Efficient source enumeration for accurate direction-of-arrival estimation in threshold region ☆

Kefei Liu [a], Hing Cheung So [a], João Paulo C.L. da Costa [b], Florian Römer [c], Lei Huang [d],*

[a] *Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong*
[b] *Department of Electrical Engineering, University of Brasília, Brasília, Brazil*
[c] *Digital Broadcasting Research Laboratory, Ilmenau University of Technology, Ilmenau, Germany*
[d] *Department of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen Graduate School, Shenzhen, China*

A R T I C L E   I N F O

A B S T R A C T

Estimation of the number of signals impinging on an array of sensors, also known as source enumeration, is usually required prior to direction-of-arrival (DOA) estimation. In challenging scenarios such as the presence of closely-spaced sources and/or high level of noise, using the true source number for nonlinear parameter estimation leads to the threshold effect which is characterized by an abnormally large mean square error (MSE). In cases that sources have distinct powers and/or are closely spaced, the error distribution among parameter estimates of different sources is unbalanced. In other words, some estimates have small errors while others may be quite inaccurate with large errors. In practice, we will be only interested in the former and have no concern on the latter. To formulate this idea, the concept of effective source number (ESN) is proposed in the context of joint source enumeration and DOA estimation. The ESN refers to the actual number of sources that are visible at a given noise level by a parameter estimator. Given the numbers of sensors and snapshots, number of sources, source parameters and noise level, a Monte Carlo method is designed to determine the ESN, which is the maximum number of available accurate estimates. The ESN has a theoretical value in that it is useful for judging what makes a good source enumerator in the threshold region and can be employed as a performance benchmark of various source enumerators. Since the number of sources is often unknown, its estimate by a source enumerator is used for DOA estimation. In an effort to automatically remove inaccurate estimates while keeping as many accurate estimates as possible, we define the matched source number (MSN) as the one which in conjunction with a parameter estimator results in the smallest MSE of the parameter estimates. We also heuristically devise a detection scheme that attains the MSN for ESPRIT based on the combination of state-of-the-art source enumerators.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Direction-of-arrival (DOA) estimation of signals impinging on an array of sensors is a fundamental problem in array processing [1], ranging from radar, sonar to wireless communications and medical imaging. Estimating the number of source signals from noisy measurements, also called source enumeration, is a required task prior to applying DOA estimation methods such as maximum-likelihood estimator (MLE), estimation of signal parameters via rotational invariance techniques (ESPRIT) [2], multiple signal classification (MUSIC) [3] and principal-singular-vector utilization for modal analysis (PUMA) [4].

The conventional performance metric for source enumeration is the probability of correct detection (PoD) [5–8]. The underlying premise of using PoD as the performance measure is that the true signal number is always the optimal choice. However, in challenging scenarios where closely-spaced sources are present and/or the signal-to-noise ratio (SNR) is low, using the true signal number for subsequent estimation incurs abnormally large mean square error (MSE) of the estimated parameters.

Fig. 1 illustrates the typical composite MSE performance curve for nonlinear parameter estimation [9]. In the asymptotic region of high SNRs, the estimates are located in the neighborhood of the true parameters and the resultant estimation errors are small. Below a specific SNR called the threshold point, the MSE of the nonlinear parameter estimates departs from the Cramér–Rao bound (CRB) and rises rapidly. As the SNR decreases, the MSE rises until
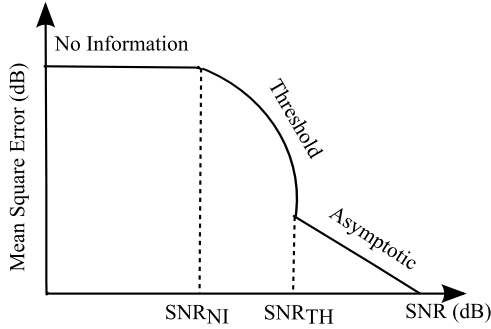
Fig. 1. Typical composite MSE curve for nonlinear parameter estimation [9].

it reaches a stable maximum value. The SNR at which the MSE enters this steady state of futility is called the no information point.

In the threshold region corresponding to closely-spaced sources and/or distinct signal powers, the estimation errors among sources are unevenly distributed: some estimates are centered about the true parameters and carry small or local errors [10], whereas other estimates carry random or global errors [10]. The latter estimates, also known as outliers [11,10], cause the threshold effect.

In practice, we are only interested in the parameter estimates with small errors but have no concern on the outliers. To formulate this idea, we propose the adaptive source number in an effort to remove unwanted outliers while keeping the accurate estimates. In particular, we have developed the concepts of effective source number (ESN) and matched source number (MSN) by considering the source enumeration and subsequent DOA estimation as a whole. The ESN refers to the actual number of sources that are visible at a given noise level by a certain parameter estimator. It is an indicator of how many parameter estimates returned by a parameter estimation scheme are accurate. The ESN is a theoretical concept that depends on the numbers of sensors and snapshots, number of sources, signal parameters and noise level, and can be determined by Monte Carlo simulation. It can be employed as a performance benchmark of various source enumerators and is useful for understanding what makes a good source enumerator in the threshold region. In practice the number of sources is often unknown, and instead its estimate by a source enumerator is used for DOA estimation. A good choice of a source number for a specific parameter estimation scheme varies and is application dependent. In this work, we focus on the scenario where false alarm entails much higher price than missed detection and hence under-enumeration is preferred, and formulate a close-to-generic definition of MSN whose combination with ESPRIT automatically removes unwanted outliers while keeping as many accurate estimates as possible. Moreover, we propose a detection scheme that attains the MSN in the threshold region based on the combination of state-of-the-art source enumerators [5,12–14,6,8].

It is worth noting that similar attempts have also been made in the literature. In [15], a detection method biased toward over-enumeration is designed to cater to radar imagery applications, where it is preferable to overestimate the number of harmonic components than underestimation. However, joint source enumeration and DOA estimation is not considered, and the extent of over-enumeration is not analyzed. More importantly, in our point of view underestimation should be preferred in a wide range of applications for challenging scenarios such as low SNR and/or presence of closely-spaced sources, since using an overestimated source number yields useless inaccurate parameter estimates which entail additional costs. In [16], a similar joint source enumeration and parameter estimation solution has been exploited based on the combination of the minimum description length (MDL) [5] criterion and ESPRIT for multipath time delay acqui-

sition. Nevertheless, the scheme is based on a very specific cost function of the parameter estimates, and the combination of MDL and ESPRIT is not sufficiently justified.

The underlying principles of ESN and MSN proposed in our work are also related to multi-model approach [17–21]. Different from the traditional single-model approach to model selection, where only one model is considered as correct and selected whereas all other models are ignored, in the multi-model approach, all fitted models are combined in a weighted manner according to their posterior likelihoods. This leads to the idea of model averaging that can be used for modeling and prediction. That is to say, our proposed model can be considered as a special case of multi-model approach. Our proposal shares similar idea to multi-model approach in the sense that alternative models (namely, underestimated number of sources) are considered in addition to the exact model (true number of sources).

The reminder of this paper is organized as follows. In Section 2, the data model for DOA estimation is formulated. In Section 3, we propose the concept of the ESN as well as a procedure to identify it using Monte Carlo simulation. In Section 4, we define the MSN, and design a detection scheme that can automatically attain the MSN. Simulation results are provided in Section 5 to evaluate the validity of the proposed concepts and detection scheme. Finally, conclusions are drawn in Section 6.

2. Data model

Consider a scenario where d far-field narrowband sources are impinging on a uniform linear array (ULA) of M sensors. For simplicity, we assume that the array and emitters are coplanar so that the DOA is characterized by the azimuth angle only.

