Spring 1965

Optimum linear and adaptive polynomial smoothers

Stanley Bruce Alterman
New Jersey Institute of Technology

65-11,673

ALTERMAN, Stanley Bruce, 1937–
OPTIMUM LINEAR AND ADAPTIVE POLYNOMIAL SMOOTHERS.

Newark College of Engineering, D.Eng.Sc., 1965
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan

OPTIMUM LINEAR AND ADAPTIVE POLYNOMIAL SMOOTHERS

BY

RUCE
STANLEY B. ALTERMAN

A THESIS

PRESENTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF ENGINEERING SCIENCE IN ELECTRICAL ENGINEERING

AT

NEWARK COLLEGE OF ENGINEERING

Newark, New Jersey

1965

ABSTRACT

The design of optimum polynomial digital data smoothers (filters) is considered for linear and adaptive processing systems. It is shown that a significant improvement in performance can be obtained by using linear smoothers that take into account known a priori constraints or distributions of the input signal. The procedure for designing optimum (minimum mean square error) adaptive polynomial data smoothers is then discussed and analyzed. The optimum smoother makes use of a priori signal statistics combined with an adaptive Bayesian weighting of a bank of conditionally optimum smoothers. Use of this technique permits large improvements in performance with a minimum of additional system complexity.

Stanley B. Alterman

Doctor of Engineering Science, Electrical Engineering,

June 1965

"Optimum Linear and Adaptive Polynomial Smoothers"

Dr. J. Padalino ____________________

Dr. P. Fox ____________________

APPROVAL OF THESIS

FOR

DEPARTMENT OF ELECTRICAL ENGINEERING

NEWARK COLLEGE OF ENGINEERING

BY

FACULTY COMMITTEE

APPROVED: ______________________
ADVISOR

____

____

____

NEWARK, NEW JERSEY

1965

DEDICATION

To my mother and father for the years of unselfish encouragement during my education.

To my wife Enid for her patience and understanding throughout the course of my graduate studies.

To my children, Betsy-Jo and Eric, for the pleasurable and relaxing interludes they provided during the preparation of this thesis.

ACKNOWLEDGMENTS

The author is indebted to Dr. Phyllis Fox of the faculty of Newark College of Engineering for her helpful suggestions and encouragement.

The financial assistance of Bell Telephone Laboratories is gratefully acknowledged.

<u>TABLE OF CONTENTS</u>

CHAPTER	TITLE	PAGE
1.0	INTRODUCTION	1
2.0	APPROACH TO THE PROBLEM	2
3.0	REVIEW OF PERTINENT PRIOR WORK	4
4.0	CLASSICAL POLYNOMIAL SMOOTHERS	8
5.0	OPTIMUM POLYNOMIAL SMOOTHERS USING A PRIORI INFORMATION	19
6.0	EXAMPLES AND SAMPLE RESULTS OF OPTIMUM POLYNOMIAL SMOOTHERS	30
	6.1 Optimum Velocity Estimate with Known Constraint on Input Acceleration	30
	6.2 Optimum Acceleration Estimate with Known Constraint on Input Acceleration	36
	6.3 Optimum Estimate of a Constant when True Value is a Sample of a Random Variable with Known Statistics	41
7.0	OPTIMUM ADAPTIVE FILTER DESIGN FOR INCOMPLETELY SPECIFIED SIGNALS	44
8.0	GENERATION OF THE OPTIMUM ADAPTIVE SUBFILTER WEIGHTS	50
9.0	EXAMPLES OF OPTIMUM ADAPTIVE FILTERING	52
	9.1 Adaptive Estimate of a Constant Signal	53
	9.2 Adaptive Estimation when A Priori Statistics are not Completely Known	66
10.0	CONCLUSIONS	79
	10.1 Discussion of Results	79
	10.2 Suggestions for Future Work	80

APPENDIX I. POLYNOMIAL SMOOTHERS

Ia. Derivation of Polynomial Smoothers

Ib. Properties of Polynomial Smoothers

APPENDIX II. PROOF OF SOME IMPORTANT THEOREMS

APPENDIX III. DERIVATION OF $W^1(t)$ POLYNOMIALS

APPENDIX IV. DERIVATION OF SUBFILTER WEIGHTS

APPENDIX V. DERIVATION OF STATISTICS FOR EXAMPLE IN SECTION 9.1

APPENDIX VI. OPTIMUM UNBIASED POLYNOMIAL SMOOTHERS

BIBLIOGRAPHY

LIST OF FIGURES AND TABLES

FIGURE	TITLE	PAGE
1	Optimum Velocity Estimate	37
2	Optimum Acceleration Estimate	40
3	Optimum Estimate of a Constant	43
4	Optimum Adaptive Smoother	47
5	Switched Adaptive Smoother	49
6	Adaptive Velocity Estimate	76
7	Suboptimum Adaptive Velocity Estimate	78
8	Confidence Limits	IV-5
9	Confidence Limits	IV-7

TABLE	TITLE	PAGE
1	Orthogonal Polynomials	I-3

LIST OF SYMBOLS

a_j	- Taylor series coefficients of input signal
A_j^k	- coefficients of j^{th} power of t (t^j) for k^{th} orthogonal polynomial, $F_k(t)$
b_j	- input signal orthonormal polynomial coefficients (related to Taylor series coefficients)
b_j^*	- estimates of b_j
B_k	- normalization coefficients for discrete orthogonal coefficients, $(F_k, F_k) = B_k$
c_{mj}	- $\left.\dfrac{d^m t^j}{dt^m}\right\vert_{t=\tau}$
D_m	- dynamic error for estimate of m^{th} derivative of input signal
D_{mJ}	- dynamic error for estimate of m^{th} derivative of input signal using a J^{th} order polynomial smoother
ε_{mj}	- dynamic error coefficients of filter estimating m^{th} derivative.
E[ ]	- expected value operator
$f_j = f_j(x) = f_j(n) = f_j(t_i/\Delta t)$	- discrete orthonormal polynomials

$F_j = F_j(x) = F_j(n)$ $= F_j(t_i/\Delta t)$	- discrete orthogonal polynomials
$f_j^{(m)}(x)$	- m^{th} derivative of discrete orthonormal polynomials
H_j^k	- coefficient of j^{th} power of t for k^{th} orthonormal polynomial, $f_k(t)$
J	- degree (order) of smoother or degree of input polynomial signal
K	- degree of input polynomial signal
m_j	- mean value of j^{th} Taylor series coefficient
mse	- mean-square error
$n(t_i) = n(i)$	- samples of input noise
N	- instantaneous value of noise at smoother output
$p(a_j)$	- probability density function of a_j
P(x)	- probability of x occurring
r	- number of data points in T
R	- square of residual error $= \left[\hat{x}(i) - y(i)\right]^2$
T	- smoothing time (seconds)
t	- time
τ	- time at which estimate is obtained
u_j	- j^{th} moment of weighting sequence, W(t) $(t^j,W) = u_j$
$U(t) = U$	- filter weighting sequence (for least squares smoother)

$W(t_i) = W(i) = W$	- filter weighting sequence
$W^j(t) = W^j$	- polynomial set defined by its moments $(t^k, W^j) = 0$ if $k \neq j$; $= 1$ if $k = j$
$W_J(i)$	- J^{th} order smoother weighting sequence
$W_J^{(m)}(i)$	- J^{th} order smoother weighting sequence for m^{th} derivative estimate
$\hat{x}(\tau) = x^*(\tau)$	- estimate of position at $t = \tau$
$x(t_i) = x(i)$	- samples of input signal
$x^{(m)}(\tau)$	- m^{th} derivative of input signal at $t = \tau$
$x^{*(m)}(\tau) = \hat{x}^{(m)}(\tau)$	- estimate of the m^{th} derivative of input signal at $t = \tau$
$\hat{x}(\alpha_1)$	- optimum estimate of x, assuming α_1 is true value of input signal parameter
$y(t_i) = y(i)$	- samples of input signal plus noise
α_1	- unknown signal parameter
σ^2_{mJ}	- variance of estimate of the m^{th} derivative of input signal using J^{th} order least squares polynomial smoother

σ_j^2	- variance of j^{th} Taylor series coefficient
σ_o^2	- variance of input noise
Δt	- spacing between data samples (seconds)
$\left(x(i),\ y(i)\right)$	- $\sum_i x(i)\ y(i)$
$\lVert x(i) \rVert$	- $\sqrt{\left(x(i),\ x(i)\right)}$
$\sum_i R = \Sigma R$	- sum of squared residual errors

1.0 INTRODUCTION

Polynomial data smoothers, because of their ease of handling, and their attractive properties, have had widespread usage in a variety of data processing applications. Heretofore, the design of these smoothers has been based only upon the minimum required a priori information necessary for their design, namely the degree of the input signal. More often than not, however, additional a priori information is available to the smoother designer. Moreover, during the course of the actual data processing, additional information about the input signal can become available. It is the purpose of this dissertation to consider the design of optimum polynomial filters either when a priori information is available to the designer or when information becomes available during the course of the data processing. The latter procedure gives rise to what is commonly called a self-adaptive processing system.

2.0 APPROACH TO THE PROBLEM

Information available to the smoother designer concerning certain characteristics of the input signal is often not effectively used. This information might be of the form, for example, that the maximum acceleration of an object, due to mechanical constraints, is equal to 100 ft/sec^2, or even less restrictively, that its velocity is less than the speed of light. We might also know that due to uncertainty in rocket design the actual acceleration can be described by some probability distribution with known mean and variance. The use of the above type of information, which is often available, can improve the estimation accuracy if properly used. This problem is considered in Chapter 5.0 where the optimum (minimum mean square error) polynomial filter design is presented when known constraints on the input signal derivatives are available. In Chapter 6.0, examples are given illustrating the improvement obtainable as compared with normal polynomial filters.*

In addition to a priori information available to the smoother designer, initial processing of the data yields information which can be used for further

*Normal polynomial filters are briefly discussed in Chapter 4.0, with complete details shown in Appendices I and II.

processing. This "learning" feature of a device is referred to as self-adaptation. In Chapters 7.0 and 8.0 the optimum adaptive filter is shown to be composed of a weighted sum of subfilters. Specific illustrations of these techniques are presented in Chapter 9.0 for several discrete, finite memory smoothers. Included are some sample results showing the improvements obtainable by using adaptive techniques.

3.0 REVIEW OF PERTINENT PRIOR WORK

The importance and usefulness of optimum filtering and prediction in our modern electronic systems environment has been clearly demonstrated during the past decade. Present-day theories of smoothing and prediction may be said to have originated with the classic papers of Wiener[31] and Kolmogoroff[23], which were written during World War II. In fact much of the research on prediction and filtering has been concerned with various extensions of the Wiener Theory. In his pioneering work, Wiener[31] showed that problems of prediction of random signals and detection of signals of known form in the presence of random noise lead to the so-called Wiener-Hopf integral equation. He gave a method (spectral factorization) for the solution of this integral equation in the practically important special case of stationary statistics and rational spectra. The method involves performing a least squares operation on observations assumed available for all past times. Under certain conditions, optimality under the least squares error criterion implies optimality under a wide class of criteria. Such conditions have been found by Benedict and Sondhi[4] and independently by Sherman[28]. Zadeh and Ragazzini[32,33] modified Wiener theory for observations which are available only over a past interval of finite duration. They also developed the useful and reasonable approach of specifying that the nonrandom part of the signal is known to be a polynomial in time of degree not greater than some fixed integer. Their continuous finite memory filters have led to many other variations on this basic approach.

Many authors, including Johnson[19], Lees[25], Bergen[6], Darlington[13], Blum[8-11], Franklin[15,16], and Alterman[1] have discussed discrete, finite-memory polynomial filters. Blum uses an orthogonal polynomial signal representation and develops a recursive discrete filter. Blackman[7] developed the technique of cascaded simple sums smoothing as a substitute for optimum smoothing using a digital computer for prediction of sampled data. Cascaded simple sums smoothers obtain estimates of the derivatives of the input signal by averaging compound differences of the noisy input data. The advantage of cascaded simple sums is the elimination of most of the multiplications and many of the arithmetic operations required in the optimum convolution type smoother. Howard and Rauch[18] consider the design of an optimum polynomial filter with a simple a priori constraint using a minimax error criterion.

One of the newest and most promising areas of investigation in prediction theory is the concept of adaptive systems. Normally, the data required by optimum filter theory are unknown a priori, so that a filter in the environment must "learn" or "adapt" to these as necessary. Adaptivity has not been precisely defined to date and many of the adaptive systems are based on parametric methods. Most of the work in this field has not been concerned with prediction and filtering as such but with control system design. There has been some limited work in the adaptive system area applied directly to filter and prediction theory. Some of the earliest work was done by Benner and Drenick[5], who were filtering a signal which could either be a ramp or a parabola in the presence of additive, zero mean, Gaussian noise. Their filter chose between two linear subfilters on the basis of an estimate of the derivative of the signal part of the input. Franklin[17] attempted to improve on Benner's work by using the optimum ramp and parabola filters as the subfilters and choosing between them in a manner which would minimize the mean square error. Shaw[29] also considers a switching two-mode filter, but rather than designing by successive optimization, he sets up a design procedure for simultaneously designing the subfilters and the switching decision rule to minimize mean square error.

Kushner[24] and Sakrison[27] have applied stochastic approximation theory to estimation problems with unspecified noise. Follin and Bucy[14] consider an adaptive scheme for a specialized case where the signal-to-noise ratio is an unknown parameter. Weaver[30] examined a linear parametric adjustment system. Balakrishnan[3] considers a nonparametric method applied to pure prediction, where no noise is assumed and no statistical assumptions made.

Magill[24] describes an adaptive approach to the problem of estimating a scalar-valued, stochastic process described by an initially unknown parameter vector. His solution is limited to those processes whose parameter vector comes from a finite set of a priori known values.

Alterman[2] describes a digital smoother technique where the known form of the differential equation for the input signal is used in conjunction with least squares polynomial smoothing to obtain an optimum finite-memory digital filter.

This paper is essentially concerned with the extension of the above work in two directions: 1) the design of optimum polynomial discrete filters when knowledge of a priori statistics and/or constraints on parameters which describe the input signal are available and 2) the design of adaptive, finite-memory digital filters when the input signal is described by one of a finite number of ranges and/or values of an unknown parameter.

4.0 CLASSICAL POLYNOMIAL SMOOTHERS

As an introduction to the problem under consideration, some of the elementary concepts and results of classical polynomial smoothing theory are presented in this chapter. Detailed discussion and derivations are included in Appendices I and II.

By classical polynomial smoothers, we mean those smoothers which are designed under the assumption that the deterministic portion of a signal is known to be a polynomial in time with degree less than or equal to a fixed integer, J.

Consider a sampled signal, $x(t_i)$, which is disturbed by noise, $n(t_i)$, such that

$$y(t_i) = x(t_i) + n(t_i) \tag{4-1}$$

where $y(t_i)$ represent noisy measurements of $x(t_i)$. We assume that samples of the input signal are obtained at uniform time intervals of Δt seconds and that $n(t_i)$ is a zero mean random variable with variance equal to σ_o^2. Further, we assume that $n(t_i)$ are independent from sample to sample (i.e. that the autocorrelation function of $n(t_i)$ is given as, $R\left(n(t_i), n(t_\ell)\right) = \sigma_o^2\delta_{i\ell}$ where $\delta_{i\ell}$ is the Kronecker delta symbol.

We are interested in determining the optimum finite memory discrete linear filter to estimate the function, $x(t_i)$, or any of its derivatives, $x^{(m)}(t_i)$, given a finite number, r, of noisy samples, $y(t_i)$, extending over a smoothing (or filtering) time, T, where $T = (r-1)\Delta t$. We say a filter is discrete linear if the transformation of a finite input sequence of numbers,

$$\ldots\ y(t_{-2}),\ y(t_{-1}),\ y(t_0),\ y(t_{+1}),\ldots$$

into a finite output sequence of estimates,

$$\ldots\ x^{*(m)}(t_{-2}),\ x^{*(m)}(t_{-1}),\ x^{*(m)}(t_0),\ x^{*(m)}(t_{+1}),\ldots$$

is given by

$$x^{*(m)}(t_i) = \sum_{i=-\frac{r-1}{2}}^{\frac{r-1}{2}} W(i)\ y(t_i)\ ^{\dagger} \qquad (4\text{-}2)$$

Hence $x^{*(m)}(t_i)$ is a linear combination (weighted average) of the input sequence, $y(t_i)$. The sequence of weights, $W(i)$, is called the weighting sequence or impulse response of the filter.

---

†The symmetrical representation will be used throughout this development for simplicity. The index $i=0$ represents the center of the smoothing interval and $i = -\frac{r-1}{2}$ and $i = \frac{r-1}{2}$ the oldest and latest points respectively.

We shall be concerned with the question of estimating the m^{th} derivative of the input signal $\left[x^{(m)}(t_1)\right]$ at some time, $t = \tau$, for the case in which $x(t)$ is either a polynomial of known degree, J, or can be approximated by a polynomial over the time interval, T.

The polynomial filter weighting function can be derived using various optimality criteria, each of which leads to identical results for the conditions described above. These criteria include least sum of squares error curve fitting, unbiased estimation and minimum variance estimation. These results can be obtained in an extremely useful form by using an orthonormal polynomial expansion signal representation rather than the usual Taylor's series approach. Let the input signal be given by

$$x(t_1) = \sum_{j=0}^{J} b_j \; f_j(t_1/\Delta t) \qquad (4\text{-}3)$$

where the b_j are coefficients related to the Taylor's series coefficients[1,2] and the $f_j(t_1)$ are certain orthonormal polynomials described in Appendix I. Let $x^*(\tau)$ be the estimate of $x(t_1)$ at time $t = \tau$ where

$$x^*(\tau) = \sum_{j=0}^{J} b_j^* \; f_j(\tau/\Delta t) \qquad (4\text{-}4)$$

in which J is called the smoother order or power and the b_j^* coefficients are to be determined. To satisfy the "least squares" error criterion, the expression for the sum of squared errors given by

$$\text{Sum of squared errors} = \Sigma R = \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} \left[y(t_i) - x^*(t_i) \right]^2 \tag{4-5}$$

is minimized with respect to the coefficients, b_j^*. This yields

$$b_k^* = \sum_i y(t_i)\, f_k(t_i/\Delta t) \tag{4-6}$$

Consequently from (4-4) the J^{th} order smoother weighting sequence for the estimate of $x^*(\tau)$, $W_J(i)$, is

$$W_J(i) = \sum_{j=0}^{J} f_j(t_i/\Delta t)\, f_j(\tau/\Delta t) \tag{4-7}$$

From Appendix I, the weighting function for the m^{th} derivative estimate of a J^{th} order smoother at time $t = \tau$ is given as,

$$W_J^{(m)}(i) = \sum_{j=0}^{J} f_j(t_i/\Delta t)\; f_j^{(m)}(\tau/\Delta t) \qquad (4\text{-}8)$$

This result gives an explicit formula for the filter weighting function (impulse response) as a function of all system parameters where

J = smoother order

m = estimated derivative

Δt = data spacing

τ = time at which estimate is obtained

$f_j(n)$ = discrete, equally spaced, finite time, orthonormal polynomials (Appendix Ia)

index i = 0 refers to center of smoothing interval

$i = -\frac{(r-1)}{2}$ refers to oldest data point

$i = \frac{r-1}{2}$ refers to latest data point

Aside from the curve fitting properties of polynomial smoothers, as illustrated by the method of derivation in Appendix I(a), polynomial filters have other desirable properties which we now consider.

As shown in Appendix I(b), the expected value of the estimate of the m^{th} derivative is given as

$$E\left[x^{*(m)}(\tau)\right] = \sum_{j=0}^{J} b_k\; f_k^{(m)}(\tau/\Delta t) = x^{(m)}(\tau) \qquad (4\text{-}9)$$

if $x(t)$ is a polynomial of degree K, which is equal or less than J. Consider the situation where the input signal is a polynomial of degree K, where $K > J$, the smoother order, or

$$x^{(m)}(t) = \sum_{j=0}^{K} b_j \, f_j^{(m)}(t) \tag{4-10}$$

Under these conditions dynamic or bias error is introduced into the estimate. Define dynamic error in the estimate of the m^{th} derivative, D_m as

$$D_m = E\left\{x^{*(m)}(\tau) - x^{(m)}(\tau)\right\} \tag{4-11}$$

It is shown in Appendix I that,

$$D_m = -\sum_{j=J+1}^{K} b_j \, f_j^{(m)}(\tau) \tag{4-12}$$

From equation (4-12), we note that the b_j coefficients for values of j from J+1 to K determine the magnitude of the dynamic errors. Further, these values of b_j are proportional[1] to the values of the $J+1^{st}$ to K^{th} derivatives of the input function at the center of the smoothing interval [i.e., Taylor's series coefficients]. Hence, polynomial smoother estimates are unbiased estimates for all derivatives if the input signal order is less than or equal to J, the smoother order.

The Taylor's series expansion of the input signal, $x(t)$, about $t = 0$ is given as,

$$x(t) = \sum_{j=0}^{K} a_j \, t^j \tag{4-13}$$

In general, D_m can be written in terms of the Taylor's series coefficients,

$$D_m = \sum_{j=0}^{K} a_j \, \varepsilon_{mj} \tag{4-14}$$

where the ε_{mj} are called dynamic error coefficients. Specifically, equation (4-12) can be written as

$$D_m = \sum_{j=J+1}^{K} a_j \, \varepsilon_{mj}$$

The ε_{mj} coefficients may be obtained by noting that for an input signal described by equation (4-13), the total dynamic error is equal to the sum [since we have a linear filter] of the dynamic errors associated with each of the terms of equation (4-13). Hence from (4-14), our definition of dynamic error coefficients, we note that ε_{mj} is simply the dynamic error in estimating the m^{th} derivative of the input for an input equal to t^j, which is

$$\varepsilon_{mJ} = \left(t^J,\ W_J^{(m)}\right) - \left.\frac{d^m t^J}{dt^m}\right|_{t=\tau} \qquad (4\text{-}15)\ ^{\dagger}$$

Also note that $\left(t^J,\ W_J^{(m)}\right)$ are the moments of the filter weighting function, $W_J^{(m)}(i)$.

