AD NUMBER
AD826128
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 30 SEP 1967. Other requests shall be referred to Air Force Technical Application Center, Washington, DC 20330.**
AUTHORITY
AFTAC USAF ltr 25 Jan 1972

AD826128

STATEMENT #2 UNCLASSIFIED

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Chief, AFTAC
Attn: VSC Wash, D.C. 20333

40

SCIENCE SERVICES DIVISION

TEXAS INSTRUMENTS INCORPORATED

VELA T/7701

STUDY OF A 1-POINT ADAPTIVE FILTER

ADVANCED ARRAY RESEARCH

Special Report No. 4

Prepared by

James E. Brown

Aaron H. Booker John P. Burg

George D. Hair, Program Manager

TEXAS INSTRUMENTS INCORPORATED
Science Services Division
P.O. Box 5621
Dallas, Texas 75222

Contract: F33657-67-C-0708-P001
Contract Date: 15 December 1966
Amount of Contract: $625,500
Contract Expiration Date: 14 December 1967

Sponsored by

ADVANCED RESEARCH PROJECTS AGENCY
Nuclear Test Detection Office
ARPA Order No. 624
ARPA Program Code No. 7F10

30 September 1967

ACKNOWLEDGMENT

This research was supported by the

ADVANCED RESEARCH PROJECTS AGENCY
Nuclear Test Detection Office

under Project VELA UNIFORM
and accomplished under the technical direction of the

AIR FORCE TECHNICAL APPLICATIONS CENTER
under Contract No. F33657-67-C-0708-P001

ABSTRACT

The design and evaluation of optimum, time-invariant filters (the classical approach to prediction problems) are subject to several limitations. Because of these limitations, an adaptive system was considered. This report describes a filtering system which can adjust to changes in either signal or noise, thereby overcoming many difficulties of time-invariant filtering. The small amount of required computational time to update the filter weights is another important feature of the scheme presented.

In this report, an adaption algorithm applied to a simple time-series model was studied. For the case of stationary data, a tradeoff between the adaption rate and the mean-square-error performance of the filter exists. The adaption rate is inversely proportional to the convergence parameter λ, while the mean-square-error is directly proportional to λ. For the case of nonstationary data, a tradeoff between adapting too slowly and adapting too rapidly exists. The optimum rate of adaption appears to be approximately 10 times faster than the average time rate of change in the input data statistics.

BLANK PAGE

TABLE OF CONTENTS

LIST OF APPENDIXES

LIST OF ILLUSTRATIONS

SECTION I

THE ADAPTIVE ALGORITHM

The classical approach to prediction problems is to design and evaluate optimum, time-invariant filters. Unfortunately, these filters are subject to several limitations which restrict their usefulness. Numerical determination of the optimum filter is often involved and costly. Even after the optimum filter is determined, it is optimum only for the signal and noise models from which it was designed. A change in the signal or noise model requires a new derivation, which frequently is considerably difficult and expensive. This report describes a filtering system which can adjust to changes in either signal or noise and thus overcome many difficulties of time-invariant filtering. The small amount of required computational time to update the filter weights is another important feature of the scheme presented.

In general, adaptive systems can be considered the combination of two systems — an adjustable "worker," with internal parameters which can be varied in value upon command from a "supervisor." The purpose of the supervisor is to measure the filter's performance and to output the necessary adjustment to the worker (Figure 1). Of particular interest are adaptive devices which perform a linear operation on the input data. The adaptive process can be viewed then as the search for a weight vector that "optimizes" the response of the filter.

There are many ways to measure how well the system performs its intended task. One of the most common (the one which shall be adopted in this paper) is the criterion of least-mean-square-error, the criterion used in the Wiener theory of signal extraction and prediction.[1] Not only is the least-mean-square-error criterion convenient for comparison to classical time-invariant filtering, but it also lends itself quite easily to an adaptive algorithm, which shall now be considered.

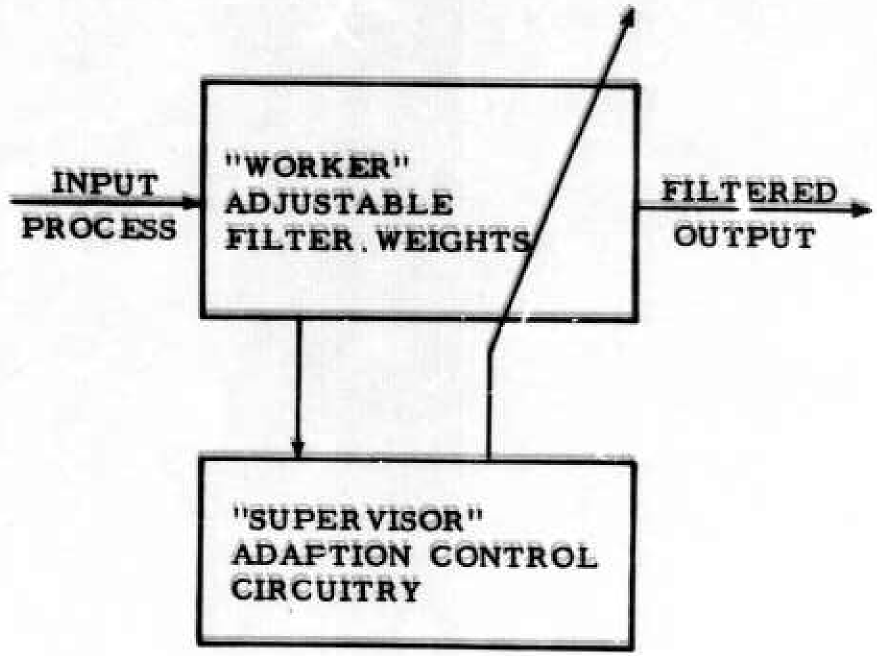

Figure 1. An Adaptive Processor

A. DERIVATION OF ADAPTIVE ALGORITHM

The adaptive algorithm can be derived by considering the Wiener theory for sampled-data systems. The theory is concerned with finding the linear, time-invariant, sample-point operator $H(t)$, which weighs the input time-series $\{x_t\}$ in such a way that the output of the filter at time t, $\hat{S}_t$, is the best mean-square approximation to the desired output S_t.

Assuming that the impulse response of the filter is of the form

$$h(t) = \sum_{j=L}^{M+L-1} a_j \, \delta(t - j\,\Delta t) \tag{1}$$

the mean-square-error can be written as

$$\overline{\epsilon^2} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (S_t - \hat{S}_t)^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \left(S_t - \sum_{j=L}^{M+L-1} a_j x_{t-j} \right)^2 \tag{2}$$

The filter weights a_j which minimize Equation 2 are those weights for which the gradient of $\overline{\epsilon^2}$ equals 0.[1] Solving for the a_j, the following matrix equation is obtained:[2]

$$\underline{\underline{R}}\ \underline{a} = \underline{\varphi}^{*} \tag{3a}$$

* In this report, a double bar under a literal expression represents a square matrix, and a single bar represents a column matrix.