The complex baseband output of the receive antennas is expressed as

$$\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S} + \boldsymbol{Z}, \tag{1}$$

where $\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_d)] \in \mathbb{C}^{M \times d}$ is the array steering matrix, and $\boldsymbol{a}(\theta_i) = [1, e^{j2\pi r \sin(\theta_i)/\lambda}, \ldots, e^{j(M-1)2\pi r \sin(\theta_i)/\lambda}]^T$, $i = 1, \ldots, d$, is the array steering vector that models the array response to a unit waveform from a certain direction. Here, $\theta_i \in [-\pi/2, \pi/2]$ is the DOA of the i-th source, measured relative to the array normal direction, r is the inter-element spacing of the receiving antenna array, and λ denotes the carrier wavelength. The $\mu_i = 2\pi r \sin(\theta_i)/\lambda$, $i = 1, \ldots, d$, is referred to as the spatial frequency. We assume that $\boldsymbol{A}(\boldsymbol{\theta})$ has full column rank of d and $d < \min(M, N)$. Furthermore, $\boldsymbol{S} = [\boldsymbol{s}_1^T, \ldots, \boldsymbol{s}_d^T]^T \in \mathbb{C}^{d \times N}$ contains the samples of all sources, and N is the number of snapshots. The $\boldsymbol{Z} \in \mathbb{C}^{M \times N}$ is the noise matrix collecting uncorrelated ZMCSCG samples with variance of σ_z^2. The noise is assumed to be uncorrelated with the signal.

Without loss of generality, in the sequel we use the following assumption: 1) $r = \lambda/2$ and hence $\mu_i = \pi \sin(\theta_i) \in [-\pi, \pi]$; 2) The samples of the i-th source $\boldsymbol{s}_i$, $i = 1, \ldots, d$, are generated from zero-mean circularly symmetric complex Gaussian (ZMCSCG) random process with variance of $\sigma_{s_i}^2$ and the samples of different sources are uncorrelated with each other; 3) The noise has unit variance, namely, $\sigma_z^2 = 1$.

2.1. Integration of source enumeration with parameter estimation

Since the ESPRIT is a closed-form parameter estimator that is accurate and computationally efficient, it is used for frequency estimation throughout the paper.

The ESPRIT algorithm uses the signal subspace to estimate the spatial frequencies. The first step is to compute the eigenvalue decomposition of the sample covariance matrix
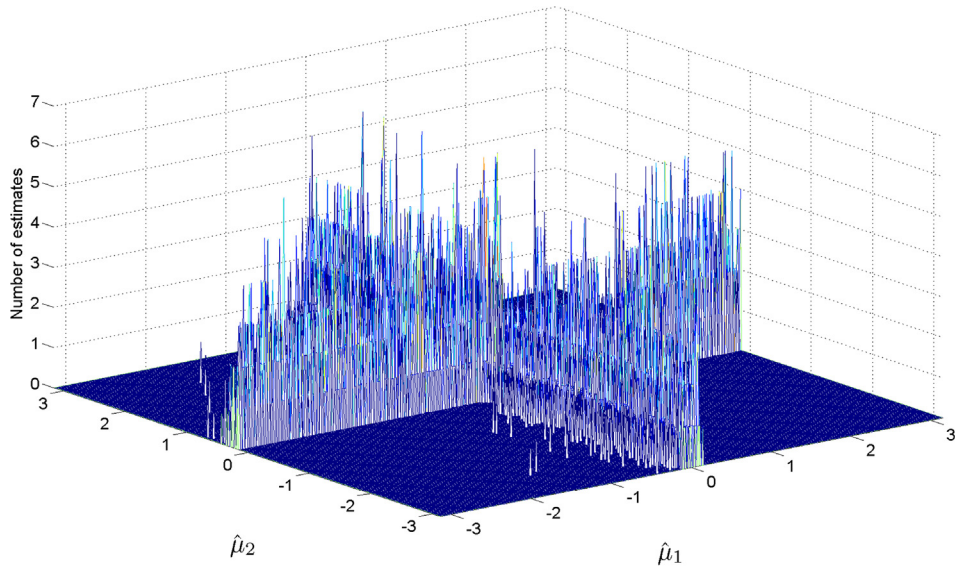
Fig. 2. Histogram of joint frequency estimate pair of two sources. $M = 10$, $N = 20$, $d = \hat{d} = 2$. $\sigma_{s_1}^2 = \sigma_{s_2}^2 = 0.1$, $\theta = [0, 5]°$.

$$\hat{R}_X = \frac{1}{N} XX^{\mathrm{H}} \in \mathbb{C}^{M \times M}, \tag{2}$$

where $^{\mathrm{H}}$ represents the Hermitian transpose. The $\hat{d}$ eigenvectors associated to the $\hat{d}$ largest eigenvalues, where $\hat{d}$ denotes the estimated source number, are assumed to form the signal subspace $U_{\hat{d}}$.

The shift invariance equation then takes the following form:

$$J_1 U_{\hat{d}} \Phi_{\hat{d}} \approx J_2 U_{\hat{d}}, \tag{3}$$

where $\Phi_{\hat{d}} \in \mathbb{C}^{\hat{d} \times \hat{d}}$ is the unknown matrix to be solved. The above sets of equations are overdetermined and can be solved by the least squares method.

From $\Phi_{\hat{d}}$, the spatial frequencies are estimated as

$$\hat{\mu}_i = \arg(\phi_i), \quad i = 1, \ldots, \hat{d}, \tag{4}$$

where ϕ_i denotes the i-th eigenvalue of $\Phi_{\hat{d}}$.

3. Effective source number

Suppose that the true number of signals d is passed to ES-PRIT, and denote the resultant d frequency estimates as $\hat{\mu}_1, \ldots, \hat{\mu}_d$, where $\hat{\mu}_i$, $i = 1, \ldots, d$, is the estimate of the spatial frequency of the i-th source, namely, μ_i.

To illustrate the concept of ESN, we consider a scenario with $d = 2$ equal-power and closely-spaced sources. Other parameter settings are $M = 10$, $N = 20$, $d = \hat{d} = 2$, and $\sigma_{s_1}^2 = \sigma_{s_2}^2 = 0.1$. 1000 Monte Carlo (MC) trials are conducted. In Fig. 2, the histogram of the joint frequency estimate pair $(\hat{\mu}_1, \hat{\mu}_2)$ is shown. Note that in almost all trials only one estimate is accurate while the other one has random errors. This indicates that only one source is visible by ESPRIT. This is due to the fact that the frequency resolution of ESPRIT is limited. For two closely-spaced sources corrupted by higher levels of noise, their small angular separation becomes invisible and they merge to one.

Loosely speaking, the ESN, denoted here as d_{eff}, refers to the number of signals that are actually visible at a given noise level by a certain parameter estimation scheme. Formally, it is defined as the number of accurate elements within the vector of parameters estimated when the true source number d is used for parameter estimation. The ESN is a function of the numbers of sensors M and snapshots N, number of sources d, the powers and DOAs of

d sources, and noise level. It is evident that in the asymptotic region $d_{\mathrm{eff}} = d$ while in the no information region $d_{\mathrm{eff}} = 0$. In the threshold region, d_{eff} falls between 0 and d. In the next section, we propose an algorithm to determine the ESN in the threshold region.

The theoretical value of ESN lies in that it is an indicator of the maximum number of available accurate parameter estimates, which helps us judge whether a source enumerator is good or not in the threshold region. As a future research direction, we plan to derive the analytical expression of the ESN and to design a source enumerator that can return a source number estimate fitting well with the ESN. This has practical significance in that with the knowledge of ESN, we may be able to decide which parameter estimates are accurate.

3.1. Proposed algorithm for identifying effective source number

The proposed algorithm relies on the empirical observation that the global errors can be approximated by random interval errors that are uniformly distributed over a certain range of values [9].

For two closely-spaced sources with equal power, however, in the threshold region the random error which is the characteristic of an outlier is contained in the parameter estimates of both sources. Fig. 3 shows the individual histograms of the frequency estimates $\hat{\mu}_1$ and $\hat{\mu}_2$, where the parameter settings are the same as those in Fig. 2. We see that although an outlier has been recognized from Fig. 2, Fig. 3 indicates that neither estimate has an approximate uniform distribution in $[-\pi, \pi]$.

