A SUBOPTIMAL SOLUTION TO IIR ADAPTIVE FILTERING USING AN EXTENDED LM ALGORITHM

EMMANUEL FERNANDEZ-GAUCHERAND AND JOHN Y. CHEUNG
School of Electrical Engineering and Computer Science,
University of Oklahoma, Norman, OK

ABSTRACT

In this paper, we present a simple updating algorithm for recursive adaptive filters. The approach used is indirect in that the error between the desired signal and the output of the adaptive recursive filter (the output error) is related to the error between the desired signal and the output of a two input-one output finite impulse response filter. The minimization of the former error is sought by minimizing the latter using a more general form of the least mean squared algorithm. Convergence analysis of the obtained algorithm shows that global minimization of the output error can be attained, however only suboptimal minimization will be possible in general.

INTRODUCTION

Widrow et. al. [1] - [5] have made comprehensive contributions in FIR adaptive filtering. A key element in their analyses is the linear combiner. Under stationary conditions, the form of the widely used mean-squared error (MSE) performance criterion is quadratic in the combiner's weight vector. In general, the mean-(2K)th error, K being a natural number, is a convex function of the mentioned weight vector and therefore cannot have local minima [5]. By using a local estimate for the expectation operator in the above type of error criterion, Widrow and Hoff [1] formulated their well known least mean-squared (LMS) algorithm to seek global minimization iteratively.

The use of an IIR structure could possibly reduce complexity and improve performance when compared to all-zero adaptive filters, in those applications where a large number of MA coefficients are required for the latter in order to obtain an acceptable degree of efficiency. Such is the case in those situations where the desired signal, d(k), is best described by an ARMA process on the input u(k), as in (i) a general system identification mode, (ii) deconvolution applications, e.g., channel equalization [6], [7] and inverse control [8], and (iii) in certain type of model following problems [9].

For algorithm development, two main approaches have been employed, namely descent methods [10] - [18], and methods based on hyperstability theory as applied to output error identifiers [19] - [25]. With the descent methods, only local minimization can be guaranteed [14], [16] due to the generally multimodal characteristics of the MSE surfaces being descended [17]. Furthermore, the computational load is heavily increased by the need of on-line stability monitoring and projection techniques for the auxiliary ARMA process approximating the gradients needed for the descent, as commented by Johnson [26]. On the other hand, the hyperstable family of adaptive recursive filters emerged as a result of the use of system identification formulations, for which a vast literature rich in theoretical results is available, laying down a pattern which has become the current trend [26]. However, a system identification mode of operation be regarded as only a less general, more structured problem than that general adaptive filtering, and the translation of results from the to the latter carries with it several yet unanswered questions [26]. rticularly, in order to guarantee convergence for the hyperstable lters, a priori knowledge of the characteristic polynomial of the process erating $d(k)$ is required to satisfy a strictly positive real (SPR) ndition. This condition has been a severely limiting factor to the use this family of algorithms. Notwithstanding these advances for the perstable family, an algorithm for IIR adaptive filtering which exhibits w computational complexity, predictable characteristics, and simple tability conditions is yet an unfulfilled need. The algorithm that we sent next tries to meet these requirements.

A TWO INPUT - ONE OUTPUT FIR SCHEME

For clarity sake, let us focus initially on a deterministic system dentification scheme, where the desired signal, $d(k)$, is assumed to be easonably modeled by the general ARMA process given below

$$d(k) = \sum_{j=0}^{M_d} b_j\, u(k-j) + \sum_{i=1}^{N_d} a_i\, d(k-i) \tag{1}$$

or using delay operator notation

$$d(k) = B_{Md}^{0}(q^{-1})\, u(k) + A_{Nd}^{1}(q^{-1})\, d(k) \tag{2}$$

where $u(k)$ is an input signal to be applied to the identifier too, and

$$B_m^n(q^{-1}) = b_n\, q^{-n} + b_{n+1}\, q^{-(n+1)} + \ldots + b_m\, q^{-m}, \qquad n \leq m \tag{3}$$

where q^{-p} denotes a delay operation of p samples, i.e., $q^{-p}\, u(k) = u(k-p)$. Other delay operators are similarly defined. Neither the coefficients, b_j's and a_i's, nor M_d and N_d are assumed to be known.

