Analysis and Implementation of Variable Step Size Adaptive Algorithms

Joseph B. Evans, *Member, IEEE*, Ping Xue, *Member, IEEE*, and Bede Liu, *Fellow, IEEE*

***Abstract*—Stochastic gradient adaptive filtering algorithms using variable step sizes are investigated in this paper. The variable step size algorithm improves the convergence rate while sacrificing little in steady-state error. Expressions describing the convergence of the mean and mean squared values of the coefficients are developed and are used to calculate the mean square error evolution. The initial convergence rate and the steady-state error are also investigated. In addition, the performance of the algorithm is studied when a power-of-two quantizer algorithm is used, and finite word-length effects are considered. The analytical results are verified with simulations encompassing various applications. Two CMOS implementations of the variable step size, power-of-two quantizer algorithm are presented to demonstrate that the performance gains are attainable with only a modest increase in circuit complexity.**

I. Introduction

VARIOUS applications of the least mean squared (LMS) gradient search adaptive algorithm have been proposed [1]–[3], [6], [10], [12]–[14], [21]–[24], [26], [27]. The LMS algorithm, while relatively simple to implement, does not always converge in an acceptable manner, particularly when the input eigenvalue spread is large.

Variable step size methods [8], [11], [28] are an attempt to improve the convergence of the LMS algorithm while preserving the steady-state performance; the increase in the complexity of the implementation is relatively low [9]. In addition to these advantages, variable step size methods are more effective than the LMS algorithm at tracking in nonstationary environments.

The structure of a generic gradient search adaptive transversal filter is shown in Fig. 1. The output error at time k is given by

$$e_k = d_k - \boldsymbol{W}_k^T \boldsymbol{X}_k \tag{1}$$

Manuscript received October 23, 1989; revised July 22, 1992. The associate editor coordinating the review of this paper and approving it for publication was Dr. Erlendur Karlsson. This work was supported by the Office of Naval Research and the Naval Research Laboratory under Grant N000014-87-k-0189, University of Kansas General Research allocations 3775-20-0038 and 3626-20-0038, and by an AT&T Bell Laboratories Graduate Fellowship.

J. B. Evans is with the Department of Electrical and Computer Engineering, University of Kansas, Lawrence, KS 66045.

P. Xue is with TriTech Microelectronics International Pte. Ltd., Singapore Science Park, Singapore 0511.

B. Liu is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544.

IEEE Log Number 9209397.

and the filter output y_k is given by

$$y_k = \boldsymbol{W}_k^T \boldsymbol{X}_k \tag{2}$$

where $\boldsymbol{W}_k$ and $\boldsymbol{X}_k$ denote the column vectors of the tap weights and the input signal samples at the kth iteration, i.e.

$$\boldsymbol{W}_k = [w_{1,k}, w_{2,k}, \cdots, w_{N,k}]^T \tag{3}$$

and

$$\boldsymbol{X}_k = [x_{1,k}, x_{2,k}, \cdots, x_{N,k}]^T. \tag{4}$$

The simple LMS algorithm updates the taps of the filter according to

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k + 2\mu e_k \boldsymbol{X}_k \tag{5}$$

where μ is the step size parameter. The step size must be selected to balance the conflicting goals of fast convergence and small steady-state error.

A. Variable Step Size Algorithms

The variable step size methods considered here use a separate step size for each tap; the step sizes are adjusted individually as adaptation progresses.

The step size for the ith coefficient is decreased if a specified number, m_0, of consecutive sign changes of its update $e_k x_{i,k}$ occur, and it is increased if a specified number, m_1, of consecutive updates are of the same sign. The step size is not changed if these thresholds are not exceeded. It should be noted that only the sign bits of the updates are involved. The parameters m_0 and m_1 are called the sign change threshold and nonsign change threshold, respectively. The step sizes are constrained to be between $\mu_{\max}$ and $\mu_{\min}$. The upper bound, $\mu_{\max}$, can be somewhat larger than the maximum allowable value of μ for the LMS algorithm, as the variable step size algorithms will attempt to reduce the step size as adaptation progresses. The lower bound, $\mu_{\min}$, is chosen to balance the desired residual error and the tracking ability. The step size update is closely linked to the convergence of the adaptive filter.

Formally, the variable step size algorithms use a coefficient update of the form

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k + 2\boldsymbol{M}_k e_k \boldsymbol{X}_k \tag{6}$$

where $\boldsymbol{M}_k$ is a diagonal matrix of step sizes for individual taps at time k

$$\boldsymbol{M}_k = \operatorname{diag}(\mu_{1,k}, \mu_{2,k}, \cdots, \mu_{N,k}). \tag{7}$$

1053-587X/93$03.00

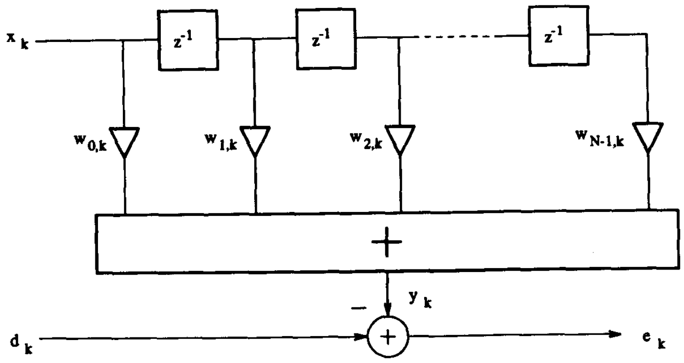


Fig. 1. Adaptive transversal filter.

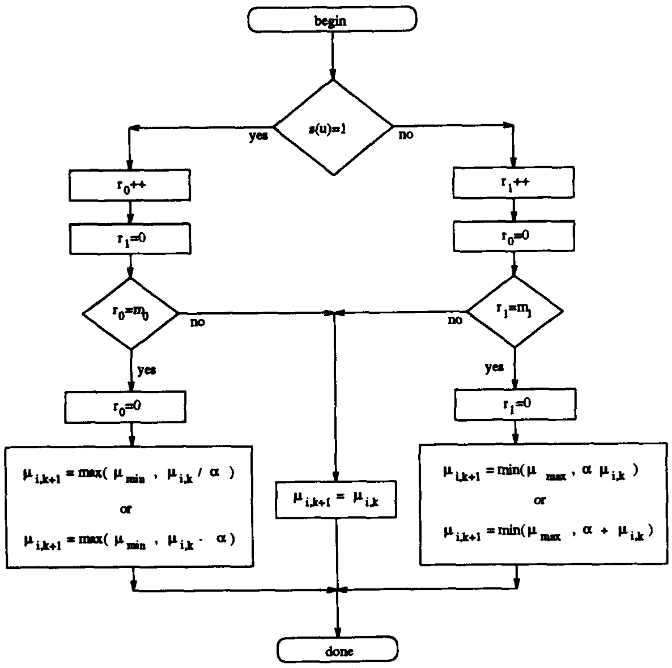


Fig. 2. Variable step size algorithms.

Two methods of updating the $\mu_{i,k}$ will be studied. The first method, discussed in [11], [28] and referred to as the VS method, increases the step size by multiplying the previous step size by a factor $\alpha > 1$ and decreases the step size by dividing the previous step size by α. That is,

$$\mu_{i,k} = \alpha\mu_{i,k-1}, \quad \text{or} \quad \mu_{i,k} = \mu_{i,k-1}/\alpha. \tag{8}$$

For each of implementation, α is usually taken to be 2, so that step size modification requires only a one-bit shift. The second method, referred to as the VSA method, adapts the step size by arithmetic steps [8]. The step size changes are accomplished via

$$\mu_{i,k} = \mu_{i,k-1} + \alpha', \quad \text{or}$$
$$\mu_{i,k} = \mu_{i,k-1} - \alpha', \quad \alpha' > 0. \tag{9}$$

Although this algorithm will be slightly more difficult to implement than the VS algorithm with $\alpha = 2$, it is comparatively easier to implement the VSA algorithm for different values of α', so the added flexibility may outweigh the slight increase in complexity.

Both the VS and the VSA algorithms are displayed in Fig. 2, where r_0 is the sign change counter, and r_1 is the nonsign change counter.

B. Algorithm Comparison and Implementation

We shall see that the VS algorithm has clear advantages in performance over the LMS algorithm; further, only shifts are required for the step size updates when $\alpha = 2$. The coefficient update operation will require no multiplications if the initial values of $\mu_{i,k}$, and $\mu_{\max}$ and $\mu_{\min}$ are chosen to be powers of two. On the other hand, $N \max(m_0, m_1)$ bits of additional memory are required to store the sign change information, and NB_μ extra bits of memory are required for the step size matrix, where B_μ is the number of bits used for each of the step sizes. The memory requirements are discussed more fully in [9].

The VSA algorithm exhibits a slight performance advantage over the VS algorithm. Unlike the VS algorithm, however, updates must be accomplished with extra adder hardware or by sacrificing speed. Additional multiplications in the coefficient update will be required, since the step size can no longer be restricted to powers of two. Additional memory is required for the VSA algorithm, as in the case of the VS algorithm.

C. Power-of-Two Quantizers

It is possible to simplify the implementation by employing a power-of-two quantizer [7], [29] to eliminate multiplication in the coefficient update. A B-bit power-of-two quantizer replaces its input u by a B-bit word having a single nontrivial bit corresponding to the most significant bit of the input; multiplication by $q(\cdot)$ becomes a simple shift operation. If we use the power-of-two-quantizer in the coefficient update in conjunction with variable step size methods, we get

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k + 2\boldsymbol{M}_k q(e_k)\boldsymbol{X}_k. \tag{10}$$

We will refer to this as the VS-PTQ algorithm.

II. Analysis of the Variable Step Size Algorithms

Various analyses of the classical LMS algorithm are based upon simplifying assumptions and approximations [18], [25], [26]. Since the variable step size algorithms contribute additional complexity, it is not unexpected that further approximations will be required in order to make the analysis tractable. Though dependent on a number of approximations, the analysis presented in this section allows the recursive prediction of the first and second moments of the coefficient misadjustment, step size evolution, and mean square error.

We will assume that the inputs $\{x_{i,k}, d_k\}$ are stationary and jointly Gaussian; both are zero mean, and $x_{i,k}$ has variance σ_x^2. The error e_k is Gaussian, with zero mean and variance σ_{ek}^2. We will also assume that the inputs $\{x_{i,k}, d_k\}$ are independent for different values of k; that is the well-known independence assumption [18]. Note that no restriction is placed on the autocorrelation matrix $\boldsymbol{R} = E[\boldsymbol{X}_k\boldsymbol{X}_k^T]$. We will further assume that the mean squared error conditioned on the tap weight vector, $E[e_k^2|\boldsymbol{W}_k]$, may be approximated by the unconditional mean squared error, that is,

$$E[e_k^2|\boldsymbol{W}_k] \approx E[e_k^2] = \sigma_{ek}^2. \tag{11}$$

This approximation has been used successfully in [16], [17] and is valid for small step size values (i.e., small coefficient updates). Finally, we will assume that the $\{\mu_{i,k}\}$ are independent of the current inputs $\{x_{i,k}, d_k\}$, tap weights $\{w_{i,k}\}$, and error e_k. The analysis of the VS and VSA algorithms are similar and thus are considered together.

