

Received December 11, 2019, accepted February 17, 2020, date of current version March 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976534

A Variable Regularized Recursive Subspace Model Identification Algorithm With Extended Instrumental Variable and Variable Forgetting Factor

JIAN-QIANG LIN^{ID}, (Student Member, IEEE), SHING-CHOW CHAN^{ID}, (Member, IEEE),
AND HAI-JUN TAN^{ID}

Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Corresponding author: Shing-Chow Chan (scchan@eee.hku.hk)

ABSTRACT This paper proposes a new smoothly clipped absolute deviation (SCAD) regularized recursive subspace model identification algorithm with square root (SR) extended instrumental variable (EIV) and locally optimal variable forgetting factor (LOFF). The proposed algorithm is based on a new local polynomial modeling (LPM)-based variable FF EIV projection approximation subspace tracking (PAST) and multivariate recursive least squares algorithms for estimating respectively the subspace of the extended observability matrix, and the input matrix and feedthrough matrix. The SREIV-PAST algorithm offers improved resilience to additive noise over the PAST algorithm and it is implemented using the more stable square-root form. The asymptotic mean square error of the corresponding LPM-based model is derived and minimized at each time instant to obtain the proposed LOFF for improving the convergence speed and steady state error. A recursive bi-iteration singular value decomposition (SVD) algorithm is also proposed for recursive computation of the pseudo-inverse of the state transition matrix and its eigenvalues. This facilitates online estimation of its model order using classical model selection criteria. Moreover, a new criterion based on the percentage of explained variance is also proposed with improved performance. The SCAD regularization is proposed for automatic model selection of the input and feedthrough matrices, as it is asymptotically unbiased. Efficient techniques for incorporating SCAD and estimating other required quantities are also developed. The proposed algorithms are evaluated using computer simulations under stationary and nonstationary environments and a real dataset on wing flutter data. Results show that the proposed algorithms offer improved convergence speed and steady state performance over the conventional algorithms and it also provides an online estimate of the system model order.

INDEX TERMS Recursive subspace model identification (RSI), variable forgetting factor (VFF), SCAD regularization, extended instrumental variable (EIV), local polynomial modeling (LPM), model order estimation.

I. INTRODUCTION

Subspace-based approaches play an important role in system identification [1]–[4] due to its theoretical and practical convenience for building state-space models directly from the input-output data of systems. Offline subspace identification techniques are characterized by the use of robust numerical tools such as the RQ factorization and the singular

values decomposition (SVD) of the data matrix [5]. However, the batch subspace model identification (SMI) algorithms are computational intensive for online implementation due to the high computational complexity of SVD. Consequently, much research on recursive subspace model identification (RSI) [6]–[9] has been devoted to update the model parameters over time with a reduced computational cost. Most RSI algorithms estimate the unknown state matrix \mathbf{A} and the output matrix \mathbf{C} from the “data equation” consisting of the input and output samples of the systems. This requires

The associate editor coordinating the review of this manuscript and approving it for publication was Anand Paul^{ID}.

the estimation of the subspace of the extended observability matrix, which requires SVD [10] or the use of more efficient subspace tracking algorithms [11], [12]. The computation of the state matrix \mathbf{A} also requires the evaluation of a pseudo-inverse and few works discussed the online estimation of the model order since it involves the computation of its eigenvalues. On the other hand, the input matrix \mathbf{B} , and feedthrough matrix \mathbf{D} can be estimated using the RLS algorithm with multiple outputs. $L1$ regularization has been proposed for automatic selection of their orders by shrinkage [13].

In the context of recursive subspace tracking, the projection approximation subspace tracking (PAST) algorithm [14] has been shown to yield good performance and efficient implementation for tracking the subspace spanned by the observability matrix. The PAST algorithm is based on the use of the recursive least squares (RLS) algorithm and the projection approximation. It assumes that the subspace is slowly time-varying and a fixed forgetting factor (FF) is usually used. Therefore, its tracking performance in time-varying systems or sudden system changes may be degraded. Furthermore, a drawback of classical recursive SMI [15]–[17] algorithms is that the disturbances acting on the system output are required to be spatially and temporally white. Biased estimates will result if such assumption is violated. The instrumental variable (IV) method [18]–[21] is an effective approach to tackle these types of noise problems by exploring the cross-correlation between the instrumental variables with the input variables rather than the covariance of the inputs as in conventional least squares method. The extension of IV to the PAST algorithm, namely the extended IV PAST (EIV-PAST) method, has been proposed in [22] to address the problem of biased estimates in the RSI algorithm when the output disturbances are not spatially and temporally white.

It is known that the EIV-PAST algorithm is able to achieve a faster convergence speed and smaller mean square error (MSE) in slowly time-varying environment if a large FF is utilized. On the other hand, a relatively small FF should be used to facilitate the tracking of fast time-varying parameters. In the context of RLS algorithms, Thus, a number of variable FF-based (VFF) algorithms have been proposed [23]–[26] to achieve a satisfactory performance in stationary, slowly or fast time-varying environment. In particular, a local polynomial modeling (LPM)-based VFF-RLS algorithm is recently developed in [26] by minimizing the MSE and its performance compares favorably with other conventional VFF-RLS algorithms tested. Note that these algorithms are applicable to the case of single output and rely on the Wiener solution involving the covariance of the inputs. On the other hand, the EIV-PAST involves multiple outputs and a system equation involving the cross-correlation between the input signal and the instrumental variables. Therefore, it is highly desirable to develop new VFF algorithms for better adaptation of the EIV-PAST algorithm in time-varying and noisy environment. Furthermore, it is important to be able to incorporate variable regularization in the EIV-PAST algorithms to improve their numerical behavior at low input signal level.

In this paper, we propose a new RSI algorithm where the subspace of the extended observability matrix (and hence the state matrix \mathbf{A}) is estimated by a new LPM-based VFF square-root (SR) EIV-PAST algorithm while the input matrix \mathbf{B} and feedthrough matrix \mathbf{D} are estimated by a VFF multivariate RLS algorithm. A recursive bi-iteration SVD algorithm is also proposed for the recursive computation of the pseudo-inverse of \mathbf{A} and its eigenvalues. This allows us to estimate online the model order of \mathbf{A} using classical Akaike information criterion (AIC) and minimum description length (MDL) criteria. Moreover, a new model order selection criterion based on the percentage of explained variance is proposed. On the other hand, the smoothly clipped absolute deviation (SCAD) regularization (SC) is proposed for automatic selection for input matrix \mathbf{B} and feedthrough matrix \mathbf{D} , as it is asymptotically unbiased unlike the $L1$ regularization [27].

The proposed VFF-SREIV-PAST algorithm utilizes the LPM channel model [26] of time-varying channels so as to derive the MSE of the EIV linear model as a function of the forgetting factor. Moreover, the resultant MSE contains a bias term which increases with the FF and a variance term which decreases with the FF. Therefore, the asymptotic MSE can be minimized at each time instant to obtain the proposed locally optimized FF for improving the convergence speed and steady state error. Techniques for estimating the various quantities required will also be introduced. The resultant SREIV-PAST algorithm offers improved resilience to additive noise over the PAST algorithm and it is implemented using the more stable square-root form.

The proposed VFF multivariate RLS algorithm is an extension of the VFF-RLS algorithm for single output [26] to multiple outputs with SCAD regularization. The VFF speeds up the convergence of the multivariate RLS algorithm whilst the SCAD regularization helps to achieve automatic model selection and the reduction of the estimation variance arising from possibly ill-conditioned covariance matrix especially at low input signal level. This is very useful during signal fading where the covariance matrix may be significantly ill-conditioned. Efficient techniques for incorporating SCAD in the numerical more stable QR decomposition (QRD) and estimating other required quantities are also developed.

Different from the LPM-based VFF-RLS algorithm in [26], which deals with single output and autocovariance matrix in the Wiener filter, the proposed algorithms are applicable to complex inputs, single and multiple outputs, and EIV systems involving cross-correlation matrix. The proposed algorithms are evaluated using computer simulations under stationary and nonstationary environments and a real dataset on wing flutter data. Experimental results show that the proposed algorithm provides faster convergence performance and more accurate estimates than conventional algorithms tested. Moreover, it also provides the model order for the state transition matrix and automatic model order selection for the input and feedthrough matrices. The convergence of the VFF-SREIV-PAST and RLS algorithms are also analyzed.

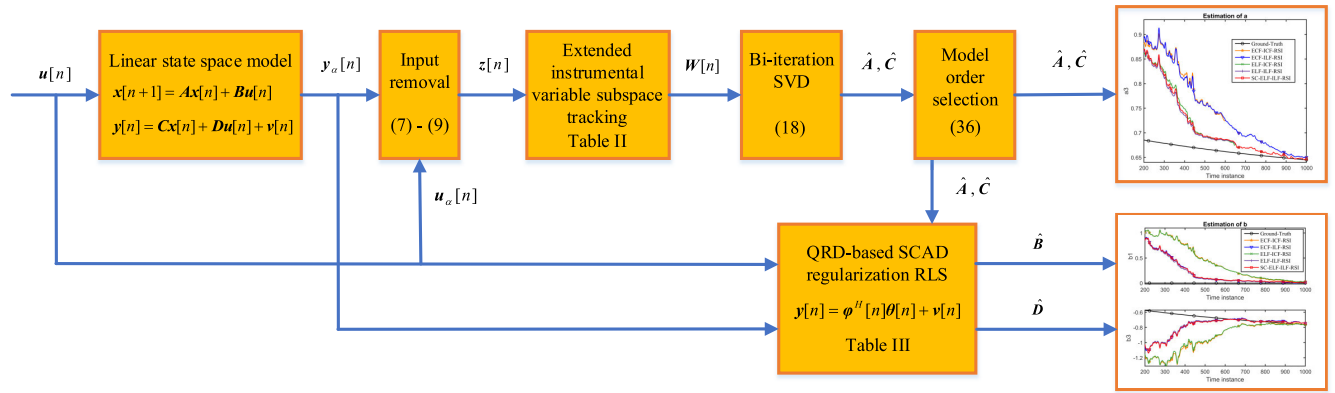


FIGURE 1. Recursive subspace model identification.

In summary, the novelties of this paper are i) the development of a new recursive RSI system identification algorithm with automatic order estimation, ii) the algorithm is based on efficient LPM based VFF EIV-PAST and multi-variate RLS algorithms for improving the tracking speed and steady state error, and they can be realized using efficient square-root form based on hyperbolic rotations and conventional QR decomposition respectively with improved its numerical stability, iii) a recursive bi-iteration singular value decomposition (SVD) algorithm is proposed for extraction of singular values of the state transition matrix for automatic order selection, iv) a new criterion based on percentage of explained variance is proposed to facilitate online estimation of the model order of A , and v) the SCAD regularization is incorporated for automatic model selection of the input matrices B and D as it is asymptotically unbiased. The regularized EIV algorithm also helps to stabilize the subspace especially during signal fading.

The rest of the paper is organized as follows. Section II will briefly review the system model, problem formulation and related works. In Section III, the proposed algorithm is introduced. Numerical results and comparison with the conventional algorithms are presented in Section IV. It suggests that the proposed algorithm has a better performance than conventional and the state-of-the-art algorithms. Finally, Section V concludes the paper.

II. PROBLEM FORMULATION

A. SYSTEM MODEL

Consider the following N -th length discrete-time linear time-invariant state-space model with M inputs and p outputs in innovation form

$$\mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{B}\mathbf{u}[n], \quad (1a)$$

$$\mathbf{y}[n] = \mathbf{C}\mathbf{x}[n] + \mathbf{D}\mathbf{u}[n] + \mathbf{v}[n] \quad (1b)$$

where $\mathbf{x}[n] \in \mathbf{R}^N$, $\mathbf{u}[n] \in \mathbf{R}^M$ and $\mathbf{y}[n] \in \mathbf{R}^p$ denote respectively the finite dimensional state vector, input vector and output vector at time instant n . The unknown state matrix \mathbf{A} , input matrix \mathbf{B} , output matrix \mathbf{C} and feedthrough matrix \mathbf{D} have appropriate dimensions. The noise term $\mathbf{v}[n]$ is

assumed to be a zero mean white process uncorrelated with the input $\mathbf{u}[n]$. The system model is summarized in Fig. 1.

For notation convenience, let us define the following stacked $p\alpha \times 1$ input, output and noise vectors

$$\mathbf{u}_\alpha = [\mathbf{u}^T[n], \dots, \mathbf{u}^T[n+\alpha-1]]^T, \quad (2a)$$

$$\mathbf{y}_\alpha = [\mathbf{y}^T[n], \dots, \mathbf{y}^T[n+\alpha-1]]^T, \quad (2b)$$

$$\mathbf{v}_\alpha = [\mathbf{v}^T[n], \dots, \mathbf{v}^T[n+\alpha-1]]^T, \quad (2c)$$

where α is a user-defined integer larger than the model order. Consequently, the input and output samples can be written in the following vector form called the “data equation” [28]

$$\mathbf{y}_\alpha[n] = \mathbf{\Gamma}_\alpha \mathbf{x}[n] + \mathbf{\Phi}_\alpha \mathbf{u}_\alpha[n] + \mathbf{v}_\alpha[n], \quad (3)$$

where $\mathbf{\Gamma}_\alpha$ is the extended observability matrix¹ defined as

$$\mathbf{\Gamma}_\alpha = [\mathbf{\Gamma}_0^T, \mathbf{\Gamma}_1^T, \dots, \mathbf{\Gamma}_{\alpha-1}^T]^T, \quad (4)$$

$\mathbf{\Gamma}_k = \mathbf{C}\mathbf{A}^k$, and $\mathbf{\Phi}_\alpha$ is a block lower triangular Toeplitz matrix consisting of the impulse response from the input to output given by

$$\mathbf{\Phi}_\alpha = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{CA}^{\alpha-2}\mathbf{B} & \dots & \mathbf{CB} & \mathbf{D} \end{bmatrix}. \quad (5)$$

It should be noted that $\mathbf{\Gamma}_\alpha$ spans an n -dimensional signal subspace, which contains the part of the output that is due to the state vector.

B. RECURSIVE ESTIMATION OF \mathbf{A} AND \mathbf{C}

One can notice from (4) that, the matrix \mathbf{C} can be obtained from the first $p \times n$ submatrix $\mathbf{\Gamma}_0$ of $\mathbf{\Gamma}_\alpha$. On the other hand, two consecutive submatrices $\mathbf{\Gamma}_{k+1}$ and $\mathbf{\Gamma}_k$ of $\mathbf{\Gamma}_\alpha$ differ only by \mathbf{A} on its right hand side: $\mathbf{\Gamma}_{k+1} = \mathbf{\Gamma}_k\mathbf{A}$. Therefore, given $\mathbf{\Gamma}_\alpha$, one can solve for \mathbf{A} as follows

$$\hat{\mathbf{A}}[n] = \mathbf{\Gamma}_a^\dagger \mathbf{\Gamma}_b \quad (6)$$

¹For notation simplicity, we have dropped the time dependent of the matrices $\mathbf{\Gamma}_k = \mathbf{C}\mathbf{A}^k$.

where $\mathbf{\Gamma}_a = [\mathbf{\Gamma}_0^T, \mathbf{\Gamma}_1^T, \dots, \mathbf{\Gamma}_{\alpha-2}^T]^T$, $\mathbf{\Gamma}_b = [\mathbf{\Gamma}_1^T, \mathbf{\Gamma}_2^T, \dots, \mathbf{\Gamma}_{\alpha-1}^T]^T$ and the superscript $(\cdot)^T$ denotes the Moore-Penrose pseudo-inverse.