For the single-channel problem of predicting L points ahead with an M-point filter, the elements $\underline{\underline{R}}$, $\underline{a}$, $\underline{\varphi}$ can be written as

$$\underline{\underline{R}} = \begin{bmatrix} r_x(0) & r_x(1) & \cdot & \cdot & \cdot & \cdot & r_x(M-1) \\ r_x(1) & r_x(0) & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ r_x(M-1) & & \cdot & \cdot & \cdot & \cdot & r_x(0) \end{bmatrix} \qquad (3b)$$

$$\underline{a} = \begin{Bmatrix} a_L \\ \cdot \\ \cdot \\ \cdot \\ a_{M+L-1} \end{Bmatrix} \qquad (3c)$$

and

$$\underline{\varphi} = \begin{Bmatrix} r_x(L) \\ \cdot \\ \cdot \\ \cdot \\ r_x(M+L-1) \end{Bmatrix} \qquad (3d)$$

where $r_x(j)$ is the j^{th} lag correlation value for the data.

Rather than solving Equation 3a for the filter $\underline{a}$ directly by inverting the $\underline{\underline{R}}$ matrix, an iterative method can be used to obtain successive approximations to the optimum filter weights. In the technique considered, a new filter matrix $\underline{a}(k+1)$ is determined from the previous filter matrix $\underline{a}(k)$ by traveling in the direction which maximizes the change of the mean-square-error for a given change of the filter matrix. This direction is given by the gradient vector of the mean-square-error:

$$\underline{\nabla}\,\overline{\epsilon^2} = \begin{bmatrix} \dfrac{\partial \overline{\epsilon^2}}{\partial a_L} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial \overline{\epsilon^2}}{\partial a_{M+L-1}} \end{bmatrix}$$

The optimum step length λ down the gradient for each iteration can be found in a systematic way from iteration to iteration. Fortunately for computational considerations, λ can also be held constant and the method will still converge.[3] Thus, our scheme can be written

$$\underline{a}(k+1) = \underline{a}(k) - \frac{\lambda}{2}\,\underline{\nabla}\,\overline{\epsilon^2}\Big|_{\underline{a}=\underline{a}(k)} \qquad (4)$$

A drawback of the iterative procedure derived in Equation 4 is that it requires knowledge of the gradient of the mean-square-error and this is seldom, if ever, known in practice.

Several authors[4,5] have shown that the preceding algorithm can be modified by using the instantaneous estimate of $\underline{\nabla}\,\overline{\epsilon^2}\,\Big|_{\underline{a}=\underline{a}(k)}$ given by

$$\nabla \epsilon_k^2 = \begin{bmatrix} \dfrac{\partial \epsilon_k^2}{\partial a_L} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial \epsilon_k^2}{\partial a_{M+L-1}} \end{bmatrix} = -2\epsilon_k \begin{bmatrix} x_{k-L} \\ \cdot \\ \cdot \\ \cdot \\ x_{k-M-L+1} \end{bmatrix} = -2\epsilon_k \underline{x}_k$$

The adaptive algorithm then becomes

$$\underline{a}(k+1) = \underline{a}(k) + \lambda \epsilon_k \underline{x}_k \qquad (5)$$

This scheme can be generalized to an NC-channel problem by treating the data matrix $\underline{x}_k$ as being composed of data from all NC channels. The filter matrix $\underline{a}(k)$ represents the NC filters applied to the data vector $\underline{x}_k$. Figure 2 shows this filter system schematically. The superscript represents the channel, and the subscript represents the time point.

B. TIME-CONSTANT CONSIDERATIONS

It is desirable to be able to relate the parameter λ to the amount of data being used to design the set of filter coefficients at each point. We propose to do this in the following manner.

