A POSTERIORI UPDATES FOR ADAPTIVE FILTERS

S.C. Douglas[1] *and M. Rupp*[2]

[1]Dept. of Electrical Engineering, University of Utah, Salt Lake City, UT 84112 USA

[2]Wireless Research Laboratory, Lucent Technologies, Holmdel, NJ 07733 USA

ABSTRACT

In adaptive FIR filters, the least-mean-square (LMS) adaptive algorithm uses the *a priori* error signal to update the filter coefficients. In this paper, we study the forms and properties of *a posteriori* adaptive filter updates in a general context. We provide a technique by which the stability of an adaptive filter's coefficient update can be easily analyzed using the relationship between the *a priori* and *a posteriori* error signals. Using this knowledge, we then develop methods for choosing the algorithm step size to guarantee the robustness and stability of the system and to provide fast adaptation behavior. Simulations verify the usefulness of *a posteriori*–error–based adaptive algorithms for unbiased adaptive IIR filtering.

1. INTRODUCTION

The normalized least-mean-square (NLMS) adaptive filter is a useful technique for adjusting the L coefficients of a finite-impulse-response (FIR) filter. The NLMS coefficient updates are

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \overline{\mu}(k)e(k)\mathbf{x}(k) \quad (1)$$

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k) \quad (2)$$

$$\overline{\mu}(k) = \frac{\mu_0(k)}{||\mathbf{x}(k)||^2}, \quad (3)$$

where $\mathbf{w}(k) = [w_0(k) \ \cdots \ w_{L-1}(k)]^T$ and $\mathbf{x}(k) = [x(k) \cdots x(k-L+1)]^T$ are the coefficient and input signal vectors at time k, respectively, $e(k)$ is the *a priori* error signal at time k, $||\mathbf{x}(k)||^2$ denotes the L_2-norm of the vector $\mathbf{x}(k)$, and $\mu_0(k)$ is a step size parameter.

The NLMS adaptive filter is a version of the LMS adaptive filter in which the effective step size is $\overline{\mu}(k) = \mu_0(k)/||\mathbf{x}(k)||^2$. Since the LMS adaptive filter is usually derived and analyzed in a statistical context [1], such a view ignores certain useful stability and robustness properties possessed by the update in (1)–(3). In particular, it can be shown that the NLMS adaptive filter is a projection-type update, and its stability and robustness can be guaranteed so long as $0 < \mu_0(k) < 2$ [2, 3]. Recent techniques relating adaptive filtering algorithms to H^∞ stability theory show that the NLMS algorithm possesses a characteristic robustness that is independent of the statistical realizations of the signals in $\mathbf{x}(k)$ and $d(k)$ [4]–[6]. Moreover, a deterministic view of the NLMS algorithm elucidates the reasons behind the fast convergence behavior of this system over that of the LMS adaptive filter. For example, when $d(k) = \mathbf{x}^T(k)\mathbf{w}_{opt}$ with $\mathbf{w}_{opt}$ being an unknown FIR coefficient vector, then $\mathbf{w}(k)$ can be made to converge to $\mathbf{w}_{opt}$ in L iterations via (1–(3) for $\mu_0(k) = 1$, so long as the sequence $\mathbf{x}(k)$, $0 \leq k \leq L-1$ spans the L-dimensional coefficient space [3]. These results complement the understanding of the NLMS adaptive algorithm obtained in various statistical contexts [8]–[11].

