A Fast Recursive Total Least Squares Algorithm for Adaptive FIR Filtering

Da-Zheng Feng, *Member, IEEE*, Xian-Da Zhang, *Senior Member, IEEE*, Dong-Xia Chang, and Wei Xing Zheng, *Senior Member, IEEE*

***Abstract*—This paper proposes a new fast recursive total least squares (N-RTLS) algorithm to recursively compute the TLS solution for adaptive finite impulse response (FIR) filtering. The N-RTLS algorithm is based on the minimization of the constrained Rayleigh quotient (c-RQ) in which the last entry of the parameter vector is constrained to the negative one. As analysis results on the convergence of the proposed algorithm, we study the properties of the stationary points of the c-RQ. The high computational efficiency of the new algorithm depends on the efficient computation of the fast gain vector (FGV) and the adaptation of the c-RQ. Since the last entry of the parameter vector in the c-RQ has been fixed as the negative one, a minimum point of the c-RQ is searched only along the input data vector, and a more efficient N-RTLS algorithm is obtained by using the FGV. As compared with Davila's RTLS algorithms, the N-RTLS algorithm saves the $6M$ number of multiplies, divides, and square roots (MADs). The global convergence of the new algorithm is studied by LaSalle's invariance principle. The performances of the relevant algorithms are compared via simulations, and the long-term numerical stability of the N-RTLS algorithm is verified.**

***Index Terms*—Adaptive filtering, fast gain vector, finite impulse response, global convergence, Rayleigh quotient, total least squares.**

I. Introduction

THE RECURSIVE least squares (RLS) algorithm has been applied extensively in adaptive signal processing areas including the adaptive filtering, online system identification, adaptive equalization, adaptive spectrum estimation, adaptive noise canceling, and so on [1]. The RLS algorithm has many desirable properties, for example, it can track variation of system parameters, and if only the system output includes a white Gaussian noise sequence, it can get unbiased estimates of system parameters. However, if both the system input and output are corrupted by white Gaussian noise, then the RLS algorithm can only provide biased estimates of the system parameters. Such a biased estimation decreases the performance of adaptive filtering. This paper considers how to find a total least squares (TLS) solution of the adaptive filtering problem in this case.

Although mentioned in [2], the TLS problems were not extensively studied for a long time. Since their basic performances were studied by Golub and Van Loan in [3], the solution of the TLS problems has been extensively applied in the domains of economics, signal processing, and automatic control [4]–[11]. As a matter of fact, however, the study on TLS solutions is still insufficient, and their application in signal processing is limited, perhaps due to a lack of efficient algorithms for solving a TLS problem online and/or offline. In general, the solution of a TLS problem can be obtained by the singular value decomposition (SVD) of a matrix [3], [12]. Since the multiplication operations of SVD for an N by N matrix are of computational complexity $O(N^3)$, the application of the TLS methods is limited in practice, especially in real-time signal processing.

To adaptively compute the generalized eigenvector associated with the smallest eigenvalue of an autocorrelation matrix, a number of algorithms have been proposed in the context of Pisarenko spectral estimation [13]. These algorithms fall into two broad categories.

The first category involves stochastic-type adaptive algorithms. Thompson [14] proposed an adaptive algorithm that is used to extract a single minor eigen-component and can be applied to find the TLS solutions of adaptive filtering and online system identification. Other similar algorithms have also been reported by several authors in [15]–[17], all leading to an adaptive implementation of Psiarenko's harmonic retrieval estimator [13]. Yang and Kaveh [16] generalized Thompson's algorithm for estimating the complete minor components with the inflation procedure. However, Yang and Kaveh's algorithm needs division operation. Oja [18], Xu *et al.* [19], and Wang and Karhunen [20] have proposed the similar algorithms that avoid the division operation but require the assumption that the smallest eigenvalue is less than unity. Recently, Luo and Unbehauen [21] presented a minor subspace analysis (MSA) algorithm to attack the above drawbacks. To solve the TLS problems in adaptive FIR and IIR filtering, Gao *et al.* [22] proposed a constrained anti-Hebbian learning algorithm that converges conditionally to the TLS solutions [23], [24]. However, the above stochastic-type adaptive algorithms have no equilibrium point under the persistent excitation condition and with the constant learning rate, as shown in [25]. In contrast, the total least mean squares (TLMS) algorithm developed in

Manuscript received August 6, 2002; revised September 25, 2003. This work was supported in part by the National Natural Science Foundation of China, the Century-Spanning Foundation for Qualified Persons in Education Province, and a research grant from the Australian Research Council. The associate editor coordinating the review of this paper and approving it for publication was Dr. Dennis R. Morgan.

D.-Z. Feng is with the Key Laboratory for Radar Signal Processing, Xidian University, 710071 Xi'an, China (e-mail: dzfeng@xidian.edu.cn).

X.-D. Zhang is with the Department of Automation, Key State Laboratory of Intelligent Technique and Systems, Tsinghua University, Beijing 100084, China.

D.-X. Chang was with the Key Laboratory for Radar Signal Processing, Xidian University, 710049 Xi'an, China. She is now with General Software Laboratory, Institute of Software, Chinese Academy of Science, Beijing, China.

W. X. Zheng is with the School of QMMS, University of Western Sydney, Sydney, NSW 1797, Australia.

Digital Object Identifier 10.1109/TSP.2004.834260

1053-587X/04$20.00 © 2004 IEEE

[26] has an equilibrium point under the persistent excitation condition [27]. In general, the stochastic-type algorithms have simple structure and require only $O(N)$ multiplication per iteration but have a relatively slow convergence speed compared with the following second category of algorithms.

A large variety of the second-category algorithms are called the recursive total least squares (RTLS) algorithms, which usually have $O(N^2)$ computational complexity per iteration. Other algorithms (such as the inverse-power method [15], the conjugate-gradient method [28], and the least squares-like method [29]) also require $O(N^2)$ multiplication operations per eigenvector update. In particular, for online solution of the TLS problems in adaptive FIR filtering, Davila [30] proposed a fast RTLS algorithm that can fast track the eigenvector associated with the smallest eigenvalue of the augmented correlation matrix. Since the Kalman gain vector can be fast estimated by utilizing the shift structure of the input data vector [31], the computational complexity of Davila's algorithm is $O(N)$ per iteration. When both the system input and output are corrupted by white Gaussian noise, the RTLS algorithms yield unbiased estimates, and thus, their performances are better than the well-known RLS algorithm [1].

It should be pointed out that computation of the Kalman gain vector may be potentially unstable [32]. Some efficient solution approaches [33], [34] were developed to overcome the potential instability of the Kalman gain vector, which results in the more complex structure and increases the computational complexity. By using the fast gain vector (FGV), the numerically stable fast transversal filter algorithms were established in [33] and [34].

The RTLS algorithm in [30] is based on fast computation of the stabilized Kalman gain vector. This paper proposes a new RTLS (N-RTLS) algorithm for adaptive FIR filtering by using the fast computation of the FGV and the adaptation minimization of the constrained Rayleigh quotient (c-RQ).

The paper is organized as follows. Section II describes briefly the TLS problems in signal processing and introduces the c-RQ criterion. Section III studies the landscape of the c-RQ criterion. The N-RTLS algorithm is developed in Section IV, whereas its global convergence is studied in Section V. In Section VI, we present computer simulations to show the performances of the N-RTLS algorithm in comparison with Davila's RTLS algorithm.

