

An MDL Algorithm for Detecting More Sources Than Sensors Using Outer-Products of Array Output

Qi Cheng, *Member, IEEE*, Piya Pal, *Member, IEEE*, Masashi Tsuji, *Member, IEEE*, and Yingbo Hua, *Fellow, IEEE*

Abstract—In this paper, we propose an algorithm for detecting the number M of Gaussian sources received by an array of a number N ($N < M$) of sensors. This algorithm is based on the minimum description length (MDL) principle using the outer-products of the array output. We show that as long as the covariance matrix of the array output has the full rank N , the covariance matrix of a vectorized outer-product of the array output has the full rank N -squared, which meets a validity condition of the MDL algorithm. We show by simulation that the MDL algorithm can perform substantially better than some relevant algorithms. A necessary identifiability condition is also obtained, for uncorrelated sources.

Index Terms—Array output, detection, MDL, outer-product, redundancy array.

I. INTRODUCTION

DETECTION and estimation of a number of sources using a smaller number of sensors are an important research topic. There are two sub-categories of this topic: signals with Gaussian properties and signals with non-Gaussian properties. For signals with Gaussian properties, one must exploit array configurations [1]–[4]. For signals with non-Gaussian properties, one could exploit the structures of the higher-order statistics of array outputs [5]–[7]. The contribution of this paper falls into the first category where only the second-order statistics are utilized and the exploitation of array configuration is essential.

The works in [1]–[3] addressed an issue known as minimum redundancy of arrays while the recent work [4] applied the notion of nested arrays. A nested array is a special type of redundancy array and is able to provide all covariance lags. It possesses the advantage of easy construction, as compared with minimum redundancy arrays [8]. The concept of subspace techniques such as MUSIC and ESPRIT was applied in [4] to the

outer-products of the array output for the estimation of the angles (-of-arrival) of plane wave sources impinging on a nested array.

In this paper, we consider the detection of the number M of sources onto a non-uniform linear array (NLA) of a number N ($N < M$) of sensors. The notation of the NLA in this paper is confined to either a nested array or a minimum redundancy array. The detection algorithm used in [4] is based on the eigen-thresholding (ET) approach proposed in [9]. The sequential noise-subspace equalization (SNSE) algorithm in [3] can also be used for such a problem. The SNSE algorithm transforms the augmented Toeplitz covariance matrix (defined by [1, (7)]) into a set of positive definite Toeplitz matrices and determines the number of sources using the likelihood ratios of those matrices. Both methods require a threshold to separate a subset of small values from another subset of large values. Generally speaking, the thresholds for the two methods depend on unknown parameters such as angles, source powers and the noise variance.

To overcome this shortcoming, in this paper, we propose a detection algorithm based on the minimum description length (MDL) principle proposed in [11], using a likelihood function of the outer-products of the array output. This method eliminates the need for a threshold. It requires more computations than the ET algorithm, but is more computationally efficient than the SNSE algorithm. We will show by simulation that the MDL algorithm performs far better than the two algorithms over a wider range of the number of snapshots (of the array output).

The most existing MDL-based algorithms are developed for the detection of sources, using uniform linear arrays (ULAs). In [20], Wax and Kailath derived an MDL criterion using the eigenvalues of the covariance matrix of the array output. In [22], Wu and Yang used the Gershgorin radii for the signal and noise subspace eigenvalues of a unitary transformed covariance matrix of the array output, to implement three algorithms. Among them, one algorithm can provide substantial improvement to the MDL algorithm in [20]. Recently, Huang and So [24] proposed an MDL criterion based on the linear shrinkage principle, aiming to improve detection accuracy of the MDL algorithm in [20] for large arrays. To handle coherent sources, Wax and Ziskind in [21] constructed signal- and noise-subspace projection matrices and used them to split the array measurement into the components in the two subspaces. They finally derived an MDL criterion which is represented using the eigenvalues of the covariance matrix of the noise subspace component. In that algorithm, a multi-dimensional search is needed to determine the maximum likelihood estimates of angles for the two projection matrices. To reduce the computational load of the algorithm in [21], Huang *et al.* in [23] proposed a scheme without requiring

Manuscript received March 09, 2014; revised June 14, 2014 and September 19, 2014; accepted September 22, 2014. Date of publication October 17, 2014; date of current version November 13, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antonio Napolitano. The work of P. Pal was supported by University of Maryland, College Park.

Q. Cheng is with the School of Engineering, University of Western Sydney, Penrith MC NSW 1797, Australia (e-mail: q.cheng@uws.edu.au).

P. Pal is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: ppal@ece.umd.edu).

M. Tsuji is with the Department of Electrical Engineering, Tokyo University of Agriculture and Technology, Tokyo 183-0057, Japan (e-mail: tsuji3@ieec.org).

Y. Hua is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: yhua@ee.ucr.edu).

Digital Object Identifier 10.1109/TSP.2014.2364019

angle estimates and eigen-decomposition. That scheme calculates the covariance matrix of the noise subspace component using a set of orthogonal matched filters. Those matched filters were derived from array measurement only and the noise covariance matrix becomes a diagonal matrix due to the orthogonality between those matched filters. Their MDL criterion was implemented using those diagonal elements.

A common feature of the algorithms mentioned in the previous paragraph, is that they all require separable signal and noise subspaces. For NLAs with fewer sensors than sources, the covariance matrix (denoted by \mathbf{R}) of the array output does not satisfy this requirement because the signal subspace expands into the noise subspace. In [16], to find the angles of sources when there are more sources than sensors, the covariance matrix (denoted by \mathbf{R}_a) of a virtual ULA is constructed from the elements of \mathbf{R} corresponding to distinct covariance lags. The virtual ULA contains more sensors than sources. Hence the conventional MUSIC can be applied.

In this paper, we also apply the algorithms in [22] and [24] to an estimate of \mathbf{R}_a and compare their performances with that of our proposed MDL algorithm in simulation. However, one should note that those algorithms no longer possess the property of the minimum description length for either the array output or the virtual array output. In contrast, our method is still of the minimum description length for the virtual array output. Since only independent sources are considered in this paper, the algorithm in [21] will not be considered. For \mathbf{R}_a of a virtual ULA, the corresponding array measurement does not seem to be readily found. Thus the algorithm in [23] will not be considered either.

The non-iterative algorithms in [22], [24] and [9] are also compared, with respect to computational complexity. Our analysis shows that the proposed MDL algorithm requires more computation than the algorithms in [22] and [9], but still offers computational advantage over the algorithm in [24].

Identifiability for parameter estimation is also investigated. An identifiability condition is derived for uncorrelated sources in terms of the number of sources and the number of sensors in the array. This condition provides a guide on the choice of array size for unique parameter estimation.

The paper is organized as follows. Section II presents the measurement model and a probability density function. Section III reveals an equivalence between the method in [4] and that in [1] for angle estimation, formulates the MDL algorithm and presents the identifiability condition. In Section IV, we provide a detailed analysis on computational complexity for the MDL algorithm, the ET, SNSE, Gershgorin-radii-based and linear-shrinkage MDL (LSMDL) algorithms, and present simulation results to compare their performances. Section V proves the full rank property of the covariance matrix of the outer-products of the array output. Section VI concludes the paper.

For readers' convenience, some frequently used vectors and matrices are listed below:

- \mathbf{I}_k the $k \times k$ identity matrix;
- $\mathbf{1}_i$ the i -th column of \mathbf{I}_N ;
- $\mathbf{I}_N^{-(n)}$ an $(N - n) \times N$ matrix by deleting the first n rows of \mathbf{I}_N ;

- $\mathbf{J}_N^{(n)}$ an $N \times n$ matrix containing the first n columns of \mathbf{I}_N ;
- $\mathbf{0}_{m \times n}$ the $m \times n$ matrix with all elements equal to zero.

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{1}_1^T & & & \\ & \mathbf{1}_2^T & & \\ & & \ddots & \\ & & & \mathbf{1}_N^T \end{bmatrix}_{N \times N^2} \quad (1)$$

$$\mathbf{P}_1 = \begin{bmatrix} \mathbf{I}_N^{-(1)} & & & \\ & \ddots & & \\ & & \mathbf{I}_N^{-(N-1)} & \mathbf{0}_{1 \times N} \end{bmatrix}_{N(N-1)/2 \times N^2} \quad (2)$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_0 & & \\ \mathbf{P}_1 & & \\ & \mathbf{P}_1 & \end{bmatrix}_{N^2 \times 2N^2} \quad (3)$$

Note that \mathbf{P}_0 , \mathbf{P}_1 and \mathbf{P} are permutation matrices.

II. MEASUREMENT MODEL

Consider an N -sensor non-uniform linear array, where the i -th sensor is located at the position d_i for $i = 1, 2, \dots, N$, $d_i < d_j$ for $i < j$ and $d_1 = 0$ is chosen for convenience. M far-field narrow-band plane wave sources are incident onto the array, from distinct angles $\theta_1, \dots, \theta_M$. The angle is measured from the normal of the (linear) array and $-90^\circ < \theta < 90^\circ$. Then the array manifold vector takes up the following form:

$$\mathbf{a}(\theta_i) = [1, e^{j(2\pi/\lambda)d_2 \sin \theta_i}, \dots, e^{j(2\pi/\lambda)d_N \sin \theta_i}]^T \quad (4)$$

where λ is the center wavelength of the waves and the superscript T denotes the transpose, and we can write the t -th snapshot of the array output as

$$\mathbf{x}_t \stackrel{def}{=} [x_{t,1}, x_{t,2}, \dots, x_{t,N}]^T = \mathbf{A}\mathbf{s}_t + \mathbf{n}_t \quad (5)$$

$t = 1, 2, \dots, T$

where T is the total number of snapshots, $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_M)]$ is an $N \times M$ array manifold matrix, \mathbf{s}_t is an $M \times 1$ source vector containing the received complex envelopes of the M waves, and \mathbf{n}_t is an $N \times 1$ measurement noise vector.

We assume that (a) \mathbf{n}_t 's are white both spatially and temporally; (b) \mathbf{s}_t 's are white temporally; (c) \mathbf{n}_t and \mathbf{s}_t are independent. Also, the source vector and noise vector are complex circular Gaussian of zero mean, and hence

$$E\{\mathbf{s}_t \mathbf{s}_k^H\} = \mathbf{R}_s \delta_{t,k}, \quad E\{\mathbf{s}_t \mathbf{s}_k^T\} = \mathbf{0} \quad (6)$$

$$E\{\mathbf{n}_t \mathbf{n}_k^H\} = \sigma \mathbf{I}_N \delta_{t,k}, \quad E\{\mathbf{n}_t \mathbf{n}_k^T\} = \mathbf{0} \quad (7)$$

where σ is the noise variance, \mathbf{R}_s is the source covariance matrix, E stands for the statistical expectation, the superscript H denotes the conjugate transpose, $\delta_{t,k} = 1$ if $t = k$ and $\delta_{t,k} = 0$ if $t \neq k$. The components in \mathbf{s}_t are allowed to be *correlated* (i.e., \mathbf{R}_s has non-zero off-diagonal elements and is of full rank)

or even *coherent* (i.e., \mathbf{R}_s is rank deficient). It then follows that \mathbf{x}_t is also complex circular Gaussian¹ with zero mean and

$$E\{\mathbf{x}_t \mathbf{x}_k^H\} = \mathbf{R} \delta_{t,k}, \quad E\{\mathbf{x}_t \mathbf{x}_k^T\} = \mathbf{0} \quad (8)$$

where

$$\mathbf{R} = E\{\mathbf{x}_t \mathbf{x}_t^H\} = \mathbf{A} \mathbf{R}_s \mathbf{A}^H + \sigma \mathbf{I}_N \quad (9)$$

is the measurement covariance matrix of the NLA.

The unknown parameters that determine the statistics of the measurement in (5) comprise the angles $\theta_1, \dots, \theta_M$, the free parameters in \mathbf{R}_s and σ . We will use the column vector $\boldsymbol{\phi}$ to denote those parameters. The free parameters in \mathbf{R}_s consists of the M real diagonal elements, and the $M(M-1)/2$ real parts and the $M(M-1)/2$ imaginary parts of the complex elements in the lower triangle of \mathbf{R}_s . Let $r_s(i, j)$ be the (i, j) -th element of \mathbf{R}_s . Then the parameter vector can be represented as

$$\begin{aligned} \boldsymbol{\phi} = & [\theta_1, \dots, \theta_M; r_s(i, i), i = 1, \dots, M; \\ & \Re\{r_s(i, j)\}, j = 1, \dots, M, i < j; \\ & \Im\{r_s(i, j)\}, j = 1, \dots, M, i < j]^T \end{aligned} \quad (10)$$

where \Re denotes the real part of a complex number and \Im the imaginary part of a complex number. The number of sources, M , is a special unknown parameter that is not included as an element in $\boldsymbol{\phi}$.

We denote the outer product of the measurement \mathbf{x}_t by $\mathbf{Z}_t = \mathbf{x}_t \mathbf{x}_t^H$ and the vectorization of all the columns of \mathbf{Z}_t by $\text{vec}(\mathbf{Z}_t)$. Obviously, the mean of $\text{vec}(\mathbf{Z}_t)$ is $\text{vec}(\mathbf{R})$.

Note that since \mathbf{Z}_t is conjugate symmetric, there are redundant elements in the real part and imaginary part of $\text{vec}(\mathbf{Z}_t)$. By removing all those redundant elements, we have the following (sufficient) (N^2 -dimensional) vector of real data:

$$\dot{\mathbf{z}}_t = \mathbf{P} \begin{bmatrix} \Re\{\text{vec}(\mathbf{Z}_t)\} \\ \Im\{\text{vec}(\mathbf{Z}_t)\} \end{bmatrix} \quad (11)$$

with its mean equal to

$$\dot{\mathbf{r}} = \mathbf{P} \begin{bmatrix} \Re\{\text{vec}(\mathbf{R})\} \\ \Im\{\text{vec}(\mathbf{R})\} \end{bmatrix}. \quad (12)$$

Let

$$\mathbf{Q} = E \left\{ \begin{bmatrix} \Re\{\text{vec}(\mathbf{Z}_t - \mathbf{R})\} \\ \Im\{\text{vec}(\mathbf{Z}_t - \mathbf{R})\} \end{bmatrix} \begin{bmatrix} \Re\{\text{vec}(\mathbf{Z}_t - \mathbf{R})\} \\ \Im\{\text{vec}(\mathbf{Z}_t - \mathbf{R})\} \end{bmatrix}^T \right\}. \quad (13)$$

Then the covariance matrix of $\dot{\mathbf{z}}_t$ can be written as

$$\mathbf{C} = E\{(\dot{\mathbf{z}}_t - \dot{\mathbf{r}})(\dot{\mathbf{z}}_t - \dot{\mathbf{r}})^T\} = \mathbf{P} \mathbf{Q} \mathbf{P}^T. \quad (14)$$

Denote

$$\mathbf{C}_{z,1} = E\{\text{vec}(\mathbf{Z}_t - \mathbf{R}) \text{vec}(\mathbf{Z}_t - \mathbf{R})^H\} \quad (15)$$

$$\mathbf{C}_{z,2} = E\{\text{vec}(\mathbf{Z}_t - \mathbf{R}) \text{vec}(\mathbf{Z}_t - \mathbf{R})^T\}. \quad (16)$$

¹If $\mathbf{x}(t)$ is relaxed to be non-Gaussian, for the MDL algorithm in Section III to hold, it has to be a second-order stationary process and satisfy the condition (98).

