Source Enumeration for High-Resolution Array Processing Using Improved Gerschgorin Radii Without Eigendecomposition

Lei Huang, *Member, IEEE*, Teng Long, *Member, IEEE*, and Shunjun Wu, *Member, IEEE*

***Abstract*—Accurate detection of sources with low complexity is of considerable interest in practical applications of high-resolution array processing. This paper addresses a new computationally efficient method for source enumeration by using enhanced Gerschgorin radii without eigendecomposition. The proposed method can calculate the Gerschgorin radii in a more efficient manner, in which the additive background noise can be efficiently suppressed and the computational complexity can be considerably reduced. Therefore, the method is more accurate and computationally attractive. Furthermore, the method does not rely on the eigenvalues of a covariance matrix or the signal/noise power, making it robust against deviations from the assumption of spatially white noise model. Numerical results are presented to demonstrate the performance of the method.**

***Index Terms*—Direction finding, eigenvalue decomposition (EVD), Gerschgorin radii, high resolution, minimum description length (MDL), multistage Wiener filter (MSWF), sensor array signal processing, signal enumeration.**

I. Introduction

HIGH-RESOLUTION methods for direction finding can be used in many areas, such as radar, sonar, wireless communications, biomedical engineering and so on [1], [2]. As is well known, the performance of the high-resolution methods, such as MUSIC [3] and ESPRIT [4], essentially relies on a priori knowledge of the number of sources. As a result, estimating the number of sources becomes an important issue [5]–[17]. In [5], Wax and Kailath originally addressed this problem by employing the Akaike information criterion (AIC) and minimum description length (MDL). As a consistent estimator, the MDL method has received more attention than the AIC estimator that tends to overestimate the number of sources. Following [5], a number of papers [6]–[14] have addressed the problem by enhancing the performance of the MDL method or by presenting the performance analysis of the MDL estimator. Recently, we addressed the issue of source enumeration by developing computationally efficient MDL methods [16], [17].

Manuscript received March 28, 2008; revised June 17, 2008. First published August 08, 2008; current version published November 19, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Olivier Besson. This work was supported by the Natural Science Foundation of China (NSFC) under Grant 60702068 and by the Guangdong Natural Science Foundation under Grant 7301317.

L. Huang and T. Long are with the Department of Electronic Engineering, Beijing Institute of Technology, Beijing 100081, China (e-mail: lhuang8sasp@hotmail.com).

S. Wu is with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: sjwu@xidian.edu.cn).

Digital Object Identifier 10.1109/TSP.2008.929331

While the classical MDL method has been extensively studied, its robustness and computational efficiency need to be further improved. As is well known, the classical MDL method is a model-dependent and eigenvalue-based estimator, in which the additive noise is assumed to be a spatially white process with equal power level. By minimizing the description length between the assumed model and the observed model characterized by the eigenvalues of a covariance matrix, the MDL method yields a consistent estimate for the number of sources. In some practical applications, however, the assumption of the spatially white noise might be unrealistic because the unknown noise environment may change slowly with time [12], [13], [18]. In these scenarios, the sensor noises become correlated from sensor to sensor and unequal in power level. While the sensor noises may be uncorrelated among all sensors in the array processing (in particular when the sparse arrays are employed), their power levels are in general unequal due to the nonidealities of the practical arrays, such as the nonideality of the receiving channel, the nonuniformity of the sensor response and the mutual coupling between sensors [12]. In the sequel, the sensor noise is a spatially inhomogeneous white process, i.e., of unequal power level and uncorrelated from sensor to sensor. When such a deviation from the spatially white noise model occurs, the multiplicity of the smallest eigenvalues equals one [5], [14] and the MDL estimator is thereby not necessarily consistent. On the other hand, the MDL method only employs the information of the eigenvalues of the covariance matrix, and does not employ any other information of the covariance matrix, such as the eigenvectors. Nevertheless, the unequal power levels of the sensor noises only perturb the eigenvalues and do not affect the eigenvectors. As a result, the MDL method is only robust to the spatially white noise, but is not robust to the spatially inhomogeneous noise. To improve the robustness of the classical MDL method, a number of methods have been addressed in [13]–[15]. In [13], Wu *et al.* developed a Gerschgorin disk estimator (GDE) for the number of sources that is more robust than the classical MDL method. Performing a unitary transformation of the covariance matrix by means of the eigenvectors, Wu *et al.* obtained the transformed Gerschgorin radii, and then used them to construct the GDE estimator to separate the signal Gerschgorin disks from the noise Gerschgorin disks. Nevertheless, while the GDE estimator is more robust against the deviation from the spatially white noise model than the MDL method, it does not outperform the MDL method in computational complexity since it, like the classical MDL method, necessarily involves the calculation of the covariance matrix and its eigenvalue-decomposition (EVD). Therefore, the computational complexities of the GDE and MDL methods need to be further reduced, in particular for the applications where the channel requires to be tracked in a real-time manner. Moreover, like the MDL estimator, as an EVD-based method, the GDE estimator also needs to use the covariance matrix whose diagonal elements are corrupted by the noise power levels. As a result, the EVD-based methods cannot efficiently eliminate the additive noise, leading to the poor performance for the case of small sample size. In this paper, we employ a spatial matched filter to calculate the Gerschogirin radii, which can significantly improve the signal-to-noise ratio (SNR) of the incident signals, equivalently efficiently suppressing the additive noise.

In the paper, motivated by the practical applications, we develop a new method for source enumeration that offers the computational simplicity and the robustness against the deviation from the spatially white noise assumption. To avoid the noise perturbation and reduce the computational complexity, we employ a procedure of multistage orthogonal projection similar to that of the multi-stage Wiener filter (MSWF) [19], [20] to calculate the Gerschgorin radii. More specifically, similar to the method in [17], the sensor data of an array is partitioned into a reference signal and an observation data of a successive refinement procedure to calculate the desired signals of the MSWF. Then, the cross-correlations between the desired signals of the adjacent stages are directly used to calculate the Gerschgorin radii instead of employing the eigenvectors of the covariance matrix. Finally, a heuristic decision rule, like the GDE method [13], is employed to yield the estimate of the number of sources. Since the proposed method can efficiently suppress the additive noise, the Gerschgorin radii can be estimated more accurately, thereby leading to the more accurate separation of the signal Gerschgorin disks from the noise Gerschgorin disks, which finally enhances the performance of the proposed method.

The remainder of the paper is organized as follows. The data model is presented in the next section. The method for source enumeration is proposed in Section III. Numerical results are presented in Section IV. Finally, conclusions are drawn in Section V.

II. Data Model

Consider an array of L sensors receiving p $(p < L-1)$ narrow-band sources $\{s_i(t)\}_{i=1}^{p}$ from distinct directions $\{\theta_i\}_{i=1}^{p}$. The p narrow-band sources are assumed to localize in the far field, and the wavefronts can be thereby approximated as planar. For simplicity, we also assume that the sources and the sensors are in the same plane. In the sequel, the sensor data of the array, excluding the last sensor data,[1] is collected as

$$\begin{aligned}\boldsymbol{x}(t) &\triangleq [x_1(t), x_2(t), \ldots, x_M(t)]^T\\ &= \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{n}(t),\ (t = 1, 2, \ldots, N)\end{aligned} \tag{1}$$

where $M = L-1$, $(\cdot)^T$ is the transpose operation and

$\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_p)]$ the $M \times p$ steering matrix
$\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_p(t)]^T$ the $p\times 1$ source waveform vector
$\boldsymbol{n}(t) = [n_1(t), n_2(t), \ldots, n_M(t)]^T$ the $M \times 1$ noise vector

[1]The last sensor data of the array is not included in (1) since, in the proposed method addressed in Section III, the first M sensor data and the last sensor data of the array are used as the observation data and, respectively, the reference signal of a successive refinement procedure, similar to the MSWF, to calculate the Gerschgorin radii.

in which $\boldsymbol{a}(\theta_i)$ $(i = 1, \ldots, p)$ is the steering vector and p represents the *unknown* number of sources. The source waveform, $s_i(t)$ $(i = 1, \ldots, p)$, is assumed to be a jointly stationary, statistically uncorrelated, zero-mean complex Gaussian random process. Meanwhile, the additive sensor noise $\boldsymbol{n}(t)$ is assumed to be an independent and identically distributed (i.i.d.) complex, zero-mean, Gaussian vector with covariance matrix $E\left[\boldsymbol{n}(t)\boldsymbol{n}^H(t)\right] = \sigma_n^2 \boldsymbol{I}_M$, where $E[\cdot]$ denotes expectation, σ_n^2 is the noise power and $\boldsymbol{I}_M$ denotes the $M \times M$ identity matrix. In addition, the noises $\boldsymbol{n}(t)$ are uncorrelated with the signals $\boldsymbol{s}(t)$. For a uniform linear array (ULA), the steering vector can be expressed as

$$\boldsymbol{a}(\theta_i) = \left[1, e^{j2\pi d \sin(\theta_i)/\lambda}, \ldots, e^{j2\pi d (M-1)\sin(\theta_i)/\lambda}\right]^T \tag{2}$$

where d represents the inter-sensor spacing and λ denotes the wavelength. It is easy to see that the steering vectors $\{\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_p)\}$ are linearly independent for any set of distinct incident angles $\{\theta_1, \theta_2, \ldots, \theta_p\}$, and the steering matrix $\boldsymbol{A}(\boldsymbol{\theta})$ is therefore full rank. In what follows, we assume that the array is a ULA for simplicity while the proposed method is not limited to this assumption.