We now concern ourselves with the effect of polynomial filters on the input noise; in particular, we desire some measure of the output noise associated with a particular estimate, $x^{*(m)}(\tau)$, using a J^{th} order smoother. From Appendix I we obtain for the variance, σ_{mJ}^2, of the estimate of the m^{th} derivative using a J^{th} smoother.

$$\left.\begin{aligned} \sigma_{mJ}^2 &= \sigma_o^2 \left\| W_J^{(m)}(i) \right\|^2 \\ &\text{where} \\ \left\| W_J^{(m)}(i) \right\|^2 &= \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} W_J^{(m)}(i)\ W_J^{(m)}(i) \end{aligned}\right\} \qquad (4\text{-}16)$$

---

† $$\left(t^J,\ W_J^{(m)}\right) = \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} t^J\ W_J^{(m)}(i) = \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} (i\Delta t)^J\ W_J^{(m)}(i$$

Using equation (4-8), equation (4-16) becomes*

$$\sigma_{mJ}^2 = \sigma_o^2 \sum_{j=0}^{J} \left[f_j^{(m)}(\tau/\Delta t) \right]^2 \qquad (4\text{-}17)$$

which is the desired result. Note that as J (the smoother order) increases, σ_{mJ}^2 increases so that the use of a higher order smoother gives rise to a noisier estimate.

We are now in a position to state an important optimality property of polynomial smoothers in the form of a theorem which is proven in Appendix II.

> Theorem: The J^{th} order polynomial smoother used to estimate the m^{th} derivative of a J^{th} degree polynomial input is that filter with zero dynamic error which minimizes the expectation of the square of the estimation error.

Although polynomial filters are the optimal filters when the signal is a polynomial of known degree, polynomial filters are of particular interest because of their applicability to the case in which the signal is not a polynomial but can be approximated by a polynomial of suitable degree, J, over a suitable smoothing time, T. Under these conditions, however, the estimation error

*See Appendix I.

consists of both a noise error and a bias error. The proper procedure for designing a polynomial filter in this case is to select the filter parameters J, and T, so as to minimize the total expectation of the square of the estimation error, which is given as

$$D_{mJ}^2 + \sigma_{mJ}^2$$

where

D_{mJ} = dynamic error of a J^{th} order smoother estimating the m^{th} derivative

σ_{mJ}^2 = output noise variance of J^{th} order smoother estimating the m^{th} derivative

The output noise variance increases with increasing J or decreasing T. On the other hand, dynamic error decreases with increasing J (due to the better fitting properties of higher order polynomials) and increases with increasing T (since more highly derivatives are required to accurately represent the signal).

Care should be taken in the selection of the estimation point in the smoothing interval [i.e., time, τ, at which estimate is made]. The parameters which must be considered in making this selection are the order of the input data, the smoother order and the allowable smoother real time delay. In general, updating smoothers [i.e.,

$\tau = \frac{r-1}{2} \Delta t]$ have the poorest accuracy but have no delay. If some delay can be tolerated, an improved accuracy of estimation of the derivatives of the input function can be obtained. Generally, the lowest estimation errors result from smoothing to the center of the smoothing interval [i.e., $\tau = 0$] which of course introduces a delay of one-half the smoothing time [i.e., T/2] for the output estimate.

5.0 OPTIMUM POLYNOMIAL SMOOTHERS USING A PRIORI INFORMATION

The design of the optimum linear polynomial smoother is considered, when statistical information or constraints on the input signal are available, a priori.

Assume the input signal, $x(t)$ is described by a polynomial of degree J, where

$$x(t) = a_o + a_1 t + \ldots + a_J t^J \tag{5-1}$$

If a finite number of discrete observations are made on the input at equally spaced time intervals and these observations are disturbed by noise $n(t_i)$, then the observations, $y(t_i)$, are given as:

$$y(t_i) = x(t_i) + n(t_i) \tag{5-2}$$

Assume the input noise consists of uncorrelated samples of a random variable with zero mean and standard deviation, σ_o. The smoothing interval to be considered is $-\frac{T}{2} \leq t \leq \frac{T}{2}$ where T is the total smoothing time. Letting Δt be the time spacing between observations and r be the total number of observations in T seconds, then

$$T = (r-1)\Delta t$$

At this point the explicit design criteria for the optimal linear smoother can be stated. Defining the linear filter in terms of its weighting function, $W(t_i)$, then the estimate of $x^m(\tau)$, the m^{th} derivative of the signal at time $= \tau$, when the signal $x(t)$ is a polynomial of known degree, J, is given as

$$\hat{x}^{(m)}(\tau) = \sum_{i=-\frac{r-1}{2}}^{\frac{r-1}{2}} y(t_i)\, W^{(m)}(t_i) \qquad (5\text{-}3)$$

It is required to determine $W^{(m)}(t_i)$, such that $\hat{x}^{(m)}(\tau)$ has a minimum mean square error when a priori knowledge about the coefficients of the input polynomial is available. This knowledge might consist of either constraints on the magnitude of the coefficients, a_j, of the input polynomial $x(t)$, or the fact that any or all of the coefficients, a_j, are described by some probability distribution with given moments.

Let the a priori probability density functions of the unknown coefficients be given as $p(a_j)$, and assume that both $E[a_j a_k]$ for $j \neq k$ and $E\left[a_j^2\right] = \sigma_j^2 + m_j^2$ are known, where σ_j^2 is the variance of the j^{th} coefficient and m_j is the mean of the j^{th} coefficient. Also assume that the observation noise and the a_j coefficients are independent.

Let N equal the noise output at $t = \tau$ and D_m, the dynamic error of the smoother. From Appendix I(b),

$$D_m = \sum_{j=0}^{J} a_j \, \varepsilon_{mj} \tag{5-4}$$

and the average square noise error is

$$E(N^2) = \sigma_o^2 \, \|W^{(m)}\|^2 \tag{5-5}$$

where $W^{(m)} \equiv W^{(m)}(t_i)$ is the filter weighting sequence

and ε_{mj} is the dynamic error in estimating the m^{th} derivative of an input, t^j, for $j = 0,1,2,\ldots J$

The total squared error at $t = \tau$ is given as

$$[N+D_m]^2 = N^2 + 2ND_m + D_m^2 = N^2 + 2N \sum_{j=0}^{J} a_j \, \varepsilon_{mj}$$

$$+ \sum_{j=0}^{J} \sum_{k=0}^{J} a_j a_k \, \varepsilon_{mj} \varepsilon_{mk} \tag{5-6}$$

Taking expectations of equation (5-6) we obtain for the total mean squared error

$$\text{mse} = E\left[(N+D_m)^2\right] = E(N^2) + 2E\left[N \sum_{j=0}^{J} a_j \, \varepsilon_{mj}\right]$$

$$+ E\left[\sum_{j=0}^{J} \sum_{k=0}^{J} a_j a_k \, \varepsilon_{mj} \varepsilon_{mk}\right] \tag{5-7}$$

Since the observation noise and the a_j's are assumed independent, equation (5-7) may be rewritten as,

$$\text{mse} = E(N)^2 + 2E[N]\ E\left[\sum_{j=0}^{J} a_j\ \varepsilon_{mj}\right]$$

$$+ E\left[\sum_{j=0}^{J}\sum_{k=0}^{J} a_j a_k\ \varepsilon_{mj}\varepsilon_{mk}\right]$$

It is easily shown using equations (5-2) and (5-3) that $E[N] = 0$ is implied by the assumption that the input noise has zero mean. Hence

$$\text{mse} = \sigma_0^2\ \|W^{(m)}\|^2 + E\left[\sum_{j=0}^{J}\sum_{k=0}^{J} a_j a_k\ \varepsilon_{mj}\varepsilon_{mk}\right]$$

which yields

$$\text{mse} = \sigma_0^2\ \|W^{(m)}\|^2 + \sum_{j=0}^{J}\sum_{k=0}^{J} \varepsilon_{mj}\varepsilon_{mk}\ E[a_j a_k] \tag{5-8}$$

Equation (5-8) must now be solved for that weighting function, $W^{(m)}$, with its associated dynamic error coefficients, ε_{mj}, which minimizes the mean square error given by $E\left[(N+D_m)^2\right]$. This result will be obtained utilizing the following theorems and definitions:

For $j = 0,1,\ldots J$, define $W^j(t)$ as a polynomial of degree J defined by its moments

$$(t^k, W^j) = \begin{cases} 0 & \text{if } k \neq j \\ 1 & \text{if } k = j \end{cases} \tag{5-9}$$

Let u_{mj} = moments of $W^{(m)}(t_1) = \left(t^j, W^{(m)}\right)$.

<u>Theorem 3</u>. (Proof in Appendix II)

Let $u_o, u_1, \ldots u_J$ be given real numbers. Then there is a unique polynomial, $W(t)$, of degree, J, such that

$$(t^j, W) = u_j \qquad 0 \leq j \leq J$$

and

$$W(t) = \sum_{j=0}^{J} u_j \, W^j(t) \tag{5-10}$$

<u>Theorem 5</u>. (Proof in Appendix II)

Of all filters with given dynamic error coefficients, that filter which minimizes the output noise has a weighting function, $W(t)$, which is a polynomial of degree J.

Define

$$c_{mj} = \left.\frac{d^m t^j}{dt^m}\right|_{t=\tau} \qquad 0 \leq j \leq J \tag{5-11}$$

From Appendix I, (I-30),

$$\varepsilon_{mj} = u_{mj} - c_{mj}$$

Hence

$$W^{(m)}(t) = U^{(m)}(t) + \sum_{j=0}^{J} \varepsilon_{mj} W^{j}(t) \qquad (5\text{-}12)$$

where

$$U^{(m)}(t) = \sum_{j=0}^{J} c_{mj} W^{j}(t) \qquad (5\text{-}13)$$

Since equation (5-13) describes the polynomial weighting function, $U^{(m)}(t)$, for $\varepsilon_{mj} = 0$ (and hence $c_{mj} = u_{mj}$), $U^{(m)}(t)$ is the J^{th} order polynomial least squares smoother weighting function which provides unbiased estimates of the m^{th} derivative of the input signal.

From equation (5-12)

$$\|W^{(m)}\|^2 = \|U^{(m)}\|^2 + \sum_{j=0}^{J} \sum_{k=0}^{J} \varepsilon_{mj}\varepsilon_{mk} (W^{j}, W^{k})$$

$$+ 2 \sum_{j=0}^{J} \varepsilon_{mj} \left(W^{j}, U^{(m)}\right) \qquad (5\text{-}14)$$

Substituting equation (5-14) into equation (5-8) and rearranging terms we obtain for the mean squared error

$$\text{mean square error} = E\left[(N+D_m)^2\right] = \sigma_o^2 \, \|U^{(m)}\|^2$$

$$+ \sum_{j=0}^{J} \sum_{k=0}^{J} \left[\sigma_o^2(W^j, W^k) + E(a_j a_k)\right] \varepsilon_{mj}\varepsilon_{mk}$$

$$+ 2\sigma_o^2 \sum_{j=0}^{J} \varepsilon_{mj} \left(W^j,\ U^{(m)}\right) \qquad (5\text{-}15)$$

Differentiating equation (5-15) with respect to each of the ε_{mj} and equating to zero yields J+1 linear equations in J+1 unknowns. The n^{th} equation of the J+1 total equations is given as

$$\sum_{k=0}^{J} \varepsilon_{mk} \left[\sigma_o^2(W^n, W^k) + E(a_n a_k)\right] = -\sigma_o^2 \left(W^n,\ U^{(m)}\right) \qquad (5\text{-}16)$$

The solution of equations (5-16) yield the dynamic error coefficients, ε_{mj} of the optimum filter.† The moments, u_{mj}, of the optimum filter are obtained since

$$u_{mj} = \varepsilon_{mj} + c_{mj}$$

---

†If σ_ℓ^2 or m_ℓ^2 is arbitrarily large (i.e. $\rightarrow \infty$), ε_ℓ is set equal to zero to obtain a minimum mean square error filter, since from (5-15) for σ_ℓ^2 or $m_\ell^2 \rightarrow \infty$, the mean square error $\rightarrow$ infinity for $\varepsilon_\ell \neq 0$.

The optimum weighting function, $W^{(m)}(t)$ is then equal to

$$W^{(m)}(t) = \sum_{j=0}^{J} u_{mj}\, W^{j}(t)$$

and the mean square error is given by equation (5-15). A procedure for obtaining the $W^{j}(t)$ polynomials is shown in Appendix III. The results are given as (III-8)

$$W^{j}(t) = \sum_{k=0}^{J} \frac{A_j^k}{B_k} F_k(t) \tag{5-17}$$

where $F_k(t)$ are the orthogonal polynomials described in Appendix I given as

$$F_k(t) = \sum_{j=0}^{J} A_j^k\, t^j \qquad 0 \leq k \leq J$$

Note that the restriction of our estimator to be strictly linear (i.e. equation 5-3) has resulted in a bias or dynamic error associated with the estimate. Specifically, the resultant dynamic error is given as the expected value of D_m (eq. 5-4).

$$E[D_m] = E\left[\sum_{j=0}^{J} a_j\, \varepsilon_{mj}\right]$$

$$E[D_m] = \sum_{j=0}^{J} E[a_j]\, \varepsilon_{mj}$$

$$E[D_m] = \sum_{j=0}^{J} m_j \, \varepsilon_{mj} \qquad (5\text{-}18)$$

where m_j is the known mean of the j^{th} Taylor's series input signal coefficient.

In Chapter 7.0, we shall have occasion to consider smoothers that have minimum mean square error, subject to the constraint that their estimates are unbiased.* For Gaussian statistics, Kalman[20] and others have shown that minimum mean square error, unbiased estimation is equivalent to conditional mean expectation estimation [i.e. using as an estimate the mean of the conditional distribution of the parameter, given a set of observations].

In Appendix VI the minimum mean square error, unbiased polynomial smoother is derived where the estimator is allowed to consist of a constant term <u>plus</u> a linear weighting sequence given as

$$\hat{x}^{(m)}(\tau) = G_m + \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} y(i)\, W^{(m)}(i) \qquad (5\text{-}19)$$

*Shaw[29] considers the very simple, special case when the highest coefficient is a zero mean Gaussian random variable.

where G_m is a constant, independent of the observations, y(i). It is shown in Appendix VI that the value of the constant, G_m, is

$$G_m = - \sum_{j=0}^{J} m_j \, \varepsilon_{mj} \tag{5-20}$$

and that the dynamic error coefficients, ε_{mj}, [and hence the filter weighting sequence] are obtained as the solution to J+1 linear equations, where the n^{th} equation is given as,

$$\sum_{k=0}^{J} \varepsilon_{mk} \left[\sigma_o^2 (W^n, W^k) + E(a_n a_k) - m_n m_k\right] = -\sigma_o^2 \left(W^n,\ U^{(m)}\right) \tag{5-21}$$

Note that if the a priori distributions of the a_j coefficients have zero mean (i.e. $m_j = 0$) then equation (5-21) is identical to (5-16) and both linear and linear plus a constant estimation yield identical results.

From Appendix VI, the mean square error of the optimum unbiased polynomial filter is (eq. VI-9)

$$\text{mse} = \sigma_o^2 \, \|U^{(m)}\|^2 + \sum_j \sum_k \left[\sigma_o^2 (W^j, W^k) + E[a_j a_k] - m_j m_k\right] \varepsilon_{mj} \varepsilon_{mk}$$

$$+ 2\sigma_o^2 \sum_j \varepsilon_{mj} \left(W^j,\ U^{(m)}\right) \tag{5-22}$$

where the ε_{mj} are obtained from the solution of (5-21). We note that "unbiased" used in the above sense indicates an average bias error equal to zero when averaged over the ensemble of all input signals. This is contrasted with the usual unbiased polynomial smoother estimates which are unbiased for each signal input.

6.0 EXAMPLES AND SAMPLE RESULTS OF OPTIMUM POLYNOMIAL SMOOTHERS

In this chapter various examples are considered which illustrate the design procedure for obtaining the optimum linear filters derived in Chapter 5.0. Results presented indicate the performance improvements obtainable using the above techniques.

6.1 Optimum Velocity Estimate With Known Constraint on the Input Acceleration

As an example of the filter design procedure derived in Chapter 5.0, consider a second degree input polynomial ($J = 2$), given by

$$x(t) = a_o + a_1 t + a_2 t^2$$

Let the a priori distribution of a_2 be given by

$$p(a_2) = \frac{1}{2a_{TH}} \qquad -a_{TH} \leq a_2 \leq a_{TH}$$

It is easily shown that the variance and mean of a_2 are,

$$\sigma^2_{a_2} = \frac{a^2_{TH}}{3} \quad \text{and} \quad m_{a_2} = 0$$

We assume that no a priori information is known about a_o or a_1 so that we may arbitrarily assign mean values of 0 and variances of ∞ to these coefficient distributions,

$\left[\text{i.e., } \sigma_{a_o}^2 = \sigma_{a_1}^2 = \infty, \qquad m_{a_o} = m_{a_1} = 0\right]$. Consequently $\varepsilon_o = \varepsilon_1 = 0$ and from equation (5-16),

$$\varepsilon_{m2} = \frac{-\sigma_o^2\left(U^{(m)},\ W^2\right)}{\sigma_o^2\ \|W^2\|^2 + \frac{a_{TH}^2}{3}} \tag{6-1}$$

Assume that we desire to find the minimum mean square estimator of the velocity ($m = 1$) at the present time, $t = \frac{T}{2} = \frac{r-1}{2}\Delta t$,

Using equation (5-11)

$$c_o = \left.\frac{d(t^o)}{dt}\right|_{t = \frac{r-1}{2}\Delta t} = 0$$

$$c_1 = \left.\frac{d(t)}{dt}\right|_{t = \frac{r-1}{2}\Delta t} = 1$$

$$c_2 = \left.\frac{d(t^2)}{dt}\right|_{t = \frac{r-1}{2}\Delta t} = \left.2t\right|_{t = \frac{r-1}{2}\Delta t} = (r-1)\Delta t$$

From Table I, Appendix I

$$F_o(x) = 1 \qquad\qquad B_o = r$$

$$F_1(x) = x \qquad\qquad B_1 = \frac{(r^2-1)r}{12}$$

$$F_2(x) = x^2 - \frac{r^2-1}{12} \qquad\qquad B_2 = \frac{(r^2-4)(r^2-1)r}{180}$$

where

$$x = t/\Delta t$$

Note that (from 5-17)

$$W^j(t) = \sum_{k=0}^{J} \frac{A_j^k}{B_k} F_k(t) \tag{6-2}$$ *

Therefore from (6-2), and Table I,

$$W^o = \frac{1}{r} + \frac{\left(x^2 - \frac{r^2-1}{12}\right)\left(-\frac{r^2-1}{12}\right)}{(r^2-4)(r^2-1)r/180}$$

In a similar manner we obtain

$$W^1 = \frac{\frac{x}{\Delta t}}{\frac{(r^2-1)r}{12}}$$

---

* $\left(\text{NOTE:} \quad A_1^1 = 1/\Delta t, \text{ since } x = t/\Delta t\right)$

and

$$W^2 = \frac{\frac{1}{\Delta t^2}\left[x^2 - \frac{r^2-1}{12}\right]}{(r^2-4)(r^2-1)r/180}$$

Using equation (5-13)

$$U(t) = \sum_{j=0}^{2} c_j \, W^j(t)$$

$$U(t) = \frac{12x}{(r^2-1)r\Delta t} + \frac{(r-1)\,\frac{1}{\Delta t}\left[x^2 - \frac{r^2-1}{12}\right]}{(r^2-4)(r^2-1)r/180}$$

which is, as mentioned in Chapter 5.0, the weighting sequence which provides the least squares estimate of velocity at $t = \frac{(r-1)}{2}\Delta t$ for a quadratic least squares smoother. Further,

$$(U,W^2) = \sum_{j=0}^{2} c_j \, (W^j,W^2)$$

$$(W^1,W^2) = 0$$

$$(W^2,W^2) = \frac{180}{\Delta t^4 (r^2-4)(r^2-1)r} = \|W^2\|^2$$

Since $c_o = 0$, and $(W^1, W^2) = 0$,

$$(U, W^2) = c_2(W^2, W^2) = (r-1)\Delta t\left[\frac{180}{\Delta t^4(r^2-4)(r^2-1)r}\right]$$

Substituting into equation (5-18)

$$\varepsilon_2 = \frac{-\sigma_o^2(r-1)180\,\Delta t \Big/ \Delta t^4(r^2-4)(r^2-1)r}{\sigma_o^2\left(\dfrac{180}{\Delta t^4(r^2-4)(r^2-1)r}\right) + \dfrac{a_{TH}^2}{3}}$$

and upon rearranging,

$$\varepsilon_2 = -(r-1)\Delta t\left[\frac{1}{1 + \dfrac{a_{TH}^2}{\left(\dfrac{540\sigma_o^2}{\Delta t^4(r^2-4)(r^2-1)r}\right)}}\right]$$

Now

$$W(t) = U(t) + \varepsilon_2 W^2(t)$$

Hence

$$W(t) = \frac{12x}{(r^2-1)r\Delta t}$$

$$+ \left[1 - \frac{1}{1 + \frac{a_{TH}^2}{\left(\frac{540\sigma_o^2}{\Delta t^4(r^2-4)(r^2-1)r}\right)}}\right] \frac{180(r-1)\left[x^2 - \frac{(r^2-1)}{12}\right]}{\Delta t(r^2-4)(r^2-1)r}$$

which is the desired filter weighting function. The mean square error is, from equation (5-15),

$$\text{mse} = \sigma_o^2 \, \|U\|^2 - \frac{\sigma_o^4 (U, W^2)^2}{\sigma_o^2 \, \|W^2\|^2 + \frac{a_{TH}^2}{3}}$$

which becomes, after some algebraic manipulations,

$$\text{mse} = \frac{\sigma_o^2}{\Delta t^2} \frac{12(8r-11)(2r-1)}{(r^2-4)(r^2-1)r}$$

$$- \frac{\sigma_o^4 \left[(r-1)^2 \Delta t^2 \left(\frac{180}{\Delta t^4(r^2-4)(r^2-1)r}\right)^2\right]}{\sigma_o^2 \frac{180}{\Delta t^4(r^2-4)(r^2-1)r} + \frac{a_{TH}^2}{3}}$$

Note that the first term in the above equation is the mean square error for the least square velocity estimate for the quadratic smoother. The fact the second term is negative shows the reduced mean square error for this filter.

