Baire function such that

$$\int_{-\infty}^{\infty} [g(u)]^2 \exp\left(-\frac{u^2}{2R_1(0)}\right) du < \infty.$$

This class of ZNL's is the most general class such that $X_2(t)$ is a well-defined second-order random process, and it is a more general class than will be encountered in practice.

The ZNL $g(\cdot)$ can be represented as

$$g(u) = \sum_{n=0}^{\infty} b_n \phi_n(u)$$

where

$$\phi_n(u) = \frac{H_n[u/\sqrt{R_1(0)}]}{\sqrt{n!}}$$

$H_n(\cdot)$ is the nth Hermite polynomial

$$b_n = \int_{-\infty}^{\infty} g(u)\,\phi_n(u)\,\frac{1}{\sqrt{2\pi R_1(0)}}\,\exp\left(-\frac{u^2}{2R_1(0)}\right) du$$

and the convergence is in an L_2 sense. It then follows [3] that the autocorrelation function of $X_2(t)$, denoted by

$$R_2(\tau) = E\{g[X_1(t+\tau)]\,g[X_1(t)]\}$$

is given by

$$R_2(\tau) = \sum_{n=0}^{\infty} b_n^2 \left(\frac{R_1(\tau)}{R_1(0)}\right)^n$$

where the convergence is absolute and uniform.

The autocorrelation function of the output $Y(t)$ is given by

$$R_Y(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_2(\tau + v - u)\,h_2(v)\,h_2(u)\,dv\,du$$

where $h_2(\cdot)$ is the Green's function characterizing the second linear time invariant system. Thus it is seen that the presence of the ZNL $g(\cdot)$ is manifested in the output autocorrelation function. However, notice that the ZNL is characterized by the coefficient sequence $\{b_n\}$, while the output autocorrelation function is coupled to the ZNL through the coefficient sequence $\{b_n^2\}$. By changing the signs of the coefficients $\{b_n\}$, ZNL's can frequently be exhibited which differ radically in functional form. For example, let $R_1(0) = \sigma^2$ and consider the following three ZNL's:

$$g_1(u) = \sin\left(\frac{u\sqrt{2}}{\sigma}\right)$$

$$g_2(u) = \sin\left(\frac{u\sqrt{2}}{\sigma}\right) + \left(\frac{2\sqrt{2}}{e\sigma}\right)u$$

$$g_3(u) = e^{-2}\sinh\left(\frac{u\sqrt{2}}{\sigma}\right).$$

It is easily shown [4] that for each of these ZNL's the coefficient sequences have the same magnitudes. Thus if one of these three ZNL's is used in the original nonlinear system, it is impossible to tell which one it is, if our decision is based only upon knowledge of the output autocorrelation function.

REFERENCES

[1] R. R. Pfeiffer, "A model for two-tone inhibition of single cochlear nerve fibers," *J. Acoust. Soc. Amer.*, vol. 48, pp. 1373–1378, 1970.
[2] E. De Boer, "Crosscorrelation function of a bandpass nonlinear network," *Proc. IEEE*, vol. 64, pp. 1443–1444, Sept. 1976.
[3] G. L. Wise, A. P. Traganitis, and J. B. Thomas, "The effect of a memoryless nonlinearity on the spectrum of a random process," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 84–89, Jan. 1977.
[4] W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics*. New York: Chelsea, 1949, p. 81.

Comments on and Additions to "An Adaptive Recursive LMS Filter"

C. RICHARD JOHNSON, JR., AND MICHAEL G. LARIMORE

In the above letter[1], the author presents an algorithm for adaptively adjusting a recursive filter similar to that for the nonrecursive LMS adaptive filter [2]. While the algorithm is of interest because in certain cases, it forces the filter to minimize a squared-error criterion, the analysis presented is basically misleading and its implied generality is incorrect. Furthermore, we will show that in general the algorithm does not minimize squared error as claimed.

In equation (4)[2] of the above letter, the expected squared error between a desired signal and the output of a recursive filter is given. Equations (5) and (6) are meant to represent the gradient of this error functional with respect to the parameters of the recursive filter. Under the stated assumption of constant correlation statistics this step is correct. However, their incorporation into (9) and (10), a crucial insertion, is invalid.

Only for the purposes of recursively estimating filter coefficients from constant correlation statistics known *a priori* as in (7) and (8) do the gradient estimates of (5) and (6) have meaning. To assume that these statistics for the optimal filter are available *a priori* is, in general, an unrealistic assumption. Only in the case where the desired signal is the output itself of some unknown filter can these statistics be given practical significance. This case degenerates to the parameter estimation of the filter transfer function from input and output statistics. The solution for the practical case of finite measurements is the well-known least-squares method [3, ch. 2]. In Feintuch's letter, (7) and (8) supply a recursive gradient search solution to this problem. Equations (7) and (8) are applicable in this case only because the parameter estimates are not used in determining the future input sequence, justifying the assumption that "the statistics . . . are not a function of the [adaptive filter] weights."

When the filter coefficients are updated during operation the requirement of constant correlation statistics is obviously violated.[3] The complete form of the gradient of the current value of the expected squared error with respect to the filter parameters is

$$\begin{aligned}\nabla_A[E\{\epsilon^2(k)\}] &= -2E\{\epsilon(k)\,(X(k) \\ &\qquad - [\nabla_A y(k-1)|\cdots|\nabla_A y(k-N_B)]\,B)\} \\ \nabla_B[E\{\epsilon^2(k)\}] &= -2E\{\epsilon(k)\,(Y(k) \\ &\qquad - [\nabla_B y(k-1)|\cdots|\nabla_B y(k-N_B)B])\}\end{aligned} \quad (A)$$

analogous to the gradient for the deterministic case given in Feintuch's first reference [4] and in [5]. The substitutions of (7) and (8) yielding (9) and (10) can be recognized as abrupt truncation of the true gradient formulas (A) leaving only the first term. This parallels higher order approximations, including several terms from the recursive expressions for $\nabla_A y(\cdot)$ and $\nabla_B y(\cdot)$, presented in [6] and [7] with limited success. A recent derivation of an adaptive recursive identifier [8] and an earlier quasi-stationary analysis of the recursive adaptive filter problem [9], both proven via hyperstability theory and then related to gradient methods, require formation of a generalized error to be used in adjusting filter parameters, as a properly weighted summation of past output errors, which is similar to these efforts with higher order gradient approximations.

Despite the mathematical invalidity of using the forms of the gradient (5) and (6) with instantaneous estimates to form (9) and (10) from (7) and (8), desirable behavior can occur as presented in the simulations of [1]. Inferring global utility from these two examples, however, is dangerous since the scheme has severe deficiencies not revealed by these examples. As noted in [6] and [10], the squared-error surface is,

Manuscript received December 27, 1976.
The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

[1] P. L. Feintuch, *Proc. IEEE*, vol. 64, pp. 1622–1624, Nov. 1976.

[2] Numbers will be used to identify the equations and figures from "An Adaptive Recursive LMS Filter" while letters will be used to designate those of this text.

[3] This algorithm, in the stochastic gradient sense, also violates the implicit condition of statistical independence between the current input and the current parameter vector required to utilize (7) and (8) in a stochastic approximation approach [3, p. 234], since the filter's output is fed back as an input as is apparent from Fig. 1 in Feintuch's letter.

