A STABLE ADAPTIVE HAMMERSTEIN FILTER EMPLOYING PARTIAL ORTHOGONALIZATION OF THE INPUT SIGNALS

Janez Jeraj, V. John Mathews

University of Utah
Dept of Electrical & Computer Engineering
50 South Central Campus Drive,
Salt Lake City, UT 84112, USA

Joel Dubow

University of Utah
Material Sci. & Eng. Dep.
122 South Central Campus Drive,
Salt Lake City, UT 84112, USA

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 work, the static nonlinearity is modeled using a polynomial system and the linear filter that follows the nonlinearity is an infinite impulse response system. The adaptation of the nonlinear components is enhanced in the algorithm by orthogonalizing the inputs to the coefficients of the polynomial system. The linear system is implemented as a recursive higher-order filter. The step sizes associated with the recursive components are constrained in such a way as to guarantee bounded-input, bounded-output stability of the overall system. Experimental results included in the paper show that the algorithm performs well and always converges to the global minimum if the input signal is white.

1. 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 Figure 1. There are many applications in which cascade nonlinear models are appropriate. Examples include the modeling of satellite communication systems [1] and biological systems [2]. The application that has motivated this work is that of modeling biological and chemical detectors. Such systems are known to be nonlinear and our prior work on modeling a US Army ICAM, an ion mobility spectrometer [3], has shown that they can be accurately modeled using a Hammerstein model in which the linear system has infinite impulse response characteristics.

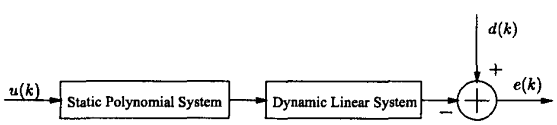

Fig. 1. Block diagram of a Hammerstein system.

In our work, we model the memoryless nonlinearity using a polynomial input-output relationship, and the linear system has infinite impulse response characteristics. The input-output relationship of the adaptive filter is given by

$$\hat{d}(k) = \sum_{i=1}^{N} a_i(k) \cdot \hat{d}(k-i) + \sum_{j=0}^{M} b_j(k) \cdot y(k-j), \quad (1)$$

where $y(k)$ is the output of a static polynomial nonlinear system

$$y(k) = p_1(k) \cdot u(k) + \ldots + p_L(k) \cdot u^L(k) \quad (2)$$

and $u(k)$ is the input to the adaptive filter. In the above equations, $p_l(k)$, $a_i(k)$ and $b_j(k)$ represent the coefficients of the adaptive filter at time k. The objective of the adaptive filter is to update the coefficients during each iteration using a stochastic gradient procedure so as to reduce the instantaneous squared estimation error during each iteration. Our approach is unique in two respects: (1) We orthogonalize the input signal to the polynomial subsystem. This improves the overall convergence behavior of the method. (2) The adaptive IIR subsystem employs a step size sequence that guarantees stability of the system. This work follows that of Carini, et. al. [4], in which the authors employ a Lyapunov stability criteria to develop stabilization algorithms for adaptive IIR filters. The derivation is also based on an adaptive linear filtering algorithm [5] that converges to the global minimum of the error surface for white input signals. Even though this paper does not contain a proof of global convergence, experimental evidence indicates that the system performs in that manner for white input signals.

2. ADAPTATION OF THE HAMMERSTEIN SYSTEM MODEL

We consider the problem of estimating the desired response signal $d(k)$ as the output of the adaptive Hammerstein filter as in (1) when its input is $u(k)$. For this purpose, we employ a stochastic gradient algorithm that attempts to reduce the mean square error at each time. In order to improve the convergence behavior of the adaptive filter, we employ an adaptive Gram-Schmidt orthogonalization procedure for the polynomial subsystem denoted by (2). Unlike the work in [4], the linear subsystem is implemented using the direct form structure, and the step size constraints that guarantee stable operation of the filter at each time are derived online for the direct form system. Finally, in order to obtain a unique solution, we must constrain one of the coefficients of the polynomial or the numerator of the linear component to a fixed value. In all the experimental results described later, we used $b_0(k) = 1$ at all times. Other constraints may be used if necessary.

