IEEE TRANSACTIONS ON

SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

www.ieee.org/sp/index.html

NOVEMBER 2001 VOLUME 49 NUMBER 11 ITPRED (ISSN 1053–587X)

PAPERS

(Contents Continued on Back Cover)

MSE-Based Regularization Approach to Direction Estimation of Coherent Narrowband Signals Using Linear Prediction

Jingmin Xin, *Member, IEEE*, and Akira Sano, *Member, IEEE*

***Abstract*—This paper addresses the problem of directions-of-arrival (DOAs) estimation of coherent narrowband signals impinging on a uniform linear array (ULA) when the number of signals is unknown. By using an overdetermined linear prediction (LP) model with a subarray scheme, the DOAs of coherent signals can be estimated from the zeros of the corresponding prediction polynomial. Although the corrected least squares (CLS) technique can be used to improve the accuracy of the LP parameters estimated from the noisy array data, the inversion of the resulting matrix in the CLS estimation is ill-conditioned, and then, the CLS estimation becomes unstable. To combat this numerical instability, we introduce multiple regularization parameters into the CLS estimation and show that determining the number of coherent signals is closely related to the truncation of the eigenvalues. An analytical expression of the mean square error (MSE) of the estimated LP parameters is derived, and it is clarified that the number of signals can be determined by comparing the optimal regularization parameters with the corresponding eigenvalues. An iterative regularization algorithm is developed for estimating directions without any *a priori* knowledge, where the number of coherent signals and the noise variance are estimated from the noise-corrupted received data simultaneously.**

***Index Terms*—Array processing, eigenvalue decomposition (EVD), linear prediction (LP), mean square error (MSE), regularization.**

I. Introduction

ARRAY signal processing is used in many fields to extract the desired information from data received at an array of sensors. In these applications, the estimation of the directions-of-arrival (DOAs) of signals from the noisy data is a major task. To estimate the DOAs of narrowband signals, maximum likelihood (ML) methods [41], [42], [51]–[53], [48], [56] and subspace-based methods are well known [59]. In general, subspace-based methods have attracted considerable attention because of their relatively high resolution capability and low computational complexity. Typical subspace-based methods include the Pisarenko method [1], multiple signal classification (MUSIC) [2], estimation of signal parameters via rotational invariance techniques (ESPRIT) [3], minimum norm (Min-Norm) [4], method of direction estimation (MODE) [50], [51], [54] [for a uniform linear array (ULA)], and weighted subspace fitting (WSF) [43], [48], [49].

In the Min-Norm linear prediction (LP) method for a ULA [4], [5], singular value decomposition (SVD) is applied to an overdetermined LP data matrix with a truncation to reduce the noise effect and to mitigate the ill-conditioned nature of the conventional LP method. The accuracy of the estimated LP parameters is obtained in the sense of least squares, and the DOAs of the incident signals are estimated from the zeros of a polynomial formed from the LP parameters. It has been shown that the Min-Norm LP method [4], [5] provides excellent resolution in moderately low signal-to-noise ratio (SNR) environments [6]. As studied in [7], in the LP-based direction finding methods, the reliable estimation of the LP parameters is clearly an important issue. To improve the accuracy of the LP parameters estimated from the noisy array data, the total least squares (TLS) technique was used to reduce the noise effect from both the observation vector and the data matrix [8].

Unfortunately, like the other subspace-based methods except MODE and WSF, the Min-Norm LP and TLS-LP methods [4], [8] suffer serious degradation when the incident signals are mutually coherent in the practical scenarios due to multipath propagation in which the rank of the source signal covariance matrix is reduced. To improve the performance of the subspace-based direction estimation algorithms, some methods such as spatial smoothing (SS) [9], smoothed eigenvector [10], and eigenvector smoothing [11] have been proposed to combat the deleterious effect due to coherency. Inspired by SS preprocessing in which the subarray covariance matrices are averaged for decorrelation, a subarray averaging was used to produce a noise-free reduced-rank approximation for an array data matrix, and a smoothed Min-Norm LP method was proposed to estimate the DOAs of coherent signals [12], [13]. By incorporating a subarray scheme with the LP model, new TLS-LP methods were developed to accurately estimate the directions of impinging signals in the presence of multipath propagation [14], [15]. However, these direction estimation methods for coherent signals require *a priori* knowledge of the number of incident signals, as do most of the subspace-based methods [1]–[4], [50]. Because the number of coherent or noncoherent signals is usually unknown, a detection procedure [16]–[23] generally must precede the direction estimation. The ML methods [18], [42], [56], [20] and the WSF method [43], [48] are generally the optimal solution to the detection and estimation problem. In these methods, the solution of direction finding is required for each of several hypotheses.

Manuscript received September 10, 1999; revised August 6, 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Vikram Krishnamurthy.

J. Xin is with YRP Mobile Telecommunications Key Technology Research Laboratories Co., Ltd., Yokosuka, Japan (e-mail: xin@yrp-ktrl.co.jp).

A. Sano is with the Department of System Design Engineering, Keio University, Yokohama, Japan (e-mail: sano@sd.keio.ac.jp).

Publisher Item Identifier S 1053-587X(01)09237-6.

In addition, MODE and WSF are the statistically efficient direction estimator in cases when either the number of snapshots or the SNR is sufficiently large [48], [52]. In general, MODE needs the number of signals and the dimension of signal subspace [50], [51], and the WSF detection scheme can be combined with MODE to provide the estimated number of signals for MODE/WSF [43], [48]. However, when the number of snapshots is finite, it will be difficult to choose an appropriate threshold for the sequential hypothesis testing of WSF detection scheme without any *a priori* knowledge [16], [20], [21], [43].

In this paper, we investigate the problem of estimating the DOAs of coherent narrowband signals impinging on a ULA when the number of signals is unknown. A new mean-square-error (MSE) based regularization approach is developed by using the LP model of array data with an appropriate overdetermined order (larger than the number of signals) and a subarray scheme. In general, the ordinary least squares (LS) estimate of the LP parameters from the noise-corrupted array data becomes biased [24], [7]. The corrected least squares (CLS) method [25], [26] is a refinement of the LS method in attempt to obtain a consistent estimate by combating the noise effect in both the data matrix and data vector, and it is asymptotically equivalent to the TLS method [27]. However, because the estimated noisy covariance matrix is subtracted by a noise covariance matrix, the resulting matrix has a reduced rank, and then, the CLS estimation of the LP parameters will be ill conditioned so that adequate truncation of the eigenvalues of the resulting matrix should be carried out [28], [29]. To perform this truncation, the noise variance and the number of signals are obviously required. Unfortunately, they are unknown in practice.

Therefore, we introduce multiple regularization parameters into the CLS estimation to stabilize the estimate of the LP parameters. We find that the regularization parameters minimizing the MSE of the estimated LP parameters give an optimal truncation of the eigenvalues. The number of signals can then be determined from the number of retained eigenvalues. An asymptotic MSE is derived through the calculations of the third- and fourth-order moments of the additive noise, and an analytical expression of the optimal regularization parameters is provided. A data-based iterative algorithm is developed to estimate the DOAs of the coherent signals without any *a priori* knowledge in which the noise variance and number of coherent signals are estimated simultaneously, and the convergence behavior of the iterative algorithm is analyzed. The performance of the proposed approach is demonstrated and compared with that of the conventional methods through numerical examples. The simulation results show that the proposed method performs better than the WSF detection scheme [43], [48] when the number of snapshots is small and that it outperforms the SS-MDL and SS-AIC methods [16], [17] in detecting the number of signals. In addition, when the number of signals is correctly estimated, the proposed method is superior to SS-based root-MUSIC [9], [58], the smoothed Min-Norm [13], and the CLS method (with the true number of signals or noise variance) in resolving closely spaced coherent signals.

The data model and decorrelation using the LP technique are presented in Section II. Section III discusses the MSE-based regularization approach to CLS estimation. In Section IV, the direction estimation from the received array data is derived. The effectiveness of the proposed approach is evaluated in Section V, and the paper is summarized in Section VI.

II. Problem Formulation

A. Data Model

We consider a linear array of M equally spaced omnidirectional sensors with spacing d. We assume that q ($q < M$) narrowband signals with center frequency f_0 are far from the array and impinge on the array from distinct directions $\{\theta_k\}$. Under the narrowband assumption, the wavefronts can be approximated as planar. Then, the propagation of a wavefront between sensors can be modeled as a simple phase delay, and all the signals are the baseband equivalents [48], [59], [60]. Then, the signal received by the ith sensor is the superimposition of q impinging wavefronts and the additive noise, which can be expressed as

$$y_i(n) = x_i(n) + w_i(n) \tag{1}$$

where $x_i(n)$ is the noiseless received signal, and $w_i(n)$ is the additive noise. The signal $x_i(n)$ is represented as

$$x_i(n) = \sum_{k=1}^{q} s_k(n) e^{j\omega_0(i-1)\tau_k} \tag{2}$$

where the incident signal $s_k(n)$ is the zero-mean random process with the direction θ_k measured relative to the normal of array, $\omega_0 = 2\pi f_0$, $\tau_k = (d/c)\sin\theta_k$, and c is the propagation speed.

The received data can thus be described by using vector-matrix notation as

$$\boldsymbol{y}(n) = \boldsymbol{A}(\theta)\boldsymbol{s}(n) + \boldsymbol{w}(n) \tag{3}$$

where $\boldsymbol{y}(n)$, $\boldsymbol{s}(n)$, and $\boldsymbol{w}(n)$ are the vectors of the received data, the incident signals, and the additive noise given by $\boldsymbol{y}(n) = [y_1(n), y_2(n), \ldots, y_M(n)]^T$, $\boldsymbol{s}(n) = [s_1(n), s_2(n), \ldots, s_q(n)]^T$, and $\boldsymbol{w}(n) = [w_1(n), w_2(n), \ldots, w_M(n)]^T$, and $\boldsymbol{A}(\theta)$ is the array response matrix given by $\boldsymbol{A}(\theta) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_q)]$, in which $\boldsymbol{a}(\theta_k) = [1, e^{j\omega_0\tau_k}, \ldots, e^{j\omega_0(M-1)\tau_k}]^T$, and $(\cdot)^T$ denotes transposition.

In this paper, we make the following assumptions in the derivation of the algorithm.

Assumption A: The ULA is calibrated, and the array response matrix $\boldsymbol{A}(\theta)$ is unambiguous, i.e., for any collection of distinct $\{\theta_1, \theta_2, \ldots, \theta_q\}$, the corresponding vectors $\{\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_q)\}$ are linearly independent. Equivalently, the matrix $\boldsymbol{A}(\theta)$ must have full rank.

Assumption B: Without loss of generality, the impinging signals are coherent and are expressed as

$$s_k(n) = \beta_k s_1(n), \qquad \text{for } k = 1, 2, \ldots, q \tag{4}$$

where β_k is the multipath coefficient representing the complex attenuation of the kth signal with respect to the first one $s_1(n)$, $\beta_k \neq 0$, and $\beta_1 = 1$.

Assumption C: The additive noise components $\{w_i(n)\}$ are temporally and spatially white complex Gaussian noise with zero-mean and variance σ^2, and they are uncorrelated with the source signals. The noise variance is given by

$$E\{w_i(n)w_k^*(n)\} = \sigma^2\delta_{i,k} \quad \text{and} \quad E\{w_i(n)w_k(n)\} = 0 \tag{5}$$

for $i, k = 1, 2, \ldots, M$, where $\delta_{i,k}$ is the Kronecker delta, and $E\{\cdot\}$ and the asterisk denote the expectation and complex conjugate.

Remark A: In this paper, a crucial assumption is that the equally spaced linear array is calibrated so that the noiseless signals received at each sensor obey a linear difference equation exactly, i.e., the calibration and other model errors are not considered. Additionally, the additive noise is assumed to be temporally and spatially white Gaussian noise. If the spatial whiteness condition is not met, we can prewhiten the array data by linearly transforming the estimated covariance matrix of the noise, which can be estimated from measurements with no signal present [48], [59]. More specifically, if the noise covariance matrix is $\boldsymbol{Q}$, the received array data are multiplied by $\boldsymbol{Q}^{-1/2}$, which denotes a Hermitian square-root factor of $\boldsymbol{Q}$. Furthermore, although we assume that the incident signals are fully coherent for simplicity, as shown in Remark B, it is straightforward to extend the proposed method to the case of partly coherent or incoherent signals. □

In array signal processing, the problem of estimating the signal parameters from the sensor measurements has received much attention, and many algorithms have been proposed. However, most of them require knowledge of the number of signals impinging on the array. In this paper, we consider the direction estimation of coherent signals from the noise-corrupted array data when the number of signals and the noise variance are unknown.

B. Linear Prediction and Decorrelation

It is well known that the source signal covariance matrix will be singular (i.e., rank deficient) in coherent situations so that the number of signals impinging on the array cannot be estimated directly from the multiplicity of the eigenvalues of the array covariance matrix. To circumvent this crucial rank deficit problem, specific modifications such as spatial smoothing [9], [30] have been suggested to decorrelate the signal coherency and restore the rank of the source signal covariance matrix to the number of arriving signals [17]. In this paper, we consider a new approach for estimating the signal parameters in the coherent case by using the LP technique and a subarray scheme.

In essence, the DOAs are estimated by using the time delays (phase differences) of signals impinging on the individual sensors in the array. From the spatial property of array data where the noiseless signal received at different sensors are phase-shifted versions of one another, the noiseless received signals $\{x_i(n)\}$ in (2) obey a linear difference equation [4], [5], [31]–[33]. By dividing the total array into L overlapping forward subarrays with m sensors where $m > q+1$ and $L = M - m + 1$, i.e., the lth subarray comprises $\{l, l+1, \ldots, l+m-1\}$ sensors, the signal $x_{l+m-1}(n)$ can be predicted from a linear combination of the other signals

$$x_{l+m-1}(n) = \boldsymbol{x}_l^T(n)\boldsymbol{a}, \qquad \text{for } l = 1, 2, \ldots, L \tag{6}$$

where $\boldsymbol{x}_l(n) = [x_l(n), x_{l+1}(n), \ldots, x_{l+m-2}(n)]^T$, $\boldsymbol{a} = [a_{m-1}, a_{m-2}, \ldots, a_1]^T$, and $\{a_i\}$ are the coefficients of the LP model. Here, m and L are the subarray size and the number of subarrays. The order of the LP model is $m-1$; therefore, the model order is larger than the number of signals as we have assumed that $m > q+1$.

By substituting (1) into (6), the noise-corrupted signal $y_{l+m-1}(n)$ can be expressed using the forward linear prediction (FLP) model

$$y_{l+m-1}(n) = \boldsymbol{y}_l^T(n)\boldsymbol{a} + \varepsilon_{l+m-1}(n) \tag{7}$$

where $\boldsymbol{y}_l(n) = [y_l(n), y_{l+1}(n), \ldots, y_{l+m-2}(n)]^T$, $\varepsilon_{l+m-1}(n)$ is the prediction error given by $\varepsilon_{l+m-1}(n) = \overline{\boldsymbol{w}}_l^T(n)\overline{\boldsymbol{a}}$, $\overline{\boldsymbol{w}}_l(n) = [\boldsymbol{w}_l^T(n), w_{l+m-1}(n)]^T$, $\boldsymbol{w}_l(n) = [w_l(n), w_{l+1}(n), \ldots, w_{l+m-2}(n)]^T$, and $\overline{\boldsymbol{a}} = [-\boldsymbol{a}^T, 1]^T$.

