Adaptive Equalization of Highly Dispersive Channels for Data Transmission

By ALLEN GERSHO

(Manuscript received May 29, 1968)

This paper analyzes an adaptive training algorithm for adjusting the tap weights of a tapped delay line filter to minimize mean-square intersymbol interference for synchronous data transmission. The significant feature of the adjustment procedure is that convergence is guaranteed for all channel response pulses, even for very severe amplitude and phase distortion.

The author examines convergence, rate of convergence, and the effect of noisy observations of the received pulses, and he shows that the noisy observations result in a random sequence of tap weight settings whose mean value converges to a suboptimal setting. The mean-square deviation of the tap weights from the suboptimal values is asymptotically bounded with a bound that can be made as small as desired by sufficiently reducing the speed of convergence.

The suboptimality arising here results from the use of isolated test pulses for the training signal. However, a training scheme using pseudorandom sequences or the actual data signal does not suffer from the suboptimality effect. Hence, although of possible utility in other pulse shaping applications, the technique presented here appears to be primarily of value in providing a conceptual framework for the closely related but more practical techniques to be examined in the sequel to this paper to be published shortly.

I. INTRODUCTION

A common approach to data transmission is to code the amplitudes of successive pulses in a periodic pulse train with a discrete set of possible amplitude levels. The coded pulse train is then linearly modulated, transmitted through the channel, demodulated, equalized, and synchronously sampled and quantized. As a result of dispersion of the pulse shape by the channel, the number of detectable amplitude levels has very often been limited by intersymbol interference rather than by additive noise.

In principle, if the channel is known precisely it is virtually always possible to design an equalizer that will make the intersymbol interference (at the sampling instants) arbitrarily small. However, in practice a channel is random in the sense of being one of an ensemble of possible channels. Consequently, a fixed equalizer designed on average channel characteristics may not adequately reduce intersymbol interference. An adaptive equalizer is then needed which can be "trained," with the guidance of a suitable training signal transmitted through the channel, to adjust its parameters to optimal values. If the channel is also time-varying, an adaptive equalizer operating in a tracking mode is needed which can update its parameter values by tracking the changing channel characteristics during the course of normal data transmission. In both cases the adaptation may be achieved by observing or estimating the error between actual and desired equalizer responses and using this error to estimate the direction in which the parameters should be changed to approach the optimal values.

A simple and effective technique for adaptive equalization was developed by Lucky using the tapped delay line filter structure for the equalizer.[1, 2] The main limitation of this technique is that convergence of the tap weight adjustment algorithm is assured only for relatively low dispersion channels. The convergence condition requires that the dispersed pulse shape have adequate quality so that, in the absence of noise, error-free binary data transmission would be possible without equalization. In other words the dispersed pulse must have an open binary "eye."

Using an approach to adaptation[3, 4] with virtually unrestricted convergence properties, Lucky and Rudin subsequently proposed and implemented an adaptive equalizer for minimizing the mean square error in frequency response of an analog channel.[5, 6] This approach was applied to synchronous data transmission by the author and independently by Lytle and by Niessen.[7–9] An implementation of the technique was described by Niessen and Drouilhet.[10] It has also been implemented for data communication at Bell Laboratories.

In this paper the approach is used for synchronous data transmission in a training mode where a sequence of isolated pulses is used as a test signal. The technique may be viewed equally as an adaptive design procedure for a sampled-data pulse shaping filter where the error criterion is to minimize the mean square error between actual and desired pulse shapes at the filter output. The important feature of the technique is that convergence is achieved for any channel pulse response whatever, thereby including highly dispersed pulses for which even binary data transmission would be impossible without equalization. Of particular interest are: (*i*) the analogous optimality condition to Lucky's zero forcing condition resulting with the change from a summed absolute error to a summed squared error criterion,[1] (*ii*) the manner in which noisy observations introduce randomness in the iterative corrections to the weights and the resulting stochastic convergence properties, (*iii*) the possibility of applying the technique where isolated pulses applied to a filter must be used to adaptively adjust the filter for optimum pulse shaping (unrelated to equalization), and (*iv*) the conceptual framework for the more practical adaptation techniques to be described in a sequel to this paper, planned for publication soon.