Let $y_u(k)$ denote the response of a FIR filter of order M to the input $u(k)$, that is

$$y_u(k) = \sum_{j=0}^{M} \hat{b}_j\, u(k-j) \tag{4}$$

Now, let $y_d(k)$ be the output of another FIR filter of order (N-1) having a one step delayed version of the desired signal as its input, as shown below

$$y_d(k) = \sum_{p=0}^{N-1} \hat{a}_{p+1}\, d((k-1) - p) \quad = \sum_{i=1}^{N} \hat{a}_i\, d(k-i) \tag{5}$$

Augmenting the output signal in (4) by (5), we define

$$\begin{aligned} y_a(k) &= y_u(k) + y_d(k) \\ &= \sum_{j=0}^{M} \hat{b}_j\, u(k-j) + \sum_{i=1}^{N} \hat{a}_i\, d(k-i) \end{aligned} \tag{6}$$

or in operator form

$$y_a(k) = \hat{B}_M^{0}(q^{-1})\ u(k) + \hat{A}_n(q^{-1})\ d(k) \tag{7}$$

The error of $y_a(k)$ with respect to $d(k)$ is given by

$$\begin{aligned} e(k) &= d(k) - y_a(k) \\ &= d(k) - \hat{B}_M^{0}(q^{-1})\ u(k) - \hat{A}_N^{1}(q^{-1})\ d(k) \\ &= [1 - \hat{A}_N^{1}(q^{-1})]\ d(k) - \hat{B}_M^{0}(q^{-1})\ u(k) \end{aligned} \tag{8}$$

Note that (8) is linear in both $\hat{A}_N^{1}(q^{-1})$ and $\hat{B}_M^{0}(q^{-1})$, thus (7) can be viewed as a two input, $u(k)$ and $d(k-1)$, one output, $y_a(k)$, FIR filter. Minimizing (8), say in a MSE sense, and given that (7) has enough complexity to span all the modes of (1), that is if $M \geq M_d$ and $N \geq N_d$, then (1) can be approximated by (7). Moreover, the coefficients of (7) will constitute estimates of those of (1), where the latter processes is viewed as having $(M - M_d)$ and $(N - N_d)$ higher order MA and AR coefficients, respectively, equal to zero.

An important fact to be realized is that (8) will be linear in the coefficients of (7) regardless of the nature of $d(k)$, and regardless whether $M \geq M_d$ and $N \geq N_d$. Hence the form of (1) is not a necessary assumption for the process $d(k)$.

Another way to realize the FIR characteristic of (7) is to compute $y_a(k)$ by means of a linear combiner. Defining the vector of parameters as

$$\begin{aligned} \underline{\hat{\Theta}} &= [\ \underline{\hat{B}}\ |\ \underline{\hat{A}}\ ]^T \\ &= [\hat{b}_0\ \hat{b}_1\ .\ .\ .\ \hat{b}_M\ |\ \hat{a}_1\ \hat{a}_2\ .\ .\ .\ \hat{a}_N]^T \end{aligned} \tag{9}$$

and the information vector as

$$\begin{aligned} \underline{X}(k) &= [\ \underline{U}(k)\ |\ \underline{D}(k-1)]^T \\ &= [u(k)\ u(k-1)\ .\ .\ .\ u(k-M)\ |\ d(k-1)\ d(k-2)\ .\ .\ .\ d(k-N)]^T \end{aligned} \tag{10}$$

then (7) can be rewritten as

$$y_a(k) = \underline{\hat{\Theta}}^T\ \underline{X}(k) = \underline{X}^T(k)\ \underline{\hat{\Theta}} \tag{11}$$

The structure of (11) above can be viewed as an extrapolated version of the linear combiner used by Widrow et. al., where two tap delay lines are used, instead of one, having as inputs $d(k)$ and $u(k)$, to form the two blocks constituting (10).

Parallel to the single tap-delay line case, the optimal vector of coefficients, denoted $\underline{\hat{\Theta}}^*$, for any values of M and N and for general $d(k)$, is given by

$$\underline{\hat{\Theta}}^* = \{E[\underline{X}(k)\ \underline{X}^T(k)]\}^{-1}\ \{E[d(k)\ \underline{X}(k)]\} \tag{12}$$

Equation (7) is sometimes called the equation error in a purely system identification setting [26] where (1) holds, which is actually an abuse since equation error formulations cover a broader scope of identification problems besides this one [27]. The applicability of (7) to more general minimization problems, as commented above, is a commonly overlooked fact. In the following discussion, the IIR adaptive filtering problem is related to the two input - one output FIR filter scheme of (7), and an extension to the LMS algorithm is presented which is capable of seeking for the optimal set of parameters of (12) iteratively.

