

A COMPARISON OF THE EQUATION ERROR AND OUTPUT ERROR APPROACHES
IN ADAPTIVE IIR FILTERING

EMMANUEL FERNANDEZ-GAUCHERAND* and JOHN Y. CHEUNG

School of Electrical Engineering and Computer Science,
University of Oklahoma, 202 W. Boyd, CEC219, Norman, OK 73019.

ABSTRACT

In a systems identification scheme, the equation error formulation of auto regressive moving average (ARMA) identifiers can be viewed in a dual way. It is shown that the equation error process generated in this way can be expressed as a recursive filtering of the inner product between the parameter estimation error vector and a suitable information vector, as it has been shown that is the case for the output error process. Thus, the similarities arising between two corresponding solutions to adaptive infinite impulse response (IIR) filtering obtained via these seemingly different approaches are exposed. Computer simulations are provided to further aid in gaining insight of the similarities, and differences, between the two approaches.

I. INTRODUCTION

The equation error approach, as applied to recursive identifiers, has been extensively studied within the systems identification community since its introduction, in implicit form, by Kalman in 1958 [1]-[3]. Though, it remained relatively unexplored by the signal processing community until recent years, when it has been employed both in batch [4] and adaptive [5]-[7] processing schemes. The concept of the equation error can be viewed from two standpoints, namely as the difference between the outputs of the unknown system and a two-input/one-output nonrecursive identifier, or as a particular smoothed version of the "output error", i.e. the difference between the output of the ARMA plant and that of an equally structured identifier. Johnson [8] used the first point of view above to argue in favor of output error techniques for IIR adaptive filtering, such as the hyperstability based solutions, because of the structural compatibility for the plant generating the desired signal and the identifier (i.e. filter). However, great similarities can be drawn between the two approaches by using the second viewpoint to equation error previously mentioned. It is interesting to note how the

equation error concept has been "rediscovered" in some occasions, e.g. [9].

II. DESCRIPTION OF THE EQUATION AND OUTPUT ERROR PROCESSES

Let us assume that the desired signal provided to the adaptive filter is satisfactorily modeled as an ARMA process of the form

$$d(k) = B_0^M(q^{-1}) u(k) + A_1^N(q^{-1}) d(k) \quad (1)$$

or

$$d(k) = \frac{B_0^M(q^{-1})}{1 - A_1^N(q^{-1})} u(k) \quad (2)$$

where $u(\cdot)$ is the exciting signal, $B_0^M(q^{-1})$ and $A_1^N(q^{-1})$ are polynomials of the form

$$B_0^M(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_M q^{-M} \quad (3)$$

$$A_1^N(q^{-1}) = a_1 q^{-1} + a_2 q^{-2} + \dots + a_N q^{-N} \quad (4)$$

and q^{-j} denotes a delay operator, i.e. $x(k) q^{-j} = x(k-j)$. Assume $f(\cdot)$ to be an equally dimensioned static identifier for the process $d(\cdot)$, that is

$$f(k) = \hat{B}_0^M(q^{-1}) u(k) + \hat{A}_1^N(q^{-1}) f(k) \quad (5)$$

or

$$f(k) = \frac{\hat{B}_0^M(q^{-1})}{1 - \hat{A}_1^N(q^{-1})} u(k) \quad (6)$$

where $\hat{B}_0^M(q^{-1})$ and $\hat{A}_1^N(q^{-1})$ are polynomials in the estimates \hat{b}_j and \hat{a}_i , defined similarly as (3) and (4), respectively. For this identification scheme, the output error is given by

$$\begin{aligned} e(k) &= d(k) - f(k) \\ &= \frac{[1 - \hat{A}_1^N(q^{-1})] d(k) - \hat{B}_0^M(q^{-1}) u(k)}{1 - \hat{A}_1^N(q^{-1})} \quad (7) \end{aligned}$$

*Now with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712.

