A Stable Adaptive Hammerstein Filter Employing Partial Orthogonalization of the Input Signals

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

Abstract—This paper presents an algorithm that adapts the parameters of a Hammerstein system model. Hammerstein systems are nonlinear systems that contain a static nonlinearity cascaded with a linear system. In this paper, the static nonlinearity is modeled using a polynomial system, and the linear filter that follows the nonlinearity is an infinite-impulse response (IIR) system. The adaptation of the nonlinear components is improved by orthogonalizing the inputs to the coefficients of the polynomial system. The step sizes associated with the recursive components are constrained in such a way as to guarantee bounded-input bounded-output (BIBO) stability of the overall system. This paper also presents experimental results that show that the algorithm performs well in a variety of operating environments, exhibiting stability and global convergence of the algorithm.

Index Terms—Adaptive Hammerstein filter, nonlinear systems, polynomial signal processing, stability analysis.

I. INTRODUCTION

THIS paper describes the derivation and experimental performance evaluation of an adaptive algorithm employing a Hammerstein system model. Hammerstein systems are cascade nonlinear systems comprising of a memoryless nonlinearity followed by a linear system as shown in Fig. 1. In our work, the nonlinearity is a memoryless polynomial model and the linear system has infinite-impulse response (IIR). There are many applications in which cascade nonlinear models are appropriate. Examples include modeling satellite communication systems [1], biological systems [2], distillation columns [3], electrical drives [4], and amplifiers [5].

A number of nonparametric procedures have been considered for Hammerstein system identification [6], [7]. Among parametric methods we mention an approach where the polynomial and the dynamic linear subsystems are iteratively estimated [2]. A stochastic gradient method was employed in [8] to derive an adaptive algorithm. An online identification method for the Hammerstein model, based on the Kalman filter design was described in [9]. It was assumed in [9] that the parameters of the nonlinear model were constant. Some work in the stability and convergence analysis has been done for Hammerstein filters with finite-impulse response (FIR) [10], [11]; however, there are
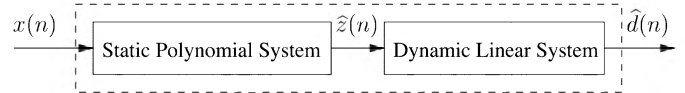
Fig. 1. Block diagram of a Hammerstein system.

limited or no stability and convergence analyzes for the algorithms with IIR.

Let the input–output relationship of the adaptive filter be 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

$$\hat{z}(n) = \hat{p}_1(n) \cdot x(n) + \hat{p}_2(n) \cdot x^2(n) + \ldots + \hat{p}_L(n) \cdot x^L(n) \quad (2)$$

is the output of a memoryless polynomial nonlinear system and $x(n)$ is the input to the adaptive filter. In the above equations, $\hat{p}_l(n), \hat{a}_i(n)$, and $\hat{b}_j(n)$ represent the coefficients of the adaptive filter at time n. Setting $\hat{b}_0(n) = 1$ for all n so as to ensure a unique solution during identification, (1) can be rewritten as

$$
\begin{aligned}
\hat{d}(n) &= \hat{z}(n) - \sum_{i=1}^{N} \hat{a}_i(n) \cdot \hat{d}(n-i) + \sum_{j=1}^{M} \hat{b}_j(n) \cdot \hat{z}(n-j) \\
&= \sum_{l=1}^{L} \hat{p}_l(n) x^l(n) - \sum_{i=1}^{N} \hat{a}_i(n) \cdot \hat{d}(n-i) \\
&\quad + \sum_{j=1}^{M} \hat{b}_j(n) \cdot \hat{z}(n-j).
\end{aligned}
\quad (3)
$$

The above representation is nonlinear in the data, but linear in the parameters, and is therefore quite useful for identifying Hammerstein systems. Such representations have been employed to derive iterative techniques for system identification of Hammerstein systems [12]. We also note that uniqueness can be ensured by fixing any one coefficient in the system.

The objective of the adaptive filter presented in this paper is to update the coefficients of the nonlinear model using a stochastic gradient procedure so as to reduce the instantaneous squared estimation error during each iteration. We orthogonalize the input signal to the polynomial subsystem. This improves the overall convergence behavior of the method, especially when the static nonlinearity is of relatively high order. For high orders of nonlinearity, the autocorrelation matrix of the vector of nonlinear signals $\mathbf{x}(n) = [x(n) \ x^2(n) \ x^3(n) \ \cdots \ x^L(n)]^T$ exhibits a large eigenvalue spread even when the input signal $x(n)$ is white. The problem for adaptation caused by this situation is avoided in our method. Furthermore, the adaptive IIR subsystem employs a step-size sequence that guarantees stability of the system. This
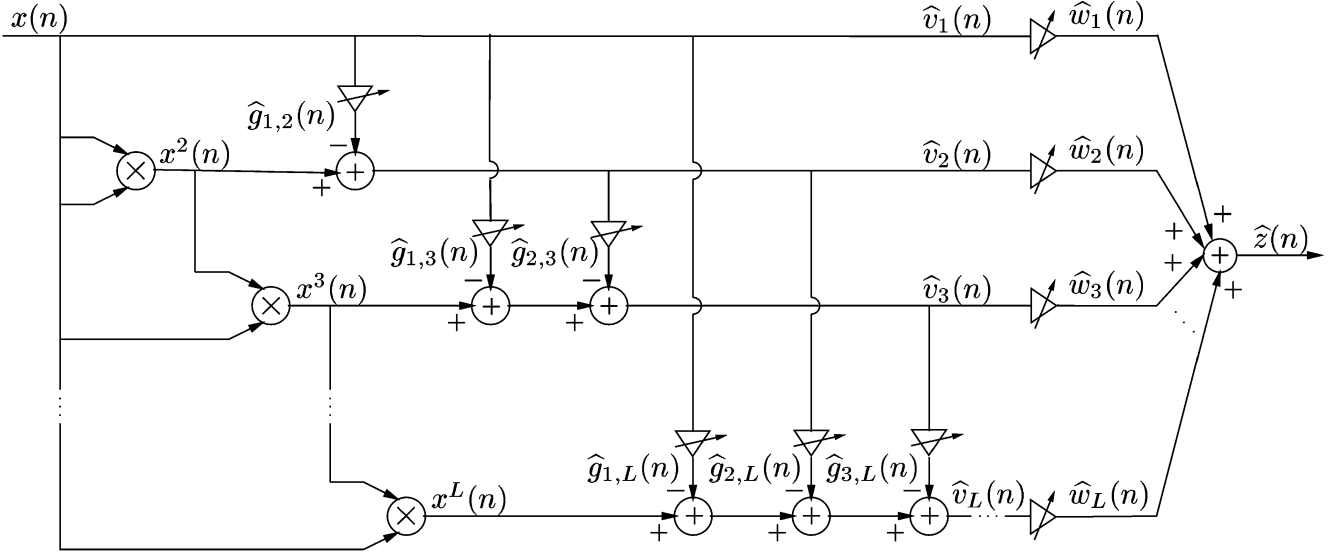
Fig. 2.   Orthogonalization of the nonlinear subsystem.

part of our work follows that of Carini *et al.* [13], in which the authors employ a Lyapunov stability criterion to ensure stability of adaptive IIR filters.

The rest of the paper is organized as follows. The next section details the derivation of the adaptive Hammerstein filter. Section II.D contains a theoretical proof of the stability of the derived algorithm. Results of experimental performance evaluation are described in Section III. A comparison of the performance of our method with the algorithm described in [9] is also provided in Section III. Finally, the concluding remarks are given in Section IV.

