A Fast Recursive Total Least Squares Algorithm for Adaptive IIR Filtering

Dong-Xia Chang, Da-Zheng Feng, *Member, IEEE*, Wei-Xing Zheng, *Senior Member, IEEE*, and Lei Li

***Abstract*—This paper develops a new fast recursive total least squares (N-RTLS) algorithm to recursively compute the total least squares (TLS) solution for adaptive infinite-impulse-response (IIR) filtering. The new algorithm is based on the minimization of the constraint Rayleigh quotient in which the first entry of the parameter vector is fixed to the negative one. The highly computational efficiency of the proposed algorithm depends on the efficient computation of the gain vector and the adaptation of the Reyleigh quotient. Using the shift structure of the input data vectors, a fast algorithm for computing the gain vector is established, which is referred to as the fast gain vector (FGV) algorithm. The computational load of the FGV algorithm is smaller than that of the fast Kalman algorithm. Moreover, the new algorithm is numerically stable since it does not use the well-known matrix inversion lemma. The computational complexity of the new algorithm per iteration is also $O(L)$. The global convergence of the new algorithm is studied. The performances of the relevant algorithms are compared via simulations.**

***Index Terms*—Adaptive filtering, fast gain vector, IIR filtering, Rayleigh quotient.**

I. Introduction

ADAPTIVE infinite-impulse-response (IIR) filters are considered as the efficient replacements for adaptive finite-impulse-response (FIR) filters when the desired filter can be more economically modeled with poles and zeros only than with the all-zero form of an FIR tapped-delay line. The possible benefits in reduced complexity and improved performance have enlarged the usability of the adaptive IIR filter. Correspondingly, adaptive IIR filters have been the subject of active research over the last three decades [1]–[3]. Examples of applications of the IIR filters methods include adaptive noise cancellation [4], spectral estimation [5], time delay estimation, and adaptive deconvolution [6].

Considerable research has been conducted to derive adaptive IIR filters in several different ways. One type of algorithms is obtained by means of the output-error method [7]–[9]. In the output error techniques, the adaptive filter operates in a recursive manner on the input signal to provide an estimate of the desired response signal. However, this type of algorithm requires a certain system transfer function to be strictly positive real in order to avoid the problems with instability and to ensure the convergence of the algorithm [10], [11]. Because the error surface of the output error is a nonlinear function of coefficients, the error function usually contains multiple local minimum points, which may not assure that filter parameters vector converges to the global minimum point associated with an unbiased solution for adaptive IIR filtering. Since the output error methods are highly nonlinear, it is difficult for them to produce the unbiased solution. The other class of algorithms is obtained by means of the equation-error technique. Since the system model employed is linear, the equation-error methods for adaptive IIR filtering can operate in a stable manner when the step size is properly selected. Moreover, they have such attractive features as a unimodal error surface, good convergence, and guaranteed stability compared with the output-error approach [2]. However, unfortunately, the equation-error adaptive algorithms usually give biased solutions for adaptive IIR filtering since the feedback of the noisy output results in the estimation error of the correlation matrix. In order to overcome the bias problem, some efficient algorithms have been developed, such as unit-norm constraint [12], [13], monic normalization [14], [15], and total least squares (TLS) approaches [16]. Among them, the total least squares (TLS) approaches have proved to be the appealing alternative for achieving the unbiased solution in adaptive IIR filtering.

This paper investigates the TLS solution of the equation-error adaptive IIR filters when only the output vector contains additive noise. Although the TLS problems were carefully proposed in 1901 [17], they were not extensively explored for a long time. Since their basic performances were studied by Golub and Van Loan in 1980 [18], the solution of the TLS problems has been widely applied in a broad class of scientific disciplines such as economics, system theory, signal processing, and automatic control [4], [19]–[24]. Nevertheless, the study of the TLS solutions is still insufficient, and their applications in signal processing are limited, perhaps due to a lack of efficient algorithms to solve the related eigenvalue problem. In general, the solution of a TLS problem can be obtained by the singular value decomposition (SVD) of a matrix [18], [25]. Since the multiplication operations of SVD for an L by L matrix are of computational complexity $O(L^3)$, the application of the TLS methods is restricted in practice, especially in real-time signal processing.

For adaptively computing the generalized eigenvector associated with the smallest eigenvalue of the autocorrelation matrix, a number of algorithms have been proposed in the context of Pisarenko spectral estimation [26]. These algorithms fall into two broad categories. The first category involves the stochastic-type adaptive algorithms [26]–[31]. However, these algorithms have

Manuscript received September 28, 2003; revised March 16, 2004. This work was supported in part by National Natural Science Foundation of China and Century-Spanning Foundation for Qualified Persons in Education Province and in part by a research grant from the Australian Research Council. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xiang-Gen Xia.

D.-X. Chang and D.-Z. Feng are with the National Laboratory for Radar Signal Processing, Xidian University, Xi'an, 710071, China (e-mail: changdongxia@163.com; dzfeng@xidian.edu.cn).

W.-X. Zheng is with the School of QMMS, University of Western Sydney, Sydney, Australia (e-mail: w.zheng@uws.edu.au).

L. Li is with the E-government Center, Institute of Software, Chinese Academy of Science, Beijing, China (e-mail: lil@intec.iscas.ac.cn).

Digital Object Identifier 10.1109/TSP.2004.842180

no equilibrium point under the persistent excitation condition and with the constant learning rate, as shown in [32]. In contrast, the total least mean squares (TLMS) algorithm developed in [33] has an equilibrium point under the persistent excitation condition [34]. The existing random algorithms have a simple structure and require $O(L)$ multiplication per iteration but have relatively slow convergence speed compared with the following second class of algorithms. Usually, we refer to all the first-category algorithms as the TLMS algorithms.

A large variety of the second-category algorithms are called the recursive total least squares (RTLS) algorithms that usually have $O(L^2)$ computational complexity per iteration. Other algorithms (such as the inverse-power method [27], the conjugate-gradient method [35], and the least squares-like method [36]) also require $O(L^2)$ multiplication operations per eigenvector update. In particular, for online solution of the TLS problems in adaptive filtering, Davila [16], [37] proposed a fast RTLS algorithm based on gradient search for the generalized Rayleigh quotient along the Kalman gain vector [38]. This algorithm can fast track the eigenvector associated with the smallest eigenvalue of the augmented autocorrelation matrix since the Kalman gain vector can be fast estimated by taking advantage of the shift structure of the input data vector. Davila's RTLS algorithm has computational complexity of $O(L)$ per iteration but is dependent on the fast computation of the Kalman gain vector. It should be pointed out that the computation of the Kalman gain vector may be potentially unstable [39], although there have been some efficient solution approaches [40], [41] for overcoming the instability of the Kalman gain vector.

This paper proposes to search the minimum point of the constraint Rayleigh quotient along the input data vector. We define a gain vector that can also be computed fast. The computational complexity of the new algorithm is lower than that of Davila's. Moreover, the proposed algorithm is independent of the recursive computation of the inversion of the autocorrelation matrix and possesses numerical stability as well.

The paper is organized as follows. In Section II, we consider the TLS problems in adaptive IIR filtering and describe the new algorithm (N-RTLS) for efficient computation of the TLS solution. Section III discusses the global convergence and unbiased property of the new algorithm. In Section IV, we present computer simulations to demonstrate the performances of the N-RTLS algorithm. Section V gives some conclusions.

The notations in this paper are as follows: A capital boldfaced letter is used to denote a matrix. A small boldfaced letter is used to denote a column vector, unless specified otherwise. An integer t is a discrete-time variable. $E\{*\}$ denotes the expectation operator. $\mathbf{0}$ is a null vector or matrix, $\mathbf{I}$ denotes an identity matrix, and $\mathbf{e}_{L,n} \in R^{L\times 1}$ represents a coordinate vector in which only the nth element is nonzero and equal to 1. If a vector $\mathbf{u} = [u_1, u_2, \ldots, u_L]^T \in R^{L\times 1}$, then $[\mathbf{u}_{1,(L-1)}] = [u_1, u_2, \ldots, u_{L-1}]^T \in R^{(L-1)\times 1}$.