Recall the facts:

$$E\{\Re\{z_1\} \Re\{z_2\}\} = \Re\{E\{z_1 z_2^*\} + E\{z_1 z_2\}\}/2 \quad (17)$$

$$E\{\Im\{z_1\} \Im\{z_2\}\} = \Re\{E\{z_1 z_2^*\} - E\{z_1 z_2\}\}/2 \quad (18)$$

$$E\{\Re\{z_1\} \Im\{z_2\}\} = \Im\{E\{z_1 z_2^*\} - E\{z_1 z_2\}\}/2 \quad (19)$$

$$E\{\Im\{z_1\} \Re\{z_2\}\} = -\Im\{E\{z_1 z_2^*\} + E\{z_1 z_2\}\}/2 \quad (20)$$

where z_1 and z_2 are two complex numbers and the superscript $*$ denotes the conjugation. It then follows that

$$\mathbf{Q} = \frac{1}{2} \begin{bmatrix} \Re\{\mathbf{C}_{z,1} + \mathbf{C}_{z,2}\} & \Im\{\mathbf{C}_{z,1} - \mathbf{C}_{z,2}\} \\ -\Im\{\mathbf{C}_{z,1} + \mathbf{C}_{z,2}\} & \Re\{\mathbf{C}_{z,1} - \mathbf{C}_{z,2}\} \end{bmatrix}. \quad (21)$$

As shown in Appendix A, we can also write $\mathbf{C}_{z,1}$ and $\mathbf{C}_{z,2}$ as

$$\mathbf{C}_{z,1} = \mathbf{R}^T \otimes \mathbf{R} \quad (22)$$

$$\mathbf{C}_{z,2} = (\mathbf{I}_N \otimes \mathbf{R}) \mathbf{B} (\mathbf{I}_N \otimes \mathbf{R}^*) \quad (23)$$

where \otimes denote the Kronecker product as defined by [15, (13.1)] and

$$\mathbf{B} = \begin{bmatrix} \mathbf{1}_1 \mathbf{1}_1^T & \mathbf{1}_2 \mathbf{1}_1^T & \cdots & \mathbf{1}_N \mathbf{1}_1^T \\ \mathbf{1}_1 \mathbf{1}_2^T & \mathbf{1}_2 \mathbf{1}_2^T & \cdots & \mathbf{1}_N \mathbf{1}_2^T \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{1}_1 \mathbf{1}_N^T & \mathbf{1}_2 \mathbf{1}_N^T & \cdots & \mathbf{1}_N \mathbf{1}_N^T \end{bmatrix}_{N^2 \times N^2}. \quad (24)$$

Note that \mathbf{B} is also a permutation matrix.

Furthermore, we have proved that \mathbf{C} always has the full rank N^2 . This property is crucial to the formulation of the probability density function of the outer-products of the array output. Since the results in the proof are not directly used in the next two sections, to avoid disruption to reading flow, the proof of the full rank property will be postponed to Section V.

Note that $\dot{\mathbf{z}}_t$, $t = 1, \dots, T$, are independent and identically distributed. Let

$$\dot{\mathbf{z}} = (1/T) \sum_{t=1}^T \dot{\mathbf{z}}_t. \quad (25)$$

Then $\dot{\mathbf{z}}$ has the mean vector $\dot{\mathbf{r}}$ and the covariance matrix \mathbf{C}/T . Applying the multivariate central limit theorem (e.g., see Theorem 1.2.17 on [12, p.19]), we know that when T is sufficiently large, $\dot{\mathbf{z}} - \dot{\mathbf{r}}$ becomes approximately a Gaussian distributed vector with the mean vector $\mathbf{0}$ and the covariance matrix \mathbf{C}/T . Conditioned on $\boldsymbol{\phi}$, an asymptotic probability density function involving $\dot{\mathbf{z}}_t$, $t = 1, \dots, T$, can be written as

$$\begin{aligned} f(\dot{\mathbf{z}}_t |_{t=1}^T | \boldsymbol{\phi}) = & \frac{1}{\sqrt{(2\pi)^{N^2} |\mathbf{C}/T|}} \\ & \cdot \exp\left(-\frac{1}{2}(\dot{\mathbf{z}} - \dot{\mathbf{r}})^T (\mathbf{C}/T)^{-1} (\dot{\mathbf{z}} - \dot{\mathbf{r}})\right) \end{aligned} \quad (26)$$

where $|\cdot|$ stands for the determinant of a matrix. Note that both $\dot{\mathbf{r}}$ and \mathbf{C} are functions of $\boldsymbol{\phi}$.

III. THE PROPOSED ALGORITHM

A. Parameter Estimation

Strictly speaking, the MDL algorithm requires the maximum likelihood (ML) estimates of the parameters in ϕ . Since the likelihood function (26) of ϕ is highly nonlinear, there does not seem to exist a computationally efficient approach for the ML parameter estimation. As an alternative, in simulation, ϕ was estimated, as shown next, by using the root-MUSIC algorithm [13] based on an NLA with N sensors. Such an algorithm was developed for nested arrays in [4] and for minimum redundancy arrays in [1].

Assume that the sensors of an NLA are located at integer multiples of a smallest spacing d . Then the covariance lags generated by this array will lie in the full range of $[-(L-1), L-1]$ where L depends on the array geometry in a complicated manner for a minimum redundancy array and is an explicit function of N for a nested array. The i -th covariance lag is $E\{x_{t,k}x_{t,l}^*\}$ (i.e., the (k, l) -th element of \mathbf{R}) where $d_k - d_l = id$. Note that there can be multiple identical elements in \mathbf{R} corresponding to the i -th covariance lag. Denote the i -th covariance by $r_a(i)$ for $i \in [-(L-1), L-1]$. Note that $r_a(i) = (r_a(-i))^*$. All the non-redundant covariance lags in \mathbf{R} can be stacked in a vector as

$$\mathbf{z}_1 = [r_a(-(L-1)), \dots, r_a(-1), r_a(0), r_a(1), \dots, r_a(L-1)]^T.$$

Define

$$\mathbf{a}_1(\theta) = [e^{-j(2\pi/\lambda)(L-1)d \sin \theta}, \dots, e^{-j(2\pi/\lambda)d \sin \theta}, 1, e^{j(2\pi/\lambda)d \sin \theta}, \dots, e^{j(2\pi/\lambda)(L-1)d \sin \theta}]^T \quad (27)$$

$$\mathbf{A}_1 = [\mathbf{a}_1(\theta_1), \dots, \mathbf{a}_1(\theta_M)]. \quad (28)$$

Then, according to [4, (11)], for *uncorrelated sources*, one can write the following $(2L-1)$ -dimensional vector

$$\mathbf{z}_1 = \mathbf{A}_1 \boldsymbol{\rho} + \sigma \mathbf{e}_L \quad (29)$$

where $\boldsymbol{\rho} = [r_s(1,1), \dots, r_s(M,M)]^T$ is the source power vector and \mathbf{e}_L is the L -th column of \mathbf{I}_{2L-1} . (Uncorrelated Gaussian sources are independent sources.)

Define $\mathbf{z}_{1,i}$ as the subvector of \mathbf{z}_1 consisting of the elements at the positions $L-i+1, \dots, 2L-i$, for $i = 1, 2, \dots, L$. The MUSIC algorithm in [4] is based on the noise subspace of $\mathbf{R}_1 = \sum_{i=1}^L \mathbf{z}_{1,i} \mathbf{z}_{1,i}^H / L$. According to the definition of $\mathbf{z}_{1,i}$, it is easy to see that

$$\begin{aligned} & [\mathbf{z}_{1,L}, \dots, \mathbf{z}_{1,1}] \\ &= \begin{bmatrix} r_a(0) & r_a(-1) & \dots & r_a(-(L-1)) \\ r_a(1) & r_a(0) & \dots & r_a(-(L-2)) \\ \vdots & \vdots & \dots & \vdots \\ r_a(L-2) & r_a(L-3) & \dots & r_a(-1) \\ r_a(L-1) & r_a(L-2) & \dots & r_a(0) \end{bmatrix} \\ &\stackrel{\text{def}}{=} \mathbf{R}_a. \end{aligned} \quad (30)$$

Then

$$\mathbf{R}_1 = \mathbf{R}_a \mathbf{R}_a^H / L = (\mathbf{R}_a)^2 / L \quad (31)$$

where the second equality of (31) follows from the fact that $\mathbf{R}_a^H = \mathbf{R}_a$. In fact, \mathbf{R}_a in (30) is the augmented (Toeplitz) covariance matrix defined in [1] and was shown there that, for *uncorrelated sources*, $\mathbf{R}_a = \mathbf{A}_{1,1} \mathbf{R}_s \mathbf{A}_{1,1}^H + \sigma \mathbf{I}_L$ where $\mathbf{A}_{1,1}$ is a submatrix of \mathbf{A}_1 (consisting of the L -th, $(L+1)$ -th, \dots , $(2L-1)$ -th rows of \mathbf{A}_1). This result coincides with Theorem 2 in [4] which proves that

$$\mathbf{R}_1 = \sum_{i=1}^L \mathbf{z}_{1,i} \mathbf{z}_{1,i}^H / L = (\mathbf{A}_{1,1} \mathbf{R}_s \mathbf{A}_{1,1}^H + \sigma \mathbf{I}_L)^2 / L. \quad (32)$$

In the MUSIC algorithm based on the noise subspace (the set of the eigenvectors corresponding to the $L-M$ smallest eigenvalues of \mathbf{R}_1), $e^{j(2\pi/\lambda)d \sin \theta_m}$ for $m = 1, 2, \dots, M$ and their conjugates are the roots of a polynomial located on the unit circle. For the unique determination of angles from \mathbf{z}_1 , it is necessary that $(0 <) d \leq \lambda/2$ so that $e^{j(2\pi/\lambda)d \sin \theta_m}$ for $m = 1, 2, \dots, M$ are located on the upper half of the unit circle. In this paper, we choose $d = \lambda/2$. (One should note that for the ESPRIT algorithm based on the signal subspace (the set of the eigenvectors corresponding to the M largest eigenvalues of \mathbf{R}_1), the range of d is relaxed to be $(0 <) d \leq \lambda$.)

In Theorem 2 of [4], it was mentioned that, to apply the MUSIC, the following condition should be satisfied

$$M \leq L - 1. \quad (33)$$

This is because, if $M = L$, there is no noise subspace for the matrix $\mathbf{A}_{1,1} \mathbf{R}_s \mathbf{A}_{1,1}^H + \sigma \mathbf{I}_L$ in (32), and hence the MUSIC can not be applied. When (33) is met, there will be a non-empty noise subspace for $\mathbf{A}_{1,1} \mathbf{R}_s \mathbf{A}_{1,1}^H + \sigma \mathbf{I}_L$. Because $\mathbf{A}_{1,1}$ resembles the array manifold of a *virtual ULA*, by the argument in the paragraph after [19, (3.6)], one knows that the MUSIC angle solution will be unique in this case.

In practice, \mathbf{R} is not available and has to be estimated. Let

$$\hat{\mathbf{R}} = (1/T) \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^H \quad (34)$$

be an estimate of \mathbf{R} , using T snapshots of measurement. The elements in $\hat{\mathbf{R}}$ corresponding to the i -th lag are no longer identical. In [1], an estimate of the i -th lag is chosen as the average of those elements in $\hat{\mathbf{R}}$. This process is called redundancy averaging in the literature. Define $N \times N$ selection matrices \mathbf{J}_i for $i \in [0, L-1]$, where the (k, l) th element of \mathbf{J}_i , $J_i(k, l) = 1$ if $d_k - d_l = -id$ and $J_i(k, l) = 0$ otherwise. Note that $\mathbf{J}_0 = \mathbf{I}_N$. Then an estimate² of $r_a(-i)$ can be written as

$$\hat{r}_a(-i) = (1/T) \sum_{t=1}^T \mathbf{x}_t^T \mathbf{J}_i \mathbf{x}_t^* / \|\mathbf{J}_i\|^2 \quad (35)$$

for $i \geq 0$ where $\|\cdot\|$ is the Frobenius norm of a matrix and $\|\mathbf{J}_i\|^2$ is actually equal to the number of 1's in \mathbf{J}_i (i.e., the number of $(-i)$ -th covariance lags in \mathbf{R}). Note that the estimated covariances are still conjugate symmetric:

$$\hat{r}_a(i) = (\hat{r}_a(-i))^* = (1/T) \sum_{t=1}^T \mathbf{x}_t^T \mathbf{J}_i^T \mathbf{x}_t^* / \|\mathbf{J}_i\|^2 \quad (36)$$

²In [4], the estimation of $r_a(-i)$ was not described. One can also estimate $r_a(-i)$ by using [10, (29)].

for $i \geq 0$. An estimate of \mathbf{R}_a in (30) is given by

$$\hat{\mathbf{R}}_a = \begin{bmatrix} \hat{r}_a(0) & \hat{r}_a(-1) & \cdots & \hat{r}_a(-(L-1)) \\ \hat{r}_a(1) & \hat{r}_a(0) & \cdots & \hat{r}_a(-(L-2)) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{r}_a(L-2) & \hat{r}_a(L-3) & \cdots & \hat{r}_a(-1) \\ \hat{r}_a(L-1) & \hat{r}_a(L-2) & \cdots & \hat{r}_a(0) \end{bmatrix}. \quad (37)$$

Then one can form an estimate of \mathbf{R}_1 as

$$\hat{\mathbf{R}}_1 = \hat{\mathbf{R}}_a \hat{\mathbf{R}}_a^H / L = (\hat{\mathbf{R}}_a)^2 / L \quad (38)$$

and obtain estimates of the eigenvalues and eigenvectors of $\hat{\mathbf{R}}_1$. Note that the second equality of (38) is due to the fact that $\hat{\mathbf{R}}_a^H = \hat{\mathbf{R}}_a$.

Let the eigenvalues of $\hat{\mathbf{R}}_1$ be $\hat{\lambda}_1, \dots, \hat{\lambda}_M, \hat{\lambda}_{M+1}, \dots, \hat{\lambda}_L$ (arranged in non-ascending order), and the corresponding orthonormal eigenvectors be $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M, \hat{\mathbf{u}}_{M+1}, \dots, \hat{\mathbf{u}}_L$. Based on the estimated noise subspace $[\hat{\mathbf{u}}_{M+1}, \dots, \hat{\mathbf{u}}_L]$ (i.e., the set of eigenvectors corresponding to the $L - M$ smallest eigenvalues of $\hat{\mathbf{R}}_1$), in [4], the (polynomial-rooting) MUSIC algorithm was employed to find estimates of the M angles as $\hat{\theta}_1, \dots, \hat{\theta}_M$. The noise variance can then be estimated by

$$\hat{\sigma} = \sum_{i=M+1}^L \sqrt{\hat{\lambda}_i / (L - M)}. \quad (39)$$

And the source covariance matrix \mathbf{R}_s can be estimated by

$$\hat{\mathbf{R}}_s = \hat{\mathbf{A}}_{1,1}^+ \hat{\mathbf{U}} (\hat{\mathbf{A}}^{1/2} - \hat{\sigma} \mathbf{I}_M) \hat{\mathbf{U}}^H \hat{\mathbf{A}}_{1,1}^{+,H} \quad (40)$$

where $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M]$ is the signal subspace of $\hat{\mathbf{R}}_1$, $\hat{\mathbf{A}} = \text{diag}[\hat{\lambda}_1, \dots, \hat{\lambda}_M]$ is a diagonal matrix containing the signal subspace eigenvalues, $\hat{\mathbf{A}}_{1,1}^+ = (\hat{\mathbf{A}}_{1,1}^H \hat{\mathbf{A}}_{1,1})^{-1} \hat{\mathbf{A}}_{1,1}^H$ is the left pseudo-inverse of $\hat{\mathbf{A}}_{1,1}$ and $\hat{\mathbf{A}}_{1,1}$ is $\mathbf{A}_{1,1}$ evaluated by using the M estimated angles. Finally, an estimate of the measurement covariance matrix is given by

$$\hat{\mathbf{R}} = \hat{\mathbf{A}} \hat{\mathbf{R}}_s \hat{\mathbf{A}}^H + \hat{\sigma} \mathbf{I}_N \quad (41)$$

where $\hat{\mathbf{A}}$ is \mathbf{A} evaluated by using the M estimated angles.

If $M = 0$ (i.e., the noise-only case), the noise variance is estimated as

$$\hat{\sigma} = \sum_{i=1}^L \sqrt{\hat{\lambda}_i / L} \quad (42)$$

and the measurement covariance matrix as

$$\hat{\mathbf{R}} = \hat{\sigma} \mathbf{I}_N. \quad (43)$$

Remark: For angle estimation, the method in [1] uses the noise subspace of \mathbf{R}_a to formulate a MUSIC algorithm. From (38), one can see that the noise subspaces used by the two MUSIC algorithms are *identical*, thus both yield the same estimates of angles. The eigenvalues of $\hat{\mathbf{R}}_a$ and $\hat{\mathbf{R}}_1$ also have one-to-one correspondence. If one rearranges the eigenvalues of $\hat{\mathbf{R}}_a$ in non-ascending order of their absolute values, the signal subspace and noise subspace of $\hat{\mathbf{R}}_1$ can be constructed from the eigen-decomposition of $\hat{\mathbf{R}}_a$. Thus the method in [4] can be regarded as the one based on the second order statistics of the array output.

