Bayesian Information Criterion for Source Enumeration in Large-Scale Adaptive Antenna Array

Lei Huang, *Senior Member, IEEE*, Yuhang Xiao, Kefei Liu, Hing Cheung So, *Fellow, IEEE*, and Jian-Kang Zhang, *Senior Member, IEEE*

***Abstract*—Subspace-based high-resolution algorithms for direction-of-arrival (DOA) estimation have been developed for large-scale adaptive antenna arrays. However, its prerequisite step, namely, source enumeration, has yet to be addressed. In this paper, a new approach is devised in the framework of the Bayesian information criterion (BIC) to provide reliable detection of the signal source number for *the general asymptotic regime*, where $m, n \to \infty$ and $m/n \to c \in (0, \infty)$, with m and n being the numbers of antennas and snapshots, respectively. In particular, the *a posteriori* probability is determined by correctly calculating the LLFs and PFs for the general asymptotic case. By means of the maximum *a posteriori* probability, we are capable of effectively finding the signal number. An accurate closed-form expression for the probability of missed detection is also derived for the proposed BIC variant. In addition, the probability of false alarm for the BIC detector is proved to converge to zero as $m, n \to \infty$ and $m/n \to c$. Simulation results are included to demonstrate the superiority of the proposed detection approach over state-of-the-art schemes and corroborate our theoretical calculations.**

***Index Terms*—Adaptive antenna array, Bayesian information criterion (BIC), direction-of-arrival (DOA) estimation, source enumeration.**

I. Introduction

As a promising technique to boost spectral efficiency, large-scale adaptive antenna arrays have received much attention in the literature [1], [2]. As the array utilizes a large number of antennas at the base station for transmission and reception, the conventional subspace-based algorithms for direction-of-arrival (DOA) estimation usually suffer serious performance degradation. This is because the subspace cannot be correctly determined for the situation where the number of antennas is comparable with the number of samples. To cope with the problem, more efficient subspace-based algorithms [3], [4] have been suggested for the large array. Nevertheless, as the prerequisite step of direction finding, source enumeration has not yet been addressed for such a situation, which turns out to be a big challenge, particularly at low signal-to-noise ratios (SNRs) or small samples.

The conventional source enumeration methodologies vary from hypothesis testing [5]–[8] to the information-theoretic criterion (ITC) [9]–[11]. Basically, the hypothesis testing, including the sphericity test [5] and the random matrix theory (RMT)-based test [8], needs to find a subjective threshold for decision making. It is shown in [8] that the RMT approach is able to provide a detection threshold that is significantly smaller than that of the classical minimum description length (MDL) method [12]. Unlike the hypothesis testing, the ITCs, such as Akaike's information criterion (AIC) [9], Schwarz's Bayesian information criterion (BIC) [13], Rissanen's MDL [14], and Kay's exponentially embedded family (EEF) [15], are derived from the perspective of information theory, and no user-defined parameter is needed. As a result, it is of considerable interest to exploit the information criterion for efficient source enumeration. Wax and Kailath [12] have employed the AIC and MDL to enumerate independent signal sources. To handle coherent signals, Wax and Ziskind [16] have combined the maximum likelihood (ML) estimates of the DOAs with the MDL principle for joint DOA estimation and source enumeration, ending up with an enhanced MDL criterion for coherent source enumeration. Valaee and Kabal [17] have proposed a predictive description length (PDL) for this task. Although the PDL method can outperform the MDL approach [16], it requires much more computational cost than the latter as ML estimation is required at each snapshot. Fishler and Poor [18] have reformulated the MDL criterion for source enumeration under nonuniform noise environment. Furthermore, they have proved the consistency of their proposed MDL variant. On the other hand, Huang *et al.* [19], [20] have developed MDL variants by using the filtered component variances or minimum mean square errors of the multistage Wiener filter rather than the sample eigenvalues corrupted by the nonuniform noise, ending up with computationally simple and robust source enumerators.

Most of the aforementioned methods are devised by utilizing the assumption that the number of antennas m is fixed while the number of snapshots n tends to infinity, which is referred to as the classical asymptotic regime. Indeed, the general asymptotic situation [21], where $m, n \to \infty$ and $m/n \to c \in (0, \infty)$, is more suitable to large-array applications since the number of antennas can be as large as the number of snapshots. On the other

Manuscript received February 26, 2015; revised May 8, 2015; accepted May 14, 2015. This work was supported by the National Natural Science Foundation of China under Grant 61222106 and Grant 61171187. The review of this paper was coordinated by Dr. T. Jiang.

L. Huang is with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: dr.lei.huang@ieee.org).

Y. Xiao is with the Department of Electronic and Information Engineering, Harbin Institute of Technology, Harbin 518055, China.

K. Liu is with the Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287-5406 USA.

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

J.-K. Zhang is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada.

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/TVT.2015.2436060

hand, it has been pointed out in [3] that the general asymptotic regime is able to provide a more accurate description for practical scenarios, where the number of snapshots and the number of antennas are finite and probably comparable in magnitude. In fact, the topics of DOA estimation and beamforming have been dealt with in [3] and [22]–[24] for the general asymptotic regime.

Basically, the ITCs have their roots in the minimization of the Kullback–Leibler (KL) information, but this minimization is carried out in the scenario where the number of antennas is fixed while the number of snapshots tends to infinity. This, in turn, means that these ITCs cannot properly work in the general asymptotic case. To enable the ITCs to properly detect the source number in this condition, Nadakuditi and Edelman [25] have devised the RMT-AIC criterion. Although the RMT-AIC is argued to be able to correctly detect the source number for the general asymptotic case, it cannot provide the consistent estimate of the source number [8]. To solve the issue of linear regression model order selection for small sample cases, variants of the AIC approach have been proposed in [26] by means of the asymptotic approximation of the bootstrap estimation of the KL information. Nevertheless, it is nontrivial to apply them to source enumeration for the large array. Therefore, it is considerably interesting to investigate the consistent methodology for source enumeration in the general asymptotic regime.

We would prefer a source enumerator that always selects the true source number, provided that the number of snapshots is large enough. It has been revealed in [12] that the BIC method offers strong consistency, whereas the AIC approach does not. As a result, the former has drawn much attention in the literature. The classical BIC criterion is composed of a likelihood function (LF) and a penalty function (PF), which correspond to data fitting and model complexity, respectively. Minimization of the BIC criterion is, in fact, a procedure trading off data fitting and model complexity, resulting in a correct estimate of the model order or source number. As previously pointed out, the existing BIC criterion does lead to the minimization of the relative KL information between the generating model and the fitted approximating model but only for the case in which m is fixed while $n \to \infty$. In the general asymptotic regime, however, there is no guarantee that what the classical BIC criterion is minimizing is exactly the relative KL divergence and that minimization of the classical BIC criterion yields a correct estimate of the source number. To circumvent this issue, we derive a variant of the BIC criterion for the general asymptotic case, in which $m, n \to \infty$ and $m/n \to c$. In particular, we reformulate the BIC criterion by calculating the LF and PF in this general asymptotic regime. Through appropriate approximations, we are able to accurately determine the LF and the PF for the BIC criterion, ending up with a new variant of the BIC criterion for source enumeration. This enables us to precisely determine the signal and noise subspaces for the subsequent DOA estimation and beamforming in large arrays. Moreover, a closed-form expression for the probability of missed detection is derived. It is also proved that the probability of false alarm converges to zero as $m, n \to \infty$ and $m/n \to c$.

The remainder of this paper is organized as follows. The data model is presented in Section II. The method for source enumeration is proposed in Section III. Statistical performance analysis is conducted in Section IV. Simulation results are presented in Section V. Finally, conclusions are drawn in Section VI.

II. Problem Formulation

Consider an array of m antennas receiving d narrowband source signals $\{s_1(t), \ldots, s_d(t)\}$ from distinct directions $\{\varphi_1, \ldots, \varphi_d\}$, respectively. Assume that the sources and array are in the same plane. In the sequel, the tth snapshot vector of the array output is written as

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{w}_t, \quad (t = 1, \ldots, n) \tag{1}$$

where $\boldsymbol{x}_t = [x_1(t), \ldots, x_m(t)]^T \in \mathbb{C}^{m\times 1}$, $\boldsymbol{A} = [\boldsymbol{a}(\varphi_1), \ldots, \boldsymbol{a}(\varphi_d)] \in \mathbb{C}^{m\times d}$, $\boldsymbol{s}_t = [s_1(t), \ldots, s_d(t)]^T \in \mathbb{C}^{d\times 1}$, and $\boldsymbol{w}_t = [w_1(t), \ldots, w_m(t)]^T \in \mathbb{C}^{m\times 1}$ are the observed snapshot vector, the steering matrix, the signal vector, and the noise vector, respectively. Here, $\boldsymbol{a}(\varphi_i), i = 1, \ldots, d$, is the steering vector, with φ_i being the DOA due to the ith source, $(\cdot)^T$ is the transpose operator, d is the *unknown* number of sources, m is the number of antennas, and n is the number of snapshots. For simplicity but without loss of generality, it is assumed that $m < n$ throughout this paper, unless stated otherwise. Moreover, the number of sources is assumed to be fixed and smaller than a constant number $\bar{m}$, which is much less than $\min(m, n)$, i.e., $\bar{m} \ll \min(m, n)$, as $m, n \to$ with $m/n \to c$. The incoherent signals are independent and identically distributed (i.i.d.) complex Gaussian distributed, i.e., $\boldsymbol{s}_t \sim \mathcal{CN}(\boldsymbol{0}_d, \boldsymbol{R}_s)$, in which $\boldsymbol{0}_d$ is the $d \times 1$ zero vector, and $\boldsymbol{R}_s \triangleq \mathbb{E}[\boldsymbol{s}_t\boldsymbol{s}_t^H] \in \mathbb{C}^{d\times d}$ has full rank, with $(\cdot)^H$ being the conjugate transpose and $\mathbb{E}[\cdot]$ being the mathematical expectation. Here, $\mathcal{CN}(\boldsymbol{\nu}, \boldsymbol{R})$ stands for the complex Gaussian distribution with mean $\boldsymbol{\nu}$ and covariance $\boldsymbol{R}$. Furthermore, the noise $\boldsymbol{w}_t$ is assumed to be an i.i.d. complex Gaussian vector with mean zero and covariance $\tau\boldsymbol{I}_m$, i.e., $\boldsymbol{w}_t \sim \mathcal{CN}(\boldsymbol{0}_m, \tau\boldsymbol{I}_m)$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, which is independent of the signals.

With the given assumptions, the observed samples can be taken as the i.i.d. Gaussian vector, i.e., $\boldsymbol{x}_t \sim \mathcal{CN}(\boldsymbol{0}_m, \boldsymbol{R})$, with $\boldsymbol{R}$ being the population covariance matrix, which is calculated as

$$\boldsymbol{R} = \mathbb{E}\left[\boldsymbol{x}_t\boldsymbol{x}_t^H\right] = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + \tau\boldsymbol{I}_m. \tag{2}$$

Recall that the signals are incoherent and $d < m$, which means that $\boldsymbol{R}_s$ is nonsingular and that $\boldsymbol{A}$ is of full column rank. Without loss of generality, we assume that the population eigenvalues of $\boldsymbol{R}$, which are denoted as $\lambda_1, \ldots, \lambda_m$, are nonincreasingly ordered, i.e.,

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

In addition, their corresponding population eigenvectors are denoted as $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$. Given (3), it is straightforward to utilize the multiplicity of τ to determine the number of signals. In practice, however, only the sample covariance matrix is accessible, which is calculated by

$$\hat{\boldsymbol{R}} = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t^H. \tag{4}$$

Let $\ell_1, \ldots, \ell_m$ and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$, be the descending eigenvalues and corresponding eigenvectors of $\hat{\boldsymbol{R}}$, respectively. Consequently, our task in this work is to infer the source number d from the noisy observations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ for $m, n \to \infty$ and $m/n \to c$.

III. Bayesian Information Criterion for Source Enumeration

A. BIC

For the i.i.d. complex Gaussian observations $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, the joint probability density function (pdf) is

$$f(\boldsymbol{X}|\boldsymbol{\theta}) = \prod_{t=1}^{n} \frac{1}{\pi^m |\boldsymbol{R}|} \exp\left(-\boldsymbol{x}_t^H \boldsymbol{R}^{-1} \boldsymbol{x}_t\right) \tag{5}$$

where $|\cdot|$ is the determinant, and $\boldsymbol{\theta}$ is the unknown parameter vector of the true model, which is specifically given by $\boldsymbol{\theta} = \left[\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_d^T, \lambda_1, \ldots, \lambda_d, \tau\right]^T$. Suppose that we have a parametric family of pdf $\{f(\boldsymbol{X}|\boldsymbol{\theta}^{(k)})\}_{k=0}^{\bar{m}-1}$ with

$$f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right) = \prod_{t=1}^{n} \frac{1}{\pi^m \left|\boldsymbol{R}^{(k)}\right|} \exp\left(-\boldsymbol{x}_t^H \left[\boldsymbol{R}^{(k)}\right]^{-1} \boldsymbol{x}_t\right) \tag{6}$$

where $\boldsymbol{\theta}^{(k)} = \left[\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_k^T, \lambda_1, \ldots, \lambda_k, \tau\right]^T$ corresponds to the kth candidate model. Let $\mathcal{H}_k$ be the hypothesis that the source number is $k \in [0, \bar{m}-1]$. It is easy to see that the hypotheses $\{\mathcal{H}_k\}_{k=0}^{\bar{m}-1}$ are nested.

According to Bayes' rule, we readily have

$$f(\mathcal{H}_k|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\mathcal{H}_k) f(\mathcal{H}_k)}{f(\boldsymbol{X})}. \tag{7}$$

Most typically, $\{\mathcal{H}_k\}_{k=0}^{\bar{m}-1}$ are assumed to be uniformly distributed, yielding $f(\mathcal{H}_k) = 1/\bar{m}$. Moreover, notice that $f(\boldsymbol{X})$ is independent of k, which, when ignored, does not affect the maximization of (7) with respect to k. As a result, we obtain from (7) that

$$\max_{k \in [0, \bar{m}-1]} f(\mathcal{H}_k|\boldsymbol{X}) = \max_{k \in [0, \bar{m}-1]} f(\boldsymbol{X}|\mathcal{H}_k). \tag{8}$$

It is indicated in (8) that maximization of the detection probability under the hypothesis $\mathcal{H}_k$ is equivalent to finding the maximum *a posteriori* probability. It follows from [27] and [28] that the *a posteriori* probability is computed as

$$\begin{aligned} f(\mathcal{H}_k|\boldsymbol{X}) &= \int f\left(\boldsymbol{X}, \boldsymbol{\theta}^{(k)}\right) d\boldsymbol{\theta}^{(k)} \\ &= \int f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right) f\left(\boldsymbol{\theta}^{(k)}\right) d\boldsymbol{\theta}^{(k)} \quad (9a) \\ &\approx (2\pi)^{\frac{\nu_k}{2}} |\hat{\boldsymbol{J}}|^{-\frac{1}{2}} f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) \quad (9b) \end{aligned}$$

where $f(\boldsymbol{X}, \boldsymbol{\theta}^{(k)})$ denotes the joint pdf of $\boldsymbol{X}$ and $\boldsymbol{\theta}^{(k)}$, $f(\boldsymbol{\theta}^{(k)})$ denotes the *a priori* pdf of $\boldsymbol{\theta}^{(k)}$, $\hat{\boldsymbol{\theta}}^{(k)}$ is the ML estimate of $\boldsymbol{\theta}^{(k)}$, ν_k is the length of $\boldsymbol{\theta}^{(k)}$, and

$$\hat{\boldsymbol{J}} = -\left.\frac{\partial^2 \log f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right)}{\partial \boldsymbol{\theta}^{(k)} \partial \left(\boldsymbol{\theta}^{(k)}\right)^H}\right|_{\boldsymbol{\theta}^{(k)} = \hat{\boldsymbol{\theta}}^{(k)}} \in \mathbb{C}^{\nu_k \times \nu_k} \tag{10}$$

is the Hessian matrix. Taking mathematical expectation of $\hat{\boldsymbol{J}}$ leads to the Fisher information matrix

$$\boldsymbol{J} = -\mathbb{E}\left[\frac{\partial^2 \log f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right)}{\partial \boldsymbol{\theta}^{(k)} \partial \left(\boldsymbol{\theta}^{(k)}\right)^H}\right]. \tag{11}$$

Note that, although [27] and [28] can arrive at the approximation in (9b), the former employs the assumption that the *a priori* pdf of $\boldsymbol{\theta}^{(k)}$ is flat around $\hat{\boldsymbol{\theta}}^{(k)}$, which means that $f(\boldsymbol{\theta}^{(k)}) \approx f(\hat{\boldsymbol{\theta}}^{(k)})$, whereas the latter utilizes Laplace's method [29] for integration. Taking the logarithm of (9b) yields

$$\begin{aligned} &\log f(\mathcal{H}_k|\boldsymbol{X}) \\ &\quad \approx \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \log f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) + \frac{\nu_k}{2} \log 2\pi - \frac{1}{2} \log |\hat{\boldsymbol{J}}| \\ &\quad = \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \log f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) + \frac{\nu_k}{2} \log 2\pi - \frac{1}{2} \log \left|n \cdot \frac{1}{n} \hat{\boldsymbol{J}}\right| \\ &\quad \approx \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) - \frac{1}{2} \nu_k \log n. \end{aligned} \tag{12}$$

The approximation in (12) is due to the fact that $\log f(\boldsymbol{\theta}^{(k)})$ and $(\nu_k/2) \log 2\pi$ are independent of n, and $\hat{\boldsymbol{J}}/n = \mathcal{O}(1)$ for the case where m is fixed while $n \to \infty$. Here, $\mathcal{O}(1)$ denotes a term that tends to a constant as $n \to \infty$. Consequently, invoking the results in [12] for log-LF (LLF) calculation, ignoring the terms independent of k and setting $\nu_k = k(2m-k)$, the classical BIC method is given as

$$\begin{aligned} \mathrm{BIC}(k) &= -2 \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \nu_k \log n \\ &= 2n(m-k) \log \frac{\frac{1}{m-k} \sum_{i=k+1}^{m} \ell_i}{\left(\prod_{i=k+1}^{m} \ell_i\right)^{\frac{1}{m-k}}} Q \\ &\quad + k(2m-k) \log n. \end{aligned} \tag{13}$$

Minimizing (13) with respect to k yields the estimate of the source number. It should be noted that the criterion in (13) can also be obtained from a different procedure based on the MDL principle [12], [14], [16].

For $m, n \to \infty$ and $m/n \to c$, however, the observed information matrix $\hat{\boldsymbol{J}}$ depends not only on n but also on m. In such a situation, the approximation in (12) is no longer valid, which considerably degrades the performance of the classical BIC method in (13), particularly when the number of snapshots is comparable with the number of antennas. To circumvent this problem, we recalculate the LLF and the PF for $m, n \to \infty$ and $m/n \to c$, ending up with a new BIC variant that is able to provide reliable detection of the source number in the large array.

B. Proposed BIC Variant

To correctly compute the *a posteriori* probability for source enumeration in the general asymptotic regime, where $m, n \to \infty$ with $m/n \to c$, we first need to determine the ML estimate of the parameter vector $\boldsymbol{\theta}^{(k)}$. It is shown in Appendix A that the ML estimate of $\boldsymbol{\theta}^{(k)}$ in the general asymptotic situation turns out to be the same as that in the classical asymptotic case. That is

$$\hat{\boldsymbol{\theta}}^{(k)} = \left[\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k\right]^T \tag{14}$$

is the ML estimate of $\boldsymbol{\theta}^{(k)}$ for $m, n \to \infty$ and $m/n \to c$. Here, $\hat{\tau}_k = (1/(m-k)) \sum_{i=k+1}^{m} \ell_i$. On the other hand, it is indicated in Appendix B that, as $m, n \to \infty$ and $m/n \to c$, the logarithm of the *a posteriori* probability can be computed as

$$\log f(\mathcal{H}_k|\boldsymbol{X}) \approx \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \log f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) + \nu_k \log \pi - \frac{1}{2}\log|\hat{\boldsymbol{J}}|. \quad (15)$$

It is pointed out in [22] that, to determine the asymptotic behavior of the sample eigenvectors, it cannot make any sense to characterize the behavior of the subspace determined by the eigenvectors as their dimension infinitely increases as $m \to \infty$. Instead, it is interesting to determine the behavior of the quadratic function of the eigenprojection matrix. Similarly, it makes little sense to discuss the parameter vector $\hat{\boldsymbol{\theta}}^{(k)}$ alone since its dimension infinitely increases as $m \to \infty$. As a result, we consider the function of $\hat{\boldsymbol{\theta}}^{(k)}$ in (15), which, when maximized, has the effect of maximizing the detection probability for source number detection.

Recall that $f(\hat{\boldsymbol{\theta}}^{(k)})$ stands for the *a priori* pdf of the parameter vector $\hat{\boldsymbol{\theta}}^{(k)}$, which is bounded as $n \to \infty$. In the sequel, $\log f(\hat{\boldsymbol{\theta}}^{(k)})$ is much less than $(1/2)\log|\hat{\boldsymbol{J}}|$ because the latter increases without bound for $m \to \infty$ or $n \to \infty$. On the other hand, $\nu_k \log \pi$ is also much less than $(1/2)\log|\hat{\boldsymbol{J}}|$ as $m, n \to \infty$ and $m/n \to c$. Hence, it follows from (15) that

$$-2\log f(\mathcal{H}_k|\boldsymbol{X}) \approx -2\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}) + \log|\hat{\boldsymbol{J}}| \quad (16)$$

which, when minimized with respect to k, is able to yield a reliable estimate of the source number, provided that $\log f \times (\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)})$ and $\log|\hat{\boldsymbol{J}}|$ can be correctly calculated for $m, n \to \infty$ and $m/n \to c$. Since $\hat{\boldsymbol{\theta}}^{(k)} = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k]^T$ is the ML estimate of $\boldsymbol{\theta}^{(k)}$ for $m, n \to \infty$ and $m/n \to c$, using the similar derivation in [12], the LLF is computed as

$$-2\log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) = 2n(m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i}{\left(\prod_{i=k+1}^{m}\ell_i\right)^{\frac{1}{m-k}}}. \quad (17)$$

On the other hand, it follows from (B.4) that the determinant of $\hat{\boldsymbol{J}}$ is

$$|\hat{\boldsymbol{J}}| = \frac{(m-k)n}{\hat{\tau}_k^2}|\boldsymbol{Q}| \quad (18)$$

where

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{bmatrix} \quad (19)$$

with $\boldsymbol{Q}_{11}$, $\boldsymbol{Q}_{12}$, $\boldsymbol{Q}_{21}$, and $\boldsymbol{Q}_{22}$ being defined in (B.8). Utilizing the formula for the determinant of partitioned matrices, we obtain

$$|\boldsymbol{Q}| = |\boldsymbol{Q}_{11}| \times \left|\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\right|. \quad (20)$$

Substituting (B.8) into (20) yields

$$|\boldsymbol{Q}| = (2n)^{mk} n^k \left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right)^{m-k+2} \hat{\tau}_k^{(m-k)k} \quad (21)$$

TABLE I
SUMMARY OF PROPOSED BIC ALGORITHM

Step 1: Perform eigenvalue decomposition on $\hat{\boldsymbol{R}}$ and obtain $\ell_1, \cdots, \ell_m$.
Step 2: Calculate $a(k), g(k)$ and $\mathcal{P}(k, m, n)$ in (25) by using $\ell_1, \cdots, \ell_m$.
Step 3: Estimate the source number according to (26).

which, when substituted into (18), leads to

$$|\hat{\boldsymbol{J}}| = (m-k)n \cdot (2n)^{mk} \cdot n^k \cdot \left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right)^{m-k+2} \cdot (\hat{\tau}_k)^{k(m-k)-2}. \quad (22)$$

Taking the logarithm of (22), we have

$$\begin{aligned}\log|\hat{\boldsymbol{J}}| &= \log[(m-k)n] + mk\log(2n) + k\log n \\ &\quad + (m-k+2)\log\left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right) + (k(m-k)-2)\log\hat{\tau}_k \\ &= m\Bigg[k\log(2n) + \frac{\log[(m-k)n]}{m} + \frac{k\log n}{m} \\ &\quad + \left(1 - \frac{k-2}{m}\right)\log\left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right) \\ &\quad + \left(k - \frac{k^2+2}{m}\right)\log\hat{\tau}_k\Bigg]. \end{aligned} \quad (23)$$

Recall that $m, n \to \infty$ while the presumed source number k can be taken as a fixed number. Consequently, we obtain $[\log((m-k)n)]/m \to 0$, $(k\log n)/m \to 0$, $(k-2)/m \to 0$, and $(k^2+2)/m \to 0$ for $m, n \to \infty$ and $m/n \to c$. It follows that, as $m, n \to \infty$ and $m/n \to c$, (23) is approximated as

$$\begin{aligned}\log|\hat{\boldsymbol{J}}| &\approx m\left[k\log(2n) + \log\left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right) + k\log\hat{\tau}_k\right] \\ &= mk\left(\log(2n) - \frac{1}{k}\sum_{i=1}^{k}\log\frac{\ell_i}{\hat{\tau}_k}\right) \\ &\triangleq \mathcal{P}(k, m, n). \end{aligned} \quad (24)$$

Therefore, substituting (17) along with (24) into (16), the proposed BIC variant is

$$\mathrm{BIC}(k) = 2n(m-k)\log\frac{a(k)}{g(k)} + \mathcal{P}(k, m, n) \quad (25)$$

where $a(k) = (1/(m-k))\sum_{i=k+1}^{m}\ell_i$ and $g(k) = \left(\prod_{i=k+1}^{m}\ell_i\right)^{1/(m-k)}$ are the arithmetic mean and the geometric mean, respectively. The source number is estimated as

$$\hat{d} = \arg\min_{k=0,\ldots,\bar{m}-1}\mathrm{BIC}(k). \quad (26)$$

Recall that $\bar{m} < \min(m, n)$, with $\bar{m}$ being the maximum presumed source number, which is fixed as $m, n \to \infty$ and $m/n \to c$. Since a pair of mutually transposed matrices shares a common set of nonzero eigenvalues up to a nuisance constant multiplication factor [30], the numbers of antennas m and samples n play symmetric roles. This implies that for $m > n$, we can swap m and n when applying the proposed BIC variant on the n nonzero eigenvalues. The proposed BIC algorithm is tabulated in Table I.