A. Weight Vector Evolution and Mean Square Error

Define the coefficient misadjustment as

$$\boldsymbol{V}_k = \boldsymbol{W}_k - \boldsymbol{W}_{\text{opt}} \tag{12}$$

where $\boldsymbol{W}_{\text{opt}} = \boldsymbol{R}^{-1}\boldsymbol{P}$ is the optimum solution, $\boldsymbol{P} = E[d_k\boldsymbol{X}_k]$ is the cross correlation of d_k and $\boldsymbol{X}_k$, and $\boldsymbol{R} = E[\boldsymbol{X}_k\boldsymbol{X}_k^T]$ is the autocorrelation matrix of the inputs $\boldsymbol{X}_k$. It is shown in Appendix A that the mean of the misadjustment satisfies

$$E[\boldsymbol{V}_{k+1}] = (\boldsymbol{I} - 2E[\boldsymbol{M}_k]\boldsymbol{R})E[\boldsymbol{V}_k]. \tag{13}$$

This may be evaluated with knowledge of the step size mean and the initial misadjustment. It is also shown in Appendix A that the second moment of the misadjustment, $\boldsymbol{K}_k = E[\boldsymbol{V}_k\boldsymbol{V}_k^T]$, satisfies

$$\boldsymbol{K}_{k+1} = \boldsymbol{K}_k - 2\boldsymbol{K}_k\boldsymbol{R}E[\boldsymbol{M}_k] - 2E[\boldsymbol{M}_k]\boldsymbol{R}\boldsymbol{K}_k + 4\sigma_{ek}^2 E[\boldsymbol{M}_k\boldsymbol{R}\boldsymbol{M}_k] + 8E[\boldsymbol{M}_k\boldsymbol{R}\boldsymbol{K}_k\boldsymbol{R}\boldsymbol{M}_k]. \tag{14}$$

This equation may be evaluated with knowledge of the first and second moments of the step size, since

$$[E[\boldsymbol{M}_k\boldsymbol{R}\boldsymbol{M}_k]]_{ij} = \begin{cases} \sigma_x^2 E[\mu_{i,k}^2] & i = j \\ E[\mu_{i,k}]r_{i-j}E[\mu_{j,k}] & i \neq j \end{cases} \tag{15}$$

$$[E[\boldsymbol{M}_k\boldsymbol{R}\boldsymbol{K}_k\boldsymbol{R}\boldsymbol{M}_k]]_{ij} = \begin{cases} \sum_l \sum_m E[\mu_{i,k}^2]r_{i-m}K_{ml,k}r_{l-i} & i = j \\ \sum_l \sum_m E[\mu_{i,k}]r_{i-m}K_{ml,k}r_{l-j}E[\mu_{j,k}] & i \neq j \end{cases} \tag{16}$$

where $r_{i-j} = E[x_{i,k}, x_{j,k}]$, and $K_{ij,k}$ is the ijth element of $\boldsymbol{K}_k$.

Finally, it is simple to show that the mean square error at the kth iteration is given by

$$\sigma_{ek}^2 = E[e_k^2] = \xi_{\min} + \operatorname{tr}(\boldsymbol{K}_k\boldsymbol{R}) \tag{17}$$

where $\xi_{\min} = E[d_k^2] - \boldsymbol{W}_{\text{opt}}^T\boldsymbol{P}$ is the Weiner error.

B. Step Size Evolution

1) Step Size Transition Model: The step size evolution will be modeled using the state transition model shown in Fig. 3, where each state consists of (r_0, r_1, ϕ_j, s), that is,

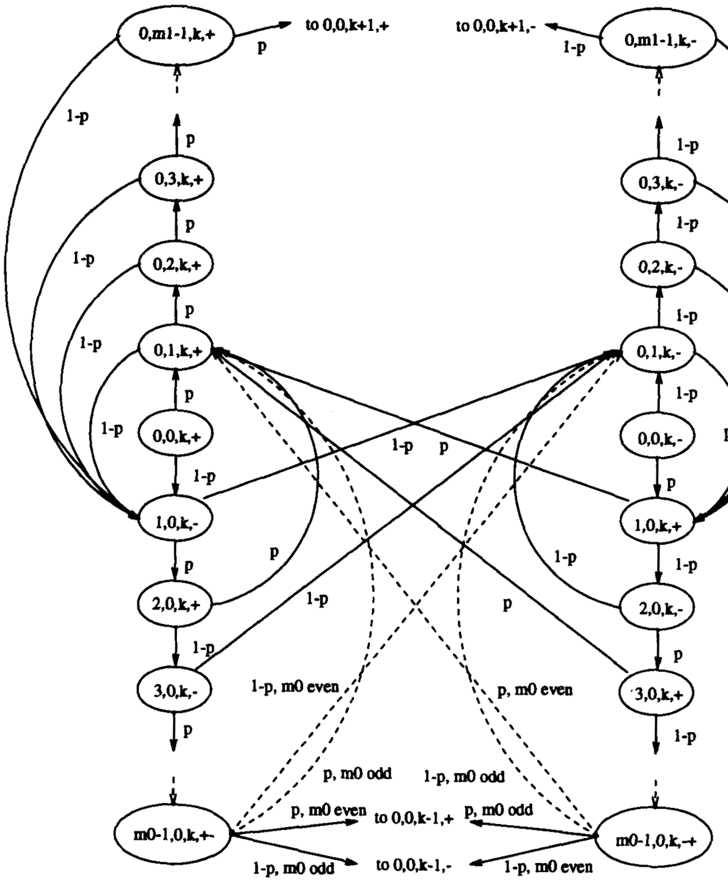


Fig. 3. State transition diagram for each step size—(r_0, r_1, ϕ_j, s).

the two sign change counter states, the step size state, and the last coefficient update sign.

Under the given assumptions on the error e_k and the input $x_{i,k}$, it is shown in Appendix A that the probability that coefficient $w_{i,k}$ will move in the direction of its steady-state value is

$$p_{i,k} = \tfrac{1}{2} + b_{i,k} \tag{18}$$

where $b_{i,k}$ is the bias due to coefficient misadjustment given by

$$b_{i,k} \approx -\frac{\rho_{i,k}}{\pi\sqrt{1-\rho_{i,k}^2}}\left\{1-\gamma-\ln\left(\frac{\rho_{i,k}}{2(1-\rho_{i,k}^2)}\right)\right\} \tag{19}$$

$\rho_{i,k} = |E[e_k x_{i,k}]|/(\sigma_{e_k}\sigma_x)$, and $\gamma = 0.5772157 \cdots$ is Euler's constant. Note that

$$E[e_k x_{i,k}] = -\sum_{j=0}^{N-1} r_{i-j} E[v_{j,k}]$$

where r_{i-j} is the ijth element of the autocorrelation $\boldsymbol{R}$, and $v_{j,k}$ is the jth element of the misadjustment $\boldsymbol{V}_k$.

In the case of the VS algorithm, the values which the step size may assume are $\{\phi_h,\ h = 0, \cdots, M-1\}$, where M is the largest integer such that $\mu_{\max}/\alpha^M > \mu_{\min}$, and

$$\phi_h = \mu_{\max}/\alpha^h \qquad 0 \le h < M. \tag{20}$$

It is assumed that all step sizes initially take on one of the ϕ_h's. In the case of the VSA algorithm, M is now the largest integer such that $\mu_{\max} - M\alpha > \mu_{\min}$, and the ϕ_h are given by

$$\phi_h = \mu_{\max} - \alpha h \qquad 0 \le h < M. \tag{21}$$

If we define the probability that the step size for the ith tap is in state (f, g, h, s) at the kth time instant by $P_{i,k,f,g,h,s}$, knowledge of the allowable state transitions as given by Fig. 3 can be used to determine the nonzero elements of the state transition matrix. In particular, the transitions to the states at middle of the branches are described by

$$P_{i,k,f,g,h,s} = \begin{cases} p_{i,k} P_{i,k,f-1,0,h,-1} & f < m_0 - 1,\ \text{even};\ g = 0 \\ & s = +1;\ h = 0, 1, \cdots, M-1 \\ p_{i,k} P_{i,k,f-1,0,h,-1} & f < m_0 - 1,\ \text{odd};\ g = 0 \\ & s = +1;\ h = 0, 1, \cdots, M-1 \\ (1 - p_{i,k}) P_{i,k,f-1,0,h,+1} & f < m_0 - 1,\ \text{even};\ g = 0 \\ & s = -1;\ h = 0, 1, \cdots, M-1 \\ (1 - p_{i,k}) P_{i,k,f-1,0,h,+1} & f < m_0 - 1,\ \text{odd};\ g = 0 \\ & s = -1;\ h = 0, 1, \cdots, M-1 \\ p_{i,k} P_{i,k,0,g-1,h,+1} & f = 0;\ g = 2, 3, \cdots, m_1 - 1 \\ & s = +1;\ h = 0, 1, \cdots, M-1 \\ (1 - p_{i,k}) P_{i,k,0,g-1,h,-1} & f = 0;\ g = 2, 3, \cdots, m_1 - 1 \\ & s = -1;\ h = 0, 1, \cdots, M-1 \end{cases} \tag{22}$$

while the transitions to the states close to the center states are

$$P_{i,k,f,g,h,s} = \begin{cases} \sum_{r=0}^{m_1-1} (1 - p_{i,k}) P_{i,k,0,r,h,+1} & f = 1;\ g = 0 \\ & s = +1;\ h = 0, 1, \cdots, M-1 \\ \sum_{r=0}^{m_1-1} p_{i,k} P_{i,k,0,r,h,-1} & f = 1;\ g = 0 \\ & s = -1;\ h = 0, 1, \cdots, M-1 \\ \sum_{r=0}^{m_0-1} p_{i,k} P_{i,k,r,0,h,+1} & f = 0;\ g = 1 \\ & s = +1;\ h = 0, 1, \cdots, M-1 \\ \sum_{r=0}^{m_0-1} (1 - p_{i,k}) P_{i,k,r,0,h,-1} & f = 0;\ g = 1 \\ & s = -1;\ h = 0, 1, \cdots, M-1 \end{cases} \tag{23}$$

and the transitions to the center states are given by

$$P_{i,k,f,g,h,s} = \begin{cases} p_{i,k} P_{i,k,0,m_1-1,h-1,+1} + p_{i,k} P_{i,k,m_0-1,0,h+1,-1} & f = 0;\ g = 0 \\ & s = +1;\ h = 1, 2, \cdots, M-2 \\ (1 - p_{i,k}) P_{i,k,m_1-1,h-1,-1} & f = 0;\ g = 0 \\ \quad + (1 - p_{i,k}) P_{i,k,m_0-1,0,h+1,+1} & s = -1;\ h = 1, 2, \cdots, M-2. \end{cases} \tag{24}$$

Finally, the transitions to the states at the upper and lower step size boundaries are described by

$$P_{i,k,f,g,h,s} = \begin{cases} (1 - p_{i,k}) P_{i,k,m_0-1,0,0,+1} & f = 0;\ g = 0 \\ \quad + (1 - p_{i,k}) P_{i,k,m_0-1,0,1,+1} & s = -1;\ h = 0 \\ p_{i,k} P_{i,k,m_0-1,0,0,-1} + p_{i,k} P_{i,k,m_0-1,0,1,-1} & f = 0;\ g = 0 \\ & s = +1;\ h = 0 \\ p_{i,k} P_{i,k,0,m_1-1,M-1,+1} & f = 0;\ g = 0 \\ \quad + p_{i,k} P_{i,k,0,m_1-1,M-2,+1} & s = +1;\ h - M - 1 \\ (1 - p_{i,k}) P_{i,k,0,m_1-1,M-1,-1} & f = 0;\ g = 0 \\ \quad + (1 - p_{i,k}) P_{i,k,0,m_1-1,M-2,-1} & s = -1;\ h = M - 1. \end{cases} \tag{25}$$

The state probabilities can be computed for any given initial conditions using this model.