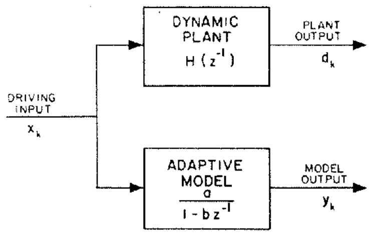

Fig. A. Approximating dynamic plant by two-parameter adaptive model.

in general, a multimodal function of the filter coefficients, therefore limiting the applicability of any gradient search algorithm. Furthermore, assuming convergence, the algorithm of (9)–(11) does not necessarily converge to any point of zero gradient, local or global, as demonstrated by the following simple example of approximating a second order dynamic plant

$$H(z^{-1}) = \frac{\alpha_0 + \alpha_1 z^{-1}}{1 - \beta_1 z^{-1} - \beta_2 z^{-2}}$$

by a simple first order system $(a)/(1 - bz^{-1})$ illustrated in Fig. A. Note that this case is plagued by insufficient degrees of freedom in the adaptive filter, unlike the examples of [1] and [6]. To avoid any question of input richness sufficiency [11] let the plant be driven by a discrete white noise process $\{x(k)\}$. Define the optimal model as that which matches its output $y(k)$ to the desired output $d(k)$ in the least squares sense. For this case the mean-squared error is

$$\begin{aligned} E\{\epsilon^2(k)\} &= E\{[d(k) - y(k)]^2\} \\ &= E\{x^2(k)\}\oint |H(z^{-1})|^2 \frac{dz}{z} \\ &\quad - 2a\, H(z^{-1})|_{z^{-1}=b}\, E\{x^2(k)\} + \frac{a^2}{1-b^2} E\{x^2(k)\}. \end{aligned} \tag{B}$$

Defining

$$\sigma_h^2 \triangleq \oint |H(z^{-1})|^2 \frac{dz}{z}$$

and dividing (B) by its first term yields a normalized mean-squared error ξ

$$\xi \triangleq 1 - \frac{2a}{\sigma_h^2} H(z^{-1})\Big|_{z^{-1}=b} + \frac{a^2}{1-b^2}\frac{1}{\sigma_h^2}.$$

Fig. B shows a contour representation of the normalized mean-squared-error surface as a function of the two variable model parameters a and b when

$$H(z^{-1}) = \frac{0.05 - 0.4\, z^{-1}}{1 - 1.1314\, z^{-1} + 0.25\, z^{-2}}.$$

The surface is bimodal with the global minimum denoted by a "$*$" in Fig. B at $(a^*, b^*) = (-0.311, 0.906)$ yielding $\xi = 0.277$ and a local minimum signified by a "+" in Fig. B at $(a^+, b^+) = (0.114, -0.519)$ achieving $\xi = 0.976$. It can be shown for this second order case that if the "unknown" filter's nonzero zero is outside the unit circle then the mean-squared-error surface is bimodal.

Suppose that the model were implemented adaptively, i.e., parameters a and b were initialized to some value and iteratively adjusted as more data $x(k)$ and $d(k)$ became available. Assuming that (a, b) were set sufficiently close to (a^*, b^*), i.e., within the "valley" surrounding the global minimum, a true gradient descent algorithm would force the model parameters to the optimal values (a^*, b^*). In vector notation

$$\begin{bmatrix} a(k+1) \\ b(k+1) \end{bmatrix} = \begin{bmatrix} a(k) \\ b(k) \end{bmatrix} + \begin{bmatrix} k_1 \dfrac{\partial}{\partial a} E[\epsilon^2] \\ k_2 \dfrac{\partial}{\partial b} E[\epsilon^2] \end{bmatrix}_{(a(k),b(k))}.$$

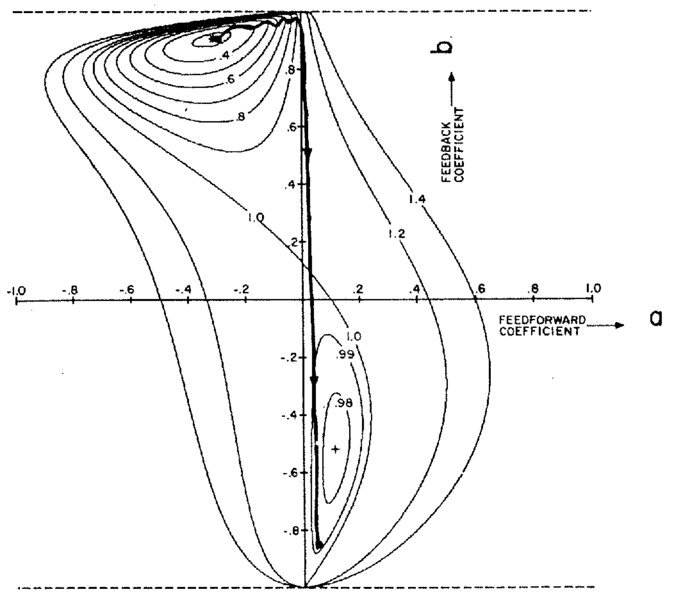

Fig. B. Bimodal case—contour representation of normalized mean-squared error showing locus of adaptive parameters (a, b): $k_1 = -0.00025$, $k_2 = -0.025$, average of 20 ensemble members.

Convergence can occur if the updating terms vanish

$$\frac{\partial}{\partial a} E[\epsilon^2]\Big|_{(a(k),b(k))} = \frac{\partial}{\partial b} E[\epsilon^2]\Big|_{(a(k),b(k))} = 0.$$

The algorithm presented in [1], however, behaves quite differently. Using equations (9) and (10)

$$\begin{bmatrix} a(k+1) \\ b(k+1) \end{bmatrix} = \begin{bmatrix} a(k) \\ b(k) \end{bmatrix} - \begin{bmatrix} 2k_1 \epsilon(k) x(k) \\ 2k_2 \epsilon(k) y(k-1) \end{bmatrix}. \tag{C}$$

Because the updating term depends directly on the random process $\{x(k)\}$ the two parameters are also random processes for this example. Thus convergence of the statistics of a and b should be considered as in [2]. For the means of a and b to converge the updating term in (C) must be zero mean

$$E[\epsilon(k)x(k)] = E[\epsilon(k)y(k-1)] = 0 \tag{D}$$

establishing the necessary conditions for a stationary point of the adaptive algorithm in [1]. However, at the global minimum of the squared-error surface in the present example

$$E\{\epsilon(k)x(k)\} = 0.361$$

$$E\{\epsilon(k)y(k-1)\} = 0.124$$

i.e. on the average the algorithm tends to push the parameters (a, b) away from the global least squares solution as evidenced by the experimentally generated locus of successive $(a(k), b(k))$ in Fig. B. In this simple case the algorithm has a unique stationary point $(0.050, -0.852)$, which is found directly by solving the conditions of (D), resulting in $\xi = 0.988$. The small square on the locus represents this stationary point, to which in this case the algorithm converges.

To illustrate that it is not the multimodal nature of the error surface in the previous example, but rather the nongradient character of the algorithm which prevents its convergence to a least squares solution, a unimodal matching problem is briefly presented. The desired signal is generated by passing $\{x(k)\}$ through

$$H(z^{-1}) = \frac{0.05}{1 - 1.75\, z^{-1} + 0.81\, z^{-2}}.$$

The single optimum of the normalized error surface is located at $(a^*, b^*) = (0.132, 0.875)$, as shown in Fig. C, with $\xi = 0.335$. The algorithm's stationary point is $(0.050, 0.967)$ for which $\xi = 0.656$. The filter's nongradient behavior is apparent from the ascending locus for successive $(a(k), b(k))$ shown in Fig. C. Both examples demonstrate an increase in the error claimed by Feintuch to be minimized.

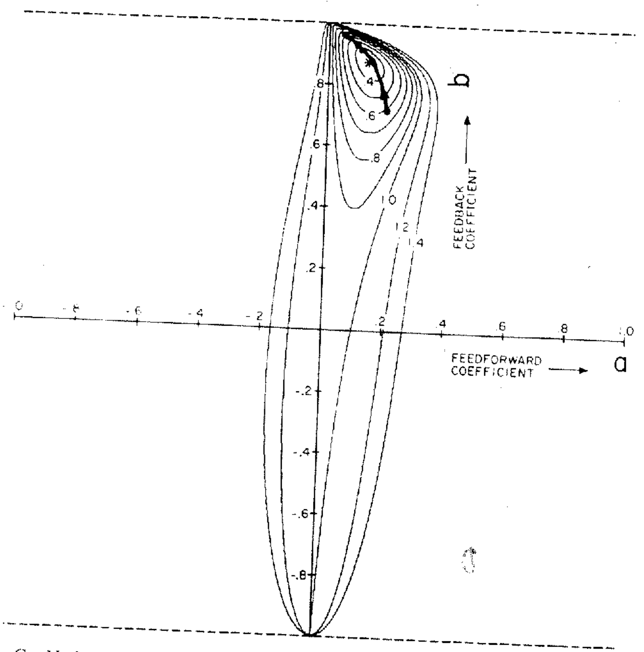

Fig. C. Unimodal case—contour representation of normalized mean-squared error showing locus of adaptive parameters (a, b): $k_1 = -0.00025$, $k_2 = -0.01$, average of 10 ensemble members.

These examples argue for the limitation of the use of the adaptive recursive filter in [1] to cases where the order of the adaptive filter is not less than the order of the minimal filter generating the desired signal from the same input. This applicability restriction has been experimentally recognized in [6] and is implicitly satisfied by the examples in [1].