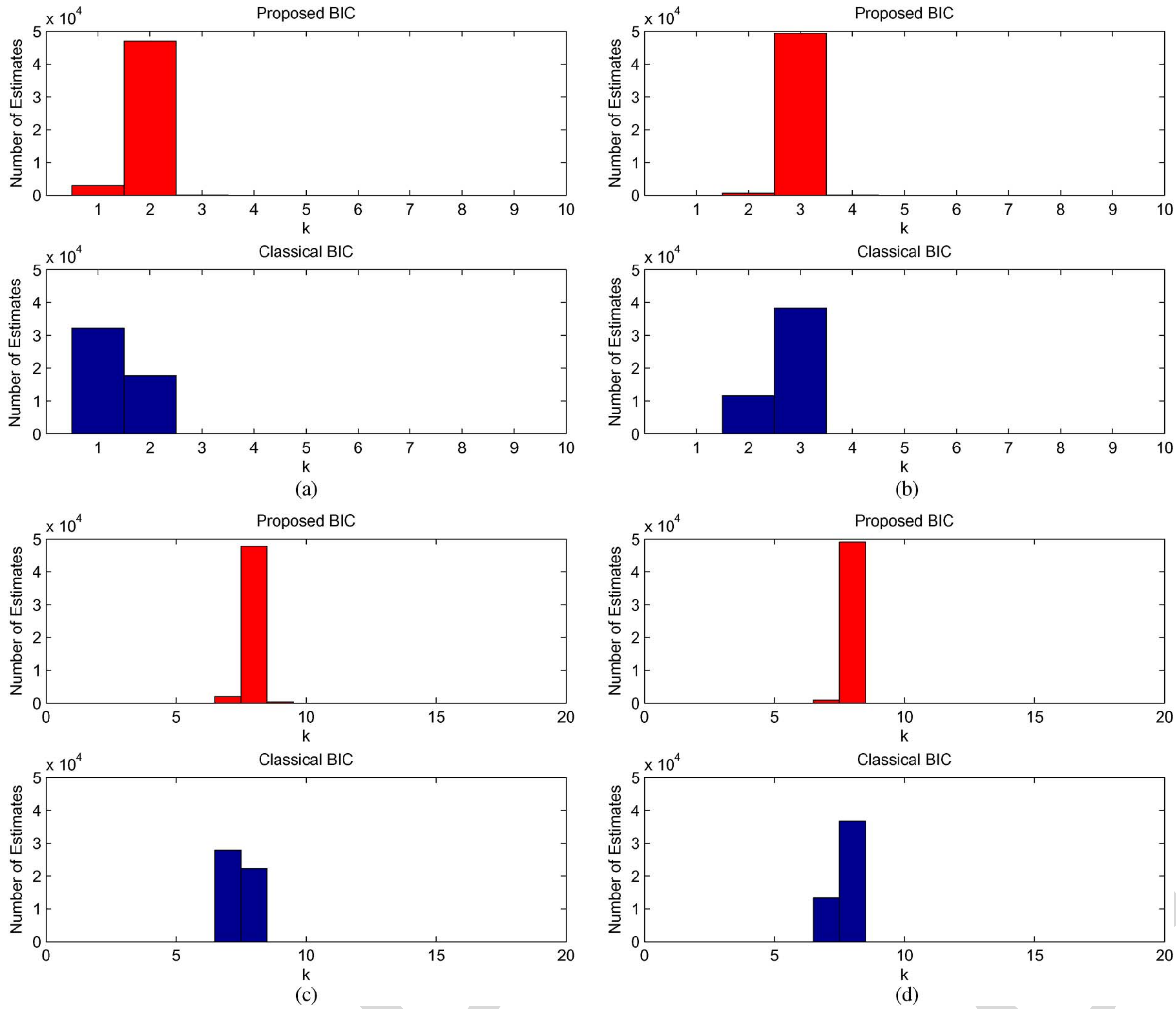


Fig. 1. Histogram plots for source enumeration. $[\varphi_1, \varphi_2] = [2°, 6.5°]$ for Fig. 1(a), $[\varphi_1, \varphi_2, \varphi_3] = [2°, 6.5°, -3°]$ for Fig. 1(b), and $[\varphi_1, \ldots, \varphi_8] = [2.5°, 22°, -4.9°, 12.3°, 7.3°, 16.7°, -9.6°, 26.7°]$ for Fig. 1(c) and (d). The signals are of equal power. (a) $m = 10$, $n = 80$, SNR $= -3$ dB, and $d = 2$. (b) $m = 10$, $n = 80$, SNR $= 5$ dB, and $d = 3$. (c) $m = 20$, $n = 60$, SNR $= -3$ dB, and $d = 8$. (d) $m = 20$, $n = 200$, SNR $= -5$ dB, and $d = 8$.

Remark: Recall that $\ell_1, \ldots, \ell_k$, are the ML estimates of the signal population eigenvalues. In the sequel, $\ell_1/\hat{\tau}_k, \ldots, \ell_k/\hat{\tau}_k$, are relative to the SNR. This, in turn, indicates that $\mathcal{P}(k, m, n)$ depends not only on the number of snapshots n but also on the number of antennas m as well as SNR. That is to say, the PF $\mathcal{P}(k, m, n)$ employs more information than that in the standard BIC [12], [28], leading to accurate computation of the PF, particularly for $m, n \to \infty$ and $m/n \to c$.

IV. Performance Analysis

Here, we derive the analytical formula for the probability of missed detection and prove that the probability of false alarm converges to zero in the general asymptotic regime.

A. Approximate Probabilities of Missed Detection and False Alarm

The statistical analysis for the performance of the classical MDL method has been widely conducted in the literature [31]–[35]. In fact, in this multiple-hypothesis test, there are two error types, namely, the probabilities of underestimating and overestimating the source number. They are also known as the probability of missed detection P_{md} and the probability of false alarm P_{fa}, respectively. P_{md} and P_{fa} for d sources are, respectively, defined as

$$P_{\mathrm{md}} = \mathrm{Prob}(\hat{d} < d | \mathcal{H}_d) \tag{27a}$$
$$P_{\mathrm{fa}} = \mathrm{Prob}(\hat{d} > d | \mathcal{H}_d). \tag{27b}$$

It has been well justified in [31]–[33] by Monte Carlo experiments that the probability of missed detection can be approximated by the probability of underestimating the source number by one, whereas the probability of false alarm can be approximated by the probability of overestimating the source number by one. That is

$$P_{\mathrm{md}} \approx \mathrm{Prob}(\hat{d} = d - 1 | \mathcal{H}_d) \tag{28a}$$
$$P_{\mathrm{fa}} \approx \mathrm{Prob}(\hat{d} = d + 1 | \mathcal{H}_d). \tag{28b}$$

Computer simulation has been carried out to verify the approximations in (28) for the proposed BIC variant in terms of the histogram of the estimated source number. Fig. 1 plots the histogram bars for source enumeration in four representative parameter settings. That is, Fig. 1(a) provides the histogram for $m=10$, $n=80$, $d=2$, and SNR$=-3$ dB; Fig. 1(b) gives the histogram for $m=10$, $n=80$, $d=3$, and SNR$=5$ dB; Fig. 1(c) shows the histogram for $m=20$, $n=60$, $d=8$, and SNR$=-3$ dB; whereas Fig. 1(d) shows the histogram for $m=20$, $n=200$, $d=8$, and SNR $=-5$ dB. Throughout this paper, the SNR is defined as $10\log_{10}(\sigma^2_{s_i}/\tau)$ with $\sigma^2_{s_i} \triangleq \mathbb{E}[|s_i(t)|^2]$ and $\tau = 1$. It is indicated in Fig. 1 that the proposed BIC variant tends to underestimate the source number, and the probability of underestimating the source number by one dominates. Moreover, compared with the classical BIC scheme, the proposed scheme considerably improves in terms of the probability of missed detection. Furthermore, the probability of false alarm is negligible. Indeed, it is proved in Appendix C that P_{fa} of the proposed BIC variant converges to zeros as $m, n \to \infty$ and $m/n \to c$. This will also be verified by the simulation results in Section V-B. Recall that $\text{Prob}(\hat{d} = d|\mathcal{H}_d) + P_{\text{md}} + P_{\text{fa}} = 1$. Therefore, it is sufficient to determine P_{md} for the proposed BIC method to evaluate its detection performance in the general asymptotic regime.

B. Analytic Probability of Missed Detection

Noticing that

$$a(d-1) = \frac{m-d}{m-d+1}a(d) + \frac{\ell_d}{m-d+1} \tag{29}$$

$$[g(d-1)]^{m-d+1} = [g(d)]^{m-d} \cdot \ell_d \tag{30}$$

we obtain

$$\begin{aligned}(m-(d-1))\log\frac{a(d-1)}{g(d-1)} &= \log\left(\frac{[a(d)]^{m-d}}{[g(d)]^{m-d}} \times \frac{\left(\frac{m-d}{m-d+1} + \frac{\frac{\ell_d}{a(d)}}{m-d+1}\right)^{m-d+1}}{\frac{\ell_d}{a(d)}}\right)\\ &= (m-d)\log\frac{a(d)}{g(d)} + \log Q_m\left[\frac{\ell_d}{a(d)}\right]\end{aligned} \tag{31}$$

where

$$Q_m\left[\frac{\ell_d}{a(d)}\right] \triangleq \frac{\left[1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right]^p}{\frac{\ell_d}{a(d)}} \tag{32}$$

with $p \triangleq m-d+1$. Recalling that $\hat{\tau}_{d-1} = a(d-1)$ and $\hat{\tau}_d = a(d)$, it is easy to obtain

$$\begin{aligned}&\mathcal{P}(d,m,n) - \mathcal{P}(d-1,m,n)\\ &= m\log(2n) - m\left(\sum_{i=1}^{d}\log\frac{\ell_i}{\hat{\tau}_d} - \sum_{i=1}^{d-1}\log\frac{\ell_i}{\hat{\tau}_{d-1}}\right)\\ &= m\log(2n) - m\log\frac{\ell_d}{\hat{\tau}_d} - m(d-1)\log\frac{\hat{\tau}_{d-1}}{\hat{\tau}_d}\\ &= m\log(2n) - m\log\frac{\ell_d}{a(d)} - m(d-1)\log\\ &\quad\times\left[1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right].\end{aligned} \tag{33}$$

Therefore, as $m, n \to \infty$ and $m/n \to c$, the probability of missed detection is calculated as

$$\begin{aligned}P_{\text{md}} &\approx \text{Prob}\left(\text{BIC}(d-1,m,n) - \text{BIC}(d,m,n) < 1|\mathcal{H}_d\right)\\ &= \text{Prob}\left(\log Q_m\left[\frac{\ell_d}{a(d)}\right] < \frac{m}{2n}\log(2n) - \frac{m}{2n}\log\frac{\ell_d}{a(d)}\right.\\ &\qquad\left.- \frac{m(d-1)}{2n}\log\left(1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right)\right|\mathcal{H}_d\Bigg)\\ &\approx \text{Prob}\Bigg(\left(p+\frac{c(d-1)}{2}\right)\log\left(1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right)\\ &\qquad\left.-\left(1-\frac{c}{2}\right)\log\frac{\ell_d}{a(d)} < \frac{c}{2}\log(2n)\right|\mathcal{H}_d\Bigg)\\ &= \text{Prob}\left(\left.\frac{\ell_d}{a(d)} < f^{-1}(\alpha)\right|\mathcal{H}_d\right)\end{aligned} \tag{34}$$

where $\alpha = c/2\log(2n)$, and $f^{-1}(z)$ is the inverse function of

$$f(z) = \left(p+\frac{c(d-1)}{2}\right)\log\left(1+\frac{z-1}{p}\right) - \left(1-\frac{c}{2}\right)\log z \tag{35}$$

with $z = \ell_d/a(d)$. Note that the last equality in (34) is due to the fact that $f(z)$ is a monotonic increasing function for $c > 0$ and $z > 1$. The function $f^{-1}(z)$ can be determined by the numerical simulation. Now, we need to determine the distribution of $\ell_d/a(d)$.

It is well known that, as $m, n \to \infty$ with $m/n \to c$, the signal eigenvalues $\lambda_i (i = 1, \ldots, d)$ are probably lower than the critical value $\tau(1+\sqrt{c})$, namely, the so-called asymptotic limit of detection due to the phase transition phenomenon [36]. In such a situation, the signal sample eigenvalue behaves similar to the noise sample eigenvalue. Note that analyzing the detection threshold for the source enumerator is also an interesting topic. It is shown in [8] that the threshold of the RMT detector can be as low as the asymptotic limit of detection when $m, n \to \infty$ with $m/n \to c$. Moreover, it is revealed in [37] that the asymptotic probability of detection for the likelihood ratio test can approach one even when the signal power is substantially lower than the asymptotic limit of detection. Additionally, the consistency of the classical BIC method with respect to the SNR has been investigated in [10] and [15]. However, these topics are beyond the scope of this paper. Consequently, we restrict our attention on the analytic probability of missed detection.

If $\lambda_d > \tau(1+\sqrt{c})$ and λ_d has multiplicity of one, it then follows from [25], [36], and [38] that, as $m, n \to \infty$ with $m/n \to c$, ℓ_d is Gaussian distributed, i.e.,

$$\sqrt{n}\left(\ell_d - \lambda_d\left(1+\frac{\tau c}{\lambda_d-\tau}\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \lambda_d^2\left(1-\frac{c}{(\lambda_d-\tau)^2}\right)\right) \tag{36}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Although this asymptotic result is correct in the general asymptotic regime, it is not accurate enough for finite m and n because it does not consider the interaction between the signals. In fact, it is verified in [22] that, as $m, n \to \infty$ with $m/n \to c$, ℓ_d almost surely (a.s.) converges to its mean derived by Lawley [39] in the classical asymptotic regime, that is

$$\mu_d = \lambda_d\left(1 - \frac{c}{m}\sum_{1\le i\ne d\le m}\frac{\lambda_i}{\lambda_i-\lambda_d}\right). \tag{37}$$

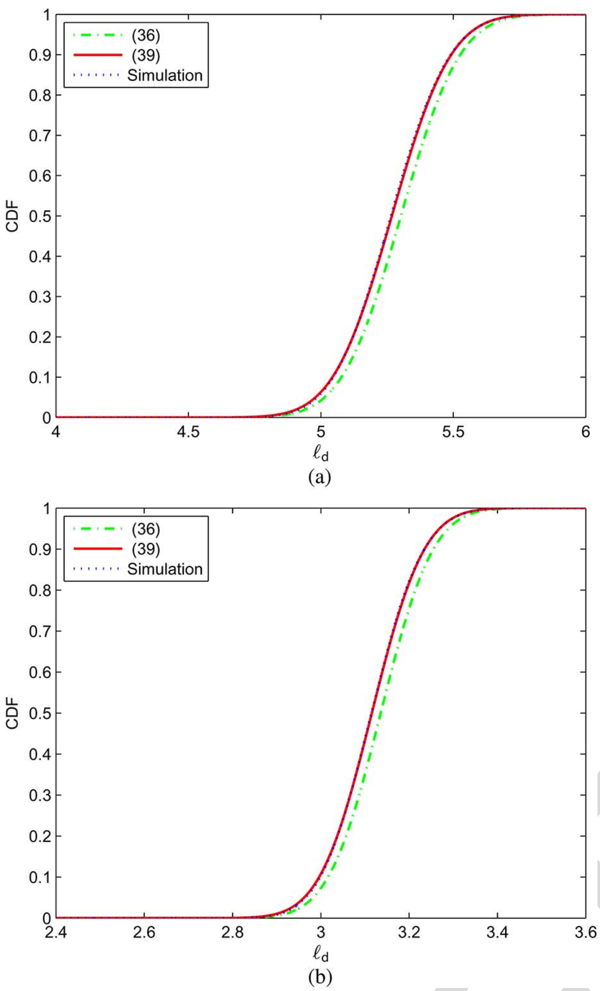


Fig. 2. CDF: Asymptotic distribution versus simulation for three sources with power values of $[-12, -17, -15]$ dB and DOA = $[-2.5°, 3.3°, 12°]$. 10^5 trials. (a) $m = 200$ and $n = 800$. (b) $m = 100$ and $n = 1000$.

As a matter of fact, (37) can be rewritten as

$$\mu_d = \lambda_d \left(1 - \frac{c}{m}\sum_{i=1}^{d-1}\frac{\lambda_i}{\lambda_i - \lambda_d} + \frac{m-d}{m}\frac{\tau c}{\lambda_d - \tau}\right) \tag{38a}$$

$$\xrightarrow{m\to\infty} \lambda_d \left(1 + \frac{\tau c}{\lambda_d - \tau}\right). \tag{38b}$$

Note that the second term within the brackets of (38a) stands for the interaction between the signals. It is implied in (38) that, although the mean of ℓ_d in (37) is the same as that in (36) as $m, n \to \infty$ with $m/n \to c$, the former is more accurate than the latter for finite m and n because it takes into account the interaction between the signals. As a consequence, the fluctuation of ℓ_d is

$$\ell_d \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mu_d, \sigma_d^2\right) \tag{39}$$

with $\sigma_d^2 = (\lambda_d^2/n)(1 - c(\lambda_d - \tau)^{-2})$. To quantitatively show the approximation accuracy between (36) and (39), we calculate their cumulative distribution functions (cdfs) and compare them with the exact distribution of ℓ_d resulted from 10^5 independent simulation trials. The results shown in Fig. 2 indicate that the asymptotic distribution in (39) is more accurate than that in (36). On the other hand, by taking into account the bias resulting from ℓ_i $(i = 1, \ldots, d)$, it follows from [31] and [35] that $a(d) \approx \varpi_d$ with

$$\varpi_d = \tau - \frac{1}{n(m-d)}\sum_{i=1}^{d}\sum_{1\le j\ne i\le m}\frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)}. \tag{40}$$

Thus, the fluctuation of z is

$$z \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mu_z, \sigma_z^2\right) \tag{41}$$

where $\mu_z = \mu_d/\varpi_d$, and $\sigma_z^2 = \sigma_d^2/\varpi_d^2$. The analytic probability of missed detection is

$$P_{\mathrm{md}} = 1 - Q\left(\frac{f^{-1}(\alpha) - \mu_z}{\sigma_z}\right) \tag{42}$$

for $\lambda_d > \tau(1 + \sqrt{c})$, where $Q(x) = \int_x^{\infty}(1/\sqrt{2\pi})e^{-t^2/2}dt$. For $\lambda_d \le \tau(1 + \sqrt{c})$, however, the signal cannot be reliably detected due to the phase transition phenomenon. In the sequel, we have $P_{\mathrm{md}} = 1$.

V. Simulation Results

A. Detection Performance

The detection performance of the proposed BIC variant is evaluated by computer simulation in this section. For the purpose of comparison, the empirical results of the representative ITCs are also presented, that is, the BIC [13], [28], linear-shrinkage-based MDL (LS-MDL) [40], EEF [41], RMT-AIC [25], and BN-AIC [42]. According to [42], the user-defined parameter C in the BN-AIC scheme is set to 2. Similar to the setting in the last section, we consider a uniform linear array with half-wavelength element separation receiving the narrow-band and equal-power stationary Gaussian signals.

The empirical probabilities of correct detection versus SNR for a relatively small sample size of $n = 60$ are plotted in Fig. 3(a), where the number of antennas is 15. We observe that the proposed BIC variant is superior to the other ITCs in terms of detection probability. When the number of snapshots is larger, e.g., $n = 150$, the gaps between the proposed BIC variant and existing ITCs become narrower, as demonstrated in Fig. 3(b). In such a large sample case, the proposed method is comparable with the EEF scheme and still outperforms the LS-MDL and RMT-AIC approaches by around 0.5 dB. Moreover, the proposed detector significantly improves compared with the standard BIC scheme. To study the behavior of the BIC variant for different angle separations, the empirical probabilities of correct detection versus the angle separation are shown in Fig. 4 for the small and large sample sizes, respectively. Here, the DOAs due to the three incident signals are set as $[0, \Delta\varphi, 2\Delta\varphi]$, and the number of antennas is 15. It is seen that the BIC variant is more accurate than the existing ITCs in source enumeration, particularly for the small sample case. It is easy to interpret the improvement of the proposed BIC variant by recalling that the standard BIC suffers from its heavy penalty term. That is, its probability of underestimating the source number dominates. As the proposed BIC offers a smaller penalty term than the standard BIC, it is able to reduce the possibility of underfitting, eventually leading to the significant enhancement in detection performance.

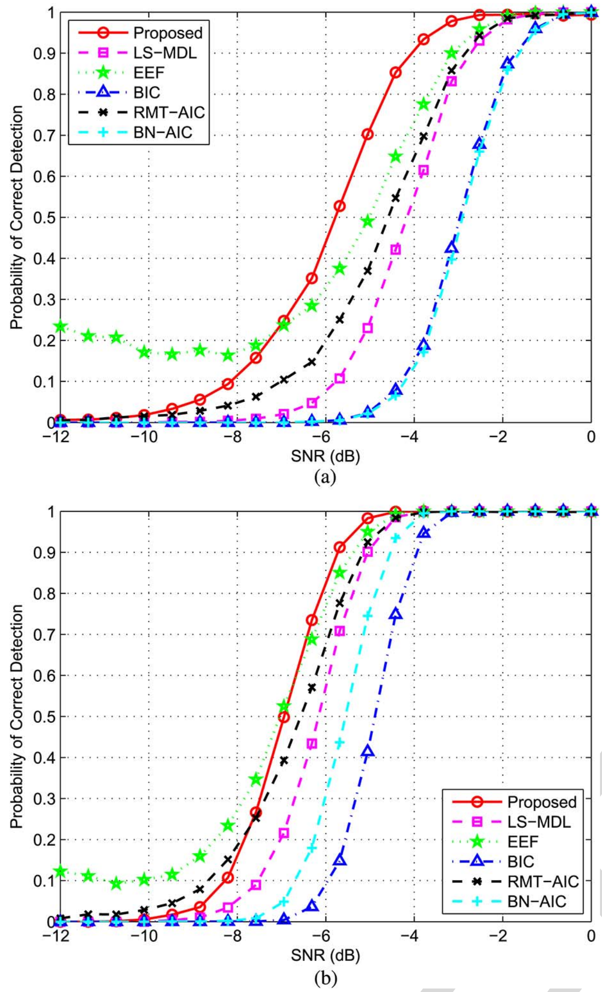


Fig. 3. Probability of correct detection versus SNR. $m = 15$, $d = 3$, $[\varphi_1, \varphi_2, \varphi_3] = [2.3^\circ, 7.5^\circ, 12^\circ]$, and 2×10^3 trials. (a) $n = 60$. (b) $n = 150$.

