# Stochastic Mean-Square Performance Analysis of an Adaptive Hammerstein Filter

Janez Jeraj, *Member, IEEE,* and V. John Mathews, *Fellow, IEEE*

***Abstract*—This paper presents an almost sure mean-square performance analysis of an adaptive Hammerstein filter for the case when the measurement noise in the desired response signal is a martingale difference sequence. The system model consists of a series connection of a memoryless nonlinearity followed by a recursive linear filter. A bound for the long-term time average of the squared *a posteriori* estimation error of the adaptive filter is derived using a basic set of assumptions on the operating environment. This bound consists of two terms, one of which is proportional to a parameter that depends on the step size sequences of the algorithm and the other that is inversely proportional to the maximum value of the increment process associated with the coefficients of the underlying system. One consequence of this result is that the long-term time average of the squared *a posteriori* estimation error can be made arbitrarily close to its minimum possible value when the underlying system is time-invariant.**

***Index Terms*—Adaptive filters, convergence analysis, Hammerstein filter, nonlinear systems.**

## I. Introduction

THIS paper describes a theoretical performance evaluation of an adaptive algorithm employing a Hammerstein system model. The system model consists of a series connection of a memoryless polynomial system followed by a recursive linear system as shown in Fig. 1. Several researchers have described algorithms for identifying cascade nonlinear models and/or analyzed their properties. Hunter and Korenberg [1] described the use of such algorithms in analysis of biological systems. The behavior of the adaptive gradient search algorithms for an LNL nonlinear system (a cascade of a linear system, a memoryless nonlinearity and another linear system) was investigated in [2] and [3]. The work was an extension of the work in [4] that considered Wiener systems (a linear system followed by a nonlinear system). Wiener systems were also considered in [5], whereas [6] and [7] considered both Wiener and Hammerstein systems. These works utilized system models with finite memory. A nonparametric algorithm to identify Hammerstein systems with a finite memory linear component was analyzed in [8]. A similar model was also used in [9] with a different recursive algorithm and a different analysis method. In spite of the existence of these and related works, there are limited or no convergence and stability analyses for adaptive algorithms employing recursive nonlinear system models. In particular, the authors are aware of no algorithm that is known to converge to the global minimum of the error surface. However, there are a number of algorithms in linear adaptive IIR filtering theory for which appropriate convergence and stability algorithms are available [10]–[17], and it may be possible to extend such results to Hammerstein models. The work in this paper is based on the analysis in [13], [14].

The input-output relationship of the adaptive filter is given by

$$\hat{d}(n) = -\sum_{i=1}^{N} \hat{a}_i(n) \cdot \hat{d}(n-i) + \sum_{j=0}^{M} \hat{b}_j(n) \cdot \hat{z}(n-j) \quad (1)$$

where $\hat{b}_0(n) = 1, \quad \forall n$, and $\hat{z}(n)$ is the output of the memoryless polynomial nonlinear system and is given by

$$\hat{z}(n) = \sum_{l=1}^{L} \hat{w}_l(n) x^l(n). \quad (2)$$

In the above equations, $\hat{w}_l(n)$, $\hat{a}_i(n)$ and $\hat{b}_j(n)$ represent the coefficients of the adaptive filter.

A detailed derivation of the algorithm as well as experimental performance evaluation can be found in [18]. The work in [18] used Lyapunov stability criterion to control the step sizes of the adaptive filter and to guarantee that the system operated in a stable manner. Experimental results presented in the paper also indicated that the system may converge to the global minimum of the error surface. The analysis of this paper shows under appropriate assumptions that for sufficiently small step sizes and stationary operating environments, the long-term time average of the excess squared estimation error can be arbitrarily close to its minimum possible value.

The rest of this paper is organized as follows. Section II provides a summary of the adaptive filter, and presents some auxiliary lemmas and assumptions. This section also contains discussion of a transformation of the adaptive filter to a functionally equivalent, but structurally different system. The main result of the paper is developed in Section III by analyzing the transformed system. Simulation results demonstrating the validity of the analysis in the paper are provided in Section IV. Finally, Section V contains the concluding remarks.

## II. Overview of the Adaptive Filter and the Analysis Model

### A. Adaptive Filter Structure

Given the input signal $x(n)$ and the desired response signal $d(n)$, the adaptive filter updates its coefficients using

Manuscript received September 3, 2004; revised June 9, 2005. A version of this paper was presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fulvio Gini.

J. Jeraj was with the Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112 USA (e-mail: jeraj@eng.utah.edu).

V. J. Mathews is with the Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112 USA.

Digital Object Identifier 10.1109/TSP.2006.873587

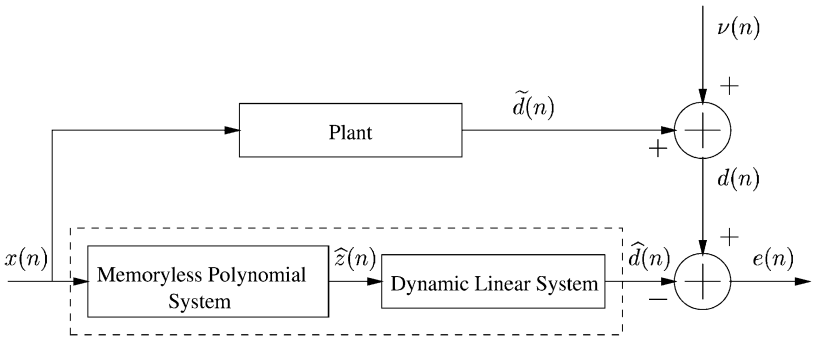


Fig. 1. Block diagram of the adaptive Hammerstein system.

TABLE I
THE ADAPTIVE HAMMERSTEIN FILTER

**Definitions**

$$
\begin{array}{lcl}
\delta & \ldots & \text{small positive constant} \\
\beta & \ldots & \text{small positive constant such that } 0 < \beta < 1 \\
\hat{\theta}(n) & = & \left[\ \hat{a}_1(n) \cdots \hat{a}_N(n) \quad \hat{b}_1(n) \cdots \hat{b}_M(n) \quad \hat{w}_1(n) \cdots \hat{w}_L(n)\ \right]^T \quad , \quad \hat{b}_0(n) = 1 \\
\hat{\mathbf{H}}(n) & = & \left[\ -\hat{d}(n-1) \cdots -\hat{d}(n-N) \quad \hat{z}(n-1) \cdots \hat{z}(n-M) \quad x(n) \cdots x^L(n)\ \right]^T \\
\hat{\mathbf{p}}(n) & = & \left[\ \hat{w}_1(n) \quad \hat{w}_2(n) \quad \ldots \quad \hat{w}_L(n)\ \right]^T \\
\mathbf{x}(n) & = & \left[\ x(n) \quad x^2(n) \quad \ldots \quad x^L(n)\ \right]^T \\
\otimes & \ldots & \text{Kronecker product} \\
\mathbf{A}(n) & = & \begin{bmatrix} -\hat{a}_1(n) & -\hat{a}_2(n) & \ldots & -\hat{a}_{N-1}(n) & -\hat{a}_N(n) \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \\
\Lambda(n) & = & diag\left[\ \mu_1(n) \quad \cdots \quad \mu_{N+M+L}(n)\ \right], \ \mu_1(n), \ldots, \mu_{N+M+L}(n) > 0
\end{array}
$$

**Main Loop**

$$e(n) = d(n) - \hat{\mathbf{H}}^T(n) \cdot \hat{\theta}(n-1)$$

$$\psi(n) = \left[\ -\hat{d}(n-1) \cdots -\hat{d}(n-N) \quad \hat{z}(n-1) \cdots \hat{z}(n-M) \quad \sum_{j=0}^{M} \hat{b}_j(n) x(n-j) \cdots \sum_{j=0}^{M} \hat{b}_j(n) x^L(n-j)\ \right]^T$$

$$\phi(n) = \psi(n) - \sum_{s=1}^{N} \hat{a}_s(n-1) \cdot \phi(n-s)$$

Verify that $\{\mu_i(n)\}$ are such that $\|vec[\mathbf{Q}(n+1)] - vec[\mathbf{Q}(n)]\| < 1$, where

$vec[\mathbf{Q}(n+1)] = -\left[\mathbf{A}^T(n) \otimes \mathbf{A}^T(n) - \mathbf{I}_{k^2}\right]^{-1} vec[\mathbf{I}_k]$, and $\hat{\mathbf{H}}^T(n)\Lambda(n)\phi(n) > -\delta\beta$. If the conditions are not satisfied, reduce elements of $\Lambda(n)$ so that they are fulfilled.

$$\hat{\theta}(n) = \hat{\theta}(n-1) + \frac{\Lambda(n)\phi(n)}{\delta + \hat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)} \cdot e(n)$$

$$\hat{z}(n) = \hat{\mathbf{p}}^T(n) \cdot \mathbf{x}(n)$$

$$\hat{d}(n) = \hat{\mathbf{H}}^T(n) \cdot \hat{\theta}(n)$$

a stochastic gradient algorithm in an attempt to reduce $E[(d(n) - \hat{d}(n))^2]$ after each iteration. The algorithm for adapting these coefficients is given in Table I.

