Maximally Reliable Exponential Prediction Equations for Data-Rate-Limited Tracking Servomechanisms*

By H. D. HELMS

(Manuscript received July 8, 1965)

Certain radars, sonars, and other sampling instruments periodically measure a variable but can measure accurately only if prediction equations provide the instrument with an accurate prediction of the next value of this variable. For these instruments, it is appropriate to define reliability as the probability that the error in this prediction does not exceed some limit. In choosing the form and parameters of the prediction equations, it is reasonable to attempt to maximize this reliability.

Assumptions that the prediction equations utilize linear error measurements, are recursive, and provide least-squares smoothing with an exponential weighting function establish a realistic basis for calculating the reliability. The first through the third orders of these equations predict the variable as reliably as possible in the presence of large initial errors in the variable and its velocity, provided that the smoothing interval of the prediction equations is sufficiently short. The dynamic error component of the prediction error of these equations is proportional to a smoothed version of the qth time-derivative of the variable, where q is the order of the prediction equations. The assumption that the measurement errors are uncorrelated and stationary makes it possible to calculate the standard deviation of the random component of the prediction error.

On the assumption that the random component of the prediction error has a normal (or Gaussian) probability distribution, there exists a safety factor which is monotonically related to the reliability. The choice of the smoothing interval of the prediction equations which maximizes this safety factor can be found, which in turn permits the optimum safety factor to be calculated. The ratio of pairs of these optimum safety factors determines which order

* This paper reflects a study supported by the Army's Nike-X Project Office, Redstone Arsenal, Alabama.

of prediction equations gives the greatest reliability in the "worst case" situation in which the first unestimated time-derivative of the variable assumes its largest possible value. Graphs containing the foregoing results make it convenient to examine the tradeoffs between the reliability, the time between measurements, and other parameters. A numerical example is given.

TABLE OF CONTENTS

I. INTRODUCTION

Certain instruments attempt to determine the value of a variable by measuring the difference between it and a prediction of it at regular intervals of time. Examples of these instruments include echo-ranging radars and sonars. These measurements enter a computer which immediately substitutes the measurements into equations which predict the next value of the variable, thereby closing the loop illustrated in Fig. 1. This prediction must be supplied to the instrument because it is assumed that the instrument can measure sufficiently accurately only if it knows approximately where to look for the next value of the variable. Conversely, large errors in this prediction are assumed to blind or otherwise confuse the instrument.

More precisely, it is assumed that positive limits L and L' on the prediction error E are given, where E is defined as the difference between the prediction $\hat{x}$ of the variable and its true value x. It is assumed that if E stays within the interval $-L' \leqq E \leqq L$, the instrument almost never grossly mismeasures the variable. However, even when E stays within this interval, it is assumed that the instrument slightly mis-

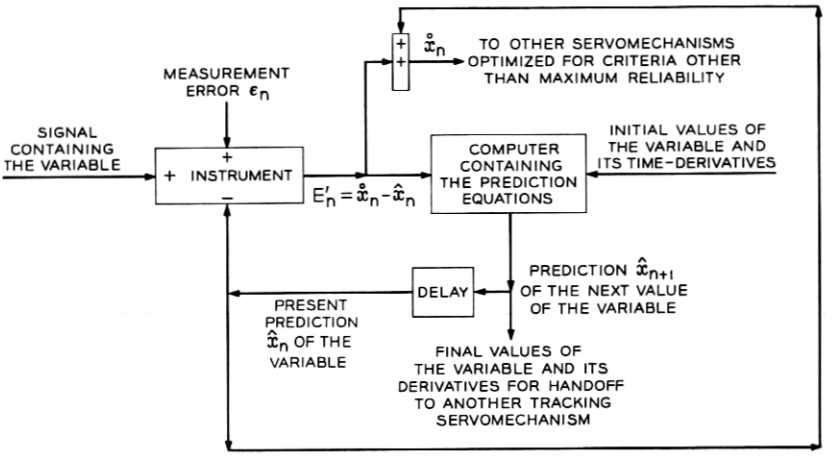

Fig. 1 — Tracking servomechanism.

measures the variable by contributing an additive random error ϵ to the measurements E' of the prediction error. The quantities L, L', E, $\hat{x}$, x, ϵ, and E' are illustrated in Fig. 2. Because E' enters the prediction equations, $\hat{x}$ contains random error caused by ϵ. These random errors help make it possible for E to exceed L or to become less than $-L'$. An example of the situation described in this paragraph is given in Ref. 1.

Reliability is defined herein as the probability that E stays within the interval $-L' \leqq E \leqq L$. Designers usually seem to feel that there is a need to maximize reliability defined in this manner. It must be admitted that different designers always seem to choose slightly different values of L and L', but it almost always turns out that the results in this paper are not affected significantly by slight differences in these values.

After selecting a relatively simple class of prediction equations which

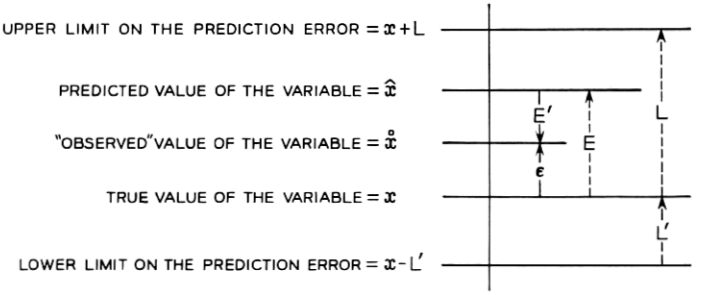

Fig. 2 — Terminology.

assume that the variable is represented in a Taylor's series, this paper attempts to explain how to choose the order and the smoothing interval of these prediction equations. The criterion used in choosing these parameters is that of maximizing reliability at the instant when the variable x is exhibiting that behavior (termed "worst case") which tends to minimize reliability. Using the same class of prediction equations (in a form less convenient for computation than the equations of this paper) and some of the assumptions employed herein, R. G. Brown[3] proposed the same objective as that of this paper, but the ill-defined character of Brown's time-series (e.g., the history of an inventory) prevented him from explicitly determining an optimum order and smoothing interval.

Reliability can be maximized by maximizing a quantity which is monotonically related to it. As discussed at the end of this section, it is assumed that the probability that $E < -L'$ can be neglected in comparison with the probability that $E > L$. Assuming also that E is normally distributed [Chapter 7 of Ref. 7], the reliability is a monotonically increasing function of the safety factor

$$\lambda = \frac{L - W}{\sigma_R} \tag{1}$$

where W is defined as the mean of E, so that $W \equiv \bar{E}$. Also, R is defined as $R \equiv E - W$ (so that $\bar{R} \equiv 0$), and σ_R denotes the standard deviation of R.

It is arbitrarily assumed that the prediction equations are linear difference equations and that the mean value $\bar{\epsilon}_n$ of each of the measurement errors is zero. In consequence, the following simplifications are evident: (*i*) W, called the dynamic error, is caused solely by the inability of the prediction equations to predict more complicated time-histories of the variable than the equations are designed to predict, and (*ii*) R, called the random error component of the prediction error, is caused solely by the measurement error. In consequence of the first simplification and of the fact that λ in (1) decreases as the dynamic error W increases, the "worst case" behavior of the variable x is that behavior which maximizes W. This maximum value of W is hereafter referred to as reaching its "worst case" value W_c.

The "worst case" value W_c can be calculated with the aid of a theorem (given later) which implies that W is essentially proportional to the qth time-derivative of the variable; $q = 1, 2$, or 3 is the order of the prediction equations used in this paper. Consequently, it is necessary to know the "worst case" bounds (i.e., the greatest lower bound or the least upper bound) on the first, second, or third time-derivatives (here-

after referred to, respectively, as the velocity v, acceleration a, or jerk j) of the variable x. These "worst case" bounds often can be deduced from constraints imposed by physics or physiology.

