STOCHASTIC MEAN-SQUARE PERFORMANCE ANALYSIS OF AN ADAPTIVE HAMMERSTEIN FILTER

J. Jeraj and V.J. Mathews*

Department of Electrical & Computer Engineering
University of Utah
Salt Lake City, UT 84112, USA

ABSTRACT

This paper presents an almost sure (a.s.) mean-square performance analysis of an adaptive Hammerstein filter for the case when the measurement noise in the desired response signal is a martingale difference sequence. The system model consists of a series connection of a memoryless nonlinearity followed by a recursive linear filter. It is shown under the conditions of the analysis that the long-term time average of the squared excess estimation error of the adaptive filter can be made arbitrarily close to zero.

1. INTRODUCTION

This paper describes a theoretical performance evaluation of an adaptive algorithm employing a Hammerstein system model [1]. The system model consists of a series connection of a memoryless polynomial system followed by a recursive linear system. Even though different algorithms that adapt the parameters of a Hammerstein model can be found in the literature, there are limited or no convergence and stability analyses for these algorithms. This work is based on the analysis of a linear adaptive IIR filter in [2].

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

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

where

$$\widehat{A}(n,q^{-1}) = 1 + \widehat{a}_1(n)q^{-1} + \cdots + \widehat{a}_N(n)q^{-N}, \tag{2}$$

$$\widehat{B}(n,q^{-1}) = 1 + \widehat{b}_1(n)q^{-1} + \cdots + \widehat{b}_M(n)q^{-M}, \tag{3}$$

and q^{-1} represents the unit delay operator. In (1), $\widehat{z}(n)$ is the output of the memoryless polynomial nonlinear system and is given by

$$\widehat{z}(n) = \sum_{l=1}^{L} \widehat{w}_l(n)x^l(n). \tag{4}$$

In the above equations, $\widehat{w}_l(n)$, $\widehat{a}_i(n)$ and $\widehat{b}_j(n)$ represent the coefficients of the adaptive filter. $x(n)$ is the input signal. The algorithm for adapting these coefficients is given in Table 1. Equation (1) can be rewritten using vector notation as

$$\widehat{d}(n) = \widehat{\theta}_c^T(n) \cdot \widehat{\mathbf{H}}_c(n), \tag{5}$$

where

$$\widehat{\theta}_c(n) = \Big[\ \widehat{a}_1(n) \ \cdots \ \widehat{a}_N(n) \ \ \widehat{b}_1(n)\widehat{\mathbf{p}}^T(n-1) \ \cdots \ \widehat{b}_M(n)\widehat{\mathbf{p}}^T(n-M) \ \ \widehat{\mathbf{p}}^T(n) \ \ \overbrace{0\cdots0}^{2L}\ \Big]^T, \tag{6}$$

$$\widehat{\mathbf{H}}_c(n) = \Big[\ -\widehat{d}(n-1) \ \cdots \ -\widehat{d}(n-N) \ \ \mathbf{x}^T(n-1) \ \cdots \ \mathbf{x}^T(n-M) \ \ \mathbf{x}^T(n) \ \ \overbrace{0\cdots0}^{2L}\ \Big]^T, \tag{7}$$

$$\widehat{\mathbf{p}}(n) = [\ \widehat{w}_1(n) \ \ \widehat{w}_2(n) \ \ \ldots \ \ \widehat{w}_L(n)\ ]^T \tag{8}$$

and

$$\mathbf{x}(n) = [\ x(n) \ \ x^2(n) \ \ \ldots \ \ x^L(n)\ ]^T. \tag{9}$$

2. PERFORMANCE ANALYSIS

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

$$\tilde{d}(n) = \frac{B(n,q^{-1})}{A(n,q^{-1})}z(n), \tag{10}$$

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

$$z(n) = \mathbf{p}^T(n)\mathbf{x}(n), \tag{11}$$

and $A(n,q^{-1})$ and $B(n,q^{-1})$ are defined appropriately. The parameters employed above have similar meanings in the context of the unknown system as the parameters of the adaptive filter. The desired response signal is a noisy version of the output of the unknown system, and is given by

$$d(n) = \tilde{d}(n) + \nu(n). \tag{12}$$

In the above equation, $\nu(n)$ is an additive noise sequence that is uncorrelated with the input signal. We also have

$$\tilde{d}(n) = \theta_c^T(n) \cdot \mathbf{H}_c(n), \tag{13}$$

where $\theta_c(n)$, $\mathbf{H}_c(n)$, and $\mathbf{p}(n)$ are defined similar to (6), (7) and (8), omitting $\widehat{(\cdot)}$ that denotes estimated values.