We can rewrite (1) in operator notation as

$$\hat{d}(n) = \frac{\hat{B}(n, q^{-1})}{\hat{A}(n, q^{-1})} \hat{z}(n) \tag{3}$$

where

$$\hat{A}(n, q^{-1}) = 1 + \hat{a}_1(n)q^{-1} + \cdots + \hat{a}_N(n)q^{-N} \tag{4}$$

$$\hat{B}(n, q^{-1}) = 1 + \hat{b}_1(n)q^{-1} + \cdots + \hat{b}_M(n)q^{-M} \tag{5}$$

and $q^{-1}$ represents the unit delay operator. Let

$$\hat{\boldsymbol{\theta}}(n) = [\hat{a}_1(n) \cdots \hat{a}_N(n) \quad \hat{b}_1(n) \cdots \hat{b}_M(n) \quad \hat{w}_1(n) \cdots \hat{w}_L(n)]^T \tag{6}$$

with $\hat{b}_0(n) = 1$ and let

$$\hat{\mathbf{H}}(n) = [-\hat{d}(n-1)\cdots-\hat{d}(n-N) \quad \hat{z}(n-1)\cdots\hat{z}(n-M) \quad x(n)\cdots x^L(n)]^T. \quad (7)$$

Then, (1) can be rewritten using vector notation as

$$\hat{d}(n) = \hat{\boldsymbol{\theta}}^T(n)\cdot\hat{\mathbf{H}}(n). \quad (8)$$

The following reformulation of the output equation will find use later in the analysis. Let us define the vectors $\hat{\boldsymbol{\theta}}_c(n)$, $\hat{\mathbf{H}}_c(n)$, $\hat{\mathbf{p}}(n)$ and $\mathbf{x}(n)$ as

$$\hat{\boldsymbol{\theta}}_c(n) = \Big[\hat{a}_1(n) \quad \cdots \quad \hat{a}_N(n)\hat{b}_1(n)\hat{\mathbf{p}}^T(n-1) \quad \cdots \quad \hat{b}_M(n)\hat{\mathbf{p}}^T(n-M) \quad \hat{\mathbf{p}}^T(n) \quad \overbrace{0\cdots0}^{2L}\Big]^T \quad (9)$$

$$\hat{\mathbf{H}}_c(n) = \Big[-\hat{d}(n-1)\cdots-\hat{d}(n-N) \quad \mathbf{x}^T(n-1) \quad \cdots \quad \mathbf{x}^T(n-M) \quad \mathbf{x}^T(n) \quad \overbrace{0\cdots0}^{2L}\Big]^T \quad (10)$$

$$\hat{\mathbf{p}}(n) = [\hat{w}_1(n) \quad \hat{w}_2(n) \quad \ldots \quad \hat{w}_L(n)]^T \quad (11)$$

and

$$\mathbf{x}(n) = [x(n) \quad x^2(n) \quad \ldots \quad x^L(n)]^T \quad (12)$$

respectively. It is straightforward to show that $\hat{d}(n)$ can be equivalently written as

$$\hat{d}(n) = \hat{\boldsymbol{\theta}}_c^T(n)\cdot\hat{\mathbf{H}}_c(n). \quad (13)$$

The need for additional zeros in the definitions of $\hat{\boldsymbol{\theta}}_c(n)$ and $\hat{\mathbf{H}}_c(n)$ will become clear later in the analysis.

The algorithm updates the parameters of the adaptive filter as

$$\hat{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n-1) + \frac{\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta + \hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}e(n) \quad (14)$$

where $e(n) = d(n) - \hat{\boldsymbol{\theta}}^T(n-1)\hat{\mathbf{H}}(n)$ denotes the *a priori* estimation error of the system and $\boldsymbol{\Lambda}(n)$ is a diagonal matrix with positive step sizes on the diagonal $\boldsymbol{\Lambda}(n) = \text{diag}[\mu_1(n), \mu_2(n), \ldots \mu_{N+M+L}(n)]$. The *a posteriori* estimation error $\epsilon(n)$ can be manipulated as

$$\epsilon(n) = d(n) - \hat{\mathbf{H}}^T(n)\hat{\boldsymbol{\theta}}(n) \quad (15)$$

$$= d(n) - \hat{\mathbf{H}}^T(n)\left\{\hat{\boldsymbol{\theta}}(n-1) + \frac{\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)e(n)}{\delta + \hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}\right\} \quad (16)$$

$$= e(n)\frac{\delta}{\delta + \hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}. \quad (17)$$

Using (17) in (14) gives us

$$\hat{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n-1) + \frac{1}{\delta}\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)\epsilon(n). \quad (18)$$

### B. Analysis Model

We assume that the adaptive filter is operating in the system identification mode and that the system model matches the unknown system exactly or overmodels it. The input-output relationship of the plant is given by

$$\tilde{d}(n) = -\sum_{i=1}^{N} a_i(n)\cdot\tilde{d}(n-i) + \sum_{j=0}^{M} b_j(n)\cdot z(n-j) \quad (19)$$

where $z(n)$ is the output of a memoryless polynomial nonlinear system

$$z(n) = \mathbf{p}^T(n)\mathbf{x}(n) \quad (20)$$

with

$$\mathbf{p}(n) = [w_1(n) \quad w_2(n) \quad \ldots \quad w_L(n)]^T \quad (21)$$

and $b_0(n) = 1$. The desired response signal is a noisy version of the output of the unknown system, and is given by

$$d(n) = \tilde{d}(n) + \nu(n) \quad (22)$$

as shown in the Fig. 1. In the above equation, $\nu(n)$ is an additive noise sequence that is uncorrelated with the input signal. We, thus, assume the following model for the unknown system:

$$\tilde{d}(n) = \frac{B(n, q^{-1})}{A(n, q^{-1})}z(n) \quad (23)$$

where

$$A(n, q^{-1}) = 1 + a_1(n)q^{-1} + \cdots + a_N(n)q^{-N} \quad (24)$$

$$B(n, q^{-1}) = 1 + b_1(n)q^{-1} + \cdots + b_M(n)q^{-M} \quad (25)$$

which gives us

$$\tilde{d}(n) = \boldsymbol{\theta}_c^T(n)\cdot\mathbf{H}_c(n) \quad (26)$$

where

$$\boldsymbol{\theta}_c(n) = [a_1(n)\cdots a_N(n) \quad b_1(n)\mathbf{p}^T(n-1) \quad \cdots \quad b_M(n)\mathbf{p}^T(n-M) \quad \mathbf{p}^T(n) \quad \overbrace{0\cdots0}^{2L}]^T \quad (27)$$

and

$$\mathbf{H}_c(n) = [-d(n-1)\cdots-d(n-N) \quad \mathbf{x}^T(n-1) \quad \cdots \quad \mathbf{x}^T(n-M) \quad \mathbf{x}^T(n) \quad \overbrace{0\cdots0}^{2L}]^T. \quad (28)$$

### C. A Transformation of the Adaptive Filter

In order to analyze the algorithm, we transform the equations into an equivalent, but different structural form. For this, we first add $2L$ zeros to the vectors $\hat{\boldsymbol{\theta}}(n)$, $\hat{\boldsymbol{\theta}}(n-1)$ and $\boldsymbol{\phi}(n)$ in (18) to get

$$\hat{\boldsymbol{\theta}}_e(n) = \hat{\boldsymbol{\theta}}_e(n-1) + \boldsymbol{\Lambda}_e(n)\boldsymbol{\phi}_e(n)\epsilon(n) \quad (29)$$

where $\hat{\boldsymbol{\theta}}_e^T(n) = [\hat{\boldsymbol{\theta}}^T(n) \quad \overbrace{0\cdots0}^{2L}]$, $\boldsymbol{\phi}_e^T(n) = [\boldsymbol{\phi}^T(n) \quad \overbrace{0\cdots0}^{2L}]$ and $\boldsymbol{\Lambda}_e(n)$ is an $(N+M+3\cdot L)\times(N+M+3\cdot L)$-element matrix defined as shown in (30) at the bottom of the next page. The expanded "step size" matrix $\boldsymbol{\Lambda}_e(n)$ contains zeroes in the last $2L$ rows, and $(N+M+L)2L$ nonzero terms $\rho_{i,l}(n)$ in the off diagonal entries as shown in (30). The $\rho_{i,l}(n)$ terms are placed

in the first through $(N+M+L)$th rows of the $(N+M+L+1)$th through $(N+M+3L)$th columns. The new elements $\rho_{i,l}(n)$ are placed at locations such that when $\mathbf{\Lambda}_e(n)$ is multiplied with $\boldsymbol{\phi}_e(n)$, the zero entries of $\boldsymbol{\phi}_e(n)$ cancel the $\rho_{i,l}(n)$'s. Thus, the choice of these variables does not affect the update equation.