For $l = 1$ to L, from (7), we have a compact vector-matrix form to express the array data as

$$\boldsymbol{z}(n) = \boldsymbol{\Phi}(n)\boldsymbol{a} + \boldsymbol{\varepsilon}(n) \tag{8}$$

where $\boldsymbol{z}(n) = [y_m(n), y_{m+1}(n), \ldots, y_M(n)]^T$, $\boldsymbol{\Phi}(n) = [\boldsymbol{y}_1(n), \boldsymbol{y}_2(n), \ldots, \boldsymbol{y}_L(n)]^T$, and $\boldsymbol{\varepsilon}(n) = [\varepsilon_m(n), \varepsilon_{m+1}(n), \ldots, \varepsilon_M(n)]^T$. From (1), we have $\boldsymbol{z}(n) = \boldsymbol{z}_s(n) + \boldsymbol{z}_w(n)$, and $\boldsymbol{\Phi}(n) = \boldsymbol{\Phi}_s(n) + \boldsymbol{W}(n)$, where $\boldsymbol{z}_s(n) = [x_m(n), x_{m+1}(n), \ldots, x_M(n)]^T$, $\boldsymbol{z}_w(n) = [w_m(n), w_{m+1}(n), \ldots, w_M(n)]^T$, $\boldsymbol{\Phi}_s(n) = [\boldsymbol{x}_1(n), \boldsymbol{x}_2(n), \ldots, \boldsymbol{x}_L(n)]^T$, and $\boldsymbol{W}(n) = [\boldsymbol{w}_1(n), \boldsymbol{w}_2(n), \ldots, \boldsymbol{w}_L(n)]^T$.

Under the assumptions of the data model, we easily see that the covariance matrix $\boldsymbol{\Sigma}$ of the data matrix $\boldsymbol{\Phi}(n)$ in (8) can be represented as the sum of the noiseless (i.e., signal) covariance matrix and noise covariance matrix

$$\boldsymbol{\Sigma} = \frac{1}{L}E\{\boldsymbol{\Phi}^H(n)\boldsymbol{\Phi}(n)\} = \boldsymbol{\Sigma}_s + \sigma^2\boldsymbol{I}_{m-1} \tag{9}$$

where $\boldsymbol{\Sigma}_s$ is the noiseless covariance matrix given by $\boldsymbol{\Sigma}_s = (1/L)E\{\boldsymbol{\Phi}_s^H(n)\boldsymbol{\Phi}_s(n)\}$, $\boldsymbol{I}_{m-1}$ denotes a $(m-1)\times(m-1)$ identity matrix, and $(\cdot)^H$ denotes Hermitian transposition. Now, we can derive the following relationship between the rank of the noiseless covariance matrix and the number of coherent signals.

Lemma: If the array is partitioned properly so that the number of subarrays and the number of coherent signals satisfy the relation $L \geq q$, then the rank of the matrix $\boldsymbol{\Sigma}_s$ will equal the number of coherent signals regardless of whether the source signals are coherent.

Proof: By defining $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ as the submatrices consisting of the first $m-1$ and L rows of the $M\times q$ array response matrix $\boldsymbol{A}(\theta)$ in (3), from (2) and (4), we can express the noiseless vector $\boldsymbol{x}_l(n)$ in (6) as [9], [30]

$$\boldsymbol{x}_l(n) = \boldsymbol{A}_1\boldsymbol{D}^{l-1}\boldsymbol{s}(n) = \boldsymbol{A}_1\boldsymbol{D}^{l-1}\boldsymbol{\beta}s_1(n) \tag{10}$$

where $\boldsymbol{D} = \text{diag}(e^{j\omega_0\tau_1}, e^{j\omega_0\tau_2}, \ldots, e^{j\omega_0\tau_q})$, and $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_q]^T$. By substituting (10) into the definition of the

noiseless matrix $\boldsymbol{\Phi}_s(n)$ in (8) and performing some manipulations, we obtain

$$\begin{aligned}\boldsymbol{\Phi}_s^T(n) &= [\boldsymbol{A}_1\boldsymbol{\beta}, \boldsymbol{A}_1\boldsymbol{D}\boldsymbol{\beta}, \ldots, \boldsymbol{A}_1\boldsymbol{D}^{L-1}\boldsymbol{\beta}]s_1(n) \\ &= \boldsymbol{A}_1\boldsymbol{B}\boldsymbol{A}_2^T s_1(n) \end{aligned} \tag{11}$$

where $\boldsymbol{B} = \text{diag}(\beta_1, \beta_2, \ldots, \beta_q)$. We can then obtain the noiseless covariance matrix $\boldsymbol{\Sigma}_s$ in (9) as

$$\boldsymbol{\Sigma}_s = \frac{1}{L} E\{\boldsymbol{\Phi}_s^H(n)\boldsymbol{\Phi}_s(n)\} = \frac{1}{L}\boldsymbol{C}^H\boldsymbol{C}r_s \tag{12}$$

where $\boldsymbol{C} = \boldsymbol{A}_2\boldsymbol{B}\boldsymbol{A}_1^T$, and $r_s = E\{s_1(n)s_1^*(n)\}$ is the autocorrelation of the source signal $s_1(n)$.

Under the assumption for the source signals that $\beta_k \neq 0$, we easily find that $\text{rank}(\boldsymbol{B}) = q$. Because the submatrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ of $\boldsymbol{A}(\theta)$ are Vandermonde matrices and we assumed $m > q+1$, we can obtain $\text{rank}(\boldsymbol{A}_1) = \min(m-1, q) = q$ and $\text{rank}(\boldsymbol{A}_2) = \min(L, q) = q$ for $L \geq q$. Hence, we have $\text{rank}(\boldsymbol{C}) = q$, i.e., the rank of matrix $\boldsymbol{\Sigma}_s$ is given by $\text{rank}(\boldsymbol{\Sigma}_s) = q$. ■

Remark B: When some incident signals are coherent and the others are uncorrelated with these signals and with each other, by assuming the first p ($1 \leq p \leq q$) signals are coherent ones as defined in (4) and performing some manipulations, we can obtain $\boldsymbol{\Sigma}_s = (1/L)\boldsymbol{A}_1^*(r_s\overline{\boldsymbol{B}}^*\boldsymbol{A}_2^H\boldsymbol{A}_2\overline{\boldsymbol{B}} + \boldsymbol{R}_u)\boldsymbol{A}_1^T$, where the $q \times q$ diagonal matrices $\overline{\boldsymbol{B}}$ and $\boldsymbol{R}_u$ are given by $\overline{\boldsymbol{B}} = \text{diag}(\beta_1, \ldots, \beta_p, 0, \ldots, 0)$, $\boldsymbol{R}_u = \text{diag}(0, \ldots, 0, r_{s_{p+1}}, \ldots, r_{s_q})$, and $r_{s_k} = E\{s_k(n)s_k^*(n)\}$. Then, we easily find that the rank of matrix $\boldsymbol{\Sigma}_s$ still equals the number of incident signals. □

Accordingly, for $L \geq q$, the eigenvalue decomposition (EVD) of the noiseless covariance matrix $\boldsymbol{\Sigma}_s$ can be expressed as

$$\boldsymbol{\Sigma}_s = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H \tag{13}$$

where $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{m-1}]$, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{m-1})$, $\{\lambda_i\}$ and $\{\boldsymbol{v}_i\}$ are the eigenvalues and corresponding eigenvectors, $\boldsymbol{V}\boldsymbol{V}^H = \boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{I}_{m-1}$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_{m-1} = 0$. Thus, the covariance matrix $\boldsymbol{\Sigma}$ of the data matrix $\boldsymbol{\Phi}(n)$ has the following EVD:

$$\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H + \sigma^2\boldsymbol{I}_{m-1} = \boldsymbol{V}_s\boldsymbol{\Lambda}_s\boldsymbol{V}_s^H + \sigma^2\boldsymbol{I}_{m-1} \tag{14}$$

where $\boldsymbol{\Lambda}_s = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$, and $\boldsymbol{V}_s = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_q]$. The dimension of the signal subspace $\boldsymbol{V}_s$ is q. This rank property is clearly useful for estimating the number of coherent signals and their directions from (8). The detection of the number of coherent signals can be formulated as the determination of the rank of the noiseless covariance matrix from the noise-corrupted array data.

From the LP parameters $\{a_i\}$, a prediction polynomial $D(z)$ can thus be formed as [4], [33]

$$D(z) = 1 - a_1 z^{-1} - \cdots - a_{m-1}z^{-(m-1)} \tag{15}$$

where $z = e^{j\omega_0(d/c)\sin\theta}$. The DOAs of the coherent signals can be estimated from the q signal zeros of $D(z)$ closest to the unit circle in the z plane. It follows that the problem of direction estimation of coherent signals is reduced to that of estimating the LP parameters $\{a_i\}$ from the noisy array data.

III. Optimal Regularization for CLS Estimation

A. CLS Estimation of LP Parameters

From (15), we can find that the reliable estimation of the LP parameters is very important for direction finding. Even though the additive noise components $\{w_i(n)\}$ in (1) are assumed to be mutually uncorrelated white Gaussian, the prediction error $\varepsilon_{l+m-1}(n)$ in (7) is no longer white. Although the ordinary LS method is simple, it is no longer an optimal estimator due to the accumulation of additive noise in $\boldsymbol{\Phi}(n)$. The LS estimate of $\boldsymbol{a}$ from (8) generally has a bias and is not consistent [8], [24], and this estimate will result in an inaccurate estimation of the angle of arrival [7]. To obtain an unbiased and consistent estimate, a variety of estimation schemes have been proposed. The CLS method [25], [26] is a modification of the LS estimation to combat the noise effects in $\boldsymbol{z}(n)$ and $\boldsymbol{\Phi}(n)$ in (8) simultaneously.

If $n = 1, 2, \ldots, N$, from (8), the CLS estimate of the LP parameters is given by [25]

$$\hat{\boldsymbol{a}}_{\text{CLS}} = (\boldsymbol{\Psi} - \sigma^2\boldsymbol{I}_{m-1})^{-1}\boldsymbol{g} \tag{16}$$

where $\boldsymbol{\Psi} = (1/LN)\sum_{n=1}^{N}\boldsymbol{\Phi}^H(n)\boldsymbol{\Phi}(n)$, $\boldsymbol{g} = (1/LN) \cdot \sum_{n=1}^{N}\boldsymbol{\Phi}^H(n)\boldsymbol{z}(n)$, and N is the number of snapshots of the array data. It should be noted that the true noise variance σ^2 is required in the CLS estimation.

In (16), we can find that the $(m-1) \times (m-1)$ matrix in the bracket approaches the noiseless covariance matrix $\boldsymbol{\Sigma}_s$ as N becomes sufficiently large, i.e.,

$$\lim_{N\to\infty}(\boldsymbol{\Psi} - \sigma^2\boldsymbol{I}_{m-1}) = \boldsymbol{\Sigma} - \sigma^2\boldsymbol{I}_{m-1} = \boldsymbol{\Sigma}_s \tag{17}$$

while the noiseless covariance matrix $\boldsymbol{\Sigma}_s$ tends to be singular because its rank is given by $\text{rank}(\boldsymbol{\Sigma}_s) = q < m-1$, where lim denotes the probability limit. The small eigenvalues of the matrix $\boldsymbol{\Psi} - \sigma^2\boldsymbol{I}_{m-1}$ will cause the estimate $\hat{\boldsymbol{a}}_{\text{CLS}}$ to be numerically unstable; therefore, the $m-q-1$ extraneous zeros of the corresponding prediction polynomial tend to fall closer to or outside the unit circle. They are usually observed as spurious peaks in the spatial spectrum [35], [36]. It is thus difficult to distinguish the q signal zeros from the extraneous zeros, and furthermore, the direction estimation will be inaccurate. Although the truncation of the eigenvalues is useful in stabilizing this ill-conditioned estimation [7], [28], [29], [36], [37], [46], the number of principal eigenvalues (i.e., the number of signals) and the noise variance are needed.

In the following, we will study ways to determine the rank of the noiseless covariance matrix (i.e., the number of coherent signals) and to estimate the noise variance simultaneously so that we can improve the estimation of the LP parameters from the noisy measurements.

B. Optimal Regularization for CLS Estimation

It is found that regularization is another approach to alleviate the ill-posed problem [34], [37], [46], [47]. In fact, regularization and truncation are intimately related. Here, we develop a

regularization scheme for combating the ill-conditioning of the CLS estimate $\hat{\boldsymbol{a}}_{\mathrm{CLS}}$ in (16).

By introducing a regularization matrix $\boldsymbol{R}$ into (16), we have a regularized CLS (RCLS) estimate of the LP parameters as

$$\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R}) = (\boldsymbol{\Psi} - \sigma^2 \boldsymbol{I}_{m-1} + \boldsymbol{R})^{-1}\boldsymbol{g} \tag{18}$$

where the regularization matrix $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{P}\boldsymbol{V}^H \tag{19}$$

in which $\boldsymbol{P} = \mathrm{diag}(\rho_1, \rho_2, \ldots, \rho_{m-1})$, and $\{\rho_i\}$ are the multiple regularization parameters.

The next problem is how to determine the regularization matrix $\boldsymbol{R}$ so that the estimation of the LP parameters can be improved. Let us consider the MSE of the estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})$ defined by

$$\mathrm{MSE} = E\{\|\boldsymbol{a} - \hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})\|^2\}. \tag{20}$$

Because the data vector $\boldsymbol{z}(n)$ and matrix $\boldsymbol{\Phi}(n)$ are perturbed by the additive noise components $\{w_i(n)\}$ as shown in (8), the derivation of the MSE of the estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})$ will be complicated because the third- and fourth-order moments of the additive noise should be taken into account [25]. For a sufficiently large number of snapshots N, the analytical expression of the asymptotic MSE is given by Theorem 1.

Theorem 1: If the number of snapshots N is sufficiently large, the asymptotic MSE of the estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})$ in (18) is given by

$$\mathrm{MSE}(\{\rho_i\}) = \frac{\sigma^2}{LN}(1 + \|\boldsymbol{a}\|^2)\sum_{i=1}^{m-1}\frac{\lambda_i + \sigma^2}{(\lambda_i + \rho_i)^2} + \sum_{i=1}^{m-1}\frac{\rho_i^2 |\boldsymbol{v}_i^H \boldsymbol{a}|^2}{(\lambda_i + \rho_i)^2}. \tag{21}$$

Proof: See Appendix A. ∎

The first term of the asymptotic MSE in (21) is the variance term, and the second one is the bias term. From (21), we can find that the small eigenvalues of the matrix $\boldsymbol{\Psi} - \sigma^2 \boldsymbol{I}_{m-1}$ will cause an excessive increase in the MSE of the estimate $\hat{\boldsymbol{a}}_{\mathrm{CLS}}$. If the regularization parameters $\{\rho_i\}$ are increased, the variance term in (21) decreases, and the bias term increases. Obviously, an accurate estimate of the LP parameters with a minimum MSE can be obtained by choosing the adequate regularization parameters $\{\rho_i\}$.