Under these assumptions, the sensor data vector $\boldsymbol{x}(t)$ is a complex Gaussian random process with zero mean and the following covariance matrix

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

where $(\cdot)^H$ is the Hermitian transpose and $\boldsymbol{R}_s = E\left[\boldsymbol{s}(t)\boldsymbol{s}^H(t)\right]$ is the signal covariance matrix. The subspace spanned by the columns of $\boldsymbol{A}(\boldsymbol{\theta})$ is called signal subspace while its orthogonal complement is called noise subspace. In practical applications, however, only finite samples are available. As a result, the sample-covariance matrix is calculated as $\hat{\boldsymbol{R}}_x = (1/N)\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}^H(t)$, where N is finite.

Remark A: By examining (3), we can observe that the covariance matrix is corrupted by the noise power levels at the diagonal elements, thereby indicating that the EVD-based methods, such as the MDL method [5] and the GDE method [13], cannot efficiently suppress the noise when calculating the eigenvalues and the eigenvectors.

III. Source Enumeration Using Improved Gerschgorin Radii

To address the proposed method more clearly, we first briefly review the Gerschgorin's disk theorem that has been employed to estimate the number of sources by Wu *et al.* [13].

A. Gerschgorin's Disk Theorem

According to the Gerschgorin's disk theorem, the Gerschgorin center and the Gerschgorin radius of a matrix can be determined by the elements of the matrix. More specifically, for a complex matrix $\boldsymbol{R}_x = (r_{i,j})_{M\times M}$, the sum of the magnitudes of all elements of the ith row vector, excluding the ith element, is defined as

$$\gamma_i = \sum_{j=1, j\neq i}^{M} |r_{i,j}|, \quad (i = 1, \ldots, M). \tag{4}$$

In the sequel, the ith Gerschgorin disk $\boldsymbol{O}_i$ of the complex matrix $\boldsymbol{R}_x$ can be defined as the collection of points z in a complex plane:

$$|z - c_i| \leq \gamma_i, \quad (i = 1, \ldots, M) \tag{5}$$

where $c_i = r_{i,i}$ denotes the center of the ith disk. As a result, the center c_i and the radius γ_i are called the Gerschgorin center and, respectively, the Gerschgorin radius. It is shown in [21] that the eigenvalues of $\boldsymbol{R}_x$ are contained in the union of the disks $\boldsymbol{O}_i$. Furthermore, if a collection of m Gerschgorin disks of $\boldsymbol{R}_x$ is isolated from the other Gerschgorin disks, there exist exactly m eigenvalues of $\boldsymbol{R}_x$ contained in this collection. In the sequel, if the Gerschgorin disks are correctly collected in a union of the disks, we can estimate the number of sources from the collection. Nevertheless, as noted in [13], since all the Gerschgorin radii are in general large and the Gerschgorin centers are close to each other, the Gerschgorin disks of the original covariance matrix cannot be directly used to estimate the number of sources.

To cure this problem, Wu *et al.* [13] introduced a unitary transformation of the covariance matrix to separate the signal Gerschgorin disks from the noise Gerschgorin disks, and developed two estimators for source number, namely the Gerschgorin likelihood estimator (GLE) and GDE. Nevertheless, the calculation of the unitary transformation matrix necessarily includes the estimate of the covariance matrix and its EVD computation, thereby making the GLE and GDE methods quite computationally intensive. In the next subsection, we will propose a new method to calculate the Gerschgorin disks, which avoids the estimate of the covariance matrix and its EVD computation, thereby making the proposed method more computationally attractive than the EVD-based methods. Meanwhile, in the calculation of the Gerschgorin radii, the proposed method can efficiently suppress the additive noise. As a result, the proposed method can yield the more accurate estimate of the number of sources than the EVD-based methods.

B. Novel Method for Gerschgorin Disk Calculation

In this subsection, we present a new method to accurately calculate the Gerschgorin centers and the Gerschgorin radii. To begin with, let $\boldsymbol{x}_0(t) = \boldsymbol{x}(t)$ be the observation data and $d_0(t) = x_{M+1}(t)$ be the reference signal of a successive refinement procedure similar to the MSWF, where $x_{M+1}(t) = \boldsymbol{s}^T(t)\mathbf{1} + n_{M+1}(t)$ is the last sensor data of the array and $\mathbf{1} = [1, 1, \ldots, 1]^T$. In the sequel, the cross-correlation between the observation data and the reference signal can be calculated as

$$\boldsymbol{r}_{x_0 d_0} = E\left[\boldsymbol{x}_0(t) d_0^*(t)\right] = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_s\mathbf{1} \triangleq \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\beta} \tag{6}$$

where $(\cdot)^*$ is the complex conjugate and $\boldsymbol{\beta} = \boldsymbol{R}_s\mathbf{1}$. Considering $\boldsymbol{R}_s$ is a nonsingular matrix yields $\boldsymbol{\beta} \neq \mathbf{0}$. As a result, the cross-correlation $\boldsymbol{r}_{x_0 d_0}$ is a linear combination of all the direction vectors $\boldsymbol{a}(\theta_i)$ $(i = 1, 2, \ldots, p)$, which thereby implies that the cross-correlation $\boldsymbol{r}_{x_0 d_0}$ can capture the signal information. Meanwhile, it is shown in (6) that the additive noise has been efficiently suppressed in the procedure of calculating $\boldsymbol{r}_{x_0 d_0}$. Actually, the cross-correlation $\boldsymbol{r}_{x_0 d_0}$ is a spatial matched filter of the incident signals, which is plotted in Fig. 1. Therefore, when used as a matched filter to extract the desired signals from the observation data, the cross-correlation can help to improve the SNR of the desired signals. A normalized version of the spatial matched filter is defined as

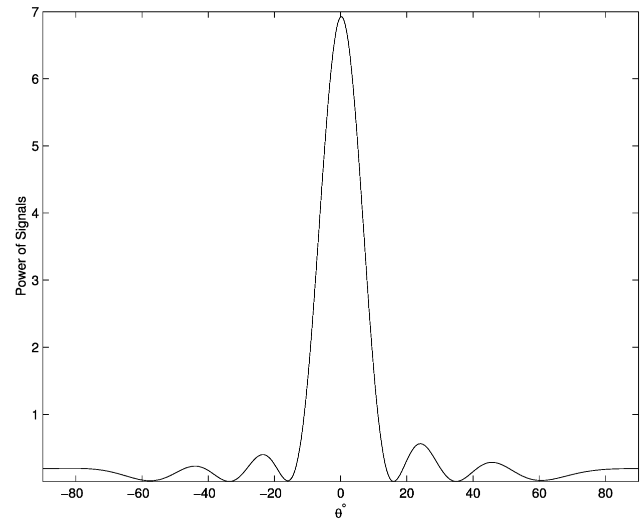

Fig. 1. Response of the spatial matched filter $\boldsymbol{r}_{x_0 d_0}$. The direction of a single incident signal is 0°, the number of sensors is 8, and SNR equals 0 dB.

