# ADAPTIVE FILTERS REQUIRING ZERO MULTIPLICATIONS

V. John Mathews

Department of Electrical Engineering
University of Utah
Salt Lake City, Utah 84112

## ABSTRACT

This paper introduces an adaptive filter structure that requires zero multiplications for its implementations. The primary input signals are quantized using DPCM and the DPCM outputs are processed by the adaptive filter. The filter coefficients are updated using the sign algorithm. We show that if the parameters are chosen properly, hardware implementation of this filter structure requires no multipliers. Under the assumption that the signals are zero mean, wide-sense stationary, and Gaussian random processes, we derive theoretical results for the mean and mean-squared behavior of the filter. A simulation example is presented that shows very good match between theoretical and empirical results.

## I. INTRODUCTION

Time invariant digital FIR filters requiring zero multiplications for their implementation were introduced by Peled and Liu [7] employing delta modulation techniques. Lee and Un [3] have studied the properties of such filters. Lee, Un and Lee [4] recently introduced an adaptive filter structure that processes a differentially pulse code modulated (DPCM) primary input signal. They used the popular least mean squared (LMS) algorithm [12] to update the coefficients of their filter and therefore requires N multiplications (N is the order of the filter) every time the coefficients are updated. The purpose of this paper is to introduce and study a new adaptive filter structure that requires no multiplications for its implementation. The structure is similar to that of Lee, Un and Lee [4], but employs the sign algorithm (SA) [5] for updating the filter coefficients, thereby eliminating the need for multiplications for its implementation.

Consider the FIR adaptive filtering problem of estimating the desired sequence d(n) using the primary input vector sequence X(n). In this paper, we consider the single channel case, where the input vector X(n) is formed by n most recent samples of the input sequence x(n), i.e., $X^T(n) = \{x(n), x(n-1), \ldots x(n-N+1)\}$. Let H(n) denote the N-vector of adaptive filter coefficients at time n. Then the following set of equations describe the multiplication-free adaptive digital filter (MADF) that we will investigate in this paper.

$$\hat{X}(n) = \beta\tilde{X}(n-1) \quad , 0<|\beta|<1, \tag{1}$$

$$\underline{\varepsilon}(n) = X(n) - \hat{X}(n) \quad , \tag{2}$$

$$B(n) = Q\{\underline{\varepsilon}(n)\} \quad , \tag{3}$$

where $Q\{\cdot\}$ is the vector consisting of quantized values of elements of $\underline{\varepsilon}(n)$. We employ uniform quantization here and the quantization step size $\Delta$ is a (possibly negative) integer power of two,

$$\tilde{X}(n) = \hat{X}(n) + B(n) \quad , \tag{4}$$

$$f(n) = H^T(n)\, B(n) \quad , \tag{5}$$

$$\tilde{g}(n) = g(n) + f(n) \quad , \tag{6}$$

$$g(n) = \beta\tilde{g}(n-1) \quad , \tag{7}$$

$$e(n) = d(n) - \tilde{g}(n) \tag{8}$$

and

$$H(n+1) = H(n) + \mu\, X(n) \operatorname{sign}\{e(n)\}. \tag{9}$$

The structure is depicted in Fig. 1. In Eq. 9,

$$\operatorname{sign}\{e(n)\} = \begin{cases} 1 & ; e(n) > 0 \\ -1 & ; e(n) < 0 \end{cases} \tag{10}$$

If in Eqs. 1-10, we choose $\beta$ and $\mu$ to be negative integer powers of two (or 1 - a power of two), we can see that the implementation of MADF requires multiplications only to compute f(n) in Eq. 5. However, since the quantization step size $\Delta$ is also an integer power of two, each multiplication in Eq. 5 can be done by a very few number of shifts and adds, as long as the number of quantization levels are relatively small. For example, if the number of quantization levels are 5 or less ($-2\Delta, -\Delta, 0, \Delta, 2\Delta$), each multiplication can be done using a single shift. Similarly, for up to 13 levels ranging from $-6\Delta$ to $6\Delta$, each multiplication can be done by at most two shifts and one addition operation. Thus, we can see that we have a system that needs practically no multipliers at all for its implementation.
It is easy to show

$$g(n) = H^T(n)\, \tilde{X}(n) - \mu\, u(n) \tag{11}$$

where

$$u(n) = \beta\left[u(n-1) + X^T(n-1)\, \tilde{X}(n-1) \operatorname{sign}\{e(n-1)\}\right]. \tag{12}$$

From Eq. 11 we can see that for small values of $\mu$


30th Midwest Symposium on Circuits and Systems
G. Glasford and K. Jabbour, eds.

and small reconstruction errors, our multiplier-free implementation is a good approximation to the direct implementation of the filter.