2) Step Size Moments: Using the results obtained above, we may express the first moment of the step size as

$$E[\mu_{i,k}] = \sum_{f=0}^{m_0-1} \sum_{g=0}^{m_1-1} \sum_{h=0}^{M-1} \phi_h (P_{i,k,f,g,h,+1} + P_{i,k,f,g,h,-1}) \tag{26}$$

and the second moment as

$$E[\mu_{i,k}^2] = \sum_{f=0}^{m_0-1} \sum_{g=0}^{m_1-1} \sum_{h=0}^{M-1} \phi_h^2 (P_{i,k,f,g,h,+1} + P_{i,k,f,g,h,-1}). \tag{27}$$

C. Learning Curve Evolution

The mean square error may be recursively computed from any known initial conditions $P_{i,0,f,g,h,s}$ by using (13), (14), (17), (19), (22)–(27). The misadjustment moments $E[V_k]$ and $E[K_k]$ are updated using (13) and (14); the mean square error estimate σ_{ek}^2 can then be obtained from (17). The sign change bias probability for each tap is found by using the mean square error estimate and the misadjustment mean via (19). This can be used to calculate the state probability $P_{i,k,f,g,h,s}$ update, via (22)–(25). The step size mean and second moment can then be found by using (26) and (27). These are used in the next iteration for updating the misadjustment moments. A similar technique can be applied in order to use the results of the following section.

III. Quantization Effects

The analysis of the VS-PTQ algorithm requires inclusion of the effects of a power-of-two quantizer on the coefficient update. Further, implementation of these algorithms requires finite word size; thus these effects should also be considered.

A. Power-of-Two Quantization Effects

Although the effect of power-of-two quantization in the coefficient update on the performance of the LMS adaptive algorithm has been investigated previously, we do not assume that the quantizer has an infinite number of bits, as in [7], or that a hard limiter is used on the data, as in [29]. It is shown in Appendix B that the mean of the coefficient misadjustment for the VS-PTQ algorithm with a B-bit power-of-two quantizer satisfies

$$E[V_{k+1}] = (I - 2A(B, \sigma_{ek})E[M_k]R)E[V_k] \tag{28}$$

where

$$A(B, \sigma_{ek}) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_{ek}}\left\{\sum_{i=1}^{B-1} 2^{-i} \exp\left(\frac{-1}{2^{2i-1}\sigma_{ek}^2}\right) + 2^{-B+1}\exp\left(\frac{-1}{2^{2B-1}\sigma_{ek}^2}\right)\right\}. \tag{29}$$

The second moments of the coefficient misadjustment for the VS-PTQ algorithm are shown in Appendix B to satisfy

$$\begin{aligned} K_{k+1} = K_k &- 2A(B, \sigma_{ek})(K_k R E[M_k] + E[M_k] R K_k) \\ &+ 4F(B, \sigma_{ek}) E[M_k R M_k] \\ &+ 4D(B, \sigma_{ek}) E[M_k R K_k R M_k] \end{aligned} \tag{30}$$

where

$$D(B, \sigma_{ek}) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_{ek}^2}\left\{\sum_{i=1}^{B-1} 2^{-3i+1} \exp\left(\frac{-1}{2^{2i-1}\sigma_{ek}^2}\right) + 2^{-3(B+1)}\exp\left(\frac{-1}{2^{2B-1}\sigma_{ek}^2}\right)\right\} \tag{31}$$

and

$$F(B, \sigma_{ek}) = \frac{3}{2}\sum_{i=1}^{B-1} 2^{-2(i-1)} \operatorname{erfc}\left(\frac{1}{2^{i-1}\sigma_{ek}}\right) + 2^{-2B+1}\operatorname{erfc}\left(\frac{1}{2^{B-1}\sigma_{ek}}\right). \tag{32}$$

B. Finite Word Length Quantization Effects

The effect of finite word length on the performance of the LMS adaptive algorithm was discussed in [5]; that analysis was modified for the power-of-two quantizer algorithm in [29].

We will assume that each data word uses B_b bits plus sign, of which B_{dx} are before the binary point. If we assume that no overflow occurs, additions will produce no error, while multiplications will introduce error when the product is quantized. For simplicity, we will assume rounding is used, so that the variance of the data quantization error is

$$\sigma_d^2 = \tfrac{1}{12} 2^{-2(B_d - B_{dx})}. \tag{33}$$

As in [5], we assume that the roundoff error is uncorrelated with the error due to adaptation and gradient noise. The mean square error is then

$$\sigma_{ek}^{2\prime} = \sigma_{ek}^2 + \bar{\sigma}_{ek}^2 \tag{34}$$

where σ_{ek}^2 is the infinite precision part of the mean square error, and $\bar{\sigma}_{ek}^2$ is due to finite precision effects.

It is shown in Appendix B that the extra mean square error due to finite-bit arithmetic effects reduces to

$$\bar{\sigma}_{ek}^2 = (\operatorname{tr}(K_k) + 2E[V_k^T]W_{\text{opt}} + W_{\text{opt}}^T W_{\text{opt}} + \rho)\sigma_d^2 \tag{35}$$

where $\rho = N$ if the N scalar products of $W_k^T X_k$ are quantized individually and then summed, and $\rho = 1$ if the scalar products are summed and the result is then quantized.

IV. VS Algorithms with White Inputs

We now consider the special case of white inputs, that is, the $\{x_{i,k}\}$ are independent, identically distributed Gaussian random variables with zero mean and variance σ_x^2. This will allow us to develop relatively simple expressions for the step size convergence rate and the steady-state step size; these, in turn, allow simple expressions for the mean square error convergence and steady-state misadjustment.

A. Misadjustment Moments and Mean Square Error

With white inputs, (13) and (14) reduce to

$$E[V_{k+1}] = (I - 2\sigma_x^2 E[M_k])E[V_k] \tag{36}$$

$$K_{k+1} = K_k - 2\sigma_x^2 M_k K_k - 2\sigma_x^2 K_k M_k + 4\sigma_{ek}^2\sigma_x^2 E[M_k^2] + 8\sigma_x^4 E[M_k K_k M_k]. \tag{37}$$

Let $K_{i,k}$ denote the ith diagonal element of K_k. Then

$$K_{i,k+1} = (1 - 2\sigma_x^2 E[\mu_{i,k}] + 2\sigma_x^4 E[\mu_{i,k}^2])K_{i,k} + \sigma_{ek}^2\sigma_x^2 E[\mu_{i,k}^2]. \tag{38}$$

Defining $\theta_k = \mathrm{tr}\,(K_k)$, the mean square error of (17) becomes

$$\sigma_{ek}^2 = \xi_{\min} + \sigma_x^2\theta_k. \tag{39}$$

B. Stability Conditions

It is shown in Appendix C that the mean of the misadjustment is guaranteed to go to zero if

$$0 < \mu_{\max} < \frac{1}{\sigma_x^2}. \tag{40}$$

Note that this bound, which holds for white noise only, is looser than the bound for the general case as given in [11].

We can observe from (39) that to find conditions for the convergence of the mean square error we need to study the behavior of θ_k. It is shown in Appendix C that convergence of the mean square error is guaranteed if

$$\frac{\mu_{\max}^2}{\mu_{\min}} < \frac{2}{(N+2)\sigma_x^2} \tag{41}$$

which implies that the minimum step size be such that

$$\mu_{\min} < \frac{1}{(N+2)\sigma_x^2}. \tag{42}$$

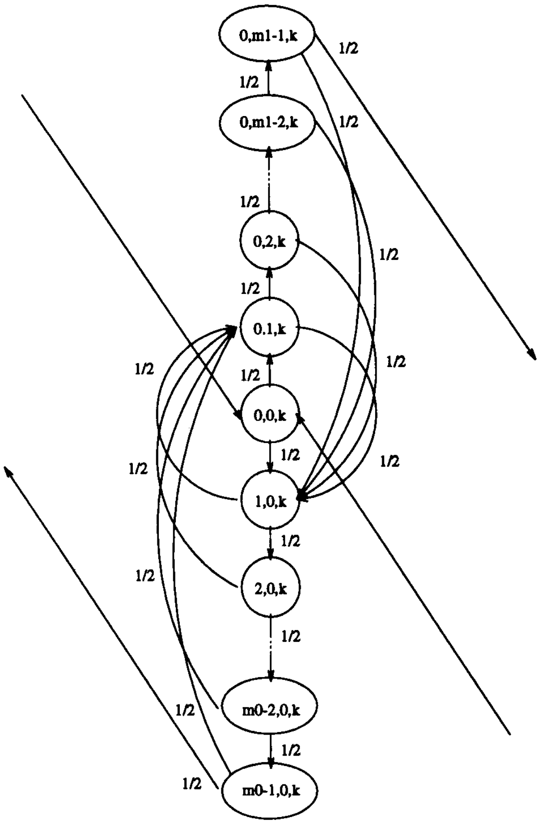


Fig. 4. Simplified state transition diagram—(r_0, r_1, ϕ_j).

C. Steady-State Error

Near steady-state, the mean of the misadjustment will approach zero; hence, from (18) and (19), we can see that $p_{i,k} \to 1/2$ as $k \to \infty$. Due to the symmetry of the state transition model, the simplified model shown in Fig. 4 can be used, where each state is now specified by the triplet (f, g, h), the counter states, and the step size state.