It is assumed that the random error component R has a normal (or Gaussian) probability distribution. This assumption is always satisfied if the measurement errors have normal probability distributions. [However, if the measurement errors are not normally distributed, the averaging effect of the prediction equations often makes the random error component distribution approximately normal (as suggested by the Central Limit Theorem)].[7] Each measurement error ϵ_n is assumed to occur in a random manner independently of every other measurement error, and all ϵ_n are assumed to have equal standard deviations. This standard deviation (or equivalently this rms value) need not be known prior to selecting the parameters of the prediction equations, but it must be known prior to calculating the reliability expected from the selected prediction equations.

Just prior to (1), it was assumed that the probability that $E < -L'$ is negligibly small compared to the probability that $E > L$. This assumption is equivalent to the statement that, measured in units of size equal to σ_R, W_c is much closer to L than to $-L'$. Alternatively, if W_c is much closer to $-L'$ than to L, the right side of (1) is replaced by $(L' + W_c)/\sigma_R$; this replacement does not change the arguments of this paper significantly. In the seemingly uncommon situation in which, after applying the results of Sections VI and VII, it turns out that W_c is not much closer to one of the limits than it is to the other, the methods of this paper do not maximize the reliability even though λ is maximum. The methods used in Section VI also assume that W_c has the same sign as the limit to which it is much closer (i.e., W_c positive if L is much closer, or W_c negative if $-L'$ is much closer); this assumption seems very likely to be satisfied.

II. Determining the Form of the Prediction Equations (Prior to Maximizing Reliability)

It is pertinent to list some reasonable objectives which the prediction equations should attain.

To make it possible to achieve a level of reliability which is satisfactory at all times, provided that the reliability is large enough at some time, it is sufficient if W_c and σ_R (and therefore also λ) do not vary as time passes. If the standard deviation of the measurement errors and the behavior of the parameter do not vary with time, this constancy of

W_c and σ_R can be achieved by using prediction equations whose form and parameters do not change as time passes. Such prediction equations effectively smooth over a constant interval of preceding measurements. (Prediction equations of this sort may exhibit an initial period, whose duration equals the smoothing interval, during which W and σ_R vary. In practice, reliability usually is not too small during this initial period, because the "worst case" behavior of the variable usually does not occur at this time.)

The tradeoffs involved in maximizing the reliability of the prediction equations should be as obvious as possible. One method of achieving this is to make the prediction equations contain only a single independent parameter, such as the length of the interval over which preceding measurements are smoothed.

The prediction equations should be easy to initialize and should readily provide information by which the track can be handed off to other prediction equations. These objectives suggest that the prediction equations should explicitly estimate the time-derivatives of the variable.

The prediction equations should occupy as little storage space in the computer as possible. This objective suggests the use of recursive equations, which require storing only the most recent values of all quantities appearing in these equations. The need for keeping computation time to a minimum implies that the prediction equations should be few and easy to calculate.

It appears that the foregoing objectives can be attained by arbitrarily assuming that the variable is represented by a Taylor's series whose coefficients satisfy the "exponential smoothing" criterion. This criterion makes the expected (or mean) values of the sum of the exponentially weighted squares of the measured prediction error as small as possible. (A source of confusion may exist because the definition of the term "exponential smoothing" in this paper is more specific than another apparently obsolete definition of this term, which a few people have used to denote *any* set of recursive equations having constant coefficients.) Stating the exponential smoothing assumption algebraically, the quantity

$$\sum_{i=o}^{\infty} {E'}_{n-i}^{2} K^i$$

[where E' denotes the measurements of prediction error, where

$$K^i = \left(\frac{N-1}{N+1}\right)^i$$

is the exponential weighting coefficient (with $N \geqq 1$ being called the

smoothing interval), and n denotes the time of the latest prediction] represents the sum of the exponentially weighted squares of E'. This sum is to be minimized by differentiating it separately with respect to each of the coefficients of the Taylor's series representing the variable, setting the differentials equal to zero, and solving the resulting set of equations simultaneously to obtain the optimum values of the Taylor's series coefficients. These steps are explained by Levine.[2]

Recursive prediction equations which satisfy this requirement can be obtained from Levine[2] by setting his weighting coefficient ω_i equal to K^{-i}, taking the limit as n approaches infinity, and substituting the results into his equations (15) through (17) or else (54) and (56). Similar prediction equations satisfying the exponential smoothing criterion appear in Refs. 3 or 11. Using Levine's notation, the prediction equations for first, second, and third-order smoothing (corresponding to a Taylor's series containing one, two, or three coefficients) can be rearranged to minimize the amount of data-storage and computation time required. The rearranged prediction equations are given in the next three subsections.

2.1 *First-Order Prediction Equations*

$$\hat{x}_{n+1} = \hat{x}_n + \alpha_1 E_n' \tag{2}$$

$$E_n' = \mathring{x}_n - \hat{x}_n \tag{3}$$

$$\alpha_1 = 1 - K \tag{4}$$

$$K = \frac{N-1}{N+1}, \qquad N \geqq 1, \tag{5}$$

where N is a constant, where $\hat{x}$ is the predicted value of the parameter, and where $\mathring{x}$ denotes the "observed" value of the variable, which can be calculated by adding $\hat{x}$ to the measured prediction error E'. (The subscript 1 in α_1 is different from Levine's subscript in that now the subscript denotes the order of the smoothing equations instead of denoting the time of the measurement being processed by Levine's equations.)

2.2 *Second-Order Prediction Equations*

$$\hat{u}_{n+1} = \hat{u}_n + \beta_2 E_n' \tag{6}$$

$$\hat{x}_{n+1} = \hat{x}_n + \hat{u}_{n+1} + \alpha_2 E_n' \tag{7}$$

$$\beta_2 = (1-K)^2 \tag{8}$$

$$\alpha_2 = 1 - K^2 \tag{9}$$

$$\hat{u}_n = \hat{v}_n T, \tag{10}$$

where K and E' are defined in (3) and (5), where $\hat{v}$ is the predicted velocity, and where T is the constant interval of time between the measurements of E'. Equation (6) is supposed to be computed immediately before (7) is computed. As a result, $\hat{u}$ and $\hat{x}$ can be stored in a single word of computer memory each. A similar statement can be made about $\hat{s}$, $\hat{u}$, and $\hat{x}$ in the next three equations.

2.3 *Third-Order Prediction Equations*

$$\hat{s}_{n+1} = \hat{s}_n + \gamma_3 E_n' \tag{11}$$

$$\hat{u}_{n+1} = \hat{u}_n + 2\hat{s}_{n+1} + \beta_3 E_n' \tag{12}$$

$$\hat{x}_{n+1} = \hat{x}_n + \hat{u}_{n+1} - \hat{s}_{n+1} + \alpha_3 E_n' \tag{13}$$

$$\gamma_3 = \tfrac{1}{2}(1 - K)^3 \tag{14}$$

$$\beta_3 = \tfrac{3}{2}(1 - K^2)(1 - K) \tag{15}$$

$$\alpha_3 = 1 - K^3 \tag{16}$$

$$\hat{s}_n = \tfrac{1}{2}\hat{a}_n T^2, \tag{17}$$

where $\hat{u}$, K, and E' are defined in (10), (5), and (3), and where $\hat{a}$ is the predicted acceleration.

For convenience, a graph of the coefficients calculated in (4), (8), and (9), and (14) through (16) is presented in Fig. 3. Equations (2), (6), and (7), or (11) through (13) are assumed to be initialized by using *a priori* estimates of the variable and (if required) its velocity and acceleration to calculate the initial value of $\hat{x}$, $\hat{u}$, and $\hat{s}$ with the aid of (10) and (17). Conversely, (2), (6), and (7), or (11) through (13) together with (10) and (17) provide estimates of $\hat{x}$, $\hat{v}$, and $\hat{a}$ which can be used to transfer (or hand off) the track to another set of prediction equations by providing initial predictions for this new set. [Of course, $\hat{a}$ is available only from (11) and $\hat{v}$ from (12) or (6).] These predictions can be extended to any future time merely by using a Taylor's series. For example,

$$\hat{x}_{n+(t/T)} = \hat{x}_n + \left(\frac{t}{T}\right)\hat{u}_n + \left(\frac{t}{T}\right)^2 \hat{s}_n . \tag{18}$$

Alternatively, if it is desired to predict $\hat{x}$, $\hat{u}$, and $\hat{s}$ at any integral multiple of T in the future, it suffices merely to calculate (2), (6), and (7), or (11) through (13) in appropriate number of times but with the coefficients α, β, γ set equal to zero.

