The Journal of Natural Science,
Pai Chai University, Korea
Volume 8: 5～15, 1995

ELIMINATION OF BIAS IN THE IIR LMS ALGORITHM

Seung-Hyon Nam and Yong-Hoh Kim*
Department of Applied Mathematics, Pai Chai University
* Department of Electronic Engineering, Pai Chai University

IIR LMS 알고리즘에서의 바이어스 제거

남승현, 김용호*
배재대학교 응용수학과
**배재대학교 전자공학과*

The equation error formulation in the adaptive IIR filtering provides convergence to a global minimum regardless a local minimum with a large stability margin. However, the equation error formulation suffers from the bias in the coefficient estimates. In this paper, a new algorithm, which does not require a prespecification of the noise variance, is proposed for the equation error formulation. This algorithm is based on the equation error smoothing and provides an unbiased parameter estimate in the presence of white noise. Through simulations, it is demonstrated that the algorithm eliminates the bias in the parameter estimate while retaining good properties of the equation error formulation such as fast convergence speed and the large stability margin.

IIR 적응 휠터의 공식오차 방식은 지역 최소값에 관계없이 전역 최소값에 수렴하며 안정성이 높다. 그러나 공식오차 방식은 입력 신호에 잡음이 섞여 경우 예측계수가 바이어스 되는 문제가 있다. 본 논문에서는 사전에 잡음에 대한 지식이 없이 바이어스가 없는 예측계수를 얻을 수 있는 새로운 공식오차 방식을 위한 알고리즘을 제안한다. 이 알고리즘은 공식오차를 스므딩하는 방식을 이용하여 입력에 추가되는 잡음이 백색잡음인 경우 바이어스 없이 계수를 예측할 수 있다. 시뮬레이션을 통해 새로운 알고리즘이 공식오차의 주요한 장범인 빠른 수렴속도와 안정성을 유지하며 바이어스를 효율적으로 제거함을 볼 수 있다.

Key words : adaptive IIR filtering, equation error, output error, LMS, bias.

I. INTRODUCTION

Over the past decade adaptive IIR filtering has been studied extensively[2~6,13] and many application areas have been considered.[8,10,16] Adaptive IIR filtering has a pole-zero structure whereas adaptive FIR filtering has only an all-zero structure. Inclusion of poles in adaptive filtering changes the filtering prodblem in many ways and adaptive IIR filtering has many advantages over its adaptive FIR counterpart. For example, in channel equalization problems, communication channels are modeled to have zeros. Thus it is necessary for an adaptive filter to have poles to remove distortion caused by the channel. In adaptive FIR filtering, however, poles are approximated using zeros since FIR filters do not have poles. Hence, the number of taps must be large enough to get satisfactory results. On the other hand, if we use adaptive IIR filtering, the zeros in the channel model can be exactly countered using poles in the adaptive filter. Therefore, the number of taps in the adaptive filter is significantly reduced and the computational burden is greatly relieved.

Although adaptive IIR filtering provides better performances in many applications, it has a stability problem. In adaptive FIR filtering, instability occurs when the coefficients get larger without bound due to a large step size beyond the upper bound. In adaptive IIR filtering, however, instability can occur without the coefficients blow-up since the poles outside the unit circle produce unbounded output. Hence, there are two sources of instability in adaptive IIR filtering. Despite this stability problem, adaptive IIR filtering would be an ultimate way to adaptive filtering since it is promising in many signal processing applications.

In general, there are two approaches to adaptive IIR filtering that minimizes the mean square error(MSE). One is the equation error formulation and the other is the output error formulation.

In the output error formulation, an adaptive algotive algorithm updates the feedback coefficient directly. Hence it is considered as a natural generalization of adaptive FIR filtering. The output-error adaptive filter is in a recursive form in that the filter output is fed back to the input. Due to this feedback, the filter output is a nonlinear function of the filter coefficients and the MSE surface is not quadratic. This nonlinearity makes the output error formulaton complicated than the equation error counterpart. Further, the MSE surface may have local minima as well as a global minimum in some cases. The MSE surface for the output error formulation has been investigated.[11,12,14] The MSE surface has local minima when the order of the adaptive filter is less than that of the signal model. Moreover, even in the exact case, the MSE surface may have local minima when the input is colored.[12] Some sufficient conditions for the error surface not to have local minima has been investigated.[14] The necessary and sufficient conditions are, however, not known to date. Convergence to a global minimum also depends on the specific algorithm. No algorithm is known to converge to the global minimum for all cases. The nonquadratic nature of the MSE surface also makes adaptive algorithms slower than their equation error counterparts.