B. Detection Algorithm

In the development of (26), and particularly for a full-rank \mathbf{C} , \mathbf{R}_s is allowed to be non-diagonal. Furthermore, (26) itself does not include any constraints to force \mathbf{R}_s to be diagonal. Hence, in this paper, the Hermitian matrix \mathbf{R}_s is parameterized in terms of M^2 (free) real numbers. Thus the unknown parameters include the M angles, the M^2 real numbers in \mathbf{R}_s and the noise variance. (One should note that a diagonal \mathbf{R}_s (i.e., for uncorrelated sources) is required by all the existing sub-optimal methods in [1]–[4].) Let \mathcal{H}_m denote the hypothesis that the number of sources is m . Then according to the MDL principle [11] and the MUSIC identifiability condition (33), the proposed MDL algorithm based on (26), estimates the number of sources as

$$\hat{M}_{mdl} = \arg \min_{m \in [0, L-1]} \left[-\ln f(\hat{\mathbf{z}}_t |_{t=1}^T | \hat{\boldsymbol{\phi}}^{(m)}, \mathcal{H}_m) + ((m + m^2 + 1)/2) \ln T \right] \quad (44)$$

where $\hat{\boldsymbol{\phi}}^{(m)}$ is an estimate of $\boldsymbol{\phi}$ under \mathcal{H}_m , and $f(\hat{\mathbf{z}}_t |_{t=1}^T | \hat{\boldsymbol{\phi}}^{(m)}, \mathcal{H}_m)$ is the value of $f(\hat{\mathbf{z}}_t |_{t=1}^T | \boldsymbol{\phi})$ (in (26)), calculated from estimates $\hat{\mathbf{r}}^{(m)}$ and $\hat{\mathbf{C}}^{(m)}$ of \mathbf{r} and \mathbf{C} respectively, under \mathcal{H}_m .

According to (12)–(14), $\hat{\mathbf{r}}^{(m)}$ and $\hat{\mathbf{C}}^{(m)}$ require an estimate of the covariance matrix (41) (and (43)), Under \mathcal{H}_m , if that estimate is denoted by $\hat{\mathbf{R}}^{(m)}$, one can obtain $\hat{\mathbf{r}}^{(m)}$ via (12) as

$$\hat{\mathbf{r}}^{(m)} = \mathbf{P} \begin{bmatrix} \Re\{\text{vec}(\hat{\mathbf{R}}^{(m)})\} \\ \Im\{\text{vec}(\hat{\mathbf{R}}^{(m)})\} \end{bmatrix} \quad (45)$$

and an estimate of \mathbf{C}_{z1} via (22) as

$$\hat{\mathbf{C}}_{z1}^{(m)} = (\hat{\mathbf{R}}^{(m)})^T \otimes \hat{\mathbf{R}}^{(m)}. \quad (46)$$

Replacing \mathbf{C}_{z1} with $\hat{\mathbf{C}}_{z1}^{(m)}$ in (79) leads to an estimate $\hat{\mathbf{C}}_{z1}^{(m)}$ of \mathbf{C}_{z1} . $\hat{\mathbf{C}}^{(m)}$ can then be constructed by substituting $\hat{\mathbf{C}}_{z1}^{(m)}$ with $\hat{\mathbf{C}}_{z1}^{(m)}$ in (95).

The determination of $\hat{\mathbf{R}}^{(m)}$ is summarized as follows:

- Step 1: Step 1: Construct the signal subspace $\hat{\mathbf{U}}^{(m)} = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_m]$ for $m > 0$ (i.e., the eigenvectors corresponding to the m largest eigenvalues of $\hat{\mathbf{R}}_1$) and the noise subspace $\hat{\mathbf{U}}_n^{(m)} = [\hat{\mathbf{u}}_{m+1}, \dots, \hat{\mathbf{u}}_L]$ (i.e., the eigenvectors corresponding to the $L - m$ smallest eigenvalues of $\hat{\mathbf{R}}_1$).
- Step 2: Step 2: For $m > 0$, determine the m angles $\hat{\theta}_1^{(m)}, \dots, \hat{\theta}_m^{(m)}$ via the MUSIC applied onto $\hat{\mathbf{U}}_n^{(m)}$ and obtain an estimate $\hat{\mathbf{A}}_{1,1}^{(m)}$ of $\mathbf{A}_{1,1}$ and an estimate $\hat{\mathbf{A}}^{(m)}$ of \mathbf{A} using the m estimated angles.
- Step 3: Step 3: Estimate the noise variance for $m \geq 0$ as

$$\hat{\sigma}^{(m)} = \sum_{i=m+1}^L \sqrt{\hat{\lambda}_i / (L - m)} \quad (47)$$

and the $m \times m$ source covariance matrix for $m > 0$ as

$$\hat{\mathbf{R}}_s^{(m)} = (\hat{\mathbf{A}}_{1,1}^{(m)})^+ \hat{\mathbf{U}}^{(m)} ((\hat{\mathbf{A}}^{(m)})^{1/2} - \hat{\sigma}^{(m)} \mathbf{I}_m) (\hat{\mathbf{U}}^{(m)})^H (\hat{\mathbf{A}}_{1,1}^{(m)})^{+,H} \quad (48)$$

where $\hat{\mathbf{A}}^{(m)} = \text{diag}[\hat{\lambda}_1, \dots, \hat{\lambda}_m]$.

Step 4: Calculate the estimated measurement covariance matrix for $m > 0$

$$\hat{\mathbf{R}}^{(m)} = \hat{\mathbf{A}}^{(m)} \hat{\mathbf{R}}_s^{(m)} (\hat{\mathbf{A}}^{(m)})^H + \hat{\sigma}^{(m)} \mathbf{I}_N \quad (49)$$

and for $m = 0$

$$\hat{\mathbf{R}}^{(m)} = \hat{\sigma}^{(m)} \mathbf{I}_N. \quad (50)$$

C. Identifiability Condition for Uncorrelated Sources

In this subsection, we present a necessary identifiability condition of all the parameters (in particular angles) from \mathbf{z}_1 in (29) for *uncorrelated sources*³, in terms of L and M .

From (27)–(29), it is easy to see that the bottom $L - 1$ elements of \mathbf{z}_1 are not independent of its top $L - 1$ elements. The two groups are conjugate to each other because the powers are positive. So we can concentrate on the top $L - 1$ elements and the middle element of \mathbf{z}_1 . Let \mathbf{z}_+ be a subvector of \mathbf{z}_1 containing the top $L - 1$ elements and \mathbf{A}_+ represent a submatrix of \mathbf{A}_1 containing the top $L - 1$ rows. Then the top L equations of \mathbf{z}_1 in (29) can be written as

$$\mathbf{z}_+ = \mathbf{A}_+ \boldsymbol{\rho} \quad (51)$$

$$\sum_{m=1}^M r_s(m, m) = \sigma. \quad (52)$$

Applying the well-known Carathéodory theorem ([18, p. 56]) to (51), one knows that the angles and source powers can be uniquely determined from the $L - 1$ elements in \mathbf{z}_+ , if $M \leq L - 1$. From the uniqueness of source power solutions, one can further uniquely determine σ from (52). Hence, given \mathbf{z}_1 , the identifiability of all the parameters is established if

$$M \leq L - 1. \quad (53)$$

Equation (53) is the minimum identifiability condition for any method (including the MUSIC) based on the data vector \mathbf{z}_1 . One should note that some methods may require values of M smaller than $L - 1$ for identifiability. Equation (53) also indicates the maximum value on the number of uncorrelated sources which can be resolved by any method based on \mathbf{z}_1 . By comparing (33) with (53), it is clear that the MUSIC achieves the full potential of resolvability provided by \mathbf{z}_1 .

Conditioned on $\boldsymbol{\phi}$, the probability density function for \mathbf{x}_t , $t = 1, \dots, T$, can be written as

$$f_1(\mathbf{x}_t |_{t=1}^T | \boldsymbol{\phi}) = \frac{1}{((2\pi)^N |\mathbf{R}|)^T} \exp \left(-\mathbf{R}^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^H \right) \quad (54)$$

where \mathbf{R} is also a function of $\boldsymbol{\phi}$. Then identifiability also implies that the maximum likelihood estimate (MLE) given by (54) is consistent, for uncorrelated sources.

For uncorrelated sources, one can represent the source covariance matrix \mathbf{R}_s using the M source powers in $\boldsymbol{\rho}$. Let us reparameterize the function (54) as following

$$f_1(\mathbf{x}_t |_{t=1}^T | \boldsymbol{\theta}', \boldsymbol{\rho}', \sigma') \quad (55)$$

where $\boldsymbol{\theta}' = [\theta'_1, \dots, \theta'_m]$, $\boldsymbol{\rho}' = [\rho'_1, \dots, \rho'_m]$ and σ' denote the unknown parameter sets of angles, powers and noise variance,

³Identifiability conditions for correlated or coherent sources remain to be a challenge.

respectively, under \mathcal{H}_m . Also denote the negative log-likelihood function (after dropping constants) by

$$\mathcal{L}^{(T)}(\boldsymbol{\theta}', \boldsymbol{\rho}', \sigma' | \mathcal{H}_m) = -(1/T) \ln f_1(\mathbf{x}_t |_{t=1}^T | \boldsymbol{\theta}', \boldsymbol{\rho}', \sigma') - N \ln(2\pi). \quad (56)$$

Then

$$\mathcal{L}^{(T)}(\boldsymbol{\theta}', \boldsymbol{\rho}', \sigma' | \mathcal{H}_m) = \ln |\mathbf{A}'(\mathbf{R}_s)'(\mathbf{A}')^H + \sigma' \mathbf{I}_N| + \text{tr} \left((\mathbf{A}'(\mathbf{R}_s)'(\mathbf{A}')^H + \sigma' \mathbf{I}_N)^{-1} \hat{\mathbf{R}} \right) \quad (57)$$

where one should keep in mind that $(\mathbf{R}_s)'$ is parameterized by its diagonal elements only. As $T \rightarrow \infty$,

$$\begin{aligned} \mathcal{L}^{(T)}(\boldsymbol{\theta}', \boldsymbol{\rho}', \sigma' | \mathcal{H}_m) &\rightarrow \mathcal{L}^\infty(\boldsymbol{\theta}', \boldsymbol{\rho}', \sigma' | \mathcal{H}_m) \\ &\stackrel{\text{def}}{=} \ln |\mathbf{A}'(\mathbf{R}_s)'(\mathbf{A}')^H + \sigma' \mathbf{I}_N| \\ &\quad + \text{tr} \left((\mathbf{A}'(\mathbf{R}_s)'(\mathbf{A}')^H + \sigma' \mathbf{I}_N)^{-1} \mathbf{R} \right). \end{aligned} \quad (58)$$

It can be shown that any optimal $(\boldsymbol{\theta}^*, \boldsymbol{\rho}^*, \sigma^*)$, under \mathcal{H}_m , satisfy

$$\mathbf{A}^*(\mathbf{R}_s)^*(\mathbf{A}^*)^H + \sigma^* \mathbf{I}_N = \mathbf{R} = \mathbf{A} \mathbf{R}_s \mathbf{A}^H + \sigma \mathbf{I}_N \quad (59)$$

where \mathbf{A}^* and $(\mathbf{R}_s)^*$ are parameterized using $\boldsymbol{\theta}^*$ and $\boldsymbol{\rho}^*$ respectively. Equation (59) follows from the result that $f(\mathbf{R}') = \ln |\mathbf{R}'| + \text{tr}((\mathbf{R}')^{-1} \mathbf{R})$ is minimized by $\mathbf{R}' = \mathbf{R}$. Our argument on identifiability ensures that there are unique $[\boldsymbol{\theta}^*, \boldsymbol{\rho}^*, \sigma^*]$ satisfying (59) and $\boldsymbol{\theta}^* = \boldsymbol{\theta} = [\theta_1, \dots, \theta_M]$, $\boldsymbol{\rho}^* = \boldsymbol{\rho}$ and $\sigma^* = \sigma$ (which also implies $m = M$). Hence, this proves that $\mathcal{L}^\infty(\boldsymbol{\theta}', \boldsymbol{\rho}', \sigma' | \mathcal{H}_m)$ exhibits a unique global minimum at $[\boldsymbol{\theta}, \boldsymbol{\rho}, \sigma]$. Let $[\boldsymbol{\theta}_{ml}(T), \boldsymbol{\rho}_{ml}(T), \sigma_{ml}(T)] = \arg \min \mathcal{L}^{(T)}(\boldsymbol{\theta}', \boldsymbol{\rho}', \sigma' | \mathcal{H}_M)$. By a continuity argument, it can be shown that $\boldsymbol{\theta}_{ml}(T)$ converges to $\boldsymbol{\theta}$.

The following theorem summarizes the result on identifiability, emphasizing that the ML estimation in particular can be used to consistently estimate the identifiable parameters (for uncorrelated sources).

Theorem 1: Consider the T independent and identically distributed data vectors \mathbf{x}_t , $t = 1, \dots, T$ with the probability density function in (54) where the matrix \mathbf{R} is defined in (9) and \mathbf{R}_s in (9) is a diagonal matrix. If $M \leq L - 1$, the angle parameters $\boldsymbol{\theta}$ are identifiable. In particular, the MLE angle estimates $\boldsymbol{\theta}_{ml}(T)$, computed using \mathbf{x}_t , $1 \leq t \leq T$, can recover $\boldsymbol{\theta}$ as $T \rightarrow \infty$, since it satisfies

$$\lim_{T \rightarrow \infty} \boldsymbol{\theta}_{ml}(T) = \boldsymbol{\theta}. \quad (60)$$

The theorem implies that for a nested array or a minimum redundancy array, up to $M \leq L - 1$ uncorrelated sources can be uniquely identified in the limit of large T .

IV. EXPERIMENTAL STUDY

In this section, we show a simulation that compares the MDL algorithm against the ET, SNSE, Gershgorin-radii-based and LSMDL algorithms, using the 2-level nested array in [4, Fig. 1]. The array has $N = 6$ sensors at positions $d_1 = 0$, $d_2 = d$, $d_3 = 2d$, $d_4 = 3d$, $d_5 = 7d$, $d_6 = 11d$ where d is the spacing of the inner array as defined in [4, Subsection III.B]. (Note that in [4], the position of the first sensor is not chosen as the phase

reference point. But the phase differences between sensors remain unchanged.) For this 2-level nested array,

$$L = N^2/4 + N/2. \quad (61)$$

The number of sources is $M = 8$ and the values of the eight angles are $-65^\circ, -45^\circ, -30^\circ, 0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ$. The sources are uncorrelated and have unity powers. The signal-to-noise ratio is defined as $\text{SNR} = -10 \log_{10} \sigma$. Before simulation results are presented, those existing algorithms will be briefly described, their computational complexity analyzed, and some related issues discussed.

A. Noise Variance Estimation

An estimate of the noise variance is needed in the ET algorithm. We first investigate the statistical properties of the estimate of the noise variance given in (39). The values of the mean and the standard deviation of $(\hat{\sigma} - \sigma)/\sigma$ for various SNR values are recorded in Table I, which were calculated from 5000 runs of simulation with $T = 5000$ snapshots of the measurement in each run.

From Table I, one can see that as the SNR increases, the estimates of the noise variance tend to be more biased and their standard deviations increase. We further investigate the cause of bias by fixing $\text{SNR} = 8$ dB and varying T . The values of the mean and the standard deviation of $(\hat{\sigma} - \sigma)/\sigma$ for different T values were recorded and listed in Table II, which were calculated from 5000 runs of simulation. According to Table II, the cause of large bias shown earlier is likely due to insufficiently large T .

In simulation, we used $T \leq 5000$. The ET algorithm, is based on [9, (3.5)], which requires an unbiased estimate of the noise variance. In favor of the ET algorithm, we considered SNR within the range of $[-5$ dB, 0 dB] for which the standard deviation remains at a low level and the absolute value of the bias is 10 times smaller than the standard deviation.

