POLYTECHNIC INSTITUTE OF BROOKLYN
DEPARTMENT OF ELECTRICAL ENGINEERING

Semi-Annual Status Report

NASA Grant NGR 33-006-040, Supplement 2

A STUDY of DIGITAL TECHNIQUES for SIGNAL PROCESSING

February 1, 1970 to July 31, 1970.

Principal Investigators:

Mischa Schwartz, Professor

Robert R. Boorstyn, Associate Professor

Introduction

We report here on activities in five areas carried out during the first six months of this grant period:

1. Adaptive array processing.

2. Dynamic programming applied to the design of adaptive equalizers.

3. Adaptive maximum-likelihood receiver for digital data transmission.

4. Recursive adaptive equalizers.

5. Finite memory communication systems.

The work on adaptive array processing has just been completed and a report describing the activities in detail is in preparation. This work was initiated during the second year of the grant, with previous work reported in the Status Reports of July 31, 1969 and January 31, 1970. In addition to the comprehensive report several papers are planned for publication.

The work on adaptive equalizers using a dynamic programming approach and on the adaptive maximum-likelihood receiver was initiated during this reporting period and is an outgrowth of previous work in adaptive equalizers, described in previous status reports. The work on recursive equalizers is a continuation of work initiated during the second year of the grant, reported on in the previous status report of January 31, 1970.

The work on finite memory communication systems was begun during this reporting period. A detailed discussion of the area and its significance to digital processing of signals was included in the Proposal leading to this third year of grant activities.

Papers delivered or appearing during the reporting period

Professor Schwartz was co-author of two papers delivered at symposia, based on the work on adaptive equalizers carried out under this grant. The first, entitled "Rapidly-Converging First Order Algorithms for Adaptive Equalization", was delivered at the Princeton Conference on Information and System Sciences in March 1970. The second, entitled "Rapidly-Converging Second Order Algorithms for Adaptive Equalization" was delivered at the 1970 International Symposium on Information Theory, Noordwijk, Holland, June 1970. The paper had previously been selected to be given as a "long" paper, a singular honor, since only 16 of the 190 presented at the Symposium were in this category. Copies of these papers were transmitted with the January 31, 1970 Status Report. The two papers have also been accepted for publication in the IEEE Transactions on Information Theory, subject to some cutting and minor revisions. Professor Schwartz is author also of the paper, "Computer Processing in Communications", keynoting the 1969 PIB International Symposium on Computer Processing in Communications, and appearing in the Proceedings of that Symposium, published in book form in June 1970. Copies of the paper are appended to this report.

Professor Boorstyn is co-author of two papers delivered at the 1970 International Symposium on Information Theory noted above. These papers are based on work done under this grant, as summarized in previous status reports, and described at length in two individual reports submitted previously under the grant. The titles of the papers are, respectively, "Properties of Real-Zero Signals", and "Optimum Differential Pulse Code Modulation". Professor Boorstyn is co-author as well of the paper, "Derivative Approximation for Recursive Signal Detection", describing work done under this grant, that also appears in the printed Proceedings of the PIB International Symposium on Computer Processing in Communications. Copies are also appended to this report. A second paper on Adaptive DPCM has been accepted for presentation at the forthcoming M. J. Kelley Symposium on Communications.