To tackle this, in each MC trial we order the $d = 2$ frequency estimates in terms of accuracy such that

$$|\hat{\mu}_{i_1} - \mu_{i_1}| \leqslant |\hat{\mu}_{i_2} - \mu_{i_2}|, \tag{5}$$

where $\{i_1, i_2\}$ is a permutation of $\{1, 2\}$. Hereafter $\hat{\mu}_{i_1}$ and $\hat{\mu}_{i_2}$ are referred to as accurate and inaccurate estimates.

The histograms of the accurate and inaccurate estimates are plotted, respectively, at the top and bottom of Fig. 4. Note that the inaccurate estimate is now approximately uniformly distributed and hence the outlier can be readily discovered.

As the SNR continues to decrease, the number of outliers is expected to increase from 1 to 2. Indeed, for an SNR of -30 dB, the union of the two estimates approximately obeys the uniform
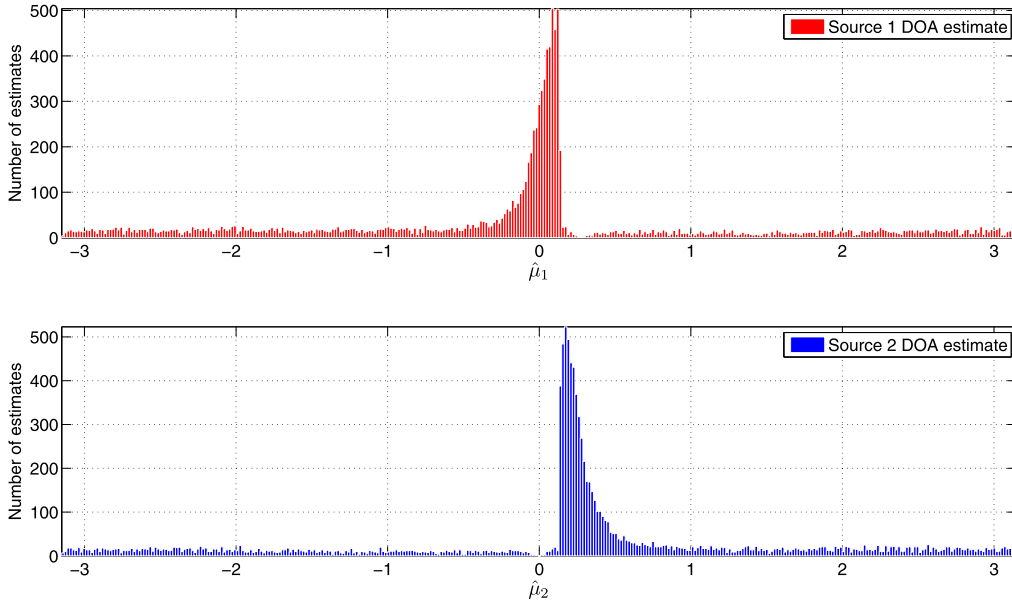
Fig. 3. Histograms of frequency estimates of individual sources. The parameter settings are the same as those in Fig. 2.
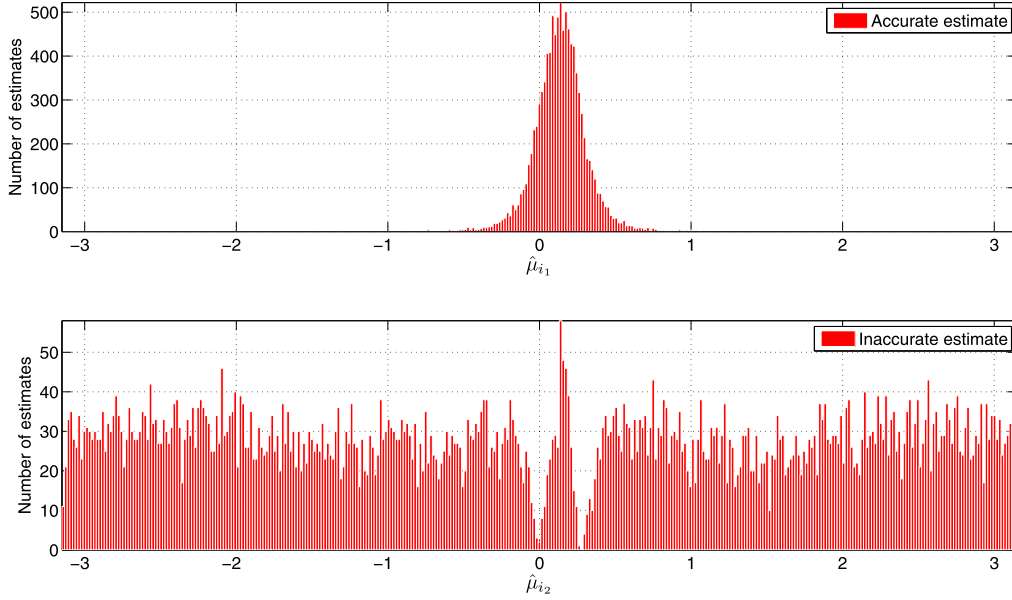

Fig. 4. Histograms of accurate and inaccurate frequency estimates by ESPRIT. The parameter settings are the same as those in Fig. 2. (Top: accurate; bottom: inaccurate.)

distribution, as shown at the bottom of Fig. 5. Note that neither of the accurate and inaccurate estimates alone follows the uniform distribution since they are ordered estimates and do not correspond to the sources. Instead, the accurate estimate is subject to a super-uniform distribution which exhibits a convex profile, while the inaccurate one is subject to a sub-uniform distribution with a concave profile.

It is worth mentioning that similar results are also obtained for multiple-source scenarios and for widely-spaced sources with imbalanced powers. In general, when d_{ol} outliers are present, the union of the d_{ol} least accurate estimates approximately follows the uniform distribution. Based on this observation, we propose an algorithm to identify the ESN by the MC method.

Given the numbers of sensors M and snapshots N, number of sources d, source powers and spatial frequencies $\mu_1, \mu_2, \ldots, \mu_d$, and noise level, the algorithm for identifying the ESN proceeds in the following steps.

1) MC simulation. Obtain Q random realizations of the signal-plus-noise measurement $\boldsymbol{X} \in \mathbb{C}^{M \times N}$ according to (1) by varying the source samples and noise. For the q-th realization, $q = 1, \ldots, Q$.

a) Pass d to ESPRIT to obtain the frequency estimates of d sources, denoted as $\hat{\mu}_1^{(q)}, \ldots, \hat{\mu}_d^{(q)}$.

b) Order the d frequency estimates in terms of accuracy as $\hat{\mu}_{i_1}^{(q)}, \ldots, \hat{\mu}_{i_d}^{(q)}$ such that

$$\left| \hat{\mu}_{i_1}^{(q)} - \mu_{i_1} \right| \leqslant \cdots \leqslant \left| \hat{\mu}_{i_d}^{(q)} - \mu_{i_d} \right|, \tag{6}$$

where $\{i_1, \ldots, i_d\}$ is a permutation of $\{1, \ldots, d\}$.

To reduce the statistical error, the number of MC trials Q should be sufficiently large, typically of the order of 1000.