Perhaps the earliest application of the tapped delay line or "transversal" filter to pulse shaping for data transmission was made by W. P. Boothroyd and E. M. Creamer.[11] Tufts and George have shown that under a mean-square error criterion the optimal receiver structure includes a tapped-delay line filter with delay between taps equal to the symbol period.[12, 13] Aaron and Tufts have also shown that the same receiver structure is needed to minimize the average error probability for binary data transmission.[14]

The basic approach to adaptive adjustment of a set of weights where a mean-square error criterion is used with a gradient search procedure was considered by Widrow and Hoff who noticed that no derivative computation is needed.[3] Narendra and McBride proposed a self-optimizing Wiener filter using a continuous-time gradient algorithm and a filter structure whose transfer function is a weighted sum of fixed functions.[4] Koford and Groner used a mean-square error criterion and a gradient learning algorithm to find an optimum set of weights for pattern classifying.[15] Widrow described a general adaptive filtering problem with the tapped delay line filter.[16] Coll and George discussed the performance of George's optimum equalizer and indicated a possible adaptive adjustment technique.[17] Lucky and Rudin were the first to apply the mean square error criterion with the gradient search procedure to the field of adaptive equalization.[5, 6] This paper expands on a short presentation given at an international symposium on information theory.[18]

II. PERFORMANCE OBJECTIVES FOR EQUALIZATION

The objective of equalization, viewed as a pulse shaping problem, is to adjust the parameters of the equalizer to a setting which minimizes a suitable measure of the error between actual and desired pulse shapes. For the usual synchronous data transmission application, the desired pulse shape is one with the Nyquist property that the sample values y_k at the sampling instant kT are given by $y_k = \delta_{kr}$ where δ_{kr} is unity for $k = r$ and zero for all other integers k. The criterion used by Lucky[1] is peak distortion, D, given by

$$D = \sum_{k \neq r} | y_k |/| y_r |.$$

An alternate criterion of interest is the mean square distortion E, defined by

$$E = \sum_{k \neq r} y_k^2/y_r^2 .$$

The physical interpretation of the peak distortion is that it is directly related to eye opening and determines the error probability for a worst case message pattern. The mean square distortion has a different interpretation. If the message pattern is such that the transmitted level for each time slot is statistically independent of the levels for other time slots, then the variance of the intersymbol interference in a given time slot is proportional to the mean square distortion. If the pulse shape has a large number of small sidelobes so that the intersymbol interference is normally distributed, then minimizing mean square distortion is equivalent to minimizing error rate.

Closely related to the mean square distortion is the mean square error

$$\mathcal{E} = \sum_k (y_k - d_k)^2 \tag{1}$$

where d_k is the desired pulse sample value at time instant kT. For the usual equalization problem where $d_k = \delta_{kr}$, the measure $\mathcal{E}$ has virtually the same interpretation as E; however, E is a normalized measure independent of pulse amplitude while $\mathcal{E}$ depends on both shape and amplitude. Optimization of the tapped delay line equalizer with respect to either criterion leads to equivalent results.

III. FORMULATION

Consider the transversal equalizer with N taps and tap spacing T equal to the symbol period. Let c_k be the weight at the kth tap for

$k = 0, 1, \cdots, N-1$ so that the input output relation of the transversal filter at the sample times is

$$y_n = \sum_{k=0}^{N-1} c_k x_{n-k} = \mathbf{c}'\mathbf{x}_n \tag{2}$$

where x_k and y_k denote the input and output pulse samples, respectively, at time instants kT, $\mathbf{c} = (c_0, c_1, \cdots, c_{N-1})$ is the tap weight vector, and $\mathbf{x}_n = (x_n, x_{n-1}, \cdots, x_{n-N+1})$ is the sample memory state of the delay line at the time instant nT; the vectors $\mathbf{c}$ and $\mathbf{x}_n$ are to be regarded as column matrices, and the prime denotes the transpose. We assume that the input sequence x_k has finite energy. Let $\epsilon_n = y_n - d_n$. Then from equation (1), using (2), the gradient of the error with respect to $\mathbf{c}$ may be written as

$$\nabla \mathcal{E} = 2 \sum_k \epsilon_k \mathbf{x}_k . \tag{3}$$

Therefore the optimality condition for minimum error $\nabla \mathcal{E} = 0$ is equivalent to the requirement that the (deterministic) corss-correlation between the input sequence x_k and output error sequence ϵ_k must have zeros for the N components with index values corresponding to the index values of the available tap weights. That is,

$$\varphi_{x\epsilon}(k) = \sum_n \epsilon_n x_{n-k} = 0 \qquad \text{for} \quad k = 0, 1, 2, \cdots, N-1.$$

This condition has an interesting similarity to Lucky's condition which states that the peak distortion, D, is minimized when the error sequence ϵ_n has zeros for the N components with index values corresponding to the index values of the available tap weights.[1] An important distinction is that Lucky's condition is generally not valid when the input pulse distortion D exceeds unity, while the mean square optimizing condition is valid for any input pulse with finite mean square distortion.