Since, the state space matrix is uniquely identifiable up to an orthogonal transformation, it is sufficient to obtain a subspace that span the column space of $\mathbf{\Gamma}_\alpha$. This problem is very similar to the following formulation of the direction of arrival (DOA) problem in array signal processing [29]

$$\mathbf{z}[n] = \mathbf{\Gamma}[\theta]\mathbf{x}[n] + \mathbf{e}[n], \quad (7)$$

where $\mathbf{z}[n] \in \mathbf{C}^m$, $\mathbf{x}[n] \in \mathbf{C}^n$ and $\mathbf{e}[n] \in \mathbf{C}^m$ represent respectively the sensor signal vector from the m sensors, the n source signals impinging the array, and additive noise. $\mathbf{\Gamma}[\theta]$ is the $m \times n$ steering vector matrix with the j -th column containing the steering vector (array response) of the j -th source. By writing $\mathbf{z}[n] = \mathbf{y}_\alpha[n] - \Phi_\alpha \mathbf{u}_\alpha$, one can see that the two models are identical with $\mathbf{\Gamma}[\theta]$ equals to $\mathbf{\Gamma}_\alpha$ and $\mathbf{v}_\alpha[n]$ equals to $\mathbf{e}[n]$. This yields

$$\mathbf{z}[n] = \mathbf{\Gamma}_\alpha \mathbf{x}[n] + \mathbf{v}_\alpha[n], \quad (8)$$

where $\mathbf{z}[n] = \mathbf{y}_\alpha[n] - \Phi_\alpha \mathbf{u}_\alpha[n]$. Since Φ_α is not available at snapshot n , it can be approximated as $\Phi_\alpha[n-1]$ consisting of previously estimated $\mathbf{A}[n-1]$, $\mathbf{B}[n-1]$, $\mathbf{C}[n-1]$ and $\mathbf{D}[n-1]$. Another approach, which is called the past inputs (PI) scheme, is to make use of the classical Givens rotations to annihilate the input $\mathbf{u}_\alpha[n]$ from the output $\mathbf{y}_\alpha[n]$ via the RQ (or QR) factorization.

$$\begin{bmatrix} \mathbf{R}_{11}[n+1] & \mathbf{0} & \mathbf{0} \\ \mathbf{R}_{21}[n+1] & \mathbf{R}_{22}[n+1] & \mathbf{z}[n] \end{bmatrix} = \mathbf{G}[n] \begin{bmatrix} \mathbf{R}_{11}[n] & \mathbf{0} & \mathbf{u}_\alpha[n] \\ \mathbf{R}_{21}[n] & \mathbf{R}_{22}[n] & \mathbf{y}_\alpha[n] \end{bmatrix}, \quad (9)$$

where $\mathbf{G}[n]$ is an appropriate orthonormal rotor, $\mathbf{R}_{11} \in \mathbf{C}^{(M\alpha) \times (M\alpha)}$ and $\mathbf{R}_{22} \in \mathbf{C}^{(p\alpha) \times (p\alpha)}$ are upper right triangular matrix and $\mathbf{R}_{21} \in \mathbf{C}^{(p\alpha) \times (M\alpha)}$.

In array signal processing and related applications, very efficient subspace tracking algorithms [14], [22], [30]–[32] have been developed to recursively update the column subspace of $\mathbf{\Gamma}[\theta]$ for each input signal vector. Therefore, [33], [34] have proposed to estimate the column space of $\mathbf{\Gamma}_\alpha$ using the PAST subspace tracking algorithm and its extensions. To this end, one usually considers the following model for $\mathbf{z}[n]$

$$\mathbf{z}[n] = \mathbf{W}[n]\mathbf{s}[n] + \boldsymbol{\eta}[n], \quad (10)$$

where $\mathbf{W}[n] \in \mathbf{C}^{(p\alpha) \times K}$ is the subspace matrix with rank K , $\mathbf{s}[n] = \mathbf{W}^H[n]\mathbf{z}[n]$ is the projection of $\mathbf{z}[n]$ onto the subspace and $\boldsymbol{\eta}[n]$ is the additive noise. Most subspace tracking algorithms such as PAST aim to find a matrix $\mathbf{W}[n]$ which spans the signal subspace by minimizing the following LS error of the reconstruction $\mathbf{z}[n] - \mathbf{W}[n]\mathbf{s}[n] = \boldsymbol{\eta}[n]$ over time:

$$\mathbb{E}_{LS}(\mathbf{W}[n]) = \sum_{m=1}^n \lambda^{n-m} \|\mathbf{z}[m] - \mathbf{W}[n]\mathbf{g}[m]\|_2^2 \quad (11)$$

where $\lambda \in (0, 1]$ is a positive forgetting factor. Moreover, the projection approximation, $\mathbf{g}[m] = \mathbf{W}^H[n-1]\mathbf{z}[m]$, is

frequently used to simply the minimization of (11), since (11) can then be reduced to a conventional recursive least squares (RLS) problem in $\mathbf{W}[n]$. Consequently, $\mathbf{W}[n]$ can be used to estimate the column space of $\mathbf{\Gamma}_\alpha$.

The use of the LS criterion in (11) implicitly assumes that the additive noise $\boldsymbol{\eta}[n]$ is spatially white. However, in the context of system identification, the additive noise $\mathbf{v}_\alpha[n]$ may not be spatially white. Consequently, the estimate tends to be biased if the noise is not spatially white [33]. In [33], a novel subspace tracking algorithm based on the concept of instrumental variable (IV) is proposed. Instead of minimizing the LS error of $\boldsymbol{\eta}[n]$, an appropriate IV vector which is uncorrelated with $\mathbf{v}_\alpha[n]$ and highly correlated with $\mathbf{x}[n]$ is used to eliminate the effects of the additive noise. The selection of the instrumental variable is problem dependent. Usually, it is chosen as neighboring signal vector over time to ensure that the noise samples are uncorrelated.

Assuming that there exists such an IV vector $\boldsymbol{\xi}[n] \in \mathbf{C}^l$ with $l \geq n$ such that

$$(C1) \quad \mathbb{E}[\mathbf{v}_\alpha[n]\boldsymbol{\xi}^H[n]] = \mathbf{0},$$

$$(C2) \quad \text{Rank}(\mathbf{C}_{x\xi}) = \text{Rank}\{\mathbb{E}[\mathbf{x}[n]\boldsymbol{\xi}^H[n]]\} = N$$

one can then multiply both sides of (8) by $\boldsymbol{\xi}^H[n]$. After taking expectation, and using the above two assumptions, one gets

$$\begin{aligned} \mathbb{E}[\mathbf{z}[n]\boldsymbol{\xi}^H[n]] &= \mathbf{\Gamma}_\alpha \mathbb{E}[\mathbf{x}[n]\boldsymbol{\xi}^H[n]] + \mathbb{E}[\mathbf{v}_\alpha[n]\boldsymbol{\xi}^H[n]] \\ &= \mathbf{\Gamma}_\alpha \mathbb{E}[\mathbf{x}[n]\boldsymbol{\xi}^H[n]]. \end{aligned} \quad (12)$$

It can be seen that assumption (C1) ensures that the IV vector $\boldsymbol{\xi}[n]$ is uncorrelated with the noise vector $\mathbf{v}_\alpha[n]$ so that the noise term on the right hand side is eliminated. This mitigates the effect of noise (which may also be correlated with $\mathbf{x}[n]$) and gives rise to a consistent unbiased estimate even for spatially color disturbance. On the other hand, (C2) is required so that the rank of $\mathbf{\Gamma}_\alpha$ is preserved in $\mathbb{E}[\mathbf{z}[n]\boldsymbol{\xi}^H[n]]$. Therefore, the space spanned by the columns of $\mathbf{\Gamma}_\alpha \mathbf{C}_{x\xi}$ is equal to the columns of $\mathbf{\Gamma}_\alpha$, i.e. $\text{span}(\mathbf{\Gamma}_\alpha \mathbf{C}_{x\xi}) = \text{span}(\mathbf{\Gamma}_\alpha)$. If $\mathbf{x}[n]$ and $\boldsymbol{\xi}[n]$ have the same dimension, then $\mathbb{E}[\mathbf{x}[n]\boldsymbol{\xi}^H[n]]$ is nonsingular and we have $\mathbf{\Gamma}_\alpha = \mathbb{E}[\mathbf{z}[n]\boldsymbol{\xi}^H[n]]\mathbb{E}[\mathbf{x}[n]\boldsymbol{\xi}^H[n]]^{-1}$. Since $\mathbf{x}[n]$ is not assessable, we resort to its alternative form in (10) where $\mathbf{s}[n]$ is now a rotated form of $\mathbf{x}[n]$. In general, there are many ways to construct instrumental variables (IVs) in IV method. For instance, additional prefiltering may be performed to derive IV for system identification and other applications [35]. The optimal instruments and prefilters have been studied in [36]. It turns out that the optimal instruments and prefiltering will depend on the true system dynamics and the statistics properties of the noise. A multistep algorithm has been proposed in [36]. In this paper, we mainly focus on the temporal IV implementation, which is chosen as time delayed measurements. Moreover, to avoid further iterations and filtering over the temporal snapshots, we do not consider prefiltering the snapshots in exchange for a simple recursive implementation. Using the same approach above with the projection approximation

TABLE 1. The ECF-ICF-RSI algorithm [22].

Initialization:
Choose $\lambda = 0.99$, $P[0] = I_{K \times K}$ and $W[0] = I_{p\alpha \times K}$.
Recursion:
Given $P[n-1]$, $W[n-1]$, $z[n]$ and $\xi[n]$, compute at time n
$u[n] = \hat{C}_{\Xi G}^H[n-1]\xi[n]$,
$g[n] = W^H[n-1]z[n]$,
$\Psi[n] = \begin{bmatrix} u[n] & g[n] \end{bmatrix}$,
$q[n] = \begin{bmatrix} \xi^H[n]\hat{C}_{Z\Xi}[n-1] \\ z[n] \end{bmatrix}$,
$\Lambda[n] = \begin{bmatrix} -\xi^H[n]\xi[n] & \lambda \\ \lambda & 0 \end{bmatrix}$,
$K[n] = P[n-1]\Psi[n](\Lambda[n] + \Psi^H[n]P[n]\Psi[n])^{-1}$,
$\hat{C}_{Z\Xi}[n] = \lambda\hat{C}_{Z\Xi}[n-1] + z[n]\xi^H[n]$,
$\hat{C}_{\Xi G}[n] = \lambda\hat{C}_{\Xi G}[n-1] + \xi[n]g^H[n]$,
$P[n] = (P[n-1] - K[n]\Psi^H[n]P[n-1])/\lambda^2$,
$W[n] = W[n-1] + K[n](q[n] - \Psi^H[n]W[n-1])$.

$g[m] = W^H[n-1]z[m]$, one gets

$$\begin{aligned} C_{z\xi} &= \mathbb{E}[z[n]\xi^H[n]] \\ &= W \cdot \mathbb{E}[g[n]\xi^H[n]] + \mathbb{E}[\eta[n]\xi^H[n]] \\ &= W \cdot C_{hg}, \end{aligned} \quad (13)$$

and an estimate of is Γ_α where $W = C_{z\xi}[n]C_{g\xi}^{-1}$, $C_{z\xi} = \mathbb{E}[z[n]\xi^H[n]]$ and $C_{g\xi} = \mathbb{E}[g[n]\xi^H[n]]$. This is called the simple IV method. In the extended IV method, the size of $\xi[n]$ is longer than $x[n]$, which leads to an over-determined system. One can then resort to the LS solution, which can be obtained by multiplying both sides of (13) by $C_{g\xi}^H$ to obtain the following normal equation for solving W :

$$(C_{z\xi}C_{g\xi}^H) = W \cdot (C_{g\xi}C_{g\xi}^H). \quad (14)$$

Note that the system matrix $(C_{g\xi}C_{g\xi}^H)$ is symmetric and also nonsingular if (A2) is met under persistent excitation.

In practice, $C_{z\xi}$ and $C_{g\xi}$ need to be recursively estimated and tracked from data using an exponentially weighting scheme as follows

$$\begin{aligned} C_{Z\Xi}[n] &= Z^H[n]D_\lambda[n]\Xi[n] \\ &= \sum_{i=1}^t \lambda^{t-i} z[i]\xi^H[i] \\ &= \lambda C_{Z\Xi}[n-1] + z[n]\xi^H[n], \end{aligned} \quad (15a)$$

$$\begin{aligned} C_{\Xi G}[n] &= C_{G\Xi}^H[n] = \Xi^H[n]D_\lambda[n]G[n] \\ &= \sum_{i=1}^t \lambda^{t-i} \xi[i]g^H[i] \\ &= \lambda C_{\Xi G}[n-1] + \xi[n]g^H[n], \end{aligned} \quad (15b)$$

where $Z[n] = [z[n], z[n-1], \dots, z[1]]^H$, $D_\lambda[n] = \text{diag}(1, \lambda, \dots, \lambda^{t-1})$, $\Xi[n] = [\xi[n], \xi[n-1], \dots, \xi[1]]^H$, $G[n] = [g[n], g[n-1], \dots, g[1]]^H$, and $\lambda \in (0, 1]$ is a positive forgetting factor (FF).

For the simple IV method, the inverse of $C_{\Xi G}[n]$ can be recursively updated from (15a) using the matrix inversion lemma. Hence, the solution of W at time n , $W[n]$,

TABLE 2. The SR-ECF-ICF-RSI algorithm [38].