Consider the following adaptive algorithm:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e_p(k)\mathbf{x}(k) \quad (4)$$

$$e_p(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k+1), \quad (5)$$

where $e_p(k)$, the *a posteriori* error signal, depends on the updated coefficients $\mathbf{w}(k)$. Using the matrix inversion lemma [12], one can show that (4)–(5) is equivalent to (1)–(2) for $\overline{\mu}(k) = \mu(k)/(1+\mu(k)||\mathbf{x}(k)||^2)$. Thus, the NLMS adaptive algorithm can be related to a coefficient update employing the *a posteriori* error signal. Although mentioned in several studies of (1)–(3), this fact has not previously been used to elucidate the useful properties of the NLMS algorithm. In addition, to our knowledge, the relationship between $e_p(k)$ and $e(k)$ has not previously been exploited to understand the behaviors of other adaptive algorithms whose updates are nonlinear with respect to the coefficient vector $\mathbf{w}(k)$. Examples of such algorithms include stochastic gradient methods based on non-mean-squared error criteria [13, 14] and recently-developed methods for adaptive IIR filters [15, 16]. Such results could have potential benefits in selecting step sizes for these algorithms to guarantee their stable operation and to provide fast convergence, particularly as such results are difficult to obtain via statistical characterizations due to the assumptions used in such analyses [17, 18].

In this paper, we develop a general theory for understanding the behavior of adaptive filtering algorithms whose updates depend on the coefficient vectors at time k and $k+1$. Our technique attempts to characterize the stability of any such algorithm using the nonlinear relationship between $e_p(k)$ and $e(k)$ as induced by the coefficient updates. So long as

$$|e_p(k)| \leq \beta|e(k)|, \qquad 0 < \beta < 1, \quad (6)$$

at each iteration, we prove that an algorithm employing non-mean-square error criteria is guaranteed to be both robust and stable. Such results are useful for several practical reasons. For example, it is possible to easily determine the form and value of the step size $\mu(k)$ that guarantees the stable operation of such an adaptive algorithm update. We

[0]This research was supported by NSF Grant No. MIP-9501680.

1058-6393/98 $10.00 © 1998 IEEE

then consider coefficient updates that, like (4)–(5), effectively employ the *a posteriori* error signal to update the filter coefficients, proving for a particular class of algorithms that such updates are stable for any positive bounded step size value. Simulations of *a posteriori* algorithms for unbiased adaptive FIR and IIR filtering indicate the useful convergence properties of the proposed schemes.

2. AN A POSTERIORI STABILITY THEORY

The technique for studying the stability of adaptive filters using the relationship between the *a priori* and *a posteriori* errors is an extension of the deterministic framework presented in [4]–[6], in which it is shown how adaptive filters can be related to well-known robustness schemes in control theory. Among other results, the theory provides conditions for L_2 stability, and it relates the noises and uncertainties within the system to the *a priori* and *a posteriori* errors. Algorithms that map uncertainties to smaller error values are inherently more robust.

In our extension, we consider the general algorithm form

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k) f(e(k))\mathbf{x}(k), \qquad (7)$$

where the odd function $f(e)$ is the derivative of a chosen convex cost function $\psi(e)$. This algorithm is a stochastic gradient method for minimizing $E\{\psi(e(k))\}$ iteratively with respect to $\mathbf{w}(k)$ [14], and its form includes many popular algorithms such as the sign-error, least-mean-K, and power-of-two quantized algorithms [7, 13].

For our analysis, we assume that $d(k)$ is generated according to the system identification model

$$d(k) = \mathbf{x}^T(k)\mathbf{w}_{opt} + \eta(k), \qquad (8)$$

where $\mathbf{w}_{opt}$ is an unknown coefficient vector and $\eta(k)$ is an observation noise signal. We also assume that $||\mathbf{x}(k)||^2 \geq \epsilon$ for some $\epsilon > 0$. Under such assumptions, we first state and then prove the following theorem:

Theorem 1: *Assume that an adaptive algorithm can be written in the form*

$$\mathbf{w}(k+1) = \mathbf{w}(k) + [e(k) - Q(e(k))]\frac{\mathbf{x}(k)}{||\mathbf{x}(k)||^2}, \qquad (9)$$

where $Q(e)$ is any function. Then, the a posteriori error is given by

$$e_p(k) = Q(e(k)). \qquad (10)$$

Moreover, if $Q(e)$ is contractive for all e such that

$$|e_p(k)| \leq \beta|e(k)|, \qquad (11)$$

for some $0 < \beta < 1$, then the algorithm is L_2-stable.