A CONCEPT OF ERROR BY KALMAN

Let f(k) be an arbitrary ARMA process as given below

$$f(k) = \sum_{j=0}^{M} \gamma_j \, u(k-j) + \sum_{i=1}^{N} \delta_i \, f(k-i) \tag{13}$$

or in operator notation

$$f(k) = T_M^0 (q^{-1}) \, u(k) + \Delta_N^1 (q^{-1}) \, f(k) \tag{14}$$

The time domain transfer function for the process of (14) is given by

$$\frac{f(k)}{u(k)} = \frac{T_M^0 (q^{-1})}{1 - \Delta_M^1 (q^{-1})} \tag{15}$$

Now, given a process d(·) that is arbitrarily correlated with u(·), the output error is defined as follows

$$e_1(k) = d(k) - f(k) = d(k) - \frac{T_M^0 (q^{-1})}{1 - \Delta^1{}_N (q^{-1})} u(k)$$

$$= \frac{[1 - \Delta^1{}_N(q^{-1})]d(k) - T_M^0(q^{-1})}{1 - \Delta_N^1 (q^{-1})} u(k) \tag{16}$$

Note that (15) and (16) share the same characteristic polynomial. Hence, minimization of $e_1(k)$ results in a nonlinear problem in the δ_i's coefficients. However, we can see that a related error, $e_2(k)$, defined as the numerator of (16)

$$e_2(k) = [1 - \Delta_N^1(q^{-1})]d(k) - T_M^0(q^{-1}) \, u(k) \tag{17}$$

is linear in both the γ_j's and δ_i's coefficients.

The minimization of (17) is a much more amenable task than that of (16), and we may use the former as an indirect approach to the latter. Clearly, if $e_2(k)$ can be made equal to zero, $e_1(k)$ will go to zero as well, since (16) results for passing $e_2(k)$ through an AR filter with denominator $1-\Delta_N^1(q^{-1})$. Similar filtering of a linear-in-the-coefficients error is found in current IIR adaptive filtering techniques [26]. This relation between (16) and (17) has been known for some years to the system identification community, specially as related to batch processing techniques [28], [29] and was first suggested implicitly by Kalman in 1958 [30]. Sometimes, (17) is called the equation error. For the sake of avoiding confusion, we will refer to it hereinafter as the Kalman error. Note again that no particular a priori assumptions about the nature of the correlation between d(·) and u(·) is needed.

THE EXTENDED LMS (ELMS) ALGORITHM

If batch processing techniques are used to minimize (17), say in a least squares sense, all that is needed is to collect sufficient measurements of $d(\cdot)$ and $u(\cdot)$ such that a nonsingular set of normal equations can be stated [29]. However, if the MSE iterative minimization of (17) is desired, for the adaptive version of (13), the problem that (17) is not directly measurable from access to $d(k)$ and $f(k)$ is encountered, but (17) is needed to estimate the performance of the minimization. Comparing (17) and (8) we realize that these two equations have identical form, hence (17) can be generated by use of an auxiliary adaptive filter of similar structure as that of (7). With this in mind, the method of steepest descent can be used, such that (12) can be sought iteratively.

Parallel to [1], the following approximation is done

$$E[(e^2(k))^2] \doteq (e_2(k))^2 \tag{18}$$

which is used in the computation of the needed gradients

$$\nabla_{\gamma_j} \{E[(e_2(k))^2]\} \doteq \nabla_{\gamma_j} \{(e_2(k))^2\}$$

$$= -2\, e_2(k)\, u(k)\, q^{-j}$$

$$= -2\, e_2(k)\, u(k-j)\ ; \quad j = 0, 1, \dots, M \tag{19a}$$

$$\nabla_{\delta_i} \{E[(e_2(k))^2]\} \doteq \nabla_{\delta_i} \{(e_2(k))^2\}$$

$$= -2\, e_2(k)\, d(k) q^{-i}$$

$$= -2\, e_2(k)\, d(k-i)\ ; \quad i = 1, 2, \dots, N \tag{19b}$$

Hence, the resultant updating equations are, with $\mu > 0$

$$\gamma_j(k+1) = \gamma_j(k) + 2\,\mu\, e_2(k)\, u(k-j)\ ; \quad j = 0, 1, \dots, M \tag{20.a}$$