Using equation (2), the gradient (3) can be expressed explicitly as a function of the tap weight vector $\mathbf{c}$, namely:

$$\nabla \mathcal{E} = 2(\mathbf{A}\mathbf{c} - \mathbf{g}) \tag{4}$$

where

$$\mathbf{A} = \sum_n \mathbf{x}_n \mathbf{x}_n' , \qquad \text{and} \qquad \mathbf{g} = \sum_n d_n \mathbf{x}_n .$$

Notice that $\mathbf{A}$ is symmetric and positive definite (see Appendix A). Setting equation (4) equal to zero yields the solution for the optimum tap weight vector $\mathbf{c}^*$,

$$\mathbf{c}^* = \mathbf{A}^{-1}\mathbf{g}.$$

Using equation (2), the error expression given by equation (1) may be expressed in the convenient form:

$$\mathcal{E}(\mathbf{c}) = \mathcal{E}(\mathbf{c}^*) + (\mathbf{c} - \mathbf{c}^*)'\mathbf{A}(\mathbf{c} - \mathbf{c}^*) \tag{5}$$

which shows explicitly the simple quadratic nature of the error surface and the unique optimality of the minimizing weight vector $\mathbf{c}^*$. It can be shown that the residual error $\mathcal{E}(\mathbf{c}^*)$ can be made as small as desired for all channels of practical interest by using a sufficiently large number, N, of taps.[19]

It is intuitively reasonable that successive corrections to the tap weight vector in the direction of steepest descent of the error surface should lead to the minimum error where $\mathbf{c} = \mathbf{c}^*$. This is the idea of the well-known[20] gradient algorithm:

$$\mathbf{c}_{i+1} = \mathbf{c}_i - \tfrac{1}{2}\alpha\nabla\mathcal{E}(\mathbf{c}_i) \tag{6}$$

where α is a suitably small positive proportionality constant, $\mathbf{c}_0$ is arbitrary, and $\mathbf{c}_i$ is the tap weight vector after the ith iteration.

The significant feature of the gradient algorithm for our quadratic error surface (5) is that the gradient can be conveniently evaluated without knowledge of the error surface itself. We have seen from equation (3) that the components of the gradient vector are values of the crosscorrelation between the input sequence and the output error sequence. This suggests the conceptually simple implementation where an isolated test pulse is transmitted through the channel and the requisite crosscorrelation values are formed by multiplying the delayed input pulse with the error pulse, sampling, and summing (or averaging). The tap weights are then incremented according to (6), the old crosscorrelation values "dumped" and a new iteration is begun with the transmission of a new test pulse.

The error pulse is formed by subtracting from the equalizer output pulse an "ideal" pulse whose sample values are the desired values d_k; the ideal pulse is locally generated at the appropriate time. The basic scheme is shown in Fig. 1. Naturally, the summation given by equation (3) cannot be performed over an infinite time interval. Suppose κT is a practical upper bound on the possible time duration of the input pulse, ξT is the time interval between successive test pulses with $\xi T > \kappa T$, ξ and κ as positive integers. Then if we include the effect of perturbing

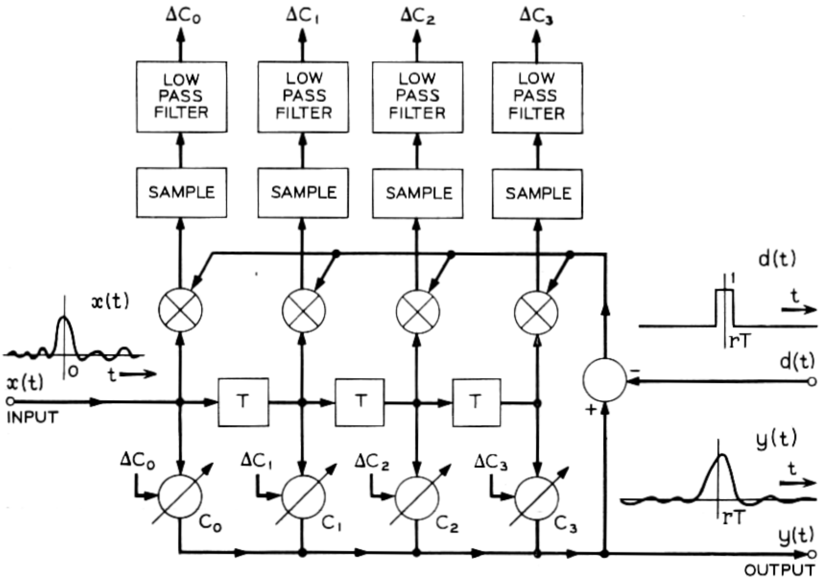

Fig. 1 — Four tap training mode adaptive equalizer.

receiver noise samples n_i and z_i at the equalizer input and output, respectively, the measured crosscorrelation vector $\hat{\boldsymbol{\varphi}}_i$ after the ith iteration is given by:

$$\hat{\boldsymbol{\varphi}}_i = \sum_{l=l_0+i\xi}^{l_0+i\xi+\kappa} (\mathbf{x}_{l-i\xi} + \mathbf{n}_l)(\epsilon_{l-i\xi} + z_l). \tag{7}$$

In the noiseless case the estimate $\hat{\boldsymbol{\varphi}}_i$ reduces to one-half the deterministic gradient, that is, $\frac{1}{2}\nabla\mathcal{E}(\mathbf{c}_i)$ under the assumption that the pulse sequence x_l and desired sequence d_l are virtually zero outside of the interval $l_0 \leqq l \leqq l_0 + \kappa - N + 1$.