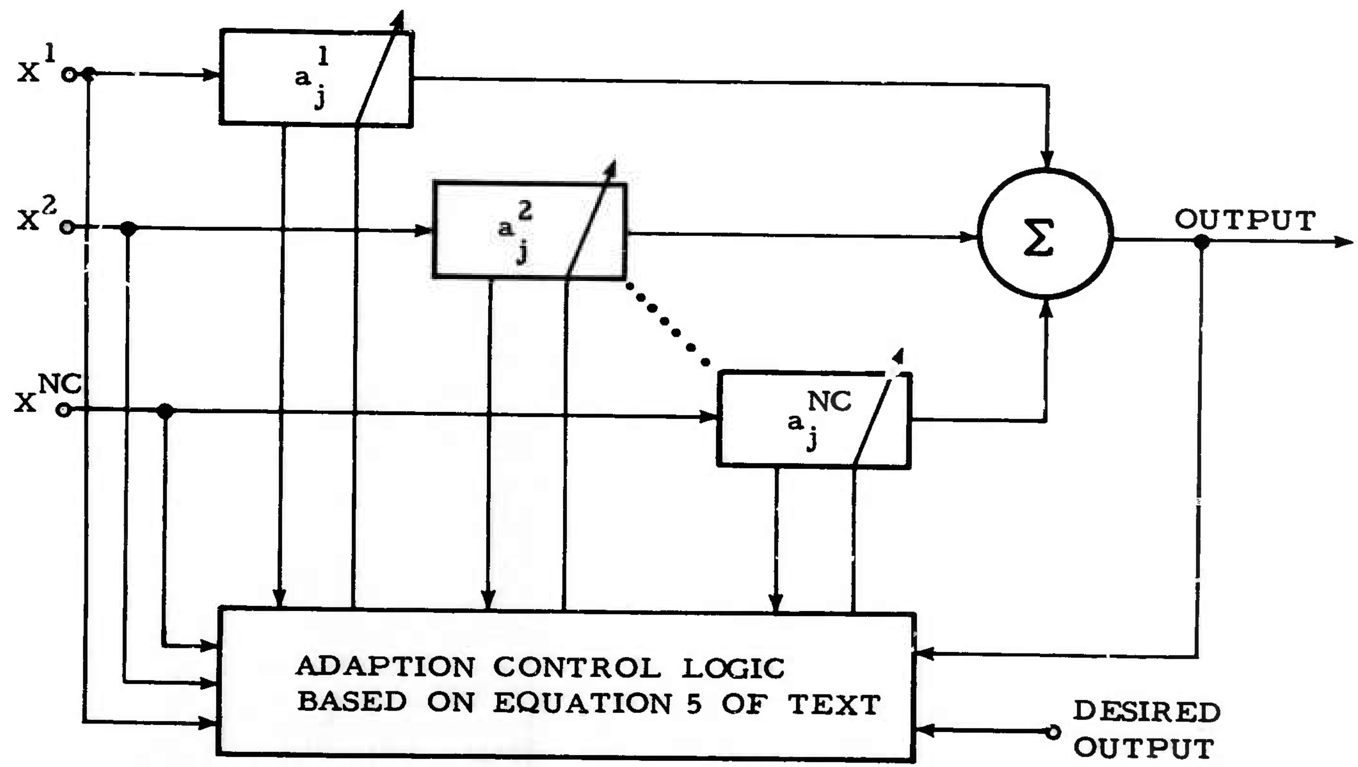

Figure 2. An Adaptive Multichannel Filter System

We define the sequence $\{\underline{a}_k'\}$ by

$$\underline{a}_k' = \underline{a}_k + \lambda_k' \epsilon_k \underline{x}_k$$

where

$$\lambda_k' = \frac{1}{\underline{x}_k^T \underline{x}_k} \quad *$$

and $\underline{a}_k$ is the adaptive filter matrix at time k. The sequence $\underline{a}_k'$ can be considered the "false gain" set of coefficients, since

$$\underline{x}_k^T \underline{a}_k' - \left(\underline{x}_k^T \underline{a}_k + \epsilon_k\right) = 0$$

We consider the sequence

$$\underline{\beta}_k = \frac{\sum_{i=1}^{\infty} e^{-i/\tau} \underline{a}_{k-i}'}{\sum_{i=1}^{\infty} e^{-i/\tau}} = \left(1 - e^{-1/\tau}\right) \underline{a}_{k-1}' + e^{-1/\tau} \underline{\beta}_{k-1}$$

$$= \underline{\beta}_{k-1} + \left(1 - e^{-1/\tau}\right)\left(\underline{a}_{k-1}' - \underline{\beta}_{k-1}\right)$$

*T denotes transpose

If we let $\underline{\beta}_k = \underline{a}_k$, then

$$\underline{a}_{k+1} = \underline{a}_k + \lambda_k' \left(1 - e^{-1/\tau}\right) \epsilon_k \underline{x}_k$$

$$= \underline{a}_k + \lambda_k \epsilon_k \underline{x}_k$$

where

$$\lambda_k \equiv \lambda_k' \left(1 - e^{-1/\tau}\right)$$

If λ_k' is constant in time, note that this algorithm is the adaptive algorithm derived earlier, with a time constant τ.

In general, λ_k' varies in time; the degree of variation affects the time constant of the adaptive algorithm in Equation 5; however, a "ball park" estimate of τ can still be obtained. Take λ_k' to be given by

$$\lambda_k' = \frac{1}{E\left[\underline{x}_k \underline{x}_k^T\right]} = \frac{1}{M \cdot NC \cdot r_x(0)} \qquad (6a)$$

where

M = number of filter points

NC = number of channels

$r_x(0)$ = zero-lag correlation of the data

The ratio between λ_k' and λ used in Equation 5, then, will give an estimate of τ; i.e.,

$$\frac{\lambda}{\lambda_k'} = \frac{\lambda}{M \cdot NC \cdot r_x(0)} = 1 - e^{-1/\tau} \tag{6b}$$

Appendix D shows that time constant τ can be associated with an "equivalent" boxcar section of data of length N by the relation

$$N = \frac{1 + e^{-1/\tau}}{1 - e^{-1/\tau}} \tag{6c}$$

SECTION II
ANALYSIS OF THE SINGLE-CHANNEL ALGORITHM

For the multichannel adaptive filter, the behavior of the algorithm becomes obscured by the complexity of the system. For this reason, the adaptive process is examined for a simple single-channel prediction problem. The results of this investigation give some intuitive feeling for the behavior of this scheme for more complicated filtering problems.

A. CASE 1, STATIONARY TIME SERIES

One of the simplest time series to which the adaptive filter could be applied is a first-order autoregressive series. The series $\{x_t\}$ is defined by the recursive relation

$$x_t = \alpha x_{t-1} + n_t \quad (7)$$

where the series $\{n_t\}$ is a sequence of independent Gaussian random variables of zero mean and unit variance. The parameter α, which must be less than 1 in magnitude for stability, determines the coherence (predictability) of the time series.

For the case where α is a constant in time, the process defined by Equation 7 is stationary. The optimum Wiener prediction filter is given by

$$h(t) = \alpha \delta (t-1)^*$$

with a mean-square-error of unity. Thus, we can analyze the 1-point adaptive filter by comparing the filter weight to the optimum time-invariant filter weight α and the mean-square-error of this process to unity.

---

* δ is the impulse function.

By starting the adaptive filter weight a_t at 0 and letting the filter adjust, we can measure the time constant of adaption. On the basis of theoretical considerations, a reasonable estimate of the time for initial transients to decrease to 1/e of their original value is

$$\frac{1}{\lambda r_x(0)} *$$

where $r_x(0)$ is the zero-lag correlation of the data and λ is the convergence parameter. The behavior of a_t has been studied experimentally by simulating the adaptive process on an IBM 7044 computer at Texas Instruments Incorporated. Results of this investigation for various values of λ for a fixed α are shown in Figure 3 where the dashed line is the theoretical curve for the expected value of a_t. We see that the theoretical and experimental curves are in close agreement

The area of primary interest is the LMS adaptive filter's steady-state behavior. To compare the steady-state performance of the adaptive system with that of the classical time-invariant filter, it is convenient to introduce a "misadjustment" parameter M. The misadjustment M is a dimensionless quantity defined by the ratio

$$M = \frac{MSE_a - MSE_o}{MSE_o} \tag{8}$$

where

MSE_a = mean-square-error of output adaption filter

MSE_o = mean-square-error of output classical time-invariant filter

---

*See Appendix A for the theoretical derivation.