II. TLS Problems in Adaptive IIR Filtering and New RTLS Algorithm

A. Signal Model

Consider an unknown system with infinite impulse response and assume that only the output is corrupted by the additive white Gaussian noise. We use an equation-error adaptive IIR filter to estimate the IIR system from the observations of the input and output, as shown in Fig. 1. The IIR vector of the unknown system is described by

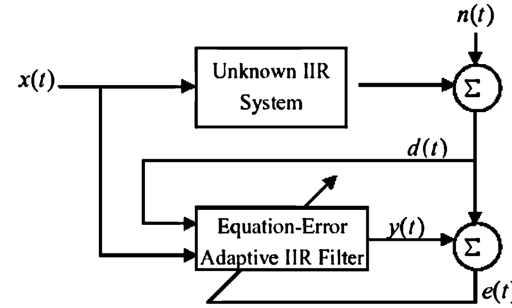

Fig. 1. Unknown adaptive IIR system with output noise $n(t)$.

$$\mathbf{h} = [a_1, a_2, \ldots, a_{N-1}, b_0, b_1, \ldots, b_{M-1}]^T \in R^{L\times 1} \tag{1}$$

where N is the autoregressive (AR) order, M is the moving-average (MA) order, and $L = N + M - 1$. $\mathbf{h}$ may be time-varying but is assumed constant here. Moreover, assume that the AR order N and the MA order M are known.

The desired output is given by

$$d(t) = \mathbf{r}^T(t)\mathbf{h} + n(t) \tag{2}$$

where the input vector $\mathbf{r}(t) \in R^{L\times 1}$ is given by

$$\mathbf{r}(t) = [d(t-1), d(t-2), \ldots, d(t-N+1), x(t), x(t-1), \ldots, x(t-M+1)]^T \tag{3}$$

and the measurement noise $n(t)$ is a zero-mean Gaussian white noise with variance σ_0^2, independent of the input vector. The output for sufficient-order equation-error adaptive IIR filter is given by

$$y(t) = \sum_{m=1}^{N-1} a_m(t)d(t-m) + \sum_{m=0}^{M-1} b_m(t)x(t-m). \tag{4}$$

In vector notations, (4) can be written as

$$y(t) = \mathbf{r}^T(t)\mathbf{w}(t)$$

where

$$\mathbf{w}(t) = [a_1(t), a_2(t), \ldots, a_{N-1}(t), b_0(t), b_1(t), \ldots, b_{M-1}(t)]^T. \tag{5}$$

At time t, the augmented data vector is defined as

$$\bar{\mathbf{r}}(t) = [d(t), \mathbf{r}^T(t)]^T. \tag{6}$$

For convenience of analysis, we define the following matrices. The autocorrelation matrix of the input vector is given by

$$\mathbf{R}(t) = E\{\mathbf{r}(t)\mathbf{r}^T(t)\}. \tag{7}$$

The autocorrelation matrix of the augmented data vector is described by

$$\bar{\mathbf{R}}(t) = E\{\bar{\mathbf{r}}(t)\bar{\mathbf{r}}^T(t)\} = \begin{bmatrix} c & \mathbf{b}^T \\ \mathbf{b} & \mathbf{R} \end{bmatrix} \tag{8}$$

where $\mathbf{b} = E\{\mathbf{r}(t)d(t)\}, c = E\{d(t)d(t)\}$. We can further show that

$$\mathbf{b} = E\{\mathbf{r}(t)d(t)\} = E\{\mathbf{r}(t)[\mathbf{r}^T(t)\mathbf{h} + n(t)]\} = \mathbf{R}(t)\mathbf{h} \tag{9}$$

$$\begin{aligned} c &= E\{d(t)d(t)\} = E\{[\mathbf{h}^T\mathbf{r}(t)+n(t)][\mathbf{r}^T(t)\mathbf{h}+n(t)]\} \\ &= \mathbf{h}^T\mathbf{R}(t)\mathbf{h}+\sigma_0^2. \end{aligned} \tag{10}$$

Therefore, the autocorrelation matrix of the augmented data vector can be written as

$$\begin{aligned} \bar{\mathbf{R}}(t) &= \begin{bmatrix} \mathbf{h}^T\mathbf{R}\mathbf{h}+\sigma_0^2 & \mathbf{h}^T\mathbf{R} \\ \mathbf{R}\mathbf{h} & \mathbf{R} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{h}^T\mathbf{R}\mathbf{h} & \mathbf{h}^T\mathbf{R} \\ \mathbf{R}\mathbf{h} & \mathbf{R} \end{bmatrix} + \begin{bmatrix} \sigma_0^2\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\ &= \mathbf{R}^* + \mathbf{R}_n. \end{aligned} \tag{11}$$

B. New RTLS Algorithm

In order to find the TLS solution for adaptive IIR filtering, Davila [16] established the following Rayleigh quotient (RQ):

$$\bar{J}(\bar{\mathbf{w}}(t)) = \frac{\bar{\mathbf{w}}^T(t)\bar{\mathbf{R}}(t)\bar{\mathbf{w}}(t)}{\bar{\mathbf{w}}^T(t)\bar{\mathbf{D}}\bar{\mathbf{w}}(t)} \tag{12}$$

where $\bar{\mathbf{w}}(t) \in R^{(L+1)\times 1}$ is the parameter vector, and $\bar{\mathbf{D}}$ is given by [42]

$$\bar{\mathbf{D}} = \begin{bmatrix} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times M} \\ \mathbf{0}_{M\times N} & \mathbf{0}_{M\times M} \end{bmatrix}. \tag{13}$$

The algorithm given in [16] searches the minimum point of $\bar{J}(\mathbf{w})$ along the Kalman gain vector. Although there are some efficient approaches that may eliminate the instability of the Kalman gain vector [40], [41], the algorithm proposed in [16] has the potential instability caused by the Kalman gain vector.

It was shown in [16] that if the parameter vector $\mathbf{w}^*$ associated with the minimization of $\bar{J}(\mathbf{w})$ is obtained, then the unbiased solution $\mathbf{w}_{\mathrm{TLS}}$ for adaptive IIR filtering is given by

$$\begin{bmatrix} -1 \\ \mathbf{w}_{\mathrm{TLS}} \end{bmatrix} = -\mathbf{w}^*/w_1^* \tag{14}$$

where w_1^* is the first element of $\mathbf{w}^*$. This shows that if the first element of $\bar{\mathbf{w}}(t)$ is constrained to the negative one, the scaling operation (14) can be avoided, which, at least, saves L multiplies, divides, and square roots (MADs). Thus, we adopt the following constraint RQ:

$$\begin{aligned} \min_{\mathbf{w}(t)} \bar{J}(\mathbf{w}(t)) &= \frac{[-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T}{[-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{D}}[-1 \quad \mathbf{w}^T(t)]^T} \\ &= \frac{[-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T}{1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)} \end{aligned} \tag{15}$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_{(N-1)\times(N-1)} & \mathbf{0}_{(N-1)\times M} \\ \mathbf{0}_{M\times(N-1)} & \mathbf{0}_{M\times M} \end{bmatrix}.$$

Notice that the number of the unknown parameters in the cost function (15) is L, whereas the number of the variable parameters in the cost function (12) is $L+1$.

An algorithm for efficiently finding the TLS solution for the adaptive IIR filtering problem is now described. This algorithm is a special gradient search method. Moreover, the selection of the update direction to be the data vector will result in the computationally efficient algorithm with computational complexity $O(L)$. The parameter vector is updated by

$$\mathbf{w}(t) = \mathbf{w}(t-1)+\beta(t)\mathbf{r}(t). \tag{16}$$

We will determine $\beta(t)$ by minimizing the Reyleigh quotient

$$\min_{\beta(t)} \bar{J}(\mathbf{w}(t)) = \frac{[-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T}{1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)}. \tag{17}$$

Remark 2.1: If the above cost function is used, it will save $2L+N+1$ MADs compared with that of [16, Table I]. In fact, since the parameter vector tracked is reduced to the L dimension from the $L+1$ dimension, more manipulations will be saved. Moreover, because the minimum point of the cost function (15) is searched along the data vector, the numerical stability of the proposed algorithm may be improved.