II. TLS Problems in Signal Processing

A. Signal Model

Consider an unknown system with finite impulse response (FIR), and assume that both the input and output are corrupted by the additive white Gaussian noise (AWGN). We use an adaptive FIR filter to estimate the FIR system from noisy observations of the input and output, as shown in Fig. 1. The FIR vector of the unknown system is described by

$$\mathbf{h} = [h_0, h_1, \ldots, h_{M-1}]^T \in R^{M\times 1} \tag{1}$$

where $\mathbf{h}$ may be time-varying. The desired output is given by

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

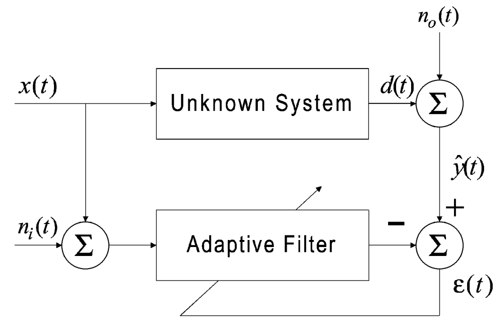

Fig. 1. Identification of unknown system $h(k)$ ($k = 0, 1, \ldots, N-1$) by using adaptive FIR filter.

where $n_o(t)$ is an AWGN with zero mean and variance σ_o^2 at the output and independent of the input signal. Moreover, the noise-free signal vector $\mathbf{x}(t) \in R^{M\times 1}$ is defined as

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

while the noisy input vector of the adaptive FIR filter $\tilde{\mathbf{x}}(t) \in R^{M\times 1}$ is given by

$$\begin{aligned}\tilde{\mathbf{x}}(t) &= [\tilde{x}(t), \tilde{x}(t-1), \ldots, \tilde{x}(t-M+1)]^T \\ &= [x(t)+n_i(t), x(t-1)+n_i(t-1), \ldots \\ &\quad x(t-M+1)+n_i(t-M+1)]^T \\ &= \mathbf{x}(t) + \mathbf{n}_i(t)\end{aligned} \tag{4}$$

where $\mathbf{n}_i(t) = [n_i(t), n_i(t-1), \ldots, n_i(t-M+1)]^T$, and $n_i(t)$ is an AWGN with zero mean and variance σ_i^2. Notice that the input noise may originate from the measured error, interference, quantized noise, and so on. Hence, we adopt the more general signal model than the adaptive least-squares-based filtering [1]. Moreover, the augmented data vector is defined as

$$\bar{\mathbf{x}}(t) = [\tilde{\mathbf{x}}^T(t), d(t)]^T \in R^{(M+1)\times 1}. \tag{5}$$

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

$$\mathbf{R} = E\{\mathbf{x}(t)\mathbf{x}^T(t)\} \tag{6}$$

and the autocorrelation matrix of the noisy input vector is described by

$$\tilde{\mathbf{R}} = E\{\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^T(t)\} = \mathbf{R} + \sigma_i^2\mathbf{I}. \tag{7}$$

Then, the autocorrelation matrix of the augmented data vector can be represented as

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

where $\mathbf{b} = E\{\tilde{\mathbf{x}}(t)d(t)\}$ and $c = E\{d(t)d(t)\}$. It is easy to show that

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

$$c = E\{[\mathbf{h}^T\mathbf{x}(t)+n_o(t)][\mathbf{x}^T(t)\mathbf{h}+n_o(t)]\} = \mathbf{h}^T\mathbf{R}\mathbf{h} + \sigma_o^2. \tag{10}$$

It is seen from (9) that if the autocorrelation matrix $\mathbf{R}$ of the noise-free input vector can be directly estimated, we can obtain an unbiased estimate of the finite impulse response $\mathbf{h}$. However, when the input vector contains the additive noise, we cannot estimate the noise-free autocorrelation matrix $\mathbf{R}$.

B. Rayleigh Quotient for Tracking TLS Solution

In order to find the TLS solution for adaptive FIR filtering, Davila [30] established the following Rayleigh quotient (RQ):

$$J(\mathbf{q}) = (\mathbf{q}^T\bar{\mathbf{R}}\mathbf{q})/(\mathbf{q}^T\bar{\mathbf{D}}\mathbf{q}) \quad (11)$$

where $\mathbf{q} \in R^{(M+1)\times 1}$ is the parameter vector, and $\bar{\mathbf{D}} = \mathrm{diag}(1,\ldots,1,\beta) \in R^{(M+1)\times(M+1)}$ is a diagonal weighting matrix with $\beta = \sigma_o^2/\sigma_i^2$. It was shown in [30] that if the parameter vector associated with the minimization of $J(\mathbf{q})$ is $\mathbf{q}^*$, then the unbiased solution $\mathbf{w}_{\mathrm{TLS}}$ for adaptive FIR filtering is given by

$$\mathbf{w}_{\mathrm{TLS}} = -[\mathbf{q}^*]_{1,M}/q^*_{M+1} \quad (12)$$

where $[\mathbf{q}^*]_{1,M}$ denotes the vector constructed by the first M entries of $\mathbf{q}^*$, and q^*_{M+1} is the last element of $\mathbf{q}^*$.

Let

$$\mathbf{q} = [\mathbf{w}^T, -1]^T. \quad (13)$$

Substituting (13) into (11), we have the cost function

$$\hat{J}(\mathbf{w}) = \{[\mathbf{w}^T,-1]\bar{\mathbf{R}}[\mathbf{w}^T,-1]^T/[\mathbf{w}^T,-1]\bar{\mathbf{D}}[\mathbf{w}^T,-1]^T\}. \quad (14)$$

Obviously, the unbiased TLS solution can also be obtained by minimizing $\hat{J}(\mathbf{w})$. Since the last entry of $\mathbf{q}$ is constrained to -1, we refer to the above cost function as the c-RQ.

Remark 2.1: If the above cost function is used, then the tenth and 11th manipulations in [30, Table I] are unnecessary and can be omitted, which saves the $2M+1$ number of multiplies, divides, and square roots (MADs). In fact, since the parameter vector tracked is reduced to M dimension from $M+1$ dimension, more manipulations will be saved.

III. Landscape of Criterion

In this section, we study the saddle points of the c-RQ.

Put $\mathbf{v} = \mathrm{diag}(1,\ldots,1,\sqrt{\beta})[\mathbf{w}^T,-1]^T = [\mathbf{w}^T,-\sqrt{\beta}]^T$. Then

$$\hat{J}(\mathbf{w}) = \{[\mathbf{w}^T,-\sqrt{\beta}]\,\breve{\mathbf{R}}\,[\mathbf{w}^T,-\sqrt{\beta}]^T/(\mathbf{w}^T\mathbf{w}+\beta)\} \quad (15)$$

where $\breve{\mathbf{R}} = \mathrm{diag}(1,\ldots,1,1/\sqrt{\beta})\bar{\mathbf{R}}\,\mathrm{diag}(1,\ldots,1,1/\sqrt{\beta})$. Let the eigenvalue decomposition (EVD) of $\breve{\mathbf{R}}$ be given by $\breve{\mathbf{R}} = \mathbf{Q}\Gamma\mathbf{Q}^T$, where $\mathbf{Q} = [\mathbf{q}_1,\ldots,\mathbf{q}_{M+1}]$, and $\Gamma = \mathrm{diag}(\gamma_1,\ldots,\gamma_{M+1})$ with the eigenvector $\mathbf{q}_i$ corresponding to the eigenvalue γ_i being arranged such that $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_M > \gamma_{M+1}$.

Differentiating $\hat{J}(\mathbf{w})$ with respect to $\mathbf{w}$, we get

$$\nabla\hat{J}(\mathbf{w}) = \left(\left[\mathbf{R}+\sigma_i^2\mathbf{I}, \frac{1}{\sqrt{\beta}}\mathbf{R}\mathbf{h}\right][\mathbf{w}^T,-\sqrt{\beta}]^T - \hat{J}(\mathbf{w})\mathbf{w}\right)\Big/(\beta+\mathbf{w}^T\mathbf{w}). \quad (16)$$

It is easy to show that the stationary points of $\hat{J}(\mathbf{w})$ are given by

$$\mathbf{w}_j = -\sqrt{\beta}[\mathbf{q}_j]_{1,M}/q_{j,(M+1)} \quad \text{for } q_{j,(M+1)} \neq 0 \quad \text{and} \quad j = 1,\ldots,M+1 \quad (17)$$

where $q_{j,(M+1)}$ is the last element of $\mathbf{q}_j$. It is easily known that $\mathbf{w}_{M+1} = \mathbf{h}$.

Theorem 3.1: If $\gamma_M > \gamma_{M+1}$ and $q_{(M+1),(M+1)} \neq 0$, then $\mathbf{w}_{M+1} = \mathbf{h}$ is a global minimum point of $\hat{J}(\mathbf{w})$. All the other stationary points are saddle (unstable) points of $\hat{J}(\mathbf{w})$.

Proof: See Appendix A. ■

The above theorem shows that one can search for the global minimum point of $\hat{J}(\mathbf{w})$ by the gradient descent method.