When the system has converged, the probability that the step size for any tap is in state (f, g, h) will reach some steady-state value $P_{\infty,f,g,h}$. Using (22)–(25) and the constraint

$$\sum_{f=0}^{m_0-1}\sum_{g=0}^{m_1-1}\sum_{h=0}^{M-1} P_{i,\infty,f,g,h} = 1 \tag{43}$$

it is shown in Appendix C that the steady-state step size for the VS algorithm is

$$E[\mu_\infty] = \begin{cases} \mu_{\min}\beta\left\{i^c\dfrac{1-(\alpha/i)^{c+1}}{1-\alpha/i} + \alpha^{c+1}d\dfrac{1-(\alpha d)(M-c-1)}{1-\alpha d}\right\} & i \neq \alpha,\quad \alpha d \neq 1 \\ \mu_{\min}\beta\left\{(c+1)\alpha^c + \alpha^{c+1}d\dfrac{1-(\alpha d)(M-c-1)}{1-\alpha d}\right\} & i = \alpha,\quad \alpha d \neq 1 \\ \mu_{\min}\beta\left\{i^c\dfrac{1-(\alpha/i)^{c+1}}{1-\alpha/i} + (M-c-1)\alpha^c\right\} & i \neq \alpha,\quad \alpha d = 1 \\ \mu_{\min}\beta\alpha^c M & i = \alpha,\quad \alpha d = 1 \end{cases} \tag{44}$$

where

$$\frac{1}{\beta} = 1 + \sum_{l=1}^{\lfloor M+1/2 \rfloor - 1} i^l + \sum_{l=1}^{\lfloor M/2 \rfloor} d^l \tag{45}$$

$$c = \left\lfloor \frac{M+1}{2} \right\rfloor - 1 \tag{46}$$

and

$$i = \frac{2^{-m_0+1}(1 - 2^{-m_1+1})}{1 - (1 - 2^{-m_0+2})(1 - 2^{-m_1+2})}$$

$$d = \frac{2^{-m_1+1}(1 - 2^{-m_0+1})}{1 - (1 - 2^{-m_0+2})(1 - 2^{-m_1+2})}. \tag{47}$$

Similarly, the steady-state step size for the VSA algorithm is

$$E[\mu_\infty] = \beta \left\{ i^c \left(\mu_{\min} \frac{1 - (1/i)^{c+1}}{1 - (1/i)} + \frac{\alpha}{i} \frac{1 - (c+1)(1/i)^c + c(1/i)^{c+1}}{(1 - (1/i))^2} \right) + d \left((\mu_{\min} + \alpha(c+1)) \frac{1 - d^{M-c-1}}{1 - d} + \alpha d \frac{1 - (M - c - 1)d^{M-c-2} + (M - c - 2)d^{M-c-1}}{(1 - d)^2} \right) \right\}. \tag{48}$$

These can be used in the expression for steady-state mean square error,

$$\sigma^2_{e_\infty} = \left(1 + \frac{N\sigma_x^2 E[\mu_\infty]}{2 - (N+2)\sigma_x^2 E[\mu_\infty]} \right) \xi_{\min}. \tag{49}$$

D. Convergence Rate

Simulation results (cf. Section V) suggest that the step size evolution of the individual taps follow roughly the same trend, so they may be approximated by a single step size. With this justification, we will develop expressions for the convergence rate of the mean square error by approximating the expressions for the step size increase and decrease probabilities. By making the appropriate assumptions about the bias probability, we can derive expressions for both the initial and the later stage convergence rates.

1) Step Size Change Probabilities, Initial Convergence: Given the initial tap weights, the expected value of the initial update term is

$$E[e_0 x_{i,0}] = \sigma_x^2 v_i \tag{50}$$

where v_i is the ith element of the initial coefficient misadjustment. Assuming the step sizes $\mu_{i,k}$ are small, this can be taken as an approximation for the expectation of the update during the early adaptation period, so that the corresponding sign change probability is

$$p = \tfrac{1}{2} + b \tag{51}$$

where

$$b = -\frac{\rho}{\pi\sqrt{1 - \rho^2}} \left\{ 1 - \gamma - \ln \left(\frac{\rho}{2(1 - \rho^2)} \right) \right\} \tag{52}$$

$$\rho = \frac{\sigma_x \sum_{j=0}^{N-1} v_j}{\xi_{\min} + \sigma_x^2 \sum_{j=0}^{N-1} v_j^2}. \tag{53}$$

The cross-correlation coefficient approximation ρ was found by averaging over all taps. The sign change probability can be used as the driving term for the step size update. Assuming that it is equally probable that the step size is at any given value, the probability of a step size increase may be approximated by $P^+ = p^{m_1+1} + (1 - p)^{m_1+1}$, and the probability of a step size decrease may be approximated by $P^- = p^{\lfloor m_0/2 \rfloor + 1}(1 - p)^{\lceil m_0/2 \rceil} + p^{\lceil m_0/2 \rceil}(1 - p)^{\lfloor m_0/2 \rfloor + 1}$.

2) Step Size Change Probabilities, Near Steady State: Given that the bias probability approaches zero at steady state, and assuming that it is equally probable that the step size is at any given value, the probability of a step size increase may be approximated by $P^+ = 2^{-m_1}$, and the probability of a step size decrease may be approximated by $P^- = 2^{-m_0}$.

3) Step Size Evolution: Using these approximations for the step size change probabilities, and ignoring the boundary effects, the VS algorithm step size evolution may be approximated by

$$E[\mu_{k+1}] = \left\{ (1 - P^+ - P^-) + \alpha P^+ + \frac{1}{\alpha} P^- \right\} E[\mu_k] \tag{54}$$

$$E[\mu^2_{k+1}] = \left\{ (1 - P^- - P^-) + \alpha^2 P^+ + \frac{1}{\alpha^2} P^- \right\} \cdot E[\mu_k^2]. \tag{55}$$

Similarly, for the VSA algorithm,

$$E[\mu_{k+1}] = E[\mu_k] + \alpha(P^- - P^-) \tag{56}$$

$$E[\mu^2_{k+1}] = E[\mu_k^2] + \alpha(P^+ - P^-)E[\mu_k] + \alpha^2(P^+ - P^-). \tag{57}$$

The expressions for the mean of the step size may be used to provide a rough estimate of the convergence rate of the step sizes for given algorithm parameters. The estimates generated for the VSA algorithm tend to be more accurate than those for the VS algorithm, as the boundary effects are less severe in the VSA method, due to smaller step size increments.

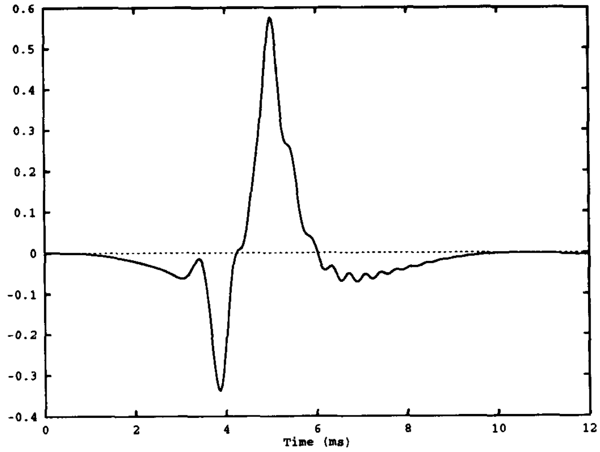


Fig. 5. Channel impulse response.

4) Mean Square Error Evolution: Simple expressions for the convergence rate of the variable step size adaptive algorithm are difficult to develop, due to the complex nature of the step size changes. Using the expressions for the step size evolution, however, we can quantify the steepness of the learning curve by defining the convergence ratio as [28]

$$R_k = \frac{\sigma_{e_{k+1}}^2 - \xi_{\min}}{\sigma_{e_k}^2 - \xi_{\min}} = 1 - 4\sigma_x^2 E[\mu_k] + 4(N+2)\sigma_x^4 E[\mu_k^2] \approx e_x^{-4E[\mu_k]\sigma^2}. \quad (58)$$

It is then clear that the initial step sizes should be chosen to be $\mu_{i,0} = \mu_{\max}$ in order to maximize the initial convergence rate. It is also clear that the convergence rate will decrease as the step size approaches its steady-state value.

E. Discussion

While the analyses of the VS and VSA algorithms are clearly very similar (the difference lies in the definition of the states ϕ_j), the algorithms differ in performance. The step size evolution of the VSA algorithm will generally be slower than that of the VS, due to the need to move through so many states to reach the steady-state value. The steady-state step size value of the VSA algorithm, however, will generally be smaller than that of the VS algorithm, as the reflections from the lower boundary will have less pronounced effects due to the tighter spacing of the step size states.

V. Simulation Results

A. Simulation Types

The simulation results presented in this section will show the potentially superior performance of the VS and VSA algorithms as compared to the LMS algorithm, the advantages of using the VS algorithm with the PTQ algorithm, and the agreement of the analysis with simulation. Results will be presented for adaptive channel equalization with a simple channel with various eigenvalue spreads, equalization of a typical telephone channel, and adaptive line enhancement.

TABLE I
Steady-State Mean Square Error Versus Step Size, Equalizer with Telephone Channel

μ	ξ_{ss}
0.25	0.00204
0.125	0.00160
0.0625	0.00138
0.03125	0.00127
0.015625	0.00122

TABLE II
Algorithm Parameters, Equalizer with Telephone Channel

Algorithm	$\mu_{\max}$	$\mu_{\min}$	m_0	m_1	α
VS	0.25	0.0625	8	9	2
VSA	0.25	0.0625	2	3	2^{-10}
LMS		$\mu = 0.0625$			

B. Adaptive Equalizer with Telephone Network Channel

In these simulations, the variable step size algorithms were tested on a channel which is an approximation to the typical voiceband telephone channel in [14]. The channel impulse response, shown in Fig. 5, has length ten at the signaling rate of 1200 Bd. The input symbol sequence is random and bipolar, and the equalizer has 20 taps. Our simulations will use an additive noise power of $\sigma_n^2 = 0.0001$, which implies $\xi_{\min} = 0.001161$.

We will restrict our choice of step size μ to a power of two. A comparison of some possible choices of μ versus the steady-state mean square error is presented in Table I. Our simulations will use a step size of $\mu = 0.00625$ as a baseline for comparing the LMS and VS methods. The parameters used in each of the algorithms are summarized in Table II.

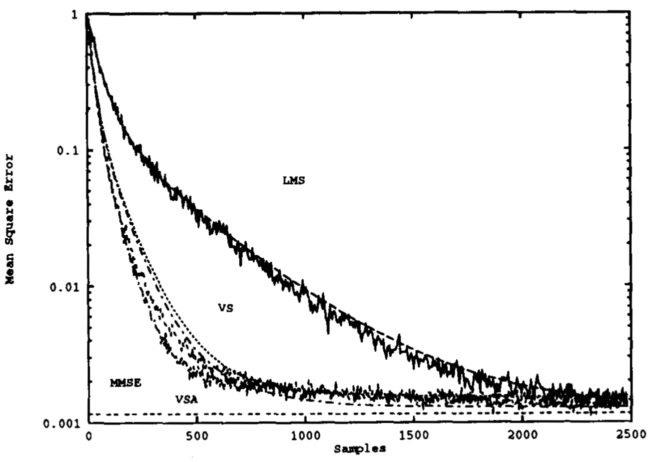


Fig. 6. Simulation and theoretical results for adaptive equalizer on telephone channel.

