Signal Processing 90 (2010) 2020–2030

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

An affine projection-based algorithm for identification of nonlinear Hammerstein systems ☆

Ifiok J. Umoh, Tokunbo Ogunfunmi *

Department of Electrical Engineering, Santa Clara University, Santa Clara, CA 95053, USA

A R T I C L E I N F O

Article history:
Received 26 March 2009
Received in revised form
20 November 2009
Accepted 4 January 2010
Available online 11 January 2010

Keywords:
Adaptive signal processing
System identification
Parametric methods
Nonlinear systems
Hammerstein systems
Wiener systems
Volterra system

A B S T R A C T

There are parametric and non-parametric methods for adaptive Hammerstein system identification. The most commonly used method is the non-parametric. In reality, the linear subsystem of a Hammerstein system is not of finite impulse response and non-parametric adaptive algorithms require large matrices and therefore increase computational complexity.

The objectives of this paper are to identify the Hammerstein system adaptively based on the affine projection criterion using a parametric algorithm. We also develop a bound for control of step size of the proposed algorithm and derive an expression for its mean square error performance. The error surface of the nonlinear Hammerstein filter was determined by examining the non-quadratic nature and the global and local minima of the mean square error cost function. A bound was determined for the adaptive step size and an expression was derived for the mean square error convergence based on energy conservation theory. Simulations of system identification applications showed that convergence speed of the proposed algorithm was faster and the convergence was superior to previously existing Hammerstein algorithms. Applying the new algorithm to the identification of the human muscles stretch reflex dynamics showed good convergence results. The proposed algorithm is of practical value in real life situations.

1. Introduction

Nonlinear Hammerstein system [1] is a special type of nonlinear filter in which a static memoryless nonlinear system is followed by a dynamic linear system as shown in Fig. 1. The Hammerstein system finds application in the modeling of signal processing problems such as the distortion in nonlinearly amplified digital communication signals, modeling the involuntary contraction of human muscles [1–3], modeling of the human heart in order to regulate the heart rate during treadmill exercises [4] and other applications can be found in [5–7]. Other nonlinear models such as Volterra and Wiener models and their applications can be found in [8,9]. Over the past decade, different methods of adaptive system identification for Hammerstein nonlinear model have been proposed. The adaptive Hammerstein system identification methods can be roughly divided into two categories: non-parametric [10–12] and parametric [13,14]. Most existing Hammerstein identification methods are not adaptive [10,15,16] and almost no results on mean square performance of adaptive Hammerstein algorithms are available except for [17].

In the non-parametric method [11], the nonlinear function $f(\cdot)$ is approximated by a polynomial and shown to converge in the mean square sense in a finite interval as the sample size n increases. However, for each n the whole input sequence has to be redefined with distribution different from that of $n-1$, and hence the approximating polynomial has to be reconstructed at each step without recursion. In [10], a frequency domain identification

☆ This work was supported by the Packard Research Fellowship at Santa Clara University.

* Corresponding author.
E-mail address: togunfunmi@scu.edu (T. Ogunfunmi).

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.01.004

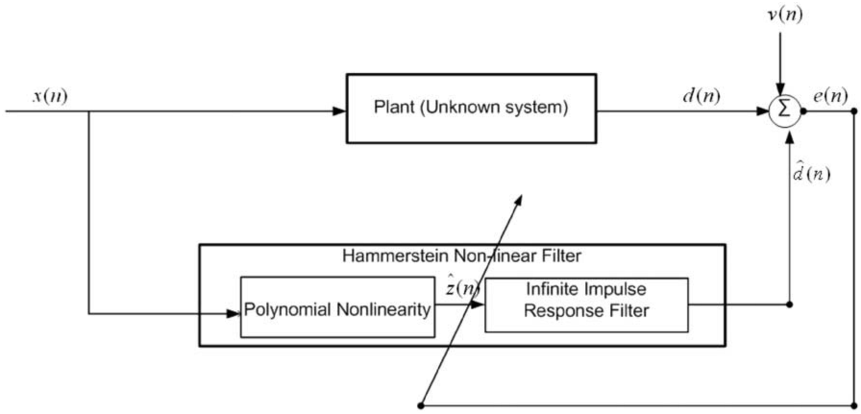

Fig. 1. Adaptive system identification of a Hammerstein system model.

was proposed. The author applied a sinusoidal input to a continuous-time Hammerstein system and the nonlinear function was expanded to a Fourier series. As a result, the nonlinearity identification reduced to estimating coefficients in the Fourier expansion. The estimates are non-recursive. In the most recent paper, [12] proposes an adaptive kernel canonical correlation analysis algorithm. The algorithm is based on a nonlinear transformation of the data from the input space to a higher-dimensional space where identification problem can be solved in a linear manner. However, this approach requires a careful choice of the constrains imposed on the identification problem to ensure convergence and is prone to overfitting especially when the dynamic linear subsystem is an infinite impulse response filter. In this paper, the system identification method proposed is parametric [18,14,13] and therefore has the advantage of lower computational complexity in comparison to non-parametric methods.

In [13], the memoryless nonlinearity is assumed to be a polynomial and the linear subsystem an infinite impulse response (IIR) filter. The authors used a Gram–Schmidt processor to linearize the polynomial nonlinearity producing a set of orthogonal linear subsystems (or a set of finite impulse response (FIR) filters with orthogonal coefficients). Applying the results in [19], the step size was constrained in such a way as to guarantee bounded-input bounded-output (BIBO) stability of the overall system. In [14], the identification of the linear and nonlinear subsystems of the Hammerstein model are done separately. The linear subsystem was identified by over sampling the output of the Hammerstein model and applying the least mean square (LMS) algorithm to minimize their proposed cost function. Two approaches for the identification of the nonlinear block were proposed, the direct approach which is based on the parametric method of adaptive system identification and the indirect approach that uses the Bezout identity in developing a cost function which is minimized using the LMS algorithm. An analysis of the performance of this algorithm was presented. The authors in [13] extended their work in [17] to present the mean square performance analysis of their proposed algorithm. The analysis was based on those done in [20,21].

The purpose of this paper is to:

(1) Identify the Hammerstein system adaptively based on the affine projection criterion, without linearizing of the Hammerstein nonlinearity and without the knowledge of the input to the linear subsystem.
(2) Develop a bound for the control of the step size of the proposed adaptive Hammerstein algorithm to achieve bounded-input bounded-output (BIBO) stability.
(3) Derive an expression for the mean-square error performance of the proposed algorithm using energy conservation arguments.

The assumption is that the unknown nonlinearity of the plant can be approximated as a finite ordered polynomial. The linear subsystem is represented as an IIR filter.

The rest of the paper is arranged in sections as follows. In Section 2, the problem to be addressed by this work is stated, in Section 3 the proposed algorithm is described. Section 4 describes the mean square error surface of the adaptive Hammerstein algorithm and in Section 5 a bound on the step size for the algorithm was determined. The mean square performance analysis and real life examples of the proposed algorithm are presented in Sections 6 and 7, respectively. Concluding remarks are given in Section 8.

