Source Enumeration Via MDL Criterion Based on Linear Shrinkage Estimation of Noise Subspace Covariance Matrix

Lei Huang, *Member, IEEE*, and H. C. So, *Senior Member, IEEE*

***Abstract*—Numerous methodologies have been investigated for source enumeration in sample-starving environments. For those having their root in the framework of random matrix theory, the involved distribution of the sample eigenvalues is required. Instead of relying on the eigenvalue distribution, this work devises a linear shrinkage based minimum description length (LS-MDL) criterion by utilizing the identity covariance matrix structure of noise subspace components. With linear shrinkage and Gaussian assumption of the observations, an accurate estimator for the covariance matrix of the noise subspace components is derived. The eigenvalues obtained from the estimator turn out to be a linear function of the corresponding sample eigenvalues, enabling the LS-MDL criterion to accurately detect the source number without incurring significantly additional computational load. Furthermore, the strong consistency of the LS-MDL criterion for $m, n \to \infty$ and $m/n \to c \in (0, \infty)$ is proved, where m and n are the antenna number and snapshot number, respectively. Simulation results are included for illustrating the effectiveness of the proposed criterion.**

***Index Terms*—Linear shrinkage, minimum description length, sample covariance matrix, source enumeration.**

I. Introduction

USING the eigenstructure of covariance matrix of sensor array data, many high-resolution approaches, such as the MUSIC [1], ESPRIT [2] and MODE [3], are able to significantly improve their performance in direction-of-arrival (DOA) estimation. Usually, the observation space of the array data is decomposed into two orthogonal components, namely, the signal subspace and noise subspace, by means of the eigenvalue decomposition (EVD) of the covariance matrix. Nevertheless, ideal assignment of the eigenvectors to the signal and noise subspaces requires the exact information of source number which is usually unknown to the receiver, calling for the study of source enumeration in the community of array signal processing. In DOA estimation, the typical signal model is

$$\boldsymbol{y}(t) = \boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{s}(t) + \boldsymbol{w}(t),\ t = 1, \ldots, n \tag{1}$$

where $\boldsymbol{y}(t) = [y_1(t), \ldots, y_m(t)]^T \in \mathbb{C}^{m\times 1}$, $\boldsymbol{A}(\boldsymbol{\phi}) = [\boldsymbol{a}(\phi_1), \ldots, \boldsymbol{a}(\phi_d)] \in \mathbb{C}^{m\times d}$, $\boldsymbol{s}(t) = [s_1(t), \ldots, s_d(t)]^T \in \mathbb{C}^{d\times 1}$, and $\boldsymbol{w}(t) = [w_1(t), \ldots, w_m(t)]^T \in \mathbb{C}^{m\times 1}$ are the observed snapshot vector, steering matrix, signal vector and noise vector, respectively. Here, $\boldsymbol{a}(\phi_i),\ i = 1, \ldots, d$, is the steering vector with ϕ_i being the DOA, $(\cdot)^T$ is the transpose operation, d is the *unknown* number of sources, m is the number of antennas, and n is the number of snapshots with $d < \min(m, n)$. Meanwhile, the incoherent signals are assumed to be zero-mean complex Gaussian, i.e., $\boldsymbol{s}(t) \sim \mathcal{CN}(\mathbf{o}_d, \boldsymbol{\Sigma}_s)$ in which $\mathbf{o}_d$ is the $d \times 1$ zero vector, $\boldsymbol{\Sigma}_s \triangleq \mathbb{E}[\boldsymbol{s}(t)\boldsymbol{s}^H(t)] \in \mathbb{C}^{d\times d}$ is a full rank diagonal matrix with $(\cdot)^H$ being the conjugate transpose and $\mathbb{E}[\cdot]$ being the mathematical expectation. Here, $\mathcal{CN}(\boldsymbol{\nu}, \boldsymbol{\Sigma})$ stands for the complex Gaussian distribution with mean $\boldsymbol{\nu}$ and covariance $\boldsymbol{\Sigma}$, and $\sim$ signifies "distributed as". Furthermore, the noise $\boldsymbol{w}(t)$ is assumed to be independent and identically distributed (IID) complex Gaussian vector with mean zero and unknown covariance $\tau\boldsymbol{I}_m$, i.e., $\boldsymbol{w}(t) \sim \mathcal{CN}(\mathbf{o}_m, \tau\boldsymbol{I}_m)$ where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, which is independent of the signals. Our task in this work is to infer the source number d from the noisy observations $\{\boldsymbol{y}(t)\}$.

According to our assumptions, the observed samples are taken as the independent Gaussian vector, i.e., $\boldsymbol{y}(t) \sim \mathcal{CN}(\mathbf{o}_m, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being the population covariance matrix, which is calculated as

$$\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{y}(t)\boldsymbol{y}^H(t)] = \boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\Sigma}_s\boldsymbol{A}^H(\boldsymbol{\phi}) + \tau\boldsymbol{I}_m. \tag{2}$$

Recall that the signals are incoherent and $d < \min(m, n)$, meaning that $\boldsymbol{\Sigma}_s$ is nonsingular and $\boldsymbol{A}(\boldsymbol{\phi})$ is of full column rank, respectively. Let λ_i and $\boldsymbol{u}_i,\ i = 1, \ldots, m$, be the population eigenvalues and corresponding population eigenvectors of $\boldsymbol{\Sigma}$ with λ_i in non-increasing order, i.e.,

$$\lambda_1 \geq \cdots \geq \lambda_d \geq \lambda_{d+1} = \cdots = \lambda_m = \tau. \tag{3}$$

It is straightforward to employ a threshold to separate the $(m - d)$ smallest eigenvalues from the d largest eigenvalues, yielding the source number information. In practice, however, only the sample covariance matrix (SCM) is accessible, calculated by $\boldsymbol{S} = (1/n)\sum_{t=1}^{n}\boldsymbol{y}(t)\boldsymbol{y}^H(t)$, and its EVD is

$$\boldsymbol{S} = \boldsymbol{E}\boldsymbol{L}\boldsymbol{E}^H \tag{4}$$

where $\boldsymbol{L} = \mathrm{diag}(\ell_1, \ldots, \ell_m)$ with $\ell_i,\ i = 1, \ldots, m$, being the sample eigenvalue and $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m]$ with $\boldsymbol{e}_i,\ i =$

Manuscript received March 14, 2013; revised June 30, 2013; accepted July 07, 2013. Date of publication July 11, 2013; date of current version September 03, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joseph Tabrikian. This work was supported in part by a grant from the NSFC/RGC Joint Research Scheme sponsored by the Research Grants Council of Hong Kong and by the National Natural Science Foundation of China (Project No.: N_CityU 104/11, 61110229/61161160564), by the National Natural Science by Grants 61222106 and 61171187, and by the Shenzhen Kongqie talent program under Grant KQC201109020061A.

L. Huang is with the Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China (e-mail: lhuang@hitsz.edu.cn; lhuang8sasp@hotmail.com).

H.C. So is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, 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/TSP.2013.2273198

$1, \ldots, m$, being the corresponding sample eigenvector. Here, the sample eigenvalues are non-increasing ordered such that

$$\ell_1 \geq \cdots \geq \ell_d \geq \ell_{d+1} \geq \cdots \geq \ell_m. \tag{5}$$

At high signal-to-noise ratio (SNR), the signal variance is significantly larger than the noise variance, leading to $\ell_d \gg \ell_{d+1}$. By using the recent result in the random matrix theory (RMT) [4]–[9], namely, the $(d+1)$th sample eigenvalue, ℓ_{d+1}, approximately tends to the Tracy-Widom distribution [7], we are able to calculate a threshold from the limiting distribution of ℓ_{d+1} so that the candidate models with sample eigenvalues smaller than the so-obtained threshold are ruled out, eventually yielding the estimate of source number. This is the fundamental idea behind the threshold-like and RMT-based detectors [10], [11]. For the case of low SNRs, however, the signal variance is usually buried by the noise variance, posing a big challenge for source enumeration. Actually, the threshold-like and RMT-based methods can only employ the information of a single eigenvalue for detection, having not used the information of all the sample eigenvalues, and thereby might not provide reliable detection of the source number at low SNRs. Furthermore, as there is no closed-form expression for calculating the inverse of the Tracy-Widom distribution, the determination of threshold undoubtedly involves the numerical evaluation procedure, incurring more overheads in the detection procedure. Finally, the threshold-like methods rely on subjective judgement, e.g., the confidence level, for decision making, thereby might not be conveniently implemented in practice [12]–[14]. To circumvent these issues, the information theoretic criteria, such as the minimum description length (MDL) [15], [16] and Akaike information criterion (AIC) [17], have been introduced by Wax and Kailath [12] for adaptive detection of source number in parameter-changing environment. It has been proved in [12] that the MDL criterion is able to yield the consistent estimate of source number while the AIC criterion tends to overestimate the source number. As a result, the MDL criterion has drawn much attention in the literature [18]–[26].

According to the MDL principle, for a given data set and a family of probabilistic models, we should select the model that yields the shortest description length of the data, which can be evaluated quantitatively. In general, given an observation data set $\boldsymbol{Y} = \{\boldsymbol{y}(t)\}_{t=1}^{n}$ and a probabilistic model $f(\boldsymbol{Y}|\boldsymbol{\mu})$ where $\boldsymbol{\mu}$ is the unknown parameter vector, the shortest code length required to encode the data using the model can be asymptotically written as [15], [16]:

$$\mathcal{L}\{\boldsymbol{y}(t)|\hat{\boldsymbol{\mu}}\} = -\log f(\boldsymbol{Y}|\hat{\boldsymbol{\mu}}) + \frac{1}{2}\nu \log n, \tag{6}$$

where $\hat{\boldsymbol{\mu}}$ is the maximum likelihood (ML) estimate of $\boldsymbol{\mu}$ and ν denotes the number of free parameters in $\hat{\boldsymbol{\mu}}$. It is revealed by Williams [27] that the parameter vector should be formed as $\boldsymbol{\mu} = [\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_d^T]^T$ in that the signal subspace is completely determined by the first d eigenvectors and the eigenvalues are just the weights for different signal components, which when ignored do not affect the signal subspace. For k presumed source signals, it follows from [28] that the ML estimate of $\boldsymbol{\mu}$ is

$$\hat{\boldsymbol{\mu}}^{(k)} = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T]^T. \tag{7}$$

As a result, the MDL for k presumed sources is given as

$$\mathcal{L}\{\boldsymbol{y}(t)|\hat{\boldsymbol{\mu}}^{(k)}\} = -\log f(\boldsymbol{Y}|\hat{\boldsymbol{\mu}}^{(k)}) + \frac{1}{2}\nu_k \log n \tag{8}$$

where ν_k is the number of free parameters in $\hat{\boldsymbol{\mu}}^{(k)}$. Minimizing (8) with respect to k gives the estimate of source number. It is revealed in [14], [17] that the log-likelihood function in (8) can be regarded as the information gained from the observed samples in the signal model. On the other hand, the second term of (8), known as the penalty function, is relative to the number of free parameters. The more parameters are used, the more accurate the signal model would be so that the information gain monotonically decreases. However, as more and more parameters are involved, the errors introduced by the estimated parameters become larger and larger, so does the penalty function. Once the presumed source number equals the order of the signal model, the decreasing information gain meets the increasing penalty function, yielding the correct estimate of source number.

It is worth pointing out that Anderson's results [28] are optimal in the ML sense only when the sample number tends to infinity while the sensor number remains fixed. This in turn means that the classical information theoretic criteria cannot provide reasonably good performance as m goes to infinity at the same speed as n, namely, the so-called general asymptotic case [29]–[31], in that the ML solution turns out to be not accurate enough. Indeed, the general asymptotic case is quite relevant to real-world applications. As an example, by properly employing the waveform diversity in multiple-input multiple-output (MIMO) radar [32]–[34], we can obtain a virtual array with extended aperture in which the number of sensors is considerably increased, probably close to or even larger than the number of snapshots. On the other hand, it has been pointed out in [35] that the general asymptotic case is able to provide a more accurate description for the practical scenarios where the number of snapshots and number of sensors are finite and probably comparable in magnitude. In fact, numerous works [35]–[38] have dealt with the topics of DOA estimation, beamforming as well as space-time adaptive processing in the general asymptotic condition. To enable the information theoretic criteria to properly detect the source number in this situation, the RMT-AIC criterion has been devised by Nadakuditi and Edelman [39]. Although the RMT-AIC is argued to be able to correctly detect the source number for the general asymptotic case, its consistency still remains to be proved. Therefore, it is considerably interested to investigate the consistent methodology for source enumeration in the general asymptotic case.

In this paper, we formulate the MDL criterion by using the linear shrinkage (LS) [29] along with the Gaussian assumption of the observations, ending up with an efficient source enumerator in the general asymptotic case. That is, the source number d does not increase while $m, n \to \infty$ and $m/n \to c$ with c being a constant and satisfying $0 < c < \infty$. In particular, we begin with the derivation of the description length of the observations, which turns out to be determined by the SCM of the noise subspace components. Then, we employ the identity structure of the covariance matrix of the IID noise subspace component along with the LS technique to derive an accurate estimator for the covariance matrix of the noise subspace components in MDL calculation, ending up with an LS-MDL criterion. Note that the covariance matrix of the noise subspace components can be estimated accurately by utilizing the LS technique. When it is employed for MDL calculation, the final detection performance can be considerably enhanced. Furthermore, the consistency of the proposed LS-MDL method in the general asymptotic case is proved. It should be pointed out herein that the proposed estimator for the covariance matrix is different from the existing methodologies based on the LS technique, such as [29]–[31]

and [36], [38], because they are either for estimating the real covariance matrix [29]–[31], [36] or based on *a priori* knowledge [38]. Moreover, the proposed method is able to employ the consistent estimates of the first- and second-order moments of noise sample eigenvalues for computing the shrinkage coefficient, leading to an accurate estimate for the covariance matrix of the noise subspace components. However, the existing approaches can only rely on the sample covariance matrix.

The rest of the paper is organized as follows. The LS-MDL criterion is proposed in Section II. That is, the LS estimator for the covariance matrix of noise subspace components is derived, its estimation accuracy is investigated and the consistency of the proposed LS-MDL criterion is proved. Extensive simulation results are presented in Section III. Finally, conclusion is drawn in Section IV.

