality, suppose that the first signal is the desired signal and we thereby obtain $d(t_\ell) = s_1(t_\ell)$. Using (4)–(6), the MMSE, denoted by ξ , is

$$\xi \triangleq E[|e(t_\ell)|^2] = \sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^H \mathbf{r}_{xd} \quad (7)$$

where $\sigma_d^2 = E[s_1(t_\ell)s_1^*(t_\ell)]$ is the power of $s_1(t_\ell)$ and

$$\mathbf{r}_{xd} = E[\mathbf{x}(t_\ell)s_1^*(t_\ell)] = \sigma_{s_1}^2 \mathbf{a}(\theta_1). \quad (8)$$

Substituting (8) into (7) and noticing that $|\mathbf{r}_{xd}^H \mathbf{u}_i|^2 = 0$ ($i = p+1, \dots, M$) yields

$$\xi^{(1)} > \xi^{(2)} > \dots > \xi^{(p+1)} = \dots = \xi^{(M)} \quad (9)$$

where

$$\xi^{(k)} \triangleq \sigma_d^2 - \sum_{i=1}^k \frac{|\mathbf{r}_{xd}^H \mathbf{u}_i|^2}{\lambda_i} \quad (k = 1, \dots, M) \quad (10)$$

is the MMSE when the number of sources is assumed to be k . In practical applications, however, only finite sample size, say N , can be accessed. Consequently, we can only obtain the maximum likelihood (ML) estimates of $\{\lambda_i, \mathbf{u}_i\}_{i=1}^M$, denoted by $\{\hat{\lambda}_i, \hat{\mathbf{u}}_i\}_{i=1}^M$, which are the eigen-pairs associated with the sample covariance matrix $\hat{\mathbf{R}}_x = (1/N) \sum_{\ell=1}^N \mathbf{x}(t_\ell) \mathbf{x}^H(t_\ell)$. In the sequel, the ML estimate of $\xi^{(k)}$ is calculated as $\hat{\xi}^{(k)} = \hat{\sigma}_d^2 - \sum_{i=1}^k |\hat{\mathbf{r}}_{xd}^H \hat{\mathbf{u}}_i|^2 / \hat{\lambda}_i$ where $\hat{\sigma}_d^2 = (1/N) \sum_{\ell=1}^N |d(t_\ell)|^2$ and $\hat{\mathbf{r}}_{xd} = (1/N) \sum_{\ell=1}^N \mathbf{x}(t_\ell) d^*(t_\ell)$. The problem here is to use the MMSEs to accurately determine the number of incident sources, including the desired and interference sources.

III. PROPOSED MMSE-BASED MDL ESTIMATOR

For a complex Gaussian process $\mathbf{X} = \{\mathbf{x}(t_\ell)\}_{\ell=1}^N$, its PDF is

$$f(\mathbf{X} | \Theta_x^{(p)}) = \prod_{\ell=1}^N \frac{1}{\pi^M \det(\mathbf{R}_x)} \exp\{-\mathbf{x}^H(t_\ell) \mathbf{R}_x \mathbf{x}(t_\ell)\} \quad (11)$$

where $\Theta_x^{(p)} = [\lambda_1, \dots, \lambda_p, \sigma_n^2, \mathbf{u}_1^T, \dots, \mathbf{u}_p^T]^T$ is the parameter vector of the model. In the sequel, the mutual information between $d(t_\ell)$ and $\mathbf{x}(t_\ell)$ is expressed as

$$I(d, \mathbf{x}) = H(d) - H(d|\mathbf{x}) = H(d) + H(\mathbf{x}) - H(d, \mathbf{x}) \quad (12)$$

where $H(z) \triangleq -\int f(z) \ln f(z) dz$ is the entropy of $z(t_\ell)$. For the complex Gaussian processes $d(t_\ell) = s_1(t_\ell)$ and $\mathbf{x}(t_\ell)$, their corresponding entropies are, respectively, given by

$$H(d) = \ln(\pi e \sigma_d^2) \quad (13)$$

$$H(\mathbf{x}) = \ln((\pi e)^M \det(\mathbf{R}_x)). \quad (14)$$

Notice that the covariance matrix of $\mathbf{y}(t_\ell) \triangleq [d(t_\ell), \mathbf{x}^T(t_\ell)]^T$ can be written as

$$\mathbf{R}_y \triangleq E[\mathbf{y}(t_\ell) \mathbf{y}^H(t_\ell)] = \begin{bmatrix} \sigma_d^2 & \mathbf{r}_{xd}^H \\ \mathbf{r}_{xd} & \mathbf{R}_x \end{bmatrix}. \quad (15)$$