2. Problem statement

Consider the Hammerstein model shown in Fig. 1, where $x(n)$, $v(n)$ and $d(n)$ are the system's input, noise and output, respectively. $\hat{z}(n)$ represents the unavailable internal signal of the adaptive Hammerstein filter model. Since the output of the memoryless nonlinear subsystem $\hat{z}(n)$ of the Hammerstein filter is unavailable, it is not possible to estimate the nonlinear subsystem without assuming an approximate nonlinear model. In this paper, a polynomial nonlinearity of order L is given by

$$\hat{z}(n) = \sum_{l=1}^{L} \hat{p}_l(n) x^l(n) \tag{1}$$

While the dynamic linear subsystem is modeled as an infinite-impulse response (IIR) filter satisfying a linear difference equation of the form

$$\hat{d}(n) = -\sum_{i=1}^{N} \hat{a}_i(n)\hat{d}(n-i) + \sum_{j=0}^{M} \hat{b}_j(n)\hat{z}(n-j) \quad (2)$$

where $\hat{p}_l(n)$, $\hat{a}_i(n)$ and $\hat{b}_j(n)$ represent the coefficient estimates of the nonlinear Hammerstein subsystem at any given time n. To ensure uniqueness of the parameterization, we normalize the dynamic linear subsystem by setting $\hat{b}_0(n)=1$. With the dynamic linear subsystem normalized, Eq. (2) can be written as

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

Eq. (3) can be rewritten in compact form as

$$\hat{d}(n) = \hat{\mathbf{s}}(n)^H \hat{\boldsymbol{\theta}}(n) \quad (4)$$

where

$$\hat{\boldsymbol{\theta}}(n) = [\hat{a}_1(n) \ \dots \ \hat{a}_N(n) \ \hat{b}_1(n) \ \dots \ \hat{b}_M(n) \ \hat{p}_1(n) \ \dots \ \hat{p}_L(n)]^H$$

$$\hat{\mathbf{s}}(n) = [-\hat{d}(n-1) \ \dots \ -\hat{d}(n-N) \ \hat{z}(n-1) \ \dots \ \hat{z}(n-M) \ x(n) \ \dots \ x^L(n)]^H$$

The goal of the proposed adaptive Hammerstein system identification algorithm, is to update the coefficient vector $\hat{\boldsymbol{\theta}}(n)$ in (4) of the nonlinear Hammerstein filter based only on the input signal $x(n)$ and output signal $d(n)$ such that $\hat{d}(n)$ is close to the desired response signal $d(n)$.

3. Adaptive Hammerstein algorithm

In this section, an adaptive Hammerstein system identification algorithm based on the theory of Affine projection [22] for estimating the parameters of the nonlinear Hammerstein subsystem represented by (3) given the input signal $x(n)$ and output signal $d(n)$ is derived. A criterion is defined for the minimization of the square Euclidean norm of the change in the weight vector is

$$\tilde{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n) - \hat{\boldsymbol{\theta}}(n-1) \quad (5)$$

subject to the set of Q constraints

$$d(n-q) = \hat{\mathbf{s}}(n-q)^H \hat{\boldsymbol{\theta}}(n), \quad q = 1, \dots, Q \quad (6)$$

Applying the method of Lagrange multipliers with multiple constraints to (5) and (6), the criterion for the affine projection filter is written as

$$J(n-1) = \|\hat{\boldsymbol{\theta}}(n) - \hat{\boldsymbol{\theta}}(n-1)\|^2 + \mathrm{Re}[(\mathbf{d}(n-1) - \hat{S}(n-1)^H \hat{\boldsymbol{\theta}}(n))^H \boldsymbol{\lambda}] \quad (7)$$

where

$$\mathbf{d}(n-1) = [d(n-1) \ \dots \ d(n-Q)]^H$$

$$\hat{S}(n-1) = [\hat{\mathbf{s}}(n-1) \ \dots \ \hat{\mathbf{s}}(n-Q)]$$

$$\boldsymbol{\lambda} = [\lambda_1 \ \dots \ \lambda_Q]^H$$

Minimizing the cost function (7) with respect to the nonlinear Hammerstein filter weight vector $\hat{\boldsymbol{\theta}}(n)$ gives

$$\frac{\partial J(n-1)}{\partial \hat{\boldsymbol{\theta}}(n)} = 2(\hat{\boldsymbol{\theta}}(n) - \hat{\boldsymbol{\theta}}(n-1)) - \frac{\partial (\mathbf{d}(n-1) - \hat{S}(n-1)^H \hat{\boldsymbol{\theta}}(n))^H \boldsymbol{\lambda}}{\partial \hat{\boldsymbol{\theta}}(n)}$$

$$\frac{\partial J(n-1)}{\partial \hat{\boldsymbol{\theta}}(n)} = 2(\hat{\boldsymbol{\theta}}(n) - \hat{\boldsymbol{\theta}}(n-1)) - \frac{\partial (\hat{\boldsymbol{\theta}}(n)^H \hat{S}(n-1))\boldsymbol{\lambda}}{\partial \hat{\boldsymbol{\theta}}(n)} \quad (8)$$

where

$$\frac{\partial (\hat{\boldsymbol{\theta}}(n)^H \hat{S}(n-1))}{\partial \hat{\boldsymbol{\theta}}(n)} = \left[\frac{\partial \hat{\boldsymbol{\theta}}(n)^H \hat{\mathbf{s}}(n-1)}{\partial \hat{\boldsymbol{\theta}}(n)} \ \cdots \ \frac{\partial \hat{\boldsymbol{\theta}}(n)^H \hat{\mathbf{s}}(n-Q)}{\partial \hat{\boldsymbol{\theta}}(n)} \right] \quad (9)$$

Since a portion of the vectors $\hat{s}(n)$ in $\hat{S}(n)$ include past $\hat{d}(n)$ which are dependent on past $\hat{\theta}(n)$ used to form the new $\hat{\theta}(n)$, the partial derivative of each element in (8) gives

$$\frac{\partial \hat{\boldsymbol{\theta}}(n)^H \hat{\mathbf{s}}(n-q)}{\partial \hat{a}_i(n)} = -\hat{d}(n-q-i) - \sum_{k=1}^{N} \hat{a}_k(n) \frac{\partial \hat{d}(n-q-k)}{\partial \hat{a}_i(n)}, \quad 1 \le i \le N \quad (10)$$

$$\frac{\partial \hat{\boldsymbol{\theta}}(n)^H \hat{\mathbf{s}}(n-q)}{\partial \hat{b}_j(n)} = \hat{z}(n-q-j) - \sum_{k=1}^{N} \hat{a}_k(n) \frac{\partial \hat{d}(n-q-k)}{\partial \hat{b}_j(n)}, \quad 1 \le j \le M \quad (11)$$

$$\frac{\partial \hat{\boldsymbol{\theta}}(n)^H \hat{\mathbf{s}}(n-q)}{\partial \hat{p}_l(n)} = x^l(n-q) + \sum_{k=1}^{M} \hat{b}_k(n) \frac{\partial \hat{z}(n-q-k)}{\partial \hat{p}_l(n)} - \sum_{k=1}^{N} \hat{a}_k(n) \frac{\partial \hat{d}(n-q-k)}{\partial \hat{p}_l(n)}, \quad 1 \le l \le L \quad (12)$$

A simplifying assumption commonly made for adaptive IIR filtering was applied to (10)–(12). The assumption is that the adaptation step size μ is sufficiently small [22,23] such that

$$\hat{\boldsymbol{\theta}}(n) \cong \hat{\boldsymbol{\theta}}(n-1) \cong \cdots \cong \hat{\boldsymbol{\theta}}(n-N)$$

and therefore,

$$\hat{a}_i(n) \cong \hat{a}_i(n-1) \cong \cdots \cong \hat{a}_i(n-N)$$

$$\frac{\partial \hat{d}(n-q)}{\partial \hat{a}_i(n)} = -\hat{d}(n-q-i) - \sum_{k=1}^{N} \hat{a}_k(n) \frac{\partial \hat{d}(n-q-k)}{\partial \hat{a}_i(n-k)} \quad (13)$$

$$\hat{b}_j(n) \cong \hat{b}_j(n-1) \cong \cdots \cong \hat{b}_j(n-N)$$

$$\frac{\partial \hat{d}(n-q)}{\partial \hat{b}_j(n)} = \hat{z}(n-q-j) - \sum_{k=1}^{N} \hat{a}_k(n) \frac{\partial \hat{d}(n-q-k)}{\partial \hat{b}_j(n-k)} \quad (14)$$

$$\hat{p}_l(n) \cong \hat{p}_l(n-1) \cong \cdots \cong \hat{p}_l(n-N)$$

$$\frac{\partial \hat{d}(n-q)}{\partial \hat{p}_l(n)} = x^l(n-q) + \sum_{k=1}^{M} \hat{b}_k(n) \frac{\partial \hat{z}(n-q-k)}{\partial \hat{p}_l(n-k)} - \sum_{k=1}^{N} \hat{a}_k(n) \frac{\partial \hat{d}(n-q-k)}{\partial \hat{p}_l(n-k)} \quad (15)$$

$$\frac{\partial \hat{p}_l(n-q-k)}{\partial \hat{p}_l(n-k)} = 1$$

therefore,

$$\frac{\partial \hat{d}(n-q)}{\partial \hat{p}_l(n)} = x^l(n-q) + \sum_{k=1}^{M} \hat{b}_k(n) x^l(n-q-k) - \sum_{k=1}^{N} \hat{a}_k(n) \frac{\partial \hat{d}(n-q-k)}{\partial \hat{p}_l(n-k)} \quad (16)$$