Proof: To show that $e_p(k)$ is the *a posteriori* error for the update in (9), assume that $d(k)$ is of the form in (8). By subtracting $\mathbf{w}_{opt}$ from both sides of (9), we obtain

$$\widetilde{\mathbf{w}}(k+1) = \widetilde{\mathbf{w}}(k) - [e(k) - Q(e(k))]\frac{\mathbf{x}(k)}{||\mathbf{x}(k)||^2}, \qquad (12)$$

where $\widetilde{\mathbf{w}}(k) = \mathbf{w}_{opt} - \mathbf{w}(k)$ is the coefficient error vector. Pre-multiplying both sides of this relation by $\mathbf{x}^T(k)$ and adding $\eta(k)$ on both sides, we obtain (10), where $e_p(k)$ is defined in (5).

To prove the second part of the theorem, we take the L_2-norms of both sides of (12). After some simplification the relationship

$$||\widetilde{\mathbf{w}}(k+1)||^2 + \frac{\overline{e}^2(k)}{||\mathbf{x}(k)||^2} = ||\widetilde{\mathbf{w}}(k)||^2 + \frac{[e_p(k) - \eta(k)]^2}{||\mathbf{x}(k)||^2} \qquad (13)$$

is obtained, where we have defined the noiseless error signal $\overline{e}(k)$ as

$$\overline{e}(k) = e(k) - \eta(k) = \mathbf{x}^T(k)\widetilde{\mathbf{w}}(k). \qquad (14)$$

It can be shown for any two real numbers a and b and any positive constant c that

$$[a+b]^2 \leq (1+c)\left[a^2 + \frac{1}{c}b^2\right]. \qquad (15)$$

By assigning $a = e_p(k)$ and $b = \eta(k)$, we obtain

$$[e_p(k) - \eta(k)]^2 \leq (1+c)\left[e_p^2(k) + \frac{1}{c}\eta^2(k)\right]. \qquad (16)$$

Using the condition in (11), (16) becomes

$$[e_p(k) - \eta(k)]^2 \leq (1+c)\left[\beta^2\{\overline{e}(k) + \eta(k)\}^2 + \frac{1}{c}\eta^2(k)\right] \qquad (17)$$

Then, employing (15) for $\{\overline{e}(k) + \eta(k)\}^2$ yields

$$[e_p(k) - \eta(k)]^2 \leq (1+c)^2\beta^2\left[\overline{e}^2(k) + \frac{1}{c}\left(1 + \frac{1}{(1+c)\beta^2}\right)\eta^2(k)\right]. \qquad (18)$$

If c is chosen such that

$$0 < c < \frac{1}{\beta} - 1, \qquad (19)$$

then

$$[e_p(k) - \eta(k)]^2 \leq \gamma_1\overline{e}^2(k) + \gamma_2\eta^2(k), \qquad (20)$$

where γ_1 and γ_2 are positive constants such that $0 < \gamma_1 < 1$. Combining (13) and (20), we obtain the relation

$$||\widetilde{\mathbf{w}}(k+1)||^2 + (1-\gamma_1)\frac{\overline{e}^2(k)}{||\mathbf{x}(k)||^2} \leq \gamma_2\frac{\eta^2(k)}{||\mathbf{x}(k)||^2} + ||\widetilde{\mathbf{w}}(k)||^2. \qquad (21)$$

