

## The Stability of Adaptive Minimum Mean Square Error Equalizers Using Delayed Adjustment

PETER KABAL, MEMBER, IEEE

**Abstract**—The necessary and sufficient conditions for the stability and convergence of adaptive minimum mean square error equalizers are developed. The analysis includes the effect of delays in the adjustment path, as occur, for instance, when the decoder in a data transmission system delays its output decisions.

### INTRODUCTION

Adaptive equalizers have become a common feature in high-speed data modems. The usual adaptation procedure is some version of the steepest descent algorithm which minimizes the mean square error. These equalizers have been used in systems employing threshold detectors. One feature of these detectors is that no decision delay is involved. Widrow's analysis [1] gives the stability and convergence conditions for these adaptive equalizers used with symbol-by-symbol detectors.

More recently, interest has been shown in using detectors that perform better than symbol-by-symbol detectors. Two examples are maximum likelihood sequence estimation using the Viterbi algorithm [2], [3] and the sequential detector described by Abend and Fritchman [4]. While these schemes perform better than conventional detectors, they also do so at the expense of complexity. Their complexity grows exponentially with the system pulse response duration.

In order to reduce the complexity of the detector, it has been suggested that a transversal equalizer be used to reduce the time dispersion of the system pulse response [5]–[7]. In an adaptive version in which the decisions from the detector involve delays, the stability and convergence criteria of the equalizer adjustment procedure must be examined anew. The analysis in this paper gives a result which replaces the empirically derived bound given in [5].

### ADAPTIVE EQUALIZATION

Fig. 1 shows a model for an adaptive equalizer. The output of the equalizer is given by

$$y_k = \mathbf{c}^{(k)T} \mathbf{r}^{(k)} \quad (1)$$

where  $\mathbf{r}^{(k)}$  is the column vector of received samples on the delay line at time  $k$  and  $\mathbf{c}^{(k)}$  is the vector of  $M$  tap coefficients

Paper approved by the Editor for Data Communication Systems of the IEEE Communications Society for publication without oral presentation. Manuscript received October 30, 1980; revised October 10, 1982.

The author is with the Department of Electrical Engineering, McGill University, Montreal, P.Q., Canada H3A 2A7, and INRS-Telecommunications, University of Quebec, Verdun, P.Q., Canada H3E 1H6.

of the transversal filter. The steepest descent algorithm for adjusting the equalizer to minimize the mean square error takes a step in the direction opposite to the direction of the gradient of the error  $D$  time units ago, that is, at time  $k + 1$  the weight vector becomes

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} - s_k \mathbf{g}^{(k-D)} \quad (2)$$

where  $s_k$  is a parameter controlling the step size and  $\mathbf{g}^{(k)}$  is the gradient vector

$$\mathbf{g}^{(k)} = E \left[ \frac{\partial e_k^2}{\partial \mathbf{c}^{(k)}} \right]. \quad (3)$$

$e_k$  is the error, that is, the difference between the desired and actual response,

$$e_k = d_k - y_k. \quad (4)$$

The stochastic version of this algorithm will be introduced later.

The weight adjustment algorithm then becomes

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \Delta_k E[e_{k-D} \mathbf{r}^{(k-D)}] \quad (5)$$

or

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} - \Delta_k (\mathbf{A} \mathbf{c}^{(k-D)} - \boldsymbol{\alpha}) \quad (6)$$

where  $\Delta_k$  is the step size parameter (equal to  $2s_k$ ),  $\mathbf{A}$  is the correlation matrix of the input sequence (assuming  $\mathbf{r}^{(k)}$  and  $d_k$  are stationary sequences)

$$\mathbf{A} = E[\mathbf{r}^{(k)} \mathbf{r}^{(k)T}] \quad (7)$$

and  $\boldsymbol{\alpha}$  is the cross-correlation vector

$$\boldsymbol{\alpha} = E[d_k \mathbf{r}^{(k)}]. \quad (8)$$

The optimum value of the weight vector (that which minimizes the mean square error) is given by

