OPTIMUM DESIGNS FOR POLYNOMIAL EXTRAPOLATION[1]

By Paul G. Hoel

University of California, Los Angeles

1. Summary. A solution is given to the problem of where to choose $k + 1$ points and what weights to assign them in order to obtain the minimum variance of what may be called an interior extrapolated value of a polynomial of degree k. It is assumed that observations can be taken in the interval $[-1, 1]$, except for a subinterval (α, β) located in its interior and within which the extrapolation occurs. This one dimensional solution is then used to solve the corresponding two dimensional problem, for a certain class of polynomials, in which it is desired to extrapolate inside a rectangular region that lies inside a larger rectangular region and within which observations can be taken, exclusive of the interior region. In addition, the two dimensional exterior extrapolation problem is solved for the same class of polynomials as those used for interior extrapolation.

2. Introduction. The problem of optimum spacing and weighting for polynomial prediction has been solved for the case of uncorrelated observational values when the prediction consists of minimax interpolation [3], [6], and when it consists of extrapolation beyond the interval of observations [4]. The present paper extends those results by determining the optimum spacing and weighting for predicting a polynomial value inside an interval (α, β), where $-1 < \alpha < \beta < 1$, when observations are restricted to be taken inside $[-1, 1]$, but outside (α, β). This type of extrapolation will be called interior extrapolation to distinguish it from the more common external extrapolation.

The notation employed in [4] will be used here also. Thus, x will denote any fixed point inside the interval (α, β) where it is desired to estimate the polynomial regression value.

$$E[y(x)] = \beta_0 + \beta_1 x + \cdots + \beta_k x^k.$$

The total number of observations to be taken, which is denoted by n and assumed fixed in advance, is to be distributed among $k + 1$ distinct points inside the two subintervals $[-1, \alpha]$ and $[\beta, 1]$. The problem is to determine how to choose those points and how to find the best set of proportions of observations to assign the points. The best proportions ordinarily will not yield integer values for the number of observations to be taken at the various points; therefore the resulting design may be optimum only in an approximate sense. In this paper an optimum design will be understood to be a choice of $k + 1$ observation points $x_0, x_1, \cdots, x_k$ and weights $w_0, w_1, \cdots, w_k$ satisfying $\sum w_i = n$ which minimizes

Received 19 October 1964; revised 29 March 1965.

[1] This research was supported by the Air Force Office of Scientific Research Grant # AF-AFOSR-459-63.

the variance of $\hat{y}(x)$, where $\hat{y}(x)$ denotes the estimated value of $E[y(x)]$ based on the traditional best unbiased linear estimates of the β's.

Since the fitted polynomial is a weighted least squares polynomial of degree k and there are but $k + 1$ observation points, it will pass through the points (x_i, y_i), $i = 0, 1, \cdots, k$, where y_i denotes the observed value of y at $x = x_i$ with weight w_i. The equation of this polynomial can be written in the form

$$\hat{y}(x) = \sum_{i=0}^{k} L_i(x) y_i$$

where $L_i(x)$ denotes the ith Lagrange interpolation polynomial of degree k, and which possesses the property that $L_i(x_j) = \delta_{ij}$. In the usual applications, y_i will be the mean of several observational values at the point $x = x_i$ and the weight w_i will be given by $w_i = np_i$, where p_i is the proportion of observations to be taken at $x = x_i$. The least squares polynomial must, of course, pass through such mean points.

Under the assumption of uncorrelated observational values, the variance of $\hat{y}(x)$ may be written in the form

$$V[\hat{y}(x)] = (\sigma^2/n) \sum_{i=0}^{k} [L_i^2(x)/p_i].$$

This assumes that the variance of the variable y_i is of the form σ^2/w_i, which can be written as σ^2/np_i whether or not np_i is an integer. Now it is readily shown by using Lagrange multiplier techniques that the minimization of $V[\hat{y}(x)]$ requires p_i to be chosen proportional to $|L_i(x)|$. In studying the problem of how to space the points it therefore suffices to consider the expression