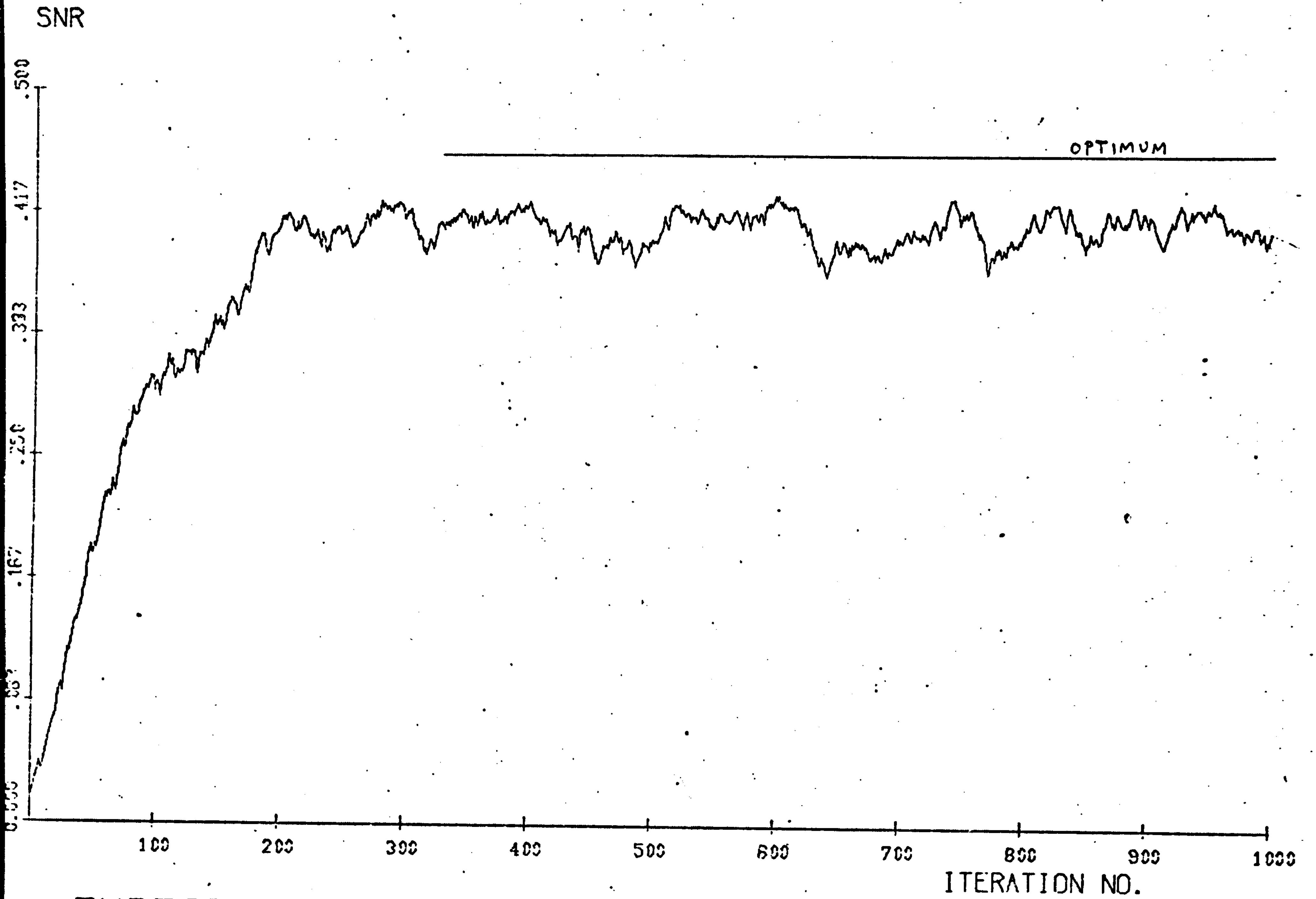
Summary of work during reporting period

1. Adaptive Array Processing: Previous work on this problem was described in the Status Reports of July 31, 1969 and January 31, 1970. The problem is to automatically make an array of isotropic detectors form a beam in a desired direction in space when unknown interfering noise is present so as to maximize the output signal-to-noise ratio (SNR). Iterative gradient techniques are used to do this.

As previously described, we have demonstrated the equivalence between the "antenna pattern" and "multichannel filter" points of view in designing optimum detector arrays. We have also reached the conclusion that, due to sensitivity considerations, we should not try to design our antenna excitations or multichannel filter coefficients on the basis of maximizing the SNR alone, but rather on the basis of maximizing the SNR subject to a constraint on the supergain ratio.

Since the last report we have been primarily concerned with the development of computer algorithms. Specifically we have completed the following:

1. Developed a computationally fast algorithm which numerically determines the optimum excitations to use in designing detector arrays subject to supergain constraints.



ENDFIRE, GRADIENT ESTIMATED, NO ADDITIVE NOISE

2. Simulated adaptive structures (see the Status Report of January 31, 1970) which minimize the mean square error (MSE) subject to a linear constraint and demonstrated agreement between the simulated and theoretical results.

3. Simulated adaptive structures which maximize the array output SNR subject to a constraint on the (nonlinear) supergain ratio, and demonstrated agreement between the simulated and theoretical results obtained in part 1 above.

In order to give some idea of how close the adaptive structure can approach the optimum structure (in terms of output SNR) after "steady state" is reached, a graph representing a typical simulation run is included.

The work in this area has now been concluded. A detailed report has been written and will be issued shortly.

2. Dynamic Programming Applied to Adaptive Equalizer Design: In previous reports on this grant we outlined a new method for the adaptive equalization of digital signals transmitted through dispersive channels. The equalizer investigated was a non-recursive digital filter with coefficients adjusted iteratively. The new method used a variable step size gradient search procedure, and the step sizes were chosen to provide the smallest possible distortion at the end of a fixed number of iterations.