Let

$$\hat{\boldsymbol{\varphi}}(n-q) = \frac{\partial \hat{d}(n-q)}{\partial \hat{\boldsymbol{\theta}}(n)} = \left[\frac{\partial \hat{d}(n-q)}{\partial \hat{a}_1(n)} \ \ldots \ \frac{\partial \hat{d}(n-q)}{\partial \hat{a}_N(n)} \ \frac{\partial \hat{d}(n-q)}{\partial \hat{b}_1(n)} \ \ldots \ \frac{\partial \hat{d}(n-q)}{\partial \hat{b}_M(n)} \ \frac{\partial \hat{d}(n-q)}{\partial \hat{p}_1(n)} \ \ldots \ \frac{\partial \hat{d}(n-q)}{\partial \hat{p}_L(n)}\right]^H$$

$$\hat{\Phi}(n-1) = \frac{\partial(\hat{\boldsymbol{\theta}}(n)^H \hat{S}(n-1))}{\partial \hat{\boldsymbol{\theta}}(n)} = [\hat{\boldsymbol{\varphi}}(n-1) \ \ldots \ \hat{\boldsymbol{\varphi}}(n-Q)]$$

$$\hat{\boldsymbol{\psi}}(n-q) = \left[-\hat{d}(n-q-1) \ \ldots \ -\hat{d}(n-q-N) \ \hat{z}(n-q-1) \ \ldots \ \hat{z}(n-q-M) \ \sum_{j=0}^{M} x(n-q-j) \ \ldots \ \sum_{j=0}^{M} x^L(n-q-j)\right]^H$$

$$\hat{\Psi}(n-1) = [\hat{\boldsymbol{\psi}}(n-1) \ \ldots \ \hat{\boldsymbol{\psi}}(n-Q)]$$

Substituting (14)–(16) into (9), gives

$$\hat{\Phi}(n-1) = \hat{\Psi}(n-1) - \sum_{k=1}^{N} \hat{a}_k(n-1) \hat{\Phi}(n-1-k) \quad (17)$$

Thus, rewriting (8)

$$\frac{\partial J(n-1)}{\partial \hat{\boldsymbol{\theta}}(n)} = 2(\hat{\boldsymbol{\theta}}(n) - \hat{\boldsymbol{\theta}}(n-1)) - \hat{\Phi}(n-1)\boldsymbol{\lambda} \quad (18)$$

Setting the partial derivative of the cost function in (18) to zero, gives

$$\tilde{\boldsymbol{\theta}}(n) = \tfrac{1}{2}\hat{\Phi}(n-1)\boldsymbol{\lambda} \quad (19)$$

From (4), the following is obtained:

$$\mathbf{d}(n-1) = \hat{S}(n-1)^H \hat{\boldsymbol{\theta}}(n) \quad (20)$$

where

$$\mathbf{d}(n-1) = [d(n-1) \ \ldots \ d(n-Q)]^H$$

$$\begin{aligned}\mathbf{d}(n-1) &= \hat{S}(n-1)^H \hat{\boldsymbol{\theta}}(n-1) + \hat{S}(n-1)^H \tilde{\boldsymbol{\theta}}(n) \\ &= \hat{S}(n-1)^H \hat{\boldsymbol{\theta}}(n-1) + \tfrac{1}{2}\hat{S}(n-1)^H \hat{\Phi}(n-1)\boldsymbol{\lambda}\end{aligned} \quad (21)$$

$$\mathbf{e}(n-1) = \mathbf{d}(n-1) - \hat{S}(n-1)^H \hat{\boldsymbol{\theta}}(n-1) \quad (22)$$

where

$$\mathbf{e}(n-1) = [e(n-1) \ \ldots \ e(n-Q)]^H$$

and

$$e(n) = d(n) - \hat{\mathbf{s}}(n)^H \hat{\boldsymbol{\theta}}(n)$$

Evaluating (21) and (22) for λ results in

$$\boldsymbol{\lambda} = 2(\hat{S}(n-1)^H \hat{\Phi}(n-1))^{-1} \mathbf{e}(n-1) \quad (23)$$

Substituting (23) into (19) yields the optimum change in the weight vector

$$\tilde{\boldsymbol{\theta}}(n) = \hat{\Phi}(n-1)(\hat{S}(n-1)^H \hat{\Phi}(n-1))^{-1} \mathbf{e}(n-1) \quad (24)$$

normalizing (24) as in [20,21] and regularizing $\hat{S}(n-1)^H \hat{\Phi}(n-1)$ matrix to guard against numerical difficulties during inversion yields

$$\tilde{\boldsymbol{\theta}}(n) = \mu \hat{\Phi}(n-1)(\delta I + \mu \hat{S}(n-1)^H \hat{\Phi}(n-1))^{-1} \mathbf{e}(n-1) \quad (25)$$

$$\hat{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n-1) - \mu \hat{\Phi}(n-1)(\delta I + \mu \hat{S}(n-1)^H \hat{\Phi}(n-1))^{-1} \mathbf{e}(n-1) \quad (26)$$

To improve the update process Newton's method is applied by scaling the update vector by $R^{-1}(n)$. The matrix $R(n)$ is an estimate of the Hessian matrix updated according to

$$R(n) = \lambda_n R(n-1) + (1-\lambda_n) \hat{\Phi}(n-1) \hat{\Phi}(n-1)^H \quad (27)$$

where λ_n is the forgetting factor and typically has values between 0 and 1. Applying the matrix inversion lemma on (27) gives

$$R(n)^{-1} = \frac{1}{\lambda_n}\left[R(n-1)^{-1} - R(n-1)^{-1}\hat{\Phi}(n-1) \left(\frac{\lambda_n}{1-\lambda_n} I - \hat{\Phi}(n-1)^H R(n-1)^{-1} \hat{\Phi}(n-1)\right)^{-1} \hat{\Phi}(n-1)^H R(n-1)^{-1}\right] \quad (28)$$

Applying (28) to (26), the new update equation is given by

$$\hat{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n-1) - \mu R(n-1)^{-1} \hat{\Phi}(n-1)(\delta I + \mu \hat{S}(n-1)^H \hat{\Phi}(n-1))^{-1} \mathbf{e}(n-1) \quad (29)$$

A summary of the proposed algorithm is shown in Algorithm 1. In the algorithm, N represents the number of feedback coefficients, M the number of feedforward coefficients for the linear subsystem and L the number of coefficients for the polynomial subsystem. Let K represent $N+M+L-2$ in the computation of the computational cost of our proposed adaptive nonlinear algorithm. We assume that the cost of inverting a $K \times K$ matrix is $\bigcirc(K^3)$ (multiplications and additions) and $\bigcirc(L^2N)$ for computing R^{-1}. Under these assumptions, the computational cost of our proposed algorithm is of $\bigcirc(QK^2)$ multiplications compared to $\bigcirc(K^2)$ in [13]. This increase in complexity due to the order of the input regression matrix in the proposed algorithm is compensated for by the algorithms' good performance.

Algorithm 1. Summary of the proposed variable stepsize Hammerstein adaptive algorithm.