Notice that $\bar{\mathbf{R}}(t)$ can be computed via an iteration formula

$$\bar{\mathbf{R}}(t) = \begin{bmatrix} c(t) & \mathbf{b}^T(t) \\ \mathbf{b}(t) & \mathbf{R}(t) \end{bmatrix} = \mu\bar{\mathbf{R}}(t-1)+\bar{\mathbf{r}}(t)\bar{\mathbf{r}}^T(t) \tag{18}$$

where

$$\mathbf{R}(t) = \mu\mathbf{R}(t-1)+\mathbf{r}(t)\mathbf{r}^T(t) \tag{19}$$

$$\mathbf{b}(t) = \mu\mathbf{b}(t-1)+\mathbf{r}(t)d(t) \tag{20}$$

$$c(t) = \mu c(t-1)+d(t)d(t). \tag{21}$$

The μ in (19)–(21) is a forgetting factor, and $0<\mu\le 1$. Substituting (16) into (17), taking the derivative of the resulting equation with respect to $\beta(t)$, and forcing it to be equal to zero yields

$$\frac{\partial \bar{J}(\mathbf{w}(t))}{\partial \beta(t)} = 0 \tag{22a}$$

or equivalently

$$\begin{aligned} &[0 \quad \mathbf{r}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T[1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)] \\ &-[-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T[\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t)] = 0. \end{aligned} \tag{22b}$$

In order to solve (17) recursively, let

$$\mathbf{k}(t) = \mathbf{R}(t)\mathbf{r}(t) \tag{23}$$

$$\lambda^0(t) = [-1 \quad \mathbf{w}^T(t-1)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t-1)]^T \tag{24}$$

$$\lambda(t) = [-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T/[1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)]. \tag{25}$$

Notice that $\lambda^0(t)$ and $\lambda(t)$ can be efficiently computed by (26) and (27), shown at the bottom of the page. Hence, the gain

$$\begin{aligned} \lambda^0(t) &= [-1 \quad \mathbf{w}^T(t-1)]\{\mu\bar{\mathbf{R}}(t-1)+\bar{\mathbf{r}}(t)\bar{\mathbf{r}}(t)\}[-1 \quad \mathbf{w}^T(t-1)]^T \\ &= \mu\lambda(t-1)\{1+\mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)\}+[\mathbf{w}^T(t-1)\mathbf{r}(t)-d(t)]^2 \end{aligned} \tag{26}$$

$$\begin{aligned} \lambda(t) &= \frac{\{[-1 \quad \mathbf{w}^T(t-1)]+\beta(t)[0 \quad \mathbf{r}^T(t)]\}\bar{\mathbf{R}}(t)\{[-1 \quad \mathbf{w}^T(t-1)]+\beta(t)[0 \quad \mathbf{r}^T(t)]\}^T}{[1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)]} \\ &= \{\lambda^0(t)+2\beta(t)[\mathbf{k}^T(t)\mathbf{w}(t-1)-\mathbf{r}^T(t)\mathbf{b}(t)]+\beta^2(t)\mathbf{r}^T(t)\mathbf{k}(t)\}/[1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)] \end{aligned} \tag{27}$$

vector $\mathbf{k}(t)$ can be efficiently computed, with its fast algorithm given in Appendix B. The fast algorithm only requires $O(L)$ multiplications.

Remark 2.2: The fast algorithm for $\mathbf{k}(t)$ is similar to that of the well-known Kalman gain vector [38]. However, the well-known Kalman gain vector is based on the matrix-inversion lemma and can be numerically unstable. Although some methods were provided in [39]–[41] to overcome this instability, the Kalman gain vector may still potentially be unstable. In contrast, the fast algorithm for computing $\mathbf{k}(t)$ is independent of the matrix-inversion lemma, thus being numerically stable.

It is shown by burdensome operations (see Appendix A) that (22), in fact, is a quadratic polynomial of $\beta(t)$ and can be written as

$$a\beta^2(t) + b\beta(t) + c = 0 \tag{28}$$

where

$$\begin{aligned} a = {} & \mathbf{k}^T(t)\mathbf{r}(t)\mathbf{w}^T(t-1)\mathbf{D}\mathbf{r}(t) \\ & - [\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t) \end{aligned} \tag{29}$$

$$b = \mathbf{k}^T(t)\mathbf{r}(t)[1 + \mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)] - \lambda^0(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t) \tag{30}$$

$$\begin{aligned} c = {} & [1 + \mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)][\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)] \\ & - \lambda^0(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1). \end{aligned} \tag{31}$$

A root of (28) is given by

$$\beta(t) = (-b + (b^2 - 4ac)^{1/2})/2a. \tag{32}$$

This root makes

$$\begin{bmatrix} -1 \\ \mathbf{w}(t) \end{bmatrix}$$

converge to the eigenvector associated with the smallest eigenvalue of $\bar{\mathbf{R}}(t)$. Another root $\beta(t) = (-b - (b^2 - 4ac)^{1/2})/2a$ will make

$$\begin{bmatrix} -1 \\ \mathbf{w}(t) \end{bmatrix}$$

converge to the eigenvector associated with the largest eigenvalue of $\bar{\mathbf{R}}(t)$. Thus, (32) represents the expected root. The new RTLS algorithm is shown in Table I.

Remark 2.3: It is worth noticing that the MADs of the N-RTLS algorithms are $17L + 2N + 27$, whereas the MADs of Davila's algorithm [16], [37] are $19L + 3N + 74$, which shows that the computational complexity of the N-RTLS algorithm is significantly lower than that of Davila's.

III. Algorithm Convergence

Several issues regarding the existence and uniqueness of the TLS solution applied to identifying pole-zero systems were discussed in [16]. In this paper, we only consider the case where the TLS solution exists and is unique. This is the case when the polynomials associated with the unknown pole-zero system

$$B(z) = \sum_{m=0}^{M-1} b_m^* z^m, \quad A(z) = 1 - \sum_{m=1}^{N-1} a_m^* z^m \tag{33}$$

have no common factor [43].

TABLE I
FAST ALGORITHM

Initialize: $\mathbf{w}(0) = [0,0,\cdots,0]$, $\lambda(0) = 0$, $\mu = 0.99 \sim 1.0$	
For $t = 1,2,\cdots$	
1. update the data vector $\mathbf{r}(t)$	
2. update the gain vector $\mathbf{k}(t)$ using the approach in Appendix B	$10L+8$
3. $\lambda^0(t) = \mu\lambda(t-1)\{1+\mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)\} + [\mathbf{w}^T(t-1)\mathbf{r}(t) - d(t)]^2$	$L+N+2$
4. $\mathbf{b}(t) = \mu\mathbf{b}(t-1) + \mathbf{r}(t)d(t)$	$2L$
5. $a = \mathbf{k}^T(t)\mathbf{r}(t)\mathbf{w}^T(t-1)\mathbf{D}\mathbf{r}(t) - [\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)$	$3L+N$
6. $b = \mathbf{k}^T(t)\mathbf{r}(t)[1+\mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)] - \lambda^0(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)$	2
7. $c = [1+\mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)][\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)] - \lambda^0(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1)$	2
8. $\beta(t) = (-b + (b^2 - 4ac)^{1/2})/2a$	6
9. $1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t) = 1+\mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1) + 2\beta(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1) + \beta^2(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)$	4
10. $\lambda(t) = \{\lambda^0(t) + 2\beta(t)[\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)] + \beta^2(t)\mathbf{r}^T(t)\mathbf{k}(t)\}/[1+\mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)]$	3
11. $\mathbf{w}(t) = \mathbf{w}(t-1) + \beta(t)\mathbf{r}(t)$	L
Total real MAD's	$17L+2N+27$

MAD's stands for the number of multiplies, divides, and square roots

Perform the following generalized eigenvalue decomposition (GEVD) of the matrix pairs $\bar{\mathbf{R}}$ and $\bar{\mathbf{D}}$:

$$\begin{aligned} &\bar{\mathbf{R}}\bar{\mathbf{V}} = \bar{\mathbf{D}}\bar{\mathbf{V}}\mathrm{diag}(\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_{L+1}) \quad \text{or} \\ &\bar{\mathbf{R}}\bar{\mathbf{v}}_j = \bar{\lambda}_j\bar{\mathbf{D}}\bar{\mathbf{v}}_j \end{aligned} \tag{34}$$

where $\bar{\mathbf{V}}$ is the generalized eigen matrix, and $\bar{\mathbf{v}}_j$ and $\bar{\lambda}_j$ are the jth generalized eigen vector and eigenvalue, respectively. Note that the eigenvalues have been arranged in the descending order $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \cdots \geq \bar{\lambda}_L > \bar{\lambda}_{L+1}$. Differentiating $\bar{J}(\mathbf{w})$ with respect to $\mathbf{w}$ yields

$$\nabla\bar{J}(\mathbf{w}) = \frac{[\mathbf{R}\mathbf{h} \quad \mathbf{R}][-1 \quad \mathbf{w}^T]^T}{1 + \mathbf{w}^T\mathbf{D}\mathbf{w}} - \frac{[-1 \quad \mathbf{w}^T]\bar{\mathbf{R}}[-1 \quad \mathbf{w}^T]^T}{(1 + \mathbf{w}^T\mathbf{D}\mathbf{w})^2}\mathbf{D}\mathbf{w}. \tag{35}$$