2.1. Adaptation of Polynomial Subsystem

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

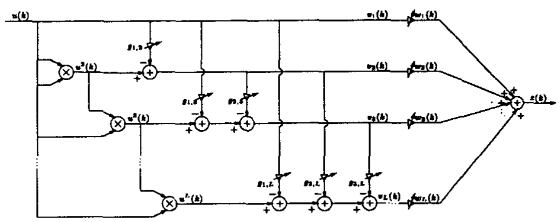

Fig. 2. Orthogonalization of the nonlinear subsystem.

$$v_1(k) = u(k) \quad (3)$$

The rest of the orthogonalized signals $v_l(k)$ are computed as

$$v_l(k) = u^l(k) - \sum_{j=1}^{l-1} g_{j,l}(k) v_j(k). \quad (4)$$

The coefficient update strategy for the coefficients of the Gram-Schmidt processor using LMS adaptive algorithm can be written as

$$g_{j,l}(k+1) = g_{j,l}(k) - \mu_{jl} \cdot \frac{\partial}{\partial g_{j,l}(k)} \left(\frac{1}{2} e_{jl}^2(k) \right), \quad (5)$$

where

$$e_{jl}(k) = u^l(k) - \sum_{s=1}^{j} g_{s,l}(k) \cdot v_s(k). \quad (6)$$

Taking the derivative in (5), results in the update equation

$$g_{j,l}(k+1) = g_{j,l}(k) + \mu_{jl} \cdot e_{jl}(k) \cdot v_j(k). \quad (7)$$

Let

$$z(k) = w_1(k) \cdot v_1(k) + \ldots + w_L(k) \cdot v_L(k). \quad (8)$$

Then,

$$\hat{d}(k) = \sum_{i=1}^{N} a_i(k) \hat{d}(k-i) + \sum_{j=0}^{M} b_j(k) z(k-j). \quad (9)$$

The coefficient update strategy for $w_l(k)$ has the following structure [6]:

$$w_l(k+1) = w_l(k) - \nu_l \frac{\partial(\frac{1}{2}\left[d(k) - \hat{d}(k)\right]^2)}{\partial w_l(k)}, \quad \nu_l > 0.$$

The parameter ν_l is the step size of the adaptive filter, and it controls the speed of convergence as well as the steady state and tracking properties of the system. Since only $\hat{d}(k)$ is a function of the coefficients, we evaluate the partial derivatives in the above expression as

$$\frac{\partial(\frac{1}{2}\left[d(k) - \hat{d}(k)\right]^2)}{\partial w_l(k)} = -\left[d(k) - \hat{d}(k)\right] \frac{\partial \hat{d}(k)}{\partial w_l(k)},$$

with $1 \leq l \leq L$. Evaluating the derivative of (9) with respect to $w_l(k)$, and using the commonly used approximation

$$\frac{\partial \hat{d}(k-s)}{\partial w_l(k-s)} \approx \frac{\partial \hat{d}(k-s)}{\partial w_l(k)}, \qquad 1 \leq l \leq L \quad (10)$$

for sufficiently small ν_l, we get

$$\frac{\partial \hat{d}(k)}{\partial w_l(k)} \approx \sum_{s=1}^{N} a_s(k) \frac{\partial \hat{d}(k-s)}{\partial w_l(k-s)} + \sum_{s=0}^{M} b_s(k) v_l(k-s).$$

Thus we can write coefficient update equations for the nonlinear subsystem as

$$\begin{aligned} \mathbf{w}(k+1) &= \left[\; w_1(k+1) \quad w_2(k+1) \ldots w_L(k+1) \;\right] \\ &= \mathbf{w}(k) + diag\left[\; \nu_1 \quad \nu_2 \ldots \nu_L \;\right] \cdot \\ &\quad \left[\; \frac{\partial \hat{d}(k)}{\partial w_1(k)} \quad \frac{\partial \hat{d}(k)}{\partial w_2(k)} \cdots \frac{\partial \hat{d}(k)}{\partial w_L(k)} \;\right] \cdot e(k). \quad (11) \end{aligned}$$

where $e(k) = d(k) - \hat{d}(k)$.