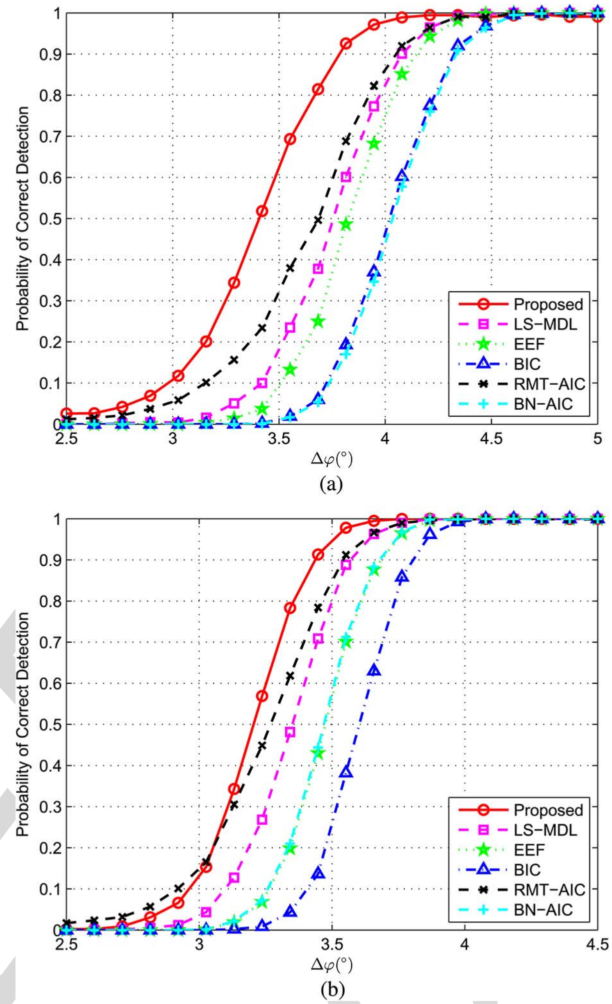


Fig. 4. Probability of correct detection versus angle separation. $m = 15$, SNR $= 0$ dB, $d = 3$, $[\varphi_1, \varphi_2, \varphi_3] = [0, \Delta\varphi, 2\Delta\varphi]$, and 2×10^3 trials. (a) $n = 60$. (b) $n = 150$.

To investigate the general asymptotic case, we enable both m and n to increase at the same speed, e.g., $m/n = 1/3$ and $m/n = 0.5$. Since the number of antennas and the number of snapshots can infinitely increase while the source number remains unchanged, we set $\bar{m} = \min(20, m)$ for all algorithms in the following simulations, where m and n increase at the same rate $c = m/n$. The empirical results shown in Fig. 5(a) indicate that the EEF and LS-MDL schemes are more accurate than the standard BIC detector while all of them are able to yield the consistent estimate of source number when the number of snapshots becomes large enough. Compared with the existing ITCs, the proposed BIC approach is capable of yielding more accurate estimate of source number. When $m/n = 0.5$ and $d = 8$, the proposed scheme converges to one in probability of correct detection much faster than the other ITCs, as indicated in Fig. 5(b).

To confirm that the probability of false alarm of the proposed BIC approach tends to zero as $m, n \to \infty$ and $m/n \to c$, the empirical probability of false alarm versus the number of antennas is shown in Fig. 6, where $m/n = 0.5$, and SNR $= -8$ dB. For comparison, the empirical results of the EEF and RMT-AIC approaches are presented as well. Fig. 6(a) corresponds to the empirical results for three incident signals with DOAs of $[2.3^\circ, 7.5^\circ, 12^\circ]$, whereas Fig. 6(b) shows the empirical results for eight incident signals with DOAs of $[\varphi_1, \ldots, \varphi_8] = [2.5^\circ, 22^\circ, -4.9^\circ, 12.3^\circ, 7.3^\circ, 16.7^\circ, -9.6^\circ, 26.7^\circ]$. It is seen that the ITCs offer different probabilities of false alarm. On the other hand, Fig. 6 implies that the probability of false alarm of the proposed BIC algorithm converges to zero as $m, n \to \infty$ and $m/n \to c$, which is in line with the theoretical analysis in Section IV-A.

To fairly compare the ITCs with the threshold-like testing methods, we need to set their probabilities of false alarm at the same level. Nevertheless, as indicated in Fig. 6, the ITCs implicitly offer different probabilities of false alarm. As a result, the threshold-like testing method should be compared with one of the ITCs at the same probability of false alarm. In the end, the empirical results of the proposed BIC variant and RMT approach [8] are plotted in Fig. 7. Here, the probability of false alarm of the RMT algorithm is equal to that of the proposed BIC approach. Moreover, to enable the RMT method to properly work, we set its probability of false alarm to 10^{-6} when the probability of false alarm of the proposed BIC variant is equal to zero. In addition, note that the minimax [43] approach is also a threshold-like algorithm, but it directly links the noise sample eigenvalue distribution, namely, the Tracy–Widom law [44], to the signal sample eigenvalue distribution, ending up with an

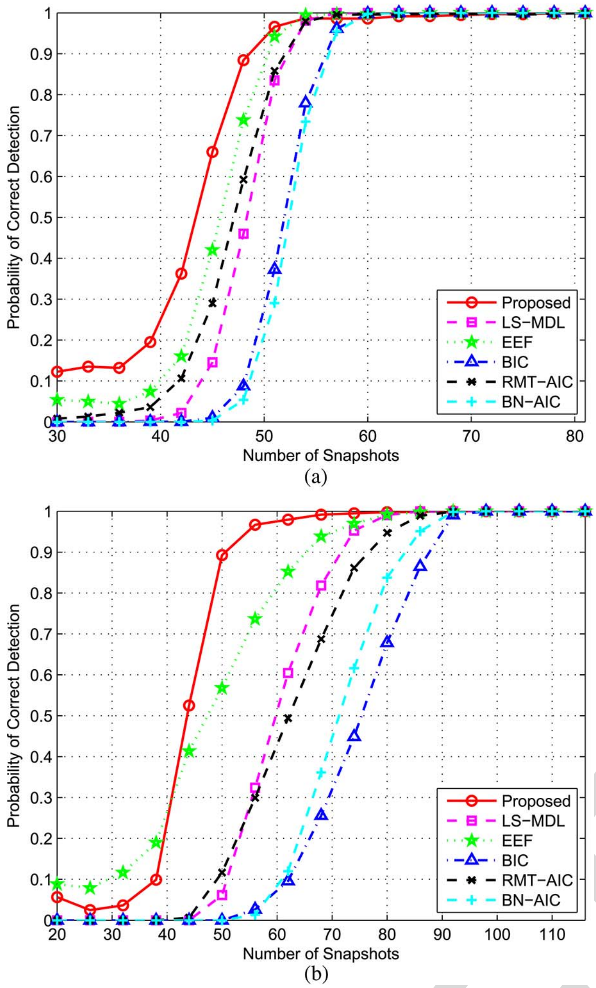


Fig. 5. Probability of correct detection versus number of snapshots. 2×10^3 trials. (a) $d=3$, $m/n=1/3$, and $[\varphi_1,\varphi_2,\varphi_3]=[2.3^\circ, 7.5^\circ, 12^\circ]$. (b) $d=8$, $m/n=0.5$, and $[\varphi_1,\ldots,\varphi_8]=[2.5^\circ, 22^\circ, -4.9^\circ, 12.3^\circ, 7.3^\circ, 16.7^\circ, -9.6^\circ, 26.7^\circ]$.

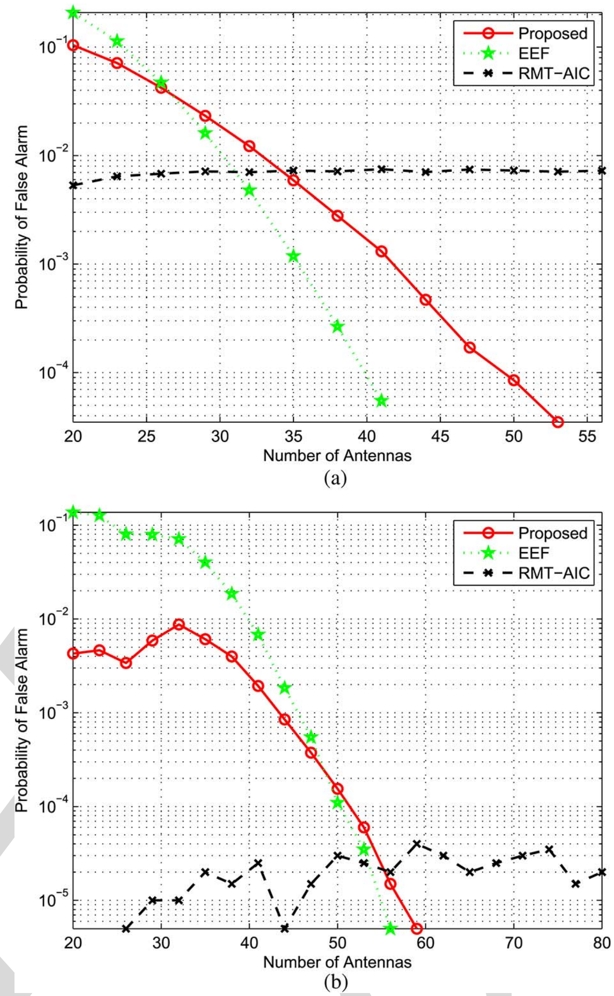


Fig. 6. Probability of false alarm versus antenna number for the proposed, EEF, and RMT-AIC approaches. SNR $= -8$ dB, $m/n = 0.5$, and 2×10^5 trials. (a) $[\varphi_1,\varphi_2,\varphi_3] = [2.3^\circ, 7.5^\circ, 12^\circ]$. (b) $[\varphi_1,\ldots,\varphi_8] = [2.5^\circ, 22^\circ, -4.9^\circ, 12.3^\circ, 7.3^\circ, 16.7^\circ, -9.6^\circ, 26.7^\circ]$.

elegant method for threshold calculation. See [43, eq. (10)] for the details. As a result, the empirical results for the minimax scheme are presented as well. Similar to [43], the "inclusion" penalty and the "exclusion" penalty are set as $c_{\mathrm{I}} = c_{\mathrm{E}}(1) = \cdots = c_{\mathrm{E}}(m)$ with $\lambda_0 = \sqrt{c} + n^{-1/3}$ and $\hat{\tau}_k = (m-k)^{-1}\sum_{i=k+1}^{m}$ for the kth threshold calculation. It is indicated in Fig. 7 that the proposed BIC variant is superior to the RMT scheme in detection accuracy and outperforms the minimax approach in consistency. Although the proposed BIC variant might not be as accurate as the minimax method, as shown in Fig. 7(b), it is able to attain correct detection probability of one in these two cases. On the other hand, it should be noted that the minimax algorithm relies on the Tracy–Widom distribution, which cannot be evaluated online, incurring more overhead in the procedure of detection.

B. *Accuracy of Analytic Probability of Missed Detection*

Here, numerical results are presented to evaluate the accuracy of the analytic probability of missed detection, which is derived in Section IV-B. To evaluate the accuracy of the analytic probability of missed detection for large arrays and large samples in large-array applications, we set the number of antennas as $m = 50$ and vary the number of samples from $n = 300$ to $n = 1000$. Fig. 8 indicates that the analytic probability of missed detection is very close to the simulated probability of missed detection. This, in turn, implies that our derived analytic probability of missed detection is able to accurately predict the detection performance.

VI. Conclusion

This paper has devised a new BIC variant for source enumeration in the general asymptotic regime, which enables us to correctly determine the signal and noise subspaces for the subsequent DOA estimation and beamforming in large-array systems. As the existing information criteria only consider the condition when the number of antennas remains unchanged while the number of snapshots tends to infinity, they cannot provide accurate detection of the source number for the large array. By correctly determining the Hessian matrix in the calculation of the PF, we have derived an efficient BIC variant for the general asymptotic regime. Moreover, a closed-form formula has been derived for calculating the probability of missed detection, and the probability of false alarm has been proved to converge

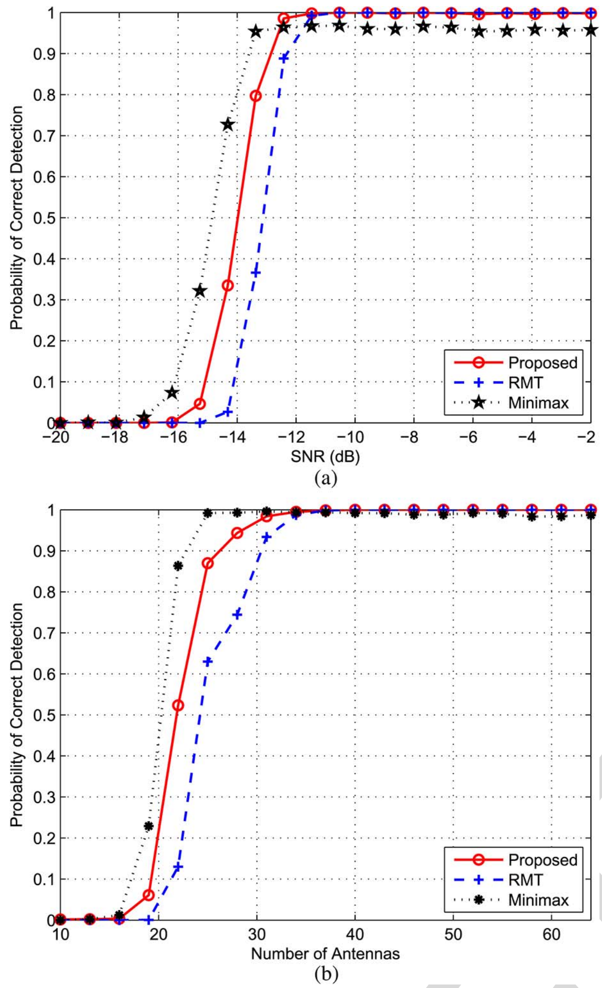


Fig. 7. Probability of correct detection for the proposed BIC, RMT, and minimax algorithms at the same probability of false alarm. $d = 3$, $[\varphi_1, \varphi_2, \varphi_3] = [2.3°, 7.5°, 12°]$, and 2×10^3 trials. (a) $m = 50$ and $n = 80$. (b) SNR $= -12$ dB and $m/n = 0.2$.

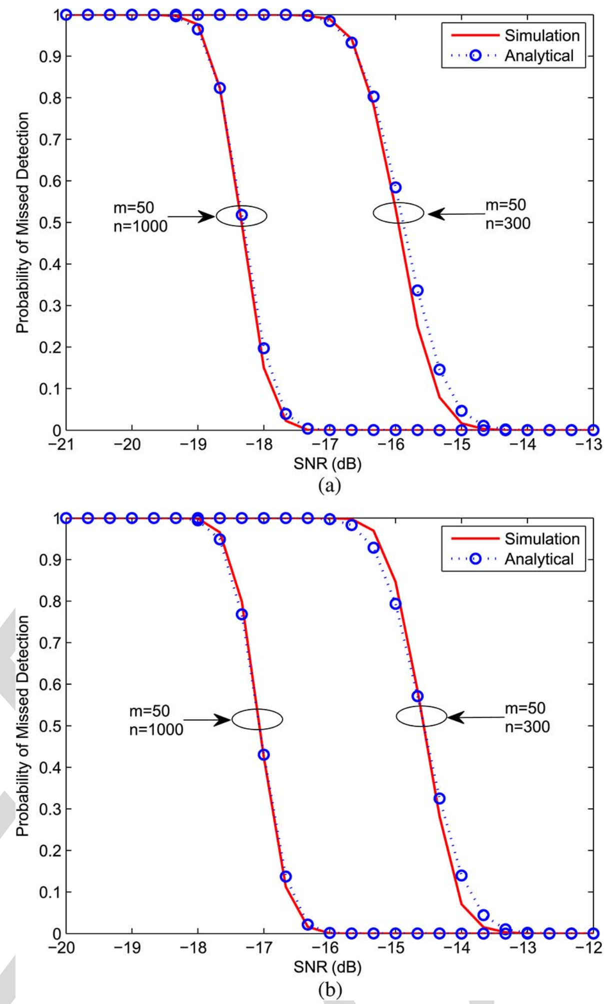


Fig. 8. Probability of missed detection versus SNR. $m = 50$ and 2×10^3 trials. (a) $d = 3$ and $[\varphi_1, \varphi_2, \varphi_3] = [-2.5°, 3.3°, 12°]$. (b) $d = 8$ and $[\varphi_1, \ldots, \varphi_8] = [-2.5°, 4.3°, 2.3°, -12°, 8.1°, 16.8°, -23.7°, 32.1°]$.

to zero as $m, n \to \infty$ and $m/n \to c$. Simulation results have verified the superiority of the proposed BIC approach over its existing counterparts and confirmed the statistical performance analysis.

Appendix A
Maximum Estimation of $\boldsymbol{\theta}^{(k)}$ in the General Asymptotic Case

For the case of k sources, let $\boldsymbol{R}^{(k)} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ and $\hat{\boldsymbol{R}} = \boldsymbol{E}\boldsymbol{L}\boldsymbol{E}^H$ be the eigenvalue decompositions of $\boldsymbol{R}^{(k)}$ and $\hat{\boldsymbol{R}}$, respectively. Here, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k, \tau, \ldots, \tau)$, $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$, $\boldsymbol{L} = \mathrm{diag}(\ell_1, \ldots, \ell_m)$, and $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m]$, with $\boldsymbol{u}_i$ and $\boldsymbol{e}_i$, $i = 1, \ldots, m$ being the population and sample eigenvectors corresponding to the population and sample eigenvalues λ_i and ℓ_i, respectively. As a result, the LLF is calculated as

$$\mathcal{L}\left(\boldsymbol{\theta}^{(k)}\right) \triangleq -n \log \left|\boldsymbol{R}^{(k)}\right| - n\mathrm{tr}\left[\left(\boldsymbol{R}^{(k)}\right)^{-1} \hat{\boldsymbol{R}}\right] - mn \log \pi$$
$$= -n \left(\sum_{i=1}^{k} \log \lambda_i + (m-k) \log \tau\right) - n\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{G}^H\boldsymbol{L}\boldsymbol{G}) - mn \log \pi \quad \text{(A.1)}$$

where $\mathrm{tr}[\cdot]$ denotes the trace operation, and $\boldsymbol{G} = \boldsymbol{E}^H\boldsymbol{U}$. Since $\boldsymbol{G}$ is orthogonal, we have the following inequality [45], [46]:

$$\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{G}^H\boldsymbol{L}\boldsymbol{G}) \geq \sum_{i=1}^{m} \frac{\ell_i}{\lambda_i}. \quad \text{(A.2)}$$

This equality in (A.2) holds for $\boldsymbol{G} = \boldsymbol{I}_m$ [46], i.e., $\boldsymbol{U} = \boldsymbol{E}$. Consequently, it follows from (A.1) and (A.2) that $\boldsymbol{e}_i$, $i = 1, \ldots, m$, is the ML estimate of $\boldsymbol{u}_i$. That is, $\hat{\boldsymbol{u}}_i = \boldsymbol{e}_i$ for $i = 1, \ldots, k$. Substituting these ML estimates into (A.1), we obtain the LLF relying on the reduced parameter vector $\boldsymbol{\vartheta}^{(k)} = [\lambda_1, \ldots, \lambda_k, \tau]$, i.e.,

$$\mathcal{L}\left(\boldsymbol{\vartheta}^{(k)}\right) = -n \left(\sum_{i=1}^{k} \log \lambda_i + (m-k) \log \tau\right) - n \left(\sum_{i=1}^{k} \frac{\ell_i}{\lambda_i} + \frac{\sum_{i=k+1}^{m} \ell_i}{\tau}\right) - mn \log \pi. \quad \text{(A.3)}$$

Maximizing $\mathcal{L}(\boldsymbol{\vartheta}^{(k)})$ with respect to $\boldsymbol{\vartheta}^{(k)}$ yields the ML estimates of $\lambda_1, \ldots, \lambda_k, \tau$, which are given as

$$\hat{\lambda}_i = \ell_i, \ i = 1, \ldots, k \quad \text{(A.4)}$$

$$\hat{\tau}_k = \frac{1}{m-k} \sum_{i=k+1}^{m} \ell_i. \quad \text{(A.5)}$$

Thus, the ML estimate of $\boldsymbol{\theta}^{(k)}$ is $\hat{\boldsymbol{\theta}}^{(k)} = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k]^T$ for the general asymptotic case, which is the same as that in the classical asymptotic situation.

APPENDIX B
DERIVATION OF (15)

It follows from [47, eq. (92)] that the Taylor series expansion of $\log f(\boldsymbol{X}|\boldsymbol{\theta}^{(k)})$ around $\hat{\boldsymbol{\theta}}^{(k)}$ is given in (B.1). Here, $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{J}}$ is the Hessian matrix defined in (10), and the superscript $(\cdot)^{(k)}$ has been dropped for simplicity. We are now at a position to prove that the zero-order term is much larger than the second-order term in (B.1) as $m, n \to \infty$ and $m/n \to c$

$$\begin{aligned}\log f(\boldsymbol{X}|\boldsymbol{\theta}) &= \log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^H \underbrace{\left.\frac{\partial \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}}_{=0} \\ &\quad + \frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^H \left[\left.\frac{\partial^2 \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^H}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + \cdots \\ &= \log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) - \frac{1}{2}\Delta\boldsymbol{\theta}^H \hat{\boldsymbol{J}} \Delta\boldsymbol{\theta} + \cdots. \end{aligned} \tag{B.1}$$

Recall that $\hat{\boldsymbol{\theta}} = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k]^T$ is the ML estimate of $\boldsymbol{\theta}$ in the general asymptotic regime. Exploiting the similar computation in [12], the zero-order term of (B.1) is

$$\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) = -n(m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i}{\left(\prod_{i=k+1}^{m}\ell_i\right)^{\frac{1}{m-k}}}. \tag{B.2}$$

To determine the second-order term of (B.1), we need to calculate the second-order partial derivative of $-\log f(\boldsymbol{X}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, which is provided in (B.3), shown at the bottom of the page. Accordingly, the Hessian matrix is calculated as (B.4), shown at the bottom of the page. To proceed, the following results are needed. If $\lambda_i > \tau(1+\sqrt{c})(i=1,\ldots,k)$ and λ_i has multiplicity 1, as $m, n \to \infty$ with $m/n \to c$, it follows from [38] that

$$\ell_i = \lambda_i + \frac{\lambda_i \tau}{\lambda_i - \tau}c + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \tag{B.5a}$$

$$\hat{\tau}_k = \tau + \mathcal{O}\left(\frac{1}{n}\right). \tag{B.5b}$$

On the other hand, under the same conditions and as $m, n \to \infty$ with $m/n \to c$, it is indicated in [38], [48], and [49] that the inner product of the largest sample and population eigenvectors converges almost surely to a deterministic value, which is given as

$$\boldsymbol{u}_i^H \boldsymbol{e}_i \xrightarrow{\text{a.s.}} \frac{1 - \frac{c\tau^2}{(\lambda_i-\tau)^2}}{1 + \frac{c\tau}{\lambda_i - \tau}}, \quad (i = 1, \ldots, k). \tag{B.5c}$$