The *a posteriori* estimation error $\epsilon(n) = d(n) - \widehat{\mathbf{H}}^T(n)\widehat{\theta}(n)$ (the variables $\widehat{\mathbf{H}}(n)$ and $\widehat{\theta}(n)$ are defined in Table 1) can be shown to be

$$\epsilon(n) = e(n)\frac{1}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}. \tag{14}$$

*Email: jeraj@eng.utah.edu

Substituting (14) in the update equation shown in Table 1 gives us

$$\widehat{\theta}(n) = \widehat{\theta}(n-1) + \Lambda(n)\phi(n)\epsilon(n). \tag{15}$$

In order to analyze the algorithm, we transform the equations into an equivalent, but different functional form. For this, we first add $2L$ zeros to the vectors $\widehat{\theta}(n)$, $\widehat{\theta}(n-1)$ and $\phi(n)$ in (15) to get

$$\widehat{\theta}_e(n) = \widehat{\theta}_e(n-1) + \Lambda_e(n)\phi_e(n)\epsilon(n), \tag{16}$$

where $\widehat{\theta}_e^T(n) = \left[\begin{array}{cc} \widehat{\theta}^T(n) & 0\cdots 0 \end{array}\right]$, $\phi_e^T(n)$ is defined similarly and $\Lambda_e(n)$ is a $(N+M+3\cdot L)\times(N+M+3\cdot L)$-element matrix. The expanded "step size" matrix $\Lambda_e(n)$ contains zeroes in the last 2 L rows, and $(M+L)2L$ non-zero terms $\rho_{i,l}(n)$ in the off diagonal entries. The $\rho_{i,l}(n)$ terms are placed in the 1 to $N+M+L$ rows of the columns $N+M+L+1$ through $N+M+3L$. The new variables $\rho_{i,l}(n)$ are finite, but not necessarily positive quantities as is the case with step sizes $\mu_i(n)$. The new elements $\rho_{i,l}(n)$ are placed at locations such that when $\Lambda_e(n)$ is multiplied with $\phi_e(n)$, the zero entries of $\phi_e(n)$ cancel the $\rho_{i,l}(n)$'s. Thus, the choice of these variables does not affect the update equation.

Table 1. The adaptive Hammerstein filter.

Definitions

$$\begin{aligned}
\widehat{\theta}(n) &= \left[\begin{array}{ccc} \widehat{a}_1(n)\cdots\widehat{a}_N(n) & \widehat{b}_1(n)\cdots\widehat{b}_M(n)\cdots\widehat{w}_L(n) \end{array}\right]^T \\
\widehat{\mathbf{H}}(n) &= \left[\begin{array}{cc} -\widehat{d}(n-1)\cdots-\widehat{d}(n-N) & \widehat{z}(n-1)\cdots \end{array}\right. \\
&\qquad\qquad \left.\begin{array}{cc} \widehat{z}(n-M) & x(n)\cdots x^L(n) \end{array}\right]^T \\
\Lambda(n) &= diag\left[\begin{array}{ccc} \mu_1(n) & \cdots & \mu_{N+M+L}(n) \end{array}\right]
\end{aligned}$$

Main Loop

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

$$\psi(n) = \left[\begin{array}{cc} -\widehat{d}(n-1)\cdots-\widehat{d}(n-N) & \widehat{z}(n-1)\cdots \\ \widehat{z}(n-M) & \sum_{j=0}^{M}\widehat{b}_j(n)x(n-j)\cdots\sum_{j=0}^{M}\widehat{b}_j(n)x^L(n-j) \end{array}\right]^T$$

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

Verify that $\{\mu(n)\cdot sign(e(n))\}$ is such that filter is stable. See [1].