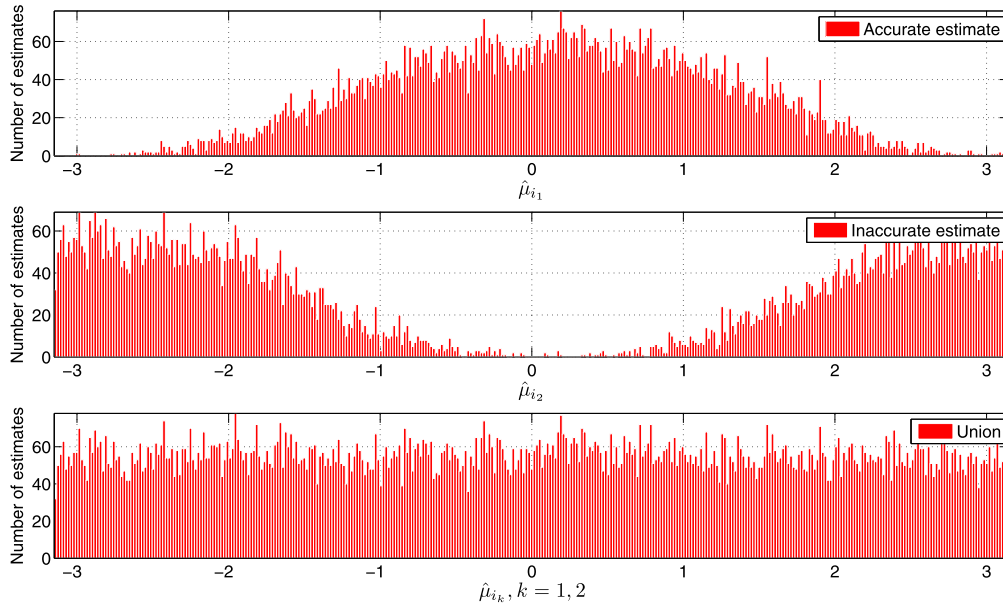2) Define the k-th ($k = 1, \ldots, d$) set of estimates as

Fig. 5. Histograms of accurate and inaccurate frequency estimates by ESPRIT and their union. $\sigma_{s_1}^2 = \sigma_{s_2}^2 = 0.001$. Other parameter settings are the same as those in Fig. 2. (Top: accurate; middle: inaccurate; bottom: union.)

$$\Omega_k = \bigcup_{q=1}^{Q} \{\hat{\mu}_{i_k}^{(q)}, \ldots, \hat{\mu}_{i_d}^{(q)}\}. \tag{7}$$

Test the goodness of fit of each set of estimates for uniform distribution using the chi-squared test. First, divide $[-\pi, \pi]$ into l equally-spaced intervals, and then calculate the chi-squared statistic [22,23]:

$$\chi^2 = \sum_{i=1}^{l} \frac{(f_i - Q/l)^2}{Q/l} = \frac{l}{Q} \sum_{i=1}^{l} f_i^2 - Q, \tag{8}$$

where f_i is the observed number of estimates that falls within the i-th interval. To maximize the power of the test, l is set as [22]

$$l = 4 \sqrt[5]{\frac{2(Q-1)^2}{\gamma^2}}, \tag{9}$$

where γ is determined from

$$\int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} dy = \alpha, \tag{10}$$

with α being the level of significance whose typical values is 1%, 5% or 10%.

The set of estimates with the smallest χ^2 is closest to being uniformly distributed.

3) Supposing Ω_j has the best fit for uniform distribution, the number of outliers is $d_{ol} = [d - (j - 1)]$, and the ESN is $d_{eff} = (j - 1)$.

Finally, we illustrate how to identify the threshold SNR and no information SNR, which is required in determination of the ESN and MSN. According to [24], the theoretical MSE expression of the ESPRIT in the asymptotic region is derived as

$$\text{MSE}_{\text{theo}} = \frac{\sum_{i=1}^{d} \|\boldsymbol{e}_i^{\text{H}} (\boldsymbol{A}^{\uparrow\dagger} \boldsymbol{I}_M^{\uparrow} - \boldsymbol{A}^{\downarrow\dagger} \boldsymbol{I}_M^{\downarrow}) \|^2 \boldsymbol{R}_S^{-1}(i, i) \sigma_z^2}{2NK}, \tag{11}$$

where $\boldsymbol{e}_i$ is a $d \times 1$ vector with 1 in the i-th position and 0 otherwise, $\boldsymbol{I}_M$ is the $M \times M$ identity matrix, $\boldsymbol{R}_S$ is the signal covariance

matrix, $\dagger$ stands for the Moore–Penrose pseudo-inverse, $[\cdot]^{\uparrow}$ denotes a matrix with the first row deleted, and $[\cdot]^{\downarrow}$ denotes a matrix with the last row deleted. Hence, the threshold SNR is identified as

$$\text{SNR}_{\text{th}} = \min \text{SNR} \quad \text{st.} \quad \sqrt{\text{MSE}_{\text{empi}}(\text{SNR})} > \sqrt{\text{MSE}_{\text{theo}}(\text{SNR})} + \eta, \tag{12}$$

where the SNR is defined as

$$\text{SNR} = \frac{\sum_{i=1}^{d} \sigma_{s_i}^2 / d}{\sigma_z^2}. \tag{13}$$

$\text{MSE}_{\text{empi}}(\text{SNR})$ denotes the empirical MSE obtained by MC simulation, and η denotes the preset deviation-tolerance for the estimation error.

To identify the no information SNR, consider that the MSE of ESPRIT in the no information region can be well approximated by the one in noise-only case, denoted here as $\text{MSE}_{\text{noise}}$. Consequently, the no information SNR is identified as

$$\text{SNR}_{\text{ni}} = \max \text{SNR} \quad \text{st.} \quad \text{MSE}_{\text{empi}}(\text{SNR}) < (1 - \rho) \text{MSE}_{\text{noise}}, \tag{14}$$

where $\text{MSE}_{\text{noise}}$ denotes the empirical MSE obtained from noise-only measurements, and ρ is a preset relative error.

4. Matched source number

In practice the number of signals d is unknown and is estimated by a source enumerator. The source number estimate is then used for DOA estimation. From Section 3, we know that in the threshold region, when d is passed to a given parameter estimator, we obtain a maximum of d_{eff} accurate estimates, and simultaneously $(d - d_{eff})$ outliers. Empirically, using an overestimated number of signals results in the same number of accurate estimates but more than $(d - d_{eff})$ outliers. For applications where missed detection incurs much more cost than false alarm, using the correct or overestimated source number is appropriate.

However, in applications where false alarm entails substantial loss and missed detection is not cared, the under-enumeration is a better choice. Empirically we observe that using an underestimated number of signals may automatically remove unwanted outliers
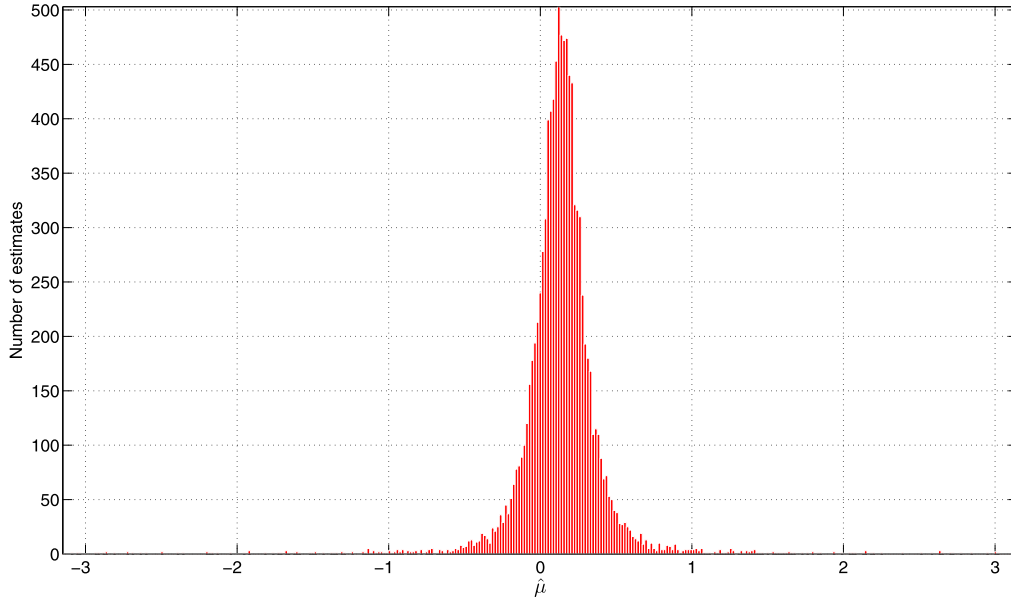
Fig. 6. Histograms of estimated frequency by ESPRIT, when the source number is underestimated as $\hat{d} = 1$. The parameter settings are the same as those in Fig. 2.

while keeping a number of accurate estimates. In Fig. 6, the parameter settings are the same as those in Fig. 2, except that an underestimated source number, namely, $\hat{d} = 1$, is used for DOA estimation instead of the true one. The histogram of the resultant frequency estimates is plotted in Fig. 6. Note that the outlier is automatically removed while an accurate estimate is obtained.