Consequently, setting $\Delta\boldsymbol{\theta} \triangleq [\boldsymbol{\epsilon}^T, \boldsymbol{\nu}^T, \varepsilon]^T$, where

$$\boldsymbol{\epsilon} = \left[(\boldsymbol{u}_1 - \boldsymbol{e}_1)^T, \ldots, (\boldsymbol{u}_k - \boldsymbol{e}_k)^T\right]^T \tag{B.6a}$$

$$\boldsymbol{\nu} = -\tau c \left[\frac{\lambda_1}{\lambda_1 - \tau}, \ldots, \frac{\lambda_k}{\lambda_k - \tau}\right]^T + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \tag{B.6b}$$

$$\varepsilon = \mathcal{O}\left(\frac{1}{n}\right) \tag{B.6c}$$

$$-\frac{\partial^2 \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^H} = \left[\begin{array}{ccc|ccc|c} \frac{2n}{\lambda_1}\hat{\boldsymbol{R}} & & & \frac{-2n\hat{\boldsymbol{R}}\boldsymbol{v}_1}{\lambda_1^2} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{2n}{\lambda_k}\hat{\boldsymbol{R}} & & & \frac{-2n\hat{\boldsymbol{R}}\boldsymbol{v}_k}{\lambda_k^2} & 0 \\ \hline \frac{-2n\boldsymbol{v}_1^H\hat{\boldsymbol{R}}}{\lambda_1^2} & & & \frac{2n\boldsymbol{v}_1^H\hat{\boldsymbol{R}}\boldsymbol{v}_1}{\lambda_1^3} - \frac{n}{\lambda_1^2} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{-2n\boldsymbol{v}_k^H\hat{\boldsymbol{R}}}{\lambda_k^2} & & & \frac{2n\boldsymbol{v}_k^H\hat{\boldsymbol{R}}\boldsymbol{v}_k}{\lambda_k^3} - \frac{n}{\lambda_k^2} & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{2n\sum_{i=k+1}^{m}\boldsymbol{v}_i^H\hat{\boldsymbol{R}}\boldsymbol{v}_i}{\tau^3} - \frac{(m-k)n}{\tau^2} \end{array}\right] \tag{B.3}$$

$$\hat{\boldsymbol{J}} = -\left.\frac{\partial^2 \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^H}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left[\begin{array}{ccc|ccc|c} \frac{2n}{\ell_1}\hat{\boldsymbol{R}} & & & \frac{-2n\boldsymbol{e}_1}{\ell_1} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{2n}{\ell_k}\hat{\boldsymbol{R}} & & & \frac{-2n\boldsymbol{e}_k}{\ell_k} & 0 \\ \hline \frac{-2n\boldsymbol{e}_1^H}{\ell_1} & & & \frac{n}{\ell_1^2} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{-2n\boldsymbol{e}_k^H}{\ell_k} & & & \frac{n}{\ell_k^2} & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{(m-k)n}{\hat{\tau}_k^2} \end{array}\right] \tag{B.4}$$

the second-order term in (B.1) can be expressed as

$$\begin{aligned}
&\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta} \\
&= \frac{1}{2}[\boldsymbol{\epsilon}^H, \boldsymbol{\nu}^H, \varepsilon]\left[\begin{array}{cc|c} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} & \mathbf{0} \\ \hline \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} & \vdots \\ \hline \mathbf{0} & \cdots & \beta \end{array}\right]\begin{bmatrix}\boldsymbol{\epsilon}\\ \boldsymbol{\nu}\\ \varepsilon\end{bmatrix} \\
&= \frac{1}{2}(\boldsymbol{\epsilon}^H\boldsymbol{Q}_{11}\boldsymbol{\epsilon}+\boldsymbol{\epsilon}^H\boldsymbol{Q}_{12}\boldsymbol{\nu}+\boldsymbol{\nu}^H\boldsymbol{Q}_{21}\boldsymbol{\epsilon}+\boldsymbol{\nu}^H\boldsymbol{Q}_{22}\boldsymbol{\nu}+\varepsilon^2\beta) \quad \text{(B.7)}
\end{aligned}$$

where

$$\boldsymbol{Q}_{11} = \text{blkdiag}\left(\frac{2n}{\ell_1}\hat{\boldsymbol{R}}, \ldots, \frac{2n}{\ell_k}\hat{\boldsymbol{R}}\right) \quad \text{(B.8a)}$$

$$\boldsymbol{Q}_{12} = \text{blkdiag}\left(\frac{-2n\boldsymbol{e}_1}{\ell_1}, \ldots, \frac{-2n\boldsymbol{e}_k}{\ell_k}\right) \quad \text{(B.8b)}$$

$$\boldsymbol{Q}_{22} = \text{diag}\left(\frac{n}{\ell_1^2}, \ldots, \frac{n}{\ell_k^2}\right) \quad \text{(B.8c)}$$

$$\beta = \frac{n(m-k)}{\hat{\tau}^2} \quad \text{(B.8d)}$$

and $\boldsymbol{Q}_{21} = \boldsymbol{Q}_{12}^H$. Here, $\text{blkdiag}(\cdot)$ denotes the block diagonal matrix. Substituting (B.6) and (B.8) into (B.7), we can calculate $\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}$. In particular, notice that

$$\boldsymbol{\epsilon}^H\boldsymbol{Q}_{11}\boldsymbol{\epsilon} = \sum_{i=1}^{k}\frac{2n}{\ell_i}\boldsymbol{\epsilon}_i^H\hat{\boldsymbol{R}}\boldsymbol{\epsilon}_i \quad \text{(B.9)}$$

with $\boldsymbol{\epsilon}_i = \boldsymbol{u}_i - \boldsymbol{e}_i$ and $\boldsymbol{\epsilon}_i^H\hat{\boldsymbol{R}}\boldsymbol{\epsilon}_i = \boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i - \boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i - \boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i + \boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i$. Since $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ span the same observation space, we assert that, for $\boldsymbol{u}_i (i = 1, \ldots, m)$, there is a nonzero set $\{\alpha_{i1}, \ldots, \alpha_{im}\}$, such that

$$\boldsymbol{u}_i = \alpha_{i1}\boldsymbol{e}_1 + \cdots + \alpha_{im}\boldsymbol{e}_m \quad \text{(B.10)}$$

which implies that

$$\boldsymbol{u}_i^H\boldsymbol{u}_i = |\alpha_{i1}|^2 + \cdots + |\alpha_{im}|^2 = 1 \quad \text{(B.11)}$$

where $|\alpha_{ij}|$ denotes the absolute value of α_{ij}. It is easy to obtain

$$\boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i = |\alpha_{i1}|^2\ell_1 + \cdots + |\alpha_{im}|^2\ell_m \quad \text{(B.12a)}$$

$$\boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i = \boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i = \alpha_{ii}\ell_i \quad \text{(B.12b)}$$

$$\boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i = \ell_i. \quad \text{(B.12c)}$$

Therefore, substituting (B.12) into (B.9) yields

$$\boldsymbol{\epsilon}^H\boldsymbol{Q}_{11}\boldsymbol{\epsilon} = n\sum_{i=1}^{k}\left(2\sum_{j=1}^{m}|\alpha_{ij}|^2\frac{\ell_j}{\ell_i} - 4\alpha_{ii} + 2\right). \quad \text{(B.13a)}$$

Moreover, the second and third terms of (B.7) are given as

$$\boldsymbol{\epsilon}^H\boldsymbol{Q}_{12}\boldsymbol{\nu} = 2n\sum_{i=1}^{k}\frac{(\alpha_{ii}-1)\lambda_i\tau c}{\ell_i(\lambda_i-\tau)} + \mathcal{O}(\sqrt{n}) = \boldsymbol{\nu}^H\boldsymbol{Q}_{21}\boldsymbol{\epsilon}. \quad \text{(B.13b)}$$

In addition, it is easy to calculate the last two terms of (B.7) as

$$\boldsymbol{\nu}^H\boldsymbol{Q}_{22}\boldsymbol{\nu} = n\sum_{i=1}^{k}\frac{(\lambda_i\tau c)^2}{\ell_i^2(\lambda_i-\tau)^2} - \mathcal{O}(\sqrt{n}) \quad \text{(B.13c)}$$

$$\varepsilon^2\beta = (m-k)\mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}(1). \quad \text{(B.13d)}$$

Consequently, substituting (B.13) into (B.7), we attain

$$\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta} = n\sum_{i=1}^{k}\left(\sum_{j=1}^{m}|\alpha_{ij}|^2\frac{\ell_j}{\ell_i} + \frac{2(\alpha_{ii}-1)\lambda_i\tau c}{\ell_i(\lambda_i-\tau)} + \frac{(\lambda_i\tau c)^2}{2\ell_i^2(\lambda_i-\tau)^2} - 2\alpha_{ii} + 1\right) + \mathcal{O}(\sqrt{n}). \quad \text{(B.14)}$$

Utilizing $\ell_1 \geq \cdots \geq \ell_m$ and $\alpha_{ii} = \boldsymbol{u}_i^H\boldsymbol{e}_i \in [0, 1]$, we assert

$$\frac{\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}}{2mn} \leq \frac{1}{m}\sum_{i=1}^{k}\left(\frac{\ell_1}{\ell_i} + 1 + \frac{(\lambda_i\tau c)^2}{2\ell_i^2(\lambda_i-\tau)^2}\right) + \mathcal{O}\left(\frac{1}{m\sqrt{n}}\right) \xrightarrow{m,n\to\infty, m/n\to c} 0. \quad \text{(B.15)}$$

However, it follows from (B.2) that $(1/nm)\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})$ is bounded as $m, n \to \infty$ and $m/n \to c$. As a result, as $m, n \to \infty$ and $m/n \to c$, by omitting the high-order terms in (B.1), we have

$$\log f(\boldsymbol{X}|\boldsymbol{\theta}) \approx \log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) - \frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}. \quad \text{(B.16)}$$

On the other hand, assuming that the *a priori* pdf of $\boldsymbol{\theta}$ is flat around $\hat{\boldsymbol{\theta}}$, we obtain $f(\boldsymbol{\theta}) \approx f(\hat{\boldsymbol{\theta}})$. Substituting this result along with (B.16) into (9a), we get

$$\begin{aligned}
f(\mathcal{H}_k|\boldsymbol{X}) &\approx f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})f(\hat{\boldsymbol{\theta}})\int \exp\left(-\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}\right)d\boldsymbol{\theta} \\
&= \frac{\pi^{\nu_k}f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})f(\hat{\boldsymbol{\theta}})}{|\hat{\boldsymbol{J}}|^{\frac{1}{2}}} \times \underbrace{\int \frac{1}{\pi^{\nu_k}|\hat{\boldsymbol{J}}^{-1}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}\right)d\boldsymbol{\theta}}_{=1} \\
&= \pi^{\nu_k}|\hat{\boldsymbol{J}}|^{\frac{1}{2}}f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})f(\hat{\boldsymbol{\theta}}). \quad \text{(B.17)}
\end{aligned}$$

Taking the logarithm of (B.17) eventually leads to (15).

Appendix C
Proof of $P_{\text{fa}} \to 0$ as $m, n \to \infty$, and $m/n \to c$

Similar to (31) and (32), we have

$$(m-d)\log\frac{a(d)}{g(d)} = (m-d-1)\log\frac{a(d+1)}{g(d+1)} + \log Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right] \quad \text{(C.1)}$$

where

$$Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right] = \frac{\left[1+\frac{1}{m-d}\left(\frac{\ell_{d+1}}{a(d+1)}-1\right)\right]^{(m-d)}}{\frac{\ell_{d+1}}{a(d+1)}}. \quad \text{(C.2)}$$

Therefore, noticing that $a(d+1) \approx a(d) \approx \tau$, it follows from (25) and (C.1) that the probability of false alarm is calculated as

$$\begin{aligned}
P_{\text{fa}} &\approx \text{Prob}\left(\text{BIC}(d+1) - \text{BIC}(d) < 0|\mathcal{H}_d\right) \\
&= \text{Prob}\left(\log Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right] > \frac{\mathcal{P}(d+1,m,n) - \mathcal{P}(d,m,n)}{2n}\,\middle|\,\mathcal{H}_d\right)
\end{aligned}$$

$$\approx \text{Prob}\left(\log Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right] > \frac{m}{2n}\log(2n) - \frac{m}{2n}\log\frac{\ell_{d+1}}{\tau}\,\middle|\,\mathcal{H}_d\right). \quad \text{(C.3)}$$

Substituting (C.2) into (C.3) and using $a(d+1) \approx \tau$ again, we can approximate P_{fa} for the proposed BIC criterion as

$$\begin{aligned} P_{\text{fa}} &\approx \text{Prob}\left(m\log\left(1+\frac{\ell_{d+1}/\tau - 1}{m}\right) - \left(1-\frac{c}{2}\right)\log\frac{\ell_{d+1}}{\tau} > \frac{c}{2}\log(2n)\,\middle|\,\mathcal{H}_d\right) \\ &\approx \text{Prob}\left(\frac{\ell_{d+1}}{\tau} - \left(1-\frac{c}{2}\right)\log\frac{\ell_{d+1}}{\tau} > \frac{c}{2}\log(2n)+1\,\middle|\,\mathcal{H}_d\right) \\ &= \text{Prob}\left(\frac{\ell_{d+1}}{\tau} > g^{-1}\left(\frac{c}{2}\log(2n)+1\right)\middle|\,\mathcal{H}_d\right) \qquad \text{(C.4)} \end{aligned}$$

where $g^{-1}(x)$ is the inverse function of $g(x) = x - (1 - c/2)\log(x)$, which is a monotonically increasing function for $x \geq 1$. We now need to determine the distribution of ℓ_{d+1}/τ. As a matter of fact, ℓ_{d+1} has the similar limiting behavior as the largest sample eigenvalue of $\hat{\boldsymbol{R}}$ in the noise-only case [50], which is described by the following lemma due to [51].

Lemma 1: Let ℓ_{max} be the largest sample eigenvalue of $\tilde{\boldsymbol{R}} \in \mathbb{C}^{(m-d)\times(m-d)}$ for the noise-only case. The normalized sample eigenvalue, i.e., ℓ_{max}/τ, is distributed as Tracy–Widom distribution of order 2. That is

$$\frac{\frac{\ell_{max}}{\tau} - \mu_{mn}}{\sigma_{mn}} \xrightarrow{\mathcal{D}} \mathcal{W} \sim F_{TW_2} \qquad \text{(C.5)}$$

where $\mu_{mn} = (1+\sqrt{c})^2$, $\sigma_{mn} = (1+\sqrt{c})^{4/3}/n\sqrt{c}$, and

$$F_{TW_2}(s) = \exp\left\{-\int_s^{\infty}(u-s)q^2(u)du\right\} \qquad \text{(C.6)}$$

with $q(u)$ being the solution to the nonlinear Painlevé II differential equation, i.e.,

$$q''(u) = uq(u) + 2q^3(u). \qquad \text{(C.7)}$$

Details concerning the analytical formula of $F_{TW_2}(s)$ can be found in [51], and the lookup table for the cdf of $F_{TW_2}(s)$ is available in [52].

As ℓ_{d+1} asymptotically has the behavior of $\ell_{\max}$, it follows from (C.4) and (C.5) that

$$\begin{aligned} P_{\text{fa}} &\approx \text{Prob}\left(\frac{\frac{\ell_{d+1}}{\tau} - \mu_{mn}}{\sigma_{mn}} > \frac{\delta - \mu_{mn}}{\sigma_{mn}}\right) \\ &= 1 - \text{Prob}\left(\frac{\frac{\ell_{d+1}}{\tau} - \mu_{mn}}{\sigma_{mn}} < \frac{\delta - \mu_{mn}}{\sigma_{mn}}\right) \\ &= 1 - F_{TW_2}\left(\frac{\delta - \mu_{mn}}{\sigma_{mn}}\right) \qquad \text{(C.8)} \end{aligned}$$

where $\delta = g^{-1}(c/2\log(2n) + 1)$.

Using [51, App. A1], we assert that $q^2(u)$ is monotonically decreasing asymptotic to $|u|/2$ as $u \to -\infty$ and to $e^{-(4/3)u^{3/2}}/(4\pi\sqrt{u})$ as $u \to \infty$. Since $g^{-1}(x)$ is the increasing function, we obtain $\delta \to \infty$ as $n \to \infty$. As a result, it follows from (C.6) that

$$\begin{aligned} \lim_{s\to\infty} F_{TW_2}(s) &= \lim_{s\to\infty}\exp\left\{-\int_s^{\infty}(u-s)\frac{e^{-\frac{4}{3}u^{\frac{3}{2}}}}{4\pi\sqrt{u}}du\right\} \\ &= \lim_{s\to\infty}\exp\left\{-\int_s^{\infty}\frac{\sqrt{u}e^{-\frac{4}{3}u^{\frac{3}{2}}}}{4\pi}du\right\} \\ &\quad \times \lim_{s\to\infty}\exp\left\{\int_s^{\infty}\frac{se^{-\frac{4}{3}u^{\frac{3}{2}}}}{4\pi\sqrt{u}}du\right\}. \qquad \text{(C.9)} \end{aligned}$$

Noting that

$$\begin{aligned} -\int_s^{\infty}\frac{\sqrt{u}}{4\pi}e^{-\frac{4}{3}u^{\frac{3}{2}}}du = -\frac{1}{6\pi}\int_{s^{\frac{3}{2}}}^{\infty}e^{-\frac{4}{3}t}dt = -\frac{\pi}{8}e^{-\frac{4}{3}s^{\frac{3}{2}}} \\ \to 0 \text{ as } s \to \infty \qquad \text{(C.10)} \end{aligned}$$

$$\begin{aligned} 0 < \int_s^{\infty}\frac{s}{4\pi\sqrt{u}}e^{-\frac{4}{3}u^{\frac{3}{2}}}du < \int_s^{\infty}\frac{\sqrt{u}}{4\pi}e^{-\frac{4}{3}u^{\frac{3}{2}}}du = \frac{\pi}{8}e^{-\frac{4}{3}s^{\frac{3}{2}}} \\ \to 0 \text{ as } s \to \infty \qquad \text{(C.11)} \end{aligned}$$

we assert that $F_{TW_2}((\delta - \mu_{mn})/\sigma_{mn}) \to 1$ as $\delta \to \infty$, which, when substituted into (C.8), establishes that the probability of false alarm converges to zero as $m, n \to \infty$ and $m/n \to c$.

References

[1] D. Gesbert, M. Kountouris, R. W. Heath, C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[3] X. Mestre and M. A. Lagunas, "Modified subspace algorithms for DOA estimation with large arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598–613, Feb. 2008.

[4] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[5] D. B. Williams and D. H. Johnson, "Using the sphericity test for source detection with narrow-band passive arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 11, pp. 2008–2014, Nov. 1990.

[6] Q. Wu and K. M. Wong, "Determination of the number of signals in unknown noise environments-PARADE," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 362–365, Jan. 1995.

[7] P.-J. Chung, J. F. Böhme, C. F. Mecklenbräuker, and A. O. Hero, "Detection of the number of signals using the Benjamini–Hochberg procedure," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2497–2508, Jun. 2007.

[8] 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.

[9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[10] D. F. Schmidt and E. Makalic, "The consistency of MDL for linear regression models with increasing signal-to-noise ratio," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1508–1510, Mar. 2011.

[11] M. Lu and A. M. Zoubir, "Generalized Bayesian information criterion for source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1470–1480, Mar. 2013.

[12] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

[13] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[14] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.

[15] C. Xu and S. Kay, "Inconsistency of the MDL: On the performance of model order selection criteria with increasing signal-to-noise ratio," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1959–1969, May 2011.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] L. Huang, T. Long, E. Mao, and H. C. So, "MMSE-based MDL method for robust estimation of number of sources without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4135–4142, Oct. 2009.

[21] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.

[22] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.

[23] 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.

[24] 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.

[25] 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.

[26] A.-K. Seghouane, "New AIC corrected variants for multivariante linear regression model selection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 2, pp. 1154–1165, Apr. 2011.

[27] P. Djurić, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2726–2735, Oct. 1998.

[28] P. Stoica and Y. Selén, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[29] O. E. Barndorff-Nielsen and D. R. Cox, *Asymptotic Techniques for Use in Statistics*. New York, NY, USA: Chapman and Hall, 1989.

[30] G. Golub and C. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[31] H. Wang and M. Kaveh, "On the performance of signal subspace processing—Part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1201–1209, Oct. 1986.

[32] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1557–1567, Oct. 1989.

[33] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413–1426, Jun. 1995.

[34] F. Haddadi, M. Malek-Mohammadi, M. M. Nayebi, and M. R. Aref, "Statistical performance analysis of MDL source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 452–457, Jan. 2010.

[35] J. P. Delmas and Y. Meurisse, "On the second-order statistics of the EVD of sample covariance matrices: Application to the detection of noncircular or/and non Gaussian components," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4017–4023, Aug. 2010.

[36] J. Baik, G. B. Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, Sep. 2005.

[37] A. Onatski, M. Moreira, and M. Hallin, "Signal detection in high dimension: The multispiked case," *arXiv preprint arXiv:1210.5663*, 2012.

[38] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statist. Sin.*, vol. 17, no. 4, pp. 1617–1642, 2007.

[39] D. N. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, vol. 43, no. 1/2, pp. 128–136, Jun. 1956.

[40] L. Huang and H. C. So, "Source enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4806–4821, Oct. 2013.

[41] C. Xu and S. Kay, "Source enumeration via the EEF criterion," *IEEE Signal Process. Lett.*, vol. 15, pp. 569–572, 2008.

[42] B. Nadler, "Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2746–2756, May 2010.

[43] P. O. Perry and P. J. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 504–513, Jun. 2010.

[44] C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Commun. Math. Phys.*, vol. 177, no. 3, pp. 727–754, 1996.

[45] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Mulrivariare Anal.*, vol. 20, no. 1, pp. 1–25, Oct. 1986.

[46] J. V. Neumann, "Some matrix inequalities and metrization of matric-space," *Tomsk. Univ. Rev.*, no. 1, pp. 286–300, 1937.

[47] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," Dept. Elect. Comput. Eng., Univ. California, San Diego, CA, USA, Tech. Rep. Course Lect. Suppl. ECE275A, Sep.–Dec. 2005. [Online]. Available: http://dsp.ucsd.edu/kreutz/PEI05.html AQ1

[48] I. M. Johnstone and A. Y. Lu, "Sparse principal components analysis," Stanford Univ., Stanford, CA, USA, Tech. Rep. (ArXiv:0901.4392v1), 2004.

[49] I. M. Johnstone, "High dimensional statistical inference and random matrices," in *Proc. Int. Congr. Math.*, M. Sanz-Solé, J. Soria, J. Varona, and J. Verdera, Eds., Zürich, Switzerland, Eur. Math. Soc., 2006, pp. 307–333.

[50] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices: A review," *Statist. Sin.*, vol. 9, no. 3, pp. 611–677, Aug. 1999. AQ2

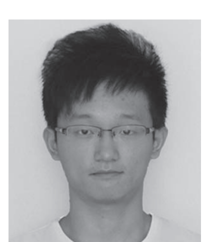
[51] I. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, Apr. 2001.

[52] A. Bejan, "Largest eigenvalues and sample covariance matrices," Tracy–Widom and Painleve II: Computational Aspects and Realization in S-Plus With Applications, 2005. [Online]. Available: http://www.vitrum.md/andrew/TWinSplus.pdf AQ3

Lei Huang (M'07–SM'14) 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, USA. From 2009 to 2010, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, and a Research Associate with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. From 2011 to 2014, he was a Professor with the Department of Electronic and Information Engineering, Shenzhen Graduate School of Harbin Institute of Technology, Shenzhen, China. In November 2014, he joined the Department of Information Engineering, Shenzhen University, where he is currently a Chair 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 serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and *Digital Signal Processing*.



Yuhang Xiao was born in Anhui, China, on January 20, 1992. He received the B.E. degree from Harbin Engineering University, Harbin, China, in 2012. He is currently working toward the Ph.D. degree in communication and information engineering with the Harbin Institute of Technology.

His research interests are in statistical signal processing and spectrum sensing.

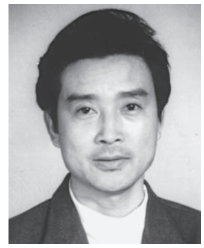
Kefei Liu received the B.Sc. degree in mathematics from Wuhan University, Wuhan, China, in 2006 and the Ph.D. degree in electronic engineering from City University of Hong Kong, Kowloon, Hong Kong, in 2013, respectively. His Ph.D. supervisor was Prof. H.-C. So, and his Ph.D. research topics were statistical and array signal processing, source enumeration, direction-of-arrival estimation, and multilinear algebra.

From September 2013 to December 2013, he was a Research Assistant of Prof. L. Huang with the Department of Electronic and Information Engineering, Shenzhen Graduate School of Harbin Institute of Technology, Shenzhen, China. Since January 2014, he has been a Postdoctoral Research Associate with the Department of Computer Science and Engineering and the Center for Evolutionary Medicine and Informatics, Biodesign Institute, Arizona State University, Tempe, AZ, USA. His cooperative supervisor is Prof. J. Ye. His current research interests are tensor decompositions for machine learning, randomized algorithms for matrix approximation, and their applications in analysis of massive biomedical data sets.

Hing Cheung So (S'90–M'95–SM'07–F'15) was born in Hong Kong. He received the B.Eng. degree in electronic engineering from City University of Hong Kong, Kowloon, Hong Kong, in 1990 and the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, Shatin, Hong Kong, in 1995.

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 was 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, robust estimation, source localization, and sparse approximation.

Dr. So has been on the Editorial Board of the IEEE SIGNAL PROCESSING MAGAZINE since 2014, *Signal Processing* since 2010, and *Digital Signal Processing* since 2011 and was on the Editoral Board of the IEEE TRANSACTIONS ON SIGNAL PROCESSING during 2010–2014. In addition, since 2011, he has been an elected member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society, where he is the Chair of the awards subcommittee. He is elected Fellow of the IEEE in recognition of his contributions to spectral analysis and source localization.



Jian-Kang Zhang (SM'09) received the B.S. degree in information science (mathematics) from Shaanxi Normal University, Xi'an, China, in 1983; the M.S. degree in information and computational science (mathematics) from Northwest University, Xi'an, in 1988; and the Ph.D. degree in electrical engineering from Xidian University, Xi'an, in 1999.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. He has held research positions at McMaster University and Harvard University, Cambridge, MA, USA. His research interests are in the general area of signal processing, digital communication, signal detection and estimation, and wavelet and time–frequency analysis, mainly emphasizing mathematics-based new-technology innovation and exploration for a variety of signal processing and practical applications, and, specifically, number theory and various linear algebra-based kinds of signal processing. His current research focuses on transceiver designs for multiuser communication systems, coherent and noncoherent space–time signal, and receiver designs for multiple-input–multiple-output and cooperative relay communications.

Dr. Zhang is the coauthor of the paper that received the IEEE Signal Processing Society Best Young Author Award in 2008. He has served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the *Journal of Electrical and Computer Engineering*.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = Provided URL in Ref. [47] was not found. Please check.
AQ2 = Note that references [16] and [50] are the same. Therefore, reference [50] was changed according to its publication details. Please confirm.
AQ3 = Provided URL in Ref. [52] was not found. Please check.

