JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 1, 215–227 (1960)

An Extrapolator and Scrutator*

R. J. DUFFIN† AND TH. W. SCHMIDT‡

Submitted by: R. J. Duffin

I. EXTRAPOLATION BY DISCOUNTED LEAST SQUARES

This paper describes a method of extrapolation of data by fitting with polynomial functions. The method permits use of polynomials of arbitrary degree. However, for the sake of being definite, let us suppose that the fitting is done with second degree polynomials of the form

$$p(x) = a_0 + a_1x + a_2x^2. \tag{1a}$$

It is supposed that the data consists of observations given at a regular sequence of values of x, say $x = 1, 2, 3, \ldots$. The observations for these values of x are denoted by $y_1, y_2, y_3, \ldots$. Then the central problem is to extrapolate this sequence of observations to obtain a predicted value at $x = 0$. This extrapolated or predicted value will be denoted by y_0^*. Thus if the polynomial $p(x)$ is suitably fitted to the data, the extrapolated value is given by $y_0^* = a_0$.

A common method of fitting is to select an integer r, termed the *range*, and to employ a least squares fit of $p(x)$ by minimizing

$$E = \sum_{n=1}^{r} [y_n - p(n)]^2. \tag{2}$$

Instead of this procedure, we minimize the expression

$$E = \sum_{n=1}^{\infty} \theta^n [y_n - p(n)]^2. \tag{3}$$

---

* Prepared under Contract DA–31–124–ORD–10, Office of Ordnance Research, United States Army.

† Director of Special Research in Applied Mathematics, Duke University and Office of Ordnance Research, Durham, North Carolina. Present address: Dublin, Institute for Advanced Studies, School of Theoretical Physics, Dublin, Ireland.

‡ Chief, Systems Research Office, Office of Ordnance Research, United States Army.

We term this method *discounted least squares.* The constant θ is termed the *discount factor.* In a qualitative way the discount factor in (3) corresponds to the range in (2). Like the range, the discount factor must be selected to suit best the character of the data. This selection requires judgment. However, typical values would be $r = 20$ and $\theta = 0.8$.

The notion of discounted least squares was suggested by certain problems in the analysis of missile trajectory data [1]. For example, from a sequence of observations relating to the position of a missile it might be desired to predict the point of impact. In such applications, x corresponds to negative time, and the sequence y_n is usually termed a time series.

The method of discounted least squares is a general method, and we believe it should have various applications besides that of rocket trajectory analysis. In particular it should suit certain problems of economics and operations research for which the progressive discount of the past seems natural [2].

When the minimization of (3) is carried out, a formula of the following type results:

$$y_0{}^* = \sum_{n=1}^{\infty} Q_n y_n \tag{4a}$$

Here Q_n denotes coefficients which depend on θ but not on the sequence y_n.

We term (3) the *long formula.* Actually the long formula is infinitely long, but the coefficients Q_n decay exponentially and are sensibly zero for large n. This also indicates that practically the sequence y_n does not have to be infinite. Roughly speaking, the practical length of the long formula is a small multiple of the range when using method (2).

We shall obtain an explicit algebraic expression for the coefficients Q_n. This expression is not absolutely necessary, because there is an equivalent prediction formula which has several advantages over (4) in most cases. This equivalent formula is termed the *short formula* and is written

$$y_0{}^* = 3(y_1 + \theta\delta_1) - 3(y_2 + \theta^2\delta_2) + (y_3 + \theta^3\delta_3). \tag{5}$$

Here $\delta_k = y_k{}^* - y$ is termed a *discrepancy.* By $y_k{}^*$ we mean the predicted value of y_k based on the previous values $y_{k+1}, y_{k+2}, \ldots$ The numerical coefficients in (5) are simply the binomial coefficients of $(a - b)^3$. More generally, if the prediction had been based on cubic polynomials, then the coefficients in (5) would be obtained from $(a - b)^4$.

Formula (5) is called the "short formula" because the predicted value is given as a function of the last three observed values and the last three discrepancies. The "shortness" is a consequence of using memory. The memory is selective because only the pertinent six numbers are used rather than the whole infinite time series.

An important feature of the short formula not manifestly evident is that prediction can now be carried as far into the future as desired. For example, suppose that y_1 is actually not known. Then in the formula, y_1 is simply replaced by the predicted value y_1^*. The general rule, then, is that if the observed values are not known, they are to be replaced by predicted values.