$$V[\hat{y}(x)] = (\sigma^2/n)\{\sum_{i=0}^{k} |L_i(x)|\}^2.$$

This in turn shows that it suffices to choose the points to minimize

$$G(x) = \sum_{i=0}^{k} |L_i(x)|.$$

3. Optimum spacing. This section is concerned with characterizing those x's that minimize $G(x)$. First, consider the restricted problem of choosing a set of x's that will satisfy the inequalities

$$(1) \quad -1 = x_0 < x_1 < \cdots < x_r < x_{r+1} = \alpha < \beta = x_{r+2} < x_{r+3} < \cdots < x_k = 1$$

and that will minimize $G(x)$. Here r is any fixed integer satisfying $1 \leqq r \leqq k - 3$. For ease of discussion it will be assumed that k is an odd integer and sufficiently large to permit the existence of such an r; however the same method of proof will apply when k is even. It will be shown later that an otimum design for this restricted set of x's is also optimum for a certain larger class of problems when the equality signs are permitted to become inequality signs also. The optimum spacing will be obtained by means of a lemma and two theorems.

Lemma. *$G(x) > 1$ for $\alpha < x < \beta$.*

Proof. Since $\sum_{i=0}^{k} L_i(x)$ is a polynomial of degree k, it follows from the property $L_i(x_j) = \delta_{ij}$ that this polynomial must assume the value 1 at the $k + 1$ points $x_0, x_1, \cdots, x_k$; consequently $\sum L_i(x) \equiv 1$. If a set of real numbers has the sum 1, then the sum of the absolute values of those numbers must exceed 1 unless all the numbers are nonnegative. In order to prove that $G(x) > 1$ it therefore suffices to show that at least two of the $L_i(x)$ possess opposite signs. Consider

$$L_r(x) = \frac{(x - x_0) \cdots (x - x_{r-1})(x - x_{r+1}) \cdots (x - x_k)}{(x_r - x_0) \cdots (x_r - x_{r-1})(x_r - x_{r+1}) \cdots (x_r - x_k)} .$$

If $L_r(x)$ is compared with $L_{r+1}(x)$ it will be observed that the two numerators will possess the same sign for $x_{r+1} < x < x_{r+2}$ but that the denominators are of opposite signs because $L_r(x)$ will contain one more negative factor, namely $(x_r - x_{r+1})$, than $L_{r+1}(x)$. Since the assumptions concerning r assure that both these terms occur in $G(x)$, this proves the lemma.

For $\alpha < x < \beta$ it is seen that $G(x)$ can be written in the form

$$G(x) = \textstyle\sum_{i=0}^{r+1} (-1)^{r+1-i} L_i(x) + \sum_{i=r+2}^{k} (-1)^{i-r-2} L_i(x).$$

This is a polynomial of degree k, which by the preceding lemma exceeds 1 throughout $\alpha < x < \beta$, and which assumes alternating values of ± 1 at the points $x_0, x_1, \cdots, x_{r+1}$ and at the points $x_{r+2}, \cdots, x_k$, with the value of 1 at x_{r+1} and x_{r+2}. The problem of determining what set of x's will minimize the variance of $\hat{y}(x)$ is now reduced to the problem of determining what set of x's will minimize this polynomial at the point x inside the interval (α, β). The solution to the latter problem is given by the following theorem which assumes the existence and uniqueness of a polynomial possessing the properties attributed to it. A proof of the existence and uniqueness of such a polynomial will be given in a later section.