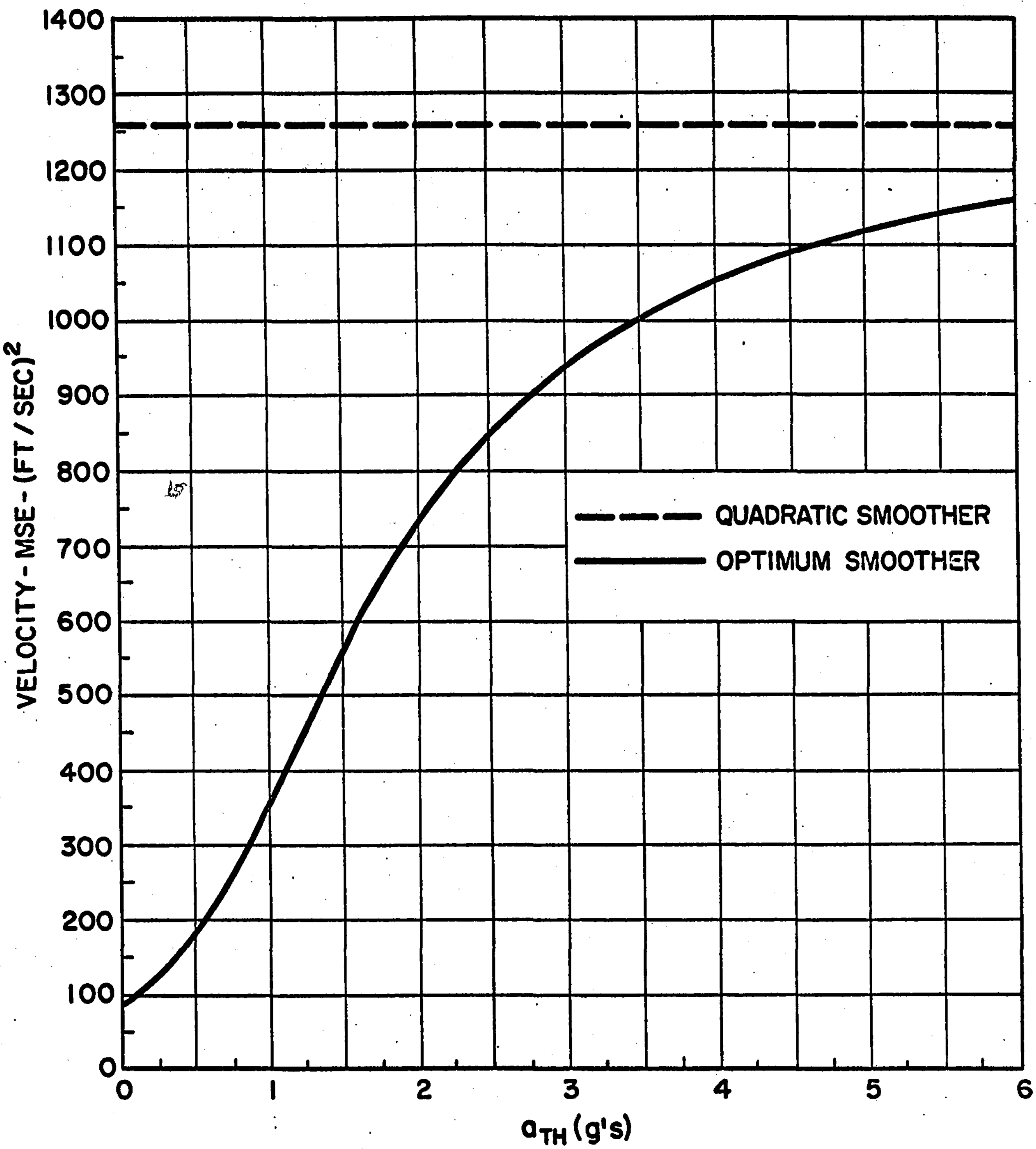
As a numerical example, consider a one-second filter with $\Delta t = .1$ sec and $r = 11$ points. Let $\sigma_o = 10$ ft. Figure 1 shows a plot of the velocity mean square error as a function of a_{TH}. The dashed line represents the asymptotic value of mean square error as $a_{TH} \to \infty$, corresponding to the normal least squares estimate. Depending on a_{TH}, improvement on the order of over 100 to 1 are obtainable using the optimum filter.*

6.2 Optimum Acceleration Estimate With Known Constraint on Input Acceleration

As a second example we consider the optimum acceleration estimate for the same conditions stated in Section 6.1. In this situation,

$$c_o = c_1 = 0; \qquad c_2 = 2; \qquad m = 2, \qquad \tau = \frac{r-1}{2} \Delta t$$

*Of course these improvements are only obtained if the a priori distributions which are used are correct. All uncertainties in the signal parameters must be reflected in these distributions in order that the results be meaningful. This situation must be kept in mind during all subsequent discussions where a priori data is utilized.



OPTIMUM VELOCITY ESTIMATE

FIG. 1

The usual quadratic filter weighting sequence, U(t) is obtained as

$$U(t) = \sum_{j=0}^{2} c_j W^j(t) = c_2 W^2(t) = \frac{360\left[x^2 - \frac{(r^2-1)}{12}\right]}{\Delta t^2 (r^2-4)(r^2-1)r}$$

From (6-1)

$$\varepsilon_2 = \frac{-\sigma_o^2 (U, W^2)}{\sigma_o^2 \|W^2\|^2 + \frac{a_{TH}^2}{3}}$$

where

$$(U, W^2) = c_2 (W^2, W^2) = \frac{360}{\Delta t^4 (r^2-4)(r^4-1)r} = 2 \|W^2\|^2$$

Therefore

$$\varepsilon_2 = \frac{-2\sigma_o^2 \frac{180}{\Delta t^4 (r^2-4)(r^2-1)r}}{\sigma_o^2 \frac{180}{\Delta t_4 (r^2-4)(r^2-1)r} + \frac{a_{TH}^2}{3}}$$

The optimum weighting sequence is given as

$$W(t) = U(t) + \varepsilon_2 W^2(t)$$

Substitution into the above equation yields,

$$W(t) = c_2 W^2(t) - \frac{2\sigma_o^2 \dfrac{180}{\Delta t^4 (r^2-4)(r^2-1)r}}{\sigma_o^2 \dfrac{180}{\Delta t^4 (r^2-4)(r^2-1)r} + \dfrac{a_{TH}^2}{3}} W^2(t)$$

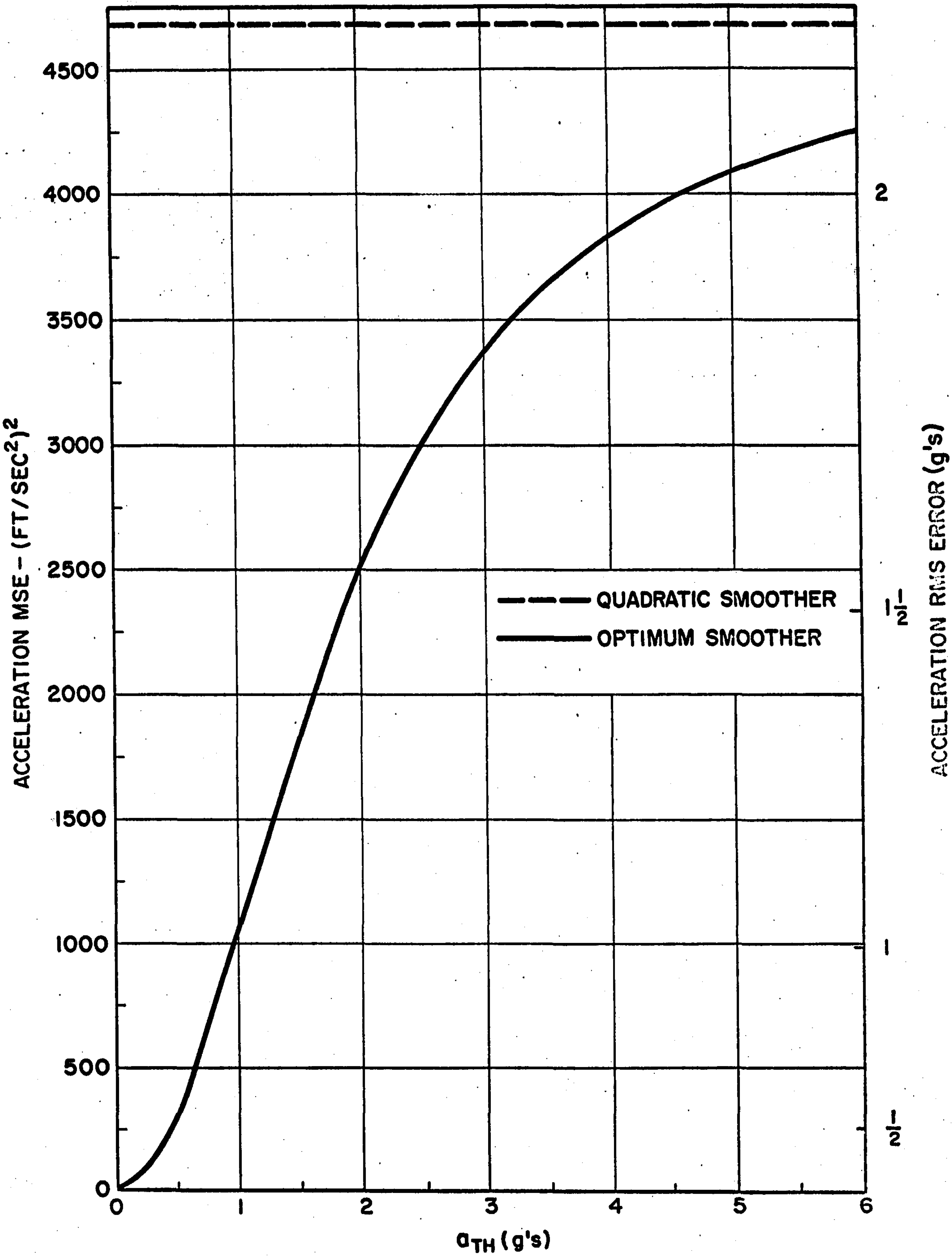
$$W(t) = \left[2 - \frac{2\sigma_o^2 \dfrac{180}{\Delta t^4 (r^2-4)(r^2-1)r}}{\sigma_o^2 \dfrac{180}{\Delta t^4 (r^2-4)(r^2-1)r} + \dfrac{a_{TH}^2}{3}} \right] \frac{180 \left[x^2 - \dfrac{(r^2-1)}{12} \right]}{\Delta t^2 (r^2-4)(r^2-1)r}$$

The mean squared error for the optimum acceleration estimate is given as

$$\text{mse} = \sigma_o^2 \, \|U\|^2 - \frac{\sigma_o^4 (U, W^2)^2}{\sigma_o^2 \, \|W^2\|^2 + \dfrac{a_{TH}^2}{3}}$$

$$= \frac{720\sigma_o^2}{\Delta t^4 (r^2-4)(r^2-1)r} - \frac{\left[\dfrac{720\sigma_o^2}{\Delta t^4 (r^2-4)(r^2-1)r} \right]^2 \Big/ 4}{\left[\dfrac{720\sigma_o^2}{\Delta t^4 (r^2-4)(r^2-1)r} \right] \Big/ 4 + \dfrac{a_{TH}^2}{3}}$$

Using the same parameter values as in 6.1 Figure 2 shows a plot of the acceleration mean square error as a function of a_{TH}. The dashed line is the mean square error for the usual quadratic smoother acceleration estimate.



OPTIMUM ACCELERATION ESTIMATE

FIG. 2

Note that for $a_{TH} = 0$, the mean square error equals zero. This is true since complete a priori knowledge of the acceleration is available.

6.3 Optimum Estimate of a Constant When True Value is a Sample of a Random Variable With Known Statistics

As a further example, consider the case of estimating a constant signal, $x(t) = a_o$, given a set of measurements of the signal. The usual procedure, assuming zero mean measurement noise, is to take an average of the observations. This corresponds to a zeroth order polynomial smoother whose weighting sequence is simply

$$W(t_i) = \frac{1}{r}$$

The mean square error of this estimate is given as σ_o^2/r. If, however, we have a priori information concerning the signal, an improvement can be obtained. Suppose we know that a_o is a sample of a random process whose variance is σ_1^2 and whose mean value is zero. Using the results developed in Chapter 5.0, the dynamic error coefficient for the optimum smoother is given as

$$\varepsilon_o = \frac{-\sigma_o^2/r}{\sigma_o^2/r + \sigma_1^2}$$

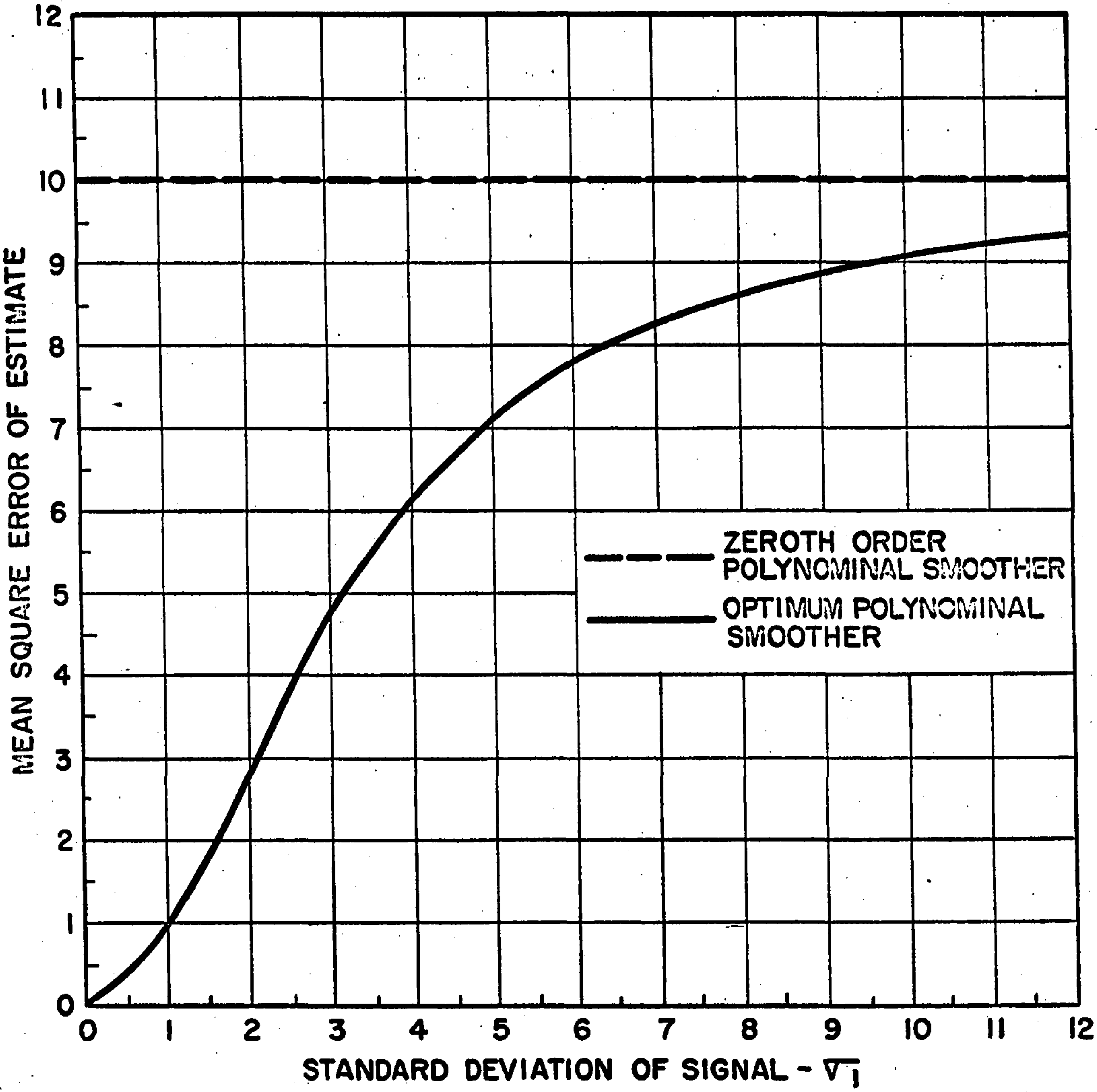
and the optimum weighting sequence is

$$W(t_1) = \frac{1}{r} - \frac{\sigma_o^2/r^2}{\sigma_o^2/r + \sigma_1^2} .$$

The mean squared error of the optimum estimate is

$$mse = \frac{\sigma_o^2}{r} - \frac{\sigma_o^4/r^2}{\sigma_o^2/r + \sigma_1^2} .$$

The first term in the above equation is the mean square error of the zeroth order polynomial smoother. Since the second term is positive, the resultant error is, as it should be, always less than for the zeroth order polynomial filter. To illustrate with some numberical results, consider a measurement accuracy of σ_1, the standard deviation of the random variable from which a_o is a sample. The dashed line represents the mean squared error obtained using the usual polynomial smoother.



OPTIMUM ESTIMATE OF A CONSTANT

FIG. 3

7.0 OPTIMUM ADAPTIVE FILTER DESIGN FOR INCOMPLETELY SPECIFIED SIGNALS

We discuss in this chapter the design of the optimum* filter given initial uncertainties about the form of the input signal. The input signal is assumed known except for some parameter, α, which might represent a specific parameter value or some range of values of a particular parameter.† It is known a priori that α can take on only a finite number of values (or ranges of values). If one knew the actual value of α, a priori, a filter could be designed to obtain a minimum mean square estimate of the input signal.

Let

x = actual value of a state of the physical world we wish to estimate

$\hat{x}$ = estimate of the state, x, given a set of measurements, Y.

The minimum mean square error estimate of x, given a set of observations, Y, is the conditional mean[20] of x given as

$$\hat{x} = E[x \mid Y] = \sum_{x} xP(x \mid Y) \qquad (7\text{-}1)$$

*Optimum is defined as the conditional mean of x which is equivalent to minimum mean square error.

†A similar approach to the problem is taken using Hilbert Space Theory in reference (24) for continuous systems using expanding memory recursive estimation. Only considered however, are cases when the unknown parameter takes on specific values. The transient behavior of these filters is not examined in (24) but is included in this paper.

Let α be a parameter of the input signal whose value is unknown a priori. Assume α can take on only a finite number of values, given as α_i, for $1 \leq i \leq L$. Each α_i may represent a range of values of a signal parameter. Summation of the joint conditional probabilities of x and α_i, given Y, yields $P(x \mid Y)$. i.e.

$$P(x \mid Y) = \sum_{i=1}^{L} P(x, \alpha_i \mid Y)$$

Using the theorem of compound probability the following result is obtained

$$P(x \mid Y) = \sum_{i=1}^{L} P(x \mid Y, \alpha_i) \cdot P(\alpha_i \mid Y) \qquad (7\text{-}2)$$

where x, Y and α_i may be vectors. Using (7-1)

$$\hat{x} = \sum_{x} x \sum_{i=1}^{L} P(x \mid Y, \alpha_i)\ P(\alpha_i \mid Y) \qquad (7\text{-}3)$$

Interchanging the order of summations yields,

$$\hat{x} = \sum_{i=1}^{L} \sum_{x} x\ P(x \mid Y, \alpha_i) \cdot P(\alpha_i \mid Y) \qquad (7\text{-}4)$$

Define

$$\hat{x}(\alpha_i) = \sum_{x} x \ P(x \mid Y,\alpha_i) \tag{7-5}$$

$\hat{x}(\alpha_i)$ is the conditional mean of x, given a set of observations, Y, assuming that α_i actually is true. Hence,

$$\hat{x} = \sum_{i=1}^{L} \hat{x}(\alpha_i) \ P(\alpha_i \mid Y) \tag{7-6}$$

The above equation states that $\hat{x}$, the conditional mean† estimate of x, is the weighted sum of the L conditional mean estimates of x (i.e. $\hat{x}(\alpha_i)$, $i = 1,2,\ldots L$). The weighting function is the conditional probability of α_i given Y (observations). Schematically, equation (7-6) corresponds to Figure 4.