In the equation error formulation, an adaptive algorithm updates the feedback coefficients in an all-zero, nonrecursive form. The feedback coefficients are then copied to a second filter implemented in an all-pole form. Therefore, the equation error formulation uses an adaptive FIR filtering technique directly. Algorithms like RLS and LMS can be employed to update filter coefficients. Adaptive algorithms based on the equation error formulation have many properties in common with the corresponding adaptive FIR algorithm. The MSE surface is quadratic with respect to the filter coefficients. Hence, it has a global minimum without lacal minima. Convergence speed is faster than that of the output error counterpart. Adaptive algorithm is in a simple form due to nonrecursive nature of the adaptive filter. Moreover, it has the hyperstability feature in that the adaptive filter can remain stable even if poles are outside the unit circle over a half of the time.[17] Unfortunately, however, its filter coefficients are biased in the presence of additive noise.[1] Thus the equation error formulation has been limited to those where the bias is not a significant problem.

Bias in the equation error formulation was investigated by Mendel.[1] The bias is a function of the signal to noise(SNR) ratio. That is, the smaller the SNR the larger the bias. He devised an unbiased algorithm assuming that the variance of noise is available. Similar approach was used by Treichler for frequency estimation of a noisy sinusoid corrupted by additive white noise.[9] However, these approaches require a priori knowledge on the variance of additive noise, which is not available in many real situations. This paper explores the bias problem in detail and proposes an algorithm the provides unbiased coefficient estimates.

II. ADAPTIVE IIR FILTERING

1. Output Error Formulation

Assume that the unknown system in Fig.1 is described by the difference equation

$$y(k)=\sum_{i=1}^{N} a_i y(K-i)+\sum_{j=0}^{M} b_j x(k-j) \quad (1)$$

where a_i's and b_j's are coefficients to be estimated. Throughout the paper, the subscript a in Fig. 1 will be replaced by others for convience. For example. e_o denotes the output error and e_o the equation error.

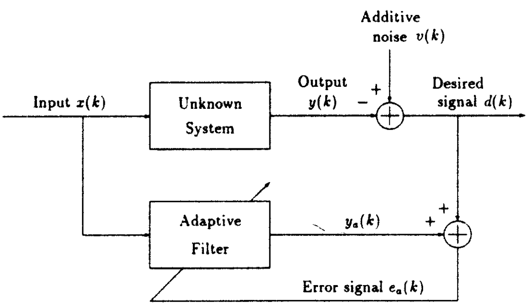

Fig. 1 System Identification Configuration.

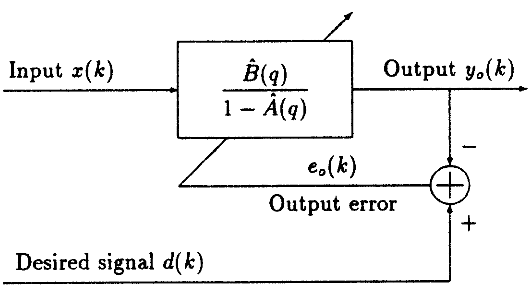

Fig. 2 Adaptive IIR Filter Based on the Output Error Formulation.

The output-error adaptive filter is described in Fig. 2. The output-error adaptive filter that approximates the unknown system is governed by the difference equation

$$y_0(k)=\sum_{i=1}^{N}\hat{a}_i(k)y_o(k-i)+\sum_{j=0}^{M}\hat{b}_j(k)x(k-j) \quad (2)$$

and the output error is given by

$$e_o(k)=d(k)-y_o(k) \quad (3)$$

where the desired signal *d(k)* is given by

$$\begin{aligned} d(k) &= y(k)+v(k) && (4)\\ &= \sum_{i=1}^{N} a_i y(k-i)+\sum_{j=0}^{M} b_j x(k-j)+v(k) && (5)\end{aligned}$$

where $v(k)$ is additive noise.

A coefficient adaptation rule that minimizes the mean square output error (MSOE)

$$\begin{aligned} J_o(k) &= E\{\, e_0^2(k)\,\}\\ &= E\{\, d(k)-y_o(K)\,\}^2 \end{aligned} \quad (6)$$

is obtained by differentiating the MSOE with respect to $\hat{\theta}(k)$. Define

$$\hat{\theta}(k)=[\ \hat{a}_1(k)\ \cdots\ \hat{a}_N(k)\ \ \hat{b}_o(k)\ \cdots\ \hat{b}_M(k)]^T \quad (7)$$

as the filter coefficient vector. Then the adaptation rule has a form

$$\hat{\theta}(k+1)=\hat{\theta}(k)+\frac{1}{2}\mu[\ -\hat{\nabla}(k)] \quad (8)$$

where $\nabla(k)$ is the instantaneous estimate of

$$\nabla(k)=\frac{\partial J_o(k)}{\partial\theta(k)}=2E\left\{e_o(k)\frac{\partial y_o(k)}{\partial\theta(k)}\right\}, \quad (9)$$

where

$$\frac{\partial y_o(k)}{\partial \theta(k)} = \left[\frac{\partial y_o(k)}{\partial \hat{a}_1(k)} \cdots \frac{\partial y_o(k)}{\partial \hat{a}_N(k)} \quad \frac{\partial y_o(k)}{\partial b_o(k)} \cdots \frac{\partial y_o(k)}{\partial b_M(k)} \right]^T \quad (10)$$