II. ADAPTATION OF THE HAMMERSTEIN SYSTEM MODEL

Let $x(n)$ and $d(n)$ represent the input signal and the desired response signal, respectively, of the adaptive Hammerstein filter. Fig. 2 depicts the approach employed to orthogonalize the input signals to the adaptive polynomial subsystem. The objective of the Gram–Schmidt orthogonalizer is to create the signals $\hat{v}_1(n), \hat{v}_2(n), \ldots, \hat{v}_L(n)$ such that

$$E\{\hat{v}_i(n)\hat{v}_j(n)\} = 0, \quad i \neq j \tag{4}$$

and the signals in the set $\{\hat{v}_i(n); i = 1, 2, \ldots, L\}$ span the space spanned by the signals in the set $\{x^i(n), i = 1, 2, \ldots, L\}$. The goal is to update the coefficients $\hat{g}_{i,j}(n)$ of the orthogonalizer such that (4) is approximately true in the steady state.

The output signals $\hat{v}_1(n), \hat{v}_2(n), \ldots, \hat{v}_L(n)$ of the Gram-Schmidt orthogonalizer are used to obtain the intermediate output $\hat{z}(n)$ of the model as

$$\hat{z}(n) = \sum_{i=1}^{L} \hat{w}_i(n)\hat{v}_i(n). \tag{5}$$

Finally, the adaptive linear subsystem estimates the desired response signal $d(n)$ as in (1).

Our goal is to update the coefficients $\hat{a}_i(n)$, $\hat{b}_j(n)$, and $\hat{w}_l(n)$ in (1) so that $\hat{d}(n)$ is close to the desired response signal $d(n)$ in

some sense. Therefore, we seek a stochastic gradient adaptation algorithm that attempts to reduce

$$J(\hat{a}_1(n), \ldots, \hat{a}_N(n), \hat{b}_0(n), \hat{b}_1(n), \ldots, \hat{b}_M(n),$$
$$\hat{w}_1(n), \ldots, \hat{w}_L(n)) = [d(n) - \hat{d}_{\mathrm{ap}}(n)]^2 \tag{6}$$

during each iteration, where the *a priori* estimate $\hat{d}_{\mathrm{ap}}(n)$ is obtained as

$$\hat{d}_{\mathrm{ap}}(n) = -\sum_{i=1}^{N} \hat{a}_i(n-1) \cdot \hat{d}(n-i) + \sum_{j=1}^{M} \hat{b}_j(n-1) \cdot \hat{z}(n-j). \tag{7}$$

We also need to update the coefficients $\hat{g}_{i,j}(n)$ of the Gram–Schmidt orthogonalizer. The coefficient updates for $\hat{a}_i(n), \hat{b}_j(n)$ and $\hat{w}_l(n)$ are given by

$$\hat{a}_i(n) = \hat{a}_i(n-1) - \mu_i(n)\frac{\partial(\frac{1}{2}[d(n) - \hat{d}_{\mathrm{ap}}(n)]^2)}{\partial\hat{a}_i(n-1)}$$

$$= \hat{a}_i(n-1) + \mu_i(n)\left[d(n) - \hat{d}_{\mathrm{ap}}(n)\right]\frac{\partial\hat{d}_{\mathrm{ap}}(n)}{\partial\hat{a}_i(n-1)};$$
$$\mu_i(n) > 0, \quad 1 \leq i \leq N \tag{8}$$

$$\hat{b}_j(n) = \hat{b}_j(n-1) - \mu_{N+j}(n)\frac{\partial(\frac{1}{2}[d(n) - \hat{d}_{\mathrm{ap}}(n)]^2)}{\partial\hat{b}_j(n-1)}$$

$$= \hat{b}_j(n-1) + \mu_{N+j}(n)[d(n) - \hat{d}_{\mathrm{ap}}(n)]$$
$$\cdot\frac{\partial\hat{d}_{\mathrm{ap}}(n)}{\partial\hat{b}_j(n-1)}; \quad \mu_{N+j}(n) > 0, \quad 0 \leq j \leq M \tag{9}$$

and

$$\hat{w}_l(n) = \hat{w}_l(n-1) - \mu_{N+M+l}(n)\frac{\partial(\frac{1}{2}[d(n) - \hat{d}_{\mathrm{ap}}(n)]^2)}{\partial\hat{w}_l(n-1)}$$

$$= \hat{w}_l(n-1) + \mu_{N+M+l}(n)[d(n) - \hat{d}_{\mathrm{ap}}(n)]$$
$$\cdot\frac{\partial\hat{d}_{\mathrm{ap}}(n)}{\partial\hat{w}_l(n-1)}; \quad \mu_{N+M+l}(n) > 0, \quad 1 \leq l \leq L \tag{10}$$

respectively. In the above equations, the variables $\mu_i(n)$ represent the step sizes employed by the adaptation algorithm and are small, positive sequences. To find explicit expressions for the updates, we evaluate the partial derivatives in (8)–(10). The

partial derivative of $\hat{d}_{\mathrm{ap}}(n)$ with respect to $\hat{a}_i(n-1)$ is given by

$$\frac{\partial \hat{d}_{\mathrm{ap}}(n)}{\partial \hat{a}_i(n-1)} = -\hat{d}(n-i) - \sum_{s=1}^{N} \hat{a}_s(n-1) \frac{\partial \hat{d}(n-s)}{\partial \hat{a}_i(n-1)}. \quad (11)$$

As is commonly done in the derivation of adaptive IIR filters [14], [15], we assume that the adaptation is slow enough, such that

$$\frac{\partial \hat{d}(n-s)}{\partial \hat{a}_i(n-1)} \approx \frac{\partial \hat{d}(n-s)}{\partial \hat{a}_i(n-s)}; \quad 1 \le i \le N \quad (12)$$

and

$$\frac{\partial \hat{d}_{\mathrm{ap}}(n)}{\partial \hat{a}_i(n-1)} \approx \frac{\partial \hat{d}(n)}{\partial \hat{a}_i(n)}; \quad 1 \le i \le N. \quad (13)$$

Substituting these approximations in (11) gives

$$\frac{\partial \hat{d}(n)}{\partial \hat{a}_i(n)} \approx -\hat{d}(n-i) - \sum_{s=1}^{N} \hat{a}_s(n-1) \frac{\partial \hat{d}(n-s)}{\partial \hat{a}_i(n-s)}; \\ 1 \le i \le N. \quad (14)$$

Using a similar procedure, we get

$$\frac{\partial \hat{d}(n)}{\partial \hat{b}_j(n)} \approx \hat{z}(n-j) - \sum_{s=1}^{N} \hat{a}_s(n-1) \frac{\partial \hat{d}(n-s)}{\partial \hat{b}_j(n-s)}; \\ 0 \le j \le M. \quad (15)$$

Finally, the update equations for the polynomial subsystem are given by

$$\frac{\partial \hat{d}_{\mathrm{ap}}(n)}{\partial \hat{w}_l(n-1)} = -\sum_{s=1}^{N} \hat{a}_s(n-1) \frac{\partial \hat{d}(n-s)}{\partial \hat{w}_l(n-s)} \\ + \sum_{s=1}^{M} \hat{b}_s(n-1) \hat{v}_l(n-s) \frac{\partial \hat{w}_l(n-s)}{\partial \hat{w}_l(n-1)}; \\ 1 \le l \le L. \quad (16)$$