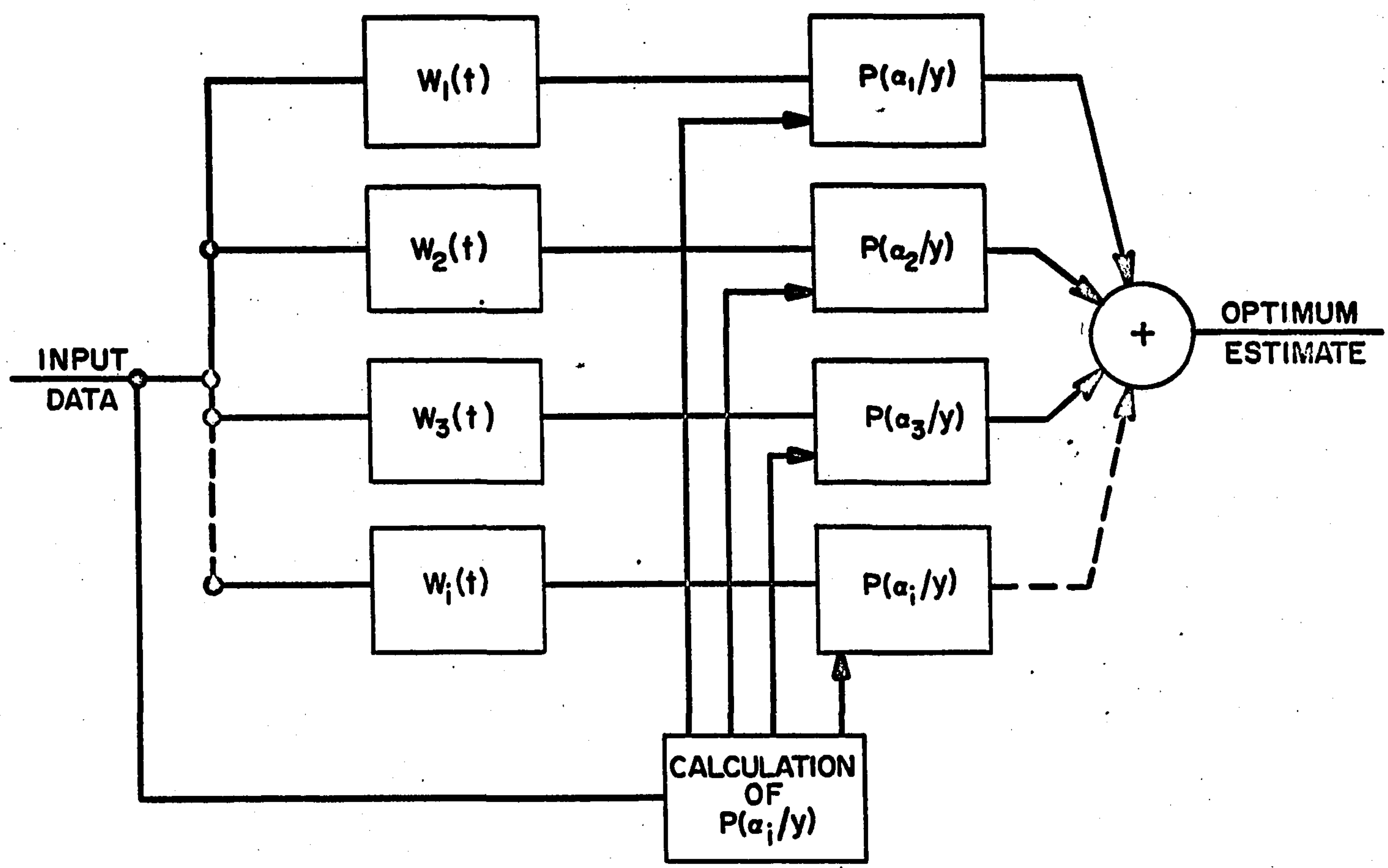
For an increasing memory filter*, as the observed data increases $P(\alpha \mid Y)$ converges[26] to $\delta(\alpha_i - \alpha_{true})$ where delta (δ) is the Kronecker symbol. Note that in the usual adaptive procedure using multiple smoothers, the estimate is taken as

$$\hat{x} = \hat{x}(\alpha_k) \ .$$

---

†Minimum mean square error.

*Smoothing time increases as more data is observed.



OPTIMUM ADAPTIVE SMOOTHER

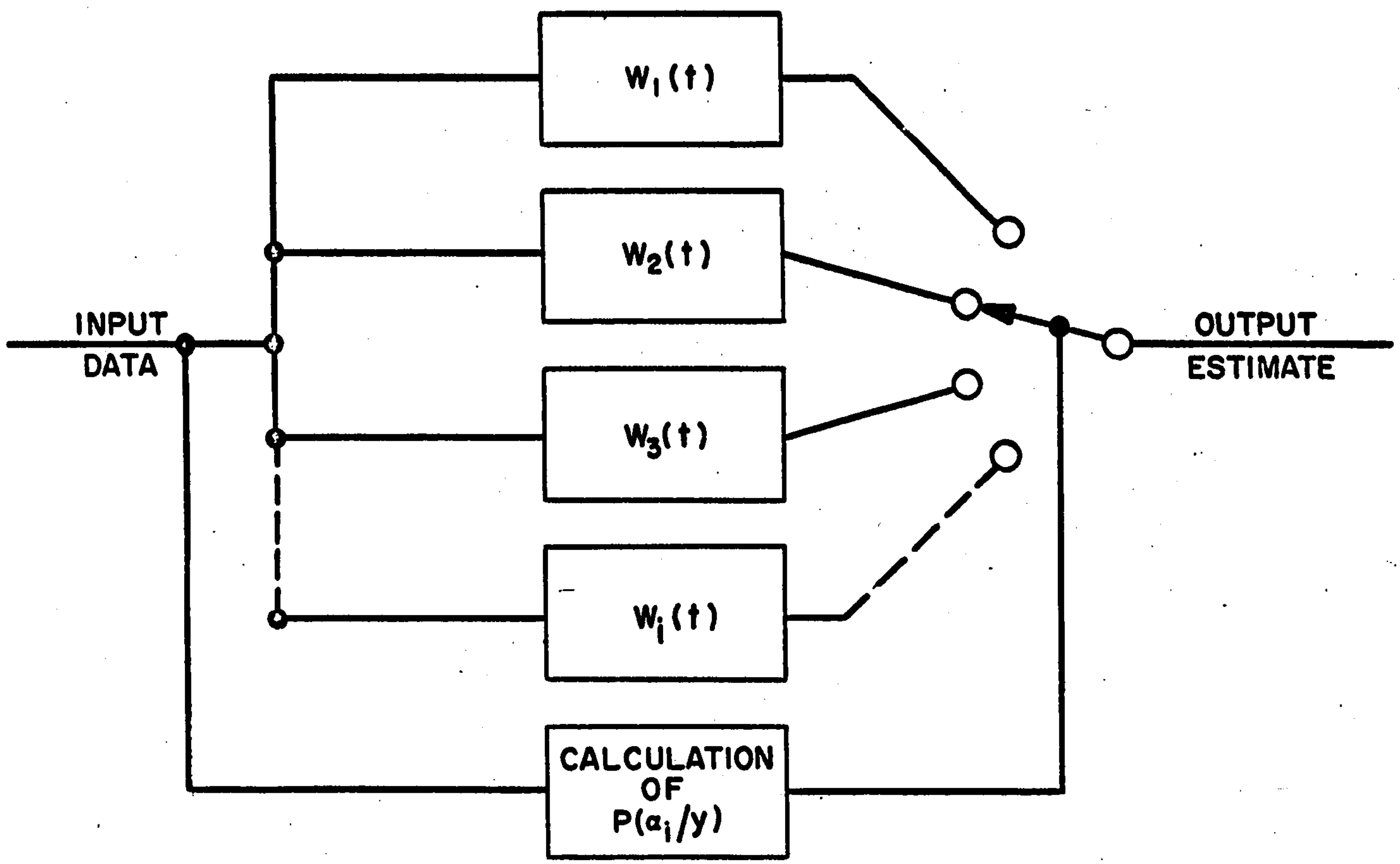
FIG. 4

where $P(\alpha_k \mid Y)$ is the maximum for all values $1 \leq i \leq L$. This estimate is optimum only when

$$P(\alpha_k \mid Y) = 1$$

The schematic representation of the usual adaptive switched smoother is shown in Figure 5.

The significant point to note with regard to the above theory is that no restrictive assumptions for distribution functions were made in the derivation. Moreover, equation (7-6) was derived in general, for any values of smoothing time or filter bandwidth, so that the estimation technique is optimum for the transient situation as well as for the infinite data steady-state condition.



SWITCHED ADAPTIVE SMOOTHER

FIG. 5

8.0 GENERATION OF THE OPTIMUM ADAPTIVE SUBFILTER WEIGHTS

The adaptive or nonlinear nature of the optimum smoothing filter, as derived in Chapter 7.0, is embodied in the weighting functions applied to the individual subfilter outputs. These weighting functions, (which are the conditional probabilities that the individual subfilters should be used, given a set of observations), are functions of the observed data, so that a nonlinear operation is introduced. Letting α_i represent a particular state of the input signal, we have from Bayes' Theorem that the a posteriori probabilities of α_i, $P(\alpha_i \mid Y)$, are given as†

$$P(\alpha_i \mid Y) = \frac{p(Y \mid \alpha_i)\ P(\alpha_i)}{\sum_j p(Y \mid \alpha_j)\ P(\alpha_j)} \qquad (8\text{-}1)$$

---

†Although Bayes' Theorem is usually written as

$$P(\alpha_i \mid Y) = \frac{P(Y \mid \alpha_i)\ P(\alpha_i)}{\sum_j P(Y \mid \alpha_j)\ P(\alpha_j)}$$

(8-1) is correct since the components of the random variable, Y, are not discrete. (See Harmon, W. W., "Principles of the Statistical Theory of Communications" equation 10-17.)

where Y is the observation vector (set of measurements) and $P(\alpha_j)$ are a priori probabilities of the parameter values α_j. If the $P(\alpha_j)$ are known, the optimum weighting functions are determined by obtaining the conditional probability density functions of the observation vector, Y, given that some α_j is true, evaluated at the particular value of the observation vector, Y. If the $P(\alpha_j)$ are unknown, then $P(\alpha_i \mid Y)$ must be estimated from the data alone. Examples are given in the next chapter assuming the $P(\alpha_j)$ are known a priori and also assuming they are unknown.

9.0 EXAMPLES OF OPTIMUM ADAPTIVE FILTERING

In Chapters 7.0 and 8.0 a generalized approach to optimum adaptive data smoothing (filtering) is discussed. It is shown that the conditional mean optimum adaptive estimate is comprised of a weighted sum of the outputs of a bank of smoothers, each designed to be optimum for some specific possible state of the input. We note that Kalman[20] shows that conditional expectation is equivalent to unbiased minimum mean square error linear estimation for Gaussian statistics. These results in conjunction with the optimum linear (and constant plus linear) filters derived in Chapter 5.0 will form the basis for the examples considered in the following paragraphs. The use of optimum linear subfilters is justified by the results of Kalman[20] who shows that "results obtainable by linear estimation can be bettered by nonlinear estimation <u>only</u> when 1) the random processes are non-Gaussian and even then only 2) by considering at least third order probability distribution functions". Also in the same paper a heuristic justification for the common use of Gaussian statistics is given. Kalman shows that: "Given any random process with known first and second order averages, we can find a Gaussian random process with the same properties. Thus Gaussian distributions and linear dynamics are natural, mutually plausible assumptions particularly when the statistical data are scant."

The examples presented in this chapter illustrate the adaptive subfilter design and determination of the system subfilter weights. In 9.1 the specific example of estimation of a constant signal when a set of noisy observations are made, is explored in detail. A comparison of the resultant mean square errors obtained using a usual polynomial smoother, an optimum linear smoother (Chapter 5.0) and an optimum adaptive smoother are given.

In 9.2 a method is presented of estimating the a posteriori probabilities (subfilter weights) using only the observed data. Various suboptimum smoothing techniques proposed in the literature are also considered for comparison with results obtained in this paper.

9.1 Adaptive Estimate of a Constant Signal

As an example consider the following problem. A set of measurements $Y(y_1, y_2, y_3 \ldots y_r)$ are made on a signal, $x(t)$, where $x(t)$ is equal to some unknown constant value a_o. There is noise, $n(i)$, associated with each measurement $y(i)$, such that

$$y(i) = a_o + n(i) \tag{9-1}$$

Assume $n(i)$ has a zero mean Gaussian probability density function with variance equal to σ_o^2, and successive samples of $n(i)$ are uncorrelated. Although a_o is constant during the set of r measurements, it is known that a_o is a sample of either one of two Gaussian random processes. The probability that a_o is a sample from process 1 is given as P_1 and from process 2 is P_2 where $P_1 + P_2 = 1$. The means and variances associated with process 1 and process 2 are $\left(0, \sigma_1^2\right)$ and $\left(0, \sigma_2^2\right)$ respectively. The optimum adaptive estimate of a_o, $\hat{a}_o$, is required.

Before obtaining a solution to this problem, let us consider some possible applications. The quantity a_o might be a transmitted voltage level in a binary communications system where the binary signals are samples of one of two random processes. The estimate $\hat{a}_o$ would be the optimum estimate of the voltage level a_o. In the solution to the above problem, by-products are estimates of the probabilities that the sample a_o is from either process 1 or process 2.

Another application might be the estimation of a reentry vehicle parameter in a missile defense system. Suppose that a particular reentry vehicle is from either one of two specific classes, either decoy or warhead, with known a priori probabilities P_D or P_W. A set of r independent measurements of the unknown parameter are obtained and the optimum estimate of the parameter is required along with an estimate of the a posteriori probability that the reentry vehicle is either a warhead or a decoy.

The solution for the optimum adaptive filter for the above problems is shown in Chapter 7.0. The optimum estimate of a_o, is obtained as the weighted average of the outputs of two subfilters designed to estimate the conditional mean (which is equivalent to unbiased minimum mean square error estimation since statistics are Gaussian) assuming that a_o is a sample of process 1 and that a_o is a sample of process 2. [The unknown parameter, α, to be learned is which random process a_o is taken from.] In this particular example, using the notation of Chapter 7.0, α_1, represents the case when a_o is a sample of process 1 and α_2 the case when a_o is a sample of process 2.

For this situation the optimum subfilters are zeroth order optimum polynomial smoothers designed with known variances of the a_o coefficient. Hence the results of Chapter 5.0 are directly applicable and since the mean of $a_o = 0$ for both processes, the weighting sequences for the subfilters are from Chapter 6.0, given as,

$$W_1(t) = \frac{1}{r} - \frac{\sigma_o^2/r^2}{\sigma_o^2/r + \sigma_1^2} \qquad (9\text{-}2)$$

and

$$W_2(t) = \frac{1}{r} - \frac{\sigma_o^2/r^2}{\sigma_o^2/r + \sigma_2^2} \qquad (9\text{-}3)$$

All that remains now is to find the optimum weights for the subfilter outputs which are, as shown in Chapter 8.0, given as

$$P(\alpha_1 \mid Y) = \frac{p(Y \mid \alpha_1)\ P_1}{p(Y \mid \alpha_1)\ P_1 + p(Y \mid \alpha_2)\ P_2} \tag{9-4}$$

and

$$P(\alpha_2 \mid Y) = \frac{p(Y \mid \alpha_2)\ P_2}{p(Y \mid \alpha_1)\ P_1 + p(Y \mid \alpha_2)\ P_2} \tag{9-5}$$

Since P_1 and P_2 are known, only $p(Y \mid \alpha_1)$ and $p(Y \mid \alpha_2)$ need be evaluated. We note that

$$p(Y \mid \alpha_1) = \int_{-\infty}^{\infty} p(Y,X_1 \mid \alpha_1)\ dX_1 \tag{9-6}$$

where $P(Y,X_1 \mid \alpha_1)$ is the joint probability density function of the observation vector, Y, and the random variable X_1,[†] given that the sample a_o is taken from process 1. Using the theorem of conditional probabilities,

$$p(Y,X_1 \mid \alpha_1) = p(Y \mid X_1,\alpha_1)\ p(X_1 \mid \alpha_1) \tag{9-7}$$

Substituting equation (9-7) into equation (9-6) we have

$$p(Y \mid \alpha_1) = \int_{-\infty}^{\infty} p(Y \mid X_1,\alpha_1)\ p(X_1 \mid \alpha_1)\ dX_1 \tag{9-8}$$

[†] X_1 and X_2 are the random variables associated with random processes 1 and 2 respectively.

We are given that process 1 is Gaussian, with zero mean and variance σ_1^2, hence

$$p(X_1) = \frac{1}{\sqrt{2\pi}\ \sigma_1} \exp\left[-\frac{X_1^2}{2\sigma_1^2}\right] \tag{9-9}$$

Since the n(i) are uncorrelated samples from a Gaussian, zero mean random process, the joint density function of the observations, given a particular sample of X_1 and α_1 is

$$p(Y \mid X_1,\alpha_1) = \prod_{i=1}^{r} \frac{1}{\sqrt{2\pi}\ \sigma_o} \exp\left[-\frac{[y_i - X_1]^2}{2\sigma_o^2}\right] \tag{9-10}$$

which can be written as,

$$p(Y \mid X_1,\alpha_1) = \frac{1}{(2\pi)^{r/2}\ \sigma_o^r} \exp\left[-\left\{\frac{\sum_i y_i^2}{2\sigma_o^2} - \frac{X_1 \sum_i y_i}{\sigma_o^2} + \frac{rX_1^2}{2\sigma_o^2}\right\}\right] \tag{9-11}$$

Substituting (9-11) and (9-9) into (9-8) and simplifying we obtain,

$$p(Y\,|\,\alpha_1) = \frac{1}{(2\pi)^{\frac{r+1}{2}}\,\sigma_o^r\sigma_1}\int_{-\infty}^{\infty}\exp\left[-\left\{\frac{\sum\limits_i y_i^2}{2\sigma_o^2} - \frac{\sum\limits_i y_i}{\sigma_o^2}X_1 + \left(\frac{r}{2\sigma_o^2} + \frac{1}{2\sigma_1^2}\right)X_1^2\right\}dX_1\right] \tag{9-12}$$

Making the substitution $\gamma_1 = \frac{r}{2\sigma_o^2} + \frac{1}{2\sigma_1^2}$ in equation (9-12) yields

$$p(Y\,|\,\alpha_1) = \frac{1}{(2\pi)^{\frac{r+1}{2}}\,\sigma_o^r\sigma_1}\exp\left[-\frac{\sum\limits_i y_i^2}{2\sigma_o^2}\right]\int_{-\infty}^{\infty}\exp\left[-\left[\gamma_1 X_1^2 - \frac{\sum y_i}{\sigma_o^2}X_1\right]dX_1\right] \tag{9-13}$$

Completing the square in the exponent of (9-13) and simplifying yields

$$p(Y\,|\,\alpha_1) = \frac{1}{(2\pi)^{\frac{r+1}{2}}\,\sigma_o^r\sigma_1}\exp\left[-\left[\frac{\sum\limits_i y_i^2}{2\sigma_o^2} - \frac{\left(\sum\limits_i y_i\right)^2}{4\sigma_o^4\gamma_1}\right]\right]\int_{-\infty}^{\infty}\exp\left[-\gamma_1\left[X_1 - \frac{\sum\limits_i y_i}{2\gamma_1\sigma_o^2}\right]^2 dX_1\right] \tag{9-14}$$

Integrating equation (9-14) yields the following results,

$$p(Y \mid \alpha_1) = \frac{1}{(2\pi)^{r/2}\sigma_o^r \sigma_1 \sqrt{2\gamma_1}} \exp\left[- \left[\frac{\sum y_i^2}{2\,{}_o^2} - \frac{\left(\sum_i y_i\right)^2}{4\sigma_o^4 \gamma_1} \right] \right] \qquad (9\text{-}15)$$

In a similar manner, we obtain

$$p(Y \mid \alpha_2) = \frac{1}{(2\pi)^{r/2}\sigma_o^r \sigma_2 \sqrt{2\gamma_2}} \exp\left[- \left[\frac{\sum_i y_i^2}{2\sigma_o^2} - \frac{\left(\sum_i y_i\right)^2}{4\sigma_o^4 \gamma_2} \right] \right] \qquad (9\text{-}16)$$

where

$$\gamma_2 = \frac{r}{2\sigma_o^2} + \frac{1}{2\sigma_2^2}$$

and

$$\sum_i y_i^2 = \sum_{i=1}^{r} y_i^2 = \text{sum of the squared observations}$$

$$\left[\sum_i y_i\right]^2 = \left[\sum_{i=1}^{r} y_i\right]^2 = \text{square of the sum of the observations.}$$

Using (9-16) and (9-15) we can obtain the optimum weights for the subfilter outputs,

$$P(\alpha_1 \mid Y) = \frac{p(Y \mid \alpha_1)\ P_1}{p(Y \mid \alpha_1)\ P_1 + p(Y \mid \alpha_2)\ P_2} \qquad (9\text{-}17)$$

and

$$P(\alpha_2 \mid Y) = \frac{p(Y \mid \alpha_2)\ P_2}{p(Y \mid \alpha_1)\ P_1 + p(Y \mid \alpha_2)\ P_2} \qquad (9\text{-}18)$$

where P_1 and P_2 are known and $p(Y \mid \alpha_1)$ and $p(Y \mid \alpha_2)$ are given by (9-15) and (9-16). Therefore $\hat{a}_o$, the optimum estimate of a_o, is given as

$$\hat{a}_o = \left[\sum_{i=-\frac{r-1}{2}}^{\frac{r-1}{2}} W_1(t)\ y_i \right] P(\alpha_1 \mid Y) + \left[\sum_{i=-\frac{r-1}{2}}^{\frac{r-1}{2}} W_2(t)\ y_i \right] P(\alpha_2 \mid Y) \qquad (9\text{-}19)$$

where $W_1(t)$ and $W_2(t)$ are given in equations (9-2) and (9-3) and $P(\alpha_1 \mid Y)$ and $P(\alpha_2 \mid Y)$ are given in equations (9-17) and (9-18). Obviously, from (9-17) and (9-18),

$$P(\alpha_1 \mid Y) + P(\alpha_2 \mid Y) = 1$$

To illustrate the improvement of the adaptive processing technique described by (9-19) over the usual nonadaptive technique let us consider the case when the number of measurements, r, approaches infinity (steady-state conditions). The nonadaptive smoother which takes no advantage of the measurements would consist of merely an averager or zeroth order smoother. The mean square error for the nonadaptive least squares smoother would be

$$\text{mse (nonadaptive)} = \sigma_o^2/r \qquad (9\text{-}20)$$

For the adaptive smoother, as $r \to \infty$, when α_1 is true, $P(\alpha_1 \mid Y) \approx 1$ and $P(\alpha_2 \mid Y) \approx 0$ and when α_2 is true, $P(\alpha_1 \mid Y) \approx 0$ and $P(\alpha_2 \mid Y) \approx 1$. From (9-19)

$$\hat{a}_o = \sum_{i=1}^{r} W_1(t)\, y_i$$

when α_1 is true and

$$\hat{a}_o = \sum_{i=1}^{r} W_2(t)\, y_i$$

when α_2 is true. Hence the total mean square error of the adaptive filter is the weighted average mean square error resulting from using $W_1(t)$ when α_1 is true and $W_1(t)$ when α_2 is true. Using the results of Chapter 5.0, the total adaptive mean square error is,

$$\text{mse (adaptive)} = P_1 \left[\frac{\sigma_o^2}{r} - \frac{\sigma_o^4/r^2}{\sigma_o^2/r + \sigma_1^2} \right]$$

$$+ P_2 \left[\frac{\sigma_o^2}{r} - \frac{\sigma_o^4/r^2}{\sigma_o^2/r + \sigma_2^2} \right] \qquad (9\text{-}21)$$

Noting that $P_1 + P_2 = 1$, equation (9-21) can be rearranged as follows,

$$\text{mse (adaptive)} = \frac{\sigma_o^2}{r} - \frac{\sigma_o^4}{r^2} \left[\frac{P_1}{\sigma_o^2/r + \sigma_1^2} + \frac{P_2}{\sigma_o^2/r + \sigma_\sigma^2} \right] \qquad (9\text{-}22)$$

Since the second term in (9-22) is always positive, the mean square error of the adaptive system is always less than $\frac{\sigma_o^2}{r}$, the nonadaptive, zeroth order smoother mean square error.