DEFINITION: $\hat{\mathbf{p}}(n) = [\hat{p}_1(n) \ \ldots \ \hat{p}_L(n)]^H$
INITIALIZE:
$R^{-1}(0) = I, \lambda_n \neq 0, 0 < \mu \leq 1, \delta \ll 1, \hat{S}(0) = \text{zeros}(M+N+L-1, Q)$
$\hat{\theta}(0) = [\underbrace{0 \ \ldots \ 0}_{\hat{a}(n)}\underbrace{0 \ \ldots \ 0}_{\hat{b}(n)}\underbrace{1 \ \ldots \ 1}_{\hat{p}(n)}]^H, \hat{b}(0) = 1$
for $n = 0$ to sample size **do**
$\mathbf{e}(n-1) = \mathbf{d}(n-1) - \hat{S}(n-1)^H \hat{\theta}(n-1)$
$\hat{\Phi}(n-1) = \hat{\Psi}(n-1) - \sum_{k=1}^{N} a_k(n-1) \hat{\Phi}(n-1-k)$
$R(n)^{-1} = \frac{1}{\lambda_n}\Big[R(n-1)^{-1} - R(n-1)^{-1}\hat{\Phi}(n-1) \left(\frac{\lambda_n}{1-\lambda_n} I - \hat{\Phi}(n-1)^H R(n-1)^{-1} \hat{\Phi}(n-1)\right)^{-1}$
$\hat{\Phi}(n-1)^H R(n-1)^{-1}\Big]$

$\hat{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n-1) - \mu R(n)^{-1}\hat{\Phi}(n-1)(\delta I + \mu \hat{S}(n-1)^H \hat{\Phi}(n-1))^{-1}\mathbf{e}(n-1)$
$\hat{z}(n) = \mathbf{x}(n)^H \hat{\mathbf{p}}(n)$
$\hat{d}(n) = \hat{\mathbf{s}}(n)^H \hat{\boldsymbol{\theta}}(n)$
end for

4. Mean square error surface

In the previous section, the update equation for an adaptive affine projection Hammerstein nonlinear filter algorithm was derived. This section provides an insight into the nature of the error surface of the nonlinear Hammerstein filter by examining the non-quadratic nature as well as the global and local minima of $E[e^2(n)]$. From (2), the transfer function of the linear dynamic subsystem of the nonlinear Hammerstein filter shown in Fig. 1 can be rewritten as

$$\hat{H}(z) = \frac{\hat{B}(z)}{\hat{A}(z)} = \frac{\hat{b}_0 + \hat{b}_1 z^{-1} + \cdots + \hat{b}_M z^{-M}}{\hat{a}_0 + \hat{a}_1 z^{-1} + \cdots + \hat{a}_N z^{-N}} \tag{30}$$

Let the dynamic linear subsystem transfer function of the plant in Fig. 1 be represented by

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}}{a_0 + a_1 z^{-1} + \cdots + a_N z^{-N}} \tag{31}$$

and the unknown nonlinear memoryless polynomial subsystem output as

$$z(n) = \sum_{l=1}^{L} p_l(n) x^l(n) \tag{32}$$

From (22), the mean square error (MSE) is given by

$$E[e^2(n)] = E\left[\left(\frac{B(z)}{A(z)}\left(\sum_{l=1}^{L} p_l(n)x^l(n)\right) - \frac{\hat{B}(z)}{\hat{A}(z)}\left(\sum_{l=1}^{L} \hat{p}_l(n)x^l(n)\right)\right)^2\right] \tag{33}$$

Taking the derivative of (33) with respect to each coefficient $\hat{a}_i(n), \hat{b}_j(n)$ and $\hat{p}_l(n)$ and setting it to zero, the stationary points of the MSE are

$$E\left[\left(\frac{B(z)}{A(z)}\left(\sum_{l=1}^{L} p_l(n)x^l(n)\right) - \frac{\hat{B}(z)}{\hat{A}(z)}\left(\sum_{l=1}^{L} \hat{p}_l(n)x^l(n)\right)\right)\frac{\hat{B}(z)}{\hat{A}^2(z)}\left(\sum_{l=1}^{L} \hat{p}_l(n)x^l(n-i)\right)\right] = 0, \quad 1 \le i \le N \tag{34}$$

$$E\left[\left(\frac{B(z)}{A(z)}\left(\sum_{l=1}^{L} p_l(n)x^l(n)\right) - \frac{\hat{B}(z)}{\hat{A}(z)}\left(\sum_{l=1}^{L} \hat{p}_l(n)x^l(n)\right)\right)\frac{1}{\hat{A}(z)}\left(\sum_{l=1}^{L} \hat{p}_l(n)x^l(n-j)\right)\right] = 0, \quad 1 \le j \le M \tag{35}$$

$$E\left[\left(\frac{B(z)}{A(z)}\left(\sum_{l=1}^{L} p_l(n)x^l(n)\right) - \frac{\hat{B}(z)}{\hat{A}(z)}\left(\sum_{l=1}^{L} \hat{p}_l(n)x^l(n)\right)\right)\left(\frac{1}{\hat{A}(z)}x^l(n) + \frac{\hat{B}(z)}{\hat{A}(z)}x^l(n)\right)\right] = 0, \quad 1 \le l \le L \tag{36}$$

It can be seen that (35) and (36) are linear with respect to b's and p's, respectively, and define a single global minimum of $E[e^2(n)]$. With respect to a's, (34) is nonlinear and therefore produces an error surface that is non-quadratic. As a result of the nonlinearity in the error surface, the mean square error surface has both a local and global minimum. It can be shown from (33), [24,25] that no local minima exists for the system identification provided the filter is normalized such that $b_0 = 1$, $a_0 = 1$, the filter is sufficiently ordered or the order of the adaptive filter numerator exceeds that of the unknown filter denominator and the input $x(n)$ is a white noise signal.

5. Step-size

In Section 3, a simplifying assumption was used in the development of the proposed algorithm. The assumption made was that the adaptation step size μ is chosen to be sufficiently small such that $\hat{\boldsymbol{\theta}}(n) \cong \hat{\boldsymbol{\theta}}(n-1) \cong \cdots \cong \hat{\boldsymbol{\theta}}(n-N)$ is true. In this section, a bound is found on the adaptive step size μ such that assumption of a sufficiently small stepsize is satisfied while guaranteeing the stability of the Hammerstein system. For the purpose of this derivation, the Hammerstein system in (3) can be expressed in its state-space form as

$$\underbrace{\begin{bmatrix} u_1(n+1) \\ u_2(n+1) \\ \vdots \\ u_N(n+1) \\ w(n) \end{bmatrix}}_{\mathbf{u}(n+1)} = \underbrace{\begin{bmatrix} -\hat{a}_1(n) & -\hat{a}_2(n) & \cdots & -\hat{a}_N(n) & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}}_{Y(n)} \underbrace{\begin{bmatrix} u_1(n) \\ u_2(n) \\ \vdots \\ u_N(n) \\ \hat{z}(n) \end{bmatrix}}_{\mathbf{u}(n)} \tag{37}$$

$$\hat{d}(n) = [\hat{b}_0(n) \;\; \hat{b}_1(n) \;\; \ldots \;\; \hat{b}_N(n)] \begin{bmatrix} u_1(n+1) \\ u_2(n+1) \\ \vdots \\ u_N(n+1) \\ w(n) \end{bmatrix} \tag{38}$$

If the assumption $\hat{\boldsymbol{\theta}}(n) \cong \hat{\boldsymbol{\theta}}(n-1) \cong \cdots \cong \hat{\boldsymbol{\theta}}(n-N)$ holds, then in particular

$$\|Y(n+1) - Y(n)\| \le \varepsilon \tag{39}$$

Therefore, for sufficiently small ε,

$$Y(n) \cong Y(n+1) \cong \cdots \cong Y(n+K-1)$$

From (37),

$$\mathbf{u}(n+K) = Y(n+K-1)\ldots Y(n+1)Y(n)\mathbf{u}(n) \tag{40}$$

which behaves approximately like the system,

$$\overline{\mathbf{u}}(n+K) = Y^K(n)\overline{\mathbf{u}}(n) \tag{41}$$

given the initialization $\mathbf{u}(n) = \overline{\mathbf{u}}(n)$. Recall the theorem proven in [26],

Theorem 1. *The linear state equation*

$$\mathbf{u}(n+1) = Y(n)\mathbf{u}(n), \quad \mathbf{u}(n_0) = \mathbf{u}_0 \tag{42}$$

is uniformly exponentially stable if and only if there exists a positive definite $(N+1)\times(N+1)$ *matrix* $D(n)$ *for which the residue*

$$Y^H(n)D(n)Y(n) - D(n) \tag{43}$$

is positive definite.