As before, we have assumed that the adaptation is slow and used the approximations

$$\frac{\partial \hat{d}(n-s)}{\partial \hat{w}_l(n-1)} \approx \frac{\partial \hat{d}(n-s)}{\partial \hat{w}_l(n-s)} \quad (17)$$

$$\frac{\partial \hat{d}_{\mathrm{ap}}(n)}{\partial \hat{w}_l(n-1)} \approx \frac{\partial \hat{d}(n)}{\partial \hat{w}_l(n)} \quad (18)$$

and

$$\frac{\partial \hat{w}_l(n-s)}{\partial \hat{w}_l(n-1)} = 1 \quad (19)$$

to derive this result. Combining (14), (15), and (16), we can write a compact expression for the coefficient updates as

$$\phi(n) = \psi(n) - \sum_{s=1}^{N} \hat{a}_s(n-1) \cdot \phi(n-s) \quad (20)$$

where

$$\phi(n) = \left[\frac{\partial \hat{d}(n)}{\partial \hat{a}_1(n)} \quad \cdots \quad \frac{\partial \hat{d}(n)}{\partial \hat{a}_N(n)} \quad \frac{\partial \hat{d}(n)}{\partial \hat{b}_1(n)} \quad \cdots \\ \frac{\partial \hat{d}(n)}{\partial \hat{b}_M(n)} \quad \frac{\partial \hat{d}(n)}{\partial \hat{p}_1(n)} \quad \cdots \quad \frac{\partial \hat{d}(n)}{\partial \hat{p}_L(n)} \right]^T. \quad (21)$$

and the information vector $\psi(n)$ is given by

$$\psi(n) = \left[-\hat{d}(n-1) \ldots -\hat{d}(n-N) \quad \hat{z}(n-1) \ldots \hat{z}(n-M) \\ \sum_{j=0}^{M} \hat{b}_j(n) \hat{v}_1(n-j) \ldots \sum_{j=0}^{M} \hat{b}_j(n) \hat{v}_L(n-j) \right]^T. \quad (22)$$

In the above equations, $[\cdot]^T$ denotes the transposition operation. Let us further define the parameter vector

$$\hat{\theta}(n) = \\ \left[\hat{a}_1(n) \ldots \hat{a}_N(n) \quad \hat{b}_1(n) \ldots \hat{b}_M(n) \quad \hat{w}_1(n) \ldots \hat{w}_L(n) \right]^T, \\ \hat{b}_0(n) = 1 \quad (23)$$

and the data vector

$$\hat{\mathbf{H}}(n) = [-\hat{d}(n-1) \ldots -\hat{d}(n-N) \\ \hat{z}(n-1) \ldots \hat{z}(n-M) \quad \hat{v}_1(n) \ldots \hat{v}_L(n)]^T. \quad (24)$$

Then, the equations for adaptation of the parameters of the filter can be written in matrix form as

$$e(n) = d(n) - \hat{\mathbf{H}}^T(n) \hat{\theta}(n-1) \quad (25)$$

and

$$\hat{\theta}(n) = \hat{\theta}(n-1) + \mathbf{\Lambda}(n) \phi(n) e(n) \quad (26)$$

where

$$\mathbf{\Lambda}(n) = \mathrm{diag}[\mu_1(n) \quad \cdots \quad \mu_{N+M+L}(n)]. \quad (27)$$

To improve the convergence behavior of the adaptive filter, we use a normalized version of the adaptive filter given by

$$\hat{\theta}(n) = \hat{\theta}(n-1) + \frac{\mathbf{\Lambda}(n) \phi(n)}{\delta + \hat{\mathbf{H}}^T(n) \mathbf{\Lambda}(n) \phi(n)} e(n) \quad (28)$$

where δ is a small positive constant included to prevent division by very small values when $\hat{\mathbf{H}}^T(n) \mathbf{\Lambda}(n) \phi(n)$ is small. It is not difficult to show that this choice of normalization with $\delta = 0$ results in zero *a posteriori* error. Furthermore, of the multitude of update strategies that result in zero *a posteriori* error, the above choice results in the smallest perturbation of the coefficient vector [16]. Finally, we employ Newton's method to improve the update strategy by scaling the coefficient increment vector by the inverse of $\mathbf{R}(n)$, the autocorrelation matrix of the information vector. The matrix $\mathbf{R}(n)$ is recursively computed as

$$\mathbf{R}(n) = \lambda \mathbf{R}(n-1) + (1-\lambda) \phi(n) \phi^T(n) \quad (29)$$

where $0 \ll \lambda < 1$. Its inverse may be evaluated recursively using the matrix inversion lemma as

$$\mathbf{R}^{-1}(n+1) \\ = \frac{1}{\lambda} \left(\mathbf{R}^{-1}(n) - \frac{\mathbf{R}^{-1}(n) \phi(n) \phi^T(n) \mathbf{R}^{-1}(n)}{\frac{\lambda}{1-\lambda} + \phi^T(n) \mathbf{R}^{-1}(n) \phi(n)} \right). \quad (30)$$

While implementing (30) care should be taken to ensure the symmetry of $\mathbf{R}^{-1}(n + 1)$ in (30). Incorporating (30) into (28) results in the update equation for our adaptive filter

$$\hat{\theta}(n) = \hat{\theta}(n - 1) + \frac{\mathbf{\Lambda}(n)\mathbf{R}^{-1}(n + 1)\boldsymbol{\phi}(n)}{\delta + \hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n)} e(n). \quad (31)$$

It might appear that the orthogonalization of the input to the polynomial subsystem in addition to the orthogonalization provided by Newton's update strategy is redundant. However, it is shown in Section III that the orthogonalization for the polynomial subsystem improves the overall performance of the algorithm for even modestly large orders of the nonlinearity.

A. Orthogonalization of the Input to the Polynomial Subsystem

Fig. 2 shows the procedure for orthogonalizing the input to the static nonlinearity in the model. It is evident that

$$\hat{v}_1(n) = x(n). \quad (32)$$

The rest of the orthogonalized signals $\hat{v}_l(n)$ are computed as

$$\hat{v}_l(n) = x^l(n) - \sum_{j=1}^{l-1} \hat{g}_{j,l}(n)\hat{v}_j(n), \quad l = 2, \ldots, L. \quad (33)$$

Since the optimal minimum mean-square error is orthogonal to the input signals, an adaptation strategy that attempts to reduce $e_{j,l}^2(n)$ given by

$$e_{jl}(n) = x^l(n) - \sum_{s=1}^{j} \hat{g}_{s,l}(n) \cdot \hat{v}_s(n),$$
$$l = 2, \ldots, L, \quad j = 1, \ldots, l - 1 \quad (34)$$

will accomplish the orthogonalization procedure. Therefore, the coefficient update strategy for the Gram–Schmidt processor can be written as

$$\hat{g}_{j,l}(n + 1) = \hat{g}_{j,l}(n) + \frac{\mu_{jl}(n)}{\delta_o + \hat{\sigma}_{v_j}^2(n)} \cdot e_{jl}(n) \cdot \hat{v}_j(n),$$
$$l = 2, \ldots, L, \quad j = 1, \ldots, l - 1 \quad (35)$$

where δ_o is a small positive constant and $\hat{\sigma}_{v_j}^2(n)$ is computed as

$$\hat{\sigma}_{v_j}^2(n) = \lambda_g \hat{\sigma}_{v_j}^2(n - 1) + (1 - \lambda_g)\hat{v}_j^2(n),$$
$$j = 1, \ldots, l - 1, \quad 0 < \lambda_g < 1. \quad (36)$$

B. Step-Size Selection

We follow the work in [13] to derive a step-size sequence that guarantees stable operation of the IIR subsystem. The basis for the step-size derivation is the following theorem [17].

