IEEE TRANSACTIONS ON

SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.ieee.org/sp/index.html

DECEMBER 2005 VOLUME 53 NUMBER 12 ITPRED (ISSN 1053-587X)

PAPERS

(Contents Continued on Back Cover)

IEEE

(Contents Continued from Front Cover)

Efficient Subspace-Based Algorithm for Adaptive Bearing Estimation and Tracking

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

***Abstract*—In some practical applications of array processing, the directions of the incident signals should be estimated adaptively, and/or the time-varying directions should be tracked promptly. In this paper, an adaptive bearing estimation and tracking (ABEST) algorithm is investigated for estimating and tracking the uncorrelated and correlated narrow-band signals impinging on a uniform linear array (ULA) based on the subspace-based method without eigendecomposition (SUMWE), where a linear operator is obtained from the array data to form a basis for the null space by exploiting the array geometry and its shift invariance property. Specifically, the null space is estimated using the least-mean-square (LMS) or normalized LMS (NLMS) algorithm, and the directions are updated using the approximate Newton method. The transient analyses of the LMS and NLMS algorithms are studied, where the "weight" (i.e., the linear operator) is in the form of a matrix and there is a correlation between the "additive noise" and "input data" that involve the instantaneous correlations of the received array data in the updating equation, and the step-size stability conditions are derived explicitly. In addition, the analytical expressions for the mean-square error (MSE) and mean-square deviation (MSD) learning curves of the LMS algorithm are clarified. The effectiveness of the ABEST algorithm is verified, and the theoretical analyses are corroborated through numerical examples. Simulation results show that the ABEST algorithm is computationally simple and has good adaptation and tracking abilities.**

***Index Terms*—Adaptive filtering, direction-of-arrival (DOA) estimation, eigendecomposition, learning curve, subspace-based methods, transient analysis.**

I. Introduction

SUBSPACE-BASED methods have been extensively studied for estimating the directions-of-arrivals (DOAs) of the incident signals in array processing or the frequencies of (complex-valued) sinusoids in time-series spectral analysis from the noisy measurements because of their high-resolution and computational simplicity (e.g., [1]–[12]). Traditionally, most subspace-based methods require either the eigenvalue decomposition (EVD) of a covariance matrix or the singular value decomposition (SVD) of a data matrix to compute the signal or noise subspace, and these methods are usually implemented in batch mode. Unfortunately, the eigendecomposition is computationally intensive and time-consuming [13]–[15], especially when the number of sensors (or the assumed order of the signal model) is large. Consequently, in some practical situations (e.g., [55]), where the directions of the incident signals (or sinusoid frequencies) should be estimated adaptively and/or the time-varying directions/frequencies should be tracked promptly, repeated EVD/SVD computation of a continually updated sample covariance or data matrix is very burdensome in general, and this heavy computational load often makes subspace-based methods difficult to implement in an online manner. To reduce the computational burden of eigendecomposition, many efficient techniques have been developed over the past two decades from different perspectives, such as computing only a few eigenvectors or a subspace basis, approximating the eigenvectors or basis, and recursively updating the eigenvectors or basis (see [15]–[18] and references therein).

Recently, computationally simple subspace-based methods have been proposed for estimating the directions of narrow-band signals efficiently [19]–[22], where the need for computation of EVD/SVD is avoided. The representative methods are the bearing estimation without eigendecomposition (BEWE) [19], propagator method (PM) and orthonormal propagator method (OPM) [57], [20], and subspace methods without eigendecomposition (SWEDE) [21], in which the exact signal/noise subspace is easily obtained from the array data based on a partition of the array response matrix. The PM, in particular, has been well studied in various aspects in the recent decade [20], [37], [58]–[65]. For improving the estimation performance, the OPM was proposed to orthonormalize the noise subspace basis in [20], and two variants of OPM were presented to estimate the propagator (or a part of that) by using three (or two) subarray covariance matrices of three nonoverlapping subarrays to eliminate the noise effect in [59], [64], and [65] in a similar way to the SWEDE. The statistical performances of the PM and OPM were analyzed and compared with the other "linear" subspace-based methods such as the BEWE and SWEDE in [60], [63], [64], and [65]. In addition, the real-time implementation of the OPM was considered, where two gradient-based algorithms and a recursive least squares (RLS)-based algorithm were presented for direction finding and tracking in [61], [62], [37], and the convergence and optimal step-size of the gradient-based adaptive algorithm were studied based on a first-order approximation for small step-size (i.e., slow variations with time) in [61] and [62]. Furthermore, the least squares (LS)-type online implementation of the SWEDE was considered in [21], and an adaptive algorithm based on the projection approximation subspace tracking with deflation (PASTd) [18] and the approximate Newton iteration for direction updating was proposed in [66B]. However, the performances of these simple subspace-based methods and algorithms [19]–[22], [37], [57]–[66] degrade severely when the incident signals are coherent (i.e., fully

Manuscript received June 1, 2004; revised February 13, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jean-Pierre Delmas.

J. Xin is with the Wireless Systems Laboratories, Fujitsu Laboratories Ltd., Yokosuka 239-0847, Japan (e-mail: jxin@jp.fujitsu.com).

A. Sano is with the Department of System Design Engineering, Keio University, Yokohama 223-8522, Japan (e-mail: sano@sd.keio.ac.jp).

Digital Object Identifier 10.1109/TSP.2005.859329

correlated), and/or the signal-to-noise ratio (SNR) is low and the data length is short. The sensor array is usually arranged in a regular and structured geometry, and this special geometry is useful in developing computationally efficient direction estimation methods [24]. By exploiting the array geometry of a uniform linear array (ULA) and its shift-invariance property, we proposed a subspace-based method without eigendecomposition (SUMWE) for estimating the directions of the uncorrelated and correlated signals impinging on the ULA [23]. We also presented an LS-type adaptive algorithm for the DOA estimation of coherent cyclostationary signals [38]; regrettably it is based on the special temporal property of signals, and its computational load is heavy.

In this paper, we address the problem of adaptive DOA estimation and tracking of narrow-band signals in a computationally efficient way and propose an adaptive bearing estimation and tracking (ABEST) algorithm for estimating and tracking the uncorrelated and correlated signals impinging on a ULA based on the SUMWE [23], where a linear operator is obtained from the array data to form a basis for the null space by exploiting the array geometry and its shift invariance property. Because the direction estimator is a complicated nonlinear function of the received array data, the real-time implementation of a subspace-based method involves the estimation of the signal/noise subspace and the minimization of a cost function for direction finding. Inspired by the computational simplicity and easy implementation of the least-mean-square (LMS) algorithm and the fast tracking capability of Newton method [25]–[28], [68], we use the LMS or normalized LMS (NLMS) algorithm for the null space estimation and use the approximate Newton method to update the direction finding in the ABEST algorithm. The ABEST algorithm inherits the advantages of the SUMWE: the reduced computational load, a less restrictive model of additive noise, and a remarkable insensitivity to the correlation of the incident signals. Although the behaviors of the LMS algorithm and its variants have been analyzed extensively in the adaptive filtering literature (e.g., [25]–[36] and references therein), these studies were generally done under different assumptions and with approximations to simplify the analysis and produce reliable results. One of the most common assumptions is the independence between the additive noise and input data. However, in the updating equation of the ABEST algorithm, the "weight" (i.e., the linear operator) is in the form of a matrix, and there is a correlation between the "additive noise" and "input data" that involve the instantaneous correlations of the received array data (and additive noise). In this paper, the laborious transient analyses of the LMS and NLMS algorithms are studied, and the step-sizes that guarantee the mean and mean-square stabilities are derived explicitly. Furthermore, the analytical expressions of the mean-square error (MSE) and mean-square deviation (MSD) learning curves of the LMS algorithm are clarified, and an analytical study of the LMS stability is performed for the case of a single signal. The estimation performance of the ABEST algorithm is demonstrated in stationary and nonstationary environments. The simulation results show that the ABEST algorithm is computationally simple and has good adaptability and tracking capabilities and that there is relatively good agreement between the theoretical analyzes and practical results.

II. Problem Formulation and Preliminaries

A. Data Model and Basic Assumptions

We consider a ULA of M identical and omnidirectional sensors with adjacent spacing d, and we assume that p narrow-band signals $\{s_k(n)\}$ with the center frequency f_0 are in the field far from the array and impinge on the array from distinct bearings (i.e., directions) $\{\theta_k(n)\}$. Under the narrow-band assumption, the received noisy signal $y_i(n)$ at the ith sensor can be expressed as [1]–[12], [19]–[24]

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

$$x_i(n) = \sum_{k=1}^{p} s_k(n)\, e^{j\omega_0(i-1)\tau(\theta_k(n))} \tag{2}$$

where $x_i(n)$ is the noiseless received signal, $w_i(n)$ is the additive noise, $\omega_0 \triangleq 2\pi f_0$, $\tau(\theta_k(n)) \triangleq (d/c)\sin\theta_k(n)$, c is the propagation speed, and the DOAs $\{\theta_k(n)\}$ are measured relative to the normal of array. The received signals can be re-expressed more compactly as

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

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

We make the following basic assumptions about the data model, which are similar to those introduced in [23].

1) The array is calibrated and the array response matrix $\boldsymbol{A}(\theta(n))$ is unambiguous, i.e., the array response vectors $\{\boldsymbol{a}(\theta_1(n)), \boldsymbol{a}(\theta_2(n)), \ldots, \boldsymbol{a}(\theta_p(n))\}$ are linearly independent for any set of distinct directions $\{\theta_1(n), \theta_2(n), \ldots, \theta_p(n)\}$. Equivalently, the matrix $\boldsymbol{A}(\theta(n))$ has full column rank.
2) Without loss of generality, the signals $\{s_k(n)\}$ are all coherent so that they are all some complex multiples of a common signal $s_1(n)$; then, under the flat-fading multipath propagation, they can be expressed as [1], [3], [12], [39]–[44]

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

where β_k is the unknown complex attenuation coefficient with $\beta_k \neq 0$ and $\beta_1 = 1$.
3) For the simplicity of theoretical performance analysis, the incident signal $s_1(n)$ is a temporally complex white Gaussian random process with zero-mean, and the variance given by

$$E\{s_1(n)s_1^*(t)\} = r_s\delta_{n,t}, \quad E\{s_1(n)s_1(t)\} = 0, \quad \forall n, t \tag{5}$$

where $E\{\cdot\}$, $(\cdot)^*$, and $\delta_{n,t}$ denote the expectation, the complex conjugate, and Kronecker delta.

4) The additive noise $\{w_i(n)\}$ is a temporally and spatially complex white Gaussian random process with zero-mean and the following covariance matrix:

$$E\{\boldsymbol{w}(n)\boldsymbol{w}^H(t)\} = \sigma^2 \boldsymbol{I}_M \delta_{n,t}, \quad \forall n,t$$
$$E\{\boldsymbol{w}(n)\boldsymbol{w}^T(t)\} = \boldsymbol{O}_{M\times M}, \quad \forall n,t \tag{6}$$

where $\boldsymbol{I}_{\mathrm{m}}$, $\boldsymbol{O}_{\mathrm{m}\times q}$, and $(\cdot)^H$ indicate the $m \times m$ identity matrix, the $m\times q$ null matrix, and the Hermitian transpose, and the additive noise is uncorrelated with the incident signals.

5) The number of incident signals p is known, and it satisfies the inequality that $p < M/2$.

Remark 1: Although the incident signals are assumed to be all coherent, the proposed ABEST algorithm can be extended to the case of partly coherent or incoherent signals (cf. [23]). The identifiability condition that guarantees the uniqueness of direction estimation is that $M > 2p$, which is less strict than the necessary condition $M \geq 3p/2$ with probability one [45]. Furthermore, if the number of incident signals is unknown, it can be estimated by some proposed techniques (see [12] and references therein). □

Remark 2: The tracking of the crossing directions is not considered in this paper. In this situation, we would need to introduce a dynamic model of the incident directions. Moreover, an additional procedure based on some kinematic parameters such as bearing velocities and accelerations could be useful in predicting the crossover points on the trajectories and blocking the tracking procedure during crossover intervals [46], [47]. Such an elaborate procedure is currently under investigation. □

In this paper, we address the adaptive estimation and tracking of the DOAs of coherent signals in a computationally simple way. An efficient subspace-based algorithm is developed, and the statistical properties of the LMS and NLMS algorithms used for the null space estimation are analyzed in details.

B. Review of SUMWE

The SUMWE is a computationally efficient and batch method for estimating the constant directions of the incident signals [23]. Here, this method is briefly summarized, where the DOAs are assumed to be time invariant (i.e., $\theta_k(n) = \theta_k$), and $\boldsymbol{A}(\theta(n))$ is denoted as $\boldsymbol{A}$ for simplicity.

Under the basic assumptions, from (3), we obtain the array covariance matrix $\boldsymbol{R}$ as

$$\boldsymbol{R} \triangleq E\{\boldsymbol{y}(n)\boldsymbol{y}^H(n)\} = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + \sigma^2\boldsymbol{I}_M \tag{7}$$

where $\boldsymbol{R}_s \triangleq E\{\boldsymbol{s}(n)\boldsymbol{s}^H(n)\}$. Due to the coherency between the incident signals, the source signal covariance matrix $\boldsymbol{R}_s$ is singular (i.e., $\mathrm{rank}(\boldsymbol{R}_s) < p$, when $p > 1$). By defining the correlation r_{ik} between the signals $y_i(n)$ and $y_k(n)$ as $r_{ik} \triangleq E\{y_i(n)y_k^*(n)\}$, where $r_{ik} = r_{ki}^*$, we find that the diagonal elements $\{r_{kk}\}$ of $\boldsymbol{R}$ are affected by the noise.

Now, we divide the full array into L overlapping subarrays with p sensors in the forward and backward directions [39], [40], where $L = M - p + 1$, and the (conjugate) signals in the lth forward/backward subarray can be expressed in a compact form for $l = 1, 2, \ldots, L$ [12], [23], [43]

$$\boldsymbol{y}_{fl}(n) \triangleq [y_l(n), y_{l+1}(n), \ldots, y_{l+p-1}(n)]^T$$
$$= \boldsymbol{A}_1\boldsymbol{D}^{l-1}\boldsymbol{s}(n) + \boldsymbol{w}_{fl}(n) \tag{8}$$
$$\boldsymbol{y}_{bl}(n) \triangleq [y_{M-l+1}(n), y_{M-l}(n), \ldots, y_{L-l+1}(n)]^H$$
$$= \boldsymbol{A}_1\boldsymbol{D}^{-(M-l)}\boldsymbol{s}^*(n) + \boldsymbol{w}_{bl}(n) \tag{9}$$

where $\boldsymbol{w}_{fl}(n) \triangleq [w_l(n), w_{l+1}(n), \ldots, w_{l+p-1}(n)]^T$, $\boldsymbol{w}_{bl}(n) \triangleq [w_{M-l+1}(n), w_{M-l}(n), \ldots, w_{L-l+1}(n)]^H$, $\boldsymbol{D} \triangleq \mathrm{diag}(e^{j\omega_0\tau(\theta_1)}, e^{j\omega_0\tau(\theta_2)}, \ldots, e^{j\omega_0\tau(\theta_p)})$, and $\boldsymbol{A}_1$ is the $p \times p$ submatrix of $\boldsymbol{A}$ consisting of the first p rows with the column $\boldsymbol{a}_1(\theta_k) \triangleq [1, e^{j\omega_0\tau(\theta_k)}, \ldots, e^{j\omega_0(p-1)\tau(\theta_k)}]^T$. By defining the correlation vectors as $\boldsymbol{\varphi}_{fl} \triangleq E\{\boldsymbol{y}_{fl}(n)y_M^*(n)\}$, $\bar{\boldsymbol{\varphi}}_{fl} \triangleq E\{\boldsymbol{y}_{fl}(n)y_1^*(n)\}$, $\boldsymbol{\varphi}_{bl} \triangleq E\{y_1(n)\boldsymbol{y}_{bl}(n)\}$, and $\bar{\boldsymbol{\varphi}}_{bl} \triangleq E\{y_M(n)\boldsymbol{y}_{bl}(n)\}$, after some manipulations, we obtain four Hankel correlation matrices [23]

$$\boldsymbol{\Phi}_f \triangleq [\boldsymbol{\varphi}_{f1}, \boldsymbol{\varphi}_{f2}, \ldots, \boldsymbol{\varphi}_{fL-1}]^T = E\{\boldsymbol{Y}_f(n)y_M^*(n)\}$$
$$= \rho_M\, r_s \bar{\boldsymbol{A}}\boldsymbol{B}\boldsymbol{A}_1^T \tag{10}$$
$$\bar{\boldsymbol{\Phi}}_f \triangleq [\bar{\boldsymbol{\varphi}}_{f2}, \bar{\boldsymbol{\varphi}}_{f3}, \ldots, \bar{\boldsymbol{\varphi}}_{fL}]^T = E\{\bar{\boldsymbol{Y}}_f(n)y_1^*(n)\}$$
$$= \rho_1\, r_s \bar{\boldsymbol{A}}\boldsymbol{B}\boldsymbol{D}\boldsymbol{A}_1^T \tag{11}$$
$$\boldsymbol{\Phi}_b \triangleq [\boldsymbol{\varphi}_{b1}, \boldsymbol{\varphi}_{b2}, \ldots, \boldsymbol{\varphi}_{bL-1}]^T = E\{\boldsymbol{Y}_b(n)y_1(n)\}$$
$$= \rho_1^*\, r_s \bar{\boldsymbol{A}}\boldsymbol{B}^*\boldsymbol{D}^{-(M-1)}\boldsymbol{A}_1^T \tag{12}$$
$$\bar{\boldsymbol{\Phi}}_b \triangleq [\bar{\boldsymbol{\varphi}}_{b2}, \bar{\boldsymbol{\varphi}}_{b3}, \ldots, \bar{\boldsymbol{\varphi}}_{bL}]^T = E\{\bar{\boldsymbol{Y}}_b(n)y_M(n)\}$$
$$= \rho_M^*\, r_s \bar{\boldsymbol{A}}\boldsymbol{B}^*\boldsymbol{D}^{-(M-2)}\boldsymbol{A}_1^T \tag{13}$$

where $\boldsymbol{Y}_f(n) \triangleq [\boldsymbol{y}_{f1}(n), \boldsymbol{y}_{f2}(n), \ldots, \boldsymbol{y}_{fL-1}(n)]^T$, $\bar{\boldsymbol{Y}}_f(n) \triangleq [\boldsymbol{y}_{f2}(n), \boldsymbol{y}_{f3}(n), \ldots, \boldsymbol{y}_{fL}(n)]^T$, $\boldsymbol{Y}_b(n) \triangleq [\boldsymbol{y}_{b1}(n), \boldsymbol{y}_{b2}(n), \ldots, \boldsymbol{y}_{bL-1}(n)]^T$, $\bar{\boldsymbol{Y}}_b(n) \triangleq [\boldsymbol{y}_{b2}(n), \boldsymbol{y}_{b3}(n), \ldots, \boldsymbol{y}_{bL}(n)]^T$, $\boldsymbol{B} \triangleq \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_p)$, $\rho_i \triangleq \boldsymbol{\beta}^H\boldsymbol{b}_i^*(\theta)$, $\boldsymbol{\beta} \triangleq [\beta_1, \beta_2, \ldots, \beta_p]^T$, $\boldsymbol{b}_i(\theta) \triangleq [e^{j\omega_0(i-1)\tau(\theta_1)}, e^{j\omega_0(i-1)\tau(\theta_2)}, \ldots, e^{j\omega_0(i-1)\tau(\theta_p)}]^T$, and $\bar{\boldsymbol{A}}$ is the $(M-p)\times p$ submatrix of $\boldsymbol{A}$ consisting of its first $M-p$ rows with the column $\bar{\boldsymbol{a}}(\theta_k) \triangleq [1, e^{j\omega_0\tau(\theta_k)}, \ldots, e^{j\omega_0(L-2)\tau(\theta_k)}]^T$.

Clearly, the correlation matrices $\boldsymbol{\Phi}_f$, $\bar{\boldsymbol{\Phi}}_f$, $\boldsymbol{\Phi}_b$, and $\bar{\boldsymbol{\Phi}}_b$ in (10)–(13) are not affected by the additive noise, $\boldsymbol{\Phi}_b = \boldsymbol{J}_{M-p}\bar{\boldsymbol{\Phi}}_f^*\boldsymbol{J}_p$, and $\bar{\boldsymbol{\Phi}}_b = \boldsymbol{J}_{M-p}\boldsymbol{\Phi}_f^*\boldsymbol{J}_p$, where $\boldsymbol{J}_{\mathrm{m}}$ is an $m \times m$ counteridentity matrix. These $(M-p)\times p$ Hankel matrices can be formed simply from the elements $\{r_{i1}\}$ and $\{r_{iM}\}$ in the Mth and first columns $\boldsymbol{\varphi}$ and $\bar{\boldsymbol{\varphi}}$ of array covaraince matrix $\boldsymbol{R}$ in (7) except for the autocorrelations r_{11} and r_{MM}, which contain the noise variance σ^2, where $\boldsymbol{\varphi} \triangleq E\{\boldsymbol{y}(n)y_M^*(n)\}$, and $\bar{\boldsymbol{\varphi}} \triangleq E\{\boldsymbol{y}(n)y_1^*(n)\}$. Furthermore, the ranks of these matrices equal p; that is, the dimension of their signal subspace equals the number of coherent signals.

Because it is assumed that $M > 2p$, we can partition the $(M-p)\times p$ matrix $\bar{\boldsymbol{A}}$, and hence the correlation matrices in (10)–(13), into two submatrices as

$$\bar{\boldsymbol{A}} \triangleq \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{bmatrix} \begin{matrix} \}p \\ \}M-2p \end{matrix} \tag{14}$$

$$\boldsymbol{\Phi}_f \triangleq \begin{bmatrix} \boldsymbol{\Phi}_{f1} \\ \boldsymbol{\Phi}_{f2} \end{bmatrix} \begin{matrix} \}p \\ \}M-2p \end{matrix}, \quad \bar{\boldsymbol{\Phi}}_f \triangleq \begin{bmatrix} \bar{\boldsymbol{\Phi}}_{f1} \\ \bar{\boldsymbol{\Phi}}_{f2} \end{bmatrix} \begin{matrix} \}p \\ \}M-2p \end{matrix}$$
$$\boldsymbol{\Phi}_b \triangleq \begin{bmatrix} \boldsymbol{\Phi}_{b1} \\ \boldsymbol{\Phi}_{b2} \end{bmatrix} \begin{matrix} \}p \\ \}M-2p \end{matrix}, \quad \bar{\boldsymbol{\Phi}}_b \triangleq \begin{bmatrix} \bar{\boldsymbol{\Phi}}_{b1} \\ \bar{\boldsymbol{\Phi}}_{b2} \end{bmatrix} \begin{matrix} \}p \\ \}M-2p \end{matrix}. \tag{15}$$

Under the assumptions for the data model, the rows of $\boldsymbol{A}_2$ can be expressed as a linear combination of the rows of $\boldsymbol{A}_1$ [20], [23], [57], [58]

$$\boldsymbol{P}^H\boldsymbol{A}_1 = \boldsymbol{A}_2 \tag{16}$$

where $\boldsymbol{P}$ denotes a $p \times (M-2p)$ linear operator. Hence, from (10)–(13) and (15), the relationship between $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ can be expressed as that between the submatrices of $\boldsymbol{\Phi}_f$, $\bar{\boldsymbol{\Phi}}_f$, $\boldsymbol{\Phi}_b$, and $\bar{\boldsymbol{\Phi}}_b$ as

$$\boldsymbol{P}^H\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2 \qquad \text{i.e.,} \qquad \boldsymbol{Q}^H\bar{\boldsymbol{A}} = \boldsymbol{O}_{(M-2p)\times p} \tag{17}$$

where $\boldsymbol{\Phi}_1 \triangleq [\boldsymbol{\Phi}_{f1}, \bar{\boldsymbol{\Phi}}_{f1}, \boldsymbol{\Phi}_{b1}, \bar{\boldsymbol{\Phi}}_{b1}]$, $\boldsymbol{\Phi}_2 \triangleq [\boldsymbol{\Phi}_{f2}, \bar{\boldsymbol{\Phi}}_{f2}, \boldsymbol{\Phi}_{b2}, \bar{\boldsymbol{\Phi}}_{b2}]$, $\boldsymbol{Q} \triangleq [\boldsymbol{P}^T, -\boldsymbol{I}_{M-2p}]^T$, and $\boldsymbol{P} = (\boldsymbol{\Phi}_1\boldsymbol{\Phi}_1^H)^{-1}\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2^H$. Clearly, the columns of matrix $\boldsymbol{Q}$ form a basis for the null space $\mathcal{N}(\bar{\boldsymbol{A}}^H)$ of $\bar{\boldsymbol{A}}^H$.