Theorem 2: The optimal regularization parameters that minimize the asymptotic MSE are given by

$$\rho_i^o = \begin{cases} \sigma^2(1 + \|\boldsymbol{a}\|^2) \cdot \dfrac{\lambda_i + \sigma^2}{LN\lambda_i |\boldsymbol{v}_i^H \boldsymbol{a}|^2}, & \text{for } i = 1, 2, \ldots, q \\ \infty, & \text{for } i = q+1, \ldots, m-1 \end{cases} \tag{22}$$

and the minimum MSE is obtained as

$$\mathrm{MSE}(\{\rho_i\})_{\min} = \frac{\sigma^2}{LN}(1 + \|\boldsymbol{a}\|^2)\sum_{i=1}^{q}\frac{\lambda_i + \sigma^2}{\lambda_i(\lambda_i + \rho_i^o)} + \sum_{i=q+1}^{m-1}|\boldsymbol{v}_i^H \boldsymbol{a}|^2. \tag{23}$$

Proof: By letting the derivative of $\mathrm{MSE}(\{\rho_i\})$ in (21) with respect to $\{\rho_i\}$ equal zero and using the fact that $\lambda_{q+1} = \cdots = \lambda_{m-1} = 0$ in (13), the optimal regularization parameters $\{\rho_i^o\}$ are easily obtained. Then, by substituting (22) into (21), we obtain the minimum MSE in (23). ∎

Remark C: Here, we consider the theoretical computation of the true LP parameters included in the definition of the MSE in (20). In the absence of additive noise, from (6) and (8), we have a compact LP model

$$\boldsymbol{z}_s(n) = \boldsymbol{\Phi}_s(n)\boldsymbol{a}. \tag{24}$$

As derived in the proof of Lemma in Section II, by some straightforward manipulations, we can obtain a succinct expression of the Yule–Walker equations

$$\boldsymbol{\zeta}_s = \boldsymbol{\Sigma}_s \boldsymbol{a} \tag{25}$$

where

$$\begin{aligned} \boldsymbol{\zeta}_s &= (1/L)E\{\boldsymbol{\Phi}_s^H(n)\boldsymbol{z}_s(n)\} = (1/L)r_s \boldsymbol{A}_1^* \boldsymbol{B}^* \boldsymbol{A}_2^H \boldsymbol{A}_2 \boldsymbol{B}\boldsymbol{\gamma} \\ \boldsymbol{\Sigma}_s &= (1/L)E\{\boldsymbol{\Phi}_s^H(n)\boldsymbol{\Phi}_s(n)\} = (1/L)r_s \boldsymbol{A}_1^* \boldsymbol{B}^* \boldsymbol{A}_2^H \boldsymbol{A}_2 \boldsymbol{B}\boldsymbol{A}_1^T \\ \boldsymbol{\gamma} &= \left[c^{j\omega_0(m-1)\tau_1}, c^{j\omega_0(m-1)\tau_2}, \ldots, c^{j\omega_0(m-1)\tau_q}\right]^T \end{aligned}$$

and r_s is the autocorrelation of signal $s_1(n)$. Because the rank of the covariance matrix $\boldsymbol{\Sigma}_s$ is q, from (25), the true LP parameters can be obtained [4], [28]

$$\boldsymbol{a} = \sum_{i=1}^{q}\frac{\boldsymbol{v}_i \boldsymbol{v}_i^H}{\lambda_i}\boldsymbol{\zeta}_s \tag{26}$$

where the principal eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_q$ and eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_q$ are given in (13). Note that (26) is the minimum norm solution of LP parameters, and we call it the "true" solution in this paper. □

C. Minimum MSE-Based Truncation and Detection

When the number of snapshots N is sufficiently large, from (17) and (19), the RCLS estimate in (18) can be expressed as

$$\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\rho_i\}) = \sum_{i=1}^{m-1}\frac{\boldsymbol{v}_i \boldsymbol{v}_i^H}{\lambda_i + \rho_i}\boldsymbol{g}. \tag{27}$$

It is known that both regularization and truncation tend to dampen the contributions of the small eigenvalues [37], [46]. From (26) and (27), we can find that the regularization with multiple parameters plays an important role in the truncation of small eigenvalues, where a regularization parameter with an infinite value implies the discarding of the corresponding small

eigenvalue. As mentioned above, the MSE of the estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\rho_i\})$ can be improved by adequately truncating the small eigenvalues that cause instability in the estimation.

To determine whether the eigenvalue λ_i should be retained or discarded, we use the MSE as the decision criterion. We can thus formulate the truncation problem as the minimization of MSE$(\{\rho_i\})$ in (21) on the condition that each regularization parameter ρ_i takes only one value: zero or infinity. The solution is given as the following theorem.

Theorem 3: The optimal truncation of the eigenvalue λ_i that minimizes the asymptotic MSE is given by

$$\rho_i^{(T)} = \begin{cases} 0, & \text{if } \lambda_i \geq \rho_i^o \quad \text{(retain)} \\ \infty, & \text{if } \lambda_i < \rho_i^o \quad \text{(discard).} \end{cases} \tag{28}$$

Proof: See Appendix B. ■

From Theorem 3, we can find that ρ_i^o works as a threshold value determining whether the eigenvalue λ_i should be retained or discarded. Therefore, if there exists a number K, and $\rho_i^{(T)}$ in (28) satisfies

$$\rho_i^{(T)} = \begin{cases} 0, & \text{for } 1 = 1, 2, \ldots, K \\ \infty, & \text{for } i = K+1, \ldots, m-1 \end{cases} \tag{29}$$

the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_K$ should be retained, whereas the others should be discarded. The dimension of the subspace spanned by the retained eigenvectors is K. Because the rank of the noiseless covariance matrix $\boldsymbol{\Sigma}_s$ equals q as shown in Lemma, i.e., the dimension of the signal subspace is q; therefore, we can decide that $q = K$, i.e., the number of coherent signals is equal to the determined truncation number.

In summary, we have shown that detecting the number of coherent signals is the same as determining the truncation of the eigenvalues. The decision rule is as follows:

1) Compute the optimal regularization parameters $\{\rho_i^o\}$ in (22).
2) Compare ρ_i^o with the eigenvalue λ_i by using (28) to find the K that satisfies (29).

This is the basic principle for detecting the number of signals. It implies that we have to obtain the optimal regularization parameters $\{\rho_i^o\}$ for determining the optimal truncation.

IV. Data-Based Iterative Algorithm for DOA Estimation

A. Regularization with Accessible Noisy Data

As shown in (22), the optimal regularization parameters $\{\rho_i^o\}$ for $i = 1, 2, \ldots, q$ depend on the LP parameters $\boldsymbol{a}$, the noise variance σ^2, and the eigenvalues $\{\lambda_i\}$ and the eigenvectors $\{\boldsymbol{v}_i\}$ of the noiseless covariance matrix $\boldsymbol{\Sigma}_s$. All of these are unknown and must be estimated from the received array data. In this section, we present an approach to calculating the regularization parameters needed to detect and estimate the coherent signals from the finite snapshots of the noisy measurements $\{y_i(n)\}_{n=1}^N$.

In most practical situations, information about the variance of the additive noise is unavailable and must be estimated from the received array data. As derived in (9) and (25), we can obtain

$$\begin{aligned} \boldsymbol{\zeta} &= \frac{1}{L} E\{\boldsymbol{\Phi}^H(n)\boldsymbol{z}(n)\} \\ &= \frac{1}{L} \{E\{\boldsymbol{\Phi}_s^H(n)\boldsymbol{z}_s(n)\} + E\{\boldsymbol{W}^H(n)\boldsymbol{z}_w(n)\}\} \\ &= \boldsymbol{\zeta}_s. \end{aligned} \tag{30}$$

Then, from (8) and (9), we easily get

$$\begin{aligned} \boldsymbol{\Pi} &= \frac{1}{L} E\{\boldsymbol{\Phi}^H(n)[\boldsymbol{\Phi}(n), \boldsymbol{z}(n)]\} \\ &= [\boldsymbol{\Sigma}, \boldsymbol{\zeta}] = \boldsymbol{\Pi}_s + \boldsymbol{\Pi}_w \end{aligned} \tag{31}$$

where

$$\begin{aligned} \boldsymbol{\Pi}_s &= (1/L)E\{\boldsymbol{\Phi}_s^H(n)[\boldsymbol{\Phi}_s(n), \boldsymbol{z}_s(n)]\} = [\boldsymbol{\Sigma}_s, \boldsymbol{\zeta}_s] \\ \boldsymbol{\Pi}_w &= (1/L)E\{\boldsymbol{W}^H(n)[\boldsymbol{W}(n), \boldsymbol{z}_w(n)]\} \\ &= \sigma^2[\boldsymbol{I}_{m-1}\boldsymbol{0}_{(m-1)\times 1}] \end{aligned}$$

and $\boldsymbol{0}_{p\times 1}$ denotes a $p \times 1$ null vector. From (25), $\boldsymbol{\Pi}_s$ can be rewritten as $\boldsymbol{\Pi}_s = (1/L)r_s\boldsymbol{A}_1^*\boldsymbol{B}^*\boldsymbol{A}_2^H\boldsymbol{A}_2\boldsymbol{B}\boldsymbol{A}_3^T$, where $\boldsymbol{A}_3$ is an $m \times q$ matrix consisting of the first m rows of the matrix $\boldsymbol{A}(\theta)$ in (3). Hence, we have rank$(\boldsymbol{\Pi}_s) = q$; therefore, the $(m-1)\times m$ matrix $\boldsymbol{\Pi}$ in (31) has the following SVD [27]:

$$\boldsymbol{\Pi} = \overline{\boldsymbol{U}}\,\overline{\boldsymbol{\Lambda}}\,\overline{\boldsymbol{V}}^H \tag{32}$$

where $\overline{\boldsymbol{U}} = [\overline{\boldsymbol{u}}_1, \overline{\boldsymbol{u}}_2, \ldots, \overline{\boldsymbol{u}}_{m-1}]$, $\overline{\boldsymbol{V}} = [\overline{\boldsymbol{v}}_1, \overline{\boldsymbol{v}}_2, \ldots, \overline{\boldsymbol{v}}_m]$, $\overline{\boldsymbol{\Lambda}} = \mathrm{diag}(\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_{m-1})$, and $\overline{\lambda}_1 \geq \overline{\lambda}_2 \geq \cdots \geq \overline{\lambda}_q > \overline{\lambda}_{q+1} = \cdots = \overline{\lambda}_{m-1} = \sigma^2$.

Therefore, by performing the SVD on the $(m-1)\times m$ matrix $[\boldsymbol{\Psi}, \boldsymbol{g}]$

$$[\boldsymbol{\Psi}, \boldsymbol{g}] = \hat{\overline{\boldsymbol{U}}}\,\hat{\overline{\boldsymbol{\Lambda}}}\,\hat{\overline{\boldsymbol{V}}}^H \tag{33}$$

where $\boldsymbol{\Psi}$ and $\boldsymbol{g}$ are given in (16), we can estimate the noise variance σ^2 as the smallest singular value $\hat{\overline{\lambda}}_{m-1}$ or as the average of the $m-q-1$ smallest singular values $\hat{\overline{\lambda}}_{q+1}, \ldots, \hat{\overline{\lambda}}_{m-1}$ [43], [48], [51], [57]. The eigenvalues and eigenvectors of the matrix $\boldsymbol{\Sigma}_s$ can then be estimated from the finite noisy array data

$$\hat{\boldsymbol{\Sigma}}_s = \frac{1}{LN}\sum_{n=1}^{N} \boldsymbol{\Phi}^H(n)\boldsymbol{\Phi}(n) - \hat{\sigma}^2\boldsymbol{I}_{m-1} = \hat{\boldsymbol{V}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}^H. \tag{34}$$

Because the RCLS estimate of $\boldsymbol{a}$ with the optimal regularization parameters is still unknown, to replace the true value of $\boldsymbol{a}$ with its estimate, we introduce another set of regularization parameters $\{\mu_p\}$, where $p = 1, 2, \ldots, m-1$. We then have an alternative regularized estimate

$$\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p\}) = \sum_{p=1}^{m-1} \frac{\hat{\boldsymbol{v}}_p\hat{\boldsymbol{v}}_p^H}{\hat{\lambda}_p + \mu_p}\,\boldsymbol{g}. \tag{35}$$

Then, by replacing the true values of $\boldsymbol{a}$, σ^2, $\{\lambda_i\}$, and $\{\boldsymbol{v}_i\}$ in (21) with their estimates $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p\})$, $\hat{\sigma}^2$, $\{\hat{\lambda}_i\}$,

and $\{\hat{\boldsymbol{v}}_i\}$ respectively, and using the property of EVD that $\boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^H = \boldsymbol{I}_{m-1}$, we get the estimated MSE (EMSE) of the estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})$

$$\begin{aligned}\mathrm{EMSE}(\{\rho_i\},\{\mu_p\}) &= \frac{\hat{\sigma}^2}{LN}\left(1+\|\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p\})\|^2\right)\sum_{i=1}^{m-1}\frac{\hat{\lambda}_i+\hat{\sigma}^2}{(\hat{\lambda}_i+\rho_i)^2} \\ &\quad + \sum_{i=1}^{m-1}\frac{\rho_i^2}{(\hat{\lambda}_i+\rho_i)^2(\hat{\lambda}_i+\mu_i)^2}\,|\hat{\boldsymbol{v}}_i^H\boldsymbol{g}|^2. \end{aligned} \tag{36}$$

Hence, we can obtain the optimal values of the regularization parameters $\{\rho_i\}$ from the available data for the given set of $\{\mu_p\}$

$$\rho_i^o(\{\mu_p\}) = \hat{\sigma}^2\left(1+\|\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p\})\|^2\right)\frac{(\hat{\lambda}_i+\hat{\sigma}^2)(\hat{\lambda}_i+\mu_i)^2}{LN\hat{\lambda}_i|\hat{\boldsymbol{v}}_i^H\boldsymbol{g}|^2} \tag{37}$$

for $i = 1, 2, \ldots, m-1$.

The alternative parameters $\{\mu_p\}$ used for calculating the estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p\})$ in (35) must have the same values as the optimal regularization parameters $\rho_i^o(\{\mu_p\})$ in order to obtain a reliable estimate of the LP parameters. That is, the parameters $\{\mu_p\}$ should be determined by solving the equation

$$\rho_i^o(\{\mu_p\}) = \mu_i, \qquad \text{for } i = 1, 2, \ldots, m-1. \tag{38}$$

As shown in (37), $\rho_i^o(\{\mu_p\})$ depends on the parameters $\{\mu_p\}$, where $p = 1, 2, \ldots, m-1$. It follows that (38) is a nonlinear equation in terms of the parameters $\mu_1, \mu_2, \ldots, \mu_{m-1}$, and it is difficult to find the analytical solutions of this equation for these parameters. However, from (37) and (38), we can determine the regularization parameters from the received array data in an iterative manner. The procedure is summarized as follows.

a) Set the initial values of parameters $\{\mu_p\}$ as $\{\mu_p^{(k)}\}$.
b) Calculate the regularization parameter $\rho_i^o(\{\mu_p^{(k)}\})$ by using (37).
c) Update the old values of $\{\mu_p^{(k)}\}$ to the new ones as $\mu_i^{(k+1)} = \rho_i^o(\{\mu_p^{(k)}\})$. Then, repeat the above steps until $\mu_i^{(k+1)}$ converges to a constant with a specified limit or exceeds a specified large value.
d) Determine the optimal regularization parameter $\hat{\rho}_i^o$ as $\hat{\rho}_i^o = \mu_i^{(k+1)}$.