$$\delta_i(k+1) = \delta_i(k) + 2\,\mu\, e_2(k)\, d(k-i)\ ; \quad i = 1, 2, \dots, N \tag{20.b}$$

such that

$$e_2(k) = d(k) - \underline{X}^T(k)\, \hat{\underline{\theta}}(k) \tag{21}$$

is minimized, with $\underline{X}(k)$ is as in (10) and

$$\hat{\underline{\theta}}(k) = [\gamma_0(k)\ .\ .\ .\ \gamma_M(k)\ |\ \delta_1(k)\ \dots\ \delta_N(k)\ ]^T \tag{22}$$

defining the output of the auxiliary two input-one output adaptive FIR filter. Note that a $\delta_0 = 0$ coefficient could have been included in (13), such that for $N = 0$ $f(k)$ becomes a MA on $u(k)$ alone and (20) reduces to (20.a), which is the classical LMS algorithm. Hence this latter algorithm is an interpolation of (20), and we will refer to (20) as a extended LMS (ELMS) algorithm. The structure of an IIR adaptive filter using (20) to tune the coefficents of (13) such that MSE minimization of (17) is sought is shown in Fig. 1.

Next, we examine the convergence properties of the ELMS algorithm studying the characteristics of the indirect MSE minimization of (16) through the minimization of (17) as described above.

CONVERGENCE ANALYSIS FOR THE ELMS ALGORITHM

As mentioned before, even though the system identification is only a particular mode of operation for an adaptive filter, it is advantageous to analyze the behaviour of the filter under these conditions so that theoretical results can be drawn from the systems identification literature. In this spirit, for the convergence analysis of the ELMS algorithm, the approach to be followed here is to show that it can be cast into a more general equation error parameter estimation formulation [27], hence allowing the interpolation of the results obtained for this broader problem.

Equation error parameter estimation

The equation error formulation of parameter estimation is amply documented in [27]. The system under identification is assumed to be realistically modeled by the linear-in-the-parameters relation

$$y(k) = \underline{\theta}^T \, \underline{X}(k) = \underline{X}^T(k) \, \underline{\theta} \tag{23}$$

between the input information vector $X(k)$ and the scalar output $y(k)$, where $\underline{\theta}$ is a vector of parameters. This formulation has been extended to the multi-output case by Johnson [31]. The identifying model is of the form

$$y(k) = \underline{X}^T(k) \, \hat{\underline{\theta}}(k) \tag{24}$$

where $\hat{\underline{\theta}}(k)$ is the k^{th} approximation of $\underline{\theta}$.

Clearly, an ARMA process such as that of (1) can be cast into the form of (23) by defining

$$\underline{\theta} = [b_0 \; b_1 \; \dots \; b_{Md} \mid a_1 \; a_2 \; \dots \; a_{Nd}]^T \tag{25.a}$$

$$\underline{X}(k) = [u(k) \; u(k-1) \; \dots \; u(k-Md) \mid d(k-1) \; d(k-2) \; \dots \; d(k-Nd)]^T \tag{25.b}$$

then $$d(k) = \underline{X}^T(k) \, \underline{\theta} \tag{26}$$

and the identifier would be

$$\hat{d}(k) = \underline{X}^T(k) \, \hat{\underline{\theta}}(k) \tag{27}$$

It is clear that (27) and the adaptive auxiliary filter used to computer $e_2(k)$ in (21), are of the same form if $M = M_d$ and $N = N_d$ for $\hat{\underline{\theta}}(k)$ in (22). Given enough complexity for this auxiliary filter, and since (25.a) can be augmented with (M - Md) MA and (N - Nd) AR zero coefficients, we may assume M = Md and N = Nd without loss of generality.