Initialization:
Choose $\lambda = 0.99$, $P[0] = I_{K \times K}$ and $W[0] = I_{p\alpha \times K}$.
Recursion:
Given $P[n-1]$, $W[n-1]$, $z[n]$ and $\xi[n]$, compute at time n
$u[n] = \hat{C}_{\Xi G}^H[n-1]ib[n]$,
$g[n] = W^H[n-1]z[n]$,
$\Psi[n] = \begin{bmatrix} u[n] & g[n] \end{bmatrix}$,
$q[n] = \begin{bmatrix} \xi^H[n]\hat{C}_{Z\Xi}[n-1] \\ z[n] \end{bmatrix}$,
$\alpha = (\xi^H[n]\xi[n])^{1/2}$,
$A = \begin{bmatrix} \alpha & 0 \\ -\lambda/\alpha & \lambda/\alpha \end{bmatrix}$,
Find Q such that
$R = \begin{bmatrix} -A & \Psi^H[n]P^{1/2}[n-1] \\ 0 & P^{1/2}[n-1] \end{bmatrix} Q = \begin{bmatrix} L_1 & 0 \\ M & L_2 \end{bmatrix}$,
$P^{1/2}[n] = \lambda^{-1}L_2$,
$K[n] = ML_1^{-1}$,
$\hat{C}_{Z\Xi}[n] = \lambda\hat{C}_{Z\Xi}[n-1] + z[n]\xi^H[n]$,
$\hat{C}_{\Xi G}[n] = \lambda\hat{C}_{\Xi G}[n-1] + \xi[n]g^H[n]$,
$W[n] = W[n-1] + K[n](q[n] - \Psi^H[n]W[n-1])$.

can be obtained as

$$W[n] = C_{Z\Xi}[n]C_{\Xi G}^{-H}[n]. \quad (16)$$

For the EIV method, (14) can be similarly solved using the RLS algorithm as shown in Table 1 [22], [33], [37]. However, since the conditioning of the system matrix $\hat{C}_{G\Xi}[n]\hat{C}_{\Xi G}[n]$ is deteriorated by the squaring operation, the accuracy of the resulting RLS solution may be compromised. We found that a recursive EIV algorithm based on the factorization of $\hat{C}_{G\Xi}[n]\hat{C}_{\Xi G}[n]$ and the Hyperbolic transformation was proposed in [38]. Due to the square root (SR) operation, the condition number of the system is considerably improved. Therefore, we have utilized this algorithm for estimating $W[n]$ in the proposed subspace-based system identification algorithm. The resultant algorithm based on the SREIV-PAST is summarized in Table 2. Interested readers are referred to [38] for the detailed derivations of the algorithm. Moreover, in Section III, we shall introduce a new VFF SREIV-PAST algorithm for estimating $W[n]$ in time-varying subspace, in which the FF is chosen to minimize the mean square error or deviation (MSD) of the estimator.

Similarly to (6), $A[n]$ and $C[n]$ can be extracted from $W[n]$ as follows

$$\hat{A}[n] = W_{1:p(\alpha-1)}^\dagger[n]W_{(p+1):p\alpha}[n], \quad (17a)$$

$$\hat{C}[n] = W_{1:p}[n]. \quad (17b)$$

where the subscript $(\cdot)_{1:p}$ represents the first row to the p -th row. Alternatively, the pseudo inverse $(\cdot)^\dagger$ can be recursively estimated with lower complexity by applying the bi-iteration singular value decomposition (SVD) [30] to the matrix $W_{1:p(\alpha-1)}$. More specifically, the iterations reads.

For $l = 1, 2, \dots$ until convergence iterate:

$$\begin{aligned} \mathbf{X}_B^{(l)} &= \mathbf{W}_{1:p(\alpha-1)} \mathbf{Q}_A^{(l)} \\ \mathbf{X}_B^{(l)} &= \mathbf{Q}_B^{(l)} \mathbf{R}_B^{(l)} \quad (\text{QR factorization}) \\ \mathbf{X}_A^{(l)} &= \mathbf{W}_{1:p(\alpha-1)}^H \mathbf{Q}_B^{(l)} \\ \mathbf{X}_A^{(l)} &= \mathbf{Q}_A^{(l+1)} \mathbf{R}_A^{(l)} \quad (\text{QR factorization}) \end{aligned} \quad (18)$$

In this recurrence, $\mathbf{Q}_A^{(l)}$ and $\mathbf{Q}_B^{(l)}$ will converge toward the N dominant left and right singular vectors of $\mathbf{W}_{1:p(\alpha-1)}$, respectively. Furthermore, both triangular matrices $\mathbf{R}_A^{(l)}$ and $\mathbf{R}_B^{(l)}$ will approach the $N \times N$ diagonal matrix composed of dominant singular values of $\mathbf{W}_{1:p(\alpha-1)}$. Instead of using the conventional identity matrix, the converged \mathbf{Q}_A at time instant n can be used as the initial value $\mathbf{Q}_A^{(0)}$ for the next time instant $n + 1$. Simulation results in section IV show that the number of iterations, can be significantly reduced over the conventional bi-iteration SVD algorithm. Usually, the iteration will converge in around five iterations.

(17a) can be recursively solved with the help of (18) using the multivariate complex RLS algorithm [32]. In Section III-C, we shall discuss how the model order of $\mathbf{A}[n]$ and $\mathbf{C}[n]$ can be recursively estimated.

C. RECURSIVE ESTIMATION OF \mathbf{B} AND \mathbf{D}

Given $\hat{\mathbf{A}}[n]$ and $\hat{\mathbf{C}}[n]$, the remaining system matrices \mathbf{B} and \mathbf{D} can be estimated from the following linear regression according to [39]

$$\mathbf{y}[n] = \boldsymbol{\phi}^H[n] \boldsymbol{\theta}[n] + \mathbf{v}[n], \quad (19)$$

where

$$\boldsymbol{\theta}[n] = \begin{bmatrix} \text{vec}^T(\mathbf{D}) & \text{vec}^T(\mathbf{B}) \end{bmatrix}^T, \quad (20a)$$

$$\boldsymbol{\phi}^H[n] = [\mathbf{u}^H[n] \otimes \mathbf{I}_p \quad (\sum_{s=1}^{t-1} \mathbf{u}^H[s] \otimes \hat{\mathbf{C}} \hat{\mathbf{A}}^{t-s-1})]. \quad (20b)$$

and $\text{vec}(\cdot)$ denotes the vectorization operator, \otimes is the Kronecker matrix product. The optimal weight vector therefore satisfies the following normal equation

$$\mathbf{R}_{\Phi\Phi}[n] \boldsymbol{\theta}_{opt}[n] = \mathbf{r}_{\Phi Y}[n], \quad (21)$$

where $\mathbf{R}_{\Phi\Phi}[n] = \sum_{i=1}^n \lambda^{n-i} \boldsymbol{\phi}[i] \boldsymbol{\phi}^H[i]$ and $\mathbf{r}_{\Phi Y}[n] = \sum_{i=1}^n \lambda^{n-i} \boldsymbol{\phi}[i] \mathbf{y}[i]$. Hence, $\hat{\boldsymbol{\theta}}[n]$ and hence $\hat{\mathbf{B}}[n]$ and $\hat{\mathbf{D}}[n]$ can be recursively updated by the multivariate RLS algorithm [40].

In this work, we shall utilize the complex VFF-RLS algorithm with multiple outputs in [32] to estimate the matrix $\hat{\mathbf{B}}[n]$ and $\hat{\mathbf{D}}[n]$. It is based on the QR decomposition and the single output VFF-RLS algorithm proposed in [26] with real-valued input. The algorithm is summarized in Table 3.

III. THE PROPOSED SC-VFF-SREIV AND -RLS RSI ALGORITHM

A. THE PROPOSED VFF-SREIV-PAST ALGORITHM

In the proposed VFF-SREIV-PAST algorithm for tracking the subspace of Γ_α , the FF is chosen to optimize the MSD of the estimator. Therefore, we need to obtain the mathematical expression of the MSD in terms of the FF used. More

precisely, we wish to analyze the MSD of the SREIV-PAST algorithm for a time-varying subspace with coefficients being modelled by a local polynomial [26], [32]. We shall first show that the MSD consists of a bias term which increases with the FF and a variance term which decreases with the FF. Therefore, it is possible to select an appropriate FF which minimizes the MSD. Then, we shall discuss the estimation of the quantities required for computing the FF. Finally, the SCAD regularization is incorporated to the VFF-SREIV-PAST algorithm to reduce the estimation variance of the subspace during signal fading.

1) LOCAL POLYNOMIAL MODELING

From (11), we can observe that the k -th element of $\mathbf{z}[m]$, $z_k[m]$, is equal to $\mathbf{w}_k^T[m] \mathbf{s}[m]$, where $\mathbf{w}_k[m]$ is the k -th row of $\mathbf{W}[n]$. To simplify the analysis and derivation, we assume that the underlying time varying vector \mathbf{h}_k that $\mathbf{w}_k[n]$ is tracking admits a local first order polynomial expansion at time index t_n as follows [32]

$$\mathbf{h}_k(t_m) = \mathbf{h}_k(t_n) + \frac{1}{1!} \mathbf{h}_k^{(1)}(t_n)(t_m - t_n) + \mathbf{r}_k(t_m - t_n) \quad (22)$$

where t_m belongs to an appropriately close neighborhood $B(t_n)$ of t_n , $\mathbf{h}_k^{(1)}(t_n)$ is the derivative of $\mathbf{h}_k(t)$ at time t_n and $\mathbf{r}_k(t_m - t_n)$ is the remainder. Both $\mathbf{h}_k^{(1)}(t_n)$ and $\mathbf{r}_k(t_m - t_n)$ can be considered as random vectors and the latter is of order $\mathcal{O}(t_m - t_n)$, which implies that $\|\mathbf{r}_k(t_m - t_n)(t_m - t_n)^{-1}\|_2$ tends to zero as $|t_m - t_n| \rightarrow 0$. For simplicity, we also assume that $\mathbf{h}_k^{(1)}(t_n)$ and $\mathbf{r}_k(t_m - t_n)$ are stationary processes inside $B(t_n)$.

Substituting (22) into (8) with $t_n = nT_s$, where T_s is the sampling period, the k -th component of the vector \mathbf{z} observed at the m -th sampling instant can be written as

$$\begin{aligned} z_k^*[m] &= \mathbf{h}^H[m] \mathbf{h}_k[m] + \eta_k[m] \\ &= \mathbf{h}^H[m] (\mathbf{h}_k[n] + \mathbf{h}_k^{(1)}[n](m-n)) + \eta_k[m] + \varsigma_k[m] \end{aligned} \quad (23)$$

where $\mathbf{h}_k[n] = \mathbf{h}_k(nT_s)$, $\mathbf{h}_k^{(1)}[n] = \mathbf{h}_k^{(1)}(nT_s)T_s$, and $\varsigma_k[m] = \mathbf{h}^H[m] \mathbf{r}_k[m-n]$ with $\mathbf{r}_k[n] = \mathbf{r}_k(nT_s)$. $\eta_k[m]$ denotes the additive noise which may arise from using the projection approximation or other sources.

Stacking the observed signal samples $z_k[m]$, $m = n, \dots, 1$, we obtain the following model for the observation signal vector

$$\mathbf{z}_k^*[n] = \mathbf{G}[n] \mathbf{h}_k[n] + \mathbf{D}_\tau[n] \mathbf{G}[n] \mathbf{h}_k^{(1)}[n] + \boldsymbol{\eta}_k[n] + \boldsymbol{\varsigma}_k[n], \quad (24)$$

where $\mathbf{D}_\tau[n] = \text{diag}(\boldsymbol{\tau}[n])$ with $\boldsymbol{\tau}[n] = [0, -1, \dots, -(n-1)]^T$, $\boldsymbol{\varsigma}_k[n] = [\varsigma_k[n], \varsigma_k[n-1], \dots, \varsigma_k[1]]^T$ is residual component vector due to higher order term, and $\boldsymbol{\eta}_k[n] = [\eta_k[n], \eta_k[n-1], \dots, \eta_k[1]]^T$ is the additive noise vector.

For simplicity of analysis, we assume that

- (A1) $\mathbf{r}_k[m-n]$, $\boldsymbol{\eta}_k[n]$ for $k = 1, \dots, K$, and $\mathbf{G}[n]$ are independent.
- (A2) $\mathbf{r}_k[m-n]$ and $\boldsymbol{\eta}_k[n]$ for $k = 1, \dots, K$, are zero mean with covariance respectively given by $\mathbb{E}[\boldsymbol{\eta}_k[n] \boldsymbol{\eta}_k^H[n]] = \sigma_{\eta_k}^2 \delta[n-m] \mathbf{I}$ and $\mathbb{E}[\mathbf{r}_k[n] \mathbf{r}_k^H[n]] =$

$\sigma_{r_k}^2 \delta[n - m] \mathbf{I}$, where $\delta[n]$ is the unit impulse function and \mathbf{I} is the identity matrix with appropriate dimension. (A3) $\mathbf{h}_k^{(1)}[n]$ has mean $\mathbb{E}[\mathbf{h}_k^{(1)}[n]] = \bar{\mathbf{h}}_k^{(1)}[n]$ and covariance $\mathbb{E}[\delta \mathbf{h}_k^{(1)}[n] \delta \mathbf{h}_k^{(1)H}[n]] = \sigma_{h_k}^2[n] \mathbf{I}$ where $\mathbf{h}_k^{(1)}[n] = \bar{\mathbf{h}}_k^{(1)}[n] + \delta \mathbf{h}_k^{(1)}[n]$.

For remarks regarding the motivation behind these assumptions, interested readers are referred to [32] and the supplementary materials for more details. From the simulation results to be presented in Section IV, we found that the proposed VFF-SREIV-PAST algorithm, which is derived from the above assumption, performs satisfactorily both in the synthetic as well as real data. This suggests that the above assumptions are reasonable.

To start with, we substituting (24) into (16) and obtain the following expression of the EIV solution under the LPM time-varying channel model:

$$\mathbf{w}_k[b] = \mathbf{h}_k[n] + \mathbf{B}[n] \mathbf{h}_k^{(1)}[n] + \mathbf{C}[n](\boldsymbol{\eta}_k[n] + \boldsymbol{\varsigma}_k[n]) \quad (25)$$

where $\mathbf{B}[n] = (\mathbf{C}_{\Xi G}^H[n] \mathbf{C}_{\Xi G}[n])^{-1} \mathbf{R}_\tau[n]$, $\mathbf{R}_\tau[n] = \mathbf{G}^H[n] \mathbf{D}_\lambda[n] \Xi[n] \Xi^H[n] \mathbf{D}_\lambda[n] \mathbf{D}_\tau[n] \mathbf{G}[n]$ and $\mathbf{C}[n] = (\mathbf{C}_{\Xi G}^H[n] \mathbf{C}_{\Xi G}[n])^{-1} \mathbf{G}^H[n] \mathbf{D}_\lambda[n] \Xi[n] \Xi^H[n] \mathbf{D}_\lambda[n]$. Moreover, we have used that fact that $(\mathbf{C}_{\Xi G}^H[n] \mathbf{C}_{\Xi G}[n])^{-1} \mathbf{G}^H[n] \mathbf{D}_\lambda[n] \Xi[n] \Xi^H[n] \mathbf{D}_\lambda[n] \mathbf{G}[n]$ is equal to the identity matrix.

2) VFF RECURSIVE EIV ALGORITHM FOR ESTIMATING \mathbf{A} AND \mathbf{C}

The mean squares deviation (MSD) of the estimator $\mathbf{W}[n]$ is equal to the sum of the MSE of all its components $\mathbf{w}_k[n]$. In Appendix, the latter is evaluated under the local polynomial time-varying model and hence the total MSD is given by (A.13) as follows

$$J_{MSD}[n] \approx \frac{1 - \lambda[n]}{1 + \lambda[n]} \sigma_\Sigma^2 \text{Tr}(\boldsymbol{\Omega}) + \frac{\lambda^2[n] \sigma_h^2[n]}{(1 - \lambda[n])^2}. \quad (26)$$

where $\boldsymbol{\Omega} = (\mathbf{C}_{\Xi G}^H \mathbf{C}_{\Xi G})^{-1} \mathbf{C}_{\Xi G}^H \mathbf{C}_{\Xi G} \mathbf{C}_{\Xi G}^H (\mathbf{C}_{\Xi G}^H \mathbf{C}_{\Xi G})^{-H}$, σ_Σ^2 and σ_h^2 denote the total error variance and the variance of the true parameter, respectively.