B. DOA Estimation with Iterative Regularization Algorithm

As described above, the regularization is the key to detecting the number of signals and estimating their DOAs. By combining the principles of the optimal truncation and the data-based regularization with the estimation of the noise variance, we can present an iterative regularization algorithm that uses only the received data $\{y_i(n)\}_{n=1}^N$. In this algorithm, the regularization parameters are updated in an iterative manner without any *a priori* information. A flowchart of the proposed algorithm is depicted in Fig. 1.

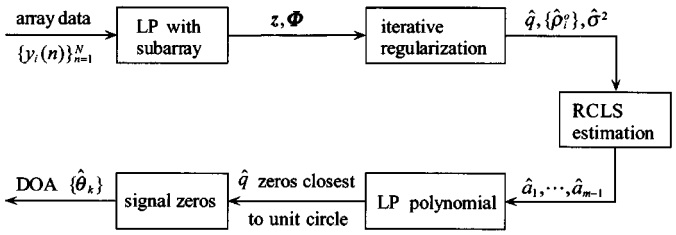

Fig. 1. Flowchart of the proposed algorithm for estimating the directions of coherent signals.

Step 1) Set the subarray size as m, where $m > q+1$ and $L = M-m+1 \geq q$, and then, form the vector $\boldsymbol{z}$ and the matrix $\boldsymbol{\Phi}$ as $\boldsymbol{z} = [\boldsymbol{z}^T(1), \boldsymbol{z}^T(2), \ldots, \boldsymbol{z}^T(N)]^T$, and $\boldsymbol{\Phi} = [\boldsymbol{\Phi}^T(1), \boldsymbol{\Phi}^T(2), \ldots, \boldsymbol{\Phi}^T(N)]^T$ by using (7) and (8).

Step 2) Estimate the noise variance σ^2 as $\hat{\sigma}_0^2 = \hat{\tilde{\lambda}}_{m-1}$ by using (33).

Step 3) Estimate the eigenvalues $\{\lambda_i\}$ and the eigenvectors $\{\boldsymbol{v}_i\}$ by using (34), i.e., $\tilde{\boldsymbol{\Sigma}}_s = (1/LN)\boldsymbol{\Phi}^H\boldsymbol{\Phi} - \hat{\sigma}_0^2\boldsymbol{I}_{m-1} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{V}}^H$.

Step 4) Let $k = 0$, and set the initial values of the regularization parameters $\{\mu_p\}$ as $\mu_p^{(0)} = \xi$, where ξ is a small positive number that expresses machine precision (e.g., $\xi = 10^{-10}$).

Step 5) Calculate the estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p^{(k)}\})$ for the given $\{\mu_p^{(k)}\}$ as

$$\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p^{(k)}\}) = \sum_{p=1}^{m-1}\frac{\tilde{\boldsymbol{v}}_p\tilde{\boldsymbol{v}}_p^H}{\tilde{\lambda}_p+\mu_p^{(k)}}\boldsymbol{g}. \tag{39}$$

Step 6) Calculate the regularization parameters $\{\rho_i^{(k)}\}$ as

$$\begin{aligned}\rho_i^{(k)} &= \hat{\sigma}_0^2\left(1+\|\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\{\mu_p^{(k)}\})\|^2\right) \\ &\quad \cdot\frac{(\tilde{\lambda}_i+\hat{\sigma}_0^2)\left(\tilde{\lambda}_i+\mu_i^{(k)}\right)^2}{LN\tilde{\lambda}_i|\tilde{\boldsymbol{v}}_i^H\boldsymbol{g}|^2}.\end{aligned} \tag{40}$$

Step 7) Replace the old values of $\{\mu_i^{(k)}\}$ with the new ones as

$$\mu_i^{(k+1)} = \rho_i^{(k)}. \tag{41}$$

If $\mu_i^{(k+1)}$ exceeds a specified large value χ (e.g., $\chi = 10^{30}$), force it to χ, and calculate the increment of $\mu_i^{(k+1)}$ as $\Delta\mu_i^{(k+1)} = \mu_i^{(k+1)} - \mu_i^{(k)}$. Then, set $k = k+1$, return to Step 5), and repeat the above procedure until any one of the following is satisfied: i) $|\Delta\mu_i^{(k+1)}| \leq \delta$ for $i = 1, 2, \ldots, m-1$ (e.g., $\delta = 10^{-5}$), or ii) the number of iterations $k > 20$.

From the results of the iterative algorithm, the optimal regularization parameters are estimated as $\hat{\rho}_i^o = \mu_i^{(k+1)}$ for $i = 1, 2, \ldots, m-1$. By comparing the eigenvalue $\tilde{\lambda}_i$ and the regularization parameter $\hat{\rho}_i^o$, the number of coherent signals is estimated to be $\hat{q} = K$, where K is the number of the converged

parameters $\{\hat{\rho}_i^o\}$. The noise variance can then be estimated as [43], [48], [51], [57]

$$\hat{\sigma}^2 = \frac{1}{m-\hat{q}-1}\sum_{i=\hat{q}+1}^{m-1}\hat{\tilde{\lambda}}_i. \tag{42}$$

Finally, by using (34), the RCLS estimate of the LP parameters is obtained as

$$\hat{\boldsymbol{a}}_{\mathrm{RCLS}} = \sum_{i=1}^{m-1}\frac{\hat{\boldsymbol{v}}_i\hat{\boldsymbol{v}}_i^H}{\hat{\tilde{\lambda}}_i+\hat{\rho}_i^o}\boldsymbol{g}. \tag{43}$$

Then, the DOAs of the coherent signals can be estimated from the $\hat{q}$ zeros of the corresponding LP polynomial $\hat{D}(z) = 1 - \hat{a}_1 z^{-1} - \cdots - \hat{a}_{m-1}z^{-(m-1)}$ that are closest to the unit circle in the z plane or from the $\hat{q}$ highest peaks of the spectrum $1/|\hat{D}(z)|^2$.

Remark D: Because a causal model is used for the prediction (i.e., FLP), it follows from the Lemma in Section II that the maximum number of coherent signals that can be detected is $q < M/2$ [9], [30]. Forward–backward averaging is a well-known method for enhancing the performance of parameter estimation in array signal processing [30], [54], [60]. Similarly, the full array can be partitioned into L backward subarrays, each with m sensors [30]; therefore, we have the following backward linear prediction (BLP) model for the noise-corrupted signal $y_i(n)$ in the lth backward subarray

$$y_{L-l+1}^*(n) = \boldsymbol{y}_{b,\,l}^T(n)\boldsymbol{a} + \varepsilon_{b,\,L-l+1}(n) \tag{44}$$

where $\boldsymbol{y}_{b,\,l}(n) = [y_{M-l+1}(n),\, y_{M-l}(n),\, \ldots,\, y_{L-l+2}(n)]^H$, the backward prediction error $\varepsilon_{b,\,L-l+1}(n)$ is given by $\varepsilon_{b,\,L-l+1}(n) = \overline{\boldsymbol{w}}_{b,\,l}^T(n)\overline{\boldsymbol{a}}$, $\overline{\boldsymbol{w}}_{b,\,l}(n) = [\boldsymbol{w}_{b,\,l}^T(n),\, w_{L-l+1}^*(n)]^T$, and $\boldsymbol{w}_{b,\,l}(n) = [w_{M-l+1}(n),\, w_{M-l}(n),\, \ldots,\, w_{L-l+2}(n)]^H$.

The above derivations and results are still valid if we use FLP and BLP simultaneously [i.e., forward–backward linear prediction (FBLP)] [5], [38], except that the number of subarrays L, the vector $\boldsymbol{z}$, and the matrix $\boldsymbol{\Phi}$ should be replaced by $\overline{L} = 2L$, $\overline{\boldsymbol{z}} = [\boldsymbol{z}^T,\, \boldsymbol{z}_b^T]^T$ and $\overline{\boldsymbol{\Phi}} = [\boldsymbol{\Phi}^T,\, \boldsymbol{\Phi}_b^T]^T$, where $\boldsymbol{z}_b(n) = [y_L(n),\, y_{L-1}(n),\, \ldots,\, y_1(n)]^H$, and $\boldsymbol{\Phi}_b(n) = [\boldsymbol{y}_{b,\,1}(n),\, \boldsymbol{y}_{b,\,2}(n),\, \ldots,\, \boldsymbol{y}_{b,\,L}(n)]^T$. In this case, the maximum number of coherent signals is $q < 2M/3$ [30], and the performance of the proposed algorithm can be improved as the number of snapshots is doubled. □

Remark E: In practice, the subarray size m should be chosen appropriately because the information on the number of signals is unavailable. From Remark D, we can find that the maximum detectable number of signals is $q < M/2$ or $q < 2M/3$ for M sensors when FLP or FBLP, respectively, is used. Therefore, for the proposed method with FLP or FBLP, we can predetermine the number of signals as $\overline{q} = M/2$ or $\overline{q} = 2M/3$, and a conservative value of the subarray size can be chosen as $m = M/2+1$ or $m = 2M/3+1$, where the total number of subarrays will be $L = M/2$ (for FLP) or $\overline{L} = 2L = 2M/3$ (for FBLP). Clearly, the inequality condition in the choice of the subarray size can be satisfied, i.e., $m > \overline{q}+1$ and $L = \overline{q}$ (for FLP) or $\overline{L} = \overline{q}$ (for FBLP); therefore, the DOAs of the coherent signals can be estimated. □

C. Convergence Behavior of Iterative Algorithm

It is rather difficult to strictly and theoretically analyze the convergence of the proposed iterative regularization algorithm. Here, we discuss the convergence behavior for a finite but large number of snapshots when the noise variance and the true noiseless covariance matrix are known.

In the iterative algorithm, the parameter $\rho_i^{(k)}$ (and, hence, $\mu_i^{(k+1)}$) in the kth iteration is determined by using the parameters $\mu_1^{(k)}, \mu_2^{(k)}, \ldots, \mu_{m-1}^{(k)}$ obtained in the $(k-1)$th iteration, where $k \geq 1$. By substituting (39) into (40) and using the fact that $\tilde{\boldsymbol{V}}^H\tilde{\boldsymbol{V}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^H = \boldsymbol{I}_{m-1}$, we can rewrite $\rho_i^{(k)}$ as

$$\begin{aligned}\rho_i^{(k)} &= \left(1+\sum_{\substack{p=1\\p\neq i}}^{m-1}\frac{|\tilde{\boldsymbol{v}}_p^H\boldsymbol{g}|^2}{\left(\tilde{\lambda}_p+\mu_p^{(k)}\right)^2}\right)\frac{\hat{\sigma}_0^2(\tilde{\lambda}_i+\hat{\sigma}_0^2)}{LN\tilde{\lambda}_i|\tilde{\boldsymbol{v}}_i^H\boldsymbol{g}|^2}\\ &\quad\cdot\left(\tilde{\lambda}_i+\mu_i^{(k)}\right)^2+\frac{\hat{\sigma}_0^2(\tilde{\lambda}_i+\hat{\sigma}_0^2)}{LN\tilde{\lambda}_i}\\ &= \frac{1}{\kappa_i^{(k)}}\left(\tilde{\lambda}_i+\mu_i^{(k)}\right)^2+\eta_i\end{aligned} \tag{45}$$

where

$$\kappa_i^{(k)} = \left(1+\sum_{\substack{p=1\\p\neq i}}^{m-1}\frac{|\tilde{\boldsymbol{v}}_p^H\boldsymbol{g}|^2}{\left(\tilde{\lambda}_p+\mu_p^{(k)}\right)^2}\right)^{-1}\frac{LN\tilde{\lambda}_i|\tilde{\boldsymbol{v}}_i^H\boldsymbol{g}|^2}{\hat{\sigma}_0^2(\tilde{\lambda}_i+\hat{\sigma}_0^2)} \tag{46}$$

$$\eta_i = \frac{\hat{\sigma}_0^2(\tilde{\lambda}_i+\hat{\sigma}_0^2)}{LN\tilde{\lambda}_i}. \tag{47}$$

Thus, we can find that $\rho_i^{(k)}$ in (40) is a quadratic function of $\mu_i^{(k)}$ that expresses a parabolic curve on the μ_i–ρ_i plane showing the trajectory of $\mu_i^{(k)}$ minimizing the EMSE for the given $\{\mu_p^{(k)}\}$, where $p \neq i$, and the value of $\mu_i^{(k)}$ on this trajectory gives the next value of $\mu_i^{(k+1)}$ [37]. If the parabolic curve does not intersect the line $\mu_i = \rho_i$, $\mu_i^{(k)}$ will diverge to infinity as $k \to \infty$. On the other hand, if the parabolic curve intersects the line $\mu_i = \rho_i$ in the limit $k \to \infty$, $\mu_i^{(k)}$ will converge to a bounded constant.

Hence, the converged regularization parameter $\rho_i^{(k)}$ is given by the intersection of the parabolic curve (40) and the straight line defined by $\mu_i = \rho_i$ [37], [40]. From (45) and $\rho_i^{(k)} = \mu_i^{(k)}$, we can obtain

$$\mu_i^{2(k)} + \left(2\tilde{\lambda}_i - \kappa_i^{(k)}\right)\mu_i^{(k)} + \tilde{\lambda}_i^2 + \kappa_i^{(k)}\eta_i = 0 \tag{48}$$

for $i = 1, 2, \ldots, K$, where $\mu_p^{(k)} = \mu_p^{(k-1)}$ for $p \neq i$. From (48), we can find that if

$$\frac{4(\tilde{\lambda}_i+\eta_i)}{\kappa_i^{(k)}} \leq 1, \qquad \text{for } i = 1, 2, \ldots, K \tag{49}$$

is satisfied, then the two intersections of (48) are given by

$$\begin{aligned}&\mu_{i,\,1}^{(k)}(\{\mu_p^{(k-1)}\}),\ \mu_{i,\,2}^{(k)}(\{\mu_p^{(k-1)}\})\\ &\quad= 0.5\left(\kappa_i^{(k)} - 2\tilde{\lambda}_i \mp \sqrt{\kappa_i^{(k)}\left(\kappa_i^{(k)} - 4(\tilde{\lambda}_i+\eta_i)\right)}\right)\end{aligned} \tag{50}$$

where $\mu_{i,1}^{(k)}(\{\mu_p^{(k-1)}\}) < \mu_{i,2}^{(k)}(\{\mu_p^{(k-1)}\})$, and $p \neq i$. From (45), the derivatives of $\rho_i^{(k)}$ in (40) with respect to $\mu_i^{(k)}$ at the these intersections can be calculated as

$$\left.\frac{d\rho_i^{(k)}}{d\mu_i^{(k)}}\right|_{\mu_i^{(k)}=\mu_{i,1}^{(k)}(\{\mu_p^{(k-1)}\})} = 2\,\frac{\tilde{\lambda}_i + \mu_{i,1}^{(k)}(\{\mu_p^{(k-1)}\})}{\kappa_i^{(k)}} = 1 - \sqrt{1 - \frac{4(\tilde{\lambda}_i + \eta_i)}{\kappa_i^{(k)}}} < 1 \tag{51}$$

$$\left.\frac{d\rho_i^{(k)}}{d\mu_i^{(k)}}\right|_{\mu_i^{(k)}=\mu_{i,2}^{(k)}(\{\mu_p^{(k-1)}\})} > 1. \tag{52}$$

The contraction mapping theorem [40] states that if the derivative at the intersection is less than one, the point is a stable convergence point. Therefore, the smaller solution $\mu_{i,1}^{(k)}(\{\mu_p^{(k-1)}\})$ will converge to a constant as $k \to \infty$. Conversely, if the inverse inequality of (49) is satisfied, i.e., $4(\tilde{\lambda}_i + \eta_i)/\kappa_i^{(k)} > 1$, a solution for (48) does not exist, i.e., the parabolic curve given by (40) and the straight line $\mu_i = \rho_i$ do not intersect, and $\mu_i^{(k)}$ will diverge to infinity.