Thus, when the finite array data are available, the directions $\{\theta_k\}$ can be estimated based on (17) without any eigendecomposition by minimizing the cost function

$$f(\theta) = \bar{\boldsymbol{a}}^H(\theta)\boldsymbol{\Pi}_{\hat{Q}}\,\bar{\boldsymbol{a}}(\theta) \tag{18}$$

where $\bar{\boldsymbol{a}}(\theta) \triangleq [1, e^{j\omega_0\tau(\theta)}, \ldots, e^{j\omega_0(L-2)\tau(\theta)}]^T$, $\boldsymbol{\Pi}_{\hat{Q}} = \hat{\boldsymbol{Q}} \cdot (\hat{\boldsymbol{Q}}^H\hat{\boldsymbol{Q}})^{-1}\hat{\boldsymbol{Q}}^H$, $\hat{\boldsymbol{Q}} = [\hat{\boldsymbol{P}}^T, -\boldsymbol{I}_{M-2p}]^T$, $\hat{\boldsymbol{P}} = (\hat{\boldsymbol{\Phi}}_1\hat{\boldsymbol{\Phi}}_1^H)^{-1}\hat{\boldsymbol{\Phi}}_1 \cdot \hat{\boldsymbol{\Phi}}_2^H$, and $\hat{x}$ denotes an estimate of the quantity x.

III. Adaptive Bearing Estimation and Tracking Algorithm

The computational load of the batch implementation of SUMWE is dominated by the computation of estimated correlation vectors $\hat{\boldsymbol{\varphi}}$ and $\hat{\bar{\boldsymbol{\varphi}}}$, the estimation of orthogonal projector $\boldsymbol{\Pi}_{\hat{Q}}$ (i.e., linear operator $\hat{\boldsymbol{P}}$), and the minimization of cost function $f(\theta)$ [23]. Hence, these major steps should be carried out efficiently, when the SUMWE is implemented in a real-time manner. Now, we consider the SUMWE-based ABEST algorithm for estimating the constant directions and for tracking the slowly time-varying directions.

A. LMS Algorithm for Null Space Estimation

First, in the case of constant directions of the incident signals, we can write the Hankel correlation matrices $\boldsymbol{\Phi}_f$, $\bar{\boldsymbol{\Phi}}_f$, $\boldsymbol{\Phi}_b$, and $\bar{\boldsymbol{\Phi}}_b$ in (10)–(13) at the instant n by using the instantaneous correlations

$$\boldsymbol{\Phi}_f(n) = \boldsymbol{Y}_f(n)y_M^*(n), \qquad \bar{\boldsymbol{\Phi}}_f(n) = \bar{\boldsymbol{Y}}_f(n)y_1^*(n) \tag{19}$$

$$\boldsymbol{\Phi}_b(n) = \boldsymbol{Y}_b(n)y_1(n), \qquad \bar{\boldsymbol{\Phi}}_b(n) = \bar{\boldsymbol{Y}}_b(n)y_M(n). \tag{20}$$

By letting $\boldsymbol{\Phi}_{f1}(n)$, $\boldsymbol{\Phi}_{f2}(n)$, $\bar{\boldsymbol{\Phi}}_{f1}(n)$, $\bar{\boldsymbol{\Phi}}_{f2}(n)$, $\boldsymbol{\Phi}_{b1}(n)$, $\boldsymbol{\Phi}_{b2}(n)$, $\bar{\boldsymbol{\Phi}}_{b1}(n)$, and $\bar{\boldsymbol{\Phi}}_{b2}(n)$ be the corresponding submatrices of the Hankel matrices in (19) and (20), from (1), (8), (9), (14), and (16), we get the following relation between these instantaneous correlation matrices after some simple algebraic manipulations:

$$\boldsymbol{\Phi}_2(n) = \boldsymbol{P}^H\boldsymbol{\Phi}_1(n) + \boldsymbol{E}_o^H(n) \tag{21}$$

where $\boldsymbol{\Phi}_1(n) \triangleq [\boldsymbol{\Phi}_{f1}(n), \bar{\boldsymbol{\Phi}}_{f1}(n), \boldsymbol{\Phi}_{b1}(n), \bar{\boldsymbol{\Phi}}_{b1}(n)]$, $\boldsymbol{\Phi}_2(n) \triangleq [\boldsymbol{\Phi}_{f2}(n), \bar{\boldsymbol{\Phi}}_{f2}(n), \boldsymbol{\Phi}_{b2}(n), \bar{\boldsymbol{\Phi}}_{b2}(n)]$, $\boldsymbol{E}_o(n) \triangleq -\boldsymbol{G}^H(n)\boldsymbol{Q}$, $\boldsymbol{G}(n) \triangleq [\boldsymbol{W}_f(n)y_M^*(n), \bar{\boldsymbol{W}}_f(n)y_1^*(n), \boldsymbol{W}_b(n)y_1(n), \bar{\boldsymbol{W}}_b(n) \cdot y_M(n)]$, $\boldsymbol{W}_f(n) \triangleq [\boldsymbol{w}_{f1}(n), \boldsymbol{w}_{f2}(n), \ldots, \boldsymbol{w}_{fL-1}(n)]^T$, $\bar{\boldsymbol{W}}_f(n) \triangleq [\boldsymbol{w}_{f2}(n), \boldsymbol{w}_{f3}(n), \ldots, \boldsymbol{w}_{fL}(n)]^T$, $\boldsymbol{W}_b(n) \triangleq [\boldsymbol{w}_{b1}(n), \boldsymbol{w}_{b2}(n), \ldots, \boldsymbol{w}_{bL-1}(n)]^T$, and $\bar{\boldsymbol{W}}_b(n) \triangleq [\boldsymbol{w}_{b2}(n), \boldsymbol{w}_{b3}(n), \ldots, \boldsymbol{w}_{bL}(n)]^T$. Note that the linear operator $\boldsymbol{P}$ is independent of the correlation matrices $\boldsymbol{\Phi}_1(n)$ and $\boldsymbol{\Phi}_2(n)$, whereas the "additive noise" $\boldsymbol{E}_o(n)$ is correlated with $\boldsymbol{\Phi}_1(n)$ and $\boldsymbol{\Phi}_2(n)$ due to the presence of $y_1(n)$ and $y_M(n)$ in $\boldsymbol{G}(n)$. Thus, from (21), we can find that the estimation of linear operator $\boldsymbol{P}(n)$ at the instant n is reduced to the minimization of the time-varying cost function $J_{\mathrm{LMS}}(n)$ given by

$$J_{\mathrm{LMS}}(n) \triangleq \|\boldsymbol{E}(n)\|_F^2 = \mathrm{tr}\{\boldsymbol{E}^H(n)\boldsymbol{E}(n)\} \tag{22}$$

where the $4p \times (M-2p)$ matrix $\boldsymbol{E}(n)$ is the estimation error given by

$$\boldsymbol{E}(n) \triangleq \boldsymbol{\Phi}_2^H(n) - \boldsymbol{\Phi}_1^H(n)\boldsymbol{P}(n-1) \tag{23}$$

while $\|\cdot\|_F^2$ and $\mathrm{tr}\{\cdot\}$ denote the square of the Frobenius norm and the trace operation.

By defining the derivative of a function $J \triangleq f(z)$ with respect to a complex variable vector $\boldsymbol{z} \triangleq [z_1, z_2, \ldots, z_m]^T$ and the gradient vector of J as (e.g., [27] and [48])

$$\frac{\partial J}{\partial \boldsymbol{z}^*} \triangleq \frac{1}{2}\left[\frac{\partial J}{\partial \bar{x}_1} + j\frac{\partial J}{\partial \bar{y}_1}, \ldots, \frac{\partial J}{\partial \bar{x}_m} + j\frac{\partial J}{\partial \bar{y}_m}\right]^T$$

and $\nabla J \triangleq 2\partial J/\partial \boldsymbol{z}^*$, where $z_k \triangleq \bar{x}_k + j\bar{y}_k$, after some manipulations (cf. [48]), we can obtain the instantaneous gradient matrix of $J_{\mathrm{LMS}}(n)$ in (22) with respect to the linear operator $\boldsymbol{P}(n-1)$

$$\begin{aligned}\nabla J_{\mathrm{LMS}}(n) &= 2\frac{\partial J_{\mathrm{LMS}}(n)}{\partial \boldsymbol{P}^*(n-1)} \\ &= -2\boldsymbol{\Phi}_1(n)(\boldsymbol{\Phi}_2^H(n) - \boldsymbol{\Phi}_1^H(n)\boldsymbol{P}(n-1)).\end{aligned} \tag{24}$$

Thus, we can easily obtain the LMS algorithm for updating the linear operator $\boldsymbol{P}(n)$ [25]–[28]

$$\begin{aligned}\boldsymbol{P}(n) &= \boldsymbol{P}(n-1) - 0.5\mu\nabla J_{\mathrm{LMS}}(n) \\ &= \boldsymbol{P}(n-1) + \mu\boldsymbol{\Phi}_1(n)\boldsymbol{E}(n)\end{aligned} \tag{25}$$

where μ is a positive step-size, which should be chosen appropriately to guarantee the stability (see Section IV for details). Henceforth, we also call the linear operator as the weight and assume that the current weight matrix $\boldsymbol{P}(n-1)$ is statistically independent of the current correlation matrices $\boldsymbol{\Phi}_1(n)$ and $\boldsymbol{\Phi}_2(n)$ as usually assumed in the adaptive filtering literature (cf. [25]–[36]).

Furthermore, by using the matrix inversion lemma (e.g., [2], [27], and [68]), from (18), we can obtain the instantaneous orthogonal projector $\boldsymbol{\Pi}(n)$ onto the null space $\mathcal{N}(\bar{\boldsymbol{A}}^H(\theta(n)))$ [23]

$$\boldsymbol{\Pi}(n) = \boldsymbol{Q}(n)\big(\boldsymbol{I}_{M-2p} - \boldsymbol{P}^H(n)(\boldsymbol{P}(n)\boldsymbol{P}^H(n) + \boldsymbol{I}_p)^{-1}\boldsymbol{P}(n)\big)\boldsymbol{Q}^H(n) \tag{26}$$

where $\boldsymbol{Q}(n) = [\boldsymbol{P}^T(n), -\boldsymbol{I}_{M-2p}]^T$. Although the computational complexity is reduced for the calculation of $\boldsymbol{\Pi}(n)$, where a $p \times p$ matrix $\boldsymbol{P}(n)\boldsymbol{P}^H(n) + \boldsymbol{I}_p$ is inversed instead of an $(M-2p) \times (M-2p)$ matrix $\boldsymbol{Q}^H(n)\boldsymbol{Q}(n)$, this inversion may be unsuitable for the real-time application. Due to the computational expediency of the Householder transformation (cf. [13] and [48]), which is a reflection operation done to annihilate all the elements of a column except for the first one, we perform the

QR decomposition of $\boldsymbol{P}(n)\boldsymbol{P}^H(n)+\boldsymbol{I}_p$ using the Householder transformation (see Appendix A for details)

$$\boldsymbol{P}(n)\boldsymbol{P}^H(n)+\boldsymbol{I}_p \triangleq \bar{\boldsymbol{P}}(n)=\bar{\boldsymbol{Q}}(n)\bar{\boldsymbol{R}}(n) \tag{27}$$

where $\bar{\boldsymbol{Q}}(n)$ and $\bar{\boldsymbol{R}}(n)$ are the $p\times p$ unitary matrix and upper triangular matrix. Then, the instantaneous orthogonal projector $\boldsymbol{\Pi}(n)$ can be obtained

$$\boldsymbol{\Pi}(n)=\boldsymbol{Q}(n)\Big(\boldsymbol{I}_{M-2p}-\boldsymbol{P}^H(n)\,\mathrm{inv}\{\bar{\boldsymbol{R}}(n)\}\bar{\boldsymbol{Q}}^H(n)\boldsymbol{P}(n)\Big)\boldsymbol{Q}^H(n) \tag{28}$$

where $\mathrm{inv}\{\bar{\boldsymbol{R}}(n)\}$ denotes the inversion operation of $\bar{\boldsymbol{R}}(n)$ with a simple back-substitution, because $\bar{\boldsymbol{R}}(n)$ is an upper triangular matrix (see Appendix B for details).

B. NLMS Algorithm for Null Space Estimation in Nonstationary Environment

The LMS algorithm in (25) has good convergence in a stationary environment, and its stability and rate of adaptation are governed by the step-size μ, whose stability region generally depends on the statistics of the signals (e.g., incident directions herein) and on the additive noise. Because the step-size impacts the performance in a rather complicated way, it is very difficult in practice to choose an appropriate step-size to track the time-varying directions [49], though some variable step-size methods have been proposed (e.g., [50]). As normalization has been used as a heuristics method for stability in numerical analysis and optimization, the NLMS algorithm can provide improved performance while maintaining the simplicity and robustness of the LMS algorithm with a fixed step-size (cf. [27], [28], and [56]). Here, we consider the NLMS algorithm for estimating the linear operator $\boldsymbol{P}(n)$, when the directions $\{\theta_k(n)\}$ are slowly time-varying (relative to the sampling rate [21]).

The NLMS algorithm can be viewed as the solution to the following optimization problem [27]: minimizing the squared Frobenius norm of the change $\Delta\boldsymbol{P}(n) \triangleq \boldsymbol{P}(n)-\boldsymbol{P}(n-1)$ under the constraint $\boldsymbol{\Phi}_2(n)=\boldsymbol{P}^H(n)\boldsymbol{\Phi}_1(n)$. We can define the time-varying cost function $J_{\mathrm{NLMS}}(n)$ as

$$J_{\mathrm{NLMS}}(n) \triangleq \|\Delta\boldsymbol{P}(n)\|_F^2+\mathrm{Re}\{\mathrm{tr}\{(\boldsymbol{\Phi}_2(n)-\boldsymbol{P}^H(n)\boldsymbol{\Phi}_1(n))\boldsymbol{\Lambda}\}\} \tag{29}$$

where $\boldsymbol{\Lambda}$ is a $4p\times(M-2p)$ Lagrange matrix, and $\mathrm{Re}\{\cdot\}$ denotes the real part of the bracketed matrix. By differentiating $J_{\mathrm{NLMS}}(n)$ with respect to the weight $\boldsymbol{P}(n)$ and letting this differentiation be zero (i.e., $\partial J_{\mathrm{NLMS}}(n)/\partial\boldsymbol{P}^*(n)=\boldsymbol{O}_{p\times(M-2p)}$), we get the optimum weight

$$\boldsymbol{P}(n)=\boldsymbol{P}(n-1)+\boldsymbol{\Phi}_1(n)\boldsymbol{\Lambda}. \tag{30}$$

In view of the constraint $\boldsymbol{\Phi}_2(n)=\boldsymbol{P}^H(n)\boldsymbol{\Phi}_1(n)$, from (30) and (23), we easily obtain

$$\boldsymbol{\Phi}_1(n)\boldsymbol{\Lambda}=(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}\boldsymbol{\Phi}_1(n)\boldsymbol{E}(n) \tag{31}$$

when the $p\times p$ matrix $\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)$ is invertible. Then, by substituting (31) into (30) to eliminate the Lagrange matrix and introducing a positive step-size $\bar{\mu}$ to control the change in $\boldsymbol{P}(n)$ from one iteration to the next, we obtain the NLMS algorithm for updating of linear operator $\boldsymbol{P}(n)$

$$\boldsymbol{P}(n)=\boldsymbol{P}(n-1)+\bar{\mu}(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}\boldsymbol{\Phi}_1(n)\boldsymbol{E}(n) \tag{32}$$

where the update term is premultiplied (normalized) by the inversed time-varying covariance matrix $(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}$ of the "input" matrix $\boldsymbol{\Phi}_1(n)$, whereas the stability region of step-size $\bar{\mu}$ is independent of the signal statistics and given by $0<\bar{\mu}<2$ (see Section IV for details).

Furthermore, with the SVD of the $4p\times p$ matrix $\boldsymbol{\Phi}_1^H(n)$ given by [13], [23]

$$\begin{aligned}\boldsymbol{\Phi}_1^H(n)&=\boldsymbol{U}(n)\tilde{\boldsymbol{\Lambda}}(n)\boldsymbol{V}^H(n)=[\underbrace{\boldsymbol{U}_1(n)}_{p},\underbrace{\boldsymbol{U}_2(n)}_{3p}]\begin{bmatrix}\tilde{\boldsymbol{\Lambda}}_1(n)\\ \boldsymbol{O}_{3p\times p}\end{bmatrix}\boldsymbol{V}^H(n)\\&=\boldsymbol{U}_1(n)\tilde{\boldsymbol{\Lambda}}_1(n)\boldsymbol{V}^H(n)\end{aligned} \tag{33}$$

where $\boldsymbol{U}(n)\boldsymbol{U}^H(n)=\boldsymbol{U}^H(n)\boldsymbol{U}(n)=\boldsymbol{I}_{4p}$, $\boldsymbol{V}(n)\boldsymbol{V}^H(n)=\boldsymbol{V}^H(n)\boldsymbol{V}(n)=\boldsymbol{I}_p$, $\tilde{\boldsymbol{\Lambda}}_1=\mathrm{diag}(\tilde{\lambda}_1(n),\tilde{\lambda}_2(n),\ldots,\tilde{\lambda}_p(n))$ with $\tilde{\lambda}_1(n)\geq\tilde{\lambda}_2(n)\geq\cdots\geq\tilde{\lambda}_p(n)>0$, we readily find that the term $\boldsymbol{V}(n)\tilde{\boldsymbol{\Lambda}}_1(n)\boldsymbol{U}_1^H(n)$ (i.e., $\boldsymbol{\Phi}_1(n)$) of the LMS algorithm in (25) is replaced by $\boldsymbol{V}(n)\tilde{\boldsymbol{\Lambda}}_1^{-1}(n)\boldsymbol{U}_1^H(n)$ (i.e., $(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}\boldsymbol{\Phi}_1(n)$) of the NLMS algorithm in (32). Thus, we can view the NLMS algorithm as the LMS algorithm with a time-varying step-size diagonal matrix $\bar{\boldsymbol{M}}(n) \triangleq \bar{\mu}\tilde{\boldsymbol{\Lambda}}_1^{-2}(n)$, whereas the latter has a constant step-size matrix $\mu\boldsymbol{I}_p$.

The normalization makes the NLMS algorithm less sensitive to the variations in matrix $\boldsymbol{\Phi}_1(n)$ at the cost of increased computational complexity, where the $p\times p$ matrix inversion is necessitated. However, this difficulty can be alleviated by using the QR decomposition given by

$$\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)=\tilde{\boldsymbol{Q}}(n)\tilde{\boldsymbol{R}}(n) \tag{34}$$

where $\tilde{\boldsymbol{Q}}(n)$ and $\tilde{\boldsymbol{R}}(n)$ are the $p\times p$ unitary matrix and upper triangular matrix, respectively. Consequently, from (34) and (32), the NLMS algorithm can be rewritten as

$$\boldsymbol{P}(n)=\boldsymbol{P}(n-1)+\bar{\mu}\,\mathrm{inv}\{\tilde{\boldsymbol{R}}(n)\}\tilde{\boldsymbol{Q}}^H(n)\boldsymbol{\Phi}_1(n)\boldsymbol{E}(n). \tag{35}$$

Then, the instantaneous orthogonal projector $\boldsymbol{\Pi}(n)$ can be estimated using (27) and (28).

Remark 3: When the matrix $\boldsymbol{\Phi}_1(n)$ is small, the numerical instability may arise due to the small singular values $\{\tilde{\lambda}_i(n)\}$ [12]. Hence, we can modify (32) by introducing a sufficiently small and positive regularization parameter ε

$$\boldsymbol{P}(n)=\boldsymbol{P}(n-1)+\bar{\mu}(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)+\varepsilon\boldsymbol{I}_p)^{-1}\boldsymbol{\Phi}_1(n)\boldsymbol{E}(n). \tag{36}$$

Thus, we call (36) the ε-NLMS algorithm. □

C. Approximate Newton Method for Direction Finding

Now, we consider the online implementation of direction finding based on the minimization of the cost function $f(\theta)$

in (18). As discussed in [23], the first-order expression for the estimation error of direction θ_k can be obtained as

$$\hat{\theta}_k - \theta_k \approx -\frac{\mathrm{Re}\{\bar{\boldsymbol{d}}^H(\theta_k)\boldsymbol{\Pi}_{\hat{Q}}\,\bar{\boldsymbol{a}}(\theta_k)\}}{\bar{\boldsymbol{d}}^H(\theta_k)\boldsymbol{\Pi}_Q\,\bar{\boldsymbol{d}}(\theta_k)} \tag{37}$$

where $\bar{\boldsymbol{d}}(\theta) \triangleq d\bar{\boldsymbol{a}}(\theta)/d\theta = j\omega_0\,(d/c)\cos\theta[0, e^{j\omega_0\tau(\theta)}, 2e^{j2\omega_0\tau(\theta)}, \ldots, (L-2)e^{j\omega_0(L-2)\tau(\theta)}]^T$. Then, by substituting the orthogonal projector $\boldsymbol{\Pi}(n)$ obtained by the LMS or NLMS algorithm into (37) to replace the estimated and true orthogonal projectors $\boldsymbol{\Pi}_{\hat{Q}}$ and $\boldsymbol{\Pi}_Q$, we get the approximate Newton iteration formula for direction updating (e.g., [25]–[28], [21], [66], and [68])

$$\hat{\theta}_k(n) = \hat{\theta}_k(n-1) - \left.\frac{\mathrm{Re}\{\bar{\boldsymbol{d}}^H(\theta)\boldsymbol{\Pi}(n)\bar{\boldsymbol{a}}(\theta)\}}{\bar{\boldsymbol{d}}^H(\theta)\boldsymbol{\Pi}(n)\bar{\boldsymbol{d}}(\theta)}\right|_{\theta=\hat{\theta}_k(n-1)}. \tag{38}$$

D. Online Bearing Estimation and Tracking Algorithm

Based on the above analyses of null space estimation and direction finding, we can summarize the online ABEST algorithm as follows.