The single independent parameter N in (5) is identified as the smooth-

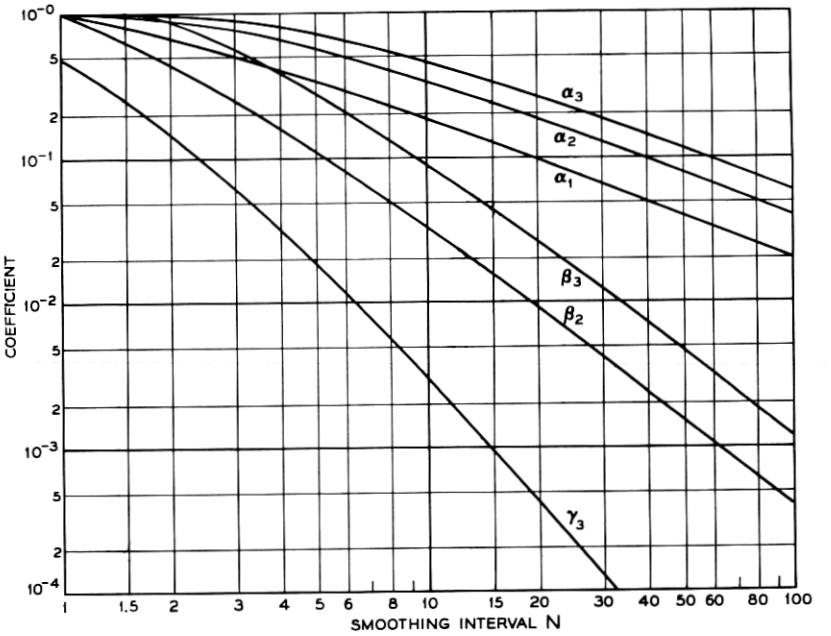

Fig. 3 — Coefficients of exponential smoothing equations.

ing interval because (2), (6), and (7), and (11) through (13) have nearly the same amounts of dynamic and random error components as polynomial smoothing equations (Ref. 4) designed to smooth over the last N samples. Additionally, K is approximated closely by exp $(-2/N)$ for all $N \geqq 1.5$, thereby making N be approximately the exponential decay constant of $\sqrt{K}$ (which in Levine's paper[2] is set equal to the reciprocal of the standard deviation).

III. COMPARISON OF THE EXPONENTIAL SMOOTHING CRITERION WITH OTHER OPTIMIZATION CRITERIA

A different set of prediction equations designed specifically for maximum reliability in acquiring the track, under the assumptions that the standard deviations of the errors in the initial values of $\hat{x}$ and $\hat{u}$ (supplied from some external source of information) are much larger than L and that the standard deviation of the measurement error is much smaller than L, would predict the second value of $\hat{x}$ by adding the *a priori* estimate of $\hat{u}$ to the first value of $\hat{x}$. This procedure permits the initial velocity to be in error by as much as L/T without E ever exceeding L.

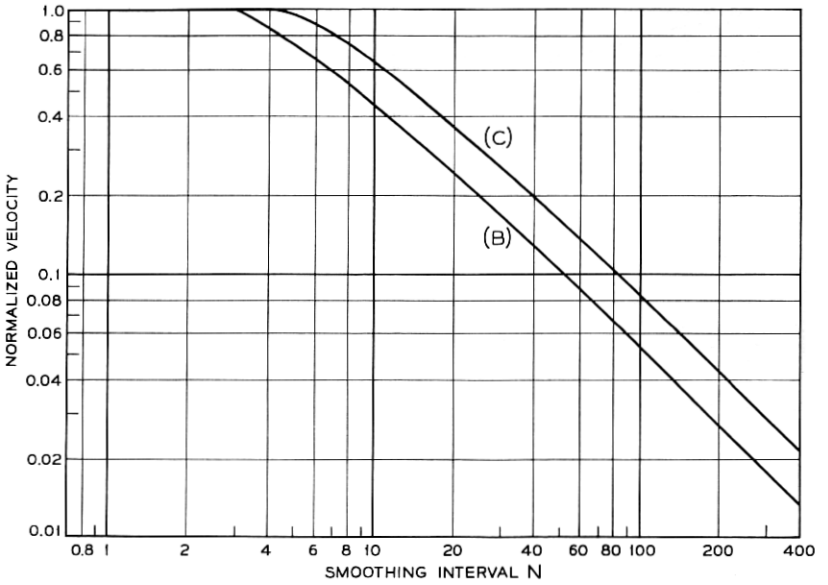

Fig. 4 — Maximum permissable initial error in velocity for (6) and (7) and (11) through (13) divided by L/T; (b) second order; (c) third order.

A comparison between the effectiveness of (6) and (7) or (11) through (13) and the above prediction equations is shown in Fig. 4, which shows the interval of permissable initial velocities of (6) and (7) or (11) through (13) divided by L/T. (Simulations reveal that the region of permissable initial positions and velocities approximates a horizontal rectangle on the prediction error phase plane, whose position error axis is intersected by the sides of the rectangle at $\pm L$.) Fig. 4 indicates that if N is small enough, (6) and (7) and (11) through (13) perform as reliably as the above set of prediction equations because the ratios are unity.

Evaluating the extent to which (6) and (7) satisfy another optimization criterion, Benedict and Bordner[5] state that critically damped equations equivalent to (6) and (7) give virtually the same accuracy as the set of recursive second-order smoothing equations which minimize a weighted sum of the random and the dynamic errors.

IV. DYNAMIC ERROR COMPONENT CALCULATIONS

The following theorem, which is illustrated in Fig. 5, provides a basis for calculating the dynamic error.

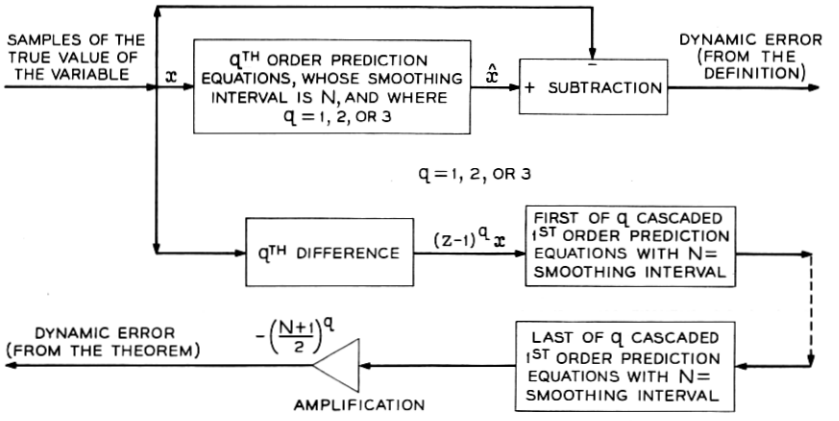

Fig. 5 — Illustration of the theorem.

Theorem: Let $q = 1, 2,$ *or* 3 *denote the order of the prediction equations, and let the* q*th difference of the variable* x *be defined as* $(z - 1)^q x$*, where* z *is the advance operator (which denotes taking the next sample of the time series on which it is operating, as in the example* $(z - 1)x_n = x_{n+1} - x_n$*). Then, the dynamic error component equals the product of* $-[(N + 1)/2]^q$ *and the final output of* q *identical first-order prediction equations having the same smoothing interval as the original prediction equations and forming a series connection in which the output of one equation is the input to the next equation, with the initial input being the* q*th difference of the variable.* A proof of the theorem is given in Appendix A.