2.2. Adaptation of Linear IIR Subsystem

The coefficient update strategy for the linear subsystem has the following structure [6]:

$$a_i(k+1) = a_i(k) - \mu_i \frac{\partial(\frac{1}{2}\left[d(k) - \hat{d}(k)\right]^2)}{\partial a_i(k)}, \quad \mu_i > 0,$$

and

$$b_j(k+1) = b_j(k) - \rho_j \frac{\partial(\frac{1}{2}\left[d(k) - \hat{d}(k)\right]^2)}{\partial b_j(k)}, \quad \rho_j > 0,$$

where the parameters μ_i and ρ_j are the step sizes of the adaptive filter. We evaluate the partial derivatives in the above expressions [7] and make similar approximations as in Section 2.1. Assuming that each of μ_i and ρ_j are sufficiently small, and therefore the changes to the coefficients over any interval of duration N or less is very small, we can derive the following expressions:

$$\frac{\partial \hat{d}(k)}{\partial a_i(k)} \approx \hat{d}(k-i) + \sum_{s=1}^{N} a_s(k) \frac{\partial \hat{d}(k-s)}{\partial a_i(k-s)} \quad (12)$$

$$\frac{\partial \hat{d}(k)}{\partial b_j(k)} \approx z(k-j) + \sum_{s=1}^{N} a_s(k) \frac{\partial \hat{d}(k-s)}{\partial b_j(k-s)}. \quad (13)$$

Recursive time-varying homogeneous linear systems are exponentially stable if 1) its instantaneous poles are inside unit circle, and 2) they are sufficiently slowly varying.

Using the results of [8] and the idea employed in [4], where a new upper bound on the maximum allowable coefficient variation for the stability of a direct-form linear recursive filter is derived, the coefficients must satisfy the inequality

$$\| vec\left[\mathbf{Q}(k+1)\right] - vec\left[\mathbf{Q}(k)\right] \| < 1, \quad (14)$$

where

$$vec\left[\mathbf{Q}(k+1)\right] = -\left[\mathbf{A}^T(k) \otimes \mathbf{A}^T(k) - \mathbf{I}_{n^2}\right]^{-1} vec\left[\mathbf{I}_n\right].$$

In the above expression, $\mathbf{A}(k)$ is the system matrix obtained when we transform the direct-form representation in (9) to the state space representation ignoring the feedforward terms (*i.e.*, the coefficients $b_j(k)$).

Let the data vector and the coefficient vector of the linear subsystem be given by

$$\mathbf{x}(k) = [\ \hat{d}(k-1)...\hat{d}(k-N) \quad z(k)...z(k-M)\ ]^T$$

and

$$\theta(k) = [\ a_1(k)...a_N(k) \quad b_0(k)...b_M(k)\ ]^T,$$

respectively. The coefficients are updated in this method as

$$\theta(k+1) = \theta(k) + \Lambda(k) \cdot \mathbf{R}^{-1}(k+1) \cdot \phi(k),$$

where $\Lambda(k)$ is a time-varying step size matrix of the adaptive filter defined as

$$\Lambda(k) = diag\ [\ \mu(k) \dots \mu(k) \quad \rho(k) \dots \rho(k)\ ]$$

and the information vector $\phi^T(k)$ is given by

$$\phi(k) = \frac{1}{1 - a_1(k)q^{-1} - \dots - a_N(k)q^{-N}}\mathbf{x}(k). \qquad (15)$$

In the above expression q^{-1} refers to a unit delay operator. The matrix $\mathbf{R}(k)$ is an estimate of the autocorrelation matrix of the information vector, and is recursively computed as

$$\mathbf{R}(k) = \lambda\mathbf{R}(k-1) + (1-\lambda)\phi(k)\phi^T(k), \qquad (16)$$

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

$$\mathbf{R}^{-1}(k+1) = \frac{1}{\lambda}\left(\mathbf{R}^{-1}(k) - \frac{\mathbf{R}^{-1}(k)\phi(k)\phi^T(k)\mathbf{R}^{-1}(k)}{\frac{\lambda}{1-\lambda} + \phi^T(k)\mathbf{R}^{-1}(k)\phi(k)}\right). \qquad (17)$$

While implementing (17) care should be taken to ensure the symmetry of $\mathbf{R}^{-1}(k+1)$ in (17). We now have all the equations necessary to implement the adaptive filter. 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 (14).