END OF ALL QUERIES

Bayesian Information Criterion for Source Enumeration in Large-Scale Adaptive Antenna Array

Lei Huang, *Senior Member, IEEE*, Yuhang Xiao, Kefei Liu, Hing Cheung So, *Fellow, IEEE*, and Jian-Kang Zhang, *Senior Member, IEEE*

***Abstract*—Subspace-based high-resolution algorithms for direction-of-arrival (DOA) estimation have been developed for large-scale adaptive antenna arrays. However, its prerequisite step, namely, source enumeration, has yet to be addressed. In this paper, a new approach is devised in the framework of the Bayesian information criterion (BIC) to provide reliable detection of the signal source number for *the general asymptotic regime*, where $m, n \to \infty$ and $m/n \to c \in (0, \infty)$, with m and n being the numbers of antennas and snapshots, respectively. In particular, the *a posteriori* probability is determined by correctly calculating the LLFs and PFs for the general asymptotic case. By means of the maximum *a posteriori* probability, we are capable of effectively finding the signal number. An accurate closed-form expression for the probability of missed detection is also derived for the proposed BIC variant. In addition, the probability of false alarm for the BIC detector is proved to converge to zero as $m, n \to \infty$ and $m/n \to c$. Simulation results are included to demonstrate the superiority of the proposed detection approach over state-of-the-art schemes and corroborate our theoretical calculations.**

***Index Terms*—Adaptive antenna array, Bayesian information criterion (BIC), direction-of-arrival (DOA) estimation, source enumeration.**

I. Introduction

As a promising technique to boost spectral efficiency, large-scale adaptive antenna arrays have received much attention in the literature [1], [2]. As the array utilizes a large number of antennas at the base station for transmission and reception, the conventional subspace-based algorithms for direction-of-arrival (DOA) estimation usually suffer serious performance degradation. This is because the subspace cannot be correctly determined for the situation where the number of antennas is comparable with the number of samples. To cope with the problem, more efficient subspace-based algorithms [3], [4] have been suggested for the large array. Nevertheless, as the prerequisite step of direction finding, source enumeration has not yet been addressed for such a situation, which turns out to be a big challenge, particularly at low signal-to-noise ratios (SNRs) or small samples.

The conventional source enumeration methodologies vary from hypothesis testing [5]–[8] to the information-theoretic criterion (ITC) [9]–[11]. Basically, the hypothesis testing, including the sphericity test [5] and the random matrix theory (RMT)-based test [8], needs to find a subjective threshold for decision making. It is shown in [8] that the RMT approach is able to provide a detection threshold that is significantly smaller than that of the classical minimum description length (MDL) method [12]. Unlike the hypothesis testing, the ITCs, such as Akaike's information criterion (AIC) [9], Schwarz's Bayesian information criterion (BIC) [13], Rissanen's MDL [14], and Kay's exponentially embedded family (EEF) [15], are derived from the perspective of information theory, and no user-defined parameter is needed. As a result, it is of considerable interest to exploit the information criterion for efficient source enumeration. Wax and Kailath [12] have employed the AIC and MDL to enumerate independent signal sources. To handle coherent signals, Wax and Ziskind [16] have combined the maximum likelihood (ML) estimates of the DOAs with the MDL principle for joint DOA estimation and source enumeration, ending up with an enhanced MDL criterion for coherent source enumeration. Valaee and Kabal [17] have proposed a predictive description length (PDL) for this task. Although the PDL method can outperform the MDL approach [16], it requires much more computational cost than the latter as ML estimation is required at each snapshot. Fishler and Poor [18] have reformulated the MDL criterion for source enumeration under nonuniform noise environment. Furthermore, they have proved the consistency of their proposed MDL variant. On the other hand, Huang *et al.* [19], [20] have developed MDL variants by using the filtered component variances or minimum mean square errors of the multistage Wiener filter rather than the sample eigenvalues corrupted by the nonuniform noise, ending up with computationally simple and robust source enumerators.

Most of the aforementioned methods are devised by utilizing the assumption that the number of antennas m is fixed while the number of snapshots n tends to infinity, which is referred to as the classical asymptotic regime. Indeed, the general asymptotic situation [21], where $m, n \to \infty$ and $m/n \to c \in (0, \infty)$, is more suitable to large-array applications since the number of antennas can be as large as the number of snapshots. On the other

Manuscript received February 26, 2015; revised May 8, 2015; accepted May 14, 2015. This work was supported by the National Natural Science Foundation of China under Grant 61222106 and Grant 61171187. The review of this paper was coordinated by Dr. T. Jiang.

L. Huang is with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: dr.lei.huang@ieee.org).

Y. Xiao is with the Department of Electronic and Information Engineering, Harbin Institute of Technology, Harbin 518055, China.

K. Liu is with the Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287-5406 USA.

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

J.-K. Zhang is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada.

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/TVT.2015.2436060

hand, it has been pointed out in [3] that the general asymptotic regime is able to provide a more accurate description for practical scenarios, where the number of snapshots and the number of antennas are finite and probably comparable in magnitude. In fact, the topics of DOA estimation and beamforming have been dealt with in [3] and [22]–[24] for the general asymptotic regime.

Basically, the ITCs have their roots in the minimization of the Kullback–Leibler (KL) information, but this minimization is carried out in the scenario where the number of antennas is fixed while the number of snapshots tends to infinity. This, in turn, means that these ITCs cannot properly work in the general asymptotic case. To enable the ITCs to properly detect the source number in this condition, Nadakuditi and Edelman [25] have devised the RMT-AIC criterion. Although the RMT-AIC is argued to be able to correctly detect the source number for the general asymptotic case, it cannot provide the consistent estimate of the source number [8]. To solve the issue of linear regression model order selection for small sample cases, variants of the AIC approach have been proposed in [26] by means of the asymptotic approximation of the bootstrap estimation of the KL information. Nevertheless, it is nontrivial to apply them to source enumeration for the large array. Therefore, it is considerably interesting to investigate the consistent methodology for source enumeration in the general asymptotic regime.

We would prefer a source enumerator that always selects the true source number, provided that the number of snapshots is large enough. It has been revealed in [12] that the BIC method offers strong consistency, whereas the AIC approach does not. As a result, the former has drawn much attention in the literature. The classical BIC criterion is composed of a likelihood function (LF) and a penalty function (PF), which correspond to data fitting and model complexity, respectively. Minimization of the BIC criterion is, in fact, a procedure trading off data fitting and model complexity, resulting in a correct estimate of the model order or source number. As previously pointed out, the existing BIC criterion does lead to the minimization of the relative KL information between the generating model and the fitted approximating model but only for the case in which m is fixed while $n \to \infty$. In the general asymptotic regime, however, there is no guarantee that what the classical BIC criterion is minimizing is exactly the relative KL divergence and that minimization of the classical BIC criterion yields a correct estimate of the source number. To circumvent this issue, we derive a variant of the BIC criterion for the general asymptotic case, in which $m, n \to \infty$ and $m/n \to c$. In particular, we reformulate the BIC criterion by calculating the LF and PF in this general asymptotic regime. Through appropriate approximations, we are able to accurately determine the LF and the PF for the BIC criterion, ending up with a new variant of the BIC criterion for source enumeration. This enables us to precisely determine the signal and noise subspaces for the subsequent DOA estimation and beamforming in large arrays. Moreover, a closed-form expression for the probability of missed detection is derived. It is also proved that the probability of false alarm converges to zero as $m, n \to \infty$ and $m/n \to c$.

The remainder of this paper is organized as follows. The data model is presented in Section II. The method for source enumeration is proposed in Section III. Statistical performance analysis is conducted in Section IV. Simulation results are presented in Section V. Finally, conclusions are drawn in Section VI.

II. Problem Formulation

Consider an array of m antennas receiving d narrowband source signals $\{s_1(t), \ldots, s_d(t)\}$ from distinct directions $\{\varphi_1, \ldots, \varphi_d\}$, respectively. Assume that the sources and array are in the same plane. In the sequel, the tth snapshot vector of the array output is written as

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{w}_t, \quad (t = 1, \ldots, n) \tag{1}$$

where $\boldsymbol{x}_t = [x_1(t), \ldots, x_m(t)]^T \in \mathbb{C}^{m\times 1}$, $\boldsymbol{A} = [\boldsymbol{a}(\varphi_1), \ldots, \boldsymbol{a}(\varphi_d)] \in \mathbb{C}^{m\times d}$, $\boldsymbol{s}_t = [s_1(t), \ldots, s_d(t)]^T \in \mathbb{C}^{d\times 1}$, and $\boldsymbol{w}_t = [w_1(t), \ldots, w_m(t)]^T \in \mathbb{C}^{m\times 1}$ are the observed snapshot vector, the steering matrix, the signal vector, and the noise vector, respectively. Here, $\boldsymbol{a}(\varphi_i), i = 1, \ldots, d$, is the steering vector, with φ_i being the DOA due to the ith source, $(\cdot)^T$ is the transpose operator, d is the *unknown* number of sources, m is the number of antennas, and n is the number of snapshots. For simplicity but without loss of generality, it is assumed that $m < n$ throughout this paper, unless stated otherwise. Moreover, the number of sources is assumed to be fixed and smaller than a constant number $\bar{m}$, which is much less than $\min(m, n)$, i.e., $\bar{m} \ll \min(m, n)$, as $m, n \to$ with $m/n \to c$. The incoherent signals are independent and identically distributed (i.i.d.) complex Gaussian distributed, i.e., $\boldsymbol{s}_t \sim \mathcal{CN}(\mathbf{0}_d, \boldsymbol{R}_s)$, in which $\mathbf{0}_d$ is the $d \times 1$ zero vector, and $\boldsymbol{R}_s \triangleq \mathbb{E}[\boldsymbol{s}_t\boldsymbol{s}_t^H] \in \mathbb{C}^{d\times d}$ has full rank, with $(\cdot)^H$ being the conjugate transpose and $\mathbb{E}[\cdot]$ being the mathematical expectation. Here, $\mathcal{CN}(\boldsymbol{\nu}, \boldsymbol{R})$ stands for the complex Gaussian distribution with mean $\boldsymbol{\nu}$ and covariance $\boldsymbol{R}$. Furthermore, the noise $\boldsymbol{w}_t$ is assumed to be an i.i.d. complex Gaussian vector with mean zero and covariance $\tau\boldsymbol{I}_m$, i.e., $\boldsymbol{w}_t \sim \mathcal{CN}(\mathbf{0}_m, \tau\boldsymbol{I}_m)$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, which is independent of the signals.

With the given assumptions, the observed samples can be taken as the i.i.d. Gaussian vector, i.e., $\boldsymbol{x}_t \sim \mathcal{CN}(\mathbf{0}_m, \boldsymbol{R})$, with $\boldsymbol{R}$ being the population covariance matrix, which is calculated as

$$\boldsymbol{R} = \mathbb{E}\left[\boldsymbol{x}_t\boldsymbol{x}_t^H\right] = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + \tau\boldsymbol{I}_m. \tag{2}$$

Recall that the signals are incoherent and $d < m$, which means that $\boldsymbol{R}_s$ is nonsingular and that $\boldsymbol{A}$ is of full column rank. Without loss of generality, we assume that the population eigenvalues of $\boldsymbol{R}$, which are denoted as $\lambda_1, \ldots, \lambda_m$, are nonincreasingly ordered, i.e.,

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

In addition, their corresponding population eigenvectors are denoted as $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$. Given (3), it is straightforward to utilize the multiplicity of τ to determine the number of signals. In practice, however, only the sample covariance matrix is accessible, which is calculated by

$$\hat{\boldsymbol{R}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t^H. \tag{4}$$

Let $\ell_1, \ldots, \ell_m$ and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$, be the descending eigenvalues and corresponding eigenvectors of $\hat{\boldsymbol{R}}$, respectively. Consequently, our task in this work is to infer the source number d from the noisy observations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ for $m, n \to \infty$ and $m/n \to c$.

III. Bayesian Information Criterion for Source Enumeration

A. BIC

For the i.i.d. complex Gaussian observations $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, the joint probability density function (pdf) is

$$f(\boldsymbol{X}|\boldsymbol{\theta}) = \prod_{t=1}^{n} \frac{1}{\pi^m |\boldsymbol{R}|} \exp\left(-\boldsymbol{x}_t^H \boldsymbol{R}^{-1} \boldsymbol{x}_t\right) \tag{5}$$

where $|\cdot|$ is the determinant, and $\boldsymbol{\theta}$ is the unknown parameter vector of the true model, which is specifically given by $\boldsymbol{\theta} = [\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_d^T, \lambda_1, \ldots, \lambda_d, \tau]^T$. Suppose that we have a parametric family of pdf $\{f(\boldsymbol{X}|\boldsymbol{\theta}^{(k)})\}_{k=0}^{\bar{m}-1}$ with

$$f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right) = \prod_{t=1}^{n} \frac{1}{\pi^m \left|\boldsymbol{R}^{(k)}\right|} \exp\left(-\boldsymbol{x}_t^H \left[\boldsymbol{R}^{(k)}\right]^{-1} \boldsymbol{x}_t\right) \tag{6}$$

where $\boldsymbol{\theta}^{(k)} = [\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_k^T, \lambda_1, \ldots, \lambda_k, \tau]^T$ corresponds to the kth candidate model. Let $\mathcal{H}_k$ be the hypothesis that the source number is $k \in [0, \bar{m}-1]$. It is easy to see that the hypotheses $\{\mathcal{H}_k\}_{k=0}^{\bar{m}-1}$ are nested.

According to Bayes' rule, we readily have

$$f(\mathcal{H}_k|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\mathcal{H}_k) f(\mathcal{H}_k)}{f(\boldsymbol{X})}. \tag{7}$$

Most typically, $\{\mathcal{H}_k\}_{k=0}^{\bar{m}-1}$ are assumed to be uniformly distributed, yielding $f(\mathcal{H}_k) = 1/\bar{m}$. Moreover, notice that $f(\boldsymbol{X})$ is independent of k, which, when ignored, does not affect the maximization of (7) with respect to k. As a result, we obtain from (7) that

$$\max_{k\in[0,\bar{m}-1]} f(\mathcal{H}_k|\boldsymbol{X}) = \max_{k\in[0,\bar{m}-1]} f(\boldsymbol{X}|\mathcal{H}_k). \tag{8}$$

It is indicated in (8) that maximization of the detection probability under the hypothesis $\mathcal{H}_k$ is equivalent to finding the maximum *a posteriori* probability. It follows from [27] and [28] that the *a posteriori* probability is computed as

$$\begin{aligned} f(\mathcal{H}_k|\boldsymbol{X}) &= \int f\left(\boldsymbol{X}, \boldsymbol{\theta}^{(k)}\right) d\boldsymbol{\theta}^{(k)} \\ &= \int f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right) f\left(\boldsymbol{\theta}^{(k)}\right) d\boldsymbol{\theta}^{(k)} \qquad \text{(9a)} \\ &\approx (2\pi)^{\frac{\nu_k}{2}} |\hat{\boldsymbol{J}}|^{-\frac{1}{2}} f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) \qquad \text{(9b)} \end{aligned}$$

where $f(\boldsymbol{X}, \boldsymbol{\theta}^{(k)})$ denotes the joint pdf of $\boldsymbol{X}$ and $\boldsymbol{\theta}^{(k)}$, $f(\boldsymbol{\theta}^{(k)})$ denotes the *a priori* pdf of $\boldsymbol{\theta}^{(k)}$, $\hat{\boldsymbol{\theta}}^{(k)}$ is the ML estimate of $\boldsymbol{\theta}^{(k)}$, ν_k is the length of $\boldsymbol{\theta}^{(k)}$, and

$$\hat{\boldsymbol{J}} = -\left.\frac{\partial^2 \log f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right)}{\partial\boldsymbol{\theta}^{(k)} \partial\left(\boldsymbol{\theta}^{(k)}\right)^H}\right|_{\boldsymbol{\theta}^{(k)}=\hat{\boldsymbol{\theta}}^{(k)}} \in \mathbb{C}^{\nu_k \times \nu_k} \tag{10}$$

is the Hessian matrix. Taking mathematical expectation of $\hat{\boldsymbol{J}}$ leads to the Fisher information matrix

$$\boldsymbol{J} = -\mathbb{E}\left[\frac{\partial^2 \log f\left(\boldsymbol{X}|\boldsymbol{\theta}^{(k)}\right)}{\partial\boldsymbol{\theta}^{(k)} \partial\left(\boldsymbol{\theta}^{(k)}\right)^H}\right]. \tag{11}$$

Note that, although [27] and [28] can arrive at the approximation in (9b), the former employs the assumption that the *a priori* pdf of $\boldsymbol{\theta}^{(k)}$ is flat around $\hat{\boldsymbol{\theta}}^{(k)}$, which means that $f(\boldsymbol{\theta}^{(k)}) \approx f(\hat{\boldsymbol{\theta}}^{(k)})$, whereas the latter utilizes Laplace's method [29] for integration. Taking the logarithm of (9b) yields

$$\begin{aligned} &\log f(\mathcal{H}_k|\boldsymbol{X}) \\ &\approx \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \log f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) + \frac{\nu_k}{2}\log 2\pi - \frac{1}{2}\log|\hat{\boldsymbol{J}}| \\ &= \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \log f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) + \frac{\nu_k}{2}\log 2\pi - \frac{1}{2}\log\left|n\cdot\frac{1}{n}\hat{\boldsymbol{J}}\right| \\ &\approx \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) - \frac{1}{2}\nu_k \log n. \end{aligned} \tag{12}$$

The approximation in (12) is due to the fact that $\log f(\boldsymbol{\theta}^{(k)})$ and $(\nu_k/2)\log 2\pi$ are independent of n, and $\hat{\boldsymbol{J}}/n = \mathcal{O}(1)$ for the case where m is fixed while $n \to \infty$. Here, $\mathcal{O}(1)$ denotes a term that tends to a constant as $n \to \infty$. Consequently, invoking the results in [12] for log-LF (LLF) calculation, ignoring the terms independent of k and setting $\nu_k = k(2m-k)$, the classical BIC method is given as

$$\begin{aligned} \mathrm{BIC}(k) &= -2\log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \nu_k \log n \\ &= 2n(m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i}{\left(\prod_{i=k+1}^{m} \ell_i\right)^{\frac{1}{m-k}}} Q \\ &\quad + k(2m-k)\log n. \end{aligned} \tag{13}$$

Minimizing (13) with respect to k yields the estimate of the source number. It should be noted that the criterion in (13) can also be obtained from a different procedure based on the MDL principle [12], [14], [16].

For $m, n \to \infty$ and $m/n \to c$, however, the observed information matrix $\hat{\boldsymbol{J}}$ depends not only on n but also on m. In such a situation, the approximation in (12) is no longer valid, which considerably degrades the performance of the classical BIC method in (13), particularly when the number of snapshots is comparable with the number of antennas. To circumvent this problem, we recalculate the LLF and the PF for $m, n \to \infty$ and $m/n \to c$, ending up with a new BIC variant that is able to provide reliable detection of the source number in the large array.

B. Proposed BIC Variant

To correctly compute the *a posteriori* probability for source enumeration in the general asymptotic regime, where $m, n \to \infty$ with $m/n \to c$, we first need to determine the ML estimate of the parameter vector $\boldsymbol{\theta}^{(k)}$. It is shown in Appendix A that the ML estimate of $\boldsymbol{\theta}^{(k)}$ in the general asymptotic situation turns out to be the same as that in the classical asymptotic case. That is

$$\hat{\boldsymbol{\theta}}^{(k)} = \left[\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k\right]^T \tag{14}$$

is the ML estimate of $\boldsymbol{\theta}^{(k)}$ for $m, n \to \infty$ and $m/n \to c$. Here, $\hat{\tau}_k = (1/(m-k)) \sum_{i=k+1}^{m} \ell_i$. On the other hand, it is indicated in Appendix B that, as $m, n \to \infty$ and $m/n \to c$, the logarithm of the *a posteriori* probability can be computed as

$$\log f(\mathcal{H}_k|\boldsymbol{X}) \approx \log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) + \log f\left(\hat{\boldsymbol{\theta}}^{(k)}\right) + \nu_k \log \pi - \frac{1}{2}\log|\hat{\boldsymbol{J}}|. \quad (15)$$

It is pointed out in [22] that, to determine the asymptotic behavior of the sample eigenvectors, it cannot make any sense to characterize the behavior of the subspace determined by the eigenvectors as their dimension infinitely increases as $m \to \infty$. Instead, it is interesting to determine the behavior of the quadratic function of the eigenprojection matrix. Similarly, it makes little sense to discuss the parameter vector $\hat{\boldsymbol{\theta}}^{(k)}$ alone since its dimension infinitely increases as $m \to \infty$. As a result, we consider the function of $\hat{\boldsymbol{\theta}}^{(k)}$ in (15), which, when maximized, has the effect of maximizing the detection probability for source number detection.

Recall that $f(\hat{\boldsymbol{\theta}}^{(k)})$ stands for the *a priori* pdf of the parameter vector $\hat{\boldsymbol{\theta}}^{(k)}$, which is bounded as $n \to \infty$. In the sequel, $\log f(\hat{\boldsymbol{\theta}}^{(k)})$ is much less than $(1/2)\log|\hat{\boldsymbol{J}}|$ because the latter increases without bound for $m \to \infty$ or $n \to \infty$. On the other hand, $\nu_k \log \pi$ is also much less than $(1/2)\log|\hat{\boldsymbol{J}}|$ as $m, n \to \infty$ and $m/n \to c$. Hence, it follows from (15) that

$$-2\log f(\mathcal{H}_k|\boldsymbol{X}) \approx -2\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}) + \log|\hat{\boldsymbol{J}}| \quad (16)$$

which, when minimized with respect to k, is able to yield a reliable estimate of the source number, provided that $\log f \times (\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)})$ and $\log|\hat{\boldsymbol{J}}|$ can be correctly calculated for $m, n \to \infty$ and $m/n \to c$. Since $\hat{\boldsymbol{\theta}}^{(k)} = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k]^T$ is the ML estimate of $\boldsymbol{\theta}^{(k)}$ for $m, n \to \infty$ and $m/n \to c$, using the similar derivation in [12], the LLF is computed as

$$-2\log f\left(\boldsymbol{X}|\hat{\boldsymbol{\theta}}^{(k)}\right) = 2n(m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m}\ell_i}{\left(\prod_{i=k+1}^{m}\ell_i\right)^{\frac{1}{m-k}}}. \quad (17)$$

On the other hand, it follows from (B.4) that the determinant of $\hat{\boldsymbol{J}}$ is

$$|\hat{\boldsymbol{J}}| = \frac{(m-k)n}{\hat{\tau}_k^2}|\boldsymbol{Q}| \quad (18)$$

where

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{bmatrix} \quad (19)$$

with $\boldsymbol{Q}_{11}$, $\boldsymbol{Q}_{12}$, $\boldsymbol{Q}_{21}$, and $\boldsymbol{Q}_{22}$ being defined in (B.8). Utilizing the formula for the determinant of partitioned matrices, we obtain

$$|\boldsymbol{Q}| = |\boldsymbol{Q}_{11}| \times \left|\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\right|. \quad (20)$$

Substituting (B.8) into (20) yields

$$|\boldsymbol{Q}| = (2n)^{mk} n^k \left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right)^{m-k+2} \hat{\tau}_k^{(m-k)k} \quad (21)$$

TABLE I
SUMMARY OF PROPOSED BIC ALGORITHM

Step 1: Perform eigenvalue decomposition on $\hat{\boldsymbol{R}}$ and obtain $\ell_1, \cdots, \ell_m$.
Step 2: Calculate $a(k), g(k)$ and $\mathcal{P}(k, m, n)$ in (25) by using $\ell_1, \cdots, \ell_m$.
Step 3: Estimate the source number according to (26).

which, when substituted into (18), leads to

$$|\hat{\boldsymbol{J}}| = (m-k)n \cdot (2n)^{mk} \cdot n^k \cdot \left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right)^{m-k+2} \cdot (\hat{\tau}_k)^{k(m-k)-2}. \quad (22)$$

Taking the logarithm of (22), we have

$$\begin{aligned}\log|\hat{\boldsymbol{J}}| &= \log[(m-k)n] + mk\log(2n) + k\log n \\ &\quad + (m-k+2)\log\left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right) + (k(m-k)-2)\log\hat{\tau}_k \\ &= m\Bigg[k\log(2n) + \frac{\log[(m-k)n]}{m} + \frac{k\log n}{m} \\ &\quad + \left(1 - \frac{k-2}{m}\right)\log\left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right) \\ &\quad + \left(k - \frac{k^2+2}{m}\right)\log\hat{\tau}_k\Bigg]. \end{aligned} \quad (23)$$

Recall that $m, n \to \infty$ while the presumed source number k can be taken as a fixed number. Consequently, we obtain $[\log((m-k)n)]/m \to 0$, $(k\log n)/m \to 0$, $(k-2)/m \to 0$, and $(k^2+2)/m \to 0$ for $m, n \to \infty$ and $m/n \to c$. It follows that, as $m, n \to \infty$ and $m/n \to c$, (23) is approximated as

$$\begin{aligned}\log|\hat{\boldsymbol{J}}| &\approx m\left[k\log(2n) + \log\left(\prod_{i=1}^{k}\frac{1}{\ell_i}\right) + k\log\hat{\tau}_k\right] \\ &= mk\left(\log(2n) - \frac{1}{k}\sum_{i=1}^{k}\log\frac{\ell_i}{\hat{\tau}_k}\right) \\ &\triangleq \mathcal{P}(k, m, n). \end{aligned} \quad (24)$$

Therefore, substituting (17) along with (24) into (16), the proposed BIC variant is

$$\mathrm{BIC}(k) = 2n(m-k)\log\frac{a(k)}{g(k)} + \mathcal{P}(k, m, n) \quad (25)$$

where $a(k) = (1/(m-k))\sum_{i=k+1}^{m}\ell_i$ and $g(k) = (\prod_{i=k+1}^{m}\ell_i)^{1/(m-k)}$ are the arithmetic mean and the geometric mean, respectively. The source number is estimated as

$$\hat{d} = \arg\min_{k=0,\ldots,\bar{m}-1} \mathrm{BIC}(k). \quad (26)$$

Recall that $\bar{m} < \min(m, n)$, with $\bar{m}$ being the maximum presumed source number, which is fixed as $m, n \to \infty$ and $m/n \to c$. Since a pair of mutually transposed matrices shares a common set of nonzero eigenvalues up to a nuisance constant multiplication factor [30], the numbers of antennas m and samples n play symmetric roles. This implies that for $m > n$, we can swap m and n when applying the proposed BIC variant on the n nonzero eigenvalues. The proposed BIC algorithm is tabulated in Table I.