It is worthwhile examining the convergence of the iterative algorithm when the noise variance σ^2 and the true noiseless covariance matrix $\mathbf{\Sigma}_s$ are known and the number of snapshots N is finite but large. From (46) and (47), we have

$$\frac{4(\lambda_i + \eta_i)}{\kappa_i^{(k)}} = 4\left(1 + \sum_{\substack{p=1\\p\neq i}}^{q} \frac{|\boldsymbol{v}_p^H \boldsymbol{g}|^2}{\left(\lambda_p + \mu_p^{(k)}\right)^2} + \sum_{\substack{p=q+1\\p\neq i}}^{m-1} \frac{|\boldsymbol{v}_p^H \boldsymbol{g}|^2}{\left(\mu_p^{(k)}\right)^2}\right) \cdot \frac{\sigma^2}{LN|\boldsymbol{v}_i^H \boldsymbol{g}|^2}\left(1 + \frac{\sigma^2}{\lambda_i}\right)\left(\lambda_i + \frac{\sigma^2}{LN}\left(1 + \frac{\sigma^2}{\lambda_i}\right)\right) \tag{53}$$

for $i = 1, 2, \ldots, K$, where $\mu_p^{(k)} = \mu_p^{(k-1)}$ for $p \neq i$, and $k \geq 1$. By using the fact that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_{m-1} = 0$ in (13), we can obviously obtain that $4(\lambda_i + \eta_i)/\kappa_i^{(k)} = \infty > 1$ for $i = q+1, \ldots, m-1$. Furthermore, we can find that the inequality $4(\lambda_i + \eta_i)/\kappa_i^{(k)} < 1$ holds for $i = 1, 2, \ldots, q$ (i.e., $K = q$) when the number of snapshots N is finite but large. Therefore, based on the above analysis, $\mu_{q+1}^{(k)}, \ldots, \mu_{m-1}^{(k)}$ will diverge to infinity, whereas $\mu_1^{(k)}, \mu_2^{(k)}, \ldots, \mu_q^{(k)}$ will converge to constants for a finite but large number of snapshots. This agrees with the theoretically analytical expression of the optimal regularization parameters derived in Theorem 2.

Unfortunately, when the number of snapshots is finite, the inequality $4(\lambda_i + \eta_i)/\kappa_i^{(k)} < 1$ will hold for $i = 1, 2, \ldots, K$, where $K \leq q$. Thus, we can easily see that the proposed iterative algorithm sometimes underestimates the number of signals due to uncertainties in the noise variance and the LP parameters and small length of data. However, from the viewpoint of stabilizing the CLS estimate of the LP parameters and reducing its MSE, the result is reasonable.

V. Numerical Examples

In this section, we evaluate the effectiveness of the proposed approach with the FBLP model in estimating the DOAs from only the observed array data. In the numerical examples, the ULA has $M = 10$ sensors with a half-wavelength spacing, and the incident signals are the coherent binary phase-shift keying (BPSK) signals with a raised cosine pulse shape with 50% excess bandwidth. The SNR is defined as the ratio of the signal power to the noise power at each sensor. For assessing the detection performance, the spatial smoothing based Akaike information criterion (AIC) and minimum description length (MDL) (i.e., SS-AIC and SS-MDL) methods [16], [17] and the WSF detection scheme (initialization by MODE) [43], [48], [51] are performed. To examine the estimation performance, the LS and CLS methods, smoothed Min-Norm [13], spatial smoothing based root-MUSIC (SS-based root-MUSIC) [9], [58], MODE (with the optimal weighting matrix and linear constraint), [48], [51], and WSF (initialization by MODE) [43], [48] are carried out. In these methods, we assume that the number of signals is known, and the dimension of the signal subspace is assumed to equal one in MODE and WSF and to equal the number of coherent signals in SS-based root-MUSIC and smoothed Min-Norm. In addition, the stochastic Cramér–Rao lower bound (CRB) [48], [52] is calculated for proper comparison. To improve the estimation accuracy, the true noise variance is used in the CLS estimate, and the last step of the two-step procedure of the MODE algorithm is iterated five times (see [55] for more details). The results shown below are all based on 200 independent trials (unless otherwise explicitly stated).

Example A—Performance versus SNR: Two coherent signals with equal power impinge on the ULA from angles $\theta_1 = 5°$ and $\theta_2 = 12°$. The number of snapshots and the subarray size are $N = 128$ and $m = 5$, where the number of subarrays is $\overline{L} = 2L = 12$.

First, we consider the detection of the number of signals by using the iterative algorithm. When the SNR is 5 dB ($\sigma^2 = 0.3162$), the calculated regularization parameters $\{\mu_i^{(k)}\}$ (i.e., $\{\rho_i^{(k-1)}\}$ for $k \geq 1$) from only the received array data are plotted in Fig. 2, in which the convergence of the regularization parameters calculated by using the iterative algorithm with the true matrix $\mathbf{\Sigma}_s$ in (12) and the actual data is shown for reference. In addition, the regularization parameters calculated by using the iterative algorithm with the true parameters such as $\mathbf{\Sigma}_s$, $\boldsymbol{\zeta}_s$, $\boldsymbol{a}$ and σ^2 are plotted, and the theoretically optimal regularization parameters $\{\rho_i^o\}$ calculated using the true matrix $\mathbf{\Sigma}_s$ in (12), the LP parameters $\boldsymbol{a}$ given in (26), and the variance σ^2 are shown. The regularization parameters $\mu_1^{(k)}$ and $\mu_2^{(k)}$ converge to constants within a few iterations, whereas $\mu_3^{(k)}$ and $\mu_4^{(k)}$ diverge to infinity. The estimated noise variance is $\hat{\sigma}^2 = 0.3390$, which is near the true value, where the squared error is 5.1706×10^{-4}. Based on these results, the optimal regularization parameters $\{\hat{\rho}_i^o\}$ and the eigenvalues $\{\tilde{\lambda}_i\}$ of the estimated noiseless covariance matrix $\tilde{\mathbf{\Sigma}}_s$ are obtained, and they are shown in Fig. 3. We find that the truncation of the eigenvalues is determined by the intersection of the two lines at $K = 2$, i.e., the number of signals is estimated to be $\hat{q} = K = 2$.

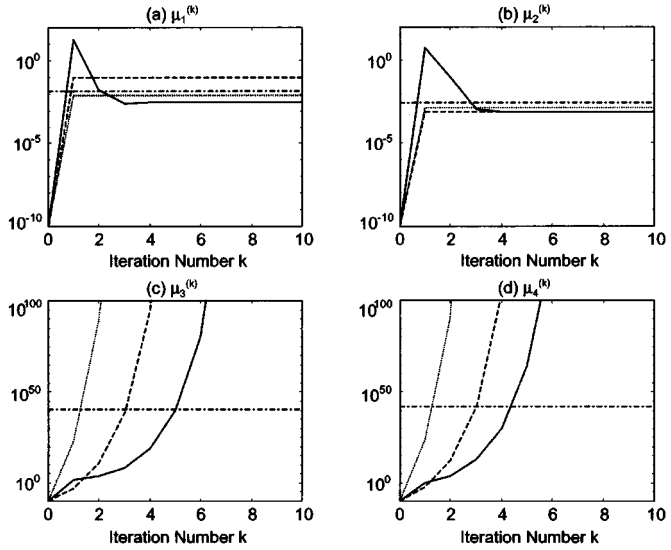

Fig. 2. Convergence of the regularization parameters $\{\mu_i^{(k)}\}$ in the iterative algorithm (solid line: values calculated with the received data; dashed line: values calculated with the true matrix $\mathbf{\Sigma}_s$ and the received data; dotted line: values calculated with the true parameters $\mathbf{\Sigma}_s$, $\boldsymbol{\zeta}_s$, $\boldsymbol{a}$ and σ^2; and dash-dot line: theoretically optimal values $\{\rho_i^o\}$ calculated with the true parameters $\mathbf{\Sigma}_s$, $\boldsymbol{a}$ and σ^2) for Example A (SNR $= 5$ dB, $N = 128$, $M = 10$, and $m = 5$).

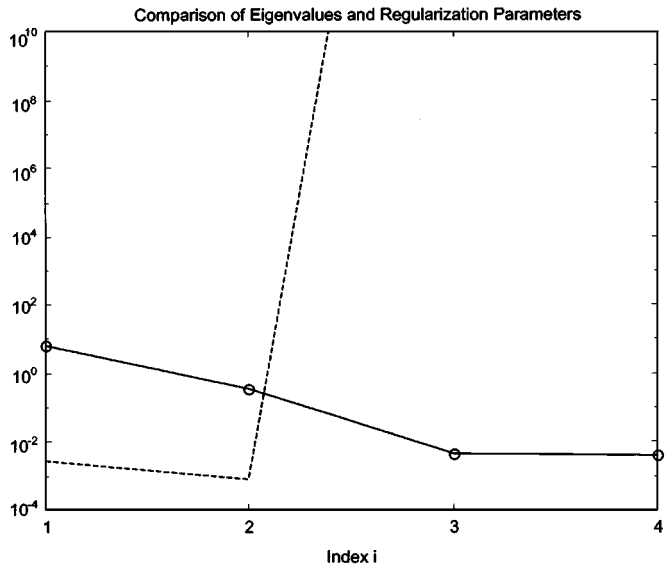

Fig. 3. Optimal truncation of eigenvalues by comparing the estimated eigenvalues of the noiseless covariance matrix (solid line with "o") with the calculated regularization parameters (dashed line) for Example A (SNR $= 5$ dB, $N = 128$, $M = 10$, and $m = 5$).

Fig. 4 shows the distribution of zeros of the LP polynomials in the z-plane for 100 trials corresponding to the LS, CLS, and proposed RCLS methods. For the LS estimation, the signal zeros and extraneous zeros deviate greatly from the true locations though the zeros are inside the unit circle. For the CLS estimation, although the estimated signal zeros cluster around the true locations in a certain sense, the extraneous ones fluctuate wildly, and some of them even wander outside the unit circle. It follows that the estimated DOAs using the LS and CLS methods will have large squared errors. However, by using the proposed regularized approach, an accurate estimate of the LP parameters is obtained. The zeros of the prediction polynomial are almost all

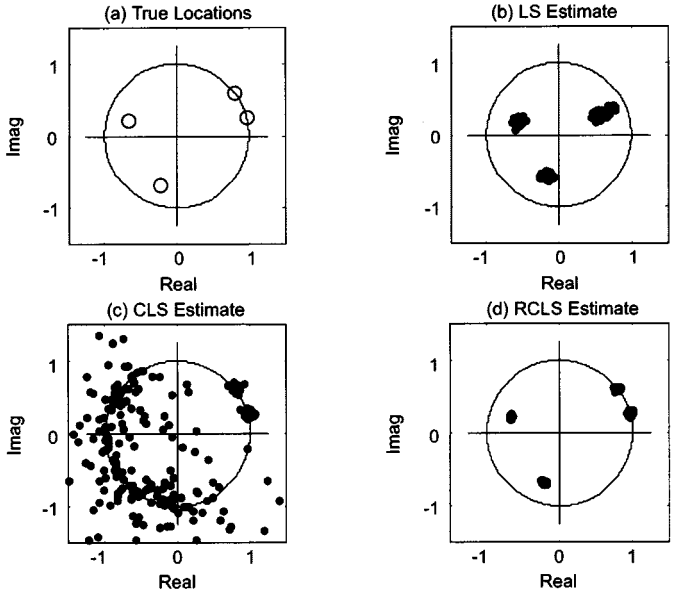

Fig. 4. Distribution of the zeros of prediction polynomials corresponding to the LS, CLS, and regularized CLS estimates of the LP parameters for Example A (SNR $= 5$ dB, $N = 128$, $M = 10$, and $m = 5$).

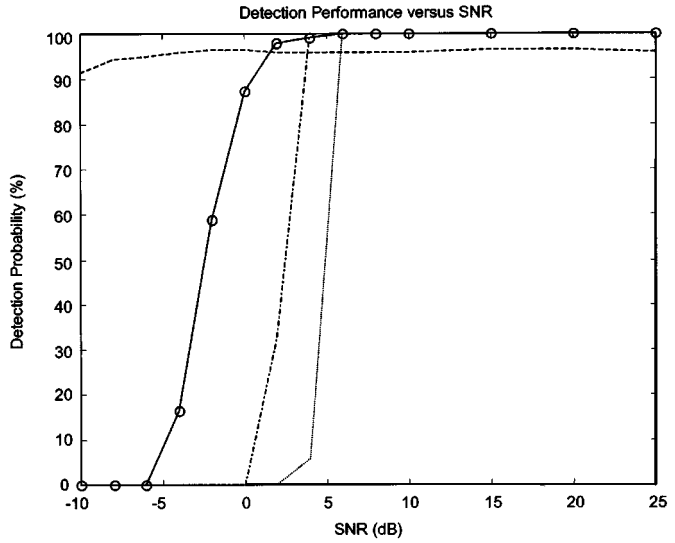

Fig. 5. Probability of correct detection versus the SNR (dotted line: SS-AIC; dash-dot line: SS-MDL; dashed line: WSF detection scheme; and solid line with "o": proposed method) for Example A ($N = 128$, $M = 10$, and $m = 5$).

inside the unit circle, and the estimated signal zeros are much less perturbed from their true locations. Thus, a more reliable estimate of direction can be obtained.

We vary the SNR from -10 to 25 dB and run 200 independent trials at each SNR. The detection performance in terms of SNR is shown in Fig. 5. In the SS-AIC and SS-MDL methods, the AIC and MDL criteria are applied to the $m \times m$ spatially smoothed covariance matrix of subarrays with m sensors [9], [16], [17], and the number of signals is determined by the multiplicity of the smallest eigenvalues. The threshold $\gamma(N)$ for the hypothesis test of the WSF detection scheme is chosen as $\gamma(N) = \gamma\sqrt{N \log N}$, which satisfies the conditions $\lim_{N\to\infty}(\gamma(N)/N) = 0$ and $\lim_{N\to\infty}(\gamma(N)/\log\log N) = \infty$ for strongly consistent

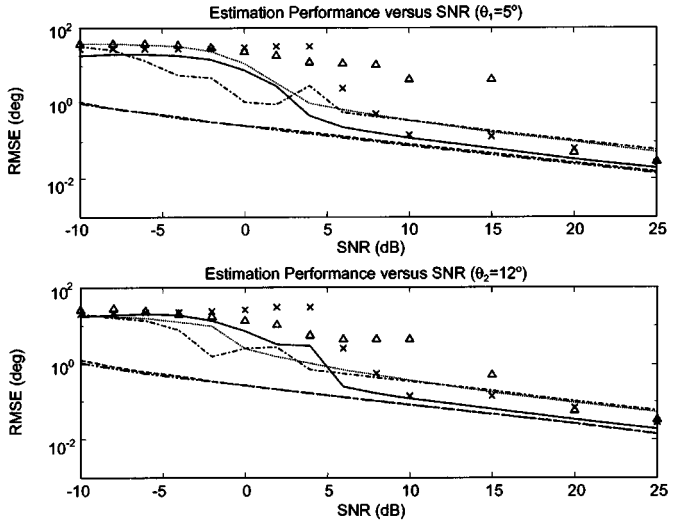

Fig. 6. RMSEs of the direction estimates versus the SNR ("x": LS; "△": CLS; dotted line: SS-based root-MUSIC; dash-dot line: smoothed Min-Norm; dashed line: MODE; solid line: proposed method; and dash-dots line: CRB) for Example A ($N = 128$, $M = 10$, and $m = 5$).