Theorem 1: The linear state equation

$$\mathbf{x}_s(n + 1) = \mathbf{A}(n)\mathbf{x}_s(n), \quad \mathbf{x}_s(n_0) = \mathbf{x}_0 \quad (37)$$

is uniformly exponentially stable if and only if there exists an $N \times N$ matrix sequence $\mathbf{Q}(n)$ that is symmetric for all n and such that

$$\eta\mathbf{I} \leq \mathbf{Q}(n) \leq \rho\mathbf{I} \quad (38)$$

and

$$\mathbf{A}^T(n)\mathbf{Q}(n + 1)\mathbf{A}(n) - \mathbf{Q}(n) \leq -\gamma\mathbf{I} \quad (39)$$

where η, ρ, and γ are finite positive constants. The condition that there exists a positive constant η such that $\eta\mathbf{I} \leq \mathbf{Q}(n)$ for all n is equivalent by definition to existence of a positive η such that $\eta\|\mathbf{x}_\eta\|^2 \leq \mathbf{x}_\eta^T\mathbf{Q}(n)\mathbf{x}_\eta$ for all n and all $N \times 1$ vectors $\mathbf{x}_\eta$, where $\|\mathbf{x}_\eta\| = \sqrt{\mathbf{x}_\eta^T\mathbf{x}_\eta}$.

These are the so-called *Lyapunov conditions*. Since our adaptation algorithm is of the form

$$\hat{\theta}(n + 1) = \hat{\theta}(n) + \mu(n)\boldsymbol{\phi}_L(n) \quad (40)$$

where $\boldsymbol{\phi}_L(n) = (\mathbf{\Lambda}(n)\mathbf{R}^{-1}(n + 1)\boldsymbol{\phi}(n)e(n))/(\delta + \hat{\mathbf{H}}^T(n)\mathbf{\Lambda}(n)\boldsymbol{\phi}(n))$, we can transform the direct-form representation of the IIR filter to the state space representation and use Theorem 1. The Hammerstein model can be written in the state–space form as

$$\mathbf{x}_s(n + 1) = \mathbf{A}(n)\mathbf{x}_s(n) + \mathbf{B}(n)\mathbf{u}_s(n) \quad (41)$$
$$\mathbf{y}_s(n) = \mathbf{C}(n)\mathbf{x}_s(n) + \mathbf{D}(n)\mathbf{u}_s(n) \quad (42)$$

where

$$\mathbf{x}_s(n) = [\hat{d}(n - 1) \quad \hat{d}(n - 2) \quad \cdots \quad \hat{d}(n - N)]^T \quad (43)$$
$$\mathbf{u}_s(n) = [\hat{z}(n - 1) \quad \hat{z}(n - 2) \quad \cdots$$
$$\hat{z}(n - M) \quad \hat{v}_1(n) \quad \cdots \quad \hat{v}_L(n)]^T \quad (44)$$
$$\mathbf{y}_s(n) = [\hat{d}(n) \quad \hat{z}(n)]^T \quad (45)$$

$$\mathbf{A}(n) = \begin{bmatrix} -\hat{a}_1(n) & -\hat{a}_2(n) & \cdots & -\hat{a}_{N-1}(n) & -\hat{a}_N(n) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$
$$(46)$$

$$\mathbf{B}(n) = \begin{bmatrix} \hat{b}_1(n) & \hat{b}_2(n) & \cdots & \hat{b}_M(n) & \hat{w}_1(n) & \cdots & \hat{w}_L(n) \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$
$$(47)$$

$$\mathbf{C}(n) = \begin{bmatrix} -\hat{a}_1(n) & -\hat{a}_2(n) & \cdots & -\hat{a}_N(n) \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (48)$$

and

$$\mathbf{D}(n) = \begin{bmatrix} \hat{b}_1(n) & \hat{b}_2(n) & \cdots & \hat{b}_M(n) & \hat{w}_1(n) & \cdots & \hat{w}_L(n) \\ 0 & 0 & \cdots & 0 & \hat{w}_1(n) & \cdots & \hat{w}_L(n) \end{bmatrix}.$$
$$(49)$$

Using the results of [17] and the ideas employed in [13], we can show that the coefficients at times n and $n + 1$ must satisfy the inequality

$$\|\text{vec}[\mathbf{Q}(n + 1)] - \text{vec}[\mathbf{Q}(n)]\| < 1 \quad (50)$$

where

$$\text{vec}[\mathbf{Q}(n + 1)] = -[\mathbf{A}^T(n) \otimes \mathbf{A}^T(n) - \mathbf{I}_{N^2}]^{-1}\text{vec}[\mathbf{I}_N]. \quad (51)$$

In the above expressions, $\text{vec}[\cdot]$ denotes an operation where the columns of a matrix are stacked together to form a (column) vector, and $\otimes$ denotes the Kronecker product. During the operation of the adaptive filter, the step sizes are selected such that it is the smaller of a pre-selected maximum or the maximum value

that maintains the inequality in (50). Selection of the step size is done in the following manner. For a given step size during any iteration, we calculate the new parameter values $\hat{a}_i$, construct $\text{vec}[\mathbf{Q}(n+1)]$ using (51) and verify if the updated coefficients meet the condition in (50). This algorithm ensures that all the instantaneous poles of the system are inside the unit circle, and therefore it is not necessary to separately check for the location of the poles.

In addition, we reduce the step sizes in $\Lambda(n)$ so that $-\delta\beta < \hat{\mathbf{H}}^T(n)\Lambda(n)\boldsymbol{\phi}(n)$, where δ is a small positive constant and $0 < \beta < 1$. This guarantees preservation of the sign for the gradient vector, i.e., ensures that $\delta + \hat{\mathbf{H}}^T(n)\Lambda(n)\boldsymbol{\phi}(n) > 0$, and that $(\Lambda(n)\mathbf{R}^{-1}(n+1)\boldsymbol{\phi}(n))/(\delta + \hat{\mathbf{H}}^T(n)\Lambda(n)\boldsymbol{\phi}(n))e(n)$ moves the coefficients in the correct direction, toward values that reduce the instantaneous squared estimation error during each iteration as intended by the derivations of the algorithm. For positive values of β that is strictly less than one, it is possible to show convergence of the algorithm to the global minimum of the error surface. We now have all the equations necessary to implement the adaptive filter. The complete algorithm is given in Table I.

C. Computational Complexity

For simplifying our discussion, we assume that N, the number of feedback coefficients of the linear subsystem, is larger than or equal to M, the number of feedforward coefficients and L, the number of coefficients of the polynomial subsystem. The algorithm of Table I can be implemented using $\bigcirc(N^{2.25} + L^2 N)$ arithmetical operations per iteration.

There are three parts of the algorithm that require more than linear complexity $\bigcirc(N + M + L)$ for implementation. The Gram–Schmidt orthogonalizer requires $\bigcirc(L^2)$ operations per iteration. The search for a step size that guarantees the stability of the algorithm requires the computation of the inverse matrix $[\mathbf{A}^T(n) \otimes \mathbf{A}^T(n) - \mathbf{I}_{N^2}]^{-1}$. Since this is a sparse matrix with no more than five nonzero entries per row, efficient algorithms to invert the matrix with a computational complexity that corresponds to $\bigcirc(N^{2.25})$ arithmetical operations are available in the literature when N is sufficiently large [18]. The third contribution to the computational complexity comes from the computation of $\mathbf{R}^{-1}(n)$ and updating the coefficients using the result. Fast algorithms for inversion of the autocorrelation matrix requiring $\bigcirc(L^2 N)$ arithmetical operations per iteration can be derived [19]. Updating the coefficients as in (31) requires an additional $\bigcirc((N + M + L)^2)$ operations. Under our assumption $L \leq N$ and $M \leq N$, this complexity can be compactly represented as that of $\bigcirc(N^2)$ arithmetical operations. Our implementations entailed somewhat higher complexity since we did not make use of the simplifications possible for the two matrix inversions.

D. Stability of the Adaptive Hammerstein Filter

We guarantee the bounded-input bounded-output (BIBO) stability of the system with the help of the following lemma [17].