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

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

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

$$\Lambda_e(n)\phi_e(n)\epsilon(n) = \Lambda_e(n)\widehat{\theta}_{r2}(n-1) + \Lambda_e(n)\widehat{\mathbf{H}}_d(n)\epsilon(n). \tag{17}$$

We only note at this time that there are more variables in the two vectors $\widehat{\theta}_{r2}(n-1)$ and $\widehat{\mathbf{H}}_d(n)$ than there are equations, and therefore a multitude of solutions exists for them. Employing (17) in (16), we get

$$\widehat{\theta}_e(n) = \widehat{\theta}_e(n-1) + \Lambda_e(n)\widehat{\theta}_{r2}(n-1) + \Lambda_e(n)\widehat{\mathbf{H}}_d(n)\epsilon(n). \tag{18}$$

Next, we multiply both sides of the $(N+1)$th through $(N+M)$th entries of (18) with $\widehat{\mathbf{p}}(n-1), \ldots, \widehat{\mathbf{p}}(n-M)$, respectively to obtain

$$\begin{aligned}
\widehat{\theta}_c(n) &= \widehat{\theta}_c(n-1) + \widehat{\theta}_r(n-1) + \Lambda_c(n)\widehat{\theta}_{r2c}(n-1) + \\
&\qquad \Lambda_c(n)\widehat{\mathbf{H}}_{dc}(n)\epsilon(n).
\end{aligned} \tag{19}$$

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

$$\Big[\overbrace{0\cdots0}^{N} \quad \widehat{b}_1(n-1)\left\{\widehat{\mathbf{p}}^T(n-1)-\widehat{\mathbf{p}}^T(n-2)\right\}\cdots$$
$$\widehat{b}_M(n-1)\left\{\widehat{\mathbf{p}}^T(n-M)-\widehat{\mathbf{p}}^T(n-M-1)\right\} \quad \overbrace{0\cdots0}^{3L}\Big]^T,$$

and $\widehat{\theta}_{r2c}(n-1)$ and $\widehat{\mathbf{H}}_{dc}(n)$ are vectors that result when we multiply $(N+1)$th through $(N+M)$th entries of $\widehat{\theta}_{r2}(n-1)$ and $\widehat{\mathbf{H}}_d(n)$ with $\widehat{\mathbf{p}}(n-1),\ldots,\widehat{\mathbf{p}}(n-M)$, respectively. The $(N+ML+3\cdot L)\times(N+ML+3\cdot L)$-element matrix $\Lambda_c(n)$ is obtained by replacing $\mu_{N+1}(n)$ through $\mu_{N+M}(n)$ with $(L\times L)$-element diagonal matrices $\mu_{N+1}(n)\mathbf{I}, \mu_{N+2}(n)\mathbf{I},\ldots,\mu_{N+M}(n)\mathbf{I}$, respectively. Also, each element $\rho_{N+1,1}(n)$ through $\rho_{N+M,2L}(n)$ of the matrix $\Lambda_e(n)$ is replaced by an $(L\times 1)$-element column vectors $[\rho_{N+1,1}(n) \ \cdots \ \rho_{N+1,1}(n)]^T$ through $[\rho_{N+M,2L}(n) \ \cdots \ \rho_{N+M,2L}(n)]^T$, respectively. Recall that the adaptive filter is composed of a polynomial with L coefficients and an IIR system with a denominator and numerator having N and M taps, respectively. With the above transformation, we expanded our original equation (15) from an $(N+M+L)\times 1$-dimensional equation to a vector equation (19) with $(N+M\cdot L+3\cdot L)$ dimensions. Let γ be a positive, finite constant of our choice. It can be shown that we can choose the "new" entries of $\Lambda_c(n)$, and the vectors $\widehat{\theta}_{r2c}(n-1)$ and $\widehat{\mathbf{H}}_{dc}(n)$ such that the following equality is satisfied in addition to the equality in (17):

$$\begin{aligned}
\Lambda_c(n)\widehat{\theta}_{r2c}(n-1) + \Lambda_c(n)\widehat{\mathbf{H}}_{dc}(n)\epsilon(n) &= -\widehat{\theta}_r(n-1) + \\
&\quad \gamma\widehat{\mathbf{H}}_c(n)\epsilon(n).
\end{aligned} \tag{20}$$