B. ET Algorithm

The ET algorithm applies a threshold (see [9, (3.12)]) to test whether an eigenvalue belongs to the noise subspace or signal subspace, under \mathcal{H}_m . The threshold is given by

$$d_m^u = \left[(m+1) \frac{1 + \frac{r}{\sqrt{T(m+1)}}}{1 - \frac{r}{\sqrt{Tm}}} - m \right] \left[\sum_{k=m+1}^L \frac{\sqrt{\hat{\lambda}_k}}{L-m} \right]. \quad (62)$$

Note that from the expression (62), one can see that d_m^u is defined only for $m > 0$. Thus the detection range of the ET algorithm is restricted to $[1, L-1]$. Then the ET estimate of M is given by

$$\hat{M}_{et} = \arg \min_{m \in [1, L-1]} \{m | \sqrt{\hat{\lambda}_m} < d_m^u\}. \quad (63)$$

The threshold (62) contains a parameter r , which was used to control the probability of correct detection in [9]. Its value needs to be determined, for our example, so that the probability of correct detection for the ET algorithm is optimized. The threshold (62) should be always positive, thus

$$r < \sqrt{T}. \quad (64)$$

TABLE I
MEAN AND STANDARD DEVIATION (SD) OF THE NORMALIZED NOISE VARIANCE ESTIMATES FOR VARIOUS SNR'S. $T = 5000$

SNR (dB)	-5	-2	0	2	5
			8	10	
mean	-0.0027	-0.0041	-0.0059	-0.0086	0.0034
			0.2022	0.5884	
sd	0.0343	0.0578	0.0857	0.1296	0.2313
			0.3627	0.5046	

TABLE II
MEAN AND STANDARD DEVIATION (SD) OF THE NORMALIZED NOISE VARIANCE ESTIMATES FOR DIFFERENT VALUES OF T . SNR = 8 dB

T	5000	10000	15000	20000	25000
mean	0.2022	0.0548	0.024	0.0128	4.73e-4
sd	0.3067	0.2938	0.256	0.2331	0.2121

Under the condition of (64), one can easily verify that the threshold is a monotonically increasing function of r . Hence, one can expect that as r increases, the threshold becomes larger, which tends to over-estimate the dimension of the noise subspace and hence under-estimate the dimension of the signal subspace. As r decreases, the opposite is true. Fig. 1(a) & (b) show the probabilities of correct detection, over-detection and under-detection of the number of sources versus r , for the ET algorithm at SNR = -5 dB, 0 dB respectively. Here, we used 1000 runs of simulation with $T = 5000$. From the figure, one can see that when $r = 25$, 100% correct detection was achieved at both SNR values.

C. SNSE Algorithm

The implementation of the SNSE algorithm involves the setup of step size and stopping criteria, etc., which depend on source parameters and array size. For readers' convenience, a more detailed review of the SNSE algorithm is needed and it is included in Appendix H to help understand the setup.

We next determine the value of α in (125). The probability of correct detection for the SNSE algorithm is shown in Fig. 2 versus α when SNR = -5 dB, 0 dB and $T = 1000, 2000$. It is obvious that $\alpha = 0.985$ gives the highest probability of correct detection.

D. Gershgorin-Radii-Based Algorithms

In [22], three algorithms were proposed, based on the AIC, MDL and a heuristic criteria, as described in [22, Section IV]. They will be called the MGAIC, MGMDL, and MGDE algorithms as in [22]. In all our simulation examples, the MGAIC algorithm performed much worse than the other two. Thus it will not be considered in the comparison with the proposed algorithm.

Decompose $\hat{\mathbf{R}}_a$ as

$$\hat{\mathbf{R}}_a = \begin{bmatrix} \hat{\mathbf{R}}_{a,-} & \hat{\mathbf{r}}_{a,-} \\ \hat{\mathbf{r}}_{a,-}^H & 1 \end{bmatrix} \quad (65)$$

where $\hat{\mathbf{R}}_{a,-}$ is a submatrix constructed from the first $L-1$ rows and $L-1$ columns and $\hat{\mathbf{r}}_{a,-}$ is a column vector containing the top $L-1$ elements of the L -th column of $\hat{\mathbf{R}}_a$. Let

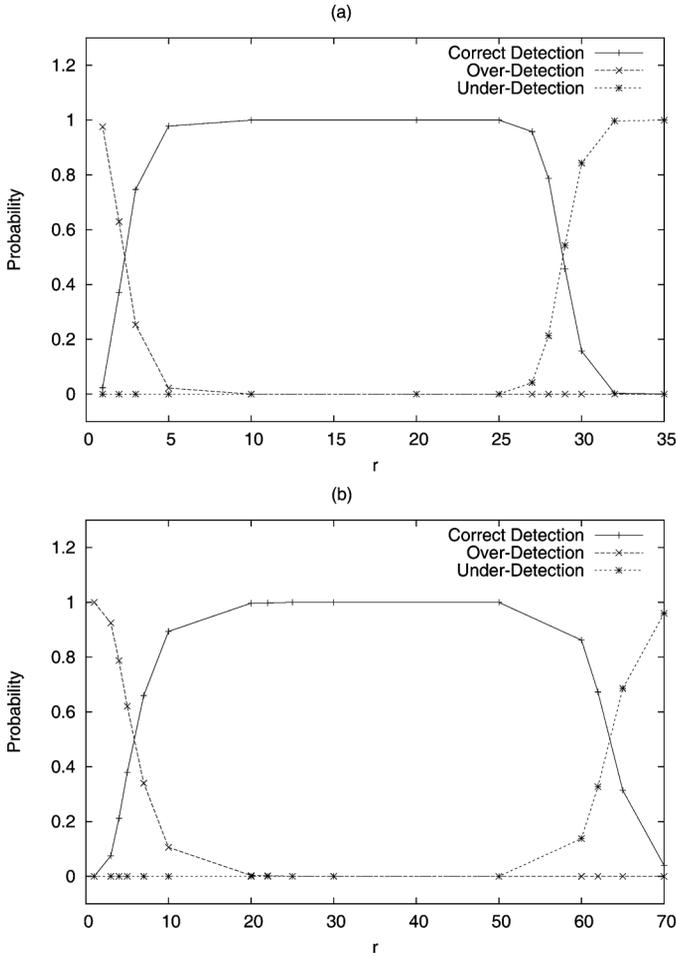


Fig. 1. Probabilities of correct detection, over-detection and under-detection versus r for the ET algorithm. $T = 5000$. (a) SNR = -5 dB. (b) SNR = 0 dB. Eight sources were present.

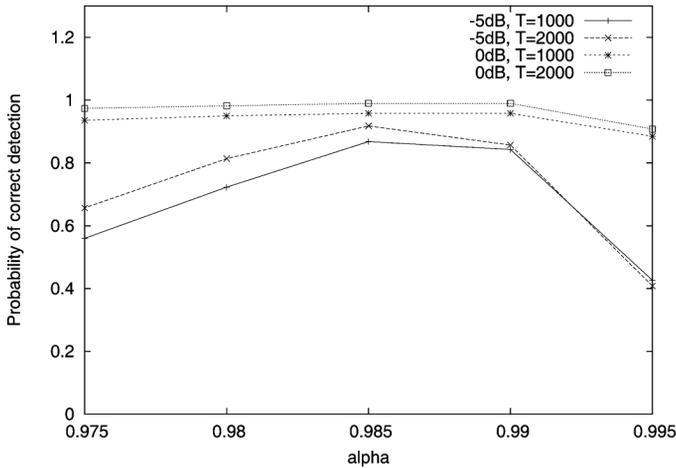


Fig. 2. Probabilities of correct detection versus α for the SNSE algorithm. SNR = -5 dB, 0 dB. Eight sources were present.

$\hat{\mathbf{U}}_g = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_{L-1}]$ contain orthonormal eigenvectors of $\hat{\mathbf{R}}_{a,-}$ where $\hat{\mathbf{u}}_i$ is an eigenvector corresponding to the eigen-

value $\hat{\eta}_i$ for $i = 1, \dots, L-1$ with $\hat{\eta}_1 \geq \hat{\eta}_2 \geq \dots \geq \hat{\eta}_{L-1}$. Further define

$$\hat{\mathbf{U}}_{g,+} = \begin{bmatrix} \hat{\mathbf{U}}_g & \mathbf{0}_{(L-1) \times 1} \\ \mathbf{0}_{1 \times (L-1)} & 1 \end{bmatrix}. \quad (66)$$

Then the matrix $\hat{\mathbf{U}}_{g,+}^H \hat{\mathbf{R}}_a \hat{\mathbf{U}}_{g,+}$ was shown [22] to have the following form

$$\begin{bmatrix} \hat{\eta}_1 & 0 & 0 & \cdots & 0 & g_1 \\ 0 & \hat{\eta}_2 & 0 & \cdots & 0 & g_2 \\ 0 & 0 & \hat{\eta}_3 & \cdots & 0 & g_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \hat{\eta}_{L-1} & g_{L-1} \\ g_1^* & g_2^* & g_3^* & \cdots & g_{L-1}^* & \hat{r}_a(0) \end{bmatrix} \quad (67)$$

where $[g_1, \dots, g_{L-1}]^T = \hat{\mathbf{U}}_g^H \hat{\mathbf{r}}_a$. The magnitudes of g_1, \dots, g_{L-1} are the Gershgorin radii for signal and noise eigenvalues.

The MGMDL algorithm basically tests the equality of the last few eigenvalues of $\hat{\mathbf{R}}_{a,-}$. This algorithm requires an estimate of the noise variance (see [22, (37)]). According to (31) of the same reference, it is defined only when $M < L-1$. Thus the detection range of the MGMDL algorithm is $[0, L-2]$. The MGDE algorithm uses a pre-selected threshold to separate the signal-eigenvalue radii from the noise-eigenvalue radii. The threshold is chosen in the same way as that in [22]: $(2.3/\ln T)[\sum_{i=1}^{L-1} |g_i|/(L-1)]$. As noted in the sentence following [22, (40)], the detection range of the MGDE algorithm is $[1, L-2]$.

E. LSMDL Algorithm

Let $\hat{\mathbf{R}}_n^{(m)} = (\hat{\mathbf{U}}_{n,1,1}^{(m)})^H \hat{\mathbf{R}}_a \hat{\mathbf{U}}_{n,1,1}^{(m)}$ be an $(L-m) \times (L-m)$ matrix where $\hat{\mathbf{U}}_{n,1,1}^{(m)}$ is an $L \times (L-m)$ matrix denoting the orthogonal subspace of $\hat{\mathbf{A}}_{1,1}^{(m)}$, under \mathcal{H}_m . (The determination of $\hat{\mathbf{A}}_{1,1}^{(m)}$ is described in Step 2 of the procedure for calculating $\hat{\mathbf{R}}_n^{(m)}$, in Section III.B.) Denote the eigenvalues of $\hat{\mathbf{R}}_n^{(m)}$ by $\hat{\mu}_1^{(m)}, \dots, \hat{\mu}_{L-m}^{(m)}$ (arranged in non-ascending order) and define the following parameters

$$\hat{\tau}^{(m)} = \sum_{i=1}^{L-m} \hat{\mu}_i^{(m)} / (L-m) \quad (68)$$

$$\hat{\alpha}^{(m)} = \frac{v_1 + v_2}{(T+1)(v_1 - v_2 / (L-m))}$$

$$\text{where } v_1 = \sum_{i=1}^{L-m} (\hat{\mu}_i^{(m)})^2, \quad v_2 = \left(\sum_{i=1}^{L-m} \hat{\mu}_i^{(m)} \right)^2 \quad (69)$$

$$\beta^{(m)} = \min\{1, \hat{\alpha}^{(m)}\} \quad (70)$$

$$\hat{\gamma}_k^{(m)} = \beta^{(m)} \hat{\tau}^{(m)} + (1 - \beta^{(m)}) \hat{\mu}_k^{(m)}, \quad k = 1, \dots, L-m. \quad (71)$$

Then the main idea of the LSMDL algorithm can be considered as the equality test of $\hat{\gamma}_1^{(m)}, \dots, \hat{\gamma}_{L-m}^{(m)}$ (which are the eigenvalues of the matrix $\beta^{(m)} \hat{\tau}^{(m)} \mathbf{I}_{L-m} + (1 - \beta^{(m)}) \hat{\mathbf{R}}_n^{(m)}$). The

definition of $\hat{\alpha}^{(m)}$ in (69) requires that $m < L - 1$. Thus the detection range of the LSMDL algorithm is $[0, L - 2]$.

F. Computational Complexity

In this subsection, the time complexity of the MDL algorithm and those in Section IV.B, Section IV.D, Section IV.E will be analyzed. The SNSE algorithm in Section IV.C is iterative. Its computational load depends on stopping criteria and does not seem to be easily found. Thus the SNSE algorithm will not further considered in this subsection.

1) MDL Algorithm:

(Step 1) To construct the signal subspace and the noise subspace of $\hat{\mathbf{R}}_1$, an EVD of $\hat{\mathbf{R}}_1$ is performed first, and requires a computational load of $O(L^3)$ flops. The notation $O(K^{l_0})$ defines a group of polynomial functions which have the same highest order l_0 but different coefficients for the highest order term K^{l_0} . All those functions share the same limiting ($K \rightarrow \infty$) rate of growth. The term ‘‘flop’’ stands for floating-point operation. A flop is defined as either of a floating-point addition, subtraction, multiplication and division. In this paper, to simplify analysis, the evaluation of a fundamental function (such as trigonometric, exponential and square-root functions) is also counted as a flop.

(Step 2) To use the MUSIC, the matrix $\hat{\mathbf{U}}_n^{(m)}(\hat{\mathbf{U}}_n^{(m)})^H$ is required. If one chooses $m = L - 1, \dots, 1$, this matrix can be computed as a rank-1 matrix update for each m , starting from a null matrix. For a given m , the updated rank-1 matrix requires $L(L + 1)/2$ multiplications to calculate and $L(L + 1)/2$ additions for it to be added to the existing matrix. The factor $(1/2)$ is due to the conjugate symmetry property of this matrix. From $\hat{\mathbf{U}}_n^{(m)}(\hat{\mathbf{U}}_n^{(m)})^H$, the MUSIC polynomial can be constructed using $(L(L - 1)/2)$ additions. To find the L roots inside the unit circle for this polynomial, a computational load of $O(L^3)$ flops is required. Angles are calculated from the m roots closest to the unit circle via m trigonometric function evaluations and then the array manifold matrix can be formed using $(L - 1)m$ multiplications. $\hat{\mathbf{A}}_{1,1}^{(m)}$ is computed from estimated angles using $(L - 1)m$ exponential function evaluations.

(Step 3) The noise variance estimate requires $L - m + 1$ square root evaluations, $L - m$ additions and 1 division. To find $\hat{\mathbf{R}}_s^{(m)}$, $[\hat{\mathbf{A}}_{1,1}^{(m)}]^H \hat{\mathbf{A}}_{1,1}^{(m)}$ is computed first, with $L(m(m + 1)/2)$ multiplications and $(L - 1)(m(m + 1)/2)$ additions. The calculation of the inverse of $[\hat{\mathbf{A}}_{1,1}^{(m)}]^H \hat{\mathbf{A}}_{1,1}^{(m)}$ requires $O(m^3)$ flops and the pseudo-inverse $(\hat{\mathbf{A}}_{1,1}^{(m)})^+$ is calculated as the product of $([\hat{\mathbf{A}}_{1,1}^{(m)}]^H \hat{\mathbf{A}}_{1,1}^{(m)})^{-1} \cdot [\hat{\mathbf{A}}_{1,1}^{(m)}]^H$ with $(L - 1)m \cdot m = (L - 1)m^2$ multiplications and $L(m - 1) \cdot m = L(m^2 - m)$ additions. To calculate $\hat{\mathbf{R}}_s^{(m)}$ in (48), one can first compute $(\hat{\mathbf{A}}_{1,1}^{(m)})^+ \hat{\mathbf{U}}^{(m)}$ which requires Lm^2 multiplications and $(L - 1)m^2$ additions. The diagonal structure of $(\hat{\mathbf{A}}_{1,1}^{(m)})^{1/2} - \hat{\sigma}^{(m)} \mathbf{I}_m$ in $\hat{\mathbf{R}}_s^{(m)}$ and the Hermitian property of $\hat{\mathbf{R}}_s^{(m)}$ can be exploited next, to obtain this matrix with $m^2 + (m(m + 1)/2)m$ multiplications and $(m(m + 1)/2)(m - 1)$ additions.