IV. New RTLS Algorithm

The purpose of this section is to develop a new algorithm for finding the TLS solution of the adaptive filtering problem. The new RTLS algorithm is a special gradient search method with computational complexity $O(M)$. The parameter vector is updated by

$$\mathbf{w}(t) = \mathbf{w}(t-1) + \alpha(t)\tilde{\mathbf{x}}(t) \quad (18)$$

where $\alpha(t)$ can efficiently be determined in $O(M)$ multiplications by

$$\min_{\alpha(t)} \hat{J}(\mathbf{w}(t)) = \min_{\alpha(t)} \frac{[\mathbf{w}^T(t),-1]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t),-1]^T}{\beta+\mathbf{w}^T(t)\mathbf{w}(t)}. \quad (19)$$

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

$$\bar{\mathbf{R}}(t) = \begin{bmatrix} \tilde{\mathbf{R}}(t) & \mathbf{b}(t) \\ \mathbf{b}^T(t) & c(t) \end{bmatrix} = \mu\bar{\mathbf{R}}(t-1) + \bar{\mathbf{x}}(t)\bar{\mathbf{x}}^T(t) \quad (20)$$

where

$$\tilde{\mathbf{R}}(t) = \mu\tilde{\mathbf{R}}(t-1) + \tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^T(t) \quad (21)$$

$$\mathbf{b}(t) = \mu\mathbf{b}(t-1) + \tilde{\mathbf{x}}(t)d(t) \quad (22)$$

$$c(t) = \mu c(t-1) + d(t)d(t). \quad (23)$$

The parameter μ in (20)–(23) is the forgetting factor. Let the gradient of $\hat{J}(\mathbf{w}(t))$ with respect to $\alpha(t)$ be equal to zero. Then

$$\frac{\partial\hat{J}(\mathbf{w}(t))}{\partial\alpha(t)} = 0 \quad (24a)$$

or equivalently

$$[\tilde{\mathbf{x}}^T(t),0]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t),-1]^T[\beta+\mathbf{w}^T(t)\mathbf{w}(t)] - [\mathbf{w}^T(t),-1]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t),-1]^T[\tilde{\mathbf{x}}^T(t)\mathbf{w}(t)] = 0. \quad (24b)$$

TABLE I
FAST ALGORITHM FOR COMPUTING THE GAIN VECTOR

Initialize $\mathbf{g}_M(0)=0$, $\tilde{\mathbf{g}}_M(0)=0$, $\pi(0)=0$.	MAD's
$\mathbf{g}_M(t)=\mu\mathbf{g}_M(t-1)+\tilde{\mathbf{x}}_M(t-1)\tilde{x}(t)$	$2M$
$\tilde{\mathbf{g}}_M(t)=\mu\tilde{\mathbf{g}}_M(t-1)+\tilde{\mathbf{x}}(t)\tilde{x}(t-M)$	$2M$
$\pi(t)=\mu\pi(t-1)+\tilde{x}(t)\tilde{x}(t)$	2
$\mathbf{k}(t)=\begin{bmatrix}\pi(t)\tilde{x}(t)+\mathbf{g}_M^T(t)\tilde{\mathbf{x}}(t-1)\\ \tilde{x}(t)[\mathbf{g}_M(t)]_{1,M-1}+[\mathbf{k}(t-1)]_{1,M-1}\end{bmatrix}-\tilde{\mathbf{g}}_M(t)\tilde{x}(t-M)$	$3M$
Total real MAD's: $7M+2$	
MAD's stands for the number of multiplies, divides, and square roots	

In order to efficiently solve (24), let

$$\mathbf{k}(t)=\tilde{\mathbf{R}}(t)\tilde{\mathbf{x}}(t) \tag{25}$$

$$\lambda^0(t)=[\mathbf{w}^T(t-1),-1]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t-1),-1]^T \tag{26}$$

$$\lambda(t)=[\mathbf{w}^T(t),-1]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t),-1]^T/[\beta+\mathbf{w}^T(t)\mathbf{w}(t)]. \tag{27}$$

By definition, $\lambda^0(t)$ and $\lambda(t)$ can efficiently be computed by

$$\begin{aligned}\lambda^0(t)&=[\mathbf{w}^T(t-1),-1]\{\mu\bar{\mathbf{R}}(t-1)+\bar{\mathbf{x}}(t)\bar{\mathbf{x}}^T(t)\}\\&\quad\times[\mathbf{w}^T(t-1),-1]^T\\&=\mu\lambda(t-1)\{\beta+\mathbf{w}^T(t-1)\mathbf{w}(t-1)\}\\&\quad+[\mathbf{w}^T(t-1)\tilde{\mathbf{x}}(t)-d(t)]^2\end{aligned} \tag{28}$$

$$\begin{aligned}\lambda(t)&=\{[\mathbf{w}^T(t-1),-1]+\alpha(t)[\tilde{\mathbf{x}}^T(t),0]\}\bar{\mathbf{R}}(t)\\&\quad\times\{[\mathbf{w}^T(t-1),-1]+\alpha(t)[\tilde{\mathbf{x}}^T(t),0]\}^T/\\&\quad[\beta+\mathbf{w}^T(t)\mathbf{w}(t)]\\&=\{\lambda^0(t)+2\alpha(t)[\mathbf{k}^T(t)\mathbf{w}(t-1)-\tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]\\&\quad+\alpha^2(t)\tilde{\mathbf{x}}^T(t)\mathbf{k}(t)\}/[\beta+\mathbf{w}^T(t)\mathbf{w}(t)].\end{aligned} \tag{29}$$

Table I shows the FGV algorithm [35], [36] for computing the gain vector $\mathbf{k}(t)$.

It can be shown (see Appendix B) that (24) can be rewritten as

$$a\alpha^2(t)+b\alpha(t)+c=0 \tag{30}$$

where

$$\begin{aligned}a&=[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)][\mathbf{w}^T(t-1)\tilde{\mathbf{x}}(t)]-\|\tilde{\mathbf{x}}(t)\|^2[\mathbf{k}^T(t)\mathbf{w}(t-1)\\&\quad-\tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]\end{aligned} \tag{31}$$

$$b=[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)][\beta+\|\mathbf{w}(t-1)\|^2]-\lambda^0(t)\|\tilde{\mathbf{x}}(t)\|^2 \tag{32}$$

$$\begin{aligned}c&=[\mathbf{k}^T(t)\mathbf{w}(t-1)-\tilde{\mathbf{x}}^T(t)\mathbf{b}(t)][\beta+\mathbf{w}^T(t-1)\mathbf{w}(t-1)]\\&\quad-\lambda^0(t)[\mathbf{w}^T(t-1)\tilde{\mathbf{x}}(t)].\end{aligned} \tag{33}$$

A solution of (30) is given by

$$\alpha(t)=(-b+\sqrt{b^2-4ac})/2a. \tag{34}$$

TABLE II
N-RTLS ALGORITHM

Initialize: $\mathbf{w}(0)=[0,0,\cdots,0]^T$, $\lambda(0)=0$, $\mu=0.99\sim1.0$	
For $t=1,2,\cdots$	
1 update the data vector $\tilde{\mathbf{x}}(t)$	
2 update gain vector $\mathbf{k}(t)$ as shown in Table 1	$7M+2$
3 $\lambda^0(t)=\mu\lambda(t-1)[\beta+\|\mathbf{w}(t-1)\|^2]+[\mathbf{w}^T(t-1)\tilde{\mathbf{x}}(t)-d(t)]^2$	$2M+3$
4 $\mathbf{b}(t)=\mu\mathbf{b}(t-1)+\tilde{\mathbf{x}}(t)d(t)$	$2M$
5 $a=[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)][\mathbf{w}^T(t-1)\tilde{\mathbf{x}}(t)]-\|\tilde{\mathbf{x}}(t)\|^2[\mathbf{k}^T(t)\mathbf{w}(t-1)-\tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]$	$4M+2$
6 $b=[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)][\beta+\|\mathbf{w}(t-1)\|^2]-\lambda^0(t)\|\tilde{\mathbf{x}}(t)\|^2$	2
7 $c=[\mathbf{k}^T(t)\mathbf{w}(t-1)-\tilde{\mathbf{x}}^T(t)\mathbf{b}(t)][\beta+\|\mathbf{w}(t-1)\|^2]-\lambda^0(t)[\mathbf{w}^T(t-1)\tilde{\mathbf{x}}(t)]$.	2
8 $\alpha(t)=\left(-b+\sqrt{b^2-4ac}\right)/2a$	6
9 $\mathbf{w}(t)=\mathbf{w}(t-1)+\alpha(t)\tilde{\mathbf{x}}(t)$	M
10 $\lambda(t)=\{\lambda^0(t)+2\alpha(t)[\mathbf{k}^T(t)\mathbf{w}(t-1)-\tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]+\alpha^2(t)\tilde{\mathbf{x}}^T(t)\mathbf{k}(t)\}/[\beta+\mathbf{w}^T(t)\mathbf{w}(t)]^T$	5
	Total real MAD's: $16M+20$

Our new RTLS (N-RTLS) algorithm is summarized in Table II.