By minimizing $J_{MSD}[n]$ with respect to $\lambda[n]$, one can equate the partial derivative of (26) with respect to $\lambda[n]$ to zero. This yields

$$\frac{\partial J_{MSD}[n]}{\partial \lambda[n]} = -\frac{2\sigma_\Sigma^2 \text{Tr}(\boldsymbol{\Omega})}{(1 + \lambda[n])^2} + \frac{2\lambda[n] \sigma_h^2[n]}{(1 - \lambda[n])^3}. \quad (27)$$

which can be solved for the desired variable FF (VFF). To this end, we let $\mu = (1 + \lambda[n])/(1 - \lambda[n])$ so that (A.3) can be reduced to $\mu^2(\mu - 1) \approx 2\sigma_\Sigma^2[n] \text{Tr}(\boldsymbol{\Omega})/\sigma_h^2[n] \approx \mu^3$. Consequently, the VFF can be determined as follows

$$\lambda[n] = \frac{\mu - 1}{\mu + 1}, \quad \text{if } \lambda[n] > 0$$

$$\mu = (2\sigma_\Sigma^2[n] \text{Tr}(\boldsymbol{\Omega})/\sigma_h^2[n])^{1/3}. \quad (28)$$

Note that $\sigma_h^2 = K \mathbb{E}[\|\delta \mathbf{h}_k^{(1)}[n]\|_2^2] + \sum_{k=1}^K \bar{\mathbf{h}}_k^{(1)}[n] = \mathbb{E}[\sum_{k=1}^K \mathbf{h}_k^{(1)H}[n] \mathbf{h}_k^{(1)}[n]]$ and they can be recursively

estimated as

$$\hat{\sigma}_h^2[n+1] = \lambda[n] \hat{\sigma}_h^2[n] + (1 - \lambda[n]) \sum_{k=1}^K (\hat{\mathbf{h}}_k^{(1)H}[n] \hat{\mathbf{h}}_k^{(1)}[n]), \quad (29)$$

where $\hat{\mathbf{h}}_k^{(1)}[n]$ is an estimate of $\mathbf{h}_k^{(1)}[n]$. For instance, $\hat{\mathbf{h}}_k^{(1)}[n]$ can be estimated by the difference of the weight vector obtained at time instances n and $n - 1$ as $\hat{\mathbf{h}}_k^{(1)}[n] = \mathbf{w}_k[n] - \mathbf{w}_k[n - 1]$. Hence, the system variance σ_h^2 can be updated as

$$\hat{\sigma}_h^2[n+1] = \lambda[n] \hat{\sigma}_h^2[n] + (1 - \lambda[n]) \|\Delta \mathbf{W}[n]\|_F^2, \quad (30)$$

where $\Delta \mathbf{W}[n] = \mathbf{W}[n] - \mathbf{W}[n - 1]$. For the noise variance σ_Σ^2 , such information can be calculated from the error term $\mathbf{e}[n]$ by using a relatively large FF as follows

$$\hat{\sigma}_\Sigma^2[n+1] = \lambda_L \hat{\sigma}_\Sigma^2[n] + (1 - \lambda_L) \|\mathbf{e}[n]\|^2. \quad (31)$$

Due to page limitation, the estimation of $\boldsymbol{\Omega}$ and other details are described in the supplementary material.

Using the VFF, one can vary the FF according to (28) in the recursive EIV algorithm in Tables 1 and 2 for tracking the subspace for computing \mathbf{A} and \mathbf{C} . It should be noted that both the SREIV-PAST in Table 2 and the EIV square root propagator method (EIV sqrtPM) proposed in [41] algorithms are derived from the classical work of Porat and Friedlander [38]. The EIVsqrtPM algorithm extends the algorithm in [38] to the multivariable case using a fixed FF and the proposed SREIV-PAST algorithm focuses on the complex-valued case and the adaption and realization of the variable FF and regularization parameter. The complex-valued case is motivated by the fact that observations in some communication systems are complex-valued. As mentioned in [42], the propagator method and the SREIV-PAST algorithm differ in the exact subspace vectors they are tracked. However, the subspace estimated are fundamentally identical. Therefore, their performances are very close to each other. The SREIV-PAST with a constant FF will exhibit a performance very similar to the one of the complex extension of the EIVsqrtPM in [41]. Therefore, we only compare our proposed algorithm with the SREIV-PAST algorithm, which is almost identical to the complex EIVsqrtPM algorithm.

The convergence of the recursive RSI algorithm has been studied in [42]. Since we are using a VFF SREIV-PAST algorithm instead of the conventional PAST algorithm, we will study the convergence of the variable forgetting factor (VFF) EIV method in the supplementary material using the ordinary differential equation (ODE) method [40]. It is found that under reasonable assumptions and conditions (C1) and (C2), the algorithm converges to the true subspace in stationary environment if $\lambda[i] \in (0, 1]$ is monotonic increasing and approaches a value of one. Like the VFF-PAST algorithm in [32], the asymptotic MSE is minimized via the forgetting factor in (28). As the recursive EIV algorithm is convergent if $\mathbf{C}_{\Xi G}^H \mathbf{C}_{\Xi G}$ is nonsingular, σ_h^2 will decrease. Hence μ and $\lambda[n]$ will increase continuously. This satisfies the monotonic

condition above if $\lambda[n]$ is allowed to change smoothly so that the recursive EIV algorithm can converge gradually to a lower MSE. If σ_h^2 is equal to zero, we immediately obtain $\lambda[n] = 1$, which is the desired result. However, in practice, σ_h^2 is estimated online and it is unlikely to be zero and $\lambda[n]$ will reach a value close to one at convergence. This is observed in the simulation results to be presented in Figs. 4 and 5 where the forgetting factor increases monotonically and reaches a value close to 0.99 at time instant 400. As suggested in [32], if a high accuracy is required in a stationary environment, one can detect the convergence of the algorithm and push the forgetting factor further.

For time-varying subspaces, the MSE mainly depends on the subspace variation and the estimate may not converge to a fixed value. In the supplementary material, the stability of the proposed VFF algorithm is analyzed. As expected, if the system is persistently excited, i.e. $\mathbf{C}_{\xi g}^H \mathbf{C}_{\xi g}$ is positive definite, under finite noise and channel variances, then the MSD of the proposed VFF algorithm as $n \rightarrow \infty$ is bounded if $0 < \lambda[n] < 1$.

Next, we consider the estimation of \mathbf{B} and \mathbf{D} using the VFF multivariate RLS algorithm.

B. THE VFF RLS ALGORITHM FOR ESTIMATING \mathbf{B} AND \mathbf{D}

Since \mathbf{B} and \mathbf{D} can be estimated using the linear regression model in (19), it can be recursively estimated using the complex-valued LOFF RLS algorithm in [32]. Moreover, it can be implemented using the QRD as shown in step (vi) normal update of Table 3.

Moreover, the MSD under the LPM modeling of $\boldsymbol{\theta}[n]$ can be similarly obtained as follows

$$\tilde{J}_{MSD}(n) \approx \frac{1 - \tilde{\lambda}[n]}{1 + \tilde{\lambda}[n]} \tilde{\sigma}_{\Sigma}^2 \text{Tr}(\mathbf{R}_{\Phi\Phi}^{-1}) + \frac{\tilde{\lambda}^2[n] \sigma_{\theta}^2[n]}{(1 - \tilde{\lambda}[n])^2}. \quad (32)$$

TABLE 3. The QRD-based SC-ELF-ILF-RSI algorithm.

Initialization:
Choose $\lambda[0] = 0.99$, $\tilde{\lambda}[0] = 0.99$, $\mathbf{P}[0] = \mathbf{I}_{K \times K}$ and $\mathbf{W}[0] = \mathbf{I}_{p\alpha \times K}$, $\mathbf{R}[0] = \delta \mathbf{I}_{L \times L}$, δ is a small positive constant, $\mathbf{f}[0] = \mathbf{0}_{L \times 1}$.
Recursion:
Given $\mathbf{P}[n-1]$, $\mathbf{W}[n-1]$, $\mathbf{R}[n-1]$, $\mathbf{f}[n-1]$, $\mathbf{z}[n]$ and $\boldsymbol{\xi}[n]$
(i). Compute $\lambda[n]$ in (28),
(ii). The update for $\mathbf{W}[n]$ in Table II using $\lambda[n]$,
(iii). Extract $\hat{\mathbf{A}}[n]$ and $\hat{\mathbf{C}}[n]$ from $\mathbf{W}[n]$ using (17a) and (17b),
(iv) Compute $\tilde{\lambda}[n]$ in (33),
(v) Compute $\mathbf{K}[n]$ in (40),
(vi) The normal update for data
$\begin{bmatrix} \mathbf{R}^{(1)}[n] & \mathbf{f}^{(1)}[n] \\ \mathbf{0}[n] & \mathbf{c}^{(1)}[n] \end{bmatrix} = \mathbf{Q}^{(1)}[n] \begin{bmatrix} \sqrt{\lambda[n]} \mathbf{R}[n-1] & \sqrt{\lambda[n]} \mathbf{f}[n-1] \\ \boldsymbol{\phi}^H[n] & \mathbf{y}[n] \end{bmatrix},$
The regularization update for the SCAD regularization
$\begin{bmatrix} \mathbf{R}[n] & \mathbf{f}[n] \\ \mathbf{0} & \mathbf{c}[n] \end{bmatrix} = \mathbf{Q}[n] \begin{bmatrix} \mathbf{R}^{(1)}[n] & \mathbf{f}^{(1)}[n] \\ \sqrt{L\mathbf{K}_m[n]} & \mathbf{0} \end{bmatrix},$
where $\mathbf{Q}^{(1)}[n]$ and $\mathbf{Q}[n]$ are calculated by Givens rotation to obtain the left hand side of each equation above.
(vii) $\boldsymbol{\theta}[n] = \mathbf{R}^{-1}[n] \mathbf{f}[n]$ (back-substitution),
(viii) Extract $\hat{\mathbf{B}}[n]$ and $\hat{\mathbf{D}}[n]$ from $\boldsymbol{\theta}[n]$.

with the following LOFF

$$\tilde{\lambda}_{opt}[n] = \frac{\tilde{\mu} - 1}{\tilde{\mu} + 1}, \quad \text{if } \tilde{\lambda}_{opt}[n] > 0$$

$$\tilde{\mu} = (2\tilde{\sigma}_{\Sigma}^2[n] \text{Tr}(\mathbf{R}_{\Phi\Phi}^{-1}) / \sigma_{\theta}^2[n])^{1/3}. \quad (33)$$

where $\tilde{\sigma}_{\Sigma}^2[n]$ and $\sigma_{\theta}^2[n]$ can be estimated as $\tilde{\sigma}_{\Sigma}^2[n+1] = \lambda_L \tilde{\sigma}_{\Sigma}^2[n] + (1 - \lambda_L) \|\tilde{\mathbf{e}}\|^2$ and $\hat{\sigma}_{\theta}^2[n+1] = \lambda[n] \hat{\sigma}_{\theta}^2[n] + (1 - \lambda[n]) \Delta \boldsymbol{\theta}^H[n] \Delta \boldsymbol{\theta}[n]$ where $\tilde{\mathbf{e}}[n] = \mathbf{y} - \boldsymbol{\phi}^H[n] \boldsymbol{\theta}[n-1]$ and $\Delta \boldsymbol{\theta}[n] = \boldsymbol{\theta}[n] - \boldsymbol{\theta}[n-1]$.

C. MODEL ORDER ESTIMATION FOR THE PROPOSED LOFF-EIV RECURSIVE SUBSPACE IDENTIFICATION (RSI) ALGORITHMS

1) THE PROPOSED MODEL ORDER ESTIMATION FOR \mathbf{A}

With the LOFF-recursive EIV algorithm, we can estimate the subspace estimate $\mathbf{W}[n]$, and hence the matrices \mathbf{A} and \mathbf{C} . On the other hand, the matrices \mathbf{B} and \mathbf{D} can be estimated using the LOFF complex RLS algorithm.

In some real applications, one may wish to estimate the order of the matrix \mathbf{A} and hence \mathbf{C} . This can be gauged from the rank of $\boldsymbol{\Gamma}_{\alpha}$, which in turns is related to the number of major eigenvalue of $\mathbf{C}_{\xi\xi}$. Since we have already estimated the K dimension subspace $\mathbf{W}[n]$, one can further estimate recursively the eigenvalues associated with $\mathbf{W}[n]$. Consequently, we can retain the major eigenvalues and estimate the model order of \mathbf{A} . More precisely, the singular values of \mathbf{A} can be computed recursively via the bi-iteration SVD in (18) with $\boldsymbol{\Gamma}_{\alpha}$ replaced by \mathbf{A} , instead of the batch SVD algorithm which requires $\mathcal{O}(N^3)$ arithmetic complexity. The corresponding matrices $\mathbf{Q}_A^{(l)}$ and $\mathbf{Q}_B^{(l)}$ will then converge toward the N dominant left and right singular vectors of \mathbf{A} , respectively. Furthermore, both triangular matrices $\mathbf{R}_A^{(l)}$ and $\mathbf{R}_B^{(l)}$ will approach the $N \times N$ diagonal matrix composed of dominant singular values of \mathbf{A} . Usually, the diagonal values of $\mathbf{R}_A^{(l)}$ are chosen as the dominant singular values. Instead of using the conventional identity matrix as the initial value $\mathbf{Q}_A^{(0)}$, the converged \mathbf{Q}_A at time instant n can be used for $\mathbf{Q}_A^{(0)}$ in the next time instant $n+1$. Simulation results in section IV show that the number of iterations can be significantly reduced from five to one over the conventional bi-iteration SVD algorithm.

Classical model order estimation algorithm such as Akaike Information Criterion (AIC) and Minimum Description Length (MDL) usually assume that the minor singular values are equal. However, this assumption may not always be met in practice. Here, we shall investigate another hypothesis proposed in [43] which aims to find the model order such that the proportion of variance explained by the minor components is less than a certain value.

Let l_1, l_2, \dots, l_P be the eigenvalues of \mathbf{A} , and denote the following estimated proportion of variance in the first K principal signal subspace

$$\hat{\Psi}_K = (\sum_{k=1}^K \hat{l}_k) / (\sum_{p=1}^P \hat{l}_p) \quad (34)$$

where \hat{l}_k is the k -th estimated eigenvalue of \mathbf{A} in descending order of their magnitudes. It can be shown

that $\hat{\Psi}_k$ is asymptotically normal with mean $\Psi_K = (\sum_{k=1}^K l_k)/(\sum_{p=1}^P l_p)$ and variance

$$\tau_K^2 = \frac{2\text{Tr}(\mathbf{A}^2)}{(N-1)(\text{Tr}(\mathbf{A}))^2} [\Psi_K^2 - 2\alpha\Psi_K + \alpha] \quad (35)$$

where $\alpha = (\sum_{k=1}^K l_k^2)/(\sum_{p=1}^P l_p^2)$ and N denotes the effective sample number, say $N = 1/(1 - \lambda[n])$. Hence (35) may be used to derive the approximate confidence intervals for Ψ_K . Consequently, we can determine the minimum $K = 1, \dots, P$ such that the lower limit of the confidence interval

$$\left[\hat{\Psi}_K - \chi \sqrt{\tau_K^2/N}, \hat{\Psi}_K + \chi \sqrt{\tau_K^2/N} \right] \quad (36)$$

contains a specified proportion of the variance. For instance, the constant χ can be selected as 1.96 for 95% confidence.