Iterating (21) from $k = 0$ to $k = n-1$ yields

$$||\widetilde{\mathbf{w}}(n)||^2 + (1-\gamma_1)\sum_{k=0}^{n-1}\frac{\overline{e}^2(k)}{||\mathbf{x}(k)||^2} \leq \gamma_2\sum_{k=0}^{n-1}\frac{\eta^2(k)}{||\mathbf{x}(k)||^2} + ||\widetilde{\mathbf{w}}(0)||^2, \qquad (22)$$

and thus

$$\sum_{k=0}^{n-1}\frac{\overline{e}^2(k)}{||\mathbf{x}(k)||^2} < \frac{||\widetilde{\mathbf{w}}(0)||^2}{1-\gamma_1} + \frac{\gamma_2}{1-\gamma_1}\sum_{k=0}^{n-1}\frac{\eta^2(k)}{||\mathbf{x}(k)||^2}. \qquad (23)$$

Therefore, as $n \to \infty$, the weighted sum-of-squared-errors $e^2(k)/||\mathbf{x}(k)||^2$ remains bounded so long as $\eta^2(k)/||\mathbf{x}(k)||^2$ is bounded □.

The above theorem indicates that the robustness of any adaptive algorithm of the form in (7) can be determined by considering the form of $\mu(k)f(e(k))$ that appears in the coefficient updates. If the function

$$Q(e(k)) \quad = \quad e(k) - \mu(k)||\mathbf{x}(k)||^2 f(e(k)) \qquad (24)$$

is contractive for all possible $e(k)$, then the algorithm is robust in the sense of the bounded error condition in (23). Such a technique can yield useful information about selecting $\mu(k)$ for the chosen algorithm. The following example indicates how to use these results.

Example: Consider the algorithm

$$\mathbf{w}(k+1) \quad = \quad \mathbf{w}(k) + \mu(k)|e(k)|^{K-2}e(k)\mathbf{x}(k). \quad (25)$$

This algorithm approximately minimizes the mean-Kth error criterion $E\{|e(k)|^K\}$ iteratively with respect to $\mathbf{w}(k)$ over time. Since the algorithm is of the form in (7), the stability of the algorithm is assured if the polynomial

$$Q(e(k)) \quad = \quad (1 - \mu(k)|e(k)|^{K-2}||\mathbf{x}(k)||^2)e(k) \quad (26)$$

is contractive for all $e(k)$. Such will be the case if

$$0 < \mu(k) < \frac{2}{|e(k)|^{K-2}||\mathbf{x}(k)||^2}. \qquad (27)$$

This condition can be used to test and adjust the value of $\mu(k)$ at each iteration, if necessary, to ensure the system's stability □.

It should be stated that the result in (27) simply constrains $\mu(k)$ so that the magnitude of $\mu(k)f(e(k))$ never exceeds that of $\overline{\mu}(k)e(k)$ in the NLMS update in (1). The true utility of these ideas is seen when one considers algorithms of a somewhat-more-general nature than that in (7), as described in the following sections.

3. A POSTERIORI ERROR ADAPTATION

Since the stability of an adaptive algorithm can be determined from the relationship between the *a priori* and *a posteriori* errors, it is reasonable to consider algorithms that by their very structure guarantee their stability for all possible parameter choices $\mu(k)$. In this section, we discuss algorithms whose forms satisfy this constraint. Such algorithms employ the *a posteriori* error directly within the coefficient updates. We first state and then prove the following theorem.

Theorem 2: *Assume that an adaptive algorithm can be written in the form*

$$\mathbf{w}(k+1) \quad = \quad \mathbf{w}(k) + \mu(k)f(e_p(k))\mathbf{x}(k), \qquad (28)$$

where the a posteriori error $e_p(k)$ is as defined in (5) and $f(e)$ is any function satisfying

$$\begin{aligned} \operatorname{sgn}(f(e)) &= \operatorname{sgn}(e) & (29) \\ |f(e)| &\geq \gamma|e| & (30) \end{aligned}$$

for some $\gamma > 0$. Then, the algorithm is L_2-stable for any bounded $\mu(k) > 0$.