II. LS-MDL Criterion for Source Enumeration

A. Minimum Description Length

We observe from (1) that, in the noise-absence case, $\boldsymbol{y}(t)$ is confined in the column space of $\boldsymbol{A}(\boldsymbol{\phi})$, namely, the so-called signal subspace, and that, in the noise-presence case, $\boldsymbol{y}(t)$ may diverge into its complement subspace, namely, the so-called noise subspace. This geometrical structure has been efficiently utilized for high-resolution DOA estimation. To facilitate the MDL computation by means of this structure, we assume that there exists a unitary coordinate transformation $\boldsymbol{U} \triangleq [\boldsymbol{U}_k, \boldsymbol{U}_{m-k}]$ which maps the first k coordinates into the signal subspace and the last $(m-k)$ coordinates into the noise subspace. For the construction of $\boldsymbol{U}$, the interested reader can refer to [13], [19]. In the sequel, $\boldsymbol{U}_k \in \mathbb{C}^{m\times k}$ is the signal subspace and $\boldsymbol{U}_{m-k} \in \mathbb{C}^{m\times(m-k)}$ is the noise subspace. Given the unitary coordinate transformation, the observed samples $\{\boldsymbol{y}(t)\}$ can be decomposed as

$$\boldsymbol{U}^H\boldsymbol{y}(t) = \begin{bmatrix} \boldsymbol{U}_k^H \\ \boldsymbol{U}_{m-k}^H \end{bmatrix} \boldsymbol{y}(t) \triangleq \begin{bmatrix} \boldsymbol{y}_{\mathcal{S}}^{(k)}(t) \\ \boldsymbol{y}_{\mathcal{N}}^{(k)}(t) \end{bmatrix} \tag{9}$$

where

$$\boldsymbol{y}_{\mathcal{S}}^{(k)}(t) = \boldsymbol{U}_k^H\boldsymbol{y}(t) \tag{10}$$

$$\boldsymbol{y}_{\mathcal{N}}^{(k)}(t) = \boldsymbol{U}_{m-k}^H\boldsymbol{y}(t) \tag{11}$$

are the signal and noise subspace components, respectively. Since $\boldsymbol{U}$ is a unitary matrix, the total code length required to encode the observation data $\{\boldsymbol{y}(t)\}$ is equivalent to encoding the signal subspace components $\{\boldsymbol{y}_{\mathcal{S}}^{(k)}(t)\}$ and the noise subspace components $\{\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)\}$. It is shown in Appendix A that the total description length encoding $\{\boldsymbol{y}(t)\}$ can be asymptotically expressed as

$$\mathcal{L}\{\boldsymbol{y}_{\mathcal{S}}^{(k)}(t), \boldsymbol{y}_{\mathcal{N}}^{(k)}(t)|\boldsymbol{U}\} = n\,(m-k)\log\frac{\frac{1}{m-k}\mathrm{tr}[\boldsymbol{S}_{nn}^{(k)}]}{\left(\det[\boldsymbol{S}_{nn}^{(k)}]\right)^{\frac{1}{m-k}}} + \frac{1}{2}k(k-1)\log n \tag{12}$$

where $\mathrm{tr}[\cdot]$ is the trace, $\det[\cdot]$ is the determinant and $\boldsymbol{S}_{nn}^{(k)} = (1/n)\sum_{t=1}^{n}\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)(\boldsymbol{y}_{\mathcal{N}}^{(k)}(t))^H$ is the SCM.

It should be pointed out that minimizing the MDL in (12) leads to the correct estimate of source number for $n \to \infty$ with m being unchanged because $\boldsymbol{S}_{nn}^{(k)}$ is the consistent estimate of $\boldsymbol{\Sigma}_{nn}^{(k)} \triangleq \mathbb{E}[\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)(\boldsymbol{y}_{\mathcal{N}}^{(k)}(t))^H]$, i.e., the population covariance matrix of noise subspace components, in this classical asymptotic case. For the general asymptotic case where $m, n \to \infty$ and $m/n \to c$, however, $\boldsymbol{S}_{nn}^{(k)}$ has a quite large variance and thereby is an inaccurate estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$. This in turn degrades the MDL criterion of (12). To circumvent this problem, one possible choice is to reformulate the MDL criterion by minimizing the Kullback-Leibler (KL) divergence, which, however, might be quite involved as the computation of the KL divergence is unknown for the case of $m, n \to \infty$ and $m/n \to c$. Another possible choice is to improve the estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$ in (12) for source enumeration. To this end, we utilize the LS technique along with the Gaussian assumption to derive a consistent estimate for $\boldsymbol{\Sigma}_{nn}^{(k)}$. It is well known that the LS technique provides a simple explicit formula for computing the covariance matrix, making the proposed source enumerator easy to implement. On the other hand, $\boldsymbol{\Sigma}_{nn}^{(k)}$ has the identity structure, that is, $\tau\boldsymbol{I}_{m-k}$, for $k \geq d$, but it does not have this structure for $k < d$. Using this identity structure, we are able to devise an accurate estimator for $\boldsymbol{\Sigma}_{nn}^{(k)}$ in the next subsection. The refined estimation, when utilized for MDL calculation, results in the LS-MDL criterion which is able to considerably improve the detection performance for $m, n \to \infty$ and $m/n \to c$.

B. Estimation of Noise Subspace Covariance Matrix

Recall that the population covariance matrix of $\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)$ is

$$\boldsymbol{\Sigma}_{nn}^{(k)} = \boldsymbol{U}_{m-k}^H\boldsymbol{\Sigma}\boldsymbol{U}_{m-k} \triangleq \boldsymbol{Q}\boldsymbol{\Lambda}_{m-k}\boldsymbol{Q}^H \tag{13}$$

where $\boldsymbol{\Lambda}_{m-k} = \mathrm{diag}(\lambda_{k+1}, \ldots, \lambda_m)$, $\boldsymbol{Q} = [\boldsymbol{q}_{k+1}, \ldots, \boldsymbol{q}_m]$ with λ_i $(i = k+1, \ldots, m)$ and $\boldsymbol{q}_i \in \mathbb{C}^{(m-k)\times 1}$ $(i = k+1, \ldots, m)$ being the eigenvalue and eigenvector of $\boldsymbol{\Sigma}_{nn}^{(k)}$. Note that the eigenvalues of $\boldsymbol{\Sigma}_{nn}^{(k)}$ are equal to each other for $k \geq d$ but unequal for $k < d$. Therefore, given $\boldsymbol{S}_{nn}^{(k)}$, our aim is to derive an optimal estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$ in the sense of minimum mean square error (MMSE) by means of the LS technique along with the IID Gaussian assumption of $\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)$. To this end, we consider the following constrained minimization of the mean square error (MSE):

$$\begin{aligned} \min_{\alpha^{(k)}} \quad & \mathbb{E}\left[\left\|\boldsymbol{R}^{(k)} - \boldsymbol{\Sigma}_{nn}^{(k)}\right\|_F^2\right] \\ \text{s.t.} \quad & \boldsymbol{R}^{(k)} = \alpha^{(k)}\boldsymbol{P}^{(k)} + (1-\alpha^{(k)})\boldsymbol{S}_{nn}^{(k)} \end{aligned} \tag{14}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\alpha^{(k)} \in [0,1]$ denotes the shrinkage coefficient [30] and $\boldsymbol{P}^{(k)} = \mathrm{tr}[\boldsymbol{\Sigma}_{nn}^{(k)}]/(m-k)\boldsymbol{I}_{m-k}$.

It is well known that $\boldsymbol{S}_{nn}^{(k)}$ is the unbiased ML estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$ for fixed m and $n \to \infty$. However, it has a large variance and usually is ill-conditioned for $m \to \infty$, as has been elaborated in [29]–[31], [36]. In contrast to the ML estimate, the structured estimate $\boldsymbol{P}^{(k)}$ is able to provide the reduced variance but with increasing bias. The estimate $\boldsymbol{R}^{(k)}$ in (14) is obtained by shrinking $\boldsymbol{S}_{nn}^{(k)}$ toward $\boldsymbol{P}^{(k)}$, having the effect of tradeoff between the low bias and low variance. Consequently, minimizing the MSE in (14) results in an accurate estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$ for $m, n \to \infty$ and $m/n \to c$.

Utilizing $\mathbb{E}[\boldsymbol{S}_{nn}^{(k)}] = \boldsymbol{\Sigma}_{nn}^{(k)}$, $\|\boldsymbol{A}\|_F^2 = \mathrm{tr}[\boldsymbol{A}\boldsymbol{A}^H]/m$ and $\|\boldsymbol{A}+\boldsymbol{B}\|_F^2 = \|\boldsymbol{A}\|_F^2 + \|\boldsymbol{B}\|_F^2 + 2\mathrm{tr}[\mathrm{Re}(\boldsymbol{A}\boldsymbol{B}^H)]/m$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{m\times m}$, the MSE cost function in (14) can be easily rewritten as

$$\begin{aligned} & h\left(\alpha^{(k)}\right) \\ &\triangleq \mathbb{E}\left[\left\|\alpha^{(k)}\boldsymbol{P} + \left(1-\alpha^{(k)}\right)\boldsymbol{S}_{nn}^{(k)} - \boldsymbol{\Sigma}_{nn}^{(k)}\right\|_F^2\right] \\ &= \mathbb{E}\left[\left\|\alpha^{(k)}\left(\boldsymbol{P} - \boldsymbol{\Sigma}_{nn}^{(k)}\right) + \left(1-\alpha^{(k)}\right)\left(\boldsymbol{S}_{nn}^{(k)} - \boldsymbol{\Sigma}_{nn}^{(k)}\right)\right\|_F^2\right] \\ &= \left(\alpha^{(k)}\right)^2\left\|\boldsymbol{P} - \boldsymbol{\Sigma}_{nn}^{(k)}\right\|_F^2 + \left(1-\alpha^{(k)}\right)^2\mathbb{E}\left[\left\|\boldsymbol{S}_{nn}^{(k)} - \boldsymbol{\Sigma}_{nn}^{(k)}\right\|_F^2\right]. \end{aligned} \tag{15}$$

Here, $\mathrm{Re}(x)$ denotes the real part of complex number x. Setting the derivative of $h(\alpha^{(k)})$ to zero and performing some straightforward algebraic manipulations, we obtain the oracle estimate of $\alpha^{(k)}$:

$$\alpha_O^{(k)} = \frac{\mathbb{E}\left[\mathrm{tr}[\boldsymbol{S}_{nn}^{(k)}(\boldsymbol{S}_{nn}^{(k)})^H]\right] - \mathrm{tr}\left[\boldsymbol{\Sigma}_{nn}^{(k)}(\boldsymbol{\Sigma}_{nn}^{(k)})^H\right]}{\mathbb{E}\left[\mathrm{tr}[\boldsymbol{S}_{nn}^{(k)}(\boldsymbol{S}_{nn}^{(k)})^H]\right] - \frac{\mathrm{tr}^2[\boldsymbol{\Sigma}_{nn}^{(k)}]}{m-k}}. \tag{16}$$

Accordingly, the oracle estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$ is

$$\boldsymbol{R}_O^{(k)} = \alpha_O^{(k)}\boldsymbol{P}^{(k)} + \left(1-\alpha_O^{(k)}\right)\boldsymbol{S}_{nn}^{(k)}. \tag{17}$$

It is easy to verify from (17) that the shrunken covariance matrix shares the same eigenvectors with the sample covariance matrix, indicating that the parameter vector remains unchanged when the former is used to replace the latter in the MDL calculation. However, the oracle estimate $\boldsymbol{R}_O^{(k)}$ in (17) cannot be computed in practice as $\alpha_O^{(k)}$ and $\boldsymbol{P}^{(k)}$ depend on $\boldsymbol{\Sigma}_{nn}^{(k)}$ whereas $\boldsymbol{S}_{nn}^{(k)}$ relies on $\boldsymbol{U}_{m-k}$. Recalling that $\hat{\tau} = (1/(m-k))\sum_{i=k+1}^m \ell_i$ and $\boldsymbol{E}_{m-k}$ are the ML estimates of τ and $\boldsymbol{U}_{m-k}$, respectively, it follows that

$$\hat{\boldsymbol{P}}^{(k)} = \hat{\tau}\boldsymbol{I}_{m-k} \tag{18}$$

$$\hat{\boldsymbol{S}}_{nn}^{(k)} = \mathrm{diag}(\ell_{k+1}, \ldots, \ell_m) \tag{19}$$

are the estimates of $\boldsymbol{P}^{(k)}$ and $\boldsymbol{S}_{nn}^{(k)}$, and from (16):

$$\hat{\alpha}_O^{(k)} = \frac{\mathbb{E}\left[\sum_{i=k+1}^m \ell_i^2\right] - \sum_{i=k+1}^m \lambda_i^2}{\mathbb{E}\left[\sum_{i=k+1}^m \ell_i^2\right] - \frac{\left(\sum_{i=k+1}^m \lambda_i\right)^2}{m-k}} \tag{20}$$

is the estimate of $\alpha_O^{(k)}$. Nevertheless, (20) still involves the population eigenvalues. Indeed, for the Gaussian observations, invoking the results in [36], [40], that is,

$$\mathbb{E}[\mathrm{tr}[\hat{\boldsymbol{S}}_{nn}^{(k)}]] = \mathrm{tr}[\boldsymbol{\Sigma}_{nn}^{(k)}] \tag{21}$$

$$\mathbb{E}[\mathrm{tr}[(\hat{\boldsymbol{S}}_{nn}^{(k)})^2]] = \frac{n+1}{n}\mathrm{tr}[(\boldsymbol{\Sigma}_{nn}^{(k)})^2] + \frac{1}{n}\mathrm{tr}^2(\boldsymbol{\Sigma}_{nn}^{(k)}) \tag{22}$$

we have

$$\sum_{i=k+1}^m \lambda_i = \mathbb{E}\left[\sum_{i=k+1}^m \ell_i\right] \tag{23}$$

$$\sum_{i=k+1}^m \lambda_i^2 = \frac{1}{n+1}\left(n\mathbb{E}\left[\sum_{i=k+1}^m \ell_i^2\right] - \left(\mathbb{E}\left[\sum_{i=k+1}^m \ell_i\right]\right)^2\right). \tag{24}$$

Substituting (23) and (24) into (20) yields

$$\hat{\alpha}_O^{(k)} = \frac{\mathbb{E}\left[\sum_{i=k+1}^m \ell_i^2\right] + \left(\mathbb{E}\left[\sum_{i=k+1}^m \ell_i\right]\right)^2}{(n+1)\left(\mathbb{E}\left[\sum_{i=k+1}^m \ell_i^2\right] - \frac{\left(\mathbb{E}\left[\sum_{i=k+1}^m \ell_i\right]\right)^2}{m-k}\right)}. \tag{25}$$

However, this estimate cannot yet be applied in practice as it requires to know $\mathbb{E}\left[\sum_{i=k+1}^m \ell_i^2\right]$ and $\mathbb{E}\left[\sum_{i=k+1}^m \ell_i\right]$. Actually, it is shown in Appendix B that, when $m, n \to \infty$ and $m/n \to c$,

$$\frac{1}{m-k}\sum_{i=k+1}^m \ell_i \xrightarrow{\text{m.s.}} \mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^m \ell_i\right] \tag{26a}$$

$$\frac{1}{m-k}\sum_{i=k+1}^m \ell_i^2 \xrightarrow{\text{m.s.}} \mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^m \ell_i^2\right] \tag{26b}$$

where $\xrightarrow{\text{m.s.}}$ means convergence in mean square. In the sequel, substituting these consistent estimates into (25) yields the consistent estimate of $\hat{\alpha}_O^{(k)}$:

$$\hat{\alpha}^{(k)} = \frac{\sum_{i=k+1}^m \ell_i^2 + \left(\sum_{i=k+1}^m \ell_i\right)^2}{(n+1)\left(\sum_{i=k+1}^m \ell_i^2 - \frac{\left(\sum_{i=k+1}^m \ell_i\right)^2}{m-k}\right)}. \tag{27}$$

Since $\hat{\alpha}^{(k)}$ can be larger than 1, we utilize $\beta^{(k)} = \min(\hat{\alpha}^{(k)}, 1)$ [29] rather than $\hat{\alpha}^{(k)}$ as the estimated shrinkage coefficient. Therefore, the covariance matrix is finally estimated as the following diagonal matrix

$$\hat{\boldsymbol{R}}^{(k)} = \beta^{(k)}\hat{\boldsymbol{P}}^{(k)} + (1-\beta^{(k)})\hat{\boldsymbol{S}}_{nn}^{(k)} \triangleq \mathrm{diag}\left(\rho_{k+1}^{(k)}, \ldots, \rho_m^{(k)}\right) \tag{28}$$

with

$$\rho_i^{(k)} = \beta^{(k)}\hat{\tau}^{(k)} + (1-\beta^{(k)})\ell_i,\ i = k+1, \ldots, m. \tag{29}$$

It is implied in (29) that the noise eigenvalues are estimated by shrinking the noise sample eigenvalues to the estimated noise variance.