For large r (as has been assumed) equation (9-22) can be approximated by

$$\text{mse (adaptive)} \approx \frac{\sigma_o^2}{r} - \frac{\sigma_o^4}{r^2} \left[\frac{P_1}{\sigma_1^2} + \frac{P_2}{\sigma_2^2} \right] \qquad (9\text{-}23)$$

Now consider the use of the optimum nonadaptive filter described in Chapter 5.0 to estimate a_o directly. Let x be a random variable defined by the mutually exclusive selection of samples of process 1 with probability P_1 and process 2 with probability P_2. It is shown in Appendix V that the variance and mean associated with the random variable, x, are given as

$$\sigma_x^2 = P_1\sigma_1^2 + P_2\sigma_2^2 \tag{9-24}$$

and

$$m_x = 0 \quad \text{when} \quad m_1 = m_2 = 0$$

The weighting function for the optimum linear filter is then (from Chapter 5.0)

$$W(t) = \frac{1}{r} - \frac{\sigma_o^2/r^2}{\sigma_o^2/r + \sigma_x^2} \tag{9-25}$$

which, after substitution of (9-24) becomes,

$$W(t) = \frac{1}{r} - \frac{\sigma_o^2/r^2}{\frac{\sigma_o^2}{r} + P_1\sigma_1^2 + P_2\sigma_2^2} \tag{9-26}$$

The mean square error of the optimum linear filter is then

$$\text{mse (optimum linear)} = \frac{\sigma_o^2}{r} - \frac{\sigma_o^4}{r^2}\left[\frac{1}{\frac{\sigma_o^2}{r} + P_1\sigma_1^2 + P_2\sigma_2^2}\right] \tag{9-27}$$

For large r (as assumed), (9-27) reduces to

$$\text{mse (optimum linear)} = \frac{\sigma_o^2}{r} - \frac{\sigma_o^4}{r^2}\left[\frac{1}{P_1\sigma_1^2 + P_2\sigma_2^2}\right] \tag{9-28}$$

In order to show that the adaptive mean square error is always less than or equal to the optimum linear filter mean square error, it is necessary to show that the bracketed quantity in equation (9-23) is always greater than or equal to the bracketed quantity in equation (9-28) i.e.

$$\frac{P_1}{\sigma_1^2} + \frac{P_2}{\sigma_2^2} \overset{?}{\geq} \frac{1}{P_1\sigma_1^2 + P_2\sigma_2^2}$$

Using the fact that $P_1 + P_2 = 1$, the following algebraic manipulations can be made.

$$\frac{P_1\sigma_2^2 + P_2\sigma_1^2}{\sigma_1^2\sigma_2^2} \overset{?}{\geq} \frac{1}{P_1\sigma_1^2 + P_2\sigma_2^2}$$

$$\frac{(1-P_2)\sigma_2^2 + P_2\left(\sigma_1^2\right)}{\sigma_1^2\sigma_2^2} \overset{?}{\geq} \frac{1}{(1-P_2)\sigma_1^2 + P_2\sigma_2^2}$$

$$\left[\sigma_2^2 + P_2\left(\sigma_1^2 - \sigma_2^2\right)\right]\left[\sigma_1^2 + P_2\left(\sigma_2^2 - \sigma_1^2\right)\right] \overset{?}{\geq} \sigma_1^2\sigma_2^2$$

$$P_2\sigma_1^2\left(\sigma_1^2 - \sigma_2^2\right) + P_2\sigma_2^2\left(\sigma_2^2 - \sigma_1^2\right)$$

$$+ P_2^2\left(\sigma_1^2 - \sigma_2^2\right)\left(\sigma_2^2 - \sigma_1^2\right) \overset{?}{\gtrless} 0$$

Dividing both sides by $P_2\left(\sigma_1^2 - \sigma_2^2\right)$ we obtain,

$$\sigma_1^2 - \sigma_2^2 + P_2\left(\sigma_2^2 - \sigma_1^2\right) \overset{?}{\gtrless} 0, \qquad P_2\left(\sigma_1^2 - \sigma_2^2\right) \neq 0$$

or

$$\sigma_1^2 - \sigma_2^2 \overset{?}{\gtrless} P_2\left(\sigma_1^2 - \sigma_2^2\right), \qquad P_2\left(\sigma_1^2 - \sigma_2^2\right) \neq 0$$

and finally

$$1 \overset{?}{\gtrless} P_2, \qquad P_2\left(\sigma_1^2 - \sigma_2^2\right) \neq 0$$

which must be true since $P_1 + P_2 = 1$, hence proving that mse (adaptive) $\leq$ mse (optimum linear). We note that for the case that $P_1 = 0$ and $P_2 = 1$ (or vice versa) the adaptive filter reduces to the optimum linear filter and no mean square error improvement is obtained. From equation (9-21) and (9-26), if $\sigma_1^2 = \sigma_2^2$, both filters are identical, as they should be.

To summarize, it has been shown that the adaptive filter mean square error is less than the optimum linear filter mean square error, and that both filters yield an improved performance over the zeroth order polynomial filter.

9.2 Adaptive Estimation When A Priori Statistics Are Not Completely Known

We now consider an example where the unknown parameter α, represents a range of values of some quantity which describes the input signal. Further, assume that the probabilities of occurrence of a signal falling within the various possible ranges of parameters are unknown a priori. It is desired to find the adaptive filter which estimates the signal parameters in an optimum fashion.

As an example, assume that the input signal may be represented by a polynomial of known degree, J. However, it is known that the highest derivative, a_J, of the signal, x(t), falls within either one of two known ranges, Δa_{J_1} or Δa_{J_2}. Let Δa_{J_1} represent those values of a_J such that $|a_J| \leq a_{TH}$ and let Δa_{J_2} represent the range of values of a_J given as $|a_J| > a_{TH}$. Suppose we wish to estimate the m^{th} derivative of the input signal at time $t = \tau$.

The solution for the conditional mean optimum adaptive filter requires that the individual subfilters for each α_i (the unknown range of the highest derivative of the input signal which is to be learned) be conditional mean estimators. Although the statistics are non-Gaussian, we shall use minimum mean square error, unbiased estimators for the subfilters, using the justification of Chapter 9.0. In particular, we must find that $W_1(t)$

such that the estimate of $x^{(m)}(\tau)$ has a minimum mean square error, subject to the constraint that

$$|a_J| \le a_{TH}$$

with all values of a_J in this range equally likely, and that $W_2(t)$ such that the estimate of $x^{(m)}(\tau)$ has a minimum mean square error, subject to the constraint that

$$|a_J| > a_{TH}$$

with all values equally likely. Note that this problem is a special case of the problem solved in Chapter 5.0. Before we consider the derivation of the optimum weights for the subfilter outputs, let us consider a specific situation and derive the optimum subfilter weighting sequences. Select the degree of the highest signal derivative to be equal to 2 (J = 2), and an estimate of the velocity (m = 1) at the latest data point $\left(\text{i.e., } i = \frac{r-1}{2}\right)$; is required. For subfilter number 1 we have that $|a_2| \le a_{TH}$, where a_2 is the signal acceleration. As shown in Chapter 6.0†, the optimum weighting sequence for subfilter number 1 is given as

---

†The results for linear estimation are equivalent to linear plus a constant estimation since the mean value of a_J is zero.

$$W_1(t) = \frac{12x}{(r^2-1)r\Delta t}$$

$$+\left[1 - \frac{1}{1 + \dfrac{a_{TH}^2}{\left(\dfrac{540\sigma_o^2}{\Delta t^4(r^2-4)(r^2-1)r}\right)}}\right] \frac{180(r-1)\left[x^2 - \dfrac{(r^2-1)}{12}\right]}{\Delta t(r^2-4)(r^2-1)r} \qquad (9\text{-}29)$$

For subfilter number 2, the variance of a_2 is infinite. Hence the optimum subfilter is the normal quadratic least squares smoother where the weighting sequence, $W_2(t)$ is

$$W_2(t) = \frac{12x}{(r^2-1)r\Delta t} + \frac{180(r-1)\left[x^2 - \dfrac{(r^2-1)}{12}\right]}{\Delta t(r^2-4)(r^2-1)r} \qquad (9\text{-}30)$$

Let us return to the derivation of the optimum weights for the subfilter outputs. Since there is no a priori information on the probabilities of occurrence of a sample of a_J being from one or another of the possible ranges, we assume that a_J [the acceleration in the above case] can take on any value, equally likely. Hence the initial minimum mean square estimate of a_J is obtained by a J^{th} order, least squares polynomial smoother. For the quadratic example, the weighting sequence for this filter is derived in Chapter 5.0, and given as U(t), where

$$U(t) = \frac{360\left[x^2 - \frac{(r^2-1)}{12}\right]}{\Delta t^2(r^2-4)(r^2-1)r} \qquad (9\text{-}31)$$

It is now shown that the optimum subfilter weights, $\left[P(\alpha_1 \mid Y),\ P(\alpha_2 \mid Y)\right]$ are given by the integrated a posteriori probability density function of a_J over the regions $|a_J| \leq a_{TH}$ and $|a_J| > a_{TH}$ respectively.

Since no a priori information is available describing the probabilities that the highest signal derivative, a_J, lies within the threshold region between $-a_{TH}$ and a_{TH}, a Bayes estimate of these probabilities must be obtained solely from the observations. The minimum mean square estimate of a_J, is obtained using a J^{th} order, least squares polynomial smoother, since there are no restrictions on the coefficients. Given this estimate of a_J, $\hat{a}_J$, the best estimate of the probability that $-a_{TH} \leq a_J \leq a_{TH}$, $p(|a_J| \leq a_{TH})$, must be obtained. Since no a priori statistics are available about a_J, the use of Bayes' Theorem to estimate $p(|a_J| \leq a_{TH})$ requires that probability density functions be assumed and then limiting arguments be used to obtain the final results. Moreover, since a_J is an unknown constant, rather than a random variable, some mathematicians[12] object to using Bayes' techniques for estimation and instead prefer the use of the method of confidence intervals. It is shown in Appendix IV that using either approach, with the assumption of independent, zero mean, Gaussian noise corrupting the observations, the best estimate of $P(|a_J| \leq a_{TH})$ is given as

$$p(|a_J| \leq a_{TH}) = \frac{1}{2}\left[\operatorname{erfc}\left(\frac{-a_{TH} - \hat{a}_J}{\sqrt{2}\,\sigma_{\hat{a}}}\right) - \operatorname{erfc}\left(\frac{a_{TH} - \hat{a}_J}{\sqrt{2}\,\sigma_{\hat{a}}}\right)\right] \tag{9-32}$$

where

$$\operatorname{erfc} x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

$\sigma_{\hat{a}}$ = standard deviation of the estimate of a_J, $\hat{a}_J$.

Therefore, the optimum adaptive filter consists of the sum of the outputs from filters one and two [equations (9-29) and (9-30)] weighted by $P(|a_J| \leq a_{TH})$ and $P(|a_J| > a_{TH})$ respectively.

Let us now investigate the improvement obtained in mean square error of the adaptive estimate compared with the nonadaptive estimate. We shall make the comparison both for the so-called steady-state solution ($r \to \infty$) and for the transient (finite r) solution for a specific example. The nonadaptive least squares quadratic polynomial estimate of present velocity yields a mean square error given as approximately[1] (for large r)

$$\sigma^2_{\hat{V}} \text{ (nonadaptive) } \cong \frac{192\sigma_o^2}{T^2 r} \;{}^{\dagger}$$

where T = smoothing time.

As $r \to \infty$ the adaptive filter reduces to either subfilter 1 or 2 depending on whether a_J belongs to Δa_{J_1}, or Δa_{J_2}. The resultant mean square error is given as

$$\text{Mean Square Error (adaptive filter)} = P\left(\Delta a_{J_1}\right) mse_1 + P\left(\Delta a_{J_2}\right) mse_2$$

The mean square error of subfilter 1, mse_1, is given as (from Chapter 6.0)

$$mse_1 \cong \frac{192\sigma_o^2}{T^2 r} - \frac{\sigma_o^4 T^2 \left(\frac{180}{T^4 r}\right)^2}{\sigma_o^2 \frac{180}{T^4 r} + \frac{a_{TH}^2}{3}}$$

---

†This result may also be obtained using I-38 and assuming $r >> 2$.

The mean square error of subfilter 2, is identical with the nonadaptive least squares filter, and given approximately (for large r)

$$mse_2 \cong \frac{192\sigma_o^2}{T^2 r}$$

The total adaptive mean square error is therefore

$$mse_{ADAPTIVE} = \left[\frac{192\sigma_o^2}{T^2 r} - \frac{\sigma_o^4 T^2 \left(\frac{180}{T^4 r}\right)^2}{\sigma_o^2 \frac{180}{T^4 r} + \frac{a_{TH}^2}{3}}\right] P\left(\Delta a_{J_1}\right)$$

$$+ \left[\frac{192\sigma_o^2}{T^2 r}\right] P\left(\Delta a_{J_2}\right)$$

Since

$$P\left(\Delta a_{J_2}\right) + P\left(\Delta a_{J_1}\right) = 1$$

and both bracketed terms are less than or equal to the nonadaptive mean square error, the total adaptive mse is always less than or equal to the nonadaptive results [equality holds for the trivial case when both filters are identical which results when $a_{TH} = \infty$].

For the transient case when only a finite amount of data is available, the closed form solution for the mean square error of the adaptive filter is extremely tedious to obtain and leads to nontabulated integral forms. Therefore a computer Monte Carlo* simulation was undertaken to obtain some results for particular examples. These examples consisted of the use of various adaptive filtering techniques for the above problem, to allow comparison of the optimum adaptive filter with other "suboptimum" filtering procedures. Franklin[17] considers the use of the least squares ramp and parabola filter as the subfilters and switching between them based on an estimate of the acceleration. We note that for the example under consideration, the normal parabola filter has zero dynamic error and a total mean square velocity error given as[1]

$$mse_{PARABOLA} = \frac{\sigma_o^2}{\Delta t^2} \frac{12(8r-11)(2r-1)}{(r-2)(r-1)r(r+1)(r+2)}$$

The ramp filter on the other hand has a bias error proportional to the value of the acceleration input, a_2, in addition to a noise error given as

*The Monte Carlo simulation consisted of generating an ensemble of observation vectors, Y, (using a random number generator computer routine) for particular values of input signal acceleration. The various proposed filter weighting sequences and operations were applied to these data and ensemble averages of the resultant estimation errors were obtained. The IBM 7094 computer was used.

$$\sigma^2_{\hat{V}(RAMP)} = \frac{12\sigma_o^2}{\Delta t^2(r-1)(r)(r+1)}$$

The total mean square error of the ramp filter is

$$mse_{(RAMP)} = \frac{a_2T^2}{4} + \frac{12\sigma_o^2}{\Delta t^2(r-1)r(r+1)}$$

where $\frac{a_2T^2}{4}$ is the contribution of the bias error. By equating mse_{RAMP} to $mse_{PARABOLA}$ we can solve for that value of a_2 for which both errors are equal. Doing this we obtain (assuming a large value of r)

$$a_{2_{TH}} = \frac{26.9\sigma_o}{T^2\sqrt{r}}$$

If $|a_2|$ is greater than $a_{2_{TH}}$ then the parabola filter yields the lower mean square error and if $|a_2| < a_{2_{TH}}$ then the ramp filter is better. Using Franklin's approach, the resultant filter consists of selecting the ramp or parabola filter depending on whether the estimate of acceleration, a_2, is less than or greater than $a_{2_{TH}}$. It is also of interest to consider the use of the ramp and parabolic subfilters but using the optimum weighting arrangement described in this chapter.

Finally, consider the use of optimum linear sub-filters, but with a switching procedure rather than a weighting technique. The following parameter values were assumed for the comparison. Assume $J = 2$, $m = 1$, $r = 100$, $\Delta t = .01$ sec, $\sigma_o = 10$ ft and that the estimate is obtained at $t = \frac{(r-1)}{2} \Delta t$. Under these conditions

$$a_{2_{TH}} = 26.9 \text{ ft/sec}^2$$

and to have a common base for comparison, we let

$$a_{TH} = a_{2_{TH}} = 26.9 \text{ ft/sec}^2$$

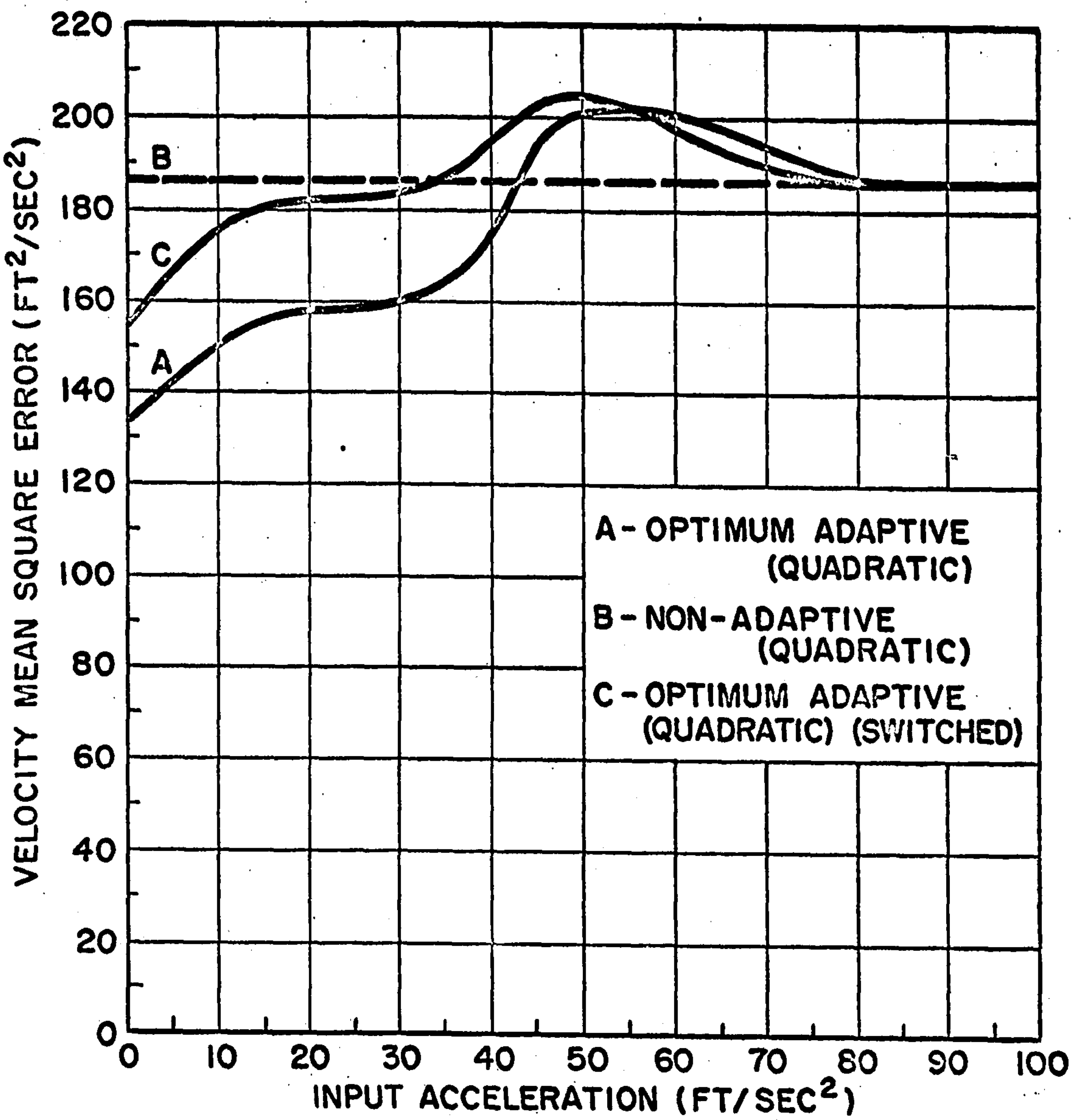
The nonadaptive least squares quadratic polynomial estimate of the present velocity yields a mean square error given as[1]

$$\sigma^2_{\hat{V}} \text{ (nonadaptive)} = \frac{\sigma_o^2}{\Delta t^2} \left[\frac{12(r-2)(r+2) + 180(r-1)^2}{(r-1)(r+1)(r)(r-2)(r+2)} \right]$$

which for the above example is

$$\sigma^2_{\hat{V}} \text{ (nonadaptive)} = 187 \text{ ft}^2/\text{sec}^2.$$

For the optimum adaptive system, the mean square error is presented as a function of the input acceleration. Figure 6 is a plot of the results, where mean square error in velocity is plotted as a function of input signal acceleration. The dashed curve is the nonadaptive

