ESD-TR-65-369

TM-03967

THE REDUCTION OF THE VARIANCE BY LEAST SQUARES POLYNOMIAL APPROXIMATION

TECHNICAL REPORT NO. ESD-TR-65-369

SEPTEMBER 1965

H. C. Joksch

Prepared for

DIRECTORATE OF COMPUTERS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 7070

Prepared by

THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF 19(628)-2390

ESD-TR-65-369 TM-03967

THE REDUCTION OF THE VARIANCE BY LEAST SQUARES POLYNOMIAL APPROXIMATION

TECHNICAL REPORT NO. ESD-TR-65-369

SEPTEMBER 1965

H. C. Joksch

Prepared for

DIRECTORATE OF COMPUTERS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 7070

Prepared by

THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF 19(628)-2390

ABSTRACT

Least squares approximations obtain parameters with a variance lower than those of the data from which they are obtained. A least squares polynomial approximation to observations may be used to obtain "smoothed" values of the observations or to make predictions. The reduction of the variance achieved by this process is determined for several special cases. Some properties of the Gram-polynomials necessary for the analysis are derived.

REVIEW AND APPROVAL

This technical report has been reviewed and is approved.

John B. Curtis

JOHN B. CURTIS
1st Lt, USAF
Project Officer

TABLE OF CONTENTS

		Page
SECTION I	INTRODUCTION	1
	APPROXIMATION TECHNIQUES	1
	LEAST SQUARES APPROXIMATIONS	1
SECTION II	APPROXIMATION BY ORTHOGONAL POLYNOMIALS	3
SECTION III	A GENERAL PROPERTY OF THE REDUCTION FACTOR	5
SECTION IV	THE REDUCTION FACTOR AT THE MIDPOINT	6
	COROLLARY	8
SECTION V	THE REDUCTION FACTOR AT THE ENDPOINT	9
	COROLLARY	10
SECTION VI	THE REDUCTION FACTOR AT THE PREDICTED POINT	11
SECTION VII	THE ASYMPTOTIC BEHAVIOR OF THE REDUCTION FACTOR	12
APPENDIX	SOME PROPERTIES OF GRAM-POLYNOMIALS	13
BIBLIOGRAPHY		17

SECTION I
INTRODUCTION

APPROXIMATION TECHNIQUES

If a function is empirically given, e.g., by discrete observations, it is frequently desirable to represent these observations by analytical expression, e.g., polynomials. If the values of the function are not exactly known, but subject to error, then it is not reasonable to obtain an exact representation. It is more appropriate to use an approximate representation, where the differences between the observations and the approximation are so small that they can be considered as errors of the observations rather than as inaccuracies of the representation. If the "true" nature of the empirical function, e.g., the degree of a polynomial, is known, it can be used to approximate the observation and determine its parameters. The differences between approximation and observation can be considered as pure errors of the observations.

One way to find the "best" approximation, which minimizes an overall measure for the error of the approximation, is a least squares approximation. The randomness of the errors, however, precludes an approximation method that will find the "true" function and separate it from the observational errors. Thus, some part of the observational errors can always influence the approximating function, and, consequently, the differences will contain not only observational errors but also inaccuracies of the approximation. The larger the number of errors, the more likely the probability that the random effects will cancel out and that the approximating function will be close to the "true" function. On the other hand, the more the parameters necessary to describe the approximating function, the less the reduction of this random error will be. These functional dependencies are the subject of this analysis.

LEAST SQUARES APPROXIMATIONS

This analysis is restricted to an important special case: the values of the empirical function ("observations") y_x are given for n equidistant values of the argument x, $x = 0 \ldots n-1$, and the approximating function is a polynomial $y(x)$.

Some important applications of least squares fitting include:

(a) smoothing a series of values;

(b) estimating more precisely the present position of an object during tracking; and

(c) determining the expected position $y(x)$ of the object being tracked and its variance at the time of the next observation.

In (a), the polynomial values at the midpoint of the components are used as "smoothed" values; in (b), a polynomial of appropriate degree is fitted to the last observations, and $y(x)$ at the endpoint is used as the last position. Thus, exact expressions and bounds for the variance of the polynomial can be given for (a), (b), and (c) at the midpoint, the endpoint, and at the first equidistant point outside the given interval, respectively.