A natural and intuitively satisfying interpretation of this theorem is obtained from the fact that the qth *difference* of the variable closely approximates the qth *derivative* of the variable multiplied by T^q. The q cascaded first-order filters, acting like simple low-pass resistance-capacitance filters, tend to attenuate all rapid fluctuations in the qth derivative and to delay the change in the dynamic error caused by a nonzero $q + 1$th derivative by $q(N + 1)/2$ samples. Thus, to estimate the worst case dynamic error of the qth order prediction equations, it suffices simply to multiply the maximum value of the variable's qth derivative (averaged over a total smoothing interval of qN samples) by $-T^q[(N + 1)/2]^q$, where $q = 1, 2,$ or 3.

To complete this treatment of dynamic error, the dynamic error component of $\hat{u}$ or $\hat{s}$ can be defined as $\hat{u}$ or $\hat{s}$ minus the first difference or one half of the second difference of the variable. [In computing the dynamic error of u for the third-order prediction equations, it was necessary to

define the true velocity operator as $\frac{1}{2}(z - 1/z)$, instead of $(z - 1)$.] It is possible to use the methods of Appendix A to prove theorems about these dynamic errors; these theorems can be interpreted in nearly the same way as in the previous paragraph. These results, along with those of the previous paragraph, are presented in Fig. 6. The dynamic error coefficients are to be multiplied by $-vT$ in the first-order case, $-aT^2$ in the second-order case, and $-jT^3$ in the third-order case. The delays through the various connections of first-order filters (each of which introduces a delay of $NT/2$) are contained in Table I. The dynamic error coefficients plotted in Fig. 6 are listed in Table II.

V. RANDOM ERROR COMPONENT CALCULATIONS

Fig. 7 shows the coefficients of the standard deviation (or rms value) of the random error component of $\hat{x}$, $\hat{u}$, and $\hat{s}$, and Table III lists the asymptotic behaviors of these coefficients for large values of N. (Fig. 7

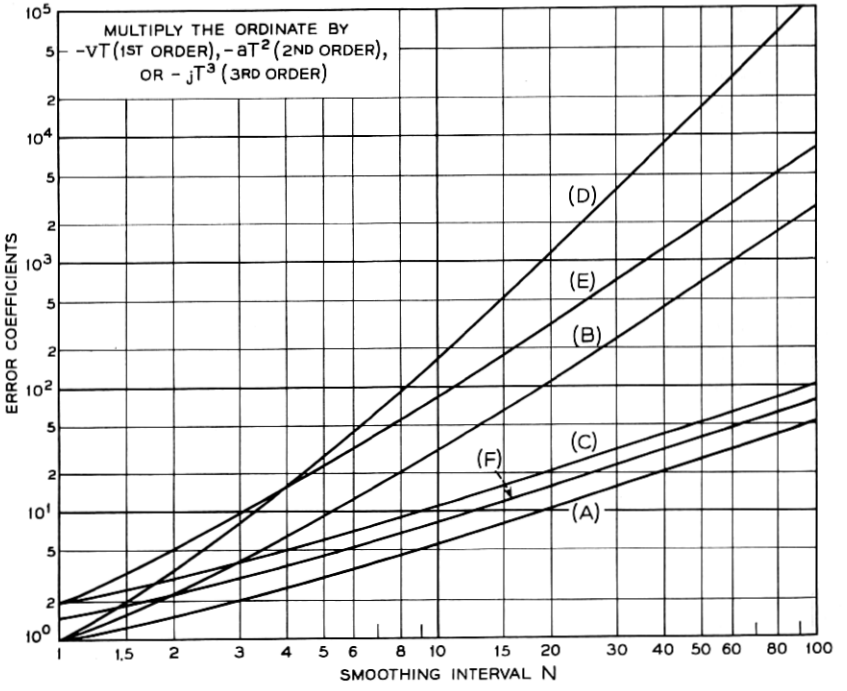

Fig. 6 — Dynamic error — first order, (a) $\hat{x}$; second order, (b) $\hat{x}$, (c) $\hat{u}$; third order, (d) $\hat{x}$, (e) $\hat{u}$, (f) $\hat{s}$.

TABLE I — DELAYS OF DYNAMIC ERROR FOR qTH ORDER PREDICTION EQUATIONS

	$q = 1$	2	3
$\hat{x}$	1	2	3
$\hat{u}$	—	$\frac{3}{2}$	$\frac{8}{3}$
$\hat{s}$	—	—	2

Multiply by $NT/2$

was obtained by numerically calculating the square root of the sums of squares of the quantities estimated by (2), (6), and (7) and (11) through (13) in response to $\mathring{x}_0 = 1$, $\mathring{x}_i = 0$, $i > 0$ [as in Ref. 5: (10) and Appendix I]. The actual standard deviations can be obtained by multiplying the coefficients shown in Fig. 7 and Table III by the standard deviation σ_ϵ of the measurement error.

Fig. 8 gives the correlation coefficients ρ (defined as the covariance divided by the product of the standard deviations of the quantities appearing in the covariance calculations), and Table IV lists their asymptotic values as N becomes very large. These correlation coefficients together with the standard deviation coefficients can be used to calculate the standard deviation of a prediction extended to any future time, because taking the expected value of the square of both sides of (18) gives

$$\begin{aligned}\sigma_{\hat{x}_{n+(t/T)}} &= [\overline{(\hat{x}_{n+(t/T)})^2}]^{\frac{1}{2}} \\ &= \left\{\sigma_{\hat{x}}^2 + \frac{t}{T}\,2\sigma_{\hat{x}}\sigma_{\hat{u}}\rho_{\hat{x}\hat{u}} + \left(\frac{t}{T}\right)^2(\sigma_{\hat{u}}^2 + 2\sigma_{\hat{x}}\sigma_{\hat{s}}\rho_{\hat{x}\hat{s}})\right. \\ &\quad \left. + \left(\frac{t}{T}\right)^3 2\sigma_{\hat{u}}\sigma_{\hat{s}}\rho_{\hat{u}\hat{s}} + \left(\frac{t}{T}\right)^4\sigma_{\hat{s}}^2\right\}^{\frac{1}{2}}.\end{aligned} \tag{19}$$

TABLE II — DYNAMIC ERROR COEFFICIENTS

	$q = 1$	2	3
$\hat{x}$	$D_1 = \frac{N+1}{2}$	$D_2 = \frac{(N+1)^2}{4}$	$D_3 = \frac{(N+1)^3}{8}$
$\hat{u}$	—	$N+1$	$\frac{(N+1)^2}{4}\left[3 - \frac{2}{N+1} + \frac{2}{(N+1)^2}\right]$
$\hat{s}$	—	—	$\frac{3(N+1)}{4}$

TABLE III — ASYMPTOTIC BEHAVIORS OF THE STANDARD DEVIATION COEFFICIENTS

	$q = 1$	2	3
$\hat{x}$	$S_1 = \frac{1}{\sqrt{N}}$	$S_2 = \frac{\sqrt{2.5}}{\sqrt{N}} = \frac{1.58}{\sqrt{N}}$	$S_3 = \frac{\sqrt{4.125}}{\sqrt{N}} = \frac{2.03}{\sqrt{N}}$
$\hat{u}$	—	$\frac{\sqrt{2}}{N^{3/2}} = \frac{1.414}{N^{3/2}}$	$\frac{\sqrt{14}}{N^{3/2}} = \frac{3.74}{N^{3/2}}$
$\hat{s}$	—	—	$\frac{\sqrt{1.5}}{N^{5/2}} = \frac{1.224}{N^{5/2}}$