We have now applied dynamic programming techniques to the adaptive equalization problem to see whether we can further speed up the convergence of the equalizer, as well as developing other equalization schemes that may be compared to these other methods for simplicity of implementation.

Specifically, let the input data samples be labeled x_h, h an integer; the output (equalized) samples be labeled y_h. For an N-tap equalizer,

$$y_h = \sum_{j=1}^{N} c_j x_{h-j+1} \tag{1}$$

To find the desired set of tap coefficients c_j we use a variable step size gradient search with the N-dimensional vector $\underline{C}$ on the $(i+1)\underline{th}$ iteration given by

$$\underline{C}_{i+1} = \underline{C}_i - \frac{\alpha_i}{2} \nabla \epsilon_i^2 \tag{2}$$

The mean-squared error ϵ^2 is given by

$$\epsilon^2 = \sum_j (y_j - m_j)^2 , \tag{3}$$

with m_j the known sequence of data transmitted.

With the dynamic programming technique we desire the appropriate sequence of step sizes α_i, $i = 1,1,\ldots M$, so that at the end of M iterations the error ϵ_M^2 is minimum, subject to the additional constraint

$$\sum_{i=0}^{M-1} \alpha_i^2 \leq E \tag{4}$$

with E a specified constraint. (This additional constraint is necessary to use dynamic programming techniques, and may be interpreted to be a constraint on the average expected excess mean-squared error).

The results are that the appropriate step sizes to be used are given very nearly by

$$\alpha_r \doteq \frac{1}{\lambda_M} \underline{\epsilon}_r^T A_r \underline{\epsilon}_r \tag{5}$$

with λ_M a constant, $\underline{\epsilon}_r$ the vector of the errors on the $r^{\underline{th}}$ iteration:

$$\underline{\epsilon}_r = \begin{bmatrix} (y_1 - m_1)_r \\ \vdots \\ (y_L - m_L)_r \end{bmatrix} \tag{6}$$

The matrix A_r is given by

$$A_r = M_r M_r^T , \tag{7}$$

with M_r in turn an input signal plus noise matrix defined as follows:

$$M_r = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & a_1 & \cdots & 0 \\ \vdots & & & \\ a_L & a_{L-1} & \cdots & a_{L-n+1} \\ \vdots & & & \\ 0 & \cdot & \cdots & a_L \end{bmatrix}_r$$

Here

$$a_i = x_i + n_i \ , \qquad (8)$$

the input signal plus noise sample on the i^{th} iteration.

The step sizes α_r thus depend on the (measured) statistics of input signal, noise, and output error.

Limited simulation has shown that this algorithm is faster than the fixed step size algorithm and that the tap gain error variance is lower than in the fixed step-size case.

We are presently deriving the convergence rate and variance equations and hope to show this analytically.

As a simplification of the dynamic programming equalizer, we are investigating a quantized step-size equalizer. The quantizer step sizes will be derived from the dynamic programming predicted step sizes. It is hoped that this equalizer performs better than the fixed step size equalizer and not much worse than the dynamic programming equalizer. It would, however, provide significant computational advantages.

3. <u>An Adaptive Maximum-Likelihood Receiver for Digital Data Transmission over Channels with Intersymbol Interference</u>: The work is concerned with the specification and design of receivers for digital data communication over channels where the rate of transmission is limited not only by the noise but also by the intersymbol interference (I.S.I.). One approach to the problem has involved a linear filter consisting of a tapped delay line with an automatically adjustable gain at each tap. Such an adaptive equalizer is capable of high speed digital signaling over slowly varying band limited channels whose impulse responses are unknown at the receiver. Different performance indices have been used to adjust the tap gains by an iterative procedure using either the steepest-descent technique or various modified versions of it, to obtain rapid convergence.

As another approach to minimize the effect of both the intersymbol interference and noise we have considered a maximum likelihood receiver. Such a receiver is known to be optimum in the sense of minimum probability of error. This receiver processes the received digital data in a sequential manner and makes decisions on a block of data of length L, where L is the number of time periods over which the intersymbol interference extends. The receiver uses a decision directed scheme to estimate the channel parameters and is therefore capable of being adaptive.