Remark 4.1: It is worth noting that the MADs of the N-RTLS algorithms is $16M+20$, whereas the MADs of Davila's algorithm [30] are $22M+74$, which shows that the computational complexity of the N-RTLS algorithm is significantly lower than that of Davila's.

V. CONVERGENCE ANALYSIS

We now study the convergence property of the N-RTLS algorithm. Since the sequence $\mathbf{w}(t)$ is a discrete-time dynamical system, its convergence can be analyzed by LaSalle's invariance principle [37].

Lemma 5.1: The following result is true:

$$\nabla\hat{J}(\mathbf{w}(t))^T\mathbf{x}(t)=0,\quad\text{for all } t\geq1. \tag{35}$$

Proof: Since

$$\frac{\partial\hat{J}(\mathbf{w}(t))}{\partial\alpha(t)}=\nabla\hat{J}(\mathbf{w}(t))^T\frac{\partial\mathbf{w}(t)}{\partial\alpha(t)}=\nabla\hat{J}(\mathbf{w}(t))^T\mathbf{x}(t). \tag{36}$$

Substituting (24) into (36), we have directly (35), which completes the proof of Lemma 5.1. ■

We point out that Lemma 5.1 was first given by Davila [30], but the above proof is much simpler.

Theorem 5.1: If t is large enough so that $\bar{\mathbf{R}}(t)\to\bar{\mathbf{R}}=E\{\bar{\mathbf{x}}(t)\mathbf{x}^T(t)\}$, then $\mathbf{w}(t)\to\mathbf{h}$ as $t\to\infty$.

Proof: Clearly, the sequence $\mathbf{w}(t)$ is a discrete-time dynamical system on a Fréchet space (see [37]) and is composed of a precompact set (also see [37]). Moreover, $\hat{J}(\mathbf{w})$ is a Lyapunov function of the sequence $\mathbf{w}(t)$ on R^M, since $\hat{J}(\mathbf{w}(t))\leq\hat{J}(\mathbf{w}(t-1))$ for all $t>0$. As the sequence $\mathbf{w}(t)$ is obviously bounded and remains in R^M for all $t>0$, we can conclude from LaSalle's invariance principle [37] that $\mathbf{w}(t)$ converges to a point in the invariance set defined by

$$F = \{\mathbf{w}(t) \,|\, \hat{J}(\mathbf{w}(t)) - \hat{J}(\mathbf{w}(t-1)) = 0 \text{ for any } t\}. \quad (37)$$

Next, we show that

$$F = \{\mathbf{w}_j \,|\, \mathbf{w}_j = -\sqrt{\beta}[\mathbf{q}_j]_{1,M}/q_{j,(M+1)} \text{ for } q_{j,(M+1)} \neq 0 \text{ and } j = 1, \ldots, M+1\}. \quad (38)$$

It follows from Lemma 5.1 that

$$\mathbf{x}^T(t)\{[\tilde{\mathbf{R}}, \mathbf{Rh}][\mathbf{w}^T(t), -1]^T - \hat{J}(\mathbf{w}(t))\mathbf{w}(t)\} = 0. \quad (39)$$

Substituting $\mathbf{w}(t-1) = \mathbf{w}(t) - a(t)\tilde{\mathbf{x}}(t)$ into $\hat{J}(\mathbf{w}(t)) - \hat{J}(\mathbf{w}(t-1)) = 0$ yields

$$\begin{aligned}
&\hat{J}(\mathbf{w}(t))[\beta + \mathbf{w}^T(t)\mathbf{w}(t) + \alpha^2(t)\tilde{\mathbf{x}}^T(t)\tilde{\mathbf{x}}(t) \\
&\quad - 2\alpha(t)\tilde{\mathbf{x}}^T(t)\mathbf{w}(t)] - [\mathbf{w}^T(t) - \alpha(t)\tilde{\mathbf{x}}^T(t), -1] \\
&\quad \times \bar{\mathbf{R}}[\mathbf{w}^T(t) - \alpha(t)\tilde{\mathbf{x}}^T(t), -1]^T \\
&\quad = -2\alpha(t)\tilde{\mathbf{x}}^T(t)[\hat{J}(\mathbf{w}(t))\mathbf{w}(t) + 2\alpha(t)\tilde{\mathbf{x}}^T(t)[\tilde{\mathbf{R}}, \mathbf{Rh}] \\
&\quad \times [\mathbf{w}^T(t), -1]^T + \hat{J}(\mathbf{w}(t))[\beta + \mathbf{w}^T(t)\mathbf{w}(t) \\
&\quad + \alpha^2(t)\tilde{\mathbf{x}}^T(t)\tilde{\mathbf{x}}(t) - [\mathbf{w}^T(t), -1]\bar{\mathbf{R}}[\mathbf{w}^T(t), -1]^T \\
&\quad - \alpha^2(t)\tilde{\mathbf{x}}^T(t)\tilde{\mathbf{R}}\tilde{\mathbf{x}}(t) \\
&\quad = \alpha^2(t)[\tilde{\mathbf{x}}^T(t)\tilde{\mathbf{x}}(t)\hat{J}(\mathbf{w}(t)) - \tilde{\mathbf{x}}^T(t)\tilde{\mathbf{R}}\tilde{\mathbf{x}}(t)] = 0 \quad (40)
\end{aligned}$$

where we have used (39) and the condition $\hat{J}(\mathbf{w}(t))[\beta + \mathbf{w}^T(t)\mathbf{w}(t)] - [\mathbf{w}^T(t), -1]\bar{\mathbf{R}}[\mathbf{w}^T(t), -1]^T = 0$. Since there normally exists $[\tilde{\mathbf{x}}^T(t)\tilde{\mathbf{x}}(t)\hat{J}(\mathbf{w}(t)) - \tilde{\mathbf{x}}^T(t)\tilde{\mathbf{R}}\tilde{\mathbf{x}}(t)] \neq 0$, we have $\alpha(t) = 0$, which usually requires

$$\nabla \hat{J}(\mathbf{w}(t)) = 0. \quad (41)$$

This implies that $F = \{\mathbf{w}(t) \,|\, \hat{J}(\mathbf{w}(t)) - \hat{J}(\mathbf{w}(t-1)) = 0 \text{ for any}; t\}$ is a stationary point set, i.e., (38) is true.

On the other hand, it has been shown in Theorem 3.1 that $\hat{J}(\mathbf{w}(t))$ has the unique stable point $\mathbf{w}(t) = \mathbf{h}$. Since the saddle point set is unstable, we deduce that $\mathbf{w}(t) \to \mathbf{h}$ as $t \to \infty$. This completes the proof. ■

VI. Simulations

We compare our N-RTLS algorithm with other three algorithms. For convenience, we refer to Davila's RTLS algorithms [30] as the original RTLS (O-RTLS for short) and use RLS and IP to represent the recursive least squares algorithm [1] and the inverse power algorithm [15], respectively. Each estimation curve is the result averaged over 30 independent runs.

Example 1—Adaptive Identification of Linear System: The unknown system impulse response is given by

$$\mathbf{h}^* = [-0.3, -0.9, 0.8, -0.7, 0.6]^T. \quad (42)$$

The noise-free input $x(k)$ is a first-order AR process and given by

$$x(k) = -0.5 * x(k-1) + \eta(t) \quad (43)$$

where $\eta(t)$ is a white Gaussian noise sequence of zero mean and unit variance. Both the white Gaussian input noise and the white Gaussian output noise have the unit variance, i.e., $\sigma_i^2 = 1$ and $\sigma_o^2 = 1$. Fig. 2 shows the averaged estimation errors for the linear time-invariant system, where the estimation error is as defined in [30].