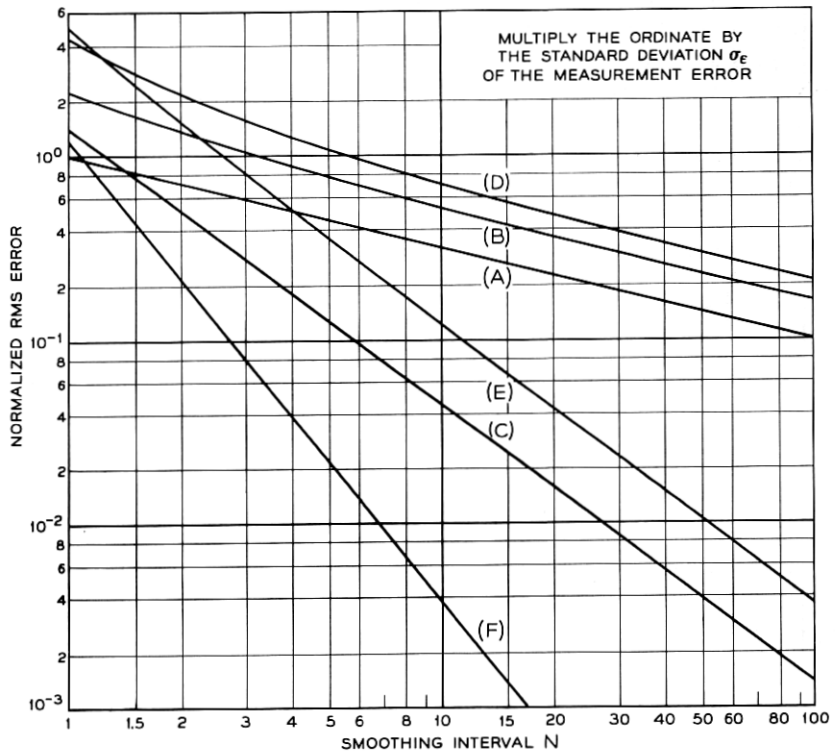

Fig. 7 — Standard deviations — first order, (a) $\hat{x}$; second order, (b) $\hat{x}$, (c) $\hat{u}$; third order, (d) $\hat{x}$, (e) $\hat{u}$, (f) $\hat{s}$.

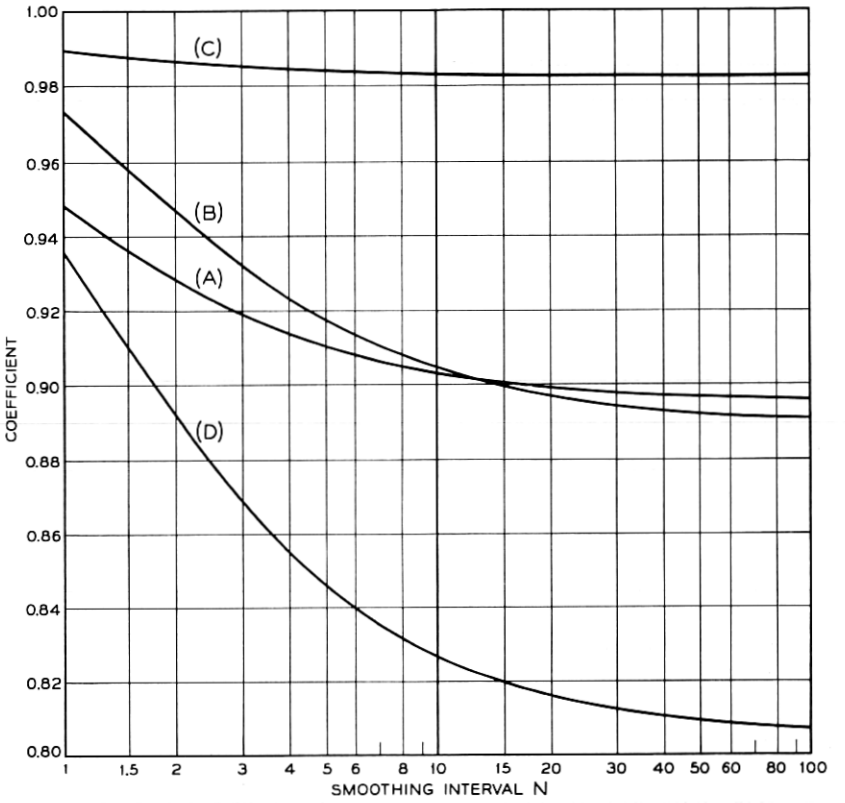

Fig. 8 — Correlation coefficients — second order, (a) $\hat{x}$, $\hat{u}$; third order, (b) $\hat{x}$, $\hat{u}$ (c) $\hat{u}$, $\hat{s}$, (d) $\hat{x}$, $\hat{s}$.

TABLE IV — ASYMPTOTIC VALUES OF THE CORRELATION COEFFICIENTS ρ

	$q = 2$	3
$\hat{x}$, $\hat{u}$	$\frac{2}{\sqrt{5}} = 0.895$	$\frac{27}{2\sqrt{231}} = 0.889$
$\hat{x}$, $\hat{s}$	—	$\frac{8}{\sqrt{99}} = 0.804$
$\hat{u}$, $\hat{s}$	—	$\frac{9}{2\sqrt{21}} = 0.983$

VI. CHOOSING THE SMOOTHING INTERVAL FOR MAXIMUM RELIABILITY

Substituting the expressions for W and σ_R shown as graphs on Figs. 6 and 7 into (1) to obtain the "worst case" safety factor produces, with v_c, a_c, and j_c denoting the "worst-case" velocity, acceleration, and jerk,

$$\lambda_1 = \frac{\left|\frac{L}{v_c T}\right| - D_1}{S_1} \cdot \frac{|v_c T|}{\sigma_\epsilon},$$

or

$$\lambda_2 = \frac{\left|\frac{L}{a_c T^2}\right| - D_2}{S_2} \cdot \frac{|a_c T^2|}{\sigma_\epsilon}, \tag{20}$$

or

$$\lambda_3 = \frac{\left|\frac{L}{j_c T^3}\right| - D_3}{S_3} \cdot \frac{|j_c T^3|}{\sigma_\epsilon},$$

where D and S, respectively, denote the dynamic and random error coefficients of the predicted value of the variable, and where the subscripts denote the order q of the prediction equations. To carry out the maximization by choosing N to maximize λ, it suffices to consider only the first factor in each of the three expressions in (20), because only D and S are functions of N. Although this statement is always true, λ does not completely determine the reliability if the dynamic error remains at its "worst case" value W_c for more than approximately N samples, because E would have two or more independent opportunities to exceed L. However, it is believed that this additional source of unreliability can be disregarded when choosing N to maximize the reliability, because as N varies the reliability depends on λ much more strongly than it depends on this additional source.

The first factors in (20) can be written as

$$\lambda_1' = \frac{P_1 - D_1}{S_1}, \quad \text{or } \lambda_2' = \frac{P_2 - D_2}{S_2}, \quad \text{or } \lambda_3' = \frac{P_3 - D_3}{S_3}, \tag{21}$$

where

$$P_1 = \left|\frac{L}{v_c T}\right|, \quad \text{or } P_2 = \left|\frac{L}{a_c T^2}\right|, \quad \text{or } P_3 = \left|\frac{L}{j_c T^3}\right|. \tag{22}$$

The values of N which maximize λ' are shown in Fig. 9 as a function of

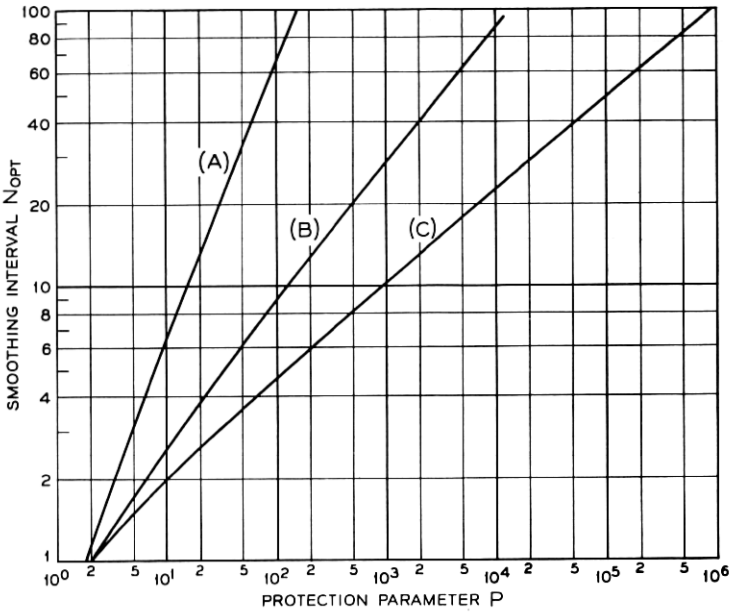

Fig. 9 — Optimum smoothing intervals versus protection parameters — (a) first order, (b) second order, (c) third order.