Note that the difference equation (2) has a recursive form in that the output $y_o(k)$ is fed back to the input. Hence $y_o(k)$ is a nonlinear function of the filter coefficient vector $\theta(k)$ and each component in (10) should be determined recursively. From (2), we have

$$\frac{\partial y_o(k)}{\partial \hat{a}_n(k)} = y_o(k-n) + \sum_{i=1}^{N} \hat{a}_i(k) \frac{\partial y_o(k-i)}{\partial \hat{a}_n(k)}, \quad 1 \le n \le N \quad (11)$$

and

$$\frac{\partial y_o(k)}{\partial \hat{b}_m(k)} = x(k-m) + \sum_{i=1}^{N} \hat{a}_i(k) \frac{\partial y_o(k-i)}{\partial \hat{b}_m(k)}, \quad o \le m \le M. \quad (12)$$

That is, post-filtering by the feedback coefficients $\hat{a}_i(k)$'s is required to obtain the gradient in(10). Equations (8), (9), (10), (11), and (12) constitute the adaptation rule of White's algorithm.[5]

Feintuch's LMS algorithm[3] does not have the second terms in the right hand sides of (11) and (12). Hence, the algorithm does not minimize the MSOE at all.

White's algorithm is quite complicated. This is due to the nonlinearity of the output $y_o(k)$ with respect to the filter coefficient vector $\hat{\theta}(k)$. Hence adaptation rules for output-error algorithms are generally complicated than their equation error counterparts. The use of a simplified gradient reduces the computational burden considerably.[4] This simplification is achieved by assuming that

$$\hat{\theta}(k) \approx \hat{\theta}(k-1) \approx \cdots \approx \hat{\theta}(k-N) \quad (13)$$

for μ sufficiently small and

$$y_o(k-1) \approx y_o(k-i), \quad \text{for } 2 \le i \le N$$
$$x_o(k) \approx x_o(k-j), \quad \text{for } 1 \le j \le M.$$

However, it is still complicated than its equation error counterpart.

The MSOE surface is not quadratic in $\hat{\theta}(k)$ due to the nonlinearity of $y_o(k)$. Since a trajectory of $\hat{\theta}(k)$ is perpendicular to contours of constant MSOE, the rate of convergence is not monotonic decreasing in trajectory. Hence for a given step size μ, the rate may vary in a wide range according to an initial values $\hat{\theta}(0)$ and the shape of the contour which depends on the autocorrelation matrix of the input data. The rate is very large if the slope of the MSOE is very steep and instability can occur. Thus if μ is chosen to be small enough not to cause instability, then the rate would be very slow where the MSOE surface is not steep. This makes output-error algorithms converge slowly.

The MSOE surface may have local minima besides a global minimum in some cases. If the order of poles and zeros in the adaptive filter are larger than or equal to those of the unknown system(overparameterized or exact cases), the global minimum corresponds to an aymptotically unbiased estimate of a true coefficient vector. If the order of poles and zeros are smaller than those of the unknown system(a reduced order case), the filter coefficient at the global minimum will be the best approximate of the true coefficient vector θ. To date necessary and sufficient conditions for the MSOE surface to have global minimum without local minima are not fully understood.[14]

2. Equation Error Formulation

Fig. 3 shows the adaptive IIR filter based on the equation error formulation. It is also called a pole-zero adaptive filter since it consists of two adaptive FIR filters, one for poles and the other for zeros. The equation-error adaptive filter is governed by the difference equation

$$\begin{aligned} y_e(k) &= \sum_{i=1}^{N} \hat{a}_i(k) d(k-i) \\ &+ \sum_{j=0}^{M} \hat{b}_j(k) x(k-1) \end{aligned} \qquad (14)$$

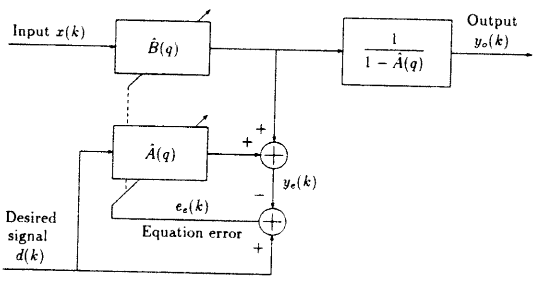

Fig. 3 Adaptive IIR Filter Based on the Equation Error Formulation.

which is nonrecursive. By defining

$$\emptyset_e(k) = [\ d(k-1) \cdots d(k-N) \ \ x(k) \cdots x(k-M)]^T, \qquad (15)$$

the output $y_e(k)$ in (14) can be rewritten in a linear regression form as

$$y_e(k) = \hat{\theta}^T(k) \emptyset_e(k). \qquad (16)$$

The equation error is then given by

$$\begin{aligned} e_e(k) &= d(k) - y_e(k) \\ &= d(k) - \hat{\theta}^T(k) \emptyset_e(k) \end{aligned} \qquad (17)$$

and the mean square equation error (MSEE) to be minimized is given by

$$\begin{aligned} J_e(k) &= E\ \{ e_e^2(k) \} \\ &= E\ \{ d(k) - y_e(k) \}^2. \end{aligned} \qquad (18)$$

Note that the output $y_e(k)$ is a linear function of the filter coefficients since $d(k)$ and $x(k)$ are independent of the filter coefficients. Hence the equation error is also linear in the filter coefficients and the MSEE is quadratic in the filter coefficients. The MSEE surface has a global minimum without local minima.