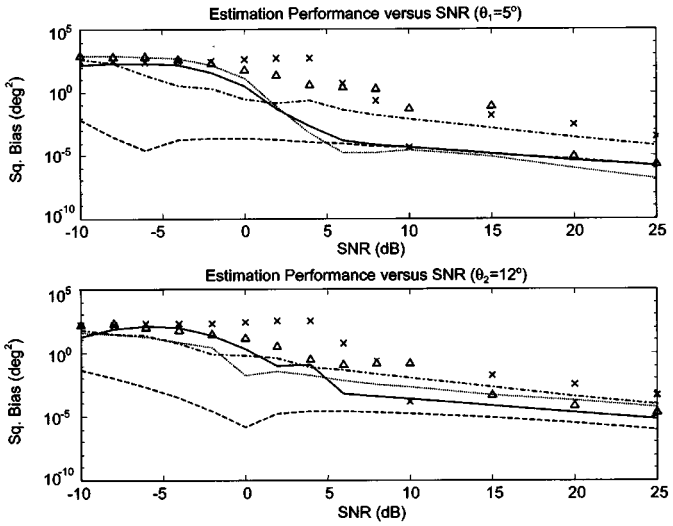

Fig. 7. Squared biases of the direction estimates versus the SNR ("x": LS; "△": CLS; dotted line: SS-based root-MUSIC; dash-dot line: smoothed Min-Norm; dashed line: MODE; and solid line: proposed method) for Example A ($N = 128$, $M = 10$, and $m = 5$).

estimation of the number of signals (see [43], [48], and references therein for details), where $\gamma = 0.995$. The AIC and MDL methods are rather simple, i.e., only the eigenvalues are needed, whereas the WSF detection scheme requires a solution of direction finding for each of several hypotheses. However, in this paper, the number of signals (i.e., the rank of the noiseless covariance matrix) is determined by minimizing the MSE of the estimated LP parameters through the truncation of the eigenvalues. In the proposed algorithm, the information about the eigenvalues and eigenvectors is jointly exploited, whereas an iterative computation is needed to estimate the regularization parameters. At low SNRs, although the proposed approach is inferior to the WSF detection scheme, it outperforms the SS-AIC and SS-MDL methods. In this example, the simulation shows that the amount of computations required by the implementation of the proposed algorithm in terms of MATLAB flops is approximately 2.6100 and 0.8667 times that of the SS-AIC (or SS-MDL) method and WSF detection scheme (averaged over 200 runs).

Next, we examine the performance of direction estimation. The root MSEs (RMSEs) and squared biases of the estimated θ_1 and θ_2 versus SNR are plotted in Figs. 6 and 7, where the CRBs of the estimates versus SNR are also plotted in Fig. 6 for comparison. Due to the noise effects in the data matrix and data vector, the LS method gives incorrect results at low to medium SNRs (e.g., at -10 to 5 dB). Because the accumulation of noise in the matrix $\hat{\boldsymbol{\Sigma}}$ (i.e., $\boldsymbol{\Psi}$) takes a similar effect as the regularization parameter, the direction estimation becomes better to a certain extent at some SNRs (e.g., at 5 to 15 dB). However, the noise variance is not the optimal parameter for minimizing the MSE of the LS estimate. The CLS method is better than the LS algorithm at a few SNRs (especially at -2 to 5 dB), but the CLS estimate of the LP parameters is easily affected by the small eigenvalues of the noiseless covariance matrix $\hat{\boldsymbol{\Sigma}}_s$. The smoothed Min-Norm exhibits better performance than the SS-based root-MUSIC, LS, and CLS methods at low to medium SNRs because two successive subspace approximations with the truncated SVD are used to remove the perturbations of additive noise and to alleviate the ill-conditioning in the estimate of the LP parameters. The MODE and WSF method provide most accurate estimates than the other methods because the true values of number of signals and dimension of signal subspace are used. Additionally, the results of MODE are almost indistinguishable from the CRB in this empirical scenario (the results of WSF are similar to that of MODE and are omitted here). Even though *a priori* information on the number of signals and the noise variance is not used, the proposed approach provides more accurate direction estimation than the LS method, and it overcomes the numerical instability of the CLS estimation. However, the proposed approach has larger errors than the smoothed Min-Norm, MODE, and WSF method at lower SNRs due to the failure in detecting the number of signals, as shown in Fig. 5. As the probability of the successful detection increases, the proposed algorithm outperforms the other methods except MODE and WSF at medium to high SNRs, and its RMSEs become closer to the CRB.

As shown in (21), the MSE of the RCLS estimate consists of the variance term and bias one. In the MSE-based regularization method, at the expense of introducing bias, we lower the variance of the estimate to reduce the overall MSE of the estimated LP parameters (and, hence, the MSE of the estimated directions). As shown in Fig. 7, in general, the actual biases of the directions estimated by using the proposed method are rather small at moderately low SNRs and, thus, do not affect the overall MSE.

Example B—Performance versus Number of Snapshots: Here, we study the effect of the number of snapshots on the detection and estimation performances. The simulation conditions are the same as in the previous example, except that the SNR is fixed at 5 dB, and the number of snapshots N is varied from 10 to 1000.

Fig. 8 shows the probability of detection versus the number of snapshots. It is found that the proposed method outperforms the

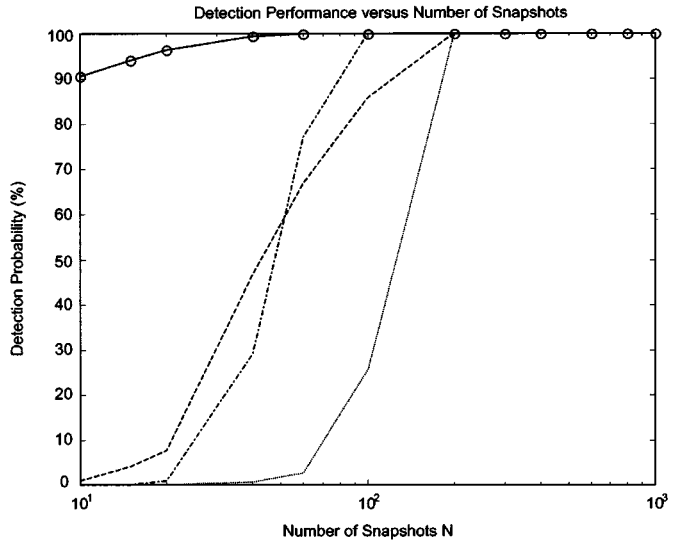

Fig. 8. Probability of correct detection versus the number of snapshots (dotted line: SS-AIC; dash-dot line: SS-MDL; dashed line: WSF detection scheme; and solid line with "o": proposed method) for Example B (SNR $= 5$ dB, $M = 10$, and $m = 5$).

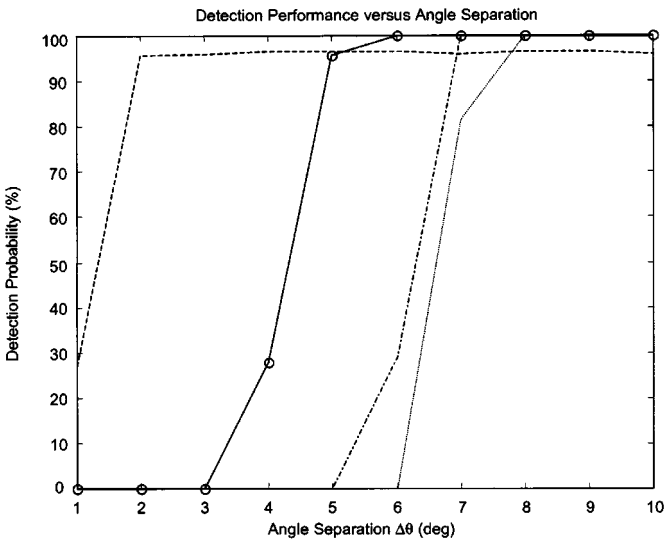

Fig. 10. Probability of correct detection versus the angle separation (dotted line: SS-AIC; dash-dot line: SS-MDL; dashed line: WSF detection scheme; and solid line with "o": proposed method) for Example C (SNR $= 5$ dB, $N = 128$, $M = 10$, and $m = 5$).

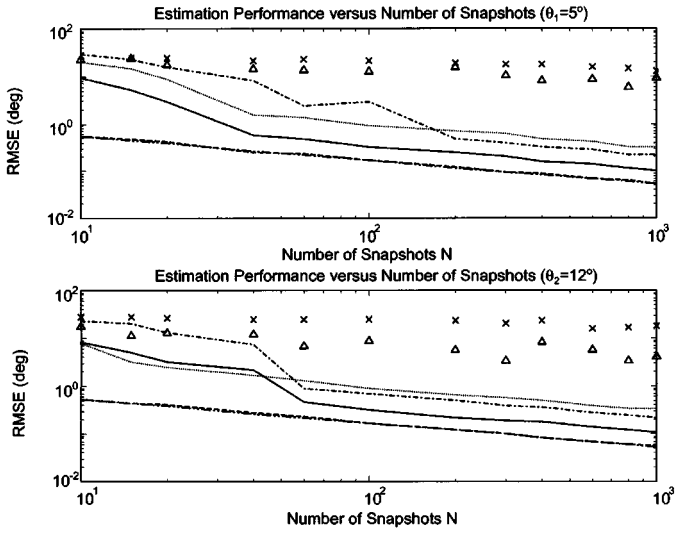

Fig. 9. RMSEs of the direction estimates versus the number of snapshots ("x": LS; "△": CLS; dotted line: SS-based root-MUSIC; dash-dot line: smoothed Min-Norm; dashed line: MODE; solid line: proposed method; and dash-dots line: CRB) for Example B (SNR $= 5$ dB, $M = 10$, and $m = 5$).

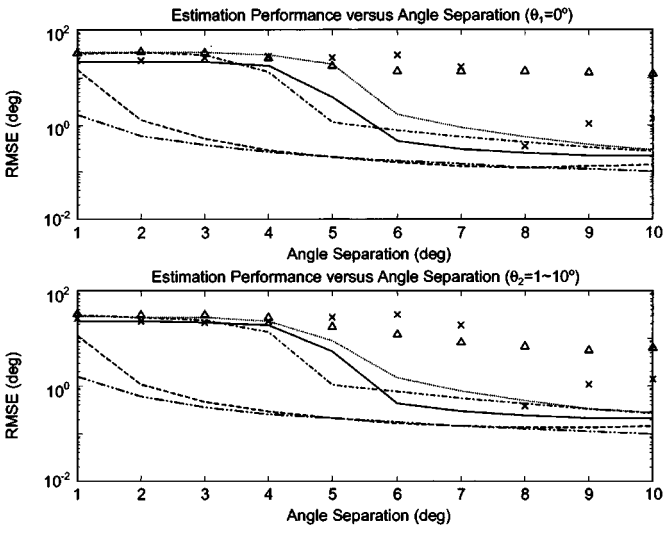

Fig. 11. RMSEs of the direction estimates versus the angle separation ("x": LS; "△": CLS; dotted line: SS-based root-MUSIC; dash-dot line: smoothed Min-Norm; dashed line: MODE; solid line: proposed method; and dash-dots line: CRB) for Example C (SNR $= 5$ dB, $N = 128$, $M = 10$, and $m = 5$).

SS-AIC, SS-MDL, and WSF detection scheme even for a small number of snapshots. As mentioned previously, it is difficult to choose an appropriate threshold $\gamma(N)$ for the hypothesis test of WSF detection scheme without any *a priori* knowledge when the number of snapshots N is finite [16], [20], [21], [43], and the performance of WSF detection scheme is degraded. In Fig. 9, the RMSEs and CRBs of the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are plotted. The proposed approach performs better than the smoothed Min-Norm, LS, and CLS methods, and it can estimate the directions with less RMSE than the SS-based root-MUSIC at relatively small lengths of data. For a larger number of snapshots, the estimated covariance matrix more closely resembles the true one and results in more precise estimation of the eigenvalues and eigenvectors. Therefore, as the number of snapshots is increased, the estimation performance of the proposed approach becomes more accurate than that of the SS-based root-MUSIC, smoothed Min-Norm, LS, and CLS methods.

Example C—Performance versus Angle Separation: In this example, we examine the performance of the proposed approach with respect to the separation between the angles of the coherent signals. The simulation conditions are similar to those in the first example, except that the SNR is fixed at 5 dB, and the two coherent signals come from $\theta_1 = 0°$ and $\theta_2 = \Delta\theta$ with $\Delta\theta$ varying from 1° to 10°.

The detection performance and the RMSEs of estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ versus angle separation $\Delta\theta$ are shown in Figs. 10 and 11, where the CRBs are also plotted for reference. As shown in Fig. 10, the SS-AIC and SS-MDL methods exhibit a sharp and sudden degradation for moderate angle separation $\Delta\theta$. The proposed approach has better detection performance for

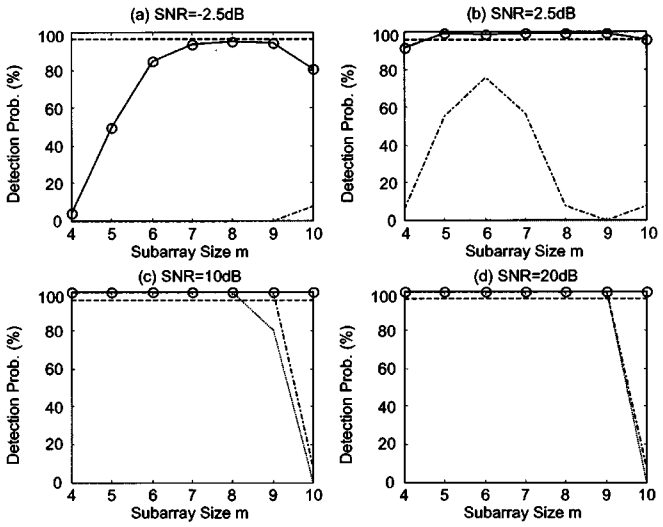

Fig. 12. Probability of correct detection versus the subarray size (dotted line: SS-AIC; dash-dot line: SS-MDL; dashed line: WSF detection scheme; and solid line with "o": proposed method) for Example D ($N = 128$, and $M = 10$).