Proof: By pre-multiplying both sides of the above equation by $\mathbf{x}^T(k)$ and subtracting $d(k)$ from both sides, we obtain the relation

$$e_p(k) \quad = \quad e(k) - \mu(k)||\mathbf{x}(k)||^2 f(e_p(k)), \qquad (31)$$

or

$$e_p(k) + \mu(k)||\mathbf{x}(k)||^2 f(e_p(k)) \quad = \quad e(k). \qquad (32)$$

Since $\operatorname{sgn}(f(e)) = \operatorname{sgn}(e)$, we have that $\operatorname{sgn}(e_p(k)) = \operatorname{sgn}(e(k))$ when $\mu(k) > 0$, and thus

$$|e_p(k)| + \mu(k)||\mathbf{x}(k)||^2|f(e_p(k))| \quad = \quad |e(k)|. \qquad (33)$$

Using (30), we obtain

$$(1 + \mu(k)\epsilon\gamma)|e_p(k)| \leq |e(k)|. \qquad (34)$$

Since the factor premultiplying $|e_p(k)|$ in (34) is strictly greater than one, then the relation in (11) is satisfied. Thus, from Theorem 1, the algorithm is stable □.

The above result is the extension of the *a posteriori* version of the NLMS algorithm in (4) to the general case of a nonlinear coefficient update. As is the case with the algorithm in (4), we can freely choose any positive step size $\mu(k)$ in (28) and obtain stable behavior, regardless of the forms of $d(k)$ and $\mathbf{x}(k)$. Because $e_p(k)$ depends on $\mathbf{w}(k+1)$, however, (28) is not an update, and thus to use this form, one must explicitly find a solution to (33) so that the RHS of (28) can be expressed in terms of $\mathbf{w}(k)$. An example illustrates this calculation.

Example: Consider the algorithm

$$\begin{aligned} &\mathbf{w}(k+1) \\ &\quad = \mathbf{w}(k) + \mu(k)[\gamma|e_p(k)| + e_p^2(k)]\operatorname{sgn}(e_p(k))\mathbf{x}(k), \quad (35) \end{aligned}$$

which is of the form in (28). Thus, from (33), the relationship between $|e_p(k)|$ and $|e(k)|$ is

$$|e_p(k)| + \alpha(k)[\gamma|e_p(k)| + |e_p(k)|^2] \quad = \quad |e(k)|, \quad (36)$$

where we have defined

$$\alpha(k) \quad = \quad \mu(k)||\mathbf{x}(k)||^2. \qquad (37)$$

This is a quadratic equation in $|e_p(k)|$ with positive root

$$\begin{aligned} &|e_p(k)| \\ &= \frac{1}{2\alpha(k)}\left(-1 - \alpha(k)\gamma + \sqrt{(1+\alpha(k)\gamma)^2 + 4\alpha(k)|e(k)|}\right). \\ &\qquad\qquad (38) \end{aligned}$$

As γ tends to zero, the algorithm becomes

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{w}(k) + \frac{h(\alpha(k)|e(k)|)}{\alpha(k)}\operatorname{sgn}(e(k))\mathbf{x}(k) (39) \\ h(e) &= \frac{1}{2}\left(1 + 2e - \sqrt{1+4e}\right). \qquad (40) \end{aligned}$$

Figure 1 plots the function $h(e)$ for positive values of e, where it is seen that it has similar behavior as e^2 near the origin and similar behavior as e when $e \to \infty$. Thus, the update in (39) guarantees that the magnitude of the update direction never exceeds $|e(k)|$ for any $\mu(k) > 0$ □.

As the previous example shows, the *a posteriori* version of any algorithm can be quite complicated when placed in standard *a priori* form. Such complexity issues must be weighed in any particular application.