General expressions for the variance are well known,[1,5,7][1] but their use requires extensive numerical calculations. The dependency of the variance on the number of observations and on the degree of the approximating polynomial is not obvious. Special studies, sometimes incidental to other problems, have been made by Cowden,[1] Guest,[3,4,5] Proschan,[9] and Smith.[10] Smith's problem is slightly different; she assumes that the observations are uniformly spread over the whole interval. Her exact results for the mid- and endpoints are, therefore, only approximations for the problem discussed in this report. Guest[4] obtains the same approximations in a much easier way by the use of Legendre polynomials. Proschan finds the same approximation for the endpoint by use of determinants, similar to Smith. However, these approximations are good for large n only; for small n, they can be very poor. Guest[3,5] and Cowden give tables for the determination of the variance of $y(x)$, based on numerical calculations. Using Guest's tables, one has to notice that $k = 1$ is not the endpoint of the interval, but, rather, $k = (n-1)/n$. Some checks showed good agreement with the results obtained from our exact formulas.

---

[1]Numbers in brackets refer to citations in the Bibliography.

SECTION II
APPROXIMATION BY ORTHOGONAL POLYNOMIALS [5,6,7,8]

Let n values y_x ("observations") be given for equidistant arguments ("times") $x = 0 \ldots n-1$. Let $y(x)$ be a polynomial of degree m to be fitted to the y_x such that

$$\sum_{x=0}^{n-1} \left[y_x - y(x) \right]^2 = \min. \qquad (1)$$

If the y_x are subject to random variations, then the coefficients of the polynomial are also subject to those variations and, consequently, the value of $y(x)$ for any given x.

Least squares approximations are simplified by the use of orthogonal polynomials $\phi_r(x)$. They are characterized by the orthogonality relations

$$\sum_x w(x)\, \phi_r(x)\, \phi_s(x) = \begin{cases} q_r > 0 & \text{if } r = s \\ 0 & \text{if } r \neq s \end{cases} . \qquad (2)$$

In the case of equal weights, $w(x) = 1$, and equidistant arguments, $x = 0 \ldots n-1$, they determine the Gram-polynomials (sometimes called Chebysheff-polynomials; but this name is more commonly used for other orthogonal polynomials). An explicit expression for them is

$$\phi_r(x,n) = \sum_{i=0}^{r} (-1)^i \binom{r}{i} \binom{r+i}{i} x^{(i)} \Big/ (n-1)^{(i)}, \qquad (3)$$

where $x^{(i)} = x(x-1) \ldots (x-i+1)$, and the sum of their squares is

$$q_r = \frac{n}{2r+1} \cdot \frac{n+1}{n-1} \cdot \frac{n+2}{n-2} \cdots \frac{n+r}{n-r} . \qquad (4)$$

The approximation which satisfies Equation (1) is given by

$$y(x) = \sum_{r=0}^{m} Y_r \phi_r(x) \Big/ q_r \;, \tag{5}$$

where

$$Y_r = \sum_{\xi=0}^{n-1} y_\xi \phi_r(\xi) \;. \tag{6}$$

From (5) and (6) we obtain

$$y(x) = \sum_{\xi=0}^{n-1} y_\xi c(x,\xi) \;, \tag{7}$$

if we define

$$c(x,\xi) = \sum_{r=0}^{m} \phi_r(x)\phi_r(\xi) / q_r \;. \tag{8}$$

Equation (7) is very convenient for practical use, since it is linear in the y_ξ. The numerical values of the $c(x,\xi)$, for several n and m, are given by Cowden,[1] Hildebrand,[6] and Milne.[8]

SECTION III
A GENERAL PROPERTY OF THE REDUCTION FACTOR

If the y_ξ are uncorrelated random variables, all with the same variance σ, then the variance of $y(x)$ is known [5,7] to be

$$V(x) = \sigma \sum_{r=0}^{m} \phi_r{}^2(x) \,/\, q_r \,. \tag{9}$$

The term $R(x) = V(x)/\sigma$ may be called the reduction factor. It can be calculated for any x from Equation (9) as, e.g., in Cowden [1] and Guest. [3] However, if one uses Equation (7) to obtain the polynomial value $y(x)$ for any of the given equidistant arguments x, it can be obtained in a simpler way. Comparison of Equations (9) and (8) shows

$$R(x) = c(x,x) \,. \tag{10}$$

Verbally expressed: the reduction factor for the variance, achieved by a least squares polynomial approximation to equally weighted equidistant observations, at any of the observation points, equals the coefficient of the observation at this point in the linear combination of the observed values giving the polynomial value at the point under consideration. This gives immediately, without calculations, the reduction factor if one uses the numerical expression for (7) as given in Cowden, [1] Hildebrand, [6] and Milne. [8]