IV. CONVERGENCE PROPERTIES

In the presence of noise the tap weight corrections contain undesired random components consisting of products of input and output noise samples and products of pulse and noise sample. As a result, the random tap weights no longer converge to the optimal values but instead approach some neighborhood of a suboptimal setting and then fluctuate randomly about this setting. The error between the optimal and suboptimal settings is small for low noise levels and decreases with increasing signal-to-noise ratios. The size of the fluctuation neighborhood about the suboptimal setting is proportional to the noise level but can be made as small as desired by making the training time sufficiently long.

Assume the noise samples n_i have zero mean and finite variance σ^2. Define the vector $\mathbf{n}_k = (n_k, n_{k-1}, \cdots, n_{k-N+1})$ to be regarded as a column matrix. Then the output noise samples of the equalizer are:

$$z_k = \mathbf{c}'\mathbf{n}_k . \tag{8}$$

Define the matrix $\mathbf{B} = E(\mathbf{n}_k\mathbf{n}_k')$, where $E(\cdots)$ denotes the expected value. Notice that $\mathbf{B}$ is symmetric and positive semidefinite.

To formulate the iterative equations describing the tap weight behavior in the presence of noise, apply equations (2) and (8) to (7) to show how the gradient estimate depends on the tap weight vector:

$$\hat{\boldsymbol{\varphi}}_i = \sum_l (\mathbf{x}_{l-i\xi} + \mathbf{n}_l)[(\mathbf{x}_{l-i\xi} + \mathbf{n}_l)'\mathbf{c}_i - d_{l-i\xi}].$$

Hence

$$\hat{\boldsymbol{\varphi}}_i = \mathbf{H}_i\mathbf{c}_i - \mathbf{g} - \mathbf{v}_i , \tag{9}$$

where $\mathbf{H}_i$ is the random symmetric matrix

$$\mathbf{H}_k = \sum_l (\mathbf{x}_{l-i\xi} + \mathbf{n}_l)(\mathbf{x}_{l-i\xi} + \mathbf{n}_l)' \tag{10}$$

and

$$\mathbf{v}_i = \sum_l \mathbf{n}_l \, d_{l-i\xi} . \tag{11}$$

Let $\mathcal{A} = E(\mathbf{H}_i)$, the expected value of $\mathbf{H}_i$. Then equation (10) yields

$$\mathcal{A} = \mathbf{A} + \kappa\mathbf{B}. \tag{12}$$

which is positive definite since $\mathbf{A}$ is positive definite and $\mathbf{B}$ is positive semidefinite.

It is convenient to examine the random variation of the tap weight vector c_k about the suboptimal setting defined by

$$\tilde{\mathbf{c}} = \mathcal{A}^{-1}\mathbf{g}, \tag{13}$$

and let $\mathbf{q}_i = \mathbf{c}_i - \tilde{\mathbf{c}}$. From equation (12) it is evident that the suboptimal setting $\tilde{\mathbf{c}}$ approaches the optimal setting $\mathbf{c}^*$ as the ratio of noise variance to input pulse sequence energy approaches zero. The iterative algorithm may be expressed in the form

$$\mathbf{q}_{i+1} = \mathbf{q}_i - \alpha\hat{\boldsymbol{\varphi}}_i , \tag{14}$$

$$\hat{\varphi}_i = \mathbf{H}_i\mathbf{q}_i + \mathbf{h}_i \tag{15}$$

where

$$\mathbf{h}_i = \mathbf{H}_i\tilde{\mathbf{c}} - \mathbf{g} - \mathbf{v}_i \,. \tag{16}$$

Equations (14) and (15) constitute a system of first-order stochastic difference equations with a forcing function $\mathbf{h}_i$ which is statistically dependent on the stochastic state matrix $\mathbf{H}_i$. We assume that the perturbing noise samples in different iterations are uncorrelated, so that $\mathbf{H}_i$ and $\mathbf{h}_i$ are independent of $\mathbf{H}_j$ and $\mathbf{h}_j$ for $i \neq j$. Notice that the expected value of any function of $\mathbf{H}_i$ and $\mathbf{h}_i$ is independent of i. Under these conditions it is proved in Appendix C that for suitably small values of α the mean value of the solution vector $\mathbf{q}_i$ approaches zero as $i \rightarrow \infty$ and the sum of the variances of the components of $\mathbf{q}_i$ is bounded with a bound that approaches zero as α approaches zero. Consequently the mean value of the tap weight vector converges to the suboptimal setting $\tilde{\mathbf{c}}$ while the actual tap weights fluctuate randomly about the converging mean values with a variability that can be made arbitrarily small.