1) Calculate the instantaneous correlation vectors $\boldsymbol{\varphi}(n)$ between $\boldsymbol{y}(n)$ and $y_M^*(n)$ and $\bar{\boldsymbol{\varphi}}(n)$ between $\boldsymbol{y}(n)$ and $y_1^*(n)$ as

$$\boldsymbol{\varphi}(n) = \boldsymbol{y}(n)y_M^*(n), \qquad \bar{\boldsymbol{\varphi}}(n) = \boldsymbol{y}(n)y_1^*(n) \tag{39}$$

where $\boldsymbol{\varphi}(n) = [\hat{r}_{1M}(n), \hat{r}_{2M}(n), \ldots, \hat{r}_{MM}(n)]^T$, and $\bar{\boldsymbol{\varphi}}(n) = [\hat{r}_{11}(n), \hat{r}_{21}(n), \ldots, \hat{r}_{M1}(n)]^T$.
..... $12M$ flops

2) Form the Hankel correlation matrices $\boldsymbol{\Phi}_f(n)$, $\bar{\boldsymbol{\Phi}}_f(n)$, $\boldsymbol{\Phi}_b(n)$, and $\bar{\boldsymbol{\Phi}}_b(n)$ from $\boldsymbol{\varphi}(n)$ and $\bar{\boldsymbol{\varphi}}(n)$ as

$$\boldsymbol{\Phi}_f(n) = \mathrm{Hank}\{\boldsymbol{h}_c, \boldsymbol{h}_r\}, \quad \bar{\boldsymbol{\Phi}}_f(n) = \mathrm{Hank}\{\bar{\boldsymbol{h}}_c, \bar{\boldsymbol{h}}_r\} \tag{40}$$

$$\boldsymbol{\Phi}_b(n) = \boldsymbol{J}_{M-p}\bar{\boldsymbol{\Phi}}_f^*(n)\boldsymbol{J}_p, \quad \bar{\boldsymbol{\Phi}}_b(n) = \boldsymbol{J}_{M-p}\boldsymbol{\Phi}_f^*(n)\boldsymbol{J}_p \tag{41}$$

where $\boldsymbol{h}_c = [\hat{r}_{1M}(n), \hat{r}_{2M}(n), \ldots, \hat{r}_{M-p,M}(n)]^T$, $\boldsymbol{h}_r = [\hat{r}_{M-p,M}(n), \hat{r}_{M-p+1,M}(n), \ldots, \hat{r}_{M-1,M}(n)]^T$, $\bar{\boldsymbol{h}}_c = [\hat{r}_{21}(n), \hat{r}_{31}(n), \ldots, \hat{r}_{L1}(n)]^T$, $\bar{\boldsymbol{h}}_r = [\hat{r}_{L1}(n), \hat{r}_{L+1,1}(n), \ldots, \hat{r}_{M1}(n)]^T$, and $\mathrm{Hank}\{\cdot\}$ denotes the Hankel operation.

3) Calculate the instantaneous estimation error $\boldsymbol{E}(n)$ by using (23).
..... $32p^2(M-2p) + 8p(M-2p)$ flops

4) LMS: Update the linear operator $\boldsymbol{P}(n)$ by using (25).
..... $32p^2(M-2p) + 2p(M-2p) + 24p^2$ flops
NLMS: Perform the QR decomposition shown in (34), and update the linear operator $\boldsymbol{P}(n)$ by using (35).
..... $32p^2(M-2p) + 2p(M-2p) + 97p^3 + 13p^2 + 22p$ flops

5) Calculate the QR decomposition of $\bar{\boldsymbol{P}}(n)$ by using (27) with Appendix A.
..... $8p^2(M-2p) + 24p^3 + 4p^2 + 23p$ flops

6) Calculate the orthogonal projector $\boldsymbol{\Pi}(n)$ by using (28) with Appendix B.
..... $8M(M-2p)^2 + 8(M-2p)((M-p)^2 + p^2) + 8p^3 + p(p+1)(p+2)$ flops

7) Update the estimates $\{\hat{\theta}_k(n)\}$ by using the approximate Newton iteration shown in (38).
..... $16(M-p)^2 + 16(M-p)$ flops

The computation complexity of each step above is roughly indicated in terms of the number of flops, where a flop is defined as a floating-point addition or multiplication operation as adopted by MATLAB software. The NLMS algorithm needs approximately $97p^3 - 11p^2 + 22p$ more flops than the LMS algorithm. Furthermore, the LMS and NLMS algorithms are initialized by $\boldsymbol{P}(0) = \boldsymbol{O}_{p\times(M-2p)}$, and the first $K_0 = 2M$ snapshots of the received data are accumulated for the offline SUMWE [23] to provide the initial values of directions $\{\hat{\theta}_k(n)\}$ for the Newton method [21].

Remark 4: Like the SUMWE, the proposed ABEST algorithm can accommodate a more general noise model of the spatially correlated noise if we choose appropriate subarrays (i.e., instantaneous cross correlations of the array data) (see [23] for reference). □

Remark 5: In practice, the subarray size should be chosen appropriately because the information on the number of signals is unavailable in some applications. When the number of signals is presumed as $\bar{p}$, if there is an error in this presumption, the estimation performance of the ABEST will be affected. If $\bar{p} > p$ with $M - \bar{p} < p$ or $\bar{p} < p$, the ranks of the $(M-\bar{p}) \times \bar{p}$ matrices $\boldsymbol{\Phi}_f$, $\bar{\boldsymbol{\Phi}}_f$, $\boldsymbol{\Phi}_b$, and $\bar{\boldsymbol{\Phi}}_b$ in (10)–(13) will be smaller than p, and apparently, the dimensions of signal subspace of these matrices cannot be restored to the number of coherent signals [12], [43]. Consequently, the $\bar{p} \times (M-2\bar{p})$ linear operator $\boldsymbol{P}$ in (17) cannot be used to estimate the directions of coherent signals accurately, and it will severely degrade the estimation performance of the ABEST algorithm. If $\bar{p} > p$ and $\bar{p}$ satisfies the condition $M - \bar{p} > \bar{p} > p$, the $(M-\bar{p}) \times \bar{p}$ matrices $\boldsymbol{\Phi}_f$, $\bar{\boldsymbol{\Phi}}_f$, $\boldsymbol{\Phi}_b$, and $\bar{\boldsymbol{\Phi}}_b$ are rank deficient, and the ordinary LS estimate $\boldsymbol{P} = (\boldsymbol{\Phi}_1\boldsymbol{\Phi}_1^H)^{-1}\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2^H$ from (17) becomes numerically unstable [12], [43]. However, due to the finite number of snapshots, the absence of additive noise in the array data will alleviate the ill-conditioning in the estimation $\hat{\boldsymbol{P}} = (\hat{\boldsymbol{\Phi}}_1\hat{\boldsymbol{\Phi}}_1^H)^{-1}\hat{\boldsymbol{\Phi}}_1\hat{\boldsymbol{\Phi}}_2^H$ to a certain extent [12], where $\hat{\boldsymbol{P}}$ is used in (18) for direction estimation from the finite and noisy array data. Thus, the ABEST algorithm still holds in this case by choosing the step-size of the LMS/NLMS algorithm properly. From Remark 1, we can see that the maximum detectable number of incident signals is $p < M/2$. Thus, if the number of signals is unknown, we can choose a conservative value of the number of incident signals (i.e., subarray size) as $\bar{p} = \lfloor M/2 \rfloor - 1$, where the inequality condition $M - \bar{p} > \bar{p} > p$ is satisfied as well, where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x.

Furthermore, by estimating the $(M-\bar{p}) \times \bar{p}$ matrices $\hat{\boldsymbol{\Phi}}_f$, $\hat{\bar{\boldsymbol{\Phi}}}_f$, $\hat{\boldsymbol{\Phi}}_b$, and $\hat{\bar{\boldsymbol{\Phi}}}_b$ from the finite array data with $\bar{p} = \lfloor M/2 \rfloor - 1$ and by performing the QR decomposition of the matrix $\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\Phi}}^H$ as $\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\Phi}}^H = \breve{\boldsymbol{Q}}\breve{\boldsymbol{R}}$, where $\hat{\boldsymbol{\Phi}} \triangleq [\hat{\boldsymbol{\Phi}}_f, \hat{\bar{\boldsymbol{\Phi}}}_f, \hat{\boldsymbol{\Phi}}_b, \hat{\bar{\boldsymbol{\Phi}}}_b]$, the number of incident signals can be determined accurately by using the elements $\{\breve{r}_{ik}\}$ of $\breve{\boldsymbol{R}}$ [75] (yet it is beyond the scope of this paper). □

TABLE I
COMPARISON BETWEEN THE COMPUTATIONAL COMPLEXITY OF SIMPLE ONLINE IMPLEMENTATION OF BENCHMARK METHODS WITH EVD AT EACH INSTANT AND THAT OF THE ABEST ALGORITHM IN MATLAB FLOPS AT EACH INSTANT

Step	MUSIC	SS-based MUSIC	ABEST
Calculation of correlations	$10M^2+6M$	$10M^2+6M$ $+2(M-m+1)m^2$	$12M$
Estimation of subspace	$O(M^3)$	$O(m^3)$	$8M(M-2p)^2+8(M-p)^2(M-2p)$ $+10(M-2p)(10p^2+p)+33p^3$ $+31p^2+25p+\kappa$
Updating of direction	$16(M+1)(M-p)$	$16(m+1)(m-p)$	$16(M-p)^2+16(M-p)$

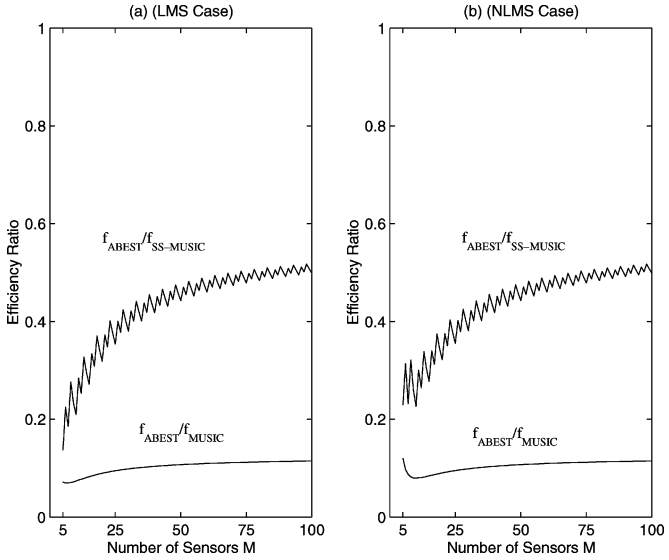


Fig. 1. Relative efficiency ratios between the estimated number of MATLAB flops required by the (a) LMS- and (b) NLMS-based ABEST algorithm and that needed by the MUSIC and SS-based MUSIC in terms of the number of sensors at each instant ($p = 2$).

Remark 6: Basically, a simple and direct online implementation of the benchmark methods of the MUSIC (for incoherent signals case) [6] and spatial smoothing (SS)-based MUSIC [39] (for the coherent signals case) with eigendecomposition involve three major steps at each instant:

1) computation of the time-varying array covariance matrix with rank-1 updating (and calculation of spatially smoothed subarray covariance matrix for the SS-based MUSIC);
2) estimation of noise subspace with EVD;
3) updating of the estimated directions with approximate Newton iteration.

The computational complexity of each step in MATLAB flops at each instant is roughly shown and compared with that of the ABEST algorithm in Table I, where m is the subarray size used in the SS-based MUSIC with $m > p$ and $M - m + 1 \geq p$ (see [39] for details), while $\kappa = 0$ or $\kappa = 97p^3 - 11p^2 + 22p$ when the LMS or NLMS is employed in the ABEST algorithm. By defining the total numbers of MATLAB flops required by the MUSIC, SS-based MUSIC, and the ABEST algorithm at each instant as f_{MUSIC}, $f_{\text{SS-MUSIC}}$, and f_{ABEST}, respectively, the relative efficiency ratios $f_{\mathrm{ABEST}}/f_{\mathrm{MUSIC}}$ and $f_{\mathrm{ABEST}}/f_{\text{SS-MUSIC}}$ in terms of the number of sensors are shown in Fig. 1 based on some examinations, where $p = 2$, M is varied from $M = 5$ to $M = 100$, the subarray size m is chosen as $m = \mathrm{round}(0.6(M+1))$ [74], and $\mathrm{round}(\cdot)$ denotes the round-off operation. Obviously, f_{ABEST} is smaller than f_{MUSIC} and $f_{\text{SS-MUSIC}}$, and these quantitative comparisons show that the ABEST algorithm is computationally efficient than the (SS-based) MUSIC method with EVD. □

IV. STATISTICAL ANALYSIS

A. Mean Behavior and Mean Stability of LMS

First by defining the weight-error $\tilde{\boldsymbol{P}}(n)$ as $\tilde{\boldsymbol{P}}(n) \triangleq \boldsymbol{P} - \boldsymbol{P}(n)$, from (21), (23), and (25), we can rewrite the weight adaptation in terms of the weight-error matrix

$$\tilde{\boldsymbol{P}}(n) = (\boldsymbol{I}_p - \mu\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))\tilde{\boldsymbol{P}}(n-1) + \mu\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}. \tag{42}$$

Under the assumption of independence between $\boldsymbol{P}(n-1)$ and $\boldsymbol{\Phi}_1(n)$, by taking the expectation on both sides of (42) and using the results shown in Appendix C, we easily get

$$\begin{aligned} E\{\tilde{\boldsymbol{P}}(n)\} &= (\boldsymbol{I}_p - \mu\bar{\boldsymbol{\Psi}}_1)E\{\tilde{\boldsymbol{P}}(n-1)\} + \mu\alpha\boldsymbol{P} \\ &= (\boldsymbol{I}_p - \mu\bar{\boldsymbol{\Psi}}_1)^n E\{\tilde{\boldsymbol{P}}(0)\} + \mu\alpha\sum_{k=0}^{n-1}(\boldsymbol{I}_p - \mu\bar{\boldsymbol{\Psi}}_1)^k\boldsymbol{P} \end{aligned} \tag{43}$$

where

$$\begin{aligned} \bar{\boldsymbol{\Psi}}_1 &\triangleq E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\} \\ &= \boldsymbol{\Psi}_1 + \sum_{l=1}^{p}\{r_{MM}(\boldsymbol{M}_{fll} + \boldsymbol{M}_{bl+1,l+1}) \\ &\quad + r_{11}(\boldsymbol{M}_{fl+1,l+1} + \boldsymbol{M}_{bll})\} \end{aligned} \tag{44}$$

$$\boldsymbol{\Omega} \triangleq E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\} = \alpha[\boldsymbol{I}_p, \boldsymbol{O}_{p\times(M-2p)}] \tag{45}$$

with $\boldsymbol{\Psi}_1 \triangleq \boldsymbol{\Phi}_1\boldsymbol{\Phi}_1^H$, $\boldsymbol{M}_{fik} \triangleq E\{\boldsymbol{y}_{fi}(n)\boldsymbol{y}_{fk}^H(n)\}$, $\boldsymbol{M}_{bik} \triangleq E\{\boldsymbol{y}_{bi}(n)\boldsymbol{y}_{bk}^H(n)\}$, and $\alpha \triangleq 2p\sigma^2(r_{11} + r_{MM})$. The recursion in (43) describes the mean behavior of the weight-error matrix $\tilde{\boldsymbol{P}}(n)$, and its first term is the natural component that governs the convergence of $\boldsymbol{P}(n)$ in (25).

Theorem 1: The stability condition of the LMS algorithm in the mean sense is that the step-size satisfies the double inequality

$$0 < \mu < \frac{2}{\lambda_{\max}(\bar{\boldsymbol{\Psi}}_1)} \tag{46}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the bracketed matrix.

Proof: From (43), the LMS algorithm converges in the mean sense if and only if the magnitudes of all eigenvalues of the matrix $\boldsymbol{I}_p - \mu\bar{\boldsymbol{\Psi}}_1$ are strictly less than one (e.g., [25]–[28]), i.e., $|1-\mu\lambda_i| < 1$ for all i, where $\{\lambda_i\}$ are the eigenvalues of $p \times p$ matrix $\bar{\boldsymbol{\Psi}}_1$, and $\lambda_i > 0$ because $\bar{\boldsymbol{\Psi}}_1$ in (44) has full rank [23]. Equivalently, the convergence condition in the mean sense is given by

$$0 < \mu < \frac{2}{\lambda_i}, \qquad \text{for} \quad i = 1, 2, \ldots, p. \tag{47}$$

Thus, the condition in (46) is established readily. ■

In addition, a sufficient condition in the mean sense is usually given by $0 < \mu < 2/\mathrm{tr}(\bar{\boldsymbol{\Psi}}_1)$, where $\mathrm{tr}\{\bar{\boldsymbol{\Psi}}_1\} = \sum_{i=1}^{p}\lambda_i$ [2]. Furthermore, from (43), the steady-state mean of the weight-error is given by $\lim_{n\to\infty} E\{\tilde{\boldsymbol{P}}(n)\} = \alpha\bar{\boldsymbol{\Psi}}_1^{-1}\boldsymbol{P}$. In view of the initialization of ABEST that $\boldsymbol{P}(0) = \boldsymbol{O}_{p\times(M-2p)}$ (i.e., $\tilde{\boldsymbol{P}}(0) = \boldsymbol{P}$), we readily get $\lim_{n\to\infty} E\{\tilde{\boldsymbol{P}}(n)\} = \boldsymbol{O}_{p\times(M-2p)}$ when μ is sufficiently small and $\sigma^2 = 0$, so the estimated linear operator $\boldsymbol{P}(n)$ is asymptotically unbiased.

Remark 7: We could get a modified LMS (MLMS) algorithm by setting $\mu = \tilde{\mu}/\mathrm{tr}\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}$

$$\boldsymbol{P}(n) = \boldsymbol{P}(n-1) + \frac{\tilde{\mu}}{\mathrm{tr}\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}}\boldsymbol{\Phi}_1(n)\boldsymbol{E}(n) \tag{48}$$

where $0 < \tilde{\mu} < 2$. We can see that (48) is the LMS algorithm with a time-varying step-size $\tilde{\mu}/\sum_{i=1}^{p}\tilde{\lambda}_i^2(n)$. □

B. Mean-Square Analysis of LMS

Now, we examine the mean-square behavior of the LMS algorithm in (25). Because the true weight $\boldsymbol{P}$ and the estimated one $\boldsymbol{P}(n-1)$ are independent of $\boldsymbol{\Phi}_1(n)$, the weight-error $\tilde{\boldsymbol{P}}(n-1)$ is also independent of $\boldsymbol{\Phi}_1(n)$. By postmultiplying (42) by its Hermitian transpose and by taking expectation on both sides, we get [51]

$$\begin{aligned}
&E\{\tilde{\boldsymbol{P}}(n)\tilde{\boldsymbol{P}}^H(n)\}\\
&= E\{\tilde{\boldsymbol{P}}(n-1)\tilde{\boldsymbol{P}}^H(n-1)\}\\
&\quad + \mu\Big(E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\}\boldsymbol{Q}E\{\tilde{\boldsymbol{P}}^H(n-1)\}\\
&\qquad + E\{\tilde{\boldsymbol{P}}(n-1)\}\boldsymbol{Q}^H(E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\})^H\\
&\qquad - E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}E\{\tilde{\boldsymbol{P}}(n-1)\tilde{\boldsymbol{P}}^H(n-1)\}\\
&\qquad - E\{\tilde{\boldsymbol{P}}(n-1)\check{\boldsymbol{P}}^H(n-1)\}E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}\Big)\\
&\quad + \mu^2\Big(E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\tilde{\boldsymbol{P}}(n-1)\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}\\
&\qquad + E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\boldsymbol{Q}^H\boldsymbol{G}(n)\boldsymbol{\Phi}_1^H(n)\}\\
&\qquad - E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}\\
&\qquad - (E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\check{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\})^H\Big).
\end{aligned} \tag{49}$$

Because of the correlation between the matrices $\boldsymbol{\Phi}_1(n)$ and $\boldsymbol{E}_o(n)$ (i.e., $\boldsymbol{G}(n)$) and the structure of the matrices $\boldsymbol{\Phi}_1(n)$ and $\boldsymbol{E}_o(n)$, which involves the instantaneous correlations of array data (and additive noise), it is mathematically formidable to evaluate the expectation $E\{\tilde{\boldsymbol{P}}(n)\tilde{\boldsymbol{P}}^H(n)\}$ in (49). From the results in Appendices D and E and by letting $\boldsymbol{K}(n) \triangleq E\{\tilde{\boldsymbol{P}}(n)\tilde{\boldsymbol{P}}^H(n)\}$ for convenience, we obtain the recursion of the mean-squared weight-error matrix $E\{\tilde{\boldsymbol{P}}(n)\tilde{\boldsymbol{P}}^H(n)\}$

$$\begin{aligned}
\boldsymbol{K}(n) &= (\boldsymbol{I}_p - \mu\bar{\boldsymbol{\Psi}}_1)\boldsymbol{K}(n-1)(\boldsymbol{I}_p - \mu\bar{\boldsymbol{\Psi}}_1)\\
&\quad + \mu\alpha(\boldsymbol{P}E\{\tilde{\boldsymbol{P}}^H(n-1)\} + E\{\tilde{\boldsymbol{P}}(n-1)\}\boldsymbol{P}^H)\\
&\quad + \mu^2(\alpha^2\boldsymbol{P}\boldsymbol{P}^H - \alpha\boldsymbol{P}E\{\tilde{\boldsymbol{P}}^H(n-1)\}\bar{\boldsymbol{\Psi}}_1\\
&\qquad - \alpha\bar{\boldsymbol{\Psi}}_1E\{\tilde{\boldsymbol{P}}(n-1)\}\boldsymbol{P}^H + \boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3\\
&\qquad + \boldsymbol{K}_4 - (\boldsymbol{K}_5 + \boldsymbol{K}_6) - (\boldsymbol{K}_5 + \boldsymbol{K}_6)^H)
\end{aligned} \tag{50}$$

where $\boldsymbol{K}_1 \sim \boldsymbol{K}_6$ are given in Appendix E. Clearly, this recursion describes the transient behavior of the LMS algorithm with (43). Although the highly nonlinear relation between $\boldsymbol{K}(n)$ and $\boldsymbol{K}(n-1)$ in (50) makes the convergence condition in the mean-square sense invisible, we can obtain the following theorem on the mean-square stability.