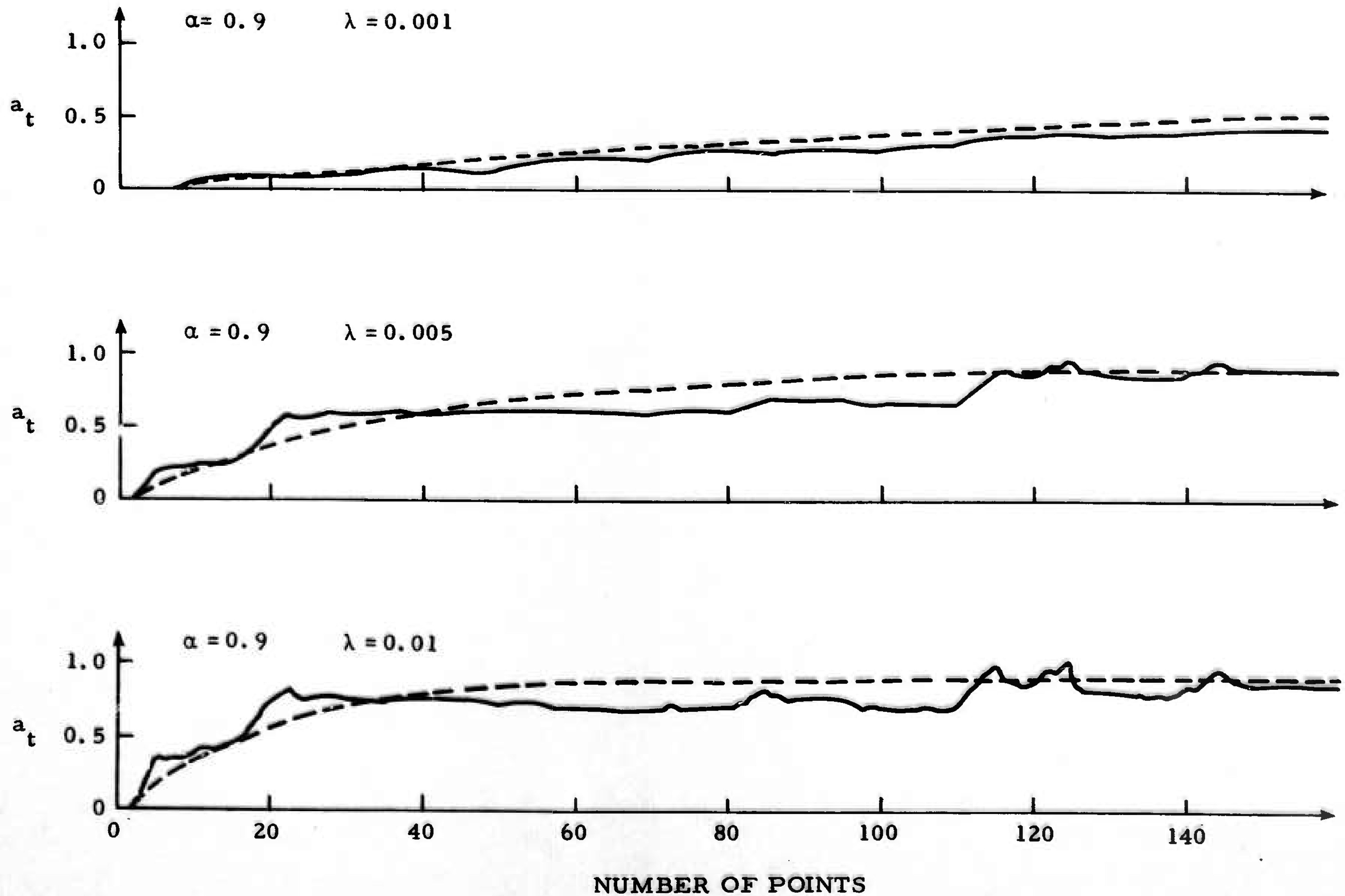

Figure 3. Behavior of the 1-Point Adaptive Filter Weight

The misadjustment coefficient M has been experimentally determined for the autoregressive time series defined by Equation 7. Results for three values of α as a function of λ are plotted in Figure 4 where the dotted lines show the theoretical curves* for the misadjustment coefficient as a function of λ. The filter weight is started at the proper value $a_1 = \alpha$, and the mean-square-error is computed over 1000 points. The curve for $\alpha = 0.2$ (Figure 4) falls below the 0-percent line because of the statistical fluctuations in the data. Had the mean-square-error been computed over an infinite length of data, this curve would have been slightly above the 0-percent line.

B. CASE 2, NONSTATIONARY TIME SERIES

Of practical importance is the behavior of the adaptive filter system when the input time series is nonstationary. Data used for this experiment are first-order time-varying autoregressive series

$$x_t = \alpha_t x_{t-1} + n_t$$

where n_t is as previously defined and $\alpha_t = A + B \sin 2\pi ft$. A is the mean value of α_t, B is the amplitude of the sinusoidal deviation about the mean, and f is the frequency of the deviation.

The experimental results of this investigation are shown in Figure 5. Figure 5a plots the mean-square-error of the adaptive filter as a function of λ for three frequencies. The autoregressive data are generated with $\alpha_t = 0.4 \sin 2\pi ft$. The mean-square-error is calculated over 5000 points. Figure 5b plots the same set of curves for $\alpha_t = 0.5 + 0.4 \sin 2\pi ft$.

*See Appendix B for the theoretical derivation.