Following the notation of (11), $y_a(k)$ instead of $d(k)$ will be used subsequently for the identifier of (27). Thus, we have that

$$\begin{aligned} e_2(k) &= d(k) - y_a(k) \\ &= \underline{X}^T(k) \, \underline{\theta} - \underline{X}^T(k) \, \hat{\underline{\theta}}(k) \\ &= \underline{X}^T(k) \, \tilde{\underline{\theta}}(k) \end{aligned} \tag{28}$$

where $\quad \underline{\tilde{\theta}}(k) = \underline{\theta} - \underline{\hat{\theta}}(k)$ (29)

Perfect Measurements case. If the input and the desired signal can be properly assumed to be measurable without any error, i.e., they are noise-free, then in [27] the following gradient descent algorithm is proposed to update $\underline{\hat{\theta}}(k)$ such that (28) is minimized

$$\underline{\hat{\theta}}(k+1) = \underline{\hat{\theta}}(k) + 2\,\underline{R}(k)\,e_2(k)\,\underline{X}(k) \tag{30}$$

where $\underline{R}(k)$ is a $(M + N + 1) \times (M + N + 1)$ symmetric weighting matrix.

By using (28), and subtracting (30) from (25.a), the following identification-error system is obtained.

$$\underline{\tilde{\theta}}(k+1) = [\underline{I} - 2\,\underline{R}(k)\,\underline{X}(k)\,\underline{X}^T(k)]\,\underline{\tilde{\theta}}(k) \tag{31}$$

where $\underline{I}$ denotes the identity matrix. For (31) above, it can be verified that $\underline{\tilde{\theta}}_e = \underline{0}$ is the only possible equilibrium point, for a general input, and thus a candidate for a potentially asymptotically stable point in the large [27].

Choosing the weighting matrix $\underline{R}(k)$ to be

$$\underline{R}(k) = \mu\,\underline{I}\,, \quad \mu \in \mathbb{R}, \quad \mu > 0 \tag{32}$$

then (30) can be decomposed into the form of the ELMS algorithm of (20). By constructing a scalar function $V(\underline{\tilde{\theta}},k)$ given by

$$\underline{V}(\underline{\tilde{\theta}}, k) = \underline{\tilde{\theta}}^T(k)\,\underline{\tilde{\theta}}(k) \tag{33}$$

it is shown in [27] that this function will be a Lyapunov function for the free system of (31) if

$$0 < \mu < \frac{1}{\underline{X}^T(k)\underline{X}(k)}\,, \quad \forall\, k \in \mathbb{N} \cup \{0\} \tag{34}$$

Then, by the Lyapunov Stability Theorem [27], we will almost always have the following convergence condition

$$\underline{\hat{\theta}}(k) \longrightarrow \underline{\theta}, \qquad k \longrightarrow \infty \tag{35}$$

regardless of the initial estimate $\hat{\theta}(0)$. This condition is almost always true because if $\underline{\tilde{\theta}}(k)$ and $\underline{X}(k)$ are orthogonal, then $e_2(k)$ in (28) will be zero when $\underline{\tilde{\theta}}(k) \neq \underline{0}$. To avoid this situation $\underline{X}(k)$ must be sufficiently rich [26], [27].

Clearly, if $M \geq M_d$ and $N \geq N_d$ for the adaptive version of (13) depicted in Fig. 1, then given (35), $e_2(k) \rightarrow 0$ asymptotically, and from (16) and (17), $f(k) \rightarrow d(k)$ asymptotically as was our ultimate goal.

Noisy measurements case. The problem of identifying $\underline{\theta}$ in (26) becomes a stochastic estimation task if $d(k)$ and $u(k)$ can only be measured in the presence of additive noise, that is, if the measured signals are of the form

$$d_m(k) = d(k) + v(k) \tag{36}$$

$$u_m(k) = u(k) + n(k) \tag{37}$$

where $v(k)$ and $n(k)$ are independent, zero-mean random sequences. Define the measured information vector as

$$\underline{X}m(k) = [u_m(k)\ u_m(k-1)\ \ldots\ u_m(k-M) \mid d_m(k-1)\ d_m(k-2)\ \ldots\ d_m\ (k-N)]^T \tag{38}$$