Proposition 1: For $k \geq d$, $\hat{\boldsymbol{R}}^{(k)}$ is the consistent estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$ as m increases to infinity at the same speed as n. That is,

$$\mathbb{E}\left(\left\|\hat{\boldsymbol{R}}^{(k)} - \boldsymbol{\Sigma}_{nn}^{(k)}\right\|_F^2\right) \to 0 \text{ as } m, n \to \infty \text{ and } m/n \to c. \tag{30}$$

Proof: The proof is in Appendix C. ■

Recall that the ML estimate, $\boldsymbol{S}_{nn}^{(k)}$, is also the consistent estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$ as $n \to \infty$ while m remains unchanged. As a result, the consistent estimate may be obtained by either minimizing the MSE in (14) or maximizing the likelihood function provided that the sample size is large enough. This thereby allows us to replace $\boldsymbol{S}_{nn}^{(k)}$ with $\hat{\boldsymbol{R}}^{(k)}$ in (12) for source enumeration. As $m, n \to \infty$ and $m/n \to c$, nevertheless, $\boldsymbol{S}_{nn}^{(k)}$ is a poor estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$, calling for the replacement of $\boldsymbol{S}_{nn}^{(k)}$ with $\hat{\boldsymbol{R}}^{(k)}$ to enhance the performance in source enumeration.

C. Source Enumeration Via LS-MDL Criterion

To carry out source enumeration, we replace $\boldsymbol{S}_{nn}^{(k)}$ in (12) with the consistent estimate of $\boldsymbol{\Sigma}_{nn}^{(k)}$, i.e., $\hat{\boldsymbol{R}}^{(k)}$, leading to a new

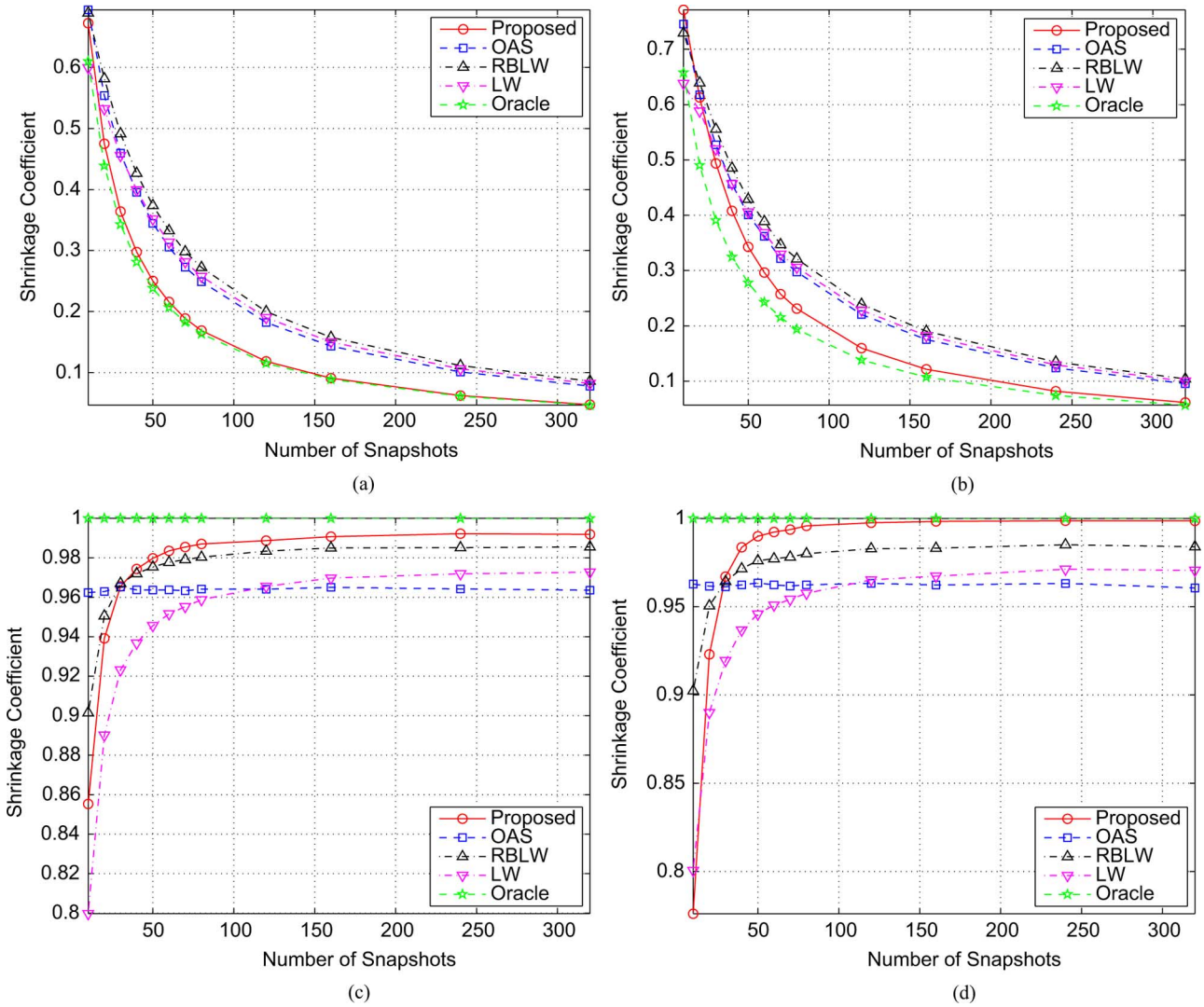


Fig. 1. Shrinkage coefficient versus number of snapshots. Parameters: $m = 30$, $d = 2$, $[\phi_1, \phi_2] = [2.7°, 10.6°]$ and $\mathrm{SNR} = -5\mathrm{dB}$.

variant of the MDL criterion due to Rissanen [15], denoted as LS-MDL:

$$\begin{aligned}
&\mathrm{LS-MDL}(k)\\
&= n\,(m-k)\log\frac{\frac{1}{m-k}\mathrm{tr}[\hat{\boldsymbol{R}}^{(k)}]}{\left(\det[\hat{\boldsymbol{R}}^{(k)}]\right)^{\frac{1}{m-k}}} + \frac{1}{2}k(k-1)\log n\\
&= n(m-k)\log\frac{\frac{1}{m-k}\sum_{i=k+1}^{m}\rho_i^{(k)}}{\left(\prod_{i=k+1}^{m}\rho_i^{(k)}\right)^{\frac{1}{m-k}}} + \frac{1}{2}k(k-1)\log n. \quad (31)
\end{aligned}$$

Minimizing (31) with respect to k yields the estimate of source number:

$$\hat{d} = \arg\min_{k=0,\ldots,\bar{m}-1} \mathrm{LS-MDL}(k) \quad (32)$$

where $\bar{m}$ is the upper limit of the source number. Usually, $\bar{m}$ can be set as $\min(m,n)-1$, just like that in [39]. In practice, however, d is a relatively small number whereas m and n may tend to infinity. In such a situation, $\bar{m}$ has to be set as a fixed number much less than $\min(m,n)$. The LS-MDL algorithm for source numeration is summarized in Table I.

D. Consistency of LS-MDL Criterion

The consistency of the LS-MDL criterion is stated in the following proposition.

Proposition 2: The LS-MDL criterion in (31) has strong consistency. That is, as $m, n \to \infty$ and $m/n \to c$, the LS-MDL criterion detects the source number with probability one.

Proof: The proof is provided in Appendix D. ■

III. Simulation Results

A. Accuracy of Estimated Eigenvalues via Linear Shrinkage

This experiment evaluates the accuracy of the shrunken eigenvalues yielded by the proposed method. To this end, we compare the MSE between $\hat{\boldsymbol{R}}^{(k)}$ and $\boldsymbol{\Sigma}_{nn}^{(k)}$ by assuming that $\boldsymbol{\Sigma}_{nn}^{(k)}$ is available for performance evaluation. Meanwhile, for the purpose of comparison, the results of the Lediot-Wolf (LW) method [31], the Rao-Blackwell [41] based LW (RBLW) approach [36] and the oracle approximating shrinkage (OAS) scheme [36] are presented in which we have employed the procedure in [36] for the conversion between the complex covariance matrix and its real counterpart. All the numerical

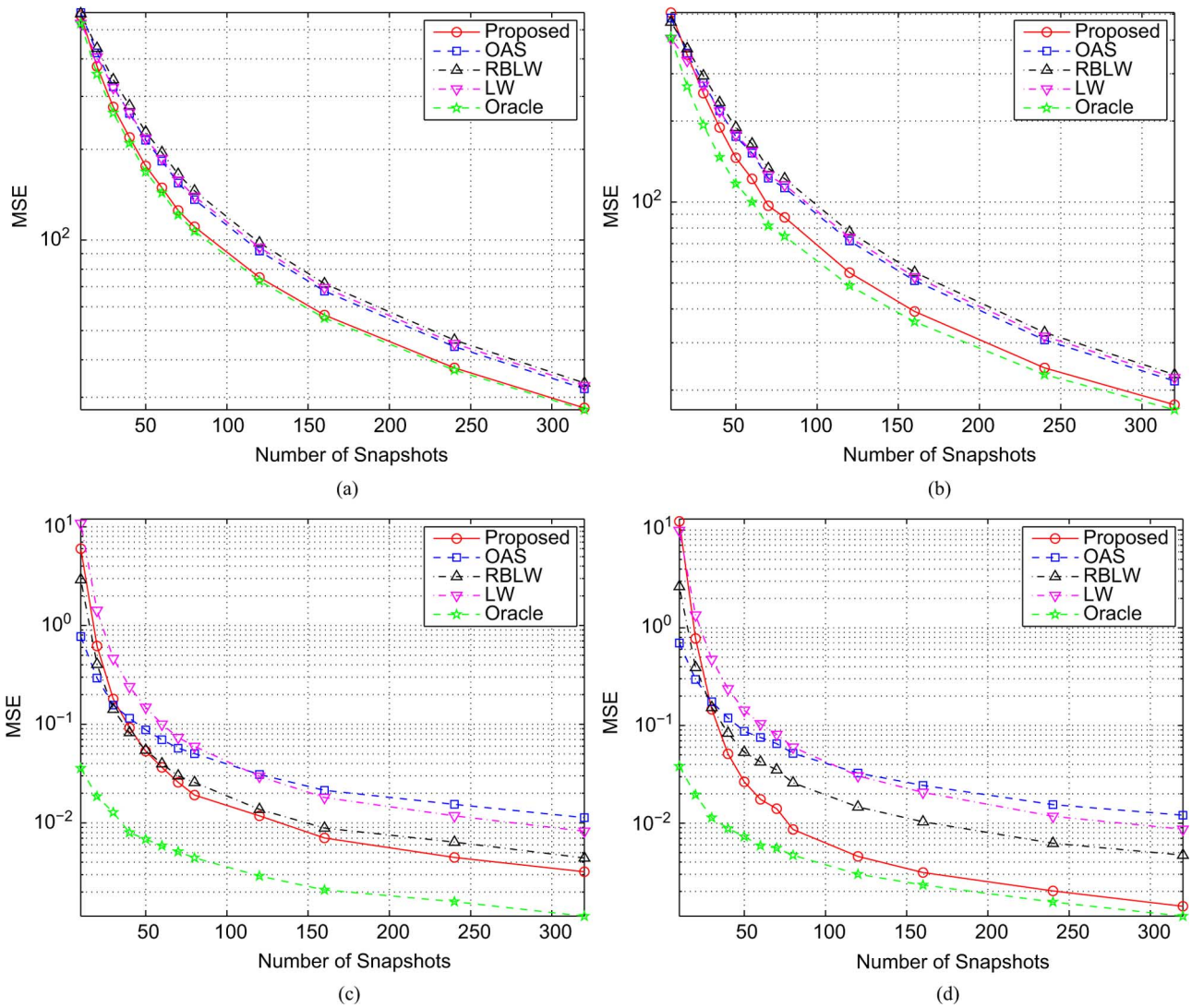


Fig. 2. MSE versus number of snapshots. Parameters: $m = 30$, $d = 2$, $[\phi_1, \phi_2] = [2.7°, 10.6°]$ and $\mathrm{SNR} = -5\mathrm{dB}$.