TABLE III
COMPUTATION COUNT OF FUNCTION CALCULATION IN THE
MDL ALGORITHM WHERE $m \in [1, L - 1]$

Function	No. of flops	Total computational load
EVD of $\hat{\mathbf{R}}_1$	$O(L^3)$	$O(L^3)$
$\hat{\mathbf{U}}_n^{(m)}(\hat{\mathbf{U}}_n^{(m)})^H$	$L(L + 1)$	$O(L^3)$
Angle estimation via root-MUSIC	$O(L^3)$	$O(L^4)$
$\hat{\mathbf{A}}_{1,1}^{(m)}$ from estimated angles	$(L - 1)m$	$O(L^3)$
$[\hat{\mathbf{A}}_{1,1}^{(m)}]^H \hat{\mathbf{A}}_{1,1}^{(m)}$	$(2L - 1) \cdot (m(m + 1)/2)$	$O(L^4)$
$([\hat{\mathbf{A}}_{1,1}^{(m)}]^H \hat{\mathbf{A}}_{1,1}^{(m)})^{-1}$	$O(m^3)$	$O(L^4)$
$(\hat{\mathbf{A}}_{1,1}^{(m)})^+$	$(2L - 1)m^2$	$O(L^4)$
$(\hat{\mathbf{A}}_{1,1}^{(m)})^+ \hat{\mathbf{U}}^{(m)}$	$(2L - 1)m^2 - Lm$	$O(L^4)$
$\hat{\sigma}^{(m)}$	$2L - 2m + 1$	$O(NL^2)$
$\hat{\mathbf{R}}_s^{(m)}$	$m^3 + (3/2)m^2 - (1/2)m$	$O(L^4)$
$\hat{\mathbf{R}}^{(m)}$	$N(N + 1) \cdot (m^2 + m - 1/2)$	$O(N^2 L^3)$
$\hat{\mathbf{C}}_{z1}^{(m)}$	$N^4/2 + N^3/2$	$O(N^4 L)$
$\hat{\mathbf{C}}^{(m)}$	$2N^4 + 2N^2 + 1$	$O(N^4 L)$
$(\hat{\mathbf{C}}^{(m)})^{-1}$	$O(N^6)$	$O(N^6 L)$
$ \hat{\mathbf{C}}^{(m)} $	$O(N^6)$	$O(N^6 L)$
$(\hat{\mathbf{z}} - \hat{\mathbf{r}}^{(m)})^T \cdot (\hat{\mathbf{C}}^{(m)}/T)^{-1} (\hat{\mathbf{z}} - \hat{\mathbf{r}}^{(m)})$	$2N^4 + 2N^2 - 1$	$O(N^4 L)$

(Step 4) $\hat{\mathbf{A}}_{1,1}^{(m)}$ can be directly obtained from some rows in $\hat{\mathbf{A}}_{1,1}^{(m)}$. $\hat{\mathbf{R}}^{(m)}$ in (49) can be calculated with $[N(N + 1)/2](m^2 + m)$ multiplications and $[N(N + 1)/2](m^2 + m - 1)$ additions.

The likelihood function (26) requires $\hat{\mathbf{r}}^{(m)}$ and $\hat{\mathbf{C}}^{(m)}$. Since \mathbf{P} in (3) is a permutation matrix, then according to (45), the calculation of $\hat{\mathbf{r}}^{(m)}$ from $\hat{\mathbf{R}}^{(m)}$ only requires row permutation. Permutation is much faster than floating-point operations, and thus it is not counted towards computational load. $\hat{\mathbf{C}}^{(m)}$ is calculated from $\hat{\mathbf{C}}_{z1}^{(m)}$ based on (95).

$\hat{\mathbf{C}}_{z1}^{(m)}$ in (46) is also Hermitian. By direct multiplication, N^2 multiplications are needed for each $N \times N$ block. Thus the computational load to obtain $\hat{\mathbf{C}}_{z1}^{(m)}$ from $\hat{\mathbf{R}}^{(m)}$ is $(N(N + 1)/2)N^2$ multiplications. According to the analysis after Theorem 4 in Section V, $\hat{\mathbf{C}}^{(m)}$ can be calculated from $\hat{\mathbf{C}}_{z1}^{(m)}$, using $[N^2(N^2 + 1)/2]4 = 2N^2(N^2 + 1)$ additions/subtractions and 1 division.

$(\hat{\mathbf{C}}^{(m)})^{-1}$ and $|\hat{\mathbf{C}}^{(m)}|$ both are also required in the likelihood function (26). Each evaluation requires $O(N^6)$ flops. Using $(\hat{\mathbf{C}}^{(m)})^{-1}$, $(\hat{\mathbf{z}} - \hat{\mathbf{r}}^{(m)})^T \cdot (\hat{\mathbf{C}}^{(m)}/T)^{-1} (\hat{\mathbf{z}} - \hat{\mathbf{r}}^{(m)})$ can be calculated, at the cost of N^2 subtractions, $N^4 + N^2$ multiplications and $N^2(N^2 - 1) + N^2 + N^2 - 1 = N^4 + N^2 - 1$ additions.

For the MDL algorithm, a summary of the computational loads of the above functions along with their respective totals summed over $m = 1, \dots, L - 1$ is presented in Table III. Thus the overall computational load for the MDL algorithm is $O(L^4) + O(NL^3) + O(N^6 L)$ flops.

2) *LSMDL Algorithm:* The LSMDL algorithm also needs the first five functions in Table III.

$\hat{\mathbf{U}}_{n,1,1}^{(m)}$ is the next function to be calculated. In our simulation, it is determined as the set of orthonormal eigenvectors,

TABLE IV
COMPUTATION COUNT OF FUNCTION CALCULATION IN THE LSMDL
ALGORITHM WHERE $m \in [1, L - 2]$

Function	No. of flops	Total computational load
EVD of $\hat{\mathbf{R}}_1$	$O(L^3)$	$O(L^3)$
$\hat{\mathbf{U}}_n^{(m)} (\hat{\mathbf{U}}_n^{(m)})^H$	$L(L+1)$	$O(L^4)$
Angle estimation via root-MUSIC	$O(L^3)$	$O(L^4)$
$\hat{\mathbf{A}}_{1,1}^{(m)}$	$(L-1)m$	$O(L^3)$
from estimated angles		
$\hat{\mathbf{A}}_{1,1}^{(m)} [\hat{\mathbf{A}}_{1,1}^{(m)}]^H$	$(L(L+1)/2)m$	$O(L^4)$
$\hat{\mathbf{U}}_{n,1,1}^{(m)}$	$O(L^2(L-m))$	$O(L^4)$
$\hat{\mathbf{R}}_n^{(m)}$	$(L^2 + L/2 - 1/2)m(m+1)$	$O(L^5)$
Eigenvalues of $\hat{\mathbf{R}}_n^{(m)}$	$O((L-m)^3)$	$O(L^4)$

associated with the $L - m$ smallest eigenvalues of the matrix $\hat{\mathbf{A}}_{1,1}^{(m)} [\hat{\mathbf{A}}_{1,1}^{(m)}]^H$. The corresponding EVD requires $O(L^3)$ flops.

$\hat{\mathbf{R}}_n^{(m)} = (\hat{\mathbf{U}}_{n,1,1}^{(m)})^H \hat{\mathbf{R}}_a \hat{\mathbf{U}}_{n,1,1}^{(m)}$ can be calculated using $(L^2 + L)[m(m+1)/2]$ multiplications and $((L-1)L + L - 1)[m(m+1)/2]$ additions.

The $L - m$ eigenvalues of $\hat{\mathbf{R}}_n^{(m)}$ are needed in (68)–(71). They can be calculated in $O((L - m)^3)$ flops.

For the LSMDL algorithm, a summary of the computational loads of the above functions along with their respective totals summed over $m = 1, \dots, L - 2$ is presented in Table IV. Thus the overall computational load for the LSMDL algorithm is $O(L^5)$ flops.

3) *Other Two Algorithms:* The major computation in Gershgorin-radii-based algorithms involves the EVD of $\hat{\mathbf{R}}_{a,-}$ and the calculation of $g_1, g_2, g_3, \dots, g_{L-1}$. An EVD of $\hat{\mathbf{R}}_{a,-}$ (introduced in (65)) requires a load of $O((L-1)^3)$ flops. The coefficients $g_1, g_2, g_3, \dots, g_{L-1}$ can be calculated using $(L-1)^2$ multiplications and $(L-1)(L-2)$ additions. Thus the computational load of Gershgorin-radii-based algorithms is $O(L^3)$ flops.

The ET algorithm requires an EVD of $\hat{\mathbf{R}}_1$ and the calculation of d_m^u for $m = 1, 2, \dots, L - 2$. The calculation of d_m^u requires $L - m$ multiplications and $L - m - 1$ additions for each m . Therefore the computational load of the ET algorithm is $O(L^3)$ flops.

4) *Concluding Remark:* It is now obvious that Gershgorin-radii-based algorithms and the ET algorithm are much more computationally efficient than the MDL and LSMDL algorithms. When $N \gg 2$, the term “ $N^2/4$ ” dominates the term “ $N/2 + 1$ ” in (61) and one can approximate L with $N^2/4$. In this case, the computational loads of the MDL and LSMDL are $O(N^8)$ and $O(N^{10})$ flops, respectively. Hence for a sufficiently large N , the proposed MDL algorithm is more efficient than the LSMDL algorithm.

G. Comparative Study

Now we compare the performance of the MDL algorithm with that of the ET, SNSE, MGDE, MGMDL and LSMDL algorithms for $r = 25$ and $\alpha = 0.985$. The probabilities of correct detection using all the algorithms versus the number T of snapshots are shown in Fig. 3, for SNR = -5 dB, -2 dB, 0 dB.

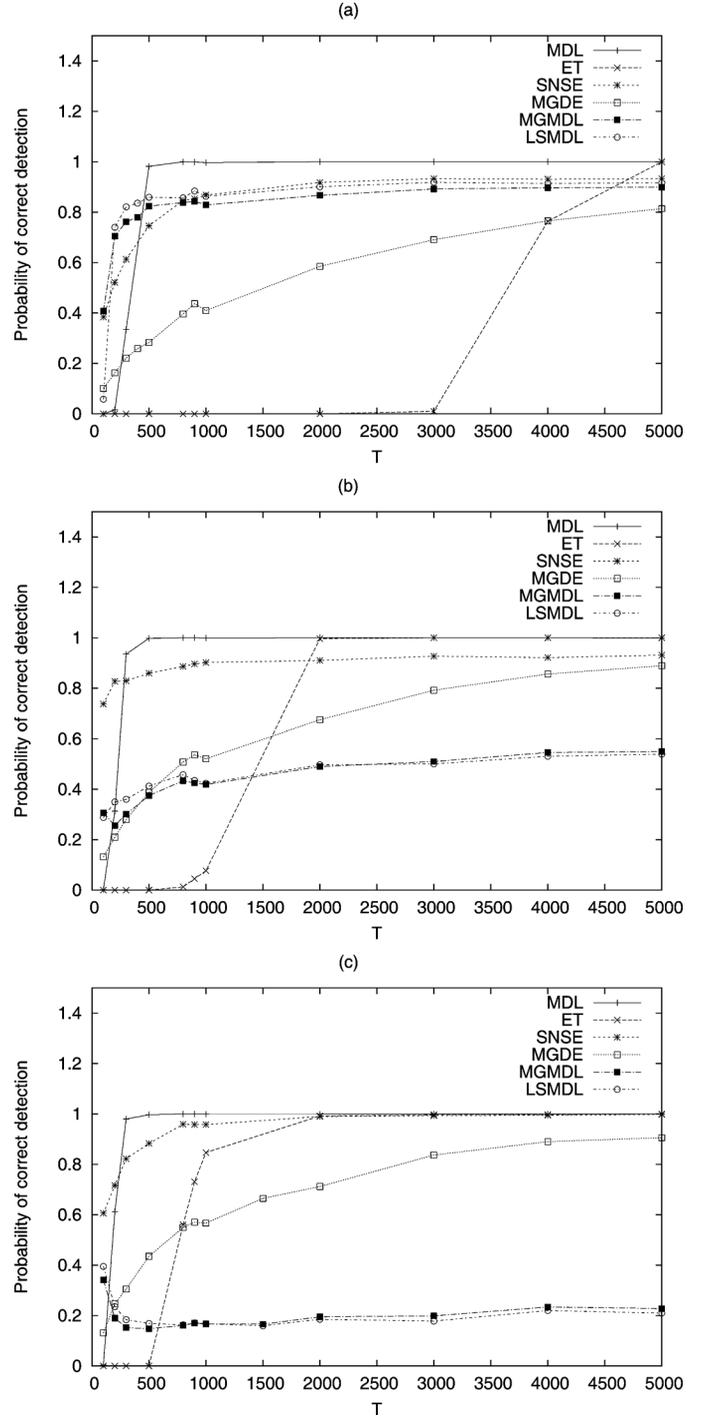


Fig. 3. Probabilities of correct detection versus T for the ET, SNSE, MGDE, MGMDL, LSMDL and MDL algorithms. $r = 25$ and $\alpha = 0.985$. (a) SNR = -5 dB. (b) SNR = -2 dB. (c) SNR = 0 dB. Eight sources were present.

The MDL algorithm can provide 100% correct detection for $T \geq 800$ at all the three SNR values. The other five algorithms have poorer performances than the MDL algorithm for most T values. The ET algorithm requires much higher SNR values to achieve 100% correct detection.

One should note that the scenario considered in this simulation is very challenging. There are only 6 data in each snapshot. This is why a minimum of 800 snapshots are needed for the perfect detection performance of the MDL algorithm. If the number N of sensors increases, this minimum value can be reduced.

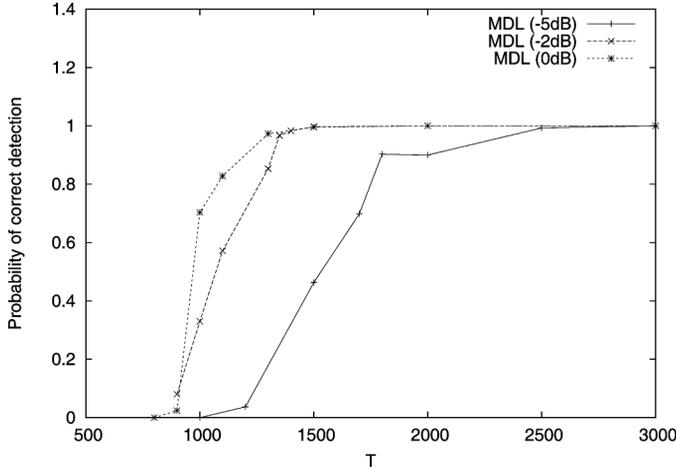


Fig. 4. Probabilities of correct detection versus T for the MDL algorithm, with SNR = -5 dB, -2 dB and 0 dB respectively. Eleven sources were present.

The MGMDL and LSMDL algorithms do not provide 100% correct detection at all the three SNR values. Detection failures of these two algorithms are related to over-estimation for $T \geq 200$. These two algorithms have another interesting behavior that both perform better at a low SNR value than at a high SNR value. Our further investigation reveals that the noise eigenvalues $\hat{\eta}_{M+1}, \dots, \hat{\eta}_{L-1}$ of $\hat{\mathbf{R}}_{a,-}$ and the noise eigenvalues $\hat{\gamma}_1^{(M)}, \dots, \hat{\gamma}_{L-M}^{(M)}$ tend to be more widely spread as the SNR increases. The spread is measured by ratios $\hat{\eta}_{M+1}/\hat{\eta}_{L-1}$ and $\hat{\gamma}_1^{(M)}/\hat{\gamma}_{L-M}^{(M)}$ respectively. As a consequence, more noise eigenvalues were mistaken as the signal eigenvalues, leading to more over-estimation, at a high SNR value.

The performance of the MGDE algorithm does not vary significantly with SNRs. However it is still unable to provide 100% correct detection even when $T = 5000$ at all the three SNR values.

Finally, we demonstrate that the proposed algorithm can correctly detect the maximum number of $N^2/4 + N/2 - 1$ sources. For $N = 6$, this maximum value is 11. In this example, the number of sources is $M = 11$ and the values of the eleven angles are $-70^\circ, -45^\circ, -30^\circ, -20^\circ, -15^\circ, 0^\circ, 15^\circ, 20^\circ, 30^\circ, 45^\circ, 60^\circ$. Probabilities of correct detection for the MDL algorithm are shown in Fig. 4. From this figure, it is obvious that the proposed MDL algorithm can still achieve 100% correct detection as long as T is large enough.

V. FULL RANK PROPERTY OF \mathbf{C}

In this section, we show that the matrix \mathbf{C} (defined in (14)) has the full rank N^2 .

While the proof shown in this section is crucial for this paper, readers may find it helpful to skip this section for the first time reading.

Lemma 1: \mathbf{B} is an $N^2 \times N^2$ permutation matrix satisfying

$$\mathbf{B}(\mathbf{1}_k \otimes \mathbf{1}_l) = (\mathbf{1}_l \otimes \mathbf{1}_k) \quad (72)$$

and

$$(\mathbf{1}_k \otimes \mathbf{1}_l)^T \mathbf{B} = (\mathbf{1}_l \otimes \mathbf{1}_k)^T \quad (73)$$

for $k, l \in [1, N]$, where $\mathbf{1}_k$ ($\mathbf{1}_l$) is defined in the last paragraph of Section I. Furthermore, \mathbf{B} is its self inverse and symmetric:

$$\mathbf{B}^{-1} = \mathbf{B}^T = \mathbf{B}. \quad (74)$$

Proof: See Appendix B. \diamond

Note that when left-multiplying a matrix with \mathbf{B} , the operation swaps the $(N(i-1)+j)$ -th row with the $(N(j-1)+i)$ -th row of that matrix and when right-multiplying a matrix with \mathbf{B} , the operation swaps the $(N(j-1)+i)$ -th column with the $(N(i-1)+j)$ -th column of that matrix.