In an effort to automatically remove inaccurate estimates while keeping as many accurate estimates as possible, we define the MSN as the source number that results in the smallest MSE when being passed to a parameter estimator, namely,

$$d_{\mathrm{mat}} = \arg\min_{k=1,\ldots,d} \mathrm{MSE}(k), \tag{15}$$

with

$$\mathrm{MSE}(k) = \frac{\sum_{j=1}^{k} |\hat{\mu}_j - \mu_{i_j}|^2}{k}, \tag{16}$$

where $\hat{\mu}_1, \ldots, \hat{\mu}_k$ are the frequency estimates obtained by passing k to a parameter estimator, and $(\hat{\mu}_j, \mu_{i_j})$, $j = 1, \ldots, k$, is the j-th estimated-true frequency pair with $\{i_1, \ldots, i_k\}$ being a combination of k elements out of $\{1, \ldots, d\}$.

As shown in Section 5, the MSN is equal to or comparable to ESN, and by inputting d_{mat} in ESPRIT the output d_{mat} estimates has a comparable accuracy to the d_{eff} estimates returned from inputting d in ESPRIT.

4.1. Design of source enumerator to attain matched source number

Since the ESPRIT is a subspace-based parameter estimation algorithm, subspace-based source enumerators such as ESTimation ERror (ESTER) [13,14], subspace-based automatic model order selection (SAMOS) [25], and order estimation using principal angles between subspaces (OE-PABS) [26] are considered as the matching detection methods for it. In this study, the ESTER is investigated due to the fact that it directly exploits the shift invariance property of the signal subspace similarly to ESPRIT.

Let $\boldsymbol{U}_k$ collect the k dominant left singular vectors associated to the largest k singular values, where k is a candidate value for the estimated number of sources, and define the residual matrix as

$$\boldsymbol{E}_k = \boldsymbol{U}_k^{\downarrow}\boldsymbol{\Phi}_k - \boldsymbol{U}_k^{\uparrow}, \tag{17}$$

where

$$\boldsymbol{\Phi}_k = \left(\boldsymbol{U}_k^{\downarrow}\right)^{\dagger}\boldsymbol{U}_k^{\uparrow}. \tag{18}$$

The ESTER estimate of the number of signals is given by

$$\hat{d}_{\mathrm{ESTER}} = \arg\min_{k=1,\ldots,\min(M-2,N)} \|\boldsymbol{E}_k\|_2^2, \tag{19}$$

where $\|\cdot\|_2$ denotes the spectral norm or 2-norm of a matrix.

Note that the cost function in the ESTER criterion, namely, $\|\boldsymbol{E}_k\|_2^2$, serves as an approximate upper bound of the MSE defined in (16). To see this, according to Corollary 3 of [13], we have

$$|\hat{\phi}_j - \phi_{i_j}| \leqslant c\|\boldsymbol{E}_k\|_2, \quad j = 1, \ldots, k, \tag{20}$$

where $\hat{\phi}_j$ is the j-th eigenvalue of $\boldsymbol{\Phi}_k$, ϕ_{i_j} is the eigenvalue of $\boldsymbol{\Phi}_d$ that is closest to $\hat{\phi}_j$ in least squares sense, and c is a constant that is independent of k.

Since

$$|\hat{\mu}_j - \mu_{i_j}| \simeq \frac{|\hat{\phi}_j - \phi_{i_j}|}{|\phi_{i_j}|} \leqslant c\frac{\|\boldsymbol{E}_k\|_2}{|\phi_{i_j}|}, \tag{21}$$

and $|\phi_{i_j}| = 1$, it follows that for any $1 \leqslant k < d$

$$\mathrm{MSE}(k) \lesssim c^2\|\boldsymbol{E}_k\|_2^2, \tag{22}$$

where $\lesssim$ denotes "less than or approximately equal to". Therefore, minimization of the ESTER cost function leads to minimization of MSE(k) in (16) and hence help us find d_{mat}.

The ESTER is robust against finite-sample constraint in the sense that it works well even when the number of snapshots is relatively small compared to the number of sensors. However, it is sensitive to the angular separation of sources and in the presence of closely-spaced sources, has a tendency to underestimate the number of signals. To deal with scenarios with closely-spaced sources, we adopt its modified version:

$$\hat{d}_{\mathrm{mat}} = \arg\min_{k=\max(1,\hat{d}_{\min}),\ldots,\min(M-2,N,\hat{d}_{\max})} \|\boldsymbol{E}_k\|_2^2, \tag{23}$$

where $\hat{d}_{\min}$ and $\hat{d}_{\max}$ are the lower and upper bounds of the candidate number of signals, respectively.

For large-sample scenarios where the number of snapshots N is much larger than the number of sensors M, we propose to establish the lower and upper bounds via the information criteria (ITCs) based detection methods, which are derived based on the maximum likelihood principle and are optimal large-sample detectors. In ITCs, the number of signals is determined by minimizing the following function [5,27]:

$$\text{ITC}(k) = N(M-k)\log\left(\frac{a(k)}{g(k)}\right) + \frac{1}{2}k(2M-k)C(N), \qquad (24)$$

where

$$g(k) = \left(\prod_{i=k+1}^{M} \ell_i\right)^{\frac{1}{M-k}}, \qquad (25)$$

$$a(k) = \frac{1}{M-k}\sum_{i=k+1}^{M}\ell_i \qquad (26)$$

are the geometric and arithmetic means of the $(M-k)$ smallest sample eigenvalues, respectively, and $C(N)$ is the penalty coefficient which may be a function of N.

The Akaike information criterion (AIC), MDL [5] and efficient detection criterion (EDC) [27] as members of the family of ITCs, have different penalty coefficients for penalizing overfitting of the model:

$$\text{AIC:} \quad C(N) = 2, \qquad (27)$$

$$\text{MDL:} \quad C(N) = \log(N), \qquad (28)$$

$$\text{EDC:} \quad C(N) = \sqrt{N \cdot \log\big(\log(N)\big)}. \qquad (29)$$

We see that in AIC the lightest penalty is imposed, whereas in EDC the heaviest penalty is imposed. Consequently, the AIC is more inclined to overestimate the number of signals, whereas the EDC is apt to underestimation.

Therefore, we propose to adopt the following estimates as the lower bound $\hat{d}_{\min}$ and upper bound $\hat{d}_{\max}$:

$$\hat{d}_{\min} = \hat{d}_{\text{EDC}}, \qquad (30)$$

$$\hat{d}_{\max} = \hat{d}_{\text{AIC}}. \qquad (31)$$

For small-sample scenarios where M and N are comparable to each other, the ITCs are no longer optimal. Instead, the source enumerators that are devised based on the large-dimensional random matrix theory have been proven better [6–8,28] and can be used to establish the lower and upper bounds.

5. Simulation results

The simulated data are generated according to (1). To determine the threshold and no information SNRs, η and ρ are set as $1°$ and 0.05, respectively. For given numbers of sensors and snapshots, number of sources, source powers and spatial frequencies, the results are plotted for various SNRs, which are defined in (13). For each SNR, the results represent an average of $Q = 1000$ independent realizations.