Since the poles of the (3) are by hypothesis always inside the unit circle, the unique, symmetric and positive definite solution of the discrete time Lyapunov equation can be used as the Lyapunov candidate

$$Y^H(n)D(n)Y(n) - D(n) = -I_{N+1} \tag{44}$$

where I_{N+1} is an $N+1\times N+1$ identity matrix. Solving (44) for $D(n)$ we have

$$vec[D(n)] = -[Y^H(n)\otimes Y^H(n) - I_{(N+1)^2}]^{-1} vec[I_{N+1}] \tag{45}$$

where $vec[D(n)]$ represents a Kronecker vector formed by stacking all the columns of $D(n)$ and $\otimes$ denotes the Kronecker product [27]

$$\overline{\mathbf{u}}^H(n)((Y^K)^H(n)D(n)(Y^K)(n) - D(n))\overline{\mathbf{u}}(n) = -\overline{\mathbf{u}}^H(n)I_{N+1}\overline{\mathbf{u}}(n) \tag{46}$$

which can be rewritten as

$$\|\overline{\mathbf{u}}(n+K)\|^2_{D(n)} - \|\overline{\mathbf{u}}(n)\|^2_{D(n)} = -\|\overline{\mathbf{u}}(n)\|^2_{I_{N+1}} \geq \varpi\|\overline{\mathbf{u}}(n)\|^2_{D(n)} \tag{47}$$

for some positive constant ϖ. Thus,

$$\|\overline{\mathbf{u}}(n+K)\|^2_{D(n)} \geq (1-\varpi)\|\overline{\mathbf{u}}(n)\|^2_{D(n)} \tag{48}$$

illustrating a strict decrease in the state vector norm. From (44),

$$\|\overline{\mathbf{u}}(n+K) - \mathbf{u}(n+K)\|^2_{D(n)} \leq \gamma \tag{49}$$

where γ is a sufficiently small constant as a result of ε being sufficiently small. Thus for a sufficiently small ε, the vector $\mathbf{u}(n+K)$ will lie within a γ- ball of $\overline{\mathbf{u}}(n+K)$ and provided the radius γ is less than the worst-case delay $\varpi\|\overline{\mathbf{u}}(n)\|^2_{D(n)}$, the inequality [23]

$$\|\overline{\mathbf{u}}(n+K)\|^2_{D(n)} \leq \|\mathbf{u}(n)\|^2_{D(n)} \tag{50}$$

should carry through.

In order to ensure the stability of the proposed adaptive nonlinear Hammerstein algorithm, an upper bound on the adaptation step-size μ is determined such that (50) is satisfied. From (49)

$$\overline{\mathbf{u}}(n+K) - \mathbf{u}(n+K) \simeq \nabla_{\hat\theta(n)}[\overline{\mathbf{u}}(n+K) - \mathbf{u}(n+K)]\Delta\hat\theta(n)$$
$$\simeq Y(n)\nabla_{\hat\theta(n-1)}[\overline{\mathbf{u}}(n+K) - \mathbf{u}(n+K)]\Delta\hat\theta \tag{51}$$

where $\nabla_{\hat\theta(n-1)}$ is the gradient operator with respect to the coefficient $\hat\theta(n)$ and $\Delta\hat\theta(n) = \hat\theta(n+1) - \hat\theta(n)$. From the update equation

$$\hat{\boldsymbol\theta}(n) - \hat{\boldsymbol\theta}(n-1) = \mu R(n-1)^{-1}\hat\Phi(n-1)(\hat S(n-1)^H\hat\Phi(n-1))^{-1}\mathbf{e}(n-1) \tag{52}$$

From (49) and (51) an explicit condition for the step size $\mu(n)$ for the stability of the proposed adaptive Hammerstein algorithm was obtained as

$$\mu \leq \frac{\gamma}{\|Y(n)\nabla_{\hat\theta(n-1)}[\overline{\mathbf{u}}(n+K) - \mathbf{u}(n+K)](R(n-1)^{-1}\hat\Phi(n-1)(\hat S(n-1)^H\hat\Phi(n-1))^{-1}\mathbf{e}(n-1))\|_{D(n)}} \tag{53}$$

6. Mean square performance

This section, presents an expression for the steady state mean square error based on energy conservation theory [28]. The expression derived is based on the assumption that with slow convergence and a stationary operating environment, the ensemble average of the excess squared error is close to its minimum possible value [22].

Introducing the *a priori* error vector $\mathbf{e}_a(n-1)$

$$\mathbf{e}_a(n-1) = \hat S(n-1)^H\hat{\boldsymbol\theta}(n-1) \tag{54}$$

Substituting $\overline{\boldsymbol\theta}(n) = \boldsymbol\theta - \hat{\boldsymbol\theta}(n)$ in (29), the equation is rewritten (29) as

$$\overline{\boldsymbol\theta}(n) = \overline{\boldsymbol\theta}(n-1) - \mu R(n-1)^{-1}\hat\Phi(n-1)(\delta I + \mu\hat S(n-1)^H\hat\Phi(n-1))^{-1}\mathbf{e}(n-1) \tag{55}$$

Multiplying both sides with $R^{1/2}(n-1)$ from the left and computing the square norm result in

$$\overline{\boldsymbol\theta}^H(n)R(n-1)\overline{\boldsymbol\theta}(n) = \|R^{1/2}(n-1)\overline{\boldsymbol\theta}(n-1) - \mu R(n-1)^{-1/2}\hat\Phi(n-1)(\delta I + \mu\hat S(n-1)^H \hat\Phi(n-1))^{-1}\mathbf{e}(n-1)\| \tag{56}$$

Expanding and applying the ensemble average gives

$$\begin{aligned} E[\overline{\boldsymbol\theta}^H(n)R(n-1)\overline{\boldsymbol\theta}(n)] &= E[\overline{\boldsymbol\theta}^H(n-1)R(n-1)\overline{\boldsymbol\theta}(n-1)] \\ &\quad - \mu E[\overline{\boldsymbol\theta}^H(n-1)\hat\Phi(n-1)(\delta I + \mu\hat S(n-1)^H\hat\Phi(n-1))^{-1}\mathbf{e}(n-1)] \\ &\quad - \mu E[\mathbf{e}^H(n-1)(\delta I + \mu\hat\Phi(n-1)^H\hat S(n-1))^{-1}\hat\Phi^H(n-1)\overline{\boldsymbol\theta}] \\ &\quad + E[\|\mu R(n-1)^{-1/2}\hat\Phi(n-1)(\delta I + \mu\hat S(n-1)^H \hat\Phi(n-1))^{-1}\mathbf{e}(n-1)\|] \end{aligned} \tag{57}$$

As $n\to\infty$, the algorithm approaches its steady state condition and $E[\overline{\boldsymbol\theta}^H(n)R(n-1)\overline{\boldsymbol\theta}(n)] = E[\overline{\boldsymbol\theta}^H(n-1)R(n-1)\overline{\boldsymbol\theta}(n-1)]$. Also, it is known that

$$\mathbf{e}(n-1) = \mathbf{e}_a(n-1) + \mathbf{v}(n-1) \tag{58}$$

The following assumptions of independence are made:

(1) the noise $v(n-1)$ is independently and identically distributed and statistically independent of the regression matrices $\hat{S}(n-1)$ and $\hat{\Phi}(n-1)$.
(2) At steady-state, $\hat{S}(n-1)$ is statistically independent of $\mathbf{e}_a(n-1)$ with $E[\mathbf{e}_a^H(n-1)\mathbf{e}_a(n-1)] = E|\mathbf{e}_a(n-1)|^2$ and $E[\mathbf{e}_a^H(n-1)] = 0$.

In the Appendix, expressions for the filters mean square error (MSE) and excess mean square error (EMSE) were derived as

$$EMSE = \frac{-\mu\sigma_v^2 Tr(E[C(n-1)])}{(Tr(\mu E[C(n-1)])-2\zeta(n)Tr(E[G(n-1)]))} \quad (59)$$

and

$$MSE = \frac{-\mu\sigma_v^2 Tr(E[C(n-1)])}{(Tr(\mu E[C(n-1)])-2\zeta(n)Tr(E[G(n-1)]))} + \sigma_v^2 \quad (60)$$

where

$$\zeta(n) = \mu - \sum_{f=1}^{n-2}(-1)^f \prod_{g=0}^{f-1} \sum_{k_{n-f+g}=1}^{N} a_{k_{n-f+g}}(n-f+g)$$

$$C(n-1) = (\delta I + \mu\hat{\Phi}^H(n-1)\hat{S}(n-1))^{-1}\hat{\Phi}^H(n-1)R^{-1}(n-1)\hat{\Phi}(n-1)(\delta I + \mu\hat{S}(n-1)^H\hat{\Phi}(n-1))^{-1}$$

$$G(n-1) = (\delta I + \mu\hat{S}(n-1)^H\hat{\Phi}(n-1))^{-1}$$

7. Results

In this section the validity of the proposed algorithm is demonstrated. Simulation results corresponding to white and colored types of signals are shown. Results are also presented for a real life Hammerstein system involving the identification of the human muscle stretch reflex dynamics.