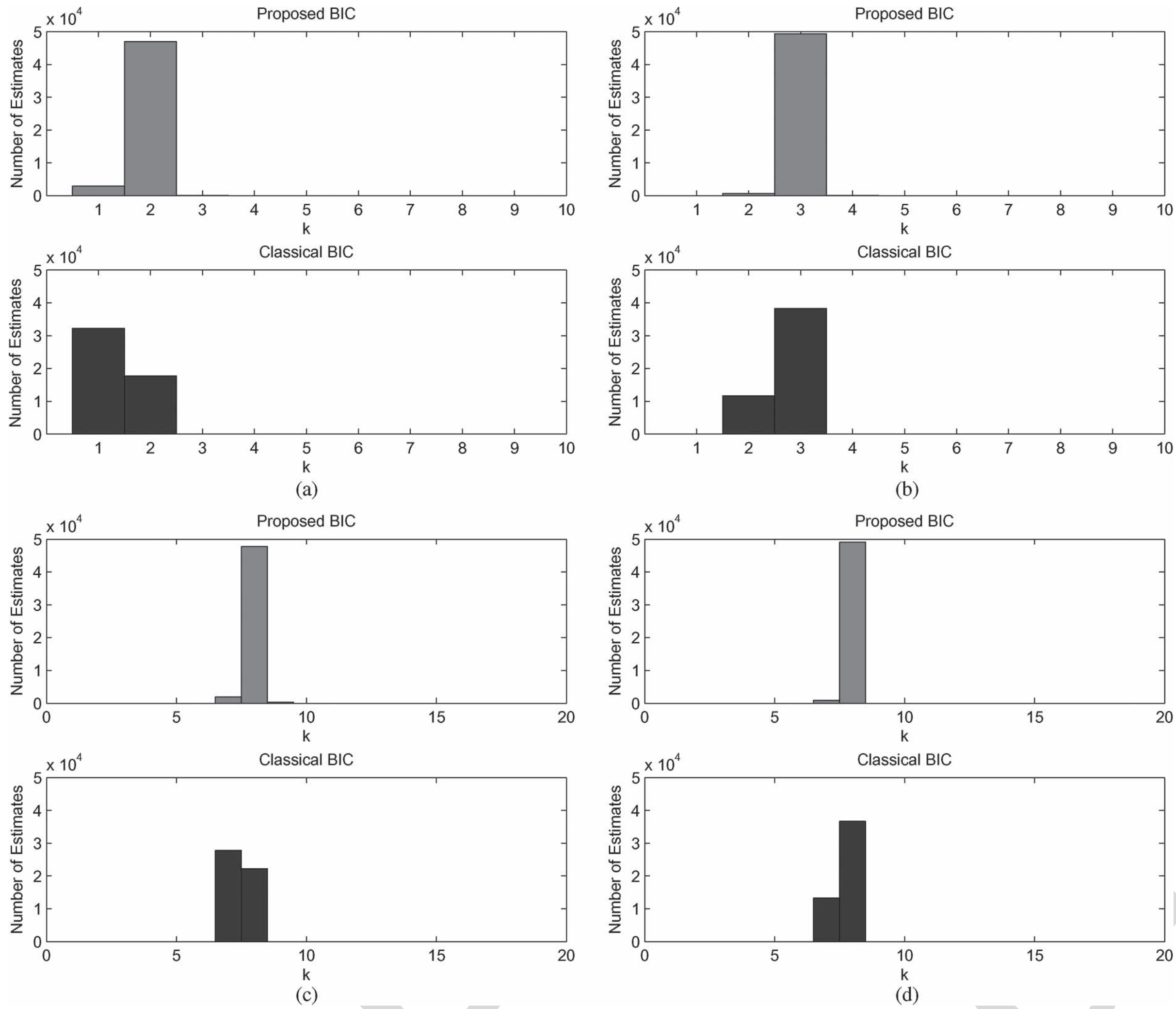


Fig. 1. Histogram plots for source enumeration. $[\varphi_1, \varphi_2] = [2°, 6.5°]$ for Fig. 1(a), $[\varphi_1, \varphi_2, \varphi_3] = [2°, 6.5°, -3°]$ for Fig. 1(b), and $[\varphi_1, \ldots, \varphi_8] = [2.5°, 22°, -4.9°, 12.3°, 7.3°, 16.7°, -9.6°, 26.7°]$ for Fig. 1(c) and (d). The signals are of equal power. (a) $m = 10$, $n = 80$, SNR $= -3$ dB, and $d = 2$. (b) $m = 10$, $n = 80$, SNR $= 5$ dB, and $d = 3$. (c) $m = 20$, $n = 60$, SNR $= -3$ dB, and $d = 8$. (d) $m = 20$, $n = 200$, SNR $= -5$ dB, and $d = 8$.

Remark: Recall that $\ell_1, \ldots, \ell_k$, are the ML estimates of the signal population eigenvalues. In the sequel, $\ell_1/\hat{\tau}_k, \ldots, \ell_k/\hat{\tau}_k$, are relative to the SNR. This, in turn, indicates that $\mathcal{P}(k, m, n)$ depends not only on the number of snapshots n but also on the number of antennas m as well as SNR. That is to say, the PF $\mathcal{P}(k, m, n)$ employs more information than that in the standard BIC [12], [28], leading to accurate computation of the PF, particularly for $m, n \to \infty$ and $m/n \to c$.

IV. Performance Analysis

Here, we derive the analytical formula for the probability of missed detection and prove that the probability of false alarm converges to zero in the general asymptotic regime.

A. Approximate Probabilities of Missed Detection and False Alarm

The statistical analysis for the performance of the classical MDL method has been widely conducted in the literature [31]–[35]. In fact, in this multiple-hypothesis test, there are two error types, namely, the probabilities of underestimating and overestimating the source number. They are also known as the probability of missed detection P_{md} and the probability of false alarm P_{fa}, respectively. P_{md} and P_{fa} for d sources are, respectively, defined as

$$P_{\mathrm{md}} = \mathrm{Prob}(\hat{d} < d | \mathcal{H}_d) \tag{27a}$$
$$P_{\mathrm{fa}} = \mathrm{Prob}(\hat{d} > d | \mathcal{H}_d). \tag{27b}$$

It has been well justified in [31]–[33] by Monte Carlo experiments that the probability of missed detection can be approximated by the probability of underestimating the source number by one, whereas the probability of false alarm can be approximated by the probability of overestimating the source number by one. That is

$$P_{\mathrm{md}} \approx \mathrm{Prob}(\hat{d} = d - 1 | \mathcal{H}_d) \tag{28a}$$
$$P_{\mathrm{fa}} \approx \mathrm{Prob}(\hat{d} = d + 1 | \mathcal{H}_d). \tag{28b}$$

Computer simulation has been carried out to verify the approximations in (28) for the proposed BIC variant in terms of the histogram of the estimated source number. Fig. 1 plots the histogram bars for source enumeration in four representative parameter settings. That is, Fig. 1(a) provides the histogram for $m = 10$, $n = 80$, $d = 2$, and SNR $= -3$ dB; Fig. 1(b) gives the histogram for $m = 10$, $n = 80$, $d = 3$, and SNR $= 5$ dB; Fig. 1(c) shows the histogram for $m = 20$, $n = 60$, $d = 8$, and SNR $= -3$ dB; whereas Fig. 1(d) shows the histogram for $m = 20, n = 200, d = 8$, and SNR $= -5$ dB. Throughout this paper, the SNR is defined as $10\log_{10}(\sigma_{s_i}^2/\tau)$ with $\sigma_{s_i}^2 \triangleq \mathbb{E}[|s_i(t)|^2]$ and $\tau = 1$. It is indicated in Fig. 1 that the proposed BIC variant tends to underestimate the source number, and the probability of underestimating the source number by one dominates. Moreover, compared with the classical BIC scheme, the proposed scheme considerably improves in terms of the probability of missed detection. Furthermore, the probability of false alarm is negligible. Indeed, it is proved in Appendix C that $P_{\rm fa}$ of the proposed BIC variant converges to zeros as $m, n \to \infty$ and $m/n \to c$. This will also be verified by the simulation results in Section V-B. Recall that $\text{Prob}(\hat{d} = d|\mathcal{H}_d) + P_{\rm md} + P_{\rm fa} = 1$. Therefore, it is sufficient to determine $P_{\rm md}$ for the proposed BIC method to evaluate its detection performance in the general asymptotic regime.

B. Analytic Probability of Missed Detection

Noticing that

$$a(d-1) = \frac{m-d}{m-d+1}a(d) + \frac{\ell_d}{m-d+1} \tag{29}$$

$$[g(d-1)]^{m-d+1} = [g(d)]^{m-d} \cdot \ell_d \tag{30}$$

we obtain

$$\begin{aligned}(m-(d-1))\log\frac{a(d-1)}{g(d-1)} &= \log\left(\frac{[a(d)]^{m-d}}{[g(d)]^{m-d}} \times \frac{\left(\frac{m-d}{m-d+1} + \frac{\frac{\ell_d}{a(d)}}{m-d+1}\right)^{m-d+1}}{\frac{\ell_d}{a(d)}}\right)\\ &= (m-d)\log\frac{a(d)}{g(d)} + \log Q_m\left[\frac{\ell_d}{a(d)}\right]\end{aligned} \tag{31}$$

where

$$Q_m\left[\frac{\ell_d}{a(d)}\right] \triangleq \frac{\left[1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right]^p}{\frac{\ell_d}{a(d)}} \tag{32}$$

with $p \triangleq m - d + 1$. Recalling that $\hat{\tau}_{d-1} = a(d-1)$ and $\hat{\tau}_d = a(d)$, it is easy to obtain

$$\begin{aligned}&\mathcal{P}(d,m,n) - \mathcal{P}(d-1,m,n)\\ &= m\log(2n) - m\left(\sum_{i=1}^{d}\log\frac{\ell_i}{\hat{\tau}_d} - \sum_{i=1}^{d-1}\log\frac{\ell_i}{\hat{\tau}_{d-1}}\right)\\ &= m\log(2n) - m\log\frac{\ell_d}{\hat{\tau}_d} - m(d-1)\log\frac{\hat{\tau}_{d-1}}{\hat{\tau}_d}\\ &= m\log(2n) - m\log\frac{\ell_d}{a(d)} - m(d-1)\log\\ &\quad\times\left[1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right].\end{aligned} \tag{33}$$

Therefore, as $m, n \to \infty$ and $m/n \to c$, the probability of missed detection is calculated as

$$\begin{aligned}P_{\rm md} &\approx \text{Prob}\left(\text{BIC}(d-1,m,n) - \text{BIC}(d,m,n) < 1|\mathcal{H}_d\right)\\ &= \text{Prob}\left(\log Q_m\left[\frac{\ell_d}{a(d)}\right] < \frac{m}{2n}\log(2n) - \frac{m}{2n}\log\frac{\ell_d}{a(d)}\right.\\ &\qquad\left. - \frac{m(d-1)}{2n}\log\left(1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right)\middle|\mathcal{H}_d\right)\\ &\approx \text{Prob}\left(\left(p+\frac{c(d-1)}{2}\right)\log\left(1+\frac{1}{p}\left(\frac{\ell_d}{a(d)}-1\right)\right)\right.\\ &\qquad\left. - \left(1-\frac{c}{2}\right)\log\frac{\ell_d}{a(d)} < \frac{c}{2}\log(2n)\middle|\mathcal{H}_d\right)\\ &= \text{Prob}\left(\frac{\ell_d}{a(d)} < f^{-1}(\alpha)\middle|\mathcal{H}_d\right)\end{aligned} \tag{34}$$

where $\alpha = c/2\log(2n)$, and $f^{-1}(z)$ is the inverse function of

$$f(z) = \left(p+\frac{c(d-1)}{2}\right)\log\left(1+\frac{z-1}{p}\right) - \left(1-\frac{c}{2}\right)\log z \tag{35}$$

with $z = \ell_d/a(d)$. Note that the last equality in (34) is due to the fact that $f(z)$ is a monotonic increasing function for $c > 0$ and $z > 1$. The function $f^{-1}(z)$ can be determined by the numerical simulation. Now, we need to determine the distribution of $\ell_d/a(d)$.

It is well known that, as $m, n \to \infty$ with $m/n \to c$, the signal eigenvalues $\lambda_i (i = 1, \ldots, d)$ are probably lower than the critical value $\tau(1+\sqrt{c})$, namely, the so-called asymptotic limit of detection due to the phase transition phenomenon [36]. In such a situation, the signal sample eigenvalue behaves similar to the noise sample eigenvalue. Note that analyzing the detection threshold for the source enumerator is also an interesting topic. It is shown in [8] that the threshold of the RMT detector can be as low as the asymptotic limit of detection when $m, n \to \infty$ with $m/n \to c$. Moreover, it is revealed in [37] that the asymptotic probability of detection for the likelihood ratio test can approach one even when the signal power is substantially lower than the asymptotic limit of detection. Additionally, the consistency of the classical BIC method with respect to the SNR has been investigated in [10] and [15]. However, these topics are beyond the scope of this paper. Consequently, we restrict our attention on the analytic probability of missed detection.

If $\lambda_d > \tau(1+\sqrt{c})$ and λ_d has multiplicity of one, it then follows from [25], [36], and [38] that, as $m, n \to \infty$ with $m/n \to c$, ℓ_d is Gaussian distributed, i.e.,

$$\sqrt{n}\left(\ell_d - \lambda_d\left(1+\frac{\tau c}{\lambda_d - \tau}\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \lambda_d^2\left(1-\frac{c}{(\lambda_d-\tau)^2}\right)\right) \tag{36}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Although this asymptotic result is correct in the general asymptotic regime, it is not accurate enough for finite m and n because it does not consider the interaction between the signals. In fact, it is verified in [22] that, as $m, n \to \infty$ with $m/n \to c$, ℓ_d almost surely (a.s.) converges to its mean derived by Lawley [39] in the classical asymptotic regime, that is

$$\mu_d = \lambda_d\left(1 - \frac{c}{m}\sum_{1\le i\ne d\le m}\frac{\lambda_i}{\lambda_i - \lambda_d}\right). \tag{37}$$

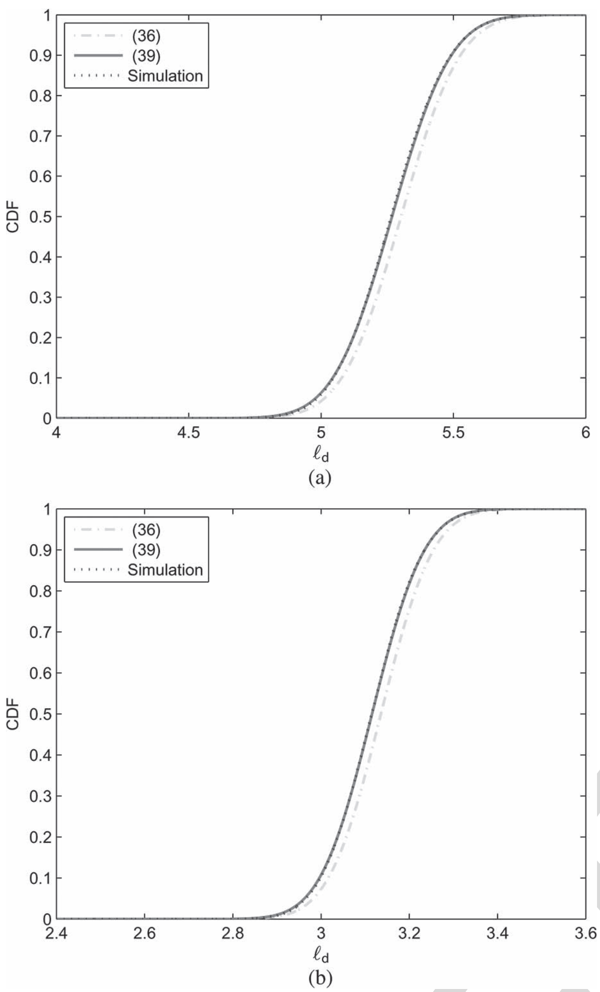


Fig. 2. CDF: Asymptotic distribution versus simulation for three sources with power values of $[-12, -17, -15]$ dB and DOA $= [-2.5°, 3.3°, 12°]$. 10^5 trials. (a) $m = 200$ and $n = 800$. (b) $m = 100$ and $n = 1000$.

As a matter of fact, (37) can be rewritten as

$$\mu_d = \lambda_d \left(1 - \frac{c}{m}\sum_{i=1}^{d-1}\frac{\lambda_i}{\lambda_i - \lambda_d} + \frac{m-d}{m}\frac{\tau c}{\lambda_d - \tau}\right) \tag{38a}$$

$$\xrightarrow{m\to\infty} \lambda_d\left(1 + \frac{\tau c}{\lambda_d - \tau}\right). \tag{38b}$$

Note that the second term within the brackets of (38a) stands for the interaction between the signals. It is implied in (38) that, although the mean of ℓ_d in (37) is the same as that in (36) as $m, n \to \infty$ with $m/n \to c$, the former is more accurate than the latter for finite m and n because it takes into account the interaction between the signals. As a consequence, the fluctuation of ℓ_d is

$$\ell_d \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mu_d, \sigma_d^2\right) \tag{39}$$

with $\sigma_d^2 = (\lambda_d^2/n)(1 - c(\lambda_d - \tau)^{-2})$. To quantitatively show the approximation accuracy between (36) and (39), we calculate their cumulative distribution functions (cdfs) and compare them with the exact distribution of ℓ_d resulted from 10^5 independent simulation trials. The results shown in Fig. 2 indicate that the asymptotic distribution in (39) is more accurate than that in (36). On the other hand, by taking into account the bias resulting from ℓ_i $(i = 1, \ldots, d)$, it follows from [31] and [35] that $a(d) \approx \varpi_d$ with

$$\varpi_d = \tau - \frac{1}{n(m-d)}\sum_{i=1}^{d}\sum_{1\le j\ne i\le m}\frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)}. \tag{40}$$

Thus, the fluctuation of z is

$$z \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mu_z, \sigma_z^2\right) \tag{41}$$

where $\mu_z = \mu_d/\varpi_d$, and $\sigma_z^2 = \sigma_d^2/\varpi_d^2$. The analytic probability of missed detection is

$$P_{\mathrm{md}} = 1 - Q\left(\frac{f^{-1}(\alpha) - \mu_z}{\sigma_z}\right) \tag{42}$$

for $\lambda_d > \tau(1+\sqrt{c})$, where $Q(x) = \int_x^\infty (1/\sqrt{2\pi})e^{-t^2/2}dt$. For $\lambda_d \le \tau(1+\sqrt{c})$, however, the signal cannot be reliably detected due to the phase transition phenomenon. In the sequel, we have $P_{\mathrm{md}} = 1$.

V. Simulation Results

A. Detection Performance

The detection performance of the proposed BIC variant is evaluated by computer simulation in this section. For the purpose of comparison, the empirical results of the representative ITCs are also presented, that is, the BIC [13], [28], linear-shrinkage-based MDL (LS-MDL) [40], EEF [41], RMT-AIC [25], and BN-AIC [42]. According to [42], the user-defined parameter C in the BN-AIC scheme is set to 2. Similar to the setting in the last section, we consider a uniform linear array with half-wavelength element separation receiving the narrowband and equal-power stationary Gaussian signals.

The empirical probabilities of correct detection versus SNR for a relatively small sample size of $n = 60$ are plotted in Fig. 3(a), where the number of antennas is 15. We observe that the proposed BIC variant is superior to the other ITCs in terms of detection probability. When the number of snapshots is larger, e.g., $n = 150$, the gaps between the proposed BIC variant and existing ITCs become narrower, as demonstrated in Fig. 3(b). In such a large sample case, the proposed method is comparable with the EEF scheme and still outperforms the LS-MDL and RMT-AIC approaches by around 0.5 dB. Moreover, the proposed detector significantly improves compared with the standard BIC scheme. To study the behavior of the BIC variant for different angle separations, the empirical probabilities of correct detection versus the angle separation are shown in Fig. 4 for the small and large sample sizes, respectively. Here, the DOAs due to the three incident signals are set as $[0, \Delta\varphi, 2\Delta\varphi]$, and the number of antennas is 15. It is seen that the BIC variant is more accurate than the existing ITCs in source enumeration, particularly for the small sample case. It is easy to interpret the improvement of the proposed BIC variant by recalling that the standard BIC suffers from its heavy penalty term. That is, its probability of underestimating the source number dominates. As the proposed BIC offers a smaller penalty term than the standard BIC, it is able to reduce the possibility of underfitting, eventually leading to the significant enhancement in detection performance.

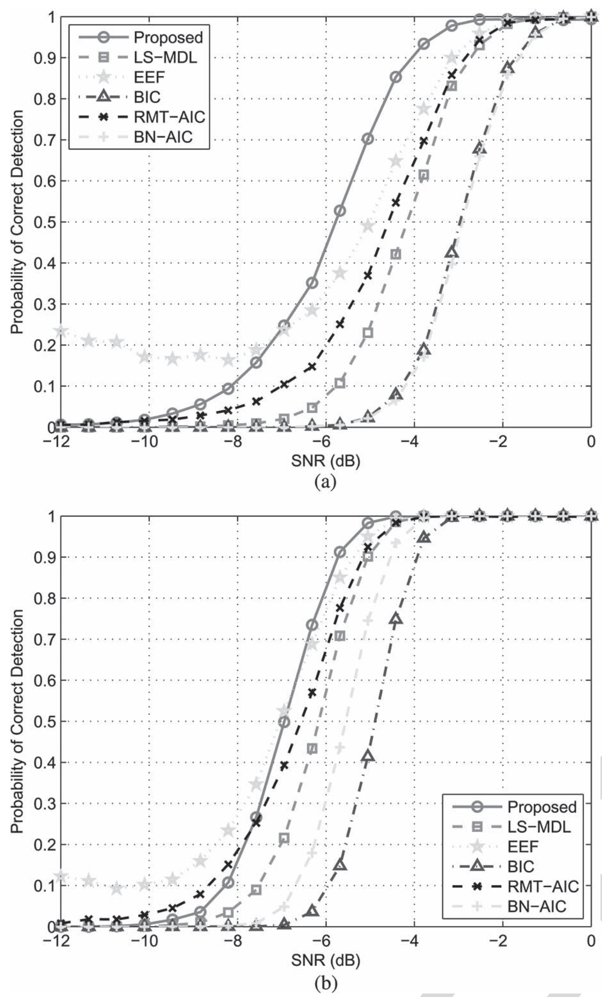


Fig. 3. Probability of correct detection versus SNR. $m = 15$, $d = 3$, $[\varphi_1, \varphi_2, \varphi_3] = [2.3^\circ, 7.5^\circ, 12^\circ]$, and 2×10^3 trials. (a) $n = 60$. (b) $n = 150$.

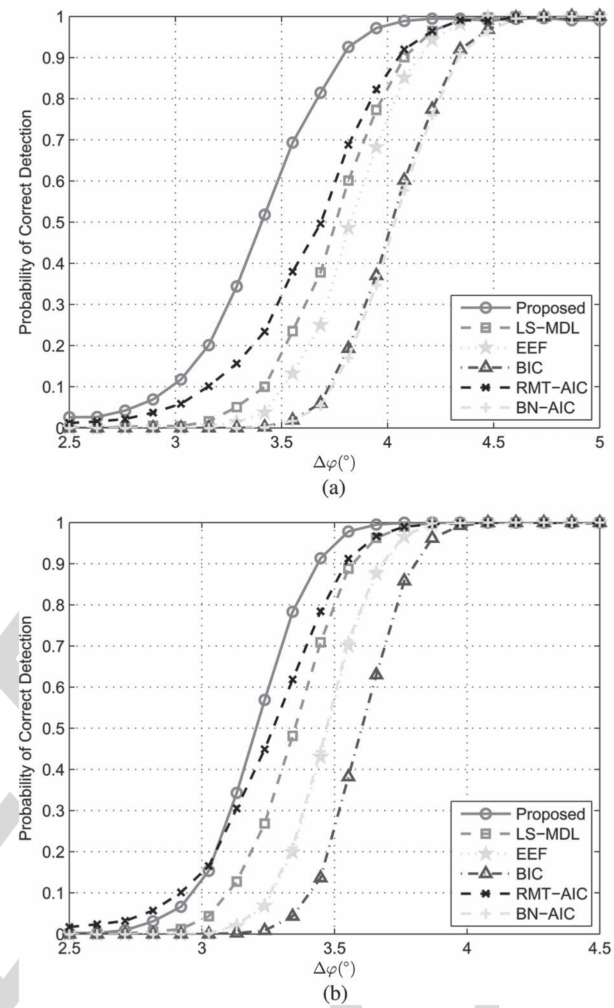


Fig. 4. Probability of correct detection versus angle separation. $m = 15$, $\mathrm{SNR} = 0$ dB, $d = 3$, $[\varphi_1, \varphi_2, \varphi_3] = [0, \Delta\varphi, 2\Delta\varphi]$, and 2×10^3 trials. (a) $n = 60$. (b) $n = 150$.