Lemma 1: If the linear state equation (37) is uniformly exponentially stable, and there exist finite constants c_B, c_C, and c_D such that

$$\|\mathbf{B}(n)\| \leq c_B, \quad \|\mathbf{C}(n)\| \leq c_C, \quad \|\mathbf{D}(n)\| \leq c_D \qquad (52)$$

TABLE I
ADAPTIVE HAMMERSTEIN FILTER WITH GRAM–SCHMIDT
ORTHOGONALIZATION OF THE POLYNOMIAL SUBSYSTEM
AND VARIABLE STEP SIZE FOR THE LINEAR SUBSYSTEM

Definitions

n ... time instant
λ_g ... positive constant, $0 < \lambda_g < 1$
δ ... small positive constant
δ_o ... small positive constant
β ... small positive constant such that $0 < \beta < 1$
$\mu_{j,l}(n)$ ... step size for $\hat{g}_{j,1}(n)$
$\Lambda(n) = diag[\ \mu_1(n) \ \cdots \ \mu_{N+M+L}(n)\ ]$
$\hat{\theta}(n) = [\ \hat{a}_1(n)\cdots\hat{a}_N(n)\ \hat{b}_1(n)\cdots\hat{b}_M(n)\ \hat{w}_1(n)\cdots\hat{w}_L(n)\ ]^T,$
$\qquad\qquad \hat{b}_0(n) = 1$
$\hat{\mathbf{H}}(n) = [\ -\hat{d}(n-1)\cdots-\hat{d}(n-N)\ \ \hat{z}(n-1)\cdots\hat{z}(n-M)$
$\qquad\qquad\qquad \hat{v}_1(n)\cdots\hat{v}_L(n)\ ]^T$
$\mathbf{x}(n) = [\ x(n)\ \ x^2(n)\ \ x^3(n)\ \ \ldots\ \ x^L(n)\ ]^T$
$\hat{\mathbf{v}}(n) = [\ \hat{v}_1(n)\ \ \hat{v}_2(n)\ \ \hat{v}_3(n)\ \ \cdots\ \ \hat{v}_L(n)\ ]^T$
$\hat{A}(n, q^{-1}) = 1 + \hat{a}_1(n)q^{-1} + \cdots + \hat{a}_N(n)q^{-N}$
$\hat{B}(n, q^{-1}) = 1 + \hat{b}_1(n)q^{-1} + \cdots + \hat{b}_M(n)q^{-M}$
$\hat{\mathbf{w}}(n) = [\ \hat{w}_1(n)\ \ \hat{w}_2(n)\ \ \ldots\ \ \hat{w}_L(n)\ ]^T$

$$\hat{\mathbf{G}}^T(n) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \hat{g}_{1,2}(n) & 1 & 0 & \ldots & 0 \\ \hat{g}_{1,3}(n) & \hat{g}_{2,3}(n) & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{g}_{1,L}(n) & \hat{g}_{2,L}(n) & \hat{g}_{3,L}(n) & \ldots & 1 \end{bmatrix}$$

$\hat{\mathbf{C}}(n) = \hat{\mathbf{G}}^{-1}(n)$

$$\mathbf{A}(n) = \begin{bmatrix} -\hat{a}_1(n) & -\hat{a}_2(n) & \ldots & -\hat{a}_{N-1}(n) & -\hat{a}_N(n) \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}_{N \times N}$$

Initialization

$$\hat{d}(0),\ \hat{\theta}(0),\ \phi(0),\ \mathbf{R}^{-1}(0) = \delta_R\mathbf{I}, \qquad \delta_R \neq 0$$

Main Loop

$\hat{\mathbf{v}}(n) = \hat{\mathbf{C}}^T(n) \cdot \mathbf{x}(n)$

$e_{jl}(n) = x^l(n) - \sum_{s=1}^{j} \hat{g}_{s,l}(n) \cdot \hat{v}_s(n), l = 2, ..., L,\ j = 1, ..., l-1$

$\hat{\sigma}_{v_j}^2(n) = \lambda_g\hat{\sigma}_{v_j}^2(n-1) + (1-\lambda_g)\hat{v}_j^2(n)$

$\hat{g}_{j,l}(n+1) = \hat{g}_{j,l}(n) + \dfrac{\mu_{jl}(n)}{\delta_o + \hat{\sigma}_{v_j}^2(n)} \cdot e_{jl}(n) \cdot \hat{v}_j(n), \quad l = 2, ..., L,$
$\qquad\qquad\qquad\qquad j = 1, ..., l-1$

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

$\psi(n) = [\ -\hat{d}(n-1)\cdots-\hat{d}(n-N)\ \ \hat{z}(n-1)\cdots\hat{z}(n-M)$
$\qquad\ \sum_{j=0}^{M}\hat{b}_j(n)\hat{v}_1(n-j)\cdots\sum_{j=0}^{M}\hat{b}_j(n)\hat{v}_L(n-j)\ ]^T$

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

$\mathbf{R}^{-1}(n+1) = \dfrac{1}{\lambda}\left(\mathbf{R}^{-1}(n) - \dfrac{\mathbf{R}^{-1}(n)\phi(n)\phi^T(n)\mathbf{R}^{-1}(n)}{\frac{\lambda}{1-\lambda} + \phi^T(n)\mathbf{R}^{-1}(n)\phi(n)}\right)$

Verify that $\{\Lambda(n)\}$ is such that $\|vec[\mathbf{Q}(n+1)] - vec[\mathbf{Q}(n)]\| < 1$,
  where $vec[\mathbf{Q}(n+1)] = -[\mathbf{A}^T(n) \otimes \mathbf{A}^T(n) - \mathbf{I}_{N^2}]^{-1}vec[\mathbf{I}_N]$.
  Reduce elements of $\Lambda(n)$ if $\hat{\mathbf{H}}^T(n)\Lambda(n)\phi(n) \leq -\delta\beta$.

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

$\hat{z}(n) = \hat{\mathbf{w}}^T(n) \cdot \hat{\mathbf{v}}(n)$

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

for all n, the state equation (41)–(42) is also uniformly bounded-input, bounded-output stable.

The selection of the step-size sequence ensures that the homogeneous system in (37) is uniformly exponentially stable. To satisfy condition (52), we limit the maximum values of the entries of $\mathbf{B}(n)$, $\mathbf{C}(n)$, and $\mathbf{D}(n)$ to some large, but finite values. In the experiments described later, this bound was set to 100; however, the parameters never reached this value. The orthogonalized signals $\hat{v}_l(n)$ are computed using (33). The input signal $x(n)$ is assumed to be bounded. The update in (35) is a classical adaptive FIR-type equation that is known to be stable, for sufficiently small step sizes $\mu_{jl}(n)$. Consequently, all of the signals $\hat{v}_l(n)$ are bounded. The intermediate output $\hat{z}(n)$ is calculated as $\hat{z}(n) = \hat{w}_1(n) \cdot \hat{v}_1(n) + \hat{w}_2(n) \cdot \hat{v}_2(n) + \cdots + \hat{w}_L(n) \cdot \hat{v}_L(n)$. Because $\hat{w}_l(n)$ are bounded by finite constants and $\hat{v}_l(n)$ are finite, $\hat{z}(n)$ is also finite.

For a linear state (41)–(42) that is uniformly BIBO stable, there exists a finite constant η such that for any fixed initial time n_0 and any input signal $\mathbf{u}_s(n)$ the corresponding zero-state (i.e., the initial state is fixed at zero) response satisfies the bound

$$\sup_{n \geq n_0} \|\mathbf{y}_s(n)\| \leq \eta \sup_{n \geq n_0} \|\mathbf{u}_s(n)\|. \tag{53}$$

The adjective "uniform" emphasizes the fact that the same η can be used for all values of n_0 and for all input signals $\mathbf{u}_s(n)$. Uniform BIBO stability thus guarantees that the input–output behavior of the adaptive Hammerstein filter exhibits finite gain in terms of the input and output suprema. Consequently, the adaptive filter of Table I operates in a stable manner.