As in the previous subsection, the coefficient adaptation rule can be found to be

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \mu e_e(k) \emptyset_e(k). \qquad (19)$$

Equations (17) and (19) constitute the LMS algorithm for the IIR case. Throughout this paper, this algorithm will be referred to as the LMS with equation error(LMS-EE) algorithm.

The LMS-EE algorithm is in fact al direct extension of the LMS algorithm. Thus many properties are common to both algorithms. The LMS-EE algorithm converges faster than its output-error counterpart due to the quadratic nature of the MSEE surface. Further, it has the hyper-stability feature in that the adaptive filter dan remain stable even though poles lie outside of the unit circle over a part of the time. Following example explains the hyper-stability feature of the LMS-EE algorithm.

Consider a first-order recursive filter

$$y_o(k) = \hat{a}_1(k) y_o(k-1) + x(k). \qquad (20)$$

Note that (20) describes the adaptive filter in the steady state condition with $\hat{b}_o(k) = 1$. Suppose that $\hat{a}_1(k)$ changes perodically with

$$\hat{a}_1(k) = \begin{cases} \hat{a}_1, & \text{if } k = 2l \\ \hat{a}_2, & \text{if } k = 2l+1 \end{cases} \qquad (21)$$

If $x(k) = 0$, for $k \neq 0$, and $x(0) = 1$, then

$$y_o(k) = \begin{cases} \hat{a}_1^k \hat{a}_2^k, & \text{if } k = 2l \\ \hat{a}_2^k \hat{a}_2^{k+1}, & \text{if } k = 2l+1 \end{cases} \qquad (22)$$

The stability condition of the flter is then $|\hat{a}_1 \hat{a}_2| < 1$ instead of $|\hat{a}_1| < 1$ and $|\hat{a}_2| < 1$. Therefore, if $\hat{a}_1 = 1.1$ and $\hat{a}_2 = 0.9$, then the filter is stable even though the pole is outside the unit circle over the half of the time. Therefore, an algorithm can remain stable even though instantaneous poles lie outside the unit circle only if the algorithm can pull them back inside the unit circle effectively. The main reason for this self-stabilizing feature is the nonrecursive nature of the equation-error adaptive filter.

Despite these advantages advantages of the equation error formulation, it has a major drawback. The filter coefficient vector $\hat{\theta}(k)$ is biased in general in the presence of additive noise.

III. BIAS IN THE EQUATION ERROR FORMULATION

The bias in the equation error formulation is analyzed by Mendel assuming that the algorithm converges.[1] Actually, Mendel considered a special FIR case using the decorrelation delay to meet independency between $\varnothing_e(k)$ and $\varnothing_e(k-1)$. Due to this delay, the analysis cannot be applied to the IIR LMS algorithm. In the following, the bias analysis by Mendel is extended to the LMS algorithm by eliminating the need of the decorrelaion delay.

Begin with the LMS-EE algorithm given in (19) as

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \mu e_e(k)\varnothing_e(k) \qquad (23)$$

where

$$\begin{aligned} e_e(k) &= d(k) - y_e(k) \\ &= d(k) - \hat{\theta}^T(k)\varnothing_e \end{aligned} \qquad (24)$$

The desired signal in (5) can be rewritten as

$$d(k) = \varnothing_y^T(k)\theta + v(k) \qquad (25)$$

where

$$\theta = [\ a_1 \cdots a_N\ b_o \cdots b_M]^T \qquad (26)$$

$$\varnothing_y(k) = [\ y(k-1) \cdots y(k-N)\ x(k) \cdots x(k-M)]^T. \qquad (27)$$

Define

$$\varnothing_v(k) = [\ v(k-1) \cdots v(k-N) 0 \cdots 0]^T. (28)$$

Then using(27), (15) can be decomposed into

$$\varnothing_e(k) = \varnothing_y(k) + \varnothing_v(k). \qquad (29)$$

Substituting(24) and (29) into (23) yields

$$\begin{aligned} \hat{\theta}(k+1) = \ & \hat{\theta}(k) \\ & + \mu[\ \varnothing_y(k) + \varnothing_v(k)]\ \varnothing_y^T(k)\theta \\ & + \mu[\ \varnothing_y(k) + \varnothing_v(k)]\ v(k) \\ & - \mu[\ \varnothing_y(k) + \varnothing_v(k)] \\ & [\ \varnothing_y(k) + \varnothing_v(k)]^T \hat{\theta}(k) \end{aligned} \qquad (30)$$

In order to make the analysis as simple as possible, the followings are assumed:[1]

(i) $x(k)$ and $v(k)$ are statistically independent.