$$\boldsymbol{h}_1 = \frac{\boldsymbol{r}_{x_0 d_0}}{\|\boldsymbol{r}_{x_0 d_0}\|} \tag{7}$$

where $\|\cdot\|$ denotes the vector norm. Partitioning the observation data $\boldsymbol{x}_0(t)$ with the matched filter $\boldsymbol{h}_1$ in a manner similar to that of the multistage Wiener filter (MSWF) [19], we attain the desired signal $d_i(t)$ and the observation data $\boldsymbol{x}_i(t)$ at the ith stage by

$$d_i(t) = \boldsymbol{h}_i^H \boldsymbol{x}_{i-1}(t) \tag{8}$$

and

$$\begin{aligned} \boldsymbol{x}_i(t) &= \boldsymbol{x}_{i-1}(t) - \boldsymbol{h}_i d_i(t) \\ &= \boldsymbol{x}_{i-1}(t) - \boldsymbol{h}_i \boldsymbol{h}_i^H \boldsymbol{x}_{i-1}(t) \\ &= \boldsymbol{B}_i \boldsymbol{x}_{i-1}(t) \end{aligned} \tag{9}$$

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

$$\boldsymbol{h}_i = \frac{E\left[\boldsymbol{x}_{i-1}(t) d_{i-1}^*(t)\right]}{\left\|E\left[\boldsymbol{x}_{i-1}(t) d_{i-1}^*(t)\right]\right\|}. \tag{10}$$

The refinement procedure (8)–(10) indicates that the desired signal $d_i(t)$ is yielded by filtering the observation data $\boldsymbol{x}_{i-1}(t)$ with the matched filters $\boldsymbol{h}_i$, but annihilated in (9). The observation data is partitioned stage-by-stage in the same refinement manner. As a result, after performing M successive recursions, we obtain M desired signals of the MSWF, which can be collected as

$$\boldsymbol{d}(t) \triangleq [d_1(t), d_2(t), \ldots, d_M(t)]^T = \boldsymbol{H}^H \boldsymbol{x}_0(t) \tag{11}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_M]$. From (8)–(10), it is easy to prove that the matched filters are orthogonal to each other, i.e., $\boldsymbol{h}_i^H \boldsymbol{h}_j = 0$ $(i, j = 1, \ldots, M,\ i \neq j)$. The proof is given in Appendix A. As a result, $\boldsymbol{H}$ is a unitary matrix.

Lemma 1: For an $M \times M$ Hermitian matrix $\boldsymbol{R}_{x_0} = E\left[\boldsymbol{x}_0(t)\boldsymbol{x}_0^H(t)\right]$, i.e., $\boldsymbol{R}_{x_0} = \boldsymbol{R}_{x_0}^H$, the covariance matrix of the desired signals $\boldsymbol{d}(t) = \boldsymbol{H}^H\boldsymbol{x}_0(t)$, i.e.,

$$\boldsymbol{R}_d = E\left[\boldsymbol{d}(t)\boldsymbol{d}^H(t)\right] = \boldsymbol{H}^H\boldsymbol{R}_{x_0}\boldsymbol{H} \tag{12}$$

is an Hermitian tridiagonal matrix, and can be explicitly expressed as

$$\boldsymbol{R}_d = \begin{pmatrix} \sigma_{d_1}^2 & \delta_{1,2} & & & & & & \\ \delta_{1,2}^* & \sigma_{d_2}^2 & \delta_{2,3} & & & & & \\ & \delta_{2,3}^* & \sigma_{d_3}^2 & \ddots & & & & \\ & & \ddots & \ddots & \delta_{p-1,p} & & & \\ & & & \delta_{p-1,p}^* & \sigma_{d_p}^2 & 0 & & \\ & & & & 0 & \sigma_n^2 & \ddots & \\ & & & & & \ddots & \ddots & 0 \\ & & & & & & 0 & \sigma_n^2 \end{pmatrix} \tag{13}$$

where $\sigma_{d_i}^2 \triangleq E\left[|d_i(t)|_{\text{abs}}^2\right]$, $\delta_{i,i+1} \triangleq E\left[d_i(t)d_{i+1}^*(t)\right]$, $\sigma_{d_{p+1}}^2 = \cdots = \sigma_{d_M}^2 = \sigma_n^2$, $\delta_{p,p+1} = \cdots = \delta_{M-1,M} = 0$ and $|\cdot|_{\text{abs}}$ denotes the absolute value.

Proof: The proof of Lemma 1 is seen in Appendix B. ■

According to the Gerschgorin's disk theorem, the Gerschgorin centers c_i and Gerschgorin radii γ_i of the tridiagonal matrix $\boldsymbol{R}_d$ can be calculated from (13) as

$$c_i = \sigma_{d_i}^2 \quad (i = 1, \ldots, M) \tag{14}$$

$$\gamma_i = |\delta_{i-1,i}^*|_{\text{abs}} + |\delta_{i,i+1}|_{\text{abs}} \quad (i = 1, \ldots, M) \tag{15}$$

where $\delta_{0,1} \triangleq 0$ and $\delta_{M,M+1} \triangleq 0$. Note in (13) that $\delta_{i,i+1} \neq 0\ (i = 1, \ldots, p-1)$ and $\delta_{i,i+1} = 0\ (i = p, \ldots, M-1)$. As a result, it follows from (15) that $\gamma_i \neq 0\ (i = 1, \ldots, p)$ and $\gamma_i = 0\ (i = p+1, \ldots, M)$.

Remark B: As noted above, the cross-correlation $\boldsymbol{r}_{x_0d_0}$ is the spatial matched filter of the incident signals, which can capture the signal information and efficiently suppress the additive noise, thereby significantly improving the SNR of the incident signals that are within the beamwidth of the matched filter (i.e., the benefit of the beamformer). In the sequel, when $\boldsymbol{r}_{x_0d_0}$ is employed as the initial information for the refinement procedure to extract the desired signals from the observation data, the desired signals can be yielded more accurately, leading to the more accurate estimates of the correlations $\sigma_{d_i}^2$ and the cross-correlations $\delta_{i,i+1}$. Meanwhile, it is shown in (30) that the noise term (i.e., $\sigma_n^2\boldsymbol{h}_{i+1}^H\boldsymbol{h}_i$) is equal to zero due to the orthogonality of the matched filters, implying that the background noise can be further suppressed in the calculation of $\delta_{i,i+1}$. Consequently, it follows from (14) and (15) that the Gerschgorin centers and radii are more accurate than that computed from the covariance matrix in the EVD-based GDE method [13], in particular when all the incident signals are within the beamwidth of the matched filter.

C. Numerical Example

For a scenario where SNR equals 5 dB, the sample size is 128 and there exist two sources impinging upon a ULA of 10 elements from distinct incident directions {0.5°, 6.2°}, the calculated original covariance matrix is given by $\hat{\boldsymbol{R}}_{x_0} = (1/128)\sum_{t=1}^{128}\boldsymbol{x}_0(t)\boldsymbol{x}_0^H(t)$. According to (4) and (5), the Gerschgorin centers and radii can be easily obtained, and the Gerschgorin disks are plotted in Fig. 2. It can be observed in Fig. 2 that all the Gerschgorin radii are large and the Gerschgorin centers are very close to each other, which indicates that the Gerschgorin radii cannot make any sense yet to estimate the number of sources.

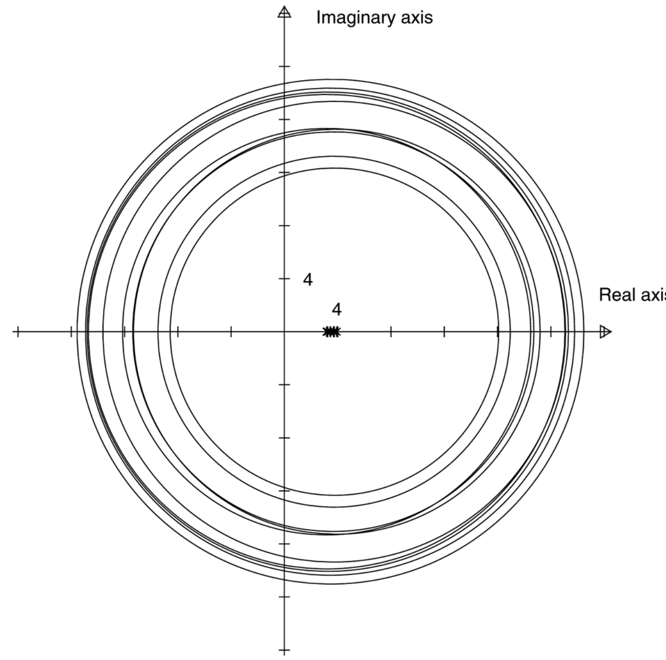

Fig. 2. Gerschgorin disks of the calculated original covariance matrix. $*$ denotes the centers of the Gerschgorin disks.