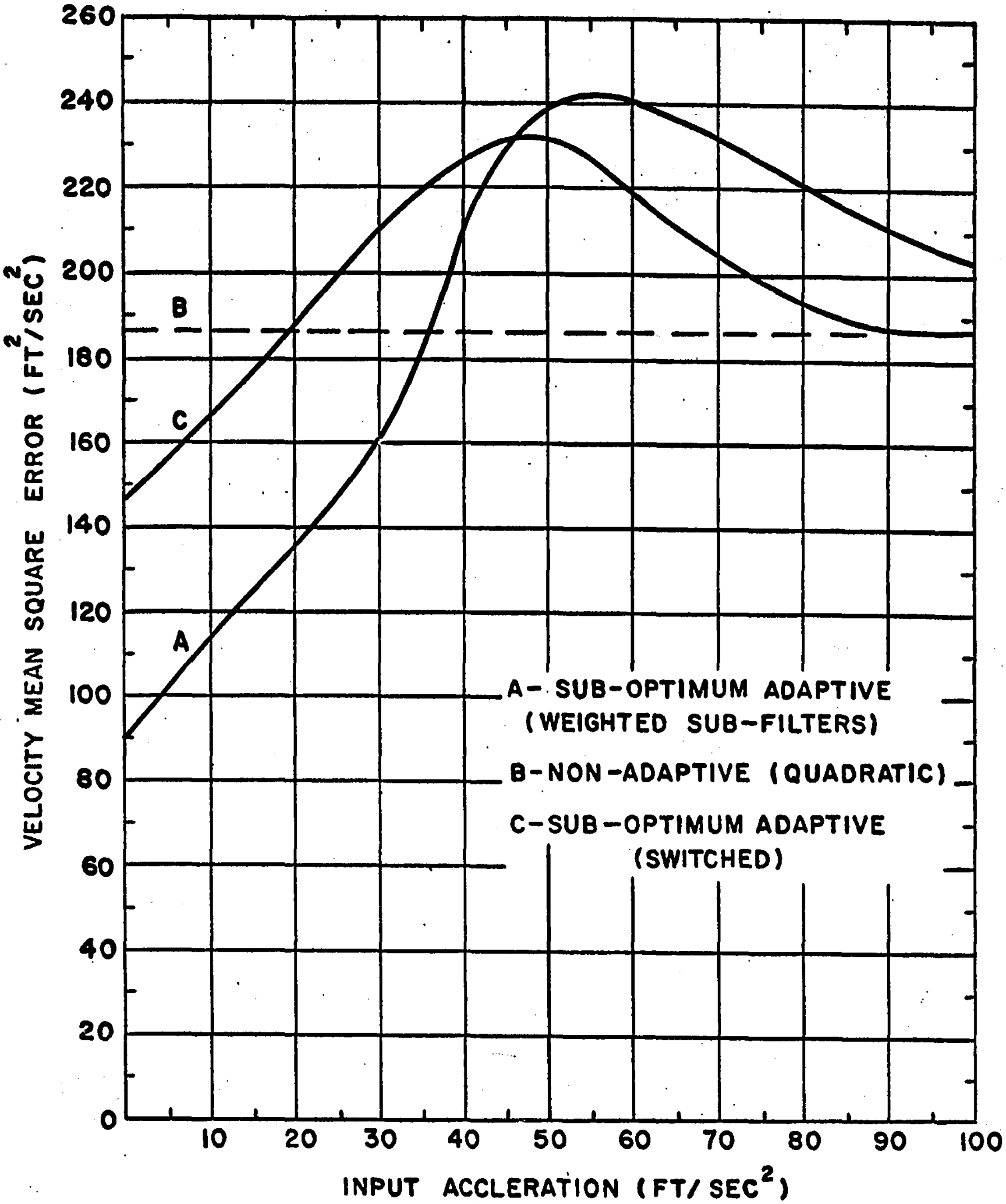


ADAPTIVE VELOCITY ESTIMATE

FIG. 6

mean square error calculated above. Although for some values of input acceleration the adaptive mse is greater than the nonadaptive mse, the average, as anticipated is lower. The actual mean square error, of course, depends on the actual probability distribution of the input acceleration. Also plotted on Figure 6 is the suboptimum switched adaptive mean square error result, using the optimum subfilters. By switched adaptive we mean that the output estimate is either that of subfilter 1 or 2 depending on which of the a posteriori probabilities, $P(|a_J| \leq a_{TH})$, or $P(|a_J| > a_{TH})$ is greater.

Figure 7 is a plot of the Franklin suboptimum switched filter and also the suboptimum weighted filter. Comparison of Figures 6 and 7 show that the optimum adaptive filter yields the lowest average mean square error. In addition, for each of the cases, the weighting approach seems to yield better results than the switching technique.



SUB-OPTIMUM ADAPTIVE VELOCITY ESTIMATE

FIG 7

10.0 CONCLUSIONS

10.1 Discussion of Results

This dissertation introduces the concept of minimum mean square error polynomial smoothing in addition to the usual methods of unbiased, minimum variance estimation. It is shown that linear polynomial smoothers can be designed, taking into account known a priori constraints or distributions of the input signal parameters, which yield substantial performance improvements with no additional system complexity. The resultant smoothers are obtained by finding that filter weighting sequence, such that the average output square error, consisting of noise and bias (dynamic error), is minimum. Closed form solutions for the optimum filter weighting sequence and the resulting mean square error are obtained, and are compared with least squares polynomial filter performance. In all cases, the optimum filter yields substantial improvements, which are illustrated by several numerical examples. Also considered is unbiased, minimum mean square error estimation using a priori information. Although polynomial signals were considered in this dissertation, the same approach would yield results for the case of a signal described by any linear combination of known functions.

The second class of problems considered is that of adaptive polynomial smoothers. The input signal is assumed known except for some parameter which can take on a finite number of values or ranges of values. It is shown that using a generalized mean square error performance index, the optimum estimate consists of the weighted sum of estimates from each of several subfilters, each designed assuming the unknown parameter takes on a different specific value or ranges of values. The weights for the individual outputs are the respective conditional probabilities that a parameter takes on a specific value, given the set of observations of the signal plus noise. Since the weights for the individual subfilter outputs are functions of the output measurements, the optimum adaptive filters are obviously nonlinear. Various examples illustrating the improved performance of adaptation are given in Chapter 9.0.

10.2 Suggestions for Future Work

It is well known[20-22] that for signal estimation problems, the minimum mean square error estimate is always obtained using the conditional mean estimator. However, when the statistics associated with the signal are non-Gaussian, the optimum filter is in general nonlinear and consequently difficult to derive or requires a complex realization. This problem is generally skirted by either one of two methods (which lead to equivalent results), namely assuming the statistics are Gaussian or finding the optimum <u>linear</u> filter.

In this dissertation a special class of non-Gaussian statistical signals was considered, consisting of signals described by a probability law obtained by the selection of samples from various Gaussian distributions with known selection probabilities. It was shown that the optimum filter consisted of the weighted sum of estimates obtained from a set of linear subfilters. Although the class of signals considered appears restricted, a slightly different point of view may lead to a more general approach to non-Gaussian signal estimation.

Specifically, the density function, p(x), of a random variable constructed by a selection procedure described above is easily (Appendix V) shown to be

$$p(x) = \sum_{i=1}^{L} P_i p_i(x), \qquad |P_i| \geq 0 \tag{10-1}$$

where P_i is the probability of selecting from the i^{th} distribution and $p_i(x)$ is the i^{th} Gaussian probability density function. $p_i(x)$ is given as

$$p_i(x) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{(x-m_i)^2}{2\sigma_i^2}\right] \tag{10-2}$$

where

σ_i^2 = variance of the i^{th} distribution

m_i = mean of the i^{th} distribution.

Since the selection procedure is mutually exclusive,

$$\sum_{i=1}^{L} P_i = 1.$$

Consider now the general estimation problem where a non-Gaussian signal with probability density function, $p_s(x)$, is given. If P_i, m_i and σ_i^2 for $1 \leq i \leq L$ and $\sum_{i=1}^{L} P_i = 1$, can be determined such that $p(x)$ equals or approximates $p_s(x)$, then $p_s(x)$ can be assumed to be of the form considered in Chapter 7.0 and 8.0 and the optimum nonlinear filter is determined.

Various questions must be considered before this approach proves useful. Specifically, how are P_i, m_i, σ_i^2 and L determined so that $p(x) \approx p_s(x)$? What are the convergence properties of the series expansion representation of (10-1)? One possibility might be to use a least squares fitting procedure, i.e. minimize

$$\int_{-\infty}^{\infty}\left[p(x) - p_s(x)\right]^2 dx = \int_{-\infty}^{\infty}\left[\sum_{i=1}^{L} P_i p_i(x) - p_s(x)\right]^2 dx \tag{10-3}$$

subject to the constraint that $\sum_{i=1}^{L} P_i = 1$. This approach seems attractive since $p_i(x)$ are linearly independent and hence can be orthogonalized, which generally simplifies the curve fitting problem. In order for this technique to be practically feasible, L must be kept small so that only a few subfilters are needed. However, under these conditions it is not clear if it is worthwhile to use this approach over the optimum linear filter technique.

Another possible approach would be that of matching the moments of $p_s(x)$ and $p(x)$. The noncentral moments of $p(x)$ are given, very simply, as

$$m_n = \sum_{i=1}^{L} P_i m_{in} \tag{10-4}$$

where m_n is the n^{th} moment of $p(x)$ and m_{in} is the n^{th} moment of $p_i(x)$. Since $p_i(x)$ is Gaussian, all of its moments can be expressed in terms of m_1 and σ_1^2. For example, the o^{th} through 5^{th} noncentral moments of $p_s(x)$ can be identically matched using L=2 (i.e. two subfilters). However, six nonlinear simultaneous equations must be solved to obtain the unknown parameters, P_1, P_2, m_1, m_2, σ_1^2, and σ_2^2.

APPENDIX I. POLYNOMIAL SMOOTHERS

a) Derivation of Polynomial Filters

Consider a sampled signal, $x(t_i)$, which is disturbed by noise, $n(t_i)$, such that

$$y(t_i) = x(t_i) + n(t_i) \tag{I-1}$$

Assume

$$E\left[n(t_i),\, n(t_\ell)\right] = \delta_{i\ell}\sigma_o^2$$

and

$$E\left[n(t_i)\right] = 0$$

and

$$t_{i+1} - t_i = \Delta t$$

We are interested in determining the optimum finite memory discrete linear filter to estimate the function, $x(t_i)$, or any of its derivatives, $x^{(m)}(t_i)$, given a finite number, r, of noisy samples, $y(t_i)$, extending over a smoothing (or filtering) time, T, where $T = (r-1)\Delta t$.

A linear estimator of the m^{th} derivative of the input signal is defined as

$$x^{*(m)}(t_i) = \sum_{i=-\frac{r-1}{2}}^{\frac{r-1}{2}} W^{(m)}(i)\; y(t_i) \tag{I-2}$$

We shall be concerned with the question of estimating the m^{th} derivative of the input signal $\left[x^{(m)}(t_1)\right]$ at some time, $t = \tau$, for the case in which $x(t)$ is either a polynomial of known degree, J, or can be approximated by a polynomial over the time interval, T. The results are obtained in an extremely useful form using an orthonormal polynomial expansion signal representation rather than the usual Taylor's series approach.

These polynomials, $f_j(n)$, are described[18] by the following recursive formulas,

$$f_j(n) = \frac{F_j(n)}{\sqrt{B_j}} \tag{I-3}$$

where

$$F_o(n) = 1$$

$$F_1(n) = n$$

$$F_{k+1}(n) = nF_k(n) - \frac{k^2(r^2-k^2)}{4(4k^2-1)} F_{k-1}(n), \quad k \geq 1 \tag{I-4}$$

where

$$n = \frac{t}{\Delta t}$$

$$B_o = r$$

and

$$B_k = \frac{k^2(r^2-k^2)}{4(4k^2-1)} B_{k-1} \quad k \geq 1 \tag{I-5}$$

Table 1 lists the first several of these polynomials.

TABLE I

Orthogonal Polynomials

$$F_o(n) = 1$$

$$F_1(n) = n$$

$$F_2(n) = n^2 - \frac{r^2-1}{12}$$

$$F_3(n) = n^3 - \left(\frac{3r^2 - 7}{20}\right) n$$

$$F_4(n) = n^4 - \left(\frac{3r^2 - 13}{14}\right) n^2 + \frac{3(r^2-1)(r^2-9)}{560}$$

$$F_5(n) = n^5 - \left[\frac{5(r^2-7)}{18}\right] n^3 + \left[\frac{15r^4 - 230r^2 + 407}{1008}\right] n$$

$$B_o = r$$

$$B_1 = \frac{(r^2-1)r}{12}$$

$$B_2 = \frac{(r^2-4)(r^2-1)r}{180}$$

$$B_3 = \frac{(r^2-9)(r^2-4)(r^2-1)r}{2800}$$

$$B_4 = \frac{(r^2-16)(r^2-9)(r^2-4)(r^2-1)r}{44,100}$$

where $n = \frac{t}{\Delta t}$

In the development of the theory we shall use the following theorem which is proven in Appendix II.

Theorem 1: Every polynomial, $x(t)$, of degree J, can be expressed as a linear combination of the orthogonal polynomials $f_j\left(\frac{t}{\Delta t}\right)$.

Using Theorem 1

$$x(t_i) = \sum_{j=0}^{J} b_j \; f_j(t_i/\Delta t) \tag{I-6}$$

where

$$\sum_{i=-\left(\frac{r-1}{2}\right)}^{\frac{r-1}{2}} f_j(t_i/\Delta t) \; f_k(t_i/\Delta t) = \delta_{jk} \tag{I-7}$$

and δ_{jk} is the Kronecker delta symbol.

Let $x^*(\tau)$ be the estimate of $x(t_i)$ at time $t = \tau$ where

$$x^*(\tau) = \sum_{j=0}^{J} b_j^* \; f_j(\tau/\Delta t) \tag{I-8}$$

where J is the smoother order and the b_j^* coefficients are to be determined. To satisfy the least squares error criterion, the expression for the sum of square errors given by

$$\text{Sum of squared errors} = \Sigma R = \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} \left[y(t_i) - x^*(t_i) \right]^2 \tag{I-9}$$

is minimized with respect to the coefficients, b_j^*. Substituting equation (I-8) into (I-9) we obtain

$$\sum_i R = \sum_i \left[y(t_i) - \sum_{j=0}^{J} b_j^* \, f_j(t_i/\Delta t) \right]^2 \tag{I-10}$$

Expanding equation (I-10) we have

$$\sum_i R = \sum_i y^2(t_i) - 2 \sum_{j=0}^{J} b_j^* \sum_i y(t_i) \, f_j(t_i/\Delta t)$$

$$+ \sum_i \sum_{j=0}^{J} \sum_{k=0}^{J} b_j^* b_k^* \, f_j\left(\frac{t_i}{\Delta t}\right) f_k\left(\frac{t_i}{\Delta t}\right) \tag{I-11}$$

Differentiating equation (I-11) with respect to b_j^* and using the orthogonality relation of equation (I-7) we obtain

$$b_k^* = \sum_i y(t_i) \, f_k(t_i/\Delta t) \tag{I-12}$$

Substituting (I-12) into (I-8) we have

$$x^*(\tau) = \sum_{j=0}^{J} \left[\sum_{i} y(t_i)\ f_j(t_i/\Delta t) \right] f_j(\tau/\Delta t) \qquad \text{(I-13)}$$

Rearranging (I-13)

$$x^*(\tau) = \sum_{i} y(t_i) \sum_{j=0}^{J} f_j(t_i/\Delta t)\ f_j(\tau/\Delta t) \qquad \text{(I-14)}$$

From (I-14) we note that the J^{th} order smoother weighting sequence for the estimate of $x^*(\tau)$, $W_J(i)$, is given as

$$W_J(i) = \sum_{j=0}^{J} f_j(t_i/\Delta t)\ f_j(\tau/\Delta t) \qquad \text{(I-15)}$$

We now consider the optimum estimate of the m^{th} derivative of the input signal, $x^{*(m)}(\tau/\Delta t)$. Blum[10] has shown that the optimum estimate is simply the m^{th} derivative of the optimum estimate of $x(\tau)$, given as

$$x^{*(m)}(\tau) = \frac{d^m}{d\tau^m}\, x^*(\tau) \qquad \text{(I-16)}$$

Substituting (I-14) into (I-16) we have

$$x^{*(m)}(\tau) = \frac{d^m}{d\tau^m} \sum_i y(t_i) \sum_{j=0}^{J} f_j(t_i/\Delta t)\; f_j(\tau/\Delta t) \tag{I-17}$$

Simplifying (I-17) and using the notation

$$\frac{d^m}{d\tau^m} f_j(\tau/\Delta t) = f_j^{(m)}(\tau/\Delta t)$$

we obtain

$$x^{*(m)}(\tau) = \sum_i y(t_i) \sum_{j=0}^{J} f_j(t_i/\Delta t)\; f_j^{(m)}(\tau/\Delta t) \tag{I-18}$$

Hence the weighting function for the m^{th} derivative estimate of a J^{th} order smoother at time $t = \tau$ is given as,

$$W_J^{(m)}(i) = \sum_{j=0}^{J} f_j(t_i/\Delta t)\; f_j^{(m)}(\tau/\Delta t) \tag{I-19}$$

b) <u>Properties of Polynomial Smoothers</u>

Aside from the curve fitting properties of polynomial smoothers, as illustrated by the method of derivation in Appendix I(a), polynomial filters have other desirable properties which we now consider.

We note that the estimates of the orthogonal polynomial coefficients, b_j^*, given in equation (I-12) are unbiased estimates of b_j, since

$$E\left[b_k^*\right] = E\left[\sum_i y(t_i)\ f_k(t_i/\Delta t)\right]$$

$$= \sum_i E\left[y(t_i)\right] f_k(t_i/\Delta t)$$

$$E\left[b_k^*\right] = \sum_i x(t_i)\ f_k(t_i/\Delta t) \qquad \text{(I-20)}$$

which from equation (I-6) yields

$$E\left[b_k^*\right] = b_k \qquad \text{(I-21)}$$

Therefore, since $x^{*(m)}(\tau) = \sum_{j=0}^{J} b_k^*\ f_k^{(m)}(\tau/\Delta t)$

$$E\left[x^{*(m)}(\tau)\right] = \sum_{j=0}^{J} b_k\ f_k^{(m)}(\tau/\Delta t) = x^{(m)}(\tau) \qquad \text{(I-22)}$$

if x(t) is a polynomial of degree, K, which is equal or less than J. Hence under these conditions polynomial estimates are unbiased estimates for all derivatives if the input signal order is less than or equal to J, the smoother order. Consider the situation where the input signal is a polynomial of degree K, where K > J. Hence

$$x^{(m)}(t) = \sum_{j=0}^{K} b_j \; f_j^{(m)}(t) \tag{I-23}$$

Under these conditions dynamic or bias error is introduced into the estimate. We shall define dynamic error in the estimate of the m^{th} derivative, D_m as

$$D_m = E\left\{x^{*(m)}(\tau) - x^{(m)}(\tau)\right\} \tag{I-24}$$

Substituting equation (I-8) and (I-23) into (I-24) we obtain

$$D_m = E\left\{\sum_{j=0}^{J} b_j^* \; f_j^{(m)}(\tau) - \sum_{j=0}^{K} b_j \; f_j(\tau)\right\}$$

Rearranging the above equation yields

$$D_m = E\left\{\sum_{j=0}^{J} \left(b_j^* - b_j\right) f_j^{(m)}(\tau) - \sum_{j=J+1}^{K} b_j \; f_j^{(m)}(\tau)\right\}$$

Since $E\left\{b_j^*\right\} = b_j$ we can simplify the above to yield

$$D_m = -\sum_{j=J+1}^{K} b_j \, f_j^{(m)}(\tau) \tag{I-25}$$

A Taylor's series expansion of the input signal, $x(t)$, about $t = 0$ yields

$$x(t) = \sum_{j=0}^{K} a_j \, t^j \tag{I-26}$$

Using (I-26), equation (I-25) can be rewritten to yield

$$D_m = \sum_{j=0}^{K} a_j \, \varepsilon_{mj} \tag{I-27}$$

where the ε_{mj} are called dynamic error coefficients and in particular for the above situation, $\varepsilon_{mj} = 0$ for $j = 0$ to $j = J$. Hence

$$D_m = \sum_{j=J+1}^{K} a_j \, \varepsilon_{mj} \tag{I-28}$$

Therefore for discrete polynomial filters, the dynamic error is zero if the input signal order, K, is less than or equal to J, the smoother order.

The ε_{mj} coefficients may be obtained by noting that for an input signal described by equation (I-26), the total dynamic error is equal to the sum of the dynamic errors [since we have a linear filter] associated with each of the terms of equation (I-26). Hence from (I-27), our definition of dynamic error coefficients, we note that ε_{mj} is simply the dynamic error due to an input equal to t^j, which is

$$\varepsilon_{mj} = \sum_{i=-\frac{r-1}{2}}^{\frac{r-1}{2}} t^j \, W_J^{(m)} - \left.\frac{d^m t^j}{dt^m}\right|_{t=\tau} \qquad \text{(I-29)}$$

where the first term on the right side is the filter output and the second term is the true value. Equation (I-29) can be rewritten in terms of the inner product notation as

$$\varepsilon_{mj} = \left(t^j,\ W_J^{(m)}\right) - \left.\frac{d^m t^j}{dt^m}\right|_{t=\tau} \qquad \text{(I-30)}$$

We also note that $\left(t^j,\ W_J^{(m)}\right)$ are the moments of the filter weighting function, $W_J^{(m)}$.