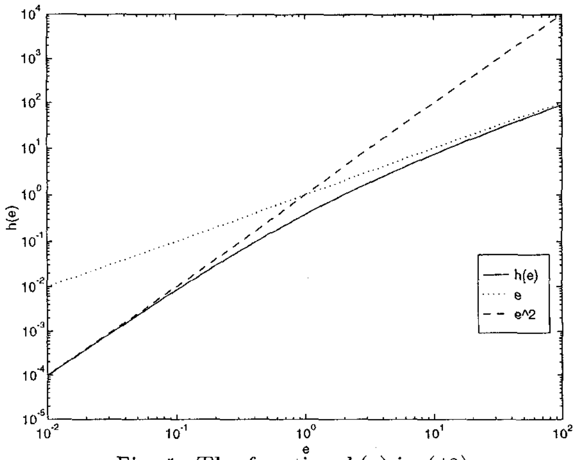

Fig. 1: The function $h(e)$ in (40).

4. EXTENSION TO BIAS REMOVAL ALGORITHMS

We now consider algorithms of a different form than that in (7). The algorithms considered here are designed to reduce coefficient biases due to noisy regressor vectors in adaptive FIR and IIR filters and are known as bias removal, unit-norm-constrained, or anti-Hebbian methods [15, 16, 19]–[22]. The coefficient updates for these algorithms are

$$\mathbf{w}(k+1) \quad = \quad \mathbf{w}(k) + \mu(k)e(k)\left[\mathbf{x}(k) + g(k)\mathbf{w}(k)\right], \tag{41}$$

where $g(k)$ is a term that depends on $d(k)$, $\mathbf{x}(k)$, and/or $\mathbf{w}(k)$. Two common choices for $g(k)$ are

$$g_1(k) = \frac{e(k)}{1+||\mathbf{w}(k)||^2} \quad \text{and} \quad g_2(k) = d(k). \tag{42}$$

We can determine the relationship between the *a priori* and *a posteriori* errors for these algorithms. Premultiplying both sides of (41) by $\mathbf{x}^T(k)$ and subtracting $d(k)$ from both sides, we obtain

$$e_p(k) \quad = \quad \{1-\mu(k)[||\mathbf{x}(k)||^2 + g(k)y(k)]\}e(k). \tag{43}$$

From this equation, we see that $|e_p(k)| < |e(k)|$ when

$$\left|1-\mu(k)[||\mathbf{x}(k)||^2 + g(k)y(k)]\right| \quad < \quad 1, \tag{44}$$

and thus we propose the step size choice

$$\mu(k) = \begin{cases} \dfrac{\mu_0}{||\mathbf{x}(k)||^2 + g(k)y(k)} & \text{if } ||\mathbf{x}(k)||^2 + g(k)y(k) > \delta, \\ 0 & \text{otherwise,} \end{cases} \tag{45}$$

where $0 < \mu_0 < 1$ and δ is a positive constant. This choice guarantees that $|e_p(k)| < |e(k)|$ and $\mathrm{sgn}(e_p(k)) = \mathrm{sgn}(e(k))$ for (41), which are two conditions that are also satisfied by the *a posteriori* NLMS update in (4).

We can also determine a bias removal algorithm that effectively employs the updated coefficient vector $\mathbf{w}(k+1)$ within the coefficient updates, as defined by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e_p(k)\left[\mathbf{x}(k) + d(k)\mathbf{w}(k+1)\right]. \tag{46}$$

By subtracting $d(k)e_p(k)\mathbf{w}(k+1)$ from both sides of (46) and dividing the resulting expression by $(1-\mu(k)d(k)e_p(k))$, we obtain

$$\mathbf{w}(k+1) = \frac{1}{1-\mu(k)d(k)e_p(k)}\left[\mathbf{w}(k) + \mu(k)e_p(k)\mathbf{x}(k)\right]. \tag{47}$$

Then, using similar manipulations as in previous cases, we obtain

$$\mu(k)d(k)e_p^2(k) + e_p(k)b(k) \;=\; e(k) \tag{48}$$

$$b(k) \;=\; 1 + \mu(k)[||\mathbf{x}(k)||^2 + d^2(k)] \tag{49}$$

such that

$$e_p(k) \quad = \quad q(\mu(k)d(k), b(k), e(k)) \tag{50}$$

$$q(a,b,c) \quad = \quad \frac{1}{2a}\left\{-b + \sqrt{b^2+4ac}\right\}. \tag{51}$$