Development of an adaptive recursive filter, a parallel model reference adaptive structure utilizing output error [12], is a complicated problem that has been avoided in the past due to its complexity [3, pp. 232-233], [10] and [3] and forsaken for a series-parallel model reference adaptive structure based on an equation error formulation. Therefore, due to its successful simulation, the simple technique presented in [1] should not be ignored. But, it should be clearly stated that the static analysis of equations (3)-(8) is invalid for the problem embodied in (9)-(11). It is meaningful only as a fortuitous stimulus. Furthermore, it should be declared that the applicability of the adaptive recursive filter formulated appears severely limited to cases where the adaptive filter is not of less order than the minimal filter necessary for exact matching; and, even then, proper convergence has not been proven by Feintuch, just implied by successful simulation.

References

[1] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, no. 11, pp. 1622-1624, Nov. 1976.

[2] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151-1162, Aug. 1976.

[3] J. M. Mendel, *Discrete Techniques of Parameter Estimation*. New York: Dekker, 1973.

[4] S. A. White, "An adaptive recursive filter," in *Proc. 9th Asilomar Conf. Circuits, Systems, and Computers*, pp. 21-25, Nov. 1975.

[5] J. A. Cadzow, "Recursive digital filter synthesis via gradient based algorithms," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-24, no. 5, pp. 349-355, Oct. 1976.

[6] G. R. Elliott, W. L. Jacklin, and S. D. Stearns, "The adaptive digital filter," Sandia Laboratories, Albuquerque, NM, Rep. SAND76-0360, Aug. 1976.

[7] D. L. Soldan, "Application of the method of least squares to adaptive systems identification," Masters Thesis, Kansas State University, Manhattan, Kansas, Sept. 1976.

[8] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Autonat. Contr.*, vol. AC-21, no. 2, pp. 194-202, Apr. 1976.

[9] E. P. Eremin, N. T. Loan, and G. S. Chkhartishvili, "Searchless identification system with a model synthesized by the hyperstability criterion," *Automat. Remote Contr.* (USSR), vol. 34, no. 5, pt. 1, pp. 728-738, May 1973.

[10] P. E. Mantey, "Convergent automatic-synthesis procedures for sampled-data networks with feedback," Stanford Electronics Laboratories, Stanford University, Stanford, CA, TR no. 6773-1, Oct. 1964.

[11] C. H. Kim and D. P. Lindorff, "Input frequency requirements for identification through Liapunov methods," *Int. J. Contr.*, vol. 20, no. 1, pp. 35-48, July 1974.

[12] I. D. Landau, "A survey of model reference adaptive techniques—theory and applications," *Automatica*, vol. 10, no. 4, pp. 353-379, July 1974.

[13] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans Automat. Contr.*, vol. AC-10, no. 5, p. 461, Oct. 1965.

Reply[4] *by Paul L. Feintuch*[5] *and Neil J. Bershad*[6]

Johnson and Larimore's comments present an interesting example which attempts to model a two-pole and two-zero network by using an adaptive recursive LMS filter that can produce one pole and one gain. Although it is agreed that the algorithm in [1] will not handle this case, the question arises as to the fairness or practicality of such a comparison. It also appears, upon a careful reading, that Johnson and Larimore's comments have not noted the significance of the development in [1]. We therefore shall discuss each step in the derivation in a manner which we hope will make clearer the conditions leading to the adaptive recursive filter algorithm. The equation numbers refer to [1], and letters refer to Johnson and Larimore's comments.

Equation (2) represents the output of a recursive filter in terms of the feedforward and feedback coefficients. For the moment, let us assume than $X(n)$ and $Y(n)$ are correlated sequences but we do not presume that $Y(n)$ is the output of a recursive filter with $X(n)$ as the input.

It is desired to minimize the mean square error between (2) and a desired sequence $d(n)$ as given by (4). If $Y(n)$ is the output of the Wiener filter, then R_{XX}, R_{XY}, R_{YY}, R_{dX} cannot be specified independently. Suppose that the statistics of R_{XX}, R_{dX} are known *a priori*. Then the Wiener filter can be obtained and R_{XY}, R_{YY}, and R_{dY} can be computed from R_{XX}, R_{dX}. Having specified the Wiener filter, say by its impulse response $h_0(t)$, we can ask the following question. Suppose we desire to synthesize the Wiener filter by using the feedback structure described by (1) and for some reason did not want to Laplace transform $h_0(t)$ to obtain the a's and b's. Instead we desire an algorithm for finding A and B from R_{XX}, R_{XY}, R_{YY}, R_{dY}, R_{dX}. Equations (7) and (8) represent a gradient search algorithm for achieving this goal. Note that the error surface in (4) is a quadratic function of A and B. $A(n)$ and $B(n)$, if they converge, will converge to the values given implicitly by (5) and (6). When A and B have sufficient degrees of freedom to model the Wiener filter as a recursive structure, then extensive simulations show that $A(n)$ and $B(n)$ converge to the weights of this recursive structure. If A and B do not have sufficient degrees of freedom to model the Wiener filter, then it is not clear in general to what $A(n)$ and $B(n)$ converge.

Before proceeding to (9) and (10), let us now consider the impact of the results obtained by Johnson and Larimore. They show that the error surface using the recursive filter network $a/(1 - bz^{-1})$ is not a quadratic function of b. They do this by evaluating the mean square error as a function of b. How does this relate to (3) and (4)? If we fix the filter structure as $a/(1 - bz^{-1})$, then we could calculate the correlations for (4). These correlations are *not* the Wiener filter correlations discussed above. Rather they are the correlations for the filter with transfer function $a/(1 - bz^{-1})$. Equations (5) and (6) will determine the values to which a and b in (7) and (8) converge. Letting $\hat{a}, \hat{b}$ denote these values, (5) and (6) yield

$$\hat{a} = H(0)$$

$$\hat{b} = (1 - b^2)b\,\frac{H(b)}{H(0)}$$

where

$$H(z^{-1}) = \frac{0.05 - 0.4z^{-1}}{1 - 1.1314z^{-1} + 0.25z^{-2}}$$

as in the above letter. Note that $\hat{a}$ and $\hat{b}$ depend on the settings for a and b that were used to generate the correlation matrices.

If the values for a and b are not the values for the Wiener filter (or

[4] Manuscript received March 16, 1977.
[5] P. L. Feintuch is with the Hughes Aircraft Co., Fullerton, CA.
[6] N. J. Bershad is with the Electrical Engineering Department, University of California, Irvine, CA, and Consultant to the Hughes Aircraft Company, Fullerton, CA.

the recursive filter does not have enough degrees of freedom to represent the Wiener filter) then $\hat{a}$ and $\hat{b}$ will not, in general, converge to a and b since the assumptions in (3) and (4) are that the correlation functions correspond to those of the Wiener filter. If, for example $H(z^{-1})$, the filter to be modeled, is given instead by

$$H(z^{-1}) = \frac{a}{1 - bz^{-1}}$$

then

$$\hat{a} = a$$

$$\hat{b} = \frac{(1 - b^2)b}{a} \cdot \frac{a}{1 - b^2} = b.$$

That is, the equilibrium values for $\hat{a}$ and $\hat{b}$ are the Wiener filter settings. Thus the recursive algorithm, (7) and (8), converges to the Wiener filter.

Now let us return to (7) and (8). Because none of the statistics are known *a priori*, (9) and (10) replace them by estimates. The estimates of R_{XX}, R_{Xd} are unbiased. The estimates of R_{XY}, R_{YY}, and R_{dY} are possibly biased but more importantly, are functions of A and B.

Johnson and Larimore assume, and are supported by simulation results, that the steady state behavior of the mean weights of the system are statistically independent of the present data. This assumption is implicit in being able to solve equation (D) explicitly for the stationary point of the algorithm in [1]. Thus by their two examples they have supported the hypothesis that (5) and (6) give the values to which the means of (9) and (10) (the recursive LMS algorithm) converge. The recursive LMS algorithm does not converge to the least mean-square error filter in this example since (7) and (8) do not converge to the MMSE weights for a and b. This does not imply that the algorithm is faulty since the examples of the above letter do not legitimately consider unconstrained LMS filters. The problem of finding the best filter within a given structural class is not the basis for deriving a Wiener filter. The Wiener filter can be derived by using the orthogonality principle, making the error orthogonal to all of the data. Equation (D) is not the orthogonality principle since we should require

$$E[\epsilon(k)x(j)] = 0, \quad \text{for all } j \leqslant k.$$

It is suspected that the second condition

$$E[\epsilon(k)y(j)] = 0, \quad \text{for all } j \leqslant k - 1$$

implies the first whenever the feedback structure has sufficient degrees of freedom to represent the Wiener filter.