TABLE I
SUMMARY OF LS-MDL ALGORITHM

Step 1: Perform EVD on $\boldsymbol{S}$ and obtain the sample eigenvalues $\ell_1, \cdots, \ell_m$.
Step 2: Estimate the noise variance by $\hat{\tau}^{(k)} = (1/(m-k))\sum_{i=k+1}^{m} \ell_i$ and calculate the shrinkage coefficient as $\beta^{(k)} = \min(\hat{\alpha}^{(k)}, 1)$ with $\hat{\alpha}^{(k)}$ being given in (27).
Step 3: The shrunken estimates of λ_i, $i = k+1, \cdots, m$, are calculated as $\rho_i^{(k)} = \beta^{(k)}\hat{\tau}^{(k)} + (1-\beta^{(k)})\ell_i$.
Step 4: The source number is estimated by minimizing the LS-MDL criterion of (31).

results in this subsection are obtained by averaging 1000 independent trials.

We assume that there are two uncorrelated narrow-band signals in the far field impinging upon a uniform linear array (ULA) with half-wavelength element separation from the directions $[2.7°, 10.6°]$, which have equal powers and are independent of the additive noise. Meanwhile, the noise is an IID Gaussian process with zero mean and unity variance, i.e., $\tau = 1$. Fig. 1 depicts the shrinkage coefficients of various methods for $k = 0, 1, 2, 3$, at $m = 30$ and $\mathrm{SNR} = -5$ dB. We observe in Figs. 1(a)–1(b) that, for $0 \le k \le d-1$, the shrinkage coefficients of the proposed method are very close to their oracle estimate which is computed from the true $\boldsymbol{\Sigma}_{nn}^{(k)}$ and works herein as the performance benchmark. However, the LW, RBLW and OAS approaches have a large gap from the benchmark, although their estimation accuracies improve as the number of snapshots increases. For $d \le k \le m-1$, the covariance matrix of the noise subspace components is $\boldsymbol{\Sigma}_{nn}^{(k)} = \mathrm{diag}\{\lambda_{k+1}^{(k)}, \ldots, \lambda_{30}^{(k)}\}$ with $\lambda_{k+1}^{(k)} = \cdots = \lambda_{30}^{(k)} = \tau = 1$. In such a case, the ideal shrinkage coefficient should be one. It is seen in Figs. 1(c)–1(d) that the shrinkage coefficient of the proposed method is closer to one than the other schemes as the sample size increases, particularly when $k = 3$. Similarly, it is indicated in Fig. 2 that the MSE of the proposed method is much closer to the benchmark than the other schemes, thereby meaning that the former is more efficient than the latter.

Figs. 3 and 4 show the numerical results for $m = 60$ and $\mathrm{SNR} = -10$ dB while the other parameters remain unchanged. We observe in Fig. 3 that the proposed method is able to provide more accurate shrinkage coefficient for the signal-bearing case. On the other hand, the proposed method is superior to the other algorithms for the signal-free case at medium and large sample sizes but is not as accurate as the OAS and RBLW schemes when m is much greater than n. In fact, it might be more interested to enumerate the source number in the cases of $m \approx n$ and $m \le n$ in array processing as m is the number of antennas whereas n is the number of snapshots. It is known

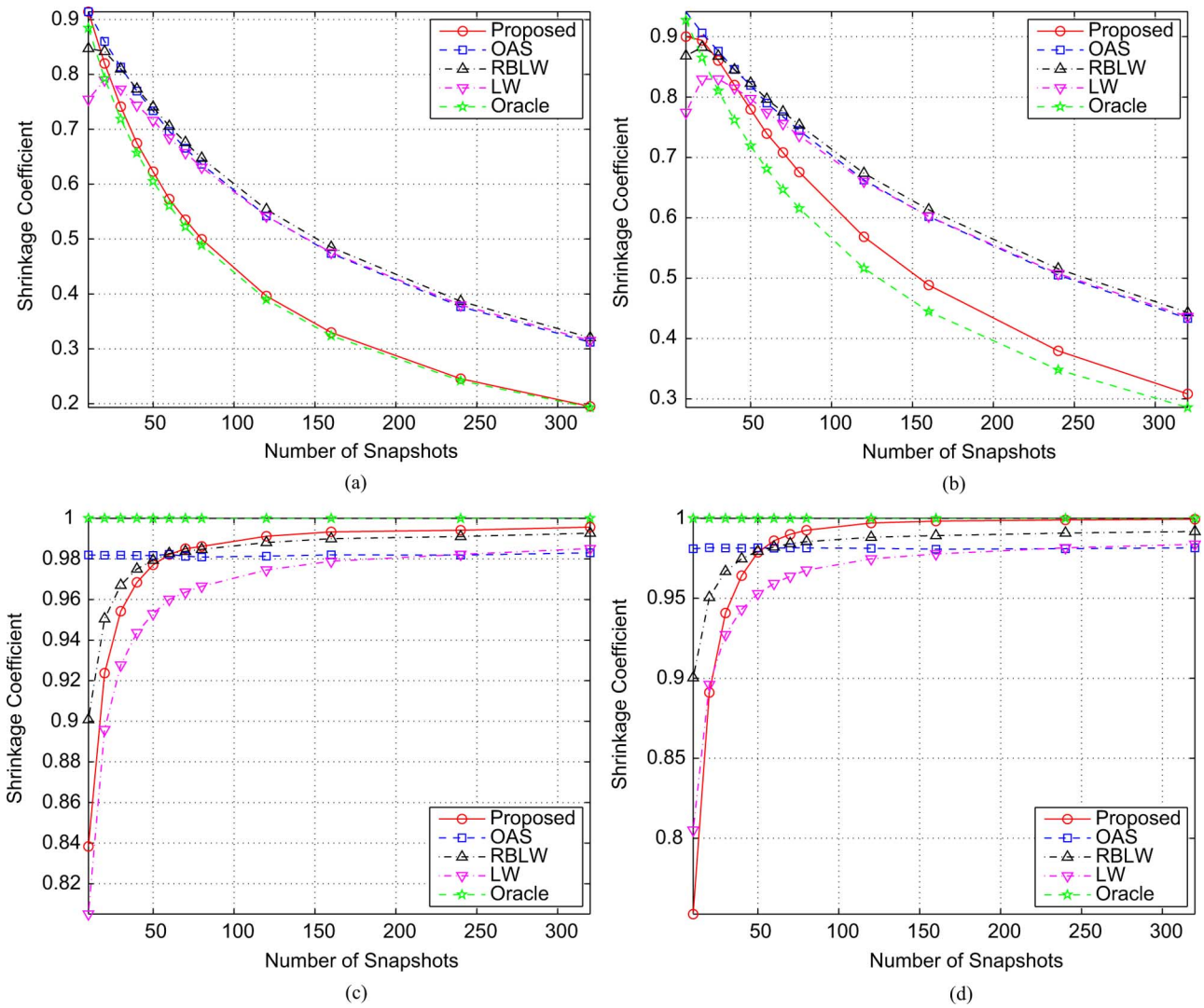


Fig. 3. Shrinkage coefficient versus number of snapshots. Parameters: $m = 60$, $d = 2$, $[\phi_1, \phi_2] = [2.7°, 10.6°]$ and $\mathrm{SNR} = -10\mathrm{dB}$.

that increasing m is much more expensive than increasing n. Meanwhile, it is seen that although the OAS method is able to accurately compute the shrinkage coefficient for $m \gg n$ and $k \geq d$, it cannot attain one as the sample size increases, thereby being an inconsistent estimator. Thus, the OAS approach is inferior to the proposed method for medium and large sample sizes. Meanwhile, Fig. 4 indicates that the proposed method is much more accurate than the other schemes when the sample number is greater than the sensor number.

B. Performance of Source Enumeration

In this experiment, we compare the performance of the proposed LS-MDL criterion with that of the existing information theoretic criteria, namely, the RMT-AIC, corrected AIC (AIC_{c_3}) [42], exponentially embedded families (EEF) [43], [43], classical MDL and AIC [12]. For self-contained presentation, the criterion functions are given in Table II.

We consider a ULA with half-wavelength element separation receiving three narrow-band stationary Gaussian signals with the DOAs: $[2.3°, 9.8°, 17.5°]$. As m and n can be large numbers, for the purpose of fair comparison, $\bar{m}$ is set to be 10 for all criteria unless stated otherwise. Meanwhile, all the numerical results are computed from 2000 independent trials.

Fig. 5(a) shows the empirical probabilities of error detection versus SNR for a sample size smaller than the sensor number,i.e., $m/n = 3$. In such a sample-starving case, only the proposed, AIC_{c_3} and RMT-AIC criteria have the ability to detect the source number. We observe that the proposed criterion is able to provide the consistent source number estimate while the RMT-AIC scheme does not. Meanwhile, the proposed criterion is considerably superior to the AIC_{c_3} method in terms of probability of error detection. As the sample number increases to m, the proposed, AIC_{c_3} EEF and MDL methods can correctly detect the source number while the RMT-AIC and AIC schemes are unable to provide reliable detection. Meanwhile, although the AIC_{c_3}, EEF and MDL approaches can correctly detect the source number, they require higher SNRs to obtain the same detection accuracy as the proposed criterion. The RMT-AIC and AIC methods, however, cannot offer consistent detection of the source number even though the SNR becomes large enough. This is demonstrated in Fig. 5(b). When n becomes much larger than m, the ML estimate turns to be optimal, leading to good detection performance of the AIC_{c_3}, EEF and MDL criteria. However, there is still a big performance gap between the proposed and MDL approaches, as illustrated in Fig. 5(c). As the AIC scheme is well known to be an inconsistent estimator, its error detection probability fails to reach 0 no matter how large the SNR is. Similarly, the RMT-AIC and AIC_{c_3} schemes cannot converge to zero in probability of error detection even at sufficiently large SNRs,

TABLE II
CRITERION FUNCTIONS OF RMT-AIC, AIC_{c_3}, EEF, MDL AND AIC

Criterion	Function
RMT-AIC	$\hat{d} = \arg\min_{k\in[0,\tilde{m}-1]} \left\{ \frac{n^2}{2} \left[\frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i^2}{\left(\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i\right)^2} - \left(1+\frac{m}{n}\right) \right]^2 + 2k \right\}$
AIC_{c_3}	$\hat{d} = \arg\min_{k\in[0,\tilde{m}-1]} \left\{ 2n\,(m-k) \log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i}{\left(\prod_{i=k+1}^{m} \ell_i\right)^{1/(m-k)}} + \frac{2(\nu_k+1)(\tilde{m}+\nu_k+2)}{\tilde{m}-\nu_k-2} - \frac{2\nu_k}{\tilde{m}-\nu_k} \right\}$ $\nu_k = k(2m-k), \tilde{m} = 2mn$
EEF	$\hat{d} = \arg\max_{k\in[0,\tilde{m}-1]} \left\{ L_{G_k} - \nu_k \left[\log\left(\frac{L_{G_k}}{\nu_k}\right) + 1 \right] \right\} \cdot u\left(\frac{L_{G_k}}{\nu_k} - 1\right)$ $\nu_k = k(2m-k)$ $u(\gamma) = 1$ for $\gamma \geq 0$ and $u(\gamma) = 0$ for $\gamma < 0$ $L_{G_k} = -2n \left[\log\left(\prod_{i=1}^{k} \ell_i\right) + (m-k)\log\left(\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i\right) - m\log\left(\frac{1}{m}\sum_{i=1}^{m} \ell_i\right) \right]$
MDL	$\hat{d} = \arg\min_{k\in[0,\tilde{m}-1]} \left\{ n\,(m-k) \log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i}{\left(\prod_{i=k+1}^{m} \ell_i\right)^{1/(m-k)}} + \frac{1}{2}k(2m-k)\log n \right\}$
AIC	$\hat{d} = \arg\min_{k\in[0,\tilde{m}-1]} \left\{ 2n\,(m-k) \log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i}{\left(\prod_{i=k+1}^{m} \ell_i\right)^{1/(m-k)}} + 2k(2m-k) \right\}$

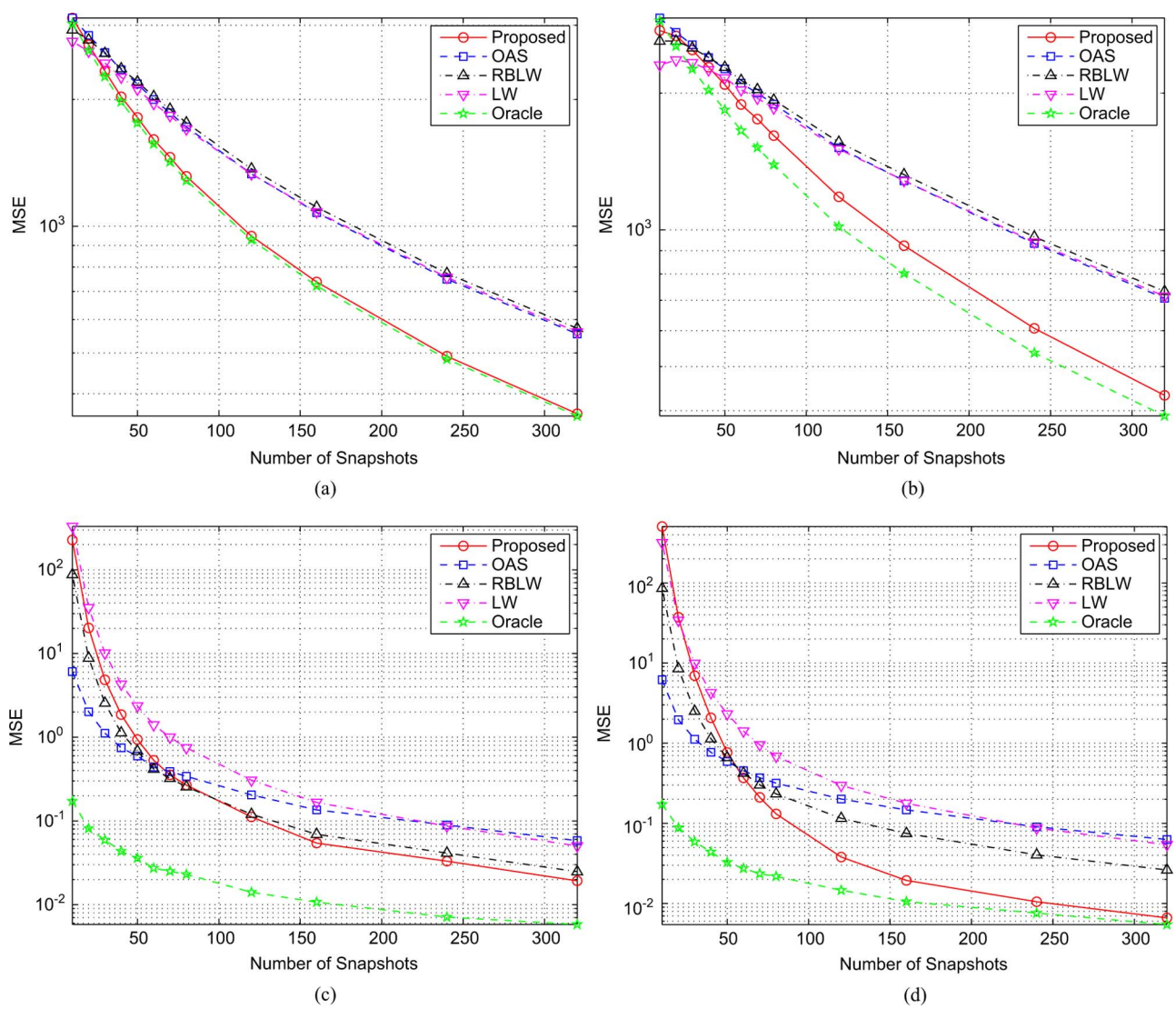
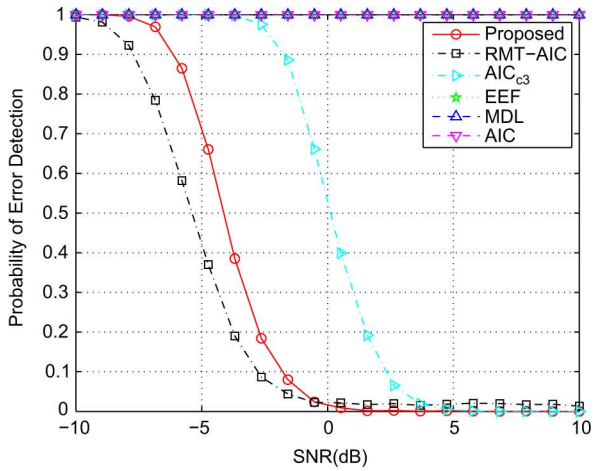