Initially we have considered the case L=1. That is, the I. S. I. extends only into the next neighboring time period. The transmission is a binary bipolar data sequence $\{a_k\}$ with white gaussian noise over the channel. It is assumed that the a_k can take the two values with equal probability. If $\{g_k\}$ are the received data in time sequence, the M.L. receiver computes the ratio

$$\lambda_k = \frac{f(\ldots g_{k-1}, g_k \mid a_k = 1)}{f(\ldots g_{k-1}, g_k \mid a_k = -1)} \tag{9}$$

where

$f(\ldots g_{k-1}, g_k | a_k = 1)$ is the conditional aposteriori probability density function. It decides that $a_k = 1$ was transmitted if $\lambda_k > 1$; otherwise $a_k = -1$ was transmitted. It is shown that in the absence of noise the g_k's form a Markov chain and the following recursive relationship can be obtained for computing λ_k:

$$\lambda_k = \ln \frac{e^{\lambda_{k-1}+a} + e^{b}}{e^{\lambda_{k-1}+a+b} + 1} + \frac{2}{\sigma^2}(g_k x_0) \tag{10}$$

where
$$a = \frac{2}{\sigma^2} g_k x_1$$
$$b = \frac{2}{\sigma^2} x_0 x_1$$

x_0 and x_1 are the impulse response of the channel at time t_k, t_{k+1}, and σ^2 is the noise power.

It is shown that the structure of such a receiver contains a tapped delay line with non-linear amplifiers at the output of each tap. The receiver can be made adaptive by a decision directed scheme in which the estimated a_k's can be used to obtain a linear estimate of the channel parameters x_0 and x_1,

For the receiver of this type probability of error was computed under the assumption that the noise is stationary and therefore the probability density function will take on a stationary value as the new data is received. It is shown that

$$Pe = \text{Prob}\left\{ \frac{\text{Sinh } g_k \frac{(x_0 - x_1)}{\sigma^2}}{\text{Sinh } g_k \frac{(x_0 - x_1)}{\sigma^2}} < -e^{\frac{-2x_0 x_1}{\sigma^2}} \right\} \tag{11}$$

Next the study was extended to take into account the case $L = 2$. Here the transmitted information is considered in a block of two symbols

$\{a_{k-1}, a_k\}$ and is denoted by a sequence $\{y_k\}$. The M.L. Receiver now computes the ratio

$$\lambda_k = \ln \frac{f(\ldots g_{k-1}, g_k | y_k = i)}{f(\ldots g_{k-1}, g_k | y_k = j)} \qquad i,j = 1,2,3,4 \tag{12}$$

and decides that the sequence $y_k = i$ was transmitted if $\lambda_k > 1$ for all $i \neq j$.

A recursive vector equation is obtained to compute λ_k. We define

$$\underline{P}_{g|y}(k) = \begin{bmatrix} f(\ldots g_{k-1}, g_k | y_k = 1) \\ f(\ldots g_{k-1}, g_k | y_k = 2) \\ f(\ldots g_{k-1}, g_k | y_k = 3) \\ f(\ldots g_{k-1}, g_k | y_k = 4) \end{bmatrix}$$

Then

$$A(k) = [a_{ij}(k)], \quad a_{ij} = P(g_k | y_k = i, y_{k-1} = j)\, P(y_k = i | y_{k-1} = j) \tag{13}$$

The first term in $a_{ij}(k)$ involves the received information, while the second term is the transitional probability for y_k's which form a Markov Chain.

Knowing the noise distribution and therefore the distribution for g_k and the transition probabilities in the Markov Chain the elements in the A(k) matrix are known. We have

$$\underline{P}_{g|y}(k) = A(k)\, \underline{P}_{g|y}(k-1) \tag{14}$$

In the present case L = 2 the matrix A(k) is a 4 x 4 matrix with only half the elements non zero, the reason being that y_k's form a Markov Chain and with the help of a state diagram we find that certain transitional probabilities are zero.

This receiver is again capable of being made adaptive by the decision directed scheme suggested previously and can easily be extended to the general case for extended I.S.I.

The scheme was simulated on a digital computer for the case of a known channel and works successfully in decoding the received information. It is planned in the future to carry out the following tasks:

1) To compute the probability of error in the general case or obtain certain bounds on it to measure the performance of this system.

2) To simulate the scheme on a computer when the channel parameters are unknown and study the effect of different decision directed schemes on the overall performance.

3) In the general case the size of the matrix A(k) will be $2^L \times 2^L$ which will involve a large number of computations. Hence it may be worthwhile to look for a suboptimum scheme with considerable saving in computation time with little loss in the performance.

4) It appears that such a minimum probability of error receiver is similar to the Viterbi decoding algorithm for convolutional codes. Therefore it would be nice to bridge these two fields and show that the techniques used to analyze the optimum decoding of convolutional codes are similar to the analysis of the optimum receiver in our case.