Notice from Appendix C that the norm of the mean solution vector $\langle \mathbf{q} \rangle_i$ is reduced at least by the factor ζ, the spectral norm[20] of $I - \alpha\mathfrak{A}$. Let ρ_1 and ρ_N denote the minimum and maximum eigenvalues, respectively, of $\mathfrak{A}$. Then

$$\zeta = \min \mid 1 - \alpha\rho_1 \mid, \quad \mid 1 - \alpha\rho_N \mid . \tag{17}$$

(For proof see p. 24 of Ref. 20.)

Then for $0 < \alpha < 2/(\rho_1 + \rho_N)$, we obtain $\zeta = 1 - \alpha\rho_1$. Consequently, while decreasing α offers a smaller bound on variability of the tap weight vector, increasing α assures a stronger bound on convergence rate. For the training mode it is likely that speed of adaptation will be relatively unimportant so that a very small value of α could be used to approach a tap weight setting that is very close to the suboptimal setting.

It is useful to obtain bounds on the eigenvalues of $\mathfrak{A}$ which can be determined without specific knowledge of the channel characteristics. If $x(t)$ denotes the channel pulse response and $n(t)$ the additive receiver noise so that the sampled values used earlier are given by $x_k = x(kT)$ and $n_k = n(kT)$, then the sampled spectrum $X^*(\omega)$ of x_k is

$$X^*(\omega) = \sum_k x_k e^{-i\omega kT} = \frac{1}{T} \sum_k X(\omega - 2\pi/T)$$

and the sampled spectral density $S^*(\omega)$ of n_k is

$$S^*(\omega) = \sum_k E(n_i n_{i+k}) e^{-i\omega kT} = \frac{1}{T} \sum_k S(\omega - 2\pi/T)$$

where $X(\omega)$ is the Fourier transform of $x(t)$ and $S(\omega)$ is the spectral density of $n(t)$. Let m and M denote the infimum and supremum, respectively, of $| X^*(\omega) |^2 + \kappa S^*(\omega)$ so that

$$m \leqq | X^*(\omega) |^2 + \kappa S^*(\omega) \leqq M. \tag{18}$$

In all cases of practical interest M will be finite; furthermore generally m will be greater than zero. It is shown in Appendix B that each eigenvalue, ρ_i, of $\mathbf{\alpha}$ will be bounded according to

$$m \leqq \rho_i \leqq M. \tag{19}$$

To illustrate the use of this bound, notice from Appendix C that the condition for convergence of the mean tap weight vector to the suboptimal solution is that $\alpha < 2/\rho_N$. Thus a sufficient condition is that

$$\alpha < 2/M. \tag{20}$$

Furthermore, the mean tap setting converges exponentially with the convergence factor ζ, given by equation (17). Hence it can be inferred that the choice of α which provides the strongest bound (least value of ζ) is $\alpha = 2/(\rho_1 + \rho_N)$ yielding

$$\zeta = \frac{p - 1}{p + 1}$$

where $p = \rho_N/\rho_1$. Using the bounds given in (19) we obtain $p \leqq M/m$, and so

$$\zeta = \frac{M - m}{M + m}. \tag{21}$$

Therefore, for the best choice of α, convergence of the mean proceeds at least at a rate given by the geometric factor $(M - m)/(M + m)$. Thus useful information regarding the convergence speed can be determined without knowledge of the channel characteristics.

V. CONCLUSION

The degree of suboptimality of the tap weight setting reached by the training algorithm may or may not be consequential, depending on the application. In applications where multilevel pulse transmission with a large number of levels could be achieved with adequate equalization, the signal-to-noise ratio is necessarily very high and therefore the degree of suboptimality is not large. Even when the noise level is fairly substantial the suboptimal setting may still be adequate if the error surface given by $\mathcal{E}(\mathbf{c})$ is "shallow" in a large neighborhood of the minimum. Then a fairly large departure of $\tilde{\mathbf{c}}$ from $\mathbf{c}^*$ may correspond to a relatively small increase in mean-square error. Also, if training mode adaptation is used as a prelude to tracking mode adaptation, a fairly large degree of suboptimality may be a tolerable starting point for a tracking mode operation such as the one we plan to describe in a future paper.

When the noise level is substantial the criterion for optimality used here becomes inadequate because it does not consider the effect of the equalizer on the receiver noise. The price of reducing intersymbol interference may be a sizable increase in noise level at the equalizer output. In our future paper the error criterion is modified to include noise with the result that the problem of suboptimality does not arise.

The random fluctuation of the tap weights which prevents true convergence to the suboptimal setting can be eliminated by reducing the proportionality constant α in each iteration using a sequence of step sizes α_k with the properties

$$\sum \alpha_k = \infty \qquad \text{and} \qquad \sum \alpha_k^2 < \infty .$$

It may then be shown that the tap weight vector converges to the suboptimal solution with probability 1. The proof uses stochastic approximation theory and follows the lines taken by Tong and Liu who considered a training mode algorithm for low dispersion channels.[21] However, this modification complicates the implementation somewhat and cannot be applied to the tracking mode adaptation problem.