The equation above can be rewritten as

$$\underline{X}_m(k) = \underline{X}(k) + \underline{N}(k) \tag{39}$$

with $\underline{X}(k)$ being the perfect measurements information vector and

$$\underline{N}(k) = [n(k)\ n(k-1)\ \ldots\ n(k-M) \mid\ v(k-1)\ v(k-2)\ \ldots\ v(k-N)]^T \tag{40}$$

then the identifier of (27) becomes

$$y_{a,m}(k) = \underline{X}_m(k)\ \hat{\underline{\theta}}(k) \tag{41}$$

and $$e_{2,m}(k) = d(k) - \underline{X}_m\ \hat{\underline{\theta}}(k) \tag{42}$$

Following the developments in [27], define

$$E[\underline{N}(k)\ \underline{N}^T(k)] = \underline{\Sigma}_n \tag{43}$$

$$E[\underline{X}(k)\ \underline{X}^T(k)] = \underline{\Omega} \tag{44}$$

If the following conditions are met

$$E[\underline{X}(k)\ \underline{N}^T(k)] = \underline{0} \tag{45}$$

$$E[v(k)\ \underline{N}(k)] = \underline{0} \tag{46}$$

$$E[\underline{X}(k)\ \underline{X}^T(k) \mid \hat{\underline{\theta}}(k)] = \underline{\Omega}$$

where (47) is a conditional expectation, and if $\hat{\underline{\theta}}(k)$ in (22) is updated by means of the following algorithm

$$\hat{\underline{\theta}}(k+p) = \hat{\underline{\theta}}(k) + 2\mu\ e_{2,m}(k)\ \underline{X}_m(k) \tag{48}$$

or equivalently

$$\gamma_j(k+p) = \gamma_j(k) + 2\ \mu\ e_{2,m}(k)\ u_m(k-j), \quad 0 \leq j \leq M \tag{49.a}$$

$$\delta_i(k+p) = \delta_i(k) + 2\ \mu\ e_{2,m}(k)\ d_m(k-i), \quad 0 < i \leq N \tag{49.b}$$

then the mean convergent parameter estimates, if they exist, are

$$\lim_{k\to\infty} \{E[\hat{\underline{\theta}}(k)]\} \triangleq \hat{\underline{\theta}}(\infty) = (\underline{\Omega} + \underline{\Sigma}_n)^{-1}\ \underline{\Omega}\ \underline{\theta} \tag{50}$$

where the value of p in (48) is chosen to uncorrelate [$\underline{X}(k)\ \underline{X}^T(k)$] with $\hat{\underline{\theta}}(k)$, so that (47) is satisfied.

Restricting to the commonly assumed situation where the input $u(k)$ is measured perfectly, and the desired signal is measured in the presence of zero-mean white noise, then (40) becomes

$$\underline{N}(k) = [0 \; 0 \; \ldots \; 0 \mid v(k-1) \; v(k-2) \; \ldots \; v(k-N)]^T \tag{51}$$

where v(k) is now a zero-mean white random sequence. Clearly (45) and (46) are satisfied, and

$$\underline{\Sigma}_n = \begin{bmatrix} \underline{0} & \underline{0} \\ \underline{0} & \underline{\Sigma}_v \end{bmatrix} \tag{52}$$

where $\underline{\Sigma}_v = \text{diag}\,\{\sigma^2\}$,

is the autocorrelation matrix of $\underline{V} = [v(k-1) \; v(k-2) \; \ldots \; v(k-N)]^T$.

Let's examine the condition in (47). First note that $\hat{\underline{\theta}}(k)$ depends on $\underline{X}_M(k-p)$ through (48), and due to the IIR characteristics of (1), d(k) will depend on all previous values d(k-1), d(k-2), ..., d(0). Thus even assuming that an appropriate choice of p in (48) uncorrelates the part due to u(·) in $\underline{X}(k)$ and $\underline{X}(k-p)$, this will not be possible for the part due to d(·). However, if the system is to be stable, the effects of the value of the output at any given instant in time over future values must die out asymptotically. Otherwise the output would diverge as $k \to \infty$. Hence, it is possible to choose p in (48) such that (47) is satisfied within a given range of accuracy. If $p \to \infty$, (47) is satisfied. By choosing a finite p, (50) would constitute an approximate formula for the biasing in $\hat{\underline{\theta}}(\infty)$.

The inportant conclusion from the above comment is that whenever there is noise present in the measurement of d(k) and u(k), the estimates for $\underline{\theta}$ obtained through (48) will be biased. Thus the adaptive filter depicted in Fig. 1 will yield a suboptimal minimization of (16) when updated using (49).