III. SIMULATION RESULTS

In this section, we present the results of several simulation experiments conducted to evaluate the performance capabilities of the adaptive filter derived in Section II. All simulations involved the identification of an unknown Hammerstein system. In our first experiment, the input–output relationship of the memoryless nonlinearity of the unknown system was

$$z(n) = 0.4x(n) - 0.3x^2(n) + 0.2x^3(n) \tag{54}$$

and the transfer function of the linear component was

$$H(z) = \frac{1 + 0.8z^{-1}}{1 - 1.2z^{-1} + 0.5z^{-2}}. \tag{55}$$

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 such that the output signal-to-noise ratio (SNR) was 20 dB. The input signal $x(n)$ of the adaptive filter was generated by filtering a uniform signal with zero mean value and 0.19 variance with the filter

$$H_c(z) = 1 + 0.95z^{-2}. \tag{56}$$

The histogram and the power spectral density (PSD) of this non-Gaussian input are shown in Fig. 3. The adaptive filter was implemented with the time-varying step size of the recursive part to be the maximum of $\mu = 1.5 \cdot 10^{-5}$ or the bound suggested
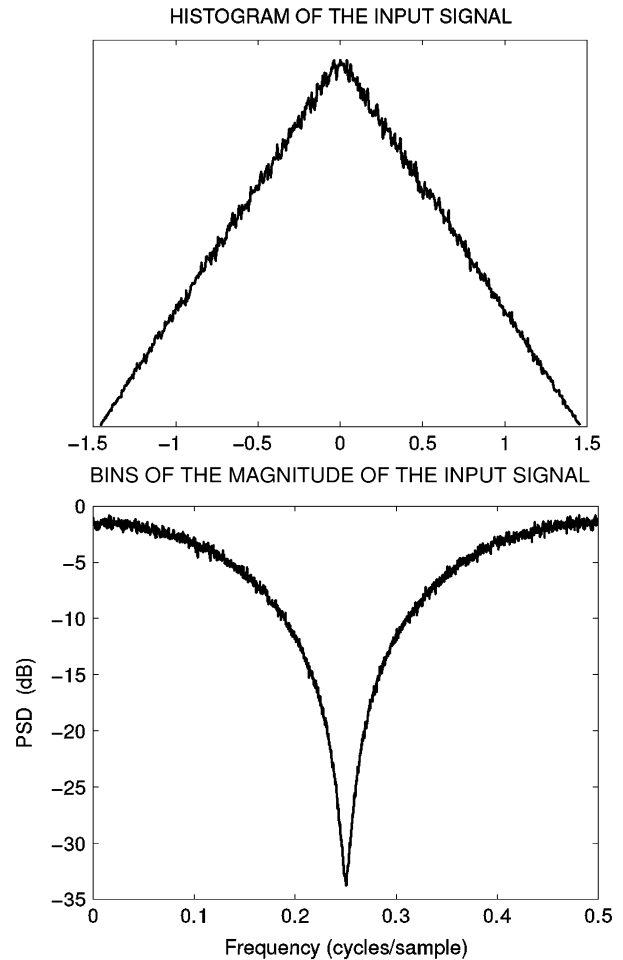

Fig. 3. Histogram and PSD of the input signal.

by the condition in (50), and the step size for the feedforward coefficients of the linear subsystem and the coefficients of the polynomial subsystem was constant and equal to $\mu = 1.5 \cdot 10^{-5}$. The step sizes for the coefficients of the orthogonalizer were constant and equal to $5 \cdot 10^{-4}$. δ_o was set to 10^{-5}, δ to 10^{-3}, λ_g to 0.99, β to 0.9, and λ to 0.95. Initial value of $\mathbf{R}^{-1}$ was a diagonal matrix with the values $\delta_R = 1$ on the diagonal entries. Initial values of $\hat{\sigma}_{v_j}^2(n)$ were set to 10. The system was initialized with poles and zeros 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. In order to ensure the BIBO stability of the orthogonalizer and the adaptive linear filter feedforward part, we also imposed an upper bound on the absolute value of orthogonalizer and feedforward coefficients. Two hundred independent experiments using 15 000 data samples each were conducted. The results presented are average values over these 200 experiments. The performance of the algorithm was compared with the one described in [9]. The initial value of the error covariance matrix $\mathbf{P}(n)$ in the Kalman filter based algorithm uses diagonal matrix with the value 0.001 on the diagonal. The parameters were chosen such that the excess mean-square errors after 15 000 iterations of the adaptive filter were approximately the same for both algorithms.
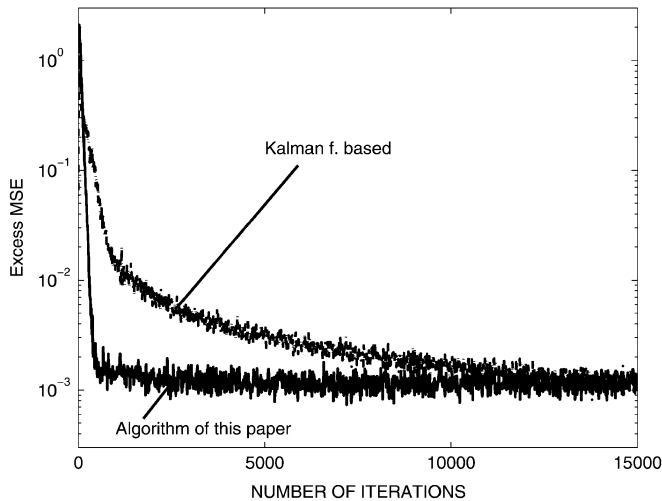
Fig. 4.   Excess MSE of the two adaptive filters in the first experiment.

Fig. 5.   Evolution of the mean-square estimation error of the two adaptive filters in the second experiment.

TABLE II
COMPARISON OF THE PERFORMANCE OF THE TWO ADAPTIVE
FILTERS IN THE FIRST EXPERIMENT

| | | Algorithm of this paper | | Kalman based [9] | |
Coeff.	True	Mean	Var. $[10^{-4}]$	Mean	Var. $[10^{-4}]$
$\widehat{a}_1$	-1.2	-1.2018	0.6931	-1.2362	0.4406
$\widehat{a}_2$	0.5	0.5012	0.3920	0.5260	0.2277
$\widehat{b}_0$	1	-	-	-	-
$\widehat{b}_1$	0.8	0.7952	2.4026	0.6577	6.9609
$\widehat{p}_1$	0.4	0.4006	0.2428	0.4175	0.1804
$\widehat{p}_2$	-0.3	-0.2998	0.6038	-0.3149	0.0768
$\widehat{p}_3$	0.2	0.1977	0.9330	0.2063	0.0681