$$\mathbf{c}_{\text{opt}} = \mathbf{A}^{-1} \boldsymbol{\alpha}. \quad (9)$$

In order to decouple the weight adjustments, define the transformation

$$\mathbf{c}' = \mathbf{P} \mathbf{c} \quad (10)$$

where  $\mathbf{P}$  is an orthonormal matrix which diagonalizes  $\mathbf{A}$

$$\mathbf{A} = \mathbf{P}^{-1} \boldsymbol{\Lambda} \mathbf{P} \quad (11)$$

and  $\boldsymbol{\Lambda}$  is the diagonal matrix of the eigenvalues of  $\lambda_j$  of  $\mathbf{A}$ . Then

$$\mathbf{c}'^{(k+1)} = \mathbf{c}'^{(k)} - \Delta_k (\boldsymbol{\Lambda} \mathbf{c}'^{(k-D)} - \mathbf{P} \boldsymbol{\alpha}). \quad (12)$$

Fig. 2 shows a feedback model for this adjustment algorithm. The optimum decoupled weight vector can be written as

$$\mathbf{c}'_{\text{opt}} = \boldsymbol{\Lambda}^{-1} \mathbf{P} \boldsymbol{\alpha}. \quad (13)$$

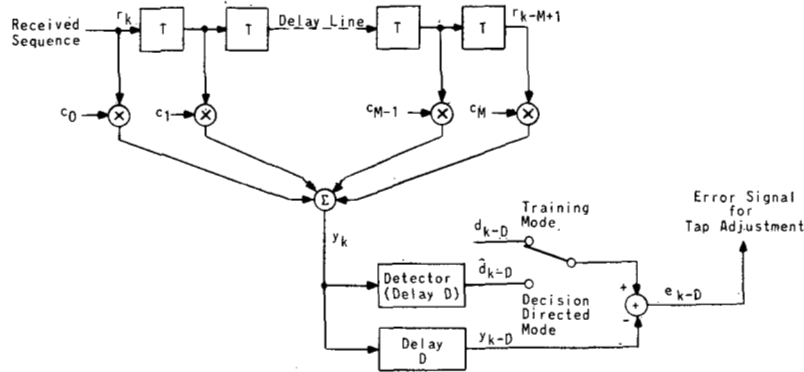


Fig. 1. An adaptive equalizer with delayed decisions.

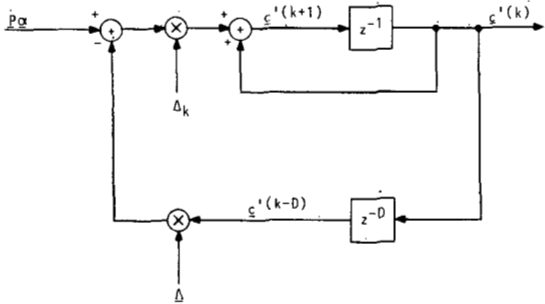


Fig. 2. Feedback model for the tap weight adjustment.

Let us define the tap weight error as

$$h^{(k)} = c^{(k)} - c'_{opt}. \tag{14}$$

Then

$$h^{(k+1)} = h^{(k)} - \Delta_k \Lambda h^{(k-D)}. \tag{15}$$

Further assume  $\Delta_k$  to be constant and equal to  $\Delta$ . Then for each  $j, j = 1, \dots, M$ , the  $z$ -transform of the  $j$ th tap coefficient error is

$$H_j(z) = \frac{z^{D+1} h_j^{(0)}}{z^{D+1} - z^D + \Delta \lambda_j}. \tag{16}$$

The limit as  $k \rightarrow \infty$  of  $|h_j^{(k)}|$  is zero if and only if all of the poles of  $H_j(z)$  are within the unit circle in the complex  $z$ -plane [8, p. 492]. If

$$\lim_{k \rightarrow \infty} |h_j^{(k)}| = 0 \quad \text{for each } j, \quad j = 1, \dots, M \tag{17}$$

then  $\lim_{k \rightarrow \infty} c^{(k)} = c'_{opt}$ .

Stability also requires that all of the poles of  $H_j(z)$  be within the unit circle [9]. Thus, we conclude that the adjustment algorithm (5) is stable and converges if and only if, for each  $j$ , all of the roots of the characteristic polynomial

$$F(z) = z^{D+1} - z^D + \Delta \lambda_j \tag{18}$$

lie within the unit circle.