2) AUTOMATIC MODEL ESTIMATION OF \mathbf{B} AND \mathbf{C} USING SCAD REGULARIZATION

For the matrices \mathbf{B} and \mathbf{C} , it is more conveniently to employ robust regularization to select automatically the model order online. Regularization is also valuable to mitigate possible ill-conditioning problem when the input is not persistently exciting when the input signal level is very low. Consequently, the covariance matrix $\mathbf{R}_{\Phi\Phi}$ may be ill-conditioned and it will result in a large estimation error variance or even divergence. To address this problem, we propose to utilize the SCAD regularization technique to avoid possible ill-conditioning and perform automatic model selection by shrinking small coefficients in \mathbf{B} and \mathbf{D} to zero. This will also reduce the variance of the estimates. Unlike L1 regularization, which is biased, robust regularization technique like SCAD is unbiased [27], [44]. The SCAD regularization can be imposed by adding the term $\sum_{i=1}^L f_\mu(\theta_i)$ to the LS objective function of the RLS algorithm

$$E_{SC}(\theta) = \sum_{i=1}^n \lambda_{n,i} (y[i] - \phi^H[i]\theta)^2 + \sum_{i=1}^L f_\mu(\theta_i), \quad (37)$$

where $\lambda_{n,i} = \lambda[n]\lambda_{n-1,i}$, $\lambda_{n,n} = 1$, $L = M(N + p)$,

$$f_\mu(\theta) = \begin{cases} \mu|\theta|, & \text{for } |\theta| \leq \mu \\ -\frac{(|\theta| - \tilde{a}\mu)^2}{2(\tilde{a} - 1)} + \frac{(\tilde{a} + 1)\mu^2}{2}, & \text{for } \mu < |\theta| \leq \tilde{a}\mu \\ (\tilde{a} + 1)\mu^2/2, & \text{for } |\theta| > \tilde{a}\mu \end{cases} \quad (38)$$

with μ is a non-negative regularization parameter and $\tilde{a} > 2$ (which is usually chosen as 3.7). It was suggested in [44] that the variable regularization parameter can be chosen as

$$\mu[n] = (1 - \tilde{\lambda}[n])L\tilde{\sigma}_\phi^2[n]\sqrt{\tilde{\sigma}_\Sigma^2[n]/(\text{Tr}(\mathbf{R}_{\Phi\Phi}) + \epsilon)M_1}, \quad (39)$$

where $\hat{\sigma}_\phi^2[n]$ is a long-term estimate of the input signal power, which can be estimated as $\hat{\sigma}_\phi^2[n] = (1/n) \sum_{i=1}^n \text{Tr}(\mathbf{R}_{\Phi\Phi}[i])$. ϵ is a small positive constant included to avoid ill-conditioning or division by zero when $\text{Tr}(\mathbf{R}_{\Phi\Phi}[n])$ is small. M_1 is the number of nonzero coefficients.

By setting its derivative of (37) with respect to θ to zero, we can obtain the equation for solving for the regularized parameter

$$(\tilde{\lambda}[n]\mathbf{R}_{\Phi\Phi}[n-1] + \phi[n]\phi^H[n] + \mathbf{K}[n])\theta_{opt}[n] = \mathbf{r}_{\Phi Y}[n] \quad (40)$$

where

$$\mathbf{K}[n] = \text{diag}(\kappa_1[n], \dots, \kappa_L[n]), \quad (41a)$$

$$\kappa_i[n] = f'_\mu(\theta_i[n-1])/\theta_i[n-1], \quad (41b)$$

$$f'_\mu(\theta) = \begin{cases} \text{sgn}(\theta)\mu, & \text{for } |\theta| \leq \mu \\ \frac{-\theta + \text{sgn}(\theta)\tilde{a}\mu}{\tilde{a} - 1}, & \text{for } \mu < |\theta| \leq \tilde{a}\mu \\ 0, & \text{for } |\theta| > \tilde{a}\mu, \end{cases} \quad (41c)$$

and $\text{sgn}(\theta)$ represents the sign function.

We note that the desired solution θ in (40) can be efficiently estimated using the QRD structure. The update of $\mathbf{R}_{\Phi\Phi}[n]$ involves two terms $\phi[n]\phi^H[n]$ and $\mathbf{K}[n]$. The first one term is a rank-one change, which can be achieved using conventional QRD. The second term is of full rank and hence its exact realization will require $\mathcal{O}(L^3)$ complexity. To reduce the complexity, we only update a row vector of $L\mathbf{K}_m[n]$ either randomly or sequentially with $\mathbf{K}_m[n]$ being the m -th row of $\mathbf{K}[n]$ and $n \bmod L = m$. Here, \bmod denotes the modulo operation. Consequently, the QRD operation is executed once for the data vector $[\phi^H[n], \mathbf{y}[n]]$ and once for the regularization term $[\sqrt{L}\mathbf{K}_m[n], \mathbf{0}]$ at each time instant. This yields the proposed SCAD regularized algorithm, as shown in Table 3.

3) ARITHMETIC COMPLEXITY COMPARISON

The arithmetic complexities of various algorithms are compared in Table 4. Here, notation convenience, we shall use ‘ECF’ and ‘ELF’ to denote the SREIV-PAST algorithm with constant and local optimal FF, respectively. ‘ICF’ and ‘ILF’ are used to denote respectively the recursive algorithm for estimating the input-related matrices \mathbf{B} and \mathbf{D} with constant FF and local optimal FF. Finally, ‘PI’ and ‘SC’ denote respectively the past-input instrumental variable and SCAD regularization schemes.

The complexities of the SREIV-PAST and multivariate RLS algorithms are respectively $\mathcal{O}(r_1^3) + 10r_1^2 + (4L_1 + 8N_1 + 9)r_1$ and $\mathcal{O}(r_2^2) + N_2r_2$, where $N_1 = p\alpha$, $r_1 = K$, $N_2 = (p + N)M$, $r_2 = p$ and L_1 denotes the dimension of the instrumental variable. The complexity of the PI scheme [34] is $\mathcal{O}(M^2) + 4N_1M$. The proposed LOFF for SREIV-PAST and multivariate RLS will require additional $(4N_1 + 6L_1)r_1^2 + (2N_1 - 2)r_1$ and $3N_2r_2 + 3N_2 + 7r_2$ operations, respectively. The SCAD regularization needs $N_2r_2 + N_2$ updates per time instance. Hence, the total arithmetic complexity of the proposed SC-ELF-ICL-RSI algorithm is $\mathcal{O}(r_1^3) + (4N_1 + 6L_1 + 10)r_1^2 + (4L_1 + 10N_1 + 7)r_1 + \mathcal{O}(r_2^2) + 5N_2r_2 + 4N_2 + 7r_2$. It can be noticed that the proposed algorithm has a comparable arithmetic complexity with those algorithms using a constant FF.

TABLE 4. Complexity comparison.

Algorithms	Arithmetic complexity
ECF-ICF-RSI [22]	$\mathcal{O}(r_1^3) + 10r_1^2 + (4L_1 + 8N_1 + 9)r_1 + \mathcal{O}(r_2^2) + N_2r_2$
PI-ECF-ICF-RSI [34]	$\mathcal{O}(r_1^3) + 10r_1^2 + (4L_1 + 8N_1 + 9)r_1 + \mathcal{O}(M^2) + 4N_1M + \mathcal{O}(r_2^2) + N_2r_2$
ECF-ILF-RSI	$\mathcal{O}(r_1^3) + 10r_1^2 + (4L_1 + 8N_1 + 9)r_1 + \mathcal{O}(r_2^2) + 4N_2r_2 + 3N_2 + 7r_2$
ELF-ICF-RSI	$\mathcal{O}(r_1^3) + (4N_1 + 6L_1 + 10)r_1^2 + (4L_1 + 10N_1 + 7)r_1 + \mathcal{O}(r_2^2) + N_2r_2$
ELF-ILF-RSI	$\mathcal{O}(r_1^3) + (4N_1 + 6L_1 + 10)r_1^2 + (4L_1 + 10N_1 + 7)r_1 + \mathcal{O}(r_2^2) + 4N_2r_2 + 3N_2 + 7r_2$
SC-ELF-ICL-RSI	$\mathcal{O}(r_1^3) + (4N_1 + 6L_1 + 10)r_1^2 + (4L_1 + 10N_1 + 7)r_1 + \mathcal{O}(r_2^2) + 5N_2r_2 + 4N_2 + 7r_2$

IV. NUMERICAL EXAMPLE

We now evaluate the performance of the proposed algorithms by comparing them with other conventional approaches using simulated experimental and real measurement data. Both stationary and dynamic environments are investigated. The SISO and MIMO systems tested are widely used in the literature for comparison [33], [34], [42]. For simplicity, we follow the notations for the proposed algorithms and other compared approaches in section III-3 and Table 4.

A. A SISO TIME-INVARIANT SYSTEM

We first present simulation results on the following SISO time invariant system studied in [33]:

$$y[n] = \frac{q^{-1} - 0.5q^{-2}}{1 - 1.5q^{-1} + 0.7q^{-2}}u[n] + \frac{1}{1 - 0.8q^{-1}}e[n], \quad (42)$$

where the numerator and denominator coefficient vectors are respectively and $a = [1 \ -1.5 \ 0.7]$, and $e[n]$ is an additive white Gaussian noise. The corresponding state-space model can be expressed as

$$\mathbf{x}[n+1] = \begin{bmatrix} 1.5 & -0.7 \\ 1 & 0 \end{bmatrix} \mathbf{x}[n] + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u[n], \quad (43a)$$

$$y[n] = [1 \ -0.5] \mathbf{x}[n] + [0] u[n] + v[n] \quad (43b)$$

The input $u[n]$ is taken as a Gaussian sequence with zero mean and unit variance. The proposed methods are compared with the RSI [33] and PI-SC-ELF-ILF-RSI algorithms at 17dB SNR. The constant forgetting factor for the ECF-based algorithms is set to 0.99. The minimum and maximum range of the forgetting factors of the proposed ELF-ICF-RSI and ECF-ILF-RSI algorithms are $\lambda_{min} = 0.97$ and $\lambda_{max} = 0.999$, respectively, while the short term and long term forgetting factors are set to $\lambda_S = 0.9$ and $\lambda_L = 0.99$, respectively. The order of the state space model is chosen as $n = 2$. All the results are obtained by averaging 100 independent trial runs.

Since the model order is unknown, we use our proposed approach to estimate the model order compared with the well-known criteria such as AIC and MDL. Different model orders of 2, 3 and 4 are used, and the result is shown in Table 5. It is shown that our proposed approach can give a consistent and accurate estimate over the AIC and MDL criteria for the tested systems.

TABLE 5. Model order estimation.

Given	AIC			MDL			Proposed		
	2	3	4	2	3	4	2	3	4
Estimate	1	2	3	1	2	3	2	2	2

The comparison of the proposed algorithms with other methods is shown in Fig. 2 and Fig. 3. It can be seen that the proposed ELF-based algorithms converge considerably faster than the RSI algorithm with a fixed FF for denominator coefficients estimation. The use of SCAD regularization also improves slightly the steady-state accuracy. The change of forgetting factor in the ELF-based algorithms is illustrated in Fig. 4. It can be seen the value of the FF gradually increases to a large value which reduces the steady state MSE. For estimating the numerator coefficients, the ILF-based algorithms converge much faster than the corresponding constant FF-based algorithms. It should be noted that the proposed variable FF can also speed up the convergence performance for PI-based algorithms. The change of the forgetting factor

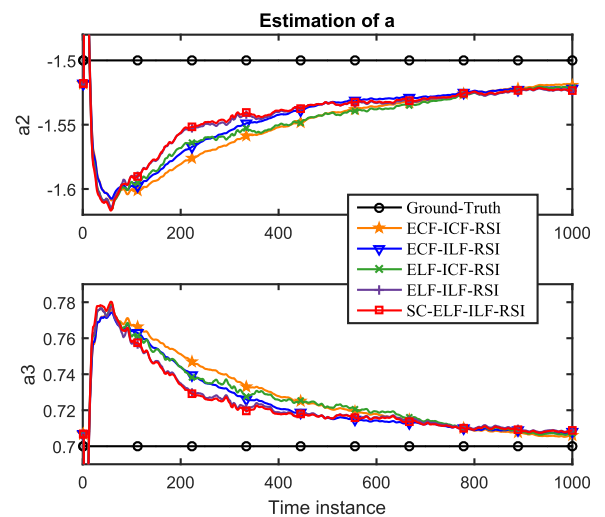


FIGURE 2. Estimated denominator coefficients using ECF-ICF-RSI in [22], and proposed ECF-ILF-RSI, ELF-ICF-RSI, ELF-ILF-RSI and SC-ELF-ILF-RSI algorithms under 17 dB SNR. 'ECF' and 'ELF' denote the SREIV-PAST algorithm with constant FF and local optimal FF, respectively. 'ICF' and 'ILF' represent the recursive algorithm to estimate the input-related matrices B and D with constant FF and local optimal FF. 'SC' is the abbreviation for SCAD regularization.

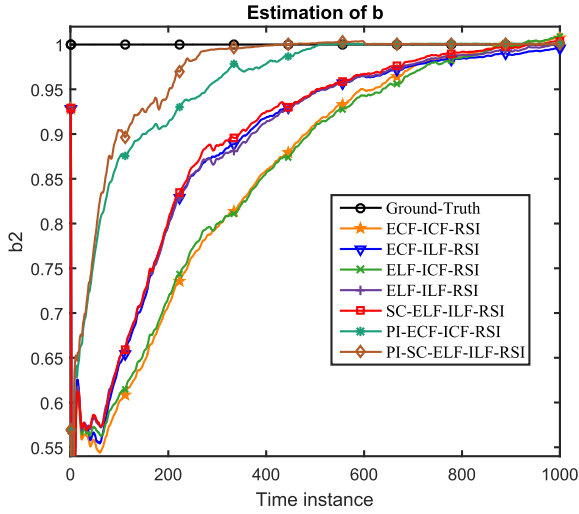


FIGURE 3. Estimated numerator coefficients using ECF-ICF-RSI in [22], and proposed ECF-ILF-RSI, ELF-ICF-RSI, ELF-ILF-RSI, SC-ELF-ILF-RSI, PI-ECF-ICF-RSI, PI-SC-ELF-ILF-RSI, algorithms under 17 dB SNR.