After filtered by $\hat{\boldsymbol{H}}$, however, the original covariance matrix $\hat{\boldsymbol{R}}_{x_0}$ is transformed to a tridiagonal matrix $\hat{\boldsymbol{R}}_d$, shown at the bottom of the next page. Here $\hat{\boldsymbol{H}} = [\hat{\boldsymbol{h}}_1, \ldots, \hat{\boldsymbol{h}}_9]$ and $\hat{\boldsymbol{h}}_i = \sum_{t=1}^{128}\boldsymbol{x}_{i-1}(t)d_{i-1}^*(t)/\|\sum_{t=1}^{128}\boldsymbol{x}_{i-1}(t)d_{i-1}^*(t)\|$. In the sequel, using (14) and (15) we can obtain the centers and radii of the improved Gerschgorin disks, which are given in Table I. The corresponding Gerschgorin disks are plotted in Fig. 3. From Table I and Fig. 3, we can observe that the enhanced Gerschgorin disks offer the quite large signal Gerschgorin radii and very small noise Gerschgorin radii, which can make more sense to estimate the number of sources.

D. Reduced-Rank Gerschgorin Disk Estimator for Source Number

Denoting the eigendecomposition of the covariance matrix $\boldsymbol{R}_{x_0}$ as

$$\boldsymbol{R}_{x_0} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H \tag{16}$$

where $\boldsymbol{D} = \text{diag}([\lambda_1, \ldots, \lambda_M])$ in which $\lambda_1 \geq \cdots \geq \lambda_{p+1} = \cdots = \lambda_M = \sigma_n^2$ are the eigenvalues and $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M]$ in which $\boldsymbol{v}_i\ (i = 1, \ldots, M)$ are the corresponding eigenvectors, and substituting (16) into (12) yields

$$\boldsymbol{R}_d = \boldsymbol{H}^H\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H\boldsymbol{H} \triangleq \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^H \tag{17}$$

where $\boldsymbol{U} = \boldsymbol{H}^H\boldsymbol{V}$. Obviously, $\boldsymbol{U}$ is also a unitary matrix. As a result, it follows from (16) and (17) that $\boldsymbol{R}_{x_0}$ and $\boldsymbol{R}_d$ share the

TABLE I
GRESCHGORIN CENTERS AND RADII OF THE CALCULATED TRIDIAGONAL MATRIX

disks	$\boldsymbol{O}_1$	$\boldsymbol{O}_2$	$\boldsymbol{O}_3$	$\boldsymbol{O}_4$	$\boldsymbol{O}_5$	$\boldsymbol{O}_6$	$\boldsymbol{O}_7$	$\boldsymbol{O}_8$	$\boldsymbol{O}_9$
c_i	14.8936	9.5463	0.8974	1.0899	1.0970	1.0340	1.1053	0.8911	1.0158
γ_i	5.8824	6.4617	0.7420	0.3788	0.4421	0.4057	0.4792	0.3135	0.0139

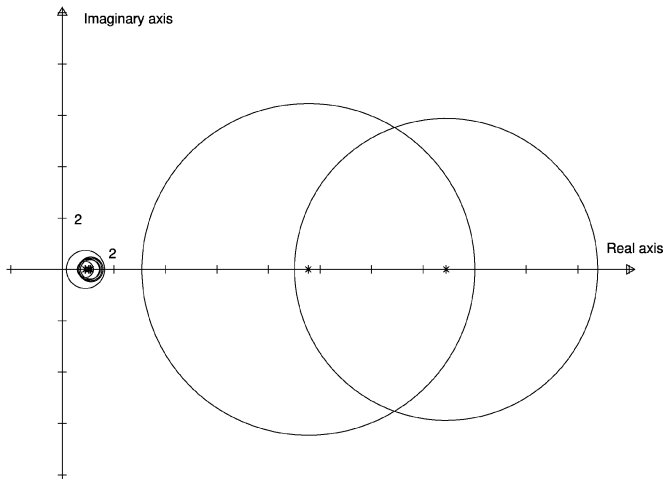

Fig. 3. Gerschgorin disks of the calculated tridiagonal matrix. $*$ denotes the centers of the Gerschgorin disks.

same eigenvalues (i.e., $\boldsymbol{D}$). Using the Gerschgorin's disk theorem discussed above, it follows, therefore, that the eigenvalues of $\boldsymbol{R}_{x_0}$ (i.e., $\lambda_1, \ldots, \lambda_M$) are contained in the collections of the Gerschgorin disks that are determined by the tridiagonal matrix $\boldsymbol{R}_d$. In other word, all the eigenvalues of $\boldsymbol{R}_{x_0}$ must be contained in the collections of the Gerschgorin disks with the centers c_i and the radii γ_i. Meanwhile, it should be noted that the eigenvalues of $\boldsymbol{R}_{x_0}$ and the diagonal elements of $\boldsymbol{R}_d$ (i.e., $\sigma_{d_i}^2$) are real. Therefore, we can estimate the locations of the eigenvalues along the real axis.

When the sample size tends to infinity, we obtain that $\gamma_i \neq 0\ (i = 1, \ldots, p)$ and $\gamma_i = 0\ (i = p+1, \ldots, M)$. Meanwhile, notice that $\lambda_i = \sigma_n^2\ (i = p+1, \ldots, M)$. Consequently, it follows that the $(M-p)$ smallest eigenvalues are exactly contained in the collection of the Gerschgorin disks centered around $c_i = \sigma_n^2\ (i = p+1, \ldots, M)$ with the zero radii $\gamma_i = 0\ (i = p+1, \ldots, M)$. Since all the eigenvalues of $\boldsymbol{R}_{x_0}$ must be contained in the collections of the Gerschgorin disks determined by $\boldsymbol{R}_d$, the remainder p largest eigenvalues must be contained in the collection of the Gerschgorin disks centered around $c_i = \sigma_{d_i}^2\ (i = 1, \ldots, p)$ with the nonzero radii $\gamma_i\ (i = 1, \ldots, p)$. The disks with zero radii (i.e., $\boldsymbol{O}_{p+1}, \boldsymbol{O}_{p+2}, \ldots, \boldsymbol{O}_M$) can be regarded as the noise Gerschgorin disks whereas the disks with nonzero radii (i.e., $\boldsymbol{O}_1, \boldsymbol{O}_2, \ldots, \boldsymbol{O}_p$) can be regarded as the signal Gerschgorin disks. Therefore, the number of sources can be determined by counting the number of nonzero Gerschgorin radii. If only finite samples are available, however, the noise Gerschgorin radii do not equal zero. In this scenario, we need to use a decision rule to determine the number of sources. Employing a decision rule similar to that of [13], a Gerschgorin disk estimator for source number without eigendecompositon, called GDEWE, can be defined as

$$\mathrm{GDEWE}(k) = \gamma_k - \frac{D(N)}{M} \sum_{i=1}^{M} \gamma_i, \quad k = 1, 2, \ldots, M \tag{18}$$

where $D(N)$ is an adjustable factor that can be selected as $D(N) = \kappa / \log(N)$ in which κ is a positive number generally not greater than $\log(N)$. It is easy to see that $D(N)$ is a nonincreasing function as the sample size increases. By detecting the first nonpositive value of $\mathrm{GDEWE}(k)$, we can obtain the estimate of the number of sources as $\hat{p} = k - 1$.

By examining (13) and (15), we can see that the last $(M-p)$ Gerschgorin disks are of zero Gerschgorin radius. In the sequel, the reduced desired signals, say $\boldsymbol{d}_J(t) = [d_1(t), d_2(t), \ldots, d_J(t)]^T\ (p < J < M)$, are enough to correctly enumerate the sources. Here J represents the dimension of the reduced-rank observation space. Therefore, a reduced-rank GDEWE estimator can be defined as

$$\mathrm{GDEWE}^{(J)}(k) = \gamma_k - \frac{D(N)}{J} \sum_{i=1}^{J} \gamma_i, \quad k = 1, 2, \ldots, J. \tag{19}$$

$$\hat{\boldsymbol{R}}_d = \begin{pmatrix} 14.8936 & 5.8824 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5.8824 & 9.5463 & 0.5792 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5792 & 0.8974 & 0.1628 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1628 & 1.0899 & 0.2160 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2160 & 1.0970 & 0.2261 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2261 & 1.0340 & 0.1796 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1796 & 1.1053 & 0.2996 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2996 & 0.8911 & 0.0139 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0139 & 1.0158 \end{pmatrix}$$

The number of sources can be estimated as $\hat{p} = k - 1$ when the first nonpositive value of GDEWE$^{(J)}(k)$ is detected.