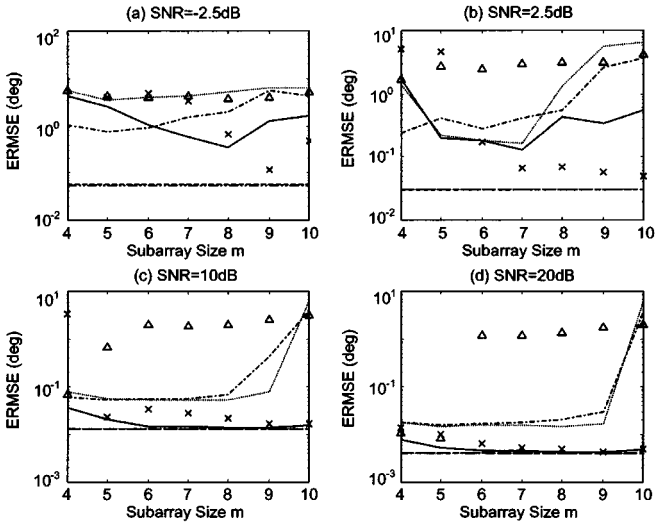

Fig. 13. ERMSEs of the direction estimates versus the subarray size ("x": LS; "△": CLS; dotted line: SS-based root-MUSIC; dash-dot line: smoothed Min-Norm; dashed line: MODE; solid line: proposed method; and dash-dots line: empirical CRB) for Example D ($N = 128$, and $M = 10$).

relatively small separation $\Delta\theta$ because the information on the eigenvectors is also exploited, though it is worse than the WSF detection scheme for smaller $\Delta\theta$. From Fig. 11, we can find that the proposed approach performs better than the LS and *a priori* knowledge-based methods such as the SS-based root-MUSIC, smoothed Min-Norm, and CLS methods for larger separation $\Delta\theta$. As we noted in the previous examples, the estimation performance is commensurate with the quality of the detected number of coherent signals when *a priori* knowledge on the number of signals and the noise variance are not available.

Example D—Performance versus Subarray Size: The impact of the subarray size on the detection and estimation is considered in this example. The parameters for the simulation are the same as in Example A, except that the subarray size m is varied from 4 to 10, where the number of subarrays is $\overline{L} = 2L = 14$ to $\overline{L} = 2$. To measure the overall estimation performance in terms of the subarray size, we define an "empirical RMSE (ERMSE)" of the direction estimates $\{\hat{\theta}_i\}$ of the coherent signals as

$$\text{ERMSE} = \sqrt{\frac{1}{q\overline{K}} \sum_{i=1}^{q} \sum_{k=1}^{\overline{K}} \left(\hat{\theta}_i^{(k)} - \theta_i\right)^2} \tag{54}$$

where $\hat{\theta}_i^{(k)}$ is the estimate obtained in the kth trial, and $\overline{K}$ is the number of trials (here, $\overline{K} = 200$).

For several SNRs, the probability of correct detection and the ERMSEs of the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ against the subarray size m are plotted in Figs. 12 and 13, in which the results of the WSF detection scheme and MODE and the "empirical CRB" are shown for comparison, where the "empirical CRB" is calculated by averaging the corresponding CRBs over the number of coherent signals. The proposed approach outperforms the SS-AIC and SS-MDL methods in terms of number detection and the CLS estimation in terms of direction estimation, regardless of the subarray size at low SNRs. Note that the choice of subarray size can significantly improve the performance of the proposed approach. For high SNR, because the number of signals is estimated correctly, the relatively minimum ERMSE is attained when the subarray size is about $m = 6$ ($\approx M/3 + 1$), and this is roughly in agreement with the results presented in [8], [15], [44], and [45]. For low SNR, the subarray size giving the minimum ERMSE is larger than this value in general. As shown in Figs. 12 and 13, the trend in detection performance versus the subarray size does not coincide with that of the estimation performance at lower SNRs. In fact, the subarray size is related to the decorrelation of the signal coherency. If the subarray size is reduced, a good decorrelation can be obtained while the resolution of the direction estimation is degraded due to the small size of the working array aperture. If the subarray size is increased, the resolution is improved, but the decorrelation effect may be unsatisfactory. The optimal subarray size generally depends on the coordinates of the coherent signals and their relative angle phases. There is a trade off between the goodness of the decorrelation and the resolution of the direction estimation; therefore, the subarray size is set at approximately $M/3+1$ in the previous examples given here.

VI. Conclusions

In most high-resolution methods of array processing for direction finding, the determination of the number of signals is an important issue. In this paper, we investigated the direction estimation of coherent narrowband signals impinging on an equally spaced linear array when the number of signals is unknown. By incorporating the linear prediction technique with a subarray scheme, we developed an MSE-based regularization approach. Analytical expressions of the asymptotic MSE and the optimal regularization parameters that minimize the MSE of the estimated LP parameters were derived, and based on them, a scheme for detecting the number of coherent signals was proposed. Furthermore, an iterative regularization algorithm was presented to estimate the arrival angles of the coherent signals from only the received noisy array data without *a priori* knowledge, where the number of signals and the noise variance are

estimated simultaneously. The effectiveness of the proposed approach was evaluated and compared with those of the conventional methods through numerical examples. The simulation results showed that the proposed method performs better than the WSF detection scheme when the number of snapshots is small and that it outperforms the SS-MDL and SS-AIC methods in detecting the number of signals. Although the proposed method is slightly inferior to MODE (with the true number of signals and dimension of signal subspace), it is superior to the other methods (with the true number of signals or noise variance) in resolving coherent signals when the number of signals has been correctly estimated.

APPENDIX A
PROOF OF THEOREM 1

The derivation of the MSE is rather tedious because the calculation of the third- and fourth-order moments of the additive noise is required [25], [39]. Here, we show only an outline of the proof.

First, by substituting (8) into the RCLS estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})$ in (18), we obtain

$$\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})-\boldsymbol{a}=(\boldsymbol{\Phi}-\sigma^2\boldsymbol{I}_{m-1}+\boldsymbol{R})^{-1}\cdot\left(\frac{1}{LN}\sum_{n=1}^{N}\boldsymbol{\Phi}^H(n)\overline{\boldsymbol{W}}(n)-\sigma^2\overline{\boldsymbol{I}}+\overline{\boldsymbol{R}}\right)\overline{\boldsymbol{a}} \tag{A1}$$

where $\overline{\boldsymbol{W}}(n)=[\overline{\boldsymbol{w}}_1(n),\overline{\boldsymbol{w}}_2(n),\ldots,\overline{\boldsymbol{w}}_L(n)]^T$, $\overline{\boldsymbol{I}}=[\boldsymbol{I}_{m-1},\boldsymbol{0}_{(m-1)\times 1}]$, $\overline{\boldsymbol{R}}=[\boldsymbol{R},\boldsymbol{0}_{(m-1)\times 1}]$, and $\boldsymbol{0}_{p\times 1}$ denotes a $p\times 1$ null vector. Note that $E\{\boldsymbol{w}_l(n)\overline{\boldsymbol{w}}_l^H(n)\}=\sigma^2\overline{\boldsymbol{I}}$.

Under the assumptions of the source signals and additive noise, by making some manipulations, we obtain the asymptotic MSE matrix $\boldsymbol{M}_a$ of the RCLS estimate $\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})$ as

$$\begin{aligned}\boldsymbol{M}_a&=\lim_{N\to\infty}\{(\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})-\boldsymbol{a})(\hat{\boldsymbol{a}}_{\mathrm{RCLS}}(\boldsymbol{R})-\boldsymbol{a})^H\}\\&=(\boldsymbol{\Sigma}_s+\boldsymbol{R})^{-1}(\boldsymbol{I}_{m-1}\otimes\overline{\boldsymbol{a}}^T)\boldsymbol{M}(\boldsymbol{I}_{m-1}\otimes\overline{\boldsymbol{a}}^*)\cdot(\boldsymbol{\Sigma}_s+\boldsymbol{R})^{-1}\end{aligned} \tag{A2}$$

where

$$\begin{aligned}\boldsymbol{M}&=\lim_{N\to\infty}\left\{\mathrm{vec}\left(\frac{1}{LN}\sum_{n=1}^{N}\overline{\boldsymbol{W}}^T(n)\boldsymbol{\Phi}^*(n)-\sigma^2\overline{\boldsymbol{I}}^T+\overline{\boldsymbol{R}}^T\right)\right.\\&\qquad\left.\cdot\mathrm{vec}^H\left(\frac{1}{LN}\sum_{n=1}^{N}\overline{\boldsymbol{W}}^T(n)\boldsymbol{\Phi}^*(n)-\sigma^2\overline{\boldsymbol{I}}^T+\overline{\boldsymbol{R}}^T\right)\right\}\\&=\boldsymbol{M}_1+\boldsymbol{M}_2+\boldsymbol{M}_3+\boldsymbol{M}_4\end{aligned} \tag{A3}$$

and we used the relation for matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$ with compatible dimensions so that $\boldsymbol{Ab}=(\boldsymbol{I}\otimes\boldsymbol{b}^T)\mathrm{vec}(\boldsymbol{A}^T)$, where $\mathrm{vec}(\boldsymbol{A}^T)$ is the vector obtained by listing the columns of $\boldsymbol{A}^T$, one beneath the other, beginning with the leftmost column, and $\otimes$ denotes the Kronecker operation.

The first term $\boldsymbol{M}_1$ in (A3) is given by

$$\begin{aligned}\boldsymbol{M}_1&=\lim_{N\to\infty}\left\{\mathrm{vec}\left(\frac{1}{LN}\sum_{n=1}^{N}\overline{\boldsymbol{W}}^T(n)\boldsymbol{\Phi}^*(n)-\sigma^2\overline{\boldsymbol{I}}^T\right)\right.\\&\qquad\left.\cdot\mathrm{vec}^H\left(\frac{1}{LN}\sum_{n=1}^{N}\overline{\boldsymbol{W}}^T(n)\boldsymbol{\Phi}^*(n)-\sigma^2\overline{\boldsymbol{I}}^T\right)\right\}\\&=\frac{1}{LN}E\{\mathrm{vec}(\overline{\boldsymbol{w}}_l(n)\boldsymbol{x}_l^H(n))\mathrm{vec}^H(\overline{\boldsymbol{w}}_l(n)\boldsymbol{x}_l^H(n))\\&\qquad+\mathrm{vec}(\overline{\boldsymbol{w}}_l(n)\boldsymbol{x}_l^H(n))\mathrm{vec}^H(\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n))\\&\qquad+\mathrm{vec}(\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n))\mathrm{vec}^H(\overline{\boldsymbol{w}}_l(n)\boldsymbol{x}_l^H(n))\\&\qquad+\mathrm{vec}(\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n)-E\{\overline{\boldsymbol{w}}_l^*(n)\boldsymbol{w}_l^T(n)\})\\&\qquad\cdot\mathrm{vec}^H(\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n)-E\{\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n)\})\}\\&=\frac{1}{LN}(\overline{\boldsymbol{M}}_1+\overline{\boldsymbol{M}}_2+\overline{\boldsymbol{M}}_3+\overline{\boldsymbol{M}}_4).\end{aligned} \tag{A4}$$

From the fact that

$$\begin{aligned}\mathrm{vec}(\boldsymbol{hb}^H)\mathrm{vec}^H(\boldsymbol{cd}^H)&=(\boldsymbol{bd}^H)\otimes(\boldsymbol{h}\cdot\boldsymbol{c}^H)\\&=\boldsymbol{b}\otimes\boldsymbol{d}^H\otimes(\boldsymbol{hc}^H)\end{aligned}$$

where $\boldsymbol{h}$ and $\boldsymbol{b}$, and $\boldsymbol{c}$ and $\boldsymbol{d}$ are vectors with compatible dimensions, we can obtain the terms $\overline{\boldsymbol{M}}_1$, $\overline{\boldsymbol{M}}_2$, and $\overline{\boldsymbol{M}}_3$ in (A4) as

$$\begin{aligned}\overline{\boldsymbol{M}}_1&=E\{\mathrm{vec}(\overline{\boldsymbol{w}}_l(n)\boldsymbol{x}_l^H(n))\mathrm{vec}^H(\overline{\boldsymbol{w}}_l(n)\boldsymbol{x}_l^H(n))\}\\&=\boldsymbol{\Sigma}_s\otimes\sigma^2\boldsymbol{I}_m\end{aligned} \tag{A5}$$

$$\begin{aligned}\overline{\boldsymbol{M}}_2&=\overline{\boldsymbol{M}}_3^H\\&=E\{\mathrm{vec}(\overline{\boldsymbol{w}}_l(n)\boldsymbol{x}_l^H(n))\mathrm{vec}^H(\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n))\}\\&=\boldsymbol{O}_{m(m-1)}\end{aligned} \tag{A6}$$

where $\boldsymbol{O}_m$ is an $m\times m$ null matrix. In addition, the fourth term $\overline{\boldsymbol{M}}_4$ in (A4) can be obtained as [25]

$$\begin{aligned}\overline{\boldsymbol{M}}_4&=E\{\mathrm{vec}(\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n)-E\{\overline{\boldsymbol{w}}_l^*(n)\boldsymbol{w}_l^T(n)\})\\&\qquad\cdot\mathrm{vec}^H(\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n)-E\{\overline{\boldsymbol{w}}_l(n)\boldsymbol{w}_l^H(n)\})\}\end{aligned} \tag{A7}$$

where the ikth $m\times m$ block element $\overline{\boldsymbol{M}}_{4ik}$ of $\overline{\boldsymbol{M}}_4$ is expressed by

$$\begin{aligned}\overline{\boldsymbol{M}}_{4ik}&=E\{(w_{l+i-1}^*(n)\overline{\boldsymbol{w}}_l(n)-E\{w_{l+i-1}(n)\overline{\boldsymbol{w}}_l^*(n)\})\\&\qquad\cdot(w_{l+k-1}(n)\overline{\boldsymbol{w}}_l^H(n)-E\{w_{l+k-1}(n)\overline{\boldsymbol{w}}_l^H(n)\})\}\\&=E\{w_{l+i-1}^*(n)w_{l+k-1}(n)\overline{\boldsymbol{w}}_l(n)\overline{\boldsymbol{w}}_l^H(n)\}\\&\qquad-E\{w_{l+i-1}^*(n)\overline{\boldsymbol{w}}_l(n)\}E\{w_{l+k-1}(n)\overline{\boldsymbol{w}}_l^H(n)\}\end{aligned} \tag{A8}$$

for $i, k = 1, 2, \ldots, m-1$. Furthermore, the hpth element $(\overline{\boldsymbol{M}}_{4ik})_{hp}$ of $\overline{\boldsymbol{M}}_{4ik}$ is given by

$$\begin{aligned}(\overline{\boldsymbol{M}}_{4ik})_{hp} =& E\{(w^*_{l+i-1}(n)w_{l+k-1}(n)\}E\{w_{l+h-1}(n) \\ & \cdot w^*_{l+p-1}(n)\} + E\{(w^*_{l+i-1}(n)w^*_{l+p-1}(n)\} \\ & \cdot E\{w_{l+k-1}(n)w_{l+h-1}(n)\} \end{aligned} \tag{A9}$$

for $h, p = 1, 2, \ldots, m$. Because $E\{w_i(n)w^*_k(n)\} = \sigma^2\delta_{i,k}$ and $E\{w_i(n)w_k(n)\} = 0$, we have

$$(\overline{\boldsymbol{M}}_{4ik})_{hp} = 0, \qquad \text{for } i \neq k \tag{A10}$$

$$(\overline{\boldsymbol{M}}_{4ik})_{hp} = \begin{cases} \sigma^4, & \text{for } h = p \\ 0, & \text{for } h \neq p \end{cases} \qquad \text{for } i = k. \tag{A11}$$