Even though (20) contains variables that are not present in the original algorithm, the components of (20) that correspond to the original adaptive filter have not changed in any way. Therefore, we can prove the convergence of the algorithm in Table 1 by proving the convergence of the algorithm given by

$$\widehat{\theta}_c(n) = \widehat{\theta}_c(n-1) + \gamma\widehat{\mathbf{H}}_c(n)\epsilon(n). \tag{21}$$

The above equation results from substituting (20) in (19).

2.1. Assumptions

Let $c_{(\cdot)}$ denote generic, finite, positive numbers.

A1: Input signal as well as the noise are bounded with bounds $|x(n)| < c_x, \forall n \geq 0$ and $|\nu(n)| < c_\nu, \forall n \geq 0$.

A2: (i) $\|\theta_c(n)\| < c_\theta$

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

(iii) The unknown plant is exponentially BIBO stable.

A3: Let $\triangle(n) = \theta_c(n) - \theta_c(n-1)$ denote the increment process associated with the unknown system. For some λ, where $0 < \lambda < 1$, there exists a constant α such that for all n, $\sum_{n=0}^{k} \lambda^{k-n} \|\triangle(n)\|^2 \leq \alpha$.

A4: Operator $A(k, q^{-1})$ is input strictly passive [3, 4], *i.e.*, there exists a positive constant ρ_0 such that

$$\sum_{k=0}^{n} u(k)\left[A(n,q^{-1})u(k)\right] \geq \rho_0 \sum_{k=0}^{n} u^2(k) \quad (22)$$

for any real sequence $\{u(k)\}, k \geq 0$. This assumption is a time-varying version of the well known strictly positive real condition. Note that ρ_0 is independent of the signal $u(n)$. It only depends on the properties of $A(n, q^{-1})$.

A5: The noise $\{\nu(n)\}$ is a martingale difference sequence, *i.e.* $E\{\nu(n+1)|F_n\} = 0$ almost surely (a.s.), and satisfies $\sup_n E\left\{|\nu(n+1)|^\delta |F_n\right\} < \infty$ (a.s.) for some $\delta > 2$ and $\lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} \nu^2(n) = \sigma_\nu^2$ (a.s.). In addition, $\{\nu(n)\}$ is independent of $\{\theta(n)\}$ and $\{x(n)\}$. F_n is the σ-algebra generated by $\{\nu(0), \nu(1), \ldots, \nu(n)\}$.

2.2. Main Result

We start by rewriting (21) as

$$\widehat{\theta}_c(n-1) = \widehat{\theta}_c(n) - \gamma \widehat{\mathbf{H}}_c(n)\epsilon(n). \quad (23)$$

Subtracting $\theta_c(n)$ from both sides, we get

$$\widetilde{\theta}_c(n-1) = \widetilde{\theta}_c(n) + \triangle(n) - \gamma \widehat{\mathbf{H}}_c(n)\epsilon(n), \quad (24)$$

where $\widetilde{\theta}_c(n) = \widehat{\theta}_c(n) - \theta_c(n)$. Pre-multiplying both sides of the above equation with their respective transposes gives

$$\begin{aligned}\|\widetilde{\theta}_c(n)\|^2 &= \|\widetilde{\theta}_c(n-1)\|^2 - 2\widetilde{\theta}_c^T(n)\triangle(n) + \\ & 2\gamma\widetilde{\theta}_c^T(n)\widehat{\mathbf{H}}_c(n)\epsilon(n) - \left\|\triangle(n) - \gamma\widehat{\mathbf{H}}_c(n)\cdot\epsilon(n)\right\|^2 . \quad (25)\end{aligned}$$