We now concern ourselves with the effect of polynomial filters on the input noise; in particular, some measure of the output noise associated with a particular estimate, $x^{*(m)}(\tau)$, using a J^{th} order smoother is desired. Consider a general linear filter which is described by some weighting function (impulse response), $W(i)$, such that the output is given by, $\hat{x}$, where

$$\hat{x} = \sum_{i} y(i)\, W(i) \tag{I-31}$$

and

$$y(i) = x(i) + n(i).$$

Equation (I-31) may be rewritten as

$$\hat{x} = \sum_{i} x(i)\, W(i) + \sum_{i} n(i)\, W(i) \tag{I-32}$$

where the second term on the right is the noise term associated with the estimate, $\hat{x}$. Let us consider the properties of this term, N, where

$$N = \sum_{i} n(i)\, W(i) \tag{I-33}$$

A measure of the "size" of N is the mean square value of N; that is to say, the expectation of $(N)^2$.

$$E(N^2) = E\left\{\left(\sum_{i} n(i)\, W(i)\right)\right\}^2$$

Therefore

$$E(N^2) = E\left\{\sum_i \sum_j n(i)\ n(j)\ W(i)\ W(j)\right\}$$

and so

$$E(N^2) = \sum_i \sum_j W(i)\ W(j)\ E\{n(i)n(j)\} \qquad \text{(I-34)}$$

By assumption, the input noise samples are mutually independent with mean value zero and variance σ_o^2. Hence

$$E\{n(i)n(j)\} = \sigma_o^2 \delta_{ij}. \qquad \text{(I-35)}$$

Equation (I-35) into (I-34) we obtain

$$E(N^2) = \sigma_o^2 \sum_i W^2(i) = \sigma_o^2\ \|W(i)\|^2 \qquad \text{(I-36)}$$

This result is significant since it states that the output noise variance is equal to the input noise variance, σ_o^2, multiplied by a constant equal to the square of the norm of the filter weighting function. Therefore, using (I-36) and (I-19) we obtain for the variance, σ_{mJ}^2, of the estimate of the m^{th} derivative using a J^{th} order smoother.

$$\sigma_{mJ}^2 = \sigma_o^2 \left\|W_J^{(m)}(i)\right\|^2$$

which is our desired result. σ_{mJ}^2 may be expressed directly in terms of the orthonormal polynomials as follows,

$$\sigma_{mJ}^2 = \sigma_o^2 \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} \sum_{j=0}^{J} \sum_{k=0}^{J} f_j^{(m)}(\tau/\ t)\ f_k^{(m)}(\tau/\Delta t)\ f_j(t_i/\Delta t)\ f_k(t_i/\Delta t)$$

$$\sigma_{mJ}^2 = \sigma_o^2 \sum_{j=0}^{J} \sum_{k=0}^{J} f_j^{(m)}(\tau/\Delta t)\ f_k^{(m)}(\tau/\Delta t) \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} f_j(t_i/\Delta t)\ f_k(t_i/\Delta t) \tag{I-37}$$

Using the orthogonality relationship of $f_j(t_i/\ t)$ given by equation (I-4), (I-37) reduces to

$$\sigma_{mJ}^2 = \sigma_o^2 \sum_{j=0}^{J} \left[f_j^{(m)}(\tau/\Delta t) \right]^2 \tag{I-38}$$

which is another useful form of the results.

APPENDIX II. PROOF OF SOME IMPORTANT THEOREMS[18]

Theorem 1. Every polynomial $x(t)$ of degree J can be expressed as a linear combination of the polynomials $f_k(t)$ [described by equations (I-3), (I-4), (I-5)]

Proof. The proof follows from the fact that $f_k(t)$ is a polynomial of degree k with nonzero coefficient of t^k, where $k = 0,1,\ldots,J$. For $k \geq 1$, $f_k(t)$ is of the form

$$At^k + R_{k-1}(t), \tag{II-1}$$

where A = constant and $R_{k-1}(t)$ is a polynomial of degree k-1. If we replace t^J in $x(t)$ by

$$\frac{f_J(t) - R_{J-1}(t)}{A} \tag{II-2}$$

the result will be of the form

$$C_J f_J(t) + U_{J-1}(t) \tag{II-3}$$

where $U_{J-1}(t)$ is a polynomial of degree J-1. Now replace t^{J-1} in $U_{J-1}(t)$ by

$$\frac{f_{J-1}(t) - R_{J-2}(t)}{B} \tag{II-4}$$

which transforms equation (II-3) into an expression of the form

$$C_J f_J(t) + C_{J-1} f_{J-1}(t) + U_{J-2}(t)$$

where $U_{J-2}(t)$ is a polynomial of degree J-1. Proceeding in this way, we will finally arrive at an expression of the form

$$x(t) = C_J f_J(t) + \ldots + C_1 f_1(t) + C_o \qquad \text{(II-5)}$$

which is the required linear combination (since $f_o(t)$ = constant).

<u>Theorem 2</u>. If U(i) is the weighting function of any discrete filter having zero dynamic error for all input polynomials of degree J, then

$$\|W(i)\|^2 \leq \|U(i)\|^2 \qquad \text{(II-6)}$$

with equality holding in (II-6) only if

$$W(i) = U(i) \qquad -\left(\frac{r-1}{2}\right) \leq i \leq \frac{r-1}{2}$$

where W(i) is the weighting function of the discrete J^{th} order polynomial filter, having zero dynamic error, described by equation (I-19).

<u>Proof</u>. Denote $U(i) - W(i)$ by $V(i)$. Thus

$$U(i) = W(i) + V(i) \tag{II-7}$$

$$\|U\|^2 = (U,U) = (W+V,W+U)$$

$$\|U\|^2 = (W,W) + (W,V) + (V,W) + (V,U) \tag{II-8}$$

$$\|U\|^2 = \|W\|^2 + 2(W,V) + \|V\|^2$$

To say that the filter with weighting function $U(i)$ has zero dynamic error for all input polynomials $x(t)$ of degree J is to say that

$$E\left[x^{*(m)}(t_1)\right] = \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} E\left[y(t_1)\right] U(i) = x^m(t_1) \tag{II-9}$$

for all input polynomials of degree J and all t_1. By hypothesis, $U(i)$ satisfies equation (II-9). By equation (I-2) for polynomial filters, (II-9) also holds when $U(i)$ is replaced by $W(i)$. By subtraction of the two equations, we obtain

$$\sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} x(t_1)[U(i) - W(i)] = 0 \tag{II-10}$$

for all i and all polynomials $x(t)$ of degree J. Thus $U(i) - W(i)$ is orthogonal over the interval $-\frac{(r-1)}{2} \leq i \leq \frac{(r-1)}{2}$ to all polynomials of degree J. In other words, by (II-7)

$$(Q,V) = 0$$

for all polynomials $Q(t)$ of degree J. But, $W(i)$ is a polynomial of degree J and hence is itself such a polynomial $Q(t)$. Hence $(W,V) = 0$ and thus by (II-8)

$$\|U\|^2 = \|W\|^2 + \|V\|^2 \qquad \text{(II-11)}$$

which proves (II-6).

Furthermore, by (II-11), equality holds in (II-6) only if $\|V\|^2 = 0$, i.e.,

$$\sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} [U(i) - W(i)]^2 = 0 \qquad \text{(II-12)}$$

The left side of equation (II-12) is the sum of a nonnegative function which can only be zero if

$$U(i) - W(i) = 0$$

Q.E.D.

From (I-36), the variance of the estimate obtained using a J^{th} order least squares polynomial smoother is less than that obtained with any other filter with zero dynamic error for a J^{th} degree input signal.

<u>Theorem 3</u>. Let $u_o, u_1, \ldots, u_J$ be given real numbers. Then there is a unique polynomial, W(t) of degree J, such that

$$(t^j, W) = u_j \qquad 0 \leq j \leq J \tag{II-13}$$

and can be represented as

$$W(t) = \sum_{j=0}^{J} u_j \, W^j(t) \tag{II-14}$$

where $W^j(t)$ is a polynomial of degree J defined by its moments

$$(t^k, W^j) = \begin{cases} 0 & \text{if } k \neq j \\ 1 & \text{if } k = j \end{cases} \tag{II-15}$$

<u>Proof</u>. Using the orthogonality relation for $f_j(t/\Delta t)$, we have, from (I-4)

$$(f_j, f_k) = 0 \tag{II-16}$$

Supposing W(t) of degree J to exist, let us prove the uniqueness of W(t). The polynomials $f_j(t)$ can be written in the form

$$f_k(t) = \sum_{j=0}^{J} H_j^k \, t^j \qquad 0 \leq k \leq J \qquad \text{(II-17)}$$

By (II-13) and (II-17) we have

$$(f_k, W) = \sum_{j=0}^{J} H_j^k \, u_j \qquad 0 \leq k \leq J \qquad \text{(II-18)}$$

From Theorem 1 we can write for any polynomial,

$$W(t) = \sum_{j=0}^{J} c_j \, f_j(t) \; .$$

From this and (II-16) we conclude

$$(f_j, W) = c_j(f_j, f_j) = c_j \qquad 0 \leq k \leq J \; .$$

Hence

$$W(t) = \sum_{k=0}^{J} (f_j, W) \; f_j(t) \qquad \text{(II-19)}$$

This proves the uniqueness of $W(t)$ since, by (II-18), (f_j, W) is uniquely determined by $u_o, u_1, \ldots, u_J$. Taking the inner product of t^k with both sides of (II-14) we have

$$(t^k, W) = \sum_{j=0}^{J} u_j \, (t^k, W^j) \qquad \text{(II-20)}$$

By (II-13) and (II-15)

$$u_k = \sum_{j=0}^{J} u_j \, \delta_{jk} = u_k$$

hence proving (II-14). The existence of W(t), is therefore demonstrated since a solution, (II-14), has been shown to satisfy the conditions of the theorem. Thus Theorem 3 is proven.

Theorem 4. Let $u_o, u_1, \ldots, u_J$ be given real numbers and let $W(t_i)$ be the polynomial of degree J such that

$$(t^j, W) = u_j \qquad 0 \leq j \leq J \qquad \text{(II-21)}$$

Then, for any function U(t) such that

$$(t^j, U) = u_j \qquad 0 \leq j \leq J \qquad \text{(II-22)}$$

holds, we have

$$\|U\|^2 = \|W\|^2 + \|U-W\|^2 \qquad \text{(II-23)}$$

<u>Proof</u>. Subtracting (II-21) from (II-22) we obtain

$$(t^j, U-W) = 0 \qquad 0 \leq j \leq J$$

Hence

$$(Q, U-W) = 0,$$

for all polynomials Q(t) of degree J. In particular, since W(t) is a polynomial of degree J, we have

$$(W, U-W) = 0 \tag{II-24}$$

In general, for any functions F(t) and G(t),

$$\|F+G\|^2 = \|F\|^2 + 2(F,G) + \|G\|^2$$

Setting F = W and G = U-W, we obtain

$$\|U\|^2 = \|W\|^2 + 2(W,U-W) + \|U-W\|^2$$

which, because of (II-28), yields the desired relation

$$\|U\|^2 = \|W\|^2 + \|U-W\|^2 .$$

<u>Theorem 5</u>. Given an input polynomial signal of degree J, then of all filters with given dynamic error coefficients, the filter which minimizes the output noise has a weighting function, $W(t_1)$ which is a polynomial of degree J.

__Proof__. The dynamic error coefficients, ε_{mj}, are given by

$$\varepsilon_{mj} = \left(t^j,\ W^{(m)}\right) - \left.\frac{d^m t^j}{dt^m}\right|_{t=\tau} \tag{II-25}$$

where

$u_j = (t^j, W) =$ the j^{th} moment of the weighting function

Hence the dynamic error coefficients determine the filter moments and therefore by Theorem 4, Theorem 5 is proven.

APPENDIX III. DERIVATION OF $W^i(t)$ POLYNOMIALS

The $W^i(t)$ polynomials are defined by their moments given as,

$$(t^j, W^i) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} = \delta_{ij} \quad \text{for} \quad 0 \leq i \leq J \qquad \text{(III-1)}$$

Using the orthogonal polynomials, $F_j(t)$, described in Appendix I, where

$$(F_j, F_k) = 0 \qquad j \neq k$$

$$(F_k, F_k) = B_k$$

and

$$F_k(t) = \sum_{j=0}^{J} A_j^k \, t^j \qquad 0 \leq k \leq J \qquad \text{(III-2)}$$

Taking the inner product of the k^{th} orthogonal polynomial with $W^i(t)$ we have

$$\left(F_k, W^i\right) = \sum_{j=0}^{J} A_j^k \, (t^j, W^i) \qquad 0 \leq k \leq J \quad \text{(III-3)}$$

Using (III-1) we have

$$\left(F_k, W^i\right) = \sum_{j=0}^{J} A_j^k \, \delta_{ji} = A_i^k \qquad \text{(III-4)}$$

where A_i^k is the coefficient of the i^{th} power of $t(t = i\Delta t)$ in the orthogonal polynomial, $F_k(t)$.

From Theorem 1, Appendix II, we can write,

$$W^i(t) = \sum_{\ell=0}^{J} g_\ell \, F_\ell(t) \tag{III-5}$$

Therefore

$$\left(F_k, W^i\right) = g_k(F_k, F_k) \tag{III-6}$$

and hence

$$W^i(t) = \sum_{k=0}^{J} \frac{\left(F_k, W^i\right)}{(F_k, F_k)} F_k(t) \tag{III-7}$$

Substituting (III-4) into (III-7) we obtain

$$W^i(t) = \sum_{k=0}^{J} \frac{A_i^k}{(F_k, F_k)} F_k(t) = \sum_{k=0}^{J} \frac{A_i^k}{B_k} F_k(t) \tag{III-8}$$

APPENDIX IV. DERIVATION OF SUBFILTER WEIGHTS

IV.1 Bayes Estimate of Prob ($|a_J| \leq a_{TH}$)

The Bayes estimate of the probability that $-a_{TH} \leq a_J \leq a_{TH}$, given an estimate of a_J, $\hat{a}_J$, is required. Since $\hat{a}_J$ is obtained using a J^{th} order least squares polynomial smoother, and the noise samples are assumed to be zero mean, independent and Gaussian, $\hat{a}_J$ is a random variable with a Gaussian distribution, whose mean value is a_J and variance, $\sigma_{\hat{a}_J}$ is given by equation (I-38). The probability density function of $\hat{a}_J$ for some given value of a_J is

$$p\left(\hat{a}_J \mid a_J\right) = \frac{1}{\sqrt{2\pi}\ \sigma_{\hat{a}_J}} e^{-\frac{\left(\hat{a}_J - a_J\right)^2}{2\sigma^2_{\hat{a}_J}}} \qquad \text{(IV-1)}$$

Since a_J is not actually known, a priori, we shall assume a_J has a uniform distribution between finite limits and finally take the limit of our results as these limits go to infinity. In particular we assume that

$$p(a_J) = \frac{1}{a} \qquad -\frac{a}{2} \leq a_J \leq \frac{a}{2} \quad \text{and} \qquad \text{(IV-2)}$$

eventually let $a \to \infty$.

With the above information as introduction we are now interested in finding the Bayes' estimate of

$$\text{Prob}\left(|a_J| \leq a_{TH} \mid \hat{a}_J\right) = \int_{-a_{TH}}^{a_{TH}} p\left(a_J \mid \hat{a}_J\right) da_J \qquad \text{(IV-3)}$$

Using the theorem of conditional probabilities,

$$p\left(a_J \mid \hat{a}_J\right) = \frac{p\left(a_J, \hat{a}_J\right)}{p\left(\hat{a}_J\right)} \qquad \text{(IV-4)}$$

and

$$p\left(a_J, \hat{a}_J\right) = p\left(\hat{a}_J \mid a_J\right) p(a_J) \qquad \text{(IV-5)}$$

Integrating equation (IV-5)

$$p\left(\hat{a}_J\right) = \int_{-\infty}^{\infty} p(a_J)\ p\left(\hat{a}_J \mid a_J\right) da_J \qquad \text{(IV-6)}$$

Substituting (IV-5) and (IV-6) into (IV-5) we obtain

$$p\left(a_J \mid \hat{a}_J\right) = \frac{p\left(\hat{a}_J \mid a_J\right) p(a_J)}{\int_{-\infty}^{\infty} p(a_J)\ p\left(\hat{a}_J \mid a_J\right) da_J} \qquad \text{(IV-7)}$$

Substituting (IV-1) and (IV-2) into (IV-7) yields

$$p\left(a_J \mid \hat{a}_J\right) = \frac{\dfrac{1}{a}\dfrac{1}{\sqrt{2\pi}\,\sigma_{\hat{a}_J}} \exp\left\{-\dfrac{\left(\hat{a}_J - a_J\right)^2}{2\sigma^2_{\hat{a}_J}}\right\}}{\dfrac{1}{2}\displaystyle\int_{-a/2}^{a/2} p\left(\hat{a}_J \mid a_J\right) da_J}$$

$$p\left(a_J \mid \hat{a}_J\right) = \frac{\dfrac{1}{\sqrt{2\pi}\,\sigma_{\hat{a}_J}} \exp\left\{-\dfrac{\left(\hat{a}_J - a_J\right)^2}{2\sigma^2_{\hat{a}_J}}\right\}}{\displaystyle\int_{-a/2}^{a/2} \dfrac{1}{\sqrt{2\pi}\,\sigma_{\hat{a}_J}} \exp\left\{-\dfrac{\left(\hat{a}_J - a_J\right)^2}{2\sigma^2_{\hat{a}_J}}\right\} da_J} \tag{IV-8}$$

Integrating (IV-8)

$$\text{Probability}\left(\left| a_J \right| \le a_{TH} \mid \hat{a}_J\right) = \int_{-a_{TH}}^{a_{TH}} p\left(a_J \mid \hat{a}_J\right) da_J$$

$$\text{Prob}\left(\left| a_J \right| \le a_{TH} \mid \hat{a}_J\right) = \frac{\dfrac{1}{2}\left[\operatorname{erfc}\left(\dfrac{-a_{TH}-\hat{a}_J}{\sqrt{2}\,\sigma_{\hat{a}_J}}\right) - \operatorname{erfc}\left(\dfrac{a_{TH}-\hat{a}_J}{\sqrt{2}\,\sigma_{\hat{a}_J}}\right)\right]}{\displaystyle\int_{-a/2}^{a/2} \dfrac{1}{\sqrt{2\pi}\,\sigma_{\hat{a}_J}} \exp\left\{-\dfrac{\left(\hat{a}_J - a_J\right)^2}{2\sigma^2_{\hat{a}}}\right\} da_J} \tag{IV-9}$$

Taking limit of (IV-9) as $a \to \infty$ yields

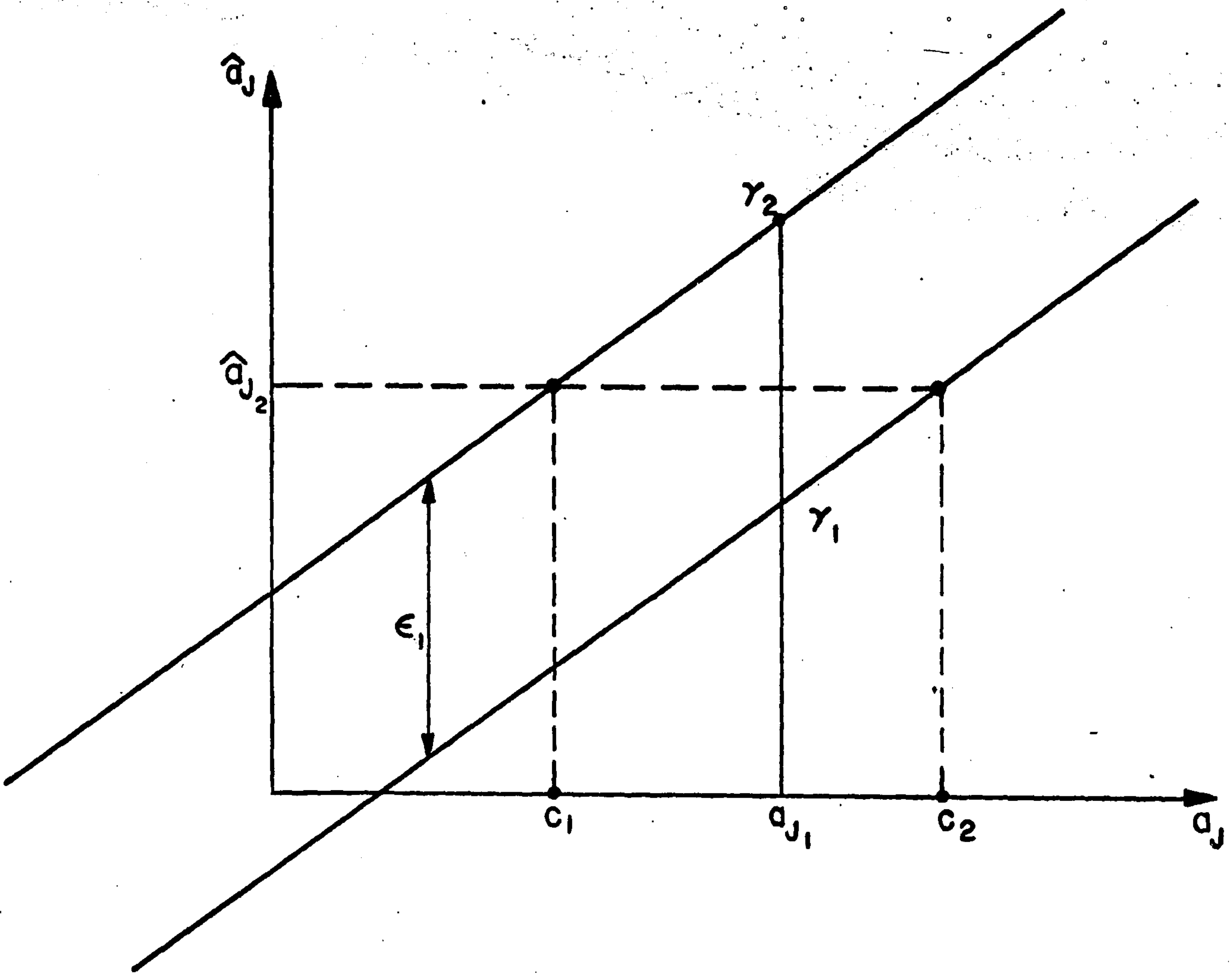
$$\text{Prob}\left(|a_J| \leq a_{TH} \mid \hat{a}_J\right) = \frac{1}{2}\left[\text{erfc}\left(\frac{-a_{TH}-\hat{a}_J}{\sqrt{2}\,\sigma_{\hat{a}_J}}\right) - \text{erfc}\left(\frac{a_{TH}-\hat{a}_J}{\sqrt{2}\,\sigma_{\hat{a}_J}}\right)\right] \tag{IV-10}$$

IV.2 Confidence Interval[12] Method of Obtaining Probability $\left(|a_J| \leq a_{TH}\right)$

The Bayes' approach used in the previous section makes the statement that "the probability of a_J being situated between given fixed limits is equal to some ε." If in fact a_J is not a random variable, questions arise as to the meaning and sense of such a statement. The method of Confidence Intervals, however, makes the statement that "the probability that some fixed limits include between them the value of the parameter, a_J, corresponding to the actual sample, is equal to ε."