Accordingly, the ikth $m \times m$ block element $\overline{\boldsymbol{M}}_{4ik}$ of $\overline{\boldsymbol{M}}_4$ in (A7) is given by

$$\overline{\boldsymbol{M}}_{4ik} = \begin{cases} \boldsymbol{O}_m, & \text{for } i \neq k \\ \sigma^4\boldsymbol{I}_m, & \text{for } i = k. \end{cases} \tag{A12}$$

From (A5), (A6), and (A12), the first term $\boldsymbol{M}_1$ of $\boldsymbol{M}$ in (A3) can be expressed as

$$\boldsymbol{M}_1 = \frac{1}{LN}(\boldsymbol{\Sigma}_s \otimes \sigma^2\boldsymbol{I}_m + \overline{\boldsymbol{M}}_4). \tag{A13}$$

Additionally, we can obtain the other terms of $\boldsymbol{M}$ in (A3) as

$$\begin{aligned}\boldsymbol{M}_2 =& \boldsymbol{M}_3^H \\ =& \text{vec}(\overline{\boldsymbol{R}}^T) \\ & \cdot \lim_{N\to\infty}\left\{\text{vec}^H\left(\frac{1}{LN}\sum_{n=1}^{N}\overline{\boldsymbol{W}}^T(n)\boldsymbol{\Phi}^*(n) - \sigma^2\overline{\boldsymbol{I}}^T\right)\right\} \\ =& \boldsymbol{O}_{m(m-1)} \end{aligned} \tag{A14}$$

$$\boldsymbol{M}_4 = \text{vec}(\overline{\boldsymbol{R}}^T)\text{vec}^H(\overline{\boldsymbol{R}}^T). \tag{A15}$$

By concatenating (A13)–(A15) into (A3), we obtain

$$\boldsymbol{M} = \frac{1}{LN}(\boldsymbol{\Sigma}_s \otimes \sigma^2\boldsymbol{I}_m + \overline{\boldsymbol{M}}_4) + \text{vec}(\overline{\boldsymbol{R}}^T)\text{vec}^H(\overline{\boldsymbol{R}}^T). \tag{A16}$$

From (A2) and (A16) and by some manipulations, the asymptotic MSE matrix $\boldsymbol{M}_a$ in (A2) of the estimate $\hat{\boldsymbol{a}}_{\text{RCLS}}(\boldsymbol{R})$ is obtained as

$$\begin{aligned}\boldsymbol{M}_a =& \frac{1}{LN}(\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1}(\boldsymbol{\Sigma}_s \otimes (\sigma^2\overline{\boldsymbol{a}}^T\overline{\boldsymbol{a}}^*) + \sigma^4\overline{\boldsymbol{a}}^T\overline{\boldsymbol{a}}^*\boldsymbol{I}_{m-1}) \\ & \cdot (\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1} + (\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1}\overline{\boldsymbol{R}}\overline{\boldsymbol{a}}\overline{\boldsymbol{a}}^H\overline{\boldsymbol{R}}^H(\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1} \\ =& \frac{\sigma^2}{LN}(1 + \|\boldsymbol{a}\|^2)(\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1}(\boldsymbol{\Sigma}_s + \sigma^2\boldsymbol{I}_{m-1}) \\ & \cdot (\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1} + (\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1}\boldsymbol{R}\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{R}^H(\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-1}. \end{aligned} \tag{A17}$$

Consequently, from the fact that $\text{MSE} = \text{tr}\{\boldsymbol{M}_a\}$, we obtain the MSE of the estimate $\hat{\boldsymbol{a}}_{\text{RCLS}}(\boldsymbol{R})$ in (18) as

$$\begin{aligned}\text{MSE} =& \frac{\sigma^2}{LN}(1 + \|\boldsymbol{a}\|^2)\text{tr}\{(\boldsymbol{\Sigma}_s + \sigma^2\boldsymbol{I}_{m-1})(\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-2}\} \\ & + \boldsymbol{a}^H\boldsymbol{R}^H(\boldsymbol{\Sigma}_s + \boldsymbol{R})^{-2}\boldsymbol{R}\boldsymbol{a} \end{aligned} \tag{A18}$$

where $\text{tr}\{\cdot\}$ denotes the trace operation. By substituting (13) and (19) into (A18), the asymptotic MSE can be obtained immediately. ■

Appendix B
Proof of Theorem 3

Because the minimization of $\text{MSE}(\{\rho_i\})$ in (21) can be performed independently for each index i, then for $\rho_i = 0$ and $\rho_i = \infty$, the ith terms of the MSE can be obtained, respectively, as

$$\text{MSE}(\{0\})_i = \sigma^2(1 + \|\boldsymbol{a}\|^2)\frac{\lambda_i + \sigma^2}{LN\lambda_i^2} \tag{B1}$$

$$\text{MSE}(\{\infty\})_i = |\boldsymbol{v}_i^H\boldsymbol{a}|^2. \tag{B2}$$

If we set the value $\rho_i = 0$ as an optimal regularization parameter rather than $\rho_i = \infty$ for $i = 1, 2, \ldots, q$, then the inequality $\text{MSE}(\{0\})_i \leq \text{MSE}(\{\infty\})_i$ should hold, i.e.,

$$\lambda_i \geq \sigma^2(1 + \|\boldsymbol{a}\|^2)\frac{\lambda_i + \sigma^2}{LN\lambda_i|\boldsymbol{v}_i^H\boldsymbol{a}|^2} = \rho_i^o. \tag{B3}$$

Hence, we have an optimal rule for retaining the eigenvalue λ_i as $\lambda_i \geq \rho_i^o$ for $i = 1, 2, \ldots, q$. If we set the value of the regularization parameter as $\rho_i = \infty$, then $\text{MSE}(\{0\})_i > \text{MSE}(\{\infty\})_i$ should be satisfied. Hence, we have an optimal rule for discarding an eigenvalue as $\lambda_i < \rho_i^o$. The optimal truncation condition can thus be established. ■

Acknowledgment

The authors would like to express their gratitude to the anonymous referees and the associate editor Prof. V. Krishnamurthy for the careful review, insightful comments, and constructive suggestions that improved the manuscript. They are particularly grateful to the referees for bringing [48]–[53] to their attention.

References

[1] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, vol. 33, pp. 247–266, 1973.
[2] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.
[3] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 984–995, July 1989.
[4] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, pp. 134–139, Jan. 1983.
[5] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoid: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, Sept. 1982.

[6] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the Minimum-Norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 331–341, Feb. 1986.

[7] B. D. Rao, "Perturbation analysis of an SVD-based linear prediction method for estimating the frequencies of multiple sinusoids," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1026–1035, July 1988.

[8] M. A. Rahman and K.-B. Yu, "Total least squares approach for frequency estimation using linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1440–1454, Oct. 1987.

[9] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806–811, Apr. 1985.

[10] J. A. Cadzow, "A high resolution direction-of-arrival algorithm for narrow-band coherent and incoherent sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 965–979, July 1988.

[11] S. S. Reddi and A. B. Gershman, "An alternative approach to coherent source location problem," *Signal Process.*, vol. 59, no. 2, pp. 221–233, 1997.

[12] H. Krim, J. H. Cozzens, and J. G. Proakis, "On spatial smoothing and linear prediction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Albuquerque, NM, Apr. 1990, pp. 2639–2643.

[13] H. Krim and J. G. Proakis, "Smoothed eigenspace-based parameter estimation," *Automatica*, vol. 30, no. 1, pp. 27–38, 1994.

[14] T. J. Abatzoglou, J. M. Mendel, and G. A. Harada, "The constrained total least squares technique and its application to harmonic superresolution," *IEEE Trans. Signal Processing*, vol. 39, pp. 1070–1087, May 1991.

[15] W. M. Steedly, C. J. Ying, and R. L. Moses, "Statistical analysis of TLS-based Prony techniques," *Automatica*, vol. 30, no. 1, pp. 115–129, 1994.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Feb. 1985.

[17] T.-J. Shan, A. Paulraj, and T. Kailath, "On smoothed rank profile tests in eigenstructure methods for directions-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1377–1385, Oct. 1987.

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

[19] W. Chen, K. M. Wong, and J. P. Reilly, "Detection of the number of signals: A predicted eigen-threshold approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 1088–1098, May 1991.

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

[21] G. Xu, R. H. Roy, III, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Processing*, vol. 42, pp. 102–112, Jan. 1994.

[22] H. Krim and J. H. Cozzens, "A data-based enumeration technique for fully correlated signals," *IEEE Trans. Signal Processing*, vol. 42, pp. 1662–1668, July 1994.

[23] C.-M. Cho and P. M. Djuric, "Detection and estimation of DOA's of signals via Bayesian predictive densities," *IEEE Trans. Signal Processing*, vol. 42, pp. 3051–3060, Nov. 1994.

[24] T. Söderström, "Identification of stochastic linear systems in presence of input noise," *Automatica*, vol. 17, no. 5, pp. 713–725, 1981.

[25] H. Schneeweiss, "Consistent estimation of a regression with errors in the variables," *Metrika*, vol. 23, pp. 101–115, 1976.

[26] R. H. Ketellapper, "On estimating parameters in a simple linear errors-in-variables model," *Technometr.*, vol. 25, no. 1, pp. 43–47, 1983.

[27] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problems: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.

[28] D. W. Tufts and R. Kumaresan, "Singular value decomposition and improved frequency estimation using linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 671–675, Apr. 1982.

[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins Univ. Press, 1989.

[30] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signals identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 8–15, Jan. 1989.

[31] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, pp. 1018–1028, Sept. 1982.

[32] F. Li, R. J. Vaccaro, and D. W. Tufts, "Min-Norm linear prediction for arbitrary sensor arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Glasgow, U.K., May 1989, pp. 2613–2616.

[33] R. Kumaresan, D. W. Tufts, and L. L. Scharf, "A Prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models," *Proc. IEEE*, vol. 72, pp. 230–233, Feb. 1984.

[34] M. Bertero, C. De Mol, and E. R. Pike, "Linear inversion problem with discrete data: Stability and regularization," *Inv. Prob.*, vol. 4, pp. 573–594, 1988.

[35] R. Kumaresan, "On the zeros of the linear prediction filter for deterministic signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 217–220, Jan. 1983.

[36] P. C. Hansen, "Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank," *SIAM J. Sci. Stat. Comput.*, vol. 11, no. 3, pp. 503–518, 1990.

[37] A. Sano, "Optimally regularized inverse of singular value decomposition and application to signal extrapolation," *Signal Process.*, vol. 30, no. 2, pp. 163–176, 1993.

[38] S. Haykin, "Radar array processing for angle of arrival estimation," in *Array Signal Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1985, pp. 194–292.

[39] J. Xin, "Minimum MSE based identification of discrete-time systems," Ph.D. dissertation, Keio University, Yokohama, Japan, Mar. 1996.

[40] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[41] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer–Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.

[42] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by altering projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553–1560, Oct. 1988.

[43] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 2436–2449, Nov. 1991.

[44] H. Clergeot, S. Tressens, and A. Ouamri, "Performance of high resolution frequencies estimation methods compared to the Cramer–Rao bounds," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1703–1720, Nov. 1989.

[45] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 814–824, May 1990.

[46] P. C. Hansen, "The truncation SVD as a method for regularization," *BIT*, vol. 27, pp. 534–553, 1987.

[47] L. Elden, "Algorithms for the regularization of ill-conditioned least squares problems," *BIT*, vol. 17, pp. 134–145, 1977.

[48] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. J. Shepherd, Eds. Berlin, Germany: Springer-Verlag, 1993, pp. 99–151.

[49] M. Kristensson, M. Jansson, and B. Ottersten, "Modified IQML and weighted subspace fitting without eigendecomposition," *Signal Process.*, vol. 79, no. 1, pp. 29–44, 1999.

[50] P. Stoica and K. C. Sharman, "Novel eigenanalysis method for direction estimation," *Proc. Inst. Elect. Eng. F*, vol. 137, no. 1, pp. 19–26, 1990.

[51] ——, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1132–1143, July 1990.

[52] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.

[53] P. Stoica, B. Ottersten, M. Viberg, and R. L. Moses, "Maximum likelihood array processing for stochastic coherent sources," *IEEE Trans. Signal Processing*, vol. 44, pp. 96–104, Jan. 1996.

[54] P. Stoica and M. Jansson, "On forward-backward MODE for array signal processing," *Digital Signal Process.*, vol. 7, pp. 239–252, 1997.

[55] J. Li, P. Stoica, and Z.-S. Liu, "Comparative study of IQML and MODE direction-of-arrival estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 149–160, Jan. 1998.

[56] M. Wax, "Detection and localization of multiple sources via the stochastic signals model," *IEEE Trans. Signal Processing*, vol. 39, pp. 2450–2456, Nov. 1991.

[57] P. Stoica, T. Söderström, and V. Simonyte, "On estimating the noise power in array processing," *Signal Process.*, vol. 26, no. 2, pp. 205–220, 1992.

[58] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Boston, MA, 1983, pp. 336–339.

[59] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Mag.*, vol. 13, pp. 67–94, July 1996.
[60] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.

Jingmin Xin (S'92–M'96) received the B.E. degree in information and control engineering from Xi'an Jiaotong University, Xi'an, China, in 1988 and the M.E. and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1993 and 1996, respectively.

From 1988 to 1990, he was with the Tenth Institute of the Ministry of Posts and Telecommunications (MPT) of China, Xi'an. He was with Communications Research Laboratory, MPT of Japan, as an Invited Research Fellow of Telecommunications Advancement Organization of Japan (TAO) from 1996 to 1997 and as a Postdoctoral Fellow of the Japan Science and Technology Corporation (JST) from 1997 to 1999. Since 1999, he has been with YRP Mobile Telecommunications Key Technology Research Laboratories Co., Ltd., Yokosuka, Japan, and he is currently a Guest Senior Researcher. His research interests are in the areas of digital signal processing and system identification and their applications to mobile communication systems.

Akira Sano (M'89) received the B.E., M.E., and Ph.D. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1966, 1968, and 1971, respectively.

He joined the Department of Electrical Engineering, Keio University, Yokohama, Japan, in 1971, where he is currently a Professor with the Department of System Design Engineering. He was a Visiting Research Fellow at the University of Salford, Salford, U.K., from 1977 to 1978. Since 1995, he has been a Visiting Research Fellow with the Communication Research Laboratory, Ministry of Posts and Telecommunications. His current research interests are in adaptive modeling and design theory in control, signal processing and communication, and applications to control of sounds and vibrations, mechanical systems, and mobile communication systems. He is a coauthor of the textbook *State Variable Methods in Automatic Control* (New York: Wiley, 1988).

Dr. Sano received the Kelvin Premium from the Institute of Electrical Engineering in 1986. He is a Fellow of the Society of Instrument and Control Engineers and is a Member of the Institute of Electrical Engineering of Japan and the Institute of Electronics, Information, and Communications Engineers of Japan. He was General Co-Chair of the 1999 IEEE Conference of Control Applications and has served as Chair of the IFAC Technical Committee on Modeling and Control of Environmental Systems since 1996. He has also been Vice Chair of IFAC Technical Committee on Adaptive Control and Learning since 1999.