Since $\left\|\triangle(n) - \gamma\widehat{\mathbf{H}}_c(n)\cdot\epsilon(n)\right\|^2$ is non-negative, we can drop this term from the RHS of (25) and then substitute $\epsilon(n) = s(n) + \nu(n)$ to get:

$$\begin{aligned}\|\widetilde{\theta}_c(n)\|^2 &\leq \|\widetilde{\theta}_c(n-1)\|^2 + 2\|\widetilde{\theta}_c(n)\|\|\triangle(n)\| + \\ & 2\gamma\widetilde{\theta}_c^T(n)\widehat{\mathbf{H}}_c(n)s(n) + 2\gamma\widetilde{\theta}_c^T(n)\widehat{\mathbf{H}}_c(n)\nu(n). \quad (26)\end{aligned}$$

In a manner similar to the derivation in [2], we can show that

$$\widetilde{\theta}_c^T(n)\cdot\widehat{\mathbf{H}}_c(n) = -A(n, q^{-1})s(n). \quad (27)$$

Substituting the above result in (26) gives

$$\begin{aligned}\|\widetilde{\theta}_c(n)\|^2 &\leq \|\widetilde{\theta}_c(n-1)\|^2 + 2\|\widetilde{\theta}_c(n)\|\|\triangle(n)\| - \\ & 2\gamma s(n)\left[A(n,q^{-1})s(n)\right] + 2\gamma\widetilde{\theta}_c^T(n)\widehat{\mathbf{H}}_c(n)\nu(n). \quad (28)\end{aligned}$$

Let $\widehat{\Upsilon}(n)$ be an $(N + M\cdot L + 3L) \times (N + M + L)$-element matrix defined in such a way that direct multiplications will show that $\widehat{\theta}_c(n) = \widehat{\Upsilon}(n)\widehat{\theta}(n)$. Pre-multiplying both sides of the adaptation equation from Table 1 with $\widehat{\Upsilon}(n)$ and simplifying using $\widehat{\theta}_c(n) = \widehat{\Upsilon}(n)\widehat{\theta}(n)$ and the definition of $\widehat{\theta}_r(n-1)$ results in

$$\widehat{\theta}_c(n) = \widehat{\theta}_c(n-1) + \widehat{\theta}_r(n-1) + \frac{\widehat{\Upsilon}(n)\Lambda(n)\phi(n)e(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}. \quad (29)$$

Subtracting $\theta_c(n)$ from both sides of (29) gives

$$\begin{aligned}\widetilde{\theta}_c(n) &= \widetilde{\theta}_c(n-1) + \widehat{\theta}_r(n-1) - \triangle(n) + \\ & \frac{\widehat{\Upsilon}(n)\Lambda(n)\phi(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)} e(n). \quad (30)\end{aligned}$$

Direct calculations will show that $\widehat{\Upsilon}(n)\widehat{\mathbf{H}}(n) = \widehat{\mathbf{H}}_c(n)$, and $\widehat{\mathbf{H}}^T(n) = \widehat{\mathbf{H}}_c^T(n)\widehat{\Upsilon}(n)$. Next we use (30) in (28) with the above equalities to get

$$\begin{aligned}\|\widetilde{\theta}_c(n)\|^2 &\leq \|\widetilde{\theta}_c(n-1)\|^2 + 2\|\widetilde{\theta}_c(n)\|\|\triangle(n)\| - 2\gamma s(n)\cdot \\ & \left[A(n,q^{-1})s(n)\right] + 2\gamma\left(\widetilde{\theta}_c^T(n-1) + \widehat{\theta}_r^T(n-1)\right)\widehat{\mathbf{H}}_c(n)\nu(n) - \\ & 2\gamma\triangle^T(n)\widehat{\mathbf{H}}_c(n)\nu(n) + 2\gamma\frac{\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)} e(n)\nu(n). \quad (31)\end{aligned}$$