For a special case, the midpoint of an uneven number of observations, this result has been obtained by Milne. [8]

SECTION IV
THE REDUCTION FACTOR AT THE MIDPOINT

The midpoint of the interval $x = 0 \ldots n-1$ is $x = (n-1)/2$. The value of ϕ_r at this point is derived in the Appendix. We substitute Equations (4) and (46) into (9), write $r = 2s$, since, in effect, we have to sum the even terms only, and obtain

$$R_m\left(\frac{n-1}{2}\right) = \sum_{s=0}^{t} \left(\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2s-1}{2s}\right)^2 \left(\frac{n+1}{n-2}\cdot\frac{n+3}{n-4}\cdots\frac{n+2s-1}{n-2s}\right)^2 \frac{4s+1}{n}\cdot\frac{n-1}{n+1}\cdot\frac{n-2s}{n+2s}\ , \qquad (11)$$

where t is the largest integer such that $2t \leq m$. Since $R_{2t} = R_{2t+1}$, we restrict our further arguments to $m = 2t$. Extensively written, Equation (11) becomcs

$$r_{2t}\left(\frac{n-1}{2}\right) = +\frac{1}{n}\left[1+5\left(\frac{1}{2}\right)^2 \frac{n^2-1}{n^2-n} + 9\left(\frac{1}{2}\cdot\frac{3}{4}\right)^2 \frac{n^2-1}{n^2-4}\cdot\frac{n^2-9}{n^2-16} + \ldots\right.$$

$$\left. + \left(4t+1\right)\left(\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2t-1}{2t}\right)^2 \frac{n^2-1}{n^2-4}\cdot\frac{n^2-9}{n^2-16}\cdots\frac{n^2-(2t-1)^2}{n^2-(2t)^2}\right]. \qquad (12)$$

This is an exact expression for the reduction factor. To obtain a simple estimate for it, we define

$$S_{2t} = 1+5\left(\frac{1}{2}\right)^2 + 9\left(\frac{1}{2}\cdot\frac{3}{4}\right)^2 + \ldots + \left(4t+1\right)\left(\frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2t-1}{2t}\right)^2 . \qquad (13)$$

We have, obviously,

$$\frac{S_{2t}}{n} < R_{2t}\left(\frac{n-1}{2}\right) < \frac{S_{2t}}{n}\cdot\frac{n^2-1}{n^2-4}\cdot\frac{n^2-9}{n^2-16}\cdots\frac{n^2-(2t-1)^2}{n^2-(2t)^2} . \qquad (14)$$

It can be shown by induction that

$$S_{2t} = \left(\frac{3}{2} \cdot \frac{5}{4} \cdots \frac{2t+1}{2t}\right)^2 . \qquad (15)$$

A comparison with Wallis' product,

$$\frac{2}{\pi} = \lim_{t=\infty} \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{3}{4} \cdot \frac{5}{4} \cdots \frac{2t-1}{2t} \cdot \frac{2t+1}{2t} ,$$

gives the estimate

$$\frac{2}{\pi}(2t+1) < S_{2t} \leq \frac{3}{4}(2t+1) . \qquad (16)$$

Combining (14) and (16), we have

$$\frac{2}{\pi} \cdot \frac{2t+1}{n} < R_{2t}\left(\frac{n-1}{2}\right) < \frac{3}{4} \cdot \frac{2t+1}{n} \cdot \frac{n^2-1}{n^2-4} \cdot \frac{n^2-9}{n^2-16} \cdots \frac{n^2-(2t-1)^2}{n^2-(2t)^2} . \qquad (17)$$