The stationary points are obtained by solving the following equation:

$$\nabla\bar{J}(\mathbf{w}) = \frac{[\mathbf{R}\mathbf{h} \quad \mathbf{R}][-1 \quad \mathbf{w}^T]^T}{1 + \mathbf{w}^T\mathbf{D}\mathbf{w}} - \frac{[-1 \quad \mathbf{w}^T]\bar{\mathbf{R}}[-1 \quad \mathbf{w}^T]^T}{(1 + \mathbf{w}^T\mathbf{D}\mathbf{w})^2}\mathbf{D}\mathbf{w} = 0 \tag{36}$$

i.e.,

$$[\mathbf{R}\mathbf{h} \quad \mathbf{h}][-1 \quad \mathbf{w}^T]^T - \bar{J}(\mathbf{w})\mathbf{D}\mathbf{w} = \mathbf{0}. \tag{37}$$

By making the substitution (34) into (37), we can get

$$\widehat{\mathbf{V}}\,(\bar{\mathbf{\Lambda}}\mathbf{v} - \bar{J}(\mathbf{v})\mathbf{D}\mathbf{v}) = 0 \tag{38a}$$

$$\text{subject to } \tilde{\mathbf{v}}\mathbf{v} = -1 \tag{38b}$$

where $\widehat{\mathbf{V}}$ consists of the last L rows of $\bar{\mathbf{V}}$, and $\tilde{\mathbf{v}}$ is the first row of $\bar{\mathbf{V}}$. Hence, $\bar{J}(\mathbf{w})$ can be denoted by the following equation:

$$\bar{J}(\mathbf{v}) = \frac{\mathbf{v}^T\bar{\Lambda}\mathbf{v}}{\mathbf{v}^T\bar{\mathbf{D}}\mathbf{v}}. \tag{39}$$

Clearly, the solution of (38) can be written as

$$\mathbf{v}_j = -\bar{\mathbf{v}}_j/\bar{v}_{1,j} \quad \bar{v}_{1,j} \neq 0 \quad j = 1, 2, \ldots, L+1 \tag{40}$$

where $\bar{v}_{1,j}$ is the jth element of the first row of $\bar{\mathbf{V}}$. Therefore, we can get the following stationary points of $\bar{J}(\mathbf{w})$:

$$\mathbf{w}_j = -\ \widehat{\mathbf{v}}_j\ /\bar{v}_{1,j} \quad \bar{v}_{1,j} \neq 0 \quad j = 1, 2, \ldots, L+1. \tag{41}$$

Lemma 3.1: If $\mathbf{R}$ is of full rank, then $\bar{\lambda}_L > \bar{\lambda}_{L+1}$.

Proof: From (34), we have

$$\bar{\lambda}_j = \frac{\bar{\mathbf{v}}_j^T \bar{\mathbf{R}} \bar{\mathbf{v}}_j}{\bar{\mathbf{v}}_j^T \bar{\mathbf{D}} \bar{\mathbf{v}}_j} = \bar{J}(\bar{\mathbf{v}}_j).$$

Considering (11), the above equation can be written as

$$\bar{\lambda}_j = \frac{\bar{\mathbf{v}}_j^T \mathbf{R}^* \bar{\mathbf{v}}_j}{\bar{\mathbf{v}}_j^T \bar{\mathbf{D}} \bar{\mathbf{v}}_j} + \sigma_0^2.$$

Since $\mathbf{R}$ is of full rank and $\mathbf{R}^*[-1, \mathbf{h}^T]^T = \mathbf{0}$, $\mathbf{R}^*$ is of rank-deficient one. Thus, if $\bar{\mathbf{v}}_j$ is not parallel to $[-1, \mathbf{h}^T]^T$, then $(\bar{\mathbf{v}}_j^T \mathbf{R}^* \bar{\mathbf{v}}_j)/(\bar{\mathbf{v}}_j^T \bar{\mathbf{D}} \bar{\mathbf{v}}_j) > 0$, which shows $\bar{\lambda}_L = (\bar{\mathbf{v}}_L^T \mathbf{R}^* \bar{\mathbf{v}}_L)/(\bar{\mathbf{v}}_L^T \bar{\mathbf{D}} \bar{\mathbf{v}}_L) + \sigma_0^2 > \bar{\lambda}_{L+1} = \sigma_0^2$. This completes the Proof of Lemma 3.1. □

Theorem 3.1: If $\bar{\lambda}_L > \bar{\lambda}_{L+1}$ and $\bar{v}_{1,L+1} \neq 0$, then $\mathbf{w}_{L+1} = -\ \widehat{\mathbf{v}}_{L+1}\ /\bar{v}_{1,L+1}$ is the global minimum point of $\bar{J}(\mathbf{w})$. All the other stationary points are the saddle (unstable) points of $\bar{J}(\mathbf{w})$.

Proof: We can directly deduce that

$$\bar{J}(\mathbf{w}_j) = \bar{\lambda}_j \quad j = 1, 2, \ldots, L+1. \tag{42}$$

Therefore, the point $\mathbf{w}_{L+1} = -\ \widehat{\mathbf{v}}_{L+1}\ /\bar{v}_{1,L+1}$ is the unique global minimum point of $\bar{J}(\mathbf{w})$. We can define a new vector as $\mathbf{u} = \bar{\mathbf{v}}_j + \varepsilon \bar{\mathbf{v}}_{L+1}$, where ε is a small positive number. Let $\mathbf{w} = -[\mathbf{u}]_{2,L+1}/u_1$. Then, we have

$$\begin{aligned}\bar{J}(\mathbf{w}) = \frac{\mathbf{u}^T \bar{\mathbf{R}} \mathbf{u}}{\mathbf{u}^T \mathbf{D} \mathbf{u}} &= \frac{(\bar{\mathbf{v}}_j + \varepsilon \bar{\mathbf{v}}_{L+1})^T \bar{\mathbf{R}} (\bar{\mathbf{v}}_j + \varepsilon \bar{\mathbf{v}}_{L+1})}{(\bar{\mathbf{v}}_j + \varepsilon \bar{\mathbf{v}}_{L+1})^T \mathbf{D} (\bar{\mathbf{v}}_j + \varepsilon \bar{\mathbf{v}}_{L+1})} \\ &= \frac{(\bar{\mathbf{v}}_j + \varepsilon \bar{\mathbf{v}}_{p+1})^T (\bar{\lambda}_j \mathbf{D} \bar{\mathbf{v}}_j + \varepsilon \bar{\lambda}_{L+1} \mathbf{D} \bar{\mathbf{v}}_{L+1})}{\bar{\mathbf{v}}_j^T \mathbf{D} \bar{\mathbf{v}}_j + \bar{\mathbf{v}}_{L+1}^T \mathbf{D} \bar{\mathbf{v}}_{L+1}} \\ &= \frac{n\bar{\lambda}_j + m\varepsilon^2 \bar{\lambda}_{L+1}}{n + m\varepsilon^2} \\ &= \bar{\lambda}_j - \frac{m\varepsilon^2}{n+m}(\bar{\lambda}_j - \bar{\lambda}_{L+1}) < \bar{J}(\mathbf{w}_j)\end{aligned}$$

where $0 < n = \bar{\mathbf{v}}_j \mathbf{D} \bar{\mathbf{v}}_j < 1, 0 < m = \bar{\mathbf{v}}_{L+1} \mathbf{D} \bar{\mathbf{v}}_{L+1} < 1$. This shows that the stationary point $\mathbf{w}_j$ is the saddle or unstable point. The Proof of Theorem 3.1 is completed. □

Remark 3.1: The above theorem guarantees that we can search the global minimum point of $\bar{J}(\mathbf{w})$ by the gradient descent method.

Lemma 3.2: For an arbitrary $t \geq 1$, there always exists

$$\nabla \bar{J}(\mathbf{w}(t))^T \mathbf{r}(t) = 0. \tag{43}$$

Proof: Since

$$\frac{\partial \bar{J}(\mathbf{w}(t))}{\partial \beta(t)} = \nabla \bar{J}(\mathbf{w}(t))^T \frac{\partial \mathbf{w}(t)}{\partial \beta(t)} = \nabla \bar{J}(\mathbf{w}(t))^T \mathbf{r}(t). \tag{44}$$

Clearly, (44) is equivalent to (22). This completes the Proof of Lemma 3.2. □

Theorem 3.2: Assuming that t is large enough so that $\bar{\mathbf{R}}(t) \to \bar{\mathbf{R}} = E\{\bar{\mathbf{r}}(t)\bar{\mathbf{r}}^T(t)\}$, then $\mathbf{w}(t) \to \mathbf{h}$ as $t \to \infty$.