It is straightforward to show that there exist vectors $\hat{\boldsymbol{\theta}}_{r2}(n-1)$ and $\hat{\mathbf{H}}_d(n)$ such that

$$\mathbf{\Lambda}_e(n)\boldsymbol{\phi}_e(n)\epsilon(n) = \mathbf{\Lambda}_e(n)\hat{\boldsymbol{\theta}}_{r2}(n-1)+\mathbf{\Lambda}_e(n)\hat{\mathbf{H}}_d(n)\epsilon(n) \quad (31)$$

for any $\epsilon(n)$ which also includes $\epsilon(n)=0$. We only note at this time that there are more variables in the two vectors than there are equations, and therefore a multitude of solutions exists for them. Appendix I contains further discussions on choosing these and other parameters employed in the analysis. Employing (31) in (29), we get

$$\hat{\boldsymbol{\theta}}_e(n) = \hat{\boldsymbol{\theta}}_e(n-1) + \mathbf{\Lambda}_e(n)\hat{\boldsymbol{\theta}}_{r2}(n-1) + \mathbf{\Lambda}_e(n)\hat{\mathbf{H}}_d(n)\epsilon(n). \quad (32)$$

Next, we multiply both sides of the $(N+1)$th through the $(N+M)$th entries of (32) with $\hat{\mathbf{p}}(n-1),\ldots,\hat{\mathbf{p}}(n-M)$, respectively, to obtain

$$\begin{bmatrix} \hat{a}_1(n) \\ \vdots \\ \hat{a}_N(n) \\ \hat{b}_1(n)\hat{\mathbf{p}}(n-1) \\ \vdots \\ \hat{b}_M(n)\hat{\mathbf{p}}(n-M) \\ \hat{\mathbf{p}}(n) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{a}_1(n-1) \\ \vdots \\ \hat{a}_N(n-1) \\ \hat{b}_1(n-1)\hat{\mathbf{p}}(n-2) \\ \vdots \\ \hat{b}_M(n-1)\hat{\mathbf{p}}(n-M-1) \\ \hat{\mathbf{p}}(n-1) \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \hat{\boldsymbol{\theta}}_r(n-1) + \mathbf{\Lambda}_c(n)\hat{\boldsymbol{\theta}}_{r2c}(n-1) + \mathbf{\Lambda}_c(n)\hat{\mathbf{H}}_{dc}(n)\epsilon(n) \quad (33)$$

where $\hat{\boldsymbol{\theta}}_r(n-1)$ is defined as

$$\hat{\boldsymbol{\theta}}_r(n-1) = \Big[\overbrace{0\cdots0}^{N} \quad \hat{b}_1(n-1)\{\hat{\mathbf{p}}^T(n-1) - \hat{\mathbf{p}}^T(n-2)\}\cdots\hat{b}_M(n-1)\{\hat{\mathbf{p}}^T(n-M) - \hat{\mathbf{p}}^T(n-M-1)\} \quad \overbrace{0\cdots0}^{L} \quad \overbrace{0\cdots0}^{2\cdot L}\Big]^T \quad (34)$$

and $\hat{\boldsymbol{\theta}}_{r2c}(n-1)$ and $\hat{\mathbf{H}}_{dc}(n)$ are vectors that result when we multiply the $(N+1)$th through the $(N+M)$th entries of $\hat{\boldsymbol{\theta}}_{r2}(n-1)$ and $\hat{\mathbf{H}}_d(n)$ with $\hat{\mathbf{p}}(n-1),\ldots,\hat{\mathbf{p}}(n-M)$, respectively. The $(N+ML+3\cdot L)\times(N+ML+3\cdot L)$-element matrix $\mathbf{\Lambda}_c(n)$ has a similar interpretation. It is obtained by replacing $\mu_{N+1}(n)$ through $\mu_{N+M}(n)$ with $(L\times L)$-element diagonal matrices $\mu_{N+1}(n)\mathbf{I}, \mu_{N+2}(n)\mathbf{I},\ldots,\mu_{N+M}(n)\mathbf{I}$, respectively. Also, each element $\rho_{N+1,1}(n)$ through $\rho_{N+M,2L}(n)$ of the matrix $\mathbf{\Lambda}_e(n)$ is replaced by an $(L\times 1)$-element column vector $[\rho_{N+1,1}(n)\cdots\rho_{N+1,1}(n)]^T$ through $[\rho_{N+M,2L}(n)\cdots\rho_{N+M,2L}(n)]^T$, respectively. Recognizing that the left-hand side (LHS) of (33) is $\hat{\boldsymbol{\theta}}_c(n)$ defined in (9), we can write this equation more compactly as

$$\hat{\boldsymbol{\theta}}_c(n) = \hat{\boldsymbol{\theta}}_c(n-1) + \hat{\boldsymbol{\theta}}_r(n-1) + \mathbf{\Lambda}_c(n)\hat{\boldsymbol{\theta}}_{r2c}(n-1) + \mathbf{\Lambda}_c(n)\hat{\mathbf{H}}_{dc}(n)\epsilon(n). \quad (35)$$

Recall that the adaptive filter is composed of a polynomial with $L$ coefficients and an IIR system with a denominator and numerator having $N$ and $M$ taps, respectively. With the above transformation, we expanded (18) from an $(N+M+L)\times 1$-dimensional equation to a vector (35) with $(N+M\cdot L+3\cdot L)$ dimensions. Let $\gamma$ be a positive, finite constant of our choice. It can be shown that we can choose the "new" entries of $\mathbf{\Lambda}_c(n)$, and the vectors $\hat{\boldsymbol{\theta}}_{r2c}(n-1)$ and $\hat{\mathbf{H}}_{dc}(n)$ such that the following equality is satisfied in addition to the equality in (31):

$$\mathbf{\Lambda}_c(n)\hat{\boldsymbol{\theta}}_{r2c}(n-1) + \mathbf{\Lambda}_c(n)\hat{\mathbf{H}}_{dc}(n)\epsilon(n) = -\hat{\boldsymbol{\theta}}_r(n-1) + \gamma\hat{\mathbf{H}}_c(n)\epsilon(n). \quad (36)$$

An outline of the arguments leading to this conclusion is provided in Appendix I. Even though (36) contains variables that are not present in the original algorithm, the components of (36) that correspond to the original adaptive filter have not changed in any way. Therefore, we can prove the bound on our original algorithm, where $\mathbf{\Lambda}(n)$ is a diagonal matrix with positive step sizes on the diagonal, by proving the bound on the algorithm obtained by substituting (36) in (35), which is given by

$$\hat{\boldsymbol{\theta}}_c(n) = \hat{\boldsymbol{\theta}}_c(n-1) + \hat{\boldsymbol{\theta}}_r(n-1) - \hat{\boldsymbol{\theta}}_r(n-1) + \gamma\hat{\mathbf{H}}_c(n)\epsilon(n) \quad (37)$$

$$= \hat{\boldsymbol{\theta}}_c(n-1) + \gamma\hat{\mathbf{H}}_c(n)\epsilon(n). \quad (38)$$

The calculations in Appendix I show that for an algorithm given by (18), an algorithm of the form (38) can always be obtained such that the equivalent coefficients of both algorithms have

$$\mathbf{\Lambda}_e(n) = \frac{1}{\delta}\begin{bmatrix} \mu_1(n) & 0 & \cdots & 0 & \cdots & 0 & \rho_{1,1}(n) & \cdots & \rho_{1,2L}(n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_N(n) & 0 & \cdots & \vdots & \rho_{N,1}(n) & \cdots & \rho_{N,2L}(n) \\ 0 & \cdots & 0 & \mu_{N+1}(n) & 0 & \vdots & \rho_{N+1,1}(n) & \cdots & \rho_{N+1,2L}(n) \\ \vdots & & \vdots & 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & & \cdots & & 0 & \mu_{N+M+L}(n) & \rho_{N+M+L,1}(n) & \cdots & \rho_{N+M+L,2L}(n) \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \ddots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}. \quad (30)$$

the same evolution. With the equivalent evolution of the corresponding coefficients, the output $\hat{d}(n)$ is the same whether obtained from (18) or (38). According to the definitions of $\hat{\boldsymbol{\theta}}_c(n)$, $\hat{\mathbf{H}}_c(n)$ and $\hat{\boldsymbol{\theta}}(n)$, $\hat{\mathbf{H}}(n)$, we can then see that the *a posteriori* error $\epsilon(n) = d(n) - \hat{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n) = d(n) - \hat{\boldsymbol{\theta}}^T(n)\hat{\mathbf{H}}(n)$ is the same for both algorithms.