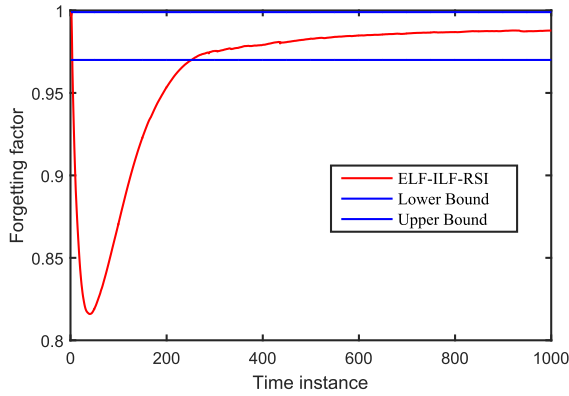


FIGURE 4. Change of the optimal forgetting factor in ELF-ILF-RSI algorithm.

in ILF-based algorithms is shown in Fig. 5. The small FF utilized in the initial stage can achieve faster convergence speed while the large FF used at the steady state yields a smaller MSE.

Next, we compare the steady state performance of the ECF-ICF-RSI, ELF-ICF-RSI, ECF-ILF-RSI and SC-ELF-ILF-RSI algorithms in stationary environment. The SNR is kept at 17dB. The Bode diagrams of different algorithms at the 1000-th time-point are shown in Fig. 6. It can be seen that the SC-ELF-ILF-RSI algorithm gives the smaller error from the ground-truth. The ELF-based algorithms are more accurate due to the more accurate estimation of the extended observability matrix Γ .

B. SISO SYSTEMS WITH TIME-VARYING A

Consider the following time-varying system studied in [15]

$$\mathbf{A}[n] = \begin{bmatrix} 1.5 + 0.05 \frac{e^{-n/1000} - 1}{e^{-1} - 1} & -0.7 \\ 1 & 0 - 0.03 \frac{e^{-n/1000} - 1}{e^{-1} - 1} \end{bmatrix} \quad (44)$$

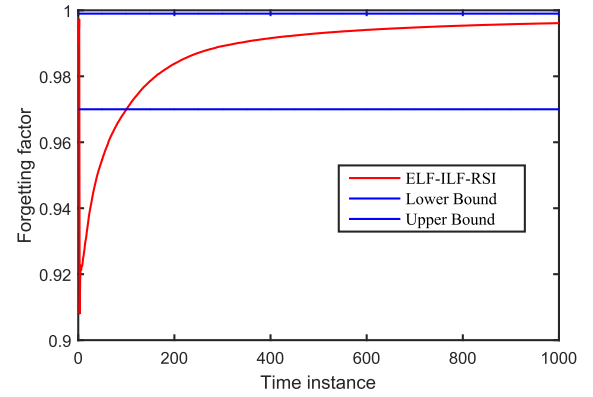


FIGURE 5. Change of the optimal forgetting factor in ELF-ILF-RSI algorithm.

All the other settings are identical to the last example. The following colored measurement noise was added to the noise-free output of the system

$$v[n] = \frac{1}{1 - 0.8q^{-1}} e[n], \quad (45)$$

where $e[n]$ is a zero mean white Gaussian noise. The output SNR was set to 25 dB.

The criterion chosen to measure and to compare the performance was subspace angle [45]–[47] in degree between Γ and $\hat{\Gamma}$ as

$$\begin{aligned} \cos(\theta_n) &= \max_{\mathbf{u} \in \Gamma} \max_{\mathbf{v} \in \hat{\Gamma}} \mathbf{u}^T \mathbf{v} = \mathbf{u}_n^T \mathbf{v}_n, \\ \text{s.t. } \|\mathbf{u}\|_2^2 &= \|\mathbf{v}\|_2^2 = 1, \quad \mathbf{u}_i^T \mathbf{u}_i = 0, \\ \mathbf{v}_i^T \mathbf{v}_i &= 0, \quad i = 1, 2, \dots, n-1 \end{aligned} \quad (46)$$

The motivation of the criterion is that the largest subspace angle is related to the distance of equidimensional subspaces. Fig. 7 compares the subspace angle of different algorithms in nonstationary environment. It can be seen that the SC-ELF-ILF-RSI algorithm achieves a faster initial convergence speed and similar steady state error as the ECF-ICF-RSI algorithm. The ELF-ICF-RSI algorithm performs similar

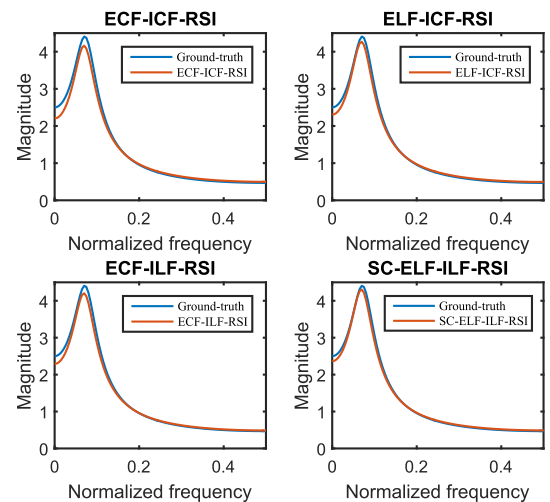


FIGURE 6. Bode diagrams of different algorithms at 1000-th time-point.

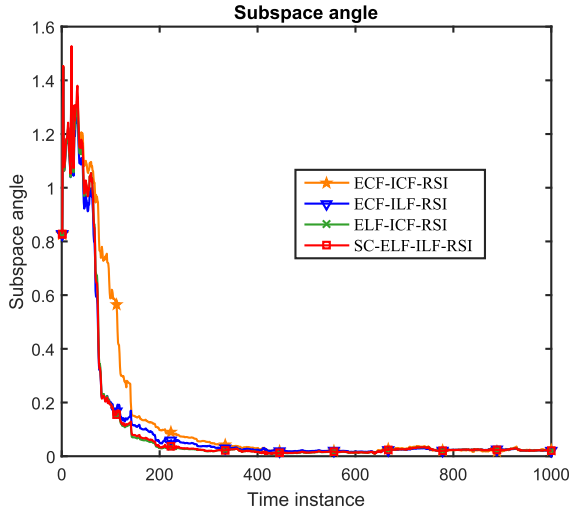


FIGURE 7. Estimated subspace angle using ECF-ICF-RSI in [22], and proposed ECF-ILF-RSI, ELF-ICF-RSI and SC-ELF-ILF-RSI algorithms under 25 dB SNR in nonstationary environment.

to the SC-ELF-ILF-RSI algorithm as a result of the variable forgetting factor employed to calculate the extended observability matrix Γ . The ECF-ILF-RSI algorithm converges a little slower than the SC-ELF-ILF-RSI algorithm, which is still faster than the ECF-ICF-RSI algorithm due to the variable forgetting factor used to compute the impulse response matrix Φ from input to output.

To better illustrate the advantage of the proposed VFF algorithms, we also consider a system with poles closer to the origin similar to the last system:

$$\mathbf{A}[n] = \begin{bmatrix} 1 + 0.1 \frac{e^{-n/1000} - 1}{e^{-1} - 1} & -0.7 \\ 1 & 0 - 0.05 \frac{e^{-n/1000} - 1}{e^{-1} - 1} \end{bmatrix} \quad (47)$$

Other settings are identical to the previous simulation. The subspace angles of different algorithms are shown in Fig. 8. It can be seen that the SC-ELF-ILF-RSI and ELF-ICF-RSI algorithms perform significantly better than the ECF-ILF-RSI and ECF-ICF-RSI algorithms. Next, we compare the tracking capability of the proposed algorithm with other methods in Fig. 9. It can be seen that SC-ELF-ILF-RSI algorithm can track the varying coefficients with the fastest speed. The ELF-ICF-RSI algorithm performs comparably with the ELF-ILF-RSI algorithm, while both of them converge faster than the ECF-ILF-RSI and ECF-ICF-RSI methods.

C. A SISO SYSTEM WITH TIME-VARYING \mathbf{B}

Next, we consider a system similar to the last time-varying system with

$$\mathbf{B}[n] = \begin{bmatrix} 1 + 0.6 \frac{e^{-n/1000} - 1}{e^{-1} - 1} & 0 - 0.2 \frac{e^{-n/1000} - 1}{e^{-1} - 1} \end{bmatrix}. \quad (48)$$

All the other settings are identical to the last example while \mathbf{A} is stationary as the first example. The tracking performances of the numerator coefficients of various algorithms are shown in Fig. 10. Both ELF-ICF-RSI and ECF-ICF-RSI algorithms

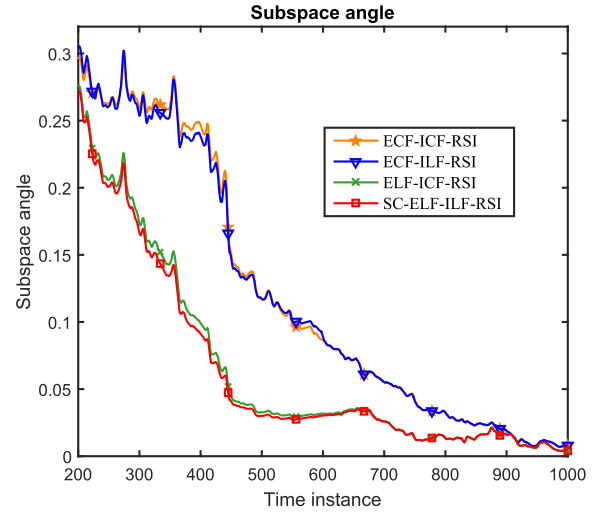


FIGURE 8. Estimated subspace angle using ECF-ICF-RSI in [22], and proposed ECF-ILF-RSI, ELF-ICF-RSI and SC-ELF-ILF-RSI algorithms under 25 dB SNR in nonstationary environment with poles closer to the origin.

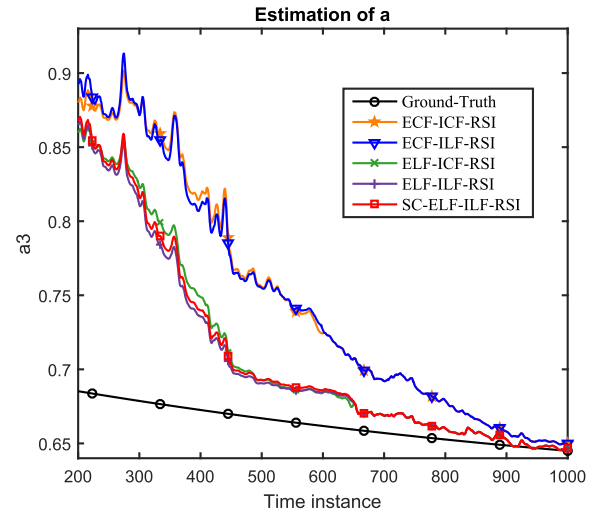


FIGURE 9. Estimated denominator coefficients using ECF-ICF-RSI in [22], and proposed ECF-ILF-RSI, ELF-ICF-RSI, ELF-ILF-RSI and SC-ELF-ILF-RSI algorithms under 25 dB SNR in nonstationary environment.

are capable of tracking the fast changing parameters, but with a slower tracking speed than the ILF-based algorithms in which local optimal FF is employed to estimate \mathbf{B} and \mathbf{D} .

D. A MIMO TIME-INVARIANT SYSTEM

Here, we consider an $N = 4$ MIMO time-invariant system studied in [48]

$$\mathbf{A} = \begin{bmatrix} 0.7 & 0.7 & 0 & 0 \\ -0.7 & 0.7 & 0 & 0 \\ 0 & 0 & -0.7 & -0.7 \\ 0 & 0 & 0.7 & -0.7 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1.1650 & -0.6965 \\ 0.6268 & 1.6961 \\ 0.0751 & 0.0591 \\ 0.3561 & 1.7971 \end{bmatrix} \quad (49a)$$

$$\mathbf{C} = \begin{bmatrix} 0.2641 & -1.4462 & 1.2460 & 0.5774 \\ 0.8717 & -0.7012 & -0.6390 & -0.3600 \end{bmatrix}, \quad (49b)$$

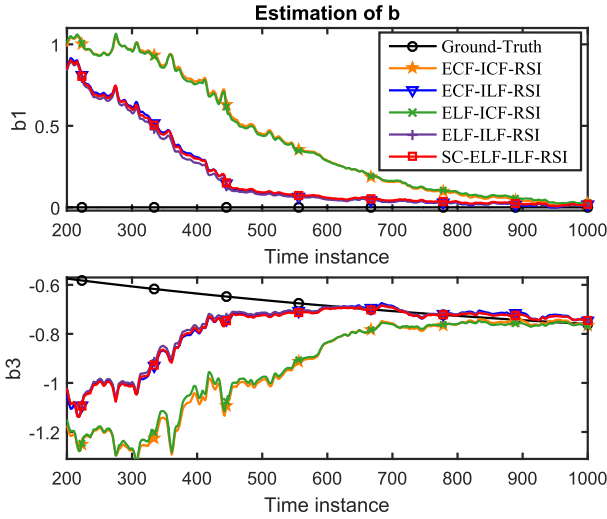


FIGURE 10. Estimated numerator coefficients using ECF-ICF-RSI in [22], and proposed ECF-ILF-RSI, ELF-ICF-RSI, ELF-ILF-RSI and SC-ELF-ILF-RSI algorithms under 25 dB SNR in nonstationary environment.

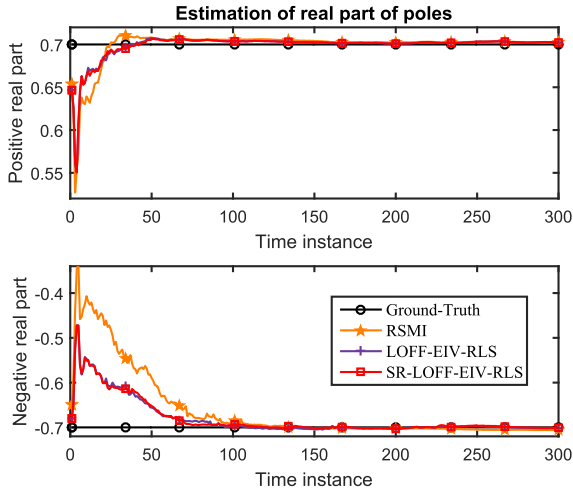


FIGURE 11. Estimated real part of poles using ECF-ICF-RSI in [22], and proposed ELF-ICF-RSI and SC-ELF-ILF-RSI algorithms in a MIMO system.

$$D = \begin{bmatrix} -0.1356 & -1.2704 \\ -1.3493 & 0.9846 \end{bmatrix} \quad (49c)$$

The measurement and process noises are respectively $Kv[n]$ and $v[n]$, where

$$K = \begin{bmatrix} 0.4968 & -0.3580 \\ -0.3312 & -0.0512 \\ 0.1560 & -0.3872 \\ -0.0900 & 0.5836 \end{bmatrix}, \quad R_v = \begin{bmatrix} 0.0176 & -0.0267 \\ -0.0267 & 0.0497 \end{bmatrix}. \quad (50)$$

$v[n]$ is a zero mean white Gaussian noise with covarianc R_v . Fig. 11 and Fig. 12 show the trajectories of the real and imaginary parts of the estimated poles using various algorithms, respectively. It can be seen from Fig. 11 that the ELF-ILF-RSI based algorithms converge much faster than

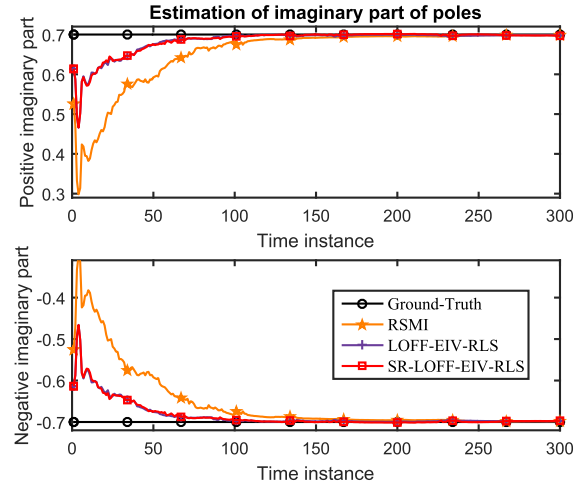


FIGURE 12. Estimated imaginary part of poles using ECF-ICF-RSI in [22], and proposed ELF-ICF-RSI and SC-ELF-ILF-RSI algorithms in a MIMO system.