To employ the short formula it is first necessary to select a starting index, say k. This selection requires judgment. Thus y_k^* is to be the first predicted value. The values of the discrepancies $\delta_{k+1},\delta_{k+2}$, and δ_{k+3} appearing in the short formula are not known. To get started let us assume them to be zero. This is equivalent to assuming that the values of y_n for $n > k$ are given by a certain quadratic polynomial. This method of starting has the desirable property of giving perfect prediction if the observations y_n are actually defined by a quadratic polynomial.

The selection of k depends on θ. To aid in this selection we define the conventional range as $R = 4(1 - \theta)^{-1}$. Then if $k \geqslant R$ an adequate estimation of y_0^* results for a class of problems.

II. Automatic Scrutation

There is no difficulty in coding the short formula for a digital computer. Moreover, provisions can be made in this coding for automatic scrutation. By *scrutation* we mean inspection of data for the purpose of rejection of obvious errors or blunders. In hand calculation, scrutation is often accomplished by looking over a graph of the data. The trend in data processing is toward elimination of human judgment. To this end a method of employing the short formula as an automatic scrutator is as follows. Suppose that by theoretical or empirical means an estimate can be made of a standard deviation of y_n, say $\sigma(y_n)$.

Then if

$$|\delta_n| > 3\sigma(y_n) \tag{6}$$

we consider y_n to be a blunder, and y_n is replaced by y_n^*. The factor 3 is merely a conventional value.

A common phenomenon in data processing is a missing or lost observation. In automatic scrutation a lost observation is treated as an "excess error." Thus if y_n is lost it is replaced by y_n^*, and the program proceeds without interruption.

In the coding, provision must be made for the contingency that the discrepancy repeatedly exceeds the prescribed tolerance. To do this an integer w is selected. If at any time relation (6) is satisfied for w consecutive values of n, then all past memory is wiped out and a fresh start is made. This process, coded into the computer, would permit extrapolation of curves made up of parabolic segments. An application would be the tracking of evasive action.

Besides gross errors it is necessary to take into account the effect of small errors. To understand this, consider a sequence y_n such that $y_n = 0$ for all n except that $y_k = 1$. We may regard y_k as a unit error deviating from the true value zero. Substituting this sequence in the long formula gives $y_0{}^* = Q_k$. The same relation would result, of course, if we used the short formula because the short formula is merely an identity which the long formula satisfies. Thus it is seen that the coefficients Q_n determine how a single error is propagated into the future. (Moreover, it is now seen that the coding of the short formula gives an automatic method of numerically evaluating the coefficients Q_n).

A knowledge of the sequence Q_n is necessary for the analysis of random errors in all the observations as well as the single error discussed above. Thus, suppose that all the observations have independent random errors with the same variance. Then the variance in the prediction is reduced by a factor $\sum Q_n{}^2$. For θ close to unity it will be shown that

$$\sum_{1}^{\infty} Q_n{}^2 \cong 2(1-\theta). \tag{7}$$

The right side approaches zero as θ approaches one, so in favorable cases one obtains appreciable smoothing as well as extrapolation.

The sections to follow will give derivations. These will be carried out for $p(x)$ of arbitrary degree.

III. The Long Formula

Let $y_1, y_2, y_3, \ldots$ be a sequence of real numbers. It is convenient to suppose that these numbers are a time series of observations and that y_x is the observation at time $-x$. The developments considered in this paper stem from approximation of the observations y_x by a polynomial

$$p(x) = \sum_{k=0}^{m-1} a_k x^k. \tag{1b}$$

The degree of $p(x)$ is taken to be less than a given integer m. As a matter of notation, it is convenient to write p_n for $p(n)$ if n is an integer.

A measure of the error of the approximation is E defined by (3) where θ satisfies the inequality $0 < \theta < 1$. It is assumed, of course, that the sequence y_n gives E a finite value. The least square approximant is taken to be the polynomial $p(x)$ of degree less than m which minimizes E. Because the weight function θ^n is an exponential, an approximation of this type is termed discounted least squares. The bilinear form associated with the quadratic form (3) is symbolized as $[y, v]$. Thus

$$[y, v] = \sum_{n=1}^{\infty} \theta^n y_n v_n.$$

It is required to choose the coefficients a_k of $p(x)$ so that $p(x)$ minimizes E. This implies that $\partial E/\partial a_k = 0$ for $k = 0, 1, \ldots, m - 1$. Carrying out the differentiation, we have $[x^k, p - y] = 0$ or

$$\sum_{j=0}^{m-1} [x^k, x^j]a_j = [x^k, y].$$

This system of m equations can be solved for the m coefficients a_j to determine the least squares polynomial $p(x)$.