### D. Assumptions

To complete our analysis, we need the following definitions and assumptions. The definitions are included here for the sake of completeness.

*Definition 1 (p. 30 [19]):* Let $\{Y_n, n \geq 1\}$ be a stochastic sequence and $\{\mathcal{F}_n, n \geq 1\}$ an increasing[1] sequence of $\sigma$ fields with $\mathcal{F}_n \subset \mathcal{F}$ for each $n \geq 1$. If $Y_n$ is $\mathcal{F}_n$ measurable for each $n \geq 1$, the $\sigma$ fields $\{\mathcal{F}_n, n \geq 1\}$ are said to be adapted to the sequence $\{Y_n, n \geq 1\}$ and $\{Y_n, \mathcal{F}_n, n \geq 1\}$ is said to be an adapted stochastic sequence.

*Definition 2 ([19]):* If $\{Y_n, \mathcal{F}_n, n \geq 1\}$ is an adapted stochastic sequence with

$$E[Y_n \,|\, \mathcal{F}_{n-1}] = 0 \quad \text{(a.s.) for each} \quad n \geq 2 \tag{39}$$

$\{Y_n, \mathcal{F}_n, n \geq 1\}$ is called a martingale difference sequence.

*Definition 3 (p. 10 [19]):* Let $\{T_n, n \geq 1\}$ and $T$ be random variables. Then $T_n$ is said to converge in probability to $T$ [19] if

$$P[|T_n - T| > \epsilon] \to 0 \quad \text{as} \quad n \to \infty \tag{40}$$

for all $\epsilon > 0$. The sequence $T_n$ is said to converge almost surely (a.s.) to $T$ if

$$P\left[T_n \to T \quad \text{as} \quad n \to \infty\right] = 1. \tag{41}$$

Almost sure convergence is denoted by $T_n \to T$ (a.s.).

Let $c_{(\cdot)}$ denote generic, finite, and positive numbers. We now enumerate the basic assumptions (A1–A5) employed to make this analysis feasible.

A1) Both the persistently exciting input signal $x(n)$ and the noise $\nu(n)$ are bounded sequences such that

$$|x(n)| < c_x, \quad \forall n \geq 0 \tag{42}$$

$$|\nu(n)| < c_\nu, \quad \forall n \geq 0. \tag{43}$$

A2) i) The coefficients of the unknown system are bounded from above such that

$$\|\boldsymbol{\theta}_c(n)\| < c_{\boldsymbol{\theta}} \tag{44}$$

ii) $N, M$, the orders of the polynomial $A(n, q^{-1})$, and $B(n, q^{-1})$, respectively, are constant, finite, and known.

iii) The unknown system (19) is exponentially BIBO stable.

A3) Let

$$\Delta(n) = \boldsymbol{\theta}_c(n) - \boldsymbol{\theta}_c(n-1) \tag{45}$$

denote the increment process associated with the unknown system. There exists a $\lambda$, where $0 < \lambda < 1$, and a constant $\alpha$ such that for all $k$,

$$\sum_{n=1}^{k} \lambda^{k-n} \|\Delta(n)\|^2 \leq \alpha. \tag{46}$$

A4) Operator $A(k, q^{-1})$ is input strictly passive [16], [20], i.e., there exists a positive constant $\kappa_0$ such that

$$\sum_{k=1}^{n} u(k)[A(n, q^{-1})u(k)] \geq \kappa_0 \sum_{k=1}^{n} u^2(k) \tag{47}$$

for any real sequence $\{u(k)\}, k \geq 0$. This assumption is a time-varying version of the well known strictly positive real condition in the case of the time varying operators. The parameter $\kappa_0$ is independent of the signal $u(n)$. It only depends on the properties of $A(n, q^{-1})$.

A5) The noise $\{\nu(n)\}$ is a martingale difference sequence, i.e., $E[\nu(n+1) \,|\, F_n] = 0$ almost surely (a.s.), and satisfies

$$\sup_n E\left[|\nu(n+1)|^{\delta_a} | F_n\right] < \infty \quad \text{(a.s.) for some } \delta_a > 2 \tag{48}$$

and

$$\lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} \nu^2(n) = \sigma_\nu^2 \quad \text{(a.s.)}. \tag{49}$$

In addition, $\{\nu(n)\}$ is independent of $\{\boldsymbol{\theta}(n)\}$ and $\{x(n)\}$. In the discussion here, $F_n$ is the $\sigma$-algebra generated by $\{\nu(0), \nu(1), \ldots, \nu(n)\}$.

## III. Main Result

We start by rewriting (38) as

$$\hat{\boldsymbol{\theta}}_c(n-1) = \hat{\boldsymbol{\theta}}_c(n) - \gamma \hat{\mathbf{H}}_c(n)\epsilon(n). \tag{50}$$

Subtracting $\boldsymbol{\theta}_c(n-1)$ from both sides, we get

$$\begin{aligned}\hat{\boldsymbol{\theta}}_c(n-1) - \boldsymbol{\theta}_c(n-1) &= \hat{\boldsymbol{\theta}}_c(n) - \boldsymbol{\theta}_c(n) + \boldsymbol{\theta}_c(n) - \boldsymbol{\theta}_c(n-1) - \gamma \hat{\mathbf{H}}_c(n)\epsilon(n) \end{aligned} \tag{51}$$

giving

$$\tilde{\boldsymbol{\theta}}_c(n-1) = \tilde{\boldsymbol{\theta}}_c(n) + \Delta(n) - \gamma \hat{\mathbf{H}}_c(n)\epsilon(n) \tag{52}$$

where $\Delta(n)$ was defined in (45) and

$$\tilde{\boldsymbol{\theta}}_c(n) = \hat{\boldsymbol{\theta}}_c(n) - \boldsymbol{\theta}_c(n). \tag{53}$$

Premultiplying both sides of (52) with their respective transposes gives

$$\begin{aligned}\|\tilde{\boldsymbol{\theta}}_c(n)\|^2 &= \|\tilde{\boldsymbol{\theta}}_c(n-1)\|^2 - 2\tilde{\boldsymbol{\theta}}_c^T(n)\Delta(n) \\ &\quad + 2\gamma \tilde{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n)\epsilon(n) - \left\|\Delta(n) - \gamma \hat{\mathbf{H}}_c(n) \cdot \epsilon(n)\right\|^2. \end{aligned} \tag{54}$$

Since $\|\Delta(n) - \gamma \hat{\mathbf{H}}_c(n) \cdot \epsilon(n)\|^2$ is nonnegative, we can drop this term from the right-hand side (RHS) of (54) to get

$$\begin{aligned}\|\tilde{\boldsymbol{\theta}}_c(n)\|^2 &\leq \|\tilde{\boldsymbol{\theta}}_c(n-1)\|^2 - 2\tilde{\boldsymbol{\theta}}_c^T(n)\Delta(n) \\ &\quad + 2\gamma \tilde{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n)\epsilon(n). \end{aligned} \tag{55}$$

[1] A sequence $\{x_1, x_2, \ldots\}$ for which $x_i \subset x_j$ if $i < j$. Note that $\{\mathcal{F}_n, n \geq 1\}$ is a sequence of $\sigma$ fields (i.e., sigma algebra) on $\Omega$ and for each $n$, $\mathcal{F}_n$ is generated by $\{Y_i, 0 \leq i \leq n-1\}$. Then it is true that $\mathcal{F}_{n-1} \subset \mathcal{F}_n$.