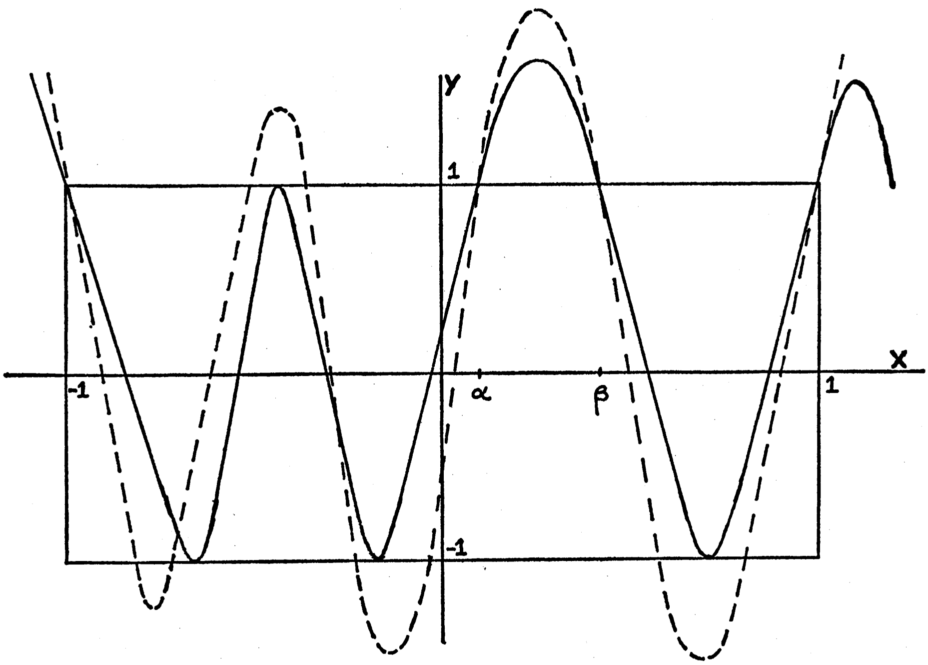
THEOREM 1. *$G(x)$ will be minimized for a set of x's satisfying* (1) *if they are chosen to satisfy the equations*

$$G'(x_i) = 0, \qquad i = 1, \cdots, r, r + 3, \cdots, k - 1. \tag{2}$$

PROOF. Consider the graphs of two polynomials, $G_1(x)$ and $G_2(x)$, possessing the properties of $G(x)$ but with $G_1(x)$ also satisfying Equations (2). In the accompanying sketch the graph of $G_1(x)$ is drawn with a solid line, whereas a broken line is used for $G_2(x)$. Here r is chosen to be an odd integer, namely 3, and k is chosen to be 7.

Any polynomial such as $G_2(x)$ that is required to pass through the $r + 2$ points $(-1, 1)$, $(x_1, -1)$, $(x_2, 1)$, $\cdots$, $(x_r, -1)$, $(\alpha, 1)$, where the x's are any numbers satisfying $-1 < x_1 < \cdots < x_r < \alpha$, will intersect $G_1(x)$ at least $r - 1$ times inside the interval $(-1, \alpha)$, if tangency is counted as two intersections. There will, however, be at least r such intersections if the graph of $G_2(x)$ lies above the graph of $G_1(x)$ immediately to the left of $x = \alpha$. Similarly, $G_2(x)$ will intersect $G_1(x)$ at least $k - r - 4$ times, or at least $k - r - 3$ times, inside the interval $(\beta, 1)$, depending upon whether the graph of $G_2(x)$ lies below, or above, the graph of $G_1(x)$ immediately to the right of $x = \beta$.

If $G_2(x) > G_1(x)$ both immediately to the left of $x = \alpha$ and to the right of $x = \beta$, a contradiction will be obtained, because $G_2(x)$ and $G_1(x)$ have four fixed



points in common and the minimum additional intersections occurring in $(-1, \alpha)$ and $(\beta, 1)$ will produce a total of $k + 1$ points in common, counting multiplicities, thereby making $G_2(x)$ identical to $G_1(x)$.

If $G_2(x) > G_1(x)$ in only one of the two preceding neighborhoods, say the one to the left of $x = \alpha$, then $G_2(x)$ must intersect $G_1(x)$ inside the interval (α, β), or be tangent to $G_1(x)$ at $x = \beta$. The total number of intersections, counting multiplicities, will then also be at least $k + 1$.

Finally, if $G_2(x) < G_1(x)$ in both neighborhoods, as is the situation shown in the sketch, it is necessary that $G_2(x) > G_1(x)$ for all x inside (α, β), unless additional intersections occur inside (α, β) or tangency occurs at either $x = \alpha$ or $x = \beta$. If tangency does not occur and there are additional intersections, there must be at least two intersections inside (α, β). If tangency occurs at one point, there must be at least one intersection inside (α, β). Tangency at both points counts as two additional intersections. In every case, therefore, at least two additional intersections, counting multiplicities, will be obtained to yield a total of at least $k + 1$ intersections and therefore making $G_2(x) = G_1(x)$. Thus, it must be true that $G_2(x) > G_1(x)$ inside (α, β) and hence that $G(x)$ is minimized by those x's that satisfy Equations (2). Although the proof was given for k and r both assumed to be odd integers, the same method of proof applies in general.