Recall that it was said that the mean of the estimates for $\underline{\theta}$ obtained by use of (48), given (45) - (47) are met, would be given by (50) if there exists a steady-state value for $\hat{\underline{\theta}}(k)$. The proof of the existence of such a value is not a trivial task, and in [27] conditions are given for the case when u is replaced by a vanishing step-size. However, preliminary simulations show the convergence of (48) for constant , at least for relatively high signal to noise ratios. Further studies on this issue are being undertaken.

SUMMARY

We have shown that the ELMS algorithm can yield global minimization of (16), given three conditions: (i) the desired signal is generated by an ARMA process as described by (1), (ii) enough complexity is provided for the adaptive filter to span all the modes of (1), and (iii) there is no noise present in the measurements. If one or both of the first two conditions are not met, then $\hat{\underline{\theta}}(k)$ will converge to (12), minimizing globally (17) but not (16). If only the last condition is not met, then $\hat{\underline{\theta}}(\infty)$ will constitute a biased estimate of (25.a). The last two situations result in a suboptimal performance of the adaptive recursive filter. However, the low computational complexity and the simple conditions required for the ELMS algorihtm makes it appealing as a substitute for the LMS algorithm in those applications where the latter has proved of use, but a large number of tap weights are necessary.

REFERENCES

[1] B. Widrow and M.E. Hoff, "Adaptive Switching Circuits," 1960 IRE WESCON Convention Record, pt. 4, pp. 96-104.

[2] B. Widrow, "Adaptive Filters," in Aspects of Network and System Theory, R. E. Kalman and N. Declaris, Eds., New York: Holt, Rinehart, and Winston, 1971, pp. 563-587.

[3] B. Widrow, J. M. McCool, and M. Ball, "The Complex LMS Algorithm," Proceedings of the IEEE, Vol. 63, no. 4, pp. 719-720, April 1975.

[4] B. Widrow, J. M. McCool, M.G. Larimore and C.R. Johnson, Jr., "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," Proceedings of the IEEE, Vol 64, no. 8, pp. 1151, 1162, Aug. 1976.

[5] E. Walach and B. Widrow, "The Least Mean Fourth (LMF) Adaptive Algorithm and Its Family," IEEE Transactions on Information Theory, Vol. IT-30, No. 2, pp. 275-283, March 1984.

[6] T. Aprille, "Filtering and Equalization for Digital Transmission," IEEE Communications Magazine, pp. 17-24, March 1983.

[7] J.G. Proakis, Digital Communications, New York: McGraw-Hill, 1983.

[8] B. Widrow and E. Walach, "Adaptive Signal Processing for Adaptive Control," Proceedings of the 1984 IEEE International Conference on Acoustic, Speech, Signal Processing, San Diego, CA, March 1984, pp. 21.1.1-21.1.4.

[9] C.R. Johnson, Jr. "Input Matching, Error Augmentation, Self-Tuning and Output Error Identification: Algorithmic Similarities in Discrete Model Following," IEEE Transactions on Automatic Control, Vol. AC-25, No. 4, pp. 697-703, August 1980.

[10] S.A. White, "An Adaptive Recursive Digital Filter," Ninth Annual Asilomar Conference on Circuits, Systems and Computers Conference Record, Pacific Grove, CA, November 1975, pp. 21-25.

[11] S.D. Stearns, G.R. Elliot and N. Ahmed, "On Adaptive Recursive Filtering," Tenth Annual Asilomar Conference on Circuits, Systems and Computers Conference Record, San Francisco, CA, November 1976, pp. 5-11.

[12] J.A. Cadzow, "Recursive Digital Filter Synthesis Via Gradient Based Algorithms," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-24, no. 5, pp. 349-355, October 1976.

[13] P.L. Feintuch, "An Adaptive Recursive LMS Filter," Proceedings of the IEEE, Vol. 64, no. 11, pp. 1622-1624, November 1976.

[14] C.R. Johnson, Jr., and M.G. Larimore, "Comments on and Additions to an Adaptive Recursive LMS Filter," Proceedings of the IEEE, Vol. 65, no. 9, pp. 1399-1402, September 1977.