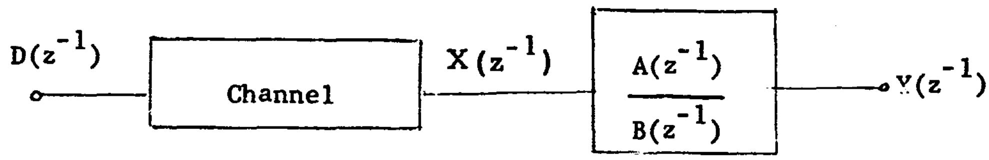
4. Recursive Adaptive Filters for Equalization: The recursive filter structure for minimizing the intersymbol distortion of a communication channel is analyzed in this work. Because of the inherent non-linear property of the recursive structure, an open-loop scheme is utilized to linearize the non-linear system.

The performance of the recursive equalizer (using the open-loop scheme) is compared with that of the non-recursive (transversal) equalizer for some known channels. Some of the problems associated with the open-loop scheme of the recursive equalizer are discussed.

A) Introduction: Non-recursive filters (transversal filters) have been satisfactorily used to equalize the intersymbol distortion of communication channels. Lucky[1] has pointed out that the performance of the transversal filter is especially good for channels which exhibit time dispersion over only a small number of symbols duration. Kaiser[2] has shown that if the amplitude-frequency characteristic of the network (in our case, an equalizer) has a sharp slope such as the sharp transition between the passband and stopband of any bandpass type network, a recursive structure representation of the network requires far less number of taps than the non-recursive structure representation. Therefore, to equalize those channels which exhibit extremely long time dispersion and require an extremely large number of taps in the non-recursive structure, one would try to use a recursive structure if one can successfully find it. In this work the recursive structure is analyzed as an equalizer.

The modified error (or linearized error) first introduced by Mantey[3] will be used in the optimization of the recursive filter parameters as an equalizer.

B) Formulation: 1) Mean Square Error



Recursive Equalizer

The mean square error is defined:

$$E \equiv \frac{1}{2\pi j} \oint \left| Y(z^{-1}) - D(z^{-1}) \right|^2 z^{-1} dz \equiv \sum_{n=0}^{\infty} (y_n - d_n)^2 \quad (15)$$

where d_n is the desired signal and, e.g.,

$$D(z^{-1}) = \sum_{i=0}^{\infty} d_i z^{-i}$$

Because of the inherent non-linearity in E, multiple minima can occur even when $A(z^{-1})/B(z^{-1}) = D(z^{-1})/X(z^{-1})$ which makes the mean square error vanish[3]. Therefore, a gradient procedure could fail to terminate at the solution that makes the error vanish even when such a possible solution exists. This deficiency is a severe one.

2) Modified Error - Open-loop Scheme: The recursive filter is basically characterized by the following difference equation:

$$\sum_{k=0}^{N} a_k\, x(nt - kt) = y(nt) + \sum_{k=1}^{M} b_k\, y(nt - kt) \tag{16}$$

For convenience the above equation can be written as:

$$\sum_{k=0}^{N} a_k\, x_{n-k} = y_n + \sum_{k=1}^{M} b_k\, y_{n-k} \tag{17}$$

For normal equalization, the objective is to choose the parameters a_k and b_k such that y_n is as close as possible to the desired signal d_n. For this purpose one may let $y_n = d_n$ in the equation (17) and choose the parameters that minimize the difference between the two sides of equation (17). In other words, one can define a convenient error function. (Mantey called this the modified error. This could be called an open-loop scheme.)

$$e_n = \sum_{k=0}^{N} a_k\, x_{n-k} - \left(d_n + \sum_{k=1}^{M} b_k\, d_{n-k}\right) \tag{18}$$

3) Minimization of the mean square error of the modified error. The mean square error of the modified error is defined

$$E_m = \sum_{n=0}^{\infty} e_n^2 \tag{19}$$

For convenience of the analysis, let us define $(N + 1 + M)$ vectors $\underline{C}$ and $\underline{q}_n$ as follows:

$$\underline{C}' = [a_0, a_1, a_2, \ldots, a_N, -b_1, -b_i, \ldots, -b_M] \tag{20}$$

$$\underline{q}_n = [x_n, x_{n-1}, x_{n-2}, \ldots, x_{n-N}, d_{n-1}, d_{n-2}, \ldots, d_{n-M}] \tag{21}$$

5. Finite Memory Communication Systems:

In converting communications systems to digital processing a recurrent problem is the resultant discrete nature of the operations. Both data and parameters (e.g. coefficients) are quantized. The subsequent round-off and truncation errors and their propagation are often studied. However, a completely different approach is to view the entire system as a finite state machine. We are concerned with the novel features of this viewpoint.