It is important to note that seeking the Wiener filter implemented in a specified recursive structure is the motivation in [1]. The resulting filter may not provide the minimum mean square error in all cases. The object was to obtain a filter that reduces mean-square error. The resulting algorithm is of interest, since in many cases it does this, and it, rather than its motivation, should be the subject of discussion. The point here is that for the case presented in the above letter, the Wiener filter cannot be implemented for the severe set of constraints on the problem. This is evidenced by the minimum-mean-square error occurring at a point where the error is *not* orthogonal to the data. Since the recursive adaptive algorithm attempts to set the error orthogonal to the data, for this example it provides a higher mean-square error.

Among the features of the adaptive recursive LMS filter is that when given sufficient degrees of freedom, i.e., two feedforward and two feedbackward adaptive taps to model the two pole-two zero fixed parameter network, then the system converges to the proper tap values. This is documented in Table I. In addition, if the adaptive algorithm is given three feedforward and three feedback taps, then it converges more rapidly to a solution with smaller mean square error, but with different tap values (Table II). Interestingly, the transfer function of the recursive adaptive filter matches the network being modeled. That is, given additional degrees of freedom, the algorithm produces a redundant pole-zero pair which cancels. The same property holds for higher order filters. On the other hand, as the number of taps in the adaptive system is reduced below that required to provide a solution, the degradation in performance is not the dramatic threshold behavior implied in the above letter. Instead, for higher order systems, it is gradual, becoming more severe as fewer taps are used, as simulations have shown.

In most adaptive system applications, the order of the system required is not known. The designer allocates more taps to the problem than the minimum number sufficient to provide a solution. The limitation pointed out does not present a restriction in practice and should not detract from the large class of problems which the processor can handle. Determination of that class is an area of active research.

TABLE I
MODELING $H(Z) = (0.05 - 0.40Z^{-1})/(1 - 1.1314Z^{-1} + 0.25Z^{-2})$ WITH THE RECURSIVE ADAPTIVE FILTER WITH TWO FEEDFORWARD AND TWO FEEDBACK TAPS; $k_1 = k_2 = -4.3 \times 10^{-4}$; WEIGHTS INITIALLY ALL ZERO

Number of Iterations, n	$a_0(n)$	$a_1(n)$	$a_2(n)$	$b_1(n)$	$b_2(n)$	$b_3(n)$	Normalized rms error
8192	.0456	-.3633	-.2968	-.3816	-.3845	-.0573	.2096
16384	.0520	-.3649	-.2701	-.4497	-.4431	.0713	.0939
24576	.0503	-.3676	-.2599	-.4805	-.4530	.1273	.0378
32768	.0503	-.3679	-.2552	-.4926	-.4613	.1466	.0143
40960	.0502	-.3682	-.2538	-.4967	-.4639	.1539	.0049
49152	.0500	-.3683	-.2533	-.4982	-.4648	.1566	.0018
57344	.0500	-.3684	-.2531	-.4988	-.4652	.1576	.0007
65536	.0500	-.3684	-.2529	-.4989	-.4653	.1580	.0003

TABLE II
MODELING $H(Z) = (0.05 - 0.40Z^{-1})/(1 - 1.134Z^{-1} + 0.25Z^{-2})$ WITH THE RECURSIVE ADAPTIVE FILTER WITH THREE FEEDFORWARD AND THREE FEEDBACK TAPS; $k_1 = k_2 = -4.3 \times 10^{-4}$; WEIGHTS INITIALLY ALL ZERO

Number of Iterations, n	$a_0(n)$	$a_1(n)$	$b_1(n)$	$b_2(n)$	Normalized Error rms
8192	.0460	-.374	-.636	-.295	.4984
16384	.0568	-.381	-.817	-.095	.3652
24576	.0503	-.393	-.945	.056	.2508
32768	.0518	-.393	-1.035	.141	.1545
40960	.5024	-.398	-1.082	.194	.0807
49152	.0497	-.398	-1.107	.223	.0416
57344	.0497	-.400	-1.121	.238	.0205
65536	.0502	-.400	--1.126	.244	.0090
"correct value"	.05	-.40	-1.131	.25	

The two simulations presented in [1] were representative of a large number of runs under diverse inputs and initial conditions. The adaptive network was given a sufficient number of taps capable to handle the problems. Significantly, it did provide effective solutions. Previous work in this area [2] indicated that an adaptive recursive filter was not technically feasible. Our results have demonstrated otherwise.

REFERENCES

[1] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, no. 11, pp. 1622–1624, Nov. 1976.
[2] P. E. Mantey, "Convergent automatic synthesis procedures for sampled data networks with feedback," Stanford Electronics Lab., Stanford University, Stanford, CA, Tech. Rep. 6773-1, Oct. 1964.

Comments on "An Adaptive Recursive LMS Filter"

BERNARD WIDROW AND JOHN M. MCCOOL

The usual application of the "least mean-square" or LMS algorithm of Widrow and Hoff [1] is to nonrecursive or feedback-free adaptive systems. An example of such a system is the adaptive transversal filter, which has been shown to be capable, when its operation is governed by the LMS algorithm, of adjusting itself to minimize mean-square error, [2], [3] where "error" is defined as the difference between the filter's

Manuscript received January 31, 1977.
B. Widrow is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305.
J. McCool is with the Fleet Engineering Department, Naval Undersea Center, San Diego, CA 92132.

output and a "desired response" or externally supplied training signal. The drawback of the nonrecursive LMS filter is that it has a finite impulse response and can realize only zeros of a digital filter transfer function.[1]

In the above letter,[2] a recursive adaptive filter based on the LMS algorithm has been described. This particular filter is structurally capable of realizing both zeros and poles of a transfer function and of having an infinite impulse response. It thus promises to be a useful and powerful tool in certain practical applications, indicated below. We must regrettably report, however, that the mathematical derivation of the recursive LMS algorithm presented in the letter is incorrect and that, contrary to Feintuch's claim, the algorithm does not in general minimize mean-square error.[3]

Feintuch specifies the recursive filter by the equation

$$y(n) = \sum_{k=0}^{N_f} a_k s(n-k) + \sum_{k=1}^{N_B} b_k y(n-k). \tag{1}$$

In vector notation he obtains

$$y(n) = A^T X(n) + B^T Y(n). \tag{2}$$

He defines the error as the difference between the desired response $d(n)$ and the actual response $y(n)$

$$\epsilon(n) = d(n) - y(n) = d(n) - A^T X(n) - B^T Y(n). \tag{3}$$

He then squares and takes expected values to obtain the mean-square error

$$E[\epsilon^2(n)] = E[d^2(n)] + A^T R_{XX} A + B^T R_{YY} B - 2A^T R_{dX} - 2B^T R_{dY} + 2A^T R_{XY} B \tag{4}$$

where the covariance terms are defined as

$$R_{XX} = E[X(n)X^T(n)], \quad R_{YY} = E[Y(n)Y^T(n)], \quad R_{dX} = E[d(n)X(n)], \quad R_{dY} = E[d(n)Y(n)], \quad \text{and } R_{XY} = E[X(n)Y^T(n)].$$

Since all algorithms in the LMS family [4]–[19] are based on optimization by the method of steepest descent, the next step required is differentiation of (4) to obtain the gradient. In taking this step, however, Feintuch argues that the covariance terms R_{XY}, R_{dY}, R_{YY} are constants when differentiated with respect to the feed-forward and feedback weights A and B. This argument is incorrect because these terms are functions of A and B. The gradient expressions given by his equations (5) and (6) are thus also incorrect and the derivation of his remaining (7)–(11) invalid.

Let us examine the recursive LMS filter from another point of view. Fig. 1 shows a nonrecursive filter comprising an adaptive transversal filter whose impulse response is controlled by adjusting its weighting coefficients. This filter is an "LMS filter" when the coefficients are adjusted through the LMS algorithm. Fig. 2 shows a recursive filter comprising two adaptive transversal filters, one providing a feed-forward network and implementing zeros and the other providing a feedback network and implementing poles.[4] When the LMS algorithm is used to adjust the weights of both filters, the result is a "recursive LMS filter" identical to the one described by Feintuch. If the input signal and

[1] The terms "zero," "pole," and "transfer function" belong to the domain of of fixed filters; they are nevertheless useful in the analysis of adaptive filters, though their meaning in this context cannot yet be precisely defined.

