Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

An Adaptive Algorithm for Fast Identification of IIR Systems

Da-Zheng Feng and Wei Xing Zheng

Abstract—**This paper considers the problem of adaptive identification of IIR systems when the system output is corrupted by noise. The standard recursive least squares algorithm is known to produce biased parameter estimates in this case. A new type of fast recursive identification algorithm is proposed which is built upon approximate inverse power iteration. The proposed adaptive algorithm can recursively compute the total least squares solution for unbiased adaptive identification of IIR systems. It is shown that the proposed adaptive algorithm has global convergence. The significant features of the proposed adaptive algorithm include efficient computation of the fast gain vector, adaptation of the inverse-power iteration, and rank-one update of the augmented covariance matrix. The proposed adaptive algorithm is superior to the standard recursive least squares algorithm and other recursive total least squares algorithms in such aspects as its ability for unbiased parameter estimation, its lower computational complexity, and its good long-term numerical stability. Computer simulation results that corroborate the theoretical findings are presented.**

I. INTRODUCTION

DURING the past decade, adaptive infinite impulse response (IIR) filters have found extensive applications in many areas, such as system identification, adaptive noise cancellation, spectral estimation, channel estimation and equalization in communication systems and so on [1], [2], [3]. Adaptive 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 enhanced the applicability of adaptive IIR filters.

This paper is concerned with adaptive identification of IIR systems when the system output is contaminated by noise. There are generally two main classes of methods for adaptive identification of IIR systems in this scenario. The first class of adaptive identification algorithms is the output-error (OE) method [4]. But the convergence as well as stability of the OE method is guaranteed only under the assumption that certain system transfer function is strictly positive real [5]. Given the fact that it is a highly nonlinear algorithm, the OE method has the difficulty in providing unbiased parameter estimates. The second class of adaptive estimation algorithms is the equation-error (EE) method [2]. Since the system model adopted is linear, the EE method for adaptive IIR identification can operate in a stable manner when the step size is properly selected. Moreover, the EE method has such attractive features as a unimodal error surface, good convergence and guaranteed stability when compared with the OE method. However, the EE method (including particularly the standard recursive least squares (RLS) algorithm) usually produces biased parameter estimates for IIR systems subject to output noise [6]. Some efficient algorithms to remove the bias in adaptive EE identification of IIR systems have been developed. For example, there are the instrumental variable algorithms [7], the hybrid algorithms [8], the unit-norm EE algorithms [9], the total least squares (TLS) algorithms [10] and the total least mean squares algorithm [11], to just mention a few. Among them, the Rayleight quotient (RQ) based TLS algorithms [10] have proved to be a viable alternative for achieving unbiased estimates for adaptive identification of IIR systems with output noise.

Regarding the computational complexity, the recursive TLS (RTLS) algorithms normally involve $O(L^2)$ operations per iteration, where L denotes the number of model parameters in an IIR system. Similarly, other adaptive algorithms (for instance, the inverse-power (IP) method [12]) also require $O(L^2)$ operations per eigenvector update. For on-line solving the TLS problem in adaptive identification, the fast RTLS algorithm proposed in [10] is based on the gradient search for the generalized RQ along the Kalman gain vector. 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. The RTLS algorithm in [10] has computational complexity of $O(L)$ per iteration, but is dependent on the fast computation of the Kalman gain vector. Unfortunately, it is a well-known fact that the computation of the Kalman gain vector may be potentially unstable [13].

In this paper, the problem of adaptive identification of IIR systems subject to output noise is investigated from a new point of view. A fast recursive identification algorithm is derived by approximating the well-known inverse power iteration along the data vector. Further, a fast scheme is developed to compute the gain vector that is used in adaptation. With guaranteed global convergence, the proposed approximate inverse power (AIP) algorithm is able to conduct recursive computation of the TLS solution so as

This work was supported in part by the National Natural Science Foundation of China and in part by a research grant from the Australian Research Council.

D.-Z. Feng is with the National Laboratory for Radar Signal Processing, Xidian University, Xi'an, 710071, China.

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

to achieve unbiased identification of IIR systems. The proposed algorithm is equipped with attractive features, such as efficient computation of the fast gain vector, adaptation of the inverse-power iteration, and rank-one update of the augmented covariance matrix. The proposed AIP algorithm is significantly advantageous over the standard RLS algorithm because the RLS algorithm produces biased estimates with computational complexity of $O(L^2)$ per iteration while the AIP algorithm gives unbiased estimates with computational complexity of $O(L)$ per iteration. Moreover, the proposed algorithm outperforms the fast RTLS algorithm in [10]. First, the computational complexity of the AIP algorithm is much less than that of the fast RTLS algorithm. Second, the long-term numerical stability of the proposed algorithm is much better than that of the fast RTLS algorithm, thanks to its efficient computation scheme for the fast gain vector and its no use of the inverse of the covariance matrix. The performance of the proposed AIP algorithm is evaluated via simulation and is compared with the standard RLS algorithm, the fast RTLS algorithm and the IP method.