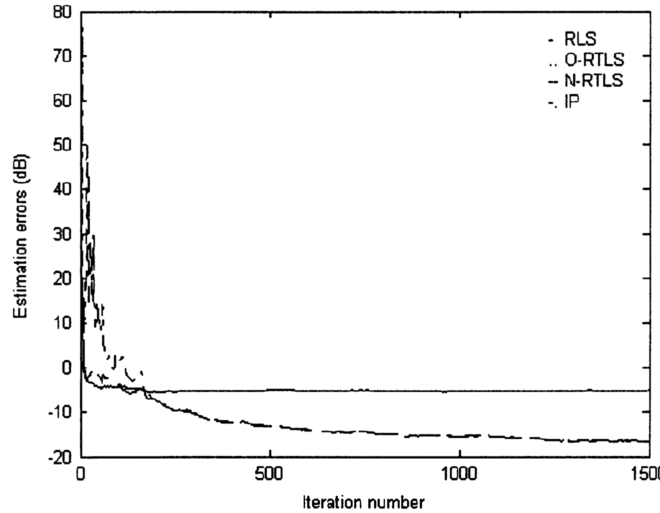

Fig. 2. Averaged estimation errors of the four methods for a time-invariant system, where the variances of the input noise and output noise are equal to 1.

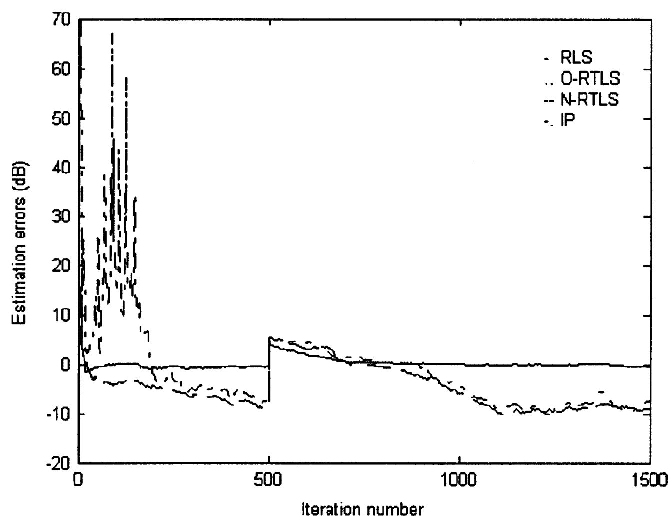

Fig. 3. Averaged estimation errors of the four methods for a time-varying system, where the variances of the input noise and output noise are equal to 1.

In order to test the tracking behavior of the four algorithms in a nonstationary environment, we simulated a time-varying system whose parameters undergo a step at time $t = 500$ and used the forgetting factor $\mu = 0.995$. Moreover, in order to make a broader comparison of the relevant algorithms under different input signals, the noise-free input $x(k)$ is also changed into the following first-order AR process

$$x(k) = 0.5 * x(k-1) + \eta(t) \quad (44)$$

where $\eta(t)$ is a white Gaussian noise sequence of zero mean and unit variance. Both the white Gaussian input noise and the white Gaussian output noise have also the unit variance. The averaged estimation errors are shown in Fig. 3.

Fig. 4 shows the long-term numerical stability of the three TLS algorithms from $t = 1$ to $t = 2 \times 10^5$ for the time-invariant

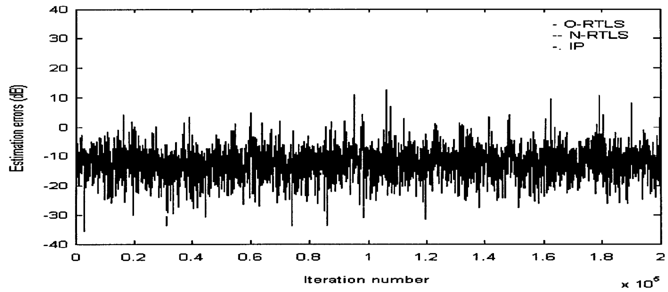

Fig. 4. Three TLS algorithms have the almost same long-time numerical stability.

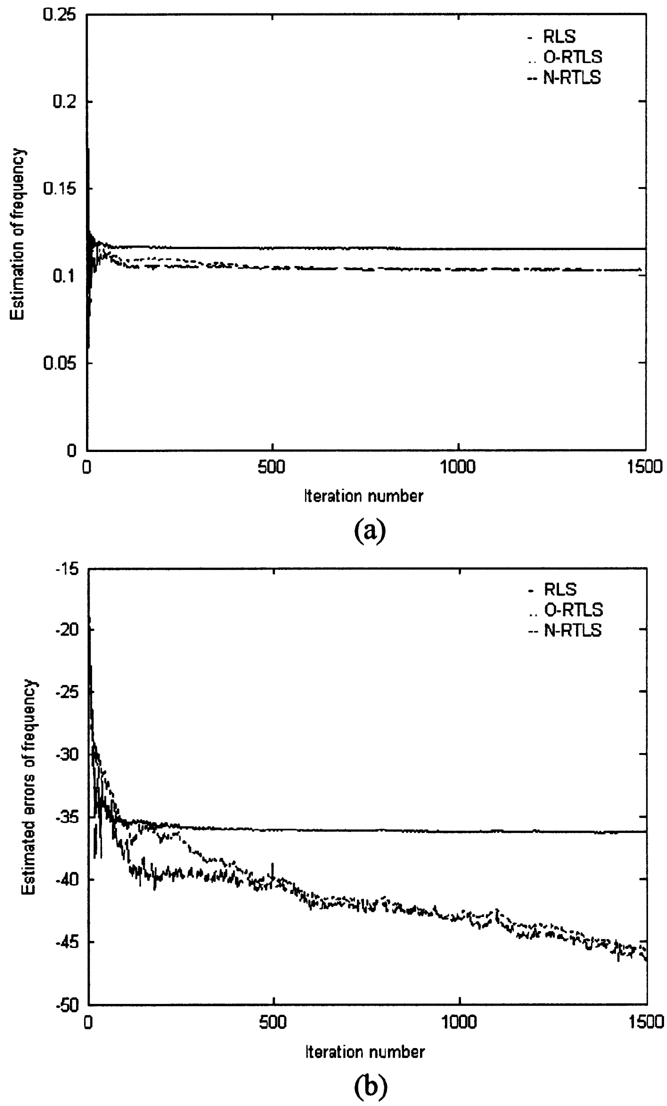

Fig. 5. Estimation results of harmonic frequency $f = 0.1$, where the linear predictive length (order) is $N = 3$. (a) Frequency estimation results. (b) Estimation errors of frequency.

system, where $\mu = 0.995$, and the input signal and noises are as given in the above.

Example 2—Adaptive Harmonic Retrieval: In this example, we consider two adaptive harmonic retrieval experiments for the N-RTLS, O-RTLS, and RLS algorithms based on the linear