[2] P. L. Feintuch, *Proc. IEEE*, vol. 64, pp. 1622–1624, Nov. 1976.

[3] As well as is known at the present time, the algorithm minimizes some function of the error. Experimental evidence indicates that this function may in general be *unimodal*, but what it is remains a subject of research. The mean-square error function is in general *multimodal*. When using the recursive LMS algorithm, in some cases mean-square error apparently is minimized but in others it clearly is not. Under certain conditions we have observed the recursive LMS algorithm, initially set at the minimum mean-square-error solution, to cause the weights to vary from this solution and stabilize elsewhere.

[4] The unit delay in the feedback network eliminates instantaneous feedback, whose only function would be gain control, which is accomplished by the feed-forward weights.

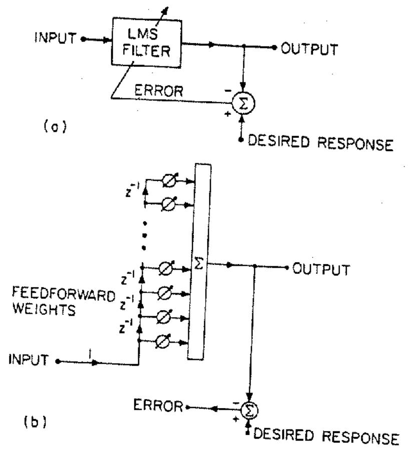

Fig. 1. Nonrecursive LMS Filter. (a) Simplified representation. (b) Schematic diagram.

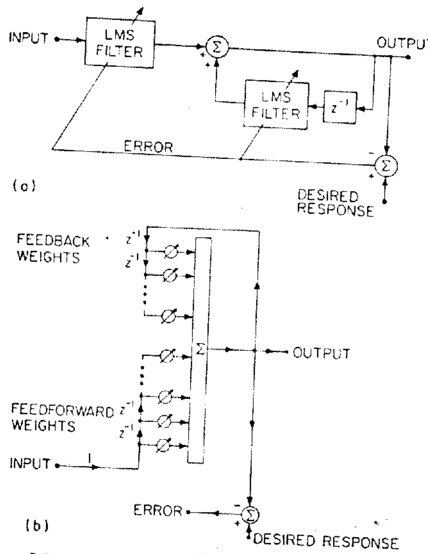

Fig. 2. Recursive LMS Filter. (a) Simplified representation. (b) Schematic diagram.

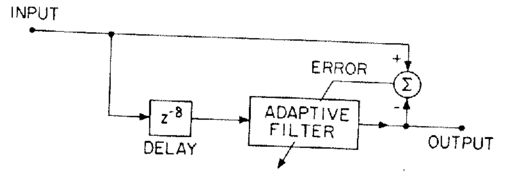

Fig. 3. The adaptive line enhancer.

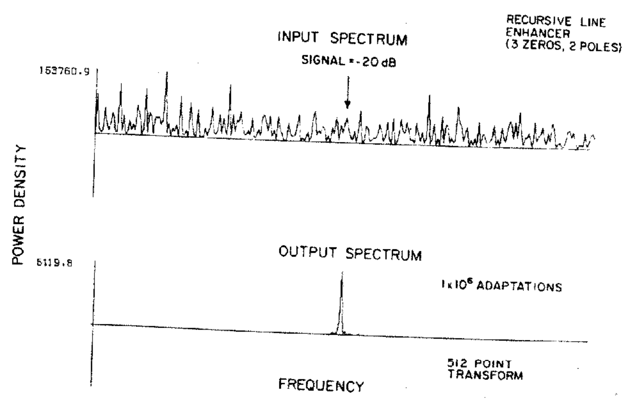

Fig. 4. Performance of a recursive adaptive line enhancer.

desired response of this filter are statistically stationary, it is clear that the covariance matrix of the inputs to the weights of the feed-forward filter is fixed and independent of the weight values, while the covariance matrix of the inputs to the feedback filter is dependent on the values of both the feed-forward and feedback filter weights. The latter dependence is characteristic of adaptive feedback systems.

The recursive adaptive filter was studied more than ten years ago by P. E. Mantey as a doctoral student at Stanford. He showed that the mean-square-error function was not quadratic and was sometimes multimodal. Mantey did not pursue his work because of the unpredictability of this filter and the difficulty of understanding its behavior. Instead he devised a recursive adaptive process using the desired response as feedback signal rather than the filter output [17]. His goal was to achieve constancy in the covariance terms and to obtain a quadratic mean-square-error function with a linear gradient. Feintuch's mathematical derivation corresponds to Mantey's second algorithm rather than to the recursive LMS algorithm.

Despite the foregoing qualifications Feintuch's work is an important contribution. He has stimulated new interest in the recursive LMS filter and has shown experimentally that it performs well as a substitute for the nonrecursive transversal filter in the self-tuning adaptive filter or "adaptive line enhancer" described by Widrow *et al.* [18] and shown in Fig. 3. In our work with the line enhancer, we have confirmed experimentally that low-level narrowband signals in noise can be effectively detected by a recursive LMS filter with poles close to the unit circle.[5] Fig. 4 shows, for example, plotted on linear scales, the input and output power spectral densities of a signal before and after processing by a filter with three zeros and two poles. The measured improvement in signal-to-noise ratio, with only five adaptive weights, is approximately 40 dB. We are thus confident that the recursive LMS filter has important potential applications in the fields of signal detection, instantaneous frequency estimation [19], and spectral analysis.

REFERENCES

[1] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," in *IRE WESCON Conv. Rec.*, pt. 4, 1960, pp. 96-104.
[2] B. Widrow, "Adaptive filters," in *Aspects of Network and System Theory*. R. Kalman and N. DeClaris, Eds. New York: Holt, Rinehart, and Winston, 1971, pp. 563-587.
[3] B. Widrow, J. McCool *et al.*, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151-1162, Aug. 1976.
[4] J. Koford and G. Groner, "The use of an adaptive threshold element to design a linear optimal pattern classifier," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 42-50, Jan. 1966.
[5] N. Nilsson, *Learning Machines.* New York: McGraw-Hill, 1965.
[6] B. Widrow, P. Mantey, L. Griffiths, and B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, pp. 2143-2159, Dec. 1967.
[7] L. J. Griffiths, "A simple adaptive algorithm for real-time processing in antenna arrays," *Proc. IEEE*, vol. 57, pp. 1696-1704, Oct. 1969.
[8] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926-935, Aug. 1972.
[9] L. J. Griffiths and P. E. Mantey, "Iterative least-squares algorithm for signal extraction," in *1969 Proc. Second Hawaii Int. Conf. System Sciences*, Western Periodicals Co., pp. 767-770.
[10] R. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, pp. 547-588, Apr. 1965.
[11] R. Lucky *et al.*, *Principles of Data Communication.* New York: McGraw-Hill, 1968.
[12] R. Riegler and R. Compton, Jr., "An adaptive array for interference rejection," *Proc. IEEE*, vol. 61, pp. 748-758, June 1973.
[13] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proc. IEEE*, vol. 63, pp. 719-720, Apr. 1975.
[14] W. F. Gabriel, "Adaptive arrays–An introduction," *Proc. IEEE*, vol. 64, pp. 239-272, Feb. 1976.
[15] B. Widrow and J. McCool, "Comparison of adaptive algorithms based on the methods of steepest descent and random search," *IEEE. Trans. Antennas Propagat.*, vol. AP-24, pp. 615-637, Sept. 1976.
[16] J. E. Timm, "The Delta-LMS adaptive algorithm," *Proc. of the 3rd IFAC Symposium*, New York: American Elsevier, June 1973, pp. 721-728.
[17] P. E. Mantey, "Convergent automatic synthesis procedures for sampled-data networks with feedback," Stanford Electronics Lab., Stanford University, Tech. Rep. 6773-1, Oct. 1964 (Ph.D. Dissertation).
[18] B. Widrow, J. Glover, J. McCool *et al.* "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.
[19] L. J. Griffiths, "Rapid measurement of instantaneous frequency," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 209-222, Apr. 1975.

[5] By experiment it appears that the recursive LMS algorithm has the extraordinary property of remaining stable even though noise in the feedback weights may occasionally push the poles outside the unit circle. The feedback of the adaptive process interacts favorably with the feedback of the filter itself to produce a "superstability" that will pull the poles back in from beyond the brink of instability.