THEOREM 2. *There exists a unique polynomial of the type $G(x)$ that satisfies Equations* (2).

PROOF. Assuming the existence of such a polynomial, consider its uniqueness. Suppose there were two such polynomials. Then the polynomial which has the lower graph inside (α, β) will necessarily intersect the other one in $k + 1$ points for the same reasons that $G_2(x)$ intersected $G_1(x)$ in $k + 1$ points when it was assumed that $G_2(x) < G_1(x)$ inside (α, β). It obviously cannot intersect the other polynomial inside (α, β) without being identical to it.

The proof of the existence of the desired polynomial can be patterned after the proof [2] of the existence of a closely related polynomial. For this purpose, let $P(x)$ be a polynomial of degree k whose graph passes through the points $(-1, 1)$, $(\alpha, 1)$, $(\beta, 1)$ and for which $P'(x)$ possesses r real zeros, counting multiplicities, in the interval $(-1, \alpha)$ and $k - r - 3$ real zeros in the interval $(\beta, 1)$. As before, it is assumed that k and r are odd integers. Such polynomials obviously exist. Let $\xi_1, \cdots, \xi_r$ denote the zeros in $(-1, \alpha)$ and $\xi_{r+3}, \cdots, \xi_{k-1}$ those in $(\beta, 1)$. Also let $\xi_0 = -1$, $\xi_{r+1} = \alpha$, $\xi_{r+2} = \beta$, $\xi_k = 1$. It is assumed that the ξ's are ordered so that $\xi_0 \leqq \xi_1 \leqq \cdots \leqq \xi_k$. Next, define

$$
\begin{aligned}
(3) \qquad \varphi_i &= (-1)^i[P(\xi_i) - P(\xi_{i-1})], \qquad i = 1, \cdots, r \\
&= (-1)^{i+1}[P(\xi_i) - P(\xi_{i-1})], \qquad i = r + 3, \cdots, k.
\end{aligned}
$$

Since the graph of $P(x)$ must pass through the point $(-1, 1)$, $P(x)$ may be written in the form

$$(4) \qquad P(x) = A \int_{-1}^{x} (a + bx + x^2)\, \pi_j\, (\xi_j - x)\, dx + 1$$

where the product π extends over $j = 1, \cdots, r, r + 3, \cdots, k - 1$, where A is an arbitrary constant and where the constants a and b are determined by the requirement that $P(\alpha) = P(\beta) = 1$.

If $\varphi_i = 2$ for all i, the polynomial $P(x)$ will possess the properties of the minimizing polynomial of Theorem 1. Therefore to prove the existence of such a polynomial, it will suffice to show that to every set of $\varphi_i > 0$ values there exists a set of values of $\xi_1, \cdots, \xi_r, \xi_{r+3}, \cdots, \xi_{k-1}$, and A that will satisfy (3). The Relations (3) may be treated as the equations of a transformation from the ξ's and A to the φ's; therefore the problem is to show that the transformation possesses the property that to every point of E^{k-2} for which the φ's > 0, there corresponds a point of E^{k-2} for which $-1 \leqq \xi_1 \leqq \cdots \leqq \xi_r \leqq \alpha < \beta \leqq \xi_{r+3} \leqq \cdots \leqq \xi_{k-1} \leqq 1$ and $-\infty < A < \infty$.