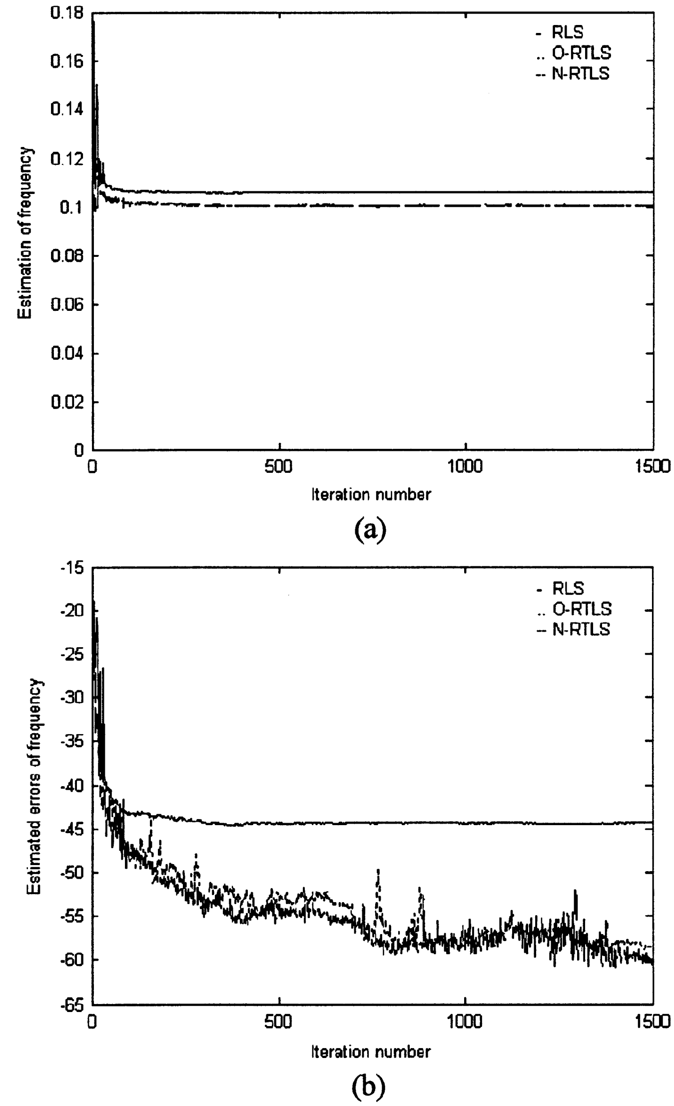

Fig. 6. Estimation results of harmonic frequency $f = 0.1$, where the linear predictive length (order) $N = 4$. (a) Frequency estimation results. (b) Estimation errors of frequency.

prediction approach [5], [11]. The estimation error is defined by

$$\varepsilon(t) = \sum_{i}^{P} (\hat{f}_i - f_i)^2 \tag{45}$$

where P is the number of harmonic waves, and f_i and $\hat{f}_i$ are the true and estimated frequency of the ith harmonic wave. Since β is difficult to be determined in advance, let simply $\beta = 1$.

Case 1) The observed data consists of a single sinusoid in the AWGN with variance 0.25, i.e., $s(t) = \cos(0.2\pi t + \phi) + n(t)$, and the phase ϕ is a random variable with uniform distribution in $[-\pi, \pi]$. The N-RTLS, O-RTLS, and RLS algorithms were used to estimate the linear prediction parameters. Figs. 5 and 6 show the estimation results of the harmonic frequency $f = 0.1$ with the linear predictive lengths (orders) $N = 3$ and $N = 4$, respectively.

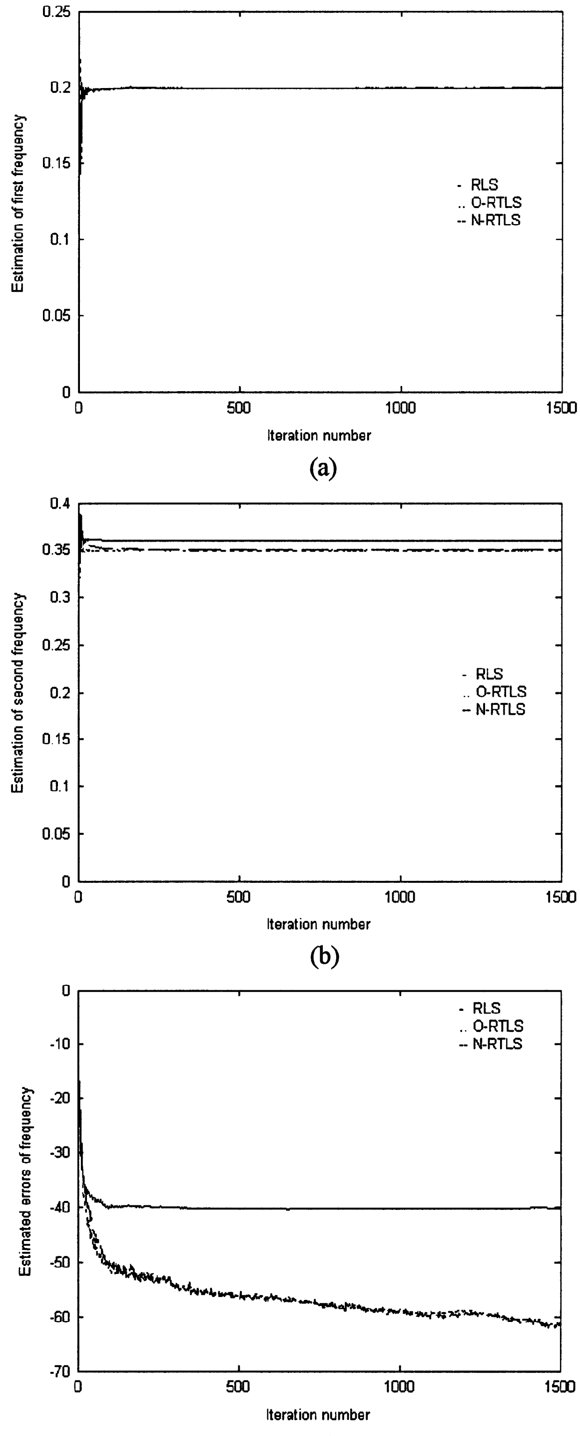

Fig. 7. Estimation results for the two sinusoids, where the linear predictive length (order) is $N = 5$. (a) Estimation results of harmonic frequency $f_1 = 0.2$. (b) Estimation results of harmonic frequency $f_2 = 0.35$. (c) Estimation errors.

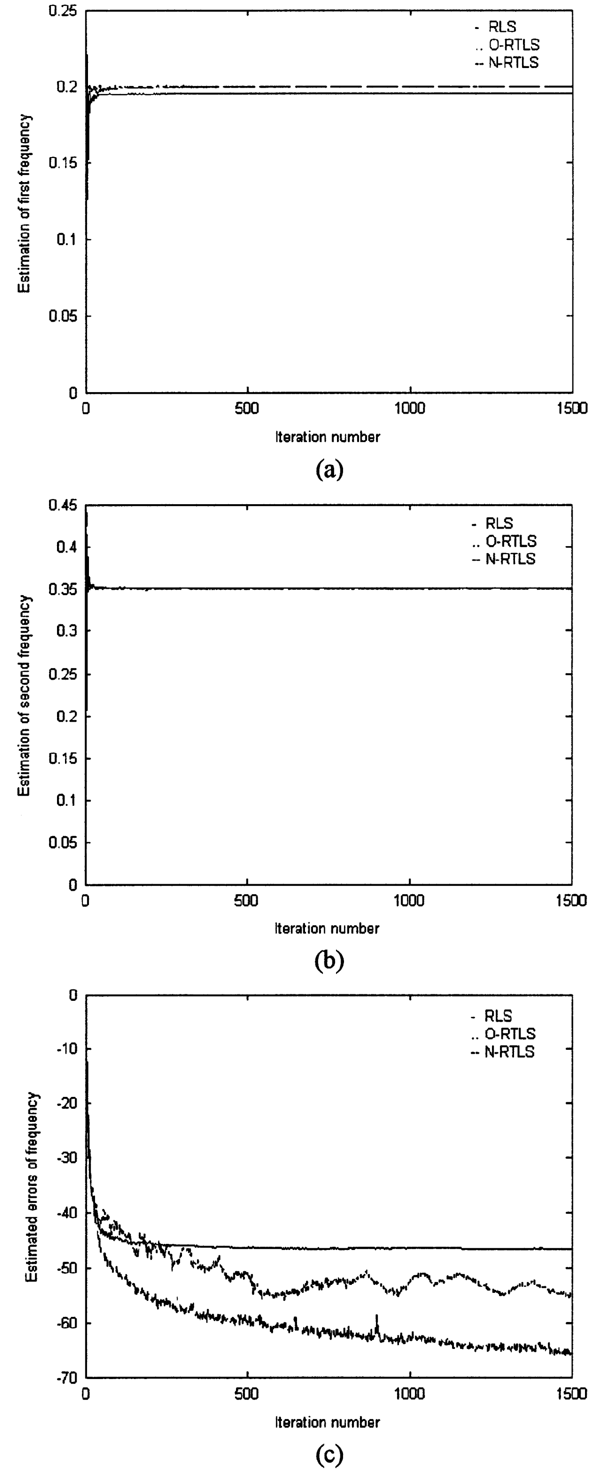

Fig. 8. Estimation results for the two sinusoids, where the linear predictive length (order) is $N = 6$. (a) Estimation results of harmonic frequency $f_1 = 0.2$. (b) Estimation results of harmonic frequency $f_2 = 0.35$. (c) Estimation errors.