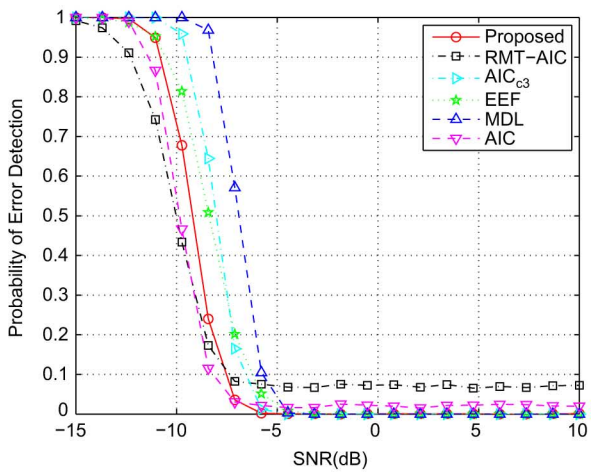
Fig. 4. MSE versus number of snapshots. Parameters: $m = 60$, $d = 2$, $[\phi_1, \phi_2] = [2.7°, 10.6°]$ and $\mathrm{SNR} = -10$ dB.

which can be verified by the enlargement of this part of graph given on the right-hand corner of Fig. 5(c).

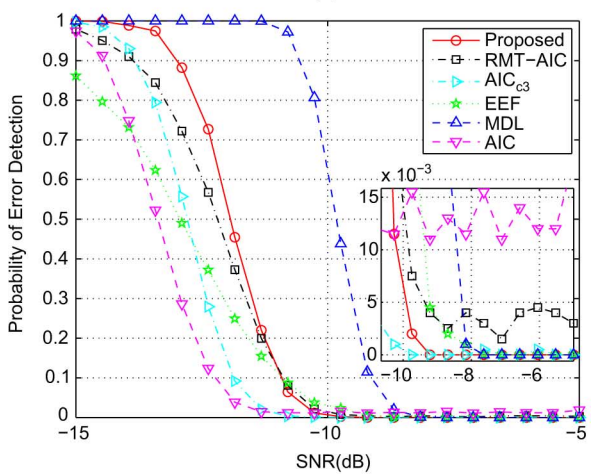
The probability of error detection versus the number of snapshots is depicted in Fig. 6. It is seen that the proposed scheme is capable of correctly enumerating the sources for $m = 3n$, $m = n$ and $m = n/3$. Meanwhile, Fig. 6(a) implies that, although the RMT-AIC and EEF methods are able to detect the source number, the former cannot attain error detection



(a)



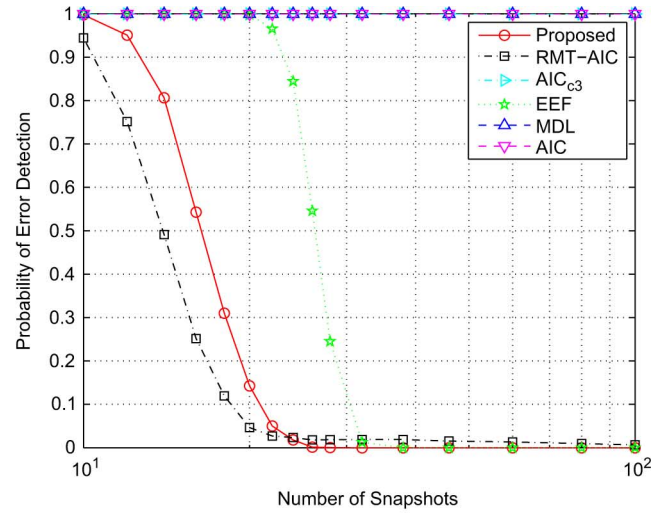
(b)



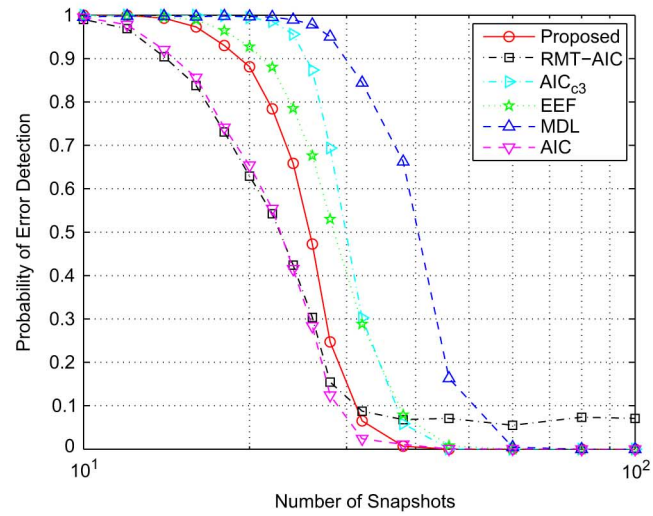
(c)

Fig. 5. Probability of error detection versus SNR. $[\phi_1, \phi_2, \phi_3] = [2.3°, 9.8°, 17.5°]$.

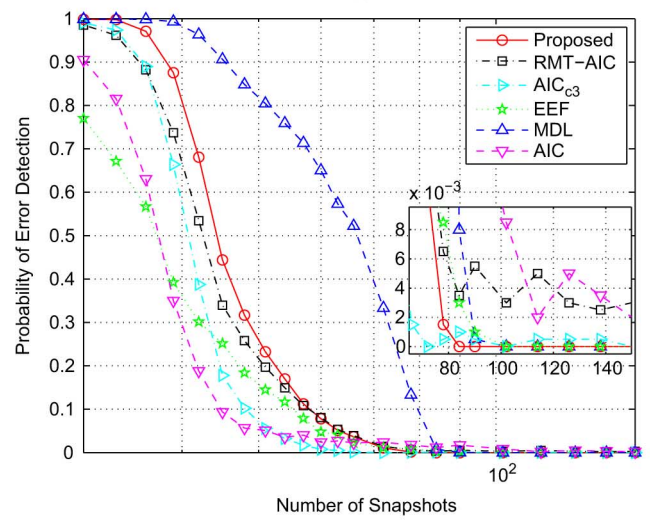
probability of zero while the latter requires around additional ten samples to yield the same detection accuracy as the proposed criterion. The other criteria, however, totally fail in this case. Although the MDL and AIC_{c_3} schemes can correctly detect the source number at $m = n$, they are still much inferior to the proposed criterion, as illustrated in Fig. 6(b). When the number of snapshots is larger than the sensor number, say, $m = n/3$, the ML-based methodologies, e.g., the MDL and EEF schemes, become optimal and thereby can accurately enumerate the sources.



(a)



(b)



(c)

Fig. 6. Probability of error detection versus number of snapshots. $[\phi_1, \phi_2, \phi_3] = [2.3°, 9.8°, 17.5°]$ and $\mathrm{SNR} = -8\mathrm{dB}$.

Again, the RMT-AIC, AIC_{c_3} and AIC schemes cannot attain error detection probability of zero even when the sample size becomes large enough. This is illustrated in the enlargement part on the right-hand corner of Fig. 6(c).

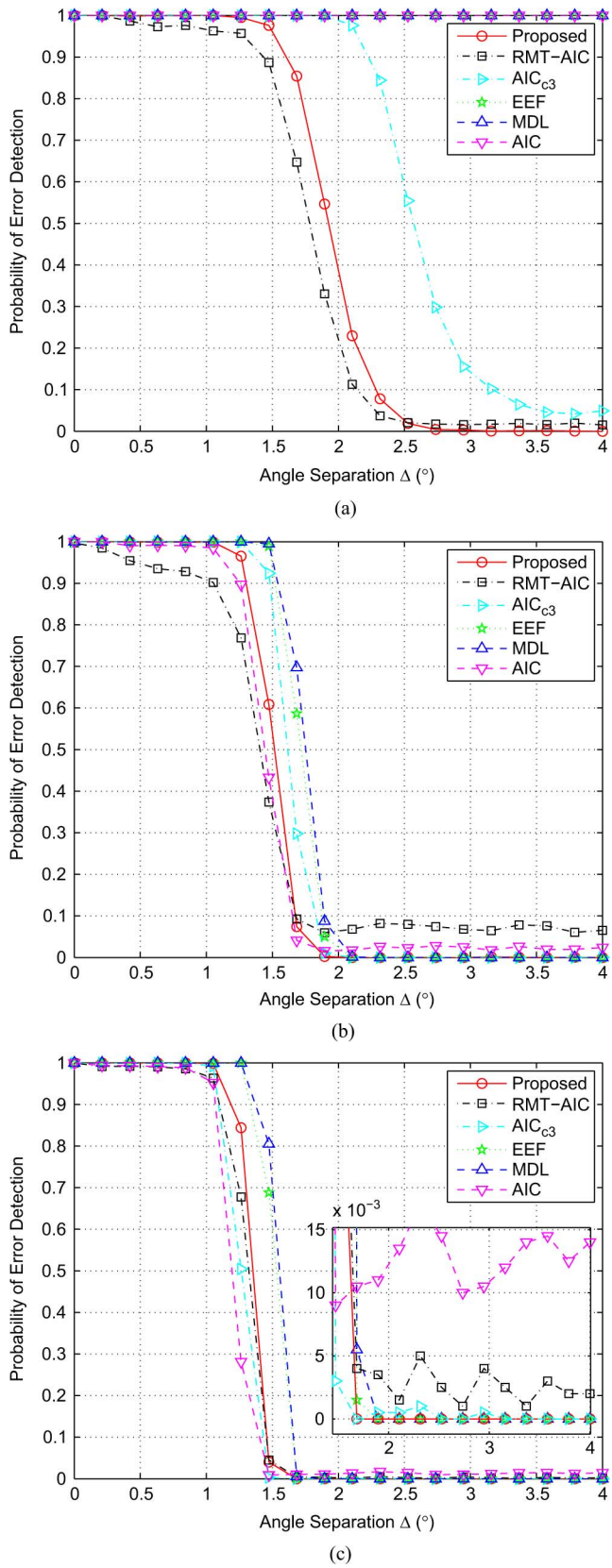


Fig. 7. Probability of error detection versus angle separation Δ. $[\phi_1, \phi_2, \phi_3] = [0, \Delta, 2\Delta]$ and $\mathrm{SNR} = 3\mathrm{dB}$.

The probability of error detection versus the angle separation is shown in Fig. 7. The impinging angles of three signals are set as $[\phi_1, \phi_2, \phi_3] = [0, \Delta, 2\Delta]$ where Δ stands for a resolution variable. It is observed that the proposed criterion is able to correctly detect the source number for $m = 3n$, $m = n$ and $m = n/3$ with $m = 30$. Although the AIC_{c_3} approach can enumerate the sources for $m = 3n$ and $m = n$, it still has a performance gap from the proposed criterion when the angle separation is small, so do the MDL and EEF schemes for $m = n$ and $m = n/3$. Though the RMT-AIC method can detect the source number with high probability, its error detection probability cannot reach zero, even for a large angle separation.

To compare the ability of the criteria to identify source number, we plot the probability of error detection versus the source number. Let $\boldsymbol{\theta} = [0°, 3.5°, 7°, \ldots, 42°, 45.5°]^T$, corresponding to 14 DOAs. The DOAs of d incident signal sources are defined as $\boldsymbol{\phi} = \boldsymbol{\theta}(1:1:d)$ where $d \leq 14$ and $(i:z:j)$ means selecting the ith to the jth elements with step size z. We observe in Figs. 8(a)–8(b) that the RMT-AIC scheme provides the best identifiability for the small sample size but cannot reach zero in probability of error detection even when the source number is sufficiently small, say, 2. Compared with the RMT-AIC method, the proposed criterion does not improve in identifiability, but attains error detection probability of zero when the source number is small. Meanwhile, the proposed method significantly outperforms the AIC_{c_3}, EEF and MDL approaches when $n \leq m$. As n becomes larger, say, $n = 90$, Fig. 8(c) implies that the proposed criterion is superior to the MDL and RMT-AIC criteria, but is inferior to the AIC_{c_3}, EEF and AIC methods.

The computational times required by the information theoretic criteria are plotted in Fig. 9 at $m = 3n$, $m = n$ and $m = n/3$. It is seen that the proposed criterion requires similar computational cost as the other criteria. It is easy to interpret the results by noticing that all the criteria involve the computation of SCM and its EVD, requiring $\mathcal{O}(m^2 n) + \mathcal{O}(m^3)$ flops. Nevertheless, the shrunken eigenvalues yielded by the proposed method is only the linear function of the sample eigenvalues, which is much computationally simpler than the computation of sample covariance matrix and its EVD. Hence, the proposed method requires the same order of computational cost as the existing information theoretic criteria.

IV. Conclusion

An LS-MDL method for source enumeration has been devised. By utilizing the identity structure of the covariance matrix and Gaussian assumption of the noise subspace components, an accurate estimator for the covariance matrix of the noise subspace components is derived. Compared with the existing source enumerators, the LS-MDL criterion is able to achieve more reliable and accurate detection of the source number in severe environments where the number of snapshots can be equal to or smaller than the number of sensors. Moreover, the LS-MDL approach is capable of providing consistent estimate of the source number in the general asymptotic case, which is not shared with the other methods. Simulation results are in line with the theoretical analysis.