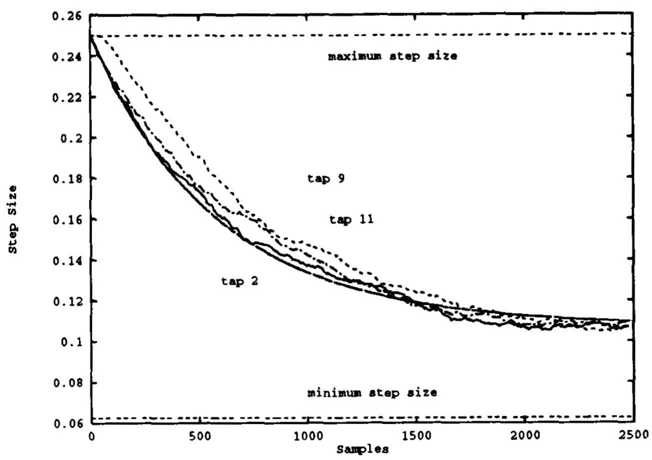


Fig. 7. Step size evolution, VS algorithm, adaptive equalizer on telephone channel.

Analytical and simulations results for the VS and VSA algorithms are shown in Fig. 6. The simulation results are obtained by averaging over 250 runs. Both the VS and VSA algorithms show improvement over the LMS algorithm; the improvement of the VSA method over the VS method is also noticeable. All of the algorithms reach approximately the same steady-state value.

Fig. 7 shows the simulation and analytical results for the step size evolution of the VS algorithm for taps 2, 9, and 11.

Fig. 8 compares the VS-PTQ algorithm against the full precision version of the LMS algorithm. Two configurations for the VS-PTQ are used; one has 8 b for data, 12 b for coefficients, and 8 b for the power-of-two quantizer, while the other has 6 b for data, 10 b for coefficients, and 6 b for the power-of-two quantizer. The theoretical curves for all cases are also shown. Both configurations of the VS-PTQ algorithm perform extremely well as compared to the LMS algorithm.

The agreement between the analytical and simulation results in Figs. 6–8 is very good.

C. Channel Equalizer with Selected Eigenvalue Spreads

This channel has been used extensively as a test case for various adaptive algorithms [12], [14], [22]. The impulse response is

$$h_i = \begin{cases} \frac{1}{2}\left(1 + \cos\left(\frac{2\pi(i-2)}{W}\right)\right) & i = 1, 2, 3 \\ 0 & \text{all other } i \end{cases} \tag{59}$$

where the eigenvalue spread is 6.08 when $W = 2.9$ and 21.71 when $W = 3.3$. The equalizers used have 11 taps. Our simulations will use an additive noise power of $\sigma_n^2 = 0.001$, which gives $\xi_{\min} = 0.001378$ for $W = 2.9$, and $\xi_{\min} = 0.002476$ for $W = 3.3$.

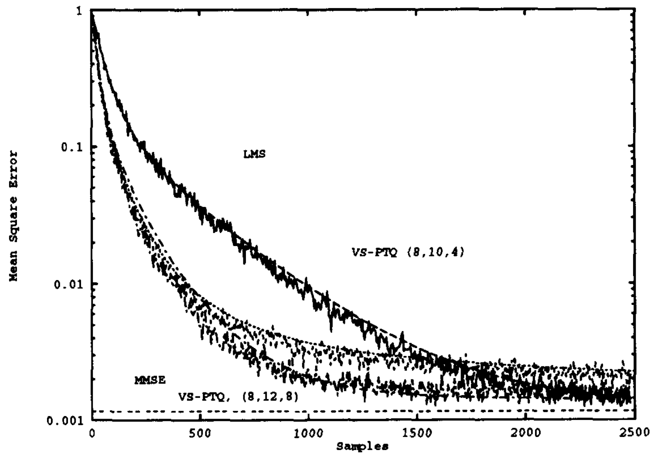


Fig. 8. Comparison of simulation and theoretical results for quantized VS-PTQ algorithm on telephone channel.

The step size μ is once again restricted to a power of two. A comparison of some possible choices of μ versus the steady-state mean square error is shown in Table III. Our simulations will use a step size of $\mu = 0.0625$ as a baseline for comparing the fixed and variable step size methods. The parameters used in each of the algorithms are as shown in Table IV.

The simulation and theoretical results are shown in Figs. 9 and 10 for the different eigenvalue spreads. The simulation curves were generated by averaging over 200 runs. In all cases, the improvement of the VS and VSA algorithms over the LMS algorithm is clearly demonstrated, and the improvement of the VSA method over the VS method is also noticeable. All of the algorithms reach approximately the same steady-state value. The degree of improvement of the variable step size algorithms over the fixed step size algorithms appears to be independent of the eigenvalue spread. The theoretical and simulation results agree reasonably well for all cases.

TABLE III
STEADY-STATE MEAN SQUARE ERROR VERSUS STEP SIZE, CHANNEL WITH SELECTED EIGENVALUES

μ	ξ_{ss}, $W = 2.9$	ξ_{ss}, $W = 3.3$
0.25	0.005535	0.010834
0.125	0.003456	0.006655
0.0625	0.002417	0.004566
0.03125	0.001897	0.003521
0.015625	0.001638	0.002999

TABLE IV
ALGORITHM PARAMETERS, CHANNEL WITH SELECTED EIGENVALUES

Algorithm	μ_{max}	μ_{min}	m_0	m_1	α
VS	0.25	0.0625	6	8	2
VSA	0.25	0.0625	2	3	2^{-11}
LMS		$\mu = 0.0625$			

D. Adaptive Line Enhancer

In this simulation the algorithm configuration is a 64 tap adaptive line enhancer (ALE). The input consists of two sines waves of power and frequency $\sigma_1^2 = 0.5, f_1 = 100$ Hz, and $\sigma_2^2 = 0.3, f_2 = 250$ Hz, corrupted by independent additive Gaussian noise. The sampling frequency is 1000 Hz, and the decorrelation delay is $\delta = 10$. The noise power is $\sigma_n^2 = 0.01$, which results in a Weiner error of $\xi_{\min} = 0.01065$.

A comparison of some possible choices of μ versus the steady-state mean square error is shown in Table V. We will use a step size of $\mu = 0.001$ as a baseline for comparing the LMS and variable step size methods. The parameters used in each of the algorithms are summarized in Table VI.

The simulation and theoretical results are shown in Fig. 11. The VS and VSA algorithms once again outperform the LMS algorithm and the improvement of the VSA method over the VS method is apparent. All of the algorithms converge to approximately the same steady-state value. The theoretical and simulation results show excellent agreement in this case.

E. Adaptive Line Enhancer with Tracking

As in the previous simulation, the algorithms are configured as 64 tap ALE's. In order to demonstrate the superior tracking capability of the variable step size algorithms, the input is chosen to be nonstationary. During the first segment, the sine wave frequency is $f_1 = 100$ Hz, while in the second segment it is $f_2 = 50$ Hz; the sine wave power is $\sigma^2 = 0.5$ in both cases. The other simulation parameters are the same as in the previous simulation. The Weiner error for the first segment is $\xi_{\min} = 0.01032$, and for the second segment $\xi_{\min} = 0.01031$.

A comparison of some possible choices of μ versus the steady-state mean square error is shown in Table VII for

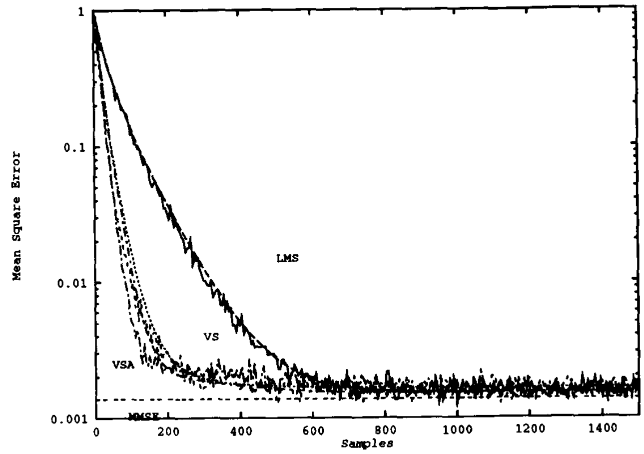


Fig. 9. Simulation and theoretical results for LMS, VS, and VSA algorithms for channel with eigenvalue spread of 6.

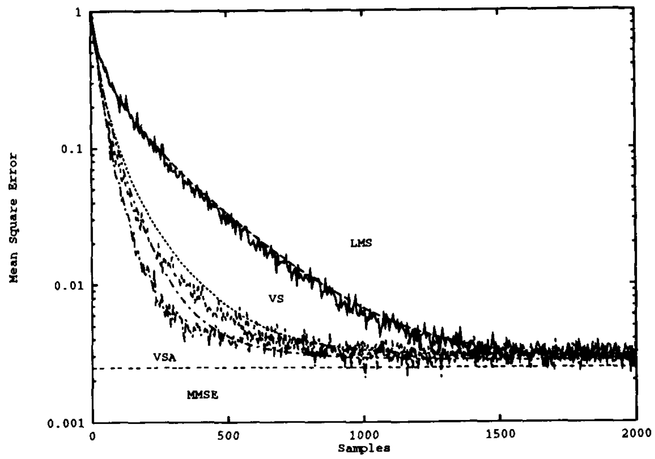


Fig. 10. Simulation and theoretical results for LMS, VS, and VSA algorithms for channel with eigenvalue spread of 21.

TABLE V
STEADY-STATE MEAN SQUARE ERROR VERSUS STEP SIZE, ADAPTIVE LINE ENHANCER

μ	ξ_{ss}
0.1	0.05685
0.01	0.01617
0.001	0.01120
0.0001	0.01070

TABLE VI
ALGORITHM PARAMETERS, ADAPTIVE LINE ENHANCER

Algorithm	μ_{max}	μ_{min}	m_0	m_1	α
VS	0.01	0.001	2	3	2
VSA	0.01	0.001	2	3	0.0001
LMS			$\mu = 0.001$		

both segments. We will use a step size of $\mu = 0.001$ as a baseline for comparing the LMS and variable step size methods. The parameters used in each of the algorithms appear in Table VI.

The simulation and theoretical results are shown in Fig. 12. The VS and VSA algorithms provide superior performance to the LMS algorithm, in both initial convergence and tracking. All of the algorithms converge to approximately the same steady-state value. The theoretical and simulation results again show good agreement.

VI. CMOS IMPLEMENTATION

The VS-PTQ algorithm was implemented in two CMOS ASIC designs, using a 2.5-μm, single level metal, single level poly process. The parameters used for these implementations of the VS-PTQ algorithm are shown in Table VIII.