Case 2) The observed data are given by $s(t) = \cos(0.4\pi t + \phi_1) + \cos(0.7\pi t + \phi_2) + n(t)$, where the variance of the AWGN is equal to 0.25, and ϕ_1 and ϕ_2 are the random variables with uniform distribution of $[-\pi, \pi]$ and are independent of each other. The estimation results are shown in Figs. 7 and 8 for $N = 5$ and $N = 6$, respectively.

The results in this example show that the performances of the N-RTLS and O-RTLS algorithms are approximately consistent.

VII. Conclusion

In this paper, a fast RTLS algorithm has been introduced for adaptive FIR filtering. The proposed algorithm has been built on the adaptation of the constrained Rayleigh quotient and the efficient computation of the fast gain vector. These represent its major differences with respect to Davila's RTLS algorithm. With the significantly reduced computational complexity, the proposed fast RTLS algorithm has demonstrated to be able to achieve the good performances that are closely consistent with those of Davila's algorithm.

APPENDIX A
PROOF OF THEOREM 3.1

We can directly deduce that

$$\hat{J}(\mathbf{w}_j) = \gamma_j, \quad j = 1, \ldots, M+1. \tag{A.1}$$

Therefore, the point $\mathbf{w}_{M+1} = -\sqrt{\beta}[\mathbf{q}]_{1,M}/q_{(M+1),(M+1)}$ is the unique global minimum point of $\hat{J}(\mathbf{w})$. Define $\mathbf{v} = \mathbf{q}_j + \varepsilon \mathbf{q}_{M+1}$ and $\mathbf{w} = -[\mathbf{v}]_{1,M}/v_{M+1}$, where ε is a positive infinitesimal, $[\mathbf{v}]_{1,M} = [v_1, \ldots, v_M]^T$, and v_{M+1} is the last entry of $\mathbf{v}$. From (15), we have

$$\begin{aligned}\hat{J}(\mathbf{w}) &= \frac{\mathbf{v}^T \breve{\mathbf{R}} \mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \frac{(\mathbf{q}_j + \varepsilon\mathbf{q}_{M+1})^T \mathbf{R} (\mathbf{q}_j + \varepsilon\mathbf{q}_{M+1})}{(\mathbf{q}_j + \varepsilon\mathbf{q}_{M+1})^T(\mathbf{q}_j + \varepsilon\mathbf{q}_{M+1})} \\ &= \frac{(\mathbf{q}_j + \varepsilon\mathbf{q}_{M+1})^T(\gamma_j \mathbf{q}_j + \varepsilon\gamma_{M+1}\mathbf{q}_{M+1})}{1+\varepsilon^2} \\ &= \frac{\gamma_j + \varepsilon^2\gamma_{M+1}}{1+\varepsilon^2} \\ &= \gamma_j - \frac{\varepsilon^2}{1+\varepsilon^2}(\gamma_j - \gamma_{M+1}) < \hat{J}(\mathbf{w}_j)\end{aligned} \tag{A.2}$$

which implies that the stationary point $\mathbf{w}_j$ is saddle or unstable. This completes the proof of Theorem 3.1. □

APPENDIX B
DERIVATION OF (30)

It is straightforward to show that

$$\begin{aligned}&[\tilde{\mathbf{x}}^T(t), 0]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t), -1]^T \\ &\quad = [\mathbf{k}^T(t), \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)][\mathbf{w}^T(t), -1]^T \\ &\quad = \mathbf{k}^T(t)\mathbf{w}(t-1) + \alpha(t)\mathbf{k}^T(t)\tilde{\mathbf{x}}(t) - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)\end{aligned} \tag{B.1}$$

$$\begin{aligned}&\beta + \mathbf{w}^T(t)\mathbf{w}(t) \\ &\quad = \beta + \|\mathbf{w}(t-1)\|^2 + 2\alpha(t)\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1) \\ &\qquad + \alpha^2(t)\|\tilde{\mathbf{x}}(t)\|^2\end{aligned} \tag{B.2}$$

$$\begin{aligned}&[\mathbf{w}^T(t), -1]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t), -1]^T \\ &\quad = \{[\mathbf{w}^T(t-1), -1] + [\alpha(t)\tilde{\mathbf{x}}^T(t), 0]\}\bar{\mathbf{R}}(t) \\ &\qquad \times \{[\mathbf{w}^T(t-1), -1]^T + [\alpha(t)\tilde{\mathbf{x}}^T(t), 0]^T\} \\ &\quad = \lambda^0(t) + 2\alpha(t)[\tilde{\mathbf{x}}^T(t), 0]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t-1), -1]^T \\ &\qquad + \alpha^2(t)[\tilde{\mathbf{x}}^T(t), 0]\bar{\mathbf{R}}(t)[\tilde{\mathbf{x}}^T(t), 0]^T \\ &\quad = \lambda^0(t) + 2\alpha(t)[\mathbf{k}^T(t), \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)][\mathbf{w}^T(t-1), -1]^T \\ &\qquad + \alpha^2(t)[\mathbf{k}^T(t), \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)][\tilde{\mathbf{x}}^T(t), 0]^T \\ &\quad = \lambda^0(t) + 2\alpha(t)[\mathbf{k}^T(t)\mathbf{w}(t-1) - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)] \\ &\qquad + \alpha^2(t)[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)],\end{aligned} \tag{B.3}$$

$$[\tilde{\mathbf{x}}^T(t)\mathbf{w}(t)] = \tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1) + \alpha(t)\|\tilde{\mathbf{x}}(t)\|^2. \tag{B.4}$$

Thus, we have

$$\begin{aligned}&[\tilde{\mathbf{x}}^T(t), 0]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t), -1][\beta + \mathbf{w}^T(t)\mathbf{w}(t)] \\ &\quad = \alpha^3(t)\|\tilde{\mathbf{x}}(t)\|^2[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)] + \alpha^2(t)\{2[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)] \\ &\qquad \times [\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1)] + \|\tilde{\mathbf{x}}(t)\|^2[\mathbf{k}^T(t)\mathbf{w}(t-1) \\ &\qquad - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]\} + \alpha(t)\{[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)][\beta \\ &\qquad + \|\mathbf{w}(t-1)\|^2] + 2[\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1)][\mathbf{k}^T(t)\mathbf{w}(t-1) \\ &\qquad - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)] + [\beta + \|\mathbf{w}(t-1)\|^2][\mathbf{k}^T(t)\mathbf{w}(t-1) \\ &\qquad - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]\end{aligned} \tag{B.5}$$

$$\begin{aligned}&[\mathbf{w}^T(t), -1]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t), -1][\tilde{\mathbf{x}}^T(t)\mathbf{w}(t)] \\ &\quad = \alpha^3(t)\|\tilde{\mathbf{x}}(t)\|^2[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)] + \alpha^2(t)\{2\|\tilde{\mathbf{x}}(t)\|^2 \\ &\qquad \times \mathbf{k}^T(t)\mathbf{w}(t-1) - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)] + [\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)] \\ &\qquad \times [\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1)]\} + \alpha(t)\{\lambda^0(t)\|\tilde{\mathbf{x}}(t)\|^2 \\ &\qquad + 2[\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1)][\mathbf{k}^T(t)\mathbf{w}(t-1) - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]\} \\ &\qquad \times \lambda^0(t)[\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1)].\end{aligned} \tag{B.6}$$

$$\begin{aligned}&[\tilde{\mathbf{x}}^T(t), 0]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t), -1]^T[\beta + \mathbf{w}^T(t)\mathbf{w}(t)] \\ &\qquad - [\mathbf{w}^T(t), -1]\bar{\mathbf{R}}(t)[\mathbf{w}^T(t), -1]^T[\tilde{\mathbf{x}}^T(t)\mathbf{w}(t)] \\ &\quad = \alpha^2(t)\{[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)][\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1)] \\ &\qquad - \|\tilde{\mathbf{x}}(t)\|^2[\mathbf{k}^T(t)\mathbf{w}(t-1) - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)]\} \\ &\qquad + \alpha(t)\{[\mathbf{k}^T(t)\tilde{\mathbf{x}}(t)][\beta + \|\mathbf{w}(t-1)\|^2] - \lambda^0(t)\|\tilde{\mathbf{x}}(t)\|^2\} \\ &\qquad + [\beta + \|\mathbf{w}(t-1)\|^2][\mathbf{k}^T(t)\mathbf{w}(t-1) - \tilde{\mathbf{x}}^T(t)\mathbf{b}(t)] \\ &\qquad - \lambda^0(t)[\tilde{\mathbf{x}}^T(t)\mathbf{w}(t-1)] = 0.\end{aligned} \tag{B.7}$$

Clearly, (B7) is just (30).