Defining $\tilde{B}_0^M(q^{-1})$ and $\tilde{A}_1^N(q^{-1})$ as polynomials of the form (3), (4), respectively, in the parameter estimates error coefficients $\tilde{b}_j = b_j - \hat{b}_j$ and $\tilde{a}_i = a_i - \hat{a}_i$, then (7) can be rewritten as [8], [10]

$$e(k) = \frac{\tilde{B}_0^M(q^{-1}) u(k) + \tilde{A}_1^N(q^{-1}) f(k)}{1 - \tilde{A}_1^N(q^{-1})} \quad (8)$$

This last expression shows the IIR nature of the output error process. Also, defining the parameter error vector as

$$\tilde{\theta} = [\tilde{b}_0 \tilde{b}_1 \dots \tilde{b}_M \mid \tilde{a}_1 \tilde{a}_2 \dots \tilde{a}_N]^T \quad (9)$$

and an information vector containing values of $f(\cdot)$ as

$$X_f(k) = [u(k) \ u(k-1) \ \dots \ u(k-M) \mid f(k-1) \ f(k-2) \ \dots \ f(k-N)]^T \quad (10)$$

then $e(\cdot)$ can be expressed as

$$e(k) = \frac{1}{1 - \tilde{A}_1^N(q^{-1})} \tilde{\theta}^T X_f(k) \quad (11)$$

which shows that $e(\cdot)$ is generated as a purely auto regressive filtering of an input signal obtained as the inner product $\tilde{\theta}^T X_f(k)$.

For the same identification scheme of (2) and (6), the equation error is defined as the numerator of (7), that is [7], [10]

$$\begin{aligned} \xi(k) &= [1 - \hat{A}_1^N(q^{-1})] e(k) \\ &= d(k) - [\hat{B}_0^M(q^{-1}) u(k) + \hat{A}_1^N(q^{-1}) d(k)] \\ &= \tilde{B}_0^M(q^{-1}) u(k) + \tilde{A}_1^N(q^{-1}) d(k) \end{aligned} \quad (12)$$

Note the dual way in which $\xi(\cdot)$ can be thought of: on the one hand as a smoothed version of $e(\cdot)$, and on the other hand as the difference between $d(\cdot)$ and the output of a two-input ($u(\cdot)$ and $d(\cdot)$) one-output transversal filter, as evidenced by the first and second expressions on the right-hand side of (12), respectively. Also, combining (11) and (12) we obtain

$$\xi(k) = \frac{[1 - \hat{A}_1^N(q^{-1})]}{1 - \tilde{A}_1^N(q^{-1})} \tilde{\theta}^T X_f(k) \quad (13)$$

Thus, both equation and output error can be viewed as the response of particular recursive systems to the same input $\tilde{\theta}^T X_f(k)$, and these recursive systems both possess the same eigenpolynomial as the system under identification.

III. ADAPTIVE ALGORITHMS

In a deterministic environment, two solutions can be found in the literature for the

problem of algorithm development in adaptive IIR filtering. The first one, based on output error techniques, is the so called simple hyperstable adaptive filter (SHARF) [8], which takes the form

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \mu X_f(k) v(k) \quad (14)$$

where $\hat{\theta}$ is a vector of parameter estimates of the form of (9), and $v(k)$ is a smoothed output error of the form

$$\begin{aligned} v(k) &= [1 - c_1^P(q^{-1})] e(k) \\ &= [1 - \sum_{k=1}^P c_k q^{-k}] e(k) \end{aligned} \quad (15)$$

The step-size μ can be chosen as some small positive constant, or for normalization as the reciprocal of X_f inner product with itself [6], [7]. Using (11), we rewrite (14) as

$$\begin{aligned} \hat{\theta}(k+1) &= \hat{\theta}(k) \\ &+ \mu X_f(k) \left[\frac{1 - c_1^P(q^{-1})}{1 - \tilde{A}_1^N(q^{-1})} \right] \tilde{\theta}^T X_f(k) \end{aligned} \quad (16)$$

As a sufficient condition for stability of (16), the c_k coefficients in (15) must be selected such that

$$H(q^{-1}) = \frac{1 - c_1^P(q^{-1})}{1 - \tilde{A}_1^N(q^{-1})} \quad (17)$$

is strictly positive real (SPR) [8].