First, we validate the concepts of the proposed ESN and MSN and their determination algorithms. We consider a system where $d = 5$ equal-power sources are impinging on a ULA of $M = 10$ elements each collecting $N = 20$ snapshots. The DOAs are $\boldsymbol{\theta} = [-90, -45, -42, -60, -70]°$ such that a pair of closely-spaced sources is present. Fig. 7 shows the identified ESN and MSN for various SNRs by MC simulation. Note that the ESN/MSN is equal to the number of sources at a low noise level in the asymptotic region, and gradually decreases as the noise power increases in the
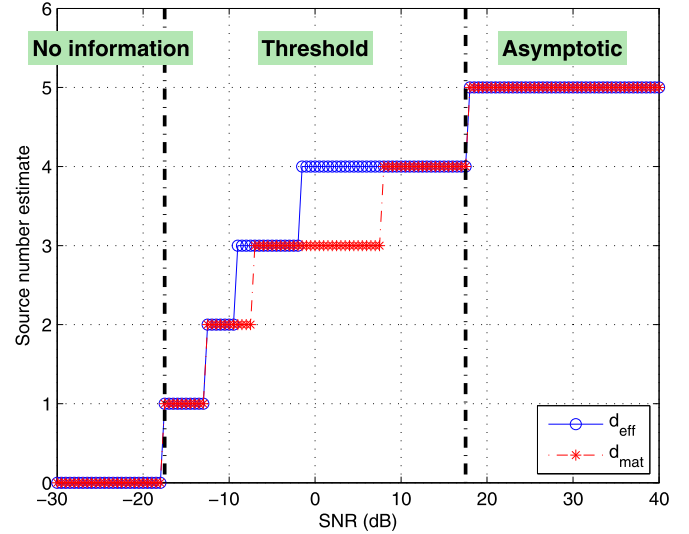

Fig. 7. Effective and matched source number versus SNR for DOA estimation in the presence of closely-spaced sources. $M = 10$, $N = 20$, $d = 5$. $\boldsymbol{\theta} = [-90, -45, -42, -60, -70]°$, $\eta = 1°$, $\rho = 0.05$.

threshold region, and finally down to zero at a high noise level in the no information region. Moreover, the gap between the MSN and ESN is small.

Fig. 8(a) shows the corresponding root MSE (RMSE) of the frequency estimates obtained by ESPRIT. In the legend, the "Matched" curve corresponds to the RMSE computed based on the d_{mat} frequency estimates that are returned when d_{mat} is sent to ESPRIT, whereas the "BEST: d_{eff}" curve corresponds to the RMSE based on the d_{eff} most accurate frequency estimates chosen from all d frequency estimates that are obtained when d is fed into ESPRIT. And the "CRB: d_{mat}" and "CRB: d_{eff}" curve respectively correspond to the square root of the mean of the d_{mat} and d_{eff} smallest diagonal elements of the CRB covariance matrix.

The result shows that the "Matched" curve is far below that of the total RMSE, which shows that under-enumeration indeed can result in improved RMSE. Likewise, the RMSE of the d_{eff} most accurate frequency estimates is well below that of the total RMSE, which indicates that the outliers have been excluded from the estimates. Furthermore, compared with the d_{eff} most accurate frequency estimates using d, there is only slight decrease in estimation accuracy of the d_{mat} frequency estimates using d_{mat}. However, for the former, other than the d_{eff} most accurate frequency estimates, $(d - d_{\text{eff}})$ outliers are also obtained, whereas for the latter, the outliers have been automatically removed, leaving d_{mat} accurate estimates. As shown in Fig. 8(b), the MSN is "optimal" in a sense that it corresponds to the smallest RMSE of the frequency estimates that could be obtained by inputting varying source number in ESPRIT. Consequently, a maximum number of accurate estimates is kept while the decline in accuracy is minimized. In this sense, the MSN is a good match for ESPRIT.

Next, we evaluate the performance of the proposed source enumerator to attain the MSN. We adopt the root mean square detection error (RMSDE) as the performance measure, which is defined as

$$\text{RMSDE} = \sqrt{\frac{\sum_{i=1}^{Q}|\hat{d} - d_{\text{mat}}|^2}{Q}}. \qquad (32)$$

The RMSDEs of AIC, MDL, EDC, SAMOS, and OE-PABS are used as the performance benchmarks.

First we consider a scenario with widely-spaced sources of imbalanced powers: $M = 10$, $N = 100$ and $d = 5$. The DOAs are
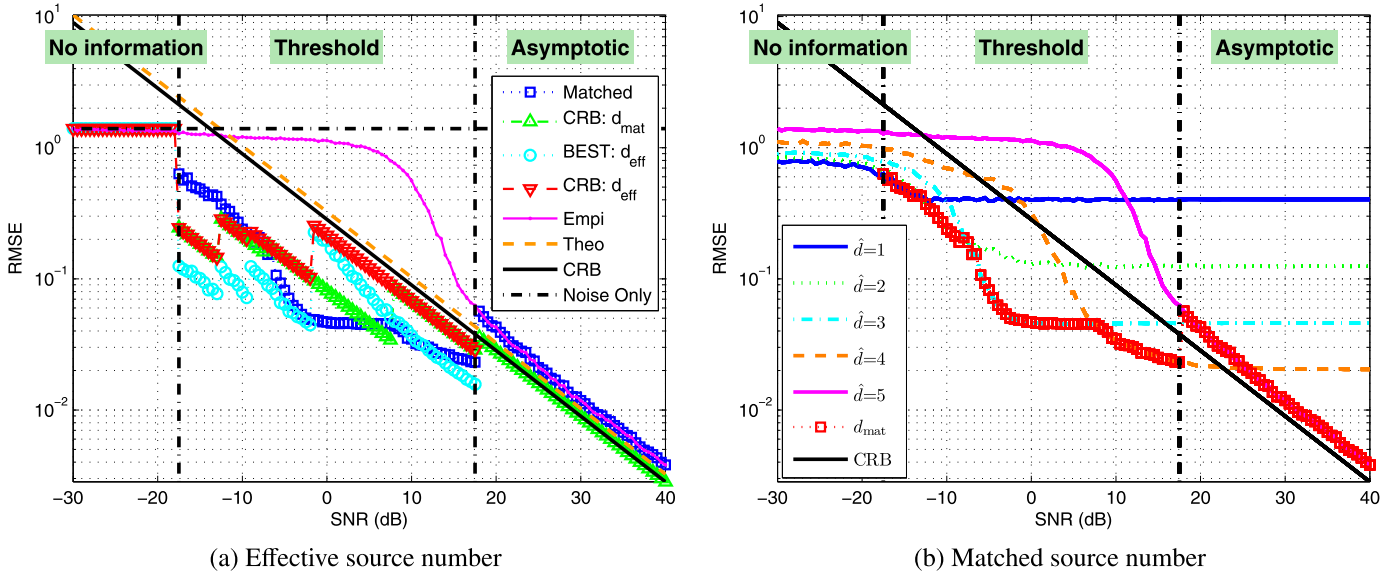
(a) Effective source number