The joint entropy of $d(t_\ell)$ and $\mathbf{x}(t_\ell)$ is calculated as

$$\begin{aligned} H(d, \mathbf{x}) &= \ln((\pi e)^{M+1} \det(\mathbf{R}_y)) \\ &= \ln(\pi e) + \ln((\pi e)^M \times \det(\mathbf{R}_x) \\ &\quad \times \det(\sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{R}_x^{-1} \mathbf{r}_{xd})) \\ &= \ln(\pi e) + \ln((\pi e)^M \det(\mathbf{R}_x)) \\ &\quad + \ln(\sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^H \mathbf{r}_{xd}). \end{aligned} \quad (16)$$

Substituting (13), (14) and (16) into (12) yields

$$\begin{aligned} I(d, \mathbf{x}) &= \ln(\sigma_d^2) - \ln(\sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^H \mathbf{r}_{xd}) \\ &= \ln(\sigma_d^2) - \ln\left(\sigma_d^2 - \sum_{i=1}^M \frac{|\mathbf{r}_{xd}^H \mathbf{u}_i|^2}{\lambda_i}\right). \end{aligned} \quad (17)$$

Let k be the *supposed* source number. Substituting (10) into (17), and considering that $|\mathbf{r}_{xd}^H \mathbf{u}_i| = 0$ ($i = k+1, \dots, M$), we obtain

$$I^{(k)}(d, \mathbf{x}) = \ln(\sigma_d^2) - \ln(\xi^{(k)}). \quad (18)$$

From [3], the Kullback–Leibler (K–L) divergence between two PDFs can be defined as

$$D(f_1 \| f_2) = \int_z f_1(z) \ln\left(\frac{f_1(z)}{f_2(z)}\right) dz. \quad (19)$$

Meanwhile, it is indicated in [9] that the mutual information and the K–L divergence have the following equality

$$\begin{aligned} I^{(k)}(d, \mathbf{x}) &= D^{(k)}\left(f(\mathbf{y}) \parallel \prod_{i=1}^{M+1} f(y_i)\right) \\ &= \sum_{i=1}^{M+1} H(y_i) - H^{(k)}(\mathbf{y}) \\ &= \sum_{i=1}^{M+1} \ln(\pi e \sigma_{y_i}^2) - H^{(k)}(\mathbf{y}) \end{aligned} \quad (20)$$

where $y_i = y_i(t_\ell)$ and $\sigma_{y_i}^2 = E[|y_i(t_\ell)|^2]$. It follows from (18) and (20) that

$$H^{(k)}(\mathbf{y}) = \ln(\xi^{(k)}) + \varrho \quad (21)$$

with $\varrho \triangleq \sum_{i=2}^{M+1} \ln(\pi e \sigma_{y_i}^2) + \ln(\pi e)$. On the other hand, as $\mathbf{y}(t_1), \dots, \mathbf{y}(t_N)$ can be considered to be statistically independent complex Gaussian random vectors with mean zero and covariance matrix \mathbf{R}_y , we obtain for $N \rightarrow \infty$

$$\begin{aligned} H^{(k)}(\mathbf{y}) &= -\frac{1}{N} \sum_{\ell=1}^N \ln\left(f(\mathbf{y}(t_\ell) | \Theta_y^{(k)})\right) \\ &= -\frac{1}{N} \ln\left(f(\mathbf{Y} | \Theta_y^{(k)})\right) \end{aligned} \quad (22)$$

