Reduced-Rank MDL Method for Source Enumeration in High-Resolution Array Processing

Lei Huang, *Member, IEEE*, Shunjun Wu, *Member, IEEE*, and Xia Li, *Member, IEEE*

***Abstract*—This paper proposes a reduced-rank minimum description length (MDL) method to enumerate the incident waves impinging on a uniform linear array (ULA). First, a new observation data and a reference signal are formed from sensor data by means of the shift-invariance property of the ULA. A cross-correlation between them is calculated, which is able to capture signal information and efficiently suppress additive noise. Second, the normalized cross-correlation is used as initial information for a recursion procedure to quickly partition the observation data into two orthogonal components in a signal subspace and a reduced-rank noise subspace. The components in the noise subspace are employed to calculate the total code length that is required to encode the observation data. Finally, the model with the shortest code length, namely the minimum description length, is chosen as the best model. Unlike the traditional MDL methods, this method partitions the observation data into the cleaner signal and noise subspace components by means of the recursion procedure, avoiding the estimation of a covariance matrix and its eigendecomposition. Thus, the method has the advantage of computational simplicity. Its performance is demonstrated via numerical results.**

***Index Terms*—Direction-of-arrival (DOA), eigenvalue decomposition (EVD), high resolution, minimum description length (MDL), multistage Wiener filter (MSWF), reduced rank, sensor array signal processing, signal enumeration, Wiener filter.**

I. Introduction

HIGH-SPEED data transmission and real-time multimedia service techniques are developing toward the fourth generation wireless communication system. As a key part of the techniques, adaptive array signal processing has received considerable attention. However, many high-resolution algorithms for parameter estimation used in the modern array signal processing depend on prior knowledge of the number of incident signals. As a result, the ability to accurately enumerate the incident signals with low computational complexity becomes more crucial.

In the community of array signal processing, the problem of source enumeration has been investigated extensively in [1]–[15]. In [1], Wax and Kailath proposed to apply the Akaike information criterion (AIC) [16] and minimum description length (MDL) [17] to source enumeration. The number of signals in the AIC and MDL criteria is determined by minimizing the Kullback-Leibler distance between the hypothesized model and the observation data. Wax and Kailath showed that the MDL criterion yields a consistent estimate of the number of signals while the AIC scheme yields an inconsistent estimate that tends, asymptotically, to overestimate the number of signals. The behavior of the information theoretic criteria for source enumeration has been well investigated by Zhang *et al.* [3] and Liavas *et al.* [4].

To enumerate the incident signals in a correlated signal environment, Di [5] and others, such as Shan *et al.* [6], Cozzens *et al.* [7], Xu *et al.* [8], and Ma *et al.* [9], employ the spatial smoothing technique [18], [19] to decorrelate the coherency of signals. Unlike these methods, Wax and Ziskind [2] have developed an MDL approach for coherent signal enumeration based on a multidimensional search. This approach first partitions the sensor data into signal and noise subspace components, and then separately calculates their MDL descriptors. Finally, the MDL descriptors are added up to attain the total description length for the sensor data. Nevertheless, while the MDL method [2] can attain an efficient estimate of the number of signals, it is rather computationally intensive since it resorts to the maximum likelihood (ML) estimates of the directions-of-arrival (DOAs) of incident signals, which are obtained by solving a multidimensional nonlinear minimization problem. Recently, Valaee and Kabal [10] developed a method for source enumeration based on predictive description length (PDL). Similar to the MDL method [2], Valaee and Kabal employed the ML estimates of the DOAs to partition the sample-covariance matrix into two orthogonal components in the signal and noise subspaces. Each component is then encoded and the results are added up to obtain the total code length. The model with the smallest code length is selected as the best model. However, while the PDL algorithm is more accurate than the MDL approach [2], it is also rather computationally expensive since it essentially involves the ML estimates of the DOAs. To reduce the computational load of the PDL method, an alternating projection (AP) method [20] may be employed to decompose the multidimensional ML estimator into several 1-D nonlinear minimization problems. Nevertheless, the PDL algorithm is still computationally cumbersome since it requires the ML estimates of the DOAs for every time instant.

Ishikawa *et al.* [12] presented a method for source enumeration that is more accurate than the traditional MDL method [1], and only depends on the estimate of the covariance matrix and its eigenvalue decomposition (EVD), requiring the computational complexity of $O(M^2N)+O(M^3)$ flops, where M and

Manuscript received August 6, 2006; revised December 29, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. A. Rahim Leyman. This work was supported by the Guangdong Natural Science Foundation under Grant 7301317.

L. Huang was with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China. He is now with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: lhuang8sasp@hotmail.com).

S. Wu is with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China.

X. Li is with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China.

Digital Object Identifier 10.1109/TSP.2007.899344

N represent the number of sensors of the array and the number of snapshots, respectively. Unfortunately, the procedure of estimating the covariance matrix and computing the eigenvalues is still computationally intensive and time consuming, which indicates that the EVD-based methods for source enumeration are unsuitable for some practical situations when the number of sensors is large and/or the number of signals requires to be tracked in an online manner. Moreover, the EVD-based methods may hardly detect the number of signals in some severe environments such as low signal-to-noise ratio (SNR) and/or small sample size where the covariance matrix cannot be calculated accurately.

All the algorithms reviewed previously require the estimated covariance matrix and its eigenvalues. Moreover, the MDL approach [2] and the PDL method [10] still rely on the ML estimates of the DOAs, besides the eigenvalues of the estimated covariance matrix. Wu and Fuhrmann [13] developed an approach for source enumeration that does not resort to the eigenvalues of the estimated covariance matrix. However, as noted in [13], this method is more computationally intensive than the common EVD-based methods since it relies on the ML solution to the direction finding problem for each of several hypotheses. While the MDL method [14] is more computationally efficient than the EVD-based methods and outperforms the methods for source enumeration [2], [12] in performance, it requires the knowledge of the training data of the desired signal. To date, there are no more successful efforts to develop low computational complexity methods for source enumeration in the literature.

To accurately detect the number of signals with a reduced computational burden, this paper proposes a reduced-rank MDL method, based on the multistage Wiener filter (MSWF) [21], [22]. The proposed method proceeds in three steps. First, a new observation data and a reference signal are attained from the sensor data by means of the shift-invariance property of a uniform linear array (ULA). A cross-correlation between them is calculated, which is capable of capturing signal information and efficiently eliminating additive noise. The normalized cross-correlation is then used to quickly partition the observation data into two orthogonal components in the $k \times M$ signal subspace and the $(D-k) \times M$ reduced-rank noise subspace, where k and D are the assumed number of signals and the dimension of the reduced-rank observation space. Meanwhile, the signal and noise subspace components are employed to determine the parameter D. Second, the variances of the noise subspace components instead of the eigenvalues associated with the estimated covariance matrix are used to calculate the total code length that is required to encode the sensor data. Therefore, the estimate of the covariance matrix and its EVD can be avoided. Finally, the number of signals is determined by minimizing the description length of the sensor data.

The remainder of this paper is organized as follows. The data model and basic assumptions are given in the next section. The reduced-rank MDL method for source enumeration is proposed in Section III. Meanwhile, its consistency is proven. Numerical results are presented in Section IV. Finally, conclusions are drawn in Section V.

II. Data Model and Basic Assumptions

Consider a ULA composed of M isotropic sensors. Impinging upon the array are p narrow-band signals $\{s_1(t), s_2(t), \ldots, s_p(t)\}$ from distinct directions $\{\theta_1, \theta_2, \ldots, \theta_p\}$. The p narrow-band signal sources, centered around a known frequency w_0, are placed in the far field, and thereby the wavefronts can be approximated as planar. For simplicity, we also assume that the sources and the sensors are in the same plane. Thus, employing complex envelope representation, the $M \times 1$ received vector of the array can be expressed as

$$\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T = \sum_{i=1}^{p} \boldsymbol{a}(\theta_i) s_i(t) + \boldsymbol{n}(t) \tag{1}$$

where $\boldsymbol{n}(t)$ is the additive noise, $\boldsymbol{a}(\theta_i)$ is the "steering vector" of the array toward direction θ_i, p represents the *unknown* number of incident signals, $(\cdot)^T$ denotes transpose, and $x_m(t) = \sum_{i=1}^{p} s_i(t) e^{j\omega_0 \tau_{m-1}(\theta_i)} + n_m(t)$ $(m = 1, \ldots, M)$ is the received noisy signal at the mth sensor in which $\tau_{m-1}(\theta_i) = (d/c)(m-1)\sin(\theta_i)$ is the propagation delay between the first sensor (the reference point) and the mth sensor to a wavefront impinging from direction θ_i, c denotes the propagation speed, and d is the distance between two adjacent sensors.

In matrix notation, (1) can be rewritten more compactly as

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{2}$$

where

$$\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_p)] \tag{3}$$

is the array response matrix with

$$\boldsymbol{a}(\theta_i) = [1, e^{j\omega_0 \tau_1(\theta_i)}, \ldots, e^{j\omega_0 \tau_{M-1}(\theta_i)}]^T.$$

Suppose that the received vector $\boldsymbol{x}(t)$ is sampled at N time instants: $t_1, t_2, \ldots, t_N$.