TABLE 6. Numerical comparison for the flutter system.

α	7	8	9	10
ECF-ICF-RSI in [22]	80.11	80.36	79.52	78.92
ECF-ILF-RSI	82.69	82.84	81.75	80.11
ELF-ICF-RSI	83.04	83.32	82.56	80.87
ELF-ILF-RSI	86.26	86.64	85.13	84.08
SC-ELF-ILF-RSI	86.12	86.27	84.96	84.05

the ECF-ICF-RSI algorithm in estimating the real part of the poles. The imaginary parts of the poles estimated also show the same characteristic in Fig. 12. The ELF-ILF-RSI algorithm performs comparably with the SC-ELF-ILF-RSI algorithm in terms of convergence speed.

Next, we compare the accuracy of the estimated poles obtained by the ECF-ICF-RSI, ELF-ILF-RSI and SC-ELF-ILF-RSI algorithms. The eigenvalues of the estimated A matrices in 100 Monte-Carlo simulations are plotted and compared with the true ones. 300 samples are used for each realization. The estimated poles obtained by the ECF-ICF-RSI, ELF-ILF-RSI and SC-ELF-ILF-RSI algorithms are presented respectively in Fig. 13, Fig. 14 and Fig. 15. It can be noted that the mean of the eigenvalue estimates are equally accurate for these algorithms. However, the variance of the eigenvalue estimates of the ECF-ICF-RSI algorithm is slightly larger than the ELF-ILF-RSI algorithm. The eigenvalue estimates obtained by the SC-ELF-ILF-RSI algorithm are more close to the true ones due to the effect of SCAD regularization.

E. APPLICATION TO A WING FLUTTER DATA

In this section, the proposed algorithms are tested using the wing flutter data [49] from the DaISy (Database for the Identification of Systems). The data consists of 1024 time

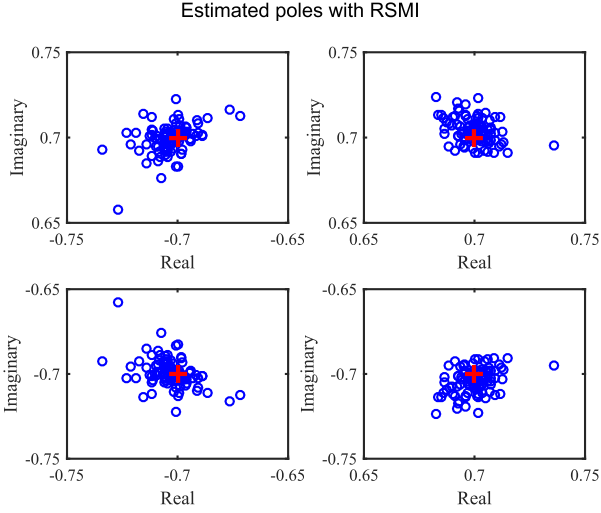


FIGURE 13. Estimated poles using ECF-ICF-RSI algorithm in [22] for a Monte Carlo simulation of size 100.

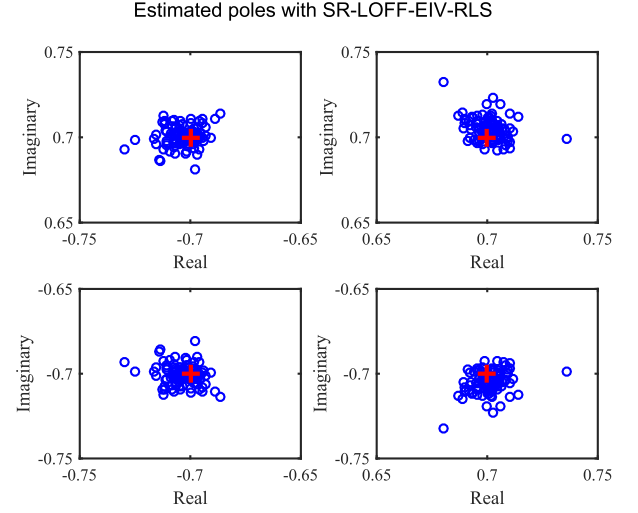


FIGURE 15. Estimated poles using SC-ELF-ILF-RSI algorithm for a Monte Carlo simulation of size 100.

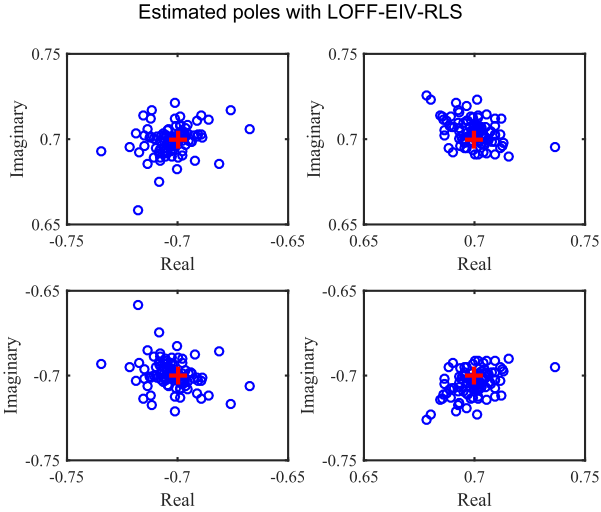


FIGURE 14. Estimated poles using ELF-ILF-RSI algorithm for a Monte Carlo simulation of size 100.

points with single input and single output. The input signal is highly colored and the response is acquired via an accelerometer sensor in real time. The mathematical model of the mechanical system is unknown. Therefore, we use our proposed approach and the model order estimated is equal to 5. We tested different numbers of block rows with $\alpha = 7, 8, 9, 10$ and the result is summarized in Table 6. The following commonly used FIT criterion [50], [51] is used to compare the performance of various algorithms

$$\text{FIT} = 100 \times \left(1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|} \right), \quad (51)$$

where y is the measured system output, \hat{y} and \bar{y} denote the estimated output of the system and the arithmetic mean of y , respectively. Note that $\|y - \hat{y}\| = \sqrt{\text{SSE}}$ and $\|y - \bar{y}\| =$

$\sqrt{\text{SST}}$, where

$$\text{SSE} = \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (52)$$

is the sum of squared errors and

$$\text{SST} = \sum_{i=1}^N (y_i - \bar{y})^2, \quad (53)$$

is the total sum of squares deviation or total variation.

It can be seen that $\alpha = 8$ gives the best results for all algorithms. The proposed ELF-ILF-RSI and SC-ELF-ILF-RSI algorithms can achieve high FIT values and outperform other compared methods. The accuracy of ELF-ICF-RSI algorithm is slightly better than the ECF-ILF-RSI algorithm. The prediction accuracy obtained with the ECF-ICF-RSI algorithm is much lower due to the constant forgetting factor.

V. CONCLUSION

A new SCAD regularized LOFF-based algorithm for recursive subspace model identification has been presented. The VFF parameter is obtained by minimizing the MSE of the SREIV-PAST algorithm when the observability matrix is modeled by local polynomial modeling. A SCAD regularization term is also incorporated to reduce the estimation variance of the subspace during signal fading. Applications of the proposed algorithms under stationary and non-stationary environments show that the proposed algorithms offer considerably improved performance over the conventional and state-of-the-art algorithms.

APPENDIX

MEAN SQUARES DEVIATION (MSD) OF THE ESTIMATOR

As mentioned, our aim is to find the FF so as to minimize the MSD of the estimator. Therefore, we need to evaluate the MSD of the estimator, which can be written as

$$J_{\text{MSD}}(\mathbf{w}_k) = \mathbb{E} \left[\|\mathbf{w}_k[n] - \mathbf{h}_k[n]\|_2^2 \right]. \quad (\text{A.1})$$

To this end, we first rewrite the estimation error as

$$\mathbf{w}_k[n] - \mathbf{h}_k[n] = \mathbf{b}_k[n] + \mathbf{v}_k[n] \quad (\text{A.2})$$

where $\mathbf{b}_k[n] = \mathbb{E}[\mathbf{w}_k[n]] - \mathbf{h}_k[n]$ is the bias and $\mathbf{v}_k[n] = \mathbf{w}_k[n] - \mathbb{E}[\mathbf{w}_k[n]]$ is associated with the variance. Consequently

$$J_{MSD}(\mathbf{w}_k) = \mathbb{E}[\|\mathbf{b}_k[n]\|_2^2] + \mathbb{E}[\|\mathbf{v}_k[n]\|_2^2]. \quad (\text{A.3})$$

We now evaluate $\mathbb{E}[\mathbf{w}_k[n]]$ and hence $\mathbf{b}_k[n]$ and $\mathbf{v}_k[n]$ from which the right hand side of (A.3) can be evaluated. From (25), we get the expected value of the estimate as

$$\mathbb{E}[\mathbf{w}_k[n]] = \mathbf{h}_k[n] + \mathbf{B}[n]\bar{\mathbf{h}}_k^{(1)}[n], \quad (\text{A.4})$$

where $\bar{\mathbf{h}}_k^{(1)}[n] = \mathbb{E}[\mathbf{h}_k^{(1)}[n]]$. Moreover, following the technique in [52], we have $(\mathbf{C}_{\Xi G}^H[n]\mathbf{C}_{\Xi G}[n])^{-1} \xrightarrow{n \rightarrow \infty} (1 - \lambda[n])^2(\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g})^{-1}$ where $\mathbf{C}_{\Xi G}[n] \approx (1 - \lambda[n])^{-1}\mathbf{C}_{\xi g}$ and $\mathbf{C}_{\xi g} = \mathbb{E}[\xi[n]\mathbf{g}^H[n]]$. For simplicity, let us define $\mathbf{R}_\tau[n] = \mathbf{G}^H[n]\mathbf{D}_\lambda[n]\Xi[n]\Xi^H[n]\mathbf{D}_\lambda[n]\mathbf{D}_\tau[n]\mathbf{G}[n]$. As $n \rightarrow \infty$,

$$\begin{aligned} \mathbf{R}_\tau[n] &\xrightarrow{n \rightarrow \infty} \mathbf{G}^H[n]\mathbf{D}_\lambda[n]\Xi[n]\Xi^H[n]\mathbf{D}_\lambda[n]\mathbf{D}_\tau[n]\mathbf{G}[n] \\ &\xrightarrow{n \rightarrow \infty} -(1 - \lambda[n])^{-1}\mathbf{C}_{\xi g}^H \sum_{i=1}^n \lambda^{n-i}[n]\xi[i]\mathbf{g}^H[i](n-i) \\ &\approx -(1 - \lambda[n])^{-1}\mathbf{C}_{\xi g}^H \sum_{i=0}^{n-1} i\lambda^i[n]\mathbf{C}_{yy} \\ &\xrightarrow{n \rightarrow \infty} -(1 - \lambda[n])^{-1}\lambda[n] \\ &\quad \times \left(\frac{1 - \lambda^{n-1}[n]}{(1 - \lambda[n])^2} - \frac{(n-1)\lambda^{n-1}[n]}{1 - \lambda[n]} \right) \mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g} \\ &\approx -\frac{\lambda[n]}{(1 - \lambda[n])^3} \mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g}. \end{aligned} \quad (\text{A.5})$$

Using (A.5) and the above results for $(\mathbf{C}_{\Xi G}^H[n]\mathbf{C}_{\Xi G}[n])^{-1}$, we have $\mathbf{B}[n] = (\mathbf{C}_{\Xi G}^H[n]\mathbf{C}_{\Xi G}[n])^{-1}\mathbf{R}_\tau[n] \xrightarrow{n \rightarrow \infty} -\frac{\lambda[n]}{1 - \lambda[n]}$ and hence

$$\mathbf{b}_k[n] = \mathbf{B}[n]\bar{\mathbf{h}}_k^{(1)}[n] \xrightarrow{n \rightarrow \infty} -\frac{\lambda[n]}{1 - \lambda[n]}\bar{\mathbf{h}}_k^{(1)}[n]. \quad (\text{A.6})$$

Note that the FF $\lambda[n]$ is slowly varying over time. However, it can be assumed to be constant inside the neighborhood $B(t_n)$, which is assumed to be the interval under consideration $i = 1, \dots, n$. From (A.7), we finally get

$$\mathbb{E}[\|\mathbf{b}_k[n]\|_2^2] = \frac{\lambda^2[n]}{(1 - \lambda[n])^2} \zeta_k, \quad (\text{A.7})$$

where $\zeta_k = \bar{\mathbf{h}}_k^{(1)H}[n]\bar{\mathbf{h}}_k^{(1)}[n]$. Similarly, we have

$$\mathbf{v}_k[n] = \mathbf{C}[n](\eta_k[n] + \zeta_k[n]) + \mathbf{B}[n]\delta\mathbf{h}_k^{(1)}[n], \quad (\text{A.8})$$

where $\mathbf{h}_k^{(1)}[n]\bar{\mathbf{h}}_k^{(1)}[n] + \delta\mathbf{h}_k^{(1)}[n]$ and

$$\begin{aligned} \mathbb{E}[\|\mathbf{v}_k[n]\|_2^2] &= \mathbb{E}[\text{Tr}(\mathbf{C}^H[n]\mathbf{C}[n]\eta_k[n]\eta_k^H[n])] \\ &\quad + \mathbb{E}[\text{Tr}(\mathbf{C}^H[n]\mathbf{C}[n]\zeta_k[n]\zeta_k^H[n])] \\ &\quad + \mathbb{E}[\text{Tr}(\mathbf{B}[n]\delta\mathbf{h}_k^{(1)}\delta\mathbf{h}_k^{(1)H}[n]\mathbf{B}[n])] \end{aligned} \quad (\text{A.9})$$