where $\mathbf{Y} = \{\mathbf{y}(t_\ell)\}_{\ell=1}^N$, $\Theta_y^{(k)} = [\tilde{\lambda}_1, \dots, \tilde{\lambda}_k, \sigma_n^2, \tilde{\mathbf{u}}_1^T, \dots, \tilde{\mathbf{u}}_k^T]^T$ and $\{\tilde{\lambda}_i, \tilde{\mathbf{u}}_i\}$ are the eigen-pairs of \mathbf{R}_y . For the finite sample size N , substituting the ML estimates of $\xi^{(k)}$, $\Theta_y^{(k)}$ and $\sigma_{y_i}^2$, i.e., $\hat{\xi}^{(k)}$, $\hat{\Theta}_y^{(k)}$ and $\hat{\sigma}_{y_i}^2$, into (21) and (22), it follows that

$$-\ln(f(\mathbf{Y} | \hat{\Theta}_y^{(k)})) = N \ln(\hat{\xi}^{(k)}) + N \hat{\varrho}. \quad (23)$$

Here $\hat{\rho} = \sum_{i=2}^{M+1} \ln(\pi e \hat{\sigma}_y^2) + \ln(\pi e)$. As a result, substituting this result into the MDL for encoding $\mathbf{y}(t_\ell)$, given by

$$\mathcal{L}\{\mathbf{y}(t_\ell)\} = -\ln f(\mathbf{Y}|\hat{\Theta}_y^{(k)}) + \frac{1}{2}K \ln N \quad (24)$$

where \mathcal{L} denotes the MDL and K equals the number of free adjusted parameters in $\hat{\Theta}_y^{(k)}$, and omitting the terms independent of k , we can rewrite the MDL as

$$\mathcal{L}\{\mathbf{y}(t_\ell)\} = N \ln(\hat{\xi}^{(k)}) + \frac{1}{2}K \ln N \quad (25)$$

It follows from (10) that the calculation of $\xi^{(k)}$ involves the projection of the k -dimensional signal subspace $\mathbf{r}_{xd}^H \mathbf{u}_i$ ($i = 1, \dots, k$) and k largest eigenvalues λ_i ($i = 1, \dots, k$). In the sequel, the number of free parameters in $\hat{\Theta}_y^{(k)}$ may be reduced when (25) is used to calculate the MDL. Note that $\mathbf{r}_{xd}^H \mathbf{u}_i$ can be regarded as the cross-correlation between $d(t_\ell)$ and the signal subspace component $x_s^{(i)}(t_\ell) \triangleq \mathbf{u}_i^H \mathbf{x}(t_\ell)$, i.e., $\mathbf{r}_{xd}^H \mathbf{u}_i = E[d(t_\ell)x_s^{(i)*}(t_\ell)]$. As a consequence, the model of the signal subspace components is given as $\mathbf{x}_s^{(k)}(t_\ell) \triangleq [\mathbf{u}_1^H \mathbf{x}(t_\ell), \dots, \mathbf{u}_k^H \mathbf{x}(t_\ell)]^T$, which can be assumed to be a statistically independent complex Gaussian random vector with mean zero and covariance matrix $\mathbf{R}_s^{(k)} = E[\mathbf{x}_s^{(k)}(t_\ell)\mathbf{x}_s^{(k),H}(t_\ell)]$. Since the family of models is necessarily described by $\mathbf{R}_s^{(k)}$, the number of free parameters within this model equals that in $\mathbf{R}_s^{(k)}$, which is k^2 . Meanwhile, considering that the number of the largest eigenvalues is k and $d(t_\ell)$ is independent of k , the total number of free parameters is calculated as $K = k^2 + k$. Therefore, the number of sources obtained by the MMSE-based MDL (MMDL) criterion, denoted by \hat{p}_{MMDL} , is

$$\hat{p}_{\text{MMDL}} = \arg \min_{k=1, \dots, M-1} \text{MMDL}(k) \quad (26)$$

where

$$\text{MMDL}(k) = N \ln(\hat{\xi}^{(k)}) + \frac{1}{2}(k^2 + k) \ln N. \quad (27)$$