Throughout this paper, we make the following basic assumptions on the sensor data model.

- A1) The array response matrix $\boldsymbol{A}(\boldsymbol{\theta})$ is unambiguous, namely the array response vectors $\{\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_p)\}$ are linearly independent for any set of distinct incident angles $\{\theta_1, \theta_2, \ldots, \theta_p\}$, which indicates that the matrix $\boldsymbol{A}(\boldsymbol{\theta})$ is of full rank.
- A2) The signals $\{\boldsymbol{s}(t_\ell)\}$ are jointly stationary, zero-mean complex Gaussian random processes, which are uncorrelated with each other.
- A3) The background noises $\{\boldsymbol{n}(t_\ell)\}$ are independent and identity distributed complex, zero-mean, Gaussian vectors with covariance matrix $\sigma_n^2 \boldsymbol{I}_M$, where $\boldsymbol{I}_M$ represents the $M \times M$ identity matrix. In addition, the noises $\{\boldsymbol{n}(t_\ell)\}$ are uncorrelated with the signals $\{\boldsymbol{s}(t_\ell)\}$.
- A4) The number of sensors is greater than the number of signals, and satisfies the inequality $M - 1 > p$.

Under these assumptions, the output of the array is a complex Gaussian random process with zero mean and the following covariance matrix:

$$\boldsymbol{R}_x = E\left[\boldsymbol{x}(t)\boldsymbol{x}^H(t)\right] = \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{R}_s \boldsymbol{A}^H(\boldsymbol{\theta}) + \sigma_n^2 \boldsymbol{I}_M \tag{4}$$

where $E[\cdot]$ denotes expectation, $(\cdot)^H$ denotes Hermition transpose, and $\boldsymbol{R}_s = E[\boldsymbol{s}(t)\boldsymbol{s}^H(t)]$ denotes the signal covariance matrix, which is nonsingular and diagonal. The subspace spanned

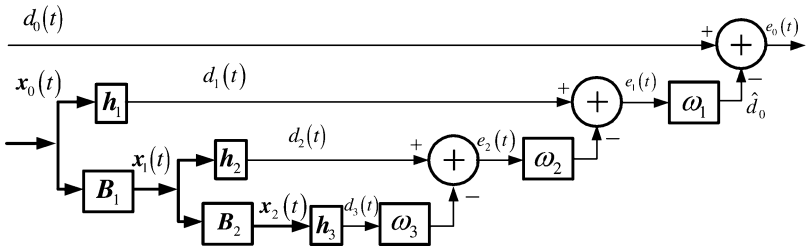


Fig. 1. MSWF.

by the columns of $\boldsymbol{A}(\boldsymbol{\theta})$ is called signal subspace, and its orthogonal complement is called noise subspace. In practical applications, however, we do not have access to the ideal covariance matrix. Instead, we, in general, only obtain its sample-covariance matrix $\hat{\boldsymbol{R}}_x = 1/N \sum_{\ell=1}^{N} \boldsymbol{x}(t_\ell)\boldsymbol{x}^H(t_\ell)$, where N is finite.

III. Reduced-Rank MDL Method for Source Enumeration

A. Subspace Decomposition Without Eigendecomposition

In this subsection, we will present a subspace decomposition method without eigendecomposition or any *priori* knowledge of signal sources. To begin with, we define a new observation data from (1) as $\boldsymbol{x}_0(t) \triangleq [x_2(t), x_3(t), \ldots, x_M(t)]^T$. Using the shift-invariance property of the ULA, we can write $\boldsymbol{x}_0(t)$ more compactly as

$$\boldsymbol{x}_0(t) = \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{D}\boldsymbol{s}(t) + \tilde{\boldsymbol{n}}(t) \tag{5}$$

where $\boldsymbol{D} = \mathrm{diag}(e^{jw_0 d/c \sin\theta_1}, \ldots, e^{jw_0 d/c \sin\theta_p})$, $\boldsymbol{A}_q(\boldsymbol{\theta})$ consists of the first q rows of $\boldsymbol{A}(\boldsymbol{\theta})$ with $q = M-1$, and $\tilde{\boldsymbol{n}}(t) = [n_2(t), n_3(t), \ldots, n_M(t)]^T$. To calculate the signal subspace by a successive refinement procedure, we also define a reference signal by

$$d_0(t) \triangleq x_1(t) = \boldsymbol{s}^T(t)\mathbf{1} + n_1(t) \tag{6}$$

where $\mathbf{1} = [1, 1, \ldots, 1]^T$. In the sequel, we can calculate the cross-correlation between the new observation data and the reference signal, namely the initial information for the refinement procedure

$$\begin{aligned} \boldsymbol{r}_{x_0 d_0} &= E\left[\boldsymbol{x}_0(t)d_0^*(t)\right] = \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{D}\boldsymbol{R}_s\mathbf{1} \\ &\triangleq \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{\beta} \\ &= \sum_{i=1}^{p} \beta_i \boldsymbol{a}_q(\theta_i) \end{aligned} \tag{7}$$

where $(\cdot)^*$ is the complex conjugate, $\boldsymbol{\beta} = \boldsymbol{D}\boldsymbol{R}_s\mathbf{1}$ and $\boldsymbol{a}_q(\theta_i) = [1, e^{jw_0\tau_1(\theta_i)}, \ldots, e^{jw_0\tau_{q-1}(\theta_i)}]^T$. Considering $\boldsymbol{R}_s$ is a nonsingular matrix yields

$$\boldsymbol{\beta} \neq \mathbf{0}. \tag{8}$$

Therefore, the cross-correlation $\boldsymbol{r}_{x_0 d_0}$ is a linear combination of all the direction vectors $\boldsymbol{a}_q(\theta_i)(i = 1, 2, \ldots, p)$. This implies that the cross-correlation $\boldsymbol{r}_{x_0 d_0}$ is able to capture the signal information. Meanwhile, it is shown in (7) that the additive noise has been efficiently eliminated in the calculation of $\boldsymbol{r}_{x_0 d_0}$. In the sequel, we use the cross-correlation to define the following matched filter:

$$\boldsymbol{h}_1 = \frac{\boldsymbol{r}_{x_0 d_0}}{\|\boldsymbol{r}_{x_0 d_0}\|} \tag{9}$$

where $\|\cdot\|$ denotes the vector norm. Partitioning the new sensor data $\boldsymbol{x}_0(t)$ with the matched filter $\boldsymbol{h}_1$ in a manner similar to that of the multistage Wiener filter (MSWF) [21], we attain the desired signal $\boldsymbol{d}_i(t)$ and its orthogonal component $\boldsymbol{x}_i(t)$ at the ith stage by

$$d_i(t) = \boldsymbol{h}_i^H \boldsymbol{x}_{i-1}(t) \tag{10}$$

and

$$\begin{aligned} \boldsymbol{x}_i(t) &= \boldsymbol{x}_{i-1}(t) - \boldsymbol{h}_i d_i(t) \\ &= \boldsymbol{x}_{i-1}(t) - \boldsymbol{h}_i \boldsymbol{h}_i^H \boldsymbol{x}_{i-1}(t) \\ &= \boldsymbol{B}_i \boldsymbol{x}_{i-1}(t) \end{aligned} \tag{11}$$

where $\boldsymbol{B}_i = \boldsymbol{I} - \boldsymbol{h}_i\boldsymbol{h}_i^H$ is the blocking matrix and $\boldsymbol{h}_i$ is the matched filter updated as

$$\boldsymbol{h}_i = \frac{E\left[\boldsymbol{x}_{i-1}(t)d_{i-1}^*(t)\right]}{\left\|E\left[\boldsymbol{x}_{i-1}(t)d_{i-1}^*(t)\right]\right\|}. \tag{12}$$

Fig. 1 shows a block diagram of a three-stage MSWF, where $e_0(t)$, $e_1(t)$, and $e_2(t)$ are the estimation errors and ω_1, ω_2, and ω_3 are the scalar weights. The desired signal $\boldsymbol{d}_i(t)$ is obtained by prefiltering the observation data $\boldsymbol{x}_{i-1}(t)$ with the matched filters $\boldsymbol{h}_i$, but annihilated by the blocking matrix $\boldsymbol{B}_i$. The observation data is partitioned stage-by-stage in the same refinement manner. As a result, we obtain the prefiltering matrix $\boldsymbol{T}_q = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_q]$ and the desired signals of the MSWF $d_i(t)(i = 1, 2, \ldots, q)$ by q successive recursions.

Lemma 1: The first p matched filters $\boldsymbol{h}_i(i = 1, 2, \ldots, p)$ span the same range subspace of $\boldsymbol{A}_q(\boldsymbol{\theta})$ while the last $q-p$ matched filters span the null subspace of $\boldsymbol{A}_q(\boldsymbol{\theta})$, namely

$$\mathcal{S}^{(p)} = \mathrm{span}\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p\} \tag{13}$$

$$\mathcal{N}^{(q-p)} = \mathrm{span}\{\boldsymbol{h}_{p+1}, \boldsymbol{h}_{p+2}, \ldots, \boldsymbol{h}_q\} \tag{14}$$

where $\mathcal{S}^{(p)}$ and $\mathcal{N}^{(q-p)}$ represent the signal subspace of rank p and the noise subspace of rank $q-p$, respectively, and $\boldsymbol{h}_i(i = 1, 2, \ldots, q)$ are the orthonormal matched filters of the MSWF.

Proof: The proof is seen in Appendix I. ■

Lemma 1 indicates that the signal and noise subspaces can be calculated by the refinement procedure. Meanwhile, note that the cross-correlation $\boldsymbol{r}_{x_0 d_0}$ is able to capture the signal information and efficiently suppress the additive noise. In the sequel,

when $\boldsymbol{r}_{x_0 d_0}$ is employed as the initial information for the refinement procedure, the signal and noise subspaces can be obtained more accurately by the refinement procedure than by the EVD-based methods, especially for the case of low SNR and/or small sample size. With Lemma 1, we can show that the proposition below is true.

Proposition 1: The last $q-p$ desired signals of the MSWF are uncorrelated with each other and their variances equal the noise variance, namely

$$\delta_{i,j} \triangleq E\left[d_i(t)d_j^*(t)\right]=0 \ (i,j=p+1,p+2,\ldots,q, i\neq j) \quad (15)$$

$$\sigma_{d_i}^2 \triangleq E[d_i(t)d_i^*(t)]=\sigma_n^2 \ (i=p+1,p+2,\ldots,q). \quad (16)$$

Proof: The proof of Proposition 1 is seen in Appendix II. ■

Notice that the first p desired signals of the MSWF do not satisfy (15) and (16). Consequently, given the $\delta_{i,j}$ and $\sigma_{d_i}^2$, we can very easily determine the number of signals from Proposition 1. In a practice situation, however, the $\delta_{i,j}$ and $\sigma_{d_i}^2$ are unknown, and thereby the number of signals cannot be obtained from (15) and (16) any more. In the following subsections, we will propose a reduced-rank MDL method for source enumeration.

B. Novel MDL Criterion for Source Enumeration

According to the MDL principal, for a given data set and a family of probabilistic models, one should select the model that yields the shortest description length of the data. The description length of the data can be evaluated quantitatively. In general, given an observation data set $\boldsymbol{X} = \{\boldsymbol{x}_i(t_\ell)\}_{\ell=1}^N$ and a probabilistic model $f(\boldsymbol{X}|\boldsymbol{\mu})$, where $\boldsymbol{\mu}$ denotes an unknown parameter vector, the shortest code length required to encode the data using the model can be asymptotically written as [2]

$$\mathcal{L}\{\boldsymbol{x}_i(t_\ell)\} = -\log f(\boldsymbol{X}|\hat{\boldsymbol{\mu}}) + \frac{1}{2}K\log N \quad (17a)$$

where $\hat{\boldsymbol{\mu}}$ is the maximum likelihood estimate of $\boldsymbol{\mu}$ and K denotes the number of free parameters in the vector $\boldsymbol{\mu}$. If the observation data $\{\boldsymbol{x}_i(t_\ell)\}$ are assumed to be statistically independent complex Gaussian random vectors with mean $\boldsymbol{m}_{x_i}$, their joint probability density is given by

$$f(\boldsymbol{X}|\boldsymbol{\mu}) = \prod_{\ell=1}^{N} \frac{1}{\pi^q |\boldsymbol{R}_{x_i}|} \times \exp\left\{-\left(\boldsymbol{x}_i(t_\ell)-\boldsymbol{m}_{x_i}\right)^H \boldsymbol{R}_{x_i}^{-1}\left(\boldsymbol{x}_i(t_\ell)-\boldsymbol{m}_{x_i}\right)\right\} \quad (17b)$$

where $|\cdot|$ is the determinant, $\boldsymbol{R}_{x_i}$ is the covariance matrix of $\boldsymbol{x}_i(t_\ell)$ and $\boldsymbol{\mu}$ is composed of $\boldsymbol{m}_{x_i}$ and $\boldsymbol{R}_{x_i}$. Taking the logarithm, omitting terms independent of $\boldsymbol{\mu}$ and substituting the result into (17a) yields the description length of the observation data

$$\mathcal{L}\{\boldsymbol{x}_i(t_\ell)\} = N\log|\boldsymbol{R}_{x_i}| + \mathrm{tr}\boldsymbol{R}_{x_i}^{-1}\hat{\boldsymbol{R}}_{x_i} + \frac{1}{2}K\log N \quad (17c)$$

where tr denotes the trace operation and $\hat{\boldsymbol{R}}_{x_i} \triangleq 1/N\sum_{\ell=1}^N (\boldsymbol{x}_i(t_\ell)-\boldsymbol{m}_{x_i})(\boldsymbol{x}_i(t_\ell)-\boldsymbol{m}_{x_i})^H$.

From (13) and (14) it follows that $\boldsymbol{T}_s^{(k)} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_k]$ and $\boldsymbol{T}_n^{(k)} = [\boldsymbol{h}_{k+1}, \boldsymbol{h}_{k+2}, \ldots, \boldsymbol{h}_q]$ span the signal and noise subspaces, respectively. Here, k is the assumed number of signals. Partitioning the observation data $\boldsymbol{x}_0(t)$ into the components of the signal and noise subspaces yields

$$\boldsymbol{T}_q^{(k),H}\boldsymbol{x}_0(t) = \begin{bmatrix}\boldsymbol{T}_s^{(k),H}\\ \boldsymbol{T}_n^{(k),H}\end{bmatrix}\boldsymbol{x}_0(t) \triangleq \begin{bmatrix}\boldsymbol{x}_s^{(k)}(t)\\ \boldsymbol{x}_n^{(k)}(t)\end{bmatrix}. \quad (18)$$

In the sequel, the components in the signal and noise subspaces can be calculated as

$$\boldsymbol{x}_s^{(k)}(t) = \boldsymbol{T}_s^{(k),H}\boldsymbol{x}_0(t) = [d_1(t), d_2(t), \ldots, d_k(t)]^T \quad (19)$$

$$\boldsymbol{x}_n^{(k)}(t) = \boldsymbol{T}_n^{(k),H}\boldsymbol{x}_0(t) = [d_{k+1}(t), d_{k+2}(t), \ldots, d_q(t)]^T. \quad (20)$$

Since the prefiltering matrix $\boldsymbol{T}_q^{(k)}$ is a unity matrix, it follows from (18) that the total code length required to encode the observation data $\{\boldsymbol{x}_0(t_\ell)\}$ is equivalent to encoding the signal subspace components $\{\boldsymbol{x}_s^{(k)}(t_\ell)\}$ and the noise subspace components $\{\boldsymbol{x}_n^{(k)}(t_\ell)\}$, respectively.

Substituting (5) into (20) yields

$$\boldsymbol{x}_n^{(k)}(t) = \boldsymbol{T}_n^{(k),H}\boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{D}\boldsymbol{s}(t) + \boldsymbol{T}_n^{(k),H}\tilde{\boldsymbol{n}}(t) = \boldsymbol{T}_n^{(k),H}\tilde{\boldsymbol{n}}(t). \quad (21)$$

Since the additive noise $\tilde{\boldsymbol{n}}(t)$ is a complex Gaussian random vector with zero mean and covariance matrix $\sigma_n^2\boldsymbol{I}_q$, it is straightforward to mode $\boldsymbol{x}_n^{(k)}(t)$ as a $(q-k)\times 1$ complex Gaussian random vector with zero mean and covariance matrix $\sigma_n^2\boldsymbol{I}_{q-k}$, namely

$$\boldsymbol{x}_n^{(k)}(t) \sim N_{q-k}\left(\boldsymbol{0}, \sigma_n^2\boldsymbol{I}_{q-k}\right). \quad (22)$$

Note that the probability mode of $\boldsymbol{x}_n^{(k)}(t)$ has only a single parameter σ_n^2. In the sequel, substituting the ML estimate of σ_n^2, namely $\hat{\sigma}_n^2 = \mathrm{tr}(\hat{\boldsymbol{R}}_{nn})/(q-k)$, and (22) into (17), we obtain the code length required to encode the noise subspace components

$$\mathcal{L}\left\{\boldsymbol{x}_n^{(k)}(t_\ell)\right\} = N\log\left(\frac{\mathrm{tr}\hat{\boldsymbol{R}}_{nn}^{(k)}}{q-k}\right)^{(q-k)} + \frac{1}{2}\log N \quad (23)$$

where $\hat{\boldsymbol{R}}_{nn}^{(k)} = 1/N\sum_{\ell=1}^N \boldsymbol{x}_n^{(k)}(t_\ell)\boldsymbol{x}_n^{(k),H}(t_\ell)$.

Given the noise subspace components $\{\boldsymbol{x}_n^{(k)}(t_\ell)\}$, the conditional distribution of the signal subspace components $\{\boldsymbol{x}_s^{(k)}(t_\ell)\}$ may be modelled as

$$\boldsymbol{x}_s^{(k)}(t)|\boldsymbol{x}_n^{(k)}(t) \sim N_k\left(\boldsymbol{C}\boldsymbol{x}_n^{(k)}(t), \boldsymbol{G}\right) \quad (24)$$

where $\boldsymbol{C}$ and $\boldsymbol{G}$ are unknown complex matrices of dimensions $k\times(q-k)$ and $k\times k$, respectively. Note that the number of free parameters in the matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ are $2(q-k)k$ and k^2, respectively, and the ML estimates of the matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ are given by

$$\hat{\boldsymbol{C}} = \hat{\boldsymbol{R}}_{sn}^{(k)}\hat{\boldsymbol{R}}_{nn}^{(k),-1} \quad (25)$$

$$\hat{\boldsymbol{G}} = \hat{\boldsymbol{R}}_{ss}^{(k)} - \hat{\boldsymbol{R}}_{sn}^{(k)}\hat{\boldsymbol{R}}_{nn}^{(k),-1}\hat{\boldsymbol{R}}_{ns}^{(k)} \quad (26)$$

where $\hat{\boldsymbol{R}}_{sn}^{(k)} = 1/N\sum_{\ell=1}^N \boldsymbol{x}_s^{(k)}(t_\ell)\boldsymbol{x}_n^{(k),H}(t_\ell)$ and $\hat{\boldsymbol{R}}_{ss}^{(k)} = 1/N\sum_{\ell=1}^N \boldsymbol{x}_s^{(k)}(t_\ell)\boldsymbol{x}_s^{(k),H}(t_\ell)$. Substituting (24)–(26) into (17a)–(17c), and performing some straightforward algebraic

manipulations, we can obtain the code length encoding the signal subspace components

$$\mathcal{L}\left\{\boldsymbol{x}_s^{(k)}(t_\ell)\right\} = N\log\frac{1}{\left|\hat{\boldsymbol{R}}_{nn}^{(k)}\right|} + \frac{1}{2}\left(2k(q-k)+k^2\right)\log N. \tag{27}$$

Therefore, combining (23) and (27) and omitting the term in (23) that is independent of k, we obtain the total code length required to encode the observation data given by

$$\mathcal{L}\left\{\boldsymbol{x}_s^{(k)}(t_\ell),\boldsymbol{x}_n^{(k)}(t_\ell)\right\} = N(q-k)\log\frac{\frac{1}{q-k}\mathrm{tr}\hat{\boldsymbol{R}}_{nn}^{(k)}}{\left|\hat{\boldsymbol{R}}_{nn}^{(k)}\right|^{1/(q-k)}} + \frac{1}{2}k(2q-k)\log N. \tag{28}$$

More details about the derivation of (28) may be found in [2].

It is indicated in Proposition 1 that the desired signals of the MSWF after the kth stage $d_i(t)(i = k+1,\ldots,q)$ are uncorrelated with each other, namely $E[d_i(t)d_j^*(t)] = 0(i,j = k+1,\ldots,q, i\neq j)$. As a result, $\boldsymbol{R}_{nn}^{(k)}$ is a diagonal matrix

$$\boldsymbol{R}_{nn}^{(k)} = \mathrm{diag}\left(\sigma_{d_{k+1}}^2,\sigma_{d_{k+2}}^2,\ldots,\sigma_{d_q}^2\right). \tag{29}$$

Thus, substituting the ML estimate of $\boldsymbol{R}_{nn}^{(k)}$ into (28), we eventually obtain a new MDL estimator for the number of signals

$$\hat{p} = \arg\min_{k=0,1,\ldots,q-1}\mathrm{MDLE}(k) \tag{30}$$

where

$$\mathrm{MDLE}(k) = N(q-k)\log\frac{\frac{1}{q-k}\sum_{i=k+1}^{q}\hat{\sigma}_{d_i}^2}{\left(\prod_{i=k+1}^{q}\hat{\sigma}_{d_i}^2\right)^{1/(q-k)}} + \frac{1}{2}k(2q-k)\log N \tag{31}$$

in which $\hat{\sigma}_{d_i}^2 = 1/N\sum_{\ell=1}^{N} d_i(t_\ell)d_i^*(t_\ell)$ is the ML estimate of $\sigma_{d_i}^2$. Note that the recursion procedure can directly generate the estimated variances of the desired signals $\hat{\sigma}_{d_i}^2(i = 1,\ldots,q)$ while attaining the estimated signal and noise subspaces.

C. Reduced-Rank MDL Criterion for Source Enumeration

Since the desired signals of the MSWF after the kth stage have the same power as the background noise, as noted in Proposition 1, reduced-rank noise subspace components may lead to the sufficient accuracy of source enumeration. The reduced-rank noise subspace components can be expressed as

$$\boldsymbol{x}_{n,D}^{(k)}(t) = \boldsymbol{T}_{n,D}^{(k),H}\boldsymbol{x}_0(t) = [d_{k+1}(t),d_{k+2}(t),\ldots,d_D(t)]^T \tag{32}$$

where $D(k < D < q)$ is the dimension of the reduced-rank observation space. Consequently, the covariance matrix of the reduced-rank noise subspace components is calculated as

$$\begin{aligned}\boldsymbol{R}_{nn}^{(k),D} &= E\left[\boldsymbol{x}_{n,D}^{(k)}(t)\boldsymbol{x}_{n,D}^{(k),H}(t)\right]\\ &= \mathrm{diag}\left(\sigma_{d_{k+1}}^2,\sigma_{d_{k+2}}^2,\ldots,\sigma_{d_D}^2\right).\end{aligned} \tag{33}$$

Substituting the ML estimate of $\boldsymbol{R}_{nn}^{(k),D}$ into (28) and noticing that the reduced-rank observation space is of dimension D, we obtain the total code length required to encode the reduced-rank observation space

$$\begin{aligned}&\mathcal{L}\left\{\boldsymbol{x}_s^{(k)}(t_\ell),\boldsymbol{x}_{n,D}^{(k)}(t_\ell)\right\}\\ &\quad= N(D-k)\log\frac{\frac{1}{D-k}\mathrm{tr}\hat{\boldsymbol{R}}_{nn}^{(k),D}}{\left|\hat{\boldsymbol{R}}_{nn}^{(k),D}\right|^{1/(D-k)}} + \frac{1}{2}k(2D-k)\log N\\ &\quad= N(D-k)\log\frac{\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\sigma}_{d_i}^2}{\left(\prod_{i=k+1}^{D}\hat{\sigma}_{d_i}^2\right)^{1/(D-k)}} + \frac{1}{2}k(2D-k)\log N\end{aligned} \tag{34}$$

where $\hat{\boldsymbol{R}}_{nn}^{(k),D} = \mathrm{diag}(\hat{\sigma}_{d_{k+1}}^2,\hat{\sigma}_{d_{k+2}}^2,\ldots,\hat{\sigma}_{d_D}^2)$ is the ML estimate of the reduced-rank covariance matrix $\boldsymbol{R}_{nn}^{(k),D}$. Thus, the reduced-rank MDL estimator for the number of signals can be eventually written as

$$\hat{p} = \arg\min_{k=0,1,\ldots,D-1}\mathrm{MDLE}^{(D)}(k) \tag{35}$$

where

$$\mathrm{MDLE}^{(D)}(k) = N(D-k)\log\frac{\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\sigma}_{d_i}^2}{\left(\prod_{i=k+1}^{D}\hat{\sigma}_{d_i}^2\right)^{1/(D-k)}} + \frac{1}{2}k(2D-k)\log N. \tag{36}$$

Remark A: Unlike the MDL approach proposed by Wax and Ziskind [2], the reduced-rank MDL method does not rely on the ML estimates of the DOAs or the explicit formation of projection matrices on the signal and noise subspaces to find the noise subspace components, but rather directly generates a set of desired signals forming the signal and noise subspace components by means of a successive refinement procedure. On the other hand, the reduced-rank MDL method directly employs the variances of the desired signals instead of the eigenvalues associated with the estimated covariance matrix to evaluate the code length of the sensor data, which is different from that of the traditional MDL methods [1], [2]. As a result, the reduced-rank MDL method is more computationally efficient than the traditional MDL approaches. Meanwhile, note that the cross-correlation between the observation data and the reference signal is able to capture the signal information and efficiently eliminate the additive noise. Since the cross-correlation is used as the initial information for the refinement procedure to partition the sensor data into the signal and noise subspace components, the signal subspace components are capable of capturing the signal information while excluding a large portion of noise. On the contrary, their orthogonal components are able to retain the power of the noise while significantly mitigating the signal subspace components, and are thereby the cleaner noise subspace components. This eventually leads to the enhanced detection performance of the reduced-rank MDL approach, especially in some severe environments such as small sample size and/or low SNR.

Remark B: From (10)–(12), we can see that the dominant computational cost among them is the calculation of the matched filter, which requires q complex multiplications and $q-1$ additions for each snapshot, equivalently approximately q floating point operations (or flops), and thereby around $O(qN) = O((M-1)N)$ flops for each recursion. Meanwhile, note that the variances of the desired signals can be directly generated by the recursion procedure, and their computational cost is less than that of the calculation of the matched filters. Consequently, the proposed method only needs around $O(D(M-1)N)$ flops to determine the number of signals. However, the classical information theoretic methods for source enumeration, such as the AIC and MDL methods [1], involve the estimate of the covariance matrix and its EVD, therefore requiring $O(M^2N) + O(M^3)$ flops that are quite heavy for a large array. Moreover, the MDL algorithm [2] and the PDL approach [10] essentially involve the ML estimates of the DOAs, the estimated covariance matrix and its eigenvalues, which imply that much more computational cost is required in these methods than in the conventional EVD-based methods [1]. Thus, the reduced-rank MDL estimator outperforms the traditional EVD-based methods and the multidimensional search-based methods [2], [10] in computational complexity, in particular for a large array.

D. Rank D Adaptation

It is easy to see from (35) and (36) that if $D \leq p$, we cannot correctly detect the number of signals. If $p \ll D \approx q$, however, the reduced-rank MDL method might not be computationally attractive. The desired D should be a little greater than the true number of signals. It is stated in Proposition 1 that the cross-correlations between the last $q-p$ desired signals of the MSWF equal zero. Meanwhile, note that the absolute values of the cross-correlations between the first p adjacent desired signals are in general greater than one. This implies that we can determine D by using the cross-correlations between the adjacent desired signals of the MSWF. In practical applications, however, the cross-correlations between the adjacent desired signals after the pth stage do not equal but are close to zero due to finite samples. To this end, we defined a detector of D as

$$D = \max\left\{i : |\hat{\delta}_i| > \epsilon\right\} \tag{37}$$

where $\hat{\delta}_i = 1/N \sum_{\ell=1}^{N} d_i(t_\ell) d_{i-1}^*(t_\ell)$, $|\mathbf{x}|$ denotes the absolute value of a complex number $\mathbf{x}$, and ϵ is a small positive constant.

Remark C: Note that $|\hat{\delta}_i| (i = 1, 2, \ldots, p)$ are generally greater than one while $|\hat{\delta}_i| (i = p+1, p+2, \ldots, q)$ are small numbers approaching to zero. Consequently, the performance of the reduced-rank MDL method may be insensitive to the selection of ϵ over a reasonable large range, as will also be indicated by simulation results in Section IV.

E. Consistency of the Reduced-Rank MDL Criterion

In this subsection, we prove the strong consistency of the reduced-rank MDL criterion. The consistency proof requires the following preliminary results.

Lemma 2: For the sample-covariance matrix $\hat{\boldsymbol{R}}_x$, the following relation holds with probability one (*w. p. 1*):

$$\hat{\boldsymbol{R}}_x = \boldsymbol{R}_x + O(\sqrt{\log\log N/N}). \tag{38}$$

Proof: The proof of Lemma 2 may be found in [2], [23]. ■

From Lemma 2, we get

$$\hat{\sigma}_{d_i}^2 = \sigma_{d_i}^2 + O(\sqrt{\log\log N/N}) \quad (i = 1, 2, \ldots, D). \tag{39}$$

It follows from (16) and (39) that:

$$\hat{\sigma}_{d_i}^2 = \sigma_n^2 + O(\sqrt{\log\log N/N}) \quad (i = p+1, p+2, \ldots, D). \tag{40}$$

Since the variances of the first p desired signals of the MSWF are the powers of signals while the variances of the desired signals of the MSWF after the pth stage are equal to the noise variance, $\hat{\sigma}_{d_i}^2 (i = k+1, k+2, \ldots, D(k < p))$ are not all identity *w. p. 1* as the number of snapshots tends to infinity. With these results, we are able to prove that the reduced-rank MDL criterion is strong consistent. That is, as the number of snapshots tends to infinity, the total code length is minimized at $k = p$ *w. p. 1*. The proof is seen in Appendix III.

IV. Numerical Results

In this section, the performance of the proposed MDL estimator is evaluated by computer simulation. For fair comparison, the results of the traditional MDL method developed by Wax and Ziskind [2] and the PDL approach proposed by Valaee and Kabal [10] are also given. The array herein is assumed to be a ULA with ten isotropic sensors whose spacings equal half-wavelength. Suppose that there are two uncorrelated signals with equal power impinging upon the ULA from the directions $\{2^\circ, 7^\circ\}$. The background noise is assumed to be a stationary Gaussian random process that is uncorrelated with the signals. The number of snapshots is 100. The SNR is defined as the ratio of the power of signals to the power of noise at each sensor.

To illustrate that the proposed reduced-rank MDL estimator can obtain its full-rank detection performance, we have run 100 independent trials to calculate the probability of detection for the proposed MDL method with different ranks: $D = 2$, $D = 3$, $D = 4$, adaptive D determined by (37) and $D = 9$, namely the full-rank case. The results are given in Fig. 2. Note that the number of signals is two and thereby the reduced-rank MDL estimator with $D = 2$ cannot calculate the noise subspace components. As a result, the reduced-rank MDL method fails to enumerate the signals for $D = 2$. If the rank D is greater than 2, the reduced-rank MDL estimator can correctly detect the number of signals. It can also be observed from Fig. 2 that the reduced-rank MDL estimator is insensitive to the choice of ϵ over a quite large range from 0.005 to 0.5 as SNR is reasonably high, say $\mathrm{SNR} > -2.5$ dB. Note that when ϵ becomes small, D increases. Consequently, more desired signals of the MSWF after the pth stage are involved, which lead to the fact that the variances of the noise subspace components are not clustered sufficiently closely especially for the case of low SNR, and thereby reducing the likelihood of correctly detecting the number of signals [4]. Meanwhile, the computational cost of the reduced-rank

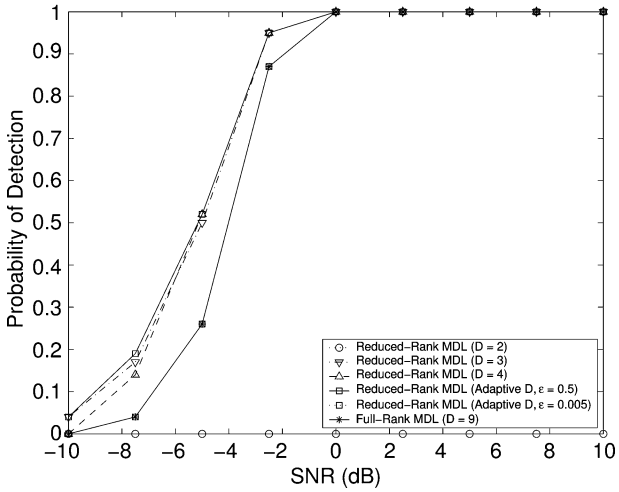


Fig. 2. Probability of detection of the proposed method for different ranks versus SNR. The number of sensors is 10 and $\mathrm{N} = 100$. SNR varies from -10 to 10 dB.

TABLE I
NUMBER OF SIGNALS DETECTED BY THE MDL (W & Z), THE PROPOSED MDL WITH ADAPTIVE D, AND THE PDL (V & K) METHODS FOR 100 INDEPENDENT RUNS

Detection method	k	SNR (dB)				
		-10	-5	0	5	10
MDL method (W & Z)	0	0	0	0	0	0
	1	100	92	0	0	0
	2	0	8	100	100	100
	3	0	0	0	0	0
Proposed method(ϵ=0.5)	0	62	0	0	0	0
	1	34	48	0	0	0
	2	4	52	100	100	100
	3	0	0	0	0	0
PDL method (V & K)	0	0	0	0	0	0
	1	83	2	0	0	0
	2	17	98	100	100	100
	3	0	0	0	0	0

MDL method will increase as D grows. As a result, ϵ should be a small constant close to 0.5.

One hundred independent trials have been run to find the number of times that the proposed reduced-rank MDL, the MDL (W & Z) [2] and the PDL (V & K) methods enumerate the signals. Table I shows the estimated number of signals. As shown in Table I, the proposed reduced-rank MDL method surpasses the traditional MDL approach, but is less accurate than the PDL method when $\mathrm{SNR} \leq -5$ dB. On the other hand, it should be noted that the PDL method requires the ML estimates of the DOAs for *each snapshot*, and is thereby much more computationally cumbersome than the MDL method (W & Z) that also involves the ML estimates of the DOAs and the EVD of the estimated covariance matrix. Thus, the reduced-rank MDL method outperforms the traditional MDL method and the PDL method in computational complexity.

Since the cross-correlation between the observation data and the reference signal is able to capture the signal information and efficiently eliminate the additive noise, when it is used as the initial information for the refinement procedure to find the signal and noise subspaces, the cleaner signal and noise subspace components can be obtained. As a result, the estimated variances of the first two desired signals $\hat{\sigma}^2_{d_i}(i = 1, 2)$, namely the powers of signals, are well separated from the estimated variances of the desired signals after the second stage $\hat{\sigma}^2_{d_i}(i = 3, 5, \ldots, 8)$, namely the powers of noises. Meanwhile, the variances of the noise subspace components (the desired signals after the second stage) are clustered sufficiently closely. Consequently, the proposed method is able to significantly reduce the likelihood of overestimating and underestimating the number of signals, eventually leading to the enhanced detection performance. Thus, the proposed MDL method is more accurate than the traditional MDL approaches [1], [2].

V. CONCLUSION

In this paper, we have proposed a reduced-rank MDL method for source enumeration in high-resolution array processing. The proposed method linearly partitions the sensor data into the signal and noise subspace components by means of the recursion procedure. The variances of the noise subspace components instead of the eigenvalues are exploited to calculate the total code length required to encode the observation data. Therefore, the proposed method does not involve the estimate of the covariance matrix or any eigendecomposition, giving it the advantage of computational simplicity. Meanwhile, notice that the cross-correlation between the observation data and the reference signal is able to capture the signal information and efficiently eliminate the additive noise. Since the cross-correlation is used as the initial information for the recursion procedure to find the signal and noise subspace components, the first p desired signals of the MSWF (the signal subspace components) are capable of capturing the signal information while suppressing a large portion of noise. On the other hand, their orthogonal components (the desired signals of the MSWF after the pth stage) become cleaner random noise. This indicates that the variances of the signal subspace components are well separated from the variances of the noise subspace components and the variances of the noise subspace components are clustered sufficiently closely. Consequently, the proposed method significantly reduces the likelihood of overestimating and underestimating the number of signals, thereby significantly improving the detection performance. Thus, the proposed MDL method surpasses the traditional MDL methods in detection performance and computational complexity, but is less accurate than the PDL method when SNR becomes low. The disadvantage of the proposed method can be balanced by the advantage of computational simplicity.

APPENDIX I
PROOF OF LEMMA 1

Since the signal covariance matrix $\boldsymbol{R}_s$ is nonsingular, the EVD of the covariance matrix $\boldsymbol{R}_{x_0}$ can be expressed

$$\boldsymbol{R}_{x_0} = \boldsymbol{V}_s \boldsymbol{\Lambda}_s \boldsymbol{V}_s^H + \sigma_n^2 \boldsymbol{V}_n \boldsymbol{V}_n^H \tag{41}$$

where $\boldsymbol{V}_s = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p]$, $\boldsymbol{V}_n = [\boldsymbol{v}_{p+1}, \boldsymbol{v}_{p+2}, \ldots, \boldsymbol{v}_q]$, $\boldsymbol{\Lambda}_s = \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\}$, $\lambda_1 > \cdots > \lambda_{p+1} = \cdots =$

$\lambda_q = \sigma_n^2$[1] are the eigenvalues and $\boldsymbol{v}_i (i = 1,2,\ldots,q)$ are the corresponding eigenvectors. Substituting $\bar{\boldsymbol{R}}_{x_0} = \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{D}\boldsymbol{R}_s\boldsymbol{D}^H\boldsymbol{A}_q^H(\boldsymbol{\theta}) + \sigma_n^2\boldsymbol{I}_q$ into (41), and performing some algebraic manipulations yields

$$\boldsymbol{V}_s = \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{Q} \tag{42}$$

where $\boldsymbol{Q} \in \mathbb{C}^{p\times p}$ is a full-rank matrix.

It is shown in [22] that the matched filters $\boldsymbol{h}_i (i = 1,2,\ldots,q)$ are orthogonal to each other. Actually, it follows from [21] and [22] that

$$\boldsymbol{h}_{i+1} = \frac{\left(\boldsymbol{I}_q - \boldsymbol{h}_i\boldsymbol{h}_i^H\right)\boldsymbol{R}_{x_{i-1}}\boldsymbol{h}_i}{\left\|\left(\boldsymbol{I}_q - \boldsymbol{h}_i\boldsymbol{h}_i^H\right)\boldsymbol{R}_{x_{i-1}}\boldsymbol{h}_i\right\|} \tag{43}$$

where

$$\boldsymbol{R}_{x_{i+1}} = \left(\boldsymbol{I}_q - \boldsymbol{h}_{i+1}\boldsymbol{h}_{i+1}^H\right)\boldsymbol{R}_{x_i}\left(\boldsymbol{I}_q - \boldsymbol{h}_{i+1}\boldsymbol{h}_{i+1}^H\right). \tag{44}$$

To prove the orthogonality of $\boldsymbol{h}_i (i = 1,2,\ldots,q)$, we employ the following induction argument. First, it is easy to verify from (43) that $\boldsymbol{h}_2$ is orthogonal to $\boldsymbol{h}_1$. Assume, now, that $\boldsymbol{h}_k$ is orthogonal to $\boldsymbol{h}_\ell$ for $k, \ell \le i$, $k \ne \ell$. Substituting (44) into (43) yields

$$\boldsymbol{h}_{i+1} = \frac{\left(\boldsymbol{I}_q - \sum_{k=i}^{1}\boldsymbol{h}_k\boldsymbol{h}_k^H\right)\boldsymbol{R}_{x_0}\boldsymbol{h}_i}{\left\|\left(\boldsymbol{I}_q - \sum_{k=i}^{1}\boldsymbol{h}_k\boldsymbol{h}_k^H\right)\boldsymbol{R}_{x_0}\boldsymbol{h}_i\right\|}. \tag{45}$$

It can then be obtained from (45) that $\boldsymbol{h}_{i+1}$ is orthogonal to $\boldsymbol{h}_k (k = 1,2,\ldots,i)$, namely $\boldsymbol{h}_k$ is orthogonal to $\boldsymbol{h}_\ell$ for $k, \ell \le i+1$, $k \ne \ell$. Therefore, $\boldsymbol{h}_i (i = 1,2,\ldots,q)$ are orthogonal to each other. In the sequel, it follows from [22] that the MSWF of rank p is completely equivalent to solving the Wiener–Hopf equation $\boldsymbol{R}_{x_0}\boldsymbol{w}_f = \boldsymbol{r}_{x_0 d_0}$ in the *Krylov* subspace $\mathcal{K}^{(p)}(\boldsymbol{R}_{x_0}, \boldsymbol{r}_{x_0 d_0}) = \operatorname{span}\{\boldsymbol{r}_{x_0 d_0}, \boldsymbol{R}_{x_0}\boldsymbol{r}_{x_0 d_0}, \ldots, \boldsymbol{R}_{x_0}^{(p-1)}\boldsymbol{r}_{x_0 d_0}\}$. Here $\boldsymbol{w}_f$ represents a linear filter. Consequently, the columns of $\boldsymbol{T}_s \triangleq [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p]$ form an orthogonal basis for the *Krylov* subspace, namely

$$\operatorname{span}\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p\} = \operatorname{span}\left\{\boldsymbol{r}_{x_0 d_0}, \boldsymbol{R}_{x_0}\boldsymbol{r}_{x_0 d_0}, \ldots, \boldsymbol{R}_{x_0}^{(p-1)}\boldsymbol{r}_{x_0 d_0}\right\}. \tag{46}$$

Employing the shift-invariance property of the *Krylov* subspace [24], we obtain

$$\begin{aligned}\mathcal{K}^{(p)}(\boldsymbol{R}_{x_0}, \boldsymbol{r}_{x_0 d_0}) &= \mathcal{K}^{(p)}\left(\boldsymbol{R}_{x_0} - \sigma_n^2\boldsymbol{I}_q, \mathbf{r}_{x_0 d_0}\right)\\ &\triangleq \mathcal{K}^{(p)}(\boldsymbol{R}_0, \boldsymbol{r}_{x_0 d_0})\end{aligned} \tag{47}$$

[1]Note that the largest eigenvalues are equal to each other provided that the incident signals have equal powers and their direction vectors are orthogonal to each other. However, this case takes place only when the angles-of-arrival (AOAs) of incident signals are not all in the beamwidth of the array. Since the work is focused on the high-resolution DOA estimations of signals, we are only interested in the AOAs of incident signals within the beamwidth of the array. Therefore, we do not consider the case of $\lambda_i = \lambda_j (i, j = 1,2,\ldots,p, i \ne j)$.

where $\boldsymbol{R}_0 = \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{D}\boldsymbol{R}_s\boldsymbol{D}^H\boldsymbol{A}_q^H(\boldsymbol{\theta})$ is the covariance matrix of noiseless sensor data. It follows that

$$\operatorname{span}\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p\} = \operatorname{span}\left\{\boldsymbol{r}_{x_0 d_0}, \boldsymbol{R}_0\boldsymbol{r}_{x_0 d_0}, \ldots, \boldsymbol{R}_0^{(p-1)}\boldsymbol{r}_{x_0 d_0}\right\}. \tag{48}$$

Therefore, there exists a full-rank matrix $\boldsymbol{K} \in \mathbb{C}^{p\times p}$ such that

$$\boldsymbol{T}_s = \left[\boldsymbol{r}_{x_0 d_0}, \boldsymbol{R}_0\boldsymbol{r}_{x_0 d_0}, \ldots, \boldsymbol{R}_0^{(p-1)}\boldsymbol{r}_{x_0 d_0}\right]\boldsymbol{K}. \tag{49}$$

Since

$$\begin{aligned}\boldsymbol{R}_0 &= \boldsymbol{R}_{x_0} - \sigma_n^2\boldsymbol{I}_q\\ &= \boldsymbol{V}_s\boldsymbol{\Lambda}_s\boldsymbol{V}_s^H + \sigma_n^2\boldsymbol{V}_n\boldsymbol{V}_n^H - \sigma_n^2\boldsymbol{I}_q\\ &= \boldsymbol{V}_s\left(\boldsymbol{\Lambda}_s - \sigma_n^2\boldsymbol{I}_p\right)\boldsymbol{V}_s^H\\ &\triangleq \boldsymbol{V}_s\tilde{\boldsymbol{\Lambda}}_s\boldsymbol{V}_s^H\end{aligned} \tag{50}$$

where $\tilde{\boldsymbol{\Lambda}}_s = \boldsymbol{\Lambda}_s - \sigma_n^2\boldsymbol{I}_p = \operatorname{diag}\{\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_p\}$ with $\tilde{\lambda}_i = \lambda_i - \sigma_n^2 (i = 1,2,\ldots,p)$, by substituting (50) into (49) and noticing that $\bar{\boldsymbol{V}}_s = \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{Q}$ and $\bar{\boldsymbol{V}}_n^H\boldsymbol{r}_{x_0 d_0} = \boldsymbol{0}$, we obtain

$$\begin{aligned}\boldsymbol{T}_s &= \Big[\left(\boldsymbol{V}_s\boldsymbol{V}_s^H + \boldsymbol{V}_n\boldsymbol{V}_n^H\right)\boldsymbol{r}_{x_0 d_0}, \boldsymbol{V}_s\tilde{\boldsymbol{\Lambda}}_s\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}, \ldots,\\ &\qquad \boldsymbol{V}_s\tilde{\boldsymbol{\Lambda}}_s^{(p-1)}\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}\Big]\boldsymbol{K}\\ &= \boldsymbol{V}_s\Big[\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}, \tilde{\boldsymbol{\Lambda}}_s\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}, \ldots, \tilde{\boldsymbol{\Lambda}}_s^{(p-1)}\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}\Big]\boldsymbol{K}\\ &= \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{Q}\Big[\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}, \tilde{\boldsymbol{\Lambda}}_s\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}, \ldots, \tilde{\boldsymbol{\Lambda}}_s^{(p-1)}\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}\Big]\boldsymbol{K}\\ &\triangleq \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{K}\\ &\triangleq \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{H}\end{aligned} \tag{51}$$

where

$$\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{K} \tag{52}$$

$$\begin{aligned}\boldsymbol{\Gamma} &= \left[\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}, \tilde{\boldsymbol{\Lambda}}_s\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}, \cdots, \tilde{\boldsymbol{\Lambda}}_s^{(p-1)}\boldsymbol{V}_s^H\boldsymbol{r}_{x_0 d_0}\right]\\ &= \begin{pmatrix}\boldsymbol{v}_1^H\boldsymbol{r}_{x_0 d_0} & & \boldsymbol{O}\\ & \boldsymbol{v}_2^H\boldsymbol{r}_{x_0 d_0} & \\ & & \ddots\\ \boldsymbol{O} & & \boldsymbol{v}_p^H\boldsymbol{r}_{x_0 d_0}\end{pmatrix}\\ &\quad\times\begin{pmatrix}1 & \tilde{\lambda}_1 & \cdots & \tilde{\lambda}_1^{p-1}\\ 1 & \tilde{\lambda}_2 & \cdots & \tilde{\lambda}_2^{p-1}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & \tilde{\lambda}_p & \cdots & \tilde{\lambda}_p^{p-1}\end{pmatrix}\\ &\triangleq \boldsymbol{E}\boldsymbol{F}.\end{aligned} \tag{53}$$

Since $\boldsymbol{R}_s$ is a nonsingular and diagonal matrix $\boldsymbol{v}_i^H\boldsymbol{r}_{x_0 d_0} = \boldsymbol{v}_i^H\boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{D}\boldsymbol{R}_s\boldsymbol{1} \ne 0 (i = 1,2,\ldots,q)$. It follows that the diagonal matrix $\boldsymbol{E}$ is full rank. Meanwhile, $\boldsymbol{F}$ is a Vandermonde matrix due to $\tilde{\lambda}_i \ne \tilde{\lambda}_j (i, j = 1,2,\ldots,p, i \ne j)$. It follows that $\boldsymbol{\Gamma}$ is also full rank. In the sequel, from (52) we can obtain that $\boldsymbol{H}$ is a full-rank matrix. Thus, the columns of $\boldsymbol{T}_s$ span the same range subspace of $\boldsymbol{A}_q(\theta)$, namely the signal subspace. Since all the matched filters $\boldsymbol{h}_i (i = 1,2,\ldots,q)$ are orthogonal

to each other, the columns of $\boldsymbol{T}_n \triangleq [\boldsymbol{h}_{p+1}, \boldsymbol{h}_{p+2}, \ldots, \boldsymbol{h}_q]$ span the noise subspace. This completes the proof of Lemma 1.

Appendix II
Proof of Proposition 1

It follows from (11) that $\boldsymbol{x}_{i-1}(t) = (\prod_{k=i-1}^{1} \boldsymbol{B}_k)\boldsymbol{x}_0(t)$. Noticing that $\boldsymbol{h}_i^H \boldsymbol{h}_j = 0 (i \neq j)$, we obtain

$$\begin{aligned}
E\left[d_i(t)d_j^*(t)\right] &= E\left[\boldsymbol{h}_i^H \boldsymbol{x}_{i-1}(t)\boldsymbol{x}_{j-1}^H(t)\boldsymbol{h}_j\right] \\
&= \boldsymbol{h}_i^H E\left[\boldsymbol{x}_{i-1}(t)\boldsymbol{x}_{j-1}^H(t)\right]\boldsymbol{h}_j \\
&= \boldsymbol{h}_i^H E\left\{\left(\prod_{k=i-1}^{1} \boldsymbol{B}_k\right)\boldsymbol{x}_0(t)\boldsymbol{x}_0^H(t) \right. \\
&\qquad \left. \times \left(\prod_{\ell=j-1}^{1} \boldsymbol{B}_\ell\right)^H\right\}\boldsymbol{h}_j \\
&= \boldsymbol{h}_i^H \left(\prod_{k=i-1}^{1} \boldsymbol{B}_k\right)\boldsymbol{R}_{x_0}\left(\prod_{\ell=j-1}^{1} \boldsymbol{B}_\ell\right)^H \boldsymbol{h}_j \\
&= \boldsymbol{h}_i^H \left(\boldsymbol{I}_q - \sum_{k=i-1}^{1} \boldsymbol{h}_k\boldsymbol{h}_k^H\right) \\
&\quad \times \left(\boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}_q^H(\boldsymbol{\theta}) + \sigma_n^2\boldsymbol{I}_q\right) \\
&\quad \times \left(\boldsymbol{I}_q - \sum_{\ell=j-1}^{1} \boldsymbol{h}_\ell\boldsymbol{h}_\ell^H\right)^H \boldsymbol{h}_j \\
&= \boldsymbol{h}_i^H \boldsymbol{A}_q(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}_q^H(\boldsymbol{\theta})\boldsymbol{h}_j + \sigma_n^2\boldsymbol{h}_i^H\boldsymbol{h}_j. \qquad (54)
\end{aligned}$$

From (14), it follows that $\boldsymbol{h}_i^H \boldsymbol{A}_q(\boldsymbol{\theta}) = \boldsymbol{0} (i = p+1, p+2, \ldots, q)$. Thus, considering $\boldsymbol{h}_i^H \boldsymbol{h}_j = 1 (i = j)$, we obtain from (54)

$$\delta_{i,j} \triangleq E\left[d_i(t)d_j^*(t)\right] = 0 (i, j = p+1, p+2, \ldots, q, i \neq j) \quad (55)$$

$$\sigma_{d_i}^2 \triangleq E\left[d_i(t)d_i^*(t)\right] = \sigma_n^2 (i = p+1, p+2, \ldots, q). \quad (56)$$

This proves Proposition 1.

Appendix III
Proof of the Consistency of the Reduced-Rank MDL Criterion

It follows from (36) that

$$\begin{aligned}
&\frac{1}{N}\left(\mathrm{MDLE}^{(D)}(k) - \mathrm{MDLE}^{(D)}(p)\right) \\
&\qquad = \gamma(k) - \gamma(p) + \frac{1}{2}(k-p)(2D-k-p)\frac{\log N}{N} \qquad (57)
\end{aligned}$$

where

$$\gamma(k) = (D-k)\log\frac{\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\sigma}_{d_i}^2}{\left(\prod_{i=k+1}^{D}\hat{\sigma}_{d_i}^2\right)^{1/(D-k)}}. \qquad (58)$$

It is easy to obtain from (58) that for $k = p$

$$\gamma(p) = \log\frac{\left(\frac{1}{D-p}\sum_{i=p+1}^{D}\hat{\sigma}_{d_i}^2\right)^{D-p}}{\prod_{i=p+1}^{D}\hat{\sigma}_{d_i}^2} = 0 \text{ a.s. as } N \to \infty. \qquad (59)$$

Here, we use the standard abbreviation "a.s." for "almost sure" to describe an event occurring with probability one.

We first consider the case $k < p$. Since $\hat{\sigma}_{d_i}^2 (i = k+1, k+2, \ldots, D)$ are not all equal with probability one when N tends to infinity, we obtain by the inequality between the arithmetic and geometric means

$$\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\sigma}_{d_i}^2 > \left(\prod_{i=k+1}^{D}\hat{\sigma}_{d_i}^2\right)^{\frac{1}{D-k}} \text{ a.s. as } N \to \infty. \quad (60)$$

As a result, it follows from (58) that as the number of snapshots grows to infinity

$$\gamma(k) > 0 \text{ a.s. as } N \to \infty. \quad (61)$$

Inserting (59) and (61) into (57), and considering $\log N/N$ approaches zero as N increases, we obtain for $k < p$

$$\mathrm{MDLE}^{(D)}(k) > \mathrm{MDLE}^{(D)}(p) \text{ a.s. as } N \to \infty. \quad (62)$$

Consider now the case $k > p$. Noticing that $log(1+x) = x - x^2/2 + x^3/3 - \cdots$ for a small number x and $\hat{\sigma}_{d_i}^2 - \sigma_n^2 = O(\sqrt{\log\log N/N})$ $(i = p+1, p+2, \ldots, D)$, we obtain

$$\begin{aligned}
\gamma(k) &= (D-k)\log\left(\frac{1}{D-k}\sum_{i=k+1}^{D}\hat{\sigma}_{d_i}^2\right) - \sum_{i=k+1}^{D}\log\hat{\sigma}_{d_i}^2 \\
&= (D-k)\log\left(\sigma_n^2\left(1 + \frac{\sum_{i=k+1}^{D}\left(\hat{\sigma}_{d_i}^2 - \sigma_n^2\right)}{(D-k)\sigma_n^2}\right)\right) \\
&\quad - \sum_{i=k+1}^{D}\log\left(\sigma_n^2\left(1 + \frac{\hat{\sigma}_{d_i}^2 - \sigma_n^2}{\sigma_n^2}\right)\right) \\
&= (D-k)\left\{\frac{\sum_{i=k+1}^{D}\left(\hat{\sigma}_{d_i}^2 - \sigma_n^2\right)}{(D-k)\sigma_n^2} \right. \\
&\qquad - \frac{1}{2}\left(\frac{\sum_{i=k+1}^{D}\left(\hat{\sigma}_{d_i}^2 - \sigma_n^2\right)}{(D-k)\sigma_n^2}\right)^2 \\
&\qquad \left. + O\left((\log\log N/N)^{3/2}\right)\right\} \\
&\quad - \sum_{i=k+1}^{D}\left(\frac{\hat{\sigma}_{d_i}^2 - \sigma_n^2}{\sigma_n^2} - \frac{1}{2}\left(\frac{\hat{\sigma}_{d_i}^2 - \sigma_n^2}{\sigma_n^2}\right)^2 \right. \\
&\qquad \left. + O\left((\log\log N/N)^{3/2}\right)\right) \\
&= \frac{1}{2}\left(\sum_{i=k+1}^{D}\left(\frac{\hat{\sigma}_{d_i}^2 - \sigma_n^2}{\sigma_n^2}\right)^2 \right. \\
&\qquad \left. - \frac{1}{D-k}\left(\sum_{i=k+1}^{D}\frac{\hat{\sigma}_{d_i}^2 - \sigma_n^2}{\sigma_n^2}\right)^2\right) \\
&\quad + O\left(\left(\frac{\log\log N}{N}\right)^{\frac{3}{2}}\right) \\
&= O(\log\log N/N)(k = p+1, \cdots, D-1), \\
&\qquad \text{a.s. as } N \to \infty. \qquad (63)
\end{aligned}$$

Substituting (59) and (63) into (57) and recalling that as $N \to \infty$, $\log N/\log\log N \to \infty$, we obtain for $k > p$

$$\begin{aligned}\frac{1}{N}\left(\mathrm{MDLE}^{(D)}(k) - \mathrm{MDLE}^{(D)}(p)\right) &= O(\log\log N/N) + \frac{1}{2}(k-p)(2D-k-p)\log N/N \\ &> 0 \text{ a.s. as } N \to \infty. \end{aligned} \tag{64}$$

It follows from (64) that for $k > p$

$$\mathrm{MDLE}^{(D)}(k) > \mathrm{MDLE}^{(D)}(p) \text{ a.s. as } N \to \infty. \tag{65}$$

This completes the proof.

ACKNOWLEDGMENT

The authors would like to thank Dr. S. Valaee for providing the codes of the PDL algorithm. They would also like to thank the anonymous referees and the associate editor, Prof. A. R. Leyman, for careful review, insightful comments, and valuable suggestions.

REFERENCES

[1] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
[2] 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.
[3] 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 signals in array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1157–1167, Oct. 1989.
[4] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, Aug. 2001.
[5] A. Di, "Multiple source location-A matrix decomposition approach," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 10, pp. 1086–1091, Oct. 1985.
[6] T. J. Shan, A. Paulraj, and T. Kailath, "On soomthed rank profile tests in eigenstructrue methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1377–1385, Oct. 1987.
[7] J. H. Cozzens and M. J. Sousa, "Source enumeration in a correlated signal environment," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 304–317, Feb. 1994.
[8] G. Xu, R. H. Roy, III, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 102–112, Jan. 1994.
[9] C. W. Ma and C. C. Teng, "Detection of coherent signals using weighted subspace smoothing," *IEEE Trans. Ant. Propag.*, vol. 44, no. 2, pp. 179–187, Feb. 1996.
[10] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, May 2004.
[11] M. Nezafat, M. Kaveh, and W. Xu, "Estimation of the number of sources based on the eigenvectors of the covariance matrix," in *Proc. Sensor Array Multichannel Signal Process. Workshop (SAM)*, 2004, pp. 465–469.
[12] Y. Ishikawa, K. Ichige, and H. Arai, "Accurate detection of the number of arrival signals using pre-estimated signal subspace," in *Proc. Signal Process. Adv. Wirel. Commun. Workshop (SPAWC)*, 2005, pp. 840–844.
[13] Q. Wu and D. R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 39, no. 8, pp. 1848–1857, Aug. 1991.
[14] L. Huang and S. Wu, "Low-complexity MDL method for accurate source enumeration," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 581–584, Sep. 2007.
[15] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, no. 1, pp. 1–25, 1986.
[16] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[17] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
[18] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 8, pp. 806–811, Aug. 1985.
[19] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 1, pp. 8–15, Jan. 1989.
[20] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.
[21] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.
[22] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1928–1946, Jul. 2001.
[23] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. J. Math. Stat.*, vol. 34, pp. 122–148, 1963.
[24] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Dept. Elect. Eng., Univ. Illinois, Urbana-Champaign, 1997.

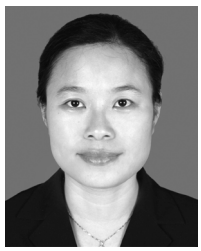
Lei Huang (M'07) was born in Guangdong, China. He received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

Since September 2006, he has been with the College of Information Engineering, Shenzhen University, Shenzhen, China. From 2005 to 2006, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Duke University, Durham, NC. He has worked in the areas of wireless communications, radar and remote sensing. His current research interests are focused on adaptive signal processing, statistical signal processing, sensor array signal processing, digital signal processing, and smart antenna.

Shunjun Wu (M'07) was born in Shanghai, China, on February 18, 1942. He received the Diploma degree from Xidian University, Xi'an, China, in 1964.

After graduation, he joined the Department of Electrical Engineering, Xidian University, where he is currently a Professor. From 1981 to 1983, he was a Visiting Scholar with the Department of Electrical Engineering, University of Hawaii, Manoa. His research interests include digital signal processing, array processing and radar signal processing.

Mr. Wu is a Fellow of the Chinese Institute of Electronics (CIE).



Xia Li (M'07) was born in Sichuan, China. She received the B.Sc. and M.Sc., degrees in electronic engineering from Xidian University, Xi'an, China, in 1989 and 1992, respectively, and the Ph.D. degree in information engineering from the Chinese University of Hong Kong, Hong Kong, in 1997.

Since 1997, she has been with the College of Information Engineering, Shenzhen University, Shenzhen, China, where she is currently a Professor. She has worked in the areas of intelligent signal processing, digital communications and image processing. Her current research interests are focused on evolutionary computation, swarm intelligence and image processing.