(b) Matched source number

Fig. 8. RMSE of frequency estimates versus SNR in the presence of closely-spaced sources. The parameter settings are the same as that in Fig. 7.
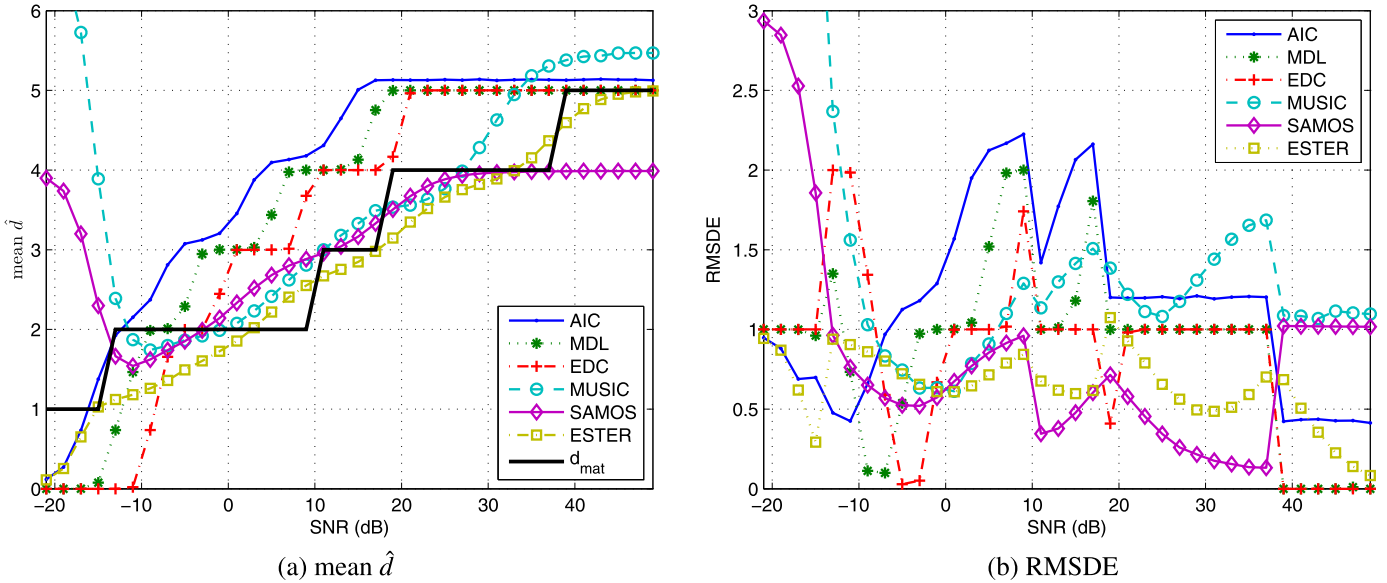

(a) mean $\hat{d}$

(b) RMSDE

Fig. 9. Goodness of fit between estimated source number by various source enumerators and matched source number, for a scenario with widely-spaced sources of imbalanced powers. $M = 10$, $N = 100$, $d = 5$. $\theta = [-81, -45, -9, 27, 63]°$. $\sigma_{s_1}^2 : \sigma_{s_2}^2 : \sigma_{s_3}^2 : \sigma_{s_4}^2 : \sigma_{s_5}^2 = 1 : 10 : 100 : 500 : 1000$.

set as $\theta = [-81, -45, -9, 27, 63]°$, and the ratio of source powers are fixed as $\sigma_{s_1}^2 : \sigma_{s_2}^2 : \sigma_{s_3}^2 : \sigma_{s_4}^2 : \sigma_{s_5}^2 = 1 : 10 : 100 : 500 : 1000$. In Fig. 9(a) and Fig. 9(b), the mean value of the estimated source number by different source enumerators and the RMSDE are shown, respectively. We see that the subspace-based source enumerators such as ESTER, SAMOS and MUSIC return source number estimates that fit much better with the MSN than that returned by ITCs. In particular, the ESTER performs consistently better than or comparably to other source enumerators for the whole SNRs in the sense that its returned source number estimate has small deviations from the MSN.

For scenarios with the presence of closely-spaced sources, the mean value of the estimated source number by different source enumerators and the RMSDE are shown in Fig. 10, where the DOAs are $\theta = [20, 59, 28, 46, 55]°$ and the sources have equal powers. We see that in the presence of closely-spaced sources, the subspace-based source enumerators such as SAMOS, ESTER and OE-PABS

tend to underestimate the MSN for high SNRs and to overestimation for low SNRs. In other words, their estimated source number is considerably deviated from the MSN. In contrast, the proposed detection scheme, which is a combination of the ESTER and ITCs, returns a source number estimate that fits well with the MSN, with higher level of fitness than that of the ITCs.

6. Conclusion

In the threshold region of nonlinear parameter estimators, the use of true source number for DOA estimation incurs large mean square error. In the presence of closely-spaced sources and/or severely imbalanced powers among sources, the estimation errors are unequally distributed among the parameter estimates of different sources. The number of signals whose parameters can be accurately estimated is referred to as the ESN. Given the numbers of sensors and snapshots, number of sources, signal parameters and

(a) mean $\hat{d}$

(b) RMSDE
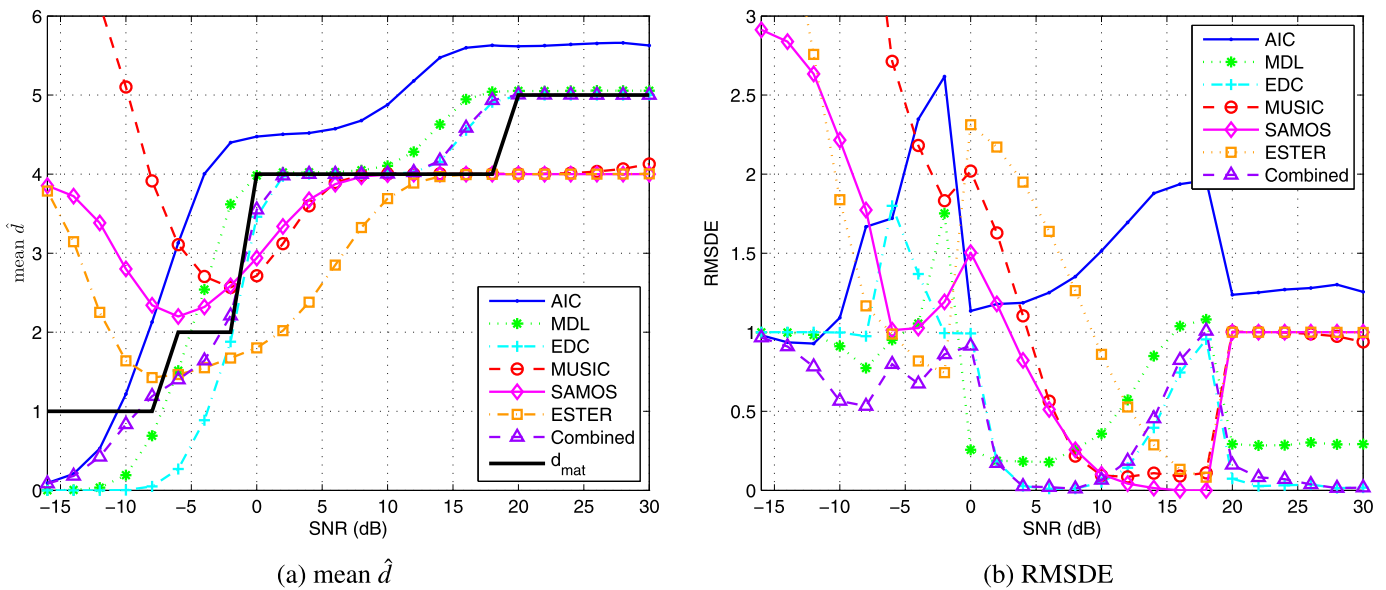
Fig. 10. Goodness of fit between estimated source number by various source enumerators and matched source number, in the presence of closely-spaced sources. The sources have equal powers. $M = 10$, $N = 100$, $d = 5$. $\theta = [20, 59, 28, 46, 55]°$.

noise level, the ESN can be determined by Monte Carlo simulation. In an effort to automatically remove the inaccurate estimates while keeping the accurate ones, we have proposed the concept of MSN for a parameter estimation scheme, which refers to the number of signals that yields the smallest MSE of the parameter estimates. Moreover, we have designed a detection scheme that can attain the MSN for ESPRIT based on the combination of state-of-the-art source enumerators. The proposed concepts and approaches may be generalized to other parameter estimation schemes. As a future research direction, we plan to derive the analytical expression of the ESN and design a source enumerator that can return a source number estimate with better fitness for the ESN than existing detection algorithms. This has practical significance in that with the knowledge of ESN, namely, the maximum number of available accurate frequency estimates, we may be able to decide which parameter estimates are accurate.

References

[1] H. Krim, M. Viberg, Two decades of array signal processing research: The parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94.

[2] R. Roy, T. Kailath, ESPRIT – Estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. ASSP-37 (1989) 984–995.

[3] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propagat. 34 (3) (1986) 276–280.