Remark A: As indicated in (10), when there is no incident signal, all the $\xi^{(k)}$ equal σ_d^2 ; otherwise, all the $\xi^{(k)}$ are smaller than σ_d^2 . In the sequel, we use the $\xi^{(k)}$ to determine whether the sources exist or not. To this end, let $\eta \triangleq |(1/M) \sum_{m=1}^M \hat{\xi}^{(m)} - \hat{\sigma}_d^2| = (1/M) \sum_{m=1}^M \sum_{i=1}^m |\hat{\mathbf{r}}_{xd}^H \hat{\mathbf{u}}_i|^2 / \hat{\lambda}_i$. It follows from the asymptotic results of ML estimation [1], that $\hat{\mathbf{r}}_{xd} = \mathbf{r}_{xd} + O(\sqrt{\ln \ln N/N})$, $\hat{\mathbf{u}}_i = \mathbf{u}_i + O(\sqrt{\ln \ln N/N})$ and $\hat{\lambda}_i = \lambda_i + O(\sqrt{\ln \ln N/N})$. In the sequel, it is easy to verify that

$$\eta = \begin{cases} O\left(\frac{\ln \ln N}{N}\right) & p = 0; \\ \left(\frac{1}{p}\right) \sum_{m=1}^p \sum_{i=1}^m \frac{|\mathbf{r}_{xd}^H \mathbf{u}_i|^2}{\lambda_i} + O\left(\sqrt{\frac{\ln \ln N}{N}}\right) & p > 0. \end{cases} \quad (28)$$

Here we have used the fact that $\mathbf{r}_{xd}^H \mathbf{u}_i = 0$ ($i = p + 1, \dots, M$). Noticing that $(1/p) \sum_{m=1}^p \sum_{i=1}^m |\mathbf{r}_{xd}^H \mathbf{u}_i|^2 / \lambda_i \gg \sqrt{\ln \ln N/N} > \ln \ln N/N$, we define an adaptive detector for determining whether there exist the incident sources or not:

$$\eta \stackrel{\mathcal{H}_0}{\leq} \gamma \quad (29)$$

where $\gamma = C\sqrt{\ln \ln N/N}$, $C \in [1, 10)$ is a constant number, and \mathcal{H}_0 and \mathcal{H}_1 denote the cases of source absence and source presence, respectively. If \mathcal{H}_0 is accepted, we conclude that there is no incident source; otherwise \mathcal{H}_1 is accepted, we then proceed

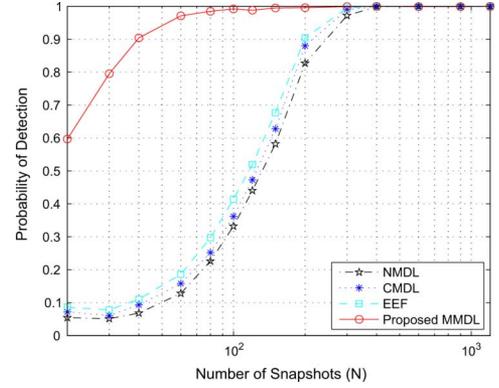


Fig. 2. Probability of correct detection versus number of snapshots. $[\theta_1, \theta_2] = [0^\circ, 5.5^\circ]$, SNR = 0 dB, $M = 6$ and $C = 2$; 500 trials.

to apply the proposed estimator in (26) to accurately obtain the number of sources.

Remark B: Following the results of Wu *et al.* [2], a nonuniform noise MDL (NMDL) method [4] has been proposed to estimate the number of sources. The MDL of the NMDL method is given as

$$\mathcal{L}\{\mathbf{x}(t_\ell)\} = N \ln \left(r_{MM} - \sum_{i=1}^k \frac{|c_i|^2}{\lambda_i} \right) + \frac{1}{2}(k^2 + k) \ln N \quad (30)$$