Fig. 4 shows the evolution of the excess mean-square error (MSE) of both the algorithms. The results show that the algorithm of this paper converged faster than the Kalman filter-based algorithm. Table II tabulates the statistics associated with the coefficients of our algorithm and the algorithm based on extended Kalman filter. Coefficients of the polynomial subsystem $\widehat{p}_l$ were obtained by conversion of the coefficients $\widehat{w}_l$ to the direct form. Detailed derivations of the conversion equations are omitted here. However, we note that the relationship between the coefficients of the direct form filter and the orthogonal signals is linear, and therefore the transformation from one set to the other is straightforward [16]. The mean values of the coefficients in the Table II for our algorithm were calculated as the mean values of the last 500 samples of the coefficients over all the runs. Variances of the coefficients were calculated in a similar way. Our algorithm operates in a stable manner as predicted by the theoretical derivations. It also converged to the global minimum in all 200 runs. We can see from these results that both algorithms work reasonably well in this set of experiments. However, we will see in the next example that the algorithm of this paper outperforms the Kalman filter-based method when the underlying system is more complex. We evaluated the capabilities of the algorithm at several noise levels. We have observed that the system performs well even in situations when SNR = 0 dB. For the parameter values that we employed in these experiments, our experience is that the behavior of the adaptive filter is less reliable at SNRs below 0 dB.
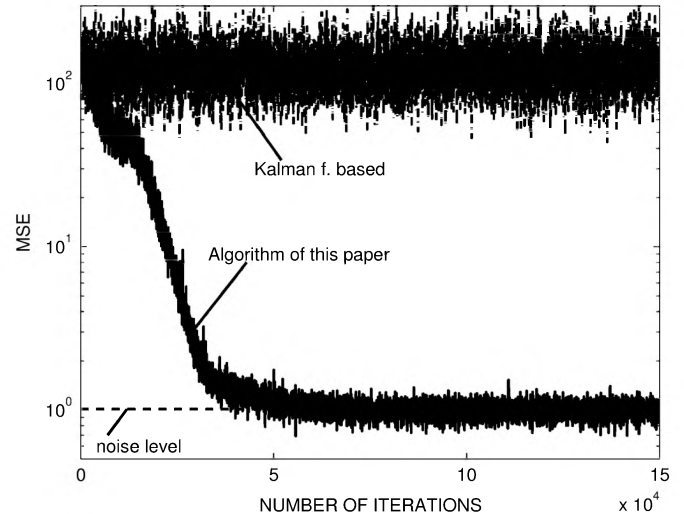
In the next set of simulations, the adaptive filter was used in a system identification problem where the linear component of the unknown system satisfied the input–output relationship

$$
\begin{aligned}
H(z) &= \tilde{D}(z)/Y(z) \\
&= \frac{1 - 1.3334z^{-1} + 1.6667z^{-2} - 2.665z^{-3} + 1.9666z^{-4}}{1 - 2.6z^{-1} + 2.78z^{-2} - 1.48z^{-3} + 0.34z^{-4}}.
\end{aligned}
\tag{57}
$$

The poles of this system are located at $z_{1,2} = 0.8 \pm j0.2, z_{3,4} = 0.5 \pm j0.5$. The input–output relationship of the memoryless nonlinearity was again given by (54). The simulation conditions were similar to those described for the previous simulation experiment except that the input was a colored Gaussian signal obtained by processing a white Gaussian signal with zero mean value and unit variance with the linear system $H_k(z) = 1 + 0.5z^{-1}$. Because of the larger dynamics of the linear subsystem, this is a more difficult overall system for the adaptive filters to converge to. Fig. 5 shows the evolution of the MSE of our algorithm and the Kalman filter based algorithm [9] using a semi-log graph. As before, the results are plotted after processing the ensemble average of the squared error over the 200 runs. The coefficients of the unknown system, mean values of the adaptive filter coefficients, and their variances after convergence are shown in Table III. Again, our algorithm operated in a stable manner as predicted by the theoretical derivations. The experiments resulted in convergence to the global minimum in all 200 runs for our algorithm, whereas the extended Kalman filter based algorithm never converged within 150 000 iterations. This may be because of slow adaptation of the extended Kalman based filter. The evolution of the mean values of the step size for the denominator coefficients over the 200 runs is shown in Fig. 6. The evolution of the coefficients of the linear subsystem, as well as the evolution of the coefficients of the nonlinear subsystem are shown in Fig. 7. We can observe that the coefficients $\widehat{b}_j(n)$ and $\widehat{w}_l(n)$ converge at about the same rate, whereas

TABLE III
STATISTICS OF THE COEFFICIENTS OF THE ADAPTIVE FILTER IN THE
SECOND EXPERIMENT WITH IDEAL COEFFICIENTS

Coefficient	True value	Mean	Variance $[10^{-4}]$
a_1	-2.6	-2.6005	0.0716
a_2	2.78	2.7810	0.4109
a_3	-1.48	-1.4807	0.3301
a_4	0.34	0.3402	0.0363
b_0	1	-	-
b_1	-1.3334	-1.3333	1.7014
b_2	1.6667	1.6643	5.9578
b_3	-2.6665	-2.6626	5.8153
b_4	1.9666	1.9646	2.1292
p_1	0.4	0.3994	0.3344
p_2	-0.3	-0.2998	0.1632
p_3	0.2	0.1999	0.0210

Fig. 7. Evolution of the mean values of the coefficients of the adaptive filter in the second experiment.
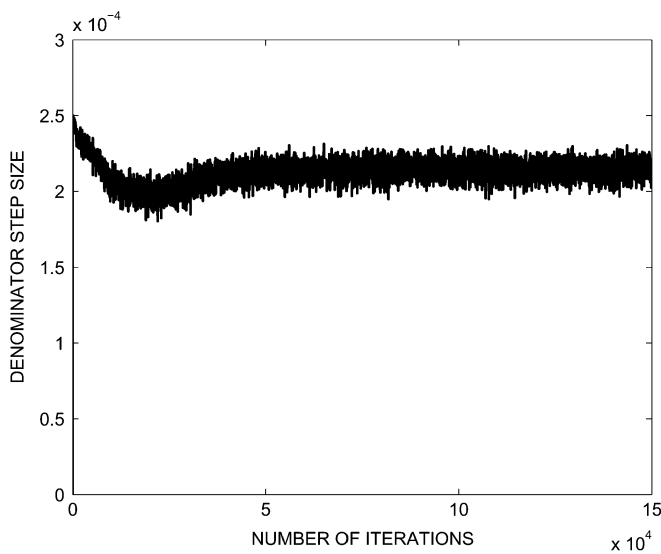

Fig. 6. Evolution of the mean values of the step size for the adaptation of the denominator coefficients of the adaptive filter in the second experiment.

the coefficients $\hat{a}_i(n)$ converged somewhat faster in this experiment. These results indicate that step-size selection using the condition in (50) results in stable operation of the linear IIR section of the adaptive filter. The initial values of step sizes are (usually) small since the initial estimation error is large. Combined with the large error, the initial values of the step size produced the largest changes possible that still maintained the exponential stability of the (linear) subsystem.

The differences between the adaptive filter with orthogonalization of the input signal to the polynomial subsystem is accentuated in situations where the polynomial subsystem requires larger model orders. This is because the eigenvalue spread of the autocorrelation matrix of the input vector to the polynomial subsystem is extremely high for large model orders. To demonstrate the differences between adaptive filters with and without the orthogonalizer, we consider a plant composed of a polynomial