Theorem 2: The step-size convergence condition that guarantees the mean-square stability of the LMS algorithm in (25) is given by (51), shown at the bottom of the page, where

$$\boldsymbol{C} \triangleq (\boldsymbol{I}_p \otimes \bar{\boldsymbol{\Psi}}_1) + (\bar{\boldsymbol{\Psi}}_1 \otimes \boldsymbol{I}_p), \quad \bar{\boldsymbol{C}} \triangleq (\bar{\boldsymbol{\Psi}}_1^T \otimes \bar{\boldsymbol{\Psi}}_1) + \bar{\boldsymbol{K}}_2 \tag{52}$$

$$\tilde{\boldsymbol{C}} \triangleq \bar{\boldsymbol{C}} + \bar{\boldsymbol{K}}_1, \quad \check{\boldsymbol{C}} \triangleq \bar{\boldsymbol{C}} - \bar{\boldsymbol{K}}_1 \tag{53}$$

$$\tilde{\boldsymbol{L}} \triangleq \begin{bmatrix} \frac{\boldsymbol{C}}{2}, & \frac{-\tilde{\boldsymbol{C}}}{2} \\ \boldsymbol{I}_{p^2}, & \boldsymbol{O}_{p^2\times p^2} \end{bmatrix}, \quad \check{\boldsymbol{L}} \triangleq \begin{bmatrix} \frac{\boldsymbol{C}}{2}, & \frac{-\check{\boldsymbol{C}}}{2} \\ \boldsymbol{I}_{p^2}, & \boldsymbol{O}_{p^2\times p^2} \end{bmatrix} \tag{54}$$

$$\bar{\boldsymbol{K}}_1 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\bar{\boldsymbol{F}}_{il,kt}^H \otimes \bar{\boldsymbol{F}}_{il,kt} \tag{55}$$

$$\bar{\boldsymbol{K}}_2 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\mathrm{vec}(\boldsymbol{F}_{il,kt})\mathrm{vec}^H(\boldsymbol{F}_{il,kt}) \tag{56}$$

in which $\bar{\boldsymbol{F}}_{il,kt} \triangleq E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{z}}_{kt}^T(n)\}$, $\boldsymbol{F}_{il,kt} \triangleq E\{\tilde{\boldsymbol{z}}_{il}(n) \cdot \tilde{\boldsymbol{z}}_{kt}^H(n)\}$, $\tilde{\boldsymbol{z}}_{1l}(n) \triangleq \boldsymbol{y}_{fl}(n)y_M^*(n)$, $\tilde{\boldsymbol{z}}_{2l}(n) \triangleq \boldsymbol{y}_{fl+1}(n)y_1^*(n)$, $\tilde{\boldsymbol{z}}_{3l}(n) \triangleq \boldsymbol{y}_{bl}(n)y_1(n)$, $\tilde{\boldsymbol{z}}_{4l}(n) \triangleq \boldsymbol{y}_{bl+1}(n)y_M(n)$, and $\otimes$ denotes the Kronecker product. Here, we assume that the real and positive eigenvalues of the $2p^2 \times 2p^2$ matrices in (54) (i.e., $\lambda(\tilde{\boldsymbol{L}}) \in \mathcal{R}^+$ and $\lambda(\check{\boldsymbol{L}}) \in \mathcal{R}^+$) exist; if they do not, the corresponding condition should be removed from (51).

Proof: See Appendix F. ■

Therefore, by combining the analyses of the exponential convergences in the mean and mean-square senses, where the former suffices for the mean weight-error to equal zero and the latter ensures a steady-state error with finite variance, from (46) and (51), we can obtain the step-size convergence condition

$$0 < \mu < \min\left\{\frac{1}{\lambda_{\max}(\boldsymbol{C}^{-1}\tilde{\boldsymbol{C}})}, \frac{1}{\lambda_{\max}(\boldsymbol{C}^{-1}\check{\boldsymbol{C}})}, \frac{1}{\max\{\lambda(\tilde{\boldsymbol{L}}) \in \mathcal{R}^+\}}, \frac{1}{\max\{\lambda(\check{\boldsymbol{L}}) \in \mathcal{R}^+\}}\right\} \tag{51}$$

that guarantees both the mean and mean-square stabilities of the LMS algorithm as in (57), shown at the bottom of the page.

Remark 8: The ordinary differential equation (ODE) method [67]–[69] is another approach for studying the performance of adaptive filtering algorithm without making any assumptions about the statistical independence of input data, where the discrete-time difference equation of stochastic algorithm is reduced to a continuous-time deterministic differential equation, and the convergence and stability properties of the resulting continuous-time system are often analyzed by using the well-studied Lyapunov stability theory [28], [67]–[69]. Unfortunately, like the averaging analysis [26], the ODE method is only applicable to study the asymptotic convergence behavior of adaptive filtering algorithm for a sufficiently small step-size (or a vanishingly small one) case (so that some approximations become possible) and does not give information on the transient behavior of the adaptive algorithm (e.g., [70]–[73]). In this paper, we do not expand on this stability issue, since one of our objectives is to derive the convergence condition of step-size that guarantees the mean and mean-square stabilities. Furthermore, albeit in a rather complex fashion, the derived exact expectation analyses hold without any approximations or assumptions, except for the assumption that the current weight matrix $\boldsymbol{P}(n-1)$ is statistically independent of the current correlation matrices $\boldsymbol{\Phi}_1(n)$ and $\boldsymbol{\Phi}_2(n)$, but this assumption is justified in the situation considered herein in light of (21) (cf. [27]). □

C. Learning Curves of LMS

Here, we study the MSE and MSD learning curves of the LMS algorithm, which provide the measures of the rate of convergence, the steady-state error, and the effect of step-size.

First, from (21) and (23), the $4p \times (M-2p)$ estimation error matrix $\boldsymbol{E}(n)$ is rewritten as

$$\boldsymbol{E}(n) = \boldsymbol{\Phi}_1^H(n)\tilde{\boldsymbol{P}}(n-1) - \boldsymbol{G}^H(n)\boldsymbol{Q}. \tag{58}$$

Then, by using the independence assumption for $\tilde{\boldsymbol{P}}(n-1)$ and $\boldsymbol{\Phi}_1(n)$, we get the MSE learning curve of the LMS algorithm [27], [28]

$$\begin{aligned} J_{\mathrm{LMS}}^{\mathrm{MSE}}(n) &\triangleq E\{J_{\mathrm{LMS}}(n)\} = \mathrm{tr}\{E\{\boldsymbol{E}^H(n)\boldsymbol{E}(n)\}\} \\ &= \mathrm{tr}\{E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}E\{\tilde{\boldsymbol{P}}(n-1)\tilde{\boldsymbol{P}}^H(n-1)\}\} \\ &\quad - \mathrm{tr}\{(E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\})^H E\{\tilde{\boldsymbol{P}}(n-1)\boldsymbol{Q}^H\} \\ &\quad - \mathrm{tr}\{E\{\tilde{\boldsymbol{P}}^H(n-1)\}E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\}\boldsymbol{Q}\} \\ &\quad + \mathrm{tr}\{\boldsymbol{Q}^H E\{\boldsymbol{G}(n)\boldsymbol{G}^H(n)\}\boldsymbol{Q}\}. \end{aligned} \tag{59}$$

In a fashion similar to the evaluation of expectation $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\}$ in Appendix C, from (E3) and (E4), we obtain from some manipulations

$$E\{\boldsymbol{G}(n)\boldsymbol{G}^H(n)\} = \sum_{l=1}^{p}\sum_{i=1}^{4} E\{\tilde{\boldsymbol{g}}_{il}(n)\tilde{\boldsymbol{g}}_{il}^H(n)\} = \alpha \boldsymbol{I}_{M-p}. \tag{60}$$

Then, by substituting (44), (45), and (60) into (59) and using the identity $\mathrm{tr}\{\boldsymbol{XY}\} = \mathrm{tr}\{\boldsymbol{YX}\}$ for two matrices with compatible dimensions, we get the MSE curve $J_{\mathrm{LMS}}^{\mathrm{MSE}}(n)$

$$\begin{aligned} J_{\mathrm{LMS}}^{\mathrm{MSE}}(n) &= \mathrm{tr}\{\bar{\boldsymbol{\Psi}}_1\boldsymbol{K}(n-1)\} - \alpha\mathrm{tr}\{\boldsymbol{P}^H E\{\tilde{\boldsymbol{P}}(n-1)\}\} \\ &\quad - \alpha\mathrm{tr}\{E\{\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{P}\}\} + \alpha\mathrm{tr}\{\boldsymbol{Q}^H\boldsymbol{Q}\} \end{aligned} \tag{61}$$

where the first- and second-order weight statistics $E\{\tilde{\boldsymbol{P}}(n-1)\}$ and $E\{\tilde{\boldsymbol{P}}(n-1)\tilde{\boldsymbol{P}}^H(n-1)\}$ of the weight-error $\tilde{\boldsymbol{P}}(n-1)$ can be determined numerically using the recursions (43) and (50). From (61) and the result of Section IV-A, the steady-state MSE (SSMSE) is given by

$$\begin{aligned} J_{\mathrm{LMS}}^{\mathrm{SSMSE}} &= \mathrm{tr}\{\bar{\boldsymbol{\Psi}}_1\boldsymbol{K}(\infty)\} - 2\alpha^2\mathrm{tr}\{\boldsymbol{P}^H\bar{\boldsymbol{\Psi}}_1^{-1}\boldsymbol{P}\} \\ &\quad + \alpha\mathrm{tr}\{\boldsymbol{P}^H\boldsymbol{P} + \boldsymbol{I}_{M-2p}\}. \end{aligned} \tag{62}$$

Furthermore, from the recursion of the mean-squared weight-error matrix in (50), we easily get the MSD learning curve $J_{\mathrm{LMS}}^{\mathrm{MSD}}(n)$ of the LMS algorithm [27], [28]

$$J_{\mathrm{LMS}}^{\mathrm{MSD}}(n) = E\{\mathrm{tr}\{\tilde{\boldsymbol{P}}^H(n)\tilde{\boldsymbol{P}}(n)\}\} = \mathrm{tr}\{\boldsymbol{K}(n)\}. \tag{63}$$

D. Stability Analysis of NLMS

Because of the high nonlinearity of the NLMS algorithm in (32) with the presence of normalization factor $(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}$ and of the correlation between the matrices $\boldsymbol{\Phi}_1(n)$ and $\boldsymbol{E}_o(n)$ (i.e., $\boldsymbol{G}(n)$) with their structure, the mathematical analysis of the NLMS algorithm is very complicated. Here, we thus only examine the convergence condition for the NLMS algorithm in the case of constant directions.

From (32), (21), and (23), after some simple manipulations, we easily obtain

$$\tilde{\boldsymbol{P}}(n) = (1-\bar{\mu})\tilde{\boldsymbol{P}}(n-1) + \bar{\mu}(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}. \tag{64}$$

Then, the expectation of the NLMS weight-error matrix is given by

$$\begin{aligned} E\{\tilde{\boldsymbol{P}}(n)\} &= (1-\bar{\mu})E\{\tilde{\boldsymbol{P}}(n-1)\} + \bar{\mu}\bar{\boldsymbol{\Omega}}\boldsymbol{Q} \\ &= (1-\bar{\mu})^n E\{\tilde{\boldsymbol{P}}(0)\} + \bar{\mu}\sum_{k=0}^{n-1}(1-\bar{\mu})^k\bar{\boldsymbol{\Omega}}\boldsymbol{Q} \\ &= (1-\bar{\mu})^n E\{\tilde{\boldsymbol{P}}(0)\} + (1-(1-\bar{\mu})^n)\bar{\boldsymbol{\Omega}}\boldsymbol{Q} \end{aligned} \tag{65}$$

where $\bar{\boldsymbol{\Omega}} = E\{(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\}$. Furthermore, in a way similar to (49) in the LMS case, from (64), we can get the recursion of mean-square weight-error $E\{\tilde{\boldsymbol{P}}(n)\tilde{\boldsymbol{P}}^H(n)\}$ of the NLMS algorithm

$$\begin{aligned} \boldsymbol{K}(n) &= (1-\bar{\mu})^2\boldsymbol{K}(n-1) + \bar{\mu}(1-\bar{\mu}) \\ &\cdot\left(\bar{\boldsymbol{\Omega}}\boldsymbol{Q}E\{\tilde{\boldsymbol{P}}^H(n-1)\} + E\{\tilde{\boldsymbol{P}}(n-1)\}\boldsymbol{Q}^H\bar{\boldsymbol{\Omega}}^H\right) + \bar{\mu}^2\boldsymbol{\Upsilon} \end{aligned} \tag{66}$$

$$0 < \mu < \min\left\{\frac{2}{\lambda_{\max}(\bar{\boldsymbol{\Psi}}_1)}, \frac{1}{\lambda_{\max}(\boldsymbol{C}^{-1}\tilde{\boldsymbol{C}})}, \frac{1}{\lambda_{\max}(\boldsymbol{C}^{-1}\breve{\boldsymbol{C}})}, \frac{1}{\max\{\lambda(\tilde{\boldsymbol{L}}) \in \mathcal{R}^+\}}, \frac{1}{\max\{\lambda(\breve{\boldsymbol{L}}) \in \mathcal{R}^+\}}\right\}. \tag{57}$$

TABLE II
RESULTS OF EXPECTATION $\bar{\mathbf{F}}_{il,kt}$ FOR $i,k = 1,2,3,4$, AND $l,t = 1,2,\ldots,p$

	$k=1$	$k=2$	$k=3$	$k=4$
$i=1$	$2\boldsymbol{\varphi}_{fl}\boldsymbol{\varphi}_{ft}^T$	$\boldsymbol{\varphi}_{fl}\bar{\boldsymbol{\varphi}}_{ft+1}^T+\bar{\boldsymbol{\varphi}}_{fl}\boldsymbol{\varphi}_{ft+1}^T$	$r_{1M}\boldsymbol{M}_{lt}+\boldsymbol{\varphi}_{fl}\boldsymbol{\varphi}_{bt}^T$	$r_{MM}\boldsymbol{M}_{l,t+1}+\boldsymbol{\varphi}_{fl}\bar{\boldsymbol{\varphi}}_{bt+1}^T$
$i=2$	$\bar{\boldsymbol{\varphi}}_{fl+1}\boldsymbol{\varphi}_{ft}^T+\boldsymbol{\varphi}_{fl+1}\bar{\boldsymbol{\varphi}}_{ft}^T$	$2\bar{\boldsymbol{\varphi}}_{fl+1}\bar{\boldsymbol{\varphi}}_{ft+1}^T$	$r_{11}\boldsymbol{M}_{l+1,t}+\bar{\boldsymbol{\varphi}}_{fl+1}\boldsymbol{\varphi}_{bt}^T$	$r_{M1}\boldsymbol{M}_{l+1,t+1}+\bar{\boldsymbol{\varphi}}_{fl+1}\bar{\boldsymbol{\varphi}}_{bt+1}^T$
$i=3$	$r_{1M}\boldsymbol{M}_{tl}^T+\boldsymbol{\varphi}_{bl}\boldsymbol{\varphi}_{ft}^T$	$r_{11}\boldsymbol{M}_{t+1,l}^T+\boldsymbol{\varphi}_{bl}\bar{\boldsymbol{\varphi}}_{ft+1}^T$	$2\boldsymbol{\varphi}_{bl}\boldsymbol{\varphi}_{bt}^T$	$\boldsymbol{\varphi}_{bl}\bar{\boldsymbol{\varphi}}_{bt+1}^T+\bar{\boldsymbol{\varphi}}_{bl}\boldsymbol{\varphi}_{bt+1}^T$
$i=4$	$r_{MM}\boldsymbol{M}_{t,l+1}^T+\bar{\boldsymbol{\varphi}}_{bl+1}\boldsymbol{\varphi}_{ft}^T$	$r_{M1}\boldsymbol{M}_{t+1,l+1}^T+\bar{\boldsymbol{\varphi}}_{bl+1}\bar{\boldsymbol{\varphi}}_{ft+1}^T$	$\bar{\boldsymbol{\varphi}}_{bl+1}\boldsymbol{\varphi}_{bt}^T+\boldsymbol{\varphi}_{bl+1}\bar{\boldsymbol{\varphi}}_{bt}^T$	$2\bar{\boldsymbol{\varphi}}_{bl+1}\bar{\boldsymbol{\varphi}}_{bt+1}^T$

TABLE III
RESULTS OF EXPECTATION $\mathbf{F}_{il,kt}$ FOR $i,k = 1,2,3,4$, AND $l,t = 1,2,\ldots,p$

	$k=1$	$k=2$	$k=3$	$k=4$
$i=1$	$r_{MM}\boldsymbol{M}_{flt}+\boldsymbol{\varphi}_{fl}\boldsymbol{\varphi}_{ft}^H$	$r_{1M}\boldsymbol{M}_{fl,t+1}+\boldsymbol{\varphi}_{fl}\bar{\boldsymbol{\varphi}}_{ft+1}^H$	$\boldsymbol{\varphi}_{fl}\boldsymbol{\varphi}_{bt}^H+\bar{\boldsymbol{\varphi}}_{fl}\bar{\boldsymbol{\varphi}}_{bt}^H$	$2\boldsymbol{\varphi}_{fl}\bar{\boldsymbol{\varphi}}_{bt+1}^H$
$i=2$	$r_{M1}\boldsymbol{M}_{fl+1,t}+\bar{\boldsymbol{\varphi}}_{fl+1}\boldsymbol{\varphi}_{ft}^H$	$r_{11}\boldsymbol{M}_{fl+1,t+1}+\bar{\boldsymbol{\varphi}}_{fl+1}\bar{\boldsymbol{\varphi}}_{ft+1}^H$	$2\bar{\boldsymbol{\varphi}}_{fl+1}\boldsymbol{\varphi}_{bt}^H$	$\bar{\boldsymbol{\varphi}}_{fl+1}\bar{\boldsymbol{\varphi}}_{bt+1}^H+\boldsymbol{\varphi}_{fl+1}\boldsymbol{\varphi}_{bt+1}^H$
$i=3$	$\boldsymbol{\varphi}_{bl}\boldsymbol{\varphi}_{ft}^H+\bar{\boldsymbol{\varphi}}_{bl}\bar{\boldsymbol{\varphi}}_{ft}^H$	$2\boldsymbol{\varphi}_{bl}\bar{\boldsymbol{\varphi}}_{ft+1}^H$	$r_{11}\boldsymbol{M}_{blt}+\boldsymbol{\varphi}_{bl}\boldsymbol{\varphi}_{bt}^H$	$r_{1M}\boldsymbol{M}_{bl,t+1}+\boldsymbol{\varphi}_{bl}\bar{\boldsymbol{\varphi}}_{bt+1}^H$
$i=4$	$2\bar{\boldsymbol{\varphi}}_{bl+1}\boldsymbol{\varphi}_{ft}^H$	$\bar{\boldsymbol{\varphi}}_{bl+1}\bar{\boldsymbol{\varphi}}_{ft+1}^H+\boldsymbol{\varphi}_{bl+1}\boldsymbol{\varphi}_{ft+1}^H$	$r_{M1}\boldsymbol{M}_{bl+1,t}+\bar{\boldsymbol{\varphi}}_{bl+1}\boldsymbol{\varphi}_{bt}^H$	$r_{MM}\boldsymbol{M}_{bl+1,t+1}+\bar{\boldsymbol{\varphi}}_{bl+1}\bar{\boldsymbol{\varphi}}_{bt+1}^H$

where $\boldsymbol{\Upsilon} \triangleq E\{(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\boldsymbol{Q}^H\boldsymbol{G}(n) \cdot \boldsymbol{\Phi}_1^H(n)(\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n))^{-1}\}$. Although it is a thorny task to evaluate the expectations $\bar{\boldsymbol{\Omega}}$ in (65) and $\boldsymbol{\Upsilon}$ in (66), we readily find that the stability region of step-size $\bar{\mu}$ is independent of the signal statistics.

Theorem 3: The convergence of the NLMS algorithm in the mean and mean-square senses is guaranteed for the step-size $\bar{\mu}$ satisfying the condition $0 < \bar{\mu} < 2$.

Proof: From (65) and (66), apparently the NLMS algorithm converges if and only if the step-size $\bar{\mu}$ satisfies the inequality $|1-\bar{\mu}| < 1$. Thus, the stability condition $0 < \bar{\mu} < 2$ is obtained straightforwardly, where the knowledge of signal statistics is not needed. ■

From (65), we also get the steady-state mean of the weight-error as $\lim_{n\to\infty} E\{\tilde{\boldsymbol{P}}(n)\} = \bar{\boldsymbol{\Omega}}\boldsymbol{Q}$, and the NLMS algorithm in (32) is asymptotically unbiased for the noiseless array data.

Remark 9: Under the noiseless scenario (i.e., $\sigma^2 = 0$), we can see that $\bar{\mu} = 1$ is the optimum step-size for the fastest convergence of the NLMS algorithm. □

E. An Analytical Study of LMS Stability

To gain insight into the convergence condition of the LMS algorithm discussed in Sections IV-A and IV-B, here we focus on the case of one signal with constant direction and study the LMS stability in more details.