APPENDIX A

Proof that **A** *is Positive Definite*

The matrix **A** is defined by

$$\mathbf{A} = \sum_{k=-\infty}^{\infty} \mathbf{x}_k \mathbf{x}_k' . \tag{22}$$

Consequently

$$\mathbf{c}'\mathbf{A}\mathbf{c} = \sum_{-\infty}^{\infty} \mathbf{c}'\mathbf{x}_k\mathbf{x}_k'\mathbf{c} = \sum_{-\infty}^{\infty} y_k^2 .$$

But the sequence y_k is the convolution of the x_k sequence with the finite tap weight sequence c_k . Hence, using Parseval's equality,

$$\mathbf{c}'\mathbf{A}\mathbf{c} = \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} | X^*(\omega) |^2 \, | C(\omega) |^2 \, d\omega, \tag{23}$$

where

$$C(\omega) = \sum_{k=0}^{N-1} c_k e^{-jk\omega T}.$$

Equation (23) shows immediately that $\mathbf{c}'\mathbf{A}\mathbf{c}$ is nonnegative for all vectors $\mathbf{c}$. Also, $C(\omega)$ can have only isolated zeros and $| X^*(\omega) |$ is square integrable since the input pulse has finite mean square distortion. It may then be inferred that $\mathbf{c}'\mathbf{A}\mathbf{c} > 0$ unless $\mathbf{c} = 0$, which proves that $\mathbf{A}$ is positive definite.

APPENDIX B

Bounds on the Eigenvalues of $\mathfrak{A}$

Since $\mathbf{B} = E(\mathbf{n}_k\mathbf{n}_k')$ the quadratic form $\mathbf{c}'\mathbf{B}\mathbf{c}$ is the mean squared value of $\mathbf{y}_k^2$ of the response of the equalizer with weight vector $\mathbf{c}$ to the input noise n_t . Consequently

$$\mathbf{c}'\mathbf{B}\mathbf{c} = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} S^*(\omega) \mid C(\omega) \mid^2 d\omega. \tag{24}$$

Combining equations (23) and (24) yields

$$\mathbf{c}'\mathfrak{A}\mathbf{c} = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} \{\mid X^*(\omega) \mid^2 + \kappa S^*(\omega)\} \mid C(\omega) \mid^2 d\omega. \tag{25}$$

Applying to equation (25) the bounds m and M given by equation (18) yields

$$m\, \mathbf{c}'\mathbf{c} \leqq \mathbf{c}'\mathfrak{A}\mathbf{c} \leqq M\mathbf{c}'\mathbf{c}. \tag{26}$$

Let $\mathbf{c}$ be the eigenvector of $\mathfrak{A}$ corresponding to eigenvalue ρ. Then $\mathfrak{A}\mathbf{c} = \rho_i\mathbf{c}$ and equation (26) yields

$$m \leqq \rho_i \leqq M \tag{27}$$

which provides a convenient bound for the largest and smallest eigenvalues of $\mathfrak{A}$.

APPENDIX C

Convergence Proof

To examine the convergence properties of the tap weight adjustment algorithm, it is convenient to define the norm of a random vector $\mathbf{u}$ as

$$\| \mathbf{u} \| = [E(\mathbf{u}'\mathbf{u})]^{1/2}, \tag{28}$$

so that the squared norm of $\mathbf{u}$ is the sum of the second moments of the components of $\mathbf{u}$. For a deterministic vector the norm reduces to the usual Euclidian norm. The norm of a deterministic matrix will denote the usual spectral norm.[20]

Theorem: Let $\mathbf{H}_k$ *be a sequence of random symmetric* $N \times N$ *matrices and* $\mathbf{h}_k$ *a sequence of random N-tuple column vectors. Suppose* $\mathbf{H}_k$ *and* $\mathbf{h}_k$ *are stationary in k with* $\mathbf{H}_k$ *and* $\mathbf{h}_k$ *independent of* $\mathbf{H}_j$ *and* $\mathbf{h}_j$ *for* $k \neq j$*. Assume* $\mathbf{h}_k$ *has zero mean, and the elements of* $\mathbf{H}_k$ *and* $\mathbf{h}_k$ *have finite variance,* $E\mathbf{H}_k = \mathcal{A}$*, independent of k with* $\mathcal{A}$ *positive definite. Define the random vector sequence* $\mathbf{q}_k$ *according to:*

$$\mathbf{q}_{k+1} = \mathbf{q}_k - \alpha \boldsymbol{\varphi}_k \tag{29}$$

where

$$\boldsymbol{\varphi}_k = \mathbf{H}_k \mathbf{q}_k + \mathbf{h}_k \tag{30}$$

for $k = 0, 1, 2, \cdots$ *and* $\mathbf{q}_0$ *is an arbitrary deterministic vector. Then for* α *positive and sufficiently small,*