Moreover, using assumptions (A1) to (A3), (A.9) can be simplified to: (A.9) can be simplified to

$$\begin{aligned} \mathbb{E}[\|\mathbf{v}_k[n]\|_2^2] &= \frac{K\sigma_{\delta h_k}^2[n]\lambda^2[n]}{(1 - \lambda[n])^2} \\ &\quad + \frac{1 - \lambda[n]}{1 + \lambda[n]}(\sigma_{\eta_k}^2 + \sigma_{r_k}^2 \text{Tr}(\mathbf{C}_{\xi\xi}))\text{Tr}(\mathbf{\Omega}), \end{aligned} \quad (\text{A.10})$$

where $\sigma_{\eta_k}^2 = \frac{1}{K}\mathbb{E}[\|\eta_k[n]\|_2^2]$, $\sigma_{r_k}^2 = \frac{1}{K}\mathbb{E}[\|r_k[n]\|_2^2]$, $\sigma_{\delta h_k}^2 = \frac{1}{K}\mathbb{E}[\|\delta\mathbf{h}_k[n]\|_2^2]$ and $\mathbf{\Omega} = (\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g})^{-1}\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi\xi}\mathbf{C}_{\xi g}(\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g})^{-H}$. Here, we have also used the following:

$$\begin{aligned} &\mathbb{E}[\text{Tr}(\mathbf{C}^H[n]\mathbf{C}[n])] \\ &= \text{Tr}((\mathbf{C}_{\Xi G}^H[n]\mathbf{C}_{\Xi G}[n])^{-H}(\mathbf{C}_{\Xi G}^H[n]\mathbf{C}_{\Xi G}[n])^{-1} \\ &\quad \cdot \mathbf{G}^H[n]\mathbf{D}_\lambda[n]\Xi[n]\Xi^H[n]\mathbf{D}_\lambda^2[n]\Xi[n]\Xi^H[n]\mathbf{D}_\lambda[n]\mathbf{G}[n]) \\ &\stackrel{n \rightarrow \infty}{=} (1 - \lambda[n])^2 \text{Tr}((\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g})^{-1}\mathbf{C}_{\xi g}^H\Xi^H[n]\mathbf{D}_\lambda^2[n]\Xi[n] \\ &\quad \cdot \mathbf{C}_{\xi g}(\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g})^{-H}) \\ &\stackrel{n \rightarrow \infty}{=} \frac{1 - \lambda[n]}{1 + \lambda[n]} \text{Tr}((\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g})^{-1}\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi\xi}\mathbf{C}_{\xi g}(\mathbf{C}_{\xi g}^H\mathbf{C}_{\xi g})^{-H}), \end{aligned} \quad (\text{A.11})$$

and the fact that $\Xi^H[n]\mathbf{D}_\lambda^2[n]\Xi[n] \approx (1 - \lambda^2[n])^{-1}\mathbf{C}_{\xi\xi}$ and $\mathbb{E}[\text{Tr}(\mathbf{B}[n]\mathbf{B}^H[n])] = \text{Tr}((\mathbf{C}_{\Xi G}^H[n]\mathbf{C}_{\Xi G}[n])^{-1}\mathbf{R}_\tau[n]\mathbf{R}_\tau^H[n] \cdot (\mathbf{C}_{\Xi G}^H[n]\mathbf{C}_{\Xi G}[n])^{-H}) \xrightarrow{n \rightarrow \infty} \frac{K\lambda^2[n]}{(1 - \lambda[n])^2}$.

Consequently,

$$J_{MSD}(\mathbf{w}_k[n]) \approx \frac{1 - \lambda[n]}{1 + \lambda[n]} \sigma_{\Sigma_k}^2 \text{Tr}(\mathbf{\Omega}) + \frac{\lambda^2[n]\sigma_{h_k}^2[n]}{(1 - \lambda[n])^2}, \quad (\text{A.12})$$

where $\sigma_{\Sigma_k}^2 = \sigma_{\eta_k}^2 + \sigma_{r_k}^2 \text{Tr}(\mathbf{C}_{\xi\xi})$, $\sigma_{h_k}^2[n] = K\sigma_{\delta h_k}^2[n] + \zeta_k$. Therefore, the total MSD is given by

$$J_{MSD}[n] \approx \frac{1 - \lambda[n]}{1 + \lambda[n]} \sigma_{\Sigma}^2 \text{Tr}(\mathbf{\Omega}) + \frac{\lambda^2[n]\sigma_h^2[n]}{(1 - \lambda[n])^2}. \quad (\text{A.13})$$

where $\sigma_{\Sigma}^2 = \sum_{k=1}^K \sigma_{\Sigma_k}^2$ and $\sigma_h^2[n] = \sum_{k=1}^K \sigma_{h_k}^2[n]$.

REFERENCES

- [1] M. B. Balogun, O. O. Oyerinde, and F. Takawira, "Adaptive subspace method for phase noise estimation in CO-OFDM systems," *IEEE Access*, vol. 6, pp. 56368–56377, 2018.
- [2] K. Jaleleddini, E. S. Tehrani, and R. E. Kearney, "A subspace approach to the structural decomposition and identification of ankle joint dynamic stiffness," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 6, pp. 1357–1368, Jun. 2017.
- [3] Y. Zhang, D. Wang, J. Wang, and X. You, "Channel estimation for massive MIMO-OFDM systems by tracking the joint angle-delay subspace," *IEEE Access*, vol. 4, pp. 10166–10179, 2016.
- [4] K. I. Diamantaras and T. Papadimitriou, "An efficient subspace method for the blind identification of multichannel FIR systems," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5833–5839, Dec. 2008.
- [5] J. R. Guerci and J. S. Bergin, "Principal components, covariance matrix tapers, and the subspace leakage problem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 152–162, Aug. 2002.
- [6] G. van der Veen, J.-W. van Wingerden, M. Lovera, M. Bergamasco, and M. Verhaegen, "Closed-loop subspace identification methods: An overview," *IET Control Theory Appl.*, vol. 7, no. 10, pp. 1339–1358, Jul. 2013.
- [7] L. Liu, F. Ding, L. Xu, J. Pan, A. Alsaedi, and T. Hayat, "Maximum likelihood recursive identification for the multivariate equation-error autoregressive moving average systems using the data filtering," *IEEE Access*, vol. 7, pp. 41154–41163, 2019.

- [8] S. A. N. Sarmadi and V. Venkatasubramanian, "Electromechanical mode estimation using recursive adaptive stochastic subspace identification," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 349–358, Jan. 2014.
- [9] Q. Song, Y. Mi, and W. Lai, "A novel variable forgetting factor recursive least square algorithm to improve the anti-interference ability of battery model parameters identification," *IEEE Access*, vol. 7, pp. 61548–61557, 2019.
- [10] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, vol. 87. Oxford, U.K.: Clarendon Press Oxford, 1965.
- [11] Q. Wu, J.-S. Zheng, Z.-C. Dong, E. Panayirci, Z.-Q. Wu, and R. Qingnuobu, "An improved adaptive subspace tracking algorithm based on approximated power iteration," *IEEE Access*, vol. 6, pp. 43136–43145, 2018.
- [12] M. Alloulah, A. Isopoussu, C. Min, and F. Kawsar, "On tracking the physicality of Wi-Fi: A subspace approach," *IEEE Access*, vol. 7, pp. 19965–19978, 2019.
- [13] Z. Yang, R. C. de Lamare, and X. Li, " L_1 -regularized STAP algorithms with a generalized sidelobe canceler architecture for airborne radar," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 674–686, Feb. 2012.
- [14] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.
- [15] Y. M. Cho and T. Kailath, "Fast subspace-based system identification: An instrumental variable approach," *Automatica*, vol. 31, no. 6, pp. 903–905, Jun. 1995.
- [16] M. Verhaegen and E. Deprettere, "A fast, recursive MIMO state space model identification algorithm," in *Proc. 30th IEEE Conf. Decis. Control*, Dec. 1991, pp. 1349–1354.
- [17] K.-L. Tung and Y. Yam, "System identification and control synthesis for a benchmark problem," *IEEE Trans. Control Syst. Technol.*, vol. 6, no. 1, pp. 103–110, Jan. 1998.
- [18] A. Janot, P.-O. Vandanjon, and M. Gautier, "Identification of physical parameters and instrumental variables validation with two-stage least squares estimator," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 4, pp. 1386–1393, Jul. 2013.
- [19] A. Janot, P.-O. Vandanjon, and M. Gautier, "A generic instrumental variable approach for industrial robot identification," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 1, pp. 132–145, Jan. 2014.
- [20] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, vol. 5. JSTOR, 1983.
- [21] T. Söderström and P. Stoica, *System Identification*. Upper Saddle River, NJ, USA: Prentice-Hall, 1988.
- [22] T. Gustafsson, "Instrumental variable subspace tracking using projection approximation," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 669–681, Mar. 1998.
- [23] S.-H. Leung and C. F. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, Aug. 2005.
- [24] Y. Lu, Q. Li, Z. Pan, and S. Y. Liang, "Prognosis of bearing degradation using gradient variable forgetting factor RLS combined with time series model," *IEEE Access*, vol. 6, pp. 10986–10995, 2018.
- [25] S. C. Chan and Y. J. Chu, "A new state-regularized QRRLS algorithm with a variable forgetting factor," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 59, no. 3, pp. 183–187, Mar. 2012.
- [26] Y. J. Chu and S. C. Chan, "A new local polynomial modeling-based variable forgetting factor RLS algorithm and its acoustic applications," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 2059–2069, Nov. 2015.
- [27] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.
- [28] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, Jan. 1994.
- [29] 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.
- [30] P. Strobach, "Bi-iteration SVD subspace tracking algorithms," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1222–1240, May 1997.
- [31] R. Badeau, B. David, and G. Richard, "Fast approximated power iteration subspace tracking," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2931–2941, Aug. 2005.
- [32] S.-C. Chan, H.-J. Tan, J.-Q. Lin, and B. Liao, "A new local polynomial modeling based variable forgetting factor and variable regularized PAST algorithm for subspace tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 3, pp. 1530–1544, Jun. 2018.
- [33] T. Gustafsson, "Recursive system identification using instrumental variable subspace tracking," *IFAC Proc. Volumes*, vol. 30, no. 11, pp. 1579–1584, 1997.
- [34] M. Lovera, T. Gustafsson, and M. Verhaegen, "Recursive subspace identification of linear and non-linear Wiener state-space models," *Automatica*, vol. 36, no. 11, pp. 1639–1650, Nov. 2000.
- [35] T. Söderström and P. Stoica, "Instrumental variable methods for system identification," *Circuits, Syst. Signal Process.*, vol. 21, no. 1, pp. 1–9, 2002.
- [36] P. Stoica and T. Söderström, "Optimal instrumental variable estimation and approximate implementations," *IEEE Trans. Autom. Control*, vol. AC-28, no. 7, pp. 757–772, Jul. 1983.
- [37] B. Friedlander, "The overdetermined recursive instrumental variable method," in *Adaptive Systems in Control and Signal Processing*. Amsterdam, The Netherlands: Elsevier, 1984, pp. 215–220.
- [38] B. Porat and B. Friedlander, "The square-root overdetermined recursive instrumental variable algorithm," *IEEE Trans. Autom. Control*, vol. AC-34, no. 6, pp. 656–658, Jun. 1989.
- [39] T. McKelvey, "Identification of state-space models from time and frequency data," Dept. Elect. Eng., Linköping Univ., Linköping, Sweden, Tech. Rep. 380, 1995.
- [40] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, 1987.
- [41] G. Mercère, L. Bako, and S. Lecœuche, "Propagator-based methods for recursive subspace model identification," *Signal Process.*, vol. 88, no. 3, pp. 468–491, Mar. 2008.
- [42] G. Mercère and M. Lovera, "Convergence analysis of instrumental variable recursive subspace identification algorithms," *Automatica*, vol. 43, no. 8, pp. 1377–1386, Aug. 2007.
- [43] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. London, U.K.: Academic, 1979.
- [44] S. C. Chan and Z. G. Zhang, "Local polynomial modeling and variable bandwidth selection for time-varying linear systems," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 3, pp. 1102–1117, Mar. 2011.
- [45] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996, pp. 374–426.
- [46] A. Björck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Math. Comput.*, vol. 27, no. 123, p. 579, Sep. 1973.
- [47] P. Å. Wedin, "On angles between subspaces of a finite dimensional inner product space," in *Matrix Pencils*. Berlin, Germany: Springer, 1983, pp. 263–285.
- [48] T. Katayama, *Subspace Methods for System Identification*. London, U.K.: Springer, 2006.
- [49] E. Feron, M. Brenner, J. Paduano, and A. Turevskiy, "Time-frequency analysis for transfer function estimation and application to flutter clearance," *J. Guid., Control, Dyn.*, vol. 21, no. 3, pp. 375–382, May 1998.
- [50] R. N. Cabrera, V. M. Alvarado, G. L. Lopez, and M. A. Medina, "An alternative approach to the inference of the extended observability matrix, and its relation with the PO-MOESP algorithm," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 3, pp. 888–898, May 2017.
- [51] L. Ljung, "Aspects and experiences of user choices in subspace identification methods," *IFAC Proc. Volumes*, vol. 36, no. 16, pp. 1765–1770, Sep. 2003.
- [52] S. C. Chan, Y. J. Chu, Z. G. Zhang, and K. M. Tsui, "A new variable regularized QR decomposition-based recursive least M-estimate algorithm—Performance analysis and acoustic applications," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 907–922, May 2013.



JIAN-QIANG LIN (Student Member, IEEE) received the bachelor's degree in information engineering from Beihang University, Beijing, China, in 2015. He is currently pursuing the Ph.D. degree in digital signal processing with the Department of Electrical and Electronic Engineering, The University of Hong Kong.

His current research interests include array signal processing, subspace identification, adaptive filtering, and mathematics.



SHING-CHOW CHAN (Member, IEEE) received the B.Sc. (Eng.) and Ph.D. degrees in electrical engineering from The University of Hong Kong, in 1986 and 1992, respectively. Since 1994, he has been with the Department of Electrical and Electronic Engineering, The University of Hong Kong, where he is currently a Professor and an Associate Head of the Department. His research interests include fast transform algorithms, filter design and realization, multirate and biomedical signal processing, communications and array signal processing, high-speed A/D converter architecture, bioinformatics, smart grid, and image-based rendering. He is also a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society and the Associate Editor of the *Journal of Signal Processing Systems*. He was the Chair of the IEEE Hong Kong Chapter of Signal Processing, from 2000 to 2002, and organizing committee members of 2003 IEEE ICASSP and 2010 IEEE ICIP. He served as the associate editors for *Digital Signal Processing*, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II, from 2012 to 2016, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I, from 2008 to 2009.



HAI-JUN TAN received the B.Eng. degree in electronic science and technology from the South China University of Technology, Guangzhou, China, in 2010, the M.Eng. degree in communication and information system from the Department of Communication and Information System, Sun Yat-sen University, Guangzhou, China, in 2013, and the Ph.D. degree in signal processing from The University of Hong Kong, Hong Kong, in 2019. His research interests include adaptive signal processing and FPGA based hardware acceleration of video processing.

• • •