When $p(x)$ is determined, it can be used to extrapolate the sequence y_n. In particular it is desired to predict the "next" value, y_0. The predicted value of y_0 will be defined as $p(0)$. Since $p(0) = a_0$, it is simply necessary to solve the system (4) for a_0. Thus if C_{jk} is the inverse matrix to $[x^k, x^j]$, then

$$a_0 = \sum_{k=0}^{m-1} C_{ok}[x^k, y] = [q(x), y].$$

Here $q(x) = \sum C_{ok} x^k$ is a polynomial of degree less than m. A predicted value is indicated by an asterisk, thus $y_0{}^*$ is the prediction of y_0. Then the relation derived for a_0 may be stated in the explicit form

$$y_0{}^* = \sum_{n=1}^{\infty} \theta^n q_n y_n. \tag{4b}$$

This proves the long formula (4a) for extrapolation by the method of discounted least squares. It also reveals that $Q_n \theta^{-n}$ is a polynomial of degree less than m.

IV. The Short Formula

A practical difficulty with prediction formula (4a) is that if θ is close to unity, a great deal of terms in the series must be retained. The procedure which is now to be developed gets around this difficulty because it is found sufficient to store only $2m$ properties of the infinite time series. To begin this developement, note that the predicted value of y_k in terms of preceding values is given by

$$y_k^* = \sum_{n=1}^{\infty} \theta^n q_n y_{n+k}. \tag{4c}$$

This is simply the analog of (4b). It is convenient to rewrite (4c) in the form

$$\theta^k y_k^* = \sum_{n=1}^{\infty} \theta^n q_{n-k} y_n - \sum_{n=1}^{k} \theta^n q_{n-k} y_n. \tag{8}$$

Let $P(x)$ be an arbitrary polynomial of degree less than m. Then the mth difference of P vanishes identically. Thus

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} P(k) = 0. \tag{9}$$

The operation (9) is applied on (8). Then the first sum on the right side vanishes because q is a polynomial so

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} \theta^k y_k^* = - \sum_{k=0}^{m} (-1)^k \binom{m}{k} \sum_{n=1}^{k} \theta^n q_{n-k} y_n.$$

The summations on the right side are interchanged, yielding

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} \theta^k y_k^* = \sum_{k=1}^{m} h_k y_k. \tag{10}$$

Here the h_k are certain absolute constants. These absolute constants are now to be evaluated explicitly.

Let $y_n = P_n$ a polynomial of degree less than m. Hence formula (7) gives perfect prediction and (10) holds with y_k^* replaced by y_k. Also the identity (9) holds for y_k. Subtracting this identity from (10) gives

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} (\theta^k - 1) y_k = \sum_{k=1}^{m} h_k y_k.$$

The term for $k = 0$ on the left can be omitted since $\theta^0 = 1$ so

$$\sum_{k=1}^{m} [(-1)^k \binom{m}{k} (\theta^k - 1) - h_k] y_k = 0. \tag{11}$$

For y_k we choose a polynomial such that $y_k = 0$ for $k = 1, \ldots j - 1, j + 1, \ldots, m$. It may be supposed that $y_j \neq 0$. Thus only the jth term in (11) need be considered. It thereby follows that

$$h_j = (-1)^j \binom{m}{j} (\theta^j - 1).$$

Substituting this evaluation of h_j in (10) gives

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} \theta^k y_k^* = \sum_{k=1}^{m} (-1)^k \binom{m}{k} (\theta^k - 1) y_k.$$

Since $\delta_k = y_k^* - y_k$ this may be written as

$$y_0^* = \sum_{j=1}^{m} (-1)^{j+1} \binom{m}{j} (y_j + \theta^j \delta_j). \tag{12}$$

This is the desired short formula for extrapolation. It is equivalent to the infinite series formula (4b). Corresponding to formula (4c), we have

$$y_k^* = \sum_{j=1}^{m} (-1)^{j+1} \binom{m}{j} (y_{j+k} + \theta^j \delta_{j+k}). \tag{13}$$

Thus it has been proved that the predicted value is given as a linear function of the last m observed values and the last m discrepancies.