Since a and b are functions of the ξ's and A, it follows from (4) that for any interior point of the domain of the ξ's and A

$$\partial P(\xi_i)/\partial \xi_j = A \int_{-1}^{\xi_i} \pi(\xi_j - x)[(p(x)/(\xi_j - x)) + a_j + b_j x]\, dx$$

where $a_j = \partial a/\partial \xi_j$, $b_j = \partial b/\partial \xi_j$, and $p(x) = a + bx + x^2$. Similarly,

$$\partial P(\xi_i)/\partial A = A \int_{-1}^{\xi_i} \pi(\xi_j - x)[(p(x)/A) + a_0 + b_0 x]\, dx$$

where a_0 and b_0 denote derivatives with respect to A. From (3), it then follows that

(5) $\partial\varphi_i/\partial\xi_j = (-1)^i A \int_{\xi_{i-1}}^{\xi_i} \pi(\xi_j - x)[(p(x)/(\xi_j - x)) + a_j + b_j x]\,dx.$

A similar formula holds for the derivative with respect to A.

Now it is readily shown by means of (4) and a and b are continuous functions of the ξ's and A, and therefore that the φ's as given by (3) are continuous functions of those variables. Furthermore, it is also easily seen that the partial derivatives given by (5) are continuous functions of those same variables. Thus, the transformation (3) is of class C'. Next, it will be shown that the Jacobian of this transformation does not vanish at any interior point of the domain.

Suppose the Jacobian $\partial(\varphi)/\partial(\xi, A)$ did vanish. Then there would exist a set of constants c_j, not all zero, such that the relation

(6) $$\sum_{j=1, j\neq r+1, r+2}^{k-1} c_j(\partial\varphi_i/\partial\xi_j) + c_k(\partial\varphi_i/\partial A) = 0$$

would hold independent of i. Consider the function

$$F(x) = (-1)^i A\pi(\xi_j - x)\{\textstyle\sum_j c_j[(p(x)/(\xi_j - x)) + a_j + b_j x] + c_k[(p(x)/A) + a_0 + b_0 x]\}.$$

From (6) and (5), it follows that

(7) $$\int_{\xi_{i-1}}^{\xi_i} F(x)\,dx = 0 \qquad i = 1, \cdots, r, r+3, \cdots, k.$$

Since $\xi_0 = -1$, this yields

(8) $$\int_{-1}^{\xi_i} F(x)\,dx = 0 \qquad i = 1, \cdots, r.$$

But from the relation $P(\alpha) = 1$, it follows that

$$\int_{-1}^{\alpha} (a + bx + x^2)\pi(\xi_j - x)\,dx = 0.$$

Differentiation with respect to ξ_j gives

$$\int_{-1}^{\alpha} \pi(\xi_j - x)[(p(x)/(\xi_j - x)) + a_j + b_j x]\,dx = 0.$$

A corresponding result holds with α replaced by β. As a consequence,

$$\int_{-1}^{\alpha} F(x)\,dx = \int_{-1}^{\beta} F(x)\,dx = 0.$$

These results when combined with (7) then show that (8) also holds for $i = r + 3, \cdots, k$.

Next, let

$$G(x) = \int_{-1}^{x} F(t)\,dt.$$

The general result (8) shows that $G(x) = 0$ for $x = -1, \xi_1, \cdots, \xi_r, \alpha, \beta, \xi_{r+3}, \cdots, \xi_k$. But since $G(x)$ is a polynomial of degree k and vanishes at $k + 1$ points, it must be identically zero; consequently $F(x) = 0$. From the definition of $F(x)$, this implies that

$$\textstyle\sum_j c_j[(p(x)/(\xi_j - x)) + a_j + b_j x] + c_k[(p(x)/A) + a_0 + b_0 x] = 0.$$

But it can be shown that the functions in brackets are not linearly dependent; therefore the assumption made in (6) that there exists a set of constants, not all zero, satisfying (6) is false. This implies that the Jacobian does not vanish at any interior point.

One can now appeal to a theorem of analysis [1] which states that a transformation of class C' possessing a non-vanishing Jacobian maps open sets into open sets to show that there cannot exist points of the space E^{k-2} determined by the conditions $\varphi_i > 0, i = 1, \cdots, r, r+3, \cdots, k$ for which there are no corresponding points in the domain of the ξ's and A. This can be done by assuming that the contrary is true and then considering the non-empty boundary between the points of E^{k-2} for which the $\varphi_i > 0$ and the points of E^{k-2} obtained under the transformation. Then by choosing a sequence of points in the transformed space that converges to a boundary point one can show that there exists a corresponding sequence in the original space possessing an interior limit point, because of the properties of the transformation, in that space which maps into the chosen boundary point. But since open sets map into open sets, this yields a contradiction.