P, and the corresponding maximum values of λ' are shown in Fig. 10. Asymptotic formulas for these optimum values of N and λ' for large values of P can be obtained with the aid of Tables II and III as

$$N_{1_{opt}} = 2/3\ P_1$$
$$N_{2_{opt}} = 0.903\ \sqrt{P_2} \tag{23}$$
$$N_{3_{opt}} = 1.045\ \sqrt[3]{P_3}$$

$$\lambda_{1_{opt}}{}' = 0.544\,(P_1)^{3/2}$$
$$\lambda_{2_{opt}}{}' = 0.478\,(P_2)^{5/4} \tag{24}$$
$$\lambda_{3_{opt}}{}' = 0.432\,(P_3)^{7/6}.$$

These formulas provide very close approximations to Figs. 9 and 10 for $N_{opt} > 2.5$ approximately.

An important relationship obtained by substituting (22) into (23) is that, for all values of N greater than approximately 2.5, the effective smoothing time $N_{opt}\,T$ equals a constant (which depends only on L and the "worst case" value of the qth time derivative of the variable).

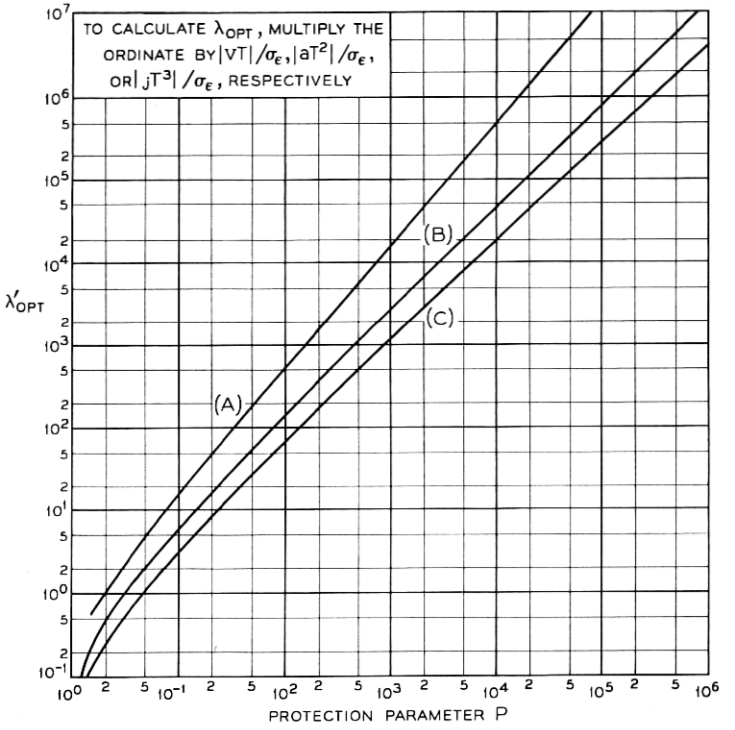

Fig. 10 — Optimum safety factors versus protection parameters — (a) first order, (b) second order, (c) third order.

A significant conclusion may be drawn from the fact that the function λ' in (21) is independent of the standard deviation of the measurement error σ_ϵ. Consequently, a single value of N (i.e., N_{opt}) simultaneously maximizes the reliability for all σ_ϵ.

VII. OPTIMIZING THE ORDER OF THE PREDICTION EQUATIONS

The order of the prediction equations can always be chosen to maximize the reliability by using the order which gives the largest value of the safety factor λ defined by (20) and shown in Fig. 10. A more general comparison between λ_1, λ_2, and λ_3 can be obtained for sufficiently large values of optimum smoothing intervals (i.e., $N_{opt} > 2.5$ approximately) by using the asymptotic formulas (24) to obtain the ratios, on the assumption that $L' = L$,

$$\frac{\lambda_1}{\lambda_2} = \frac{1.14 \mid L \mid^{1/4} \mid a_c \mid^{1/4}}{\mid v_c \mid^{1/2}}, \tag{25}$$

where v_c and a_c are the "worst case" velocity and acceleration, and

$$\frac{\lambda_2}{\lambda_3} = \frac{1.11 \mid L \mid^{1/12} \mid j_c \mid^{1/6}}{\mid a_c \mid^{1/4}}, \tag{26}$$

where j_c is the "worst case" jerk.

If (25) is greater than unity, then the optimum first-order prediction equations [which are optimum in the sense that they use the values N_{opt} given asymptotically by (23)] track more reliably than the optimum second-order prediction equations. Likewise, if (26) is greater than unity, then the optimum second-order prediction equations track more reliably than the optimum third-order prediction equations. The converse holds also.

It is significant that, if the time T between measurements is not so large that N is too small for (25) and (26) to hold, these equations indicate that the optimum order q_{opt} of the smoothing equations is optimum for all T (and σ_ϵ).

Examples illustrating these methods are given in Appendix B.

VIII. REDUCTION OF THE SAFETY FACTOR AND THE RELIABILITY DUE TO USING A NON-OPTIMUM SMOOTHING INTERVAL

Ratios of the actual safety factor obtained by using an arbitrary value of N to the optimum safety factor obtained by using N_{opt} from (23) can be obtained for large values of P by using the values of λ_{opt}' from (24) and using Tables II and III in conjunction with (21). A graph of these ratios is given in Fig. 11 as a function of N/N_{opt}. This graph indicates that it is worse to use a value of N larger than N_{opt} than it is to use a value proportionately smaller than N_{opt}.

The reduction in reliability caused by accepting a safety factor smaller than optimum can be calculated from tables of the normal distribution function [pp. 966–72 of Ref. 8]. For example, if $\lambda = 3$, the unreliability (i.e., unity minus the reliability) is only 0.135 percent, but the unreliability rises to 0.35, 0.82, 1.79, or 6.68 percent if λ decreases by 10, 20, 30, or 50 percent, respectively.

IX. CHOOSING THE TIME T BETWEEN MEASUREMENTS

The assumptions of this paper are not sufficient to determine an optimum value of T, but the methods of this paper do permit describ-

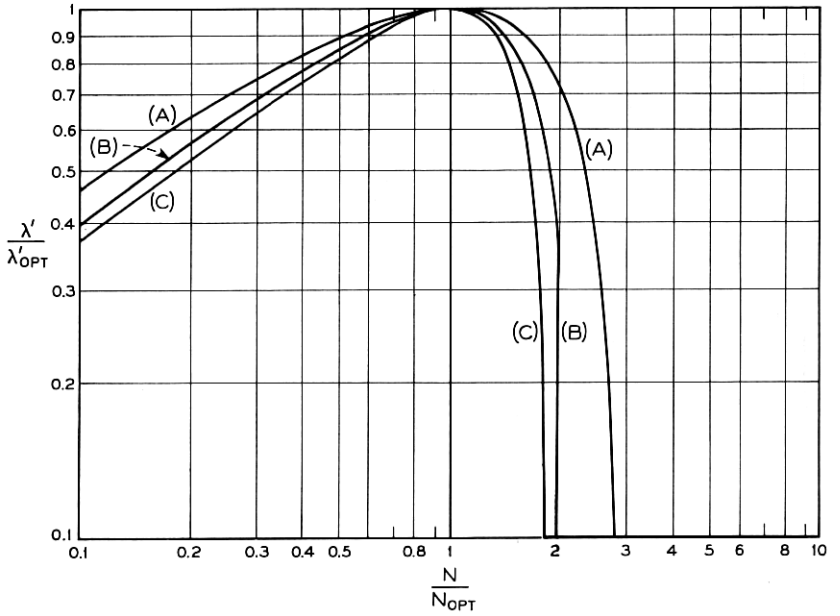

Fig. 11 — Relative safety factor versus the ratio of actual decay constant to the optimum decay constant.

ing the effects of various choices of T, if it is assumed that the optimum value of N is used and that N_{opt} is large enough (i.e., $N_{opt} > 2.5$ approximately) so that $N_{opt}\ T$ equals a constant.