We have simulated the behaviors of the proposed bias removal algorithms for a particular autoregressive modeling task. In these simulations, the input and desired response signals are generated as

$$x(k) \quad = \quad \overline{x}(k) + \nu(k) \tag{52}$$

$$\overline{x}(k) \quad = \quad 1.317\overline{x}(k-1) - 0.81\overline{x}(k-2) + s(k) \tag{53}$$

$$d(k) \quad = \quad x(k+1) - s(k+1), \tag{54}$$

where $s(k)$ and $\nu(k)$ are zero-mean Gaussian signals with variances of unity and 0.01, respectively. In each case, we have chosen $L=3$, such that $\mathbf{w}_{opt} = [1.317\ -0.81\ 0]^T$ in the absence of any observation noise, and $\mathbf{w}(0) = [0\ 0\ 0]^T$ in all cases. For each algorithm, we compute the average value of the total coefficient error power $||\widetilde{\mathbf{w}}(k)||^2$ as obtained from ensemble averages of one hundred different simulation runs on pseudo-random signals. We have chosen step sizes of $\mu_0(k) = 0.1$, $\mu_0 = 0.1$, $\mu_0 = 0.1$, and $\mu(k) = 0.05$ in the NLMS algorithm in (1)–(3), the two bias removal algorithms employing $g_1(k)$ and $g_2(k)$ in [(41),(42),(45)], and the *a posteriori* bias removal algorithm in [(47),(49)–(51)], respectively.

Figure 2 shows the average behaviors of the various methods on these signals, where it is seen that the bias removal algorithms perform better than the NLMS algorithm in estimating the coefficients of the unknown model. In particular, the *a posteriori* algorithm of [(47),(49)–(51)] is found to provide much-faster convergence than that provided by the other schemes for equal or lower coefficient error powers in steady-state. Although simulations are not shown, various tests confirmed that one cannot choose a fixed step size for the bias removal algorithms in [(41),(42),(45)] and obtain fast convergence behavior as exhibited by the schemes shown. Clearly, the proposed methods provide faster, more accurate convergence to a lower error state through their choices of step sizes.

It should be noted that at the time of writing, the general stability theory of the previous sections, and Theorems 1 and 2 in particular, have not been directly extended to these cases. Even so, the simulated performance of the algorithms appears to be quite good, and the stability of the algorithms appears to be maintained in situations where they are appropriate. Further analysis and study is needed before a definitive statement regarding the robustness of these algorithms can be given.

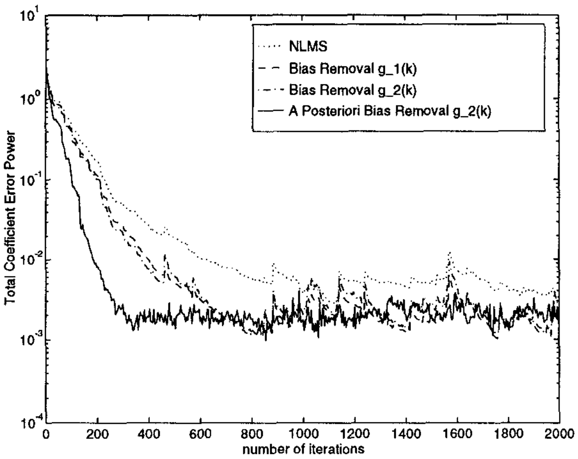

Fig. 2: Convergence of the averaged total coefficient error powers for the various algorithms in the simulation example.