Let $\epsilon(n) = s(n)+\nu(n)$. Substituting this for $\epsilon(n)$ and replacing $2\widetilde{\boldsymbol{\theta}}_c^T(n)\Delta(n)$ with $2\|\widetilde{\boldsymbol{\theta}}_c(n)\|\|\Delta(n)\|$ in (55) gives

$$\|\widetilde{\boldsymbol{\theta}}_c(n)\|^2 \le \|\widetilde{\boldsymbol{\theta}}_c(n-1)\|^2 + 2\|\widetilde{\boldsymbol{\theta}}_c(n)\|\|\Delta(n)\| + 2\gamma\widetilde{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n)s(n) + 2\gamma\widetilde{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n)\nu(n). \quad (56)$$

It is shown in Appendix II that [13]

$$\widetilde{\boldsymbol{\theta}}_c^T(n)\cdot\hat{\mathbf{H}}_c(n) = -A(n,q^{-1})s(n). \quad (57)$$

Substituting the above result in (56) gives

$$\|\widetilde{\boldsymbol{\theta}}_c(n)\|^2 \le \|\widetilde{\boldsymbol{\theta}}_c(n-1)\|^2 + 2\|\widetilde{\boldsymbol{\theta}}_c(n)\|\|\Delta(n)\| - 2\gamma s(n)\left[A(n,q^{-1})s(n)\right] + 2\gamma\widetilde{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n)\nu(n). \quad (58)$$

Let $\hat{\boldsymbol{\Upsilon}}(n)$ be an $(N+M\cdot L+3L)\times(N+M+L)$-element matrix defined as

$$\hat{\gamma}\begin{bmatrix} \mathbf{I}_{N\times N} & 0 & 0 & 0 & \cdots & 0 \\ 0 & \hat{\mathbf{p}}(n-1) & 0 & \vdots & \cdots & 0 \\ 0 & 0 & \hat{\mathbf{p}}(n-2) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & \hat{\mathbf{p}}(n-M) & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \mathbf{I}_{L\times L} \\ \mathbf{0}_{2L\times 1} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{2L\times 1} \end{bmatrix}. \quad (59)$$

Direct multiplications will show that

$$\hat{\boldsymbol{\theta}}_c(n) = \hat{\boldsymbol{\Upsilon}}(n)\hat{\boldsymbol{\theta}}(n), \quad (60)$$

$$\hat{\mathbf{H}}_c(n) = \hat{\boldsymbol{\Upsilon}}(n)\hat{\mathbf{H}}(n) \quad (61)$$

and that

$$\hat{\mathbf{H}}^T(n) = \hat{\mathbf{H}}_c^T(n)\hat{\boldsymbol{\Upsilon}}(n). \quad (62)$$

Premultiplying both sides of (14) with $\hat{\boldsymbol{\Upsilon}}(n)$ and simplifying using (60) and the definition of $\hat{\boldsymbol{\theta}}_r(n-1)$ from (34) results in

$$\hat{\boldsymbol{\theta}}_c(n) = \hat{\boldsymbol{\theta}}_c(n-1) + \hat{\boldsymbol{\theta}}_r(n-1) + \frac{\hat{\boldsymbol{\Upsilon}}(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}e(n). \quad (63)$$

Subtracting $\boldsymbol{\theta}_c(n)$ from both sides of (63) gives

$$\widetilde{\boldsymbol{\theta}}_c(n) = \widetilde{\boldsymbol{\theta}}_c(n-1) + \hat{\boldsymbol{\theta}}_r(n-1) - \Delta(n) + \frac{\hat{\boldsymbol{\Upsilon}}(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}e(n). \quad (64)$$

Next we use (64) in (58) to get

$$\|\widetilde{\boldsymbol{\theta}}_c(n)\|^2 \le \|\widetilde{\boldsymbol{\theta}}_c(n-1)\|^2 + 2\|\widetilde{\boldsymbol{\theta}}_c(n)\|\|\Delta(n)\| - 2\gamma s(n)\left[A(n,q^{-1})s(n)\right] + 2\gamma\left(\widetilde{\boldsymbol{\theta}}_c^T(n-1) + \hat{\boldsymbol{\theta}}_r^T(n-1)\right)\cdot\hat{\mathbf{H}}_c(n)\nu(n) - 2\gamma\Delta^T(n)\hat{\mathbf{H}}_c(n)\nu(n) + 2\gamma\frac{\hat{\mathbf{H}}_c^T(n)\hat{\boldsymbol{\Upsilon}}(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}e(n)\nu(n). \quad (65)$$

Using (62), we can write

$$\hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n) = \hat{\mathbf{H}}_c^T(n)\hat{\boldsymbol{\Upsilon}}(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n). \quad (66)$$

Using (66) in (65) we get

$$\|\widetilde{\boldsymbol{\theta}}_c(n)\|^2 \le \|\widetilde{\boldsymbol{\theta}}_c(n-1)\|^2 + 2\|\widetilde{\boldsymbol{\theta}}_c(n)\|\|\Delta(n)\| - 2\gamma s(n)\left[A(n,q^{-1})s(n)\right] + 2\gamma\left(\widetilde{\boldsymbol{\theta}}_c^T(n-1) + \hat{\boldsymbol{\theta}}_r^T(n-1)\right)\cdot\hat{\mathbf{H}}_c(n)\nu(n) - 2\gamma\Delta^T(n)\hat{\mathbf{H}}_c(n)\nu(n) + 2\gamma\frac{\hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{\phi}(n)}e(n)\nu(n). \quad (67)$$

Let us now present the Martingale limit theorem [19].

*Theorem 1 (Martingale Limit Theorem [19]):* Let assumption A5 hold, and let $f(n-1)$ be an $F_{n-1}$ measurable sequence. Then

$$\left|\sum_{n=1}^{m} f(n-1)\nu(n)\right| = \circ\left(\sum_{n=1}^{m} f^2(n-1)\right) + \bigcirc(1)\ \text{(a.s.)} \quad (68)$$

where the symbols $\circ(\cdot)$ and $\bigcirc(\cdot)$ denote the "order of the magnitude."

For $f(n-1)$ to be $F_{n-1}$ measurable, we require that $f(n-1)$ can be only a function of $\nu(k)$, where $k<n$. In more lax words, $F_{n-1}$ measurability implies that $f(n-1)$ is a nonanticipative function of a signal $\nu(n)$.

Let $g$ be a positive function. The concepts of $f=\circ(g)$ and $f=\bigcirc(g)$ are defined as follows [21].

*Definition 4:* Function $f$ is said to be of lower order than $g$ in the neighborhood of $x=x_0$ if

$$\lim_{x\to x_0}\frac{f(x)}{g(x)} = 0. \quad (69)$$

We use the notation $f=\circ(g)$ to represent this relationship. The statement $f=\circ(1)$ is equivalent to $f\to 0$.

*Definition 5:* We say that $f$ is of the order of $g$ on a set $S$ if there is a positive number $c_B$ such that

$$\left|\frac{f(x)}{g(x)}\right| < c_B \text{ if } x\in S. \quad (70)$$

This is denoted by $f=\bigcirc(g)$. Stating "$f=\bigcirc(1)$ on $S$" is the same as saying that "$f$ is bounded on $S$." Clearly, $f=\circ(g)$ implies, and is stronger than $f=\bigcirc(g)$ [21].

Since the step size sequence satisfies the Lyapunov conditions for stability of the system [18], $\widetilde{\boldsymbol{\theta}}(n)$ and $\hat{d}(n)$ are bounded sequences. Bounded $\hat{d}(n)$ implies that $\hat{\mathbf{H}}(n)$ is also bounded. Let $c_{\widetilde{\theta}}$ and $c_{\hat{H}}$ denote the upper bounds of $\|\widetilde{\boldsymbol{\theta}}_c(n)\| = \sqrt{\widetilde{\boldsymbol{\theta}}_c^T(n)\widetilde{\boldsymbol{\theta}}_c(n)}$, and $\|\hat{\mathbf{H}}_c(n)\|$, respectively, i.e.,

$$\|\widetilde{\boldsymbol{\theta}}_c(n)\| \le c_{\widetilde{\theta}} \quad (71)$$

$$\|\hat{\mathbf{H}}_c(n)\| \le c_{\hat{H}}. \quad (72)$$

*Theorem 2:* Let assumptions A1-A5 hold. Then

$$\lim_{m\to\infty}\sup\frac{1}{m}\sum_{n=0}^{m}(d(n)-\hat{d}(n)-\nu(n))^2 \le \alpha\frac{c_{\tilde{\theta}}}{\gamma\kappa_0}+c_\Lambda\frac{1}{\kappa_0}\sigma_\nu^2 \quad \text{(a.s.)} \quad (73)$$

where $\kappa_0$ is a parameter from Assumption A4, while $\alpha$ was introduced in Assumption A3 and $c_\Lambda$ is a bound such that $|(\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n))/(\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n))| \le c_\Lambda$. Algorithm requires that $-\delta\beta < \hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)$, where $\delta$ is a small positive constant and $0<\beta<1$.