One effect is that as T decreases (and N_{opt} increases), the maximum initial velocity error (for which E barely keeps from exceeding L) increases because it equals L/T until the appropriate break point shown in Fig. 4 is reached, beyond which point the maximum tolerable initial velocity error remains constant as T decreases.

Another effect is that as T increases, the standard deviations of $\hat{x}$, $\hat{v}$, and $\hat{a}$ [the latter two quantities being related to $\hat{u}$ and $\hat{s}$ by (10) and (17)] increase proportionately to $\sqrt{T}$. In consequence, an increase of T causes the safety factor to decrease as $1/\sqrt{T}$, because σ_R is proportional to $\sqrt{T}$, and because W is independent of T (due to the assumption that N_{opt} is large).

X. CONCLUSION

Except for the previous section, the objective of this paper has been the combining of limits on the maximum tolerable prediction error L

and the "worst case" time-derivatives v_c, a_c, or j_c with a time T between measurements, for the purpose of determining the optimal order q_{opt} and smoothing interval N_{opt} of the prediction equations. The results of this paper make it possible to perform parameter-variation studies in which q_{opt} and N_{opt} are assumed to be used, and in which tradeoffs between the following quantities are examined: (*i*) L, (*ii*) T, (*iii*) the standard deviation σ_ϵ of the measurement error, (*iv*) the reliability (defined as the probability that $-L' \leqq E \leqq L$ at the time of the "worst case" value v_c, a_c, or j_c), and (*v*) the "worst case" value v_c, a_c, or j_c.

XI. ACKNOWLEDGMENTS

A. R. Eckler provided sponsorship and guidance. R. T. Piotrowski and Miss A. T. Seery programmed the digital computer to prepare the graphs, which are plotted directly on microfilm with the aid of J. F. Kaiser's subroutines. A. J. Pellecchia (of Western Electric) and G. A. Jones (now of Pacific Gas and Electric Company) assisted. W. L. Nelson, F. A. Russell, and R. A. McDonald suggested useful revisions. The author is grateful to all of the above.

APPENDIX A

Proof of Dynamic Error Theorem

The first-order prediction (2) can be written in the operator notation as

$$\hat{x} = \frac{\alpha_1}{z + \alpha_1 - 1} \mathring{x} \tag{27}$$

where z is the advance operator for which $z(x_n) = x_{n+1}$. Calculating the dynamic error component of the prediction error E is accomplished by subtracting the true value x of the variable from $\hat{x}$ and then by setting the observed value $\mathring{x}$ of the variable equal to its true value x, because setting the (random) measurement error $\epsilon = \mathring{x} - x$ equal to zero removes the random error component from E. Thus, the dynamic error component of E equals

$$\begin{aligned} \hat{x} - x &= \frac{\alpha_1}{z + \alpha_1 - 1} \mathring{x} - x = \left[\frac{\alpha_1}{z + \alpha_1 - 1} - 1 \right] x \\ &= \frac{1}{\alpha_1} \cdot \left(\frac{\alpha_1}{z + \alpha_1 - 1} \right) [-(z - 1)x]. \end{aligned} \tag{28}$$

According to (4) and (5), $1/\alpha_1 = (N + 1)/2$, so that the dynamic error equals

$$\frac{N+1}{2}\left(\frac{\alpha_1}{z + \alpha_1 - 1}\right)[-(z-1)x]. \tag{29}$$

Equation (27) shows that (29) denotes passing the negative of the first difference of x through a first-order prediction equation which is identical to the first-order prediction equation being considered.

The second-order equations (6) and (7) can be written in operator notation and simultaneously solved for $\hat{x}$ to give

$$\hat{x} = \frac{[\beta_2 z + \alpha_2(z-1)]\mathring{x}}{\beta_2 z + \alpha_2(z-1) + (z-1)^2}. \tag{30}$$

Computing the dynamic error by subtracting x from both sides and setting $\mathring{x} = x$ as before makes (30) become

$$\hat{x} - x = \frac{-(z-1)^2 x}{\beta_2 z + \alpha_2(z-1) + (z-1)^2}. \tag{31}$$

Using (8) and (9) and then (4) and (5) gives

$$\begin{aligned} \hat{x} - x &= \frac{-(z-1)^2 x}{(z-K)^2} \\ &= \frac{[-(z-1)^2 x]}{\alpha_1^2}\left[\frac{\alpha_1}{z - 1 + \alpha_1}\right]^2 \\ &= \left(\frac{N+1}{2}\right)^2\left(\frac{\alpha_1}{z - 1 + \alpha_1}\right)^2 [-(z-1)^2 x]. \end{aligned} \tag{32}$$

Equation (32) represents two cascaded first-order prediction equations, whose smoothing interval N is the same as the smoothing interval of the original second-order prediction equations, operating on the negative of the second difference of the variable, as stated in the theorem.

The third-order prediction equations (11) through (13) can be written in operator notation and solved simultaneously to give

$$\hat{x} = \left[\frac{\alpha_3(z-1)^2 + \beta_3 z(z-1) + \gamma_3 z(z+1)}{(z-1)^3 + \alpha_3(z-1)^2 + \beta_3 z(z-1) + \gamma_3 z(z+1)}\right]\mathring{x}, \tag{33}$$

so that the dynamic error equals

$$\hat{x} - x = \frac{[-(z-1)^3 x]}{(z-1)^3 + \alpha_3(z-1)^2 + \beta_3 z(z-1) + \gamma_3 z(z+1)}. \tag{34}$$

Using (14) through (16) gives

$$\begin{aligned} \text{-- } x &= \frac{[-(z-1)^3 x]}{(z-K)^3} \\ &= \left(\frac{N+1}{2}\right)^3 \left(\frac{\alpha_1}{z-1+\alpha_1}\right)^3 [-(z-1)^3 x], \end{aligned} \tag{35}$$

where the smoothing interval N of the three cascaded first-order prediction equations is the same smoothing interval as that which appears in the original third-order prediction equations, as stated in the theorem.*

The abridgment of this theorem obtained by omitting the q cascaded first-order prediction equations {i.e., the dynamic error is approximated by multiplying the *unsmoothed* qth difference by $-[(N+1)/2]^q$} has approximated the dynamic error very accurately in simulations of the slowdown of ballistic devices re-entering the earth's atmosphere.

As an application of the theorem, the maximum dynamic error component for the second-order prediction equations (6) and (7) for an initial error u_0' in $\hat{u}$ can be approximated by performing the steps listed in the theorem. Thus, the second difference of $x = u_0' n$ for $n \geqq 0$ and $x = 0$ for $x < 0$ is an impulse of height u_0'. This impulse enters two cascaded first-order prediction equations, whose combined impulse response {obtained by convolving two exponentials together and multiplying by $[(N+1)/2]^2$} is approximately equal to $u_0' n \exp(-2n/N)$. The maximum value of this impulse response occurs at $n = N/2$, so that the maximum value of the dynamic error component of the prediction error E equals approximately $u_0'(N/2e)$.