In particular we have been investigating finite-memory systems for the two hypothesis testing problem based on an approach suggested by Cover and Hellman[4]. Let $x_1, x_2, \ldots, x_n$ be a sequence of independent identically distributed gaussian random variables with unit variance and mean -1 under hypothesis H_0, and +1 under hypothesis H_1. Hellman and Cover have shown that the optimum m state memory for detecting which hypothesis is true under the constraint that the data sequence be infinite is as follows: labeling the state of the memory 1,2,...,m, the machine will move from any state j, j=1,...,m, to state j + 1 or j - 1 if the current data sample is such that it maximizes or minimizes the likelihood ratio of the two hypotheses. If the current data sample does not maximize or minimize the likelihood ratio, the memory remains in its present state. No other transitions are necessary. For our gaussian example, the long term occupation probability of the states of the memory is zero for states 2,...,m-1, and finite for states 1 and m. Therefore if we are in state 1, we assume H_0 is true and if we are in state m we assume H_1 is true. To build this machine we would set two thresholds, one at -D and the other at +D where $D \to \infty$. Every time the input sample was greater than D or less than -D we would move to the next higher or lower state from our previous state.

The machine described above is fine for the infinite data constraint. However, if we limit the data samples to say n, where n is finite, the optimum infinite data machine is not even a close approximation to the optimum finite data machine, the reason being that in the infinite data case the machine can wait for and use only that information that maximizes or minimizes the likelihood ratio. However, for the finite data case there is no guarantee that this data will occur. Therefore, for the finite data case, the machine must make use of more of the information.

In an attempt to gain insight into the memory structure for the finite data case machines with only a few states are being investigated. Below we describe these machines and develop expressions for their performance. Subsequent work will include calculations of performance.

Two State Memory:

Again consider the gaussian example. Under hypothesis H_0 the data has mean -1 and unity variance. Under hypothesis H_1 the data has mean +1 and unity variance. The machine consists of two states, 0 and 1, and the data is compared to two thresholds $\pm D$. If the memory is in state 0 and the input is greater than D we move to state 1. If the memory is in state 1 we move to state 0 if the input is less than -D. For all other inputs we stay in the previous state. If it is assumed both hypotheses are equally probable, then the probability of error can be calculated assuming H_1 is true and calculating the probability of being in state 0. For this case

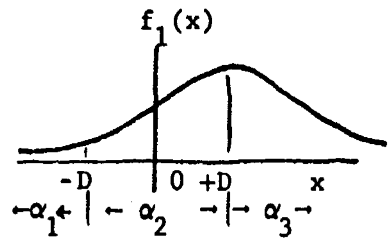
$$Pe^{(2)}(n) = \frac{1}{2}\left\{1 - \frac{(\alpha_3-\alpha_1)}{\alpha_1+\alpha_3}(1 - \alpha_2^{\,n})\right\} = \frac{1}{2}\left\{1 - (\alpha_3-\alpha_1)\left[\frac{(1-\alpha_2^{\,n})}{1-\alpha_2}\right]\right\} \quad (26)$$

where

$\alpha_1 = \text{prob}\ \{x < -D\}$

$\alpha_2 = \text{prob}\ \{-D < x < D\}$

$\alpha_3 = \text{prob}\ \{x > D\}$



If $|D| = \infty$, which is the case for infinite data, then $\alpha_2 = 1$ and $Pe^{(2)}(\infty) = \frac{1}{2}$, showing that the infinite data machine is far from optimum in the finite data case.

Taking $\frac{\partial Pe^{(2)}(n)}{\partial D}$ and setting it equal to zero, we get

$$[2\alpha_3(1-\alpha_2^{\,n}) - n(\alpha_3^{\,2}-\alpha_1^{\,2})\alpha_2^{\,n-1}]\ \alpha_1' = [2\alpha_1(1-\alpha_2^{\,n}) - n(\alpha_1^{\,2}-\alpha_3^{\,2})\ \alpha_2^{\,n-1}]\ \alpha_3'. \quad (27)$$

This is the equation that the thresholds must satisfy for the optimum setting. Since $\alpha_2 < 1$, if $n \to \infty$ we get