In this case (i.e., $p = 1$ and $L = M$), we readily have

$$\boldsymbol{A} = \boldsymbol{a}(\theta_1) = [1, e^{j\omega_0\tau(\theta_1)}, \ldots, e^{j\omega_0(M-1)\tau(\theta_1)}]^T \quad (67)$$

$$r_{ik} = r_s e^{j\omega_0(i-k)\tau(\theta_1)} + \sigma^2\delta_{i,k} \quad (68)$$

$$\boldsymbol{A}_2 = \boldsymbol{a}_2(\theta_1) = [e^{j\omega_0\tau(\theta_1)}, \ldots, e^{j\omega_0(M-2)\tau(\theta_1)}]^T \quad (69)$$

and $\boldsymbol{A}_1 = a_1(\theta_1) = 1$, so the linear operator becomes a $1\times(M-2)$ vector quantity $\boldsymbol{p}$ given by $\boldsymbol{p} = \boldsymbol{a}_2^H(\theta_1)$, and the LMS algorithm can be rewritten as

$$\boldsymbol{p}(n) = \boldsymbol{p}(n-1) + \mu\boldsymbol{\varphi}_1(n)\boldsymbol{E}(n) \quad (70)$$

where

$$\begin{aligned}\boldsymbol{\Phi}_1(n) = \boldsymbol{\varphi}_1(n) &= [y_1(n)y_M^*(n), y_2(n)y_1^*(n), y_1(n)y_M^*(n), \\ &\quad y_M(n)y_{M-1}^*(n)] \\ &\triangleq [\tilde{z}_{11}(n), \tilde{z}_{21}(n), \tilde{z}_{31}(n), \tilde{z}_{41}(n)] \quad (71)\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Phi}_2(n) &= [\bar{\boldsymbol{y}}_{f1}(n)y_M^*(n), \bar{\boldsymbol{y}}_{f2}(n)y_1^*(n), y_1(n)\bar{\boldsymbol{y}}_{b1}(n), \\ &\quad y_M(n)\bar{\boldsymbol{y}}_{b2}(n)] \quad (72)\end{aligned}$$

with $\bar{\boldsymbol{y}}_{f1}(n) \triangleq [y_2(n), y_3(n), \ldots, y_{M-1}(n)]^T$, $\bar{\boldsymbol{y}}_{f2}(n) \triangleq [y_3(n), y_4(n), \ldots, y_M(n)]^T$, $\bar{\boldsymbol{y}}_{b1}(n) \triangleq [y_{M-1}(n), y_{M-2}(n), \ldots, y_2(n)]^T$, and $\bar{\boldsymbol{y}}_{b2}(n) \triangleq [y_{M-2}(n), y_{M-3}(n), \ldots, y_1(n)]^T$, while $\boldsymbol{E}(n) \triangleq \boldsymbol{\Phi}_2^H(n) - \boldsymbol{\varphi}_1^H(n)\boldsymbol{p}(n-1)$. Furthermore, the correlation vectors $\boldsymbol{\varphi}_{fl}$, $\bar{\boldsymbol{\varphi}}_{fl}$, $\boldsymbol{\varphi}_{bl}$, and $\bar{\boldsymbol{\varphi}}_{bl}$, and matrices $\boldsymbol{M}_{fik}$, $\boldsymbol{M}_{bik}$, and $\boldsymbol{M}_{ik}$ are reduced to scalar quantities given by

$$\varphi_{fl} = r_{lM},\ \varphi_{fl} = r_{l1},\ \varphi_{bl} = r_{1,M-l+1},\ \varphi_{bl} = r_{M,M-l+1} \quad (73)$$

$$M_{fik} = r_{ik},\ M_{bik} = r_{M-k+1,M-i+1},\ M_{ik} = r_{i,M-k+1}. \quad (74)$$

By some simple calculations, from (44), (C5), (73), and (74), we obtain

$$\bar{\Psi}_1 = E\{\boldsymbol{\varphi}_1(n)\boldsymbol{\varphi}_1^H(n)\} = 4r_s^2 + 4(r_s+\sigma^2)^2. \quad (75)$$

Moreover, by substituting (73) and (74) into the expectations shown in Tables II and III, we can get the expectations $\bar{F}_{i1,k1} \triangleq E\{\tilde{z}_{i1}(n)\tilde{z}_{k1}(n)\}$ and $F_{i1,k1} \triangleq E\{\tilde{z}_{i1}(n)\tilde{z}_{k1}^*(n)\}$ as

$$\bar{F}_{11,11} = \bar{F}_{11,31} = \bar{F}_{31,11} = \bar{F}_{31,31} = 2r_s^2 e^{j\omega_0 2(M-1)\tau(\theta_1)} \quad (76)$$

$$\begin{aligned}\bar{F}_{11,21} &= \bar{F}_{11,41} = \bar{F}_{31,21} = \bar{F}_{31,41} = \bar{F}_{21,11} = \bar{F}_{21,31} \\ &= \bar{F}_{41,11} = \bar{F}_{41,31} = r_s(2r_s+\sigma^2)e^{j\omega_0(2-M)\tau(\theta_1)} \quad (77)\end{aligned}$$

$$\bar{F}_{21,21} = \bar{F}_{21,41} = \bar{F}_{41,21} = \bar{F}_{41,41} = 2r_s^2 e^{j\omega_0 2\tau(\theta_1)} \quad (78)$$

$$\begin{aligned}F_{11,11} &= F_{11,31} = F_{21,21} = F_{31,11} = F_{31,31} = F_{41,41} \\ &= r_s^2 + (r_s+\sigma^2)^2 \quad (79)\end{aligned}$$

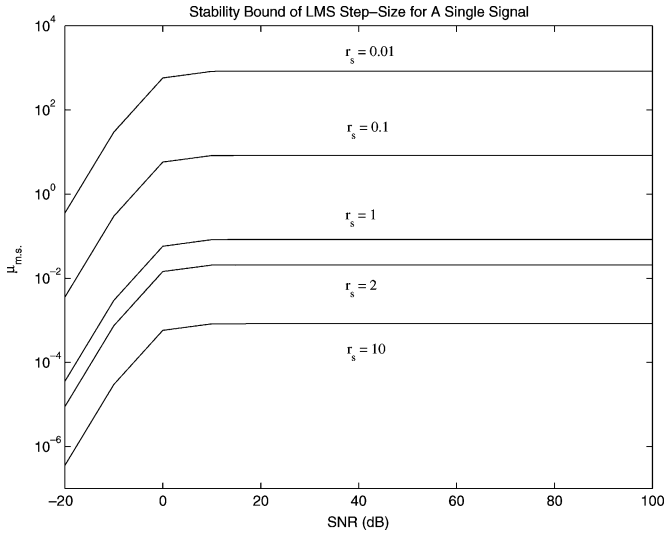


Fig. 2. Stability bounds of LMS step-size versus the SNR for several signal power in the case of one incident signal.

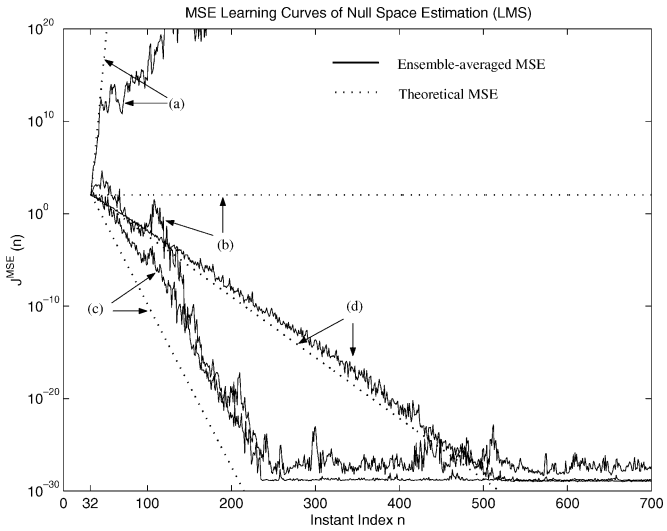


Fig. 3. MSE learning curves of LMS algorithm for null space estimation in the case of one signal without additive noise (solid line: ensemble-averaged MSE; dotted line: theoretical MSE for (a) $\mu = 0.25$, (b) $\mu = 1/12$, (c) $\mu = 1/24$, and (d) $\mu = 0.01$) for Example 1 ($M = 16$, $r_s = 1$, and $\theta_1 = 10^\circ$).

$$F_{11,21} = F_{11,41} = F_{31,21} = F_{31,41} = 2r_s^2 e^{-j\omega_0 M\tau(\theta_1)} \quad (80)$$

$$F_{21,11} = F_{21,31} = F_{41,11} = F_{41,31} = 2r_s^2 e^{j\omega_0 M\tau(\theta_1)} \quad (81)$$

$$F_{21,41} = F_{41,21} = 2r_s^2. \quad (82)$$

From (F3), (52), (55), (56), (F14), and (75)–(82) and some straightforward calculations, we obtain

$$\bar{K}_1 = \sum_{i=1}^{4}\sum_{k=1}^{4} |\bar{F}_{i1,k1}|^2 = 8r_s^2(4r_s^2 + (2r_s + \sigma^2)^2) \quad (83)$$

$$\bar{K}_2 = \sum_{i=1}^{4}\sum_{k=1}^{4} |F_{i1,k1}|^2 = 6(r_s^2 + (r_s + \sigma^2)^2)^2 + 40r_s^4 \quad (84)$$

$$C = 2\bar{\Psi}_1, \quad \tilde{C} = \bar{\Psi}_1^2 + \bar{K}_1 + \bar{K}_2. \quad (85)$$

Therefore, from (43) and (47), the convergence of LMS is governed by $F_{\text{mean}} \triangleq 1 - \mu\bar{\Psi}_1$ in the mean sense, and the step-size stability region is determined by $0 < \mu < \mu_{\text{mean}}$, where $\mu_{\text{mean}} \triangleq 2/\bar{\Psi}_1$. From (50), (F3), and (F12), the mean-square stability is controlled by $F_{\text{m.s.}} \triangleq 1 - \mu C + \mu^2\tilde{C}$, and the convergence condition is given by $0 < \mu < \mu_{\text{m.s.}}$, where $\mu_{\text{m.s.}} \triangleq C/\tilde{C}$, while $\mu_o = \mu_{\text{m.s.}}/2$ is the optimum step-size for the fastest convergence in the noiseless case. It follows from (85) that $C/\tilde{C} < 2/\bar{\Psi}_1$, so the step-size stability bound in both the mean and mean-square senses is obtained as $0 < \mu < \min\{\mu_{\text{mean}}, \mu_{\text{m.s.}}\} = \mu_{\text{m.s.}}$.

Furthermore, by setting SNR $\triangleq r_s/\sigma^2$, we get the stability bound $\mu_{\text{m.s.}}$ in terms of the SNR as shown in Fig. 2 for $r_s =$ 0.01, 0.1, 1, 2, and 10. From (75) and (83)–(85), we easily get $\mu_{\text{m.s.}} = 1/17.35r_s^2$ when SNR $= 0$ dB (i.e., $r_s = \sigma^2$). Obviously, for a given signal power r_s, $\mu_{\text{m.s.}}$ decreases monotonically with the decreasing SNR when SNR < 0 dB, while it quickly nears the supremum $1/12r_s^2$ when SNR > 0 dB; for a given SNR, $\mu_{\text{m.s.}}$ decreases monotonically with an increasing r_s^2.

V. Numerical Examples

We evaluate the performance of the proposed ABEST algorithm and demonstrate the validity of the analytical results through several numerical examples. The ULA with M sensors is separated by a half-wavelength, and the SNR is defined as the ratio of the power of the source signals to that of the additive noise at each sensor. In the simulations, the first $K_0 = 2M$ snapshots of the array data are accumulated for the offline SUMWE [23] to provide the initial values of estimated directions $\{\hat{\theta}_k(K_0)\}$ for the Newton iteration in (38), the LMS and NLMS algorithms in (25) and (35) are initialized by $\boldsymbol{P}(K_0) = \boldsymbol{O}_{p\times(M-2p)}$, and the online algorithm starts at the instant $n = K_0 + 1$. The simulation results shown subsequently are obtained by ensemble-averaging over 1000 independent trials.

A. Example 1—Verification of Stability Analysis

First, we examine the stability analysis of the LMS and NLMS algorithms for null space estimation. The number of sensors is $M = 16$, and one signal impinges the array along $\theta_1 = 10^\circ$ with the signal power $r_s = 1$. The additive noise is assumed to be absent. The step-size of the LMS algorithm is set to $\mu = 0.25$, 1/12, 1/24, and 1, while that of the NLMS algorithm is chosen as $\bar{\mu} = 2$, 1.5, 1, and 0.1.

From the analyses described in Section IV, the stability bounds of the LMS step-size in the mean and mean-square senses are $\mu_{\text{mean}} = 1/4r_s^2 = 0.25$ and $\mu_{\text{m.s.}} = 1/12r_s^2 = 1/12$, respectively, and the optimum step-size is $\mu_o = \mu_{\text{m.s.}}/2 = 1/24$. Figs. 3 and 4 show the ensemble-averaged MSE learning curves of the LMS and NLMS algorithms, respectively, which are computed by averaging the sample curves of $\bar{N}$ trials (here $\bar{N} = 1000$) (cf. [35]). From Fig. 3, we can see that there is an almost perfect agreement between the LMS theoretical MSE learning curve given by (61) and the simulation results for $\mu = 1/24$ and $\mu = 0.01$, which are smaller than the stability supremum $\mu_{\text{m.s.}}$. These simulation results demonstrate that the convergence rate depends highly on the step-size and that the fastest convergence is achieved with the optimum step-size $\mu = \mu_o = 1/24$. However, there are appreciable differences between the behaviors of the ensemble-averaged curves and those of the theoretical ones for

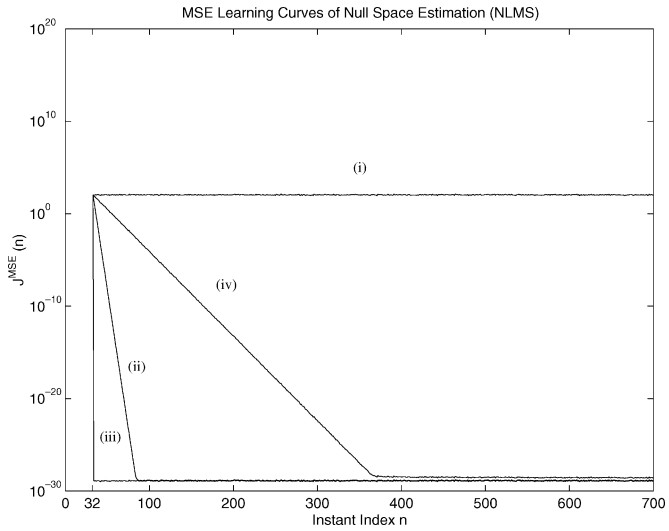


Fig. 4. Ensemble-averaged MSE learning curves of NLMS algorithm for null space estimation in the case of one signal without additive noise ((i) $\bar{\mu} = 2$, (ii) $\bar{\mu} = 1.5$, (iii) $\bar{\mu} = 1$, and (iv) $\bar{\mu} = 0.1$) for Example 1 ($M = 16$, $r_s = 1$, and $\theta_1 = 10^\circ$).

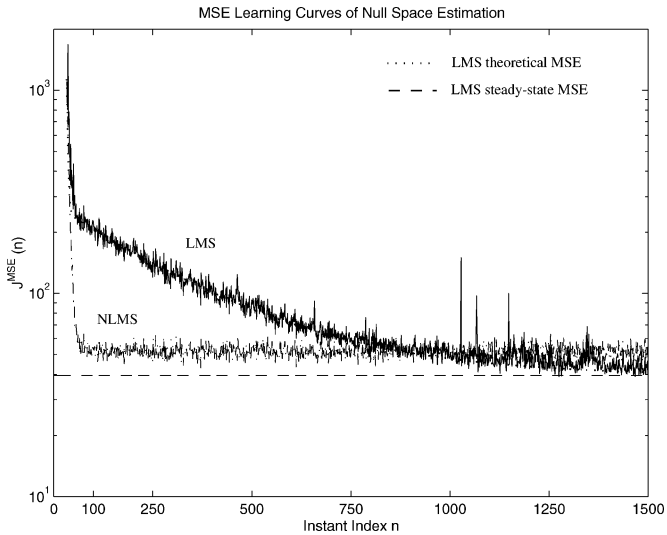


Fig. 5. MSE learning curves of null space estimation in the case of two coherent signals with constant directions (solid line: LMS ensemble-averaged MSE; dotted line: LMS theoretical MSE; dashed line: LMS steady-state MSE; and dashed–dotted line: NLMS ensemble-averaged MSE) for Example 2 ($M = 16$, SNR $= 20$ dB ($r_s = 1$), $\theta_1 = 10^\circ$, $\theta_2 = 20^\circ$, $\mu = 10^{-3}$, and $\bar{\mu} = 0.1$).

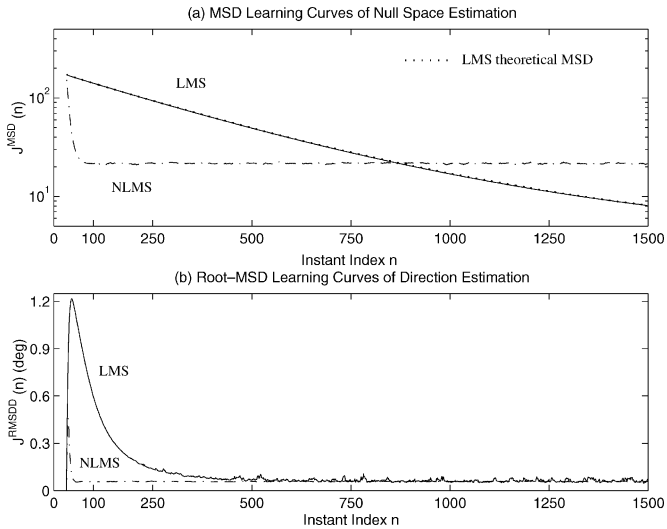


Fig. 6. (a) MSD learning curves of null space estimation and (b) root-MSD learning curves of direction estimation in the case of two coherent signals with constant directions (solid line: LMS ensemble-averaged MSD/root-MSD; dotted line: LMS theoretical MSD; and dashed–dotted line: NLMS ensemble-averaged MSD/root-MSD) for Example 2 ($M = 16$, SNR $= 20$ dB ($r_s = 1$), $\theta_1 = 10^\circ$, $\theta_2 = 20^\circ$, $\mu = 10^{-3}$, and $\bar{\mu} = 0.1$).

$\mu = \mu_{\text{mean}} = 0.25$ and $\mu = \mu_{\text{m.s.}} = 1/12$, which are out of the stability region $0 < \mu < \mu_{\text{m.s.}}$, where the averaged curve coincides well with the theoretical one during the initial time instants and deviates from it thereafter, and the convergence occurs even though the divergence is predicted by the mean-square stability analysis. These phenomena essentially conform with the learning mechanism clarified and studied in [35], where a combination of the mean-square stability and almost sure (a.s.) stability would be a more appropriate performance measure, especially for larger step-sizes (see [35] and [28] for more details).

In addition, Fig. 4 shows that the theoretical results for the NLMS stability better match the simulation results, in which the convergence of the NLMS ensemble-averaged MSE learning curves is guaranteed for the step-size $\bar{\mu}$ satisfying $0 < \bar{\mu} < 2$, and the learning curve converges fastest with $\bar{\mu} = 1$. Furthermore, the simulation results show that the convergence condition for the NLMS algorithm is independent of the statistics of the incident signal and that the NLMS algorithm converges faster than the LMS one. The behaviors of the MSD learning curves of the LMS and NLMS algorithms are similar to that of the MSE curves and are omitted here.

B. Example 2—Adaptive Estimation of Constant Directions

We then assess the estimation performance of the proposed ABEST algorithm when the directions of the incident signals are constant. Two coherent signals with equal power r_s come from $\theta_1 = 10^\circ$ and $\theta_2 = 20^\circ$, where $r_s = 1$, and the SNR is 20 dB. The number of sensors is $M = 16$, and the step-sizes of the LMS and NLMS algorithms are set to $\mu = 10^{-3}$ and $\bar{\mu} = 0.1$.

The bounds in (70) are evaluated and given by $2/\lambda_{\max}(\bar{\boldsymbol{\Psi}}_1) = 1.9871 \times 10^{-3}$, $2/\lambda_{\max}(\boldsymbol{C}^{-1}\tilde{\boldsymbol{C}}) = 1.9871 \times 10^{-3}$, and $1/\lambda_{\max}(\boldsymbol{C}^{-1}\breve{\boldsymbol{C}}) = 2.6998 \times 10^{-3}$, while the real and positive eigenvalues of $\tilde{\boldsymbol{L}}$ and $\breve{\boldsymbol{L}}$ do not exist. Hence, the stability bound on the step-size μ is $0 < \mu < \mu_{\max}$, where $\mu_{\max} = 1.9871 \times 10^{-3}$. The ensemble-averaged MSE learning curves of the LMS and NLMS algorithms are shown in Fig. 5, where the LMS theoretical MSE curve given by (61) and the SSMSE given by (62) are also depicted, in which the steady-state mean-squared weight-error matrix $\boldsymbol{K}(\infty)$ in (62) is evaluated by ensemble-averaging. As shown in Fig. 5, the theoretical results on the MSE curve and SSMSE of the LMS algorithm are in better agreement with the simulation results, while the ensemble-averaged MSE curve of the NLMS algorithm has faster convergence than that of the LMS algorithm, but it has a slightly larger steady-state MSE. Furthermore, the MSD learning curves of the LMS and NLMS algorithms are shown in Fig. 6(a). Obviously, the LMS ensemble-averaged

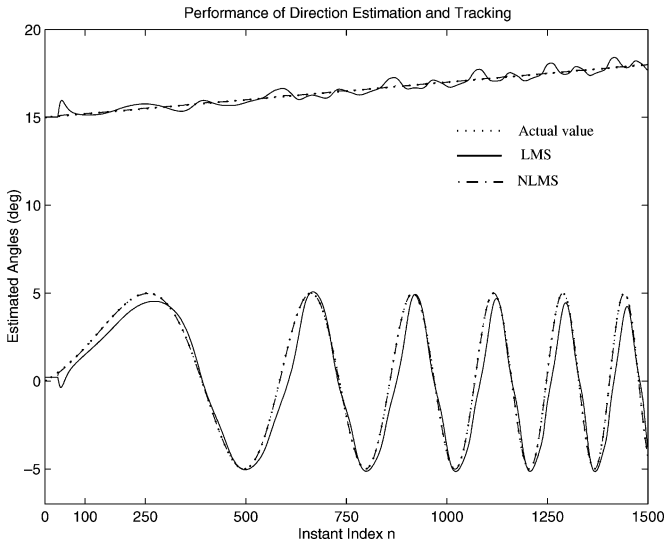


Fig. 7. Averaged direction estimates for tracking time-varying directions of two coherent signals (solid line: LMS; dashed–dotted line: NLMS; and dotted line: actual values) for Example 3 ($M = 16$, SNR $= 20$ dB ($r_s = 1$), $\theta_1(n) = 15^\circ + 0.002^\circ(n-1)$, $\theta_2(n) = 5^\circ \sin(2\pi(4 \times 10^{-4}n + 2.25 \times 10^{-6}n^2))$, $\mu = 2 \times 10^{-3}$, and $\bar{\mu} = 1$).

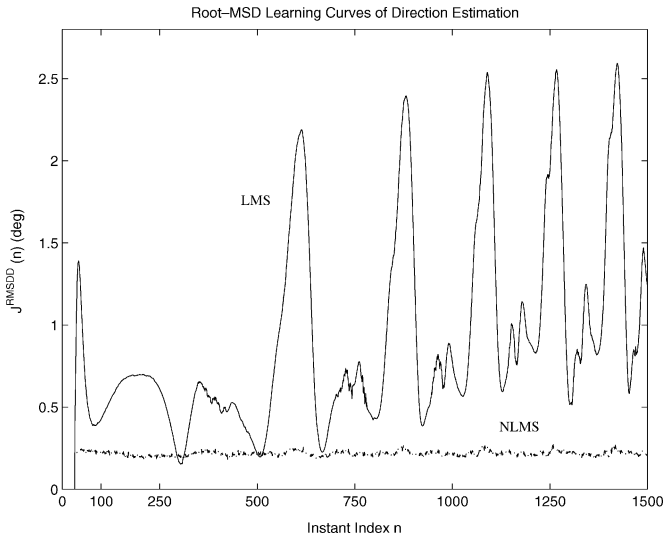


Fig. 8. Root-MSD learning curves of direction estimation for tracking time-varying directions of two coherent signals (solid line: LMS; and dashed–dotted line: NLMS) for Example 3 ($M = 16$, SNR $= 20$ dB ($r_s = 1$), $\theta_1(n) = 15^\circ + 0.002^\circ(n-1)$, $\theta_2(n) = 5^\circ \sin(2\pi(4 \times 10^{-4}n + 2.25 \times 10^{-6}n^2))$, $\mu = 2 \times 10^{-3}$, and $\bar{\mu} = 1$).

MSD curve is very close to the theoretical one, and the NLMS ensemble-averaged MSD curve converges faster than that of the LMS algorithm but with a larger steady-state MSD.

To measure the overall performance of estimating the directions, we define a root-MSD learning curve of estimated directions (RMSDD) as

$$J^{\mathrm{RMSDD}}(n) \triangleq \sqrt{\frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} \sum_{k=1}^{p} \left(\hat{\theta}_k^{(i)}(n) - \theta_k(n)\right)^2} \qquad (86)$$

where $\hat{\theta}_k^{(i)}(n)$ is the estimate obtained in the ith trial at the instant n, and $\bar{N}$ is the number of trials. The $J^{\mathrm{RMSDD}}(n)$ learning curves of the LMS- and NLMS-based ABEST algorithms are plotted in Fig. 6(b), where we can see that the NLMS-based algorithm provides better direction estimation than the LMS-based one with faster convergence and less fluctuation in this empirical scenario.

C. Example 3—Direction Tracking of Coherent Signals

Next, we evaluate the performance of the proposed ABEST algorithm for tracking time-varying directions of two coherent signals. The simulation conditions are similar to those of Example 2, except that the incident directions are linearly and nonlinearly time-varying, respectively, as $\theta_1(n) = 15^\circ + 0.002^\circ(n-1)$ and $\theta_2(n) = 5^\circ \sin(2\pi(4 \times 10^{-4}n + 2.25 \times 10^{-6}n^2))$, and the step-sizes are $\mu = 2 \times 10^{-3}$ and $\bar{\mu} = 1$.