## II. ANALYSIS OF THE MADF

To make the analysis mathematically tractable, we need to make several simplifying assumptions. They are summarized below.

1. The primary input sequence $X(n)$, the reconstructed sequence $\tilde{X}(n)$ and the desired output sequence $d(n)$ are all jointly wide sense stationary, zero mean, Gaussian random processes.

2. The reconstruction error vector $\eta(n) = X(n) - \tilde{X}(n)$ is a zero mean white process with covariance matrix $\sigma_\eta^2 I$. Furthermore $\eta(n)$ and $X(n)$ are mutually uncorrelated. We will also assume that the elements of $\eta(n)$ are uniformly distributed in $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right)$ so that

$$\sigma_\eta^2 = \frac{\Delta^2}{12}. \tag{13}$$

3. In inputs $\{X(n), \tilde{X}(n), d(n)\}$ are uncorrelated for different values of time.

4. The error sequence $e(n)$ is zero mean and Gaussian with variance $\sigma^2_{e|H}(n)$ conditioned on the coefficient vector $H(n)$. Because of the presence of $\mu\, u(n)$ in the error sequence, its conditional probability distribution function is non-Gaussian. However, for small values of $\mu$, the deviations from Gaussian distribution is very small. Moreover, the variations in the statistics of $H(n)$ are small enough so that the conditional expectation $E\{e^2(n)|H(n)\}$ can be approximated as the unconditional expectation $E\{e^2(n)\}$.

Taking the statistical expectation of both sides of Eq. 9, we get

$$E\{H(n+1)\} = E\{H(n)\} + \mu\, E\{X(n) \text{ sign } \{e(n)\}\}. \tag{14}$$

Since both $X(n)$ and $e(n)$ are assumed to be zero mean and Gaussian, we can write the second expectation on the right-hand side of Eq. 14 as [2]

$$E\{X(n) \text{ sign } e(n)\} = E\{E\{X(n) \text{ sign } \{e(n)\}\ |H(n)\}\}$$

$$= \frac{E\{X(n)\, e(n)\}\, E\{|e(n)|\,|H(n)\}}{E\{e^2(n)|H(n)\}} \tag{15a}$$

$$= \sqrt{\frac{2}{\pi}}\, \frac{1}{\sigma_{e|H}(n)}\, E\{X(n)\, e(n)\}, \tag{15b}$$

$$\approx \sqrt{\frac{2}{\pi}}\, \frac{1}{\sigma_e(n)}\, E\,\{X(n)\, e(n)\} \tag{15c}$$

Now,

$$E\{X(n)\, e(n)\} = E\{X(n)\, [d(n) - \tilde{X}^T(n)\, H(n) + \mu\, u(n)]\} \tag{16a}$$

$$= P - R\, E\{H(n)\} \tag{16b}$$

where $R$ is the autocorrelation matrix of $X(n)$ and $P$ is the cross-correlation vector of $X(n)$ and $d(n)$.

In deriving Eq. 16, we have made use of the uncorrelated input signals assumption, which is turn implies that $X(n)$ and $H(n)$ are uncorrelated and also that $X(n)$ and $u(n)$ are uncorrelated. In addition, we have also made use of the fact that the reconstruction error vector $\eta(n)$ is uncorrelated with $X(n)$. Using Eq. 14-16 and substituting $V(n) + H_{opt}$ for $H(n)$, we get

$$E\{V(n+1)\} = \left[I - \mu \sqrt{\frac{2}{\pi}}\, \frac{1}{\sigma_e(n)}\, R\right] E\{V(n)\} \tag{17}$$

where $I$ denotes the $N \times N$ identity matrix. It is easy to show that the above system converges if

$$0 < \mu < \frac{\sqrt{2\pi}\ \sigma_e(n)}{\lambda_{max}} \tag{18}$$

for all $n$. A more conservative but simpler condition can be obtained by replacing $\sigma_e(n)$ by $\sqrt{\xi_{min}}$ in Eq. 18. This will give