For large n, the terms containing n^2 are very nearly 1, and, since $2/\pi = 0.64$, the range for R_{2t} given by Equation (17) is fairly small. However, if it is not small compared to n, an approximation of Equation (12) may be worthwhile. We notice that the factors

$$\left(4t+1\right)\left(\frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2t-1}{2t}\right)$$

in Equation (12) approach $4/\pi = 1.27$ very rapidly. The first of them, $5(1/2)^2 = 1.25$, is already very close to the limit. Therefore,

$$nR_{2t}\left(\frac{n-1}{2}\right) \approx 1 + \frac{5}{4}\left[\frac{n^2-1}{n^2-4} + \frac{n^2-1}{n^2-4} \cdot \frac{n^2-9}{n^2-16} + \cdots + \frac{n^2-1}{n^2-4} \cdots \frac{n^2-(2t-1)^2}{n^2-(2t)^2}\right]. \qquad (18)$$

If we expand the terms in the bracket and neglect all terms of higher order than n, we obtain

$$nR_{2t}\left(\frac{n-1}{2}\right) \approx 1 + \frac{5}{4}\left[t + \frac{t(t+1)(4t+5)}{6n^2}\right] . \qquad (19)$$

Guest [4] and Smith [10] obtain the estimate $R_{2t} \approx S_{2t} / n$. It is quite good, except when m is comparable to n. For a first estimate, one may even make use of Equation (16) and approximate $R_{2t} \approx 0.7\,(2t+1) / n$.

COROLLARY

If a polynomial of $(n-1)^{th}$ degree is fitted to n observations, an exact fit is achieved, and, therefore, $R = 1$ for the arguments of the observations. For odd n, the midpoint $(n-1)/2$ is one of the given points. Therefore, we have the identity

$$n = 1+5\left(\frac{1}{2}\right)^2 \frac{n^2-1}{n^2-4} + 9\left(\frac{1}{2} \cdot \frac{3}{4}\right)^2 \frac{n^2-1}{n^2-4} \cdot \frac{n^2-9}{n^2-16} + \ldots$$

$$+ (2n-1)\left(\frac{1}{2} \cdot \frac{3}{4} \cdots \frac{n-2}{n-1}\right)^2 \frac{n^2-1}{n^2-4} \cdot \frac{n^2-9}{n^2-16} \cdots \frac{4n-4}{2n-1} \qquad (20)$$

for all odd n.

SECTION V
THE REDUCTION FACTOR AT THE ENDPOINT

At the endpoint of the interval, $x = 0$ (the other, $x = n-1$, gives the same results), we have $\phi_r = +1$; therefore, Equation (9) combined with Equation (4) immediately gives

$$R_m(0) = \frac{1}{n}\left[1+3\,\frac{n-1}{n+1} + 5\,\frac{n-1}{n+1}\cdot\frac{n-2}{n+2} + \ldots + (2m+1)\,\frac{n-1}{n+1}\cdots\frac{n-m}{n+m}\right] . \qquad (21)$$

If we consider that $1+3+5 \ldots + (2m+1) = (m+1)^2$, we have, obviously,

$$\frac{(m+1)^2}{n} > R_m(0) > \frac{(m+1)^2}{n}\cdot\frac{n-1}{n+1}\cdot\frac{n-2}{n+2}\cdots\frac{n-m}{n+m} . \qquad (22)$$

The argument used in the corollary of SECTION IV gives $R_{n-1}(0) = 1$, since the endpoint is always one of the given points. Therefore, $R_m(0) \leq 1$ is obvious from Equation (21). This, combined with Equation (22), gives

$$\text{Min}\left\{1, \frac{(m+1)^2}{n}\right\} \geq R_m(0) > \frac{(m+1)^2}{n}\cdot\frac{n-1}{m+1}\cdot\frac{n-2}{n+2}\cdots\frac{n-m}{n+m} . \qquad (23)$$

For small n, these bounds can be quite far apart. To obtain an approximation, we expand the fractions in the bracket in Equation (21) in series and omit terms containing n and higher powers. This gives

$$(m+1)^2 - \frac{2}{n}\left[3\cdot 1 + 5\,(1+2) + \ldots + (2m+1)(1+2+\ldots+m)\right] . \qquad (24)$$

The bracket can easily be evaluated, and we obtain, by combining Equations (21) and (24),

$$R_m(0) \approx \frac{(m+1)^2}{n}\left[1 - \frac{m\,(m+2)}{2n}\right] . \qquad (25)$$

Guest,[4] Proschan,[9] and Smith[10] give the approximation, $R_m(0) \approx (m+1)^2/n$.[2] This approximation is good for very large n only. It is very poor for small n. For example, if $m = 3$ and $n = 10$, it gives $R \approx 1.6$, whereas the exact value is $R = 0.67$. It is much closer to the lower bound $R > 0.47$ in Equation (22). Even the approximation, (23), gives the poor estimate $R_m \approx 0.4$. Therefore, it is advisable to determine the bounds given by Equation (21), and to use the exact expression, (20), if they are too far apart.