Theorem 1 *[5]*
Let assumption A5 hold, and let $f(n-1)$ be an F_{n-1} measurable sequence. Then

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

For $f(n-1)$ to be an F_{n-1} measurable, we require that $f(n-1)$ can be only a function of $\nu(k)$, where $k < n$. In more lax words, F_{n-1} measurability implies that $f(n-1)$ is a non-anticipative function of a signal $\nu(n)$. Note that $a_n = \circ(b_n)$ implies that $\lim_{n\to\infty} a_n/b_n = 0$ and $a_n = \bigcirc(b_n)$ implies that $|a_n/b_n| < c_B$, where c_B is a positive number. Then it follows that $a_n = \bigcirc(1)$ as $n \to +\infty$ means that $\{a_n\}$ is a bounded sequence.

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

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

$$\begin{aligned}\lim_{m\to\infty}\sup \frac{1}{m}\sum_{n=0}^{m}\left(d(n) - \widehat{d}(n) - \nu(n)\right)^2 \leq \\ \alpha\frac{c_{\widetilde{\theta}}}{\gamma\rho_0} + \min\left(1, |c_\Lambda|\right)\frac{1}{\rho_0}\sigma_\nu^2 \quad \textit{(a.s.)}, \quad (32)\end{aligned}$$

where ρ_0 is a parameter from assumption A4, while α was introduced in assumption A3. c_Λ is a bound such that

$$-1 < \widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n) \leq c_\Lambda. \quad (33)$$

Proof 1 *Summing both sides of (31) from $n = 1$ to m, it follows that*

$$\|\widetilde{\theta}_c(m)\|^2 + 2\gamma \sum_{n=1}^{m} s(n)\left[A(n,q^{-1})s(n)\right] \le \|\widetilde{\theta}_c(0)\|^2 + 2\sum_{n=1}^{m} \|\widetilde{\theta}_c(n)\|\|\triangle(n)\| + 2\gamma \sum_{n=1}^{m}\left(\widetilde{\theta}_c^T(n-1) + \widehat{\theta}_r^T(n-1)\right)\cdot \widehat{\mathbf{H}}_c(n)\nu(n) - 2\gamma\sum_{n=1}^{m}\triangle^T(n)\widehat{\mathbf{H}}_c(n)\nu(n) + 2\gamma\sum_{n=1}^{m}\frac{\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}e(n)\nu(n). \quad (34)$$

Since $\left(\widetilde{\theta}_c^T(n-1) + \widehat{\theta}_r^T(n-1)\right)\widehat{\mathbf{H}}_c(n)$ is F_{n-1} measurable we have that

$$\left|\sum_{n=1}^{m}\left(\widetilde{\theta}_c^T(n-1) + \widehat{\theta}_r^T(n-1)\right)\widehat{\mathbf{H}}_c(n)\nu(n)\right| \le o(m) \quad (35)$$

almost surely, where we have used the fact that $\left\|\left(\widetilde{\theta}_c^T(n-1) + \widehat{\theta}_r^T(n-1)\right)\right\|$ and $\left\|\widehat{\mathbf{H}}_c(n)\right\|$ are bounded for all $n \ge 0$. This is true since $\widetilde{\theta}_c(n-1)$, $\widehat{\theta}_r(n-1)$, $\widehat{\mathbf{H}}_c(n)$ are finite, and because $\widehat{A}(n-1,q^{-1})$ is guaranteed to be stable by our algorithm. By assumption A5, $\triangle(n)$ is independent of $\nu(n)$. It follows that $\triangle^T(n)\widehat{\mathbf{H}}_c(n)$ is F_{n-1} measurable, and by application of Theorem 1 to this sequence gives

$$\left|\sum_{n=1}^{m}\triangle^T(n)\widehat{\mathbf{H}}_c(n)\nu(n)\right| \le o(m) \quad (a.s.). \quad (36)$$