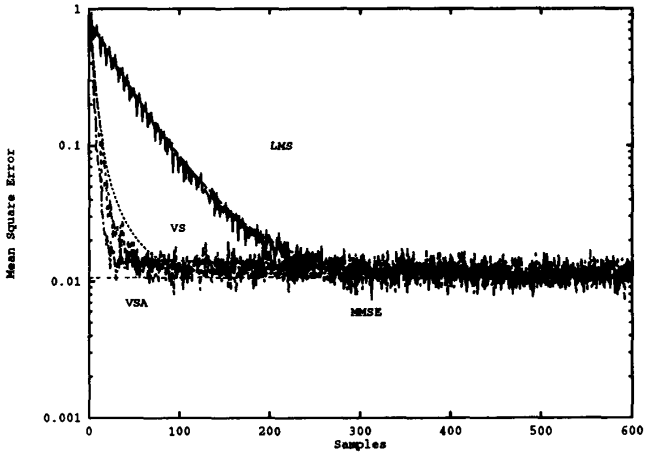


Fig. 11. Simulation and theoretical results for adaptive line enhancer.

TABLE VII
STEADY-STATE MEAN SQUARE ERROR VERSUS STEP SIZE, ALE WITH TRACKING

μ	$\xi_{ss,1}$	$\xi_{ss,2}$
0.1	0.05509	0.05504
0.01	0.01567	0.01565
0.001	0.01085	0.01084
0.0001	0.01037	0.01036

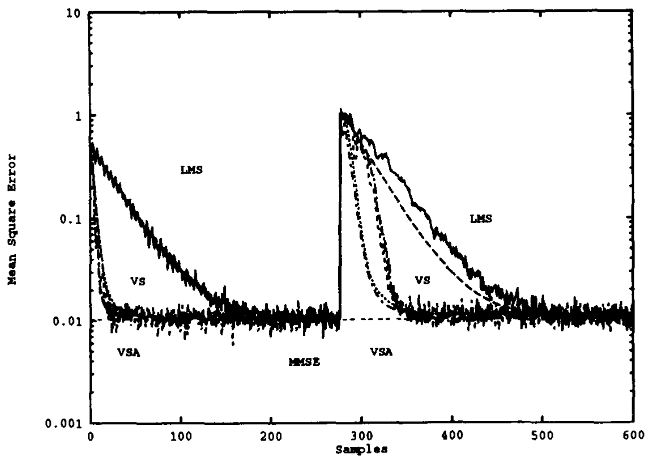


Fig. 12. Simulation and theoretical results for ALE with tracking.

TABLE VIII
PARAMETERS FOR VS-PTQ IMPLEMENTATION

Input Data	8 b
Coefficients	12 b
$\mu_{\max}$	2^{-2}
$\mu_{\min}$	2^{-4}
m_0	8
m_1	9

In the first design, each tap has its own processing element (PE). This allows high speed operation, at the cost of area. The second design has 63 taps on the chip, using a single multiplexed PE to attain area efficiency. While these designs are not for any particular commercial application, they do demonstrate that the VS algorithm is comparable to the conventional LMS algorithm in terms of chip area and speed.

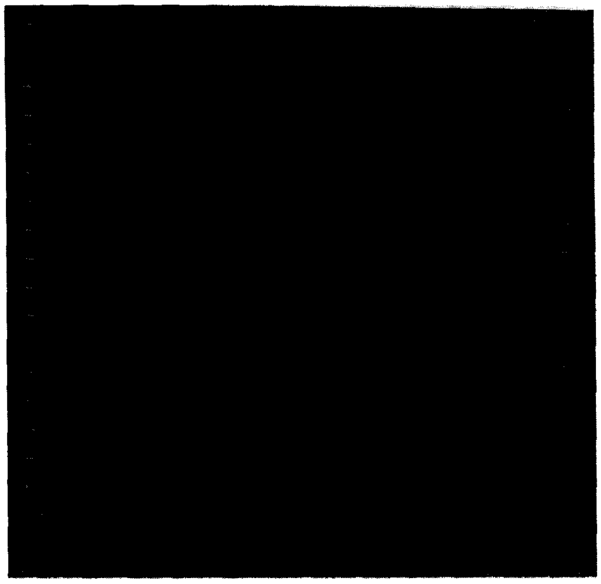


Fig. 13. Variable step size adaptive filtering chip (dedicated PE).

A. VS-PTQ with Dedicated PE per Tap

The high speed VS-PTQ chip has four identical sections, each corresponding to a filter tap. Each tap uses a 9 by 8 b two's complement array multiplier and an 8-b Brent–Kung adder for the filter section. The coefficient update circuitry consists of a 12-b Brent–Kung adder for accumulation, an 8 by 8 b power-of-two quantizer-multiplier, and variable step size shift and update circuitry. Shift registers are used for data storage; each tap has two 8-b registers for data, a 12-b register for coefficients, and a 12-b combination serial/parallel register for step size data. Pseudo two-phase clocking is used, and a global reset signal is provided.

A photograph of the layout of an implementation of the VS-PTQ algorithm in 2.5-μm single level metal CMOS is shown in Fig. 13. The chip is expandable by cascading beyond the integral four taps. Some details of the implementation are shown in Table IX. With suitable λ-rule scaling, this chip can be shown to be comparable to that reported in [4] in terms of area and speed, while providing the superior performance of the VS-PTQ algorithm.

B. VS-PTQ with Multiplexed PE

This chip shares the use of a single processing element so as to be more area efficient. It uses a 9 by 8 b two's complement array multiplier, and an 8-b Brent–Kung adder for the FIR filter. The coefficient update section has a 12-b Brent–Kung adder, an 8 by 8 b power-of-two quantizer-multiplier, and the variable step size circuitry. Data storage consists of static RAM; the data bank is 64 by 8 b, the coefficient bank is 64 by 12 b, and the variable step size bank is 64 by 12 b. Pseudo two-phase clocking is used, and linear feedback shift registers are used as counters. A global reset signal is provided.

TABLE IX
VS-PTQ Chip, Dedicated PE

Filter length	4 Taps
Maximum operating speed	7.14 MHz
Active area	5.77 by 5.72 mm
Die size	7 by 7 mm
Transistors	19992

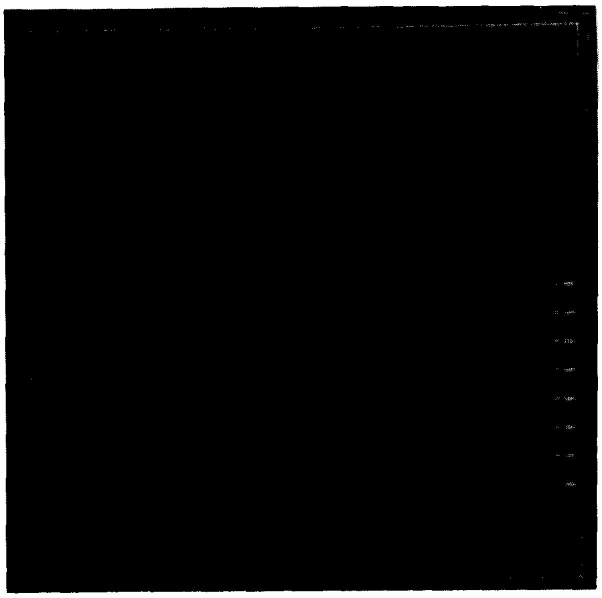


Fig. 14. Variable step size adaptive filtering chip (multiplexed PE).

TABLE X
VS-PTQ Chip, Multiplexed PE

Filter length	63 Taps
Maximum operating speed	32 kHz
Active area	5.25 by 5.25 mm
Die size	7 by 7 mm
Transistors	24635

The layout of this implementation is shown in Fig. 14. Some details of the implementation are shown in Table X.

VII. Conclusion

We have studied variable step size gradient search algorithms which use relatively simple methods for step size selection. In addition to studying a previously proposed algorithm (VS) which uses geometric step size updates, we have also presented a new variable step size algorithm which uses arithmetic step size updates (VSA).

The variable step size algorithms were analyzed under a number of simplifying assumptions. Expressions for the step size moments and the first and second moments of the coefficient misadjustment were obtained. These were used to find the mean square error and convergence rate.

Conditions for the convergence of the algorithms were also provided.

We have compared analytical and simulation results for the variable step size algorithms in the contexts of channel equalization and adaptive line enhancement; these have shown good agreement. The VS and VSA algorithms display convergence speed advantages over the simple LMS algorithm, and the VSA algorithms outperforms the VS algorithm at the cost of increased implementation complexity. Both variable step size methods are sensitive to parameter selections, however, and may perform poorly, as compared to the LMS algorithm, if the step size parameters are chosen unwisely. Further, the variable step size algorithms may show only minor convergence speed advantages over the LMS algorithm in low signal-to-noise ratio environments.

CMOS ASIC implementations of the VS-PTQ algorithm were designed. One of these employs dedicated processing elements to attain high speed, while the other uses a multiplexed processing element for area efficiency. These designs demonstrated that performance may be increased through variable step size methods, with only a modest increase in complexity.

In a favorable environment, and with reasonable parameter selection, these variable step size methods will provide significantly improved convergence performance, at little cost in steady-state error, and with low implementation costs.

Appendix A
Derivations for Misadjustment and Step Size Evolution

1. *Mean of Coefficient Misadjustment (13)*

From the definition of the weight vector update, (6), and the coefficient misadjustment, (12),

$$V_{k+1} = V_k + 2M_k e_k X_k. \tag{A.1}$$

The evolution of the coefficient misadjustment mean is thus given by

$$E[V_{k+1}|V_k] = V_k + 2E[M_k e_k X_k | V_k]. \tag{A.2}$$

Equation (13) follows from the assumption that the current step size is independent of the current error, misadjustment, and input data.

2. *Second Moment of Coefficient Misadjustment (14)*

Using the definition of K_k and (A.1), it is simple to show that

$$K_{k+1} = K_k + 2E[M_k e_k X_k V_k^T] + 2E[V_k X_k^T e_k M_k] + 4E[M_k e_k X_k X_k^T e_k M_k]. \tag{A.3}$$

Assuming the step size matrix M_k is independent of the coefficient misadjustment vector V_k, the update $e_k X_k$, and the error e_k, the second term becomes

$$\begin{aligned} E[M_k e_k X_k V_k^T] &= E[M_k] E[e_k X_k V_k^T] \\ &= -E[M_k] E[R V_k V_k^T] \\ &= -E[M_k] R K_k. \end{aligned} \tag{A.4}$$

Similarly,

$$E[V_k X_k^T e_k M_k] = -K_k R E[M_k]. \tag{A.5}$$

To simplify the last term of (A.3), we condition on M_k and V_k, apply assumption (11), and use Price's theorem [20] to find

$$\begin{aligned} &E[M_k e_k X_k X_k^T e_k M_k | M_k, V_k] \\ &\quad = M_k E[e_k X_k X_k^T e_k | V_k] M_k. \\ &\quad = M_k(\sigma_{ek}^2 R + 2R V_k V_k^T R) M_k. \end{aligned} \tag{A.6}$$

Thus

$$E[M_k e_k X_k X_k^T e_k M_k] = \sigma_{ek}^2 E[M_k R M_k] + 2E[M_k R K_k R M_k]. \tag{A.7}$$

Combining (A.4), (A.5), and (A.7), we get (14).