$$0 < \mu < \frac{\sqrt{2\pi\ \xi_{min}}}{\lambda_{max}} \tag{19}$$

as a sufficient condition for convergence of the mean behavior of the adaptive filter coefficients. In Eq. 18 and 19, $\lambda_{max}$ denotes the maximum eigenvalue of the autocorrelation matrix $R$, and if convergence does occur,

$$\lim_{n\to\infty} E\{H(n)\} = R^{-1}\, P = H_{opt}. \tag{20}$$

To complete the analysis of the MADF we need to develop an expression for $\sigma_e(n)$ which in turn will require us to develop an expression for $K(n)$. This is done next.

It is straightforward to show that

$$\sigma_e^2(n) = E\{e^2(n)\} = \xi_{min} + tr\{R\, K(n)\} + \mu^2\, E\{u^2(n)\} + \sigma_\eta^2\, \{\|H_{opt}\|^2 + 2\, H_{opt}\, E\{V(n)\} + tr\,\{K(n)\}\}. \tag{21}$$

where

$$\xi_{min} = E\{d^2(n)\} - H_{opt}^T\, P \tag{22}$$

is the minimum mean-squared estimation error. We can see from Eq. 21 that the steady state, mean squared error will be larger than the usual $\xi_{min} + tr\{R\, K(\infty)\}$ by $\mu^2\, E_{ss}\, \{u^2(n)\} + \sigma_\eta^2\, \{\|H_{opt}\|^2 + tr\{K(\infty)\}\}$. By properly designing the MADF, in many applications we can assure that this extra error is small compared with $\xi_{min}$ ($E_{ss}\, \{u^2(n)\}$ is the steady-state value of $E\{u^2(n)\}$). Now, from Eq. 12 and using the uncorrelatedness assumption,

$$E\{u^2(n)\} = \beta^2\, [E\{u^2(n-1)\} + E\{X^T(n-1)\, \tilde{X}(n-1)\, X^T n-1)\, \tilde{X}\, (n-1)\}]. \tag{23}$$

Even though assuming that $u(n)$ and $X^T(n)\tilde{X}(n)$ are uncorrelated is gross, the effect of this assumption on the overall mean squared error calculation is very small since the dependence of $E\{u^2(n)\}$ on $\sigma_e^2(n)$ is only proportional to $\mu^2$. The fourth order expectation in Eq. 23 can be simplified using the Gaussian assumption and one can easily show that

$$E\{X^T(n)\ \tilde{X}(n)\ X^T(n)\ \tilde{X}(n)\} = tr^2\{R\} + tr\{R(R + \sigma_\eta^2\ I)\} + \|R\|_F^2. \qquad (24)$$

In Eq. 24 $\|R\|_F^2$ is the Frobenuis norm of the matrix R with elements $r_{i,j}$ given by

$$\|R\|_F^2 = \sum_{i,j} r_{i,j}^2\ . \qquad (25)$$

In the limit, $E\{u^2(n)\}$ will converge to

$$E_{ss}\{u^2(n)\} = \lim_{n\to\infty} E\{u^2(n)\} = \frac{\beta^2}{1-\beta^2}\left[tr^2(R) + \|R\|_F^2 + tr\ R(R + \sigma_\eta^2 I)\right]. \qquad (26)$$

A recursive expression for $K(n)$ is given next. Details of the derivation can be found in [6].

$$K(n+1) = K(n) + \mu^2 R - \mu\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_e(n)}\left[R\ K(n) + K(n)\ R\right] \qquad (27)$$

By looking at the individual elements of the matrix Eq. 27 it is very easy to show that the system converges if $\mu$ satisfies at every time (see [5] for similar derivations)

$$0 < \mu < \frac{\sqrt{\frac{\pi}{2}}\ \sigma_e(n)}{\lambda_{max}}\ . \qquad (28)$$

However, $\sigma_e(n)$. depends on $K(n)$. A sufficient and simpler condition for mean-squared convergence of the MADF is

$$0 < \mu < \frac{\sqrt{\frac{\pi}{2}}\ \xi_{min}}{\lambda_{max}}\ . \qquad (29)$$

Let $K(\infty)$ and $\sigma_e^2(\infty)$ denote the steady-state values of $K(n)$ and $\sigma_e^2(n)$. (Note that if $K(n)$ converges, so does $\sigma_e^2(n)$). Some manipulations of (21) and (27) will give the steady-state values

$$K(\infty) = \mu\sqrt{\frac{\pi}{2}}\frac{\sigma_e(\infty)}{2}\ I \qquad (30)$$

and

$$\sigma_e^2(\infty) = \alpha\ \sigma_e(\infty) + \gamma \qquad (31)$$

where

$$\alpha = \frac{\mu}{2}\sqrt{\frac{\pi}{2}}\left[\left(\sum_{i=1}^{N}\lambda_i\right) + N\ \sigma_\eta^2\right] \qquad (32)$$

and

$$\gamma = \xi_{min} + \mu^2\ E_{ss}\{u^2(n)\} + \sigma_\eta^2\ \{\|H_{opt}\|^2\}. \qquad (33)$$

Solving for $\sigma_e(\infty)$ in Eq. 31 and retaining the positive root, we get

$$\sigma_e(\infty) = \frac{\alpha + \sqrt{\alpha^2 + 4\gamma}}{2}\ . \qquad 34)$$

By substituting for $\alpha$ and $\gamma$ in Eq. 34, squaring and neglecting all terms involving $\mu^2$, we get an approximate expression for the steady-state mean-squared estimation error as

$$\sigma_e^2(\infty) \approx \xi_{min} + \sigma_\eta^2\ \|H_{opt}\|^2 + \frac{\mu}{2}\sqrt{\frac{\pi}{2}}\left(N\ \sigma_\eta^2 + \sum_{i=1}^{N}\lambda_i\right)\sqrt{\xi_{min} + \sigma_\eta^2\ \|H_{opt}\|^2}. \qquad (35)$$

Remarks

1. Comparing this result with that for the sign algorithm [5], we can see that the reconstruction error in DPCM produces additional excess mean-squared estimation error in the MADF. However, in many applications of interest, the minimum mean-squared error $\xi_{min}$ will be fairly large compared with the reconstruction error power $\sigma_\eta^2$ and we can design the MADF such that the overall performance of the system is still very good.

2. The choices of quantization step size $\Delta$ and the predictor coefficient $\beta$ are very important in the performance of the MADF. In our work we optimized these two parameters individually. Details of this aspect may be found in [1].

3. For the MADF to work well, the input signal must be sufficiently correlated for small lags. Otherwise, the performances of the linear predictor and the quantizer will suffer. However, if we use a relatively large number of quantization levels, fairly small amounts of oversampling will produce adequate results. Our experiments have indicated that if we use 13 quantization levels and sample the input sequences at twice the Nyquist rate, signal-to-reconstruction noise ratios around 20 dB can be obtained.

## III. A SIMULATION EXAMPLE

For our experiments, we chose a fourth-order autoregressive signal described by

$$x(n) = 1.79\ x(n-1) - 1.9425\ x(n-2) + 1.27\ x(n-3) - 0.5\ x(n-4) + 0.4\ \ \xi(n)\quad , \qquad (36)$$

where $\xi(n)$ is a zero mean, white, Gaussian signal with unit variance. The MADF was operated as a fourth-order linear predictor. Note that the ratio of the maximum and minimum eigenvalues of the autocorrelation matrix of the input signal is more than 80. The following parameters were selected for our system:

$$\beta = 0.75,\ \Delta = 0.5 \text{ and } \mu = 2^{-7}\ .$$

The results presented here are averaged over 200 independent runs using 8000 data samples each. Figure 2b displays plots of the mean behavior of the third predictor coefficient obtained from the simulations and from the theoretical results. Figure 2 displays the corresponding second moment behavior. We can see that in spite of the fairly large eigenvalue spread of the input correlation matrix, the theoretical results show fairly close match to the empirical results.

## IV. CONCLUSIONS

In this paper we presented and analyzed an efficient algorithm for adaptive FIR filtering. The performance of the MADF indicates that it is a viable and attractive alternative to traditional adaptive filtering techniques, especially when the input signals are sufficiently lowpass.

## REFERENCES


1. N. S. Jayant and P. Noll, Digital Coding of Waveforms, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.
2. T. Koh and E. J. Powers, "Efficient methods to estimate correlation functions of Gaussian processes and their performance analysis," IEEE Trans. Acoust., Speech, Signal Proc., Vol. ASSP-33, No. 4, pp. 1032-1035, August 1985.
3. J. C. Lee and C. K. Un, "On FIR delta modulation digital filters," IEEE Trans. Acoust., Speech, Signal Proc., Vol. ASSP-29, No. 6, pp. 1194-1201, December 1981.
4. J. W. Lee, C. K. Un and J. C. Lee, "Adaptive digital filtering of differentially coded signals," Proc. Int. Conf. Acoust., Speech, Signal Proc., pp. 1257-1260, Tampa, Florida, March 1985.
5. V. J. Mathews and S. H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," IEEE Trans. Acoust., Speech, Signal Proc., Vol. ASSP-35, No. 4, pp. 450-454, April 1987.
6. V. J. Mathews, "An efficient FIR adaptive filter using DPCM and the sign algorithm," Submitted to the IEEE Trans. Acoust., Speech, Signal Proc..
7. A. Peled and B. Liu, "A new approach to the realization of nonrecursive digital filters," IEEE Trans. Audio Electro Acoust., AU-21, No. 6, pp. 477-484, December 1973.
8. B. Widrow, et al., "Adaptive noise cancelling: Principles and applications," Proc. IEEE, Vol. 63, pp. 1692-1716, December 1975.


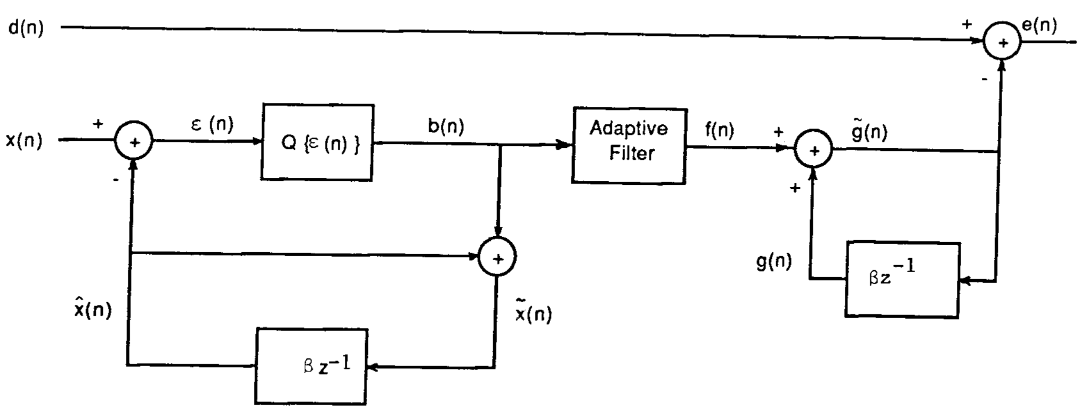


Fig. 1. Block diagram of the multiplication free adaptive filter.

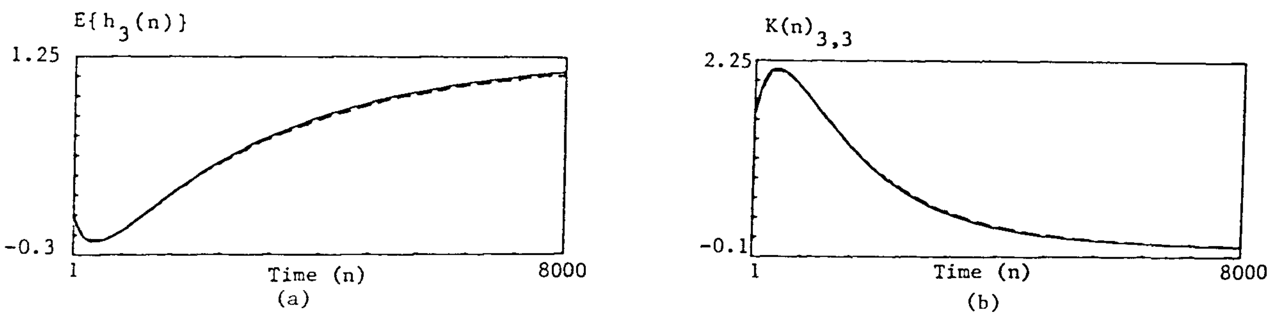


Fig. 2. Predicted (——) and simulated (- - - -) (a) mean and (b) mean squared behavior of the third coefficient ($h_3(n)$) in the example.