$$\lim_{k \to \infty} \| E\, \mathbf{q}_k \| = 0 \tag{31a}$$

and

$$\limsup_{k \to \infty} \| \mathbf{q}_k \| \leqq V(\alpha) \tag{31b}$$

with $V(\alpha)$*, given in* (47), *satisfying:*

$$\lim_{\alpha \to 0} V(\alpha) = 0. \tag{32}$$

Proof: Combining equations (29) and (30) yields

$$\mathbf{q}_{k+1} = (\mathbf{I} - \alpha \mathbf{H}_k)\mathbf{q}_k - \alpha \mathbf{h}_k\,. \tag{33}$$

Noting that $\mathbf{q}_k$ is independent of $\mathbf{H}_k$, taking the expected value in equation (33), we find

$$E(\mathbf{q}_{k+1}) = (\mathbf{I} - \alpha \mathcal{A}) E(\mathbf{q}_k). \tag{34}$$

It follows then that

$$\| E(\mathbf{q}_k) \| \leqq \zeta^k \| E\, \mathbf{q}_0 \| \tag{35}$$

where

$$\zeta = \| I - \alpha \mathcal{A} \|\,. \tag{36}$$

Hence equation (31a) follows when $\zeta < 1$, or equivalently, for

$$0 < \alpha < 2/\rho_N \tag{37}$$

where ρ_N is the largest eigenvalue of $\mathcal{A}$.

To prove equation (31b), observe that

$$E(\mathbf{q}'_{k+1}\mathbf{q}_{k+1}) = E[\mathbf{q}'_k(I - \alpha H_k)^2\mathbf{q}_k] - E[2\alpha\mathbf{q}'_k(I - \alpha H_k)\mathbf{h}_k] + \alpha^2 \,||\, \mathbf{h}_k \,||^2 \tag{38}$$

from equation (33). Noting again that $\mathbf{q}_k$ is independent of H_k, we have

$$E[\mathbf{q}'_k(I - \alpha\mathbf{H}_k)^2\mathbf{q}_k] = E\{\mathbf{q}'_k E[(I - \alpha\mathbf{H}_k)^2]\mathbf{q}_k\} \leqq \mu \,||\, \mathbf{q}_k \,||^2, \tag{39}$$

where

$$\mu = ||\, E[(I - \alpha\mathbf{H}_k)^2] \,|| . \tag{40}$$

Also, using the Schwarz inequality,

$$E[-\mathbf{q}'_k(I - \alpha\mathbf{H}_k)\mathbf{h}_k] = \alpha\mathbf{q}'_k E(\mathbf{H}_k\mathbf{h}_k) \leqq \alpha \,||\, \mathbf{q}'_k \,||\, f$$

where $f = ||\, E(\mathbf{H}_k\mathbf{h}_k) \,||$. Using equation (35) we obtain

$$-E[\mathbf{q}_k(I - \alpha\mathbf{H}_k)\mathbf{h}_k] \leqq \alpha\zeta^k \,||\, E(\mathbf{q}_0) \,||\, f. \tag{41}$$

The bounds (39) and (41) may be applied to equation (38), yielding

$$||\, \mathbf{q}_{k+1} \,||^2 \leqq \mu \,||\, \mathbf{q}_k \,||^2 + \alpha^2 f \,||\, E\, \mathbf{q}_0 \,||\, \zeta^k + \alpha^2 \,||\, \mathbf{h}_k \,||^2 . \tag{42}$$

If we now define the bounding sequence of positive numbers Q_k according to

$$Q_0 = ||\, E\, \mathbf{q}_0 \,||^2$$

and

$$Q_{k+1} = \mu Q_k + \alpha^2 f \,||\, E(\mathbf{q}_0) \,||\, \zeta^k + \alpha^2 \,||\mathbf{h}_k \,||^2 , \tag{43}$$

then it follows from (42) that

$$||\, \mathbf{q}_k \,||^2 \leqq Q_k .$$

But the difference equation given by (43) has the asymptotic solution

$$\lim_{k\to\infty} Q_k = \frac{\alpha^2 \,||\, \mathbf{h}_k \,||^2}{1 - \mu} ,$$

for $\zeta < 1$ and $\mu < 1$. Then

$$\limsup_{k\to\infty} \,||\, q_k \,||^2 \leqq \frac{\alpha^2 \,||\, \mathbf{h}_k \,||^2}{1 - \mu}. \tag{44}$$

Notice that $\| h_k \|$ is independent of k by the hypothesis of stationarity.