V. Multi-Step Extrapolation

Suppose now that it is desired to predict the value of y_0 given only the observations y_n for $n \geqslant h$ a positive integer. In analogy to the case $h = 1$, we seek the polynomial $p(x)$ of degree less than m which minimizes

$$E_h = \sum_{n=h}^{\infty} \theta^n (p_n - y_n)^2. \tag{14}$$

Then the predicted value of y_0 is defined to be $p(0)$.

Let $\quad y_n' = y_n$ if $n \geqslant h \quad$ and let $\quad y_n' = p_n$ if $n < h$.

Consider

$$E_1 = \sum_{n=1}^{\infty} \theta^n (P_n - y_n')^2. \tag{15}$$

Again we seek a polynomial $P(x)$ of degree less than m which minimizes E_1. Actually, $P(x) = p(x)$, the minimizing polynomial for (14). To see this, one notes that $E_1 \geqslant E_h$ in general, but if $P = p$, then $E_1 = E_h$. Comparing (15) and (3) it is seen that the problem has been reduced to the case of one step prediction. We can therefore write

$$y_0^* = \sum_{j=1}^{m} (-1)^{j+1} \binom{m}{j} [y_j' + \theta^j (y_j^* - y_j')]. \tag{16}$$

Here $y_j' = y_j$ if $j \geqslant h$ and $y_j' = y_j^*$ if $j < h$. In (16) y_j^* denotes one step prediction if $j \geqslant h - 1$, and for $j < h - 1$, it denotes multi-step prediction. Of course this notation is ambiguous if the value of h is not stated.

Formula (16) is a solution of the problem of multi-step prediction. The computational procedure is essentially the same as in the one-step case. The only difference is that predicted values are substituted for observed values when observed values are not available.

It is of interest to consider the special case of two-step prediction. Then putting $h = 2$ in (16) gives

$$y_0^* = m y_1^* + \sum_{j=2}^{m} (-1)^{j+1} \binom{m}{j} (y_j + \theta_j \delta_j). \tag{17}$$

In using this relation only the one-step prediction would be saved. Of course the one-step predictions are obtained from formula (12). In some cases it would be desirable to save only the two-step predictions, and the following formula holds

$$y_0'' = m\theta y_1'' - (1 - \theta) m y_{m+1} + \tag{18}$$

$$\sum_{j=2}^{m} (-1)^{j+1} \left[\binom{m}{j} (\theta^j y'' - \theta^j y_j + y_j) + (1 - \theta) m \binom{m}{j+1} y_j \right].$$

To avoid ambiguity, the two-step predictions are denoted as y_j''. The proof of (18) is similar to the proof of (12).

VI. Prediction of Intermediate Values and Derivatives

Again suppose that the series $y_1, y_2, \ldots$ is known but that values of y are to be predicted at intermediate points such as $x = -2/3$. By formula (16) we can obtain multi-step predicted values $y_0^*, y_{-1}^*, y_{-2}^*, \ldots, y_{m+1}^*$. Of course these are the values of $p(x)$ at the points $x = 0, -1, \ldots, -m+1$. Let $f(x)$ be the Lagrange interpolation polynomial relative to these points. Thus

$$f(x) = x(x+1)(x+2)\ldots(x+m-1)$$

and the Lagrange interpolation formula is

$$p(x) = \sum_{j=0}^{m-1} [y_{-j}^* f(x)]/[f'(-j)(x+j)] \tag{19}$$

Formula (19) can be used to predict values of y at intermediate points. Thus $y^*(-2/3) = p(-2/3)$.

Prediction of the derivative dy/dx can be obtained by differentiating (19). An equivalent method is to form the polynomial $g(x) = [p(0) - p(-x)]/x$. Then $g(x)$ is of degree less than $m - 1$ and $g(0) = p'(0) = (dy/dx)_0^*$. Thus

$$\left(\frac{dy}{dx}\right)_0^* = \sum_{j=1}^{m} \frac{(-1)^j}{j}\binom{m-1}{j}(y_{-j}^* - y_0^*). \tag{20}$$

This follows by applying (9) to $g(x)$.