Remark C: It is indicated in (15) that the Gerschgorin radii can be directly computed in the refinement procedure [i.e., (8)–(10)], avoiding the estimate of the covariance matrix and its EVD computation. Meanwhile, it is shown in (8)–(10) that the refinement procedure only involves complex vector-vector products, and does not include any complex matrix-vector products, thereby only requiring around $O(M)$ flops for each snapshot and each stage. In the sequel, the computational cost of the full-rank GDEWE estimator is about $O(M^2N)$ flops, and the computational cost of the reduced-rank GDEWE estimator can be reduced to around $O(JMN)$ flops. However, the EVD-based methods, such as the GDE method of [13], necessarily involve the estimate of the covariance matrix and its EVD computation, which require around $O(M^2N) + O(M^3)$ flops. In addition, to calculate the transformed Gerschgorin radii, the GDE method still needs to use the projection of the last column of the estimated covariance matrix onto the unitary matrix consisting of the eigenvectors, requiring additional (MN) flops. Therefore, the proposed method is more computationally attractive than the EVD-based methods, in particular when the sensor array is large.

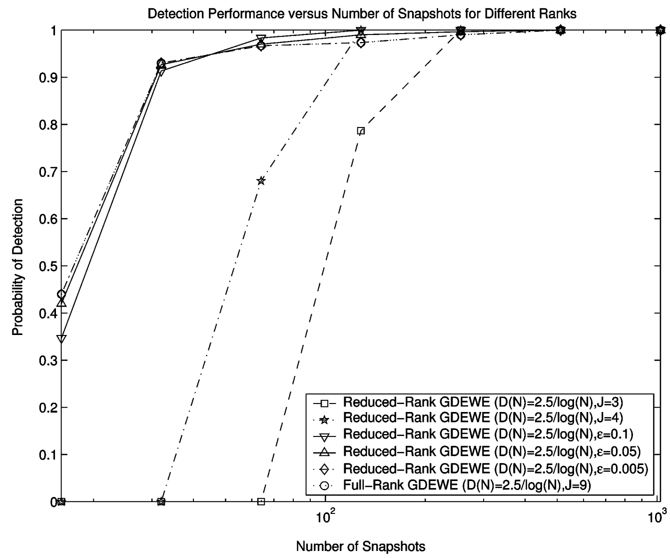

Fig. 4. Probability of correct detection of the GDEWE method versus number of snapshots for different ranks in the case of spatially white noise. $[\theta_1, \theta_2] = [0.5°, 6.2°]$, SNR $= 0$ dB, $L = 10$, and N varies from 16 to 1024.

E. Rank J Adaptation

It follows from (19) that the performance of the reduced-rank GDEWE method relies on the dimension of the reduced-rank observation space (i.e., J). When J is less than the true number of sources (i.e., p), the reduced-rank GDEWE cannot correctly detect the sources while it may be computationally attractive. When $J \approx M$, however, the reduced-rank GDEWE method does can detect the number of sources but might be not computationally simple any more provided that $p \ll M$. Therefore, the correct selection of J is also important for the reduced-rank GDEWE method.

Note that, in the practical applications, $|\hat{\delta}_{i,i+1}|$ $(i = p, \ldots, M-1)$ are small number unequal to zero due to the finite sample size N while $|\hat{\delta}_{i,i+1}|$ $(i = 1, \ldots, p-1)$ are generally greater than one. In the sequel, to make the dimension of the reduced-rank observation space adaptively adjustable, we define the detector of J similar to that in [17] as

$$J = \max\left\{i : |\hat{\delta}_{i,i+1}|_{\text{abs}} > \epsilon\right\} \tag{20}$$

where ϵ is a small positive constant.

IV. Numerical Results

The performance of the proposed GDEWE estimator is evaluated by computer simulation in this section. For fair comparison, the empirical results of the GDE method developed by Wu *et al.* [13], the traditional MDL estimator proposed by Wax and Kailath [5] and the reduced-rank MDL approach proposed by Huang *et al.* [17] are also presented. We consider a ULA with ten sensors whose spacings equal half-wavelength. There exist two uncorrelated signals in the far field with equal power impinging upon the ULA from the directions $\{0.5°, 6.2°\}$. The background noise is assumed to be a stationary Gaussian random process, which is uncorrelated with the incident signals.

We first consider the case of spatially white noise. To calculate the probability of correct detection, three hundred Monte Carlo trials have been run. Fig. 4 depicts the detection performance of the GDEWE method as a function of the number of snapshots for different ranks. In this case, the SNR is defined as the ratio of the power of signals to the power of noise at each sensor. It can be observed that the full-rank GDEWE method (i.e., $J = M = 9$) offers the best detection performance among the cases, whereas when $D(N) = 2.5/\log(N)$ and $\epsilon = 0.1$ the adaptive reduced-rank GDEWE method is more robust than the full-rank GDEWE method. More specifically, the adaptive reduced-rank GDEWE method surpasses the full-rank GDEWE method for a moderate sample size, say $N = 100$ and is less accurate than the latter when the number of snapshots becomes smaller than 32. It can also be observed that the reduced-rank GDEWE method becomes less accurate when J is fixed to 3 or 4. This is due to the fact that when the sample size becomes small, the noise Gerschgorin radii are very close to the signal Gerschgorin radii, thereby increasing the possibility of error detection of the reduced-rank GDEWE method, in particular when the rank J is very close to the true number of sources. The detection performance of the reduced-rank GDEWE method versus the number of snapshots for the different κ in $D(N)$ is illustrated in Fig. 5. It is indicated in Fig. 5 that the larger the κ is, the earlier the reduced-rank GDEWE method reaches 1. Nevertheless, as κ increases, the reduced-rank GDEWE method becomes less accurate for small sample size. As a result, κ should be selected as 2.5.

The empirical results of the reduced-rank GDEWE method, the GDE method, the reduced-rank MDL approach and the traditional MDL estimator varying with the number of snapshots are depicted in Fig. 6. In the reduced-rank GDEWE method and the reduced-rank MDL method, the threshold ϵ is set to 0.1 to determine the dimension of the reduced-rank observation space (i.e., J). Meanwhile, for comparison, the adaptively adjustable factors of the reduced-rank GDEWE method and the GDE method [13] are set as $D(N) = 2.5/\log(N)$. In addition, note that the MDL method [5] needs to use the observation data of 10 sensors (i.e., $L = 10$), whereas the reduced-rank GDEWE method and the reduced-rank MDL method only use the observation data of 9 sensors (i.e., $M = 9$). It is shown in Fig. 6 that all the methods approach to 1 when the number of snapshots

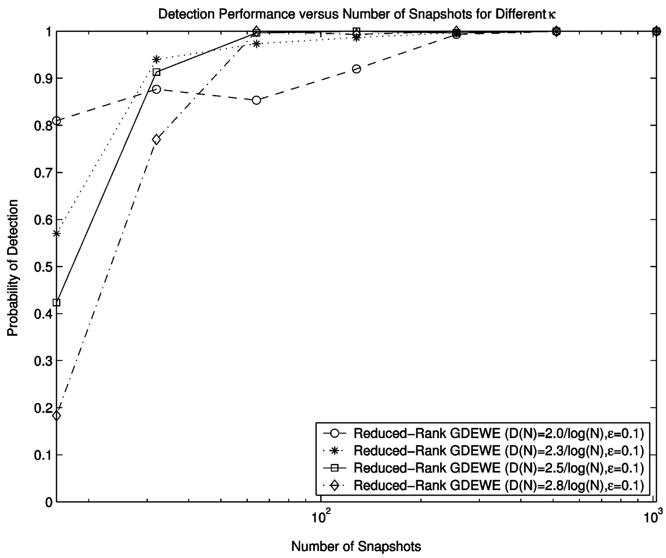

Fig. 5. Probability of correct detection of the GDEWE versus number of snapshots for different κ in the case of spatially white noise. $[\theta_1,\ \theta_2] = [0.5^\circ,\ 6.2^\circ]$, SNR $= 0$ dB, $L = 10$, and N varies from 16 to 1024.