To investigate the general asymptotic case, we enable both m and n to increase at the same speed, e.g., $m/n = 1/3$ and $m/n = 0.5$. Since the number of antennas and the number of snapshots can infinitely increase while the source number remains unchanged, we set $\bar{m} = \min(20, m)$ for all algorithms in the following simulations, where m and n increase at the same rate $c = m/n$. The empirical results shown in Fig. 5(a) indicate that the EEF and LS-MDL schemes are more accurate than the standard BIC detector while all of them are able to yield the consistent estimate of source number when the number of snapshots becomes large enough. Compared with the existing ITCs, the proposed BIC approach is capable of yielding more accurate estimate of source number. When $m/n = 0.5$ and $d = 8$, the proposed scheme converges to one in probability of correct detection much faster than the other ITCs, as indicated in Fig. 5(b).

To confirm that the probability of false alarm of the proposed BIC approach tends to zero as $m, n \to \infty$ and $m/n \to c$, the empirical probability of false alarm versus the number of antennas is shown in Fig. 6, where $m/n = 0.5$, and $\mathrm{SNR} = -8$ dB. For comparison, the empirical results of the EEF and RMT-AIC approaches are presented as well. Fig. 6(a) corresponds to the empirical results for three incident signals with DOAs of $[2.3^\circ, 7.5^\circ, 12^\circ]$, whereas Fig. 6(b) shows the empirical results for eight incident signals with DOAs of $[\varphi_1, \ldots, \varphi_8] = [2.5^\circ, 22^\circ, -4.9^\circ, 12.3^\circ, 7.3^\circ, 16.7^\circ, -9.6^\circ, 26.7^\circ]$. It is seen that the ITCs offer different probabilities of false alarm. On the other hand, Fig. 6 implies that the probability of false alarm of the proposed BIC algorithm converges to zero as $m, n \to \infty$ and $m/n \to c$, which is in line with the theoretical analysis in Section IV-A.

To fairly compare the ITCs with the threshold-like testing methods, we need to set their probabilities of false alarm at the same level. Nevertheless, as indicated in Fig. 6, the ITCs implicitly offer different probabilities of false alarm. As a result, the threshold-like testing method should be compared with one of the ITCs at the same probability of false alarm. In the end, the empirical results of the proposed BIC variant and RMT approach [8] are plotted in Fig. 7. Here, the probability of false alarm of the RMT algorithm is equal to that of the proposed BIC approach. Moreover, to enable the RMT method to properly work, we set its probability of false alarm to 10^{-6} when the probability of false alarm of the proposed BIC variant is equal to zero. In addition, note that the minimax [43] approach is also a threshold-like algorithm, but it directly links the noise sample eigenvalue distribution, namely, the Tracy–Widom law [44], to the signal sample eigenvalue distribution, ending up with an

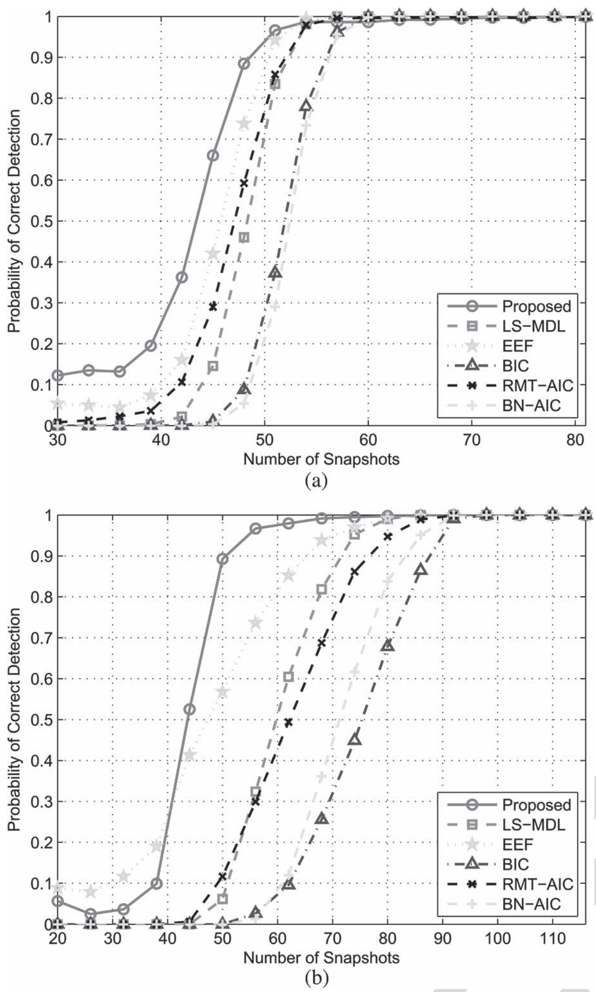


Fig. 5. Probability of correct detection versus number of snapshots. 2×10^3 trials. (a) $d=3$, $m/n=1/3$, and $[\varphi_1, \varphi_2, \varphi_3]=[2.3°, 7.5°, 12°]$. (b) $d=8$, $m/n=0.5$, and $[\varphi_1, \ldots, \varphi_8]=[2.5°, 22°, -4.9°, 12.3°, 7.3°, 16.7°, -9.6°, 26.7°]$.

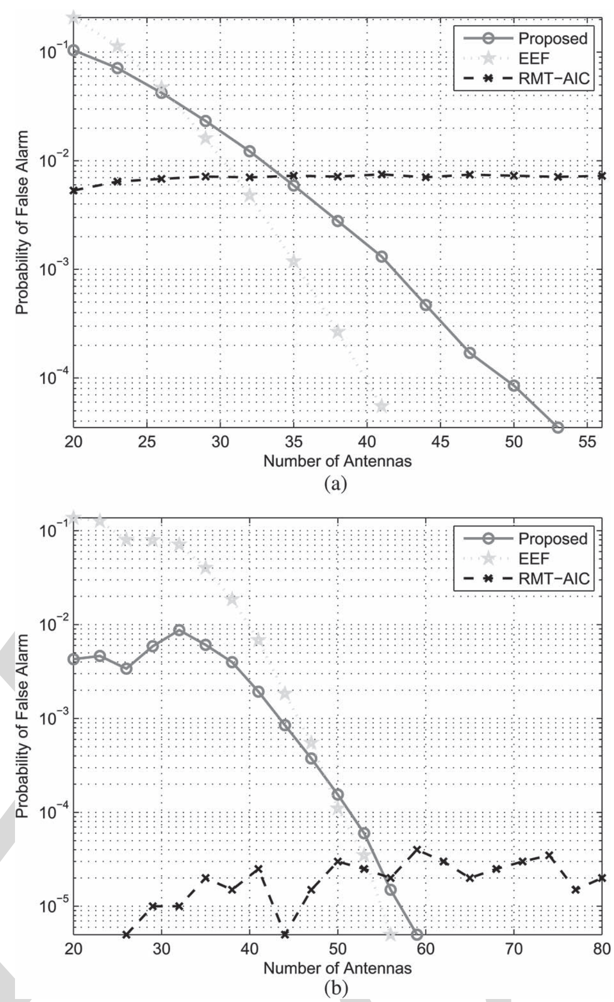


Fig. 6. Probability of false alarm versus antenna number for the proposed, EEF, and RMT-AIC approaches. SNR $= -8$ dB, $m/n = 0.5$, and 2×10^5 trials. (a) $[\varphi_1, \varphi_2, \varphi_3] = [2.3°, 7.5°, 12°]$. (b) $[\varphi_1, \ldots, \varphi_8] = [2.5°, 22°, -4.9°, 12.3°, 7.3°, 16.7°, -9.6°, 26.7°]$.

elegant method for threshold calculation. See [43, eq. (10)] for the details. As a result, the empirical results for the minimax scheme are presented as well. Similar to [43], the "inclusion" penalty and the "exclusion" penalty are set as $c_{\mathrm{I}} = c_{\mathrm{E}}(1) = \cdots = c_{\mathrm{E}}(m)$ with $\lambda_0 = \sqrt{c} + n^{-1/3}$ and $\hat{\tau}_k = (m - k)^{-1} \sum_{i=k+1}^{m}$ for the kth threshold calculation. It is indicated in Fig. 7 that the proposed BIC variant is superior to the RMT scheme in detection accuracy and outperforms the minimax approach in consistency. Although the proposed BIC variant might not be as accurate as the minimax method, as shown in Fig. 7(b), it is able to attain correct detection probability of one in these two cases. On the other hand, it should be noted that the minimax algorithm relies on the Tracy–Widom distribution, which cannot be evaluated online, incurring more overhead in the procedure of detection.

B. Accuracy of Analytic Probability of Missed Detection

Here, numerical results are presented to evaluate the accuracy of the analytic probability of missed detection, which is derived in Section IV-B. To evaluate the accuracy of the analytic probability of missed detection for large arrays and large samples in large-array applications, we set the number of antennas as $m = 50$ and vary the number of samples from $n = 300$ to $n = 1000$. Fig. 8 indicates that the analytic probability of missed detection is very close to the simulated probability of missed detection. This, in turn, implies that our derived analytic probability of missed detection is able to accurately predict the detection performance.

VI. Conclusion

This paper has devised a new BIC variant for source enumeration in the general asymptotic regime, which enables us to correctly determine the signal and noise subspaces for the subsequent DOA estimation and beamforming in large-array systems. As the existing information criteria only consider the condition when the number of antennas remains unchanged while the number of snapshots tends to infinity, they cannot provide accurate detection of the source number for the large array. By correctly determining the Hessian matrix in the calculation of the PF, we have derived an efficient BIC variant for the general asymptotic regime. Moreover, a closed-form formula has been derived for calculating the probability of missed detection, and the probability of false alarm has been proved to converge

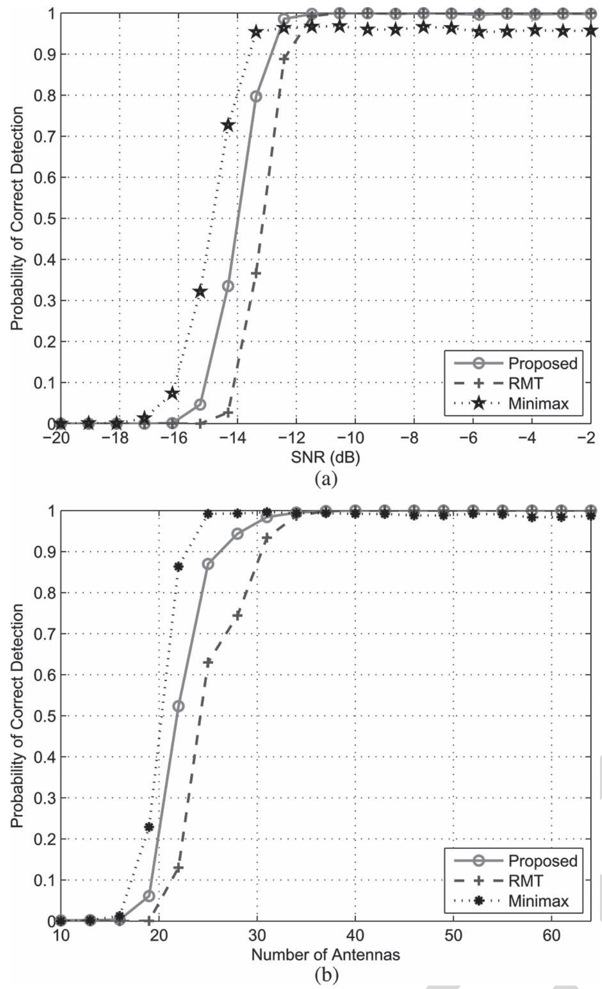


Fig. 7. Probability of correct detection for the proposed BIC, RMT, and minimax algorithms at the same probability of false alarm. $d = 3$, $[\varphi_1, \varphi_2, \varphi_3] = [2.3°, 7.5°, 12°]$, and 2×10^3 trials. (a) $m = 50$ and $n = 80$. (b) SNR $= -12$ dB and $m/n = 0.2$.

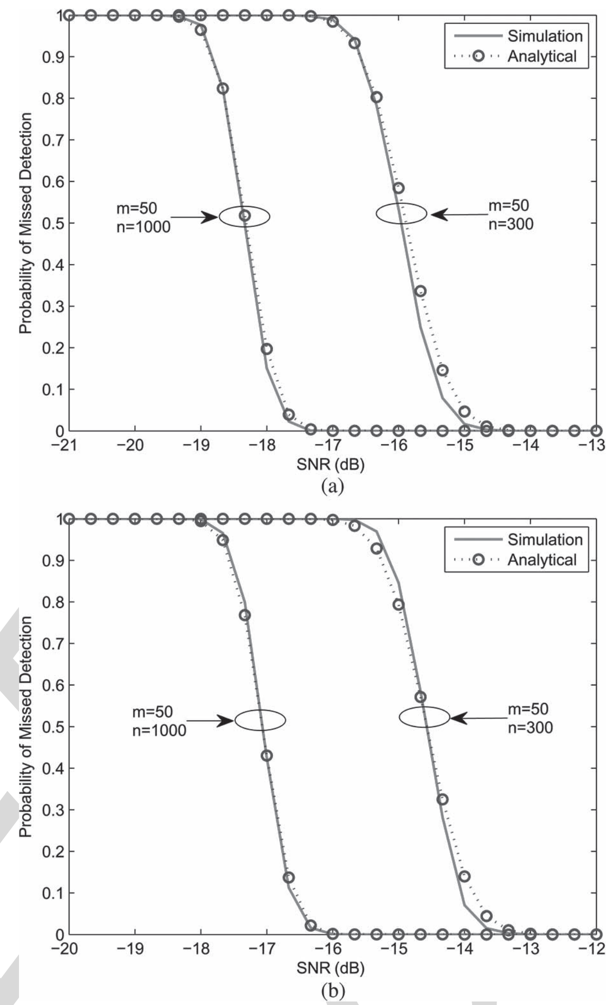


Fig. 8. Probability of missed detection versus SNR. $m = 50$ and 2×10^3 trials. (a) $d = 3$ and $[\varphi_1, \varphi_2, \varphi_3] = [-2.5°, 3.3°, 12°]$. (b) $d = 8$ and $[\varphi_1, \ldots, \varphi_8] = [-2.5°, 4.3°, 2.3°, -12°, 8.1°, 16.8°, -23.7°, 32.1°]$.

to zero as $m, n \to \infty$ and $m/n \to c$. Simulation results have verified the superiority of the proposed BIC approach over its existing counterparts and confirmed the statistical performance analysis.

Appendix A
Maximum Estimation of $\boldsymbol{\theta}^{(k)}$ in the General Asymptotic Case

For the case of k sources, let $\boldsymbol{R}^{(k)} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ and $\hat{\boldsymbol{R}} = \boldsymbol{E}\boldsymbol{L}\boldsymbol{E}^H$ be the eigenvalue decompositions of $\boldsymbol{R}^{(k)}$ and $\hat{\boldsymbol{R}}$, respectively. Here, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k, \tau, \ldots, \tau)$, $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$, $\boldsymbol{L} = \mathrm{diag}(\ell_1, \ldots, \ell_m)$, and $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m]$, with $\boldsymbol{u}_i$ and $\boldsymbol{e}_i$, $i = 1, \ldots, m$ being the population and sample eigenvectors corresponding to the population and sample eigenvalues λ_i and ℓ_i, respectively. As a result, the LLF is calculated as

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\theta}^{(k)}\right) &\triangleq -n \log \left|\boldsymbol{R}^{(k)}\right| - n\mathrm{tr}\left[\left(\boldsymbol{R}^{(k)}\right)^{-1} \hat{\boldsymbol{R}}\right] - mn \log \pi \\ &= -n \left(\sum_{i=1}^{k} \log \lambda_i + (m-k) \log \tau\right) \\ &\quad - n\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{G}^H\boldsymbol{L}\boldsymbol{G}) - mn \log \pi \qquad \text{(A.1)}\end{aligned}$$

where $\mathrm{tr}[\cdot]$ denotes the trace operation, and $\boldsymbol{G} = \boldsymbol{E}^H\boldsymbol{U}$. Since $\boldsymbol{G}$ is orthogonal, we have the following inequality [45], [46]:

$$\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{G}^H\boldsymbol{L}\boldsymbol{G}) \geq \sum_{i=1}^{m} \frac{\ell_i}{\lambda_i}. \qquad \text{(A.2)}$$

This equality in (A.2) holds for $\boldsymbol{G} = \boldsymbol{I}_m$ [46], i.e., $\boldsymbol{U} = \boldsymbol{E}$. Consequently, it follows from (A.1) and (A.2) that $\boldsymbol{e}_i$, $i = 1, \ldots, m$, is the ML estimate of $\boldsymbol{u}_i$. That is, $\hat{\boldsymbol{u}}_i = \boldsymbol{e}_i$ for $i = 1, \ldots, k$. Substituting these ML estimates into (A.1), we obtain the LLF relying on the reduced parameter vector $\boldsymbol{\vartheta}^{(k)} = [\lambda_1, \ldots, \lambda_k, \tau]$, i.e.,

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\vartheta}^{(k)}\right) &= -n \left(\sum_{i=1}^{k} \log \lambda_i + (m-k) \log \tau\right) \\ &\quad - n \left(\sum_{i=1}^{k} \frac{\ell_i}{\lambda_i} + \frac{\sum_{i=k+1}^{m} \ell_i}{\tau}\right) - mn \log \pi. \qquad \text{(A.3)}\end{aligned}$$

Maximizing $\mathcal{L}(\boldsymbol{\vartheta}^{(k)})$ with respect to $\boldsymbol{\vartheta}^{(k)}$ yields the ML estimates of $\lambda_1, \ldots, \lambda_k, \tau$, which are given as

$$\hat{\lambda}_i = \ell_i, \ i = 1, \ldots, k \qquad \text{(A.4)}$$

$$\hat{\tau}_k = \frac{1}{m-k} \sum_{i=k+1}^{m} \ell_i. \qquad \text{(A.5)}$$

Thus, the ML estimate of $\boldsymbol{\theta}^{(k)}$ is $\hat{\boldsymbol{\theta}}^{(k)} = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k]^T$ for the general asymptotic case, which is the same as that in the classical asymptotic situation.

APPENDIX B
DERIVATION OF (15)

It follows from [47, eq. (92)] that the Taylor series expansion of $\log f(\boldsymbol{X}|\boldsymbol{\theta}^{(k)})$ around $\hat{\boldsymbol{\theta}}^{(k)}$ is given in (B.1). Here, $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{J}}$ is the Hessian matrix defined in (10), and the superscript $(\cdot)^{(k)}$ has been dropped for simplicity. We are now at a position to prove that the zero-order term is much larger than the second-order term in (B.1) as $m, n \to \infty$ and $m/n \to c$

$$\begin{aligned}\log f(\boldsymbol{X}|\boldsymbol{\theta}) &= \log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^H \underbrace{\left.\frac{\partial \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}}_{=0} \\ &\quad + \frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^H\left[\left.\frac{\partial^2 \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^H}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + \cdots \\ &= \log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) - \frac{1}{2}\Delta\boldsymbol{\theta}^H \hat{\boldsymbol{J}} \Delta\boldsymbol{\theta} + \cdots. \end{aligned} \quad \text{(B.1)}$$

Recall that $\hat{\boldsymbol{\theta}} = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_k^T, \ell_1, \ldots, \ell_k, \hat{\tau}_k]^T$ is the ML estimate of $\boldsymbol{\theta}$ in the general asymptotic regime. Exploiting the similar computation in [12], the zero-order term of (B.1) is

$$\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) = -n(m-k)\log \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} \ell_i}{\left(\prod_{i=k+1}^{m} \ell_i\right)^{\frac{1}{m-k}}}. \quad \text{(B.2)}$$

To determine the second-order term of (B.1), we need to calculate the second-order partial derivative of $-\log f(\boldsymbol{X}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, which is provided in (B.3), shown at the bottom of the page. Accordingly, the Hessian matrix is calculated as (B.4), shown at the bottom of the page. To proceed, the following results are needed. If $\lambda_i > \tau(1+\sqrt{c})(i=1,\ldots,k)$ and λ_i has multiplicity 1, as $m, n \to \infty$ with $m/n \to c$, it follows from [38] that

$$\ell_i = \lambda_i + \frac{\lambda_i \tau}{\lambda_i - \tau}c + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \quad \text{(B.5a)}$$

$$\hat{\tau}_k = \tau + \mathcal{O}\left(\frac{1}{n}\right). \quad \text{(B.5b)}$$

On the other hand, under the same conditions and as $m, n \to \infty$ with $m/n \to c$, it is indicated in [38], [48], and [49] that the inner product of the largest sample and population eigenvectors converges almost surely to a deterministic value, which is given as

$$\boldsymbol{u}_i^H \boldsymbol{e}_i \xrightarrow{\text{a.s.}} \frac{1 - \frac{c\tau^2}{(\lambda_i - \tau)^2}}{1 + \frac{c\tau}{\lambda_i - \tau}}, \quad (i = 1, \ldots, k). \quad \text{(B.5c)}$$