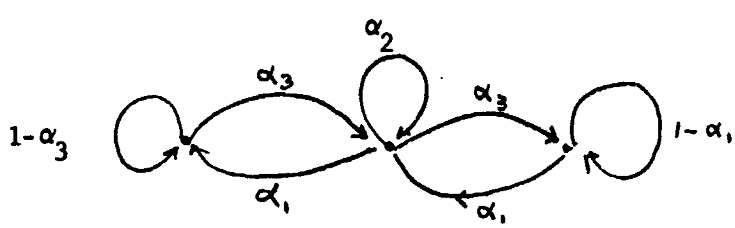
$$\alpha_3\alpha_1' = \alpha_1\alpha_3' \quad (28)$$

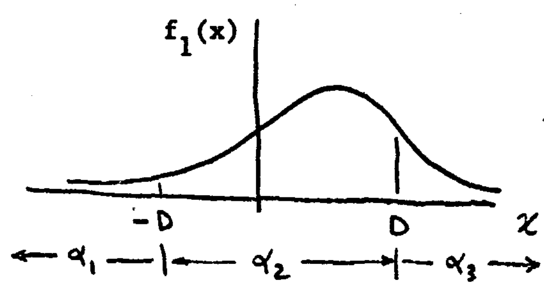
which is satisfied if $|D| = \infty$.

Three State Memory:

Again considering the gaussian example and assuming H_1 is true and that H_0 is detected we have the following cases:

(1) Single Threshold





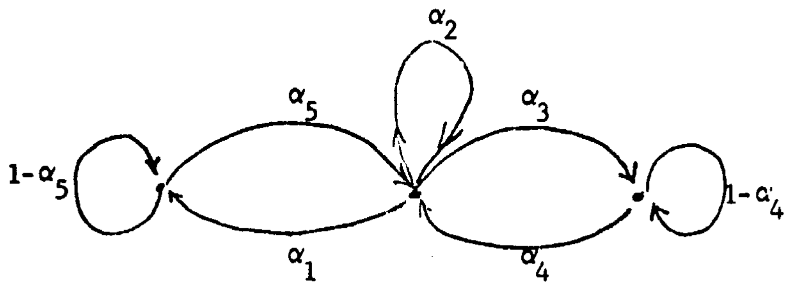
The resulting probability of error is

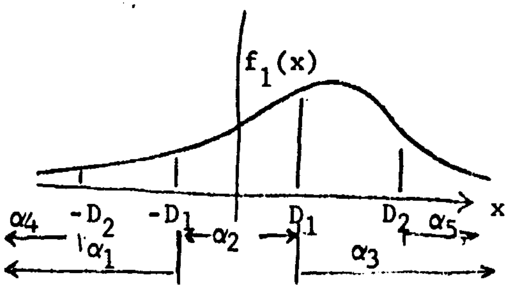
$$Pe^{(3)}(n) = \frac{1}{2}\left\{1 - \frac{(\alpha_3-\alpha_1)}{2}\left[\frac{(1-\lambda_2^{\,n})}{1-\lambda_2} + \frac{(1-\lambda_3^{\,n})}{1-\lambda_3}\right]\right\} \quad (29)$$

where $\lambda_2 = \alpha_2 + \sqrt{\alpha_1\alpha_3}$

$\lambda_3 = \alpha_2 - \sqrt{\alpha_1\alpha_3}$

(2) Double Threshold





the resulting probability of error is

$$Pe^{(3)}(n) = \frac{1}{2}\left\{1 - \frac{(\alpha_3-\alpha_1)}{2}\left(\frac{1-\lambda_2{}^n}{1-\lambda_2} + \frac{1-\lambda_3{}^n}{1-\lambda_3}\right)\right\} \tag{30}$$

$$+ \frac{(\alpha_1+\alpha_3)\left(\frac{x-y}{2}\right)}{2(\lambda_2-\lambda_3)}\left(\frac{1-\lambda_2{}^n}{1-\lambda_2} - \frac{1-\lambda_3{}^n}{1-\lambda_3}\right)$$

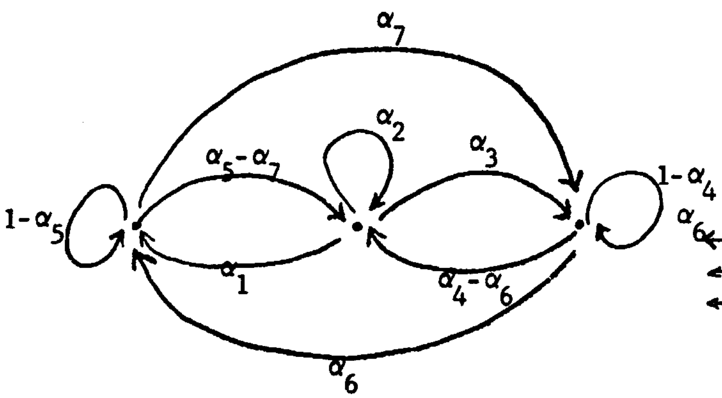
where $\lambda_2 = \alpha_2 + \left(\frac{x+y}{2}\right) + \sqrt{\alpha_1\alpha_3 + \left(\frac{x-y}{2}\right)^2}$