COROLLARY

The term $R_{n-1}(0) = 1$, combined with Equation (21), gives the identity

$$n = 1+3\,\frac{n-1}{n+1} + 5\,\frac{n-1}{n+1}\cdot\frac{n-2}{n+2} + \ldots + (2n-1)\,\frac{n-1}{n+1}\cdots\frac{1}{2n-1} \qquad (26)$$

for all integer n.

[2] This corresponds to the upper bound in Equation (21).

SECTION VI
THE REDUCTION FACTOR AT THE PREDICTED POINT

The predicted point is the first point outside the given interval, $x = -1$ ($x = n$ gives the same results). The value $\phi_r(-1, n)$ is derived in the appendix. Substitution of Equations (52) and (4) into (9) gives

$$R_m(-1) = \frac{1}{n}\left[1 + 3\,\frac{n-1}{n+1}\cdot\frac{(n+1)^2}{(n-1)^2} + 5\,\frac{n-1}{n+1}\cdot\frac{n-2}{n+2}\cdot\frac{(n+1)^2}{(n-1)^2}\cdot\frac{(n+2)^2}{(n-2)^2} + \dots \right.$$

$$\left. + (2m+1)\,\frac{n-1}{n+1}\cdots\frac{n-m}{n+m}\left(\frac{n+1}{n-1}\cdots\frac{n+m}{n-m}\right)^2\right], \qquad (27)$$

and

$$R_m(-1) = \frac{1}{n}\left[1 + 3\,\frac{n+1}{n-1} + 5\,\frac{n+1}{n-1}\cdot\frac{n+2}{n-2} + \dots + (2m+1)\,\frac{n+1}{n-1}\cdots\frac{n+m}{n-m}\right]. \qquad (28)$$

Equation (28) has a remarkable symmetry to (21). In a similar way, we derive the bounds

$$\frac{(m+1)^2}{n} < R_m(-1) < \frac{(m+1)^2}{m}\,\frac{m+1}{n-1}\cdots\frac{n+m}{n-m}, \qquad (29)$$

and the approximation

$$R_m(-1) \approx \frac{(m+1)^2}{n}\left[1 + \frac{m(m+2)}{2n}\right]. \qquad (30)$$

SECTION VII
THE ASYMPTOTIC BEHAVIOR OF THE REDUCTION FACTOR

Since $\phi_r(x,n)$ is rapidly increasing with x for $x > n$, the variance of the extrapolated polynomial $y(x)$ is rapidly increasing. Exact values can be obtained by lengthy computations only. An asymptotic estimate, however, is quite easy. For large x, the highest term becomes overwhelming in $\phi_r(x,n)$, therefore

$$\phi_r(x) \sim (-1)^{(r)} \frac{1 \cdot 2 \ldots (2r)}{(1 \cdot 2 \ldots r)^2} \cdot \frac{x^r}{(n-1)^{(r)}} \quad . \qquad (31)$$

Similarly, in Equation (9) the term with $r = m$ becomes overwhelming. Therefore,

$$R_m(x) \sim x^{2r} \left[\frac{1 \cdot 2 \ldots (2m)}{(1 \cdot 2 \ldots m)^2} \cdot \frac{1}{(n-1) \ldots (n-m)} \right]^2 \frac{2m+1}{n} \cdot \frac{n-1}{n+1} \cdots \frac{n-m}{n+m} \quad , \qquad (32)$$

or

$$R_m(x) \sim \left(\frac{4x}{n} \right)^{2r} \left[\left(\frac{1 \cdot 3 \ldots (2m-1)}{2 \cdot 4 \ldots 2m} \right)^2 \left(2m+1 \right) \right] \frac{n^2}{n^2-1} \cdots \frac{n^2}{n^2-(m^2)} \quad . \qquad (33)$$

This gives the asymptotic estimate

$$R_m(d) \sim f(m) \left(\frac{4x}{n} \right)^{2r} \frac{n^2}{n^2-1} \cdots \frac{n^2}{n^2-m^2} \quad . \qquad (34)$$

The factor, $f(m)$, in Equation (33), is bounded by $2/\pi$ and $3/4$. For small m, it is closer to $3/4$; for large m, closer to $2/\pi$. Often the approximation $f(m) = 0.7$ will be sufficient.