Consequently, setting $\Delta\boldsymbol{\theta} \triangleq [\boldsymbol{\epsilon}^T, \boldsymbol{\nu}^T, \varepsilon]^T$, where

$$\boldsymbol{\epsilon} = \left[(\boldsymbol{u}_1 - \boldsymbol{e}_1)^T, \ldots, (\boldsymbol{u}_k - \boldsymbol{e}_k)^T\right]^T \quad \text{(B.6a)}$$

$$\boldsymbol{\nu} = -\tau c\left[\frac{\lambda_1}{\lambda_1 - \tau}, \ldots, \frac{\lambda_k}{\lambda_k - \tau}\right]^T + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \quad \text{(B.6b)}$$

$$\varepsilon = \mathcal{O}\left(\frac{1}{n}\right) \quad \text{(B.6c)}$$

$$-\frac{\partial^2 \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^H} = \left[\begin{array}{ccc|ccc|c} \frac{2n}{\lambda_1}\hat{\boldsymbol{R}} & & & \frac{-2n\hat{\boldsymbol{R}}\boldsymbol{v}_1}{\lambda_1^2} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{2n}{\lambda_k}\hat{\boldsymbol{R}} & & & \frac{-2n\hat{\boldsymbol{R}}\boldsymbol{v}_k}{\lambda_k^2} & 0 \\ \hline \frac{-2n\boldsymbol{v}_1^H\hat{\boldsymbol{R}}}{\lambda_1^2} & & & \frac{2n\boldsymbol{v}_1^H\hat{\boldsymbol{R}}\boldsymbol{v}_1}{\lambda_1^3} - \frac{n}{\lambda_1^2} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{-2n\boldsymbol{v}_k^H\hat{\boldsymbol{R}}}{\lambda_k^2} & & & \frac{2n\boldsymbol{v}_k^H\hat{\boldsymbol{R}}\boldsymbol{v}_k}{\lambda_k^3} - \frac{n}{\lambda_k^2} & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & \begin{array}{c}\frac{2n\sum_{i=k+1}^{m}\boldsymbol{v}_i^H\hat{\boldsymbol{R}}\boldsymbol{v}_i}{\tau^3} \\ -\frac{(m-k)n}{\tau^2}\end{array} \end{array}\right] \quad \text{(B.3)}$$

$$\hat{\boldsymbol{J}} = -\left.\frac{\partial^2 \log f(\boldsymbol{X}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^H}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left[\begin{array}{ccc|ccc|c} \frac{2n}{\ell_1}\hat{\boldsymbol{R}} & & & \frac{-2n\boldsymbol{e}_1}{\ell_1} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{2n}{\ell_k}\hat{\boldsymbol{R}} & & & \frac{-2n\boldsymbol{e}_k}{\ell_k} & 0 \\ \hline \frac{-2n\boldsymbol{e}_1^H}{\ell_1} & & & \frac{n}{\ell_1^2} & & & 0 \\ & \ddots & & & \ddots & & \vdots \\ & & \frac{-2n\boldsymbol{e}_k^H}{\ell_k} & & & \frac{n}{\ell_k^2} & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{(m-k)n}{\hat{\tau}_k^2} \end{array}\right] \quad \text{(B.4)}$$

the second-order term in (B.1) can be expressed as

$$\begin{aligned}\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta} &= \frac{1}{2}[\boldsymbol{\epsilon}^H,\boldsymbol{\nu}^H,\varepsilon]\left[\begin{array}{c|c|c}\boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} & \mathbf{0}\\ \hline \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} & \vdots \\ \hline \mathbf{0} & \cdots & \beta\end{array}\right]\begin{bmatrix}\boldsymbol{\epsilon}\\ \boldsymbol{\nu}\\ \varepsilon\end{bmatrix}\\ &= \frac{1}{2}(\boldsymbol{\epsilon}^H\boldsymbol{Q}_{11}\boldsymbol{\epsilon}+\boldsymbol{\epsilon}^H\boldsymbol{Q}_{12}\boldsymbol{\nu}+\boldsymbol{\nu}^H\boldsymbol{Q}_{21}\boldsymbol{\epsilon}+\boldsymbol{\nu}^H\boldsymbol{Q}_{22}\boldsymbol{\nu}+\varepsilon^2\beta) \quad \text{(B.7)}\end{aligned}$$

where

$$\boldsymbol{Q}_{11} = \text{blkdiag}\left(\frac{2n}{\ell_1}\hat{\boldsymbol{R}},\ldots,\frac{2n}{\ell_k}\hat{\boldsymbol{R}}\right) \quad \text{(B.8a)}$$

$$\boldsymbol{Q}_{12} = \text{blkdiag}\left(\frac{-2n\boldsymbol{e}_1}{\ell_1},\ldots,\frac{-2n\boldsymbol{e}_k}{\ell_k}\right) \quad \text{(B.8b)}$$

$$\boldsymbol{Q}_{22} = \text{diag}\left(\frac{n}{\ell_1^2},\ldots,\frac{n}{\ell_k^2}\right) \quad \text{(B.8c)}$$

$$\beta = \frac{n(m-k)}{\hat{\tau}^2} \quad \text{(B.8d)}$$

and $\boldsymbol{Q}_{21} = \boldsymbol{Q}_{12}^H$. Here, $\text{blkdiag}(\cdot)$ denotes the block diagonal matrix. Substituting (B.6) and (B.8) into (B.7), we can calculate $\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}$. In particular, notice that

$$\boldsymbol{\epsilon}^H\boldsymbol{Q}_{11}\boldsymbol{\epsilon} = \sum_{i=1}^{k}\frac{2n}{\ell_i}\boldsymbol{\epsilon}_i^H\hat{\boldsymbol{R}}\boldsymbol{\epsilon}_i \quad \text{(B.9)}$$

with $\boldsymbol{\epsilon}_i=\boldsymbol{u}_i-\boldsymbol{e}_i$ and $\boldsymbol{\epsilon}_i^H\hat{\boldsymbol{R}}\boldsymbol{\epsilon}_i=\boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i-\boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i-\boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i+\boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i$. Since $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m$ and $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m$ span the same observation space, we assert that, for $\boldsymbol{u}_i (i=1,\ldots,m)$, there is a nonzero set $\{\alpha_{i1},\ldots,\alpha_{im}\}$, such that

$$\boldsymbol{u}_i = \alpha_{i1}\boldsymbol{e}_1+\cdots+\alpha_{im}\boldsymbol{e}_m \quad \text{(B.10)}$$

which implies that

$$\boldsymbol{u}_i^H\boldsymbol{u}_i = |\alpha_{i1}|^2+\cdots+|\alpha_{im}|^2 = 1 \quad \text{(B.11)}$$

where $|\alpha_{ij}|$ denotes the absolute value of α_{ij}. It is easy to obtain

$$\boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i = |\alpha_{i1}|^2\ell_1+\cdots+|\alpha_{im}|^2\ell_m \quad \text{(B.12a)}$$

$$\boldsymbol{u}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i = \boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{u}_i = \alpha_{ii}\ell_i \quad \text{(B.12b)}$$

$$\boldsymbol{e}_i^H\hat{\boldsymbol{R}}\boldsymbol{e}_i = \ell_i. \quad \text{(B.12c)}$$

Therefore, substituting (B.12) into (B.9) yields

$$\boldsymbol{\epsilon}^H\boldsymbol{Q}_{11}\boldsymbol{\epsilon} = n\sum_{i=1}^{k}\left(2\sum_{j=1}^{m}|\alpha_{ij}|^2\frac{\ell_j}{\ell_i}-4\alpha_{ii}+2\right). \quad \text{(B.13a)}$$

Moreover, the second and third terms of (B.7) are given as

$$\boldsymbol{\epsilon}^H\boldsymbol{Q}_{12}\boldsymbol{\nu}=2n\sum_{i=1}^{k}\frac{(\alpha_{ii}-1)\lambda_i\tau c}{\ell_i(\lambda_i-\tau)}+\mathcal{O}(\sqrt{n})=\boldsymbol{\nu}^H\boldsymbol{Q}_{21}\boldsymbol{\epsilon}. \quad \text{(B.13b)}$$

In addition, it is easy to calculate the last two terms of (B.7) as

$$\boldsymbol{\nu}^H\boldsymbol{Q}_{22}\boldsymbol{\nu} = n\sum_{i=1}^{k}\frac{(\lambda_i\tau c)^2}{\ell_i^2(\lambda_i-\tau)^2}-\mathcal{O}(\sqrt{n}) \quad \text{(B.13c)}$$

$$\varepsilon^2\beta = (m-k)\mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}(1). \quad \text{(B.13d)}$$

Consequently, substituting (B.13) into (B.7), we attain

$$\begin{aligned}\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}=n\sum_{i=1}^{k}\Bigg(&\sum_{j=1}^{m}|\alpha_{ij}|^2\frac{\ell_j}{\ell_i}+\frac{2(\alpha_{ii}-1)\lambda_i\tau c}{\ell_i(\lambda_i-\tau)}\\ &+\frac{(\lambda_i\tau c)^2}{2\ell_i^2(\lambda_i-\tau)^2}-2\alpha_{ii}+1\Bigg)+\mathcal{O}(\sqrt{n}). \quad \text{(B.14)}\end{aligned}$$

Utilizing $\ell_1\geq\cdots\geq\ell_m$ and $\alpha_{ii}=\boldsymbol{u}_i^H\boldsymbol{e}_i\in[0,1]$, we assert

$$\begin{aligned}\frac{\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}}{2mn} &\leq \frac{1}{m}\sum_{i=1}^{k}\left(\frac{\ell_1}{\ell_i}+1+\frac{(\lambda_i\tau c)^2}{2\ell_i^2(\lambda_i-\tau)^2}\right)+\mathcal{O}\left(\frac{1}{m\sqrt{n}}\right)\\ &\xrightarrow{m,n\to\infty,m/n\to c} 0. \quad \text{(B.15)}\end{aligned}$$

However, it follows from (B.2) that $(1/nm)\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})$ is bounded as $m,n\to\infty$ and $m/n\to c$. As a result, as $m,n\to\infty$ and $m/n\to c$, by omitting the high-order terms in (B.1), we have

$$\log f(\boldsymbol{X}|\boldsymbol{\theta}) \approx \log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})-\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}. \quad \text{(B.16)}$$

On the other hand, assuming that the *a priori* pdf of $\boldsymbol{\theta}$ is flat around $\hat{\boldsymbol{\theta}}$, we obtain $f(\boldsymbol{\theta})\approx f(\hat{\boldsymbol{\theta}})$. Substituting this result along with (B.16) into (9a), we get

$$\begin{aligned}f(\mathcal{H}_k|\boldsymbol{X}) &\approx f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})f(\hat{\boldsymbol{\theta}})\int\exp\left(-\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}\right)d\boldsymbol{\theta}\\ &= \frac{\pi^{\nu_k}f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})f(\hat{\boldsymbol{\theta}})}{|\hat{\boldsymbol{J}}|^{\frac{1}{2}}}\\ &\quad\times\underbrace{\int\frac{1}{\pi^{\nu_k}|\hat{\boldsymbol{J}}^{-1}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\Delta\boldsymbol{\theta}^H\hat{\boldsymbol{J}}\Delta\boldsymbol{\theta}\right)d\boldsymbol{\theta}}_{=1}\\ &= \pi^{\nu_k}|\hat{\boldsymbol{J}}|^{\frac{1}{2}}f(\boldsymbol{X}|\hat{\boldsymbol{\theta}})f(\hat{\boldsymbol{\theta}}). \quad \text{(B.17)}\end{aligned}$$

Taking the logarithm of (B.17) eventually leads to (15).

Appendix C
Proof of $P_{\text{fa}}\to 0$ as $m,n\to\infty$, and $m/n\to c$

Similar to (31) and (32), we have

$$(m-d)\log\frac{a(d)}{g(d)}=(m-d-1)\log\frac{a(d+1)}{g(d+1)}+\log Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right] \quad \text{(C.1)}$$

where

$$Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right] = \frac{\left[1+\frac{1}{m-d}\left(\frac{\ell_{d+1}}{a(d+1)}-1\right)\right]^{(m-d)}}{\frac{\ell_{d+1}}{a(d+1)}}. \quad \text{(C.2)}$$

Therefore, noticing that $a(d+1)\approx a(d)\approx\tau$, it follows from (25) and (C.1) that the probability of false alarm is calculated as

$$\begin{aligned}P_{\text{fa}} &\approx \text{Prob}\left(\text{BIC}(d+1)-\text{BIC}(d)<0|\mathcal{H}_d\right)\\ &= \text{Prob}\left(\log Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right]\right.\\ &\qquad\left.>\frac{\mathcal{P}(d+1,m,n)-\mathcal{P}(d,m,n)}{2n}\right|\mathcal{H}_d\Bigg)\end{aligned}$$

$$\approx \text{Prob}\left(\log Q_m\left[\frac{\ell_{d+1}}{a(d+1)}\right] > \frac{m}{2n}\log(2n) - \frac{m}{2n}\log\frac{\ell_{d+1}}{\tau}\middle|\,\mathcal{H}_d\right). \tag{C.3}$$

Substituting (C.2) into (C.3) and using $a(d+1) \approx \tau$ again, we can approximate P_{fa} for the proposed BIC criterion as

$$\begin{aligned}P_{\text{fa}} &\approx \text{Prob}\left(m\log\left(1+\frac{\ell_{d+1}/\tau-1}{m}\right) - \left(1-\frac{c}{2}\right)\log\frac{\ell_{d+1}}{\tau} > \frac{c}{2}\log(2n)\middle|\,\mathcal{H}_d\right)\\ &\approx \text{Prob}\left(\frac{\ell_{d+1}}{\tau} - \left(1-\frac{c}{2}\right)\log\frac{\ell_{d+1}}{\tau} > \frac{c}{2}\log(2n)+1\middle|\,\mathcal{H}_d\right)\\ &= \text{Prob}\left(\frac{\ell_{d+1}}{\tau} > g^{-1}\left(\frac{c}{2}\log(2n)+1\right)\middle|\,\mathcal{H}_d\right)\end{aligned} \tag{C.4}$$

where $g^{-1}(x)$ is the inverse function of $g(x) = x - (1 - c/2)\log(x)$, which is a monotonically increasing function for $x \geq 1$. We now need to determine the distribution of ℓ_{d+1}/τ. As a matter of fact, ℓ_{d+1} has the similar limiting behavior as the largest sample eigenvalue of $\hat{\boldsymbol{R}}$ in the noise-only case [50], which is described by the following lemma due to [51].

Lemma 1: Let ℓ_{max} be the largest sample eigenvalue of $\tilde{\boldsymbol{R}} \in \mathbb{C}^{(m-d)\times(m-d)}$ for the noise-only case. The normalized sample eigenvalue, i.e., ℓ_{max}/τ, is distributed as Tracy–Widom distribution of order 2. That is

$$\frac{\frac{\ell_{max}}{\tau} - \mu_{mn}}{\sigma_{mn}} \xrightarrow{\mathcal{D}} \mathcal{W} \sim F_{TW_2} \tag{C.5}$$

where $\mu_{mn} = (1+\sqrt{c})^2$, $\sigma_{mn} = (1+\sqrt{c})^{4/3}/n\sqrt{c}$, and

$$F_{TW_2}(s) = \exp\left\{-\int_s^\infty (u-s)q^2(u)du\right\} \tag{C.6}$$

with $q(u)$ being the solution to the nonlinear Painlevé II differential equation, i.e.,

$$q''(u) = uq(u) + 2q^3(u). \tag{C.7}$$

Details concerning the analytical formula of $F_{TW_2}(s)$ can be found in [51], and the lookup table for the cdf of $F_{TW_2}(s)$ is available in [52].

As ℓ_{d+1} asymptotically has the behavior of $\ell_{\max}$, it follows from (C.4) and (C.5) that

$$\begin{aligned}P_{\text{fa}} &\approx \text{Prob}\left(\frac{\frac{\ell_{d+1}}{\tau}-\mu_{mn}}{\sigma_{mn}} > \frac{\delta-\mu_{mn}}{\sigma_{mn}}\right)\\ &= 1-\text{Prob}\left(\frac{\frac{\ell_{d+1}}{\tau}-\mu_{mn}}{\sigma_{mn}} < \frac{\delta-\mu_{mn}}{\sigma_{mn}}\right)\\ &= 1 - F_{TW_2}\left(\frac{\delta-\mu_{mn}}{\sigma_{mn}}\right)\end{aligned} \tag{C.8}$$

where $\delta = g^{-1}(c/2\log(2n)+1)$.

Using [51, App. A1], we assert that $q^2(u)$ is monotonically decreasing asymptotic to $|u|/2$ as $u \to -\infty$ and to $e^{-(4/3)u^{3/2}}/(4\pi\sqrt{u})$ as $u \to \infty$. Since $g^{-1}(x)$ is the increasing function, we obtain $\delta \to \infty$ as $n \to \infty$. As a result, it follows from (C.6) that

$$\begin{aligned}\lim_{s\to\infty} F_{TW_2}(s) &= \lim_{s\to\infty}\exp\left\{-\int_s^\infty (u-s)\frac{e^{-\frac{4}{3}u^{\frac{3}{2}}}}{4\pi\sqrt{u}}du\right\}\\ &= \lim_{s\to\infty}\exp\left\{-\int_s^\infty \frac{\sqrt{u}e^{-\frac{4}{3}u^{\frac{3}{2}}}}{4\pi}du\right\}\\ &\quad\times\lim_{s\to\infty}\exp\left\{\int_s^\infty \frac{se^{-\frac{4}{3}u^{\frac{3}{2}}}}{4\pi\sqrt{u}}du\right\}.\end{aligned} \tag{C.9}$$

Noting that

$$\begin{aligned}-\int_s^\infty \frac{\sqrt{u}}{4\pi}e^{-\frac{4}{3}u^{\frac{3}{2}}}du &= -\frac{1}{6\pi}\int_{s^{\frac{3}{2}}}^\infty e^{-\frac{4}{3}t}dt = -\frac{\pi}{8}e^{-\frac{4}{3}s^{\frac{3}{2}}}\\ &\to 0 \text{ as } s\to\infty\end{aligned} \tag{C.10}$$

$$\begin{aligned}0 < \int_s^\infty \frac{s}{4\pi\sqrt{u}}e^{-\frac{4}{3}u^{\frac{3}{2}}}du &< \int_s^\infty \frac{\sqrt{u}}{4\pi}e^{-\frac{4}{3}u^{\frac{3}{2}}}du = \frac{\pi}{8}e^{-\frac{4}{3}s^{\frac{3}{2}}}\\ &\to 0 \text{ as } s\to\infty\end{aligned} \tag{C.11}$$

we assert that $F_{TW_2}((\delta-\mu_{mn})/\sigma_{mn}) \to 1$ as $\delta \to \infty$, which, when substituted into (C.8), establishes that the probability of false alarm converges to zero as $m, n \to \infty$ and $m/n \to c$.

REFERENCES

[1] D. Gesbert, M. Kountouris, R. W. Heath, C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[3] X. Mestre and M. A. Lagunas, "Modified subspace algorithms for DOA estimation with large arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598–613, Feb. 2008.

[4] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[5] D. B. Williams and D. H. Johnson, "Using the sphericity test for source detection with narrow-band passive arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 11, pp. 2008–2014, Nov. 1990.

[6] Q. Wu and K. M. Wong, "Determination of the number of signals in unknown noise environments-PARADE," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 362–365, Jan. 1995.

[7] P.-J. Chung, J. F. Böhme, C. F. Mecklenbräuker, and A. O. Hero, "Detection of the number of signals using the Benjamini–Hochberg procedure," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2497–2508, Jun. 2007.

[8] 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.

[9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[10] D. F. Schmidt and E. Makalic, "The consistency of MDL for linear regression models with increasing signal-to-noise ratio," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1508–1510, Mar. 2011.

[11] M. Lu and A. M. Zoubir, "Generalized Bayesian information criterion for source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1470–1480, Mar. 2013.

[12] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

[13] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

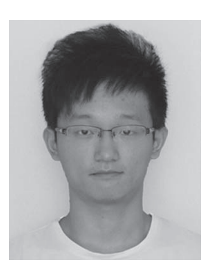
[14] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.

[15] C. Xu and S. Kay, "Inconsistency of the MDL: On the performance of model order selection criteria with increasing signal-to-noise ratio," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1959–1969, May 2011.
[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] L. Huang, T. Long, E. Mao, and H. C. So, "MMSE-based MDL method for robust estimation of number of sources without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4135–4142, Oct. 2009.
[21] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.
[22] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.
[23] 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.
[24] 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.
[25] 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.
[26] A.-K. Seghouane, "New AIC corrected variants for multivariante linear regression model selection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 2, pp. 1154–1165, Apr. 2011.
[27] P. Djurić, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2726–2735, Oct. 1998.
[28] P. Stoica and Y. Selén, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
[29] O. E. Barndorff-Nielsen and D. R. Cox, *Asymptotic Techniques for Use in Statistics*. New York, NY, USA: Chapman and Hall, 1989.
[30] G. Golub and C. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.
[31] H. Wang and M. Kaveh, "On the performance of signal subspace processing—Part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1201–1209, Oct. 1986.
[32] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1557–1567, Oct. 1989.
[33] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413–1426, Jun. 1995.
[34] F. Haddadi, M. Malek-Mohammadi, M. M. Nayebi, and M. R. Aref, "Statistical performance analysis of MDL source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 452–457, Jan. 2010.
[35] J. P. Delmas and Y. Meurisse, "On the second-order statistics of the EVD of sample covariance matrices: Application to the detection of noncircular or/and non Gaussian components," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4017–4023, Aug. 2010.
[36] J. Baik, G. B. Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, Sep. 2005.
[37] A. Onatski, M. Moreira, and M. Hallin, "Signal detection in high dimension: The multispiked case," *arXiv preprint arXiv:1210.5663*, 2012.
[38] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statist. Sin.*, vol. 17, no. 4, pp. 1617–1642, 2007.
[39] D. N. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, vol. 43, no. 1/2, pp. 128–136, Jun. 1956.
[40] L. Huang and H. C. So, "Source enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4806–4821, Oct. 2013.
[41] C. Xu and S. Kay, "Source enumeration via the EEF criterion," *IEEE Signal Process. Lett.*, vol. 15, pp. 569–572, 2008.
[42] B. Nadler, "Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2746–2756, May 2010.
[43] P. O. Perry and P. J. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 504–513, Jun. 2010.
[44] C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Commun. Math. Phys.*, vol. 177, no. 3, pp. 727–754, 1996.
[45] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Mulrivariare Anal.*, vol. 20, no. 1, pp. 1–25, Oct. 1986.
[46] J. V. Neumann, "Some matrix inequalities and metrization of matric-space," *Tomsk. Univ. Rev.*, no. 1, pp. 286–300, 1937.
[47] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," Dept. Elect. Comput. Eng., Univ. California, San Diego, CA, USA, Tech. Rep. Course Lect. Suppl. ECE275A, Sep.–Dec. 2005. [Online]. Available: http://dsp.ucsd.edu/kreutz/PEI05.html AQ1
[48] I. M. Johnstone and A. Y. Lu, "Sparse principal components analysis," Stanford Univ., Stanford, CA, USA, Tech. Rep. (ArXiv:0901.4392v1), 2004.
[49] I. M. Johnstone, "High dimensional statistical inference and random matrices," in *Proc. Int. Congr. Math.*, M. Sanz-Solé, J. Soria, J. Varona, and J. Verdera, Eds., Zürich, Switzerland, Eur. Math. Soc., 2006, pp. 307–333.
[50] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices: A review," *Statist. Sin.*, vol. 9, no. 3, pp. 611–677, Aug. 1999. AQ2
[51] I. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, Apr. 2001.
[52] A. Bejan, "Largest eigenvalues and sample covariance matrices," Tracy–Widom and Painleve II: Computational Aspects and Realization in S-Plus With Applications, 2005. [Online]. Available: http://www.vitrum.md/andrew/TWinSplus.pdf AQ3

Lei Huang (M'07–SM'14) 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, USA. From 2009 to 2010, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, and a Research Associate with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. From 2011 to 2014, he was a Professor with the Department of Electronic and Information Engineering, Shenzhen Graduate School of Harbin Institute of Technology, Shenzhen, China. In November 2014, he joined the Department of Information Engineering, Shenzhen University, where he is currently a Chair 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 serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and *Digital Signal Processing*.



Yuhang Xiao was born in Anhui, China, on January 20, 1992. He received the B.E. degree from Harbin Engineering University, Harbin, China, in 2012. He is currently working toward the Ph.D. degree in communication and information engineering with the Harbin Institute of Technology.

His research interests are in statistical signal processing and spectrum sensing.

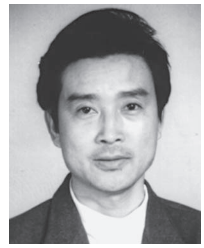
Kefei Liu received the B.Sc. degree in mathematics from Wuhan University, Wuhan, China, in 2006 and the Ph.D. degree in electronic engineering from City University of Hong Kong, Kowloon, Hong Kong, in 2013, respectively. His Ph.D. supervisor was Prof. H.-C. So, and his Ph.D. research topics were statistical and array signal processing, source enumeration, direction-of-arrival estimation, and multilinear algebra.

From September 2013 to December 2013, he was a Research Assistant of Prof. L. Huang with the Department of Electronic and Information Engineering, Shenzhen Graduate School of Harbin Institute of Technology, Shenzhen, China. Since January 2014, he has been a Postdoctoral Research Associate with the Department of Computer Science and Engineering and the Center for Evolutionary Medicine and Informatics, Biodesign Institute, Arizona State University, Tempe, AZ, USA. His cooperative supervisor is Prof. J. Ye. His current research interests are tensor decompositions for machine learning, randomized algorithms for matrix approximation, and their applications in analysis of massive biomedical data sets.

Hing Cheung So (S'90–M'95–SM'07–F'15) was born in Hong Kong. He received the B.Eng. degree in electronic engineering from City University of Hong Kong, Kowloon, Hong Kong, in 1990 and the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, Shatin, Hong Kong, in 1995.

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 was 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, robust estimation, source localization, and sparse approximation.

Dr. So has been on the Editorial Board of the IEEE Signal Processing Magazine since 2014, *Signal Processing* since 2010, and *Digital Signal Processing* since 2011 and was on the Editoral Board of the IEEE Transactions on Signal Processing during 2010–2014. In addition, since 2011, he has been an elected member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society, where he is the Chair of the awards subcommittee. He is elected Fellow of the IEEE in recognition of his contributions to spectral analysis and source localization.



Jian-Kang Zhang (SM'09) received the B.S. degree in information science (mathematics) from Shaanxi Normal University, Xi'an, China, in 1983; the M.S. degree in information and computational science (mathematics) from Northwest University, Xi'an, in 1988; and the Ph.D. degree in electrical engineering from Xidian University, Xi'an, in 1999.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. He has held research positions at McMaster University and Harvard University, Cambridge, MA, USA. His research interests are in the general area of signal processing, digital communication, signal detection and estimation, and wavelet and time–frequency analysis, mainly emphasizing mathematics-based new-technology innovation and exploration for a variety of signal processing and practical applications, and, specifically, number theory and various linear algebra-based kinds of signal processing. His current research focuses on transceiver designs for multiuser communication systems, coherent and noncoherent space–time signal, and receiver designs for multiple-input–multiple-output and cooperative relay communications.

Dr. Zhang is the coauthor of the paper that received the IEEE Signal Processing Society Best Young Author Award in 2008. He has served as an Associate Editor for the IEEE Signal Processing Letters. He is currently serving as an Associate Editor for the IEEE Transactions on Signal Processing and the *Journal of Electrical and Computer Engineering*.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = Provided URL in Ref. [47] was not found. Please check.
AQ2 = Note that references [16] and [50] are the same. Therefore, reference [50] was changed according to its publication details. Please confirm.
AQ3 = Provided URL in Ref. [52] was not found. Please check.

END OF ALL QUERIES