Theorem 2: Through \mathbf{B} , $\mathbf{C}_{z,1}$ and $\mathbf{C}_{z,2}$ are related to each other by

$$\mathbf{C}_{z,2} = \mathbf{B}\mathbf{C}_{z,1}^* \quad (75)$$

$$\mathbf{C}_{z,2} = \mathbf{C}_{z,1}\mathbf{B}. \quad (76)$$

Proof: See Appendix C. \diamond

From (75), $\Re\{\mathbf{C}_{z,2}\} = \mathbf{B}\Re\{\mathbf{C}_{z,1}\}$ and $\Im\{\mathbf{C}_{z,2}\} = -\mathbf{B}\Im\{\mathbf{C}_{z,1}\}$, and \mathbf{Q} can be rewritten as

$$\begin{aligned} \mathbf{Q} &= \frac{1}{2} \begin{bmatrix} (\mathbf{I}_{N^2} + \mathbf{B})\Re\{\mathbf{C}_{z,1}\} & (\mathbf{I}_{N^2} + \mathbf{B})\Im\{\mathbf{C}_{z,1}\} \\ -(\mathbf{I}_{N^2} - \mathbf{B})\Im\{\mathbf{C}_{z,1}\} & (\mathbf{I}_{N^2} - \mathbf{B})\Re\{\mathbf{C}_{z,1}\} \end{bmatrix} \\ &= \dot{\mathbf{B}}\dot{\mathbf{C}}_{z,1} \end{aligned} \quad (77)$$

where

$$\dot{\mathbf{B}} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_{N^2} + \mathbf{B} & \\ & \mathbf{I}_{N^2} - \mathbf{B} \end{bmatrix} \quad (78)$$

$$\dot{\mathbf{C}}_{z,1} = \begin{bmatrix} \Re\{\mathbf{C}_{z,1}\} & \Im\{\mathbf{C}_{z,1}\} \\ -\Im\{\mathbf{C}_{z,1}\} & \Re\{\mathbf{C}_{z,1}\} \end{bmatrix}_{2N^2 \times 2N^2}. \quad (79)$$

Similarly, from (76), one can obtain $\Re\{\mathbf{C}_{z,2}\} = \Re\{\mathbf{C}_{z,1}\}\mathbf{B}$ and $\Im\{\mathbf{C}_{z,2}\} = \Im\{\mathbf{C}_{z,1}\}\mathbf{B}$, and \mathbf{Q} can also be expressed as

$$\mathbf{Q} = \dot{\mathbf{C}}_{z,1}\dot{\mathbf{B}}. \quad (80)$$

Thus from (77) and (80), one obtains

$$\dot{\mathbf{B}}\dot{\mathbf{C}}_{z,1} = \dot{\mathbf{C}}_{z,1}\dot{\mathbf{B}} \rightarrow \dot{\mathbf{B}}\dot{\mathbf{C}}_{z,1}\dot{\mathbf{B}}^T = \dot{\mathbf{C}}_{z,1}\dot{\mathbf{B}}\dot{\mathbf{B}}^T. \quad (81)$$

From the second equation of (74), $\dot{\mathbf{B}}^T = \dot{\mathbf{B}}$, which is equivalent to $\dot{\mathbf{B}}\dot{\mathbf{B}}^T = \dot{\mathbf{B}}\dot{\mathbf{B}}$. Equation (74) implies that $\mathbf{B}\mathbf{B}^T = \mathbf{I}_{N^2}$, and using this property, it is easy to verify that $\dot{\mathbf{B}}\dot{\mathbf{B}} = \dot{\mathbf{B}}$. Hence the second equation of (81) can be written as

$$\dot{\mathbf{B}}\dot{\mathbf{C}}_{z,1}\dot{\mathbf{B}}^T = \dot{\mathbf{C}}_{z,1}\dot{\mathbf{B}}. \quad (82)$$

Combination of (82) and (80) leads to

$$\mathbf{Q} = \dot{\mathbf{B}}\dot{\mathbf{C}}_{z,1}\dot{\mathbf{B}}^T. \quad (83)$$

From Lemma F1, $\dot{\mathbf{C}}_{z,1}$ is shown to be positive-definite. Thus \mathbf{Q} is now represented in a semi-positive (or positive) definite form.

Next, we aim to find the eigen-decomposition of \mathbf{B} . Let

$$\mathbf{U}_{1,k} = (1/\sqrt{2}) \begin{bmatrix} \mathbf{I}_k \otimes \mathbf{1}_{k+1} \\ \mathbf{J}_N^{(k)} \\ \mathbf{0}_{(N^2-N(k+1)) \times k} \end{bmatrix}_{N^2 \times k} \quad (84)$$

$$\mathbf{U}_{2,k} = (1/\sqrt{2}) \begin{bmatrix} \mathbf{I}_k \otimes \mathbf{1}_{k+1} \\ -\mathbf{J}_N^{(k)} \\ \mathbf{0}_{(N^2-N(k+1)) \times k} \end{bmatrix}_{N^2 \times k} \quad (85)$$

$k = 1, 2, \dots, N-1$

where $\mathbf{J}_N^{(k)}$ is defined in the last paragraph of Section I. Also let

$$\mathbf{U}_1 = [\mathbf{P}_0^T \quad \mathbf{U}_{1,1} \quad \mathbf{U}_{1,2} \quad \cdots \quad \mathbf{U}_{1,N-1}]_{N^2 \times r_1} \quad (86)$$

$$\mathbf{U}_2 = [\mathbf{U}_{2,1} \quad \mathbf{U}_{2,2} \quad \cdots \quad \mathbf{U}_{2,N-1}]_{N^2 \times r_2} \quad (87)$$

where \mathbf{P}_0 is defined in (1), $r_1 = N(N+1)/2$ and $r_2 = N(N-1)/2$.

Lemma 2: Both \mathbf{U}_1 and \mathbf{U}_2 contain orthonormal columns, and they are orthogonal complements to each other.

Proof: See Appendix D. \diamond

Theorem 3: \mathbf{U}_1 and \mathbf{U}_2 are the matrices of the orthonormal eigenvectors of \mathbf{B} corresponding to the eigenvalues 1 and -1 respectively, i.e.,

$$\mathbf{B}\mathbf{U}_1 = \mathbf{U}_1 \quad (88)$$

$$\mathbf{B}\mathbf{U}_2 = -\mathbf{U}_2. \quad (89)$$

Proof: See Appendix E. \diamond

Since \mathbf{U}_1 and \mathbf{U}_2 altogether contain N^2 orthonormal columns in the N^2 -dimensional space, then from Lemma 2, one can write

$$\mathbf{U}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{U}_2^T = \mathbf{I}_{N^2}. \quad (90)$$

Using (90) along with Theorem 3, one knows that the eigen-decomposition of \mathbf{B} can be written as

$$\mathbf{B} = \mathbf{U}_1\mathbf{U}_1^T - \mathbf{U}_2\mathbf{U}_2^T. \quad (91)$$

Based on (90) and (91), one obtains that $\mathbf{I}_{N^2} + \mathbf{B} = 2\mathbf{U}_1\mathbf{U}_1^T$ and $\mathbf{I}_{N^2} - \mathbf{B} = 2\mathbf{U}_2\mathbf{U}_2^T$ and

$$\dot{\mathbf{B}} = \dot{\mathbf{U}}\dot{\mathbf{U}}^T \quad (92)$$

where

$$\dot{\mathbf{U}} = \begin{bmatrix} \mathbf{U}_1 & \\ & \mathbf{U}_2 \end{bmatrix}_{2N^2 \times N^2}. \quad (93)$$

Plugging (92) into (83) gives another expression of \mathbf{Q} :

$$\mathbf{Q} = \dot{\mathbf{U}}(\dot{\mathbf{U}}^T \dot{\mathbf{C}}_{z,1} \dot{\mathbf{U}}) \dot{\mathbf{U}}^T. \quad (94)$$

Substitution of (94) into (14) leads to

$$\mathbf{C} = \mathbf{P}\dot{\mathbf{U}} \cdot (\dot{\mathbf{U}}^T \dot{\mathbf{C}}_{z,1} \dot{\mathbf{U}}) \cdot \dot{\mathbf{U}}^T \mathbf{P}^T. \quad (95)$$

Note that $\mathbf{P}\dot{\mathbf{U}}$ is square. The next theorem shows that both $\mathbf{P}\dot{\mathbf{U}}$ and $\dot{\mathbf{U}}^T \dot{\mathbf{C}}_{z,1} \dot{\mathbf{U}}$ are of the full rank N^2 .

Theorem 4: (1) $\mathbf{P}\dot{\mathbf{U}}$ is of full rank N^2 , and (2) $\dot{\mathbf{U}}^T \dot{\mathbf{C}}_{z,1} \dot{\mathbf{U}}$ is of full rank N^2 .

Proof: See Appendix F. \diamond

Therefore, the (square) matrix \mathbf{C} has been shown to have the full rank N^2 .

The structure of (94) and the properties of the matrices \mathbf{P} , \mathbf{U}_1 , \mathbf{U}_2 can be used to develop an efficient procedure to calculate \mathbf{C} from $\mathbf{C}_{z,1}$.

From (79) and (93), one obtains

$$\dot{\mathbf{U}}^T \dot{\mathbf{C}}_{z,1} \dot{\mathbf{U}} = \begin{bmatrix} \Re\{\mathbf{U}_1^T \mathbf{C}_{z,1} \mathbf{U}_1\} & \Im\{\mathbf{U}_1^T \mathbf{C}_{z,1} \mathbf{U}_2\} \\ -\Im\{\mathbf{U}_2^T \mathbf{C}_{z,1} \mathbf{U}_1\} & \Re\{\mathbf{U}_2^T \mathbf{C}_{z,1} \mathbf{U}_2\} \end{bmatrix} \quad (96)$$

and knows that $\dot{\mathbf{U}}^T \dot{\mathbf{C}}_{z,1} \dot{\mathbf{U}}$ can be obtained by taking real and imaginary parts of relevant blocks of $[\mathbf{U}_1, \mathbf{U}_2]^T \mathbf{C}_{z,1} [\mathbf{U}_1, \mathbf{U}_2]$.

$\sqrt{2}[\mathbf{U}_1, \mathbf{U}_2]$ is a matrix with elements $(0, 1, -1)$ and only contains two nonzero elements in each column. Thus $[\mathbf{U}_1, \mathbf{U}_2]^T \mathbf{C}_{z,1} [\mathbf{U}_1, \mathbf{U}_2]$ can be calculated using $[N^2(N^2+1)/2]4 = 2N^2(N^2+1)$ additions and 1 division. Note that each row of $\sqrt{2}\dot{\mathbf{U}}$ only contains a nonzero element 1 or -1 and \mathbf{P} is a permutation matrix. Then $\sqrt{2}\mathbf{P}\dot{\mathbf{U}}$ also has one nonzero element in each row. Thus further computations only involve sign change for some elements and row/column permutation. Sign change and permutation are much faster than floating-point operations, and thus are not counted towards computational load.

In summary, the calculation of \mathbf{C} from $\mathbf{C}_{z,1}$ requires $2N^2(N^2+1)$ additions/subtractions and 1 division.

VI. CONCLUSIONS

We have developed an MDL-based detection algorithm using the outer-products of the array output for a nested array or a minimum redundancy array. We have applied some of the properties of a nested array shown in [4] for estimating the unknown parameters (other than the number of sources) that govern the probability density function of a vectorized outer-product of the array output. With a nested array of N sensors, the MDL algorithm can detect up to $N^2/4 + N/2 - 1$ uncorrelated Gaussian sources. The resulting MDL detection algorithm has been shown, via simulation results, to substantially outperform the ET detection algorithm applied in [4], the SNSE detection algorithm in [3] and the other detection algorithms proposed in [22] and [24]. Computationally, the proposed MDL algorithm is more efficient than that in [24] when N is large enough.

APPENDIX A

EXPRESSIONS OF TWO COVARIANCE MATRICES

Derivation of $\mathbf{C}_{z,1}$: Note that $\mathbf{C}_{z,1} = E\{\text{vec}(\mathbf{Z}_t)\text{vec}(\mathbf{Z}_t)\} - \text{vec}(\mathbf{R})\text{vec}(\mathbf{R})^H$. The (i, j) -th $N \times N$ block of $\text{vec}(\mathbf{Z}_t)\text{vec}(\mathbf{Z}_t)^H$ is $\mathbf{x}_t \mathbf{x}_t^H x_{t,i}^* x_{t,j}$. Then $x_{t,i} = \mathbf{1}_i^T \mathbf{x}_t$. The (i, j) -th $N \times N$ block of $\text{vec}(\mathbf{R})\text{vec}(\mathbf{R})^H$ is $\mathbf{R}\mathbf{1}_i(\mathbf{R}\mathbf{1}_j)^H$. Since $\mathbf{R}^H = \mathbf{R}$, $(\mathbf{R}\mathbf{1}_j)^H = \mathbf{1}_j^T \mathbf{R}$. Then the (i, j) -th $N \times N$ block of $\text{vec}(\mathbf{R})\text{vec}(\mathbf{R})^H$ is equal to $\mathbf{R}\mathbf{1}_i \mathbf{1}_j^T \mathbf{R}$. Then the (i, j) -th $N \times N$ block of $\mathbf{C}_{z,1}$ is given by

$$\begin{aligned} \mathbf{C}_{z,1}(i, j) &= E\{\mathbf{x}_t \mathbf{x}_t^H x_{t,i}^* x_{t,j}\} - \mathbf{R}\mathbf{1}_i \mathbf{1}_j^T \mathbf{R} \\ &= E\{\mathbf{x}_t \mathbf{x}_t^H \cdot \mathbf{x}_t^H \mathbf{1}_i \cdot \mathbf{1}_j^T \mathbf{x}_t\} - \mathbf{R}\mathbf{1}_i \mathbf{1}_j^T \mathbf{R}. \end{aligned} \quad (97)$$

It is shown that (see [16], for example), for any four jointly distributed Gaussian random variables x_1, x_2, x_3, x_4 with zero means, the following equation

$$E[x_1 x_2^* x_3 x_4^*] = E[x_1 x_2^*] E[x_3 x_4^*] + E[x_1 x_4^*] E[x_2 x_3^*] \quad (98)$$

holds. Using the Gaussianity of the measurement and (8),

$$\begin{aligned} &E\{\mathbf{x}_t \mathbf{x}_t^H \cdot \mathbf{x}_t^H \mathbf{1}_i \cdot \mathbf{1}_j^T \mathbf{x}_t\} \\ &= E\{\mathbf{x}_t \mathbf{x}_t^H\} E\{\mathbf{1}_j^T \mathbf{x}_t \mathbf{x}_t^H \mathbf{1}_i\} + E\{\mathbf{x}_t \mathbf{x}_t^H \mathbf{1}_i\} E\{\mathbf{1}_j^T \mathbf{x}_t \mathbf{x}_t^H\} \\ &= \mathbf{R} \cdot \mathbf{1}_j^T \mathbf{R} \mathbf{1}_i + \mathbf{R} \mathbf{1}_i \mathbf{1}_j^T \mathbf{R}. \end{aligned} \quad (99)$$

Thus from (99) and (97), $\mathbf{C}_{z,1}(i, j) = \mathbf{R} \cdot \mathbf{1}_j^T \mathbf{R} \mathbf{1}_i$ and (22) is proved.