VII. The Standard Deviation of the Extrapolation

For the purpose of analyzing the variance in the extrapolation it is useful to evaluate the series

$$S = \sum_{n=1}^{m} Q_n^2 c^n \tag{21}$$

where $0 < c < \theta^{-2}$. To this end let z be a complex number and let $y_j = z^j$ in the long formula (4b). Then it is seen that y_0^* is a power series in z and that the series converges if $|z|\theta < 1$. From formula (4c) it is seen that $y_k^* = z^k y_0^*$.

To obtain a closed form for the power series y_0^* we employ the short formula thus

$$y_0^* = \sum_{j=1}^{m} (-1)^{j+1}\binom{m}{j}[z^j + \theta^j(z^j y_0^* - z^j)].$$

$$y_0^*(1-\theta z)^m = (1-\theta z)^m - (1-z)^m.$$

Solving for y_0^* yields the identity

$$\sum_0^\infty Q_n z^n = -\left(\frac{1-z}{1-\theta z}\right)^m. \tag{22}$$

Here we have defined $Q_0 = -1$ to simplify the formula.

It follows from (22) that if $|z| > c\theta$

$$\sum_0^\infty Q_k \frac{c^k}{z^k} = -\left(\frac{z-c}{z-\theta c}\right)^m. \tag{23}$$

Taking the product of (22) and (23) gives the Laurent series

$$\sum_0^\infty \sum_0^\infty Q_n Q_k c^k z^{n-k} = \left(\frac{1-z}{1-\theta z}\right)^m \left(\frac{z-c}{z-\theta c}\right)^m.$$

The coefficients of a Laurent series are given by a contour integral formula. In particular the constant term of our Laurent series may be expressed as

$$\sum_0^\infty Q_n{}^2 c^n = \frac{1}{2\pi i}\int \left(\frac{1-z}{1-\theta z}\right)^m \left(\frac{z-c}{z-\theta c}\right)^m \frac{(dz)}{z}. \tag{24}$$

Here the contour could be a circle of radius r provided $\theta c < r < \theta^{-1}$. The contour integral may be evaluated by residues to give

$$\sum_1^\infty Q_n{}^2 c^n = -1 + \theta^{-m} + \frac{1}{(m-1)!}\left(\frac{d}{dz}\right)^{m-1}\left[\left(\frac{1-z}{1-\theta z}\right)^m \frac{(z-c)^m}{z}\right]. \tag{25}$$

Here after the differentiations have been carried out z is to be set equal to θc.

First suppose that the errors in the observations are independent but have the same variance. Then the variance in the extrapolation is decreased by the factor S defined by (21) with $c = 1$. In the case of linear extrapolation $m = 2$ and formula (25) yields

$$S = (1 - \theta)(1 + \theta)^{-3}(\theta^2 + 4\theta + 5). \tag{26}$$

In quadratic extrapolation $m = 3$ and formula (25) yields

$$S = (1 - \theta)(1 + \theta)^{-5}(\theta^4 + 6\theta^3 + 16\theta^2 + 24\theta + 19). \tag{27}$$

In case θ is close to unity it is seen that (27) takes on the approximate form already stated in (7).

Another case of interest is when the "more distant" observations have the greater error. In particular suppose that the standard deviation of y_n say $\sigma(y_n)$ obeys a law of the form

$$\sigma^2(y_n) = K\theta^{-n} \tag{28}$$

for some constant K. Then it is seen that the standard deviation of $y_0{}^*$ satisfies $\sigma^2(y_0{}^*) = KS$ for $c = \theta^{-1}$. Substituting $c = \theta^{-1}$ in (25) it is clear that the derivative term vanishes. Thus $\sigma^2(y_0{}^*) = K(\theta^{-m} - 1)$. This may be written in the form

$$\sigma^2(y_0{}^*) = \sigma^2(y_m)[1 - \theta^m]. \tag{29}$$

The following hypothetical situation is worth noting: (a) The y_n are actually observations of a polynomial of degree less than m. (b) The errors in the y_n are independent. (c) The errors have a standard deviation satisfying (28). (d) The errors have a Gaussian distribution. Under these hypotheses it is seen that the determination of the polynomial p by the method of discounted least squares is in accord with Fisher's principle of maximum likelihood.

VIII. Associated Laguerre Polynomials

An explicit determination of the polynomial $q(x)$ appearing in the long formula will now be given.