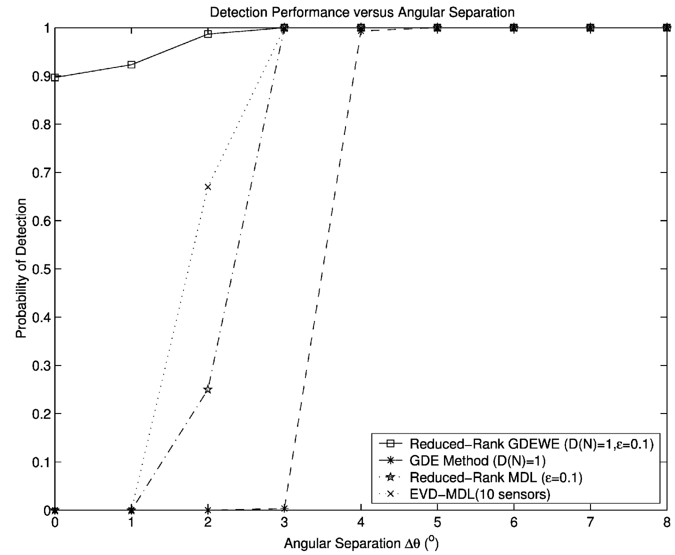

Fig. 7. Probability of correct detection versus angular separation $(\Delta\theta)$ for the case of spatially white noise. $[\theta_1, \theta_2] = [0.5^\circ, \theta_1 + \Delta\theta]$, SNR $= 5$ dB, $N = 256$, $L = 10$, and the angular separation varies from 0° to 0°.

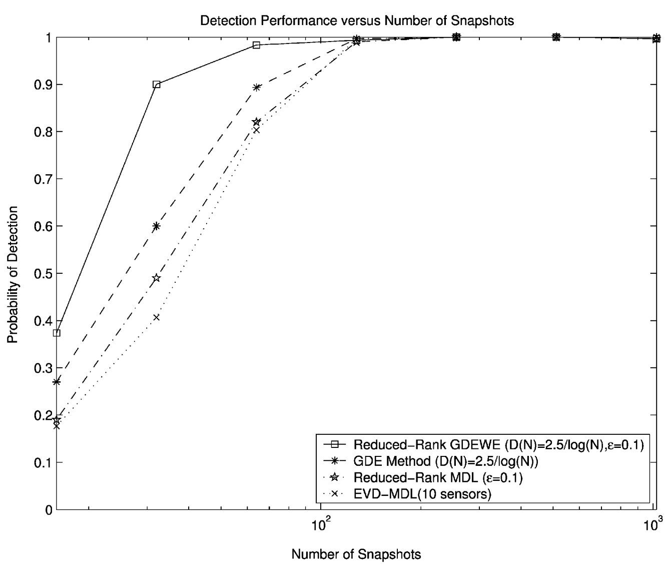

Fig. 6. Probability of correct detection versus number of snapshots for the case of spatially white noise. $[\theta_1, \theta_2] = [0.5^\circ, 6.2^\circ]$, SNR $= 0$ dB, $L = 10$, and N varies from 16 to 1024.

tends to infinity. This thereby indicates that all the methods are robust to the spatially white noise. Meanwhile, from Fig. 6 we can observe that the reduced-rank GDEWE method surpasses the other three methods in detection performance, in particular when the number of snapshots becomes small.

To study the angular separation between sources needed for reliable detection, we assume that the directions of the two sources are given as $[\theta_1, \theta_2] = [0.5^\circ, \theta_1 + \Delta\theta]$ in which $\Delta\theta$ denotes the angular separation. Fig. 7 depicts the empirical probabilities of correct detection versus the angular separation, in which the number of snapshots is 256, SNR equals 5 dB, and the angular separation varies from 0° to 8°. Since the number of snapshots is fixed, the adaptively adjustable factor is set as $D(N) = 1$ for the GDEWE and GDE methods. Fig. 7 indicates that the GDEWE method is more accurate than the GDE method, especially when the angular separation is less than 4°. An interesting observation herein is that the reduced-rank GDEWE method can successfully enumerate the spatially close sources. This is due to the fact that in this case the spatial matched filter can significantly enhance the SNR of the incident signals that are within the beamwidth of the matched filter, and thereby can considerably reduce the possibility of error detection of the GDEWE method.

Consider now the scenario of spatially inhomogeneous noise. In this scenario, we assume that the sensor noises are the spatially inhomogeneous white Gaussian processes, whose power levels are subject to a uniform distribution over the interval $[\sigma^2_{\min}, \sigma^2_{\max}]$, given as

$$\begin{aligned}&[\sigma_1^2, \ldots, \sigma_{10}^2]\\ &\quad = [\,5.0 \quad 0.6 \quad 1.9 \quad 0.6 \quad 2.2 \quad 3.6 \quad 0.9 \quad 0.7 \quad 3.3 \quad 0.5\,].\end{aligned}$$

In the sequel, exploiting the definition of the worst noise power ratio (WNPR) in [18], we obtain WNPR $\triangleq \sigma^2_{\max}/\sigma^2_{\min} = 10$ (Note that WNPR $= 1$ represents the scenario of spatially white noise.) Similarly, to calculate the empirical probabilities of correct detection of the four methods, three hundred Monte Carlo trials have been run. The thresholds of the reduced-rank GDEWE method and the reduced-rank MDL method are set as $\epsilon = 0.5$ to determine the dimension of the reduced-rank observation space. Meanwhile, the adaptively adjustable factor is still set as $2.5/\log(N)$ for the GDEWE and GDE estimators.

Fig. 8 depicts the empirical results of the four methods versus the number of snapshots in the case of spatially inhomogeneous noise. In this case, the SNR is defined as the power of signals to the *average* power of noises. From Fig. 8, we can observe that the probabilities of correction detection of the classical EVD-based MDL method and the reduced-rank MDL method converge to zero as the number of snapshots increases. It is easy to interpret this phenomenon. Note that the EVD-based MDL method is based on the equality of the smallest eigenvalues, and the reduced-rank MDL method is based on the equality of the variances of the last $(M - p)$ desired signals of the MSWF. On the other hand, note that the unequal power levels of the sensor noises lead to the differences in the smallest eigenvalues and

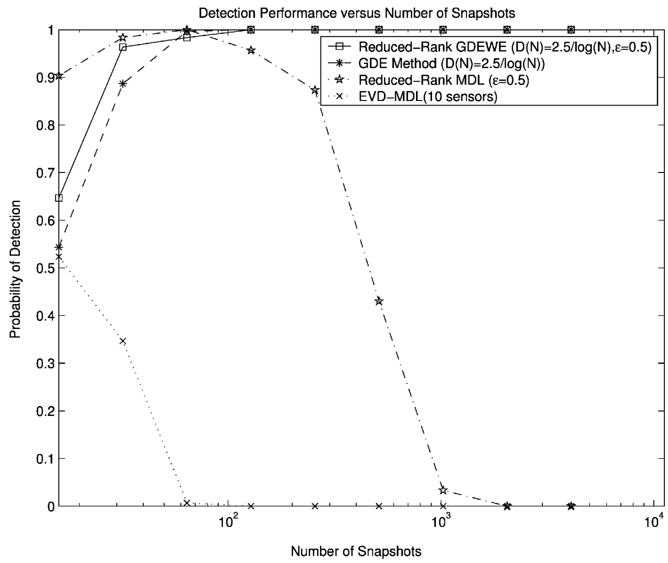

Fig. 8. Probability of correct detection versus number of snapshots in the case of spatially inhomogeneous noise. $[\theta_1, \theta_2] = [0.5°, 6.2°]$, SNR $= 5$ dB, WNPR $= 10$, $L = 10$, and N varies from 16 to 4096.

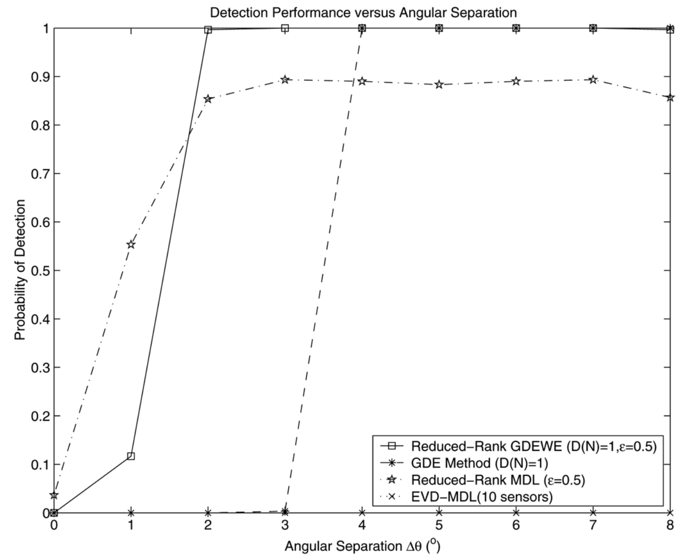

Fig. 9. Probability of correct detection versus angular separation in the case of spatially inhomogeneous noise. $[\theta_1, \theta_2] = [0.5°, \theta_1 + \Delta\theta]$, SNR $= 10$ dB, $N = 256$, WNPR $= 10$, $L = 10$, and $\Delta\theta$ varies from $0°$ to $8°$.