Appendix A
Derivation of (12)

It follows from [19] (Eq. (17)) that the total code length encoding $\{\boldsymbol{y}(t)\}$ can be expressed as

$$\mathcal{L}_a\{\boldsymbol{y}_{\mathcal{S}}^{(k)}(t), \boldsymbol{y}_{\mathcal{N}}^{(k)}(t) | \boldsymbol{U}\} = n \log\left(\left(\det[\boldsymbol{S}_{ss}^{(k)}]\right)\left(\frac{\mathrm{tr}[\boldsymbol{S}_{nn}^{(k)}]}{m-k}\right)^{(m-k)}\right) + \frac{1}{2}(k^2+1)\log n \quad (33)$$

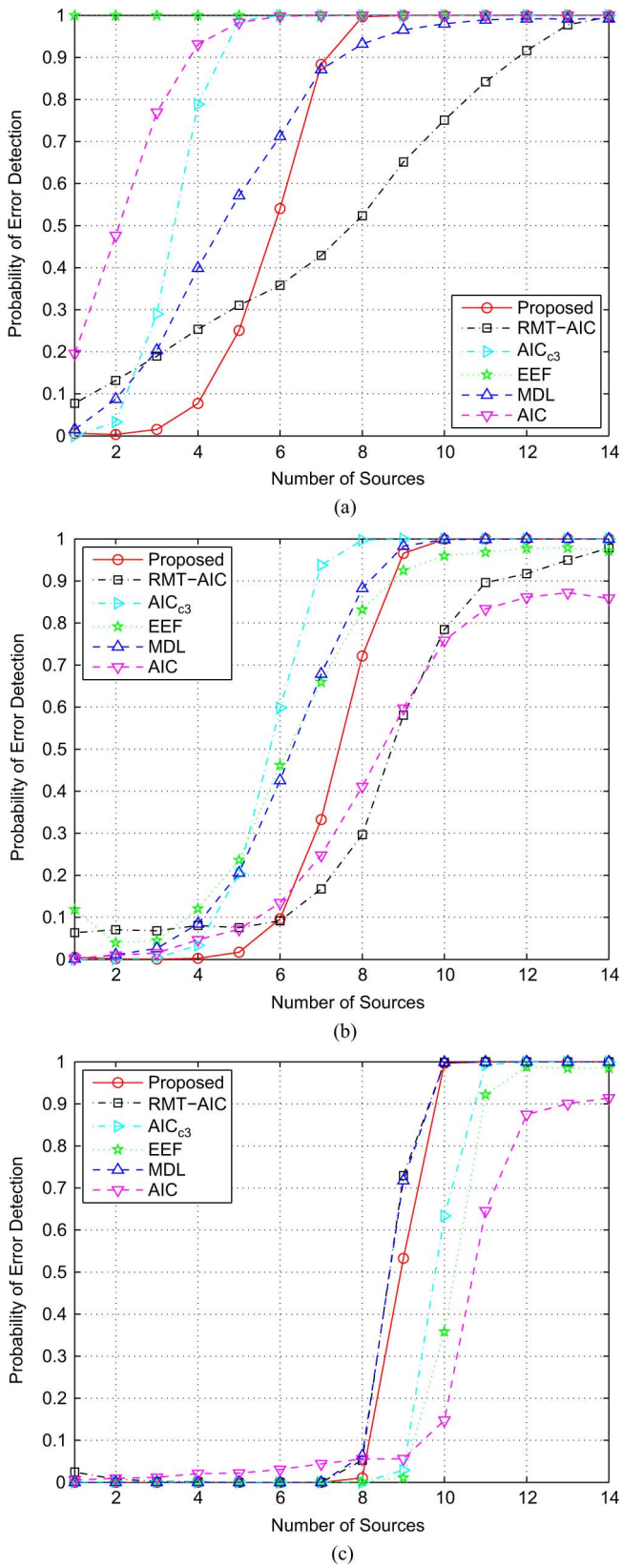


Fig. 8. Probability of error detection versus number of sources. $\bar{m} = 16$ and $\mathrm{SNR} = -5$ dB.

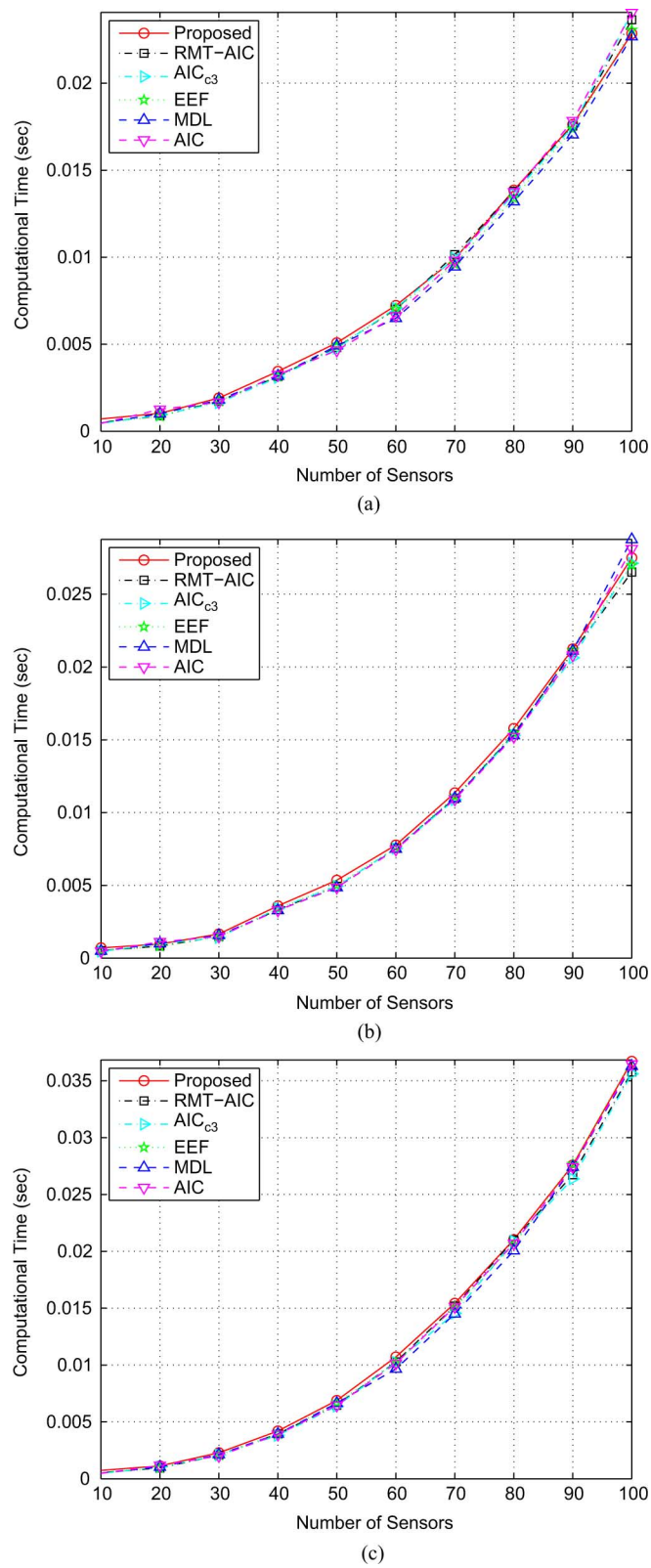


Fig. 9. Computational time versus number of sensors $[\phi_1, \phi_2, \phi_3] = [2.3°, 9.8°, 17.5°]$ and $\mathrm{SNR} = 3\mathrm{dB}$.

where

$$\boldsymbol{S}_{ss}^{(k)} = \boldsymbol{U}_k^H \boldsymbol{S} \boldsymbol{U}_k \tag{34}$$

$$\boldsymbol{S}_{nn}^{(k)} = \boldsymbol{U}_{m-k}^H \boldsymbol{S} \boldsymbol{U}_{m-k}. \tag{35}$$

In fact, it is indicated in [27] that, to further improve the detection performance of the information theoretic criteria by accurately determining the parameter vector, the k eigenvalues of $\boldsymbol{S}_{ss}^{(k)}$ should be excluded in the parameter vector. This in turn means that the number of free parameters should be reduced to $(k^2 - k + 1)$.

Notice that $\boldsymbol{y}_{\mathcal{S}}^{(k)}(t)$ is statistically independent of $\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)$. Invoking the results in [13], [45], that is,

$$\boldsymbol{S}_{nn}^{(k)} = \tau \boldsymbol{I}_{m-k} + \mathcal{O}\left(\sqrt{\frac{\log\log n}{n}}\right) \tag{36}$$

$$\boldsymbol{S}_{sn}^{(k)} \triangleq \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{y}_{\mathcal{S}}^{(k)}(t)\left(\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)\right)^H = \mathcal{O}\left(\sqrt{\frac{\log\log n}{n}}\right) \tag{37}$$

we obtain

$$\boldsymbol{S}_{sn}^{(k)}\left(\boldsymbol{S}_{nn}^{(k)}\right)^{-1}\boldsymbol{S}_{ns}^{(k)} = \mathcal{O}\left(\frac{\log\log n}{n}\right) \tag{38}$$

where $\boldsymbol{S}_{ns}^{(k)} = (\boldsymbol{S}_{sn}^{(k)})^H$. It follows from (38) that $\boldsymbol{S}_{sn}^{(k)}(\boldsymbol{S}_{nn}^{(k)})^{-1}\boldsymbol{S}_{ns}^{(k)}$ approaches zero at the rate of $\log\log n/n$, which allows us to express $\det[\boldsymbol{S}]$ as

$$\begin{aligned}\det[\boldsymbol{S}] &= \det\left(\begin{bmatrix}\boldsymbol{S}_{ss}^{(k)} & \boldsymbol{S}_{sn}^{(k)}\\ \boldsymbol{S}_{ns}^{(k)} & \boldsymbol{S}_{nn}^{(k)}\end{bmatrix}\right)\\ &= \det\left[\boldsymbol{S}_{nn}^{(k)}\right] \times \det\left(\boldsymbol{S}_{ss}^{(k)} - \boldsymbol{S}_{sn}^{(k)}\left(\boldsymbol{S}_{nn}^{(k)}\right)^{-1}\boldsymbol{S}_{ns}^{(k)}\right)\\ &\approx \det\left[\boldsymbol{S}_{nn}^{(k)}\right]\det\left[\boldsymbol{S}_{ss}^{(k)}\right]\end{aligned} \tag{39}$$

in which we have used the fact that $\det[\boldsymbol{U}^H\boldsymbol{S}\boldsymbol{U}] = \det[\boldsymbol{S}]$. Note that the asymptotic (approximate) equality in (39) is valid because $\boldsymbol{S}_{ss}^{(k)}$ is much larger than $\boldsymbol{S}_{sn}^{(k)}(\boldsymbol{S}_{nn}^{(k)})^{-1}\boldsymbol{S}_{ns}^{(k)}$ which converges to zero at the rate of $\log\log n/n$. Substituting $\det[\boldsymbol{S}_{ss}^{(k)}] \approx \det[\boldsymbol{S}]/\det[\boldsymbol{S}_{nn}^{(k)}]$ into (33) and recalling that the number of free parameters in the parameter vector is (k^2-k+1) yield

$$\mathcal{L}_a\{\boldsymbol{y}_{\mathcal{S}}^{(k)}(t),\boldsymbol{y}_{\mathcal{N}}^{(k)}(t)|\boldsymbol{U}\} \approx n\log(\det[\boldsymbol{S}]) + n\log\left(\frac{\left(\frac{\operatorname{tr}\left[\boldsymbol{S}_{nn}^{(k)}\right]}{m-k}\right)^{(m-k)}}{\det\left[\boldsymbol{S}_{nn}^{(k)}\right]}\right) + \frac{1}{2}(k^2-k+1)\log n. \tag{40}$$

Note that $\det[\boldsymbol{S}]$ is not a function of k. As a result, ignoring the terms independent of k, i.e., $n\log(\det[\boldsymbol{S}])$ and $(1/2)\log n$, it follows from (40) that the total code length for encoding $\{\boldsymbol{y}(t)\}$ is asymptotically expressed as (12).

Appendix B
Derivation of (26)

To complete the derivation of (26), we need the following results.

Lemma 1: Let $\ell_{k+1} \geq \cdots \geq \ell_m$ be sample eigenvalues associated with the SCM of the $(m-k)\times n$ IID Gaussian observations with mean zero and variance τ. As $m, n \to \infty$ and $m/n \to c \in (0,\infty)$, we have

$$(m-k)\left(\begin{bmatrix}\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i\\ \frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i^2\end{bmatrix} - \begin{bmatrix}\tau\\ \tau^2(1+c)\end{bmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{o}_2, \boldsymbol{D}) \tag{41}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and

$$\boldsymbol{D} = \begin{bmatrix}\tau^2 c & 2\tau^3 c(1+c)\\ 2\tau^3 c(1+c) & 2\tau^4 c(2c^2+5c+2)\end{bmatrix}. \tag{42}$$

Proof: Noting that $m-k \to m$ as $m \to \infty$, Lemma 1 turns out to be Proposition 3.2 in [39], which has been proved in [4] by means of the method of moments for real number and in [9] by means of the Stieltjes transform for complex number. ■

It is indicated in Lemma 1 that $\sum_{i=k+1}^{m}\ell_i/(m-k)$ and $\sum_{i=k+1}^{m}\ell_i^2/(m-k)$ are the unbiased estimates of τ and $\tau^2(1+c)$, respectively. Meanwhile, their variances converge to zero as $m, n \to \infty$ and $m/n \to c \in (0,\infty)$, that is,

$$\operatorname{var}\left(\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i - \tau\right) = \frac{\tau^2 c}{(m-k)^2} \to 0 \tag{43a}$$

$$\operatorname{var}\left(\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i^2 - \tau^2(1+c)\right) = \frac{2\tau^4 c(2c^2+5c+2)}{(m-k)^2} \to 0 \tag{43b}$$

which results in (26).