The direction estimates $\hat{\theta}_1(n)$ and $\hat{\theta}_2(n)$ obtained by the LMS- and NLMS-based ABEST algorithm are shown in Fig. 7, and the root-MSD learning curves corresponding to the direction estimates are plotted in Fig. 8. While it is difficult to set an appropriate LMS step-size μ to track two time-varying directions with different dynamics accurately and promptly, the NLMS step-size $\bar{\mu}$ can be easily set, enabling the incident directions to be well tracked. The estimates obtained with the NLMS-based algorithm are almost indistinguishable from the actual values of the incident directions as shown in Fig. 7. Furthermore, careful examinations show that a small NLMS step-size $\bar{\mu}$ satisfying $0 < \bar{\mu} \leq 1$ should be used to achieve good tracking with a small steady-state MSE from the noisy array data in stationary and nonstationary environments, though the NLMS stability region is given by $0 < \bar{\mu} < 2$, and the fastest convergence occurs at $\bar{\mu} = 1$ in the noiseless case, as shown in Fig. 4.

D. Example 4—Direction Tracking of Signals With Time-Varying Correlation Factor

Here, we verify the performance of the ABEST algorithm when the incident signals have a time-varying correlation factor. Two signals arrive from $\theta_1 = 12^\circ$ (constant) and $\theta_2(n) = 5^\circ \sin(2\pi n/3000)$ (nonlinear and slowly time-varying), where $n = 1, 2, \ldots, 1500$, and the signal $s_2(n)$ is a superposition of two uncorrelated signals $s_1(n)$ and $s_{2o}(n)$ with equal power r_s given by [23]

$$s_2(n) = \rho^*(n)s_1(n) + \sqrt{1 - |\rho(n)|^2}\, s_{2o}(n) \qquad (87)$$

where $\rho(n)$ denotes the correlation factor. The other simulation parameters are similar to those in Example 2.

First, we consider the situation where two incident signals are uncorrelated (i.e., $\rho(n) = 0$) and then we test the direction estimation when the magnitude of the correlation factor $\rho(n)$ between 0 and 1 is

$$\rho(n) = \begin{cases} 0, & \text{for } 1 \leq n \leq 250 \\ \sin(2\pi \times 10^{-3}(n - 250)), & \text{for } 251 \leq n \leq 500 \\ 1, & \text{for } 501 \leq n \leq 1000 \\ \sin(2\pi \times 10^{-3}(n - 750)), & \text{for } 1001 \leq n \leq 1250 \\ 0, & \text{for } 1251 \leq n \leq 1500 \end{cases} \qquad (88)$$

where the phase of the correlation factor is assumed to be zero for simplicity. The direction estimates and the corresponding root-MSE learning curves are depicted in Figs. 9 and 10.

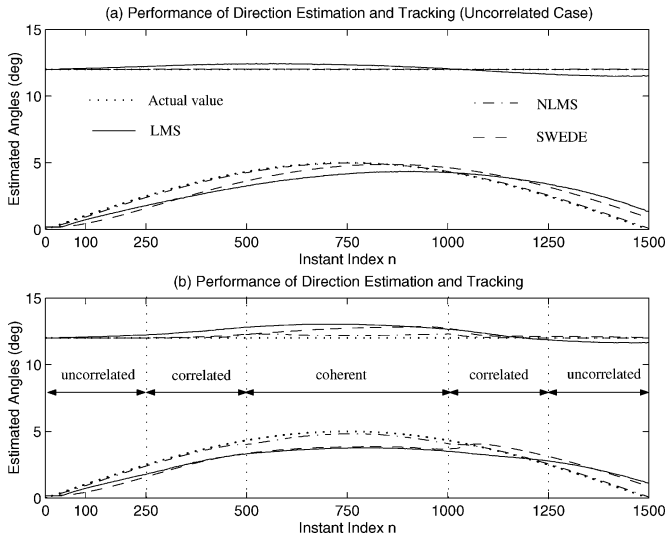


Fig. 9. Averaged direction estimates for tracking time-varying directions of (a) two uncorrelated signals and (b) two signals with time-varying correlation factor (solid line: LMS; dashed–dotted line: NLMS; dashed line: SWEDE; and dotted line: actual values) for Example 4 ($M = 16$, SNR $= 20$ dB ($r_s = 1$), $\theta_1 = 12^\circ$, $\theta_2(n) = 5^\circ \sin(2\pi n/3000)$, $\mu = 10^{-3}$, and $\bar{\mu} = 0.1$).

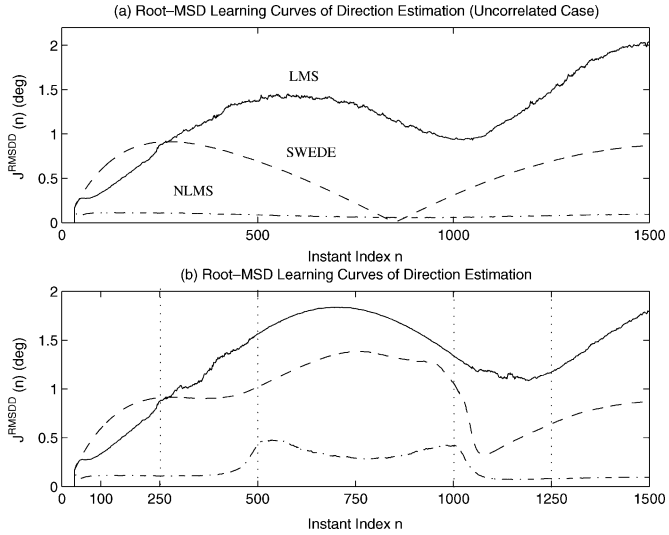


Fig. 10. Root-MSD learning curves of direction estimation for tracking time-varying directions of (a) two uncorrelated signals and (b) two signals with time-varying correlation factor (solid line: LMS; dashed–dotted line: NLMS; and dashed line: SWEDE) for Example 4 ($M = 16$, SNR $= 20$ dB ($r_s = 1$), $\theta_1 = 12^\circ$, $\theta_2(n) = 5^\circ \sin(2\pi n/3000)$, $\mu = 10^{-3}$, and $\bar{\mu} = 0.1$).

To compare the estimation performance, the online SWEDE without eigendecomposition [21] is also carried out, where the forgetting factor for rank-1 updating of the subarray covariance matrices is $\alpha = 0.01$, the null space is obtained in a QR–LS way, and the Newton step for direction updating is not reiterated at each time instant (see [21] for more details). In addition, the number of MATLAB flops required by the SWEDE is roughly $8(p+4)M^2+8p^2(M-2p)+24(M+p^3)+O(p^3)+O(M^2p)$, which is larger than that needed by the proposed ABEST algorithm. Clearly, the coherency of the incident signals leads to significantly degraded SWEDE performance during the time instant $501 \leq n \leq 1000$, and the different dynamics of the incident directions make it difficult to optimize the step-size μ

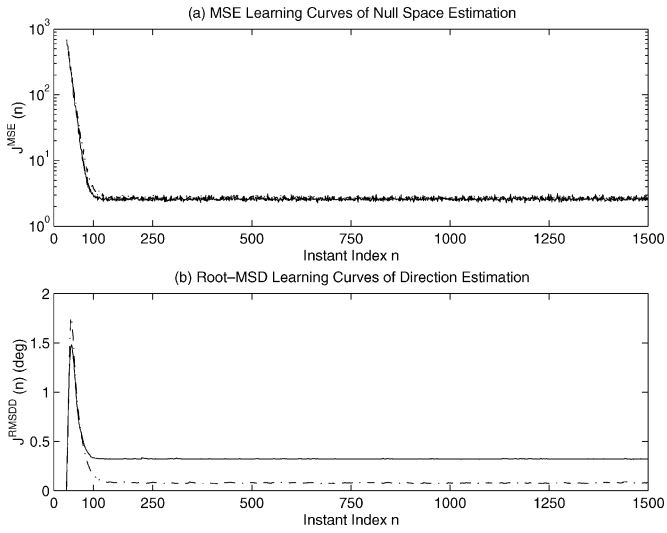


Fig. 11. (a) Ensemble-averaged MSE learning curves of null space estimation and (b) root-MSD learning curves of direction estimation for adaptive estimation of two constant directions (solid line: LMS; and dashed–dotted line: NLMS) for Example 5 ($M = 16$, SNR $= 20$ dB, $\theta_1 = 10^\circ$, $\theta_2 = 20^\circ$, $\bar{p} = 7$, $\mu = 10^{-4}$, and $\bar{\mu} = 0.05$).

for the LMS-based ABEST algorithm. However, as shown in Figs. 9 and 10, the tracking performance of the NLMS-based ABEST algorithm appears to be (significantly) less sensitive to the signal correlations and direction dynamics than that of the SWEDE and the LMS-based ABEST algorithm.

E. Example 5—Insensitivity to Presumed Number of Signals

Finally, we study the insensitivity of the proposed ABEST algorithm to an overdetermined number of signals in the scenarios similar to Examples 2 and 3. Here, the number of signals is presumed as $\bar{p} = \lfloor M/2 \rfloor - 1 = 7 > p$, and the parameter p in Steps $2 \sim 6$ shown in Section III-D is replaced with $\bar{p}$.

First, we consider the adaptive estimation of two constant directions, where the simulation conditions are similar to that of Example 2, except that the step-sizes of the LMS and NLMS algorithms are set to $\mu = 10^{-4}$ and $\bar{\mu} = 0.05$. The ensemble-averaged MSE learning curves $J^{\rm MSE}_{\rm LMS}$ and $J^{\rm MSE}_{\rm NLMS}$ of null space estimation are plotted in Fig. 11(a), while the root-MSD learning curves of direction estimation for the LMS and NLMS algorithms are shown in Fig. 11(b). Then, we test the tracking of time-varying directions of two coherent signals with the other simulation parameters being identical to those in Example 3, except that the step-sizes are $\mu = 1.2 \times 10^{-4}$ and $\bar{\mu} = 0.6$. The averaged direction estimates and the corresponding root-MSD learning curves for the LMS and NLMS algorithms are shown in Fig. 12(a) and (b), respectively. From Figs. 11 and 12, we can see that the ABEST algorithm with a presumed number of signals (i.e., $\bar{p} = \lfloor M/2 \rfloor - 1$) exhibits comparable estimation and tracking performance to that with the true number of signals in the stationary and nonstationary environment, as shown in Figs. 5, 6(b), 7, and Fig. 8 by choosing the step-size of the LMS/NLMS algorithm properly to dampen the noise effect. Furthermore, careful examinations indicate that the stability region of step-size becomes narrower than that shown in (57) and Theorem 3.

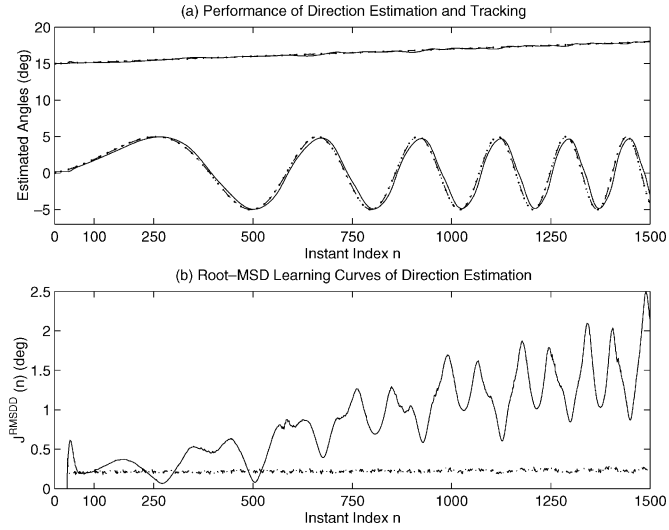


Fig. 12. (a) Averaged direction estimates and (b) root-MSD learning curves of direction estimation for tracking of two time-varying directions (solid line: LMS; dashed–dotted line: NLMS; and dotted line: actual value) for Example 5 ($M = 16$, SNR $= 20$ dB, $\theta_1(n) = 15^\circ + 0.002^\circ(n-1)$, $\theta_2(n) = 5^\circ \sin(2\pi(4 \times 10^{-4}n + 2.25 \times 10^{-6}n^2))$, $\bar{p} = 7$, $\mu = 1.2 \times 10^{-4}$, and $\bar{\mu} = 0.6$).

VI. Conclusion

In this paper, a computationally efficient subspace-based algorithm was developed for adaptive direction estimation and tracking of uncorrelated and correlated narrow-band signals impinging on a ULA. In this ABEST algorithm, the null space is estimated using the LMS or NLMS algorithm, and the directions are updated using the approximate Newton method. The ABEST algorithm has a reduced computational load, a less restrictive model of additive noise, and a remarkable insensitivity to the correlation of the incident signals. The transient analyses of the LMS and NLMS algorithms were studied, and the convergence conditions for the step-size that guarantee the mean and mean-square stabilities were explicitly derived. The analytical expressions of the MSE and MSD learning curves of the LMS algorithm were also clarified. The estimation effectiveness of the ABEST algorithm was verified, and the theoretical analyses were substantiated through numerical examples, and the simulation results showed that the ABEST algorithm is computationally simple and has good adaptation and tracking abilities.

Appendix A
Computation of QR Decomposition $\bar{\boldsymbol{P}} = \bar{\boldsymbol{Q}}\bar{\boldsymbol{R}}$ With Householder Transformation

We assume that $\bar{\boldsymbol{P}}$ is a $p \times q$ complex matrix and set $K = q$ for $p > q$; otherwise, $K = p - 1$. The algorithm for QR decomposition via Householder transformation is given as follows (e.g., [48]):

```
Initialization: Q̄ = I_p and R̄ = P̄
for k = 1: K
  v = R̄(k : p, k),  v_1 = v_1 + sign(v_1) ||v||
  ṽ = [0_{1×(k−1)}, v^T]^T,  γ = −2/v^H v
  Q̄ = Q̄ + γ Q̄ ṽ ṽ_k^H
  R̄(k : p, k : q) = R̄(k : p, k : q) + γ v v^H R̄(k : p, k : q)
end
```

This algorithm requires about $24p^3 + 4p^2 + 22p - 5$ MATLAB flops for $p = q$. ■

Appendix B
Back-Substitution for Inversion of Upper Triangular Square Matrix

For a $p \times p$ upper triangular complex matrix $\boldsymbol{U}$ with the ikth element as $u_{ik} = 0$ for $i < k$ while $u_{ik} \neq 0$ for $i \leq k$, after some simple manipulations, we find that the inverse matrix $\bar{\boldsymbol{U}}$ of $\boldsymbol{U}$ is also an upper triangular matrix with its ikth nonzero element $\bar{u}_{ik}$ given by

$$\bar{u}_{ik} = \begin{cases} \frac{1}{u_{ii}}, & \text{for } i = k \\ -\frac{1}{u_{ii}} \sum_{l=1}^{k-i} u_{i,l+i}\bar{u}_{l+i,k}, & \text{for } i < k \end{cases} \tag{B1}$$

for $i, k = p, p-1, \ldots, 1$. Furthermore, it takes nearly $p(p+1)(p+2)$ MATLAB flops. ■

Appendix C
Evaluation of Expectations $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}$ and $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\}$

First, the $p \times p$ Hankel matrices given by (19) and (20) can be rewritten as $\boldsymbol{\Phi}_{f1}(n) = [\boldsymbol{z}_{f1}(n), \boldsymbol{z}_{f2}(n), \ldots, \boldsymbol{z}_{fp}(n)]$, $\bar{\boldsymbol{\Phi}}_{f1}(n) = [\bar{\boldsymbol{z}}_{f1}(n), \bar{\boldsymbol{z}}_{f2}(n), \ldots, \bar{\boldsymbol{z}}_{fp}(n)]$, $\boldsymbol{\Phi}_{b1}(n) = [\boldsymbol{z}_{b1}(n), \boldsymbol{z}_{b2}(n), \ldots, \boldsymbol{z}_{bp}(n)]$, and $\bar{\boldsymbol{\Phi}}_{b1}(n) = [\bar{\boldsymbol{z}}_{b1}(n), \bar{\boldsymbol{z}}_{b2}(n), \ldots, \bar{\boldsymbol{z}}_{bp}(n)]$, where $\boldsymbol{z}_{fl}(n) \triangleq \boldsymbol{y}_{fl}(n)y_M^*(n) \triangleq \tilde{\boldsymbol{z}}_{1l}(n)$, $\bar{\boldsymbol{z}}_{fl}(n) \triangleq \boldsymbol{y}_{fl+1}(n)y_1^*(n) \triangleq \tilde{\boldsymbol{z}}_{2l}(n)$, $\boldsymbol{z}_{bl}(n) \triangleq \boldsymbol{y}_{bl}(n)y_1(n) \triangleq \tilde{\boldsymbol{z}}_{3l}(n)$, and $\bar{\boldsymbol{z}}_{bl}(n) \triangleq \boldsymbol{y}_{bl+1}(n)y_M(n) \triangleq \tilde{\boldsymbol{z}}_{4l}(n)$. Under the assumptions for the data model, we can obtain [51]

$$\begin{aligned} E\{\tilde{\boldsymbol{z}}_{1l}(n)\tilde{\boldsymbol{z}}_{1t}^H(n)\} &= E\{y_M^*(n)y_M(n)\boldsymbol{y}_{fl}(n)\boldsymbol{y}_{ft}^H(n)\} \\ &= E\{y_M^*(n)y_M(n)\}E\{\boldsymbol{y}_{fl}(n)\boldsymbol{y}_{ft}^H(n)\} \\ &\quad + E\{y_M^*(n)\boldsymbol{y}_{fl}(n)\}E\{y_M(n)\boldsymbol{y}_{ft}^H(n)\} \\ &\quad + E\{y_M(n)\boldsymbol{y}_{fl}(n)\}E\{y_M^*(n)\boldsymbol{y}_{ft}^H(n)\} \\ &= r_{MM}\boldsymbol{M}_{flt} + \boldsymbol{\varphi}_{fl}\boldsymbol{\varphi}_{ft}^H \end{aligned} \tag{C1}$$

for $l, t = 1, 2, \ldots, p$. Similarly, we get

$$E\{\tilde{\boldsymbol{z}}_{2l}(n)\tilde{\boldsymbol{z}}_{2t}^H(n)\} = r_{11}\boldsymbol{M}_{fl+1,t+1} + \bar{\boldsymbol{\varphi}}_{fl+1}\bar{\boldsymbol{\varphi}}_{ft+1}^H \tag{C2}$$

$$E\{\tilde{\boldsymbol{z}}_{3l}(n)\tilde{\boldsymbol{z}}_{3t}^H(n)\} = r_{11}\boldsymbol{M}_{blt} + \boldsymbol{\varphi}_{bl}\boldsymbol{\varphi}_{bt}^H \tag{C3}$$

$$E\{\tilde{\boldsymbol{z}}_{4l}(n)\tilde{\boldsymbol{z}}_{4t}^H(n)\} = r_{MM}\boldsymbol{M}_{bl+1,t+1} + \bar{\boldsymbol{\varphi}}_{bl+1}\bar{\boldsymbol{\varphi}}_{bt+1}^H. \tag{C4}$$

Thus, from (C1)–(C4), the expectation $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}$ is given by

$$\begin{aligned} &E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\} \\ &= \sum_{l=1}^{p}\sum_{i=1}^{4} E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{z}}_{il}^H(n)\} \\ &= \sum_{l=1}^{p}(\boldsymbol{\varphi}_{fl}\boldsymbol{\varphi}_{fl}^H + \bar{\boldsymbol{\varphi}}_{fl+1}\bar{\boldsymbol{\varphi}}_{fl+1}^H + \boldsymbol{\varphi}_{bl}\boldsymbol{\varphi}_{bl}^H + \bar{\boldsymbol{\varphi}}_{bl+1}\bar{\boldsymbol{\varphi}}_{bl+1}^H) \\ &\quad + \sum_{l=1}^{p}\{r_{MM}(\boldsymbol{M}_{fll} + \boldsymbol{M}_{bl+1,l+1}) \\ &\qquad + r_{11}(\boldsymbol{M}_{fl+1,l+1} + \boldsymbol{M}_{bll})\} \\ &= \bar{\boldsymbol{\Psi}}_1 + \sum_{l=1}^{p}\{r_{MM}(\boldsymbol{M}_{fll} + \boldsymbol{M}_{bl+1,l+1}) \\ &\qquad + r_{11}(\boldsymbol{M}_{fl+1,l+1} + \boldsymbol{M}_{bll})\}. \end{aligned} \tag{C5}$$

Next, we reexpress the $(M-p) \times p$ matrices $\boldsymbol{W}_f(n)$, $\bar{\boldsymbol{W}}_f(n)$, $\boldsymbol{W}_b(n)$, and $\bar{\boldsymbol{W}}_b(n)$ given by (21) as $\boldsymbol{W}_f(n) =$

$[\tilde{\boldsymbol{w}}_{f1}(n), \tilde{\boldsymbol{w}}_{f2}(n), \ldots, \tilde{\boldsymbol{w}}_{fp}(n)]$, $\bar{\boldsymbol{W}}_f(n) = [\tilde{\boldsymbol{w}}_{f2}(n), \tilde{\boldsymbol{w}}_{f3}(n), \ldots, \tilde{\boldsymbol{w}}_{fp+1}(n)]$, $\boldsymbol{W}_b(n) = [\tilde{\boldsymbol{w}}_{b1}(n), \tilde{\boldsymbol{w}}_{b2}(n), \ldots, \tilde{\boldsymbol{w}}_{bp}(n)]$, and $\bar{\boldsymbol{W}}_b(n) = [\tilde{\boldsymbol{w}}_{b2}(n), \tilde{\boldsymbol{w}}_{b3}(n), \ldots, \tilde{\boldsymbol{w}}_{bp+1}(n)]$, where $\tilde{\boldsymbol{w}}_{fl}(n) \triangleq [w_l(n), w_{l+1}(n), \ldots, w_{l+L-2}(n)]^T$, and $\tilde{\boldsymbol{w}}_{bl}(n) \triangleq [w_{M-l+1}(n), w_{M-l}(n), \ldots, w_{p+2-l}(n)]^H$. By letting $\tilde{\boldsymbol{g}}_{1l}(n) \triangleq \tilde{\boldsymbol{w}}_{fl}(n) y_M^*(n)$, $\tilde{\boldsymbol{g}}_{2l}(n) \triangleq \tilde{\boldsymbol{w}}_{fl+1}(n) y_1^*(n)$, $\tilde{\boldsymbol{g}}_{3l}(n) \triangleq \tilde{\boldsymbol{w}}_{bl}(n) \cdot y_1(n)$, and $\tilde{\boldsymbol{g}}_{4l}(n) \triangleq \tilde{\boldsymbol{w}}_{bl+1}(n) y_M(n)$, in a way similar to that for (C1), we can obtain

$$E\{\tilde{\boldsymbol{z}}_{1l}(n)\tilde{\boldsymbol{g}}_{1l}^H(n)\} = r_{MM} E\{\boldsymbol{w}_{fl}(n)\tilde{\boldsymbol{w}}_{fl}^H(n)\} \quad \text{(C6)}$$

$$E\{\tilde{\boldsymbol{z}}_{2l}(n)\tilde{\boldsymbol{g}}_{2l}^H(n)\} = r_{11} E\{\boldsymbol{w}_{fl+1}(n)\tilde{\boldsymbol{w}}_{fl+1}^H(n)\} \quad \text{(C7)}$$

$$E\{\tilde{\boldsymbol{z}}_{3l}(n)\tilde{\boldsymbol{g}}_{3l}^H(n)\} = r_{11} E\{\boldsymbol{w}_{bl}(n)\tilde{\boldsymbol{w}}_{bl}^H(n)\} \quad \text{(C8)}$$

$$E\{\tilde{\boldsymbol{z}}_{4l}(n)\tilde{\boldsymbol{g}}_{4l}^H(n)\} = r_{MM} E\{\boldsymbol{w}_{bl+1}(n)\tilde{\boldsymbol{w}}_{bl+1}^H(n)\} \quad \text{(C9)}$$

for $l = 1, 2, \ldots, p$. Furthermore, we easily get

$$E\{\boldsymbol{w}_{fk}(n)\tilde{\boldsymbol{w}}_{fk}^H(n)\} = E\{\boldsymbol{w}_{bk}(n)\tilde{\boldsymbol{w}}_{bk}^H(n)\} = \sigma^2[\boldsymbol{I}_p, \boldsymbol{O}_{p\times(M-2p)}] \quad \text{(C10)}$$

for $k = 1, 2, \ldots, p+1$. Hence, from (C6)–(C10), the expectation $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\}$ is given by

$$\begin{aligned} E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\} &= \sum_{l=1}^{p}\sum_{i=1}^{4} E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{g}}_{il}^H(n)\} \\ &= 2p\sigma^2(r_{11} + r_{MM})[\boldsymbol{I}_p, \boldsymbol{O}_{p\times(M-2p)}]. \end{aligned} \quad \text{(C11)}$$