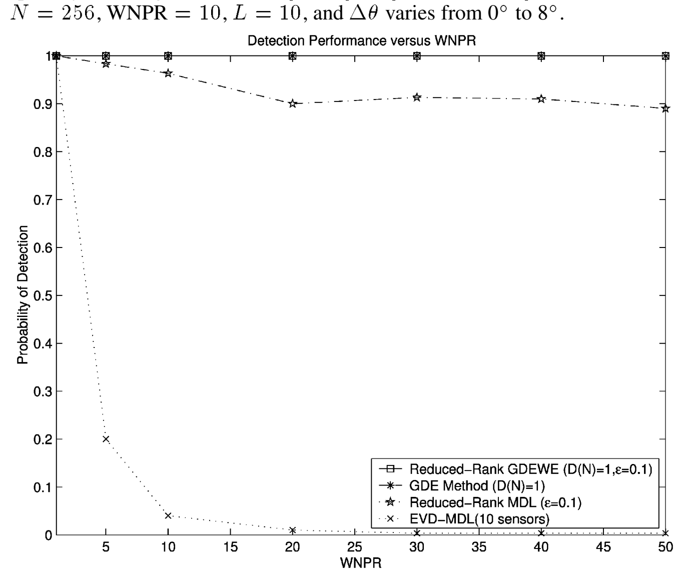

Fig. 10. Probability of correct detection versus WNPR. $[\theta_1, \theta_2] = [0.5°, 6.2°]$, SNR $= 5$ dB, $N = 256$, $L = 10$, and WNPR varies from 1 to 50.

also result in the differences in the variances of the last $(M-p)$ desired signals of the MSWF. This thereby causes more sources (namely the so-called "virtual" sources addressed in [14]) to be detected in the EVD-based MDL and reduced-rank MDL methods. When the number of snapshots is not large enough, the EVD-based MDL method cannot detect the differences in the smallest eigenvalues, and the reduced-rank MDL method cannot detect the differences in the variances of the last $(M-p)$ desired signals either. As a result, neither the EVD-based MDL method nor the reduced-rank MDL method can detect the "virtual" sources resulted from the differences. As the number of snapshots becomes large enough, however, both of them detect the "virtual" sources as valid sources, thereby leading to an error event. Consequently, neither the EVD-based MDL method nor the reduced-rank MDL method are robust to the spatially inhomogeneous noise. Nevertheless, the GDE method can successfully enumerate the sources even when the number of snapshots becomes large enough because it is only based on the eigenvectors, and independent of the eigenvalues. Also, the reduced-rank GDEWE method can correctly estimate the number of sources since it only relies on the cross-correlations of the desired signals of the MWSF, and does not need the eigenvalues or the signal/noise power levels. Furthermore, Fig. 8 implies that the reduced-rank GDEWE method outperforms the GDE method in detection performance, particularly when the number of snapshots is less than 64. As noted previously, in the calculation of the Gerschgorin radii, the GDEWE method can efficiently eliminate the additive noise and capture the information of the incident signals. As a result, the Gerschgorin radii are more accurate than that derived from the eigenvectors in [13]. This eventually leads to the improved performance of the reduced-rank GDEWE method.

Fig. 9 shows the empirical probabilities of correct detection of the four methods varying with the angular separation. Again, since there is no variation in the number of snapshots, the adaptively adjustable factors of the GDEWE and GDE estimators are set to 1. As an eigenvalue-based method, the classical MDL method fails to correctly estimate the number of sources. The reduced-rank MDL method cannot successfully enumerate the sources either since it relies on the assumption of spatially white noise. However, since the reduced-rank GDEWE method and the GDE method are robust to the spatially inhomogeneous noise, they can correctly enumerate the sources in this case. Meanwhile, the reduced-rank GDEWE method surpasses the GDE method in detection performance, in particular when the angular separation becomes smaller than $4°$, as indicated in Fig. 9. Therefore, the GDEWE method is more accurate than the GDE method, especially when all the incident signals are within the beamwidth of the spatial matched filter.

To study the robustness of the methods against different WNPRs, we calculate their probabilities of correct detection versus the WNPR, which are shown in Fig. 10. Since the GDE and GDEWE methods are robust to the spatially white noise and the spatially inhomogeneous noise, they can correctly estimate the number of sources when the WNPR varies from 1 to 50, as is indicated in Fig. 10. However, as an eigenvalue-based method, the classical MDL method can merely detect the signals for WNPR $= 1$ that represents the case of the spatially white noise, but cannot correctly estimate the number of sources when WNPR becomes greater than 1, which represents the case of the spatially inhomogeneous noise. Similarly, since the reduced-rank MDL estimator is also only robust to the spatially white noise, it cannot correctly enumerate the sources as WNPR > 1.

V. Conclusion

In this paper, we have developed a method for source enumeration by using the improved Gerschgorin radii without eigendecomposition. In the proposed method, the Gerschgorin radii can be directly calculated in the recursion procedure of the MSWF, avoiding the estimate of the covariance matrix and its EVD computation, and thereby making the proposed method a favorable choice for the practical implementations. Meanwhile, the proposed method can efficiently suppress the additive noise and capture the signal information when calculating the Gerschgorin radii, thereby leading to the more accurate separation of the signal Gerschgorin disks from the noise Gerschgorin disks. This results in the more accurate estimate of the number of sources for the proposed method.

Appendix A
Proof of the Orthogonality of the Matched Filters

Substituting (8) and (9) into (10) and recalling that $\boldsymbol{B}_i = \boldsymbol{I}_M - \boldsymbol{h}_i\boldsymbol{h}_i^H$, we obtain

$$\boldsymbol{h}_{i+1} = \frac{\left(\boldsymbol{I}_M - \boldsymbol{h}_i\boldsymbol{h}_i^H\right)\boldsymbol{R}_{x_{i-1}}\boldsymbol{h}_i}{\left\|\left(\boldsymbol{I}_M - \boldsymbol{h}_i\boldsymbol{h}_i^H\right)\boldsymbol{R}_{x_{i-1}}\boldsymbol{h}_i\right\|} \tag{21}$$

where

$$\begin{aligned}\boldsymbol{R}_{x_{i-1}} &= E\left[\boldsymbol{x}_{i-1}(t)\boldsymbol{x}_{i-1}^H(t)\right]\\ &= \boldsymbol{B}_{i-1}\boldsymbol{R}_{x_{i-2}}\boldsymbol{B}_{i-1}^H = \cdots\\ &= \left(\prod_{k=i-1}^{1}\boldsymbol{B}_k\right)\boldsymbol{R}_{x_0}\left(\prod_{k=i-1}^{1}\boldsymbol{B}_k^H\right)\\ &= \left(\boldsymbol{I}_M - \sum_{k=1}^{i-1}\boldsymbol{h}_k\boldsymbol{h}_k^H\right)\boldsymbol{R}_{x_0}\left(\boldsymbol{I}_M - \sum_{k=1}^{i-1}\boldsymbol{h}_k\boldsymbol{h}_k^H\right).\end{aligned} \tag{22}$$

To prove the orthogonality of $\boldsymbol{h}_i$ $(i = 1, 2, \ldots, M)$, we employ the following induction argument. First, it is easy to verify from (21) that $\boldsymbol{h}_2$ is orthogonal to $\boldsymbol{h}_1$. Suppose, now, that $\boldsymbol{h}_k$ is orthogonal to $\boldsymbol{h}_\ell$ for $k, \ell \leq i$, $k \neq \ell$. Substituting (22) into (21) yields

$$\boldsymbol{h}_{i+1} = \frac{\left(\boldsymbol{I}_M - \sum_{k=1}^{i}\boldsymbol{h}_k\boldsymbol{h}_k^H\right)\boldsymbol{R}_{x_0}\boldsymbol{h}_i}{\left\|\left(\boldsymbol{I}_M - \sum_{k=1}^{i}\boldsymbol{h}_k\boldsymbol{h}_k^H\right)\boldsymbol{R}_{x_0}\boldsymbol{h}_i\right\|}. \tag{23}$$

It can then be obtained from (23) that $\boldsymbol{h}_{i+1}$ is orthogonal to $\boldsymbol{h}_k$ $(k = 1, 2, \ldots, i)$, namely $\boldsymbol{h}_k$ is orthogonal to $\boldsymbol{h}_\ell$ for $k, \ell \leq i+1$, $k \neq \ell$. Therefore, $\boldsymbol{h}_i$ $(i = 1, 2, \ldots, M)$ are orthogonal to each other.