Proof: Since the (constraint) RQ function $\bar{J}(\mathbf{w})$ is bounded, $\bar{J}(\mathbf{w})$ can be defined as the energy function associated with the discrete-time sequence $\mathbf{w}(t)$ [44]. On the other hand, since $\mathbf{w}(t)$ is a gradient-descent sequence associated with $\bar{J}(\mathbf{w})$, there always exists that $J(\mathbf{w}(t)) \leq J(\mathbf{w}(t-1))$ for all $t > 0$, which shows that the discrete-time sequence $\mathbf{w}(t)$ will converge to a point in the following invariance set:

$$F = \{\mathbf{w}(t) \,|\, J(\mathbf{w}(t)) - J(\mathbf{w}(t-1)) = 0, \forall t\}. \tag{45}$$

Next, we will show that the above invariance set corresponds to the following stationary point set of $\bar{J}(\mathbf{w})$:

$$F = \{\mathbf{w}_j \,|\, \mathbf{w}_j = -\ \widehat{\mathbf{v}}_j\ /\bar{v}_{1,j}, \bar{v}_{1,j} \neq 0\ j = 1, 2, \ldots, L+1\}. \tag{46}$$

From Lemma 3.2, we can directly deduce that

$$\mathbf{r}^T(t)\{[\mathbf{hR} \quad \mathbf{R}][-1 \quad \mathbf{w}^T(t)]^T - \bar{J}(\mathbf{w}(t))\mathbf{D}\mathbf{w}(t)\} = 0 \tag{47}$$

and we can write $J(\mathbf{w}(t)) - J(\mathbf{w}(t-1)) = 0$ as

$$\begin{aligned}J(\mathbf{w}(t))[1 + \mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)] - [-1 \quad \mathbf{w}^T(t) - \beta(t)\mathbf{r}^T(t)] \\ \times \bar{\mathbf{R}}[-1 \quad \mathbf{w}^T(t) - \beta(t)\mathbf{r}^T(t)]^T = 0.\end{aligned} \tag{48}$$

Substituting $\mathbf{w}(t) = \mathbf{w}(t-1) + \beta(t)\mathbf{r}(t)$ into (48), we can get

$$\begin{aligned}&-2\beta(t)\mathbf{r}^T(t)\{\bar{J}(\mathbf{w}(t))\mathbf{D}\mathbf{w}(t) - [\mathbf{Rh} \quad \mathbf{R}][-1 \quad \mathbf{w}^T(t)]^T\} \\ &\quad + \bar{J}(\mathbf{w}(t))[1 + \mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t) + \beta^2(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)] \\ &\quad - [-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}[-1 \quad \mathbf{w}^T(t)]^T - \beta^2(t)\mathbf{r}^T(t)\mathbf{R}\mathbf{r}(t) = 0.\end{aligned} \tag{49}$$

Since $\bar{J}(\mathbf{w}(t))[1 + \mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)] - [\mathbf{w}^T(t), -1]\bar{\mathbf{R}}[\mathbf{w}^T(t), -1]^T = 0$, (49) can be written by using (47) as

$$\beta^2(t)[\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)\bar{J}(\mathbf{w}(t)) - \mathbf{r}^T(t)\mathbf{R}\mathbf{r}(t)] = 0. \tag{50}$$

Note that there always exists $[\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)\bar{J}(\mathbf{w}(t)) - \mathbf{r}^T(t)\mathbf{R}\mathbf{r}(t)] \neq 0$. Therefore, we can deduce $\beta(t) = 0$. This usually requires that

$$\nabla \bar{J}(\mathbf{w}(t)) = 0 \tag{51}$$

which implies that the invariance set $F = \{\mathbf{w}(t) \,|\, \bar{J}(\mathbf{w}(t)) - \bar{J}(\mathbf{w}(t-1)) = 0, \forall t\}$ is the stationary point set of $\bar{J}(\mathbf{w})$, i.e., we have (46).

Since the saddle point set is unstable and $\mathbf{h}$ is the unique stable point (see Theorem 3.1), we conclude that $\mathbf{w}(t) \to h$ as $t \to \infty$. This completes the Proof of Theorem 3.2. □

IV. SIMULATIONS

Some simulation results are now presented to support the theoretical analysis of the new fast recursive total least squares IIR algorithm. Here, the proposed algorithm and Davila's algorithm [16] are simply called N-RTLS and O-RTLS, respectively.

Example 1—System Identification: In this example, the RLS, O-RTLS, and the N-RTLS are applied to the system identification experiment. The unknown system impulse response is defined by

$$\mathbf{h} = [0.5 \quad -.05 \quad 0.9 \quad -0.3 \quad -0.9 \quad 0.8 \quad -0.7 \quad 0.6]^T \tag{52}$$

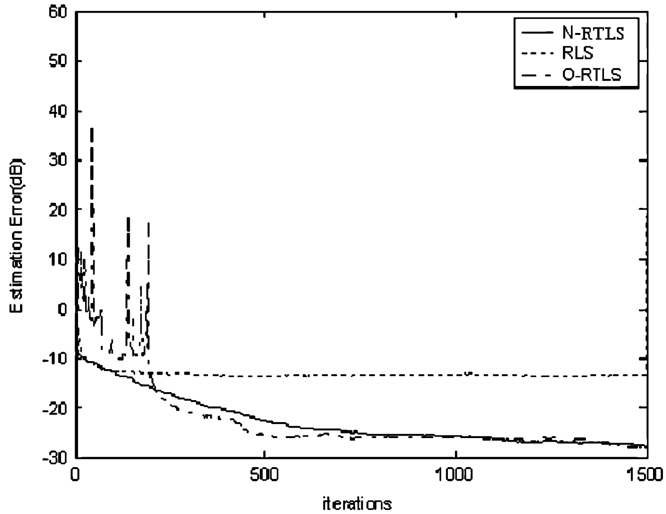

Fig. 2. N-RTLS, O-RTLS, and RLS are used for identifying a linear time-invariant system.

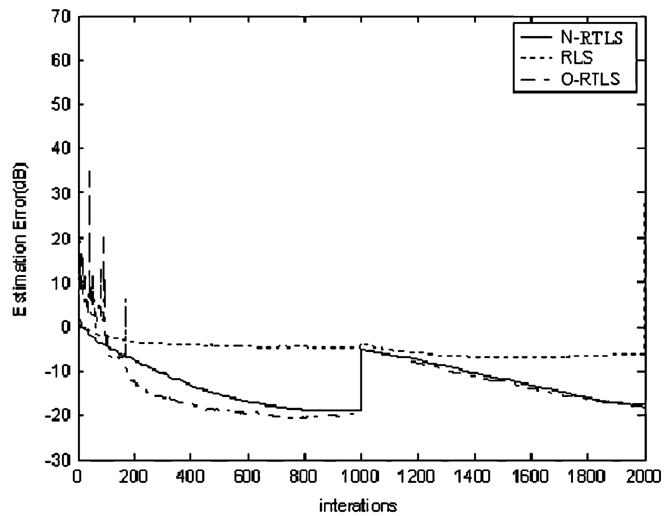

Fig. 3. N-RTLS, O-RTLS, and RLS are used for identifying a linear time-varying system.

where $N = 4, M = 5$, and $L = N + M - 1 = 8$. For the convenience of comparison and computation, the system is only stimulated by the white noise with a unit variance, as shown in Fig. 1. Only the system output is contaminated by additive, zero-mean, and white Gaussian noise that is statistically independent of the input signal. The forgetting factor μ in the evaluation was chosen to be 0.998. The estimate error is defined by

$$E(t) = \|\mathbf{w(t)} - \mathbf{h}\|_2. \tag{53}$$

When the system is linear time-invariant, the estimation results are shown in Fig. 2 for the RLS, N-RTLS, and O-RTLS algorithms. Notice that although the performance of N-RTLS is similar to that of O-RTLS, the computational complexity of N-RTLS is significantly lower than that of O-RTLS. Here, all the results are averaged over 50 independent tests.

In order to test the tracking behavior of the relative algorithms in a nonstationary environment, the parameter estimation experiment is repeated, but the unknown system parameters undergo a step change at $t = 1000$. The obtained results are shown in Fig. 3.