Similarly, it is possible to calculate the dynamic error due to an initial error s_0' in $\hat{s}$ in the third-order equations (11) through (13) since the third difference of $x = s_0' n^2$ for $n \geqq 0$ and $x = 0$ for $n < 0$ is two successive impulses of height equal s_0'. These two impulses can be regarded as having the effect of a single impulse of height $2s_0'$ if the smoothing interval N is sufficiently large compared to two. The impulse response of three cascaded first-order prediction equations is

$$(n^2/2) \exp[-2(n-1)/(N+1)],$$

---

* For pth order exponential *smoothing* (as opposed to *prediction*) equations in which p is any integer, it is possible to prove that the dynamic error defined as $\bar{x} - x$ is equal to $[(N-1)/2]^p$ times the negative of the pth difference {defined as $[1-(1/z)]^p$} and passed through p first-order exponential *smoothing* equations of the same smoothing interval. (This proof starts with (7) of Ref. 10 and uses the lemma

$$(z-1)S\hat{x} = \frac{\alpha_1 z}{1-\alpha_1}[e - S]\hat{x},$$

where S is an operator denoting a first-order exponential smoothing equation.)

after multiplying by the $[(N + 1)/2]^3$ factor. Thus, the dynamic error component of the prediction error E can be approximated by $s_0'n^2 \exp [-2(n - 1)/(N + 1)]$. The maximum of this approximation occurs at $n = N + 1$, so that the maximum dynamic error equals approximately $s_0'(N + 1)^2 \exp [-2N/(N + 1)]$. (This approximation is never more than 12 percent larger than the true value of the dynamic error for any $N \geqq 7$. The expression $s_0'(N + 1)^2 \exp (-2)$ provides an approximation which is always smaller than the true value of the dynamic error.)

To conclude these examples of the use of the theorem, the dynamic error components of the prediction error E for the third-order prediction equations (11) through (13) caused by an initial error u_0' or x_0' in $\hat{u}$ or $\hat{x}$ can be approximated respectively by the first or second derivative with respect to n of $\frac{1}{2}n^2 \exp (-2n/N)$ multiplied by u_0' or x_0'.

APPENDIX B

Examples

It is necessary to illustrate how the assumptions and results discussed in the text apply to an actual instrument. Thus, echo-ranging radars have measurement errors which occur independently of the measurement error on any other pulse. The reason for this independence is that the interval T between measurements is always much larger than the reciprocal of the bandwidth of the instrument, because the time for the echo to return from the target always greatly exceeds the duration of the echo.

In monopulse radars measuring the range and angles of a target, thermal noise originating in the receiver often results in σ_ϵ being equal to the product of the following two quantities: (*i*) a large fraction of the pulsewidth or beamwidth, and (*ii*) the reciprocal of the square root of the ratio of the peak signal power to the average noise power [Chapter 10 of Ref. 6]. If quantization errors are present because the instrument measures the prediction error digitally, σ_ϵ can usually be calculated by taking the square root of the sum of squares of the standard deviations of the thermal and quantization errors.

At the end of Section VI of the text, it is stated that a single value of the smoothing interval N simultaneously maximizes the reliability for all σ_ϵ. In radar terminology, this value of N (i.e., N_{opt}) is optimum for any ratio of signal power to noise power. Likewise, Section VII concludes with a result which can be interpreted as stating that the optimum value q_{opt} of the order of the prediction equations is optimum for any data rate and any ratio of signal power to noise power.

The use of some of the equations and graphs presented previously is exemplified by the following: A radar emitting a 100 foot-wide pulse every 0.27 seconds is tracking the range of a descending spaceship whose maximum velocity is $-35{,}000$ ft/sec, whose maximum acceleration is 320 ft/(sec)^2, and whose maximum jerk is simulated to be 8.3 ft/(sec).[3] (These maxima occur at different times during the re-entry of the spaceship into the atmosphere.) It is discovered that the radar can measure the prediction error sufficiently accurately only if the prediction error $|E|$ is smaller than half the pulsewidth, so that $L = 50$ feet.* The standard deviation σ_ϵ of the measurement error ϵ is found to be approximately equal to $L/\sqrt{S/N}$, where S/N denotes the ratio of the instantaneously maximum signal power to the average noise power in the output of the radar's IF strip.

Equation (25) reveals that second-order smoothing is more reliable than first-order smoothing if $L < 2.26 \times 10^6$ feet, and (26) reveals that third-order smoothing is more reliable than second if $L < 4.36 \times 10^5$ feet. In consequence of assuming L to be only 50 feet, third-order smoothing should be used.

Using (22) to calculate P_3 gives $P_3 = 306$. Fig. 9 indicates that $N_{\text{opt}} = 7$, and Fig. 10 indicates that $\lambda' = 275$. Computing λ by multiplying λ' by $|j_c T^3|/\sigma_\epsilon$ gives $\lambda = 0.9\sqrt{S/N}$. If $S/N = 16 = 12$ db, $\lambda = 3.6$ and the "worst-case" reliability is 99.98 percent, according to pp. 966–972 of Ref. 8. Similarly, if $S/N = 4 = 6$ db, $\lambda = 1.8$, and the "worst-case" reliability is 96.4 percent. The values of α_3, β_3, and γ_3 corresponding to $N_{\text{opt}} = 7$ can be calculated with the aid of (14) through (16) or looked up on Fig. 3 as $\alpha_3 = 0.578$, $\beta_3 = 0.164$, and $\gamma_3 = 0.0078$, or, in exact octal fractions, $\alpha_3 = (0.45)_8$, $\beta_3 = (0.124)_8$, $\gamma_3 = (0.004)_8$.

The "worst case" dynamic error component W_c can be calculated with

* Because of this assumption that $L' = L$, it is also assumed that the dynamic error $|W_c|$ is large enough and σ_R is small enough so that

$$\frac{L + |W_c|}{\sigma_R} \gg \frac{L - |W_c|}{\sigma_R} \equiv \lambda,$$

thereby making the reliability be significantly affected only by changes in the value of λ. This assertion can be verified by calculating

$$\frac{L + |W_c|}{\sigma_R} \quad \text{and} \quad \frac{L - |W_c|}{\sigma_R}$$

with the aid of the values of W_c and σ_R given in the last paragraph of this appendix and using the tables of the normal distribution function (pp. 966–972 of Ref. 8) to compare the effects of changes in

$$\frac{L + |W_c|}{\sigma_R} \quad \text{and} \quad \frac{L - |W_c|}{\sigma_R}.$$

the aid of Fig. 6 as 64 times $j_c T^3$, or 10.5 feet. The standard deviation σ_R of the random component can be calculated with the aid of Fig. 7 as $0.88\ \sigma_\epsilon = 44'/\sqrt{S/N}$. If $S/N = 4 = 6$ db, $\sigma_R = 22$ feet and if $S/N = 16 = 12$ db, $\sigma_R = 11$ feet. As a check on the consistency of the results, (1) gives values of λ which are identical to those calculated in the previous paragraph.

REFERENCES

1. Helms, H. D., Random Walk Theory Probabilities That a Radar Servo Will Fail to Lock On, IEEE Transactions on Automatic Control, *AC-9*, No. 3, July, 1964, pp. 288–289.
2. Levine, N., Increasing the Flexibility of Recursive Least Squares Data Smoothing, B.S.T.J., *40*, May, 1961, pp. 821–840. Also Bell Telephone Laboratories Monograph, 3892.
3. Brown, R. G., *Smoothing, Forecasting, and Prediction of Discrete Time Series*, Prentice-Hall, Englewood Cliffs, N. J., 1963.
4. Nesline, F. William, Jr., Polynomial Filtering of Signals, Proc. Conv. Mil. Elec., 1961, pp. 531–542 (Tables VI and VIII).
5. Benedict, T. R. and Bordner, G. W., Synthesis of an Optimal Set of Radar Track-While-Scan Smoothing Equations, IRE Trans. on Automatic Control, *AC-7*, July, 1962, pp. 27–32.
6. Skolnik, Merrill I., *Introduction to Radar Systems*, McGraw-Hill Book Co., New York, 1962.
7. Feller, William, *An Introduction to Probability Theory and Its Applications*, 2nd ed., John Wiley & Sons, Inc., New York.
8. Abramowitz, Milton and Stegun, Irene C., Handbook of Mathematical Functions, Nat. Bur. of Stds., App. Math. Series No. 55, U. S. Dept. of Commerce, U. S. Govt. Printing Office, Washington, D. C., June, 1964.
9. Whittaker, E. T. and Robinson, G., *The Calculus of Observations*, 4th ed., Blackie & Son Ltd., London, 1944, p. 308.
10. Brown, R. A. and Meyer, R. F., The Fundamental Theorem of Exponential Smoothing, Operations Research, *9*, 1961, pp. 673–687.
11. Morrison, Norman, Recursive Numerical Filtering with Exponentially Decaying Weights, to be published.