■

Appendix D
Expectation Computation of Product of Six Complex Vector/Matrix-Valued Gaussian Random Variables $E\{\boldsymbol{x}\boldsymbol{z}^H\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\bar{\boldsymbol{z}}\bar{\boldsymbol{x}}^H\}$

Let $\boldsymbol{x}$, $\boldsymbol{z}$, $\bar{\boldsymbol{x}}$, and $\bar{\boldsymbol{z}}$ be four $M\times 1$ complex Gaussian random vectors while $\boldsymbol{Y}$ and $\bar{\boldsymbol{Y}}$ be two $M \times K$ complex Gaussian random matrices with zero-mean, and assume that these matrices are independent from these vectors. First, by expressing the matrix $\boldsymbol{Y}\bar{\boldsymbol{Y}}^H \triangleq \tilde{\boldsymbol{Y}}$ in terms of its column vectors $\{\tilde{\boldsymbol{y}}_{*k}\}$, the qrth element of the product $\tilde{\boldsymbol{Z}} \triangleq \boldsymbol{x}\boldsymbol{z}^H\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\bar{\boldsymbol{z}}\bar{\boldsymbol{x}}^H$ can be rewritten as

$$\begin{aligned} \tilde{z}_{qr} &= x_q \boldsymbol{z}^H\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\bar{\boldsymbol{z}}\bar{x}_r^* \\ &= x_q\left(\sum_{l=1}^{M} \boldsymbol{z}^H\tilde{\boldsymbol{y}}_{*l}\bar{z}_l\right)\bar{x}_r^* = \sum_{l=1}^{M}\sum_{k=1}^{M} x_q z_k^*\tilde{y}_{kl}\bar{z}_l\bar{x}_r^* \end{aligned} \quad \text{(D1)}$$

where $(\cdot)_{ik}$ and $(\cdot)_k$ denote the ikth and kth element of the bracketed matrix and vector. By invoking the independence assumption for the matrices and vectors, we get [51]

$$\begin{aligned} &E\{\tilde{z}_{qr}\} \\ &\quad= \sum_{l=1}^{M}\sum_{k=1}^{M} E\{\tilde{y}_{kl}\}(E\{x_q z_k^*\}E\{\bar{z}_l\bar{x}_r^*\} \\ &\qquad + E\{x_q\bar{z}_l\}E\{z_k^*\bar{x}_r^*\} + E\{x_q\bar{x}_r^*\}E\{z_k^*\bar{z}_l\}) \\ &\quad= \sum_{l=1}^{M}\sum_{k=1}^{M}\{(E\{\boldsymbol{x}\boldsymbol{z}^H\})_{qk}E\{\tilde{y}_{kl}\}(E\{\bar{\boldsymbol{z}}\bar{\boldsymbol{x}}^H\})_{lr} \\ &\qquad\qquad + (E\{\bar{\boldsymbol{x}}^*\boldsymbol{z}^H\})_{rk}E\{\tilde{y}_{kl}\}(E\{\bar{\boldsymbol{z}}\boldsymbol{x}^T\})_{lq}\} \\ &\qquad + (E\{\boldsymbol{x}\bar{\boldsymbol{x}}^H\})_{qr}\sum_{l=1}^{M}\sum_{k=1}^{M}(E\{\bar{\boldsymbol{z}}\boldsymbol{z}^H\})_{lk}E\{\tilde{y}_{kl}\} \\ &\quad= (E\{\boldsymbol{x}\boldsymbol{z}^H\}E\{\tilde{\boldsymbol{Y}}\}E\{\bar{\boldsymbol{z}}\bar{\boldsymbol{x}}^H\})_{qr} \\ &\qquad + (E\{\bar{\boldsymbol{x}}^*\boldsymbol{z}^H\}E\{\tilde{\boldsymbol{Y}}\}E\{\bar{\boldsymbol{z}}\boldsymbol{x}^T\})_{rq} \\ &\qquad + (E\{\boldsymbol{x}\bar{\boldsymbol{x}}^H\})_{qr}\mathrm{tr}\{E\{\bar{\boldsymbol{z}}\boldsymbol{z}^H\}E\{\tilde{\boldsymbol{Y}}\}\}. \end{aligned} \quad \text{(D2)}$$

Thus, we can obtain the expectation $E\{\boldsymbol{x}\boldsymbol{z}^H\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\bar{\boldsymbol{z}}\bar{\boldsymbol{x}}^H\}$ in matrix notation

$$\begin{aligned} E\{\boldsymbol{x}\boldsymbol{z}^H\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\bar{\boldsymbol{z}}\bar{\boldsymbol{x}}^H\} &= E\{\boldsymbol{x}\boldsymbol{z}^H\}E\{\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\}E\{\bar{\boldsymbol{z}}\bar{\boldsymbol{x}}^H\} \\ &\quad + E\{\boldsymbol{x}\bar{\boldsymbol{z}}^T\}(E\{\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\})^T(E\{\bar{\boldsymbol{x}}\boldsymbol{z}^T\})^H \\ &\quad + E\{\boldsymbol{x}\bar{\boldsymbol{x}}^H\}\mathrm{tr}\{E\{\bar{\boldsymbol{z}}\boldsymbol{z}^H\}E\{\boldsymbol{Y}\bar{\boldsymbol{Y}}^H\}\}. \end{aligned} \quad \text{(D3)}$$

Remark: The expectation formulas for the produce of the six Gaussian random vectors given in [29], [30], and [53] are special cases of the general expression (D3). ■

Appendix E
Evaluation of Expectations

$E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\tilde{\boldsymbol{P}}(n-1)\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}$, $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\boldsymbol{Q}^H\boldsymbol{G}(n)\boldsymbol{\Phi}_1^H(n)\}$, AND $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}$

The evaluation of these expectations is rather burdensome, so we omit some trivial details and only highlight the main steps and results.

First, we evaluate some correlations invoking the additive noise $\{w_i(n)\}$. Under the assumption for the additive noise given in (6), we can get the expectations $E\{\boldsymbol{w}_{fi}(n)\tilde{\boldsymbol{w}}_{fk}^H(n)\}$ and $E\{\boldsymbol{w}_{bi}(n)\tilde{\boldsymbol{w}}_{bk}^H(n)\}$ as (E1) and (E2), shown at the bottom of the page, where $\bar{\boldsymbol{I}}_m^{(\pm q)}$ denotes an $m \times m$ matrix with unity elements along the qth ($1 \le q \le m-1$) upper (for $+q$) or lower

$$E\{\boldsymbol{w}_{fi}(n)\tilde{\boldsymbol{w}}_{fk}^H(n)\} = \begin{cases} \boldsymbol{O}_{p\times(M-p)}, & \text{for } i-k < -(p-1) \\ \sigma^2[\bar{\boldsymbol{I}}_p^{(i-k)}, \boldsymbol{O}_{p\times(M-2p)}], & \text{for } -(p-1) \le i-k < 0 \\ \sigma^2[\boldsymbol{I}_p, \boldsymbol{O}_{p\times(M-2p)}], & \text{for } i-k = 0 \\ \sigma^2[\boldsymbol{O}_{p\times(i-k)}, \boldsymbol{I}_p, \boldsymbol{O}_{p\times(M-2p-(i-k))}], & \text{for } 0 < i-k \le M-2p \\ \sigma^2[\boldsymbol{O}_{p\times(M-2p)}, \bar{\boldsymbol{I}}_p^{(i-k-M+2p)}], & \text{for } M-2p < i-k \le M-p-1 \\ \boldsymbol{O}_{p\times(M-p)}, & \text{for } i-k > M-p-1 \end{cases} \triangleq \bar{\boldsymbol{\Sigma}}_{ik} \quad \text{(E1)}$$

$$E\{\boldsymbol{w}_{bi}(n)\tilde{\boldsymbol{w}}_{bk}^H(n)\} = \bar{\boldsymbol{\Sigma}}_{ik} \quad \text{(E2)}$$

(for $-q$) diagonal off the major diagonal and zeros elsewhere, and their elements are given by [23], [54]

$$(\bar{\boldsymbol{I}}_m^{(-q)})_{ik} = \begin{cases} 1, & \text{for } i-k=q \\ 0, & \text{others} \end{cases}$$
$$(\bar{\boldsymbol{I}}_m^{(+q)})_{ik} = \begin{cases} 1, & \text{for } k-i=q \\ 0, & \text{others.} \end{cases}$$

Similarly, we easily obtain [54]

$$E\{\tilde{\boldsymbol{w}}_{fi}(n)\tilde{\boldsymbol{w}}_{fk}^H(n)\} = \begin{cases} \sigma^2\boldsymbol{I}_{M-p}, & \text{for } i-k=0 \\ \sigma^2\bar{\boldsymbol{I}}_{M-p}^{(i-k)}, & \text{for } 0<|i-k|\le M-p-1 \\ \boldsymbol{O}_{(M-p)\times(M-p)}, & \text{for } |i-k|>M-p-1 \end{cases} \triangleq \tilde{\boldsymbol{\Sigma}}_{ik} \tag{E3}$$

$$E\{\tilde{\boldsymbol{w}}_{bi}(n)\tilde{\boldsymbol{w}}_{bk}^H(n)\} = \tilde{\boldsymbol{\Sigma}}_{ik}. \tag{E4}$$

Furthermore by defining $\bar{\boldsymbol{w}}_{fi}(n)$ and $\bar{\boldsymbol{w}}_{bi}(n)$ as element-reversed versions of $\tilde{\boldsymbol{w}}_{fi}(n)$ and $\tilde{\boldsymbol{w}}_{bi}(n)$, i.e., $\bar{\boldsymbol{w}}_{fl}(n) \triangleq [w_{l+L-2}(n),\ldots,w_{l+1}(n),w_l(n)]^T$, and $\bar{\boldsymbol{w}}_{bi}(n) \triangleq [w_{p+2-l}(n),\ldots,w_{M-l}(n),w_{M-l+1}(n)]^H$, we obtain [54]

$$\tilde{\boldsymbol{w}}_{fl}(n) = \bar{\boldsymbol{w}}_{bp+2-l}^*(n), \qquad \tilde{\boldsymbol{w}}_{bl}(n) = \bar{\boldsymbol{w}}_{fp+2-l}^*(n). \tag{E5}$$

Then, under the assumption for the additive noise, we get (E6) and (E7), shown at the bottom of the page, where $\bar{\boldsymbol{J}}_m^{(\pm q)}$ represents an $m\times m$ matrix with unity elements along the qth ($1\le q\le m-1$) upper (for $+q$) or lower (for $-q$) diagonal off the major cross-diagonal and zeros elsewhere, and their elements are given by [23], [54]

$$(\bar{\boldsymbol{J}}_m^{(+q)})_{ik} = \begin{cases} 1, & \text{for } i+k=m-q+1 \\ 0, & \text{others} \end{cases}$$
$$(\bar{\boldsymbol{J}}_m^{(-q)})_{ik} = \begin{cases} 1, & \text{for } i+k=m+q+1 \\ 0, & \text{others} \end{cases}.$$

Next, under the independence assumption for $\boldsymbol{P}(n-1)$ and $\boldsymbol{\Phi}_1(n)$, by using (C5) and (D3) and after some straightforward manipulations, we obtain

$$\begin{aligned} &E\{\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\tilde{\boldsymbol{P}}(n-1)\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\} \\ &\quad=\sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{z}}_{il}^H(n)\tilde{\boldsymbol{P}}(n-1) \\ &\qquad\cdot\tilde{\boldsymbol{P}}^H(n-1)\tilde{\boldsymbol{z}}_{kt}(n)\tilde{\boldsymbol{z}}_{kt}^H(n)\} \\ &\quad\triangleq\bar{\boldsymbol{\Psi}}_1\boldsymbol{K}(n-1)\bar{\boldsymbol{\Psi}}_1+\boldsymbol{K}_1+\boldsymbol{K}_2 \end{aligned} \tag{E8}$$

where

$$\boldsymbol{K}_1 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\bar{\boldsymbol{F}}_{il,kt}\boldsymbol{K}^T(n-1)\bar{\boldsymbol{F}}_{il,kt}^*$$
$$\boldsymbol{K}_2 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\boldsymbol{F}_{il,kt}\mathrm{tr}\{\boldsymbol{F}_{il,kt}^H\boldsymbol{K}(n-1)\}$$

in which $\bar{\boldsymbol{F}}_{il,kt} \triangleq E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{z}}_{kt}^T(n)\}$, and $\boldsymbol{F}_{il,kt} \triangleq E\{\tilde{\boldsymbol{z}}_{il}(n)\cdot\tilde{\boldsymbol{z}}_{kt}^H(n)\}$. For $i,k=1,2,3,4$ and $l,t=1,2,\ldots,p$, the expectations $\bar{\boldsymbol{F}}_{il,kt}$ and $\boldsymbol{F}_{il,kt}$ are summarized in Tables II and III, where $\boldsymbol{M}_{ik} \triangleq E\{\boldsymbol{y}_{fi}(n)\boldsymbol{y}_{bk}^T(n)\}$.

Then, in a way similar to (E8), by using (C11) and some calculations, we get

$$\begin{aligned} &E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\boldsymbol{Q}^H\boldsymbol{G}(n)\boldsymbol{\Phi}_1^H(n)\} \\ &\quad=\sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{g}}_{il}^H(n)\boldsymbol{Q}\boldsymbol{Q}^H\tilde{\boldsymbol{g}}_{kt}(n)\tilde{\boldsymbol{z}}_{kt}^H(n)\} \\ &\quad\triangleq\alpha^2\boldsymbol{P}\boldsymbol{P}^H+\boldsymbol{K}_3+\boldsymbol{K}_4 \end{aligned} \tag{E9}$$

where

$$\boldsymbol{K}_3 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\boldsymbol{\Gamma}_{il,kt}\boldsymbol{Q}^*\boldsymbol{Q}^T\boldsymbol{\Gamma}_{kt,il}^H$$
$$\boldsymbol{K}_4 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\boldsymbol{F}_{il,kt}\mathrm{tr}\{\tilde{\boldsymbol{\Gamma}}_{il,kt}^H\boldsymbol{Q}\boldsymbol{Q}^H\}$$

with $\boldsymbol{\Gamma}_{il,kt} \triangleq E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{g}}_{kt}^T(n)\}$, and $\tilde{\boldsymbol{\Gamma}}_{il,kt} \triangleq E\{\tilde{\boldsymbol{g}}_{il}(n)\cdot\tilde{\boldsymbol{g}}_{kt}^H(n)\}$, and $\boldsymbol{\Omega}\boldsymbol{Q}=\alpha\boldsymbol{P}$ is used implicitly. Furthermore, the expectations $\boldsymbol{\Gamma}_{il,kt}$ and $\tilde{\boldsymbol{\Gamma}}_{il,kt}$ are summarized in Tables IV and V for $i,k=1,2,3,4$ and $l,t=1,2,\ldots,p$, where $\boldsymbol{e}_k$ is an $(M-p)\times 1$ unit vector with one as the kth element whereas zeros elsewhere, and the results in (E3), (E4), (E6), and (E7) are used implicitly.

Finally, we can get the expectation $E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\cdot\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\}$ as

$$\begin{aligned} &E\{\boldsymbol{\Phi}_1(n)\boldsymbol{G}^H(n)\boldsymbol{Q}\tilde{\boldsymbol{P}}^H(n-1)\boldsymbol{\Phi}_1(n)\boldsymbol{\Phi}_1^H(n)\} \\ &\quad=\sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}E\{\tilde{\boldsymbol{z}}_{il}(n)\tilde{\boldsymbol{g}}_{il}^H(n)\boldsymbol{Q}\tilde{\boldsymbol{P}}^H(n-1)\cdot\tilde{\boldsymbol{z}}_{kt}(n)\tilde{\boldsymbol{z}}_{kt}^H(n)\} \\ &\quad\triangleq\alpha\boldsymbol{P}E\{\tilde{\boldsymbol{P}}^H(n-1)\}\bar{\boldsymbol{\Psi}}_1+\boldsymbol{K}_5+\boldsymbol{K}_6 \end{aligned} \tag{E10}$$

$$E\{\boldsymbol{w}_{fi}(n)\tilde{\boldsymbol{w}}_{bk}^H(n)\} = E\{\boldsymbol{w}_{fi}(n)\bar{\boldsymbol{w}}_{fp+2-k}^H(n)\} = \begin{cases} \boldsymbol{O}_{p\times(M-p)}, & \text{for } i+k<3 \\ \sigma^2[\boldsymbol{O}_{p\times(M-2p)},\bar{\boldsymbol{J}}_p^{(i+k-p-2)}], & \text{for } 3\le i+k<p+2 \\ \sigma^2[\boldsymbol{O}_{p\times(M-2p)},\boldsymbol{J}_p], & \text{for } i+k=p+2 \\ \sigma^2[\boldsymbol{O}_{p\times(M-p-(i+k)+2)},\boldsymbol{J}_p,\boldsymbol{O}_{p\times(i+k-(p+2))}], & \text{for } p+2<i+k\le M-p+2 \\ \sigma^2[\bar{\boldsymbol{J}}_p^{(i-k-M+p-2)},\boldsymbol{O}_{p\times(M-2p)}], & \text{for } M-p+2<i+k\le M+1 \\ \boldsymbol{O}_{p\times(M-p)}, & \text{for } i+k>M+1 \end{cases} \triangleq \boldsymbol{\Sigma}_{ik} \tag{E6}$$

$$E\{\boldsymbol{w}_{bi}(n)\tilde{\boldsymbol{w}}_{fk}^H(n)\} = E\{\boldsymbol{w}_{bi}(n)\bar{\boldsymbol{w}}_{bp+2-k}^H(n)\} = \boldsymbol{\Sigma}_{ik} \tag{E7}$$

TABLE IV
RESULTS OF EXPECTATION $\boldsymbol{\Gamma}_{il,kt}$ FOR $i,k = 1,2,3,4$, AND $l,t = 1,2,\ldots,p$

	$k=1$	$k=2$	$k=3$	$k=4$
$i=1$	$\boldsymbol{O}_{p\times(M-p)}$	$\begin{cases}\boldsymbol{O}_{p\times(M-p)}, & t=1,\cdots,p-1 \\ & \& \ p\neq 1 \\ \sigma^2\overline{\varphi}_{fl}\boldsymbol{e}_{M-p}^T, & t=p\end{cases}$	$r_{1M}\boldsymbol{\Sigma}_{lt}$	$r_{MM}\boldsymbol{\Sigma}_{l,t+1}$
$i=2$	$\begin{cases}\sigma^2\varphi_{fl+1}\boldsymbol{e}_1^T, & t=1 \\ \boldsymbol{O}_{p\times(M-p)}, & t=2,\cdots,p \\ & \& \ p\neq 1\end{cases}$	$\boldsymbol{O}_{p\times(M-p)}$	$r_{11}\boldsymbol{\Sigma}_{l+1,t}$	$r_{M1}\boldsymbol{\Sigma}_{l+1,t+1}$
$i=3$	$r_{1M}\boldsymbol{\Sigma}_{lt}$	$r_{11}\boldsymbol{\Sigma}_{l,t+1}$	$\boldsymbol{O}_{p\times(M-p)}$	$\begin{cases}\boldsymbol{O}_{p\times(M-p)}, & t=1,\cdots,p-1 \\ & \& \ p\neq 1 \\ \sigma^2\overline{\varphi}_{bl}\boldsymbol{e}_{M-p}^T, & t=p\end{cases}$
$i=4$	$r_{MM}\boldsymbol{\Sigma}_{l+1,t}$	$r_{M1}\boldsymbol{\Sigma}_{l+1,t+1}$	$\begin{cases}\sigma^2\varphi_{bl+1}\boldsymbol{e}_1^T, & t=1 \\ \boldsymbol{O}_{p\times(M-p)}, & t=2,\cdots,p \\ & \& \ p\neq 1\end{cases}$	$\boldsymbol{O}_{p\times(M-p)}$

TABLE V
RESULTS OF EXPECTATION $\bar{\boldsymbol{\Gamma}}_{il,kt}$ FOR $i,k = 1,2,3,4$, AND $l,t = 1,2,\ldots,p$

	$k=1$	$k=2$	$k=3$	$k=4$
$i=1$	$r_{MM}\tilde{\boldsymbol{\Sigma}}_{lt}$	$r_{1M}\tilde{\boldsymbol{\Sigma}}_{l,t+1}$	$\begin{cases}\sigma^4\overline{\boldsymbol{J}}_{M-p}^{(M-p-1)}, & l=t=1 \\ \boldsymbol{O}_{(M-p)\times(M-p)}, & \text{others}\end{cases}$	$\boldsymbol{O}_{(M-p)\times(M-p)}$
$i=2$	$r_{M1}\tilde{\boldsymbol{\Sigma}}_{l+1,t}$	$r_{11}\tilde{\boldsymbol{\Sigma}}_{l+1,t+1}$	$\boldsymbol{O}_{(M-p)\times(M-p)}$	$\begin{cases}\sigma^4\overline{\boldsymbol{J}}_{M-p}^{(-(M-p-1))}, & l=t=p \\ \boldsymbol{O}_{(M-p)\times(M-p)}, & \text{others}\end{cases}$
$i=3$	$\begin{cases}\sigma^4\overline{\boldsymbol{J}}_{M-p}^{(M-p-1)}, & l=t=1 \\ \boldsymbol{O}_{(M-p)\times(M-p)}, & \text{others}\end{cases}$	$\boldsymbol{O}_{(M-p)\times(M-p)}$	$r_{11}\tilde{\boldsymbol{\Sigma}}_{lt}$	$r_{1M}\tilde{\boldsymbol{\Sigma}}_{l,t+1}$
$i=4$	$\boldsymbol{O}_{(M-p)\times(M-p)}$	$\begin{cases}\sigma^4\overline{\boldsymbol{J}}_{M-p}^{(-(M-p-1))}, & l=t=p \\ \boldsymbol{O}_{(M-p)\times(M-p)}, & \text{others}\end{cases}$	$r_{M1}\tilde{\boldsymbol{\Sigma}}_{l+1,t}$	$r_{MM}\tilde{\boldsymbol{\Sigma}}_{l+1,t+1}$

where

$$\boldsymbol{K}_5 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\bar{\boldsymbol{F}}_{il,kt}E\{\tilde{\boldsymbol{P}}^*(n-1)\}\boldsymbol{Q}^T\boldsymbol{\Gamma}_{kt,il}^H$$

$$\boldsymbol{K}_6 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\boldsymbol{F}_{il,kt}\mathrm{tr}\{\bar{\boldsymbol{\Gamma}}_{kt,il}\boldsymbol{Q}E\{\tilde{\boldsymbol{P}}^H(n-1)\}\}$$

with $\bar{\boldsymbol{\Gamma}}_{kt,il} \triangleq E\{\tilde{\boldsymbol{z}}_{kt}(n)\tilde{\boldsymbol{g}}_{il}^H(n)\}$, and the expectation $\bar{\boldsymbol{\Gamma}}_{kt,il}$ are summarized in Table VI for $i,k = 1, 2, 3, 4$ and $l,t = 1,2,\ldots,p$, where the results in (E1) and (E2) are used implicitly. ■