Derivation of $\mathbf{C}_{z,2}$: Note that $\mathbf{C}_{z,2} = E\{\text{vec}(\mathbf{Z}_t)\text{vec}(\mathbf{Z}_t)^T\} - \text{vec}(\mathbf{R})\text{vec}(\mathbf{R})^T$. The (i, j) -th $N \times N$ block of $\text{vec}(\mathbf{Z}_t)\text{vec}(\mathbf{Z}_t)^T$ is $\mathbf{x}_t \mathbf{x}_t^T x_{t,i}^* x_{t,j}$. The (i, j) -th $N \times N$ block of $\text{vec}(\mathbf{R})\text{vec}(\mathbf{R})^T$ is

$\mathbf{R}\mathbf{1}_i(\mathbf{R}\mathbf{1}_j)^T = \mathbf{R}\mathbf{1}_i\mathbf{1}_j^T\mathbf{R}^T = \mathbf{R}\mathbf{1}_i\mathbf{1}_j^T\mathbf{R}^*$. Then the (i, j) -th $N \times N$ block of $\mathbf{C}_{z,2}$ is given by

$$\begin{aligned} \mathbf{C}_{z,2}(i, j) &= E\{\mathbf{x}_t\mathbf{x}_t^T x_{t,i}^* x_{t,j}^*\} - \mathbf{R}\mathbf{1}_i\mathbf{1}_j^T\mathbf{R}^* \\ &= E\{\mathbf{x}_t\mathbf{x}_t^T \cdot \mathbf{x}_t^H \mathbf{1}_i \cdot \mathbf{1}_j^T \mathbf{x}_t^*\} - \mathbf{R}\mathbf{1}_i\mathbf{1}_j^T\mathbf{R}^*. \end{aligned} \quad (100)$$

In a similar way to (99), one obtains

$$\begin{aligned} &E\{\mathbf{x}_t\mathbf{x}_t^T \cdot \mathbf{x}_t^H \mathbf{1}_i \cdot \mathbf{1}_j^T \mathbf{x}_t^*\} \\ &= E\{\mathbf{x}_t\mathbf{x}_t^H \mathbf{1}_i\}E\{\mathbf{1}_j^T \mathbf{x}_t^* \mathbf{x}_t^T\} + E\{\mathbf{x}_t\mathbf{x}_t^H \mathbf{1}_j\}E\{\mathbf{1}_i^T \mathbf{x}_t^* \mathbf{x}_t^T\} \\ &= \mathbf{R}\mathbf{1}_i\mathbf{1}_j^T\mathbf{R}^* + \mathbf{R}\mathbf{1}_j\mathbf{1}_i^T\mathbf{R}^*, \end{aligned} \quad (101)$$

then $\mathbf{C}_{z,2}(i, j) = \mathbf{R}\mathbf{1}_j\mathbf{1}_i^T\mathbf{R}^*$. Using \mathbf{B} defined in (24), (23) is proved.

APPENDIX B PROOF OF LEMMA 1

The following Lemma B1 is easy to prove. It will be used extensively in this and later appendices.

Lemma B1: $\mathbf{1}_i^T \mathbf{1}_j = 0$ if $i \neq j$ and $\mathbf{1}_i^T \mathbf{1}_j = 1$ if $i = j$. \diamond

Proof of (72): The column vector $\mathbf{1}_k \otimes \mathbf{1}_l$ has the structure: $\mathbf{1}_k \otimes \mathbf{1}_l = [\mathbf{0}^T, \dots, \mathbf{1}_l^T, \dots, \mathbf{0}^T]^T$ where $\mathbf{1}_l^T$ only appears in the k -th segment. (Each segment has N elements.) Right-multiplying it to \mathbf{B} gives

$$\mathbf{B}(\mathbf{1}_k \otimes \mathbf{1}_l) = [(\mathbf{1}_k \cdot \mathbf{1}_1^T \mathbf{1}_l)^T, \dots, (\mathbf{1}_k \cdot \mathbf{1}_N^T \mathbf{1}_l)^T]^T \quad (102)$$

where only the l -th segment is nonzero and equal to $\mathbf{1}_k$. Thus (72) is proved.

Proof of (73): The result can be proved similarly.

Proof of (74): The (i, j) -th $N \times N$ block of $\mathbf{B}^T \mathbf{B}$ is equal to $\sum_{n=1}^N \mathbf{1}_n \cdot \mathbf{1}_i^T \mathbf{1}_j \cdot \mathbf{1}_n^T$. From Lemma B1, one knows that: when $i \neq j$, the block is $\mathbf{0}$; when $i = j$, the block is equal to $\sum_{n=1}^N \mathbf{1}_n \mathbf{1}_n^T = \mathbf{I}_N$. Thus $\mathbf{B}^T \mathbf{B} = \mathbf{I}_{N^2}$. Similarly one can prove that $\mathbf{B}\mathbf{B}^T = \mathbf{I}_{N^2}$. Thus the first equation is proved. The second equation is obvious by the definition of \mathbf{B} .

APPENDIX C PROOF OF THEOREM 2

The following lemma is obvious from the definitions of $\mathbf{C}_{z,1}$ and $\mathbf{C}_{z,2}$ in (15) and (16).

Lemma C1: (1) $\mathbf{C}_{z,1} = \mathbf{C}_{z,1}^H$; (2) $\mathbf{C}_{z,2} = \mathbf{C}_{z,2}^T$.

Proof of Theorem 2: Denote $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N]$. Then the n -th row of \mathbf{R} can be written as \mathbf{r}_n^H (because $\mathbf{R}^H = \mathbf{R}$).

According to (22), the $(N(i-1) + j)$ -th row of $\mathbf{C}_{z,1}$ is $\mathbf{r}_i^T \otimes \mathbf{r}_j^H$. According to (23), the $(N(i-1) + j)$ -th row of $\mathbf{C}_{z,2}$ is $\mathbf{r}_j^H \otimes \mathbf{r}_i^T$. Further using (73) of Lemma 1, (76) is proved.

Using Part (1) of Lemma C1, the $(N(i-1) + j)$ -th column of $\mathbf{C}_{z,1}$ is $(\mathbf{r}_i^T \otimes \mathbf{r}_j^H)^H = \mathbf{r}_i^* \otimes \mathbf{r}_j$. Using Part (2) of Lemma C1, the $(N(i-1) + j)$ -th column of $\mathbf{C}_{z,2}$ is $(\mathbf{r}_j^H \otimes \mathbf{r}_i^T)^T = \mathbf{r}_j^* \otimes \mathbf{r}_i$. From the above two results and (72) of Lemma 1, (75) is proved as well.

APPENDIX D PROOF OF LEMMA 2

Note that the first nonzero blocks of $\mathbf{U}_{1(2),k}$ have dimensions $kN \times k$ and the second one $N \times k$. Another representation of $\mathbf{J}_N^{(k)} = [\mathbf{1}_1, \dots, \mathbf{1}_k]_{N \times k}$ using $\mathbf{1}_l$, helps to understand the proof. *Orthonormal Columns of \mathbf{U}_1 :*

Case 1: $l < k$. In this case,

$$\begin{aligned} \mathbf{U}_{1,l}^T \mathbf{U}_{1,k} &= (1/2) [c_1 \mathbf{I}_l \quad (\mathbf{J}_N^{(l)})^T \mathbf{1}_{k+1} \quad \mathbf{0}_{l \times (k-l-1)}] \\ &\quad + (1/2) \mathbf{0}_{l \times N} \mathbf{J}_N^{(k)} \end{aligned} \quad (103)$$

where $c_1 = \mathbf{1}_{l+1}^T \mathbf{1}_{k+1}$ and $\mathbf{0}_{l \times N} \mathbf{J}_N^{(k)} = \mathbf{0}_{l \times k}$. From Lemma B1, $c_1 = 0$ and $(\mathbf{J}_N^{(l)})^T \mathbf{1}_{k+1} = \mathbf{0}_{l \times 1}$, then $\mathbf{U}_{1,l}^T \mathbf{U}_{1,k} = \mathbf{0}_{l \times k}$.

Case 2: $l = k$. In this case,

$$\mathbf{U}_{1,k}^T \mathbf{U}_{1,k} = (1/2)(\mathbf{I}_k + (\mathbf{J}_N^{(k)})^T \mathbf{J}_N^{(k)}) = \mathbf{I}_k. \quad (104)$$

Case 3: $l > k$. This part can be proved by transposing (103).

Further note that

$$\begin{aligned} (\mathbf{P}_0^T)^T \mathbf{U}_{1,k} &= (1/\sqrt{2}) \begin{bmatrix} \text{diag} [\mathbf{1}_1^T, \mathbf{1}_2^T, \dots, \mathbf{1}_k^T] (\mathbf{I}_N \otimes \mathbf{1}_{k+1}) \\ \mathbf{1}_{k+1}^T \mathbf{J}_N^{(k)} \\ \text{diag} [\mathbf{1}_{k+2}^T, \dots, \mathbf{1}_N^T] \mathbf{0}_{N(N-(k+1)) \times k} \end{bmatrix} \\ &= \mathbf{0}_{N \times k} \end{aligned} \quad (105)$$

and

$$\mathbf{P}_0 \mathbf{P}_0^T = \mathbf{I}_N. \quad (106)$$

Thus $\mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I}_{r_1}$ and the columns of \mathbf{U}_1 are orthonormal.

Orthonormal Columns of \mathbf{U}_2 : Due to a similar structure of \mathbf{U}_2 , (103) and (104) still hold true for \mathbf{U}_2 . It can then be similarly shown that \mathbf{U}_2 satisfies $\mathbf{U}_2^T \mathbf{U}_2 = \mathbf{I}_{r_2}$.

Orthogonality of \mathbf{U}_1 and \mathbf{U}_2 :

Case 1: $l < k$. In this case, the expression of $\mathbf{U}_{1,l}^T \mathbf{U}_{2,k}$ differs from (103) only in the sign for the last term, and thus the result is zero.

Case 2: $l = k$. In this case,

$$\mathbf{U}_{1,k}^T \mathbf{U}_{2,k} = (1/2)(\mathbf{I}_k - (\mathbf{J}_N^{(k)})^T \mathbf{J}_N^{(k)}) = \mathbf{0}_{k \times k}. \quad (107)$$

Case 3: $l > k$. The result can still be proved to be zero as in the case of $l < k$.

By the same argument as in (105), one can prove that

$$(\mathbf{P}_0^T)^T \mathbf{U}_{2,k} = \mathbf{0}_{N \times k}. \quad (108)$$

Combination of (107), (108) and the conclusion of Case 3, proves that $\mathbf{U}_1^T \mathbf{U}_2 = \mathbf{0}_{r_1 \times r_2}$.

Note that there are r_1 orthonormal (independent) vectors in \mathbf{U}_1 and r_2 orthonormal (independent) vectors in \mathbf{U}_2 . Since $r_1 + r_2 = N^2$, thus \mathbf{U}_1 and \mathbf{U}_2 are complements in a N^2 -dimensional space.

APPENDIX E
PROOF OF THEOREM 3

Eigenvectors With Respect to the Eigenvalue 1: Let \mathbf{B}_l be the l -th row block of \mathbf{B} (containing $((l-1)N+1)$ -st to (lN) -th rows). Then

$$\mathbf{B}_l \mathbf{U}_{1,k} = (1/\sqrt{2})(\mathbf{J}_N^{(k)} \cdot \mathbf{1}_l^T \mathbf{1}_{k+1} + \mathbf{1}_{k+1} \cdot \mathbf{1}_l^T \mathbf{J}_N^{(k)}). \quad (109)$$

Equation (109) will be analyzed for three different cases.

Case 1: $l \leq k$. In this case, from Lemma B1, $\mathbf{1}_l^T \mathbf{1}_{k+1} = 0$ and $\mathbf{1}_l^T \mathbf{J}_N^{(k)} \neq \mathbf{0}_{1 \times k}$, and then

$$\begin{aligned} \mathbf{B}_l \mathbf{U}_{1,k} &= (1/\sqrt{2}) \mathbf{1}_{k+1} [0, 0, \dots, \underbrace{1}_{l\text{-th position}}, \dots, 0] \\ &\rightarrow \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix} \mathbf{U}_{1,k} = (1/\sqrt{2}) \mathbf{I}_k \otimes \mathbf{1}_{k+1}. \end{aligned} \quad (110)$$

Case 2: $l = k+1$. In this case, from Lemma B1, $\mathbf{1}_l^T \mathbf{1}_{k+1} = 1$ and $\mathbf{1}_l^T \mathbf{J}_N^{(k)} = \mathbf{0}_{1 \times k}$, and then

$$\mathbf{B}_{k+1} \mathbf{U}_{1,k} = (1/\sqrt{2}) \mathbf{J}_N^{(k)}. \quad (111)$$

Case 3: $l > k+1$. In this case, also based on Lemma B1, both $\mathbf{1}_l^T \mathbf{1}_{k+1} = 0$ and $\mathbf{1}_l^T \mathbf{J}_N^{(k)} = \mathbf{0}_{1 \times k}$, and then

$$\mathbf{B}_l \mathbf{U}_{1,k} = \mathbf{0}_{N \times k} \rightarrow \begin{bmatrix} \mathbf{B}_{k+2} \\ \vdots \\ \mathbf{B}_N \end{bmatrix} \mathbf{U}_{1,k} = \mathbf{0}_{N(N-k-1) \times k}. \quad (112)$$

Combining (110), (111), (112) along with (84) proves that $\mathbf{B} \mathbf{U}_{1,k} = \mathbf{U}_{1,k}$. Additionally, it is easy to verify that $\mathbf{B} \mathbf{P}_0^T = \mathbf{P}_0^T$. Thus (88) is proved.

Eigenvectors With Respect to the Eigenvalue -1:

$$\mathbf{B}_l \mathbf{U}_{2,k} = (1/\sqrt{2})(\mathbf{J}_N^{(k)} \cdot \mathbf{1}_l^T \mathbf{1}_{k+1} - \mathbf{1}_{k+1} \cdot \mathbf{1}_l^T \mathbf{J}_N^{(k)}). \quad (113)$$

By comparing (109) and (113), one can see that the analysis for this part differs only in the negative sign of the second term in (113). Therefore one can also prove (89).

APPENDIX F
PROOF OF THEOREM 4

Proof of Part (1): Note that

$$\mathbf{P} \dot{\mathbf{U}} = \begin{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \end{bmatrix} \mathbf{U}_1 \\ \mathbf{P}_1 \mathbf{U}_2 \end{bmatrix}_{N^2 \times N^2}. \quad (114)$$

The top-left block is a $r_1 \times r_1$ matrix and the bottom one is a $r_2 \times r_2$ matrix.

$\mathbf{P}_1 \mathbf{U}_{1,k}$ has the following structure:

$$\begin{aligned} \mathbf{P}_1 \mathbf{U}_{1,k} &= \begin{bmatrix} \text{diag} \left[\mathbf{1}_k^{(N-1)}, \mathbf{1}_{k-1}^{(N-2)}, \dots, \mathbf{1}_1^{(N-k)} \right] \\ \mathbf{I}_N^{-(k+1)} \mathbf{J}_N^{(k)} \\ \mathbf{0}_{r_3 \times k} \end{bmatrix}_{r_2 \times k} \\ r_3 &= r_2 - [2N - (k+2)](k+1)/2 \end{aligned} \quad (115)$$

for $2 \leq k \leq N-2$ and

$$\mathbf{P}_1 \mathbf{U}_{1,N-1} = \text{diag} \left[\mathbf{1}_{N-1}^{(N-1)}, \mathbf{1}_{N-2}^{(N-2)}, \dots, \mathbf{1}_1^1 \right]_{r_2 \times (N-1)} \quad (116)$$

where $\mathbf{1}_k^{(N-l)}$ ($1 \leq l \leq k \leq N-1$) is the k -th column of the identity matrix \mathbf{I}_{N-l} and $\mathbf{I}_N^{-(k+1)} \mathbf{J}_N^{(k)} = \mathbf{0}_{(N-k) \times k}$.

Let us divide the rows of $\mathbf{P}_1 \mathbf{U}_1$ into $N-1$ block rows with the first block row containing the first $N-1$ rows, the second block row containing the following $N-2$ rows, and so on, until the last block row containing the last row only. According to the structures of (115) and (116), the l -th ($1 \leq l \leq k$) column of $\mathbf{P}_1 \mathbf{U}_{1,k}$ has non-zero segment $\mathbf{1}_k^{(k-l+1)}$ which falls within the l -th block row; and therefore for the l -th block row, $\mathbf{P}_1 \mathbf{U}_{1,k}$ contains non-zero column segments $\mathbf{1}_1^{(N-l)}, \mathbf{1}_2^{(N-l)}, \dots, \mathbf{1}_{N-l}^{(N-l)}$. Those column segments are all the (independent) columns of \mathbf{I}_{N-l} . Let $\mathbf{U}'_1 = [\mathbf{U}_{1,1}, \dots, \mathbf{U}_{1,N-1}]$. Then $\mathbf{P}_1 \mathbf{U}'_1$ has the same rank of r_2 as the matrix

$$\text{diag} [\mathbf{I}_{N-1}, \mathbf{I}_{N-2}, \dots, \mathbf{I}_2, 1] \quad (117)$$

which is a column exchanged version of it and has a full rank of r_2 . From (105) and (106), $\mathbf{P}_0 \mathbf{U}_1 = [\mathbf{P}_0 \mathbf{P}_0^T, \mathbf{P}_0 \mathbf{U}'_1] = [\mathbf{I}_N, \mathbf{0}_{N \times r_2}]$; and $\mathbf{P}_1 \mathbf{U}_1 = [\mathbf{P}_1 \mathbf{P}_0^T, \mathbf{P}_1 \mathbf{U}'_1] = [\mathbf{0}_{r_2 \times N}, \mathbf{P}_1 \mathbf{U}'_1]$. Hence the top-left block (square) matrix of (114) has a full rank of $r_2 + N = r_1$.