where $r_{MM} \triangleq E[|x_M(t_\ell)|^2]$, $c_i \triangleq |\mathbf{r}_{xd}^H \mathbf{u}_i|$ ($i = 1, \dots, k$) are the so-called Gerschgorin radii which satisfy $c_1 \geq \dots \geq c_k \geq c_{k+1} = \dots = c_{M-1} = 0$, and the eigenvalues are arranged in a decreasing order as $\lambda_1 \geq \dots \geq \lambda_k \geq \lambda_{k+1} = \dots = \lambda_{M-1}$. It is worth stressing that the MMDL method does not use the ordered $|\mathbf{r}_{xd}^H \mathbf{u}_i|$ ($i = 1, \dots, k$) in the MMSE calculation which is the essential difference between the proposed method and the NMDL method although they take similar formulations. Actually, the nonnegative term $|c_k|^2 / \lambda_k$ is the well-known cross-spectral (CS) energy of the cross-spectral metric (CSM) [10] for adaptive reduced-rank filtering. If the CS energy terms are arranged in a decreasing order and then used to calculate the MMSE, one may obtain the faster convergence of the MMSE. This is why the CS energy can be used in CSM to obtain the reduced rank of the observed space in adaptive reduced-rank filtering. However, when the CS energy is employed to determine the number of sources, i.e., the rank of the principal eigen-subspace, the CS energy terms should be calculated by directly using the eigen-pairs $\{\lambda_i, \mathbf{u}_i\}$ instead of the ordered Gerschgorin radii so that the CS energy terms are not in a decreasing order. Otherwise, the estimated number of sources is the rank of the CS subspace which is generally less than the true number of sources. Therefore, the NMDL method tends to underestimate the number of sources, particularly for large-scale systems where both the number of sensors and number of sources are large.

IV. NUMERICAL RESULTS

Consider first a ULA of six sensors with $d = \lambda/2$ receiving two equal-power sources. The empirical probabilities of correct detection versus number of snapshots are plotted in Fig. 2. Note that a correct detection means that the source of interest as well as the interference source are successfully detected. Since the MMDL method can use the training sequence of the desired signal, it is much more accurate than the classical MDL (CMDL), NMDL and exponentially embedded families (EEF)

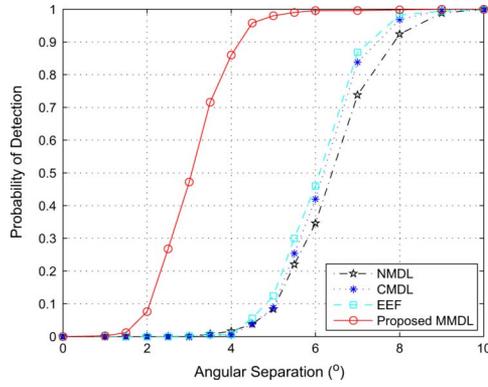


Fig. 3. Probability of correct detection versus angular separation. $[\theta_1, \theta_2] = [0^\circ, \Delta\theta]$, SNR = 0 dB, $N = 80$, $M = 6$ and $C = 2$; 500 trials.

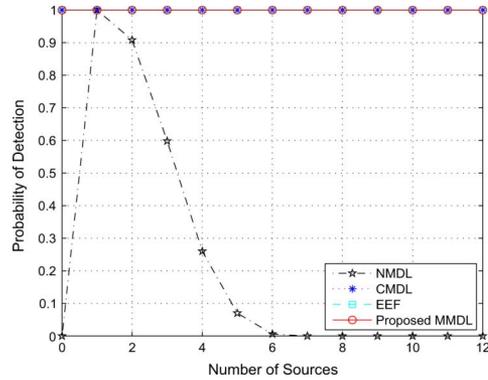


Fig. 4. Probability of correct detection versus number of sources in large-scale system. SNR = 10 dB, $N = 100$, $M = 16$ and $C = 2$; 500 trials.

[5] methods. To obtain the same accuracy as the MMDL, the NMDL, CMDL and EEF methods need around additional 150 samples. Fig. 3 shows the empirical results versus angular separation. It is seen that the MMDL method is much more accurate than the CMDL, NMDL and EEF methods for small angular separation.