(ii) v(k) has a zero mean with finite variance.

(iii) $\varnothing_e(k)$ and $\hat{\theta}(k)$ are independent and $E\{\ \varnothing_e(k)\varnothing_e^T(k) \mid \hat{\theta}(k)\ \}$ is finite and positive definite.

Note that assumption (iii) above requires statistical independencd of $\varnothing_e(k)$ and $\hat{\theta}(k)$.[1] This requires that $\varnothing_e(k)$ and $\hat{\theta}(k)$ do not contain any common element, which is clearly not valid. In,[1] Mendel used the decorrelation delay to meet this assumption. However, it should be noticed that assumption (iii) is invalid despite the use of decorrelation delay since $y(k-1)$ is a function of $\hat{\theta}(k-1)$ and $y(k-2)$, and $y(k-2)$ is in turn a function of $\theta(k-2)$ and $y(k-3)$ and so on. Assumption (iii), however, works quite well for the unit decorrelation delay, which is the case of the LMS-EE algorithm under investigation. In the analysis of adaptive FIR filtering algorithms, assumption (iii) is accepted as a consequence of socalled "fundamental assumption".[18]

Taking expectation of both sides of (30) and using assumption (iii) gives

$$\begin{aligned} E[\hat{\theta}(k+1)] = & E[\hat{\theta}(k)] \\ & +\mu E\{[\varnothing_y(k)+\varnothing_v(k)] \\ & \varnothing_y^T(k)\}\theta \\ & +\mu E\{[\varnothing_y(k)+\varnothing_v(k)] \\ & v(k)\} \\ & -\mu E\{[\varnothing_y(k)+\varnothing_v(k)] \\ & [\varnothing_y(k)+\varnothing_v(k)]^T\} \\ & E\{\hat{\theta}(k)\}. \end{aligned} \tag{31}$$

Simplify (31) using assumptions (i) and (ii) to obtain

$$\begin{aligned} E[\hat{\theta}(k+1)] = & E[\hat{\theta}(k)] \\ & +\mu E\{[\varnothing_y(k)\varnothing_y^T(k)]\}\theta \\ & +\mu E\{\varnothing_v(k)v(k)\}\theta \\ & -\mu\{E[\varnothing_y(k)\varnothing_y^T(k)] \\ & +E[\varnothing_v(k)\varnothing_v^T(k)]\} \\ & E[\hat{\theta}(k)]. \end{aligned} \tag{32}$$

Assume that the algorithm converges. Then, at steady state,

$$\lim_{k\to\infty} E[\hat{\theta}(k+1)] = \lim_{k\to\infty} E[\hat{\theta}(k)]. \tag{33}$$

Therefore, taking limit of both sides of (32) and using (33) gives

$$\begin{aligned} \lim_{k\to\infty} E[\hat{\theta}(k)] = & \{E[\varnothing_y(k)\varnothing_y^T(k)] \\ & +E[\varnothing_v(k)\varnothing_v^T(k)]\}^{-1} \\ & \{E[\varnothing_y(k)\varnothing_y^T(k)]\theta \\ & +E[\varnothing_v(k)v(k)]\}. \end{aligned} \tag{34}$$

Equation (34) shows the bias in the coefficient estimate in the presence of additive noise $v(k)$. The bias expression is very complicated. However, if $v(k)$ is white

$$E\{\phi_v(k)v(k)\} = 0 \tag{35}$$

and (34) is simplified to

$$\begin{aligned} \lim_{k\to\infty} E[\hat{\theta}(k)] = & \{E[\phi_y(k)\phi_y^T(k)] \\ & +E[\phi_v(k)\phi_v^T(k)]\}^{-1} \\ & E[\phi_y(k)\phi_y^T(k)]\theta. \end{aligned} \tag{36}$$

The bias increases as noise power increases. Further, the bias is a function of the true coefficient θ. Note that only first N components of $\phi_v(k)$ are nonzero. Hence the bias appears in the feedback coefficients $\hat{a}_i(k)$'s. This propagates to the feedforward coefficients $\hat{b}_j(k)$'s since $\hat{a}_i(k)$'s and $\hat{b}_j(k)$'s are coupled through correlation between $y(k-i)$) and $x(k-j)$ where $i \le j$.

Ⅳ. BIAS ELIMINATION FOR THE EQUATION ERROR FORMULATION

From (36), it is clear that the bias can be removed if the term $E\{\phi_v(k)\phi_v^T(k)\}$ is cancelled somehow. There has been several afforts to cancel this term.[1,9]. In,[1] Mendel devised the unbiased algorithm with the adaptation rule

$$\begin{aligned} \hat{\theta}(k+1) = & [I+\mu R_v]\hat{\theta}(k) \\ & +\mu e_e(k)\phi_e(k) \end{aligned} \tag{37}$$

where $R_v = E\{\phi_v(k)\phi_v^T(k)\}$. A variant of this algorithm called the γ-LMS algorithm was suggested by Frost, and it has been used by Treichler to estimate the frequency of a noisy sinusoid corrupted by white noise.[9] These efforts, however, resort to a priori knowledge of the noise variance, which is not available in real situations. Further, if an estimate of R_v is used, then the accuracy of the filter coefficient would depend strongly on the accuracy of the estimate of R_v.