Although the author of [2], did not explain his reasoning in arriving at his conclusion from the properties of his transformation, whatever it was it applies equally well to the present problem. Just as for his problem, interior points map into interior points and boundary points into boundary points, and the same properties of his transformations hold.

Thus, by choosing the φ's to be equal to 2, this proves that there must exist a polynomial of the type desired, namely one that passes through the four points $(-1, 1)$, $(\alpha, 1)$, $(\beta, 1)$, $(1, 1)$ and which possesses r extrema of alternating values ± 1 in the interval $(-1, \alpha)$ and $k - r - 3$ extrema of alternating values ± 1 in the interval $(\beta, 1)$, with the last extremum in the first interval and the first extremum in the second interval possessing the value -1.

It remains to be shown that the optimum design obtained under the restriction that the x values -1, α, β, and 1 are always chosen as observation points is also optimum without this restriction for a useful class of problems.

Consider a class of problems for which it is known that at least two points must be chosen in each of the two intervals $[-1, \alpha]$ and $[\beta, 1]$. For such problems the preceding restricted design with the four fixed points is optimum in general. This fact is verified by the same type of arguments as those used to prove Theorems 1 and 2. Thus, suppose a set of values $x_0, \cdots, x_k$ has been selected and that, say, α was not selected for the value of x_{r+1}, in which case $x_{r+1} < \alpha$. Now the arguments of Theorem 1 will show that whatever the choice of x values, they must satisfy Conditions (2) if they are to yield an optimum design. Let $G_1(x)$ denote the $G(x)$ that satisfies Conditions (2) for the selected values of x_0, x_{r+1}, x_{r+2}, and x_k, and construct $G_2(x)$, also satisfying (2), based on the same values of x_0, x_{r+2}, and x_k but passing through the point $(\alpha, 1)$ instead of through $(x_{r+1}, 1)$. It is now easily verified by the usual argument that $G_2(x) < G_1(x)$ inside (α, β), otherwise $G_2(x) \equiv G_1(x)$. The same reasoning requires the selection of $x_{r+2} = \beta$, and the same type of reasoning will show that it is necessary to choose $x_0 = -1$ and $x_k = 1$.

If only one point is to be chosen in one of the two subintervals, say in $[\beta, 1]$, then it must be chosen at $x = \beta$, otherwise a $G(x)$ with $x_{r+3} = \beta$ can be constructed that will be superior to one based on a different point in $[\beta, 1]$.

4. Choice of r. In the preceding section it was assumed that r could be specified in advance. Now consider the problem of how r should be chosen.

The choice of r will obviously depend upon the value of k and upon the location of the subinterval (α, β) inside $(-1, 1)$. If, for example, α is fixed and β is allowed to approach 1, it is easily shown that the variance will become infinite if two or more points are chosen in the interval $[\beta, 1]$. Thus, as β approaches 1, there can be but one point chosen in that interval, and therefore from the results at the end of the preceding section it must be chosen at $x = \beta$.

If the subinterval (α, β) is not located symmetrically inside $(-1, 1)$, that is if $\alpha \neq -\beta$, and if neither end point is close to the boundary of $(-1, 1)$, the number of points to be chosen in each subinterval will depend upon the location of (α, β). The larger of the two subintervals would be expected to contain the larger number of points if an unequal choice is made; however the decision as to whether an equal or unequal choice should be made (when k is odd) will again depend upon the size and location of (α, β).

If the subinterval (α, β) is located symmetrically and if k is odd, there will be an equal number of points in the two subintervals $[-1, \alpha]$ and $[\beta, 1]$, which will be symmetrically located. The polynomial $G(x)$ then degenerates into a polynomial of degree $k - 1$. If k is even there will necessarily be an unequal number of points in the two subintervals and the optimum allocation will then depend upon where x was chosen in the interval (α, β).