As noted in Remark B, the NMDL method generally underestimates the number of sources because it requires the *ordered* Gerschgorin radii, equivalently selecting the CS subspace. This becomes more true for a large-scale system. Consider now a ULA of 16 sensors which receives five equal-power sources. Table I depicts the number of sources detected by the four methods at SNR = 5 dB, $[\theta_1, \dots, \theta_5] = [4.03^\circ, -5.08^\circ, 12.97^\circ, 8.51^\circ, -0.47^\circ]$ and $C = 2$. It is seen that the MMDL scheme outperforms the EEF and CMDL methods when $N \leq 100$. The NMDL method, however, tends to underestimate the source number even at $N = 3000$. As the number of sources increases, the probability of correct detection of the NMDL method converges to zero, as shown in Fig. 4, thereby indicating that it can not provide a reliable estimate of the number of sources in this large-scale system.

TABLE I
NUMBER OF SOURCES DETECTED BY THE NMDL, CMDL, EEF AND PROPOSED MMDL METHODS FOR 100 INDEPENDENT RUNS

Detection Method	k	Number of Snapshots					
		40	70	100	500	1000	3000
NMDL	0	0	0	0	0	0	0
	1	47	14	1	0	0	0
	2	17	19	21	15	6	0
	3	3	9	25	18	8	2
	4	33	58	52	31	0	0
	5	0	0	1	36	86	98
CMDL	0	0	0	0	0	0	0
	1	0	0	0	0	0	0
	2	0	0	0	0	0	0
	3	0	0	0	0	0	0
	4	87	54	23	0	0	0
	5	13	46	77	100	100	100
EEF	0	0	0	0	0	0	0
	1	0	0	0	0	0	0
	2	0	0	0	0	0	0
	3	0	0	0	0	0	0
	4	78	48	17	0	0	0
	5	22	52	83	100	100	100
Proposed MMDL	0	0	0	0	0	0	0
	1	5	0	0	0	0	0
	2	28	2	0	0	0	0
	3	32	3	0	0	0	0
	4	0	0	0	0	0	0
	5	35	95	100	100	100	100
6	0	0	0	0	0	0	

V. CONCLUSION

An MMDL estimator for source number has been developed. Using the training sequence of the desired signal, the MMDL method can accurately calculate the MMSEs and thereby results in a more accurate estimation of the source number than the CMDL, NMDL and EEF methods. Meanwhile, we show that the NMDL method tends to underestimate the source number in large-scale systems due to the use of the *ordered* Gerschgorin radii while the MMDL, CMDL and EEF methods retain their insensitiveness.

REFERENCES

- [1] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 8, pp. 1190–1196, Aug. 1989.
- [2] H. T. Wu, J. F. Yang, and F. K. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1325–1333, 1995.
- [3] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.
- [4] S. Aouada, A. M. Zoubir, and C. M. S. See, "Source detection in the presence of nonuniform noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Montreal, QC, Canada, May 2004, vol. 2, pp. 165–168.
- [5] C. Xu and S. Kay, "Source enumeration via the EEF criterion," *IEEE Signal Process. Lett.*, vol. 15, pp. 569–572, 2008.
- [6] L. Huang, T. Long, and S. Wu, "Source enumeration for high-resolution array processing using improved Gerschgorin radii without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5916–5925, Dec. 2008.
- [7] L. Huang and S. Wu, "Low-complexity MDL method for accurate source enumeration," *IEEE Signal Process. Lett.*, vol. 14, no. 4, pp. 581–584, Sep. 2007.
- [8] L. Huang, S. Wu, and X. Li, "Reduced-rank MDL method for source enumeration in high-resolution array processing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5658–5667, Dec. 2007.
- [9] M. Novey and T. Adali, "Complex ICA by negentropy maximization," *IEEE Trans. Neural Networks*, vol. 19, no. 4, pp. 596–609, Apr. 2008.
- [10] J. S. Goldstein and I. S. Reed, "Subspace selection for partially adaptive sensor array processing," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. 33, no. 2, pp. 539–544, Apr. 1997.