Although previous research efforts have not ruled out the requirement of the prespecification of the noise variance, those suggest one clear path to bias elimination for the equation error formulation. What we need now is to cancel the noise term R_v in an adaptive fashion without the prespecification of the noise variance. This goal can be achieved by comparing both equation error and output error formulations discussed in the previous section noting that the output error formulation provides an unbiased coefficient estimate.

1. Bias Elimination Technique

Define

$$1-A(q) = 1-a_1q^{-1}-a_2q^{-2}-\cdots -aNq^{-N} \quad (38)$$

$$1-\hat{A}(q) = 1-\hat{a}(k)q^{-1}-\hat{a}_2(k)q^{-2}-\cdots -\hat{a} \quad (39)$$

$$B(q) = b_o+b_1q^{-1}+b_2q^{-2}+\cdots +b_Mq^{-M} \quad (40)$$

$$\hat{B}(q) = \hat{b}_o(k)+\hat{b}_1(k)q^{-1}+\hat{b}_2(k)q^{-2} +\cdots+\hat{b}_M(k)q^{-M} \quad (41)$$

where q^{-1} is a delay operator for a time varying process. Then the output error $e_0(k)$ in (3) is given by

$$e_o(k) = d(k)-y_o(k) \quad (42)$$

$$= d(k)-\hat{A}(q)y_o(k) -\hat{B}(q)x(k) \quad (43)$$

and the equation error in (17) by

$$e_e(k) = d(k)-y_e(k) \quad (44)$$

$$= d(k)-\hat{A}(q)d(k) -\hat{B}(q)x(k). \quad (45)$$

Thus

$$e_e(k)-e_o(k) = -\hat{A}(q)[\ d(k)-y_o(k)] = -\hat{A}(q)e_o(k) \quad (46)$$

or

$$e_e(k) = [\ 1-\hat{A}(q)]\ e_o(k) \quad (47)$$

$$= [\ 1-\hat{A}(q)]\ [\ d(k)-y_o(k)] \quad (48)$$

This is the well-known relationship between the equation error and the output error.

In adaptive filtering, the ultimate goal is to minimize the MSOE. However, the equation error formulation takes a round-about approach. As a result, in the equation error formulation, the adaptive filter should minimize the MSEE, which is a function of both the output error and the feedback coefficients. In other word, the equation-error adaptive filter should minimize the filtered version of the MSOE. This filtering of the noisy process $d(k)$ by the feedback polynomial $[1-\hat{A}(q)]$ produces bias. This facts suggests one bias elimination technique: counterbalancing of the polynomial $[1-\hat{A}(q)]$.

Counterbalancing may be achieved by AR filtering of the output error $e_0(k)$ using a proper polynomial. Clearly, the ideal case is to use $[1-\hat{A}(q)]$. The resulting algorithm will be very similar to the output-error algorithm by Feintuch [3], which may not converge to either a gloval minimum or a local minimum.[15] Hence, MA filtering (or smoothing) is desirable.

Define

$$e(k) = [\ 1+C(q)]\ e_e(k) \quad (49)$$

where the polynomial $1+C(q)$ is chosen as

$$\frac{1}{1-\hat{A}(q)} \approx 1+C(q) \quad (50)$$

$$= 1+c_1(k)q^{-1}+C_2(k)q^{-2} +\cdots+c_L(k)q^{-L}. \quad (51)$$

for some L. Note that, for a second order $\hat{A}(q)$,

$$\frac{1}{1-\hat{A}(z^{-1})} = \frac{1}{1-\hat{a}_1(k)z^{-1}-\hat{a}_2(k)z^{-2}} = 1+\hat{a}_1(k)z^{-1} +[\ \hat{a}_2(k)+\hat{a}_1^2(k)]\ z^{-2}+\cdots \quad (52)$$

and we can take

$$1+C(q) = 1+\hat{a}_1(k)q^{-1} + [\ \hat{a}_1(k)+\hat{a}_1^2(k)]\ q^{-2} \quad (53)$$

using a second order approximation. Similarly, for a third order [$\hat{A}(q)$], we may have

$$1+C(q) = 1+\hat{a}_1(k)q^{-1} + [\ \hat{a}_2(k)+\hat{a}_1^2(k)]\ q^{-2} + [\ \hat{a}_3(k)+2\hat{a}_1(k)\hat{a}_2(k) + \hat{a}_1^3(k)]\ q^{-3} \quad (54)$$

using a third order approximation.

From the above idea, the LMS-EE algorithm is modified to have the adaptation rule

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \mu e(k)\phi_e(k). \quad (55)$$

The algorithm with the modified adaptation rule (55) will be referred to as the LMS with smoothed equation error (LMS-SEE) algorithm. Fig. 4 describes the adaptive filter based on the smoothed equation error formulation.