The second solution is based on a gradient descent of the quadratic hypersurface generated by (12), and it takes the form [7], [10]

$$\begin{aligned} \hat{\theta}(k+1) &= \hat{\theta}(k) + \mu X_d(k) \xi(k) \\ &= \hat{\theta}(k) + \mu X_d(k) \left[\frac{1 - \hat{A}_1^N(q^{-1})}{1 - \tilde{A}_1^N(q^{-1})} \right] \tilde{\theta}^T X_f(k) \end{aligned} \quad (18)$$

where $X_d(k)$ is defined similarly as (10), but with values of $d(\cdot)$ instead of $f(\cdot)$. Since the component-wise equations arising from (18) use only input signals for the estimated gradients, they resemble those obtained for the LMS algorithm. Thus, we may call (18) an extended LMS algorithm (ELMS).

The similarities between the two approaches presented above are clear by comparing (11) and (13) for the dynamic description of each process, and by comparing (16) and (18) for the corresponding adaptive algorithms. Also, from (12) and (15) we see that both algorithms seek to minimize a smoothed version of the output error.

IV. COMPUTER SIMULATIONS

For the algorithms of (16) and (18), several

aspects can be examined, e.g. convergence characteristics, input sufficiency, sensitivity to measurement noise, etc. [7]. To do this in a somewhat formal way, the examples presented are drawn from the literature pertaining to the SHARF solution.

Example I: the case to be considered here is taken from a paper by Johnson et al. [11]. A white Gaussian zero mean sequence $u(\cdot)$ is filtered by a second-order system with poles located at $z = 0.9 \pm j0.3$, in order to generate the desired signal. The time-domain transfer function of the filter used is given by

$$\frac{d(k)}{u(k)} = \frac{1}{1 - 1.5588 q^{-1} + 0.81 q^{-2}} \quad (19)$$

An equally dimensioned, equally structured adaptive filter (identifier) is initialized with its poles at $z = \pm 0.5$, i.e. $\hat{a}_1(0) = 0$ and $\hat{a}_2(0) = -0.25$. Figure 1 shows the parameter estimate trajectories for both SHARF and ELMS, where one smoothing coefficient $c_1 = 0.99$ was provided for the former to aid in its convergence [7], while obeying the SPR condition for (17). Though, ELMS still converges faster and smoother than SHARF. Figure 2 shows the case when the SPR condition for SHARF is trivially satisfied, i.e. $H(q^{-1}) = 1$ in (17). Even though the trajectories obtained using SHARF are smoother than before, its convergence rate worsens. An arithmetic average of 500 learning curves for the squared errors associated with Figure 1 are shown in Figure 3. Both algorithms are seen to minimize the mean squared error to zero, but ELMS does it faster. A motivating factor for using IIR adaptive filters instead of basic transversal ones is the possibility of drastically reducing the number of adaptive coefficients, while maintaining a similar performance. Figure 4 shows the learning curves for ELMS and a 9 tap transversal filter adapted using LMS. It is clear from this figure that the advantages of the first method over the second one are numerous. For the case when the available measurements of $d(\cdot)$ are corrupted by (white) additive noise, both ELMS [7] and SHARF [8] yield biased estimates. However, the latter shows a greater insensitivity to noisy measurements than the former. As an example, the "learning" curves for the ratio of the squared errors are shown in Figure 5. As it can be seen, for high signal/noise ratios, both algorithms perform very similarly; but as the signal/noise ratio drops, the performance of ELMS worsens as compared to that of SHARF.

Example II: Johnson and Larimore [12] studied the reduced order modeling case of a second order plant and a first order identifier. For this case, the mean squared error surface is multimodal, i.e. it has both local and global minimums. Thus, gradient search techniques are ill-fated in general. Specifically, let the plant have transfer function

$$\frac{d(k)}{u(k)} = \frac{0.05 - 0.4 q^{-1}}{1 - 1.1314 q^{-1} + 0.25 q^{-2}} \quad (20)$$

and let the adaptive filter/identifier have the following structure

$$\frac{f(k)}{u(k)} = \frac{\hat{b}(k)}{1 - \hat{a}(k) q^{-1}} \quad (21)$$

Figure 6 shows the contour of the normalized mean squared error surface, as well as the points to which both ELMS and SHARF converge [7]. It is clear then that both are nonrobust to undermodelling of the plant.