[15] B. Widrow and J.M. McCool, "Comments on An Adaptive Recursive LMS Filter," Proceedings of the IEEE, Vol. 65, no. 9, pp. 1402-1404, September 1977.

[16] T.C. Parikh and N. Ahmed, "On an Adaptive Algorithm for IIR Filters," Proceedings of the IEEE, Vol. 66, no. 5, pp. 585-588, May 1978.

[17] S.D. Stearns, "Error Surfaces of Recursive Adaptive Filters," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-29, no. 3, pp. 763-766, June 1981.

[18] T.C. Hsia, "A Simplified Adaptive Recursive Filter Design," Proceedings of the IEEE, Vol. 69, No.9, pp. 1153-1155, September 1981.

[19] C.R. Johnson, Jr., "A Convergence Proof for a Hyperstable Adaptive Recursive Filter," IEEE Transaction on Informaton Theory, Vol. IT-25, no. 6, pp. 745-749, November 1979.

[20] J.R. Treichler, M.G. Larimore and C.R. Johnson, Jr., "On the Convergence Properties of the Simple Hyperstable Adaptive Recursive Filter (SHARF)," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Denver, Co, April 1980, pp. 997-1000.

[21] C.R. Johnson, Jr., "A Stable Family of Adaptive IIR Filters" Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Denver, CO, April 1980, pp. 1001-1004.

[22] M.G. Larimore, J.R. Treichler and C.R. Johnson, Jr., "SHARF: An Algorithm for Adapting IIR Digital Filters," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-28, no. 4, pp. 428-440, August 1980.

[23] C.R. Johnson, Jr., I.D. Landau, T. Taylor and L. Dugard, "On Adaptive IIR Filters and Parallel Adaptive Idenfifiers with Adaptive Error Filtering," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, GA, March 1981, pp. 538-541.

[24] C.R. Johnson, Jr., M.G. Larimore, J.R. Treichler and B.D.O. Anderson, "SHARF Convergence Properties," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-29, no.3, pp. 659-670, June 1981.

[25] M.G. Larimore, "Hyperstability and Adaptive Filtering," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-29, no.2, pp. 319-325, April 1981.

[26] C.R. Johnson, Jr., "Adaptive IIR Filtering: Current Results and Open Issues," IEEE Transactions on Information Theory, Vol. IT-30, No. 2, pp. 237-250, March 1984.

[27] J.M. Mendel, Discrete Techniques of Parameter Estimation: The Equation Error Formulation. New York: Marcel Dekker, 1973.

[28] T.K. Sarkar, S.A. Dianat, and D.D. Weiner, "A Discussion of Various Approaches to The Linear Systems Identification Problem," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No. 3, pp. 654-656, June 1984.

[29] K. Steiglitz, "On the Simultaneous Estimation of Poles and Zeros in Speech Analysis," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-25, no. 3, pp. 229-234, June 1977.

[30] R.E. Kalman, "Design of Self-Optimizing Control Systems," Transactions of the ASME, vol. 80, pp. 465-478, Feb. 1958.

[31] C.R. Johnson, Jr., "Adaptive Parameter Matrix and Output vector Estimation via an Equation Error Formulation," IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-9, no. 7, pp. 392-3297, July 1979.

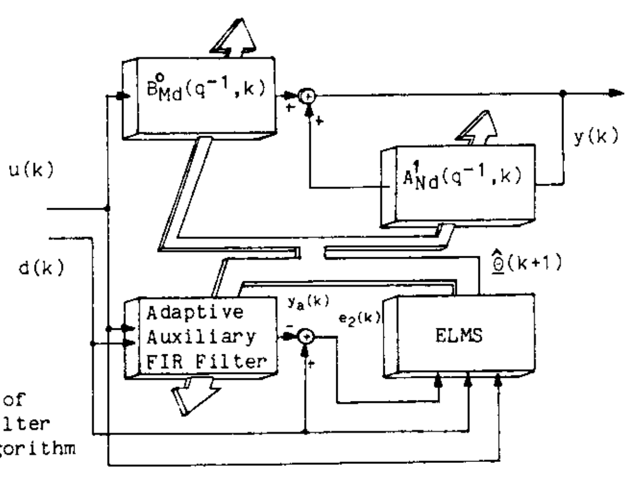

Fig. 1 Structure of an adaptive IIR filter using the ELMS algorithm