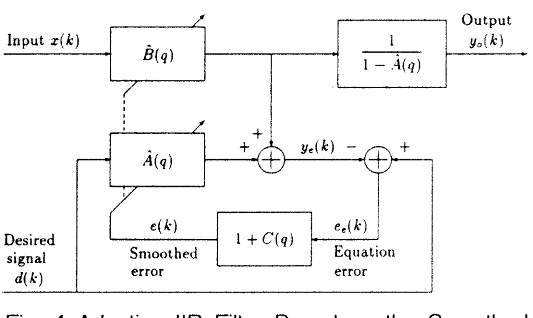

Fig. 4 Adaptive IIR Filter Based on the Smoothed Equation Error Formulation.

V. SIMULATIONS

Refer back to the system identification configuration in Fig. 1. The additive noise $v(k)$ is assumed to be white and is uncorrelated with the input $x(k)$. Both colored or white noise are used as an the input $x(k)$. As mentioned in previous sections, a stability monitoring and projection scheme is required to guarantee stability. However, if the step size is taken small enough, the adaptive filter would be stable without the scheme. All simulations in this section are performed without a stability monitoring and projection scheme.

The problem is to identify the unknown system

$$H_d(q) = \frac{0.8-1.6q^{-1}}{(1-1.5q^{-1}+0.8125q^{-2})} \quad (56)$$

using the adaptive filter

$$H_a(q) = \frac{b_o(k)+b_1(k)q^{-1}}{(1-a_1(k)q^{-1}-a_2(k)q^{-2})}\ . \quad (57)$$

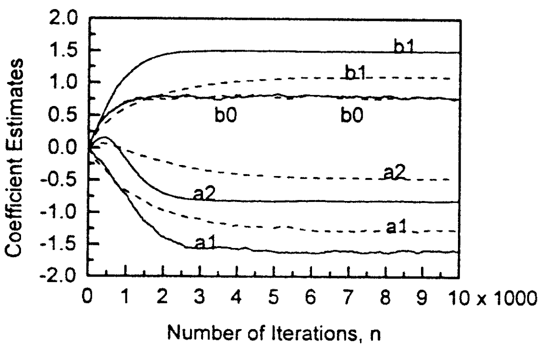

Fig. 5 Coefficient Convergence of the LMS-EE (dashed) and LMS-SEE(Solid) Algorithms for SNR=0 dB.

with the white noise input $x(k)$ at SNR = 0 dB. Both the LMS-EE and LMS-SEE algorithms are simulated. Step size parameters used are μ_F = 0.002 for feedforward coef- ficients b_j's and μ_B = 0.0002 for feedback coefficients a_i's for all algorithms. It is interesting to note that μ_F is chosen much larger than μ_B. This is because convergence of the feedfoward coefficients tends to stabilize the convegence process of the feedback coefficents. Fig 5 shows convergence of coefficients for the LMS-EE and LMS-SEE algorithms averaging 10 independent runs. Clearly, all coefficient estimates are biased except b_0. The bias is caused by filtering of noisy process $d(k)$ which is applied to the feedback section of the adaptive filter. Hence, the bias in the feedback coefficents is directly caused by the noise process. This bias in turn affects feedforward coefficients via coupling between $y(k-1)$ and $x(k-j)$ for $i < j$. The coefficient estimate for b_0 is unbiased since $x(k)$ is uncorrelated with $y(k-i)$ for $i \geq 1$. The LMS-SEE algorithem, however, provides unbiased coefficient estimates for all coefficients as expected.

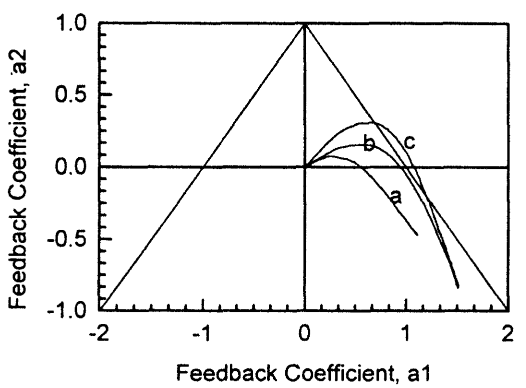

Fig. 6 Convergence of the Feedback Coefficients (a) without Error Smoothing (the LMS-EE algorithm), (b) with Time Varying $1+C(q)$, and (c) with Constant $1+C(q)$ for SNR=0 dB.