II. PRELIMINARIES

A. System Description

Consider an unknown IIR system and assume that only the output is corrupted by the additive white Gaussian noise. Our task is to use an EE adaptive IIR filter to estimate the IIR system from the observations of the input and output. The parameter vector of the unknown IIR system is given by

$$\mathbf{h} = [a_1, a_2, \cdots, a_{N-1}, b_0, b_1, \cdots, b_{M-1}]^T \in R^{L\times 1} \tag{2.1}$$

where N and M are the denominator order and the numerator order of the IIR system transfer function, respectively, and $L = N + M - 1$. $\mathbf{h}$ may be time varying. Assume that the system orders N and M are known.

The noise-free system output is given by

$$\hat{d}(t) = \hat{\mathbf{r}}^T(t)\mathbf{h} \tag{2.2}$$

where the noise-free data vector $\hat{\mathbf{r}}(t)$ is given by $\hat{\mathbf{r}}(t) = [\hat{\mathbf{d}}^T(t), \mathbf{x}^T(t)]^T$ with $\hat{\mathbf{d}}(t) = [\hat{d}(t-1), \cdots, \hat{d}(t-N+1)]^T$ and $\mathbf{x}(t) = [x(t), \cdots, x(t-M+1)]^T$.

The observation output can be represented as

$$d(t) = \hat{d}(t) + n(t) = \hat{\mathbf{r}}^T(t)\mathbf{h} + n(t) \tag{2.3}$$

where the measurement noise $n(t)$ is a zero-mean Gaussian white noise with variance σ_o^2, independent of the input signal $x(t)$. The output for a sufficient-order EE adaptive IIR filter is given by

$$y(t) = \sum_{m=1}^{N-1} a_m d(t-m) + \sum_{m=0}^{M-1} b_m x(t-m) = \mathbf{r}^T(t)\mathbf{w} \tag{2.4}$$

where $\mathbf{w} = [\mathbf{a}^T, \mathbf{b}^T]^T$ is the adjustable parameter vector, $\mathbf{r}(t) = [\mathbf{d}^T(t), \mathbf{x}^T(t)]^T$ is the noisy data vector, and

$$\mathbf{a} = [a_1, \cdots, a_{N-1}]^T, \qquad \mathbf{b} = [b_0, \cdots, b_{M-1}]^T \tag{2.5}$$

$$\mathbf{d}(t) = [d(t-1), \cdots, d(t-N+1)]^T \tag{2.6}$$

Moreover, we have

$$\mathbf{d}(t) = \hat{\mathbf{d}}_{N-1}(t-1) + \mathbf{n}_{N-1}(t-1) \tag{2.7a}$$

$$\mathbf{n}_{N-1}(t-1) = [n(t-1), n(t-2), \cdots, n(t-N+1)]^T \tag{2.7b}$$

$$\hat{\mathbf{d}}_{N-1}(t-1) = [\hat{d}(t-1), \hat{d}(t-2), \cdots, \hat{d}(t-N+1)]^T \tag{2.7c}$$

At time t the augmented data vector is defined as

$$\bar{\mathbf{r}}(t) = [d(t), \mathbf{r}^T(t)]^T = [\bar{\mathbf{d}}^T(t), \mathbf{x}^T(t)]^T \tag{2.8a}$$

$$\bar{\mathbf{d}}(t) = \hat{\mathbf{d}}_N(t) + \mathbf{n}_N(t) = [d(t), \mathbf{d}^T(t)]^T \tag{2.8b}$$

The covariance matrix of the data vector is given by

$$\mathbf{R} = E\{\mathbf{r}(t)\mathbf{r}^T(t)\} = \begin{bmatrix} \mathbf{R}_{dd} & \mathbf{R}_{dx} \\ \mathbf{R}_{dx}^T & \mathbf{R}_{xx} \end{bmatrix} = \hat{\mathbf{R}} + \sigma_o^2 \begin{bmatrix} \mathbf{I}_{N-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{2.9}$$

where

$$\hat{\mathbf{R}} = E\{\hat{\mathbf{r}}(t)\hat{\mathbf{r}}^T(t)\} = \begin{bmatrix} \mathbf{R}_{\hat{d}\hat{d}} & \mathbf{R}_{\hat{d}x} \\ \mathbf{R}_{\hat{d}x}^T & \mathbf{R}_{xx} \end{bmatrix} \tag{2.10}$$

The covariance matrix of the augmented data vector is described by

$$\bar{\mathbf{R}} = E\{\bar{\mathbf{r}}(t)\bar{\mathbf{r}}^T(t)\} = \begin{bmatrix} c & \mathbf{g}^T \\ \mathbf{g} & \mathbf{R} \end{bmatrix} \tag{2.11}$$

where

$$\mathbf{g} = E\{\mathbf{r}(t)d(t)\} = \begin{bmatrix} \mathbf{g}_{dd} \\ \mathbf{g}_{dx} \end{bmatrix} = \hat{\mathbf{R}}\mathbf{h} \tag{2.12}$$

$$c = E\{d(t)d(t)\} = \mathbf{h}^T\hat{\mathbf{R}}\mathbf{h} + \sigma_o^2 \tag{2.13}$$