Keeping these statements in mind, we now find the probability that the range of values between $-a_{TH}$ and a_{TH} include the value of a_J which corresponds to the actual sample.

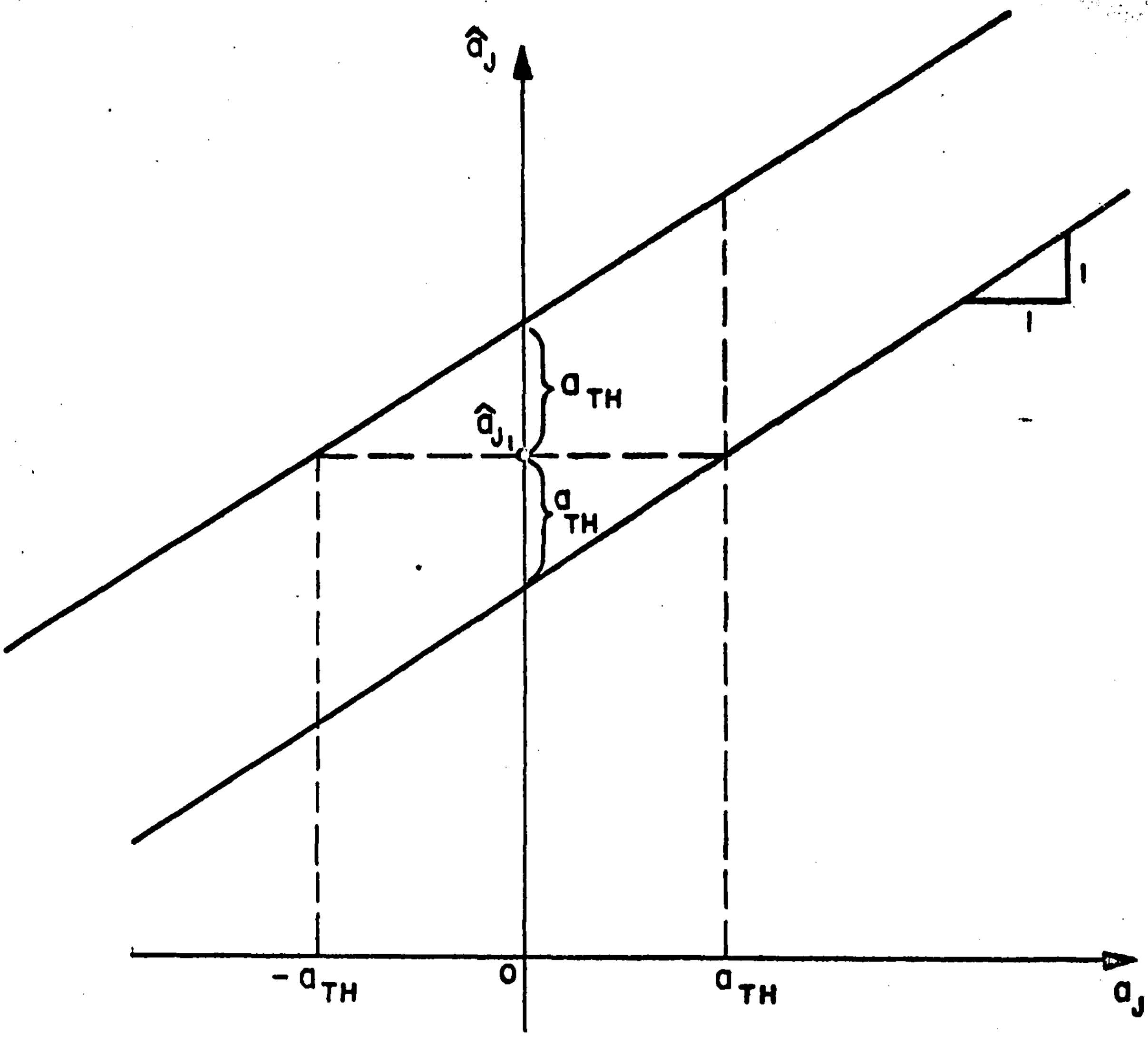
Consider the a_J vs. $\hat{a}_J$ plane shown in Figure (8). For some value a_{J_1} of a_J, two limits of $\hat{a}_J$, γ_1 and γ_2, are



CONFIDENCE LIMITS

FIG 8

selected, such that the probability of $\hat{a}_J$ falling within this region (γ_1,γ_2) is equal to ε. The curves obtained when this is done for all a_J are called confidence curves for some confidence level equal to ε. For a different ε, different curves are obtained. The relationships between γ_1, γ_2, ε and a_J, are obtained using $p\left(\hat{a}_J \mid a_J\right)$ given by equation (IV-1). If an estimate of a_J, $\hat{a}_{J_2}$, is obtained, it may be stated[12] that the unknown value of the parameter, a_J, lies within the confidence interval (C_1,C_2), or between the confidence limits C_1 and C_2 with a confidence level equal to ε, where $\varepsilon = \int_{\gamma_1}^{\gamma_2} p\left(\hat{a}_J \mid a_J\right) d\hat{a}_J$. Since in equation (IV-1), a_J appears only as the mean of a normal distribution, the (γ_1,γ_2) interval will shift linearly with unity slope for different values of a_J if (γ_1,γ_2) is selected over the same portion of $p\left(\hat{a}_J \mid a_J\right)$ relative to the mean (a_J) for all values of a_J. Consider Figure (9) where $\hat{a}_{J_1}$ is an estimate of a_J for a given set of observations. We are interested in determining with that confidence level we can say that a_J lies within the confidence limits of $a_J = -a_{TH}$ and $a_J = +a_{TH}$. Confidence curves can be constructed with unity slope passing through the points $\left(\hat{a}_{J_1}, a_{TH}\right)$ and $\left(\hat{a}_{J_1}, -a_{TH}\right)$. It is now required to find the appropriate confidence level, ε, for the resultant curves. This is easily accomplished by selecting any



CONFIDENCE LIMITS

FIG 9

arbitrary value of a_J, say $a_J = 0$, and finding the probability that $\hat{a}_J$, lies between the intersection of the confidence curves with the $a_J = 0$ line. Using equation (IV-1)

$$\varepsilon = \int_{\hat{a}_{J_1} - a_{TH}}^{\hat{a}_{J_1} + a_{TH}} \frac{1}{\sqrt{2\pi}\,\sigma_{\hat{a}_J}} e^{-\frac{\left[\hat{a}_J\right]^2}{2\sigma^2_{\hat{a}_J}}} da_J = \int_{-a_{TH}}^{a_{TH}} e^{-\frac{\left[\hat{a}_J - \hat{a}_{J_1}\right]^2}{2\sigma^2_{\hat{a}_J}}} d\hat{a}_J \tag{IV-11}$$

Integration yields

$$\varepsilon = \frac{1}{2}\left[\operatorname{erfc}\left(\frac{-a_{TH} - \hat{a}_{J_1}}{\sqrt{2}\,\sigma_{\hat{a}_J}}\right) - \operatorname{erfc}\left(\frac{a_{TH} - \hat{a}_{J_1}}{\sqrt{2}\,\sigma_{\hat{a}_J}}\right)\right] \tag{IV-12}$$

which is the same result as equation (IV-10) which was found using Bayes' Theorem.

APPENDIX V. DERIVATION OF STATISTICS FOR EXAMPLE IN SECTION 9.1

Define a_o to be a sample value of a random variable, x. Let x be the random variable with cumulative probability distribution, $P(x)$ = Probability $(x \leq X)$, defined by the following model (function of random variables, x_1 and x_2): a_o is a sample of either one of two Gaussian random variables, x_1 and x_2. The probability that a_o is a sample of x_1 is P_1 and of x_2 is P_2, where $P_1 + P_2 = 1$. The means and variances of x_1 and x_2 are $\left(m_1, \sigma_1^2\right)$ and $\left(m_2, \sigma_2^2\right)$ respectively and $p_1(x_1)$ and $p_2(x_2)$ represent their probability density functions. The mean, m_x, and variance, σ_x^2, of x is required.

$$P(x) = \text{Probability } (x \leq X)$$

which is equivalently the joint probability of x_1 being selected and $x_1 \leq X$, and that x_2 is selected and $x_2 \leq X$ i.e.

$$P(x) = \text{Probability } \left[(1, x_1 \leq X) \text{ and } (2, x_2 \leq X)\right].$$

Since the selection of a sample from x_1 and x_2 are mutually exclusive,

$$P(x) = \text{Probability } (1, x_1 \leq X) + \text{Probability } (2, x_2 \leq X)$$

The selection of a random variable (x_1 or x_2) is independent from the sample values of the random variables. Hence

$$\text{Probability } (1, x_1 \leq X) = \text{Probability } (1) \cdot \text{Probability } (x_1 \leq X)$$

or

$$Pr(1, x_1 \leq X) = P_1 Pr(x_1 \leq X)$$

and

$$Pr(2, x_2 \leq X) = P_2 Pr(x_2 \leq X)$$

Therefore

$$P(x) = P_1 Pr(x_1 \leq X) + P_2 Pr(x_2 \leq X)$$

Differentiating $P(x)$ we obtain the probability density function, $p(x)$, of x.

$$p(x) = P_1 p_1(x) + P_2 p_2(x)$$

The mean of x, m_x is given as $E[x]$ where

$$E[x] = m_x = \int_{-\infty}^{\infty} x \ p(x) \ dx = \int_{-\infty}^{\infty} x \left[P_1 p_1(x) + P_2 p_2(x) \right] dx$$

$$m_x = P_1 \int_{-\infty}^{\infty} x \; p_1(x) \; dx + P_2 \int_{-\infty}^{\infty} x \; p_2(x) \; dx$$

which is

$$m_x = P_1 m_1 + P_2 m_2$$

The variance of x, σ_x^2 is given as

$$\sigma_x^2 = E[x^2] - [E(x)]^2 = E[x^2] - m_x^2$$

where

$$E[x^2] = \int_{-\infty}^{\infty} x^2 \; p(x) \; dx = \int_{-\infty}^{\infty} x^2 \left[P_1 p_1(x) + P_2 p_2(x) \right] dx$$

$$E[x^2] = P_1 \int_{-\infty}^{\infty} x^2 \; p_1(x) \; dx + P_2 \int_{-\infty}^{\infty} x^2 \; p_2(x) \; dx$$

$$E[x^2] = P_1 \left[\sigma_1^2 + m_1^2 \right] + P_2 \left[\sigma_2^2 + m_2^2 \right]$$

Therefore

$$\sigma_x^2 = P_1 \left[\sigma_1^2 + m_1^2 \right] + P_2 \left[\sigma_2^2 + m_2^2 \right] - \left[P_1 m_1 + P_2 m_2 \right]^2$$

If

$$m_1 = m_2 = 0$$

then $$m_x = 0$$

and

$$\sigma_x^2 = P_1\sigma_1^2 + P_2\sigma_2^2$$

APPENDIX VI. OPTIMUM UNBIASED POLYNOMIAL SMOOTHERS

In order to obtain an unbiased minimum mean square error estimate, we allow the form of the estimate of the m^{th} derivative of the input signal to be a constant term, G_m, plus a linear combination of the observed data, $y(i)$.

$$\hat{x}^{(m)}(\tau) = G_m + \sum_{i=-\frac{r-1}{2}}^{\frac{r-1}{2}} y(i)\, W^{(m)}(i) \qquad \text{(VI-1)}$$

Using the identical procedure as in Chapter 1.0, the dynamic error is given as,

$$D_m = G_m + \sum_{j=0}^{J} a_j\, \varepsilon_{mj} \qquad \text{(VI-2)}$$

To insure an unbiased estimate,

$$E[D_m] = E\left[G_m + \sum_{j=0}^{J} a_j\, \varepsilon_{mj}\right] = 0 \qquad \text{(VI-3)}$$

Rearranging (VI-3) and noting that $E[a_j] = m_j$,

$$\bar{G}_m = - \sum_{j=0}^{J} m_j\, \varepsilon_{mj} \qquad \text{(VI-4)}$$

Substituting (VI-4) into (VI-1) we obtain

$$\hat{x}^{(m)}(\tau) = - \sum_{j=0}^{J} m_j \, \varepsilon_{mj} + \sum_{i=-\frac{(r-1)}{2}}^{\frac{r-1}{2}} y(i) \, W^{(m)}(i) \qquad \text{(VI-5)}$$

The dynamic error associated with the above estimate is, from (VI-2) and (VI-4), given as,

$$D_m = \sum_{j=0}^{J} (a_j - m_j) \, \varepsilon_{mj} \qquad \text{(VI-6)}$$

Using (VI-6) and the fact that the output noise variance is given as

$$E[N^2] = \sigma_o^2 \, \|W^{(m)}\|^2$$

we obtain for the resultant mean square error,

$$\begin{aligned} E\left\{[N+D_m]^2\right\} &= E[N^2] + E[2ND_m] + E\left[D_m^2\right] \\ &= \sigma_o^2 \, \|W^{(m)}\|^2 + E\left[2N \sum_{j=0}^{J} (a_j - m_j) \, \varepsilon_{mj}\right] \\ &\quad + E\left[\sum_{j=0}^{J} \sum_{k=0}^{J} (a_j - m_j)(a_k - m_k) \, \varepsilon_{mj} \varepsilon_{mk}\right] \end{aligned} \qquad \text{(VI-7)}$$

Noting that $E[N] = 0$ and that the observation noise has been assumed uncorrelated with the a_j random variable, equation (VI-7) can be simplified to

$$E\left[[N+D_m]^2\right] = \sigma_o^2 \|W^{(m)}\|^2 + \sum_{j=0}^{J} \sum_{k=0}^{J} \varepsilon_{mj}\, \varepsilon_{mk} \left(E[a_j a_k] - m_j m_k\right) \tag{VI-8}$$

Substituting (5-14) into (VI-8) we obtain,

$$E\left[[N+D_m]^2\right] = \sigma_o^2 \|U^{(m)}\|^2 + \sum_{j=0}^{J} \sum_{k=0}^{J} \varepsilon_{mj}\, \varepsilon_{mk} \left(E[a_j a_k] - m_j m_k + \sigma_o^2 (W^j, W^k)\right) + 2\sigma_o^2 \sum_{j=0}^{J} \varepsilon_{mj} \left(W^j,\ U^{(m)}\right) \tag{VI-9}$$

Differentiating (VI-9) with respect to each of the ε_{mj} and equating to zero yields J+1 linear equations in J+1 unknowns. The n^{th} equation of the J+1 total equations is given as

$$\sum_{j=0}^{J} \varepsilon_{mj} \left[\sigma_o^2 (W^n, W^j) + E(a_n a_j) - m_n m_j\right] = -\sigma_o^2 \left(W^n,\ U^{(m)}\right) \tag{VI-10}$$

The solution of (VI-10) yield the dynamic error coefficients, ε_{mj} of the optimum unbiased filter. The mean square error of this filter is then given by (VI-9).

BIBLIOGRAPHY

1. Alterman, S. B., "Dynamic and Noise Errors in Least Square Polynomial Data Smoothers", Bell Telephone Labs., MM-62-4212-13, Oct. 1962.

2. Alterman, S. B., "Dynamic Correction of Smoothed Data for Re-entering ICBM Tracking", Bell Telephone Labs., MM-63-4212-7, April 1963.

3. Balakrishnan, A. V., "An Adaptive Non-linear Data Predictor", Proceed. of the National Telemetry Symposium, 1962.

4. Benedict, T. R. and M. M. Sondhi, "On a Property of Wiener Filters", Proc. IRE, Vol. 45, pp. 1021-1022, July 1957.

5. Benner, A. H. and Drenick, R., "An Adaptive Servo System", IRE Convention Record, Part 4, 1955.

6. Bergen, A. R., "Optimum Filtering of Sampled Polynomial Message Plus Random Noise", Columbia Univ. Electronics Research Labs., TR T-1/133, May 1956.

7. Blackman, R. B., "Smoothing and Prediction of Time Series by Cascaded Simple Averages", IRE International Convention Record, Vol. 8, p. 2, pp. 47-54, 1960.

8. Blum, M., "An Extension of the Minimum Mean Square Prediction Theory for Sampled Input Signals, IRE Trans. on Information Theory, IT-2,3, Sept. 1956, pp. 176-184.

9. Blum, M., "Fixed Memory Least Squares Filters Using Recursion Methods", IRE Trans. on Information Theory, Vol. IT-3, pp. 178-181, Sept. 1957.

10. Blum, M., "On the Mean Square Noise Power of an Optimum Linear Discrete Filter Operating on Polynomial Plus White Noise Input", IRE Trans. on Information Theory, Vol. IT-3, pp. 225-231, Dec. 1957.

11. Blum, M., "Recursion Formulas for Growing Memory Digital Filters", IRE Trans. on Info. Theory, Vol. IT-4, pp. 24-30, March 1958.

12. Cramér, Harold, "Methods of Mathematical Statistics", Princeton University Press, 1951.

13. Darlington, S., "Linear Least-Squares Smoothing and Prediction, with Applications", Bell Sys. Tech. J., Vol. 37, pp. 1221-1294, September 1958.

14. Follin, J. W. and Bucy, R. S., "Adaptive Finite Time Filtering", Trans. PG-AC, July 1962, pp. 169-178.

15. Franklin, G. F., "Linear Filtering of Sampled Data", 1955 IRE 4, Convention Record, Part 4, pp. 119-128.

16. Franklin, G. F., "Filtering Deterministic Messages from Sampled Data", Columbia University Electronics Research Laboratory Technical Note, TN-3/127, June 1956.

17. Franklin, G. F., "Is This an Adaptive System?", AIEE Paper CP 59-345, Feb. 1959.

18. Howard, W. A. and Rausch, R. G., "Polynomial Filters", Unpublished Book.

19. Johnson, K. R., "Optimum, Linear, Discrete, Filtering of Signals Containing a Nonrandom Component", IRE Trans., Vol. IT-2 (June 1956), pp. 49-55.

20. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, Series D, Journal of Basic Engineering, Vol. 82, 1960, pp. 35-45.

21. Kalman, R. E., "New Methods and Results in Linear Prediction and Filtering Theory", RIAS Technical Rept., 61-1.

22. Kalman, R. E. and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory", J. Basic Engineering (ASME Trans.) 83D (1961), 95-108.

23. Kolmogoroff, A., "Interpolation and Extrapolation von statinaren zufalligen Folgen". Bull. Ac. Sci. USSR Ser. Math. 5, 3-14, 1942.

24. Kushner, "Efficient Iterative Methods for Optimizing the Performance of Multiparameter Noisy Systems", Lincoln Labs. Rept. 22G-0043, October 1960.

25. Lees, A. B., "Interpolation and Extrapolation of Sampled Data", IRE Trans. on Information Theory, IT-2, 1 March 1956, pp. 12-17.

26. Magill, D. T., "Optimal Adaptive Estimation of Sampled Stochastic Processes", Tech. Rept. No. 6302-3 Stanford Electronics Lab., Dec. 1963.

27. Sakrison, D. J., "Application of Stochastic Approximation Methods to System Optimization", D.S.C. Thesis M.I.T., May 1961.

28. Sherman, S., "Non-Mean-Square Error Criteria", IRE Trans. on Information Theory, Vol. IT-4, pp. 125-127, September 1958.

29. Shaw, L. G., "Dual Mode Filtering", Tech. Rept. #2106-2, Stanford Electronics Laboratories, Sept. 1960.

30. Weaver, C. S., "Adaptive Communication Filtering", Trans. PGIT, Sept. 1962, pp. 169-178.

31. Wiener, N., "The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications", New York, Wiley, 1949.

32. Zadeh, L. A. and J. R. Ragazzini, "An Extension of Wiener's Theory of Prediction", Journal of Applied Physics, Vol. 21, 1950, pp. 645-655.

33. Zadeh, L. A. and Ragazzini, J. R., "Optimum Filters for the Detection of Signals in Noise", Proc. IRE 40, 1223-1231, 1952.

BIOGRAPHY

Stanley B. Alterman was born in in . He obtained his primary and secondary education in the New York City school system and in 1958 received the B.E.E. degree, Cum Laude, from the City College of New York. After moving to New Jersey in 1958, Mr. Alterman received the M.S.E.E. degree from Stevens Institute of Technology and the Sc. D.E.E. degree from Newark College of Engineering in 1961 and 1965 respectively.

During the period from 1958 to 1961, he worked for I.T.T. Laboratories where his principal efforts were in the fields of radar and communication countermeasures system studies and design. Since 1961 he has been a member of the technical staff of Bell Telephone Laboratories, where he has been engaged in theoretical studies of data processing and discrimination techniques for ballistic missile defense systems. In 1962 he was appointed to the position of Adjunct Professor of Electrical Engineering in the Graduate Division of Newark College of Engineering.

Mr. Alterman is an active member of the Wayne Jaycee's and was founder and past president of a local civic association. He and his wife, Enid, with their two children, Betsy-Jo and Eric Dwight, presently live in Wayne, New Jersey.