Reply[6] by Paul L. Feintuch[7]

I should like to thank Professor Widrow and Dr. McCool for their comments and simulation examples which agree with mine [1]. I too have found that the device is stable and has potential for signal detection and spectral analysis applications. However, I must disagree with their claim that the analysis is incorrect. A heuristic derivation was presented for which the key assumptions were clearly pointed out. What is appearing here is a differing viewpoint rather than mathematical error. There is no question that when viewed as in the above letter, the output correlation statistics are functions of the weights. I was aware of this by referencing White [2]. However, the approach, instead, was that of a Wiener filter for which *a priori* statistical information is used to dictate the filter parameters rather than the reverse. For the fixed parameter case, it is certainly valid to view the problem in this way. The procedure was to assume that we have the Wiener filter and are using its input and output statistics to determine the parameters in the recursive digital filter structure. At this stage the correlation matrices are not functions of the filter weights and are thus constants when forming the gradient vectors. The gradient search procedure removes the need to invert matrices, and the problem reduces to one of obtaining the statistics, just as it did in the transversal LMS case. The open question, as was noted, is the validity of replacing the output correlations, which the Wiener filter would produce, by estimates using the instantaneous output values. These estimates are biased at the outset but have asymptotic properties which, though not yet understood, must be desirable to produce the simulation results that we have both been observing. The heuristic derivation was presented to show that a logical procedure suggested the processor structure, rather than its being an ad-hoc hookup. I suspect that when the properties of the assumed estimates are understood, the behavior of the entire filter will be understood as well.

I have also carefully examined Mantey's results [3] and do not see how his work could lead to my algorithm. Instead, his results indicate that an adaptive feedback structure could *not* have desirable stability or steady-state properties. He thus pursued network modeling in a feedforward manner only, terminating further research on a recursive adaptive filter.

The processor is by no means completely understood and its analysis is complicated. However, the device produces exciting results.

REFERENCES

[1] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, pp. 1622-1624, Nov. 1976.
[2] S. A. White, "An adaptive versive digital filter," presented at Proc. 9th Asilomar Conf. Circuits, Systems, and Computers, Nov. 3, 1975.
[3] P. E. Mantey, "Convergent automatic synthesis procedures for sampled data networks with feedback," Stanford Electronics Lab., Stanford University, Tech. Rep. 6773-1, Oct. 1964.

[6] Manuscript received March 16, 1977.
[7] The author is with Hughes Aircraft Company, Fullerton, CA.

The Analysis of a Third-Order System

VIMAL SINGH

Abstract–An example of third-order nonlinear feedback system is considered. Its previously known sector for global asymptotic stability is (0, 1]. In the present letter, several sectors for global asymptotic stability are obtained.

We consider a third-order nonlinear feedback system whose linear part $\tilde{G}(s)$ is given by

$$\tilde{G}(s) = \frac{s^2}{(s^2+1)(s+1)}. \tag{1a}$$

Manuscript received March 21, 1977.
The author is with the Department of Electrical Engineering, M.N.R. Engineering College, Allahabad 211004, India.

$f(x)$ and $g(x)$ describe the taper of the line, and x is the distance along the line. The dual L_d of L is characterized by (2):

$$r_d(x) = r_0 g(x) \qquad c_d(x) = c_0 f(x), \qquad 0 \leqslant x \leqslant l. \tag{2}$$

Let $z_1(s)$ be the input impedance of the short-circuited line characterized by (1) and let $z_2(s)$ be that of an ideal gyrator, of gyration resistance r_g, terminated by the open-circuited line (2). Let

$$R_T = r_0 l \qquad C_T = c_0 l \qquad S = R_T C_T s \tag{3}$$

so that

$$z_1(s) = f(0) R_T p(s) \qquad Z_2(s) = f(0)\left(\frac{r_g^2}{R_T}\right) q(s) \tag{4}$$

where, from [2],

$$q(s) = Sp(s). \tag{5}$$

Let

$$u(s) = p(s)/q(s) \tag{6a}$$

so that, from (5),

$$u(s) = 1/S = 1/R_T C_T s. \tag{6b}$$

It is easy to see that when the line L is a $\overline{\text{URC}}$, the lines (L, L_d) form a pair of commensurate $\overline{\text{URC}}$'s and 6(b) reduces to (3) of [1]. Thus the synthesis procedure outlined in [1] can be employed to realize a rational transfer function using any RC line and its dual.

III. The Effect of Departure From Duality

The proposed synthesis procedure is based on the assumption that the lines are duals of one another, which ensures that $u(s)$ is rational in s. Therefore, the effect of nonduality due to the variation of any of the tapers or distributed parameters should be studied. It is not possible to study the effect of these two factors on $u(s)$ for arbitrary distributions of the lines, since solutions of the telegrapher's equations are not then available. Even for some solvable lines, it may neither be desirable nor practical to study this effect. However, it has been shown that a tapered RC line is equivalent to an infinite cascade of commensurate $\overline{\text{URC}}$'s [3]. A study of the effect of the nonduality between a cascade of two $\overline{\text{URC}}$'s and its dual on $u(s)$ has consequently been made. The two $\overline{\text{URC}}$'s considered were commensurate but had different total resistances and capacitances. Further, the study was made by examining the classical sensitivity [4] $S_{x_i}^{u(jw)}$ of $u(jw)$ with respect to x_i, where x_i is any one of the distributed parameters. In this study, the frequency range over which $S_{x_i}^{\angle u(jw)}$ and $S_{x_i}^{|u(jw)|}$ are both zero is of interest. In this case, even if the distributed parameters deviate from their nominal values, u remains rational in s over this range. However, the frequency range over which $S_{x_i}^{|u(jw)|}$ is constant and small is also of interest, provided $S_{x_i}^{\angle u(jw)}$ is very small over this range. Over such a range, instead of being proportional to $(1/jw)$, $u(jw)$ will be proportional to $((1+\epsilon)/jw)$, where ϵ is a real but small number. Hence, it is seen that $u(jw)$ will also remain rational over this range.

The above study strongly led to the following *conjecture*.

i) If the lines are not exactly duals of each other, there are two regions of w, namely, a lower and an upper region, over which u will remain rational in w. In between these two regions u cannot remain rational.

ii) The range of the lower region will increase with decreasing values of the taper.

In order to justify the conjecture, the following two exponential lines were used to form the impedances $z_1(s)$ and $z_2(s)$ given by (4):

$$\begin{aligned} r_1(x) &= r_{01}\, e^{2k_1 x} \qquad c_1(x) = c_{01}\, e^{-2k_1 x}, \qquad 0 \leqslant x \leqslant l_1 \\ r_2(x) &= r_{02}\, e^{2k_2 x} \qquad c_2(x) = c_{02}\, e^{-2k_2 x}, \qquad 0 \leqslant x \leqslant l_2. \end{aligned} \tag{7}$$

If the lines are duals of each other,

$$r_{01} = r_{02} = r_0 \qquad c_{01} = c_{02} = c_0 \qquad l_1 = l_2 = l \qquad k_1 = -k_2 = k \tag{8}$$

and it is readily shown that

$$u = \left(\frac{1}{R_T C_T s}\right) = 1/S \qquad R_T C_T = (r_0 l)(c_0 l). \tag{9}$$

If the lines are not dual, we have

$$U(s) = \frac{1}{\alpha_2 S} \frac{\sinh\sqrt{(k_1 l_1)^2 + \alpha_1 S}}{\begin{aligned}&(\cosh\sqrt{(k_1 l_1)^2 + \alpha_1 S}) \cdot (\sqrt{(k_1 l_1)^2 + \alpha_1 S}) - (k_1 l_1)\sinh\sqrt{(k_1 l_1)^2 + \alpha_1 S} \\ &\cdot (\cosh\sqrt{(k_2 l_2)^2 + \alpha_2 S}) \cdot (\sqrt{(k_2 l_2)^2 + \alpha_2 S}) - (k_2 l_2)\sinh\sqrt{(k_2 l_2)^2 + \alpha_2 S}\end{aligned}} \tag{10}$$

where

$$(r_{0j} l_j) \cdot (c_{0j} l_j) = R_{jT} C_{jT} = \alpha_j R_T C_T (j = 1, 2). \tag{11}$$

We note that $S_{k_1 l_1}^u = -S_{k_2 l_2}^u$ and $S_{\alpha_2}^u = -1 - S_{\alpha_1}^u$. A close study of the plots of $S_{x_i}^u$ (where x_i is any one of the parameters of the lines) indicates strong agreement with the conjecture. For lack of space, only some of the plots are presented in Fig. 1. Note that the curves labelled 1, which correspond to the case of $\overline{\text{URC}}$'s, are identical to Figs. 1(c) and (d) in [1].