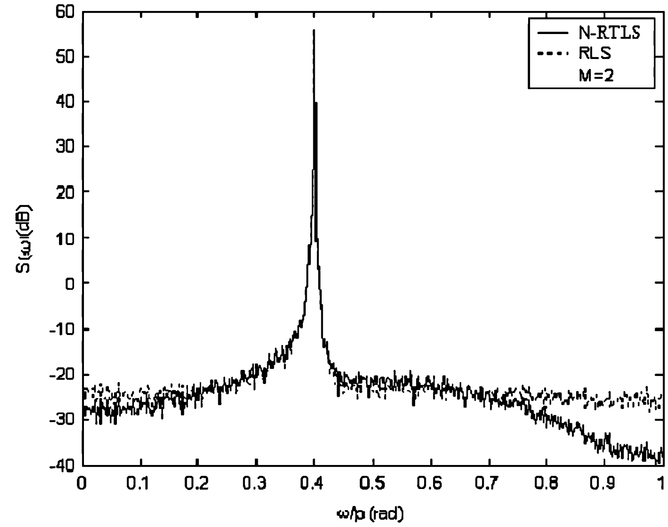

Fig. 4. Averaged periodograms of ALE outputs for N-RTLS and RLS for one sinusoid ALE lengths of 2.

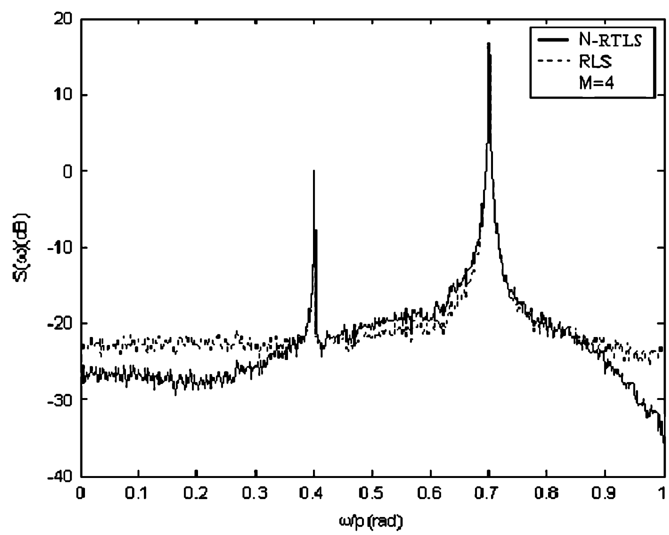

Fig. 5. Averaged periodograms of ALE outputs for N-RTLS and RLS for two sinusoids, ALE lengths of 4.

Example 2—Adaptive Line Enhancement: The N-RTLS algorithm is then applied to an adaptive line enhancement (ALE) experiment. The adaptive line enhancers based on IIR were first discussed by Rao and Kung [45]. In Experiment 1, the periodic signal consists of a single sinusoid $s(t) = \cos(0.3\pi t + \phi) + n(t)$ in additive Gaussian white noise with variance of 0.25. The phase angle ϕ is a pseudo-random variable uniformly distributed on $[0, 2\pi]$ and is held constant for each experiment. To determine the relative performance of the ALE, periodograms are computed from successive output of the ALE corresponding to $t = 1$ to 1024. These periodograms are consequently averaged over 50 independent tests. The average periodograms $S(\omega)$ are shown in Fig. 4 for ALE lengths of $M = 2$.

In Experiment 2, the observed signal includes two sinusoids in additive white noise with variance of 0.25. It can be represented as $s(t) = \cos(0.4\pi t + \phi_1) + \cos(0.6\pi t + \phi_2) + n(t)$, where ϕ_1 and ϕ_2 are randomly produced on a uniform distribution of $[-\pi, \pi]$. The obtained results are shown in Fig. 5 for ALE lengths of $M = 4$.

V. CONCLUSION

In this paper, an algorithm for efficiently computing the eigenvector associated with the smallest eigenvalue of the sample covariance matrix has been described and applied to recursively solving the total least squares (TLS) solution to the adaptive IIR filtering problem. When only the output vector contains additive noise, it has been shown that the filter coefficients produced by the N-RTLS algorithm are unbiased. The experiment results have been provided to confirm the efficiency of the proposed algorithm. The computational complexity of the N-RTLS algorithm is significantly lower than the existing algorithms.

APPENDIX A
DERIVATION OF (28)

It is straightforward to show the following equations:

$$\begin{aligned}
&[0 \quad \mathbf{r}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T \\
&\quad = [\mathbf{r}^T(t)\mathbf{b}(t) \quad \mathbf{k}(t)][-1 \quad \mathbf{w}^T(t)]^T \\
&\quad = \mathbf{k}^T(t)\mathbf{w}(t-1) + \beta(t)\mathbf{k}^T(t)\mathbf{r}(t) - \mathbf{r}^T(t)\mathbf{b}(t) \qquad (54)
\end{aligned}$$

$$\begin{aligned}
&1 + \mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t) \\
&\quad = 1 + \mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1) + 2\beta(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1) \\
&\qquad + \beta^2(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t) \qquad (55)
\end{aligned}$$

$$\begin{aligned}
&[-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T \\
&\quad = \{[-1 \quad \mathbf{w}^T(t-1)] + [0 \quad \alpha(t)\mathbf{r}^T(t)]\}\bar{\mathbf{R}}(t) \\
&\qquad \times \{[-1 \quad \mathbf{w}^T(t-1)]^T + [0 \quad \beta(t)\mathbf{r}^T(t)]^T\} \\
&\quad = \lambda^0(t) + 2\beta(t)[0 \quad \mathbf{r}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t-1)]^T \\
&\qquad + \beta^2(t)[0 \quad \mathbf{r}^T(t)]\bar{\mathbf{R}}(t)[0 \quad \mathbf{r}^T(t)]^T \\
&\quad = \lambda^0(t) + 2\beta(t)[\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)] \\
&\qquad + \beta^2(t)\mathbf{k}^T(t)\mathbf{r}(t) \qquad (56)
\end{aligned}$$

$$\begin{aligned}
&\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t) \\
&\quad = \mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1) + \beta(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t). \qquad (57)
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&[0 \quad \mathbf{r}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T[1 + \mathbf{w}^T(t)\mathbf{D}\mathbf{w}(t)] \\
&\quad = \beta^3(t)[\mathbf{k}^T(t)\mathbf{r}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t) \\
&\qquad + \beta^2(t)\{2\mathbf{k}^T(t)\mathbf{r}(t)[\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1)] \\
&\qquad + [\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)\} \\
&\qquad + \beta(t)\{\mathbf{k}^T(t)\mathbf{r}(t)[1 + \mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)] \\
&\qquad + 2[\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1)\} \\
&\qquad + [1 + \mathbf{w}^T(t-1)\mathbf{D}\mathbf{w}(t-1)][\mathbf{k}^T(t)\mathbf{w}(t-1) \\
&\quad - \mathbf{r}^T(t)\mathbf{b}(t)] \qquad (58)
\end{aligned}$$

$$\begin{aligned}
&[-1 \quad \mathbf{w}^T(t)]\bar{\mathbf{R}}(t)[-1 \quad \mathbf{w}^T(t)]^T\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t) \\
&\quad = \beta^3(t)[\mathbf{k}^T(t)\mathbf{r}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t) \\
&\qquad + \beta^2(t)\{2[\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t) \\
&\qquad + \mathbf{k}^T(t)\mathbf{r}(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1)\} \\
&\qquad + \beta(t)\{2[\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{b}(t)]\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1) \\
&\qquad + \lambda^0(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{r}(t)\} + \lambda^0(t)\mathbf{r}^T(t)\mathbf{D}\mathbf{w}(t-1). \qquad (59)
\end{aligned}$$

Finally, subtracting (59) from (58) yields (28).

APPENDIX B
FGA FOR COMPUTING GAIN VECTOR

The fast algorithm for computing the gain vector is deduced by the approach similar to [38]. For convenience of analysis, we will repeatedly define the related variables. Let $p = L + 2$, and introduce the p-dimensional vector

$$\begin{aligned}
\bar{\mathbf{r}}_p(t) = [&d(t-1), \ldots, d(t-N+1), d(t-N), x(t) \\
&\ldots, x(t-M+1), x(t-M)]^T. \qquad (60)
\end{aligned}$$

Define the permutation matrix $\mathbf{S}_{\mathrm{pp}}$ and $\mathbf{Q}_{\mathrm{pp}}$

$$\mathbf{S}_{\mathrm{pp}}\bar{\mathbf{r}}_p(t) = \left[\xi_2^T(t) \quad \mathbf{r}^T(t-1)\right]^T \qquad (61)$$

$$\mathbf{Q}_{\mathrm{pp}}\bar{\mathbf{r}}_p(t) = \left[\mathbf{r}^T(t) \quad \rho_2^T(t)\right]^T \qquad (62)$$

where

$$\boldsymbol{\xi}_2(t) = [d(t-1) \quad x(t)]^T \qquad (63)$$

$$\boldsymbol{\rho}_2(t) = [d(t-N) \quad x(t-M)]^T. \qquad (64)$$

Obviously, $\mathbf{S}_{\mathrm{pp}}^{-1} = \mathbf{S}_{\mathrm{pp}}^T, \mathbf{Q}_{\mathrm{pp}}^{-1} = \mathbf{Q}_{\mathrm{pp}}^T$. Note that vector and matrix dimensions are indicated by the subscripts.