Similar calculations on the last term of (34) gives the following result:

$$\sum_{n=1}^{m}\frac{\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}e(n)\nu(n) = \sum_{n=1}^{m}\frac{\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)\nu(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}\left(\tilde{d}(n) + \nu(n) - \widehat{\mathbf{H}}_c^T(n)\widehat{\theta}_c(n-1)\right) \le \min(1,|c_\Lambda|)\sum_{n=1}^{m}\nu^2(n) + o(m) \quad (a.s.). \quad (37)$$

Note that $\Lambda(n)$, which is directly related to γ as described above, is chosen at time instant n, but independently of the value of $\nu(n)$. We obtained the result in (37) applying similar procedures as used to obtain (35) and (36). We only comment here on the term $\sum_{n=1}^{m}\frac{\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}\nu(n)\nu(n)$. Since the algorithm requires that $\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n) > -1$, and $\widehat{\mathbf{H}}^T(n)$, $\Lambda(n)$ and $\phi(n)$ are all finite due to the algorithm, we can introduce a bound c_Λ such that

$$-1 < \widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n) \le c_\Lambda. \quad (38)$$

Using (38) we have

$$\left|\sum_{n=1}^{m}\frac{\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}{1+\widehat{\mathbf{H}}^T(n)\Lambda(n)\phi(n)}\nu(n)\nu(n)\right| \le \left|\sum_{n=1}^{m}\frac{c_\Lambda}{1+c_\Lambda}\nu(n)\nu(n)\right|. \quad (39)$$

Noting that $\frac{c_\Lambda}{1+c_\Lambda} \le \min(1, c_\Lambda)$, we get $\left|\sum_{n=1}^{m}\frac{c_\Lambda}{1+c_\Lambda}\nu(n)\nu(n)\right| \le \min(1,|c_\Lambda|)\sum_{n=1}^{m}\nu^2(n)$. By assumption A3, $\|\triangle(n)\| \le \alpha$. Since $\|\widetilde{\theta}_c(n)\| \le c_{\widetilde{\theta}}$, we have

$$\sum_{n=1}^{m}\|\widetilde{\theta}_c^T(n)\|\|\triangle(n)\| \le \alpha c_{\widetilde{\theta}} m. \quad (40)$$

Using (35), (36), (37), (40) and assumption A4 in (34), results in

$$\|\widetilde{\theta}_c(m)\|^2 + 2\gamma\rho_0\sum_{n=1}^{m}s^2(n) \le \|\widetilde{\theta}_c(0)\|^2 + 2\alpha c_{\widetilde{\theta}} m + 2\gamma\min(1,|c_\Lambda|)\sum_{n=1}^{m}\nu^2(n) + o(m). \quad (41)$$

Dividing the entire equation (41) by $2\gamma\rho_0 m$ and taking the limit as m goes toward infinity, the Theorem 2 is proven. Q.E.D.

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

3. CONCLUDING REMARKS

A theoretical treatment of a recursive nonlinear adaptive filter developed in [1] was given in this paper. The convergence behavior of this algorithm was studied in a stochastic framework and in a non-stationary environment, and in the presence of a possibly colored and non-stationary measurement noise that is a martingale difference sequence. Using the martingale limit theorem, we showed that the global minimum on the error surface of our adaptive Hammerstein filter can be achieved with arbitrary precision when the rate of change of the parameters of the underlying plant is zero. The adaptive system analyzed in this paper does not account for the Gram-Schmidt orthogonalization of the input signal as done in [1]. Extension of the analysis to this case is straightforward.

4. REFERENCES

[1] J. Jeraj, V.J. Mathews, and J. Dubow, "A stable adaptive Hammerstein filter employing partial orthogonalization of the input signals," in *Proc. Int. Conf Acoustics, Speech and Sig. Process. (ICASSP'02)*, 2002, vol. 2, pp. II–1349–52.

[2] M. Radenkovic and T. Bose, "Adaptive IIR filtering of non-stationary signals," *Signal Processing*, vol. 81, no. 1, pp. 183–195, January 2001.

[3] G.C. Goodwin and K.S. Sin, *Adaptive Filtering Prediction and Control*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1984.

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

[5] W.F. Stout, *Almost Sure Convergence*, Academic Presss, New York, 1974.