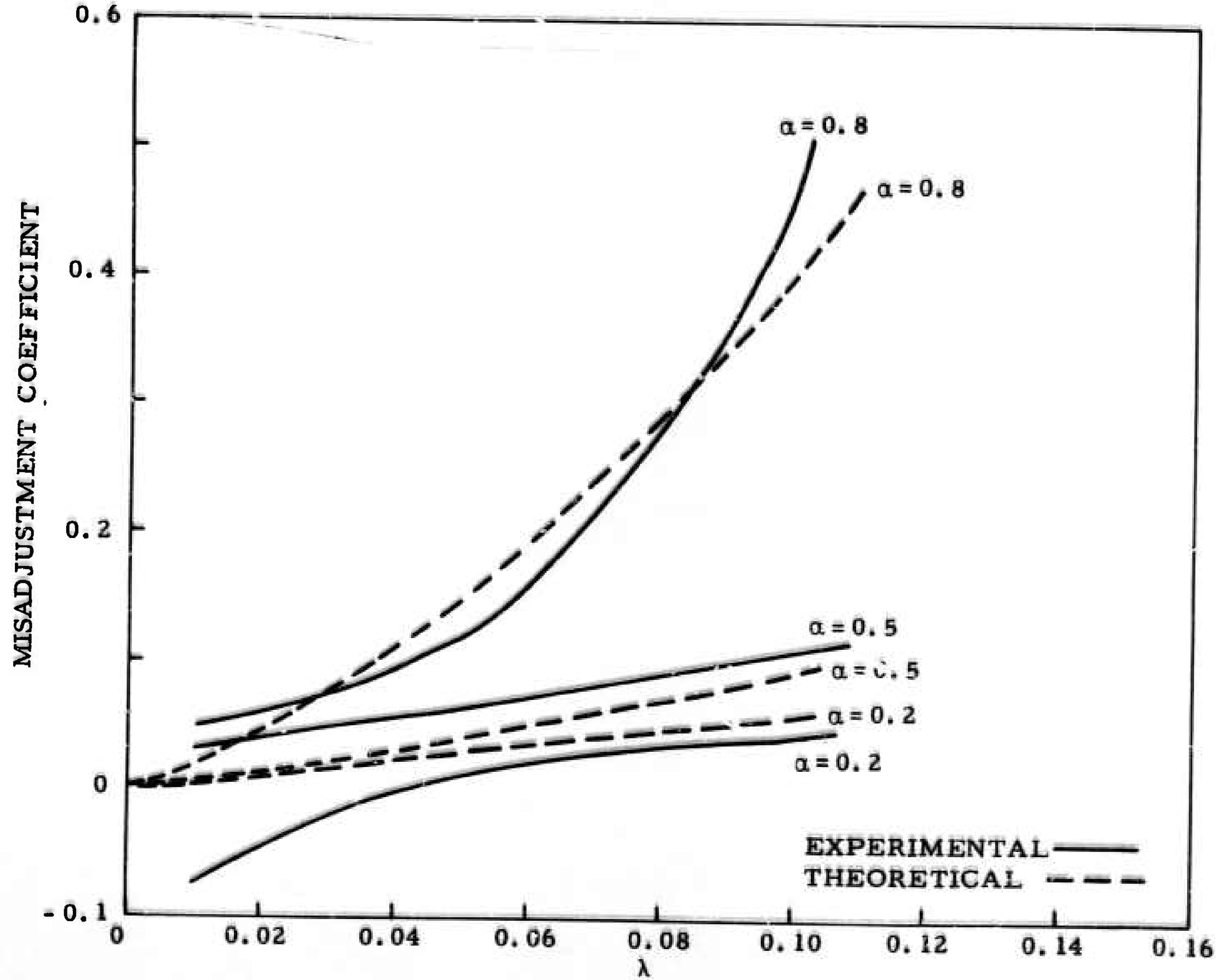

Figure 4. Experimental and Theoretical Misadjustment Coefficients as a Function of λ and α

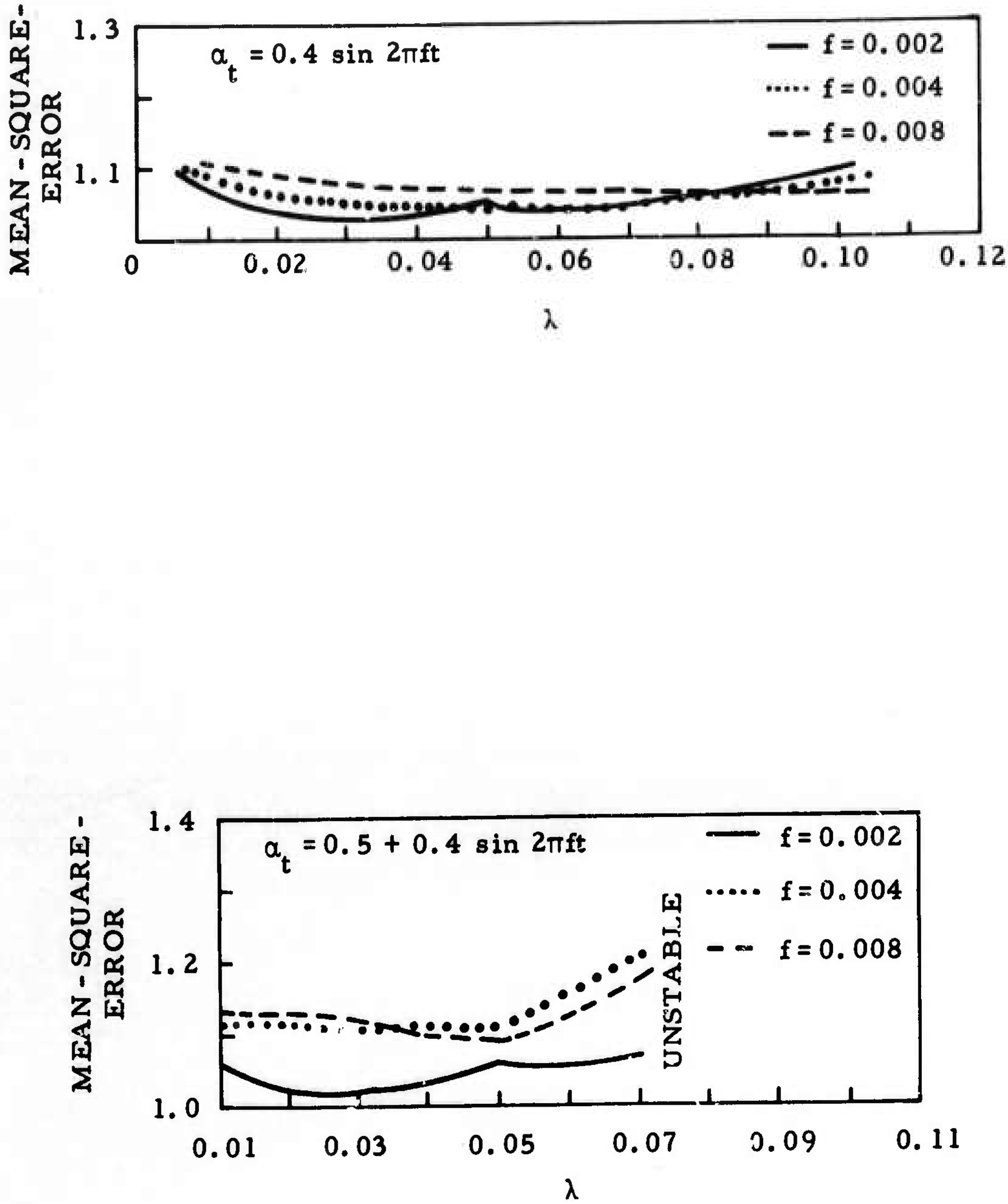

Figure 5. Mean-Square-Error as a Function of λ for Three Frequencies

As might be intuitively expected, the optimum value of λ increases with the increasing time variability of the data (increasing frequency of α_t). Also, on the basis of the results with the stationary data, one would expect the time-varying data with the largest mean value of α (largest average power) to be the most sensitive to changes in λ. This is indicated by the results.

The performance of the adaptive filter appears to be optimized when the filter can adjust to the data at a rate 10 times faster than the rate of change in the data statistics. In other words, the optimum value of λ can be found from the empirical relation

$$10 \left| \frac{d\alpha_t}{dt} \right|_{avg} = \lambda_{opt} \, r_x(0)_{avg} \qquad (9)$$

where $\left| \frac{d\alpha_t}{dt} \right|_{avg} = 4Bf$ is the average rate of change of the autoregressive coefficient,* $r_x(0)_{avg}$ is the average power in the trace, and λ_{opt} is the optimum value of the convergence parameter. In Table 1, the λ_{opt} given by Equation 9 is compared with the true λ_{opt} for each of the six experiments.