ACKNOWLEDGMENT

The authors would like to thank Associate Editor D. R. Morgan and the reviewers for their constructive comments and suggestions that have significantly improved the manuscript.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[2] K. Pearson, "On lines and planes of closest fit to points in space," *Phil. Mag.*, vol. 2, pp. 559–572, 1901.
[3] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," in *SIAM Number. Anal.*, vol. 17, Dec. 1980, pp. 883–893.
[4] S. Van Huffel and J. Vanderwalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.
[5] M. A. Rahman and K. B. Yu, "Total least square approach for frequency estimation using linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1440–1454, Oct. 1987.
[6] T. J. Abatzoglou and J. M. Mendel, "Constrained total least squares," in *Proc. ICASSP*, 1987, pp. 1485–1488.
[7] T. J. Abatzoglou, "Frequency super-resolution performance of the constrained total least squares method," in *Proc. ICASSP*, 1990, pp. 2603–2605.
[8] 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, July 1993.
[9] S. Valaee, B. Champagne, and P. Kabal, "Localization of wide-band signals using least-squares and total least-squares approaches," *IEEE Trans. Signal Processing*, vol. 47, pp. 1213–1222, May 1999.
[10] M. G. Larimore, "Adaptation convergence of spectral estimation based on Pisarenko harmonic retrieval," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 955–962, Aug. 1983.
[11] Y. B. Hua and T. K. Sarkar, "On the total least squares linear prediction method for frequency estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2186–2189, Dec. 1990.
[12] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[13] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. Astonom. Soc.*, vol. 33, pp. 347–366, 1973.
[14] P. A. Thompson, "Adaptive spectral analysis technique for unbiased frequency estimation in the presence white noise," in *Proc. 13th Asilomar Conf. Circuits, Syst. Comput.*, 1979, pp. 529–533.
[15] R. Vaccaro, "On adaptive implementations of Pisarenko's harmonic retrieval method," in *Proc. ICASSP*, 1984, pp. 224–227.

[16] J. F. Yang and M. Kaveh, "Adaptive eigen-subspace algorithms for direction or frequency estimation and tracking," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 241–251, Feb. 1988.
[17] N. L. Owsley, "Adaptive data orthogonalization," in *Proc. ICASSP*, 1978, pp. 109–112.
[18] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Networks*, vol. 5, pp. 927–935, 1992.
[19] L. Xu, E. Oja, and C. Y. Suen, "Modified Hebbian learning for curve and surface fitting," *Neural Networks*, vol. 5, pp. 441–457, 1992.
[20] L. Wang and J. Karhunen, "A unified neural bi-gradient algorithm for robust PCA and MCA," *Int. J. Neural Syst.*, vol. 7, pp. 53–67, 1996.
[21] F. L. Luo, R. Unbehauen, and A. Cichocki, "A minor component analysis algorithm," *Neural Networks*, vol. 10, no. 2, pp. 291–297, 1997.
[22] K. Q. Gao, M. O. Ahmad, and M. N. S. Swamy, "A constrained anti-Hebbian learning algorithm for total least-squares estimation with applications to adaptive FIR and IIR filtering," *IEEE Trans. Circuits Syst. II*, vol. 41, pp. 718–729, Nov. 1994.
[23] C. S.-T. Choy and W.-C. Siu, "Analysis of the convergence and divergence of a constrained anti-Hebbian learning algorithm," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 1494–1502, Nov. 1998.
[24] Q. Zhang and Y. W. Leung, "Convergence of a Hebbian-type learning algorithm," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 12, pp. 1599–1601, Dec. 1999.
[25] V. Solo and X. Kong, "Performance analysis of adaptive eigen-analysis algorithms," *IEEE Trans. Signal Processing*, vol. 46, pp. 636–646, Mar. 1998.
[26] D.-Z. Feng, Z. Bao, and L.-C. Jiao, "Total least mean squares algorithm," *IEEE Trans. Signal Processing*, vol. 46, pp. 2212–2220, Aug. 1998.
[27] 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.
[28] H. Chen, T. Sarkar, S. Dianat, and J. Brule, "Adaptive spectral estimation by the conjugate gradient method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 272–284, Apr. 1986.
[29] V. U. Reddy, B. Edgardt, and T. Kailath, "Least squares-type algorithm for adaptive implementations of Pisarenko's harmonic retrieval method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 399–405, June 1982.
[30] C. E. Davila, "An efficient recursive total least squares algorithm for FIR adaptive filtering," *IEEE Trans. Signal Processing*, vol. 42, pp. 268–280, Feb. 1994.
[31] L. Ljung, M. Morf, and D. Falconer, "Fast calculation of gain matrices for recursive estimation schemes," *Int. J. Contr.*, vol. 7, no. 1, pp. 1–19, 1978.
[32] S. Ljung and L. Ljung, "Error propagation properties of recursive least-squares adaptation algorithms," *Automatica*, vol. 21, no. 2, pp. 157–167, 1985.
[33] J. L. Botto and G. V. Moustakides, "Stabilization of fast Kalman algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1342–1348, Sept. 1989.
[34] D. T. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Processing*, vol. 39, pp. 92–114, Jan. 1991.
[35] C. E. Davila, "Line search algorithms for adaptive filtering," *IEEE Trans. Signal Processing*, vol. 41, pp. 2490–2494, July 1993.
[36] ——, "Efficient, high performance, subspace tracking for time-domain data," *IEEE Trans. Signal Processing*, vol. 48, pp. 3307–3315, Dec. 2000.
[37] J. P. LaSalle, *The Stability of Dynamical Systems*. Philadelphia, PA: SIAM, 1976.

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 from Xidian University, Xi'an, in 1995.

From May 1996 to May 1998, he was a post-doctoral research affiliate and an associate professor with Xi'an Jiaotong University. From May 1998 to June 2000, he was an associate professor with Xidian University. Since July 2000, he has been a Professor with Xidian University. He has published more than 40 journal papers. His research interests include signal processing, intelligence information processing, and InSAR.

Xian-Da Zhang (SM'93) was born in 1946. He graduated from Xidian University, Xi'an, China, in 1970 and received the M.S. degree in instrument engineering from Harbin Institiue of Technology, Harbin, China, in 1982 and the Ph.D. degree in elelctrical engineering from Tohoku University, Sendai, Japan, in 1987.

From 1987 to 1992, he was a Research Professor with the Changcheng Institute of Metrology and Measurement, Beijing, China. From August 1990 to August 1991, he was a postdoctoral Resarcher with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla. Since 1992, he has been a Professor with the Department of Automatation, Tsinghua University, Beijing. Since April 1999, he has been with Key Laboratory for Radar Signal Processing, Xidian University as a specially Appointed Professor awarded by the Ministry of Education and the Cheung Kong Scholars Programme. He has published more than 70 papers and is the author of four books. His resarch interests include signal processing for communications, radar signal processing, nonstationary signal processing, and intelligent signal and information processing.

Dong-Xia Chang received the B.S. and M.S. degrees in applied mathematics from Xidian University, Xi'an, China, in 2000 and 2003, respectively.

Since April 2003, she has been with the General Software Laboratory, Institute of Software, Chinese Academy of Science, Beijing, China.

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 at the 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. He was also 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 with 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: SEU, 1991).

Dr. Zheng has received several science prizes, including 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) and the 41st IEEE Conference on Decision and Control (CDC'2002). 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 has been an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I since 2002 and an Associate Editor of the IEEE Control Systems Society's Conference Editorial Board since 2000.