5. Illustration. Consider the following numerical example of the preceding theory. Let $\alpha = 0$, $\beta = \frac{1}{2}$, and $x = \frac{1}{4}$, and assume that $k = 5$. First, choose one point in each of the subintervals $(-1, 0)$ and $(\frac{1}{2}, 1)$. The values that minimized $G(\frac{1}{4})$ were found to be $x_1 = -.44$ and $x_4 = .82$. Next, consider the possibility of choosing only one point in $[\frac{1}{2}, 1]$, which must then be at $\frac{1}{2}$, and three points in $(-1, 0)$. The three points that minimized $G(x)$ were found to be $x_1 = -.88$, $x_2 = -.56$, and $x_3 = -.19$. The values of $G(\frac{1}{4})$ for the two cases being considered were 1.9 and 8, respectively; therefore the first choice is better, which of course was to be expected. Choosing three points in the right interval will also obviously lead to a larger value of $G(\frac{1}{4})$; therefore the first case considered and its solution is the optimum design for this problem.

The iterative method proposed in [5] for a closely related problem was used here to arrive at these numerical answers. It consists of first choosing the required number of points of the form $(x_i, \pm 1)$, in which the y values alternate in sign in the two subintervals $(-1, \alpha)$ and $(\beta, 1)$, and then forming the polynomial of degree k that passes through those points and through the four fixed points. A uniform spacing of the initial x_i values in the two subintervals will usually lead to rapid covergence of the iterative process. Next, the zeros of the derivative of this polynomial inside the two subintervals, except possibly for one near -1 or 1 if a total of $k - 2$ zeros is obtained inside these subintervals, are chosen to yield a new set of points $(x_i', \pm 1)$; and the process is repeated until sufficient accuracy is attained.

6. Two dimensional interior extrapolation. The preceding theory will now be used to obtain optimum designs for the corresponding two-dimensional problem but restricted to polynomials in two variables of the type given by the formula

$$E[z(x, y)] = \sum_{\alpha=0}^{l} \sum_{\beta=0}^{m} c_{\alpha\beta} x^{\alpha} y^{\beta}. \tag{9}$$

Assume that observations can be taken inside the rectangle $\{0 \leqq x \leqq X, 0 \leqq y \leqq Y\}$, exclusive of the rectangle $\{a < x < b, c < y < d\}$ which lies inside it. Let (x_i, y_j), $i = 1, \cdots, l + 1, j = 1, \cdots, m + 1$ denote $(l + 1)(m + 1)$ points that have been selected in the permissible domain and let $w_{ij} = np_{ij}$ denote the corresponding weight assigned to (x_i, y_j). Since the regression polynomial (9) contains $(l + 1)(m + 1)$ unknown coefficients and that same number of points has been selected, the least squares estimated polynomial may be written in the form

$$\hat{z}(x, y) = \sum_{i=1}^{l+1} \sum_{j=1}^{m+1} L_i(x) L_j(y) z_{ij}.$$

Here z_{ij} denotes the observed value of z at the point (x_i, y_j) with weight w_{ij}. The variance of this estimate for a point (x, y) chosen inside the interior rectangle is therefore given by

$$V[\hat{z}(x, y)] = (\sigma^2/n) \sum_{i=1}^{l+1} \sum_{j=1}^{m+1} L_i^2(x) L_j^2(y)/p_{ij} \tag{10}$$

where it is assumed that the z_{ij} are uncorrelated and that z_{ij} possesses the variance $\sigma^2/w_{ij} = \sigma^2/np_{ij}$. If n_{ij} observations are taken at (x_i, y_j) and all observations possess the same variance σ^2, then z_{ij} denotes the mean of those observations and np_{ij} is an integer. As in the case of the one-dimensional theory, the optimizing weights may not turn out to be integers; therefore only approximate optimality may be attained here.