3. EXPERIMENTAL RESULTS

In this section we present the results of an experiment conducted to evaluate the performance capabilities of the adaptive filter derived in the previous section. The adaptive filter was used in a system identification problem where the linear component of the unknown system satisfied the input-output relationship

$$H(z) = \frac{1 - 1.3334z^{-1} + 1.6667z^{-2} - 2.665z^{-3} + 1.9666z^{-4}}{1 - 1.2z^{-1} + 0.74z^{-2} - 0.14z^{-3} + 0.02z^{-4}} \qquad (18)$$

$H(z) = \widetilde{D}(z)/Y(z)$ and the input-output relationship of the memoryless nonlinearity was $y(k) = 0.4u(k) - 0.3u^2(k) + 0.2u^3(k)$. The input signal $u(k)$ of the adaptive filter was white with zero mean value and unit variance. The desired response signal $d(k)$ of the adaptive filter was obtained by corrupting the output of the unknown system $\tilde{d}(k)$ in (18) with additive white noise with zero mean value and variance such that the output SNR was 30 dB. Twenty independent experiments using 8000 data samples each were conducted. The results presented are average values over these twenty experiments.

Table 1. Coefficients of the unknown system, mean values of the adaptive filter coefficients, and their variances after convergence.

Coefficient	True value	Mean	Variance
a_1	1.2	1.2011	0.0039
a_2	-0.74	-0.7404	0.0035
a_3	0.14	0.1392	0.0030
a_4	-0.02	-0.0192	0.0019
b_0	1	-	-
b_1	-1.3334	-1.3338	0.0092
b_2	1.6667	1.6677	0.0144
b_3	-2.6665	-2.6677	0.0242
b_4	1.9666	1.9672	0.0207
p_1	0.4	0.3982	0.0137
p_2	-0.3	-0.2986	0.0028
p_3	0.2	0.2011	0.0021

The adaptive filter was implemented with the step size of the recursive part to be the minimum of $\mu = 0.01$ or the bound suggested by the conditions. The moving average part of the linear subsystem was normalized with forgetting factor 0.95. Similarly, adaptation for the nonlinear part was normalized with forgetting factor 0.99. The step sizes for the $g_{j,l}(k)$ coefficients were constant and equal to 10^{-4}. The system was initialized with $H_{init}(z) = 1/(1+0.0064z^{-4})$ and $y_{init}(k) = 0 \cdot u(k) + 0 \cdot u^2(k) + 0 \cdot u^3(k)$. The coefficient b_0 is set to a value 1 and is not changed throughout the simulation. This ensures the uniqueness of the solution. Coefficients of the unknown system, mean values of the adaptive filter coefficients, and their variances after convergence are shown in Table 1. Coefficients of the polynomial subsystem p_l were obtained by conversion of the coefficients w_l. Figure 3 depicts the evolution of the mean square estimation error of the adaptive filter using a semi-log graph. All experiments resulted in convergence

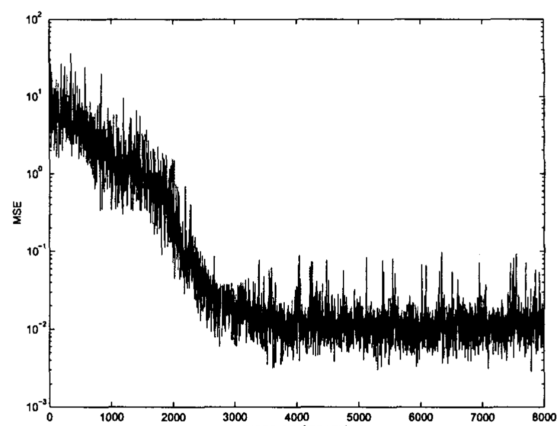

Fig. 3. Evolution of the mean-square estimation error.