The Hammerstein nonlinear subsystem to be identified has a linear subsystem of infinite impulse response given by

$$H_1(z) = \frac{1.0000-1.8000z^{-1}+1.6200z^{-2}-1.4580z^{-3}+0.6561z^{-4}}{1.0000-0.2314z^{-1}+0.4318z^{-2}-0.3404z^{-3}+0.5184z^{-4}} \quad (61)$$

and a nonlinear subsystem of polynomial nonlinearity given by

$$z(n) = x(n) - 0.3x(n)^2 + 0.2x(n)^3 \quad (62)$$

Since it is known that $b(0) = 1$, the update weight vector for all simulations were initialized to

$$\hat{\theta}(0) = [0 \ \ldots \ 0 \ 0 \ \ldots \ 0 \ 1 \ \ldots \ 1]^H$$

Results shown were obtained by ensemble averaging over 100 independent trials of the experiment.

7.1. System identification with white input

System identification simulations were performed to identify the nonlinear Hammerstein type plant with impulse response given in (61) and (62). The desired response signal $d(n)$ of the adaptive Hammerstein filter was obtained by corrupting the output of the unknown system with additive white noise signals of zero mean and variance such that the output signal to noise ratio was 30 dB. The input signal $x(n)$ of the adaptive filter was an additive white noise signal with zero mean and unit variance. The adaptive Hammerstein filter was

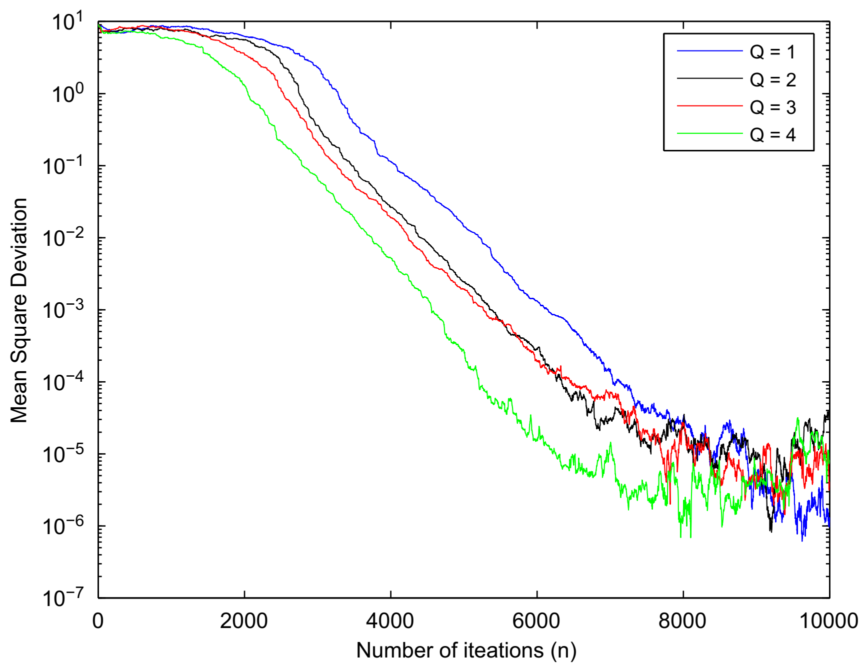

Fig. 2. Mean square deviation (MSD) learning curve of proposed algorithm for white input and $Q = 1, 2, 3, 4$.

simulated with a fixed step size initialized based on (53) to 1.674e−7. Other parameter settings were as follows λ_n the forgetting factor was set to 0.995, δ for the regularization matrix was set to 1e−4 and $\gamma = 1e{-3}$ based on the noise variance.

Fig. 2 shows the learning curve describing the mean square deviation of the adaptive Hammerstein filter weights from the optimum weight of the Hammerstein model plant. From this figure, it can be seen that the convergence speed of the proposed algorithm in this paper increased with an increase in the number of constraints Q. Fig. 3 shows that the mean square error learning curve for the case where the number of constraints Q was set to 3 compared very well with the

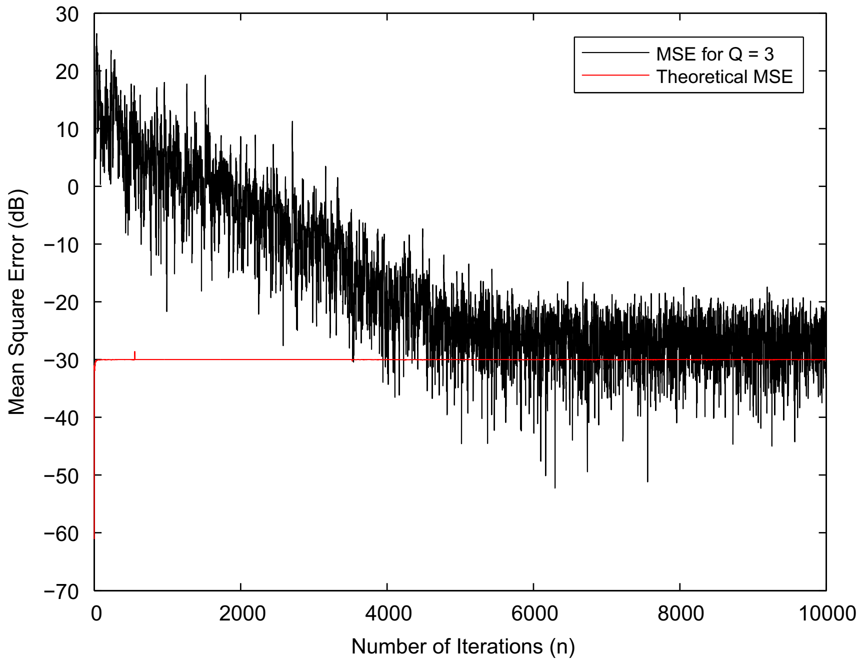

Fig. 3. Mean square error learning curve of proposed algorithm for white noise input, $Q = 3$ and theoretical MSE line.

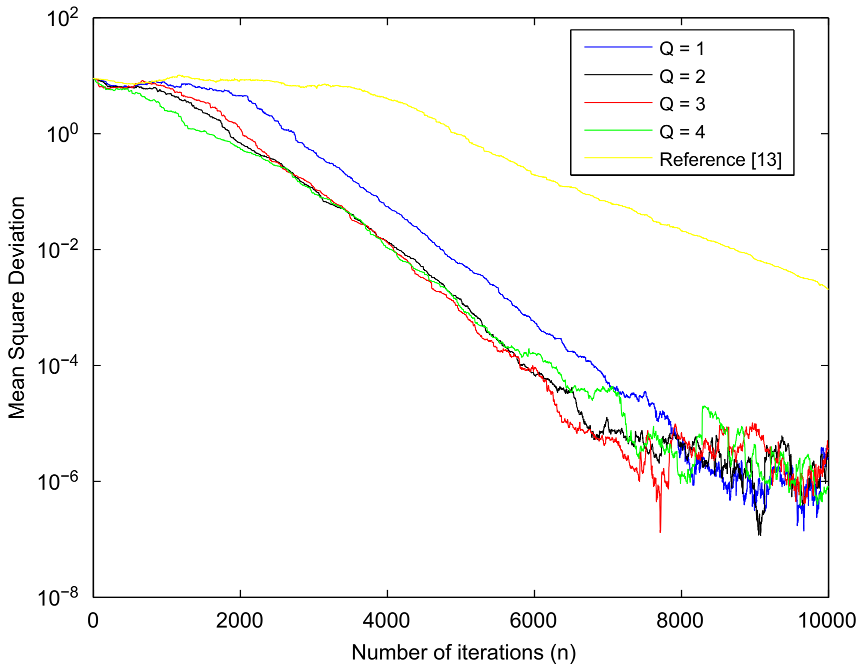

Fig. 4. Mean square deviation (MSD) learning curve of proposed algorithm for colored input and $Q = 1,2,3,4$ compared with results from [13].