Hans Joksch

APPENDIX
SOME PROPERTIES OF GRAM-POLYNOMIALS

A RECURSION FORMULA

For any orthogonal polynomial, a recursion formula,

$$\phi_r(x) = a_r x \phi_{r-1}(x) + b_r \phi_{r-1}(x) + c_r \phi_{r-2}(x) \quad , \tag{35}$$

exists. The coefficients a_r, b_r and c_r can be derived by the following arguments: we note that $\phi_r(0) = 1$ for all r; therefore,

$$1 = b_r + c_r \quad . \tag{36}$$

Multiplication of Equation (35) with $\phi_{r-1}(x)$ and summation over all x gives

$$0 = a_r \sum_{x=0}^{n-1} x \phi_r^2(x) + b_r q_{r-1} \quad . \tag{37}$$

The same operation with ϕ_{r-2} gives

$$0 = a_r \sum_{x=0}^{n-1} x \phi_{r-1}(x) \phi_{r-2}(x) + c_r q_{r-2} \quad . \tag{38}$$

Since $\phi_r(x)$ is a symmetric or antisymmetric function with respect to the midpoint $(n-1)/2$ of the interval, $\phi_r^2(x)$ is symmetric and it holds that

$$\sum_{x=0}^{n-1} \left(x - \frac{n-1}{2} \right) \phi_r^2(x) = 0 \quad . \tag{39}$$

This gives

$$\sum_{x=0}^{n-1} \phi_r^2(x) = \frac{n-1}{2} q_r \qquad (40)$$

for the sum in Equation (37).

To evaluate the sum in Equation (38), we develop the polynomial $x\phi_{r-2}(x)$ in a series of polynomials:

$$x\phi_{r-2}(x) = \alpha\phi_{r-1}(x) + \text{polynomials of lower degree.} \qquad (41)$$

From this it follows that

$$\sum_{x=0}^{n-1} x\phi_{r-1}(x)\,\phi_{r-2}(x) = \alpha q_{r-1} \quad . \qquad (42)$$

The term α is obviously the quotient of the highest coefficients of x in ϕ_{r-2} and ϕ_{r-1}, which, from Equation (3), can be found to be

$$\alpha = \frac{r-1}{2(2r-3)}(n-r+1) \quad . \qquad (43)$$

From Equations (36), (37), (38), (40), (42), and (43), we obtain a_r, b_r, and c_r, and have the recursion formula for Gram-polynomials:

$$\phi_r(x,n) = -\frac{2(2r-1)}{r(n-r)}\left(x - \frac{n-1}{2}\right)\phi_{r-1}(x,n) - \frac{r-1}{r}\,\frac{n+r-1}{n-r}\,\phi_{r-2}(x,n) \quad . \qquad (44)$$

THE VALUE AT THE MIDPOINT

For the midpoint, $x = (n-1)/2$, Equation (44) simplifies to

$$\phi_r\left(\frac{n-1}{2}, n\right) = -\frac{r-1}{r}\,\frac{n+r-1}{n-r}\,\phi_{r-2}\left(\frac{n-1}{2}, n\right) . \qquad (45)$$

This allows recursive calculation of the even-order polynomials for the midpoint; namely,

$$\phi_{2s}\left(\frac{n-1}{2}, n\right) = (-1)^s \frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2s-1}{2s}\cdot\frac{n+1}{n-2}\cdot\frac{n+3}{n-4}\cdots\frac{n+2s-1}{n-2s} \quad . \qquad (46)$$

The odd-order polynomials equal zero at the midpoint.