[4] F.K.W. Chan, H.C. So, W. Sun, Subspace approach for two-dimensional parameter estimation of multiple damped sinusoids, Signal Process. 92 (9) (2012) 2172–2179.

[5] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, IEEE Trans. Acoust. Speech Signal Process. 33 (2) (1985) 387–392.

[6] A. Quinlan, J.-P. Barbot, P. Larzabal, M. Haardt, Model order selection for short data: An exponential fitting test (EFT), EURASIP J. Adv. Signal Process. 2007 (2007) 071953.

[7] R.R. Nadakuditi, A. Edelman, Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples, IEEE Trans. Signal Process. 56 (7) (2008) 2625–2638.

[8] S. Kritchman, B. Nadler, Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory, IEEE Trans. Signal Process. 57 (10) (2009) 3930–3941.

[9] C.D. Richmond, Mean-squared error and threshold SNR prediction of maximum-likelihood signal parameter estimation with estimated colored noise covariances, IEEE Trans. Inform. Theory 52 (5) (2006) 2146–2164.

[10] C.D. Richmond, Capon algorithm mean-squared error threshold SNR prediction and probability of resolution, IEEE Trans. Signal Process. 53 (8) (2005) 2748–2764.

[11] F. Athley, Performance analysis of DOA estimation in the threshold region, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP '02), Orlando, FL, USA, 2002, vol. III, pp. 3017–3021.

[12] L.C. Zhao, P.R. Krishnaiah, Z.D. Bai, On detection of number of signals in presence of white noise, J. Multivar. Anal. 20 (1986) 1–25.

[13] R. Badeau, B. David, G. Richard, Selecting the modeling order for the ESPRIT high resolution method: An alternative approach, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP '04), Montreal, Canada, 2004, vol. 2, pp. 1025–1028.

[14] R. Badeau, B. David, G. Richard, A new perturbation analysis for signal enumeration in rotational invariance techniques, IEEE Trans. Signal Process. 54 (2) (2006) 450–458.

[15] E. Radoi, A. Quinquis, A new method for estimating the number of harmonic components in noise with application in high resolution radar, EURASIP J. Appl. Signal Process. 8 (2004) 1177–1188.

[16] B. Yang, K.B. Letaief, R.S. Cheng, Z. Cao, Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling, IEEE Trans. Commun. 49 (3) (2001) 467–479.

[17] J.A. Hoeting, D. Madigan, A.E. Raftery, C.T. Volinsky, Bayesian model averaging: A tutorial, Stat. Sci. (1999) 382–401.

[18] K.P. Burnham, D.R. Anderson, Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach, Springer Verlag, 2002.

[19] P. Stoica, Y. Selén, J. Li, Multi-model approach to model selection, Digital Signal Process. 14 (5) (2004) 399–412.

[20] J.K. Nielsen, M.G. Christensen, A.T. Cemgil, S.J. Godsill, S.J. Jensen, Bayesian interpolation and parameter estimation in a dynamic sinusoidal model, IEEE Trans. Audio Speech Language Process. 19 (7) (2011) 1986–1998.

[21] J.K. Nielsen, M.G. Christensen, S.H. Jensen, Bayesian model comparison and the BIC for regression models, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP '13), Vancouver, Canada, 2013, in press.

[22] H.B. Mann, A. Wald, On the choice of the number of class intervals in the application of the Chi Square test, Ann. Math. Stat. 13 (1942) 306–317.

[23] C.A. Williams Jr., On the choice of the number and width of classes for the Chi-square test of goodness of fit, J. Am. Stat. Assoc. 45 (1950) 77–86.

[24] F. Li, H. Liu, R.J. Vaccaro, Performance analysis for DOA estimation algorithms: Unification, simplification, and observations, IEEE Trans. Aerospace Electron. Syst. 29 (4) (1993) 1170–1184.

[25] J.-M. Papy, L. de Lathauwer, S. van Huffel, A shift invariancebased order-selection technique for exponential data modeling, IEEE Signal Process. Lett. 14 (2007) 473–476.

[26] M.G. Christensen, A. Jakobsson, S.H. Jensen, Sinusoidal order estimation using angles between subspaces, EURASIP J. Adv. Signal Process. 2009 (2009) 62.

[27] Z.D. Bai, P.R. Krishnaiah, L.C. Zhao, On rates of convergence of efficient detection criteria in signal processing with white noise, IEEE Trans. Inf. Theory 35 (2) (1989) 380–388.

[28] M.O. Ulfarsson, V. Solo, Rank selection in noist PCA with SURE and random matrix theory, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP '08), Nevada, USA, 2008, pp. 3317–3320.

Kefei Liu is currently a PhD candidate in the Department of Electronic Engineering, City University of Hong Kong, Hong Kong. He received the B.Sc. degree in applied mathematics from Wuhan University in 2006 and M.Sc. degree in mathematics from Beihang University in January 2009, both in China. From February 2012 to August 2012, he studied as a visiting PhD student at the Department of Electrical Engineering, University of Brasilia, Brazil. His research interests are statistical and array signal processing, with particular focus on source enumeration, prewhitening, and parameter estimation.

Hing Cheung So was born in Hong Kong. He obtained the B.Eng. degree from City University of Hong Kong and the Ph.D. degree from The Chinese University of Hong Kong, both in electronic engineering, in 1990 and 1995, respectively.

From 1990 to 1991, he was an Electronic Engineer at the Research and Development Division of Everex Systems Engineering Ltd., Hong Kong. During 1995–1996, he worked as a Post-Doctoral Fellow at The Chinese University of Hong Kong. From 1996 to 1999, he was a Research Assistant Professor at the Department of Electronic Engineering, City University of Hong Kong, where he is currently an Associate Professor. His research interests include statistical signal processing, fast and adaptive algorithms, signal detection, parameter estimation, and source localization. He has been on the editorial boards of IEEE Transactions on Signal Processing, Signal Processing, Digital Signal Processing and ISRN Applied Mathematics as well as a member in Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.

João Paulo Carvalho Lustosa da Costa received the Diploma degree in electronic engineering in 2003 from the Military Institute of Engineering (IME) in Rio de Janeiro, Brazil, his M.S. degree in 2006 from University of Brasília (UnB) in Brazil, and his Doktor-Ingenieur (Ph.D.) degree with Magna cum Laude in 2010 at Ilmenau University of Technology (TU Ilmenau) in Germany.

Currently, he is a professor at the Electrical Engineering Department, University of Brasília (UnB), and he cooperates with the Laboratory of Technologies for Decision Making (LATITUDE) supported by DELL computers of Brazil, with the Laboratory of Automation and Robotics (LARA), and with the Microwave and Wireless Systems Laboratory (MWSL). He coordinates the Laboratory of Array Signal Processing (LASP) at UnB. His research interests are in the areas of multi-dimensional array signal processing, model order selection, principal component analysis, MIMO communications systems, multilinear algebra and parameter estimation schemes.

Florian Römer studied computer engineering at the Ilmenau University of Technology, Germany, and McMaster University, Hamilton, ON, Canada. He received the Diplom-Ingenieur (M.Sc.) degree in communications engineering and the Dr.-Ing. (Ph.D.) degree in electrical engineering from the Ilmenau University of Technology in October 2006 and October 2012, respectively. From December 2006 until September 2012, he has been a Research Assistant in the Communications Research Laboratory at Ilmenau University of Technology. In October 2012 he has joined the Digital Broadcasting Research Group, a joint research activity between the Fraunhofer Institute of Integrated Circuits IIS and Ilmenau University of Technology, as a postdoctoral research fellow. His research interests include multi-dimensional signal processing, high-resolution parameter estimation, compressive sensing, as well as two-way relaying. Mr. Römer received the Siemens Communications Academic Award in 2006 for his diploma thesis and the best dissertation award from the Förder- und Freundeskreis (FFK) of the Ilmenau University of Technology for his dissertation.

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

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

He currently is an editorial board member of Digital Signal Processing.