STABILITY AND CONVERGENCE CRITERIA

The particular form of  $F(z)$  allows us to determine the location of its roots. Define

$$\beta = \Delta \lambda_j. \tag{19}$$

The values of  $\beta$  for which  $F(z)$  has roots on the unit circle can be determined by setting

$$z = e^{j\phi}. \tag{20}$$

Since  $\beta$  is real, from (18),

$$\beta = \cos D\phi - \cos (D+1)\phi \tag{21}$$

$$\sin D\phi - \sin (D+1)\phi = 0. \tag{22}$$

The values of  $\phi$  satisfying the latter equation define the points at which roots can lie on the unit circle, namely

$$\phi = \frac{l\pi}{2D+1} \tag{23}$$

where  $l$  is odd and  $|l| \leq 2D+1$  or  $l=0$ .

The values of  $\beta$  corresponding to  $F(z)$  having roots on the unit circle can be determined from (21) and (23),

$$\beta = 2 \sin \frac{l\pi}{2(2D+1)}. \tag{24}$$

The overall behavior of the system can be ascertained by using root locus techniques [10]. For  $\beta = 0$ ,  $D$  roots lie at the origin and one root lies at  $z = 1$ . As  $\beta$  is increased, the root at  $z = 1$  and one of the other roots move toward each other on the real axis to meet at  $z = D/(D+1)$ . The other roots move radially from the origin. For larger values of  $\beta$ , the two real axis roots split to form a complex conjugate pair which cross the unit circle at

$$\phi = \pm \frac{\pi}{2D+1}. \tag{25}$$

The values of  $\beta$  for which all roots lie inside the unit circle correspond to the interval  $(0, \beta_{max})$  where  $\beta_{max}$  corresponds to  $l = 1$  in (24). In terms of  $\Delta$  and  $\lambda_j$  this interval is

$$0 < \Delta < \frac{2}{\lambda_j} \sin \frac{\pi}{2(2D+1)}. \tag{26}$$

The step size  $\Delta$  must satisfy (26) for each  $\lambda_j$ . Thus, the system is stable and converges if and only if

$$0 < \Delta < \frac{2}{\lambda_{max}} \sin \frac{\pi}{2(2D+1)}. \tag{27}$$

where  $\lambda_{\max}$  is the largest eigenvalue of the correlation matrix  $A$ .

### CONVERGENCE OF A PRACTICAL ALGORITHM

In practice, the weight adjustment algorithm of (5) must be modified. Since a large observation time is necessary in order to get a good estimate of the gradient vector, the algorithm is modified to become

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \Delta_k e_{k-D} \mathbf{r}^{(k-D)}. \quad (28)$$

In the stochastic version of the algorithm, a noisy unbiased estimate replaces the gradient. For this algorithm, it will be shown that (27) also guarantees that

$$\lim_{k \rightarrow \infty} E[\mathbf{c}^{(k)}] = \mathbf{c}_{\text{opt}} \quad (29)$$

if the vector of weights  $\mathbf{c}^{(k)}$  (now a random vector) is independent of the received vector  $\mathbf{r}^{(k)}$ . The independence assumption has been justified by Mazo [11] using a perturbation solution. From (28), taking expected values and fixing  $\Delta_k = \Delta$ ,

$$\begin{aligned} E[\mathbf{c}^{(k+1)}] &= E[\mathbf{c}^{(k)}] + \Delta E[e_{k-D} \mathbf{r}^{(k-D)}] \\ &= E[\mathbf{c}^{(k)}] + \Delta(AE[\mathbf{c}^{(k-D)}] - \alpha). \end{aligned} \quad (30)$$

This set of equations can be decoupled as before to give an equation of the same form as (12).

Thus, applying the same reasoning, the weight vector converges in the mean to the optimum value if and only if the step size  $\Delta$  satisfies (27).