Acknowledgment

The authors would like to thank one of the anonymous reviewers for suggestions to improve the presentation of the material.

References

[1] M. S. Abougabal, B. B. Bhattacharyya, and M. N. S. Swamy, "A low sensitivity active distributed realization of rational transfer functions," *IEEE Trans. Circuits Systems*, vol. CAS-21, pp. 391–395, May 1974.

[2] B. B. Bhattacharyya and M. N. S. Swamy, "Dual distributions of solvable nonuniform lines," *Proc. IEEE*, vol. 54, pp. 1979–1980, Dec. 1966.

[3] J. C. Giguere, "The approximation of a non-uniform transmission line by a cascade of uniform lines," *Electron. Lett.*, vol. 7, pp. 511–512, Sept. 1971.

[4] S. K. Mitra, *Analysis and Synthesis of Linear Active Networks*. New York: Wiley, 1969.

An Adaptive Recursive LMS Filter

P. L. FEINTUCH

***Abstract*—An adaptive, recursive, least mean-square digital filter is heuristically derived that has the computational simplicity of existing transversal adaptive filters, with the additional capability of producing poles in the filter transfer function. Simulation results are presented to demonstrate its capability.**

Introduction

Adaptive digital filters that are constrained to a transversal tapped delay structure appear in the literature [1]. These filters converge to optimal processors in the mean-square-error sense, do not require *a priori* knowledge of second-order statistics of the observed and desired waveforms, are easily implemented in real-time with little storage, and can track parameters that vary slowly with respect to the convergence time of the iterative process. Such filters have a finite impulse response, i.e., they can produce only zeros with no poles in the filter transfer function. This limits the capability of transversal adaptive filters in many applications. To overcome this limitation, a new adaptive filter structure is described which is capable of producing poles in the transfer function and is easily implemented using two transversal adaptive filters.

Analysis

The recursive filter structure, in the time domain, is described by the input–output relationship

$$y(n) = \sum_{k=0}^{N_F} a_k x(n-k) + \sum_{k=1}^{N_B} b_k y(n-k). \tag{1}$$

Manuscript received February 9, 1976; revised June 21, 1976.
The author is with the Marine Systems Division, Rockwell International, Anaheim, CA 92803.

The set $\{a_k\}$ is referred to as the set of feed-forward coefficients and the set $\{b_k\}$ are the feedback coefficients. In vector notation, let $A^T = [a_0, a_1, \cdots, a_{N_F}]$, $B^T = [b_1, b_2, \cdots, b_{N_B}]$, $X(n)^T = [x(n), x(n-1), \cdots, x(n-N_F)]$, and $Y(n)^T = [y(n-1), \cdots, y(n-N_B)]$. Then equation (1) can be rewritten as

$$y(n) = A^T X(n) + B^T Y(n). \tag{2}$$

Assume that the observables are wide-sense-stationary and zero mean. The filter is used to estimate a desired waveform, $d(n)$, in a minimum mean square error sense. Let $\epsilon(n)$ denote the error waveform at the nth time sample, and $E(\cdot)$ denote the expectation operation. Then

$$\epsilon(n) = d(n) - y(n) = d(n) - A^T X(n) - B^T Y(n) \tag{3}$$

$$E[\epsilon^2(n)] = E[d^2(n)] + A^T R_{XX} A + B^T R_{YY} B - 2A^T R_{dX} - 2B^T R_{dY} + 2A^T R_{XY} B \tag{4}$$

where

$$R_{XX} = E[X(n)X^T(n)], \quad R_{YY} = E[Y(n)Y^T(n)], \quad R_{dX} = E[d(n)X(n)]$$
$$R_{dY} = E[d(n)Y(n)], \quad \text{and} \quad R_{XY} = E[X(n)Y^T(n)].$$

The theory of Wiener filtering employs known second order input statistics to dictate the impulse response, $h(t)$, of the linear filter that minimizes the mean-square error. Once $h(t)$ is known, output statistics can be calculated *a priori*. The algorithm for the transversal adaptive filter uses prior knowledge of the cross-correlation between the observed data and the desired waveform (R_{dX}), and the autocorrelation of the observed data (R_{XX}). The recursive adaptive algorithm requires, in addition, the autocorrelation of the output (R_{YY}), the cross-correlation of the output and the input (R_{XY}), and the cross-correlation of the output with the desired waveform (R_{dY}), which can be calculated using $h(t)$ as mentioned above. Thus the set of statistics at the Wiener filter output is assumed, for the moment, to be known, and will be used to determine the weights in the feedforward and feedback filters. The statistics for the fixed parameter network are not a function of the weights, but instead the weights are a function of these statistics. Therefore, R_{XY}, R_{dX}, and R_{YY} are constants when differentiating with respect to the vectors A and B. The set of weights that minimize the mean squared error is found by setting the gradient vector with respect to the filter parameters equal to zero;

$$\nabla_A [E(\epsilon^2(n))] = 2R_{XX}A - 2R_{dX} + 2R_{XY}B = 0$$
$$A = R_{XX}^{-1}(R_{dX} - R_{XY}B) \tag{5}$$
$$\nabla_B [E(\epsilon^2(n))] = 2R_{YY}B - 2R_{dY} + 2R_{XY}^T A = 0$$
$$B = R_{YY}^{-1}(R_{dY} - R_{XY}^T A). \tag{6}$$

Thus, one can solve for the filter coefficients *if all the second-order statistics are known*. These statistics are not known in general, and the matrix inversions are intractable even if the matrices are known. The filter is made adaptive to estimate the unknown statistics. An iterative gradient search technique (the method of steepest descent) is used. It updates the filter coefficients with steps proportional to the gradient vector;

$$A(n+1) = A(n) + k_1 \nabla_A [E(\epsilon^2(n))]$$
$$= A(n) + 2k_1 [R_{XX}A(n) - R_{dX} + R_{XY}B(n)] \tag{7}$$
$$B(n+1) = B(n) + k_2 \nabla_B [E(\epsilon^2(n))]$$
$$= B(n) + 2k_2 [R_{YY}B(n) - R_{dY} + R_{XY}^T A(n)]. \tag{8}$$

The LMS algorithm [1] replaces the unknown matrices with instantaneous estimates of their values. The *a priori* statistic R_{XX} is estimated by $X(n)X^T(n)$ at the nth iteration. Similarly R_{dX} is estimated by $d(n)X(n)$, R_{XY} by $X(n)Y^T(n)$, R_{dY} by $d(n)Y(n)$, and R_{YY} by $Y(n)Y^T(n)$. It is shown in [1] that $d(n)X(n)$, and $X(n)X^T(n)$ provide unbiased estimates of R_{dX} and R_{XX}, respectively. The quality of the estimates of output correlations using the instantaneous estimates is currently an open question. Proceeding formally with the LMS procedure results in the following vector difference equations:

$$A(n+1) = A(n) + 2k_1[X(n)X^T(n)A(n) - d(n)X(n) + X(n)Y^T(n)B(n)]$$
$$= A(n) - 2k_1 \epsilon(n) X(n) \tag{9}$$

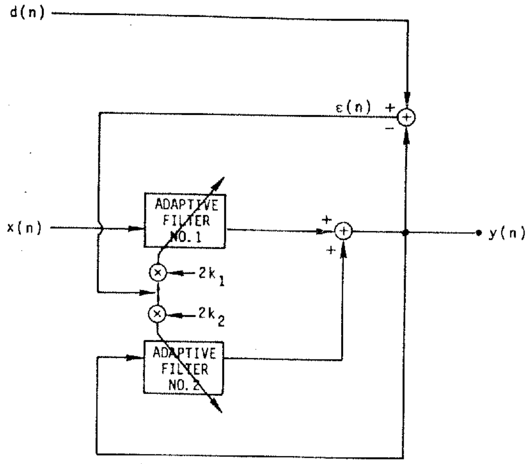

Fig. 1. Adaptive recursive LMS filter using two transversal adaptive filters.

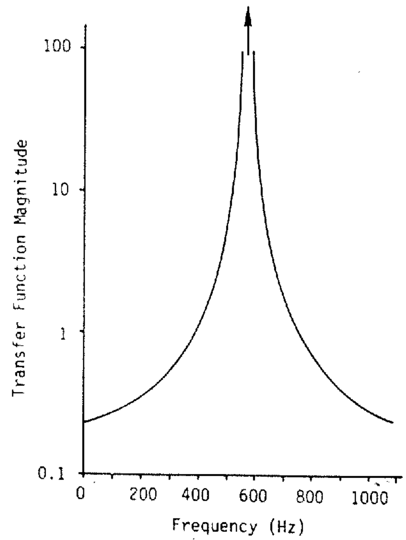

Fig. 2. Example producing a pole in the steady-state transfer function.