THE VALUE AT THE PREDICTED POINT

We call $x = -1$ the "predicted point." Since $(-1)^{(i)} = (-1)^i i!$, Equation (3) simplifies to

$$_r(-1) = 1 + \frac{r(r+1)}{1(n-1)} + \frac{r(r-1)\cdot(r+1)(r+2)}{1\cdot 2\cdot(n-1)(n-2)}\cdots$$

$$+ \frac{r(r-1)\ldots 1\cdot(r+1)\ldots(2r)}{1\cdot 2\ldots r\cdot(n-1)\ldots(n-r)} \quad . \qquad (47)$$

This is a hypergeometric series, namely,

$$\phi_r(-1,n) = F(r+1; -r; -n+1; 1) \quad . \qquad (48)$$

For Gauss' relation, [2][3]

$$(c-a-b)\,F(a,b;c;x) - (c-a)\,F(a-1,b;c;x) + b\,(1-x)\,F(a,b+1;c;x) = 0 \quad , \qquad (49)$$

we obtain, for $x = 0$, since F is finite,

$$F(a,b;c;1) = \frac{c-a}{c-a-b}\,F(a-1,b;c;1) \quad . \qquad (50)$$

And, if $a > 0$, by induction,

$$F(a,b;c;1) = \frac{c-a}{c-a-b} = \frac{c-a+1}{c-a-b+1}\cdots\frac{c-1}{c-b-1} \quad , \qquad (51)$$

[3] See Formula 6.

since $F(0,b;c;1) = 1$. Substitution of $c = -n+1$, $a = r+1$ and $b = -r$ and re-arrangement gives

$$\phi_r(-1,n) = F(r+1, -r; -n+1; 1) = \frac{n+1}{n-1} \cdot \frac{n+2}{n-1} \cdots \frac{n+r}{n-r} \quad . \tag{52}$$

BIBLIOGRAPHY

1. Cowden, D. J., "The Perils of Polynomials," Management Science, 9, 1963, 542-550.

2. Gauss, C. F., "Disquisitiones generales circa seriem infinitam... ," Comm. Soc. Reg. Sci. Gottingensis rec., II, MDCCCXIII, reprinted Werke, III, Gottingen 1866.

3. Guest, P. G., "Estimation of the Error at a Point on a Least Squares Curve," Australian J. of Sci. Res., Ser. A. 3, 1950, 173-182.

4. Guest, P. G., "The Spacing of Observations in Polynomial Regression," Ann. Math. Stat., 29, 1958, 294-299.

5. Guest, P. G., Numerical Methods of Curve Fitting, Cambridge, at the University Press, 1961.

6. Hildebrand, F. B., Introduction to Numerical Analysis, McGraw-Hill, New York, Toronto, London, 1956.

7. Kimball, B. F., "Orthogonal Polynomials Applied to Least Square Fitting of Weighted Observation," Ann. Math. Stat., 11, 1940, 348-352.

8. Milne, W. E., Numerical Calculus, Princeton University Press, Princeton, 1949.

9. Proschan, F., "Precision of Least Squares Polynomial Estimates," SIAM Review, 3, 1961, 230-236.

10. Smith, K., "On the Standard Deviation of Adjusted and Interpolated Values of an Observed Polynomial Function and Its Constants and the Guidance They Give towards a Proper Choice of the Distribution of Observations," Biometrika, 12, 1918, 1-86.

Security Classification

DOCUMENT CONTROL DATA - R&D

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

1. ORIGINATING ACTIVITY *(Corporate author)* The MITRE Corporation Bedford Massachusetts	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

The Reduction Of The Variance By Least Squares Polynomial Approximations

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

N/A

5. AUTHOR(S) *(Last name, first name, initial)*

Joksch Hans C.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1965	20	10

8a. CONTRACT OR GRANT NO. AF19(628)-2390 b. PROJECT NO. 7070 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ESD-TR-65-´69 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* TM-03967

10. AVAILABILITY/LIMITATION NOTICES

Qualified requestors may obtain from DDC. DDC release to CFSTI formerly OTS) authorized.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Directorate of Computers Electronic Systems Division L. G. Hanscom Field, Bedford, Mass.

13. ABSTRACT

Least squares approximations obtain parameters with a variance lower than those of the data from which they are obtained. At least squares polynomial approximation to observations may be used to obtain "smoothed" values of the observations or to make predictions. The reduction of the variance achieved by this process is determined for several special cases. Some properties of the Gram-polynomials necessary for the analysis are derived.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
TRACKING Measurements; Numerical Analysis NUMERICAL ANALYSIS Smoothing Data Polynomial Arc Smoothing						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
______________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
______________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
______________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional