Low-Complexity MDL Method for Accurate Source Enumeration

Lei Huang and Shunjun Wu, *Member, IEEE*

***Abstract*—A low-complexity method for source enumeration is proposed in this letter. Given the training data of a desired signal, an array data matrix is partitioned into orthogonal signal and noise components. The noise components are then used to calculate the total description length required to encode the array data. The model with the minimum description length (MDL) is chosen as the best model. Unlike the traditional MDL methods, the proposed method linearly partitions the array data into the cleaner signal and noise components and thereby is more accurate and computationally efficient. Its performance is demonstrated via numerical results.**

***Index Terms*—Array signal processing, direction of arrival (DOA), eigenvalue decomposition (EVD), minimum description length (MDL), multistage Wiener filter (MSWF), Wiener filter.**

I. Introduction

SENSOR array signal processing can be applied to many areas, such as radar, sonar, remote sensing, and wireless communications. In most applications, one important objective might be to detect the number of signal sources. In the community of array signal processing, this problem has been widely investigated in [1]–[5]. In [1], Wax and Kailath originally introduced the information-theoretic criteria, such as the Akaike information criterion (AIC) and minimum description length (MDL), into source enumeration. To date, the methods for source enumeration have been widely studied in [2]–[5]. However, while these methods are efficient in detection performance, they essentially involve the estimation of a covariance matrix and its eigenvalue decomposition [1]–[5] or multidimensional nonlinear minimization problems [2], [3] and are thereby generally rather computationally intensive.

To reduce the computational complexity and attain accurate detection performance, a low-complexity MDL method is developed in this letter. This method first employs the training data of the desired signal to quickly partition the array data into two orthogonal components in signal and noise subspaces. Similar to the method proposed in [2], the components in the noise subspace are then used to calculate the total code length required to encode the array data. Finally, the model with the shortest code length is selected as the best model. Since the method only involves the forward recursions of the multistage Wiener filter (MSWF) [6] to find the signal and noise components and does not calculate the covariance matrix or its eigenvalues, it is computationally efficient. Meanwhile, the proposed method exploits the training data of the desired signal to partition the array data into the cleaner signal and noise components, thus obtaining a significant enhancement in performance, particularly for the case of small sample size and/or low signal-to-noise ratio (SNR).

Manuscript received June 5, 2006; revised August 24, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yimin Zhang.

L. Huang is with the College of Information Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China (e-mail: huanglei8rsp@yahoo.com.cn).

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

Digital Object Identifier 10.1109/LSP.2006.885286

II. Preliminaries

A. Data Model

Consider an array consisting of M isotropic sensors with arbitrary locations. This array received a desired signal $s_1(t)$ with a known training data from an *unknown* direction, say, θ_1, and $P-1 (P<M)$ jammer signals $\{s_2(t), s_3(t), \ldots, s_P(t)\}$ from *unknown* distinct directions $\{\theta_2, \theta_3, \ldots, \theta_P\}$ that are different from θ_1. The P narrowband 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 $M \times 1$ measurements of the output of the array corrupted by additive noise can be expressed as

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

where $\boldsymbol{x}(t), \boldsymbol{n}(t)$, and $\boldsymbol{s}(t)$ are the vectors of the received signals, the additive noise, and the signal waveforms, respectively, and

$$\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_P)] \tag{2}$$

is the array response matrix with $\boldsymbol{a}(\theta_i) = [1, e^{-j\omega_0\tau_1(\theta_i)}, \ldots, e^{-j\omega_0\tau_{M-1}(\theta_i)}]^T$, where $\tau_k(\theta_i) (k = 1, 2, \ldots, M-1)$ is the propagation delay between the first sensor (the reference point) and the $(k+1)$th sensor to a wavefront impinging from direction θ_i, and $(\cdot)^T$ denotes transpose.

Throughout this letter, we assume that the array response matrix $\boldsymbol{A}(\boldsymbol{\theta})$ is unambiguous. Meanwhile, the additive noise is a stationary, temporally, and spatially white Gaussian random process with zero mean and variance σ_n^2. Furthermore, all signals are jointly stationary, temporally white, zero-mean complex Gaussian random processes, which are uncorrelated with the background noise $\boldsymbol{n}(t)$. Under these assumptions, the output of the array is complex Gaussian with zero mean and the following covariance matrix:

$$\boldsymbol{R}_x = E\left[\boldsymbol{x}(t)\boldsymbol{x}^H(t)\right] = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}^H(\boldsymbol{\theta}) + \sigma_n^2\boldsymbol{I}_M \tag{3}$$

where $\boldsymbol{R}_s = E[\boldsymbol{s}(t)\boldsymbol{s}^H(t)]$, $\boldsymbol{I}_M$ denotes the $M \times M$ identity matrix, and $(\cdot)^H$ is Hermitian transpose. In actual applications, however, we do not have access to the ideal covariance matrix.

Instead, we in general only obtain its sample covariance matrix $\hat{\boldsymbol{R}}_x = 1/N \sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}^H(t)$, where N is finite.

B. Common 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 $\boldsymbol{X} = \{\boldsymbol{x}(t)\}_{t=1}^{N}$ 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]

$$L\{\boldsymbol{x}(t)\} = -\log f(\boldsymbol{X}|\hat{\boldsymbol{\mu}}) + \frac{1}{2}K \log N \tag{4}$$

where $\hat{\boldsymbol{\mu}}$ is the maximum likelihood estimate of $\boldsymbol{\mu}$, and K denotes the number of free parameters in the vector $\boldsymbol{\mu}$. To implement the MDL principal, Wax and Ziskind partitioned the observation data into two orthogonal components in the signal and noise subspaces, separately calculated the code lengths for the two orthogonal components, and finally added them up to attain the total code length for the data. The total code length of the signal and noise components for k unknown direction parameters $\boldsymbol{\theta}^{(k)}$ derived by Wax and Ziskind [2] can be expressed as

$$\begin{aligned} &L\{\boldsymbol{x}_s^{(k)}(t), \boldsymbol{x}_n^{(k)}(t)\} \\ &\quad = N(M-k)\log \frac{\frac{1}{M-k}\mathrm{tr}\hat{\boldsymbol{R}}_{nn}(\boldsymbol{\theta}^{(k)})}{\left|\hat{\boldsymbol{R}}_{nn}(\boldsymbol{\theta}^{(k)})\right|^{1/(M-k)}} \\ &\qquad + \frac{1}{2}k(2M-k+1)\log N \\ &\quad = N(M-k)\log \frac{\frac{1}{M-k}\sum_{i=1}^{M-k}\hat{\lambda}_i(\boldsymbol{\theta}^{(k)})}{\left(\prod_{i=1}^{M-k}\hat{\lambda}_i(\boldsymbol{\theta}^{(k)})\right)^{1/(M-k)}} \\ &\qquad + \frac{1}{2}k(2M-k+1)\log N \end{aligned} \tag{5}$$

where $\boldsymbol{x}_s^{(k)}(t)$ is the $k \times 1$ component in the signal subspace, $\boldsymbol{x}_n^{(k)}(t)$ is the $(M-k)\times 1$ component in the noise subspace, $\hat{\boldsymbol{R}}_{nn}(\boldsymbol{\theta}^{(k)}) = 1/N\sum_{t=1}^{N}\boldsymbol{x}_n^{(k)}(t)\boldsymbol{x}_n^{(k),H}(t)$, k is the assumed number of signals, $\hat{\lambda}_1(\boldsymbol{\theta}^{(k)}) \geq \cdots \geq \hat{\lambda}_{M-k}(\boldsymbol{\theta}^{(k)})$ denote the nonzero eigenvalues of the $M \times M$ matrix $\boldsymbol{P}_{\boldsymbol{A}}^{\perp}(\theta^{(k)})\hat{\boldsymbol{R}}_x\boldsymbol{P}_{\boldsymbol{A}}^{\perp}(\theta^{(k)})$, and $\boldsymbol{P}_{\boldsymbol{A}(\theta^{(k)})}^{\perp}$ is the projection matrix on the noise subspace. To obtain the two orthogonal components in the signal and noise subspaces, Wax and Ziskind had to solve a multidimensional nonlinear minimization problem to find the directions $\boldsymbol{\theta}^{(k)}$, using them to formulate the projection matrices on the signal and noise subspaces: $\boldsymbol{P}_{\boldsymbol{A}(\theta^{(k)})} = \boldsymbol{A}(\boldsymbol{\theta}^{(k)})[\boldsymbol{A}^H(\boldsymbol{\theta}^{(k)})\boldsymbol{A}(\boldsymbol{\theta}^{(k)})]^{-1}\boldsymbol{A}^H(\boldsymbol{\theta}^{(k)})$, $\boldsymbol{P}_{\boldsymbol{A}(\theta^{(k)})}^{\perp} = \boldsymbol{I}_M - \boldsymbol{P}_{\boldsymbol{A}(\theta^{(k)})}$. By projecting the array data onto the signal and noise subspaces, one may obtain the signal and noise components. However, this procedure is rather computationally complex, particularly for solving the multidimensional nonlinear minimization problem. While using the alternating projection (AP) algorithm may reduce the computational cost, the MDL method essentially involves the estimated covariance matrix and its eigendecomposition, which are still computationally expensive.

In this letter, we propose an alternative method to find the two orthogonal components in the signal and noise subspaces. This method is much more computationally efficient than the traditional MDL methods due to avoidance of the eigendecomposition of the covariance matrix and the multidimensional nonlinear minimization problem. In the mean time, the method can be used in the case of small sample size and/or low SNR since it uses the training data of the desired signal to partition the array data into the cleaner signal and noise components that result in the improved performance.

III. Low-Complexity MDL Criterion for Source Enumeration

A. Subspace Decomposition

Given a reference signal, the MSWF [6] partitions the observation data into two directions: the desired signal $d_i(t) = \boldsymbol{h}_i^H\boldsymbol{x}_{i-1}(t)$ and its orthogonal component $\boldsymbol{x}_i(t) = \boldsymbol{B}_i\boldsymbol{x}_{i-1}(t)$ at the ith stage. The desired signal $d_i(t)$ is obtained by pre-filtering the observation data $\boldsymbol{x}_{i-1}(t)$ with the matched filter $\boldsymbol{h}_i$ but annihilated by the blocking matrix $\boldsymbol{B}_i = \boldsymbol{I} - \boldsymbol{h}_i\boldsymbol{h}_i^H$. The observation data are partitioned stage by stage in the same manner. As a result, we obtain the pre-filtering matrix: $\boldsymbol{T}_M = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_M]$ and the desired signals of the MSWF: $d_1(t), d_2(t), \ldots, d_M(t)$. The MSWF based on the data-level lattice structure is given by the following set of recursions.

- *Initialization*: $d_0(t)$ and $\boldsymbol{x}_0(t) = \boldsymbol{x}(t)$.
- *Forward Recursion*: For $i = 1, 2, \ldots, M$

$$\begin{aligned} \boldsymbol{h}_i &= E[\boldsymbol{x}_{i-1}(t)d_{i-1}^*(t)]/\|E[\boldsymbol{x}_{i-1}(t)d_{i-1}^*(t)]\|_2 \\ d_i(t) &= \boldsymbol{h}_i^H\boldsymbol{x}_{i-1}(t); \\ \boldsymbol{x}_i(t) &= \boldsymbol{x}_{i-1}(t) - \boldsymbol{h}_i d_i(t). \end{aligned}$$

- *Backward Recursion*: For $i = M, M-1, \ldots, 1$ with $e_M(t) = d_M(t)$

$$\begin{aligned} w_i &= E[d_{i-1}(t)e_i^*(t)]/E[|e_i(t)|^2]; \\ e_{i-1}(t) &= d_{i-1}(t) - w_i^* e_i(t). \end{aligned}$$

In the algorithm above, the reference signal $d_0(t)$ is the training data of the desired signal, which is added to the message for training purposes [7]. The observation here is that the multistage decomposition gives a computationally simple method for estimating the signal and noise subspaces. As noted in [8], for P uncorrelated narrowband signals impinging upon the array, the signal and noise subspaces can be spanned by the M orthonormal matched filters of the MSWF, namely,

$$\mathcal{S}^{(P)} = \mathrm{span}\,\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_P\} \tag{6}$$

$$\mathcal{N}^{(M-P)} = \mathrm{span}\,\{\boldsymbol{h}_{P+1}, \boldsymbol{h}_{P+2}, \ldots, \boldsymbol{h}_M\} \tag{7}$$

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

B. Novel MDL Criterion Function for Source Enumeration

From (6) and (7), it follows that $\boldsymbol{T}_s^{(k)} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_k]$ and $\boldsymbol{T}_n^{(k)} = [\boldsymbol{h}_{k+1}, \boldsymbol{h}_{k+2}, \ldots, \boldsymbol{h}_M]$ span the signal and noise subspaces, respectively. Here, k is the assumed number of signals.

In the sequel, the component in the noise subspace can be calculated as

$$\boldsymbol{x}_n^{(k)}(t) = \boldsymbol{T}_n^{(k),H}\boldsymbol{x}(t) = [d_{k+1}(t), d_{k+2}(t), \ldots, d_M(t)]^T. \quad (8)$$

Accordingly, the covariance matrix of the noise component can be written as

$$\boldsymbol{R}_{nn}^{(k)} = E\left[\boldsymbol{x}_n^{(k)}(t)\boldsymbol{x}_n^{(k),H}(t)\right]. \quad (9)$$

By substituting the estimate of (9) into (5) and noticing that the k direction parameters $\boldsymbol{\theta}^{(k)}$ are not included in the parameter vector, we can obtain a new MDL criterion for source enumeration

$$L\{\boldsymbol{x}_s^{(k)}(t), \boldsymbol{x}_n^{(k)}(t)\} = N(M-k)\log\frac{\frac{1}{M-k}\mathrm{tr}\hat{\boldsymbol{R}}_{nn}^{(k)}}{|\hat{\boldsymbol{R}}_{nn}^{(k)}|^{1/(M-k)}} + \frac{1}{2}k(2M-k)\log N \quad (10)$$

where $\hat{\boldsymbol{R}}_{nn}^{(k)} = 1/N\sum_{t=1}^{N}\boldsymbol{x}_n^{(k)}(t)\boldsymbol{x}_n^{(k),H}(t)$, which is independent of the direction estimates $\hat{\theta}$.

Nevertheless, the calculation of $|\hat{\boldsymbol{R}}_{nn}^{(k)}|$ indicates that the estimation of the covariance matrix $\hat{\boldsymbol{R}}_{nn}^{(k)}$ and the additional computational cost of $O((M-k)^3)$ flops are required, besides calculating the signal and noise subspaces. Actually, the desired signals of the MSWF after the kth stage $d_i(t)(i = k+1, k+2, \ldots, M)$ are uncorrelated with each other, namely, $\delta_{i,j} \triangleq E[d_i(t)d_j^*(t)] = 0$, for$i \neq j, i, j = k+1, k+2, \ldots, M$. The proof may be found in [9]. As a result, $\boldsymbol{R}_{nn}^{(k)}$ is a diagonal matrix:

$$\boldsymbol{R}_{nn}^{(k)} = \mathrm{diag}\left(\sigma_{d_{k+1}}^2, \sigma_{d_{k+2}}^2, \ldots, \sigma_{d_M}^2\right) \quad (11)$$

where $\sigma_{d_i}^2 = E[d_i(t)d_i^*(t)](i = k+1, k+2, \ldots, M)$. Thus, substituting the estimate of $\boldsymbol{R}_{nn}^{(k)}$ into (10), we eventually obtain the new MDL estimator for the number of sources P

$$\hat{P} = \arg\min_{k=0,1,\ldots,M-1}\mathrm{MDLE}(k) \quad (12)$$

where

$$\mathrm{MDLE}(k) = N(M-k)\log\frac{\frac{1}{M-k}\sum_{i=k+1}^{M}\hat{\sigma}_{d_i}^2}{\left(\prod_{i=k+1}^{M}\hat{\sigma}_{d_i}^2\right)^{1/(M-k)}} + \frac{1}{2}k(2M-k)\log N \quad (13)$$

with $\hat{\sigma}_{d_i}^2 = 1/N\sum_{t=1}^{N}d_i(t)d_i^*(t)$, and $d_i(t)$ is the ith desired signal of the MSWF. That is to say, the estimated variances of the desired signals $\hat{\sigma}_{d_i}^2$ can be directly calculated by the forward recursions of the MSWF while attaining the signal and noise subspaces.

Remarks: The traditional methods for source enumeration generally involve the estimated covariance matrix and its eigenvalue decomposition, therefore requiring $O(M^2N) + O(M^3)$ flops. Note that for each forward recursion of the MSWF, the dominant computational cost is the calculation of the matched filter, which requires M complex multiplications and $M-1$ additions for each snapshots, equivalently approximately M flops, and thereby around $O(MN)$ flops for each matched filter. Since M desired signals are used in the new MDL criterion, the proposed MDL method only requires a computational complexity of $O(M^2N)$, which is equivalent to that of calculating the sample covariance matrix, saving the computational complexity of $O(M^3)$. Meanwhile, the proposed method employs the training data of the desired signal to partition the array data into the cleaner signal and noise components and can thereby be used in the case of small sample size and/or low SNR where the traditional methods may fail. Thus, the proposed MDL method outperforms the traditional MDL methods in computational complexity for a large array and in detection performance for the case of small sample size and/or low SNR.

IV. Numerical Results

In this section, the performance of the proposed MDL method is evaluated by computer simulation. For comparison purpose, the results of the traditional MDL method developed by Wax and Ziskind [2] and the method proposed by Ishikawa *et al.* [5] are also given. In the method of Ishikawa *et al.*, the beamformer threshold and the test function threshold are -10 dB and 0.5, respectively. The dimension of the pre-estimated signal subspace is selected based on the eigenvalues of the pre-estimated covariance matrix. For simplicity, the array herein is assumed to be a ULA with eight isotropic sensors whose spacings equal half-wavelength. Suppose that there are three uncorrelated signals with equal power impinging upon the ULA. The true DOAs are $\{3°, 9°, 18°\}$. The background noise is assumed to be a stationary Gaussian white random process that is uncorrelated with the signals.

One hundred independent trials have been made to compute the probabilities of detection for the proposed MDL approach, the traditional MDL method (W & Z), and the method proposed by Ishikawa *et al.* The probability of detection versus SNR is shown in Fig. 1, where $\mathrm{N} = 80$. From Fig. 1, we can observe that the MDL approach with the maximum likelihood (ML) estimates of the DOAs nearly fails to correctly detect the number of signals when SNR is lower than 2 dB. Its probability of detection is only around 0.65 for SNR $= 2$ dB and less than 0.1 for SNR ≤ 0 dB. The proposed MDL method, however, is capable of yielding an accurate detection of the number of signals. Its probability of detection is about 1 for SNR $= -2$ dB and around 0.75 for SNR $= -4$ dB, where the traditional MDL method completely fails. It is interesting to observe that the result of the proposed MDL method is quite close to that of the MDL algorithm (W & Z), where the true DOA information is assumed to be exactly known. From Fig. 1, we can also observe that the proposed method outperforms the method developed by Ishikawa *et al.*, which uses the bearing information to improve the detection performance.

Fig. 2 displays the probability of detection versus the number of snapshots, where SNR $= -2$ dB. From Fig. 2, we can observe that the proposed MDL approach is able to obtain a high probability for detecting the number of signals, even for a quite small sample size. Its probability of detection essentially reaches 1 for $\mathrm{N} = 160$ and is around 0.98 even for $\mathrm{N} = 80$. However, the MDL method (W & Z) with the ML estimates of the DOAs only has the probability of detection of about 0.2 for $\mathrm{N} = 320$ and fails to correctly detect the number of signals when $\mathrm{N} \leq 320$. As N becomes large, the probabilities of the traditional and proposed MDL methods approach 1. Again,

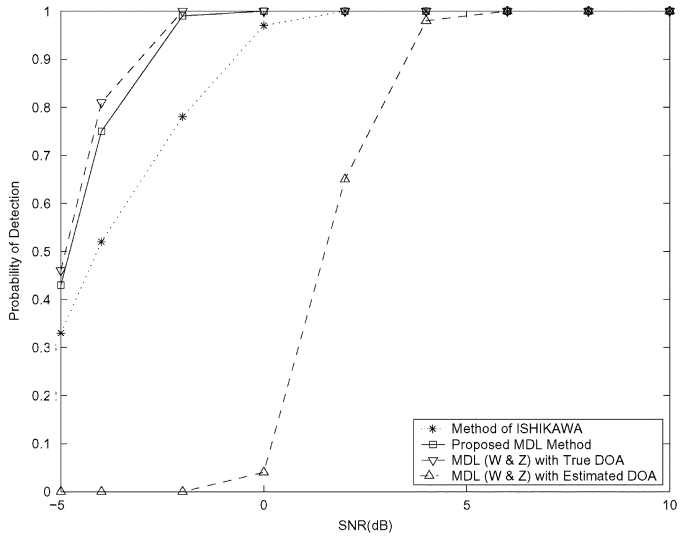

Fig. 1. Probability of detection versus SNR. The number of sensors is eight and $\mathrm{N} = 80$.

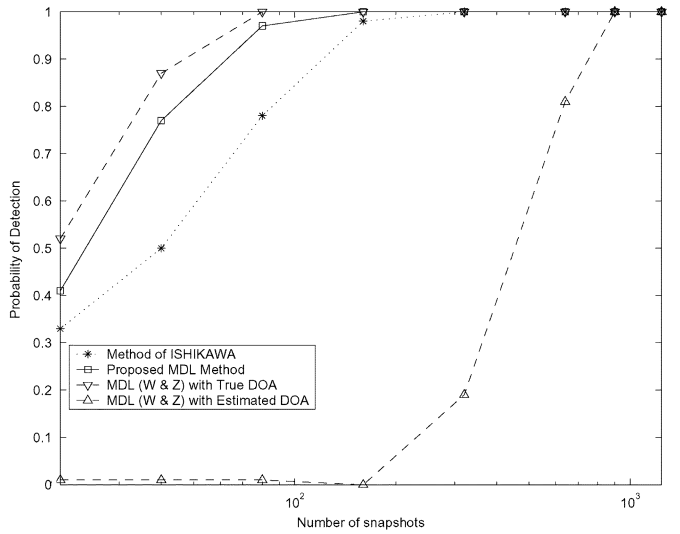

Fig. 2. Probability of detection versus number of snapshots. The number of sensors is eight, and SNR equals -2 dB.

the proposed MDL algorithm is close to the traditional MDL method (W & Z) with the true DOA information in detection performance even for small samples, as is illustrated in Fig. 2. We can also observe from Fig. 2 that the proposed method surpasses the method of Ishikawa *et al.* in detection accuracy, in particular as N becomes small.

Since the training data of the desired signal $d_0(t)$ are used to calculate the matched filters $\boldsymbol{h}_i (i = 1, 2, \ldots, 8)$, and the matched filters are then employed to extract the desired signals of the MSWF $d_i(t) (i = 1, 2, \ldots, 8)$ from the observation data, the first three matched filters of the MSWF, namely, $\boldsymbol{h}_1$, $\boldsymbol{h}_2$, and $\boldsymbol{h}_3$, are capable of capturing the signal information while excluding a large portion of noise. On the contrary, their orthogonal complements are able to mitigate the signal subspace components more efficiently from the noisy data and, therefore, form a cleaner noise subspace. That is to say, the basis vectors of the noise subspace $\boldsymbol{h}_i (i = 4, 5, \ldots, 8)$ have the ability to efficiently eliminate the powers of signals while retaining the powers of noises in calculating the desired signals of the MSWF after the third stage. As a result, the variances of the first three desired signals $\hat{\sigma}^2_{d_i} (i = 1, 2, 3)$, namely, the powers of signals, are well separated from the variances of the desired signals after the third stage $\hat{\sigma}^2_{d_i} (i = 4, 5, \ldots, 8)$, namely, the powers of noises. Meanwhile, the variances of the desired signals after the third stage are clustered sufficiently closely. Thus, the proposed method is capable of significantly reducing the likelihood of overestimating and underestimating the number of signals, eventually leading to the enhanced detection performance.

V. Conclusions and Discussions

In this letter, we have addressed a low-complexity MDL method for accurate source enumeration. This method employs the training data of the desired signal to quickly partition the array data into two orthogonal components in the signal and noise subspaces, and then uses the noise components to calculate the total code length. The model with the shortest code length, namely, the minimum description length, is selected as the best one. Since the estimation of the covariance matrix and its eigendecomposition are not involved, the proposed method is computationally efficient. Meanwhile, the proposed method uses the training data of the desired signal to partition the array data into the cleaner signal and noise components that lead to the enhanced performance and is thereby suitable for the case of low SNR and/or small sample size where the traditional MDL methods fail.

The proposed MDL method cannot be directly applied to completely correlated signals, namely, coherent signals, since in this case, the signal and noise subspaces may not be correctly calculated. Applying the spatial smoothing technique to the array data to decorrelate the coherency of the signals prior to subspace decomposition may be a solution to the problem.

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] 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.

[4] 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 Processing Workshop*, Jul. 2004, pp. 465–469.

[5] Y. Ishikawa, K. Ichige, and H. Arai, "Accurate detection of the number of arrival signals using pre-estimated signal subspace," in *Proc. Signal Processing Advances Wireless Communications Workshop*, May 2005, pp. 840–844.

[6] 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.

[7] J. Li, B. Halder, P. Stoica, and M. Verberg, "Computationally efficient angle estimation for signal with known waveforms," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2154–2163, Sep. 1995.

[8] L. Huang, S. Wu, D. Feng, and L. Zhang, "Computationally efficient direction-of-arrival estimation based on partial a priori knowledge of signal sources," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–7, Article ID 19514.

[9] L. Huang, S. Wu, and L. Zhang, "Low-complexity method of weighted subspace fitting for direction-of-arrival estimation," in *Proc. IEEE Int. Radar Conf.*, Washington, DC, May 2005, pp. 491–496.