$\mathbf{P}_1 \mathbf{U}_{2,k}$ has the block $-\mathbf{I}_N^{-(k+1)} \mathbf{J}_N^{(k)}$ (which is zero) and $\mathbf{P}_1 \mathbf{U}_2$ has the same structure as $\mathbf{P}_1 \mathbf{U}'_1$. Thus the bottom-right block (square) matrix of (114) has a full rank of r_2 .

Hence, $\mathbf{P} \dot{\mathbf{U}}$ is of full rank N^2 . \diamond

The following lemma is needed for the proof of Part (2).

Lemma F1: $\mathbf{C}_{z,1}$ is of full rank and is positive definite.

Proof: Since \mathbf{R} has a full rank, by Corollary 13.11 of [15], $\mathbf{C}_{z,1}$ is of full rank. Let the eigenvalue decomposition (EVD) of \mathbf{R} be $\mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^H$ where $\boldsymbol{\Sigma}$ is a diagonal matrix containing N positive eigenvalues of \mathbf{R} and \mathbf{U} contains the corresponding orthonormal eigenvectors. Then the EVD of $\mathbf{C}_{z,1}$ can be written as $(\mathbf{U}^* \otimes \mathbf{U})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})(\mathbf{U}^T \otimes \mathbf{U}^H)$. Further from the full rank property of $\mathbf{C}_{z,1}$, one knows that $\mathbf{C}_{z,1}$ is positive definite. \diamond

Proof of Part (2): We will first show that $\dot{\mathbf{C}}_{z,1}$ is of full rank $2N^2$. According to Lemma F1, $\mathbf{C}_{z,1}$ can be written as $\mathbf{G}^H \mathbf{G}$ where \mathbf{G} is an $N^2 \times N^2$ complex matrix with full rank. Let

$$\dot{\mathbf{G}} = \begin{bmatrix} \Re\{\mathbf{G}\} & -\Im\{\mathbf{G}\} \\ \Im\{\mathbf{G}\} & \Re\{\mathbf{G}\} \end{bmatrix}. \quad (118)$$

Then one can write $\dot{\mathbf{C}}_{z,1} = \dot{\mathbf{G}}^T \dot{\mathbf{G}}$. In a similarly way to the proof of the full column rank of the matrix in [17, (65)], it can

be shown that $\hat{\mathbf{G}}$ has a full rank. So does $\hat{\mathbf{C}}_{z,1}$. Further note that $\hat{\mathbf{U}}$ has a full column rank of N^2 . Thus $\hat{\mathbf{U}}^T \hat{\mathbf{C}}_{z,1} \hat{\mathbf{U}}$ is of full rank N^2 . \diamond

APPENDIX G A COMPACT EXPRESSION OF $\hat{\mathbf{R}}_a$

Let $\bar{\mathbf{X}}_t = \mathbf{I}_L \otimes \mathbf{x}_t$ be an $NL \times L$ matrix. Define $\mathbf{B}_i = \mathbf{J}_i / \|\mathbf{J}_i\|^2$ and the $NL \times NL$ (real) matrix

$$\bar{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \cdots & \mathbf{B}_{L-2} & \mathbf{B}_{L-1} \\ \mathbf{B}_1^T & \mathbf{B}_0 & \cdots & \mathbf{B}_{L-3} & \mathbf{B}_{L-2} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{B}_{L-2}^T & \mathbf{B}_{L-3}^T & \cdots & \mathbf{B}_0 & \mathbf{B}_1 \\ \mathbf{B}_{L-1}^T & \mathbf{B}_{L-2}^T & \cdots & \mathbf{B}_1^T & \mathbf{B}_0 \end{bmatrix}. \quad (119)$$

Form (35)–(37), one can write

$$\hat{\mathbf{R}}_a = (1/T) \sum_{t=1}^T \bar{\mathbf{X}}_t^T \bar{\mathbf{B}} \bar{\mathbf{X}}_t^*. \quad (120)$$

For the 2-level nested array in [4], $N = 6$ and $L = 12$. By MATLAB evaluation, it is found that $\bar{\mathbf{B}}$ has 42 negative eigenvalues. Thus the estimated augmented covariance matrix $\hat{\mathbf{R}}_a$ for this type of array is generally in-definite.

APPENDIX H IMPLEMENTATION OF THE SNSE ALGORITHM

The SNSE algorithm consists of five steps (in the last paragraph of [3, p. 966]) which are implemented by the iterative minimization procedures [3, (68), (51), (62) and (39)]. Each minimization procedure includes two nested loops: the inner-loop (ϵ -loop) and the outer-loop (t -loop). ϵ in the ϵ -loop is the step size of the maximum perturbation. Each loop is allowed to iterate depending on certain stopping criteria, which are generally application dependent. Thus in this appendix, the choice of stopping criteria is described.

For the convenience of the exposition in the remaining part of this appendix, the five steps of the SNSE algorithm are summarized below where all equation numbers referred to are from [3]:

- Step 1: From $\hat{\mathbf{R}}_a$, solve the linear programming (LP) problem (68)–(72) to obtain the positive definite Toeplitz matrix $\hat{\mathbf{T}}_0$ as given in (72). (To our understanding, $-p_0$ in (71) is a typo. It was replaced by p_0 (an estimate of the noise variance) in our implementation.)
- Step 2: From $\hat{\mathbf{T}}_0$, solve the LP problems (51)–(55) and (62)–(66) to obtain the Toeplitz matrix $\hat{\mathbf{T}}_{ml}^u$.
- Step 3: From $\hat{\mathbf{T}}_{ml}^u$, solve the LP problem (68)–(72) to obtain the Toeplitz matrix $\hat{\mathbf{T}}_0^{opt}$.
- Step 4: From $\hat{\mathbf{T}}_0^{opt}$, solve the LP problems (51)–(55)+(59) and (62)–(67) to obtain the positive definite Toeplitz matrix $\hat{\mathbf{T}}_{ml}$.
- Step 5: From $\hat{\mathbf{T}}_{ml}$, solve the LP problem (39)–(42) to obtain the Toeplitz matrices $\hat{\mathbf{T}}_\mu$, for $\mu = 0, 1, 2, \dots, L-2$, and $\hat{\mathbf{T}}_{L-1} = \hat{\mathbf{T}}_{ml}$.

The main idea of the SNSE algorithm is to introduce a small perturbation to an existing Toeplitz matrix as described in [3, (26)], so that the noise eigenvalues of the updated Toeplitz matrix (on the left-hand side of [3, (26)]), have a smaller span, as

compared with that for the existing matrix. Due to this reason, the ϵ -loop should be performed for small ϵ 's. If a large ϵ is used, those minimization procedures may finish with a feasible solution, but the perturbation to the Toeplitz matrix can become too large and the updated eigenvalues may not be close to those as indicated by the minimization. (In that case, we found in simulations, that the span of the noise eigenvalues is not reduced.) Thus in our implementation, the initial value of ϵ is chosen as the minimum absolute value of all the elements of the starting Toeplitz matrix in a step.

According to [3, (45)], the eigenvalues calculated from the updated Toeplitz matrix should be close to that indicated by the first order perturbation. This means that eigenvalue increments (perturbations) should be small enough. This requirement can be used as a stopping criterion of the ϵ -loop. In our implementation, the ϵ -loop stops iteration if the increment ratio is smaller than τ_ϵ , where the increment ratio is defined as the norm of the vector of the eigenvalue increments against the norm of the vector of the existing eigenvalues. If the increment ratio is greater than τ_ϵ , the ϵ -loop is iterated with the current ϵ being halved. The t -loop is iterated until the norm of the difference between the Toeplitz matrices obtained in two successive t -loops is less than τ_t . The matrices obtained in all the t -loops are compared and the one with the largest likelihood ratio is chosen as the final Toeplitz matrix in each step. Both the ϵ -loop and the t -loop were repeated for no more than 10 times. Through extensive simulation investigation, we found that $\tau_\epsilon = 0.2$ and $\tau_t = 0.1$ can provide better Toeplitz matrices (in the likelihood ratio sense) and maintain the first-order approximation at the same time.

In the first four steps, the increment ratio is defined for all the eigenvalues of the matrix

$$\mathbf{G} \stackrel{def}{=} (\hat{\mathbf{R}})^{-1/2} \mathcal{R}(\tilde{\mathbf{T}}) (\hat{\mathbf{R}})^{-1/2} \quad (121)$$

where $(\hat{\mathbf{R}})^{-1/2}$ is the inverse square-root of $\hat{\mathbf{R}}$, \mathcal{R} stands for transformation from an $L \times L$ Toeplitz matrix to an $N \times N$ measurement covariance matrix and $\tilde{\mathbf{T}}$ is a Toeplitz matrix given by any of those minimization procedures. The transformation \mathcal{R} is defined as

$$\mathcal{R}(\tilde{\mathbf{T}}) = r_0 \mathbf{I} + \sum_{l=1}^{11} r_{-l} \mathbf{J}_l + \sum_{l=1}^{11} r_l \mathbf{J}_l^T \quad (122)$$

where r_0, r_1, \dots, r_{L-1} are the L covariance lags of $\tilde{\mathbf{T}}$ and covariance lags satisfy the property $r_{-l} = r_l^*$ for $l \in [1, L-1]$. In Step 5, only the last $(L-\mu)$ eigenvalues of the Toeplitz matrix $\hat{\mathbf{T}}_\mu$ were involved in the noise eigenvalue equalization. Thus in Step 5, the increment ratio is defined for the last $(L-\mu)$ eigenvalues of the Toeplitz matrices $\hat{\mathbf{T}}_\mu$ for $\mu = 1, 2, \dots, L-1$. The likelihood ratio of a Toeplitz matrix $\tilde{\mathbf{T}}$ is defined as

$$LR(\tilde{\mathbf{T}}) = \prod_{i=1}^N \gamma_i^{-1} / \left(\sum_{i=1}^N \gamma_i^{-1} / N \right) \quad (123)$$

where $\gamma_1, \dots, \gamma_N$ are the eigenvalues of the corresponding matrix \mathbf{G} defined in (121). p_0 in [3, (59), (67), (71) and (72)] is the noise variance estimate. It was chosen as $\sqrt{\hat{\lambda}_L}$ in Steps 1–4 and was given by (39) with M being replaced by m under \mathcal{H}_m in Step 5.

Based on the aforementioned initial value of ϵ and stopping criteria for the ϵ -loop and the t -loop, and the angles in the following section, we ran 1000 simulations and discovered, from a statistical point of view, that (1) $LR(\tilde{\mathbf{T}}_{ml})$ did not significantly improve $LR(\tilde{\mathbf{T}}_0^{opt})$; and (2) $LR(\tilde{\mathbf{T}}_0^{opt})$ was very close to $LR(\tilde{\mathbf{T}}_0)$. Thus in our further simulation examples (as presented in the following section), Step 3 and Step 4 were omitted; and $\tilde{\mathbf{T}}_{ml}$ in Step 5 is replaced by $\tilde{\mathbf{T}}_{ml}^u$ obtained in Step 2. Note that $\tilde{\mathbf{T}}_{ml}^u$ is not guaranteed to be positive definite. Whenever it is not positive definite, the loading operation as given by the second expression of [3, (72)] is applied to $\tilde{\mathbf{T}}_{ml}^u$. The loading operation only slightly affects the likelihood ratio values of $\tilde{\mathbf{T}}_{ml}^u$. The matrix transformation sequence is thus given by $\hat{\mathbf{R}}_a \rightarrow \tilde{\mathbf{T}}_0 \rightarrow \tilde{\mathbf{T}}_{ml}^u \rightarrow \tilde{\mathbf{T}}_\mu$, $\mu = 0, 1, 2, \dots, L-2$ and $\tilde{\mathbf{T}}_{L-1} = \tilde{\mathbf{T}}_{ml}^u$.

Let the eigenvalues of $\tilde{\mathbf{T}}_\mu$ be $\gamma_1^\mu, \gamma_2^\mu, \dots, \gamma_{\mu+1}^\mu, \dots, \gamma_L^\mu$ (organized in non-ascending order) and define the admissible set of the hypothesized number of sources by

$$\mathcal{M} = \{\mu | \gamma_{\mu+1}^\mu - \gamma_L^\mu \leq \hat{\sigma}^{(\mu)}, \mu \in [0, L-2]\}. \quad (124)$$

The detection step of the SNSE algorithm produces an estimate of the number of sources as

$$\hat{M}_{snse} = \arg \min_{\mu \in \mathcal{M}} \{\mu | LR(\tilde{\mathbf{T}}_\mu) \geq \alpha LR(\tilde{\mathbf{T}}_{ml}^u)\} \quad (125)$$

and $\hat{M}_{snse} = L-1$ if the solution given by (125) is null. In [3], α was determined based on an over-estimation probability. We found that, for very large T (say, $T = 5000$), the under-estimation probability is very low, and over-estimation accounts for almost all incorrect detections; but for not very large T (say, $T = 1000$), the under-estimation probability is not negligible, and thus in simulation, we chose α to achieve a maximum probability of correct detection. Note that the detection range of the SNSE algorithm is $[0, L-1]$.

ACKNOWLEDGMENT

The authors would like to thank two anonymous reviewers for their insightful comments and suggestions which have helped to improve the presentation of this paper.

REFERENCES

- [1] S. U. Pillai, Y. Bar-Ness, and F. Haber, "A new approach to array geometry for improved spatial spectrum estimation," *Proc. IEEE*, vol. 73, pp. 1522–1524, Oct. 1985.
- [2] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer, "Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays. I. Fully augmentable arrays," *IEEE Trans. Signal Process.*, vol. 46, pp. 2458–2471, Sep. 1998.
- [3] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Detection-estimation of more uncorrelated Gaussian sources than sensors in nonuniform linear arrays. I. Fully augmentable arrays," *IEEE Trans. Signal Process.*, vol. 49, pp. 959–971, May 2001.
- [4] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, pp. 4167–4181, Aug. 2010.

- [5] B. Porat and B. Friedlander, "Direction finding algorithms based on high-order statistics," *IEEE Trans. Signal Process.*, vol. 39, pp. 2016–2024, Sep. 1991.
- [6] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, "On the virtual array concept for higher order array processing," *IEEE Trans. Signal Process.*, vol. 53, pp. 1254–1271, Apr. 2005.
- [7] P. Pal and P. P. Vaidyanathan, "Multiple level nested arrays: An efficient geometry for $2q$ th order cumulant based array processing," *IEEE Trans. Signal Process.*, vol. 60, pp. 1253–1269, Mar. 2012.
- [8] A. T. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, pp. 172–175, Mar. 1968.
- [9] W. Chen, K. M. Kong, and J. P. Reilly, "Detection of the number of signals: A predicted eigen-threshold approach," *IEEE Trans. Signal Process.*, vol. 39, pp. 1088–1098, May 1991.
- [10] H. Li, P. Sotica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrix," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, May 1999.
- [11] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
- [12] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York, NY, USA: Wiley, 2005.
- [13] B. D. Rao and K. V. S. Hari, "Performance of MUSIC," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1939–1949, Dec. 1989.
- [14] A. Graham, *Kronecker Products and Matrix Calculus With Applications*. Chichester, U.K.: Ellis Horwood, 1981.
- [15] A. Laub, *Matrix Analysis for Scientists and Engineers*. Philadelphia, PA, USA: SIAM, 2005.
- [16] S. U. Pillai and B. H. Kwon, "Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1176–1189, Aug. 1989.
- [17] Q. Cheng and Y. Hua, "Detection of cisoids using least square error function," *IEEE Trans. Signal Process.*, vol. 45, no. 6, pp. 1584–1590, Jun. 1997.
- [18] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*. Berkeley, CA, USA: Univ. of California Press, 1958.
- [19] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
- [20] 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.
- [21] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 8, pp. 1190–1196, Aug. 1989.
- [22] 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.
- [23] 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.
- [24] L. Huang and H. C. So, "Source Enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4806–4821, Oct. 2013.

Qi Cheng (M'95) received the Ph.D. degree in 1995. He is currently a Lecturer at the University of Western Sydney.

Piya Pal (S'08–M'12) received her Ph.D. degree in 2013. She is currently an Assistant Professor at the University of Maryland, College Park.

Masashi Tsuji (S'08–M'13) received the Ph.D. degree in 2013 and is currently working in Toshiba Corporation.

Yingbo Hua (F'02) received the Ph.D. degree in 1988. He is currently a Professor at the University of California at Riverside.