Appendix C
Proof of Proposition 1

Substituting (28) into (30) and recalling $\boldsymbol{\Sigma}_{nn} = \tau\boldsymbol{I}_{m-k}$ for $k \geq d$ yield

$$\mathbb{E}\left[\left\|\hat{\boldsymbol{R}} - \boldsymbol{\Sigma}_{nn}\right\|_F^2\right] = \mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\beta\hat{\tau} + (1-\beta)\ell_i - \tau)^2\right] \tag{44}$$

where we have suppressed the superscript $(\cdot)^{(k)}$ for notation simplicity. As $\hat{\alpha} \geq 0$ and

$$\beta = \begin{cases}\hat{\alpha} & 0 \leq \hat{\alpha} < 1\\ 1 & \hat{\alpha} \geq 1\end{cases} \tag{45}$$

it follows from (44) that

$$\begin{aligned}&\mathbb{E}\left[\|\hat{\boldsymbol{R}} - \boldsymbol{\Sigma}_{nn}\|_F^2\right]\\ &= \int_0^{+\infty} f(\hat{\alpha})\mathbb{E}\left[\|\hat{\boldsymbol{R}} - \boldsymbol{\Sigma}_{nn}\|_F^2 \middle| \hat{\alpha}\right] d\hat{\alpha}\\ &= \int_0^{1} f(\hat{\alpha})\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\beta\hat{\tau} + (1-\beta)\ell_i - \tau)^2 \middle| \hat{\alpha}\right] d\hat{\alpha}\\ &+ \int_1^{+\infty} f(\hat{\alpha})\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\beta\hat{\tau} + (1-\beta)\ell_i - \tau)^2 \middle| \hat{\alpha}\right] d\hat{\alpha}\\ &= \int_1^{+\infty} f(\hat{\alpha})\mathbb{E}\left[(\hat{\tau}-\tau)^2 \middle| \hat{\alpha}\right] d\hat{\alpha}\\ &\quad + \int_0^{1} f(\hat{\alpha})\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\hat{\alpha}\hat{\tau} + (1-\hat{\alpha})\ell_i - \tau)^2 \middle| \hat{\alpha}\right] d\hat{\alpha}\end{aligned} \tag{46}$$

where $f(\hat{\alpha})$ denotes the probability density function of $\hat{\alpha}$ and we have used the equality $\mathbb{E}[X] = \mathbb{E}\{\mathbb{E}[X|Y]\}$ for random variables X and Y. Using again this equality along with the results in Lemma 1, that is, $(m-k)(\hat{\tau}-\tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2 c)$, the first term of (46) is calculated as

$$\begin{aligned}\int_1^{+\infty} f(\hat{\alpha})\mathbb{E}\left[(\hat{\tau}-\tau)^2 \middle| \hat{\alpha}\right] d\hat{\alpha} &\leq \int_0^{+\infty} f(\hat{\alpha})\mathbb{E}\left[(\hat{\tau}-\tau)^2 \middle| \hat{\alpha}\right] d\hat{\alpha}\\ &= \mathbb{E}\left[(\hat{\tau}-\tau)^2\right] = \frac{\tau^2 c}{(m-k)^2}.\end{aligned} \tag{47}$$

Now let us examine the second term of (46). Setting $e \triangleq \tau - \hat{\tau}$, we readily obtain from Lemma 1 that $e = \mathcal{O}(1/(m-k))$. As a result, it is computed as

$$\begin{aligned}
&\int_0^1 f(\hat{\alpha})\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\hat{\alpha}\hat{\tau}+(1-\hat{\alpha})\ell_i-\tau)^2\middle|\hat{\alpha}\right]d\hat{\alpha}\\
&=\int_0^1 f(\hat{\alpha})\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}((1-\hat{\alpha})(\ell_i-\hat{\tau})-e)^2\middle|\hat{\alpha}\right]d\hat{\alpha}\\
&=\int_0^1 f(\hat{\alpha})(1-\hat{\alpha})^2\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\ell_i-\hat{\tau})^2\right]d\hat{\alpha}+\int_0^1 f(\hat{\alpha})\mathbb{E}[e^2]d\hat{\alpha}.
\end{aligned} \tag{48}$$

It follows from Lemma 1 that

$$\mathbb{E}\left[1/(m-k)\sum_{i=k+1}^{m}(\ell_i-\hat{\tau})^2\right]=\tau^2c(1-(m-k)^{-2})$$

and $\mathbb{E}[e^2]=\tau^2c/(m-k)^2$. Therefore, substituting these results into (48) and noticing that $\int_0^1 f(\hat{\alpha})d\hat{\alpha}\le\int_0^{+\infty}f(\hat{\alpha})d\hat{\alpha}=1$, we obtain

$$\begin{aligned}
&\int_0^1 f(\hat{\alpha})\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\hat{\alpha}\hat{\tau}+(1-\hat{\alpha})\ell_i-\tau)^2\middle|\hat{\alpha}\right]d\hat{\alpha}\\
&\le\tau^2c\left(1-\frac{1}{(m-k)^2}\right)\int_0^1 f(\hat{\alpha})(1-\hat{\alpha})^2d\hat{\alpha}+\frac{\tau^2c}{(m-k)^2}\\
&<\tau^2c\int_0^{+\infty}f(\hat{\alpha})(1-\hat{\alpha})^2d\hat{\alpha}+\frac{\tau^2c}{(m-k)^2}\\
&=\tau^2c\mathbb{E}\left[(1-\hat{\alpha})^2\right]+\frac{\tau^2c}{(m-k)^2}.
\end{aligned} \tag{49}$$

It follows from (27) that $1-\hat{\alpha}$ can be expressed as

$$\begin{aligned}
1-\hat{\alpha}&=1-\frac{\frac{\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i^2}{\left(\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i\right)^2}+(m-k)}{(n+1)\left(\frac{\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i^2}{\left(\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i\right)^2}-1\right)}\\
&\triangleq 1-\frac{z+(m-k)}{(n+1)(z-1)}\triangleq g(z)
\end{aligned} \tag{50}$$

where $z=(\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i^2)/(\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i)^2$. It is shown in [39] that $z-z_0\xrightarrow{\mathcal{D}}\mathcal{N}(0,2c^2/(m-k)^2)$ with $z_0=1+c$. In the sequel, using the Taylor series expansion of $g(z)$ around z_0 yields $g(z)=g(z_0)+g'(z_0)(z-z_0)+\frac{1}{2}g''(z_0)(z-z_0)^2+\cdots$, where $g'(z)$ and $g''(z)$ denote the first-order and second-order derivatives of $g(z)$. It is easy to obtain $g(z_0)=(cn-(m-k+1))/(c(n+1))$, $g'(z_0)=(m-k+1)/((n+1)c^2)$ and $g''(z_0)=-2(m-k+1)/((n+1)c^3)$. With these results, $g(z)$ is calculated as

$$\begin{aligned}
g(z)&=\frac{cn-(m-k+1)}{c(n+1)}+\frac{m-k+1}{c^2(n+1)}(z-z_0)\\
&\quad-\frac{m-k+1}{c^3(n+1)}(z-z_0)^2+\cdots\\
&=\frac{-1}{n+1}+\frac{1}{c}(z-z_0)-\frac{1}{c^2}(z-z_0)^2+\cdots
\end{aligned} \tag{51}$$

where we have set $c=(m-k+1)/(n+1)$ due to [9] which argues that the unknown limiting parameter c can be replaced by $(m-k+1)/(n+1)$ for large m and n. It follows from (51) that the mean and variance of $g(z)$ are, respectively, given as

$$\mathbb{E}[g(z)]=\frac{-1}{n+1}-\frac{2}{(m-k)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right) \tag{52}$$

and

$$\mathrm{var}[g(z)]=\frac{2}{(m-k)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right). \tag{53}$$

As a result, using $\mathbb{E}[g^2(z)]=\mathrm{var}[g(z)]+\mathbb{E}^2[g(z)]$ and $c=(m-k+1)/(n+1)$, we obtain

$$\begin{aligned}
\mathbb{E}\left[(1-\hat{\alpha})^2\right]&=\frac{2}{(m-k)^2}+\frac{1}{(n+1)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right)\\
&=\frac{2}{(m-k)^2}+\frac{c^2}{(m-k+1)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right)\\
&<\frac{c^2+2}{(m-k)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right).
\end{aligned} \tag{54}$$

It follows from (49) and (54) that

$$\begin{aligned}
&\int_0^1 f(\hat{\alpha})\mathbb{E}\left[\frac{1}{m-k}\sum_{i=k+1}^{m}(\hat{\alpha}\hat{\tau}+(1-\hat{\alpha})\ell_i-\tau)^2\middle|\hat{\alpha}\right]d\hat{\alpha}\\
&<\frac{\tau^2c(c^2+3)}{(m-k)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right).
\end{aligned} \tag{55}$$

Thus, substituting (47) along with (55) into (46) and considering that $1/(m-k)^2\to 0$ as $m\to\infty$, we have

$$\mathbb{E}\left[\|\hat{\boldsymbol{R}}-\boldsymbol{\Sigma}_{nn}\|_F^2\right]<\frac{\tau^2c(c^2+4)}{(m-k)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right)\to 0. \tag{56}$$

This completes the proof of Proposition 1.

Appendix D
Proof of Proposition 2

It follows from (56) that

$$\frac{1}{m-k}\sum_{i=k+1}^{m}\mathbb{E}\left[\left(\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2\right]<\frac{\tau^2c(c^2+4)}{(m-k)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right). \tag{57}$$

Note that

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2\right]&<\left[\frac{\tau^2c(c^2+4)}{(m-k)^2}+\mathcal{O}\left(\frac{1}{(m-k)^3}\right)\right]\\
&\times\frac{\max\limits_{i\in[d+1,m]}\mathbb{E}\left[\left(\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2\right]}{\frac{1}{m-k}\sum_{i=k+1}^{m}\mathbb{E}\left[\left(\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2\right]}.
\end{aligned} \tag{58}$$

Meanwhile, for $k\ge d$ and $m,n\to\infty$ with $m/n\to c$, it is easy to verify from (27) along with (41) that $\hat{\alpha}^{(k)}\to 1$, which, when inserting into (29), leads to $\rho_i^{(k)}\to\hat{\tau}$ for $i=d+1,\ldots,m$. This thereby indicates that $\max_{i\in[d+1,m]}\mathbb{E}[(\frac{\rho_i^{(k)}-\tau}{\tau})^2]$ and

$(\frac{1}{m-k}\sum_{i=k+1}^{m}\mathbb{E}[(\frac{\rho_i^{(k)}-\tau}{\tau})^2])$ are in the same order of $(m-k)^{-2}$, whose ratio is a bounded constant number. Utilizing $(\mathbb{E}[X])^2 \le \mathbb{E}[X^2]$ for random variable X, it follows from (58) that

$$\mathbb{E}\left[\frac{\rho_i^{(k)}-\tau}{\tau}\right] = \mathcal{O}\left(\frac{1}{m-k}\right), \tag{59}$$

$$\mathbb{E}\left[\left(\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2\right] = \mathcal{O}\left(\frac{1}{(m-k)^2}\right) \tag{60}$$

where $i = k+1, \ldots, m,\ k = d, \ldots, \bar{m}$. In the sequel, setting $\mathbb{E}[\frac{\rho_i^{(k)}-\tau}{\tau}] = c_1/(m-k)$ and $\mathbb{E}[(\frac{\rho_i^{(k)}-\tau}{\tau})^2] = c_2/(m-k)^2$ with c_1 and c_2 being constant numbers, we can assert that $\frac{\rho_i^{(k)}-\tau}{\tau}$ is Gaussian distributed with mean $c_1/(m-k)$ and variance $(c_2-c_1^2)/(m-k)^2$ due to central limit theorem, i.e.,

$$(m-k)\left(\frac{\rho_i^{(k)}-\tau}{\tau} - \frac{c_1}{m-k}\right) \sim \mathcal{N}(0, c_2 - c_1^2). \tag{61}$$

As a result, we have

$$\frac{\rho_i^{(k)}-\tau}{\tau} = \frac{c_1}{m-k} + \mathcal{O}\left(\frac{1}{m-k}\right) = \mathcal{O}\left(\frac{1}{m-k}\right). \tag{62}$$

Consider first the probability of overestimating the source number, which is expressed as

$$\begin{aligned}
&\Pr\{\mathrm{LS-MDL}(k) < \mathrm{LS-MDL}(d) | k > d\} \\
&= \Pr\left\{\frac{1}{2}((k-d)(k+d-1))\frac{\log n}{n} < (m-d)\log\frac{\frac{1}{m-d}\sum_{i=d+1}^{m}\rho_i^{(d)}}{\left(\prod_{i=d+1}^{m}\rho_i^{(d)}\right)^{\frac{1}{m-d}}}\right. \\
&\qquad \left. - (m-k)\log\frac{\frac{1}{m-k}\sum_{i=k+1}^{m}\rho_i^{(k)}}{\left(\prod_{i=k+1}^{m}\rho_i^{(k)}\right)^{\frac{1}{m-k}}}\middle| k > d\right\}.
\end{aligned} \tag{63}$$