Appendix B
Proof of Lemma 1

It is straightforward to verify from (12) that $\boldsymbol{R}_d$ is an Hermitian matrix since $\boldsymbol{R}_{x_0}$ is also an Hermitian matrix. Consider now the tridiagonal property of $\boldsymbol{R}_d$. By setting $\alpha_{i+1,i} = \left\|\left(\boldsymbol{I}_M - \sum_{k=1}^{i}\boldsymbol{h}_k\boldsymbol{h}_k^H\right)\boldsymbol{R}_{x_0}\boldsymbol{h}_i\right\|$, (23) can be rewritten as

$$\begin{aligned}\boldsymbol{h}_{i+1} &= \frac{\boldsymbol{R}_{x_0}\boldsymbol{h}_i - \sum_{k=1}^{i}\boldsymbol{h}_k\left(\boldsymbol{h}_k^H\boldsymbol{R}_{x_0}\boldsymbol{h}_i\right)}{\alpha_{i+1,i}}\\ &= \frac{\boldsymbol{R}_{x_0}\boldsymbol{h}_i - \sum_{k=1}^{i}\alpha_{k,i}\boldsymbol{h}_k}{\alpha_{i+1,i}}\end{aligned} \tag{24}$$

where $\alpha_{k,i} = \boldsymbol{h}_k^H\boldsymbol{R}_{x_0}\boldsymbol{h}_i$. It follows that

$$\boldsymbol{R}_{x_0}\boldsymbol{h}_i = \alpha_{i+1,i}\boldsymbol{h}_{i+1} + \sum_{k=1}^{i}\alpha_{k,i}\boldsymbol{h}_k. \tag{25}$$

In the sequel, for $j \geq i$, we obtain the (j,i)th element of $\boldsymbol{R}_d$ as

$$R_{d,j,i} \triangleq E[d_j(t)d_i^*(t)] = \boldsymbol{h}_j^H\boldsymbol{R}_{x_0}\boldsymbol{h}_i \tag{26}$$

$$\begin{aligned}&= \alpha_{i+1,i}\boldsymbol{h}_j^H\boldsymbol{h}_{i+1} + \sum_{k=1}^{i}\alpha_{k,i}\boldsymbol{h}_j^H\boldsymbol{h}_k\\ &= \begin{cases}\alpha_{i,i} \triangleq \sigma_{d_i}^2, & j = i\\ \alpha_{i+1,i} \triangleq \delta_{i+1,i} = \delta_{i,i+1}^*, & j = i+1\\ 0, & j \geq i+2.\end{cases}\end{aligned} \tag{27}$$

Considering that $\boldsymbol{R}_d$ is an Hermitian matrix, it follows, thereby, from (27) that $\boldsymbol{R}_d$ is a tridiagonal matrix. As a result, $\boldsymbol{R}_d$ is an Hermitian tridiagonal matrix.

Employing the results of [17], i.e., the first p matched filters $\{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p\}$ span the signal subspace while the last $(M-p)$ matched filters $\{\boldsymbol{h}_{p+1}, \ldots, \boldsymbol{h}_M\}$ span the noise subspace, we can obtain $\boldsymbol{A}^H(\boldsymbol{\theta})\boldsymbol{h}_i = \boldsymbol{0}$ $(i = p+1, \ldots, M)$. In the sequel, substituting (3) into (26) yields

$$R_{d,j,i} = \boldsymbol{h}_j^H\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}^H(\boldsymbol{\theta})\boldsymbol{h}_i + \sigma_n^2\boldsymbol{h}_j^H\boldsymbol{h}_i. \tag{28}$$

For $j = i$, it follows from (27) and (28) that

$$\begin{aligned}\sigma_{d_i}^2 &= \boldsymbol{h}_i^H\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}^H(\boldsymbol{\theta})\boldsymbol{h}_i + \sigma_n^2\boldsymbol{h}_i^H\boldsymbol{h}_i\\ &= \begin{cases}\boldsymbol{h}_i^H\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}^H(\boldsymbol{\theta})\boldsymbol{h}_i + \sigma_n^2, & 1 \leq i \leq p\\ \sigma_n^2, & p+1 \leq i \leq M.\end{cases}\end{aligned} \tag{29}$$

For $j = i+1$, we obtain from (27) and (28) that

$$\begin{aligned}\delta_{i,i+1}^* &= \boldsymbol{h}_{i+1}^H\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}^H(\boldsymbol{\theta})\boldsymbol{h}_i + \sigma_n^2\boldsymbol{h}_{i+1}^H\boldsymbol{h}_i\\ &= \begin{cases}\boldsymbol{h}_{i+1}^H\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_s\boldsymbol{A}^H(\boldsymbol{\theta})\boldsymbol{h}_i, & 1 \leq i \leq p-1\\ 0, & p \leq i \leq M-1.\end{cases}\end{aligned} \tag{30}$$

Therefore, recalling that $\boldsymbol{R}_d$ is an Hermitian matrix, it follows from (29) and (30) that the tridiagonal matrix $\boldsymbol{R}_d$ can be written as (13). This completes the proof of Lemma 1.

Acknowledgment

The authors would like to thank the anonymous reviewers and the associate editor Prof. O. Besson for their suggestions which improved the presentation of the material in this paper.

REFERENCES

[1] L. C. Godara, "Application of antenna arrays to mobile communications, Part ii: Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, pp. 1195–1245, Aug. 1997.

[2] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[3] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1981.

[4] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, Jul. 1989.

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

[6] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1190–1196, Aug. 1989.

[7] 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, pp. 1157–1167, Oct. 1989.

[8] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, Aug. 2001.

[9] Q. Wu and D. R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 1848–1857, Aug. 1991.

[10] E. Fishler and H. Messer, "On the use of order statistics for improved detection of signals by the MDL criterion," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2242–2247, Aug. 2000.

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

[12] 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.

[13] H. T. Wu, J. F. Yang, and F. K. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1325–1333, Jun. 1995.

[14] 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.

[15] M. Nezafat, M. Kaveh, and W. Xu, "Estimation of the number of sources based on the eigenvectors of the covariance matrix," in *Proc. Sensor Array Multichannel Signal Processing Workshop (SAM 2004)*, Jul. 2004, pp. 465–469.

[16] L. Huang and S. Wu, "Low-complexity MDL method for accurate source enumeration," *IEEE Signal Process. Lett.*, vol. 14, pp. 581–584, Sep. 2007.

[17] 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.

[18] M. Pesavento and A. B. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1310–1324, Jul. 2001.

[19] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.

[20] M. Joham and M. D. Zoltowski, "Interpretation of the multistage nested Wiener filter in the Krylov subspace framework," Purdue Univ., West Lafayette, IN, Tech. Rep. TR-ECE-00-51, 2000.

[21] J. H. Wilkinso, *The Algebraic Eigenvalue Problem*. Oxford, U.K.: Clarendon, 1965, pp. 71–72.

Lei Huang (M'07) 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 May 2005 to April 2006, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Duke University, Durham, NC. From May 2006 to July 2008, he was with the College of Information Engineering, Shenzhen University, Shenzhen, China, first as a Lecturer and then as an Associate Professor. In August 2008, he joined the Department of Electronic Engineering, Beijing Institute of Technology, Beijing, China, as an Associate Professor. His current research interests include spectral estimation, array signal processing, statistical signal processing, and their applications in radar and wireless communication systems.

Teng Long (M'99) received the B.Sc. degree in electronic engineering from University of Science and Technology of China (USTC), Hefei, China, in 1989 and the M.Sc. and Ph.D. degrees in electrical engineering from the Beijing Institute of Technology (BIT), Beijing, China, in 1991 and 1995, respectively.

From January 1999 to August 1999, he was a Visiting Associate Professor with the Department of Electrical Engineering, Stanford University, Stanford, CA. From March 2002 to September 2002, he was a Visiting Scholar with the Department of Electric Engineering, London University, London, U.K. Since September 2000, he has been a Full Professor with the Department of Electronic Engineering and the Director of the Institute of Radar Technology, Beijing Institute of Technology, Beijing, China. His research interests include radar system, array processing and synthetic aperture radar.

Shunjun Wu (M'04) was born in Shanghai, China, on February 18, 1942. He graduated from Xidian University, Xi'an, China, in 1964.

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

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