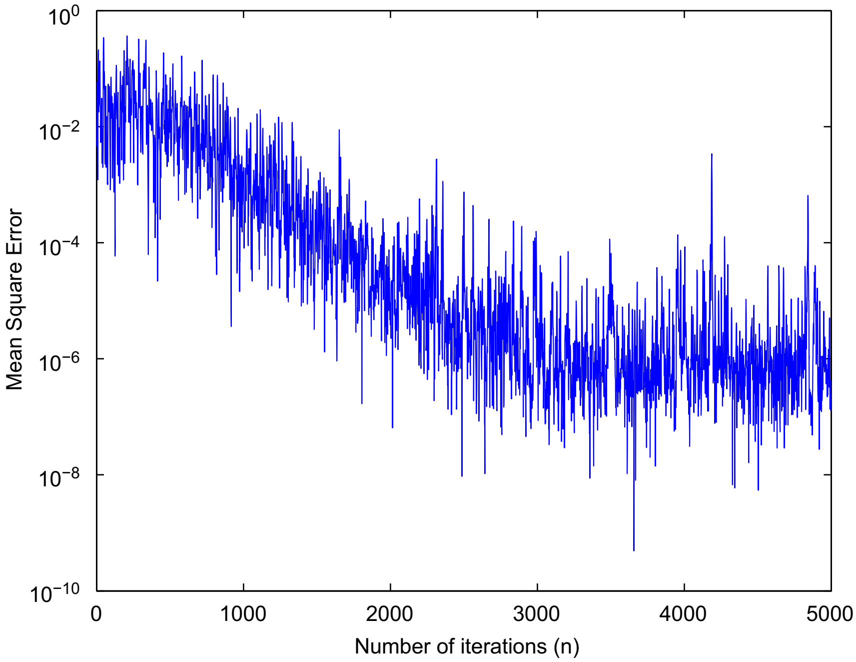

Fig. 5. Mean square error learning curve of our proposed algorithm used to identify the human muscle stretch reflex dynamics.

steady state expression in (60). This behavior is expected since (60) was based on the assumption of independence.

7.2. System identification with colored input

For a colored input signal $x(n)$ the adaptive filter was generated by filtering an additive white noise signal with zero mean and unit variance with the filter

$$H_2(z) = 1+0.5z^{-1} \tag{63}$$

All the parameter settings for the proposed algorithm and [13] were kept the same as described in Section 7.1. In the case of [13], δ_0 was set 10^{-5} and the initial value of $\hat{\sigma}^2_{v_j}(n)$ was set to 10. The results obtained using the reasonably colored signal as input are shown in Fig. 4. The proposed algorithm converged much faster than the algorithm proposed in [13] even though the authors had used the Gram–Schmidt process to better enhance the algorithms' performance in a colored input environment. Also, when the number of constraints was set above 1, the mean square convergence in the weight was independent of the number of constraints Q.

7.3. A practical example

It is interesting to apply the affine projection adaptive Hammerstein filter developed in this paper to a practical example which is the identification of the human muscles stretch reflex dynamics. The stretch reflex is the involuntary contraction of a muscle which results from perturbation of its lengths. In this subsection, real life data obtained from [1] was used. A description of how the input and output data was collected for system identification problem is given in [1]. In this experiment, the linear subsystem of the Hammerstein filter was chosen as an eighth order IIR filter and the nonlinearity was selected to be of the fifth order. The Hammerstein adaptive filter parameters were set as follows: $\lambda_n = 0.99$, $\delta = 1\text{e}{-3}$, $\gamma = 1\text{e}{-3}$ and $\mu = 1.674\text{e}{-6}$. The step-size μ was carefully chosen to satisfy the condition in (53) for stability. The mean square error learning curve for this experiment is shown in Fig. 5. This figure shows the convergence of the algorithm in the mean square error sense. The result also shows sensitivity of the proposed algorithm to variation in the weight parameter. This is due to the use of a polynomial nonlinearity to approximate the half-wave rectifier type nonlinearity of the stretch reflex dynamic shown in [1].

8. Conclusion

In this paper, an adaptive affine projection nonlinear Hammerstein algorithm for the identification of Hammerstein type nonlinear subsystems is proposed. The Hammerstein model considered was a cascade of a polynomial nonlinearity and an infinite impulse response filter. Due to the presence of an IIR filter, a recursive bound on the adaptation step-size was derived to achieve bounded-input bounded-output stability of the adaptive Hammerstein algorithm. Also, a theoretical expression for the convergence behavior of the mean square error was derived based on energy conservation theory. Results for colored inputs showed the superior convergence of the proposed algorithm when compared to an existing adaptive Hammerstein algorithm. We demonstrated good convergence by the proposed algorithm when used under real life, colored and white input data environments.

Appendix A

In this appendix, (59) and (60) are derived. From Eq. (57)

$$E[\mu^2 e_a^H(n-1)C(n-1)e_a(n-1)]+E[\mu^2|v(n-1)|]E[C(n-1)] = 2\mu E[\overline{\theta}^H(n-1)\hat{\Phi}(n-1)G(n-1)e_a(n-1)] \quad (64)$$

where

$$C(n-1)=(\delta I+\mu\hat{\Phi}^H(n-1)\hat{S}(n-1))^{-1}\hat{\Phi}^H(n-1)R^{-1}(n-1)\hat{\Phi}(n-1)(\delta I+\mu\hat{S}(n-1)^H\hat{\Phi}(n-1))^{-1}$$

$$G(n-1)=(\delta I+\mu\hat{S}(n-1)^H\hat{\Phi}(n-1))^{-1}$$

It is clear that Eq. (17) can also be written as

$$\hat{\Phi}(n-1)=\hat{S}(n-1)+\hat{\Delta}(n-1)-\sum_{k=1}^{N}a_k(n-1)\hat{\Phi}(n-1-k) \quad (65)$$

where

$$\hat{\kappa}(n-q)=\left[0\ \dots\ 0\ 0\ \dots\ 0\ \sum_{j=1}^{M}x(n-q-j)\ \dots\ \sum_{j=1}^{M}x^L(n-q-j)\right]^H$$

$$\hat{\Delta}(n-1)=[\hat{\kappa}(n-1)\ \dots\ \hat{\kappa}(n-Q)]$$

Thus

$$\mu^2(E[e_a^H(n-1)C(n-1)e_a(n-1)]+E[|v(n-1)|]E[C(n-1)]) = 2\mu E\Big[e_a^H(n-1)G(n-1)e_a(n-1)+\overline{\theta}^H(n-1)\left(\hat{\Delta}(n-1)-\sum_{k=1}^{N}a_k(n-1)\hat{\Phi}(n-1-k)\right)G(n-1)e_a(n-1)\Big] \quad (66)$$

Expanding the expression for $\hat{\Phi}(n-1)$ taking into account the effects of the inputs within the time interval 0 to $n-1$, we have

$$\hat{\Phi}(n-1)=\hat{S}(n-1)+\hat{\Delta}(n-1)+\sum_{f=1}^{n-2}(-1)^f\prod_{g=0}^{f-1}\times\sum_{k_{n-f+g}=1}^{N}a_{k_{n-f+g}}(n-f+g)\hat{S}(n-f-1)+\sum_{f=1}^{n-2}(-1)^f\prod_{g=0}^{f-1}\sum_{k_{n-f+g}=1}^{N}a_{k_{n-f+g}}(n-f+g)\hat{\Delta}(n-f-1) \quad (67)$$

From Section 4, the requirement on $\hat{S}(n-1)$ is that it be sufficiently exciting in the sense that $\hat{S}(n-1)^H\overline{\theta}(n-1)=0$ only if $\overline{\theta}=0$. Thus at steady state

$$\hat{S}(n-1)^H\overline{\theta}(n-1)=\hat{S}(n-2)^H\overline{\boldsymbol{\theta}}(n-1)=\hat{S}(n-3)^H\overline{\theta}(n-1)=\cdots=e_a(n-1) \quad (68)$$