APPENDIX F
PROOF OF THEOREM 2

By using the algebraic identities for the matrices with compatible dimensions (e.g., [52] and [48])

$$\mathrm{vec}(\boldsymbol{XYZ}) = (\boldsymbol{Z}^T \otimes \boldsymbol{X})\mathrm{vec}(\boldsymbol{Y}) \tag{F1}$$

$$\mathrm{tr}\{\boldsymbol{XY}\} = \mathrm{vec}^H(\boldsymbol{X}^H)\mathrm{vec}(\boldsymbol{Y}) \tag{F2}$$

where $\mathrm{vec}(\cdot)$ converts the matrix into a column vector by stacking the matrix columns one beneath the other beginning with the leftmost column, from (50), we obtain the following vector relation after some manipulations:

$$\begin{aligned}\mathrm{vec}(\boldsymbol{K}(n)) =& (\boldsymbol{I}_{p^2} - \mu\boldsymbol{C} + \mu^2\bar{\boldsymbol{C}})\mathrm{vec}(\boldsymbol{K}(n-1)) \\ &+ \mu^2\bar{\boldsymbol{K}}_1\mathrm{vec}(\boldsymbol{K}^T(n-1)) \\ &+ \mu\boldsymbol{c}_1(n-1) + \mu^2\boldsymbol{c}_2(n-1)\end{aligned} \tag{F3}$$

where

$$\begin{aligned}\boldsymbol{c}_1(n-1) \triangleq& \alpha(\boldsymbol{I}_p \otimes \boldsymbol{P})\mathrm{vec}(E\{\tilde{\boldsymbol{P}}^H(n-1)\}) \\ &+ \alpha(\boldsymbol{P}^* \otimes \boldsymbol{I}_p)\mathrm{vec}(E\{\tilde{\boldsymbol{P}}(n-1)\})\end{aligned} \tag{F4}$$

$$\begin{aligned}\boldsymbol{c}_2(n-1) \triangleq& \alpha^2\mathrm{vec}(\boldsymbol{P}\boldsymbol{P}^H) - \alpha(\boldsymbol{P}^* \otimes \bar{\boldsymbol{\Psi}}_1)\mathrm{vec}(E\{\tilde{\boldsymbol{P}}(n-1)\}) \\ &- \alpha(\bar{\boldsymbol{\Psi}}_1^T \otimes \boldsymbol{P})\mathrm{vec}(E\{\tilde{\boldsymbol{P}}^H(n-1)\}) \\ &+ \bar{\boldsymbol{K}}_3\mathrm{vec}(\boldsymbol{Q}^*\boldsymbol{Q}^T) + \bar{\boldsymbol{K}}_4\mathrm{vec}(\boldsymbol{Q}\boldsymbol{Q}^H) \\ &- \bar{\boldsymbol{K}}_5(\boldsymbol{Q} \otimes \boldsymbol{I}_p)\}\mathrm{vec}(E\{\tilde{\boldsymbol{P}}^*(n-1)\}) \\ &- \bar{\boldsymbol{K}}_6(\boldsymbol{I}_p \otimes \boldsymbol{Q})\mathrm{vec}(E\{\tilde{\boldsymbol{P}}^H(n-1)\}) \\ &- \tilde{\boldsymbol{K}}_5(\boldsymbol{I}_p \otimes \boldsymbol{Q}^*)\mathrm{vec}(E\{\tilde{\boldsymbol{P}}^T(n-1)\}) \\ &- \tilde{\boldsymbol{K}}_6(\boldsymbol{Q}^* \otimes \boldsymbol{I}_p)\mathrm{vec}(E\{\tilde{\boldsymbol{P}}(n-1)\})\end{aligned} \tag{F5}$$

TABLE VI
RESULTS OF EXPECTATION $\bar{\boldsymbol{\Gamma}}_{kt,il}$ FOR $i,k = 1,2,3,4$, AND $l,t = 1,2,\ldots,p$

	$i=1$	$i=2$	$i=3$	$i=4$
$k=1$	$r_{MM}\bar{\boldsymbol{\Sigma}}_{tl}$	$r_{1M}\bar{\boldsymbol{\Sigma}}_{t,l+1}$	$\begin{cases}\sigma^2\bar{\boldsymbol{\varphi}}_{ft}\boldsymbol{e}_1^T, & l=1\\ \boldsymbol{O}_{p\times(M-p)}, & \text{others}\end{cases}$	$\boldsymbol{O}_{p\times(M-p)}$
$k=2$	$r_{M1}\bar{\boldsymbol{\Sigma}}_{t+1,l}$	$r_{11}\bar{\boldsymbol{\Sigma}}_{t+1,l+1}$	$\boldsymbol{O}_{p\times(M-p)}$	$\begin{cases}\sigma^2\boldsymbol{\varphi}_{ft+1}\boldsymbol{e}_{M-p}^T, & l=p\\ \boldsymbol{O}_{p\times(M-p)}, & \text{others}\end{cases}$
$k=3$	$\begin{cases}\sigma^2\bar{\boldsymbol{\varphi}}_{bt}\boldsymbol{e}_1^T, & l=1\\ \boldsymbol{O}_{p\times(M-p)}, & \text{others}\end{cases}$	$\boldsymbol{O}_{p\times(M-p)}$	$r_{11}\bar{\boldsymbol{\Sigma}}_{tl}$	$r_{1M}\bar{\boldsymbol{\Sigma}}_{t,l+1}$
$k=4$	$\boldsymbol{O}_{p\times(M-p)}$	$\begin{cases}\sigma^2\boldsymbol{\varphi}_{bt+1}\boldsymbol{e}_{M-p}^T, & l=p\\ \boldsymbol{O}_{p\times(M-p)}, & \text{others}\end{cases}$	$r_{M1}\bar{\boldsymbol{\Sigma}}_{t+1,l}$	$r_{MM}\bar{\boldsymbol{\Sigma}}_{t+1,l+1}$

$$\bar{\boldsymbol{K}}_3 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\boldsymbol{\Gamma}_{kt,il}^{*}\otimes\boldsymbol{\Gamma}_{il,kt} \tag{F6}$$

$$\bar{\boldsymbol{K}}_4 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\mathrm{vec}(\boldsymbol{F}_{il,kt})\mathrm{vec}^H(\tilde{\boldsymbol{\Gamma}}_{il,kt}) \tag{F7}$$

$$\bar{\boldsymbol{K}}_5 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\boldsymbol{\Gamma}_{kt,il}^{*}\otimes\bar{\boldsymbol{F}}_{il,kt} \tag{F8}$$

$$\bar{\boldsymbol{K}}_6 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\mathrm{vec}(\boldsymbol{F}_{il,kt})\mathrm{vec}^H(\bar{\boldsymbol{\Gamma}}_{kt,il}^{H}) \tag{F9}$$

$$\check{\boldsymbol{K}}_5 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\bar{\boldsymbol{F}}_{il,kt}^{H}\otimes\boldsymbol{\Gamma}_{il,kt} \tag{F10}$$

$$\check{\boldsymbol{K}}_6 \triangleq \sum_{l=1}^{p}\sum_{t=1}^{p}\sum_{i=1}^{4}\sum_{k=1}^{4}\mathrm{vec}(\boldsymbol{F}_{il,kt})\mathrm{vec}^H(\bar{\boldsymbol{\Gamma}}_{kt,il}) \tag{F11}$$

in which the $p\times p$ matrices $\boldsymbol{F}_{il,kt}$ and $\bar{\boldsymbol{F}}_{il,kt}$, $p\times(M-p)$ matrices $\boldsymbol{\Gamma}_{il,kt}$ and $\bar{\boldsymbol{\Gamma}}_{kt,il}$, and $(M-p)\times(M-p)$ matrix $\tilde{\boldsymbol{\Gamma}}_{il,kt}$ are given in Tables II–VI of Appendix E. Since the $p\times p$ matrix $\boldsymbol{K}(n)$ is Hermitian and its diagonal elements involve the mean-squared weight-errors $\{|(\tilde{\boldsymbol{P}}(n))_{ik}|^2\}$ of the matrix $\boldsymbol{P}(n)$, to facilitate the analysis of mean-square stability, we write the matrix $\boldsymbol{K}(n)$ and vectors $\boldsymbol{c}_1(n-1)$ and $\boldsymbol{c}_2(n-1)$ in terms of their real and imaginary parts as $\boldsymbol{K}(n)\triangleq\boldsymbol{K}_r(n)+j\boldsymbol{K}_i(n)$, $\boldsymbol{c}_1(n-1)\triangleq\boldsymbol{c}_{1r}(n-1)+j\boldsymbol{c}_{1i}(n-1)$, and $\boldsymbol{c}_2(n-1)\triangleq\boldsymbol{c}_{2r}(n-1)+j\boldsymbol{c}_{2i}(n-1)$. Then, the vector recursion (F3) can be expressed in its real and imaginary parts

$$\mathrm{vec}(\boldsymbol{K}_r(n))=\boldsymbol{F}_r\mathrm{vec}(\boldsymbol{K}_r(n-1))+\mu\boldsymbol{c}_{1r}(n-1)+\mu^2\boldsymbol{c}_{2r}(n-1) \tag{F12}$$

$$\mathrm{vec}(\boldsymbol{K}_i(n))=\boldsymbol{F}_i\mathrm{vec}(\boldsymbol{K}_i(n-1))+\mu\boldsymbol{c}_{1i}(n-1)+\mu^2\boldsymbol{c}_{2i}(n-1) \tag{F13}$$

where

$$\boldsymbol{F}_r=\boldsymbol{I}_{p^2}-\mu\boldsymbol{C}+\mu^2\tilde{\boldsymbol{C}},\qquad \boldsymbol{F}_i=\boldsymbol{I}_{p^2}-\mu\boldsymbol{C}+\mu^2\breve{\boldsymbol{C}}. \tag{F14}$$

Hence, we conclude that the convergence of recursion (F3) is governed by the stability of the $p^2\times p^2$ matrices $\boldsymbol{F}_r$ and $\boldsymbol{F}_i$ (i.e., all their eigenvalues $\lambda(\boldsymbol{F}_r)$ and $\lambda(\boldsymbol{F}_i)$ should lie inside the unit circle) for the multiple signal case [36], [28], while the parabola $F_r=F_{\mathrm{m.s.}}\triangleq 1-\mu C+\mu^2\tilde{C}$ dominates the stability of (F3) (i.e., $F_{\mathrm{m.s.}}$ should satisfy the inequality $0<F_{\mathrm{m.s.}}<1$) for the single signal case (see Section IV-E for more details). From (52), (55), (56), and (53), we easily see that the $p^2\times p^2$ matrices $\boldsymbol{C}$ and $\tilde{\boldsymbol{C}}$ are Hermitian. Because the eigenvalues of $\boldsymbol{C}$ are all the combinations $\{\lambda_i+\lambda_k\}$ of the eigenvalues $\{\lambda_i\}$ of matrix $\bar{\boldsymbol{\Psi}}_1$ for $1\le i,k\le p$ [28], we can find that the matrix $\boldsymbol{C}$ is positive definite and invertible in view of that $\lambda_i>0$ for $i=1,2,\ldots,p$. We can also see that the Hermitian matrices $\tilde{\boldsymbol{C}}$ and $\breve{\boldsymbol{C}}$ are nonnegative definite. Thus, we can find that the matrices $\boldsymbol{F}_r$ and $\boldsymbol{F}_i$ in (F14) are stable if and only if μ satisfies (51) [36], [28]. As a result, the stability condition of the LMS algorithm in the mean-square sense in (51) is obtained immediately. ■

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the Associate Editor Prof. J. P. Delmas for their critical review, incisive comments, and constructive suggestions and especially for bringing [59]–[66] and [69] to their attention.

REFERENCES

[1] S. Haykin, Ed., *Advances in Spectrum Analysis and Array Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995, vol. I–III.
[2] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[3] H. L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*. New York: Wiley, 2002.
[4] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
[5] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. Astron. Soc.*, vol. 33, pp. 347–366, 1973.
[6] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, Rome, NY, Oct. 1979, pp. 243–258.
[7] 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, 1982.
[8] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 19, no. , pp. 134–139, 1983.

[9] R. Roy and T. Kailath, "ESPRIT—Estimation for signal parameters via rational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.

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

[11] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[12] J. Xin and A. Sano, "MSE-based regularization approach to direction estimation of coherent narrow-band signals using linear prediction," *IEEE Trans. Signal Process.*, vol. 49, pp. 2481–2497, 2001.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: The John Hopkins Univ. Press, 1989.

[14] J. P. Reilly, "A real-time high-resolution technique for angle-of-arrival estimation," *Proc. IEEE*, vol. 75, no. 12, pp. 1692–1694, Dec. 1987.

[15] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, no. 8, pp. 1327–1343, Aug. 1990.

[16] V. U. Reddy, G. Mathew, and A. Paulraj, "Some algorithms for eigensubspace estimation," *Dig. Signal Process.*, vol. 5, pp. 97–115, 1995.

[17] R. D. DeGroat, E. M. Dowling, and D. A. Linebarger, "Subspace tracking," in *The Digital Signal Processing Handbook*, V. K. Madisetti and D. B. Williams, Eds. Boca Raton, FL: CRC, 1998.

[18] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, pp. 95–107, 1995.

[19] C.-C. Yeh, "Simple computation of projection matrix for bearing estimations," *Proc. Inst. Elect. Eng.*, pt. F, vol. 134, no. 2, pp. 146–150, 1987.

[20] S. Marcos, A. Marsal, and M. Benidir, "The propagator method for source bearing estimation," *Signal Process.*, vol. 42, no. 2, pp. 121–138, 1995.

[21] A. Eriksson, P. Stoica, and T. Söderström, "On-line subspace algorithms for tracking moving sources," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2319–2330, Sep. 1994.

[22] 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.

[23] J. Xin and A. Sano, "Computationally efficient subspace-based method for direction-of-arrival estimation without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 876–893, Apr. 2004.

[24] A. L. Swindlehurst, P. Stoica, and M. Jansson, "Exploiting arrays with multiple invariance using MUSIC and MODE," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2511–2521, Nov. 2001.

[25] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[26] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[27] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[28] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[29] L. L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 3, pp. 722–736, Jun. 1981.

[30] B. Fisher and N. J. Bershad, "The complex LMS adaptive algorithm-transient weight mean and covariance with applications to the ALE," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 1, pp. 34–44, Feb. 1983.

[31] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 1, pp. 222–230, Feb. 1985.

[32] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 680–691, Jul. 1988.

[33] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 4, pp. 793–806, Aug. 1986.

[34] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2811–2825, Sep. 1993.

[35] V. H. Nascimento and A. H. Sayed, "On the learning mechanism of adaptive filters," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1609–1625, Jun. 2000.

[36] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639–652, Mar. 2003.

[37] J. Sanchez-Araujo and S. Marcos, "Tracking moving sources using subspace-based adaptive linear methods," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Munich, Germany, Apr. 1997, pp. 3497–3500.

[38] J. Xin and A. Sano, "Directions-of-arrival tracking of coherent cyclostationary signals in array processing," *IEICE Trans. Fundamentals*, vol. E86-A, no. 8, pp. 2037–2046, 2003.

[39] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 806–811, Aug. 1985.

[40] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signals identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 1, pp. 8–15, Jan. 1989.

[41] J. Sheinvald, M. Wax, and A. J. Weiss, "On maximum-likelihood localization of coherent signals," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2475–2482, Oct. 1996.

[42] J. C. Liberti Jr and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[43] J. Xin and A. Sano, "Linear prediction approach to direction estimation of cyclostationary signals in multipath environment," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 710–720, Apr. 2001.

[44] ——, "Direction estimation of coherent signals using spatial signature," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 414–417, Dec. 2002.

[45] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 996–1000, Jul. 1989.

[46] C. R. Rao, C. R. Sastry, and B. Zhou, "Tracking the direction of arrival of multiple moving targets," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1133–1144, May 1994.

[47] S. Affes, S. Gazor, and Y. Grenier, "An algorithm for multisource beamforming and multitarget tracking," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1512–1522, Jun. 1996.

[48] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[49] N. J. Bershad, "On the optimum gain parameter in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 7, pp. 1065–1068, Jul. 1987.

[50] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Mar. 1997.

[51] P. H. M. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Trans. Autom. Control*, vol. 33, no. 9, pp. 867–870, Sep. 1988.

[52] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sep. 1978.

[53] P. C. Wei, J. Han, J. R. Zeidlei, and W. H. Ku, "Comparative tracking performance of the LMS and RLS algorithms for chirped narrow-band signal recovery," *IEEE Trans. Signal Process.*, vol. 50, pp. 1602–1609, 2002.

[54] J. Xin and A. Sano, "Effect of subarray size on direction estimation of coherent cyclostationary signals based on forward-backward linear prediction," *IEICE Trans. Fundamentals*, vol. E85-A, no. 8, pp. 1807–1821, 2002.

[55] S. Haykin and A. Steinhardt, Eds., *Adaptive Radar Detection and Estimation*. New York: Wiley, 1992.

[56] D. R. Morgan and S. G. Kratzer, "On a class of computationally efficient, rapidly converging, generalized NLMS algorithms," *IEEE Signal Process. Lett.*, vol. 3, no. 8, pp. 245–247, Aug. 1996.

[57] J. Munier and G. Y. Delisle, "Spatial analysis using new properties of the cross-spectral matrix," *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 746–749, Mar. 1991.

[58] S. Marcos and M. Benidir, "On a high resolution array processing method nonbased on the eigenanalysis approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Albuquerque, NM, Apr. 1990, pp. 2955–2958.

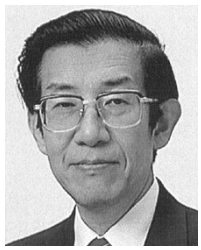
[59] ——, "Source bearing estimation and sensor positioning with the propagator method," in *Proc. SPIE Advanced Signal Processing Algorithms, Architectures, Implementations*, vol. 1348, San Diego, CA, Jul. 1990, pp. 312–323.

[60] S. Marcos, A. Marsal, and M. Benidir, "Performance analysis of the propagation method for source bearing estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Adelaide, Australia, Apr. 1994, pp. 237–240.

[61] S. Marcos and M. Benidir, "An adaptive subspace algorithm for direction finding and tracking," in *Proc. SPIE Advanced Signal Processing Algorithms, Architectures, Implementations*, vol. 2563, Hong Kong, Jun. 1995, pp. 230–239.
[62] S. Marcos, M. Benidir, and J. Sanchez-Araujo, "An adaptive tracking algorithm for direction finding and array shape estimation in a nonstationary environment," *J. VLSI Signal Process. Syst.*, vol. 14, no. 1, pp. 107–118, 1996.
[63] J. Sanchez-Araujo and S. Marcos, "Statistical analysis of the propagator method for DOA estimation without eigendecomposition," in *Proc. IEEE Workshop Statistical Signal Array Processing*, Corfu, Greece, Jun. 1996, pp. 570–573.
[64] ——, "A class of subspace-based methods without eigendecomposition," in *Proc. IASTED Int. Conf. Signal Image Processing Applications*, Annecy, France, Jun. 1996, pp. 202–205.
[65] S. Marcos and J. Sanchez-Araujo, "High resolution 'linear' methods for direction of arrival estimation: Performance and complexity," *Traitement du Signal*, vol. 14, no. 2, pp. 99–116, 1997.
[66] J. Sanchez-Araujo and S. Marcos, "An efficient PASTd-algorithm implementation for multiple direction of arrival tracking," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2321–2324, Aug. 1999.
[67] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. 22, no. 4, pp. 551–575, Aug. 1977.
[68] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
[69] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer-Verlag, 1990.
[70] B. Yang, "Asymptotic convergence analysis of the projection approximation subspace tracking algorithms," *Signal Process.*, vol. 50, pp. 123–136, 1996.
[71] J. P. Delmas and J. F. Cardoso, "Performance analysis of an adaptive algorithm for tracking dominant subspaces," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3045–3057, Nov. 1998.
[72] J. P. Delmas and F. Alberge, "Asymptotic performance analysis of subspace adaptive algorithms introduced in the neural network literature," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 170–182, Jan. 1998.
[73] J. P. Delmas, "On adaptive EVD asymptotic distribution of centro-symmetric covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1402–1406, May 1999.
[74] A. B. Gershman and V. T. Ermolaev, "Optimal subarray size for spatial smoothing," *IEEE Signal Process. Lett.*, vol. 2, pp. 28–30, Feb. 1995.
[75] J. Xin, Y. Ohashi, and A. Sano, "Computational efficient method for estimation of the number of signals in multipath environment," presented at the IEEE 13th Workshop on Statistical Signal Processing, Bordeaux, France, Jul. 2005.

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 Ministry of Posts and Telecommunications (MPT) of China, Xi'an. He was with Communications Research Laboratory (currently the National Institute of Information and Communications Technology), Japan, as an Invited Research Fellow of the Telecommunications Advancement Organization of Japan (TAO) (currently the National Institute of Information and Communications Technology) from 1996 to 1997 and as a Postdoctoral Fellow of the Japan Science and Technology Corporation (JST) (currently the Japan Science and Technology Agency) from 1997 to 1999. He also served as a Guest (Senior) Researcher with YRP Mobile Telecommunications Key Technology Research Laboratories Company, Ltd., Yokosuka, Japan, from 1999 to 2001. Since 2002, he has been with Fujitsu Laboratories, Ltd., Yokosuka, Japan. His research interests are in the areas of signal processing, including adaptive filtering, statistical signal processing, sensor array processing and spectral estimation, 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, 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 (currently the National Institute of Information and Communications Technology), Japan. He is a coauthor of the textbook *State Variable Methods in Automatic Control* (New York: Wiley, 1988). 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.

Dr. Sano 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 (IEICE) of Japan. He received the Kelvin Premium from the Institute of Electrical Engineering in 1986. He was General Co-Chair of 1999 IEEE Conference of Control Applications and an IPC Chair of 2004 IFAC Workshop on Adaptation and Learning in Control and Signal Processing. He served as Chair of IFAC Technical Committee on Modeling and Control of Environmental Systems from 1996 to 2001. He has also been Vice Chair of IFAC Technical Committee on Adaptive Control and Learning since 1999 and has been Chair of IFAC Technical Committee on Adaptive and Learning Systems since 2002. He is also presently on the Editorial Board of *Signal Processing*.