The autocorrelation matrix of $\bar{\mathbf{r}}_p(t)$ is estimated by

$$\bar{\mathbf{R}}_{\mathrm{pp}}(t) = \mathbf{S}_{\mathrm{pp}}^T \begin{bmatrix} \xi_2(t)\xi_2^T(t) & \xi_2(t)\mathbf{r}^T(t-1) \\ \mathbf{r}(t-1)\xi_2^T(t) & \mathbf{R}(t-1) \end{bmatrix} \mathbf{S}_{\mathrm{pp}} \qquad (65)$$

or equivalently

$$\bar{\mathbf{R}}_{\mathrm{pp}}(t) = \mathbf{Q}_{\mathrm{pp}}^T \begin{bmatrix} \mathbf{R}(t) & \mathbf{r}(t)\rho_2^T(t) \\ \rho_2(t)\mathbf{r}^T(t) & \rho_2(t)\rho_2^T(t) \end{bmatrix} \mathbf{Q}_{\mathrm{pp}}. \qquad (66)$$

From (65), we can obtain

$$\begin{aligned}
\bar{\mathbf{k}}(t) &= \bar{\mathbf{R}}_{\mathrm{pp}}(t)\bar{\mathbf{r}}_p(t) \\
&= \mathbf{S}_{\mathrm{pp}}^T \begin{bmatrix} \pi_{22}(t)\xi_2(t) + \mathbf{B}_{L2}^T(t)\mathbf{r}(t-1) \\ \mathbf{B}_{L2}(t-1)\xi_2(t) + \mathbf{k}(t-1) \end{bmatrix} \qquad (67)
\end{aligned}$$

where

$$\mathbf{B}_{L2}(t) = \mathbf{B}_{L2}(t-1) + \mathbf{r}(t-1)\boldsymbol{\xi}_2^T(t) \qquad (68)$$

$$\boldsymbol{\pi}_{22}(t) = \boldsymbol{\pi}_{22}(t-1) + \boldsymbol{\xi}_2(t)\boldsymbol{\xi}_2^T(t). \qquad (69)$$

From (66), we deduce

$$\bar{\mathbf{k}}(t) = \bar{\mathbf{R}}_{\mathrm{pp}}(t)\bar{\mathbf{r}}_p(t) = \mathbf{Q}_{\mathrm{pp}}^T \begin{bmatrix} \mathbf{k}(t) + \tilde{\mathbf{B}}_{L2}(t)\boldsymbol{\rho}_2(t) \\ * \end{bmatrix} \qquad (70)$$

where

$$\tilde{\mathbf{B}}_{L2}(t) = \tilde{\mathbf{B}}_{L2}(t-1) + \mathbf{r}(t)\rho_2^T(t). \qquad (71)$$

Furthermore, comparing (67) with (70) yields

$$\begin{aligned}
&\mathbf{S}_{\mathrm{pp}}^T \begin{bmatrix} \boldsymbol{\pi}_{22}(t)\xi_2(t) + \mathbf{B}_{L2}^T(t)\mathbf{r}(t-1) \\ \mathbf{B}_{L2}(t-1)\xi_2(t) + \mathbf{k}(t-1) \end{bmatrix} \\
&\qquad = \mathbf{Q}_{\mathrm{pp}}^T \begin{bmatrix} \mathbf{k}(t) + \tilde{\mathbf{B}}_{L2}(t)\rho_2(t) \\ * \end{bmatrix}. \qquad (72)
\end{aligned}$$

Define a vector as

$$\begin{aligned}
\hat{\mathbf{k}}(t) &= \mathbf{Q}_{\mathrm{pp}}\mathbf{S}_{\mathrm{pp}}^T \begin{bmatrix} \boldsymbol{\pi}_{22}(t)\xi_2(t) + \mathbf{B}_{L2}^T(t)\mathbf{r}(t-1) \\ \mathbf{B}_{L2}(t-1)\xi_2(t) + \mathbf{k}(t-1) \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{k}(t) + \tilde{\mathbf{B}}_{L2}(t)\rho_2(t) \\ * \end{bmatrix}. \qquad (73)
\end{aligned}$$

Then, (73) gives rise to

$$\mathbf{k}(t) = [\hat{\mathbf{k}}(t)]_{1,L} - \tilde{\mathbf{B}}_{L2}(t)\rho_2(t) \qquad (74)$$

TABLE II
FGA FOR COMPUTING THE GAIN VECTOR

Algorithm	MAD's (see Table1)
Initialize $\mathbf{B}_{L2}(0)=\mathbf{0}, \tilde{\mathbf{B}}_{L2}(0)=\mathbf{0}, \boldsymbol{\pi}_{22}(0)=\mathbf{0}$.	
$\mathbf{B}_{L2}(t)=\mathbf{B}_{L2}(t-1)+\mathbf{r}(t-1)\boldsymbol{\xi}_2^T(t)$	$2L$
$\tilde{\mathbf{B}}_{L2}(t)=\tilde{\mathbf{B}}_{L2}(t-1)+\mathbf{r}(t)\boldsymbol{\rho}_2^T(t)$	$2L$
$\boldsymbol{\pi}_{22}(t)=\boldsymbol{\pi}_{22}(t-1)+\boldsymbol{\xi}_2(t)\boldsymbol{\xi}_2^T(t)$	4
$\bar{\mathbf{k}}(t)=\mathbf{S}_{pp}^T\begin{bmatrix}\boldsymbol{\pi}_{22}(t)\boldsymbol{\xi}_2(t)+\mathbf{B}_{L2}^T(t)\mathbf{r}(t-1)\\ \mathbf{B}_{L2}(t-1)\boldsymbol{\xi}_2(t)+\mathbf{k}(t-1)\end{bmatrix}$	$4L+4$
$\begin{bmatrix}\mathbf{m}(t)\\ \boldsymbol{\eta}(t)\end{bmatrix}=\mathbf{Q}_{pp}\bar{\mathbf{k}}(t)$	
$\mathbf{k}(t)=\mathbf{m}(t)-\tilde{\mathbf{B}}_{L2}(t)\boldsymbol{\rho}_2(t)$	$2L$
Total real MAD's:	$10L+8$

where $[\hat{\mathbf{k}}(t)]_{1,L}$ denotes the vector constructed by the first L elements of $\hat{\mathbf{k}}(t)$. The FGA for computing the gain vector $\mathbf{k}(t)$ is given in Table II.

ACKNOWLEDGMENT

The authors would like to thank Associate Editor Prof. X.-G. Xia and the reviewers for their valuable comments and suggestions for significantly improving the manuscript.

REFERENCES

[1] A. Carini, V. J. Mathews, and G. L. Sicuranza, "Sufficient stability bounds for slowly varying direct-form recursive linear filters and their applications in adaptive IIR filters," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2561–2567, Sep. 1999.

[2] P. A. Regalia, *Adaptive IIR Filtering in Signal Process. and Control*. New York: Marcel Dekker, 1995.

[3] S. L. Netto, P. S. R. Diniz, and P. Agathoklis, "Adaptive IIR filtering algorithms for system identification: A general framework," *IEEE Trans. Educ.*, vol. 38, pp. 54–66, Feb. 1995.

[4] B. Friedlander, "System identification techniques for adaptive noise canceling," in *Proc. 14th Asilomar Conf. Circuits, Syst., Comput.*, Pacific Grove, CA, Nov. 1980.

[5] ——, "A recursive maximum likelihood algorithm for ARMA spectral estimation," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 4, pp. 639–646, Jul. 1982.

[6] ——, "System identification techniques for adaptive signal processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, pp. 240–246, Apr. 1982.

[7] M. P. Shinde and S. N. Gupta, "Signal detection in the presence of atmospheric noise in tropics," *IEEE Trans. Commun.*, vol. COM-22, pp. 1055–1063, Aug. 1974.

[8] M. Bouvet and S. C. Schwartz, "Comparison of adaptive and robust receivers for signal detection in ambient underwater noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 621–626, May 1989.

[9] B. Mandelbrot, "The variation of certain speculative prices," *J. Business*, vol. 36, pp. 394–419, 1963.

[10] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 194–202, Apr. 1976.