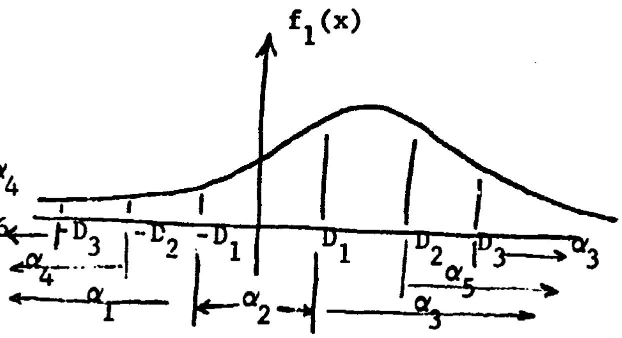
$\lambda_3 = \alpha_2 + \left(\frac{x+y}{2}\right) - \sqrt{\alpha_1\alpha_3 + \left(\frac{x-y}{2}\right)^2}$

$x = \alpha_3 - \alpha_5$

$y = \alpha_1 - \alpha_4$

(3) Triple Threshold





The resulting probability of error is

$$Pe^{(3)}(n) = \frac{1}{2}\left\{1 - \frac{(\alpha_3-\alpha_1)}{2}\left(\frac{1-\lambda_2{}^n}{1-\lambda_2}\right) + \left(\frac{1-\lambda_3{}^n}{1-\lambda_3}\right)\right\}$$
$$+ \left\{\frac{(\alpha_1+\alpha_3)\left(\frac{x-y}{2}\right) \; \alpha_3\alpha_6 - \alpha_1\alpha_7}{2(\lambda_2-\lambda_3)}\right\}\left(\frac{1-\lambda_2{}^n}{1-\lambda_2} - \frac{1-\lambda_3{}^n}{1-\lambda_3}\right) \tag{31}$$

where

$$\lambda_2 = \alpha_2 + \frac{(x-y)}{2} + \sqrt{\alpha_1\alpha_3 + \left(\frac{x-y}{2}\right)^2 - \alpha_1\alpha_7 \cdot \alpha_6\alpha_7 - \alpha_3\alpha_6}$$

$$\lambda_3 = \alpha_2 + \frac{(x-y)}{2} - \sqrt{\alpha_1\alpha_3 + \left(\frac{x-y}{2}\right)^2 - \alpha_1\alpha_7 + \alpha_6\alpha_7 - \alpha_3\alpha_6}$$

$$x = \alpha_3 - \alpha_5$$

$$y = \alpha_1 \;\; \alpha_4$$

Case (1) is just a simple extension of the two state machine. Case (2) tries to take advantage of the fact that when the machine is in an end state it favors either H_0 or H_1. Therefore it will not leave this state unless the next input is greater than some second threshold. Case (3) says if the machine is in an end state and if the input is sufficiently large, favoring one hypothesis much more than the other, that the memory should completely switch in the machine to the other end state unless it is already there.

Cases (1), (2), and (3) seem to fairly completely cover the three state machine. Computer analysis is being done to find the optimum thresholds for each case. With these thresholds the probabilities of errors can be compared to determine if any advantage is gained by the added complexity in each case.

Work is continuing to try and extend these results to a general m state machine.

References

1. R. W. Lucky, J. Salz and E. J. Weldon, Jr., Principles of Data Communication, New York, McGraw-Hill, 1968.

2. J. F. Kaiser and F. F. Kuo, System Analysis by Digital Computer, New York, Wiley and Sons, 1966.

3. P. Mantey, "Convergent Automatic-Synthesis Procedures for Sampled-Data Networks with Feedback", Stanford Electronics Laboratories, Report No. 6773-1, Stanford, California, October 1964.

4. T. M. Cover, "Hypothesis Testing with Finite Statistics", Annals of Math. Stat., Vol. 40, No. 3, pp. 828-835, June 1969.

M. Hellman, "Learning with Finite Memory", Ph.D. Thesis, Stanford University, Stanford, California, 1969.