V. SUMMARY

The equation and output error methods for ARMA identifiers, as well as their related solutions to adaptive IIR filtering, have been succinctly expressed. Their intrinsic structural similarities have been exposed, and computer simulations have been provided to gain insight in this respect. Extensive computer simulations and analysis of these and other application areas can be found elsewhere [7], [13]; as well as a block formulation of the equation error-based IIR adaptive filter [7] (see also an accompanying paper).

REFERENCES

- [1] R.E. Kalman, "Design of Self-Optimizing Control Systems", Trans. ASME, Vol. 80, pp. 468-478, 1958.
- [2] K.J. Åström and P. Eykhoff, "System Identification - A Survey", Automatica, Vol. 7, pp. 123-162, 1971.
- [3] J.M. Mendel, Discrete Techniques of Parameter Estimation: The Equation Error Formulation, Marcel Dekker, New York, 1973.
- [4] K. Steiglitz, "On the Simultaneous Estimation of Poles and Zeros in Speech Analysis", IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-25, No. 3, pp. 229-234, 1977.
- [5] R.P. Gooch, "Adaptive Pole-Zero Filtering: The Equation Error Approach", Ph.D. dissertation, Department of Electrical Engineering, Stanford Univ., 1983.
- [6] E. Fernández-Gaucherand and J.Y. Chung, "A Suboptimal Solution to IIR Adaptive Filtering using an Extended LMS Algorithm", Proc. 22nd Annual Allerton Conf. Comm., Control, and Computing, Univ. of Illinois at Urbana-Champaign, pp. 132-141, 1984.
- [7] E. Fernández-Gaucherand, "Adaptive IIR Filtering: Current Results and the Equation Error Approach", M.Sc. Thesis, School of Elect. Eng. and Comp. Sci., The Univ. of Oklahoma, 1985.
- [8] C.R. Johnson, Jr., "Adaptive IIR Filtering: Current Results and Open Issues", IEEE Trans. Inf. Theory, Vol. IT-30, No. 2, pp. 237-250, 1984.
- [9] W.A. Gardner, "A Stable Two-Stage Recursive-Like Adaptive Noise Canceller", Proc. IEEE, Vol. 69, No. 4, pp. 487-489, 1981.
- [10] E. Fernández-Gaucherand and J.Y. Chung, "On the Similarities between the Equation Error and the Output Error Approaches in Adaptive IIR Filtering", submitted for publication, 1986.
- [11] C.R. Johnson, Jr., M.G. Larimore, J.R. Treichler and B.D.O. Anderson, "SHARF Convergence Properties", IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-29, No. 3, pp. 659-670, 1981.
- [12] C.R. Johnson, Jr. and M.G. Larimore, "Comments on and Additions to an Adaptive Recursive LMS Filter", IEEE Proc., Vol. 65, No. 9, pp. 1399-1402, 1977.
- [13] E. Fernández-Gaucherand and J.R. Cruz, "Equalization for Transmission Line Channels - A Discussion of Three IIR Adaptive Filtering Algorithms", 1985 IEEE ICASSP, Tampa, FL, pp. 1249-1252, 1985.