### DISCUSSION

It has been shown that (27) gives the necessary and sufficient conditions to ensure convergence of the mean tap weights. This condition reduces to that given by Widrow [10] for no delay. The delay in the update procedure can be seen to reduce the maximum allowable step size by a factor  $\sin(\pi/(4D + 2))$ .

In the models depicted in Figs. 1 and 2, the delay has been lumped at one point. However, the characteristic polynomial for the system is not affected if the delay is moved to another point or even distributed in the adjustment loop. Thus, the same analysis applies if, for instance, the equalizer is remote from the decision unit (resulting in delay both in the decision path and the update path).

The convergence condition shows the tradeoff between convergence rate and stability even with delayed adjustments. As the loop gain or delay is varied, the root loci will give insight into the system behavior. For very small step sizes the system has an exponential response with time constant

$$\tau = \ln \frac{1}{1 - \sigma} \quad (31)$$

where  $\sigma$  is the position of the real axis root nearest the unit circle, satisfying

$$\Delta \lambda_{\max} = \sigma^D (1 - \sigma). \quad (32)$$

This exponential behavior is valid for  $\Delta \lambda_{\max} \ll 1/(D + 1)$ . For

larger values of  $\sigma$ , the response becomes oscillatory as more than one root begins to affect the response.

Since oscillatory behavior is usually undesirable and because of the uncertainty in estimating  $\lambda_{\max}$ , the step size parameter will generally be small enough to give an exponential convergence. A small step size also assures that the fluctuations about the optimal values will be small.

The analysis has examined the mean tap weights. The real performance measure is mean square error. Ungerboeck's work [12] indicates that (for no delay) the same paths closely follow the mean values if the step-size parameter is reduced from the maximum value given here by a factor  $M$ .

Other algorithms have been suggested that vary the step size parameter  $\Delta_k$  dynamically to speed convergence (e.g., [13], [12]). These schemes attempt to achieve rapid convergence (by using a relatively large initial step size) and a small excess error due to tap weight fluctuations around the optimal values (by using a small final step size). For these schemes, the given stability criterion remains a sufficient condition if it is applied at each time instant to  $\Delta_k$ .

In some circumstances, the adjustment delay may be reduced to allow for stability along with a reasonable rate of convergence. For instance, in the Viterbi detector, premature decisions are available. These, although they contain more errors than the final decision, might be suitable for equalizer adjustment. In the case of other detectors, an auxiliary detector with no delay or less delay might be used expressly to adjust the equalizer. The stability and convergence limits presented here allow one to evaluate such alternatives.

### REFERENCES

- [1] B. Widrow, "Adaptive filters, I: Fundamentals," Stanford Electron. Lab., Stanford Univ., Stanford, CA, Tech. Rep. 6764-6, Dec. 1966; see also "Adaptive filters," in *Aspects of Network and System Theory, Part IV*, R. E. Kalman and N. DeClaris, Eds. New York: Holt, Rinehart and Winston, 1971, pp. 563-587.
- [2] H. Kobayashi, "Correlative level coding and maximum likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 586-594, Sept. 1971.
- [3] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363-378, May 1972.
- [4] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, pp. 779-785, May 1970.
- [5] S. U. H. Qureshi and E. E. Newhall, "An adaptive receiver for data transmission over time-dispersive channels," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 448-457, July 1973.
- [6] D. D. Falconer and F. R. Magee, Jr., "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, vol. 52, pp. 1541-1562, Nov. 1973.
- [7] A. Cantoni and K. Kwong, "Further results on the Viterbi algorithm equalizers," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 764-767, Nov. 1974.
- [8] L. A. Zadeh and C. A. Desoer, *Linear Systems Theory—The State Space Approach*. New York: McGraw-Hill, 1963.
- [9] E. I. Jury, *Theory and Application of the z-Transform Method*. Huntington, NY: Krieger, 1973.
- [10] B. C. Kuo, *Automatic Control Systems*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1967.
- [11] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963-993, May-June 1979.
- [12] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communication," *IBM J. Res. Develop.*, vol. 16, pp. 546-555, Nov. 1972.
- [13] T. J. Schonfeld and M. Schwartz, "A rapidly converging first-order training algorithm for an adaptive equalizer," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 431-439, July 1971.