Since the choice of p_{ij} in (10) is independent of the dimensionality of the problem, it follows from the one-dimensional result that the minimization of $V[\hat{z}(x, y)]$ requires that

$$p_{ij} = |L_i(x) L_j(y)| / \sum \sum |L_i(x) L_j(y)|. \tag{11}$$

As a result, (10) reduces to

$$\begin{aligned} V[\hat{z}(x, y)] &= (\sigma^2/n)[\sum \sum |L_i(x) L_j(y)|]^2 \\ &= (\sigma^2/n)[\sum |L_i(x)|]^2 [\sum |L_j(y)|]^2. \end{aligned} \tag{12}$$

From the one-dimensional result it now follows that $V[\hat{z}(x, y)]$ will be minimized if, and only if, the x_i and the y_j are selected to satisfy Theorem 1. The determination of the number of points to be chosen in each of the x and y subintervals depends upon the location and size of the extrapolation subregion just as in the one-dimensional problem.

It should be noted that the polynomial (9) is not the traditional polynomia. of total degree $l + m$ because not all possible terms of this degree are present

7. Two-dimensional exterior extrapolation. The technique employed in the preceding section will also yield a solution to the problem of optimum two-dimensional exterior extrapolation for polynomials of the type given by (9).

Assume that observations can be taken inside the rectangle $\{a \leqq x \leqq b, c \leqq y \leqq d\}$ and that $(l + 1)(m + 1)$ points are to be selected in that domain. The problem of optimum spacing and weighting is, as before, reduced to the problem of choosing a set of x's and a set of y's to minimize (12) and choosing weights to satisfy (11). These choices will depend upon the location of the exterior extrapolation point, which will be denoted by (x_0, y_0).

If neither of the inequalities $a < x < b$, $c < y < d$ is satisfied by x_0, y_0, the x_i and the y_j should be selected as the Chebychev points for each variable as given in [4] because the problem is reduced to two one-dimensional exterior extrapolation problems.

If one of the inequalities $a < x < b$, $c < y < d$ is satisfied by x_0, y_0, say $c < y_0 < d$, the x_i should be selected as the Chebychev points but the y_j should all be selected at y_0. This last fact can be demonstrated as follows.

It follows from (12) that the y_j must be chosen to minimize $\sum |L_j(y_0)|$. But if y_0 is any value other than y_j, $j = 1, \cdots, m + 1$ inside the interval (c, d) it follows from the lemma of Section 3 that $\sum |L_j(y_0)| > 1$, whereas $\sum |L_j(y_i)| = 1$ for $i = 1, \cdots, m + 1$. Hence, at least one of the y_j values must be chosen as y_0. It remains to be shown that all the y_j values must be chosen in this manner. Suppose they were not. Then $V[\hat{z}(x_0, y_0)]$ as given by (10) would reduce to

$$V[\hat{z}(x_0, y_0)] = (\sigma^2/n) \textstyle\sum_{i=1}^{l+1} [L_i^2(x_0)/p_{i0}] \tag{13}$$

where p_{i0} is the value of p_{ij} assigned to the point (x_i, y_0), because $L_j(y_t) = \delta_{jt}$. Whatever the relative values of the p_{i0}, $i = 1, \cdots, l + 1$, may be, it is clear that (13) will be minimized only if these values are made as large as possible. Hence, a minimum will be attained if, and only if, $\sum_i p_{i0} = 1$. But this requires that $p_{ij} = 0$ for $j \neq 0$, which implies that all points must have the y coordinate y_0.

The preceding methods can obviously be extended to solve the problems of optimum interior and exterior extrapolation for generalized polynomials of the type (9) in any number of dimensions, provided the regions are rectangular with sides parallel to the coordinate planes.

8. Acknowledgment. I wish to express my appreciation to my colleague Professor Motzkin for his many helpful discussions of certain aspects of this problem.

REFERENCES

[1] Buck, R. C. (1965). *Advanced Calculus* (2d ed.). McGraw-Hill, New York.
[2] Davis, C. (1957). Extrema of a polynomial. *Amer. Math. Monthly* **64** 679–680.

[3] GUEST, P. G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist.* **29** 294–299.

[4] HOEL, P. G. and LEVINE, A. (1964). Optimal spacing and weighting in polynomial prediction. *Ann. Math. Statist.* **35** 1553–1560.

[5] KAMMERER, W. J. (1961). Polynomial approximations to finitely oscillating functions. *Math. Comput.* **15** 115–119.

[6] KIEFER, J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.