Substituting Eqs. (67) and (66) into (65) and simplifying give

$$\begin{aligned}\mu^2(E[e_a^H(n-1)C(n-1)e_a(n-1)]+E[|v(n-1)|]E[C(n-1)]) &= 2\mu E[e_a^H(n-1)G(n-1)e_a(n-1)] + 2\mu E[\overline{\boldsymbol{\theta}}^H(n-1)\hat{\Delta}(n-1)G(n-1)e_a(n-1)]\\ &+2\mu E\left[\sum_{f=1}^{n-2}(-1)^f\prod_{g=0}^{f-1}\sum_{k_{n-f+g}=1}^{N}a_{k_{n-f+g}}(n-f+g)e_a^H(n-1)G(n-1)e_a(n-1)\right]\\ &+2\mu E\left[\overline{\theta}^H(n-1)\sum_{f=1}^{n-2}(-1)^f\prod_{g=0}^{f-1}\sum_{k_{n-f+g}=1}^{N}a_{k_{n-f+g}}(n-f+g)\hat{\Delta}(n-f-1)G(n-1)e_a(n-1)\right]\end{aligned} \quad (69)$$

using the second assumption made in Section 5, Eq. (69) is reduced to

$$\mu^2E|e_a(n-1)|\mathrm{Tr}(E[C(n-1)])-2\mu E|e_a(n-1)|\mathrm{Tr}(E[G(n-1)])-2\mu\sum_{f=1}^{n-2}(-1)^f\prod_{g=0}^{f-1}\sum_{k_{n-f+g}=1}^{N}a_{k_{n-f+g}}(n-f+g)E|e_a(n-1)|\mathrm{Tr}(E[G(n-1)])=-\mu^2E[|v(n-1)|]\mathrm{Tr}(E[C(n-1)]) \quad (70)$$

$$E|e_a(n-1)|\Big(\mu^2\,\mathrm{Tr}(E[C(n-1)])-2\mu\,\mathrm{Tr}(E[G(n-1)])-2\mu\sum_{f=1}^{n-2}(-1)^f\prod_{g=0}^{f-1}\sum_{k_{n-f+g}=1}^{N}a_{k_{n-f+g}}(n-f+g)\mathrm{Tr}(E[G(n-1)])\Big)=-\mu^2E[|v(n-1)|]\mathrm{Tr}(E[C(n-1)]) \quad (71)$$

Eq. (71) can be used to develop an expression for the filter mean square error (MSE) or equivalent, for the filter excess mean square error (EMSE), which is defined by

$$EMSE=\lim_{n\to\infty}|e_a(n)| \quad (72)$$

From Eqs. (58) and (72), the MSE can be written as

$$MSE=EMSE+\sigma_v^2 \quad (73)$$

where $\sigma_v^2=E[|v(n-1)|]$. From Eq. (71), the MSE of the filter is given in Eq. (60)

$$EMSE=\frac{-\mu\sigma_v\,\mathrm{Tr}(E[C(n-1)])}{\mu Tr(E[C(n-1)])-2(\mu+\sum_{f=1}^{n-2}(-1)^f\prod_{g=0}^{f-1}\sum_{k_{n-f+g}=1}^{N}a_{k_{n-f+g}}(n-f+g)\mathrm{Tr}(E[G(n-1)])} \quad (74)$$

References

[1] D.T. Westwick, R.E. Kearney, Identification of Nonlinear Physiological Systems, Wiley, New York, 2003.
[2] K.J. Hunt, M. Munih, N.de.N. Donaldson, F.M.D. Barr, Investigation of the Hammerstein hypothesis in the modeling of electrically stimulated muscle, IEEE Trans. Biomed. Eng. 45 (8) (1998) 998–1009.
[3] I.W. Hunter, M.J. Korenberg, The identification of nonlinear biological systems: Wiener and Hammerstein cascade models, Biol. Cybern. 55 (1986) 135–144.

[4] S.W. Su, L. Wang, B.G. Celler, A.V. Savkin, Y. Guo, Identification and control of heart rate regulation during treadmill exercise, IEEE Trans. Biomed. Eng. 54 (2007) 1238–1246.
[5] B. Kwak, A.E. Yagle, J.A. Levitt, Nonlinear system identification of hydraulic actuator friction dynamics using a Hammerstein model, in: Proceedings of the IEEE ASSP'98, vol. 4, Seattle, WA, May 1998, pp. 1933–1936.
[6] E. Eskinat, S.H. Johnson, W.L. Luyben, Use of Hammerstein models in identification of nonlinear systems, AIChE J. 37 (1991) 255–268.
[7] R.C. Emerson, M.J. Korenberg, M.C. Citron, Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex, Biol. Cybern. 66 (1992) 291–300.
[8] T. Ogunfunmi, Adaptive Nonlinear System Identification: Volterra and Wiener Model Approaches, Springer, Berlin, 2007.
[9] V.J. Mathews, G.L. Sicuranza, Polynomial Signal Processing, Wiley, New York, 2000.
[10] E. Bai, Frequency domain identification of Hammerstein models, IEEE Trans. Autom. Control 48 (4) (April 2003) 530–542.
[11] Z.Q. Lang, A nonparametric polynomial identification algorithm for the Hammerstein system, IEEE Trans. Autom. Control 42 (October 1997) 1435–1441.
[12] S.V. Vaerenbergh, J. Via, I. Santamaria, Adaptive kernel canonical correlation analysis algorithms for nonparametric identification of Wiener and Hammerstein systems, EURASIP J. Adv. Signal Process. 2008 (2008) Article ID 875351.
[13] J. Jeraj, V.J. Mathews, A stable adaptive Hammerstein filter employing partial orthogonalization of the input signals, IEEE Trans. Signal Process. 54 (April 2006) 1412–1420.
[14] E.W. Bai, A blind approach to Hammerstein model identification, IEEE Trans. Signal Process. 50 (2002) 1610–1619.
[15] H. Chen, Pathwise convergence of recursive identification algorithms for Hammerstein systems, IEEE Trans. Autom. Control 49 (October 2004) 1641–1649.
[16] Y. Liu, E. Bai, Iterative identification of Hammerstein systems, Automatica 43 (2007) 346–354.
[17] J. Jeraj, V.J. Mathews, Stochastic mean square performance analysis of an adaptive Hammerstein filter, IEEE Trans. Signal Process. 54 (June 2006) 2168–2177.
[18] I. Umoh, T. Ogunfunmi, An adaptive algorithm for Hammerstein filter system identification, in: European Signal Processing Conference, Lausanne, Switzerland, 2008.
[19] A. Carini, V.J. Mathews, G.L. Sicuranza, Sufficient stability bounds for slowly varying direct-form recursive linear filters and their application in adaptive IIR filters, IEEE Trans. Signal Process. 47 (9) (September 1999) 2561–2567.
[20] M. Radenkovic, T. Bose, T. Mathurasai, Optimality and almost sure convergence of adaptive IIR filters with output error recursion, Digital Signal Process. 9 (October 1999) 315–328.
[21] M. Randenkovic, T. Bose, Adaptive IIR filtering of nonstationary signals, Signal Processing 81 (2001) 183–195.
[22] S. Haykin, Adaptive Filter Theory, fourth ed., Prentice-Hall, New Jersey, USA, 2002.
[23] P.A. Regalia, Adaptive IIR Filtering in Signal Processing and Control, Marcel Dekker Inc., New York, USA, 1995.
[24] J.J. Shynk, Adaptive IIR filtering, IEEE Signal Process. Mag. (April 1989) 4–21.
[25] S.D. Stearns, Error surface of recursive adaptive filters, IEEE Trans. Circuits Syst. 28 (June 1989) 603–606.
[26] B.D.O. Anderson, J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1979.
[27] J.W. Brewer, Kronecker products and matrix calculus in system theory, IEEE Trans. Circuits Syst. 25 (September 1978) 772–781.
[28] A.H. Sayed, Fundamentals of Adaptive Filtering, Wiley, New York, USA, 2003.