In Section III, convergence of the LMS-SEE algorithm is conjected in the case of the constant error smoothing polynomial $1 + C(q)$. The effect of $1 + C(q)$ can be seen in Fig. 6. Trajectory (a) is for the LMS-EE algorithm, (b) for the LMS-SEE algorithm with the time varying $1 + C(q)$, and (c) with the constant $1 + C(q)$. Note that trajectory (b) lies between trajectories (a) and (c). Loosely speaking, using the constant $1 + C(q)$ has and effect of using larger step sizes especially in the transient period. Therefore, trajectory (c) is pushed outside the stability region. On the other hand, if $1 + C(q)$ is time varying, the effect of $1 + C(q)$ may be negligible initially since $C(q)$ is small in magnitude. Thus trajectory (b) follows (a) closely in the early part of the trajectory. As coefficients are updated, the magnitude of $C(q)$ grows and the effect of $1 + C(q)$ becomes evident. Thus trajectory (b) approaches trajectory (c) as filter coefficients approaches true values. Therefore, the trajectory with time varying $1 + C(q)$ lies between those without and with the constant $1 + C(q)$. This is valid for step size small enough. Further, it is noted that the part of the trajectory (c) is outside the stability region and the algorithm still converges. This is the hyper-stability feature of the equation error formulation. However, if the trajectory is far outside the stability region, the algorithm will not be stable any longer. The trajectory of the LMS-SEE algorithm moves inside the stability region as smaller step size parameters are used.

Simulations in this section shows that the LMS-SEE algorithm eliminated the bias in the coefficient estimates if aditive noise is white. Furthermore, it is confirmed that the LMS -SEE algorithm retains advantages of the equation error formulation such as faster convergence speed, and a self stability featu- res. Hence, the LMS-SEE algorithm makes the equation error formulation viable even in the presence of the additive white noise.

VI. CONCLUSIONS

This paper studied the bias problem in the IIR LMS algorithm. The IIR LMS algorithm produces biased estimates for all coefficients except the zeroth order MA coefficient if there is additive noise in the input. An algorithm is devised to get unbiased coefficient estimates when the additive noise is white. Simulations are performed to demonstrated the performance of the new algorithm. The new algorithm retaines properties such as the hyper-stability feature, global convergence, and fast convergence speed while provides unbiased coefficient estimates.

Acknowlegement

This study was financially supported by a central research fund in 1995 from PaiChai university.

REFERENCES

1. J.M.Mendel, *Discrete Techniques of Parameter Estimation: The Equation Error Formulation.* Marcel Dekker, New York, 1973.
2. R.P.Gooch, "Adaptive pole-zero filtering: the equation error approach,"Ph.D. diss. Stanford University, 1983.
3. P.L.Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE,* vol. 64, no. 11, pp. 1622 -1624, Nov. 1976
4. S.Horvath, Jr., "A new adaptive recursive LMS filter," in *Digital Signal Processing,* V.Cappellini and A.G. Constantinides, eds. Academic Press, New York, 1980, pp. 21-26.
5. S.A.White, "An adaptive recursive digital f ilter," in *Proc. 9th Asiloma Conf. Circuits, Systems, Computers,* Pacific Grove, California, Nov. 1975, pp. 21-25.
6. J.R. Treichler, M.G. Larimore, and C.R. Johnson, Jr., "Simple adaptive IIR filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Proc.,* Tulsa, Oklahoma, Apr. 1978, pp. 118 -112.
7. J.R.Treichler, C.R.Johnson, Jr., and M.G. Larimore, *Theory and Design of Adaptive Filters.* Wiley, New York, 1987.
8. J.R.Treichler, "Adaptive algorithm for infinite impulse response filters," in *Adaptive Filters,* C.F.N. Cowen and P.M. Grant, eds. Prentice-Hall, Englewood Cliffs, N.J., 1985, ch. 4, pp. 60-90.
9. J.R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," in *IEEE Trans. Acoust., Speech, Signal. Proc.,* vol. ASSP-27, pp. 53-62, Feb. 1979.
10. John J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Magazine,* pp. 4-21, April 1989.
11. S.D. Stearns, "Error surface of recursive adaptive filters," *IEEE Trans, Circuits Systems,* vol. CAS-28, no. 6, pp. 603-606, June 1981.
12. T.Söderström and P.Stoica, "Some properties of the output error method," *Automatica,* vol. 18, pp. 93-99, Jan. 1982
13. H. Fan and W.K. Jenkins, "A new adaptive IIR filter," *IEEE Trans. Circuits Syst.,* vol. CAS-33, pp. 939-947, Oct. 1986.
14. H. Fan and M. Nayeri, "On error surface of sufficient order adaptive IIR filters: proofs and counterexamples to a unimodality conjecture," *IEEE Trans. Acoust., Speech, Sig. Proc.,* vol.37, no.9, pp. 1436-1442, Sept. 1989.
16. C.R.Johnson, Jr. and A.L.Hamm, "Debiasing equation error frequency detection of a noisy sinusoid without noise variance prespecifica- tion," *IEEE Trans. Acoust., Speech, Sig. Proc.,* vol. ASSP-27, no. 6, pp. 654-656, Dec. 1979.
17. G.Long, D.Shewd, and D.D.Falconer, "Study of a pole-zero adaptive echo canceller," IEEE Trans. Circuits, Ssytems, vol. CAS-34, no.7, pp. 765-769, July 1987.