3. *Sign Change Probability (18)*

Assume that e_k is Gaussian with zero mean and variance σ_{ek}^2, and the $x_{i,k}$ are Gaussian with zero mean and variance σ_x^2. Further, assume that the error e_k and the inputs $x_{i,k}$ are jointly Gaussian with cross-correlation coefficient $\rho_{i,k} = E[e_k x_{i,k}]/(\sigma_x \sigma_{ek})$. The distribution of the product $e_k x_{i,k}$ is then [19]

$$\begin{aligned} f(z) &= \frac{1}{\pi \sigma_x \sigma_{ek} \sqrt{1-\rho_{i,k}^2}} \exp\left(\frac{z\rho_{i,k}}{\sigma_x \sigma_{ek}(1-\rho_{i,k}^2)}\right) \int_0^\infty \frac{1}{w} \\ &\quad \cdot \exp\left[-\frac{1}{2(1-\rho_{i,k}^2)}\left(\frac{w^2}{2\sigma_{ek}^2} + \frac{z^2}{2\sigma_x^2 w^2}\right)\right] dw \\ &= \frac{1}{\pi \sigma_x \sigma_{ek} \sqrt{1-\rho_{i,k}^2}} \exp\left(\frac{z\rho_{i,k}}{\sigma_x \sigma_{ek}(1-\rho_{i,k}^2)}\right) \\ &\quad \cdot K_0\left(\frac{z}{(1-\rho_{i,k}^2)\sigma_x \sigma_{ek}}\right) \end{aligned} \tag{A.8}$$

where $K_0(\cdot)$ is the zero-order modified Bessel function of the second kind. Since we know that the mean of $e_k x_{i,k}$ will converge toward zero, the probability that coefficient $w_{i,k}$ will move in the direction of its steady-state solution is one-half plus the area under $f(z)$ from 0 to $E[e_k x_{i,k}]$, that is,

$$p_{i,k} = \tfrac{1}{2} + b_{i,k} \tag{A.9}$$

where

$$\begin{aligned} b_{i,k} &= \frac{1}{\pi \sigma_x \sigma_{ek} \sqrt{1-\rho_{i,k}^2}} \int_0^{E[e_k x_{i,k}]} \exp\left(\frac{z\rho_{i,k}}{\sigma_x \sigma_{ek}(1-\rho_{i,k}^2)}\right) \\ &\quad \cdot K_0\left(\frac{z}{(1-\rho_{i,k}^2)\sigma_x \sigma_{ek}}\right) dz \end{aligned} \tag{A.10}$$

is the bias due to the coefficient misadjustment. This integral may be evaluated numerically, or preferably, the kernel of the bias integral may be expressed as a series expansion [15]. For small $e_k x_{i,k}$, it is reasonable to approximate the integral using the first term, yielding

$$b_{i,k} \approx -\frac{|E[e_k x_{i,k}]|}{\pi \sigma_x \sigma_{e_k} \sqrt{1-\rho_{i,k}^2}} \cdot \left\{1 - \gamma - \ln\left(\frac{|E[e_k x_{i,k}]|}{2\sigma_x \sigma_{e_k}(1-\rho_{i,k}^2)}\right)\right\}. \quad \text{(A.11)}$$

Additional terms of the expansion may be necessary where $e_k x_{i,k}$ is large. Equations (18) and (19) then follow directly.

Appendix B
VS-PTQ Derivations

We first prove two lemmas for later use. The approach used in these proofs is similar to that used for a different nonlinearity in [16].

Lemma 1: Assume that X_1 and X_2 are jointly Gaussian, zero-mean random variables, with autocorrelation matrix

$$R = \begin{bmatrix} \sigma_1^2 & r_{12} \\ r_{12} & \sigma_2^2 \end{bmatrix} \quad \text{(B.1)}$$

and $q(\cdot)$ is defined in [29]. Then

$$E[q(X_2)X_1] = A(B, \sigma_2) r_{12} \quad \text{(B.2)}$$

where $A(\cdot, \cdot)$ is given by (29).

Proof: From Price's theorem [20],

$$\begin{aligned} \frac{\partial E[q(X_2)X_1]}{\partial r_{12}} &= E\left[\frac{\partial q(X_2)}{\partial X_2}\frac{\partial X_1}{\partial X_1}\right] \\ &= E\Bigg[\sum_{i=1}^{B-1} 2^{-i}\{\delta(X_2 - 2^{-i+1}) \\ &\quad + \delta(X_2 + 2^{-i+1})\} \\ &\quad + 2^{-B+1}\{\delta(X_2 - 2^{-B+1}) \\ &\quad + \delta(X_2 + 2^{-B+1})\}\Bigg] \\ &= A(B, \sigma_2) \end{aligned} \quad \text{(B.3)}$$

where $\delta(\cdot)$ is the Dirac delta function. Equations (B.2) and (29) follow from integration, noting that the constant of integration is zero, since $E[q(X_1)X_2] = 0$ when $r_{12} = 0$.

Lemma 2: Assume that X_1, X_2, and X_3 are jointly Gaussian zero-mean random variables, with autocorrelation matrix

$$R = \begin{bmatrix} \sigma_1^2 & r_{12} & r_{13} \\ r_{12} & \sigma_2^2 & r_{23} \\ r_{13} & r_{23} & \sigma_3^2 \end{bmatrix}. \quad \text{(B.4)}$$

Then

$$E[q^2(X_3)X_1X_2] = F(B, \sigma_3) r_{12} + D(B, \sigma_3)|r_{13} r_{23} \quad \text{(B.5)}$$

where $D(\cdot, \cdot)$ and $F(\cdot, \cdot)$ are given by (31) and (32), and erfc$(\cdot)$ is the complementary error function.

Proof: Proceeding in the same manner as for Lemma 1, and using results from [29], we find that

$$\frac{\partial E[q^2(X_3)X_1X_2]}{\partial r_{12}} = E[q^2(X_3)] = F(B, \sigma_3). \quad \text{(B.6)}$$

Integrating with respect to r_{12}, we get

$$E[q^2(X_3)X_1X_2] = F(B, \sigma_3) r_{12} + C_2. \quad \text{(B.7)}$$

In order to determine the constant of integration C_2, we apply Price's theorem again,

$$\begin{aligned} \frac{\partial E[q^2(X_3)X_1X_2]}{\partial r_{13}} &= E\left[X_2 \frac{\partial q^2(X_3)}{\partial r_{13}}\right] \\ &= E\Bigg[X_2 \sum_{i=1}^{B-1} 2^{-2i}\{\delta(X_3 - 2^{-i+1}) \\ &\quad - \delta(X_3 + 2^{-i+1})\} \\ &\quad + 2^{-2(B-1)}\{\delta(X_3 - 2^{-B+1}) \\ &\quad - \delta(X_3 + 2^{-B+1})\}\Bigg] \\ &= D(B, \sigma_3) r_{23} \end{aligned} \quad \text{(B.8)}$$

where the assumption that X_2 and X_3 are jointly Gaussian has been used. We can then integrate with respect to r_{13}, to find

$$E[q^2(X_3)X_1X_2] = r_{13} r_{23} D(B, \sigma_3) + C_3 \quad \text{(B.9)}$$

where C_3 is another constant of integration. Equation (B.5) follows from (B.7), (B.9), and

$$E[q^2(X_3)X_1X_2]|_{r_{13}=0 \text{ and } r_{23}=0} = 0. \quad \text{(B.10)}$$

1. Mean of Coefficient Misadjustment (28)

From (10) and (12),

$$\boldsymbol{V}_{k+1} = \boldsymbol{V}_k + 2\boldsymbol{M}_k q(e_k)\boldsymbol{X}_k. \quad \text{(B.11)}$$

Taking the conditional and expectation, and assuming that the current step sizes are independent of the data and error, we find

$$E[\boldsymbol{V}_{k+1}|\boldsymbol{V}_k] = \boldsymbol{V}_k + 2E[\boldsymbol{M}_k]E[q(e_k)\boldsymbol{X}_k|\boldsymbol{V}_k]. \quad \text{(B.12)}$$

We then apply Lemma 1, and (11) and (12), to yield

$$E[\boldsymbol{V}_{k+1}|\boldsymbol{V}_k] = \boldsymbol{V}_k + 2A(B, \sigma_{e_k})E[\boldsymbol{M}_k]E[e_k\boldsymbol{X}_k|\boldsymbol{V}_k]. \quad \text{(B.13)}$$

Next note that $E[e_k X_k | V_k] = -RV_k$; substituting this into the previous equation, and taking the expectation, will yield (28).

2. *Second Moment of Coefficient Misadjustment (30)*

From $K_k = E[V_k V_k^T]$ and (B.11) we have

$$K_{k+1} = K_k + 2E[M_k V_k X_k^T q(e_k)] + 2E[q(e_k) X_k V_k^T M_k] + 4E[M_k q^2(e_k) X_k X_k^T M_k]. \tag{B.14}$$

Applying Lemma 1 and (11), and assuming that the current step size is independent of the current data and error signals, we get

$$E[M_k V_k X_k^T q(e_k)] = E[M_k] E[E[V_k X_k^T q(e_k) | V_k]] = -A(B, \sigma_{e_k}) E[M_k] K_k R \tag{B.15}$$

and

$$E[q(e_k) X_k V_k^T M_k] = -A(B, \sigma_{e_k}) R K_k E[M_k]. \tag{B.16}$$

Using Lemma 2 and (11), and the independence assumption once again, we find

$$\begin{aligned} E[M_k q^2(e_k) X_k X_k^T M_k] &= E[E[M_k q^2(e_k) X_k X_k^T M_k | M_k, V_k]] \\ &= E[M_k E[q^2(e_k) X_k X_k^T | V_k] M_k] \\ &= D(B, \sigma_{e_k}) E[M_k R K_k R M_k] \\ &\quad + \sigma_{e_k}^2 F(B, \sigma_{e_k}) E[M_k R M_k] \end{aligned} \tag{B.17}$$

where $D(B, \sigma_{e_k})$ and $F(B, \sigma_{e_k})$ are given in (31) and (32), respectively. Substituting (B.15), (B.16), and (B.17) into (B.14), we obtain (30).

3. *Finite Word Length (35)*

In [5], [29] it is shown that the primary contributions to the finite word-length effects in the mean-square error can be expressed as $\bar{\sigma}_{e_k}^2 = \nu_k + \eta$, where $\eta = \rho\sigma_d^2$, $\nu_k = E[W_k^T W_k]\sigma_d^2$, and ρ depends on the order in which quantization and accumulation are performed. We can rewrite ν_k as

$$\nu_k = (E[V_k^T V_k] + E[V_k^T] W_{opt} + W_{opt} E[V_k^T] + W_{opt}^T W_{opt})\sigma_d^2. \tag{B.18}$$

But $E[V_k^T V_k] = \mathrm{tr}(K_k)$, so

$$E[W_k^T W_k] = \mathrm{tr}(K_k) + 2E[V_k^T] W_{opt} + W_{opt}^T W_{opt}. \tag{B.19}$$

Combining these results, we get (35).