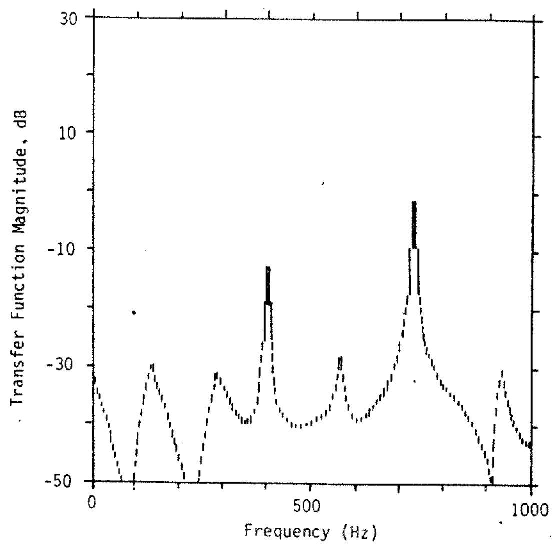

Fig. 3. Simulation example detecting two line components of signal-to-noise ratios −10 dB at 750 Hz and −26 dB at 400 Hz.

$$B(n+1) = B(n) + 2k_2[Y(n)Y^T(n)B(n) - d(n)Y(n) + Y(n)X^T(n)A(n)]$$
$$= B(n) - 2k_2\epsilon(n)Y(n) \quad (10)$$
$$y(n) = A(n)^T X(n) + B(n)^T Y(n). \quad (11)$$

The filter structure is easily implemented, as shown in Fig. 1. Note that the storage required for the weight coefficients is significantly less than previous adaptive recursive filters [2].

Simulations

Fig. 2 shows the steady-state transfer function of the recursive adaptive filter with 8 forward and 8 feedback taps. The desired waveform was white noise low-pass band-limited to 1150 Hz. The filter input was the desired signal passed through a filter, with transfer function $1 + Z^{-2}$ to produce a zero at 575 Hz. The Wiener filter for this problem is the inverse of the fixed parameter filter. As shown, the adaptive filter approximated a pole at the correct frequency.

Fig. 3 shows the steady-state transfer function of the recursive adaptive filter with 16 forward and 16 feedback taps. The desired waveform consists of two sinusoids in flat broad-band noise, of signal-to-noise ratio −26 dB at 400 Hz, and −10 dB at 750 Hz, in a 1-kHz band. The filter input was a delayed replica of the desired waveform, such that the noise was decorrelated due to the delay. The figure shows that the adaptive filter correctly sensed the correlated waveforms. The 20-dB difference between response at 400 and 750 Hz, and spurious peaks, shows that the filter attempted to produce poles at these frequencies.

References

[1] B. Widrow, "Adaptive filters," in *Aspects of Network and System Theory*, R. E. Kalman and N. DeClaris, Eds. New York: Holt, Rinehart, and Winston, 1970.

[2] S. A. White, "An adaptive recursive digital filter," in *Proc. 9th Asilomar Conf. on Circuits, Systems, and Computers*, Nov. 3, 1975.

On the Probability of Success in a Routing Process

PRATHIMA AGRAWAL

Abstract—Lee's path search mechanism has been identified as a percolation process and the probability of a successful search is obtained analytically.

Routing of connections on a printed circuit board (PCB) requires finding nonintersecting paths between electrically common points. Digital computers are used whenever the number of connections to be routed is large. Among the many routing methods, the one using Lee's algorithm [1], [2] has the advantage of guaranteeing a path if it exists. Some authors [2], [3] have observed that for the Lee router, as the routing of wires progresses on a PCB, it becomes more and more difficult to route new connections, and that the probability of finding paths becomes practically zero after a certain cutoff density of wiring has been reached. It appears that the way in which a path is found by the Lee router is similar to the manner in which a fluid introduced at a source atom in a porous medium percolates through the surrounding atoms. The study of such phenomena is known as percolation theory [4]–[6]. For a given probability of blocked cells in the medium there is a probability, known as the percolation probability, with which the fluid from a single source atom will wet infinitely many other atoms. The critical probability is the lowest probability of blockings at which the percolation probability becomes zero. This is similar to the cutoff density of wiring observed in the Lee router.

Lee's algorithm [1] is a systematic procedure of searching for shortest Manhattan paths between two given points. The routing area, such as a PCB, is divided into a square grid. The search for a path between a source cell S and a target cell T proceeds by expanding about S until T is found or further expansion is not possible. When blockings are few, the expanded cells form diamonds centered around S, as shown in Fig. 1. Let us assume that each cell on the PCB is blocked with a probability q and that blocked cells are uniformly distributed

Manuscript received March 22, 1976. This work was supported by NSF Grant ENG74-18647.

P. Agrawal is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90007.

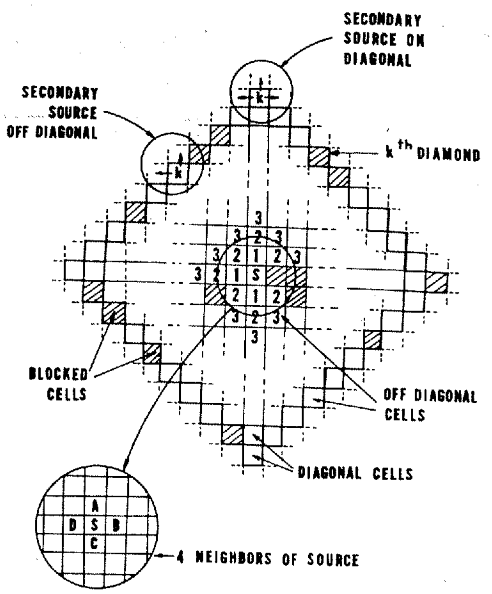

Fig. 1. Growth of search pattern generated by Lee router.

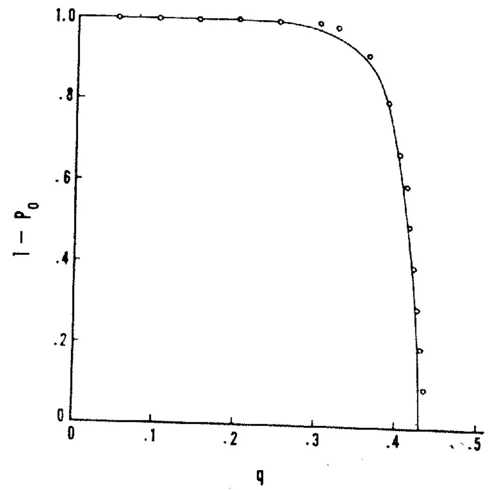

Fig. 2. Probability of successful search as a function of q.

throughout the PCB. The cells on a given diamond can be classified as either diagonal cells with three neighbors to expand, or off-diagonal cells with no more than two cells as neighbors. The source cell has four neighboring diagonal cells.

We now obtain the probability that the Lee's search process, starting at a source cell, will terminate. The complement of this probability corresponds to the event that the search can continue indefinitely. Notice that a path between two cells will not exist if and only if at least one of the cells is surrounded by obstacles. Similar probability for percolation process has only been obtained by the Monte Carlo method [5]. Consider the cells forming a diamond-shaped boundary around the source cell at a Manhattan distance k. All the cells in this boundary, at which the search has successfully arrived, are termed as secondary sources. We define p_k as the probability that the search starting at a secondary off-diagonal source cell will terminate. Let P_k be the similar probability for a secondary diagonal source cell. It can be shown that p_k depends on $p_{k+1}, p_{k+2}, \cdots$, and that $p_k \geq p_{k+1} \geq p_{k+2} \cdots$. In particular,

$$p_k = q^2 + 2q(1-q)p_{k+1} + (1-q)^2[q + (1-q)p_{k+2}]p_{k+1}. \quad (1)$$

As we decrease k from a large value, p_k converges to some finite value $p \leq 1$. This is obtained by substituting $p_{k+2} = p_{k+1} = p_k = p$ in (1). Thus

$$p = q^2/(1-q)^3, \qquad 0 \leq q \leq 0.43$$
$$= 1, \qquad 0.43 \leq q \leq 1. \quad (2)$$