5. CONCLUSIONS

In this paper, we have presented a general theory for understanding the robustness of non-mean-square-error-based adaptive filtering algorithms using the relationship between the *a priori* and *a posteriori* error signals. The theory provides a simple method for characterizing the stability properties of these adaptive filtering schemes that complements more-complicated statistical analyses of the nonlinear coefficient updates. In addition, we have indicated how to express and derive adaptive filtering algorithms that are based on *a posteriori* error minimization, and we have verified their stability and robustness properties. Simulations indicate the useful convergence properties of the resulting algorithms as applied to bias removal techniques. The proposed methods are expected to be useful for a number of adaptive filtering algorithms and applications.

REFERENCES

[1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing* (New Jersey: Prentice-Hall, 1985).

[2] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 282-287, June 1967.

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

[4] M. Rupp and A.H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Processing,* vol. 44, pp. 1428-1440, June 1996.

[5] A.H. Sayed and M. Rupp, "Error energy bounds for adaptive gradient algorithms," *IEEE Trans. Signal Processing,* vol. 44, pp. 1982-1989, Aug. 1996.

[6] A.H. Sayed and M. Rupp, "An l_2-stable feedback structure for nonlinear H^∞ adaptive filtering," *Automatica,* vol. 33, no. 1, pp. 13-30, Jan. 1997.

[7] D.L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-30, pp. 578-586, Aug. 1982.

[8] N.J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 793-806, Aug. 1986.

[9] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data," *IEEE Trans. Inform. Theory,* vol. 34, pp. 680-691, July 1988.

[10] D.T.M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Processing,* vol. 41, pp. 2811-2825, Sept. 1993.

[11] S.C. Douglas and T.H.-Y. Meng, "Normalized data nonlinearities for LMS adaptation," *IEEE Trans. Signal Processing,* vol. 42, pp. 1352-1365, June 1994.

[12] T. Kailath, *Linear Systems* (Englewood Cliffs, NJ: Prentice-Hall, 1980).

[13] E. Walach and B. Widrow, "The least-mean-fourth (LMF) adaptive algorithm and its family," *IEEE Trans. Inform. Theory,* vol. IT-30, pp. 275-283, Mar. 1984.

[14] S.C. Douglas and T.H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Processing,* vol. 42, pp. 1335-1351, June 1994.

[15] S.C. Douglas and M. Rupp, "On bias removal and unit norm constraints in equation-error adaptive IIR filters," *Proc. 30th Asilomar Conf. Signals, Systems, Comput.*, Pacific Grove, CA, vol. 2, pp. 1093-1097, Nov. 1996.

[16] M. Rupp and S.C. Douglas, "Deterministic stability analyses of unit-norm constrained algorithms for unbiased adaptive IIR filtering," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Munich, Germany, vol. 3, pp. 1937-1940, Apr. 1997.

[17] J.E. Mazo, "On the independence theory of equalizer convergence," *Bell Sys. Tech. J.* , vol. 58, no. 3, pp. 963-993, May-June 1979.

[18] S.C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Trans. Signal Processing,* vol. 43, pp. 2863-2871, Dec. 1995.

[19] V.U. Reddy, B. Egardt, and T. Kailath, "Least squares type algorithm for adaptive implementation of Pisarenko's harmonic retrieval method," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-30, pp. 399-405, June 1982.

[20] P.A. Regalia, "An unbiased equation error identifier and reduced-order approximations," *IEEE Trans. Signal Processing,* vol. 42, pp. 1297-1412, June 1994.

[21] K. Gao, M.O. Ahmad, and M.N.S. Swamy, "A constrained anti-Hebbian learning algorithm for total least-squares estimation with applications to adaptive FIR and IIR filtering," *IEEE Trans. Circuits Syst. II: Anal. Dig. Signal Processing,* vol. 41, pp. 718-729, Nov. 1994.

[22] S.C. Douglas, "Analysis of an anti-Hebbian adaptive FIR filtering algorithm," *IEEE Trans. Circ. Syst. II: Anal. Dig. Signal Processing,* vol. 43, pp. 777-780, Nov. 1996.