The relation (4b) may be written as $y_0{}^* = [q,y]$. Suppose $p(x)$ is a polynomial of degree less than m, then

$$p(o) = [q(x),p(x)]. \tag{30}$$

This relation determines $q(x)$ in terms of certain orthogonal polynomials. To show this, we first consider the sequence of polynomials $L_0(x),L_1(x)$, $L_2(x),\ldots$ of degree $0,1,2,\ldots$ respectively, which are orthogonal with respect to the scalar product $[y,v]$. Thus $[L_i(x), L_j(x)] = 0$ if $i \neq j$. These are the *discrete Laguerre polynomials*. Their properties were given by Gottlieb [3]. An account of related work is to be found in the treatise of Szegö [4]. Gottlieb gives the following explicit expression for $L_m(x)$

$$L_m(x) = \theta^{-x} \Delta^m \left[\theta^x \binom{x-1}{m} \right]. \tag{31}$$

Here Δ denotes the usual right difference operator.

Actually we need here the discrete analog of the *associated Laguerre polynomials*, say $L_m{}^1(x)$. No references to such polynomials were found in the literature, but their properties can be developed in analogy to $L_m(x)$. Thus define

$$L_m{}^1(x) = \Delta L_m(x). \tag{32}$$

Then this function is clearly a polynomial of degree $m - 1$. It is readily found by partial summation that

$$[L_m{}^1(x), x^k] = 0 \qquad \text{for} \qquad k = 1,2,\ldots, m-1.$$

It is an immediate consequence of these relations that the associated Laguerre polynomials form an orthogonal sequence, but relative to the weight function $x\theta^x$ instead of θ^x. (More generally, associated Laguerre polynomials can be developed relative to the weight function $x^j\theta^x$ but they will not be considered here).

Let $p(x)$ be a polynomial of degree less than m. We can write

$$p(x) = p(o) + xr(x) \tag{33}$$

where $r(x)$ is of degree less than $m - 1$. Then

$$[L_m{}^1(x), p(x)] = [L_m{}^1(x), p(o)] + [L_m{}^1(x), xr(x)].$$

The second term on the right vanishes by orthogonality. Then comparison with (30) gives

$$q(x) = L_m{}^1(x)/[L_m{}^1, 1]. \tag{34}$$

To simplify the numerator and the denominator in (34) it is convenient to make certain transformations on the definitions. Gottlieb gives the relation

$$L_m(x) = \theta^m \sum_{k=0}^{m} (1 - \theta^{-1})^k \binom{m}{k} \binom{x-1}{k}. \tag{35}$$

Then (32) gives

$$L_m{}^1(x) = \theta^m \sum_{k=0}^{m} (1 - \theta^{-1})^k \binom{m}{k}\binom{x-1}{k-1}. \tag{36}$$

By definition the denominator in (34) is

$$[L_m{}^1,1] = \sum_{n=1}^{\infty} \theta^n \, [L_m(n+1) - L_m(n)]$$

$$= \sum_{n=1}^{\infty} \theta^n L_m(n+1)$$

$$= \theta^{-1} \sum_{n=2}^{\infty} \theta^n L_m(n) = - L_m(1).$$

Substituting $x = 1$ in (35) gives

$$[L_m{}^1, 1] = - \theta^m. \tag{37}$$

This together with (34) and (37) yields

$$q(x) = - \sum_{k=0}^{m} (1 - \theta^{-1})^k \binom{m}{k}\binom{x-1}{k-1}. \tag{38}$$

This relation gives an explicit algebraic evaluation of the coefficients of the extrapolation formula.

References

1. Duffin, R. J., and Schmidt, Th. W. A simple formula for prediction and automatic scrutation, *Am. Rocket Soc., Flight Testing Conf.*, Daytona Beach, Florida, March 24, 1959.
2. Duffin, R. J., and Schmidt, Th. W. Discounted least squares, *Proc. Fourth Ordnance Conf. on Operations Research*, Redstone Arsenal, April 1, 1959.
3. Gottlieb, M. J. Polynomials orthogonal on a finite or enumerable set of points, *Am. J. Math.*, **60**, 453–458 (1938).
4. Szegö, G. Orthogonal polynomials, *Am. Math. Soc. Colloquium Publications*, **23**, Chapter II (1939).