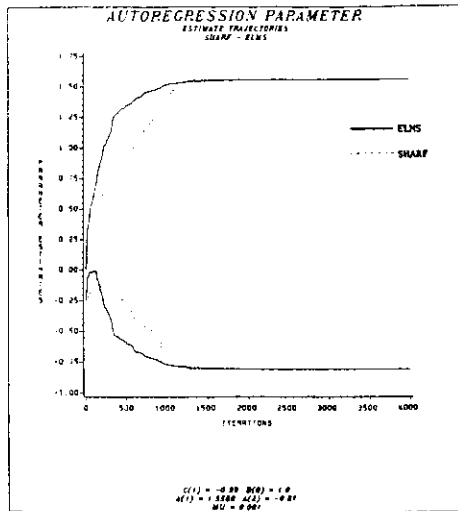


Fig. 1

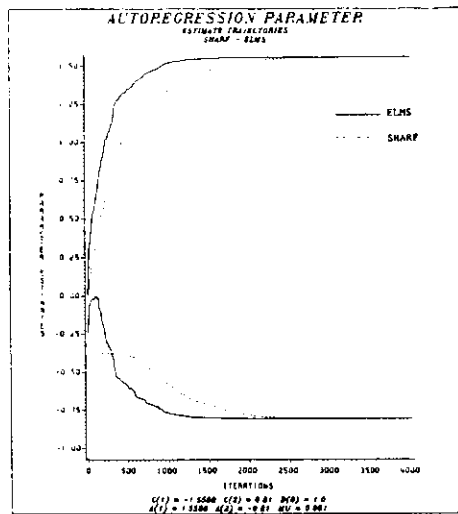


Fig. 2

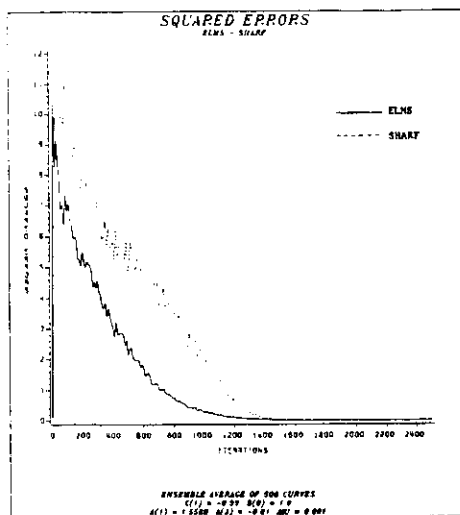


Fig. 3

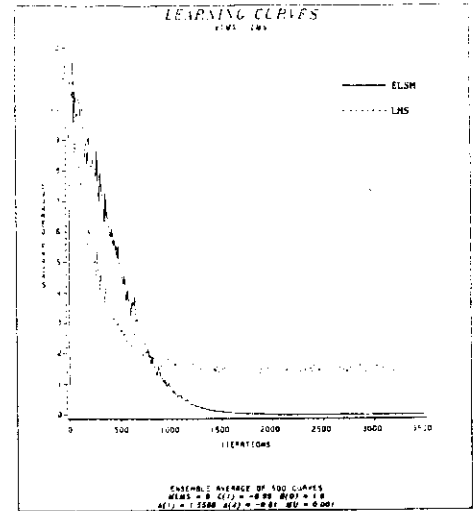


Fig. 4

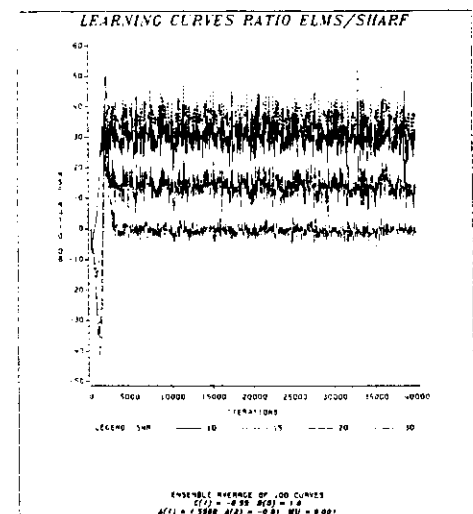


Fig. 5

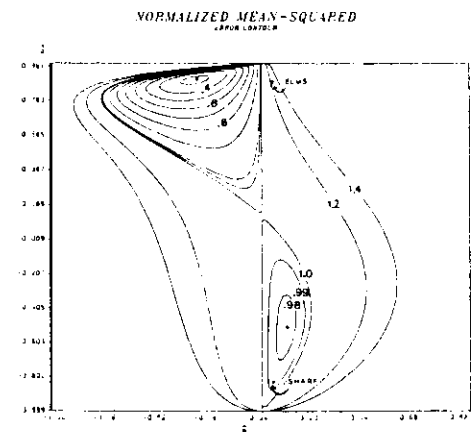


Fig. 6