Since

$$(I - \alpha \mathbf{H}_k)^2 = (I - \alpha \mathbf{Q})^2 + \alpha^2 E(\mathbf{G}_k^2).$$

where

$$\mathbf{G}_k = \mathbf{H}_k - \mathbf{Q}, \tag{45}$$

we find that

$$\begin{aligned} \mu &\leqq \| I - \alpha \mathbf{Q} \|^2 + \alpha^2 \| E(\mathbf{G}_k^2) \| \\ \mu &\leqq \zeta^2 + \alpha^2 \gamma \end{aligned} \tag{46}$$

where $\gamma = \| \mathbf{G}_k^2 \|$. Furthermore for $\alpha < 2/(\rho_1 + \rho_N)$, we have $\zeta = 1 - \alpha\rho_1$. Then, using (46), we see that

$$\frac{\alpha^2}{1 - \mu} \leqq \frac{\alpha^2}{2\alpha\rho_1 + \alpha^2(\rho_1^2 + \gamma)}.$$

We have therefore shown that for positive and sufficiently small α, equations (31b) and (32) are valid where

$$V(\alpha) = \frac{\alpha}{2\rho_1 + \alpha^2(\rho_1^2 + \gamma)}. \tag{47}$$

REFERENCES

1. Lucky, R. W., "Automatic Equalization for Digital Communication," B.S.T.J., *44,* No. 4 (April 1965), pp. 547–588.
2. Lucky, R. W., "Techniques for Adaptive Equalization of Digital Communication Systems," B.S.T.J., *45,* No. 2 (February 1966), pp. 255–286.
3. Widrow, B., and Hoff, M. E., Jr., "Adaptive Switching Circuits," IRE Wescon Conv. Record, Pt. 4 (August 1960), pp. 96–104.
4. Narenda, K. S. and McBride, L. E., "Multiparameter Self-Optimizing Systems Using Correlation Techniques," IEEE Trans. Automatic Control, *AC-9,* No. 1 (January 1964), pp. 31–38.
5. Lucky, R. W. and Rudin, H. R., Generalized Automatic Equalization for Communication Channels", Proc. IEEE (Letters), *54,* No. 3, pp. 439–40, March 1966.
6. Lucky, R. W. and Rudin, H. R. An Automatic Equalizer for General-Purpose Communication Channels, B.S.T.J., *46,* No. 9 (November 1967), pp. 2179–2208.
7. Gersho, A., unpublished work.
8. Lytle, D. W., unpublished work.
9. Niessen, C. W., "Automatic Channel Equalization Algorithm," Proc. IEEE, *55,* No. 5 (May 1967), p. 689.
10. Niessen, C. W., and Drouilhet, P. R., "Adaptive Equalizer for Pulse Transmission," Conf. Digest, IEEE Int. Conf. on Commun. (June 1967), p. 117.
11. Boothroyd, W. P. and Creamer, E. M., Jr., "A Time Division Multiplexing System," Trans. AIEE, *68,* Pt. I, (1949) pp. 92–97.
12. Tufts, D. W., "Nyquist's Problem—the Joint Optimization of Transmitter

and Receiver in Pulse Amplitude Modulation," Proc. IEEE, *53,* No. 3 (March 1965), pp. 248–259.
13. George, D. A., "Matched Filters for Interfering Signals," IEEE Trans. Inform. Theory *IT-11,* No. 1 (January 1965), pp. 153–154.
14. Aaron, M. R. and Tufts, D. W., "Intersymbol Interference and Error Probability," IEEE Trans. Inform. Theory *IT-12,* No. 1 (January 1966), pp. 26–34.
15. Koford, J. S., and Groner, G. F., "The Use of an Adaptive Threshold Element to Design a Linear Optimal Pattern Classifier," IEEE Trans. Inform. Theory, *IT-12,* No. 1 (January 1966), pp. 42–50.
16. Widrow, B., "Adaptive Filters I: Fundamentals," Stanford Electronics Lab., Stanford, Calif., Report SEL-60-126 (Technical Report 6764-6), December 1966.
17. Coll, D. C. and George, D. A., "A Receiver for Time-Dispersed Pulses, Conf. Record, 195, IEEE Annual Commun. Conv., pp. 753–757.
18. Gersho, A., "Automatic Equalization of Highly Dispersive Channels for Data Transmission," Int. Symp. on Inform. Theory, San Remo, Italy, September 1967.
19. Gersho, A. and Freeny, S. L., "Performance Capabilities of Transversal Equalizers for Digital Communication," Conf. Digest, IEEE Int. Conf. on Commun. 1967, p. 88.
20. Goldstein, A. A., *Constructive Real Analysis,* New York: Harper and Row, 1967, Chap. 1.
21. Tong, P. S. and Liu, B., "Automatic Time Domain Equalization in the Presence of Noise," Proc. Nat. Elec. Conf., *23,* (October 1967), pp. 262–266.
22. Varga, R. S., *Matrix Iterative Analysis,* New York: Prentice-Hall, 1962.