*Proof:* Summing both sides of (67) from $n=1$ to $m$, it follows that

$$\begin{aligned}&\|\tilde{\boldsymbol{\theta}}_c(m)\|^2+2\gamma\sum_{n=1}^{m}s(n)[A(n,q^{-1})s(n)]\\&\le\|\tilde{\boldsymbol{\theta}}_c(0)\|^2+2\sum_{n=1}^{m}\|\tilde{\boldsymbol{\theta}}_c(n)\|\|\Delta(n)\|+2\gamma\sum_{n=1}^{m}\Big(\tilde{\boldsymbol{\theta}}_c^T(n-1)\\&+\hat{\boldsymbol{\theta}}_r^T(n-1)\Big)\hat{\mathbf{H}}_c(n)\nu(n)-2\gamma\sum_{n=1}^{m}\Delta^T(n)\hat{\mathbf{H}}_c(n)\nu(n)\\&+2\gamma\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}e(n)\nu(n). \quad (74)\end{aligned}$$

Note that $(\tilde{\boldsymbol{\theta}}_c^T(n-1)+\hat{\boldsymbol{\theta}}_r^T(n-1))\hat{\mathbf{H}}_c(n)$ is independent of $\nu(n)$ and $F_{n-1}$ measurable. Then from Theorem 1 we have

$$\begin{aligned}&\left|\sum_{n=1}^{m}\Big(\tilde{\boldsymbol{\theta}}_c^T(n-1)+\hat{\boldsymbol{\theta}}_r^T(n-1)\Big)\hat{\mathbf{H}}_c(n)\nu(n)\right|\\&=\circ\left(\sum_{n=1}^{m}\Big(\Big(\tilde{\boldsymbol{\theta}}_c^T(n-1)+\hat{\boldsymbol{\theta}}_r^T(n-1)\Big)\hat{\mathbf{H}}_c(n)\Big)^2\right)+\bigcirc(1)\\&=\circ\left(\sum_{n=1}^{m}\Big\|\Big(\tilde{\boldsymbol{\theta}}_c^T(n-1)+\hat{\boldsymbol{\theta}}_r^T(n-1)\Big)\Big\|^2\Big\|\hat{\mathbf{H}}_c(n)\Big\|^2\right)+\bigcirc(1)\\&=\circ(m)+\bigcirc(1)\\&=\circ(m)\ \text{(a.s.)} \quad (75)\end{aligned}$$

where we have used the Cauchy-Schwartz inequality

$$\Big(\Big(\tilde{\boldsymbol{\theta}}_c^T(n-1)+\hat{\boldsymbol{\theta}}_r^T(n-1)\Big)\hat{\mathbf{H}}_c(n)\Big)^2 \le\Big\|\Big(\tilde{\boldsymbol{\theta}}_c^T(n-1)+\hat{\boldsymbol{\theta}}_r^T(n-1)\Big)\Big\|^2\cdot\Big\|\hat{\mathbf{H}}_c(n)\Big\|^2 \quad (76)$$

and the fact that $\|(\tilde{\boldsymbol{\theta}}_c^T(n-1)+\hat{\boldsymbol{\theta}}_r^T(n-1))\|$ and $\|\hat{\mathbf{H}}_c(n)\|$ are bounded for all $n\ge 0$. This is true since $\tilde{\boldsymbol{\theta}}_c(n-1)$, $\hat{\boldsymbol{\theta}}_r(n-1)$, $\hat{\mathbf{H}}_c(n)$ are finite, and because $\hat{A}(n-1,q^{-1})$ is guaranteed to be stable by our algorithm. By Assumption A5, $\Delta(n)$ is independent of $\nu(n)$. It follows that $\Delta^T(n)\hat{\mathbf{H}}_c(n)$ is $F_{n-1}$ measurable, and by application of Theorem 1 to this sequence gives

$$\begin{aligned}&\left|\sum_{n=1}^{m}\Delta^T(n)\hat{\mathbf{H}}_c(n)\nu(n)\right|\\&=\circ\left(\sum_{n=1}^{m}\Big(\Delta^T(n)\hat{\mathbf{H}}_c(n)\Big)^2\right)+\bigcirc(1)\\&=\circ\left(\sum_{n=1}^{m}\Big\|\Delta^T(n)\Big\|^2\Big\|\hat{\mathbf{H}}_c(n)\Big\|^2\right)+\bigcirc(1)\ \text{(a.s.)}. \quad (77)\end{aligned}$$

Note that by Assumption A3, $\|\Delta(n)\|$ is finite for all $n\ge 0$. Since $\|\hat{\mathbf{H}}_c(n)\|$ is bounded, (77) yields

$$\left|\sum_{n=1}^{m}\Delta^T(n)\hat{\mathbf{H}}_c(n)\nu(n)\right|=\circ(m)\ \text{(a.s.)}. \quad (78)$$

Similar calculations on the last term of (74) gives the following result:

$$\begin{aligned}&\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}e(n)\nu(n)\\&=\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\Big(\tilde{d}(n)+\nu(n)\\&\quad-\hat{\mathbf{H}}_c^T(n)\hat{\boldsymbol{\theta}}_c(n-1)\Big)\nu(n)\\&=\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\tilde{d}(n)\nu(n)\\&\quad+\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\nu(n)\nu(n)\\&\quad-\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\Big(\hat{\mathbf{H}}_c^T(n)\hat{\boldsymbol{\theta}}_c(n-1)\Big)\nu(n)\\&\le c_\Lambda\sum_{n=1}^{m}\nu^2(n)+\circ(m)\ \text{(a.s.)}. \quad (79)\end{aligned}$$

Note that $\mathbf{\Lambda}(n)$, which is directly related to $\gamma$ as aforementioned, is chosen at time instant $n$, but independently of the value of $\nu(n)$. We obtained the result in (79) applying similar procedures as used to obtain (75) and (78). We only comment here on the term $\sum_{n=1}^{m}(\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n))/(\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n))\nu(n)\nu(n)$. It is obvious that

$$\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\nu(n)\nu(n) \le\sum_{n=1}^{m}\left|\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\right|\nu^2(n). \quad (80)$$

Since the algorithm requires that $\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n) > -\delta\beta$, and $\hat{\mathbf{H}}^T(n)$, $\mathbf{\Lambda}(n)$ and $\boldsymbol{\phi}(n)$ are all finite due to the algorithm, we can introduce a positive bound $c_\Lambda$ such that

$$\left|\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\right| \le c_\Lambda. \tag{81}$$

Using (81) we have

$$\sum_{n=1}^{m}\frac{\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}{\delta+\hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)}\nu^2(n) \le c_\Lambda\sum_{n=1}^{m}\nu^2(n). \tag{82}$$

By Assumption A3, $\|\mathbf{\Delta}(n)\| \le \alpha$. Since $\|\tilde{\boldsymbol{\theta}}_c(n)\| \le c_{\tilde{\theta}}$ [see (71)], we have

$$\sum_{n=1}^{m}\|\tilde{\boldsymbol{\theta}}_c^T(n)\|\|\mathbf{\Delta}(n)\| \le \alpha c_{\tilde{\theta}} m. \tag{83}$$

Using (75), (78), (79), and (83), (74) simplifies to

$$\|\tilde{\boldsymbol{\theta}}_c(m)\|^2 + 2\gamma\sum_{n=1}^{m}s(n)A(n,q^{-1})s(n) \le \|\tilde{\boldsymbol{\theta}}_c(0)\|^2 + 2\alpha c_{\tilde{\theta}} m + 2\gamma c_\Lambda\sum_{n=1}^{m}\nu^2(n) + o(m). \tag{84}$$

Using Assumption A4, (84) becomes

$$\|\tilde{\boldsymbol{\theta}}_c(m)\|^2 + 2\gamma\kappa_0\sum_{n=1}^{m}s^2(n) \le \|\tilde{\boldsymbol{\theta}}_c(0)\|^2 + 2\alpha c_{\tilde{\theta}} m + 2\gamma c_\Lambda\sum_{n=1}^{m}\nu^2(n) + o(m). \tag{85}$$

Dividing the entire (85) by $2\gamma\kappa_0 m$ and taking the limit as $m$ goes toward infinity, we have

$$\lim_{m\to\infty}\sup\frac{1}{m}\sum_{n=0}^{m}\Big(d(n)-\hat{d}(n)-\nu(n)\Big)^2 \le \alpha\frac{c_{\tilde{\theta}}}{\gamma\kappa_0} + c_\Lambda\frac{1}{\kappa_0}\sigma_\nu^2 \quad \text{(a.s.)}. \tag{86}$$

This proves Theorem 2.

$c_\Lambda$ depends on $\Lambda(n)$ and can be made arbitrarily small. Assuming that the underlying system is time-invariant (i.e., $\alpha = 0$), Theorem 2 implies that the long-term time average of the square of the *a posteriori* excess estimation error can be arbitrarily close to zero. That is, the system can approach the global minimum of the performance surface with arbitrarily small error. As one would expect the long-term average of the squared error contributed by the variations of the parameters of the underlying time-varying system depends on the strength of coefficient increment process ($\alpha$), and is inversely proportional to the step sizes in $\Lambda(n)$.

## IV. Simulation Results

In this section, we present the results of a simulation experiment conducted to demonstrate the global convergence capabilities of the adaptive filter analyzed in this paper. The adaptive filter was employed in the system identification mode in the simulations. An unknown Hammerstein system composed of the memoryless nonlinearity with an input-output relationship

$$z(n) = 0.1x(n) - 0.075x^2(n) + 0.05x^3(n) \tag{87}$$

and a linear component with the transfer function

$$H(z) = \frac{0.25}{1-0.4z^{-1}+0.2z^{-2}} \tag{88}$$

was identified using the adaptive filter. The parameters of both the linear and the nonlinear part were time-invariant in this simulation example. The linear system $H(z)$ is a strictly positive real (SPR) [20], [22] and satisfies the constraint (47) with $\kappa_0 = 0.7$. The desired response signal $d(n)$ of the adaptive filter was obtained by corrupting the output of the unknown system with additive white noise with zero mean value and variance 0.0166, such that output SNR was 20 dB. The input signal $x(n)$ of the adaptive filter was generated by filtering a Gaussian signal with zero mean value and unit variance with the filter

$$H_c(z) = 1 + 0.5z^{-1}. \tag{89}$$

The adaptive filter was implemented with the time-varying step size of the recursive component of the linear subsystem to be the maximum of $\mu = 10^{-3}$ or the bound suggested by the condition in $\|\text{vec}[\mathbf{Q}(n+1)] - \text{vec}[\mathbf{Q}(n)]\| < 1$, and the step sizes for the coefficients of the first, second and third order terms of the polynomial subsystem were constant and equal to $10^{-3}$, $5\cdot 10^{-4}$ and $10^{-4}$, respectively. The step-sizes corresponding to the linear part were time-varying so that the stability of the IIR filter was assured according to the condition $\|\text{vec}[\mathbf{Q}(n+1)] - \text{vec}[[\mathbf{Q}(n)]\| < 1$. The step-sizes corresponding to the nonlinear part can be time-varying as well as long as they are small enough to ensure stability. However, unlike the recursive part of the linear subsystem, stability of the adaptation process does not require a time varying step size. Because of this, we chose to use a fixed step size for all components other than the step sizes associated with the feedback coefficients of the linear subsystem. The parameter $\delta$ was set to $10^{-3}$ and $\beta$ to 0.99. The system was initialized with poles of the linear subsystem at the origin, and the initial values of polynomial coefficients at zero. The coefficient $\hat{b}_0(n)$ was set to 1 and was not changed throughout the simulation. This ensured the uniqueness of the solution. Two hundred independent experiments using 50 000 data samples each were conducted. The results presented are average values over these 200 experiments.

Fig. 2 shows the evolution of the excess mean-square error (MSE) in the simulations. Our algorithm operates in a stable manner as predicted by the theoretical derivations. We can see

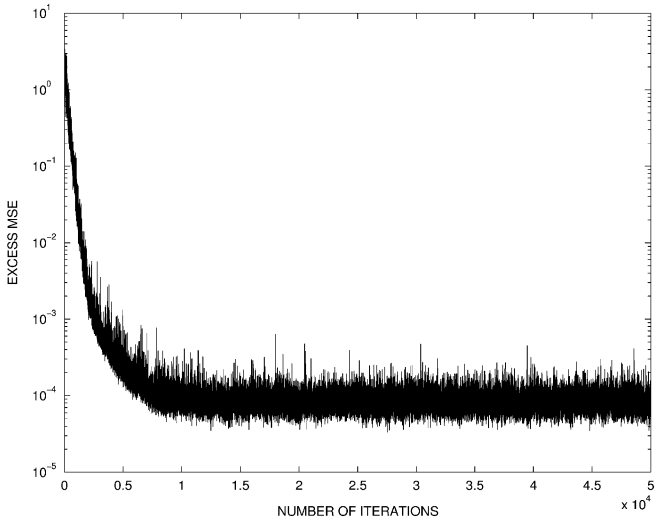


Fig. 2. Excess MSE of the adaptive filter employed in simulations.

from these results that the algorithm works reasonably well in this set of experiments.

It is difficult to obtain a tight bound for the LHS of (81) for small step sizes. Consequently we do not attempt to provide a comparison of the theoretical bound in (73) with simulation results. However the simulation results provide additional demonstration of the global convergence result implied by the theorem. To see this, we evaluated the long-term time average excess squared *a posteriori* error for each run over the last 15 000 samples. By only computing the averages over the last 15 000 samples, we eliminated the effect of the initial excess errors from the calculations. This would be also the case if we averaged over an arbitrarily large number of samples. The mean value of these 200 time averages was $8.75 \cdot 10^{-5}$. The largest value of the time averaged *a posteriori* excess squared error was $1.26 \cdot 10^{-4}$. These results indicate that the adaptive filter converged to locations that are close to the global minimum of the error surface in each of the runs, again validating the theoretical derivations in the paper.

We have evaluated the algorithm of this paper in a large number of synthetic scenarios, including those employing random initialization of the coefficients. The method provided good performance in terms of stable operation and global convergence in each case as predicted by the theoretical derivations.

## V. Concluding Remarks

A theoretical treatment of a recursive nonlinear adaptive filter developed in [18] was given in this paper. The convergence behavior of this algorithm was studied in a stochastic framework in a nonstationary environment, and in the presence of a possibly colored and nonstationary measurement noise that is a martingale difference sequence. Using the martingale limit theorem, we showed that the global minimum on the error surface of our adaptive Hammerstein filter can be achieved with arbitrary precision when the rate of change of the parameters of the underlying plant is zero. The adaptive system analyzed in this paper does not account for the Gram-Schmidt orthogonalization of the input signal as done in [18]. Extension of the analysis to this case is straightforward [23].

## Appendix I
## Proof of (31) and (36)

*Proof:* We will choose the variables $\rho_{i,j}(n)$ in such a way that $\mathbf{\Lambda}_c(n)\hat{\boldsymbol{\theta}}_{r2c}(n-1) = -\hat{\boldsymbol{\theta}}_r(n-1)$ and $\mathbf{\Lambda}_c(n)\hat{\mathbf{H}}_{dc}(n) = \gamma\hat{\mathbf{H}}_c(n)$. The proof consists of two parts.

1) In the first part we show that there are at least as many variables $\rho_{i,j}(n)$ as there are equations that we need to solve to establish (36).
2) In the second part, we show how (31) and (36) are satisfied by appropriate selection of the vector $\hat{\mathbf{H}}_d(n)$, and appropriate choice of the variables $\rho_{i,j}(n)$.

The calculations described below are not done in the implementation of the algorithm, but rather serve as a proof that for any algorithm (18) there exists an algorithm (38). Since equivalent coefficients of (18) and (38) evolve in an identical manner, the *a posteriori* errors of the two algorithms also evolve in the same way, implying that analyzing (38) is equivalent to analyzing (18).

The dimension of the step size matrix $\mathbf{\Lambda}(n)$ is $(N+M+L)\times(N+M+L)$, and therefore has $(N+M+L)$ nonzero step sizes $\mu_i(n)$. On the other hand, $\mathbf{\Lambda}_c(n)$ is an $(N+ML+3L)\times(N+ML+3L)$-element matrix which has the same $(N+M+L)$ step size entries $\mu_i(n)$ along the diagonal and an additional $2\cdot L\cdot(N+M+L)$ variables $\rho_{i,j}(n)$ that are off-diagonal. Having $2\cdot L\cdot(N+M+L)$ variables $\rho_{i,j}(n)$, and the total number of $2\cdot(N/2+M\cdot L+3\cdot L)$ equations, the latter can be satisfied as long as $L\geq 3$, since for this case $2\cdot L\cdot(N+M+L)\geq 2\cdot(N/2+M\cdot L+3\cdot L)$. We can make this proof valid for $L\geq 1$ by adding another $L_L$ column(s) to matrix $\mathbf{\Lambda}_c(n)$ so that $(N+M+L)\cdot(2L+L_L)$, the new number of unknown variables $\rho_{i,j}(n)$ is larger than the number of equations that must be satisfied.