to the global minimum. The evolution of the mean values of the

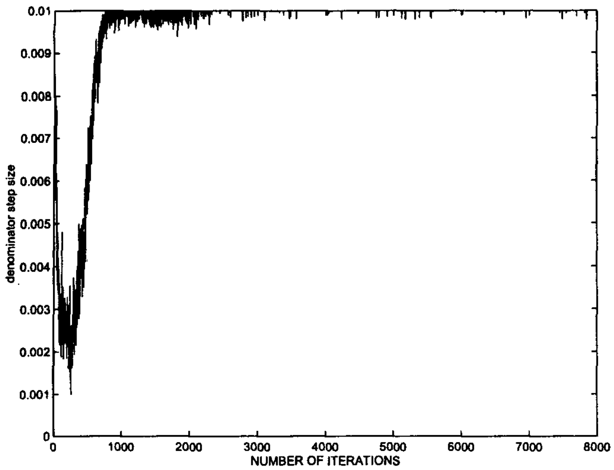

Fig. 4. Evolution of the mean values of the step size for the adaptation of the denominator coefficients.

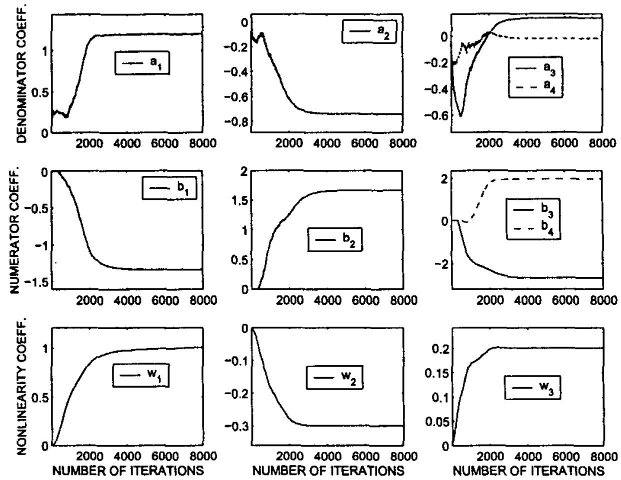

Fig. 5. Evolution of the mean values of the coefficients.

step size for the denominator coefficients is shown in Figure 4. The evolution of the coefficients of the linear subsystem, as well as the evolution of the coefficients of the nonlinear subsystem is shown in Figure 5.

The results indicate that step size selection using the condition in (14) 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. For the stability of the entire Hammerstein filter, one should be also concerned about the stability of the polynomial subsystem. This is ensured by constraining the step sizes associated with the nonlinear subsystem to be sufficiently small.

4. CONCLUDING REMARKS

This paper presented the derivation and preliminary 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. First, an algorithm that uses constant step sizes was derived. Our system employs stability bounds on the step sizes for adapting the coefficients of the recursive linear part. The resulting time-varying step sizes guarantee stable operation of the adaptive filter.

The error surface of the adaptive filter employing our model is not unimodal, and therefore our algorithm may converge to local minima of its error surface if we do not use white input signal. However, the algorithm derivation is based on the derivation of an adaptive linear recursive filter that is globally convergent for stationary and white input signals [5]. We believe that the adaptive filter of this paper is also globally convergent for the same set of conditions on the input signals. Simulation results so far have not contradicted this conjecture. More detailed analysis of this problem is necessary. This, as well as means of avoiding local minima for colored input signals are currently under investigation.

5. REFERENCES

[1] J. G. Proakis, *Digital Communications,* McGraw-Hill, New York, second edition, 1989.

[2] 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.

[3] J. Dubow, J. Jeraj, E. Kholmovski, N. Arnold, H. Meuzelaar, and V. J. Mathews, "A dynamic system model for ion mobility spectrometry," in *Proc. Of First Joint Conference on Point Detectors for Chemical and Bilogical Defence*, Williamsburg, Virginia, October 2000.

[4] A. Carini, V. J. Mathews, and G. L. Sicuranza, "Sufficient stability bounds for slowly-varying discrete-time recursive linear filters and their applications in adaptive IIR filtering," *IEEE Trans. Signal Proc*, vol. 47, no. 9, pp. 2561–7, September 1999.

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

[6] S. Haykin, *Adaptive Filter Theory,* Prentice Hall, Englewood Cliffs, N. J., third edition, 1996.

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

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