Appendix C
White Input Derivations

1. *Convergence of Misadjustment Mean (40)*

From (36), it is clear that the mean of the misadjustment will go to zero if

$$0 < E[\mu_{i,k}] < \frac{1}{\sigma_x^2} \qquad \forall i, k. \tag{C.1}$$

Since the $\mu_{i,k}$ are bounded by μ_{max}, (40) follows.

2. *Convergence of Mean Square Error (41)*

From (38) and (39), we have

$$\theta_{k+1} = \sum_{i=0}^{N-1} (1 - 4\sigma_x^2 E[\mu_{i,k}] + 8\sigma_x^4 E[\mu_{i,k}^2]) K_{i,k} + 4\sigma_x^4 \theta_k \sum_{i=0}^{N-1} E[\mu_{i,k}^2] + 4\xi_{min}\sigma_x^2 \sum_{i=0}^{N-1} E[\mu_{i,k}^2]. \tag{C.2}$$

We note that

$$\sum_{i=0}^{N-1} (1 - 4\sigma_x^2 E[\mu_{i,k}] + 2\sigma_x^4 E[\mu_{i,k}^2]) K_{i,k} \le (1 - 4\sigma_x^2 \mu_{min} + 8\sigma_x^4 \mu_{max}^2)\theta_k \tag{C.3}$$

and

$$\sum_{i=0}^{N-1} E[\mu_{i,k}^2] \le N\mu_{max}^2. \tag{C.4}$$

Define ψ_k by

$$\psi_{k+1} = R\psi_k + 4N\xi_{min}\sigma_x^2 \mu_{max}^2 \tag{C.5}$$

where

$$R = 1 - 4\sigma_x^2 \mu_{min} + 4(N + 2)\sigma_x^4 \mu_{max}^2. \tag{C.6}$$

We require that $R < 1$ for the convergence of the ψ_k. It is clear from these definitions that if $\psi_{k_0} > \theta_{k_0}$ for some k_0, then $\theta_k \le \psi_k$ for all $k > k_0$; hence, we require that $R < 1$ for the convergence of the θ_k. Equation (41) then follows directly.

3. *Steady-State Step Sizes (44), (48)*

Given $p = 1/2$, the constraint (43), and (22)–(25), the steady-state condition of the step size system may be reduced to the set of equations

$$\begin{aligned} P_{\infty,0,0,h} &= dP_{\infty,0,0,h-1} + iP_{\infty,0,0,h+1}, \quad h = 1, \cdots, M-2 \\ P_{\infty,0,0,0} &= \frac{i}{1-i} P_{\infty,0,0,1} \\ P_{\infty,0,0M-1} &= \frac{d}{1-d} P_{\infty,0,0,M-2} \end{aligned} \tag{C.7}$$

where i and d are given by (47). The solution to this system of equations is

$$P_{\infty,0,0,h} = \begin{cases} \dfrac{1}{H\left(1 + \sum_{l=1}^{c} \prod_{r=l-1}^{c-1} F(i, d, r) + \sum_{l=1}^{c+1} \prod_{r=l-1}^{c} F(d, i, r)\right)} & h = c \\ \prod_{r=h}^{c-1} F(i, d, r) P_{\infty,0,0,c} & h = 0, 1, \cdots, c-1 \\ \prod_{r=M-1-h}^{c} F(d, i, r) P_{\infty,0,0,c} & h = c+1, c+2, \cdots, M-1 \end{cases} \quad \text{(C.8)}$$

where c is given by (46)

$$H = \frac{1}{\sum_{h=0}^{M-1} P_{\infty,0,0,h}} \quad \text{(C.9)}$$

and

$$F(\vartheta, \varrho, n) = \begin{cases} \dfrac{\vartheta}{1 - \varrho F(\vartheta, \varrho, n-1)} & n > 0 \\ \dfrac{\vartheta}{1 - \vartheta} & n = 0. \end{cases} \quad \text{(C.10)}$$

The steady-state step size may be expressed as

$$E[\mu_\infty] = \sum_{f=0}^{m_0-1} \sum_{g=0}^{m_1-1} \sum_{h=0}^{M-1} \phi_h P_{\infty,f,g,h} = H \sum_{h=0}^{M-1} \phi_h P_{\infty,0,0,h}. \quad \text{(C.11)}$$

For the values encountered in this system, the approximation $F(\vartheta, \varrho, n) = \vartheta$ can be successfully applied; (44) and (48) then follow.

References

[1] S. T. Alexander, *Adaptive Signal Processing: Theory and Applications.* Springer-Verlag, 1986.

[2] M. Bellanger, *Digital Processing of Signals: Theory and Practice.* New York: Wiley, 1984.

[3] M. Bellanger, *Adaptive Digital Filters and Signal Analysis.* M. Dekker, 1987.

[4] T. H. Bui, "A high speed LMS adaptive filter," in *Proc. IEEE Custom IC Conf.*, May 1987, pp. 227–229.

[5] C. Caraiscos and B. Liu, "A roundoff error analysis of the LMS adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 1, pp. 34–41, Feb. 1984.

[6] C. F. N. Cowan and P. M. Grant, *Adaptive Filters.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[7] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 4, pp. 578–586, Aug. 1982.

[8] J. B. Evans and B. Liu, "Variable step size methods for the LMS adaptive algorithm," in *Proc. IEEE Int. Symp. Circuits Syst.*, Apr. 1987, pp. 422–425.

[9] J. B. Evans and B. Liu, "A CMOS implementation of a variable step size digital adaptive filter," in *Proc. IEEE Trans. Acoust., Speech, Signal Processing*, 1989, pp. 2489–2492.

[10] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

[11] R. W. Harris, D. M. Chabries, and F. A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 2, pp. 309–316, Apr. 1986.

[12] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1986.

[13] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications.* Kluwer Academic, 1984.

[14] R. W. Lucky and H. Rudin, "An automatic equalizer for general-purpose communication channels," *Bell Syst. Tech. J.*, vol. 46, pp. 2179–2208, Nov. 1967.

[15] Y. L. Luke, *Integrals of Bessel Functions.* New York: McGraw-Hill, 1962.

[16] V. J. Mathews, "Performance analysis of adaptive filters with the dual sign algorithm," *IEEE Trans. Signal Processing*, vol. 39, no. 1, pp. 85–91, Jan. 1991.

[17] V. J. Mathews and S. H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 4, pp. 450–454, Apr. 1987.

[18] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963–993, May–June 1979.

[19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 1965.

[20] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT-4, no. 4, pp. 66–72, June 1958.

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

[22] E. H. Satoris and J. Pack, "Application of least squares lattice algorithm to adaptive equalization," *IEEE Trans. Commun.*, vol. COM-29, no. 2, pp. 136–142, Feb. 1981.

[23] S. D. Stearns and R. A. David, *Signal Processing Algorithms.* Englewood Cliffs, NJ: Prentice-Hall, 1988.

[24] J. R. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters.* New York: Wiley, 1986.

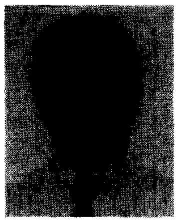
[25] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communications," *IBM J. Res. Develop.*, pp. 546–555, Nov. 1972.

[26] 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.

[27] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

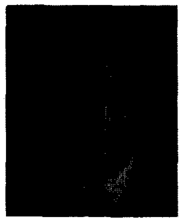
[28] P. Xue and B. Liu, "On the convergence of a variable step size adaptive equalizer," in *Proc. IEEE Int. Symp. Circuits Syst.*, June 1985, pp. 1161–1162.

[29] P. Xue and B. Liu, "Adaptive equalizer using finite-bit power-of-two quantizer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 6, pp. 1603–1611, Dec. 1986.



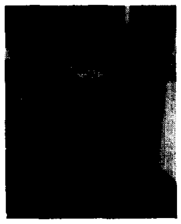
Joseph B. Evans (S'80–M'83) was born in New Jersey on March 24, 1961. He received the B.S.E.E. degree from Lafayette College in 1983, and the M.S.E., M.A., and Ph.D. degrees from Princeton University in 1984, 1986, and 1989, respectively.

Since 1989 he has been an Assistant Professor of Electrical and Computer Engineering at the University of Kansas. Prior to this, he held a postdoctoral position in the Network Systems Research Department of AT&T Bell Laboratories in Holmdel, NJ, where he was involved in the design of a high performance integrated network. While at Princeton, he was awarded an AT&T Bell Laboratories Graduate Fellowship for 1984–1988; during this time, he was also a part-time employee of Bell Labs, working in the field of speech processing algorithms for packet networks. His current research interests include adaptive signal processing, digital filtering, digital speech processing, special-purpose computer architecture, high speed (gigabit) networks, and VLSI implementations.



Ping Xue (S'81–M'84) received the M.S.E., M.A., and Ph.D. degrees, all in electrical engineering and computer science, from Princeton University, Princeton, NJ, in 1981, 1982, and 1985, respectively.

He was a Member of the Technical Staff at David Sarnoff Research Center from 1984 to 1986, and on the faculty of Jiao Tong University, Shanghai, China, from 1986 to 1990. He visited Princeton again at the beginning of 1991. Since April 1991, he has been a Principal Engineer of TriTech Microelectronics International Pte. Ltd., a subsidiary of Singapore Technologies. His research interests are in digital audio/video processing, data communication, and VLSI DSP implementations.



Bede Liu (S'55–M'62–F'72) was born in Shanghai, China. He received the B.S.E.E. degree from the National Taiwan University, and the M.E.E. and D.E.E. degrees from the Polytechnic Institute of Brooklyn.

Prior to joining the Princeton Faculty in 1962, he had been with Bell Laboratories, Allen B. DuMont Laboratory, and Western Electric Company. He has also been a visiting faculty member at the National Taiwan University, Taipei, Taiwan; the University of California, Berkeley; the Imperial College, London; and the Jiao Tong University, Shanghai. He is currently a Professor of Electrical Engineering at Princeton University. He edited the research volume *Digital Filters and the Fast Fourier Transform* (DHR, 1975) and coauthored the senior level textbook *Digital Signal Processing: Theory, Design, and Implementation* (Wiley, 1976). His current research interests lie mainly in image and signal processing.

Dr. Liu was an administrative committee (ADCOM) member of the IEEE Acoustics, Speech, Signal Processing Society, an ADCOM member of the IEEE Circuits and Systems Society, the Technical Program Chairman for the 1978 International Symposium on Circuits and Systems, the Vice President (1979) and President (1982) of the IEEE Circuits and Systems Society, and the IEEE Division I Director for 1984 and 1985. He was a recipient of an IEEE Centennial Medal, the 1985 Technical Achievement Award of the IEEE Acoustics, Speech, Signal Processing Society, and the 1988 Eduction Award of the IEEE Circuits and Systems Society. In September 1988 he was named an honorary professor by the Institute of Acoustics and the Institute of Electronics of the Academia Sinica.