In the second part, we show how (31) may be satisfied. Even though we do not know $\epsilon(n)$, choosing the vectors $\hat{\boldsymbol{\theta}}_{r2}(n-1)$ and $\hat{\mathbf{H}}_d(n)$ in (31) is always possible. One approach is to choose the first $N$ entries of $\hat{\mathbf{H}}_d(n)$ to be identical to the first $N$ entries of $\boldsymbol{\phi}_e(n)$. This makes the first $N$ elements of vector $\hat{\boldsymbol{\theta}}_{r2}(n-1)$ zero. For the rest of the elements of the vector $\hat{\mathbf{H}}_d(n)$, we only need to maintain $\hat{\mathbf{H}}_d(n)\neq\boldsymbol{\phi}_e(n)$ implying that the rest of the elements of $\hat{\boldsymbol{\theta}}_{r2}(n-1)$ are not zero. In particular, it is important that the last $2L$ elements of $\hat{\mathbf{H}}_d(n)$ and $\hat{\boldsymbol{\theta}}_{r2}(n-1)$ are not zero.

Finally, since we have more free variables $\rho_{i,j}(n)$ than there are equations, we can simultaneously solve for $\rho_{i,j}(n)$ in the equations $\mathbf{\Lambda}_c(n)\hat{\boldsymbol{\theta}}_{r2c}(n-1) = -\hat{\boldsymbol{\theta}}_r(n-1)$ and $\mathbf{\Lambda}_c(n)\hat{\mathbf{H}}_{dc}(n) = \gamma\hat{\mathbf{H}}_c(n)$ to satisfy (36). This completes the proof. ■

## Appendix II
## Proof of (57)

*Proof:* Note that $\epsilon(n) = s(n)+\nu(n) = d(n) - \hat{\mathbf{H}}^T(n)\hat{\boldsymbol{\theta}}(n)$ and

$$\begin{aligned} A(n,q^{-1})s(n) &= A(n,q^{-1})\left[\epsilon(n)-\nu(n)\right] \\ &= A(n,q^{-1})\left[d(n)-\hat{d}(n)-\nu(n)\right] \\ &= A(n,q^{-1})d(n) - A(n,q^{-1})\hat{d}(n) \\ &\quad - A(n,q^{-1})\nu(n). \end{aligned} \tag{90}$$

By using (20), (22), and (23), we have

$$\begin{aligned} A(n,q^{-1})d(n) &= A(n,q^{-1})\tilde{d}(n) + A(n,q^{-1})\nu(n) \\ &= B(n,q^{-1})\mathbf{p}^T(n)\mathbf{x}(n) \\ &\quad + A(n,q^{-1})\nu(n). \end{aligned} \tag{91}$$

Combining (24), (90), and (91) gives

$$\begin{aligned} & A(n,q^{-1})s(n) \\ &= B(n,q^{-1})\mathbf{p}^T(n)\mathbf{x}(n) \\ &\quad + A(n,q^{-1})\nu(n) - A(n,q^{-1})\hat{d}(n) - A(n,q^{-1})\nu(n) \\ &= B(n,q^{-1})\mathbf{p}^T(n)\mathbf{x}(n) - \hat{d}(n) \\ &\quad - a_1(n)\hat{d}(n-1) - \cdots - a_N(n)\hat{d}(n-N) \\ &= \boldsymbol{\theta}_c^T(n)\hat{\mathbf{H}}_c(n) - \hat{d}(n) \\ &= \boldsymbol{\theta}_c^T(n)\hat{\mathbf{H}}_c(n) - \hat{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n) \\ &= -\tilde{\boldsymbol{\theta}}_c^T(n)\hat{\mathbf{H}}_c(n). \end{aligned} \tag{92}$$

■

## ACKNOWLEDGMENT

The authors wish to thank to T. Bose and M. Radenkovic for their review of the paper and valuable comments.

## REFERENCES

[1] I. W. Hunter and M. J. Korenberg, "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models," *Biol. Cybern.*, vol. 55, pp. 135–144, 1986.

[2] N. Bershad, S. Bouchired, and F. Castanie, "Stochastic analysis of adaptive gradient identification of Wiener–Hammerstein systems for Gaussian inputs," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 557–560, Feb. 2000.

[3] B. Nollett and N. Bershad, "Errata to Stochastic analysis of adaptive gradient identification of Wiener–Hammerstein systems for Gaussian inputs," *IEEE Trans. Signal Process.*, vol. 49, no. 9, p. 2162, Sep. 2001.

[4] N. Bershad, P. Celka, and J.-M. Vesin, "Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for Gaussian data and noisy input and output measurements," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 675–689, Mar. 1999.

[5] T. Wigren, "Output error convergence of adaptive filters with compensation for output nonlinearities," *IEEE Trans. Autom. Control*, vol. 43, no. 7, pp. 975–978, Jul. 1998.

[6] A. Nordsjö and L. Zetterberg, "Identification of certain time-varying nonlinear Wiener and Hammerstein systems," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 577–592, Mar. 2001.

[7] A. E. Nordsjö, "Recursive prediction error algorithms for joint time delay and parameter estimation of certain classes of nonlinear systems," in *Proc. 37th IEEE Conf. Decision and Control*, vol. 3, Tampa, FL, 1998, pp. 3429–3438.

[8] W. Greblicki, "Stochastic approximation in nonparametric identification of Hammerstein systems," *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1800–1810, Nov. 2002.

[9] H.-F. Chen, "Pathwise convergence of recursive identification algorithms for Hammerstein systems," *IEEE Trans. Autom. Control*, vol. 49, no. 10, pp. 1641–1649, Oct. 2004.

[10] C. R. Johnson Jr., "A convergence proof for a hyperstable adaptive recursive filter," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 745–749, Dec. 1979.

[11] M. G. Larimore, J. R. Treichler, and C. R. Johnson Jr., "SHARF: An algorithm for adapting IIR digital filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 4, pp. 428–440, Aug. 1980.

[12] C. R. Johnson Jr., M. G. Larimore, J. R. Treichler, and B. D. O. Anderson, "SHARF convergence properties," *IEEE Trans. Circuits Syst.*, vol. CAS-28, no. 6, pp. 499–510, Jun. 1981.

[13] M. Radenkovic and T. Bose, "Adaptive IIR filtering of nonstationary signals," *Signal Process.*, vol. 81, no. 1, pp. 183–195, Jan. 2001.

[14] M. Radenkovic, T. Bose, and T. Mathurasai, "Optimality and almost sure convergence of adaptive IIR filters with output error recursion," *Digital Signal Process.*, vol. 9, no. 4, pp. 315–328, Oct. 1999.

[15] H. Fan, "Application of Benveniste's convergence results in the study of adaptive IIR filtering algorithms," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 692–709, Jul. 1988.

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

[17] K. X. Miao, H. Fan, and M. Doroslovački, "Cascade lattice IIR adaptive filters," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 721–742, Apr. 1994.

[18] J. Jeraj and V. J. Mathews, "A stable adaptive Hammerstein filter employing partial orthogonalization of the input signals," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1412–1420, Apr. 2005.

[19] W. F. Stout, *Almost Sure Convergence*. New York: Academic, 1974.

[20] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[21] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*. Oxford, Oxfordshire, U.K.: Clarendon, 1956.

[22] A. Betser and E. Zeheb, "Modified output error identification—Elimination of the SPR condition," *IEEE Trans. Autom. Control*, vol. 40, no. 1, pp. 190–193, Jan. 1995.

[23] J. Jeraj, "Adaptive estimation and equalization of nonlinear systems," Ph.D. dissertation, Univ. Utah, Dept. Elect. Comp. Eng., 2005.

**Janez Jeraj** (S'01–M'03) received the B.S. and M.S. degrees in electrical engineering from the University of Ljubljana, Slovenia. He received the Ph.D. degree in electrical engineering from the University of Utah, Salt Lake City.

**V. John Mathews** (S'82–M'84–SM'90–F'02) received the B.E. (Hons.) degree in electronics and communication engineering from the University of Madras, India, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Iowa, Iowa City, in 1981 and 1984, respectively.

At the University of Iowa, he was a Teaching/Research Fellow from 1980 to 1984 and a Visiting Assistant Professor with the Department of Electrical and Computer Engineering during the 1984–1985 academic year. He joined the University of Utah in 1985, where he is a Professor with the Department of Electrical and Computer Engineering. He served as Chairman of the department from 1999 to 2003. His research interests are in adaptive filtering, nonlinear filtering, image compression and application of signal processing techniques in communication systems and biomedical engineering. He is the author of *Polynomial Signal Processing* (New York: Wiley). He has published more than 100 technical papers.

Dr. Mathews served as the Vice President-Finance of the IEEE Signal Processing Society during 2003–2005, and has served on the Publication Board and the Conference Board of the Society. He is a Past Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE SIGNAL PROCESSING LETTERS. He serves on the editorial board of the IEEE *Signal Processing Magazine*. He was a member of the Signal Processing Theory and Methods and Education Technical Committees and served as the General Chairman of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2001.