Without additional theoretical results, further experimental study of the adaptive algorithm as applied to nonstationary autoregressive time series is hardly merited. It would be of interest to study its behavior, however, if the convergence parameter were allowed to change in some optimum manner as the filter "tracked" the data. Comparison of the results of such an investigation with those presented in this paper would give some indication as to whether this approach would significantly improve the performance of the adaptive filter.

---

*See Appendix C for the derivation.

Table 1

COMPARISON BETWEEN ESTIMATED λ_{opt} AND EXPERIMENTALLY DETERMINED λ_{opt}

A	B	f	$r_x(0)_{avg}$	Estimated λ_{opt}	Experimental λ_{opt}
0.0	0.4	0.002	1.11	0.029	0.03
0.0	0.4	0.004	1.08	0.059	0.06
0.0	0.4	0.008	1.14	0.112	0.1+
0.5	0.4	0.002	1.97	0.016	0.025
0.5	0.4	0.004	2.01	0.032	0.03
0.5	0.4	0.008	2.23	0.057	0.05

SECTION III

CONCLUSIONS

For this report, we have studied an adaption algorithm applied to a simple time-series model.

For the case of stationary data, it is shown that a tradeoff between the adaption rate and the mean-square-error performance of the filter exists and that the adaption rate is inversely proportional to the convergence parameter λ, while the mean-square-error is directly proportional to λ.

For the case of nonstationary data, a tradeoff between adapting too slowly and adapting too rapidly exists. The optimum rate of adaption appears to be approximately 10 times faster than the average time rate of change in the input data statistics.

SECTION IV

REFERENCES AND BIBLIOGRAPHY

1. Wiener, N., 1949: Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications, Wiley & Sons, New York, N.Y.

2. Burg, J.P., 1964, Three-Dimensional Filtering with an Array of Seismometers: Geophysics, Oct., p. 693-713.

3. Bodewig, E., 1956: Matrix Calculus, Interscience Publishers, New York, N.Y., p. 165-179.

4. Widrow, B. and M.E. Hoff Jr., 1960: Adaptive Switching Circuits, WESCON Conv. Rec., Inst. Radio Engnrs., Pt. 4, p. 96-104.

5. Widrow, B., 1966: Adaptive Filters, I-Fundamentals, Tech. Rpt. 6764-6, Systems Theory Lab., Stanford Elec. Lab., Stanford Univ., Stanford, Calif., Dec.

— Widrow, B., 1960: Adaptive Sampled-Data Systems, Proc., 1st International Congress of International Fed. of Automatic Control, Moscow.

6. Lee, Y.W., 1960: Statistical Theory of Communication, Wiley & Sons, New York, N.Y.

— Koford, J.S., 1964: Adaptive Pattern Dichotomization, Tech. Rpt. 6201-1, System Theory Lab., Stanford Electronics Lab., Stanford Univ., Stanford, Calif., Apr.

SECTION IV

REFERENCES AND BIBLIOGRAPHY

1. Wiener, N., 1949: Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications, Wiley & Sons, New York, N.Y.

2. Burg, J.P., 1964, Three-Dimensional Filtering with an Array of Seismometers: Geophysics, Oct., p. 693-713.

3. Bodewig, E., 1956: Matrix Calculus, Interscience Publishers, New York, N.Y., p. 165-179.

4. Widrow, B. and M.E. Hoff Jr., 1960: Adaptive Switching Circuits, WESCON Conv. Rec., Inst. Radio Engnrs., Pt. 4, p. 96-104.

5. Widrow, B., 1966: Adaptive Filters, I-Fundamentals, Tech. Rpt. 6764-6, Systems Theory Lab., Stanford Elec. Lab., Stanford Univ., Stanford, Calif., Dec.

— Widrow, B., 1960: Adaptive Sampled-Data Systems, Proc., 1st International Congress of International Fed. of Automatic Control, Moscow.

6. Lee, Y.W., 1960: Statistical Theory of Communication, Wiley & Sons, New York, N.Y.

— Koford, J.S., 1964: Adaptive Pattern Dichotomization, Tech. Rpt. 6201-1, System Theory Lab., Stanford Electronics Lab., Stanford Univ., Stanford, Calif., Apr.

APPENDIX A
THEORETICAL BEHAVIOR FOR EXPECTED VALUE OF ADAPTIVE FILTER WEIGHT

APPENDIX A

THEORETICAL BEHAVIOR FOR EXPECTED VALUE OF ADAPTIVE FILTER WEIGHT

For a 1-point prediction filter, the adaptive algorithm of Equation 5 becomes

$$a(k+1) = a(k) + \lambda \epsilon_k x_{k-1} \tag{A-1}$$

where

$$\epsilon_k = x_k - a(k)\, x_{k-1} \tag{A-2}$$

Substituting Equation A-2 into Equation A-1 and rearranging terms, we obtain

$$a(k+1) = \left(1 - \lambda x_{k-1}^2\right) a(k) + \lambda x_k x_{k-1} \tag{A-3}$$

The relation between the k + 1 iteration and the initial filter weight a (1) is obtained by recursive substitution into Equation A-3, resulting in

$$\begin{aligned} a(k+1) &= \left(1 - \lambda x_{k-1}^2\right) a(k) + \lambda x_k x_{k-1} \qquad \text{(A-4)} \\ &= \left(1 - \lambda x_{k-1}^2\right)\left[\left(1 - \lambda x_{k-2}^2\right) a(k-1) + \lambda x_{k-1} x_{k-2}\right] \\ &+ \lambda x_k x_{k-1} = \left(1 - \lambda x_{k-1}^2\right)\left(1 - \lambda x_{k-2}^2\right) a(k-1) \\ &+ \lambda x_k x_{k-1} + \lambda \left(1 - \lambda x_{k-1}^2\right) x_{k-1} x_{k-2} \\ &\;\;\vdots \\ &= \left[\prod_{i=1}^{k} \left(1 - \lambda x_{i-1}^2\right)\right] a(1) + \lambda \sum_{i=1}^{k} \left[x_i x_{i-1} \prod_{j=i+1}^{k} \left(1 - \lambda x_{j-1}^2\right)\right] \end{aligned}$$

using the convention $\prod_{i=p}^{k} \left(1 - \lambda x_{i-1}^2\right) = 1$ for $p > k$.