So the covariance matrix of the augmented data vector can be written as

$$\bar{\mathbf{R}} = \begin{bmatrix} \mathbf{h}^T\hat{\mathbf{R}}\mathbf{h} + \sigma_0^2 & \mathbf{h}^T\hat{\mathbf{R}} \\ \hat{\mathbf{R}}\mathbf{h} & \mathbf{R} \end{bmatrix} = \mathbf{R}^* + \mathbf{R}_o \tag{2.14}$$

where $\mathbf{R}^* = \begin{bmatrix} \mathbf{h}^T\hat{\mathbf{R}}\mathbf{h} & \mathbf{h}^T\hat{\mathbf{R}} \\ \hat{\mathbf{R}}\mathbf{h} & \hat{\mathbf{R}} \end{bmatrix}$ and $\mathbf{R}_o = \sigma_o^2 \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. It is easy to verify that $\mathbf{R}^* \begin{bmatrix} -1 \\ \mathbf{h} \end{bmatrix} = \mathbf{0}$. This means that if $\hat{\mathbf{R}}$ has full rank, then $\mathbf{R}^*$ is rank-one deficient.

B. Equation Error Method

The equation error (EE) [6], [9] is defined as

$$e(t) = y(t) - d(t) = \mathbf{w}^T\mathbf{r}(t) - d(t) = \bar{\mathbf{w}}^T\bar{\mathbf{r}}(t) \tag{2.15}$$

where $\bar{\mathbf{w}} = [-1, \mathbf{w}^T]^T = [-1, \mathbf{a}^T, \mathbf{b}^T]^T \in R^{(L+1)\times 1}$. Under the natural 'monic' constraint, the variance of the EE is given by

$$E\{e^2(t)\} = \bar{\mathbf{w}}^T\bar{\mathbf{R}}\bar{\mathbf{w}} = \bar{\mathbf{w}}^T\mathbf{R}^*\bar{\mathbf{w}} + \sigma_o^2[\bar{\mathbf{a}}^T\bar{\mathbf{a}}] \tag{2.16}$$

where $\bar{\mathbf{a}} = [a_0, \mathbf{a}^T]^T$. The first term on the right-hand side of (2.16) is associated with the noise-free situation, while the second term is seen to add an interference that will produce the estimation bias. This shows that the least-squares-type cost function (2.16) based on the 'monic' constraint does not yield the unbiased estimate for IIR systems subject to output noise. If $\bar{\mathbf{a}}^T\bar{\mathbf{a}}$ is constrained to 1, then (2.16) becomes

$$E\{e^2(t)\} = \overline{\mathbf{w}}^T \mathbf{R}^* \overline{\mathbf{w}} + \sigma_o^2 \tag{2.17}$$

Since the presence of the measurement noise adds only a constant term to the EE variance, the solution obtained by minimizing the cost function (2.17) does not vary with the noise variance. However, the final solution has to be obtained by the following scaling operation

$$\mathbf{w} = -[\overline{\mathbf{w}}]_{2,L+1} / a_0 \tag{2.18}$$

Note that for a vector $\mathbf{u} = [u_1, u_2, \cdots, u_L]^T \in R^{L\times 1}$, we define $[\mathbf{u}]_{m,n} = [u_m, u_{m+1}, \cdots, u_n]^T \in R^{(n-m+1)\times 1}$ for $1 \le m \le n \le L$. In order to find the solution for adaptive IIR filtering, the least-squares-type cost function in [6], [9] is to minimize

$$E\{e^2(t)\} = \overline{\mathbf{w}}^T \overline{\mathbf{R}} \overline{\mathbf{w}} \quad \text{subject to} \quad \overline{\mathbf{a}}^T \overline{\mathbf{a}} = 1 \tag{2.19}$$

On the other hand, to efficiently seek the TLS solution for adaptive estimation of IIR systems, the following Rayleigh quotient (RQ) is established in [10]:

$$J\{\overline{\mathbf{w}}(t)\} = \frac{\overline{\mathbf{w}}^T \overline{\mathbf{R}} \overline{\mathbf{w}}}{\overline{\mathbf{w}}^T \overline{\mathbf{D}} \overline{\mathbf{w}}} \tag{2.20}$$

where $\overline{\mathbf{D}} = diag\{\mathbf{I}_{N\times N}, \mathbf{0}_{M\times M}\}$. Clearly, (2.20) is like (2.19). Now consider minimizing the RQ $J(\overline{\mathbf{w}})$. Forcing the gradient of $J(\overline{\mathbf{w}})$ with respect to $\overline{\mathbf{w}}$ to be equal to zero leads to the special generalized eigenvalue decomposition associated with the matrix pair $(\overline{\mathbf{R}}, \overline{\mathbf{D}})$, that is,

$$\overline{\mathbf{R}}\overline{\mathbf{w}} - \lambda \overline{\mathbf{D}}\overline{\mathbf{a}} = 0, \quad \overline{\mathbf{a}}^T \overline{\mathbf{a}} = 1 \tag{2.21}$$

This shows that the unbiased parameter estimate of IIR systems can be also achieved by finding the generalized eigenvector associated with the smallest generalized eigenvalue of the matrix pair $(\overline{\mathbf{R}}, \overline{\mathbf{D}})$.

C. Inverse Power Iteration

The inverse-power (IP) iteration for finding the eigenvector associated with the smallest generalized eigenvalue of the matrix pair $(\overline{\mathbf{R}}, \overline{\mathbf{D}})$ is as follows [14].

Randomly produce the initial value $\overline{\mathbf{w}}(0)$*, for* $t = 1, 2, \cdots$

solve a set of linear equations

$$\overline{\mathbf{R}}\hat{\mathbf{w}}(t) = \overline{\mathbf{D}}\overline{\mathbf{w}}(t-1) \tag{2.22}$$

and perform the following normalization

$$\overline{\mathbf{w}}(t) = \hat{\mathbf{w}}(t) / \|\hat{\mathbf{w}}(t)\|. \tag{2.23}$$

It can be seen that the core of the IP iteration is (2.22). Moreover, it is shown in [14] that the IP iteration globally exponentially converges to the eigenvector associated with the smallest generalized eigenvalue of the matrix pair $(\overline{\mathbf{R}}, \overline{\mathbf{D}})$. Unfortunately, the IP iteration is an algorithm with computational complexity $O(L^3)$. Note that if the inverse update formula of $\overline{\mathbf{R}}$ is used like the RLS algorithms, the IP iteration will become an algorithm with computational complexity $O(L^2)$. However, the potential instability of the inverse update formula of the covariance matrix may cause the potential instability of such IP iteration.

Since we expect that the non-normalized eigenvector associated with the smallest generalized eigenvalue of the matrix pair $(\overline{\mathbf{R}}, \overline{\mathbf{D}})$ has the first entry to be fixed as -1, we may implement an IP iteration in conjunction with monic normalization, which is given as follows.

Randomly produce the initial value $\widetilde{\mathbf{w}}(0)$ *and let* $[\widetilde{\mathbf{w}}(0)]_{1,1} = -1$*, for* $t = 1, 2, \cdots$

solve a set of linear equations

$$\overline{\mathbf{R}}\hat{\mathbf{w}}(t) = \overline{\mathbf{D}}\widetilde{\mathbf{w}}(t-1) \tag{2.24}$$

and perform the following monic normalization

$$\widetilde{\mathbf{w}}(t) = \hat{\mathbf{w}}(t) / [\hat{\mathbf{w}}(0)]_{1,1}. \tag{2.25}$$

Using the eigenvalue decomposition (EVD), it can be shown that $\widetilde{\mathbf{w}}(t)$ given by the simple IP iteration (2.24) and (2.25) globally exponentially converges to the true augmented parameter vector $[-1, \mathbf{h}^T]^T$ as $t \to \infty$.

For implementation convenience, we may combine (2.24) and (2.25) together to get the following compact IP iteration.

Randomly produce the initial value $\mathbf{w}(0)$*, for* $t = 1, 2, \cdots$

solve a set of linear equations

$$\overline{\mathbf{R}}[-1, \mathbf{w}^T(t)]^T = \gamma(t)\overline{\mathbf{D}}[-1, \mathbf{w}^T(t-1)]^T. \tag{2.26}$$

Note that it is straightforward to establish the equivalence of (2.26) to (2.24) and (2.25).

III. The Approximate Inverse Power Algorithm

In this section, we develop an efficient algorithm for finding the TLS solution of the adaptive IIR filtering problem. This algorithm is an approximate inverse-power (AIP) iteration. Choosing update direction to be the data vector will also give rise to the computationally efficient algorithm with computational complexity $O(L)$.

The basic idea is to update the parameter vector in (2.26) by the following rank-one updating formula

$$\mathbf{w}(t) = \mathbf{w}(t-1) + \beta(t)\mathbf{r}(t) \tag{3.1a}$$

or equivalently

$$[-1, \mathbf{w}^T(t)]^T = [-1, \mathbf{w}^T(t-1)]^T + \beta(t)[0, \mathbf{r}^T(t)]^T \tag{3.1b}$$

Substituting (3.1) into (2.26) yields the approximate formula

$$\overline{\mathbf{R}}(t)[-1, \mathbf{w}^T(t-1) + \beta(t)\mathbf{r}^T(t)]^T \approx \gamma(t)\overline{\mathbf{D}}[-1, \mathbf{w}^T(t-1)]^T \tag{3.2}$$

where $\beta(t)$ and $\gamma(t)$ can be efficiently determined in $O(L)$ multiplications by the following equations

$$\begin{aligned}[0, \mathbf{r}^T(t)]\{\overline{\mathbf{R}}(t)[-1, (\mathbf{w}(t-1) + \beta(t)\mathbf{r}(t))^T]^T \\ - \gamma(t)\overline{\mathbf{D}}[-1, \mathbf{w}^T(t-1)]^T\} = 0\end{aligned} \tag{3.3a}$$

$$\begin{aligned}[-1, \mathbf{w}^T(t-1)]\{\overline{\mathbf{R}}(t)[-1, (\mathbf{w}(t-1) + \beta(t)\mathbf{r}(t))^T]^T \\ - \gamma(t)\overline{\mathbf{D}}[-1, \mathbf{w}^T(t-1)]^T\} = 0\end{aligned} \tag{3.3b}$$

Notice that $\overline{\mathbf{R}}(t)$ in adaptive identification is computed by a recursive formula

$$\overline{\mathbf{R}}(t) = \begin{bmatrix} c(t) & \mathbf{g}^T(t) \\ \mathbf{g}(t) & \mathbf{R}(t) \end{bmatrix} = \mu\overline{\mathbf{R}}(t-1) + \overline{\mathbf{r}}(t)\overline{\mathbf{r}}^T(t) \tag{3.4}$$

where

$$c(t) = \mu c(t-1) + d(t)d(t) \tag{3.5a}$$

$$\mathbf{g}(t) = \mu\mathbf{g}(t-1) + d(t)\mathbf{r}(t) \quad (3.5b)$$

$$\mathbf{R}(t) = \mu\mathbf{R}(t-1) + \mathbf{r}(t)\mathbf{r}^T(t) \quad (3.5c)$$

The μ in (3.4) and (3.5) is the forgetting factor. Substituting (3.4) and (3.5) into (3.3) yields

$$[\mathbf{r}^T(t)\mathbf{g}(t), \mathbf{r}^T(t)\mathbf{R}(t)][-1, (\mathbf{w}(t-1) + \beta(t)\mathbf{r}(t))^T]^T - \gamma(t)y_a(t) = 0 \quad (3.6a)$$

$$\begin{aligned} &-[-1, \mathbf{w}^T(t-1)]\overline{\mathbf{R}}(t)[-1, \mathbf{w}^T(t-1)]^T \\ &-\beta(t)[-1, \mathbf{w}^T(t-1)][\mathbf{r}^T(t)\mathbf{g}(t), \mathbf{r}^T(t)\mathbf{R}(t)]^T \\ &+\gamma(t)[1 + \mathbf{a}^T(t-1)\mathbf{a}(t-1)] = 0 \end{aligned} \quad (3.6b)$$

where

$$y_a(t) = \mathbf{a}^T(t-1)\mathbf{d}(t) \quad (3.7)$$

After some manipulations, (3.6) can be further represented as

$$\begin{aligned} &\beta(t)[\mathbf{r}^T(t)\mathbf{R}(t)\mathbf{r}(t)] - \gamma(t)y_a(t) \\ &= -\mathbf{r}^T(t)\mathbf{R}(t)\mathbf{w}(t-1) + \mathbf{r}^T(t)\mathbf{g}(t) \end{aligned} \quad (3.8a)$$

$$\begin{aligned} &-\beta(t)[\mathbf{r}^T(t)\mathbf{R}(t)\mathbf{w}(t-1) - \mathbf{r}^T(t)\mathbf{g}(t)] \\ &+\gamma(t)[1 + \mathbf{a}^T(t-1)\mathbf{a}(t-1)] \\ &= [-1, \mathbf{w}^T(t-1)]\overline{\mathbf{R}}(t)[-1, \mathbf{w}^T(t-1)]^T \end{aligned} \quad (3.8b)$$

Let

$$a_{11}(t) = \mathbf{r}^T(t)\mathbf{R}(t)\mathbf{w}(t-1) \quad (3.9a)$$

$$b_1(t) = -\mathbf{r}^T(t)\mathbf{R}(t)\mathbf{w}(t-1) + \mathbf{r}^T(t)\mathbf{g}(t) \quad (3.9b)$$

$$a_{22}(t) = 1 + \mathbf{a}^T(t-1)\mathbf{a}(t-1) \quad (3.9c)$$

$$b_2(t) = [-1, \mathbf{w}^T(t-1)]\overline{\mathbf{R}}(t)[-1, \mathbf{w}^T(t-1)]^T \quad (3.9d)$$

The matrix form of (3.8) is then given by

$$\begin{bmatrix} a_{11}(t) & -y_a(t) \\ b_1(t) & a_{22}(t) \end{bmatrix}\begin{bmatrix} \beta(t) \\ \gamma(t) \end{bmatrix} = \begin{bmatrix} b_1(t) \\ b_2(t) \end{bmatrix} \quad (3.10)$$

In order to compute efficiently the coefficients of (3.10), let

$$\mathbf{k}(t) = \mathbf{R}(t)\mathbf{r}(t) \quad (3.11)$$

$$b_2^0(t) = [-1, \mathbf{w}^T(t)]\overline{\mathbf{R}}(t)[-1, \mathbf{w}^T(t)]^T \quad (3.12)$$

Here the gain vector $\mathbf{k}(t)$ can be fast computed as shown in Table I.

TABLE I
FAST SCHEME FOR COMPUTING THE GAIN VECTOR

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

Note that the fast scheme for the gain vector $\mathbf{k}(t)$ shown in Table I is derived by the approach similar to [15]. It should be emphasized that unlike the well-known Kalman gain vector that is based on the matrix-inversion lemma and can be numerically unstable, the fast scheme for computing $\mathbf{k}(t)$ is independent of the matrix-inversion lemma, thereby being numerically stable.

Using the gain vector $\mathbf{k}(t)$, the coefficients $a_{11}(t)$ and $b_1(t)$ can be efficiently computed by

$$a_{11}(t) = \mathbf{r}^T(t)\mathbf{k}(\mathbf{t}) \quad (3.13)$$

$$b_1(t) = -\mathbf{k}^T(t)\mathbf{w}(t-1) + \mathbf{r}^T(t)\mathbf{g}(t) \quad (3.14)$$

Note that with the gain vector $\mathbf{k}(t)$ and the coefficients $a_{11}(t)$, $b_1(t)$ and $y(t)$, $b_2(t)$ and $b_2^0(t)$ can be efficiently computed by

$$\begin{aligned} b_2(t) &= [-1, \mathbf{w}^T(t-1)][\mu\overline{\mathbf{R}}(t-1) + \overline{\mathbf{r}}(t)\overline{\mathbf{r}}^T(t)][-1, \mathbf{w}^T(t-1)]^T \\ &= \mu b_2^0(t-1) + [y(t) - d(t)]^2 \end{aligned} \quad (3.15)$$

$$\begin{aligned} b_2^0(t) &= [-1, \mathbf{w}^T(t-1) + \beta(t)\mathbf{r}^T(t)]^T\overline{\mathbf{R}}(t) \\ &\quad [-1, \mathbf{w}^T(t-1) + \beta(t)\mathbf{r}^T(t)]^T \\ &= b_2(t) + 2\beta(t)[\mathbf{k}^T(t)\mathbf{w}(t-1) - \mathbf{g}^T(t)\mathbf{r}(t)] \\ &\quad + \beta^2(t)\mathbf{k}^T(t)\mathbf{r}(t) \\ &= b_2(t) - 2\beta(t)b_1(t) + \beta^2(t)a_{11}(t) \end{aligned} \quad (3.16)$$

Solving (3.10), it follows that

$$\beta(t) = [a_{22}(t)b_1(t) + y_a(t)b_2(t)]/[a_{11}(t)a_{22}(t) + y_a(t)b_1(t)] \quad (3.17)$$

TABLE II
AIP ALGORITHM

Initialize: $\mathbf{w}(0) = [0, 0, \cdots, 0]^T$, $b_2^0(0) = 0$, $\mu = 0.99 \sim 1.0$	
For $t = 1, 2, \cdots$	MAD's
1 update the data vector $\mathbf{r}(t)$	
2 update gain $\mathbf{k}(t)$ (see Table I)	$10L+8$
$y_a(t) = \mathbf{a}^T(t-1)\mathbf{d}(t)$	
3 $y_b(t) = \mathbf{b}^T(t)\mathbf{x}(t)$	L
$y(t) = y_a(t) + y_b(t)$	
4 $\mathbf{g}(t) = \mu\mathbf{g}(t-1) + \mathbf{r}(t)d(t)$	$2L$
5 $a_{11}(t) = \mathbf{r}^T(t)\mathbf{k}(t)$	L
6 $b_1(t) = -\mathbf{k}^T(t)\mathbf{w}(t-1) + \mathbf{r}^T(t)\mathbf{g}(t)$	$2L$
7 $a_{22}(t) = 1 + \mathbf{a}^T(t-1)\mathbf{a}(t-1)$	$N-1$
8 $b_2(t) = \mu b_2^0(t-1) + (y(t) - d(t))^2$	2
8 $\beta(t) = [a_{22}(t)b_1(t) + y_a(t)b_2(t)]$ $/[a_{11}(t)a_{22}(t) + y_a(t)b_1(t)]$	5
9 $\mathbf{w}(t) = \mathbf{w}(t-1) + \beta(t)\mathbf{r}(t)$	L
10 $b_2^0(t) = b_2(t) + \beta(t)[\beta(t)a_{11}(t) - 2b_1(t)]$	3
Total real MAD's:	$17L+N+17$

Notice that $\gamma(t)$ is not required for finding the TLS solution of adaptive IIR filtering. The above algorithm is called the approximate inverse-power (AIP) algorithm and is summarized in Table II.

Since the monic normalization is adopted, the above table does not include the tenth and eleventh manipulations in Table 1 in [10], which saves about $2L$ MAD's, where MAD's stands for the number of multiples and divides. Moreover, since the parameter vector being tracked is reduced to L dimensions from $L+1$ dimensions, more manipulations have been saved. It can be shown by computing the MAD's of each manipulations that the MAD's of the AIP algorithms is $17L+N+17$, while the MAD's of the RTLS algorithm [10] is $19L+3N+74$. This indicates that the AIP algorithm is able to achieve a substantial reduction in the computational complexity from the RTLS algorithm.

IV. Theoretical Analysis

Performing the generalized eigenvalue decomposition (GEVD) of the matrix pair $(\overline{\mathbf{R}},\overline{\mathbf{D}})$ gives

$$\overline{\mathbf{R}}\,\overline{\mathbf{V}}=\overline{\mathbf{D}}\overline{\mathbf{V}}\mathrm{diag}(\bar{\lambda}_1,\bar{\lambda}_2,\cdots,\bar{\lambda}_{L+1}) \text{ or } \overline{\mathbf{R}}\overline{\mathbf{v}}_j=\bar{\lambda}_j\overline{\mathbf{D}}\overline{\mathbf{v}}_j \quad (4.1)$$

where $\overline{\mathbf{V}}$ is the generalized eigenvector matrix, and $\overline{\mathbf{v}}_j$ and $\bar{\lambda}_j$ are the *j*-th generalized eigenvector and eigenvalue, respectively. Note that the eigenvalues are arranged in a descending order $\bar{\lambda}_1\ge\bar{\lambda}_2\ge\cdots\ge\bar{\lambda}_L>\bar{\lambda}_{L+1}$. This means that $\overline{\mathbf{D}}$ and $\overline{\mathbf{R}}$ have the EVD: $\overline{\mathbf{V}}^T\overline{\mathbf{D}}\overline{\mathbf{V}}=\mathbf{\Gamma}$ and $\overline{\mathbf{V}}^T\overline{\mathbf{R}}\overline{\mathbf{V}}=\mathbf{\Phi}$, where $\mathbf{\Gamma}=\mathrm{diag}(\rho_1,\rho_2,\cdots,\rho_{L+1})$, $\mathbf{\Phi}=\mathrm{diag}(\varphi_1,\varphi_2,\cdots,\varphi_{L+1})$, and $\bar{\lambda}_i=\varphi_i/\rho_i$ ($i=1,\ldots,L+1$).

The following lemma is obvious.

Lemma 4.1 If $\hat{\mathbf{R}}$ is of full rank, then $\bar{\lambda}_L>\bar{\lambda}_{L+1}$.

By substituting $\overline{\mathbf{w}}=[-1,\mathbf{w}^T]^T$ into (2.20), we have the cost function

$$\bar{J}(\mathbf{w})=[-1,\mathbf{w}^T]\overline{\mathbf{R}}[-1,\mathbf{w}^T]^T/(1+\mathbf{a}^T\mathbf{a}) \quad (4.1)$$

The next Theorem 4.1 guarantees that we can search the global minimum point of $\bar{J}(\mathbf{w})$ by the gradient descent method. Its proof is omitted here due to limited space.

Theorem 4.1 If $\bar{\lambda}_L>\bar{\lambda}_{L+1}$ and $\bar{v}_{1,L+1}\neq 0$, where $\bar{v}_{1,L+1}$ is the $(L+1)$-th element of the first row of $\overline{\mathbf{V}}$, then $\mathbf{w}_{L+1}=-\hat{\mathbf{v}}_{L+1}/\bar{v}_{1,L+1}$ is the global minimum point of $\bar{J}(\mathbf{w})$. All the other stationary points are the saddle points of $\bar{J}(\mathbf{w})$.

We can show that the proposed AIP algorithm is globally convergent. To this end, a lemma is first stated.

Lemma 4.2 If t is large enough so that $\overline{\mathbf{R}}(t)\to\overline{\mathbf{R}}$, then it always holds that $\bar{J}(\mathbf{w}(t))-\bar{J}(\mathbf{w}(t-1))\le 0$.

Theorem 4.2 Assuming that t is large enough so that $\overline{\mathbf{R}}(t)\to\overline{\mathbf{R}}$, then $\mathbf{w}(t)\to\mathbf{h}$ with probability 1 as $t\to\infty$.

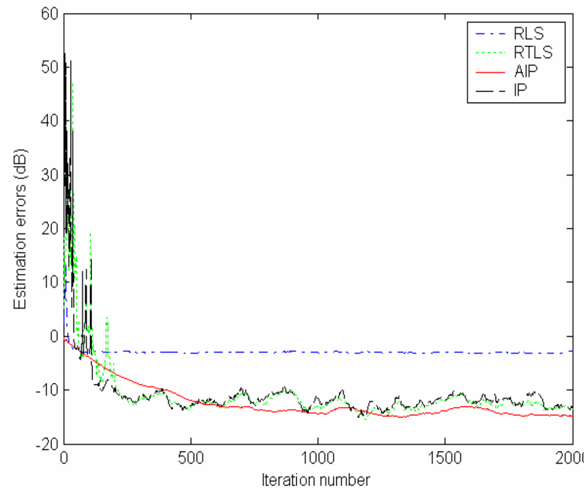

Fig. 1. AIP, IP, RLS and RTLS are used to identify a linear time-invariant IIR system. The errors denoted by dash-dot line, dashed line, dotted line and solid line are obtained by AIP, IP, RLS and RTLS, respectively.

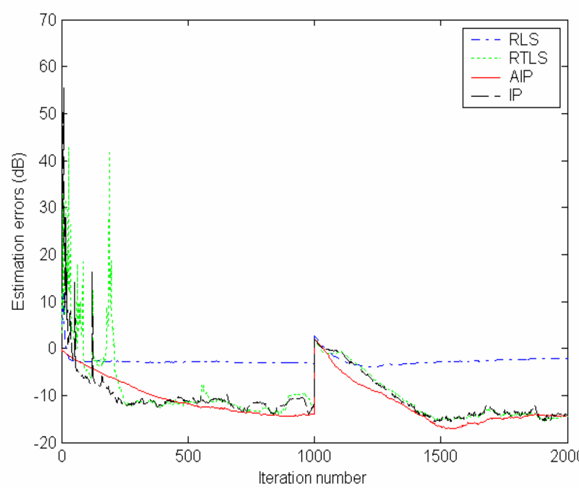

Fig. 2. AIP, IP, RLS and RTLS are used to identify a linear time-varying IIR system. The errors denoted by dash-dot line, dashed line, dotted line and solid line are obtained by AIP, IP, RLS and RTLS, respectively.

The proofs of Lemma 4.2 and Theorem 4.2 are omitted here due to limited space.

V. Simulation Results

Computer simulations have been conducted to verify the performance of the proposed AIP algorithm. Here all the simulation results are averaged over 30 independent tests.

The unknown linear IIR system is defined by [16]

$$H(z)=\frac{1+0.4z^{-2}-0.5z^{-3}+0.7z^{-4}}{1-1.1z^{-1}+0.8z^{-2}-0.88z^{-3}+0.64z^{-4}}.$$

Thus, $N=5$, $M=5$ and $L=N+M-1=9$. The standard RLS, the RTLS in [10], the IP in [12], and the proposed AIP algorithms are applied to the system identification experiment for comparison. The IIR system is excited by the colored input signal $x(t)$, which is generated by the first-order AR model $x(t)-0.25x(t-1)=\omega(t)$, where $\omega(t)$ is a white noise with unit variance. The IIR system output is contaminated by additive, zero-mean and white Gaussian noise with a unit variance, which is statistically independent of the input signal $x(t)$. The forgetting factor μ used in simulation is chosen to be 0.99. The estimation error is defined by

$$E(t)=\|\mathbf{w}(t)-\mathbf{h}\|_2/\|\mathbf{h}\|_2$$

where $\mathbf{h}=[1.1,-0.8,\ 0.88,-0.64,\ 1.0,\ 0.0,\ 0.4,-0.5,0.7]^T$.

When the system is linear time-invariant, the estimation results of the RLS, RTLS, IP and AIP algorithms are shown in Fig. 1. In order to test the tracking behavior of the relative algorithms in a nonstationary environment, the parameter estimation experiment is repeated, but each of the numerator parameters of the unknown IIR system will undergo a sign change at $t=1001$. The obtained simulation results are plotted in Fig. 2. It is interesting to see from Figs. 1 and 2 that the AIP algorithm has the slightly slow convergence in the first segment of the learning curves, but it has the significantly small statistical fluctuation in comparison with the IP and RTLS algorithms. Note that the AIP algorithm is implemented at a numerical cost much lower than the RTLS algorithm. As expected, the standard RLS algorithm fails to work because it produces biased parameter estimates. The numerical stability of the AIP algorithm has also been tested via a long-time ($t=2\times10^5$) experiment with satisfactory results as shown in Fig. 3.

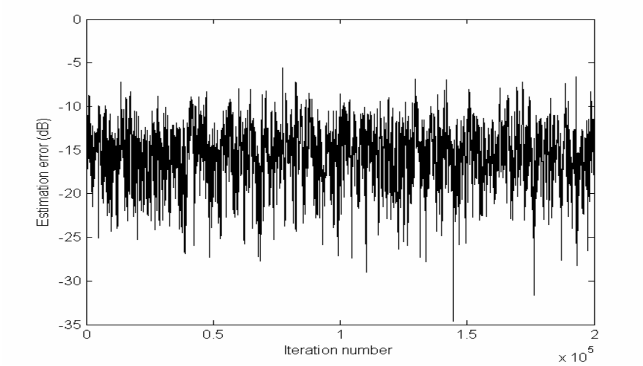

Fig. 3. The long-time numerical stability of the proposed AIP algorithm is tested.

VI. Concluding Remarks

The problem of adaptive identification of IIR systems subject to output noise has been studied in this paper. A fast algorithm has been proposed to recursively compute the TLS solution for unbiased estimation of IIR systems. The main idea is to efficiently compute the non-normalized eigenvector associated with the smallest eigenvalue of the sample covariance matrix via approximate inverse-power iteration and fast gain vector computation. It has been demonstrated that in addition to its good long-term numerical stability, the proposed AIP algorithm has a significant computational advantage over the fast RTLS algorithm in [10] and the IP algorithm in [12]. The global convergence and the estimation unbiasedness of the proposed algorithm have been established. The good performance of the AIP algorithm has been verified via computation simulations.

References

[1] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 1996.

[2] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. New York: Wiley, 1999.

[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, no. 1, pp. 54-66, Feb. 1995.

[4] M. Bouvet and S. C. Schwartz, "Comparison of adaptive and robust receivers for signal detection in ambient underwater noise," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-37, no. 5, pp. 621-626, May 1989.

[5] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Automat. Contr.*, vol. AC-21, no. 2, pp.194-202, Apr. 1976.

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

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

[8] J. E. Cousseau and P. S. R. Diniz, "A general consistent equation-error algorithm for adaptive IIR filtering," *Signal Processing*, vol. 56, no. 2, pp. 121-134, Jan. 1997.

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

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

[11] D.-Z. Feng, Z. Bao, and L.-C. Jiao, "Total Least Mean Squares Algorithm," *IEEE Trans. Signal Processing*, vol. 46, no. 8, pp. 2212-2220, Aug. 1998.

[12] R. Vaccaro, "On adaptive implementations of Pisarenko's harmonic retrieval method," in *Proc. ICASSP'84*, March 1984, pp. 224-227

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

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

[15] 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.

[16] R. Lopez-Valcare, "An algorithm for unit-norm equation error system identification based on the method of multipliers," *IEEE Trans. Signal Processing*, vol. 51, no. 12, pp. 3080-3085, Dec. 2003.

[17] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci. USA*, vol. 79, pp. 2554-2558, 1962.