[11] M. G. Larimore, "Adaptation convergence of spectral estimation based on Pisarenko harmonic retrieval," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, pp. 955–962, 1983.

[12] P. A. Regalia, "An unbiased equation error identifier and reduced-order approximations," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1397–1412, Jun. 1994.

[13] K. C. Ho and Y. T. Chan, "Bias removal in equation-error adaptive IIR filers," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 51–62, Jan. 1995.

[14] H.-N. Kim and W.-J. Song, "Unbiased equation-error adaptive IIR filtering based on monic normalization," *IEEE Signal Process. Lett.*, vol. 6, no. 2, pp. 35–37, Feb. 1999.

[15] ——, "Adaptive IIR filtering with monic normalization: Reduced-order approximation," *IEEE Signal Process. Lett.*, vol. 7, no. 3, pp. 54–56, Mar. 2000.

[16] C. E. Davila, "An algorithm for efficient, unbiased, equation-error infinite impulse response adaptive filtering," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1221–1226, May 1994.

[17] K. Pearson, "On lines and planes of closest fit to points in space," *Phil. Mag.*, vol. 2, pp. 559–572, 1901.

[18] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM. J. Numer. Anal.*, vol. 17, no. 6, pp. 883–893, Dec. 1980.

[19] S. Van Huffel and J. Vanderwalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.

[20] M. A. Rahman and K. B. Yu, "Total least square approach for frequency estimation using linear prediction," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1440–1454, Oct. 1987.

[21] T. J. Abatzoglou and J. M. Mendel, "Constrained total least squares," in *Proc. ICASSP*, Dallas, TX, Apr. 6–9, 1987, pp. 1485–1488.

[22] T. J. Abatzoglou, "Frequency super-resolution performance of the constrained total least squares method," in *Proc. ICASSP*, 1990, pp. 2603–2605.

[23] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive total least square algorithm for image reconstruction," *Multidimensional Syst. Signal Process.*, vol. 4, pp. 253–268, Jul. 1993.

[24] S. Van Huffel and P. Lemmerling, *Total Least Squares and Error-in-Variables Modeling: Analysis, Algorithms, and Applications*. Boston, MA: Kluwer, 2002.

[25] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[26] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. Astonom. Soc.*, vol. 33, pp. 347–366, 1973.

[27] R. Vaccaro, "On adaptive implementations of Pisarenko's harmonic retrieval method," in *Proc. ICASSP*, Mar. 1984, pp. 224–227.

[28] J. F. Yang and M. Kaveh, "Adaptive eigen-subspace algorithms for direction or frequency estimation and tracking," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 6, pp. 241–251, Jun. 1988.

[29] N. L. Owsley, "Adaptive data orthogonalization," in *Proc. IEEE Conf. Acoust., Speech Signal Process.*, 1978, pp. 109–112.

[30] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Networks*, vol. 5, pp. 927–935, 1992.

[31] L. Xu, E. Oja, and C. Y. Suen, "Modified Hebbian learning for curve and surface fitting," *Neural Networks*, vol. 5, pp. 441–457, 1992.

[32] V. Solo and X. Kong, "Performance analysis of adaptive eigen-analysis algorithms," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 636–646, Mar. 1998.

[33] D.-Z. Feng, Z. Bao, and L.-C. Jiao, "Total least mean squares algorithm," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2212–2220, Aug. 1998.

[34] S. Ouyang, Z. Bao, and G. Liao, "Adaptive step-size minor component extraction algorithm," *Electron. Lett.*, vol. 35, no. 6, pp. 443–444, Mar. 1999.

[35] H. Chen, T. Sarker, S. Dianant, and J. Brule, "Adaptive spectral estimation by the conjugate gradient method," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-34, pp. 272–284, 1986.

[36] V. U. Reddy, B. Egardt, and T. Kailath, "Least squares-type algorithm for adaptive implementations of Pisarenko's harmonic retrieval method," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-34, no. 3, pp. 399–405, Jun. 1982.

[37] C. E. Davila, "An efficient recursive total least squares algorithm for FIR adaptive filtering," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 268–280, Feb. 1994.

[38] L. Ljung, M. Morf, and D. Falconer, "Fast calculation of gain matrices for recursive estimation schemes," *Int. J. Contr.*, vol. 27, no. 1, pp. 1–19, 1978.

[39] S. Ljung and L. Ljung, "Error propagation properties of recursive least-squares adaptation algorithms," *Automatica*, vol. 21, no. 2, pp. 157–167, 1985.

[40] J. L. Botto and G. V. Moustakides, "Stabilizating the fast Kalman algorithms," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 9, pp. 1342–1348, Sep. 1989.

[41] D. T. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 92–114, Jan. 1991.

[42] S. Sagara and K. Wada, "On-line modified least-squares parameter estimation of linear discrete dynamic systems," *Int. J. Contr.*, vol. 25, no. 3, pp. 329–3343, 1977.

[43] T. Soderstrom and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[44] J. J. Hopfield, "Neural networks and physical systems with emergent collective computionalabilities," in *Proc. Nat. Acad. Sci. USA*, vol. 79, 1982, pp. 2554–2558.

[45] D. V. R. Rao and S. Y. Kung, "Adaptive notch filtering for the retrieval of sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 791–802, Aug. 1984.

Dong-Xia Chang was born in December 1977. She received the B.S. and M.S. degree in applied mathematic from Xidian University, Xi'an, China, in July 2000 and April 2003, respectively.

She was with the General Software Laboratory, Institute of Software, Chinese Academy of Science, Beijing, China. She is now with the National Laboratory for Radar Signal Processing, Xidian University, Xi'an, China. Her research interests include adaptive signal processing and intelligence information processing.

Da-Zheng Feng (M'02) was born in December 1959. He graduated from Xi'an Science and Technique University, Xi'an, China, in 1982 and received the M.S. degree from Xi'an Jiaotong University, in 1986, and the Ph.D. degree in electronic engineering in from Xidian University, Xi'an, in 1995.

From May 1996 to May 1998, he was a post-doctoral research affiliate and Associate Professor with Xi'an Jiaotong University. From May 1998 to June 2000, he was an associate professor with Xidian University. Since Jul. 2000, he has been a Professor with Xidian University. He has published about 50 journal papers. His research interests include signal processing, intelligence information processing, and InSAR.

Wei-Xing Zheng (M'93–SM'98) was born in Nanjing, China. He received the B.Sc. degree in applied mathematics and the M.Sc. and Ph.D. degrees in electrical engineering, in January 1982, July 1984, and February 1989, respectively, all from the Southeast University, Nanjing, China.

From 1984 to 1991, he was with the Institute of Automation, Southeast University, first as a Lecturer and later as an Associate Professor. From 1991 to 1994, he was a Research Fellow with the Department of Electrical and Electronic Engineering, Imperial College of Science, Technology, and Medicine, University of London, London, U.K.; with the Department of Mathematics, University of Western Australia, Perth, Australia; and with the Australian Telecommunications Research Institute, Curtin University of Technology, Perth. In 1994, he joined the University of Western Sydney, Sydney, Australia, where he has been an Associate Professor since 2001. He has also held various visiting positions in the Institute for Network Theory and Circuit Design, Munich University of Technology, Munich, Germany; with the Department of Electrical Engineering, University of Virginia, Charlottesville; and with the Department of Electrical and Computer Engineering, University of California, Davis. His research interests are in the areas of systems and controls, signal processing, and communications. He coauthored the book *Linear Multivariable Systems: Theory and Design* (Nanjing, China: Southeast Univ. Press, 1991).

Dr. Zheng received the Chinese National Natural Science Prize awarded by the Chinese Government in 1991. He has served on the technical program committee of several conferences including the 34th IEEE International Symposium on Circuits and Systems (ISCAS'2001), the 41st IEEE Conference on Decision and Control (CDC'2002), and the 43rd IEEE Conference on Decision and Control (CDC'2004). He has also served on several technical committees, including the IEEE Circuits and Systems Society's Technical Committee on Digital Signal Processing (since 2001) and the IFAC Technical Committee on Modeling, Identification, and Signal Processing (since 1999). He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, from 2002 to 2004. He has been an Associate Editor of the IEEE Control Systems Society's Conference Editorial Board since 2000 and an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL since 2005.

Lei Li was born in May 1971. He received the B.S. and the Ph.D. degrees in computer system structure in June 1994 and October, 1999, respectively, all from the National University of Defense Technology, Beijing, China.

Since January 2001, he has served as the Director of the E-government Center, Institute of Software, Chinese Academy of Science, Beijing. His research interests include computer operation system, network security, and management, neural networks, artificial intelligence in e-government.