Consider Equation A-4 for the case where λx_{j-1}^2 is small compared with one for all j. We can then approximate a (k + 1) by

$$a(k+1) \approx \left[\prod_{i=1}^{k} e^{-\lambda x_{i-1}^2}\right] a(1) + \lambda \sum_{i=1}^{k} \left[x_i\, x_{i-1} \prod_{j=i+1}^{k} e^{-\lambda x_{j-1}^2}\right] \quad \text{(A-5)}$$

$$= \left[\prod_{i=1}^{k} e^{-\lambda x_{i-1}^2}\right] a\ (1) + \sum_{i=1}^{k} \left[\left(1 - e^{-\lambda x_i\, x_{i-1}}\right) \prod_{j=i+1}^{k} e^{-\lambda x_{j-1}^2}\right]$$

The expected value of a (k + 1) is approximated by

$$E\left[a\ (k+1)\right] \approx \left[\prod_{i=1}^{k} e^{-\lambda r_x(0)}\right] a\ (1) + \sum_{i=1}^{k} \left[\left(1 - e^{-\lambda \alpha r_x(0)}\right) \prod_{j=i+1}^{k} e^{-\lambda r_x(0)}\right] \quad \text{(A-6)}$$

assuming that the data are stationary where $r_x(0)$ is the zero-lag correlation of the data and $\alpha r_x(0)$ is the first-lag correlation value.

Simplifying Equation A-6, we obtain

$$E\left[a(k+1)\right] \approx a(1)\, e^{-\lambda k\, r_x(0)} + \frac{1 - e^{-\lambda \alpha r_x(0)}}{1 - e^{-\lambda r_x(0)}} \left[1 - e^{-\lambda k\, r_x(0)}\right]$$

$$\approx a(1)\, e^{-\lambda k\, r_x(0)} + \alpha \left[1 - e^{-\lambda k\, r_x(0)}\right]$$

This expression shows that, on the average, the initial adaption transient should decay with a time constant of $\frac{1}{\lambda r_x(0)}$ per iteration. Also there is the result that, in the limit as k approaches infinity, the expected value of our filter weight is the optimum value α; i.e.,

$$\lim_{k \to \infty} E\left[a(k+1)\right] = \alpha$$

APPENDIX B
DERIVATION OF MISADJUSTMENT COEFFICIENT FOR 1-POINT ADAPTIVE FILTER

APPENDIX B

DERIVATION OF MISADJUSTMENT COEFFICIENT FOR 1-POINT ADAPTIVE FILTER

In the text, we define the misadjustment of an adaptive filter as

$$M = \frac{MSE_a - MSE_o}{MSE_o} \qquad \text{(B-1)}$$

where

MSE_a = mean-square-error for the output of the adaptive filter

MSE_o = mean-square-error for the output of the classical time-invariant filter

We now evaluate this quantity on the basis of theoretical considerations for the 1-point prediction filter.

The error term of the adaptive filter for the k^{th} output is given by

$$\epsilon_k = x_k - a(k)\, x_{k-1}$$

$$= x_k - \alpha x_{k-1} - \Delta_k x_{k-1} = N_k - \Delta_k x_{k-1}$$

where Δ_k is the difference between the adaptive filter weight and the optimum time-invariant filter weight. The expected value for the mean-square-error for the k^{th} output is

$$E\left[\epsilon_k^2\right] = E\left[\left(x_k - \alpha x_{k-1}\right)^2 - 2\Delta_k x_{k-1}\left(x_k - \alpha x_{k-1}\right) + \Delta_k^2 x_{k-1}^2\right] \qquad \text{(B-2)}$$

$$= MSE_o + E\left[\Delta_k^2 x_{k-1}^2\right]$$

To evaluate the second term in Equation B-2, we can obtain, by recursive substitution,

$$\Delta_k = \Delta_{k-1} + \lambda \epsilon_{k-1} x_{k-2} = \Delta_{k-1}\left(1 - \lambda x_{k-2}^2\right) + \lambda N_{k-1} x_{k-2}$$

$$= \Delta_1 \prod_{i=2}^{k} \left(1 - \lambda x_{i-2}^2\right) + \lambda \sum_{i=3}^{k} x_{i-2}\left(x_{i-1} - \alpha x_{i-2}\right) \prod_{j=i+1}^{k} \left(1 - \lambda x_{j-2}^2\right)$$

where

$$\epsilon_{k-1} = x_{k-1} - \alpha(k-1)\, x_{k-2} = \left(x_{k-1} - \alpha x_{k-2}\right) - \Delta_{k-1}\, x_{k-2}$$

We simplify this expression by considering the steady-state behavior of Δ_k (or, $\Delta_1 = 0$). Thus,

$$\Delta_k = \lambda \sum_{i=3}^{k} x_{i-2}\, N_{i-1} \prod_{j=i+1}^{k} \left(1 - \lambda x_{j-2}^2\right) \qquad \text{(B-3)}$$

where N_{i-1} is as defined in the text. We then have

$$E\left[\Delta_k^2\, x_{k-1}^2\right] = \lambda^2\, E\left[\sum_{i=3}^{k} \sum_{m=3}^{k} N_{i-1}\, N_{m-1}\, x_{i-2}\, x_{m-2}\, x_{k-1}^2 \prod_{j=i+1}^{k} \prod_{n=m+1}^{k} \left(1 - \lambda x_{j-2}^2\right)\left(1 - \lambda x_{n-2}^2\right)\right]$$

$$= \lambda^2 \sum_{i=3}^{k} \sum_{m=3}^{k} E\left[N_{i-1}\, N_{m-1}\, x_{i-2}\, x_{m-2}\, x_{k-1}^2 \prod_{j=i+1}^{k} \prod_{n=m+1}^{k} \left(1 - \lambda x_{j-2}^2\right)\left(1 - \lambda x_{n-2}^2\right)\right]$$

For small λ, we can treat terms like $1 - \lambda x_t^2$ as being independent of the other random variables. Therefore,

$$E\left[\Delta_k^2 \, x_{k-1}^2\right] \approx \lambda^2 \sum_{i=3}^{k} \sum_{m=3}^{k} \left\{ E\left[N_{i-1} \, N_{m-1} \, x_{i-2} \, x_{m-2} \, x_{k-1}^2\right] \left(1 - \lambda r_x(0)\right)^{k-i} \left(1 - \lambda r_x(0)\right)^{k-m} \right\}$$

where $r_x(0)$ is the average power in the data.