$$z(n) = 1.4x(n) + 0.7x^2(n) + 1.2x^3(n)$$
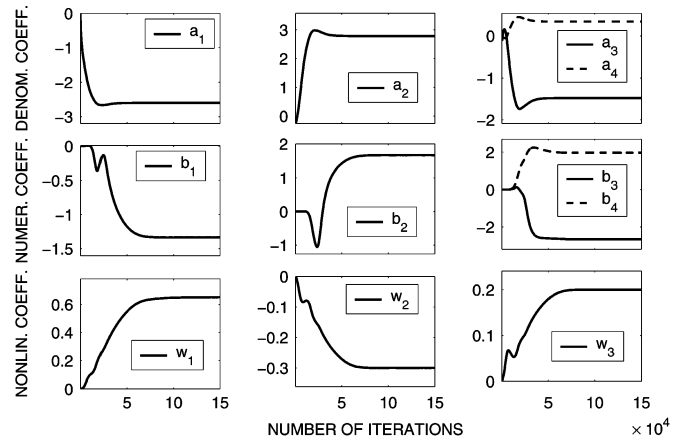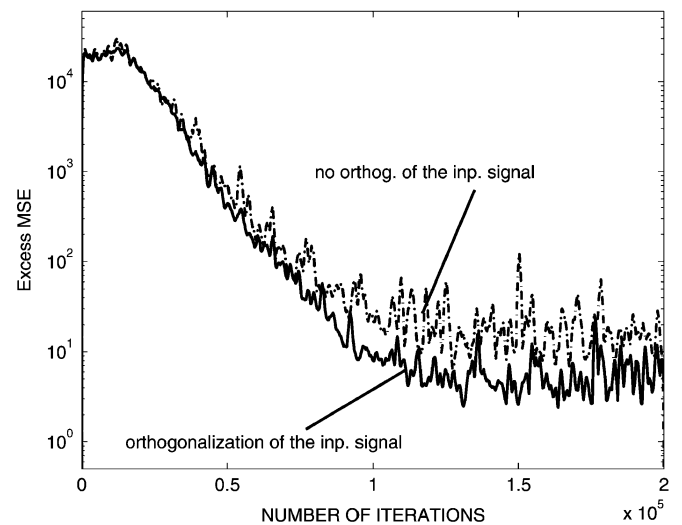$$+ 0.9x^4(n) + 1.2x^5(n) \quad (58)$$

Fig. 8. MSE comparing the orthogonalization case with the case where orthogonalization was not performed.

and a linear component whose transfer function is given by (55). Fig. 8 shows the evolution of the sliding window averaged excess MSE of the two filters with a window length of 50 samples. The time averaging was done to reduce the variability of the two curves so that the differences in performance can be better observed. Comparing the results of the orthogonal case with the nonorthogonal case, we can see that the convergence speed of the two methods is comparable in this example. However, the algorithm without orthogonalization exhibits higher excess MSE indicating that the algorithm equipped with the orthogonalizer performs better than the algorithm that is not so equipped. Both algorithms converged to the global minimum in all 200 experiments.

We have evaluated the algorithm of this paper in a large number of synthetic scenarios, including in situations that involve model mismatch, random (but stable) initial values of the adaptive filter and time-varying operating environments. The method provided good performance in terms of stable operation and global convergence in each case.

IV. CONCLUDING REMARKS

This work presented the derivation and performance evaluation of an adaptive nonlinear filter employing the Hammerstein system model. The model, consisting of a static nonlinearity followed by a recursive linear system, is useful in many applications including in the modeling of communications systems, biological systems, and chemical and biological detectors. Our system employs a step-size sequence that guarantees stable operation of the adaptive filter.

The convergence behavior of this algorithm is studied in a stochastic framework and in a nonstationary environment in [20]. Using the martingale limit theorem, it is shown there, that under the conditions of the analysis the long-term time average of the squared estimation error of the adaptive filter can be made arbitrarily close to its minimum possible value when the underlying system is time-invariant. The ability to converge to the the global minimum as well as its stability properties make the adaptive filter presented in this paper an attractive candidate in applications in which Hammerstein models are appropriate.

REFERENCES

[1] S. Benedetto and E. Biglieri. "Nonlinear equalization of digital satellite channels." *IEEE J. Sel. Areas Commun.*, vol. SAC-1, no. 1, pp. 57–62. Jan. 1983.

[2] I. W. Hunter and M. J. Korenberg. "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models." *Biol. Cybern.*, vol. 55, no. 2–3, pp. 135–144, 1986.

[3] E. Eskinat, S. H. Johnson, and W. L. Luyben. "Use of Hammerstein models in identification of nonlinear systems." *AIChE J.*, vol. 37, pp. 255–268, 1991.

[4] A. Balestrino, A. Landi, M. Ould-Zmirli, and L. Sani. "Automatic nonlinear auto-tuning method for Hammerstein modeling of electrical drives." *IEEE Trans. Ind. Electron.*, vol. 48, no. 3, pp. 645–655, Jun. 2001.

[5] J. Kim and K. Konstantinou. "Digital predistortion of wide-band signals based on power amplifier model with memory." *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418. Nov. 8, 2001.

[6] W. Greblicki and M. Pawlak. "Nonparametric identification of Hammerstein systems." *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 409–418, Mar. 1989.

[7] Z. Hasiewicz. "Non-parametric estimation of nonlinearity in a cascade time-series system by multiscale approximation." *Signal Process.*, vol. 81, no. 4, pp. 791–807, 2001.

[8] H. W. Kang, Y. S. Cho, and D. H. Youn. "Adaptive precompensation of Wiener systems." *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2825–2829, Oct. 1998.

[9] M. Boutayeb and M. Darouach. "Recursive identification method for the Hammerstein model: Extension to nonlinear MISO systems." *Control Theory Adv. Technol.*, vol. 10, no. 1, pp. 57–72, Mar. 1994.

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

[11] N. J. Bershad, P. Celka, and S. McLaughlin. "Analysis of stochastic gradient identification of Wiener–Hammerstein systems for nonlinearities with Hermite polynomial expansions." *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 1060–1072. May 2001.

[12] J. Vörös. "Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities." *IEEE Trans. Autom. Control.*, vol. 44, no. 11, pp. 2145–2149. Nov. 1999.

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

[14] J. R. Treichler, C. R. Johnson Jr., and M. G. Larimore. *Theory and Design of Adaptive Filters.* Upper Saddle River, NJ: Prentice-Hall PTR, 2001.

[15] S. Haykin. *Adaptive Filter Theory.* 4th ed. Englewood Cliffs, N.J.: Prentice-Hall, 2001.

[16] J. Jeraj. "Adaptive estimation and equalization of nonlinear systems." Ph.D. dissertation, Electrical and Computer Eng. Dept., Univ. of Utah, Salt Lake City, UT, 2005.

[17] W. J. Rugh. *Linear System Theory.* Upper Saddle River, NJ: Prentice-Hall, 1996.

[18] G. Villard. "Computation of the inverse and determinant of a matrix." in *Institut Nat. de Recherche en Informatique et en Automatique (INRIA).* F. Chyzak, Ed., Nov. 2003. Rapport de recherche No. 5003, pp. 29–32.

[19] J. M. Cioffi and T. Kailath. "Fast, recursive-least-squares transversal filters for adaptive filtering." *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 304–337, Apr. 1984.

[20] J. Jeraj and V. J. Mathews. "Stochastic mean-square performance analysis of an adaptive Hammerstein filter." *IEEE Trans. Sig. Process.*, vol. 54, no. 6. Jun. 2005, to be published.

Janez Jeraj (S'01–M'03) received the B.S. and M.S. degrees in electrical engineering from the University of Ljubljana, Slovenia, in 1996 and 1999, respectively, and the Ph.D. degree in electrical engineering from the University of Utah, Salt Lake City, in 2005.

Currently, he works at Agilent Technologies, Palo Alto, CA.

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. degree and Ph.D. degree in electrical and computer engineering both from the University of Iowa, Iowa City, in 1981 and 1984, respectively.

From 1980 to 1984, he was a Teaching/Research Fellow and during the 1984–1985 academic year a Visiting Assistant Professor at the University of Iowa in the Department of Electrical and Computer Engineering . In 1985, he joined the University of Utah, Salt Lake City, where he is currently a Professor in the Department of Electrical and Computer Engineering. From 1999 to 2003, he served as the chairman of the department. 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 coauthor (with Prof. G. L. Sicuranza, University of Trieste, Italy) of the book *Polynomial Signal Processing* (New York: Wiley), and he has published more than 100 technical papers.

Dr. Mathews served as the Vice President—Finance of the IEEE Signal Processing Society from 2003 to 2005 and has served on the Publication Board and the Conference Board of the Society. 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. 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.