Utilizing (62), the right hand side (RHS) within the braces of (63) is calculated as shown in (64) at the bottom of the page,

$$\begin{aligned}
\mathrm{RHS} &= (m-d)\log\left(\frac{1}{m-d}\sum_{i=d+1}^{m}\rho_i^{(d)}\right) - \log\prod_{i=d+1}^{m}\rho_i^{(d)} - (m-k)\log\left(\frac{1}{m-k}\sum_{i=k+1}^{m}\rho_i^{(k)}\right) + \log\prod_{i=k+1}^{m}\rho_i^{(k)} \\
&= (m-d)\log\tau\left(1+\frac{1}{m-d}\sum_{i=d+1}^{m}\frac{\rho_i^{(d)}-\tau}{\tau}\right) - \sum_{i=d+1}^{m}\log\tau\left(1+\frac{\rho_i^{(d)}-\tau}{\tau}\right) \\
&\quad -(m-k)\log\tau\left(1+\frac{1}{m-k}\sum_{i=k+1}^{m}\frac{\rho_i^{(k)}-\tau}{\tau}\right) + \sum_{i=k+1}^{m}\log\tau\left(1+\frac{\rho_i^{(k)}-\tau}{\tau}\right) \\
&\overset{(a)}{=} (m-d)\left(\frac{1}{m-d}\sum_{i=d+1}^{m}\frac{\rho_i^{(d)}-\tau}{\tau} - \frac{1}{2}\left(\frac{1}{m-d}\sum_{i=d+1}^{m}\frac{\rho_i^{(d)}-\tau}{\tau}\right)^2 + \mathcal{O}\big((m-d)^{-3}\big)\right) \\
&\quad -\sum_{i=d+1}^{m}\frac{\rho_i^{(d)}-\tau}{\tau} + \frac{1}{2}\sum_{i=d+1}^{m}\left(\frac{\rho_i^{(d)}-\tau}{\tau}\right)^2 - \sum_{i=d+1}^{m}\mathcal{O}\big((m-d)^{-3}\big) \\
&\quad -(m-k)\left(\frac{1}{m-k}\sum_{i=k+1}^{m}\frac{\rho_i^{(k)}-\tau}{\tau} - \frac{1}{2}\left(\frac{1}{m-k}\sum_{i=k+1}^{m}\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2 + \mathcal{O}\big((m-k)^{-3}\big)\right) \\
&\quad +\sum_{i=k+1}^{m}\frac{\rho_i^{(k)}-\tau}{\tau} - \frac{1}{2}\sum_{i=k+1}^{m}\left(\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2 + \sum_{i=k+1}^{m}\mathcal{O}\big((m-k)^{-3}\big) \\
&= \frac{1}{2}\left(\sum_{i=d+1}^{m}\left(\frac{\rho_i^{(d)}-\tau}{\tau}\right)^2 - \frac{1}{m-d}\left(\sum_{i=d+1}^{m}\frac{\rho_i^{(d)}-\tau}{\tau}\right)^2\right) - \mathcal{O}\big((m-d)^{-2}\big) \\
&\quad -\frac{1}{2}\left(\sum_{i=k+1}^{m}\left(\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2 - \frac{1}{m-k}\left(\sum_{i=k+1}^{m}\frac{\rho_i^{(k)}-\tau}{\tau}\right)^2\right) + \mathcal{O}\big((m-k)^{-2}\big) \\
&= \mathcal{O}\left(\frac{1}{m-d}\right) - \mathcal{O}\left(\frac{1}{m-k}\right) \xrightarrow{(b)} \mathcal{O}\left(\frac{1}{m}\right)
\end{aligned} \tag{64}$$

where (a) is due to $\log(1+x) = x - x^2/2 + x^3/3 - \cdots$ for a small number x whereas (b) is due to the fact that d and k are fixed numbers while m tends to infinity. As a result, substituting (64) into (63) yields

$$\begin{aligned}
&\Pr\{\mathrm{LS-MDL}(k) < \mathrm{LS-MDL}(d) | \, k > d\} \\
&\xrightarrow{m \to \infty} \Pr\left\{ \frac{1}{2}(k-d)(k+d-1)\log n < \mathcal{O}\left(\frac{n}{m}\right) \middle| \, k > d \right\} \\
&\xrightarrow{m/n \to c} \Pr\left\{ \frac{1}{2}(k-d)(k+d-1)\log n < \mathcal{O}\left(\frac{1}{c}\right) \middle| \, k > d \right\} \\
&\xrightarrow{n \to \infty} 0. \qquad (65)
\end{aligned}$$

Consider now the probability of underestimating the source number, which turns out to be

$$\begin{aligned}
&\Pr\{\mathrm{LS-MDL}(k) < \mathrm{LS-MDL}(d) | k < d\} \\
&= \Pr\left\{ (m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \rho_i^{(k)}}{\left(\prod_{i=k+1}^{m} \rho_i^{(k)}\right)^{\frac{1}{m-k}}} - (m-d)\log \frac{\frac{1}{m-d}\sum_{i=d+1}^{m} \rho_i^{(d)}}{\left(\prod_{i=d+1}^{m} \rho_i^{(d)}\right)^{\frac{1}{m-d}}} \right. \\
&\left. < \frac{1}{2}(d-k)(d+k-1)\frac{\log n}{n} \middle| k < d \right\}. \qquad (66)
\end{aligned}$$

Employing the results in (64), we can easily verify that

$$(m-d)\log \frac{\frac{1}{m-d}\sum_{i=d+1}^{m} \rho_i^{(d)}}{\left(\prod_{i=d+1}^{m} \rho_i^{(d)}\right)^{\frac{1}{m-d}}} \to \mathcal{O}\left(\frac{1}{m}\right), \text{ for } m \to \infty. \qquad (67)$$

As a result, for $m, n \to \infty$ and $m/n \to c$, the probability of underestimation can be rewritten as

$$\begin{aligned}
&\Pr\{\mathrm{LS-MDL}(k) < \mathrm{LS-MDL}(d) | \, k < d\} \\
&\xrightarrow{m \to \infty} \Pr\left\{ \frac{n}{\log n}\gamma - \mathcal{O}\left(\frac{1}{m/n \log n}\right) < \frac{1}{2}(d-k)(d+k-1) \middle| k < d \right\} \\
&\xrightarrow{m/n \to c} \Pr\left\{ \frac{n}{\log n}\gamma - \mathcal{O}\left(\frac{1}{c \log n}\right) < \frac{1}{2}(d-k)(d+k-1) \middle| k < d \right\} \qquad (68)
\end{aligned}$$

where

$$\gamma = (m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \rho_i^{(k)}}{\left(\prod_{i=k+1}^{m} \rho_i^{(k)}\right)^{\frac{1}{m-k}}}. \qquad (69)$$

Obviously, the probability of underestimating the source number converges to zero provided that $\gamma > 0$ because $n/\log n \to \infty$ and $1/(c \log n) \to 0$ as $n \to \infty$. Actually, noticing that $\mathcal{O}(1/m) \to 0$ as $m \to \infty$, it follows from (67) that

$$\left(\prod_{i=d+1}^{m} \rho_i^{(k)}\right)^{\frac{1}{m-d}} \to \frac{1}{m-d}\sum_{i=d+1}^{m} \rho_i^{(k)}, \text{ as } m \to \infty. \qquad (70)$$

Meanwhile, for $k < d$, it is easy to verify from (29) that

$$\rho_{k+1}^{(k)} > \cdots > \rho_d^{(k)} \qquad (71)$$

indicating

$$\frac{1}{d-k}\sum_{i=k+1}^{d} \rho_i^{(k)} > \left(\prod_{i=k+1}^{d} \rho_i^{(k)}\right)^{\frac{1}{d-k}}. \qquad (72)$$

As a result, with the use of (70) and (72), we can calculate γ as

$$\begin{aligned}
\gamma &= (m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{d} \rho_i^{(k)} + \frac{m-d}{m-k}\left(\frac{1}{m-d}\sum_{i=d+1}^{m} \rho_i^{(k)}\right)}{\left(\prod_{i=k+1}^{d} \rho_i^{(k)}\right)^{\frac{1}{m-k}} \cdot \left(\prod_{i=d+1}^{m} \rho_i^{(k)}\right)^{\frac{1}{m-d}\cdot\frac{m-d}{m-k}}} \\
&> (m-k)\log \frac{\frac{d-k}{m-k}\left(\prod_{i=k+1}^{d} \rho_i^{(k)}\right)^{\frac{1}{d-k}} + \frac{m-d}{m-k}\left(\frac{1}{m-d}\sum_{i=d+1}^{m} \rho_i^{(k)}\right)}{\left(\left(\prod_{i=k+1}^{d} \rho_i^{(k)}\right)^{\frac{1}{d-k}}\right)^{\frac{d-k}{m-k}} \cdot \left(\prod_{i=d+1}^{m} \rho_i^{(k)}\right)^{\frac{1}{m-d}\cdot\frac{m-d}{m-k}}} \\
&\xrightarrow{m \to \infty} (m-k)\log \frac{\frac{d-k}{m-k}\left(\prod_{i=k+1}^{d} \rho_i^{(k)}\right)^{\frac{1}{d-k}} + \frac{m-d}{m-k}\left(\frac{1}{m-d}\sum_{i=d+1}^{m} \rho_i^{(k)}\right)}{\left(\left(\prod_{i=k+1}^{d} \rho_i^{(k)}\right)^{\frac{1}{d-k}}\right)^{\frac{d-k}{m-k}} \cdot \left(\frac{1}{m-d}\sum_{i=d+1}^{m} \rho_i^{(k)}\right)^{\frac{m-d}{m-k}}}.
\end{aligned} \qquad (73)$$

Invoking the generalized arithmetic-mean geometric-mean inequality, that is,

$$\frac{\omega_1 A_1 + \omega_2 A_2}{A_1^{\omega_1} A_2^{\omega_2}} \geq 1, \quad \omega_1 + \omega_2 = 1, \qquad (74)$$

we obtain from (73) that $\gamma > 0$ which, when inserted into (68), leads to

$$\Pr\{\mathrm{LS-MDL}(k) < \mathrm{LS-MDL}(d) | k < d\} \to 0, \quad \text{as } m, n \to \infty \text{ and } m/n \to c. \qquad (75)$$

Thus, it follows from (65) and (75) that

$$\Pr\left\{\hat{d} = d\right\} \to 1, \text{as } m, n \to \infty \text{ and } m/n \to c. \qquad (76)$$

This completes the proof of Proposition 2.

References

[1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276–280, 1986.

[2] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT-A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[3] P. Stoica and K. C. Sharman, "Novel eigenanalysis method for direction estimation," *Proc. IEEE*, vol. 137, no. 1, pp. 19–26, Feb. 1990.

[4] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivar. Anal.*, vol. 12, pp. 1–38, 1982.

[5] K. Johansson, "On fluctuations of random Hermitian matrices," *Duke Math. J.*, vol. 91, pp. 151–203, 1998.

[6] I. Johnstone, "On the distribution of the largest eigenvalues in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, 2001.

[7] C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Commun. Math. Phys.*, vol. 177, pp. 727–754, 1996.

[8] Z. D. Bai, "Methodologies in spectral analysis of large-dimensional random matrices, a review," *Statistica Sinica*, vol. 9, pp. 611–677, 1999.

[9] Z. D. Bai and J. W. Silverstein, "CLT for linear spectral statistics of a large dimensional sample covariance matrix," *Ann. Prob.*, vol. 32, pp. 553–605, 2004.

[10] R. R. Nadakuditi and J. W. Silverstein, "Fundamental limit of sample generalized eigenvalue baseddetection of signals in noise using relatively few signal-bearing and noise-only samples," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 468–480, Jun. 2010.

[11] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.
[12] 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.
[13] 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.
[14] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 11, pp. 1959–1971, Nov. 1990.
[15] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
[16] J. Rissanen, *Information and Complexity in Statistical Modeling*. New York, NY, USA: Springer, 2007, p. 144.
[17] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 716–723, 1974.
[18] D. M. A. Haughton, "On the choice of a model to fit data from an exponential family," *Ann. Statist.*, vol. 16, pp. 342–355, 1988.
[19] M. Wax, "Detection and localization of multiple sources via the stochastic signals model," *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2450–2456, Nov. 1991.
[20] 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.
[21] G. Xu, R. H. I. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Process.*, vol. 42, pp. 102–112, Jan. 1994.
[22] S. Valaee and P. Kabal, "An inform. theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, May 2004.
[23] E. Fishler and H. V. Poor, "Estimation of the number of sources in unbalanced arrays via information theoretic criteria," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3543–3553, Sep. 2005.
[24] 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.
[25] L. Huang, T. Long, E. Mao, and H. C. So, "MMSE-based MDL method for robust estimation of number of sources without eigen decomposition," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4135–4142, Oct. 2009.
[26] C. D. Giurcaneanu, S. A. Razavi, and A. Liski, "Variable selection in linear regression: Several approaches based on normalized maximum likelihood," *Signal Process.*, vol. 91, no. 8, pp. 1671–1692, 2011.
[27] D. B. Williams, "Counting the degrees of freedom when using AIC and MDL to detect signals," *IEEE Trans. Signal Process.*, vol. 42, no. 11, Nov. 1994.
[28] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. J. Math. Statist.*, vol. 34, pp. 122–148, 1963.
[29] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, pp. 365–411, 2004.
[30] O. Ledoit and M. Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *J. Empir. Finance*, vol. 10, no. 5, pp. 603–621, Dec. 2003.
[31] O. Ledoit and M. Wolf, "Honey, I shrunk the sample covariance matrix," *J. Portfolio Manage.* , vol. 31, no. 1, 2004.
[32] E. Fishler, A. Haimovich, R. Blum, L. Cimini, D. Chizhik, and R. Valenzuela, "Spatial diversity in radars-Models and detection performance," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 823–838, Mar. 2006.
[33] J. Li, P. Stoica, L. Xu, and W. Roberts, "On parameter identifiability of MIMO radar," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 968–971, Dec. 2007.
[34] I. Bekkerman and J. Tabrikian, "Target detection and localization using MIMO radars and sonars," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3873–3883, Oct. 2006.
[35] X. Mestre and M. Lagunas, "Modified subspace algorithms for DOA estimation with large arrays," *IEEE Signal Process. Lett.*, vol. 56, no. 2, pp. 598–613, Feb. 2008.
[36] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.
[37] L. Du, J. Li, and P. Stoica, "Fully automatic computation of diagonal loading levels for robust adaptive beamforming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 449–458, Jan. 2010.
[38] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On using *a priori* knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598–2602, Jun. 2008.
[39] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, Jul. 2008.
[40] G. Letac and H. Massam, "All invariant moments of the Wishart distribution," *Scand. J. Statist.*, vol. 31, no. 2, pp. 285–318, 2004.
[41] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 1971, pt. I.
[42] A.-K. Seghouane, "Asymptotic bootstrap corrections of AIC for linear regression models," *Signal Process.*, vol. 90, no. 1, pp. 217–224, 2010.
[43] S. M. Kay, "Exponentially embedded families: A new approach to model order estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 333–345, Jan. 2005.
[44] C. Xu and S. Kay, "Source enumeration via the EEF criterion," *IEEE Signal Process. Lett.*, vol. 15, pp. 569–572, 2008.
[45] W. F. Stout, *Almost Sure Convergence*. New York, NY, USA: Academic, 1974.

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.

From 2005 to 2006, he was a Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC. From 2009 to 2010, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong and a Research Associate with the Department of Electronic Engineering, The Chinese University of Hong Kong. Since 2011, he has been with the Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School, where he is currently a Professor. His research interests include spectral estimation, array signal processing, statistical signal processing, and their applications in radar and wireless communication systems.

Dr. Huang is currently an editorial board member of Digital Signal Processing.

H. C. So (S'90–M'95–SM'07) was born in Hong Kong. He received the B.Eng. degree from the City University of Hong Kong and the Ph.D. degree from The Chinese University of Hong Kong, both in electronic engineering, in 1990 and 1995, respectively.

From 1990 to 1991, he was an Electronic Engineer with the Research and Development Division, Everex Systems Engineering Ltd., Hong Kong. During 1995–1996, he worked as a Postdoctoral Fellow with The Chinese University of Hong Kong. From 1996 to 1999, he was a Research Assistant Professor with the Department of Electronic Engineering, City University of Hong Kong, where he is currently an Associate Professor. His research interests include statistical signal processing, fast and adaptive algorithms, signal detection, parameter estimation, and source localization.

Dr. So has been on the editorial boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *Signal Processing*, *Digital Signal Processing*, and ISRN *Applied Mathematics*, as well as a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.