Evaluating this expression, we obtain

$$E\left[\Delta_k^2 \, x_{k-1}^2\right] \approx \left[\frac{MSE_o \, r_x(0)}{2 - \lambda r_x(0)}\right] \lambda + \left[\frac{4\alpha^2 MSE_o \, r_x^2(0) \left(1 - \alpha^2\right)}{\left[1 - \alpha^2 \left(1 - \lambda r_x^2(0)\right)\right] \left[1 - \alpha^2 \left(1 - \lambda r_x(0)\right)\right]} \lambda^2\right]$$

The misadjustment for the 1-point adaptive filter is then given by

$$M \approx \frac{\lambda r_x(0)}{2 - \lambda r_x(0)} + \frac{4\alpha^2 \, r_x^2(0) \left(1 - \alpha^2\right) \lambda^2}{\left[1 - \alpha^2 \left(1 - \lambda r_x^2(0)\right)\right] \left[1 - \alpha^2 \left(1 - \lambda r_x(0)\right)\right]} \qquad \text{(B-4)}$$

APPENDIX C
DERIVATION OF AVERAGE RATE OF CHANGE FOR TIME-VARYING AUTOREGRESSIVE COEFFICIENT α_t

APPENDIX C

DERIVATION OF AVERAGE RATE OF CHANGE FOR TIME-VARYING AUTOREGRESSIVE COEFFICIENT α_t

In the text, we define the time-varying autoregressive coefficient to be of the form

$$\alpha_t = A + B \sin 2\pi ft \tag{C-1}$$

where A is the mean value of α_t, B is the amplitude of the sinusoidal deviation about the mean, and f is the frequency of the deviation. The average rate of change of α_t is given

$$\left|\frac{d\alpha_t}{dt}\right|_{avg} = 2\pi fB \quad |\sin 2\pi ft|_{avg} \tag{C-2}$$

where $| \quad |_{avg}$ indicates the average absolute value. Clearly,

$$|\sin 2\pi ft|_{avg} = \frac{2}{\pi} \int_0^{\pi/2} \sin x dx = \frac{2}{\pi} \tag{C-3}$$

Substituting the above value into Equation C-2, we obtain the result stated in the text:

$$\left|\frac{d\alpha_t}{dt}\right|_{avg} = 4Bf \tag{C-4}$$

APPENDIX D

DERIVATION OF "EQUIVALENT" BOXCAR LENGTH N WITH EXPONENTIAL TIME CONSTANT τ

APPENDIX D

DERIVATION OF "EQUIVALENT" BOXCAR LENGTH N WITH EXPONENTIAL TIME CONSTANT τ

In the text, we define the exponential weighting scheme as

$$\underline{\beta}_k = \frac{\sum_{i=1}^{\infty} e^{-i/\tau} \underline{a}'_{k-i}}{\sum_{i=1}^{\infty} e^{-i/\tau}} = \left(1 - e^{-1/\tau}\right) \sum_{i=1}^{\infty} e^{-i/\tau} \underline{a}'_{k-i}$$

The variance of the random variable $\underline{\beta}_k$ is given by

$$\mathrm{var}\left[\underline{\beta}_k\right] = E\left[\left(\underline{\beta}_k - E\left[\underline{\beta}_k\right]\right)^2\right]$$

$$= \left(1 - e^{-1/\tau}\right)^2 E\left[\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} e^{-i/\tau} e^{-j/\tau} \left(\underline{a}'_{k-i} - E\left[\underline{a}'_{k-i}\right]\right)\left(\underline{a}'_{k-j} - E\left[\underline{a}'_{k-j}\right]\right)\right]$$

If the $\underline{a}'_{k-i}$ are uncorrelated time-stationary random variables, then

$$E\left[\left(\underline{\beta}_k - E\left[\underline{\beta}_k\right]\right)^2\right]$$

$$= \left(1 - e^{-1/\tau}\right)^2 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} e^{-i/\tau} e^{-j/\tau} \delta_{ij} E\left[\left(\underline{a}'_{k-i} - E\left[\underline{a}'_{k-j}\right]\right)^2\right]$$

$$= \left(1 - e^{-1/\tau}\right)^2 \sum_{i=1}^{\infty} e^{-2i/\tau} E\left[\left(\underline{a}'_{k-i} - E\left[\underline{a}'_{k-i}\right]\right)^2\right]$$

$$= \left(1 - e^{-1/\tau}\right)^2 \frac{1}{1 - e^{-2/\tau}} \mathrm{var}\left[\underline{a}'_k\right]$$

$$= \frac{1 - e^{-1/\tau}}{1 + e^{-1/\tau}} \mathrm{var}\left[\underline{a}'_k\right]$$

It is a well-known result that the variance of the average of N random variables, each with the same variance, is given by[6]

$$\sigma_N^2 = \frac{1}{N} \sigma_\epsilon^2$$

Making the identification

$$\frac{1}{N} = \frac{1 - e^{-1/\tau}}{1 + e^{-1/\tau}}$$

we have the result that the time constant τ may be associated with an "equivalent" boxcar section of data of length

$$N = \frac{1 + e^{-1/\tau}}{1 - e^{-1/\tau}}$$

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Texas Instruments Incorporated Science Serivces Division P.O. Box 5621, Dallas, Texas 75222	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP —

3. REPORT TITLE

ADVANCED ARRAY RESEARCH-SPECIAL REPORT NO. 4
STUDY OF A 1-POINT ADAPTIVE FILTER

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Special

5. AUTHOR(S) (First name, middle initial, last name)

Brown, James E. Booker, Aaron H.
Burg, John P.

6. REPORT DATE 30 September 1967	7a. TOTAL NO. OF PAGES 33	7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. Contract F 33657-67-C-0708-P001 b. PROJECT NO. VELA T/7701 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) — 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) —	

10. DISTRIBUTION STATEMENT

This document is subject to special export controls, and each transmittal to foreign governments or foreign nationalsmay be made only with prior approval of Chief, AFTAC.

11. SUPPLEMENTARY NOTES ARPA Order No. 624 ARPA Program Code No. 7 F10	12. SPONSORING MILITARY ACTIVITY Advanced Research Projects Agency Department of Defense The Pentagon, Washington, D.C. 20301

13. ABSTRACT

The design and evaluation of optimum, time-invariant filters (the classical approach to prediction problems) are subject to several limitations. Because of these limitations, an adaptive system was considered. This report describes a filtering system which can adjust to changes in either signal or noise, thereby overcoming many difficulties of time-invariant filtering. The small amount of required computational time to update the filter weights is another important feature of the scheme presented.

In this report, an adaption algorithm applied to a simple time-series model was studied. For the case of stationary data, a tradeoff between the adaption rate and the mean-square-error performance of the filter exists. The adaption rate is inversely proportional to the convergence parameter λ, while the mean-square-error is directly proportional to λ. For the case of non-stationary data, a tradeoff between adapting too slowly and adapting too rapidly exists. The optimum rate of adaption appears to be approximately 10 times faster than the average time rate of change in the input data statistics.

lambda

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Advanced Array Research						
1-Point Adaptive Filter						

UNCLASSIFIED
Security Classification